Draft Technical Report	2005

Development of a Benthic Index of Biotic Integrity for Wadeable Streams in Northern Coastal California and its Application to Regional 305(b) Assessment

September 2005

www.waterboards.ca.gov/swamp

Development of a Benthic Index of Biotic Integrity (B-IBI) for Wadeable Streams in Northern Coastal California and its Application to Regional 305(b) Assessment

Andrew C. Rehn and Peter R. Ode

California Dept. of Fish and Game
Aquatic Bioassessment Laboratory
2005 Nimbus Road
Rancho Cordova, California 95670

Draft Report—Do not cite or distribute

Introduction

Bioassessment, or the use of biological assemblages to evaluate ecological condition, has become a widely used technique in water quality monitoring programs around the world. Benthic macroinvertebrates (BMIs) are the most commonly used organisms in bioassessment; in the United States, they are used in over 90% of bioassessment programs (Diamond *et al*. 1996). The index of biotic integrity (IBI) was first introduced by Karr (1981) as a measure of stream condition based on fish assemblages, but IBIs also have been developed for BMI (e.g. Kerans & Karr 1994) and periphyton (e.g. Pan *et al*. 1996) assemblages. An IBI is typically composed of a set of metrics that together represent different attributes of assemblage composition, structure and function such as species richness, tolerance guilds and trophic guilds. Metrics are selected for inclusion in an IBI based on their responsiveness to anthropogenic stressor gradients and/or their ability to discriminate between minimally disturbed reference sites and test sites that are known or suspected to have been exposed to stressors of interest. Adoption of a consistent and quantifiable method for defining reference condition is fundamental to any bioassessment program (Hughes 1995, Bailey *et al*. 2004).

In California, several state and federal agencies have become increasingly involved in developing analytical tools that can be used to assess the biological and physical condition of streams and rivers. For example, the US Environmental Protection Agency (EPA), the California EPA, the US Forest Service (USFS) and California's state and regional Water Quality Control Boards (WQCBs) have collected fish, periphyton and BMIs from California streams and rivers as a critical component of regional water quality assessment and management programs. Together, these agencies have sampled thousands of sites in California, but until recently, no analysis of BMI assemblage data sets based on comprehensively defined regional reference conditions has been undertaken (Ode *et al*. 2005). The purpose of this paper is to develop a benthic IBI (B-IBI) for the northern coastal region of California based on BMI assemblage data and apply the B-IBI to an assessment of all mapped wadeable streams in the region based on the probability stream survey recently completed by EPAs Environmental Monitoring and Assessment Program Western Pilot (WEMAP).

Methods

Study Area

The northern coastal California B-IBI (NorCal B-IBI) was developed for the region that drains directly west to the Pacific Ocean from Marin County in the south to the Oregon border in the north (Fig. 1). This area contains 3 Level III ecoregions (Omernik 1987) and receives the highest annual rainfall totals in California with areas near the Oregon border receiving nearly 200 inches and areas in the south receiving >50 inches over mountain ranges. High rainfall totals combined with rapid uplift due to tectonic subduction and compression, unstable soil types, and land use practices such as logging and grazing that promote erosion result in the state's highest total sediment yields (Mount 1995). The estimated human population in the region, although relatively low compared to other regions in California, exceeded 1,040,000 in 2004 and is concentrated in Marin

and Sonoma counties (California Department of Finance, Demographic Research Unit, www.dof.ca.gov). The region currently supplies two-thirds of the state's total timber production and during the 19th century suffered large-scale deforestation and degradation of entire watersheds as timber harvest and mercury mining operations expanded to meet the demands of hydraulic gold mining in the Sierra Nevada.

Field Protocols and Combining Datasets

The NorCal B-IBI is based on BMI and physical habitat data collected from 257 sites (Fig. 1) using the three protocols described below. Sites were sampled during base flow periods between April and early October of 2000-2003.

California Stream Bioassessment Protocol (CSBP, 150 sites) - Both regional WQCBs in northern coastal California have implemented biomonitoring programs in their respective jurisdictions and have collected BMIs according to the CSBP (Harrington 1999). At CSBP sites, three riffles within a 100m reach were randomly selected for sampling. At each riffle, a transect was established perpendicular to the flow, from which three separate areas of 0.18 m^2 each were sampled upstream of a 0.3m wide D-frame net and composited by transect. A total of 1.82m^2 of substrate was sampled per reach and 900 organisms were subsampled from this material (300 organisms were processed separately from each of three transects). Water chemistry data was collected in accordance with the protocols of the different regional WQCBs (Puckett 2002) and qualitative physical habitat characteristics were measured according to Barbour *et al*. (1999) and Harrington (1999).

USFS (32 sites)- The USFS sampled streams on national forest lands in northern California in 2000 and 2001 using the Hawkins *et al*. (2001) targeted riffle protocol. All study reaches were selected non-randomly as part of a program to develop an interpretive (reference) framework for the results of stream biomonitoring studies on national forests in California. BMIs were sampled at study reaches (containing at least 4 fast water habitat units) by disturbing 2 separate 0.09m^2 areas of substrate upstream of a 0.3m wide D-frame net in each of four separate fast water units; a total of 0.72m^2 was disturbed and all sample material from a reach was composited. Field crews used a combination of qualitative and quantitative measures to collect physical habitat and water chemistry data (Hawkins *et al*. 2001). A 500 organism subsample was processed from the composite sample and identified following methods described by Vinson and Hawkins (1996).

WEMAP (75 sites) - The EPA sampled study reaches in northern coastal California from 2000 through 2003 as part of its WEMAP pilot project. A sampling reach was defined as 40 times the average stream width at the center of the reach, with a minimum reach length of 150m and maximum length of 500m. A BMI sample was collected at each site using the USFS methodology described above (Hawkins *et al*. 2001) in addition to a standard WEMAP BMI sample (not used in this analysis). A 500 organism subsample was processed in the laboratory according to WEMAP standard taxonomic effort levels (Klemm *et al*. 1990). Water chemistry samples were collected from the mid-point of each reach and analyzed using WEMAP protocols (Klemm and Lazorchak 1994). Field

crews recorded physical habitat data using EPA qualitative methods (Barbour *et al*. 1999) and quantitative methods (Kaufmann *et al*. 1999).

Because USFS style riffle samples were collected at all WEMAP sites, only two field methods were combined in this study. All 900 count CSBP samples were standardized to 500 individuals per reach using a randomized rarefaction technique based on previous analyses that demonstrated no difference between the two methods once counts are adjusted (Ode *et al*. 2005). All WEMAP and CSBP samples were collected and processed by the California Department of Fish and Game's Aquatic Bioassessment Laboratory (ABL) and all USFS samples were processed by the US Bureau of Land Management's Bug Lab in Logan, Utah. Taxonomic data from both labs were combined in an MS Access© database that standardized BMI taxonomic effort levels and metric calculations allowing us to minimize any differences between the two labs that processed samples. Taxonomic effort followed standards defined by the California Aquatic Macroinvertebrate Laboratory Network (2002; www.dfg.ca.gov/cabw/camlnetste.pdf). Sites with fewer than 450 organisms sampled were omitted from the analyses.

Screening Reference Sites

We followed an objective and quantitative reference site selection procedure in which potential reference sites were first screened with quantitative GIS land use analysis at several spatial scales, and then were screened with local condition assessments (in-stream and riparian) to quantify stressors acting within study reaches. We calculated the proportions of different land cover classes and other measures of human activity upstream of each site at four spatial scales that give unique information about potential stressors acting on each site: 1) within polygons delimiting the entire watershed upstream of each sampling site; 2) within polygons representing local regions (defined as the intersection of a 1km radius circle around each site and the primary watershed polygon); 3) within a 100m riparian zone on each side of all streams within each watershed; 4) within a 100m riparian zone in the local region. We used the ArcView® (ESRI 1999) extension ATtILA (Ebert and Wade 2002) to calculate the percentage of various landcover classes (urban, agriculture, natural, etc.) and other measures of human activity (population density, road density, etc.) in each of the four spatial areas defined for each site. Landcover analyses were based on the California Department of Forestry and Fire Protection (CDF) Multisource California Land Cover Mapping and Monitoring Program (LCMMP, http://frap.cdf.ca.gov/data/frapgisdata/select.asp). Where available, these data were supplemented with development footprint layers derived from 2000 census data (CDF) and the California Department of Conservation's 2002 Farmland Mapping data (DOC- FMMP, http://www.consrv.ca.gov/dlrp/fmmp). Population data were derived from the 2000 migrated TIGER dataset (CDF). Stream layers were obtained from the USGS National Hydrography Dataset (NHD). The road network was obtained from the USFS Remote Sensing Lab (http://fsweb/gis/gis_data/calcovs/fs/nwctran03_2.html). The USGS National Elevation Dataset (NED) was used for elevation data. Frequency histograms of land use percentages for all sites were used to establish subjective thresholds for eliminating sites from the potential reference pool (Table 1). Sites were further screened from the reference pool on the basis of reach scale conditions (obvious

bank stability or erosion/ sedimentation problems, evidence of mining, dams, grazing, recent fire, recent logging). Once the pool of reference sites was defined, we randomly divided the full set of sites into a development set that was used to screen metrics and establish scoring ranges for component B-IBI metrics and a validation set that was used for independent evaluation of B-IBI performance.

Screening Metrics and Assembling the B-IBI

Seventy-seven metrics were evaluated for possible use in the NorCal B-IBI (Table 2). A multi-step screening process was used to evaluate each metric for: 1) sufficient range to be used in scoring; 2) responsiveness to watershed scale and reach scale disturbance variables; 3) discrimination between reference and test sites; 4) lack of correlation with other responsive metrics.

Pearson correlations between all watershed scale and reach scale disturbance gradients were used to define the smallest suite of independent (non-redundant) disturbance variables against which to test biological metric response. Disturbance variables with correlation coefficients $|r| \geq 0.7$ were considered redundant. Responsiveness was assessed using visual inspection of biotic metric vs. disturbance gradient scatter plots and linear regression coefficients. Metrics were selected as responsive if they showed either a linear or a wedge-shaped relationship with disturbance gradients. Biological metrics often show a wedge-shaped relationship with single disturbance gradients where the upper boundary represents a threshold of biological response. Multiple limiting factors may result in lower metric values than expected if response were to the single gradient alone. For wedge-shaped relationships, we used a method similar to Blackburn *et al*. (1992) and Rankin & Yoder (1999) to characterize response thresholds. The x-axis was divided into 10 equal categories and the three largest metric values were selected from each category (the three smallest values were selected for negative metrics). Ordinary least-squares regressions were calculated for the subsets of data to estimate the upper bound slopes of wedge-shaped polygons. Metrics that passed the range and responsiveness tests were tested for redundancy. Pairs of metrics with Product-Moment correlation coefficients $|r| \geq 0.7$ were considered redundant and the least responsive metric of the pair was eliminated.

Metrics that were significantly correlated ($p<0.05$) with watershed area were normalized to the mean reference watershed area of 35km^2 following the method described by Urquart (1982). We calculated the regression equation of the metrics with $\log_{10}$ watershed area in km^2 for reference sites. We then applied that reference regression equation to all sites and calculated their residuals. The predicted metric value for a reference site with the mean watershed area of 35km^2 was determined and this constant was added to all residuals. The sum of the residual plus the constant at each site resulted in a corrected metric value that was unrelated to watershed area with some sites having negative values.

Scoring ranges were defined for each metric using techniques described in Hughes *et al*. (1998), McCormick *et al*. (2001) and Klemm *et al*. (2003). Metrics were scored on a 0-10 scale using statistical properties of the raw metric values from both reference and non-

reference sites to define upper and lower thresholds. For positive metrics (those that increase as disturbance decreases), any site with a metric value equal to or greater than the 80th percentile of reference sites received a score of 10; any site with a metric value equal to or less than the 5th percentile of the non-reference sites received a score of 0; these thresholds were reversed for negative metrics (20th percentile of reference and 95th percentile of non-reference). In both cases, the remaining range of intermediate metric values was divided equally and assigned scores of 1 through 9. Before assembling the B-IBI, we used Kruskal-Wallace tests to determine whether any of the final metrics were significantly different between Klamath, coastal and chaparral reference sites in the northern California coastal region, in which case they would require separate scoring ranges in the B-IBI. Finally, an overall B-IBI score was calculated for each site by summing the constituent metric scores and adjusting the B-IBI to a 100 point scale.

Validation of B-IBI and Measurement of Performance Characteristics

To test whether the distribution of B-IBI scores in reference and test sites might have resulted from chance, we compared score distributions in the development set to those in the validation set. We also investigated a separate performance issue that ambient bioassessment studies often neglect: spatial variation at the reach scale. Although our use of a validation dataset tested whether the B-IBI scoring range is repeatable (Fore *et al*. 1996, McCormick *et al*. 2001), we designed a separate experiment to explicitly measure within-site index precision. In September 2004, we estimated within-site variance in B-IBI scores due to sampling error by taking 3 nested, replicate samples from a single 150m reach at 15 different streams following the USFS protocol. B-IBI scores were then calculated for each replicate. The mean squared error from an ANOVA with site as the independent variable was used as the variance among replicates to calculate the minimum detectable difference (MDD) between two B-IBI scores based on a two-sample *t*-test model (Zar 1999). The index range was divided by the MDD to estimate the number of stream condition categories detectable by the B-IBI (Doberstein *et al*. 2000, Fore *et al*. 2001).

Regional Assessment of Stream Condition

We used the B-IBI in conjunction with WEMAP's probabilistic sampling design and weighted frequency distribution of streams (Herlihy *et. al*. 2000) to estimate the total length of streams in the region achieving a particular condition. We calculated 95% confidence bounds for these measures over the entire study area. A stratified random sampling design was developed wherein each stream segment in EPA's 1:100,000 scale "River Reach File Version 3" (RF3) was given a probability of selection that was roughly inverse to its percent contribution to the total estimated resource population. First order streams were assigned a relatively low probability of selection, whereas larger order streams (fourth order and higher) were assigned a relatively high probability of selection to ensure that the final stratified random sample would contain sample reaches across all stream orders. Each potential sampling site was assigned an associated weight equal to the number of stream kilometers represented by that sample reach.

Results

Reference Sites

Ninety-one sites passed all the land use and local condition screens and were selected as reference sites, leaving 164 sites in the test group (Fig. 1). The development set comprised 190 sites (66 reference/124 test) and the validation set comprised 67 sites (24 reference/43 test).

Selected metrics

Seven non-redundant stressor gradients were selected for metric screening: percent watershed unnatural, percent watershed in agriculture, road density in local watershed, qualitative channel alteration score, percent sand and fine substrates, conductivity and total phosphorous. Thirteen biological metrics failed the range test, 14 metrics were unresponsive to stressor gradients, 22 metrics were redundant with other more responsive metrics ($|r| \geq 0.7$), 5 metrics showed poor discrimination between reference and test sites, and 15 metrics were rejected because they were biologically redundant, but not statistically redundant, with selected metrics (Table 2). A final set of 8 minimally correlated metrics was selected for the B-IBI: EPT richness, Coleoptera richness, Diptera Richness, percent intolerant individuals, percent non-gastropod scraper individuals, percent predator individuals, percent shredder taxa, and percent non-insect taxa (Table 3). All metrics rejected as statistically redundant were derived from taxa similar to those of selected metrics, but had weaker relationships with stressor gradients. Regression coefficients were significant between all 8 selected metrics and at least two stressor gradients: road density in local watershed and percent sand and fine sediment ($p \leq 0.0008$ when a Bonferroni correction for multiple tests is applied; Table 3). The final eight metrics included several metric types: richness, composition, tolerance measures and functional feeding groups.

Six of the final 8 metrics were significantly different between reference sites in the Klamath, northern coast range and chaparral ecoregions (Kruskal-Wallis $p < 0.05$; Fig. 2). We adjusted for these differences by creating separate scoring scales for the six metrics in the three ecoregions (Table 4). The metric percent intolerant individuals was significantly correlated with watershed area ($|r| = -0.371$; $p < 0.0001$) and was adjusted by the following stepwise procedure:

1. The predicted metric at each site (y) = $-0.089(\log_{10}$ watershed area$) + 0.433$
2. The difference (residual) between the observed metric value and the predicted metric value was calculated.
3. The constant 0.296 (the predicted proportion of intolerant individuals at the mean watershed area of $35km^2$) was added to each site's residual and multiplied by 100 to convert to percent.

After adjustment, the metric percent intolerant individuals was unrelated to watershed area. Each site's final B-IBI score was multiplied by 1.25 to adjust the scoring range to a 100 point scale.

Validation of B-IBI and Measurement of Performance Characteristics

The distribution of B-IBI scores at reference and non-reference sites was nearly identical between the development and validation data sets (Figure 3), indicating that our characterization of reference conditions and subsequent B-IBI scoring was repeatable and not likely due to chance. Although IBI scores were significantly different between reference and test sites in both the development and validation sets (Mann-Whitney U tests: $p < 0.0001$ and $p = 0.001$, respectively), there was overlap between the reference and test quartiles in both data sets. We speculated that this overlap might be due to the large number of test sites in the Klamath and northern coast ranges that were omitted from the reference pool but were relatively unaffected by human land use. No overlap in quartiles was observed when we compared the reference distributions to sites with ≥ 25% upstream watershed unnatural, indicating that the B-IBI provides good discrimination between reference sites and highly degraded sites (Fig. 3).

Based on a two-sample *t*-test model, setting $\alpha = 0.05$ and $\beta = 0.10$, the MDD for the NorCal IBI is 19.7. Thus, we have a 90% chance of detecting a 19.7 point difference between sites at the $p = 0.05$ level. Dividing the 100-point B-IBI scoring range by the MDD indicates that the NorCal B-IBI can detect approximately 5 biological condition categories, a result similar to other recent estimates of B-IBI precision (Doberstein *et al*. 2000, Fore *et al*. 2001, Ode *et al*. 2005). The B-IBI scoring range can simply be divided into 5 equal categories as follows: 0-20 = "very poor", 21-40 = "poor", 41-60 = "fair", 61-80 = "good" and 81-100 = "very good" (Figure 4). By contrast, a threshold of biological impairment can be established at 2 SDs below the mean reference site score (B-IBI score= 52) with the consequence that some "fair" sites would be considered impaired and others would be considered unimpaired (Fig. 4).

We ran a Principle Components analyses (PCA) on the environmental stressor values used for testing metric responsiveness plus several additional variables that quantified stressor and natural gradients in study watersheds. The PCA was restricted to a subset of 97 sites from which we had data for 13 variables that together defined a multi-factorial axis of watershed condition. Only the first PCA axis was significant, having eigenvalues larger than those predicted from the broken stick model (McCune and Grace 2002). The first PCA axis accounted for 43% of the variance in the environmental data and was highly correlated with B-IBI score ($r = -0.774$, $p < 0.0001$), which decreased with increasing human disturbance (Fig. 4). The axis clearly reflects a water quality, land cover and habitat gradient; percent watershed unnatural and nutrient concentrations had the highest positive loadings, and percent forest in local watershed and qualitative epifaunal substrate score had the highest negative loadings.

Finally, we tested whether our scoring adjustments removed relationships between B-IBI scores and several natural gradients (Fig. 5). We found no significant relationship between reference site B-IBI scores and ecoregion (Kruskal-Wallace test, $p = 0.09$), $\log_{10}$ watershed area ($r^2 = 0.02$, $p = 0.09$), or elevation ($r^2 = 0.007$, $p = 0.53$). There was a significant relationship between B-IBI and Julian date ($r^2 = 0.06$, $p = 0.009$). However,

this relationship was driven by two low scoring chaparral reference sites that were sampled early in the year, and was not observed when these two sites were removed from the test ($r^2 = 0.03$, $p = 0.06$). Moreover, the low value of r^2 indicates a weak relationship that is significance only because the regression slope $\neq 0$ (Fig. 5), thus does not indicate that B-IBI score is affected by sampling date.

Regional Assessment of Stream Condition

The total target sampleable, wadeable stream length mapped at a 1:100,000 scale in northern coastal California was estimated to be 7317 km. A total of 7451 km was not assessed because of land-owner denial and physical inaccessibility, but if assumed to be perennial brings the estimated target stream length to 14,768 km (Table 5). Over 50% of the total target stream length was estimated to be in Good condition based on our B-IBI, with 95% confidence intervals ranging between 47% and 72% (Table 6). The second most common condition was Very Good (between 12% and 31%). Between 6% and 27% of the total target stream length was estimated to be in Fair condition, and between 0% and 5% was estimated to be in Poor condition. None of the probability sites were in Very Poor condition.

Discussion

The NorCal B-IBI is the first quantitative index that allows assessment of biological condition of streams in northern coastal California in relation to multiple anthropogenic stressors. Our B-IBI can be a valuable tool for resource management when used to compare biological condition among sites and has enough precision to distinguish 5 categories of biological condition. However, when making regional assessments, it is also necessary to define what IBI scores constitute "acceptable" versus "impaired" conditions. Section 303(d) of the Clean Water Act requires states to list impaired waters, establish total maximum daily loads (TMDLs) for pollutants causing the impairment, and establish plans to rehabilitate those waters. Various options for determining impaired waters have been proposed (e.g. Hughes *et al*. 1998; McCormick *et al*. 2001; Ode *et al*. 2005), but the process is inherently subjective. Because of its widespread statistical acceptance, we set our threshold at 2 SDs below the mean reference site, or a B-IBI score of 52. Using this threshold, only 6% of the mapped, wadeable, sampleable stream length in northern coastal California is impaired (Fig. 6a). Herlihy *et al*. (2005) presented similar results based on a benthic IBI for headwater streams in western Oregon, a region that shares 2 of the 3 ecoregions included in the present study. According to those authors, only 6% of sites in western Oregon were impaired and 62% of sites had no impairment.

We also used a vertebrate (amphibians + fish) IBI recently developed for coldwater streams of western Oregon and Washington (Hughes *et al*. 2004) to score vertebrate assemblages collected at north coast probability sites. Use of this IBI is appropriate for northern coastal California since two of the ecoregions (Klamath Mtns. and Northern Coastal Ranges) in our study area also occur in western Oregon, and since sampling protocols used at northern California sites were similar or identical to those used in

Oregon and Washington. Using the same threshold of impairment for the vertebrate IBI (2 SDs below mean of reference), 7.5% of the mapped, wadeable, sampleable stream length in northern coastal California is impaired (Fig. 6b). By contrast, Hughes *et al*. (2004) found that 45% of stream kilometers in western Oregon and Washington were impaired based on the vertebrate IBI.

References

Barbour, M.T., J. Gerritsen, B.D. Snyder and J.B. Stribling. 1999. Revision to rapid bioassessment protocols for use in stream and rivers: periphyton, BMIs and fish. EPA 841-D-97-002. U.S. Environmental Protection Agency. Washington DC.

Bailey, C.R., R.H. Norris and T.B. Reynoldson. 2004. *Bioassessment of Freshwater Ecosystems: Using the Reference Condition Approach*. Kluwer, The Netherlands.

Blackburn, T.M., J.H. Lawton and J.N. Perry. 1992. A method for estimating the slope of upper bounds of plots of body size and abundance in natural animal assemblages. Oikos 65: 107-112.

Diamond, J.M., M.T. Barbour and J.B. Stribling. 1996. Characterizing and comparing bioassessment methods and their results: a perspective. Journal of the North American Benthological Society 15:713-727.

Doberstein, C.P., J.R. Karr and L.L. Conquest. 2000. The effect of fixed-count subsampling on macroinvertebrate biomonitoring in small streams. Freshwater Biology 44:355-371.

Ebert, D.W. and T.G. Wade. 2002. Analytical Tools Interface for Landscape Assessments (ATtILA), Version 3.0. US EPA, Office of Research and Development.

ESRI. 1999. ArcView GIS, Version 3.2., Spatial Analyst Extension. Environmental Systems Research Institute, Inc.

Fore, L.S., J.R. Karr, and R.W. Wisseman. 1996. Assessing invertebrate responses to human activities: Evaluating alternative approaches. Journal of the North American Benthological Society 15: 212-231.

Fore, L.S., K. Paulsen and K. O'Laughlin. 2001. Assessing the performance of volunteers in monitoring streams. Freshwater Biology 46:109-123.

Harrington, J.M. 1999. California stream bioassessment procedures. California Department of Fish and Game, Water Pollution Control Laboratory. Rancho Cordova, CA.

Hawkins, C.P., J. Ostermiller and M. Vinson. 2001. Stream invertebrate, periphyton and environmental sampling associated with biological water quality assessments. field protocols. Utah State University, Logan.

Herlihy, A.T. D.P Larsen, S.G. Paulsen, N.S. Urquart and B.J. Rosenbaum. 2000. Designing a spatially balanced, randomized site selection process for regional stream condition surveys: the EMAP Mid-Atlantic pilot study. Environmental Monitoring and Assessment 63:95-113.

Herlihy, A.T. W.J. Gerth, J. Li and J.L. Banks. 2005. Macroinvertebrate community response to natural and forest harvest gradients in western Oregon headwater streams. Freshwater Biology: doi:10.1111/j.1365-2427.2005.01363.x

Hughes, R.M. 1995. Defining acceptable biological status by comparing with reference conditions, pp 31-47, *In* W.S. Davis and T.P. Simon (eds) *Biological Assessment and Criteria: Tools for Water Resource Planning and Decision Making*. CRC Press, Inc. Boca Raton.

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.

Hughes, R.M., S. Howlin and P.R. Kaufmann. 2004. A biointegrity index (IBI) for coldwater streams of western Oregon and Washington. Transactions of the American Fisheries Society 133:1497-1515.

Karr, J.R. 1981. Assessment of biotic integrity using fish communities. Fisheries 6: 21-27.

Kaufmann, P.R., P. Levine, E.G. Robison, C. Seeliger and D.V. Peck. 1999. Surface waters: quantifying physical habitat in wadeable streams. US EPA, Office of Research and Development. EPA/620/R-99/003.

Kerans, B.L. and J.R. Karr. 1994. A benthic index of biotic integrity (B-IBI) for rivers of the Tennessee Valley. Ecological Applications 4: 768-785.

Klemm, D.J., P.A. Lewis, F.A. Fulk, and J.M. Lazorchak. 1990. Macroinvertebrate field and laboratory methods for evaluating the biological integrity of surface waters. U.S. Environmental Protection Agency, Cinncinati, OH. EPA. 600/4-90/030.

Klemm, D.J. and J.M. Lazorchak. 1994. Environmental monitoring and assessment program, surface water and Region 3 regional monitoring and assessment program, 1994 pilot laboratory methods manual for streams. US EPA, Office of Research and Development. EPA/620/R-94/003.

Klemm, D.J., K.A. Blocksom, F.A. Fulk, A.T. Herlihy, R.M. Hughes, P.R. Kaufmann, D.V. Peck, J.L. Stoddard, W.T. Thoeny, M.B. Griffith and W.S. Davis. 2003. Development and evaluation of a macroinvertebrate biotic integrity index (MBII) for regionally assessing Mid-Atlantic Highland streams. Environmental Management 31:656-669.

McCormick, F.H., R.M. Hughes, P.R. Kaufmann, D.V. Peck, J.L. Stoddard, and A.T. Herlihy. 2001. Development of an index of biotic integrity for the Mid-Atlantic Highlands Region. Transactions of the American Fisheries Society 130: 857-877.

McCune, B. and J.B. Grace. 2002. Analysis of Ecological Communities. MjM Software Design. Gleneden Beach, Oregon.

Mount, J.F. 1995. *California Rivers and Streams: The Conflict Between Fluvial Processes and Land Use*. University of California Press, Berkeley.

Ode, P.R., A.C. Rehn and J.T. May. 2005. A quantitative tool for assessing the integrity of southern coastal California streams. Environmental Management 35:493-504

Omernik, J.M. 1987. Ecoregions of the conterminous United States. Map (scale 1:7,500,000). Annals of the Association of American Geographers 77(1):118-125.

Pan, Y., R.J. Stevenson, B.H. Hill, A.T. Herlihy and C.B. Collins. (1996). Using diatoms as indicators of ecological conditions in lotic systems: a regional assessment. Journal of the North American Benthological Society 15:481-494.

Puckett, M. 2002. Quality Assurance Management Plan for the State of California's Surface Water Ambient Monitoring Program (SWAMP). Prepared for California State Water Resources Control Board, Division of Water Quality, Sacramento, CA. First version. December 2002. (http://www.swrcb.ca.gov/swamp/qapp.html)

Rankin, E.T. and C.O. Yoder. 1999. Methods for deriving maximum species richness lines and other threshold relationships in biological field data, pp 611-621, *In* T.P. Simon [ed.] *Assessing the Sustainability and Biological Integrity of Water Resources Using Fish Communities*. CRC Press, Inc., Boca Raton.

Urquhart, N.S. 1982. Adjustment in covariance when one factor affects the covariate. Biometrics 38:651-660.

Vinson, M.R. and C.P. Hawkins. 1996. Effects of sampling area and subsampling procedures on comparisons of taxonomic richness among streams. Journal of the North American Benthological Society 15: 392-399.

Zar, J.H. 1999. Biostatistical Analysis. 4th Edition. Prentice-Hall, Inc., Upper Saddle River, NJ. 931 pp.

Table 1. List of minimum or maximum land use thresholds used for rejecting potential reference sites.

Stressor	Threshold
Percentage of unnatural land use at the local scale	> 5%
Percental of urban land use at the local scale	> 3%
Percentage of total agriculture at the local scale	> 5%
Road density at the local scale	> 1.5 km/km^2
Population density (2000 census) at the local scale	> 25 ind./ km^2
Percentage of unnatural land use at the watershed scale	> 5%
Percentage of urban land use at the watershed scale	> 3%
Percentage of total agriculture at the watershed scale	> 5 %
Road density at the watershed scale	> 2.0 km/km^2
Population density (2000 census) at the watershed scale	> 50 ind./ km^2

Table 2. Seventy-seven metrics screened for used in the NorCal B-IBI. Metrics that failed the range test are marked with an asterisk (*); metrics that failed the responsiveness test ($r^2 < 0.2$, $p > 0.05$) are in italics; metrics that were adopted for use in the B-IBI are in bold. Poor discrimination between reference and test sites, statistical redundancy and biological redundancy with selected metrics are indicated in the second column.

Metric	**Redundancy/Discrimination**
CF + CG Richness	poor
Coleoptera Richness	
Collector Filterer Richness*	
Collector Gatherer Richness	poor
Diptera Richness	
Elmidae Richness*	
Ephemerellidae Richness*	
Ephemeroptera Richness	EPT Richness
EPT Richness	
Hydropsychidae Richness*	
Intolerant EPT Richness	
Intolerant Richness	EPT Richness
Mollusca Richness*	
Non-insect Richness	
Crustacea + Mollusca Richness*	
Plecoptera Richness	EPT Richness
Predator Richness	EPT Richness
Scraper Richness	EPT Richness
Shredder Richness	EPT Richness

Table 2. Continued.

Metric	**Redundancy/Discrimination**
Taxonomic Richness	EPT Richness
Trichoptera Richness	EPT Richness
% Baetidae Individuals	
% CF + CG Individuals	% Intolerant Individuals
% CF + CG Taxa	poor
% CF Taxa	
% CG Taxa	% Intolerant Individuals
% Chironomidae Individuals	
% Collector-Filterer Individuals	poor
% Collectors Gatherer Individuals	% Non-Gastropod Scraper Individuals
% Corbicula Individuals *	
% Crustacea Individuals	
% Diptera Individuals	
% Diptera Taxa	
% Dominant Taxon	poor
% Elmidae Individuals	biologically redundant
% Ephemeroptera Individuals	biologically redundant
% Ephemeroptera Taxa	biologically redundant
% EPT Individuals	% Intolerant Individuals
% EPT Taxa	EPT Richness
% Gastropoda Individuals	
% Glossosomatidae Individuals *	
% Hydropsychidae Individuals	biologically redundant
% Hydroptilidae Individuals *	
% Intolerant Individuals	
% Intolerant Diptera Individuals *	
% Intolerant Ephemeroptera Individuals	biologically redundant
% Intolerant Scraper Individuals	biologically redundant
% Intolerant Taxa	EPT Richness
% Intolerant Trichoptera Individuals	% Intolerant Individuals
% Mollusca Individuals	
% Non Baetis Fallceon Ephemeroptera Individuals	biologically redundant
% Non Hydro Cheumato Trichoptera Individuals	biologically redundant
% Non-Gastropoda Scraper Individuals	
% Non-Hydropsyche Hydropsychidae Individuals *	
% Non-Insecta Taxa	
% of Ephemeroptera Individuals that are Intolerant	biologically redundant
% of Trichoptera Individuals that are Intolerant	biologically redundant
% Oligochaeta Individuals	
% Perlodidae Individuals *	
% Philopotamidae Individuals *	

Table 2. Continued.

Metric	**Redundancy/Discrimination**
% Plecoptera Individuals	biologically redundant
% Plecoptera Taxa	biologically redundant
% Predator Taxa	
% Predator Individuals	
% Rhyacophildae Individuals	biologically redundant
% Scraper Taxa	
% Scraper Individuals	% Non-Gastropod Scraper Individuals
% Sensitive EPT Individuals	% Intolerant Individuals
% Shredder Taxa	
% Shredder Individuals	biologically redundant
% Simuliidae Individuals	
% Tolerant Individuals	
% Tolerant Taxa	% Non-Insect Taxa
% Trichoptera Individuals	biologically redundant
% Trichoptera Taxa	% Non-Insect Taxa
Shannon Diversity	EPT Richness
Tolerance Value	EPT Richness

Table 3. Values of r^2 from mean and upper bound regressions between metrics adopted for use in the NorCal B-IBI and stressor gradients used in metric screening. Significant values ($p < 0.0008$ after Bonferroni correction) are indicated in bold.

Stressor	EPT Richness	EPT Richness upper bound	Coleoptera Richness	Coleoptera Richness upper bound	Diptera Richness	Diptera Richness upper bound	% Intolerant Individuals	% Intolerant Individuals upper bound	% Non-Gastropod Scraper Individuals	% Non-Gastropod Scraper Individuals upper bound	% Predator Individuals	% Predator Individuals upper bound	% Shredder Taxa	% Shredder Taxa upper bound	% Non-Insect Taxa	% Non-Insect Taxa upper bound
% of watershed unnatural	**-0.366**	**-0.702**	**-0.17**	-0.235	**-0.172**	**-0.499**	**-0.231**	**-0.618**	**-0.157**	**-0.417**	**-0.087**	-0.191	**-0.121**	**-0.468**	**0.473**	**0.826**
% of watershed in agriculture	**-0.137**	-0.328	**-0.078**	-0.238	**-0.064**	-0.053	**-0.062**	-0.296	**-0.105**	-0.291	**-0.045**	**-0.476**	-0.021	-0.019	**0.097**	0.197
road density in local watershed	**-0.266**	**-0.677**	**-0.119**	**-0.689**	**-0.171**	**-0.697**	**-0.19**	**-0.722**	**-0.116**	**-0.656**	**-0.106**	**-0.593**	**-0.112**	**-0.597**	**0.297**	**0.51**
% sand and fine substrates	**-0.341**	**-0.793**	**-0.189**	**-0.665**	**-0.169**	**-0.658**	**-0.163**	**-0.744**	**-0.108**	**-0.653**	**-0.082**	**-0.629**	**-0.09**	**-0.513**	**0.452**	**0.707**
conductivity	**-0.263**	**-0.467**	0.005	-0.128	-0.028	-0.152	**-0.189**	-0.239	**-0.093**	**-0.457**	**-0.17**	**-0.517**	**-0.108**	-0.355	**0.274**	0.249
total phosphorous	**-0.332**	**-0.629**	**-0.159**	-0.518	-0.114	-0.343	**-0.173**	-0.488	-0.114	-0.438	**-0.203**	-0.413	-0.108	-0.603	0.107	0.472
qualitative channel alteration	**0.133**	**0.807**	0.053	**0.750**	**0.09**	**0.803**	0.071	**0.644**	**0.079**	**0.633**	0.053	**0.647**	0.063	**0.7**	**-0.171**	**-0.629**

Table 4. Scoring ranges for 8 component metrics in the NorCal B-IBI. Six metrics have separate scoring ranges for the three Omernik Level III ecoregions in northern coastal California region (1= Coast Ranges, 6=Chaparral and Oak Woodlands, 78=Klamath Mountains).

Metric Score	EPT Richness		Coleoptera Richness		Diptera Richness	% Intolerant Individuals		% Non-Gastropoda Scraper Individuals		% Predator Individuals		% Shredder Taxa		% Non-Insect Taxa
	1 & 78	**6**	**1 & 78**	**6**	**All Sites**	**1 & 78**	**6**	**1**	**6 & 78**	**78**	**1 & 6**	**1**	**6& 78**	**All Sites**
10	>25	>20	≥6	≥8	≥10	≥41	≥28	≥41	≥18	≥22	≥16	≥20	≥16	0-7
9	23-25	19-20	5	7	9	36-40	24-27	37-40	17	19-21	14-15	18-19	14-15	8-13
8	21-22	17-18		6	8	31-35	21-23	33-36	15-16	17-18	12-13	16-17	12-13	14-18
7	18-20	15-16	4		7	26-30	17-20	29-32	13-14	15-16	11	14-15	11	19-24
6	16-17	13-14		5	6	21-25	14-16	25-28	11-12	13-14	9-10	12-13	9-10	25-29
5	13-15	11-12	3	4	5	16-20	10-13	21-24	9-10	10-12	8	10-11	8	30-35
4	11-12	9-10		3	4	11-15	7-9	17-20	7-8	8-9	6-7	8-9	6-7	36-40
3	8-10	7-8	2		3	6-10	3-6	13-16	5-6	6-7	5	6-7	5	41-46
2	6-7	5-6		2	2	1-5	0-2	9-12	3-4	4-5	3-4	4-5	3-4	47-51
1	3-5	3-4	1	1	1	-4 to 0	-4 to -1	5-8	1-2	2-3	2	2-3	2	52-56
0	0-2	0-2	0	0	0	≤-5	≤-5	0-4	0	0-1	0-1	0-1	0-1	≥57

Table 5. Estimated percentage and length of stream kilometers in each evaluation category.

Status	n	Estimated % of stream km (95% confidence interval)	Length of stream km
Landowner Denied	33	20.0 ± 5.4	4529
Non-Target	40	29.5 ± 6.4	6668
Physical Barrier	21	12.9 ± 4.4	2922
Target Sampled	59	32.4 ± 6.3	7317

Table 6. Estimated percentage and length of stream kilometers in each condition category based on B-IBI.

Condition Category	n	Estimated % of stream km (95% confidence interval)	Length of stream km
Very Poor	0		0
Poor	4	2.1 ± 2.6	154
Fair	10	16.8 ± 10.5	1227
Good	31	59.7 ± 12.4	4369
Very Good	14	21.4 ± 9.8	1569

Figure 1. Map of study area and Omernik Level III ecoregions.

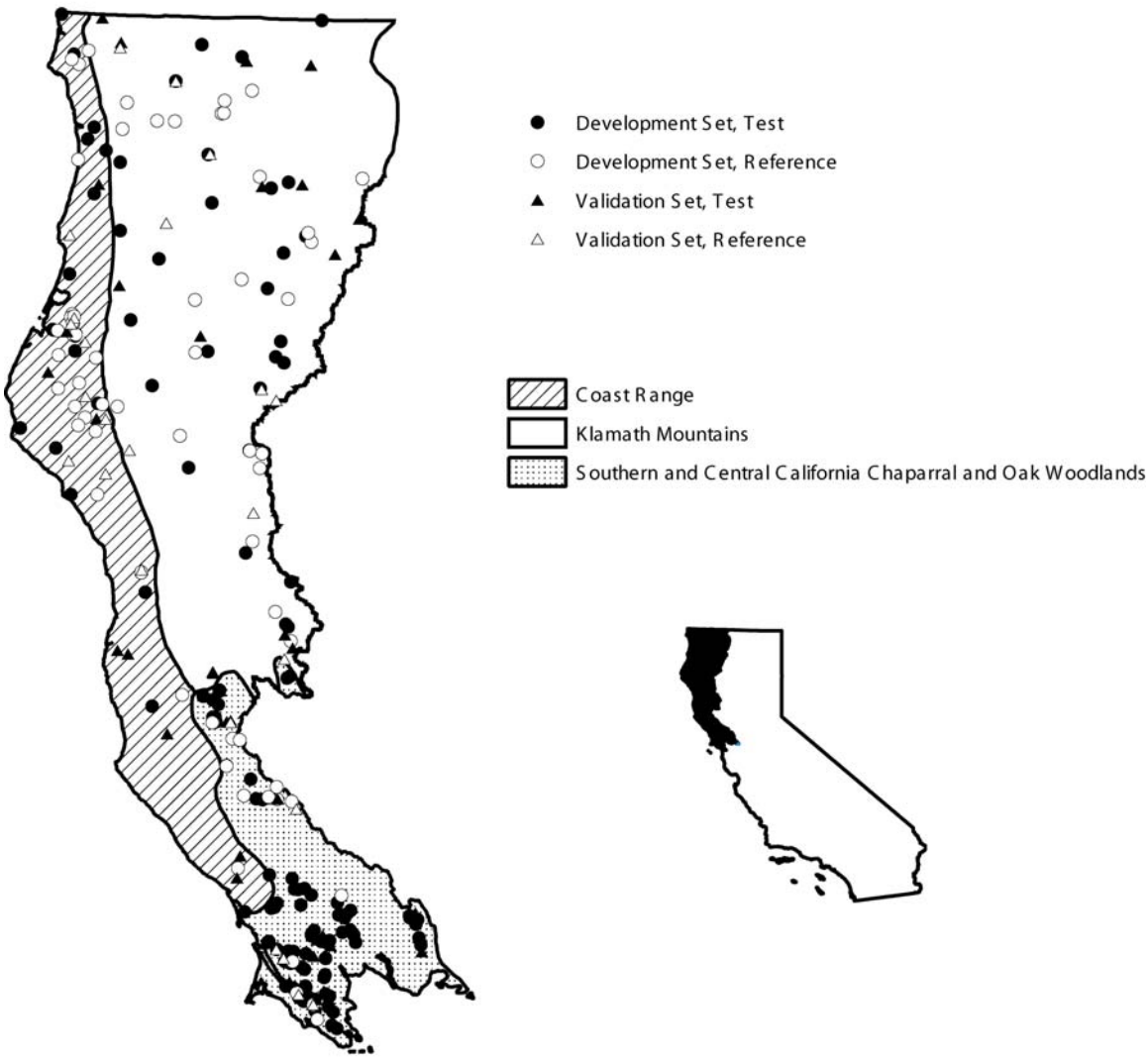


Figure 2. Box plots of metric distributions in reference sites in each Omernik Level III ecoregion in northern coastal California. Separate scoring scales were developed for metrics that differed significantly (Kruskal-Wallis p <0.05) between ecoregions.

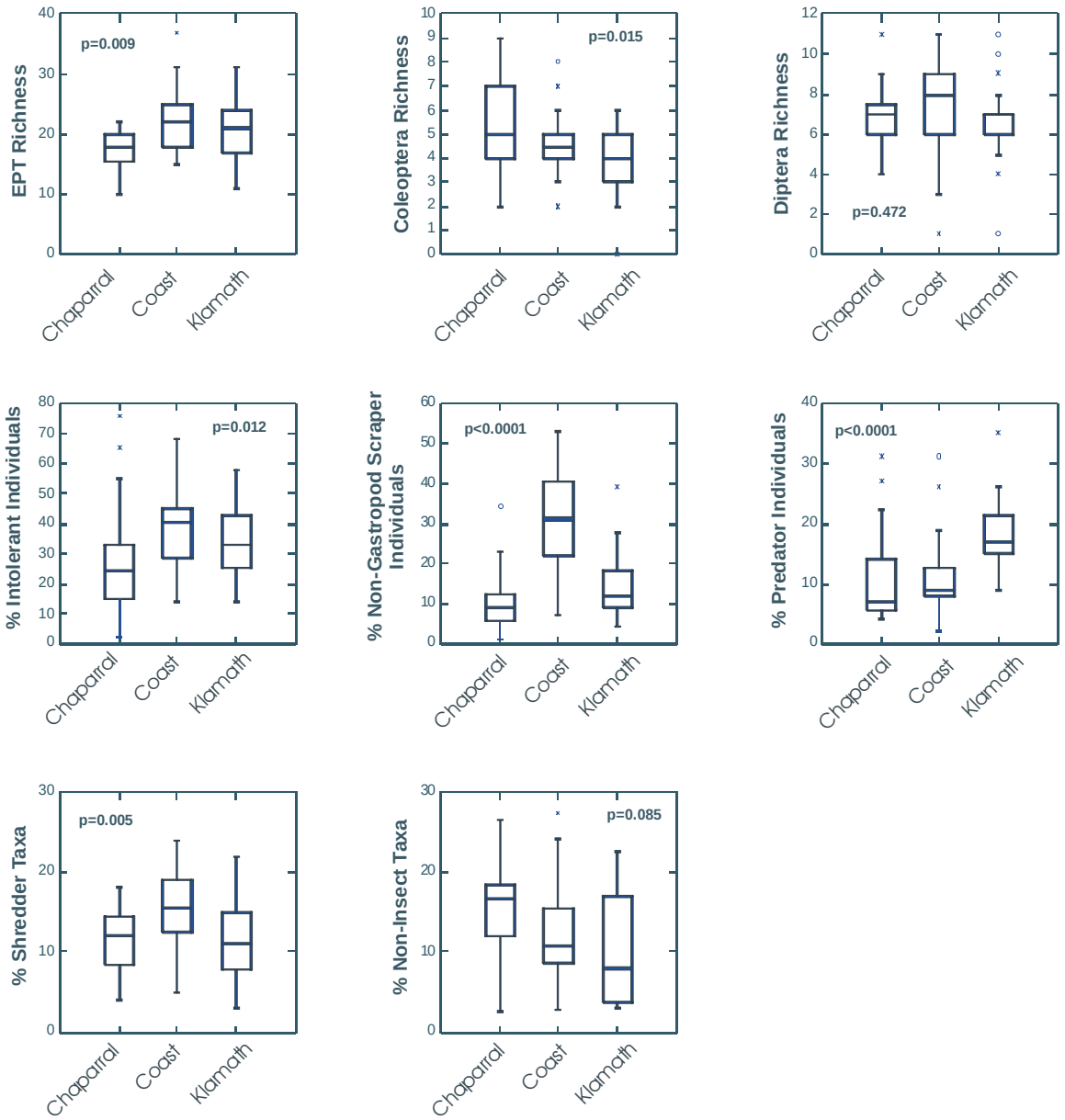


Figure 3. Box plots of B-IBI scores for reference (R), test (T) and "worst" (W) groups in development and validation data sets. "Worst" sites have ≥ 25% upstream watershed in unnatural land use (agriculture and urban). Dotted lines indicate condition category boundaries.

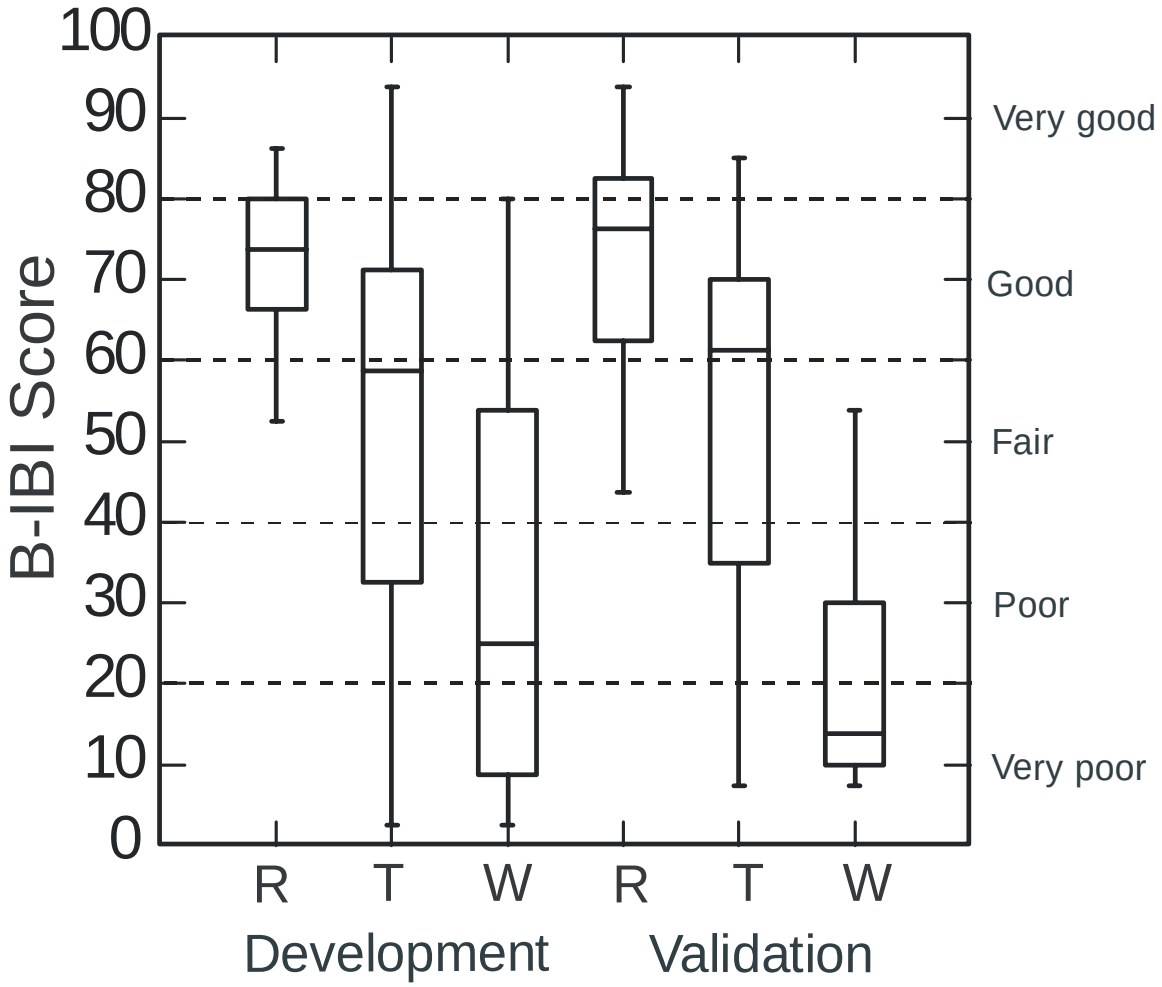
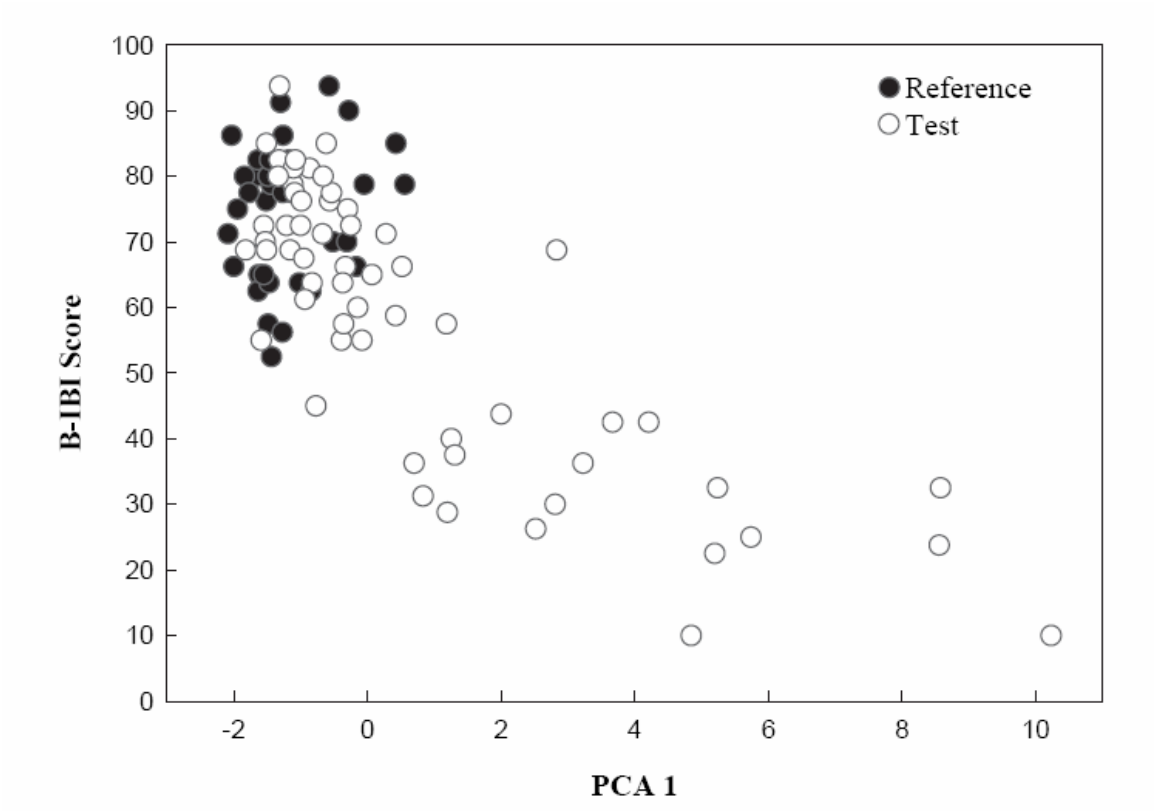


Figure 4. B-IBI score as a function of a multivariate watershed condition axis (PCA 1). Pearson correlation = -0.774, $p < 0.0001$.



Axis 1 loadings

% watershed unnatural	0.368
% watershed in agriculture	0.344
% local watershed forested	-0.336
% local watershed in agriculture	0.229
road density in local watershed	0.224
population density in local watershed	0.210
mid-channel canopy density	-0.117
% sand and fine sediment	0.252
conductivity	0.281
total phosporous	0.336
chloride	0.352

Figure 5. Relationship between B-IBI scores at 91 reference sites and (a) Omernik Level III ecoregion, (b) Julian date (day of year), (c) $\log_{10}$ watershed area, and (d) elevation.

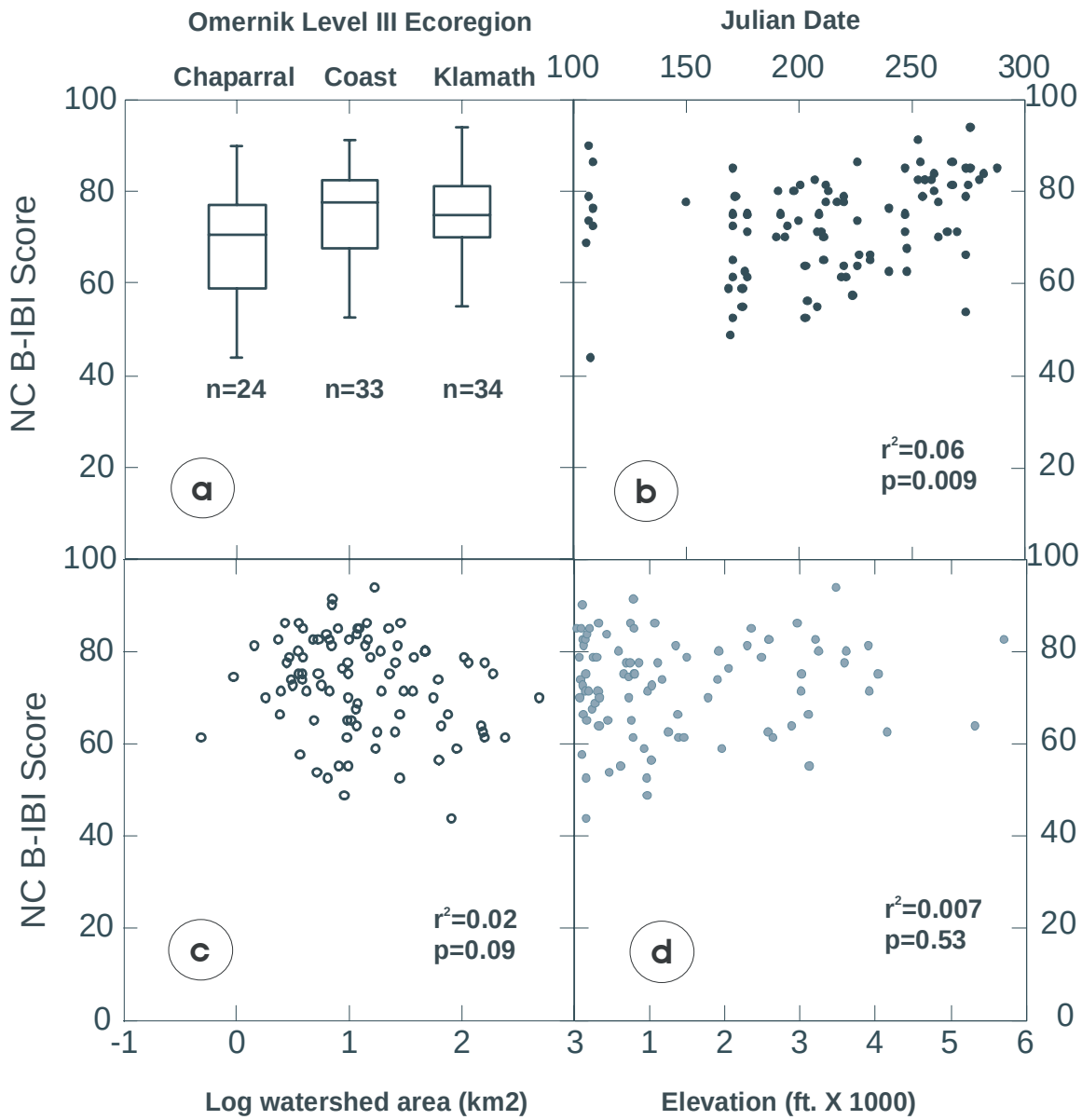
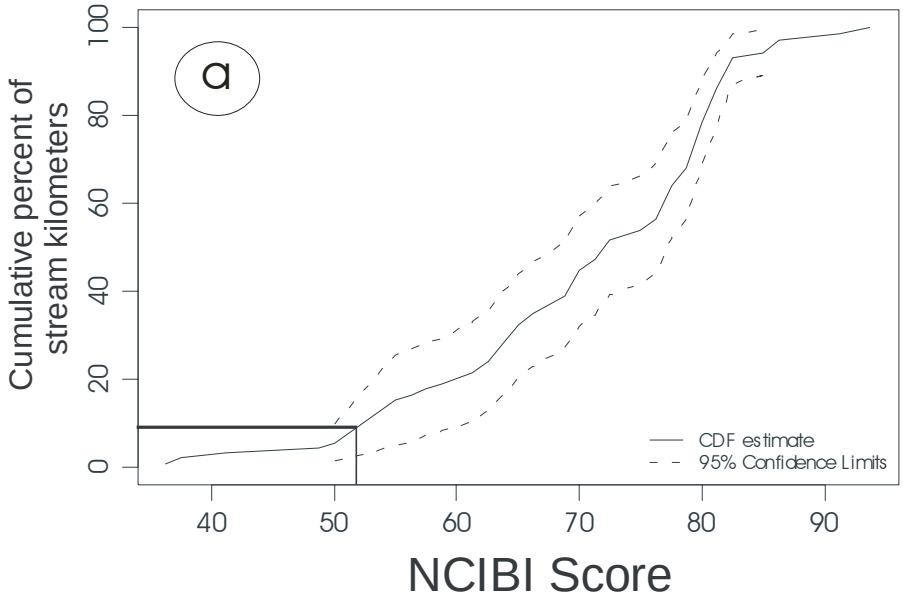
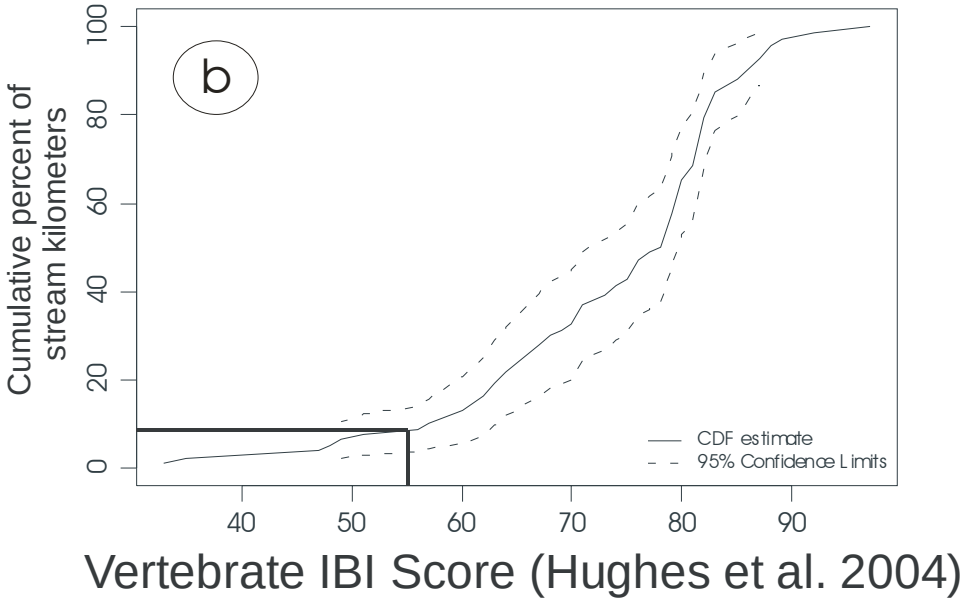



Figure 6. Cumulative distribution function of B-IBI scores (a) and vertebrate IBI scores (b) estimated from 59 probability sites in northern coastal California. Impairment thresholds 2 SDs below the mean reference score (B-IBI =52; vertebrate IBI=55) are indicated.