



**Assessing Invertebrate Responses to Human Activities: Evaluating
Alternative Approaches**

Leska S. Fore; James R. Karr; Robert W. Wisseman

Journal of the North American Benthological Society, Vol. 15, No. 2. (Jun., 1996), pp.
212-231.

Stable URL:

<http://links.jstor.org/sici?sici=0887-3593%28199606%2915%3A2%3C212%3AAIRTHA%3E2.0.CO%3B2-O>

Journal of the North American Benthological Society is currently published by The North American Benthological Society.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/nabs.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.

Assessing invertebrate responses to human activities: evaluating alternative approaches

LESKA S. FORE AND JAMES R. KARR

School of Fisheries, Box 352200, University of Washington, Seattle, Washington 98195 USA

ROBERT W. WISEMAN

Aquatic Biology Associates, 3490 NW Deer Run Road, Corvallis, Oregon 97330 USA

Abstract. The goal of biological monitoring is to evaluate the effect of human activities on biological resources. In this study, we linked human activities across landscapes to specific changes in assemblages of benthic macroinvertebrates in streams that drain those landscapes. We used data from 2nd- to 4th-order streams in southwestern Oregon to test approximately 30 hypotheses about how macroinvertebrates respond to several common human actions, especially logging and associated road construction. We found 10 attributes of macroinvertebrate assemblages to be reliable indicators of disturbance. Data from a subsequent year confirmed those results. We used simple graphical methods to evaluate land-use data and to relate the data to observed responses of invertebrates.

We constructed a multimetric index (benthic index of biological integrity, or B-IBI) from component metrics that distinguished disturbed stream sites from minimally disturbed sites. Using an independent data set, we found that B-IBI scores were significantly lower for streams whose watersheds were more degraded by human activities. We also tested rapid bioassessment protocol (RBP) III as modified by Oregon Department of Environmental Quality. RBP III failed to detect differences among sites that B-IBI did detect.

Because biologists continue to debate the relative merits of a multimetric vs. a multivariate approach for interpreting biomonitoring data, we also used principal components analysis (PCA) to explore patterns in our multidimensional data. We chose PCA based on species abundance because it has been the most common technique used by state and federal agency biologists to interpret biomonitoring data. PCA failed to detect clear differences in our data set (e.g., between most and least disturbed sites).

We suggest that multivariate statistical analyses are most appropriate for exploratory analysis when the investigator has limited knowledge of an ecological system and wants to generate testable hypotheses. We demonstrate that sufficient information about the natural history of stream invertebrates is available to support more direct tests of how invertebrates respond to human disturbance. The components of a good multimetric index are founded on tested hypotheses and thus provide a sound scientific assessment of a stream site. We argue that a method that incorporates ecological information is more suitable for biomonitoring than one relying on statistical algorithms.

Key words: biological monitoring, benthic macroinvertebrates, assessment, disturbance, streams, B-IBI, biological integrity, multivariate statistics, Oregon.

The Pacific Northwest region of the USA is on the brink of losing its salmon (Nehlsen et al. 1991, Frissell 1993). In 1994, the salmon fishery was closed along the West Coast from Mexico to Washington because of low returns of spawning adults. The decline in salmon and closure of sport and commercial fisheries (Wilderness Society 1993) represents an annual economic loss in excess of \$1 billion and 60,000 jobs (Pacific Rivers Council 1995). This trend is simply the most recent chapter in a history of decline and loss of regional aquatic resources that spans the continent (Hughes and Noss 1992, Allan and Flecker 1993), including Atlantic salmon (*Salmo*

salar) in rivers of northeastern North America, sport and commercial species in midwestern USA (Karr et al. 1985), and native fishes of California (Moyle and Williams 1990). Salmon and other aquatic organisms are disappearing because dams, timber harvest, agriculture, urbanization, and other human actions alter watersheds and degrade rivers that support populations of aquatic species. These trends are likely to continue as long as society undervalues water, water bodies, and the aquatic ecosystems they sustain (Karr 1995).

Widespread recognition of this degradation has stimulated numerous efforts to improve our

ability to track the condition of aquatic ecosystems (Davis and Simon 1995). Comprehensive, multimetric indexes (Barbour et al. 1995) were first developed in the midwestern USA for use with fishes (Karr 1981, Fausch et al. 1984, Karr et al. 1986) and modified for use with invertebrates (Ohio EPA 1988, Plafkin et al. 1989, Kerans and Karr 1994, DeShon 1995). Although the original multimetric index—the index of biotic integrity (IBI; Karr 1981)—has been adapted for fish in the Willamette River (a large Oregon river; Hughes and Gammon 1987), broad application of a fish IBI to the Pacific Northwest is restricted by 3 problems: 1) streams in the Pacific Northwest have few species of fish; 2) many fish spend a substantial portion of their life cycle in marine waters; and 3) salmonids, the dominant fish in most streams, are affected by harvesting and stocking.

In contrast, streams west of the Cascade Mountains in Oregon and Washington host several hundred easily identifiable taxa of macroinvertebrates that spend all or most of their lives in fresh water. The variety of taxa and knowledge of their natural histories make it possible to distinguish different types and levels of human-induced degradation. Thus, we evaluate the biotic integrity of streams based on analysis of the benthic invertebrate assemblage.

In this paper, we advocate simple graphical and statistical analyses rather than multivariate statistics to relate human activities to biological changes. We argue that simple graphs are easier to explain to an audience of non-scientists than are multidimensional plots derived from multivariate analyses such as principal components analysis (PCA). We test specific hypotheses about how invertebrates respond to human disturbance, primarily logging, with data from 2 different years. Based on those attributes (or metrics), we modify the original benthic IBI (B-IBI) developed for Tennessee Valley streams (Kerans and Karr 1994). Using an independent data set, we test whether B-IBI and a similar index, Rapid Bioassessment Protocol (RBP) III (Plafkin et al. 1989) as modified by Oregon Department of Environmental Quality (DEQ) (Mulvey et al. 1992), can detect differences in intensity of logging. Finally, we discuss why multimetric indexes with known statistical properties and constructed from tested hypotheses are better for ranking and assessing resource condition than is the output from PCA.

Methods

Study location

No new data were collected for this study. Of the data available from Alaska, Idaho, Oregon, and Washington (US EPA's Region 10), we chose streams from southwestern Oregon for our study because many sites had been sampled during 3 y and land-use data were some of the best available. Macroinvertebrates had been collected from mid-order streams within the drainages of the Umpqua and Rogue rivers in the western foothills of the Cascade Mountains south of Eugene, Oregon. Watershed area ranged from 15 km² to 400 km² with a median area of 54 km². Original vegetation was primarily Douglas Fir (*Pseudotsuga menziesii*).

Our study area was in the Umpqua National Forest (NF) and in the Roseburg and Medford Districts of the Bureau of Land Management (BLM), the latter in a checkerboard pattern of interspersed public and private lands. Within the National Forest, clearcuts have increased the frequency and intensity of runoff which has simplified channel morphology and increased debris torrents (USFS 1992). Logging and road-building have increased sediment and significantly widened stream channels since 1937 (Dose and Roper 1994). Watersheds under BLM and private ownership are modified by timber harvest as well as grazing, agriculture, and urbanization.

Data sets

An ideal data set for testing metrics would include sites that were physically identical and differed only in their degree of human influence. Data available for the Pacific Northwest were not ideal, and so we had to first identify a subset of the data that included a sufficient number of physically similar sites. We controlled for the effect of large-scale geographic features by selecting sites within southwest Oregon. We minimized seasonal effects by using fall-sampling data and avoided year effects by testing within 1 year. We selected 2nd- to 4th-order streams and excluded Umpqua River sites because the mainstem is a much larger river with greater taxa richness, different species composition, and markedly different biological dynamics. We divided the data into subsets for

TABLE 1. Number of sites available, number used in B-IBI analysis (in parentheses), and analysis for which data were used (footnotes) for each data set. High elevation and large river sites were excluded for metric and index testing and the 2nd principal component analysis (PCA); after exclusion of these sites, some data sets had too few sites remaining for analysis.

Year	Umpqua NF	Roseburg BLM	Medford BLM
1990	46 ^a (32 ^{a,b,c})		
1991	30 (20 ^d)	30 (30 ^{c,d})	22 (18 ^e)
1992	30 (24 ^e)	12 (0)	21 (0)

^a PCA

^b Testing metrics

^c Determining metric ranges

^d Testing B-IBI and RBP

^e Validating metrics

each phase of analysis according to year and location (Table 1):

Metric testing. We evaluated benthic invertebrate responses to timber harvest and road building with data collected in 1990 from 32 sites in the Umpqua NF.

Metric validation. We validated patterns observed in 1990 metric scores with data collected in 1992 from 24 Umpqua NF sites.

Index testing. We compared overall index scores of B-IBI and RBP for 20 Umpqua sites (1991) and 30 Roseburg sites (1991). The different levels of human activity in these 2 areas allowed us to compare B-IBI and RBP.

Sites sampled in the Umpqua NF in 1991 and 1992 were a subset of those originally sampled in 1990 except for a few (<5) new sites sampled each year. An ideal data set would have included sites independent in both space and time. In our case, the best available data sets for testing metrics were independent in time only (1990 and 1992). We saved the "best" data sets (Roseburg 1991 and Umpqua 1991) for testing the B-IBI and RBP indexes—data sets independent in time and space from data sets used to test and validate the metrics used in B-IBI. For index testing, rather than comparing sites within the Umpqua NF with each other, Umpqua NF sites were compared as a group with sites in the Roseburg District of the BLM.

Data collection

Invertebrates.—The data sets described above were originally collected for the US Forest Service and the Bureau of Land Management to assess watershed condition. Consistent data quality was assured because all samples were collected and identified by one of us (RWW) according to the same sampling and lab protocols. Macroinvertebrates were collected using a kicknet (500- μ m mesh) to sample an approximately 30-cm $\times$ 70-cm area from 5 erosional habitats (typically riffles) along an approximately 150-m reach. The 5 collections were composited into a single collection representing a total area of 1 m². All individuals collected were identified and counted, unless the sample was greater than 1000 individuals, in which case a subsample was processed. Insects were dominant in the benthic assemblages, and were usually identified to genus or species. Level of taxonomic identifications for non-insects varied from species to phyla. Invertebrates were collected in the fall because the largest number of taxa are typically present at this time and it is a practical and safe season for field crews.

A single location in each watershed yielded 1 collection of invertebrates for each stream. We analyzed riffle data for 3 reasons: (1) riffles are easier to define by field crews than pools or margins; (2) riffles are more uniform than other stream microenvironments and are thus easier to compare across watersheds; and (3) riffles have high current velocity and shallow depth, facilitating sampling with kicknets or Surber samplers. Pools provide information that riffles do not and there is some evidence that they are degraded by sediment before riffles (Kerans et al. 1992); however, our priority was to evaluate a sampling protocol that could be easily applied by others.

Land use.—We evaluated human influence in the Umpqua NF at 2 landscape scales: 1) over the entire watershed and 2) within the stream channel. At the watershed scale, the US Forest Service (USFS) provided estimates of total watershed area, watershed area logged, and lengths of road within each watershed for Umpqua sites. Unfortunately, information about the type, location, and date of timber harvest was not available. At the riparian scale, we worked with USFS hydrologist Mikeal Jones to estimate the instream condition of sites within the Ump-

qua NF. Over the past 15 y he has walked the stream channels in order to place instruments, collect water chemistry data, and recommend buffer sizes for timber sales. Based on his observations of riparian corridor, stream bed, bank stability, and influence of road building and logging on the stream channel, Jones provided us with a relative ranking of streams within each of the 7 sub-basins of the Umpqua River. In addition, he described an overall, general condition for each sub-basin. We assigned a value from 1 to 7 to each stream according to its condition relative to other streams in its sub-basin and then adjusted the scores up or down if the entire basin was unusually protected or disturbed.

Water chemistry data were available for very few sites. Supplementary information regarding elevation, location of roads and urban areas, location of private and federal land, and stream size was obtained from topographic maps of the US Geological Survey. For Medford and Roseburg BLM sites west of the National Forest, specific information about land use was available for only a few watersheds.

Data analysis

Our goal was to extract relevant biological pattern from taxa lists and counts of macroinvertebrates and relate those patterns to human-induced changes in the watersheds. We evaluated data sets with a multimetric approach (Karr et al. 1986, Kerans and Karr 1994, Barbour et al. 1995) and a multivariate approach (principal components analysis) (Norris and Georges 1993) to identify patterns in macroinvertebrate assemblages.

Evaluating human influence.—Initially we had to determine the level of human disturbance at each site. First, we selected a subset of physically similar sites. Second, we compared the rankings of sites according to percent area logged and instream condition to identify data errors and inconsistencies. After corrections, and after taking into consideration ancillary data such as road location and type of logging, we determined that the land use data only had the precision to define 3 categories—least, moderately, and most disturbed.

Sites with both high percentages of logged area and poorest instream condition were rated as most disturbed. For 1990, we selected 4 sites

as the most disturbed (Brownie and Tom Creek, lower Cow Creek, upper Elk Creek and lower Elk Creek) and 4 sites as the least disturbed (Calf Creek, South Fork and East Fork of Cow Creek, and Quartz Creek). Sites with the best instream condition had been logged, but had been left to recover (upper forks of Cow Creek) or had been logged from ridgeline roads rather than roads along the valley floor (Quartz Creek). The 1992 data set was smaller and we selected 3 best sites (South Fork Cow Creek, North Boulder Creek, and Quartz Creek) and 2 worst sites (lower Elk Creek, lower Jackson Creek).

Least disturbed sites were not pristine; because all watersheds have been logged, we chose sites that were the best available. Best sites did not have roads paralleling the mainstem channel, had been logged least or longest ago, and were not easily accessible to humans. The most disturbed sites were affected by roads along the bank of the stream, heavy logging, grazing, and urbanization.

Metric testing.—For the multimetric analysis, metrics started out as hypotheses about how invertebrates respond to disturbance (Table 2); for example, we expected the number of Plecoptera taxa to decline as human influence increased. The hypotheses we tested came from published literature (Cummins et al. 1989, Karr and Kerans 1991), existing protocols and multimetric indexes (Plafkin et al. 1989, Hayslip 1993, Kerans and Karr 1994, DeShon 1995), and our field experience.

Metrics were selected from 4 general classes: taxonomic richness and composition, tolerance and intolerance, feeding ecology, and population attributes. We tested metrics based on the taxonomic richnesses of the major orders of insects as well as other specific groups, such as intolerant taxa. Classification of tolerant and intolerant taxa has a long history from the early Saprobien system (Kolkwitz and Marsson 1909) to more recent efforts (Chutter 1972, Hilsenhoff 1982, Lenat 1993). Our approach differs somewhat in that we identify only the most and least tolerant taxa rather than assigning a tolerance score to every taxon. In addition, we identify the first and last taxa to disappear in response to a broad array of disturbances rather than any single influence, such as organic pollution. Tolerant and intolerant taxa were identified based on 20 y of author RWW's experience while collect-

TABLE 2. Attributes of the invertebrate assemblage (metrics), their hypothesized response to human disturbance, and their ability to distinguish most and least disturbed Umpqua NF sites in 1990 and 1992.

Metric	Predicted response	Metrics that could distinguish sites
Taxa richness & composition		
Total number of taxa ^a	Decrease	✓
Ephemeroptera ^a	Decrease	✓
Plecoptera ^a	Decrease	✓
<i>Pteronarcys</i> species ^a	Decrease	✓
Trichoptera ^a	Decrease	✓
Diptera	Decrease	
Chironomidae	Increase	
Tolerance/intolerance		
Intolerant taxa ^a	Decrease	✓
Sediment intolerant taxa ^a	Decrease	✓
Sediment tolerant taxa	Increase	✓
% Tolerant ^a	Increase	✓
% Sediment tolerant ^a	Increase	✓
% Sediment intolerant	Decrease	
% Oligochaeta	Increase	
% Chironomidae	Increase	
Feeding ecology		
% Scraper	Variable	
% Predator	Decrease	
% Gatherer	Variable	
% Shredder	Variable	
% Filterer	Variable	
% Omnivore	Increase	
Scraper taxa	Variable	
Shredder taxa	Variable	
Omnivore taxa	Increase	
Population attributes		
% Dominance (3 taxa) ^{a,b}	Increase	✓
Abundance ^a	Decrease	

^a Component metric included in benthic index of biotic integrity (B-IBI)
^b Also tested 1, 2, 4, and 5 dominant taxa

ing and identifying invertebrates from streams throughout the western USA. Invertebrate taxa were classified as scraper, predator, gatherer, shredder, or filterer according to the primary category documented by Merrit and Cummins (1984) and in some cases modified according to RWW's experience with local taxa. When ambiguities about feeding ecology could not be resolved, taxa were identified as omnivore or unknown.

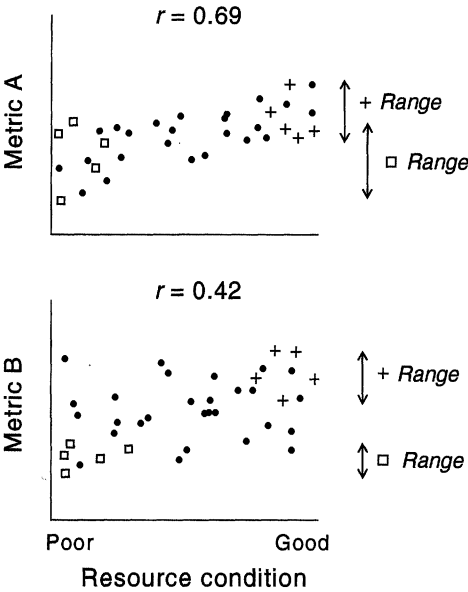


FIG. 1. Hypothetical relationships between resource condition and candidate biological metrics. Metric A is more strongly correlated with resource condition (or r^2 is higher if using regression) than metric B, suggesting it is a better metric. A more reliable test compares the metric's ability to distinguish between minimally disturbed (+) and severely degraded (□) sites (ranges noted by arrows). Note that moderately degraded sites (●) span the range between the extremes. Metric B is the more effective metric in spite of its smaller statistical correlation.

We used graphical analysis to judge candidate metrics because it provides insight into how invertebrates respond to different levels of human influence (e.g., monotonically or at some threshold) and because graphs reveal outliers that stimulate fresh hypotheses. Metrics that had little or no overlap between most disturbed and least disturbed sites and that arrayed moderately disturbed sites between the extremes (Fig. 1) were selected and tested again with data from a subsequent year. Only metrics that responded to a gradient of human influence in both 1990 and 1992 were included in B-IBI. Because 1992 was a drought year, we had greater confidence that metrics selected were minimally influenced by this source of natural variability.

Statistical considerations.—We could have used a statistical test (e.g., Mann-Whitney test for 2 independent samples) to decide if the most and least disturbed sites were significantly different. Instead, we based our decisions on graphical

analysis because it provided more insight into biology than a simple p -value could. For example, we were interested in whether scores for best and worst sites clumped together tightly or were spread more evenly. Graphs allowed us to examine each metric's range and to evaluate over what range of human disturbance a metric was most sensitive.

A typical alternative approach might select metrics according to statistical correlations with water chemistry, percent area logged, or similar variables. In our study area, although logging was the primary land use, the percentage area of the watershed logged did not differentiate between sites that had been clearcut to the river bank and those sites that had riparian areas left intact. Another limitation was that the year of the cut and the frequency of cutting were also unknown. Our situation was not unique; rarely is a single variable available that can summarize all the human activities that affect streams and provide a linear ranking of sites. Typically multiple human activities influence watersheds simultaneously. Watersheds may be affected by mining, in which case water chemistry data are appropriate, or by channel modification by dams or roads, in which case physical habitat data are relevant. The narrow use of statistical correlation, or any other statistical model to make decisions, can lead to a dependence on p -values. Tests of significance are overused by ecologists (Yoccoz 1991); they focus on detection of impacts rather than on their magnitude or importance (Stewart-Oaten 1996). Routine statistical testing can short-circuit the process of looking at the data and asking whether they make sense and what they show.

Index construction and testing.—Following the method developed for the fish IBI (Karr 1981, Karr et al. 1986), metrics were combined into an index by transforming the metric values to a score of 5 (similar to or deviating slightly from minimally disturbed, or reference, condition), 3 (moderate deviation), or 1 (strong deviation from reference condition). The sum of the metric scores gave a B-IBI score for each site.

Simple, general rules for setting metric scoring criteria are difficult to define because they depend on the sampling design that generated the data. If sites from a region are sampled such that $\frac{1}{3}$ are in pristine areas, $\frac{1}{3}$ in moderately disturbed areas, and $\frac{1}{3}$ in very disturbed areas, then the range of each metric can be divided at

the 33rd and 67th percentiles. In our case, we suspected that our data sets included neither very disturbed sites (e.g., sites from urban areas) nor pristine sites with no human influence. Therefore, we tried to err on the conservative side by broadening the middle category (3) to include more sites, thus making it more difficult for a site to score a 5 or 1. Natural breakpoints in the distribution probably reflect relevant biology. Where these occurred, we adjusted the scoring criteria to fall at these points. Setting scoring criteria is an iterative process and should be reconsidered as more regional data become available across a broader range of human influences.

We tested the Oregon DEQ version of the RBP because it resolves some of the problems of the original RBP III. Ratio metrics in the original RBP, such as scrapers/filterers are replaced with metrics such as percentage scrapers and percentage filterers. The original RBP calculated scoring criteria as a percentage of a reference site, which we could not do because reference sites were not sampled. In general, setting scoring criteria as a percentage of a reference site is a poor approach because it fails to recognize that reference condition is more meaningful when defined as a range of values rather than as a single value. Oregon DEQ recognized this problem and defined scoring criteria that assign a score of 5, 3, or 1 to metric values (Mulvey et al. 1992) similar to B-IBI. We used these scoring criteria to calculate RBP.

To summarize our multimetric approach, we started with metrics that varied systematically along a gradient of human influence for fish (Miller et al. 1988) and invertebrates in other regions (Ohio EPA 1988, Kerans and Karr 1994). We also included new metrics suggested for invertebrates. We tested metrics with data from the Umpqua NF, selected metrics that varied along a gradient of human influence, and tested those successful metrics with data from a subsequent year. Finally, we tested B-IBI and RBP by comparing index scores for sites from yet another year and from 2 areas (Umpqua NF and Roseburg District of the BLM) with different levels of human influence.

Principal components analysis.—We chose PCA because it is the most common multivariate technique in use by state and federal biologists to interpret monitoring data; in the late 1980s it was also the most common multivariate method

reported in ecological journals (James and McCulloch 1990). The typical application of this procedure uses taxa lists and abundances to interpret differences between stream sites (Norris and Georges 1993). We present 2 PCAs based on the 1990 Umpqua NF data: all 46 sites and a subset of 32 sites (used in the initial round of metric testing).

Principal components analysis calculates the line ("the component") that extracts the maximum amount of statistical variance from a cloud of points. In our case each point in the cloud represents a stream site. The number of dimensions through which the line passes is equal to the number of taxa collected. PCA solutions are degraded by departures from multivariate normality (Tabachnick and Fidell 1989). In particular, outliers, such as sites with high abundances or many zeros, can cause problems (Kenkel and Orlóci 1986). To address this issue, abundance data are often transformed by normalizing (i.e., subtracting the mean and dividing by the standard deviation) or by taking the logarithm. For our data, neither approach was satisfactory because neither transformation resulted in normally distributed variables. Three versions of the same data were used in PCA: normalized, log transformation, and no transformation. We present only results from the normalized data because that transformation best supported the underlying assumptions of the model.

Other ordination techniques, such as correspondence analysis, are recommended for species abundance—or count—data (Manly 1994). These ordination methods rely on the same mathematical rules to extract maximum statistical variance from the variance-covariance matrices of species and/or sites (Ludwig and Reynolds 1988, p. 255). In addition, correspondence analysis assumes that species abundance changes along an environmental gradient by increasing, peaking, and then declining (the Gaussian model) (Gauch 1982, p. 243). Our data did not support this assumption of an underlying modal optimum for species abundance. Another multivariate technique, canonical correlation analysis, can be used to relate two sets of variables to each other, for example, land use variables and taxa abundances. For our land use data, too many variables had missing values; therefore, we simply identified important site characteristics afterward on the PCA plots.

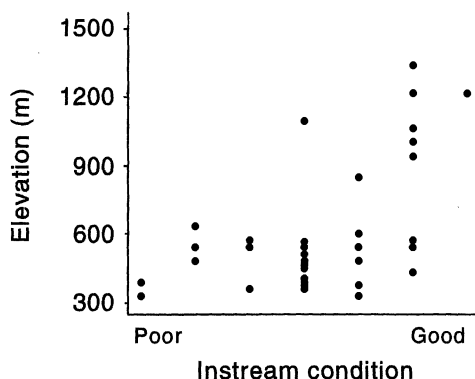


FIG. 2. Relationship between elevation and instream condition (based on channel morphology, bank erosion and stability, and riparian condition) for the Umpqua NF ($n = 41$). Sites above 750 m were excluded from analysis because very little logging occurred at high elevation; thus, elevation was confounded with disturbance.

Results

Classifying sites and evaluating human influence

Our goal in classifying sites was to obtain a subset of sites that were physically similar and did not confound human influence with other physical attributes. For example, high elevation sites were logged less (Fig. 2) and had fewer roads in their watershed; thus, high elevation sites were biased toward more pristine conditions. In addition, high elevation sites had a much higher number of taxa, but we could not determine if high taxa richness was a result of pristine conditions or elevation. We excluded high elevation sites from the multimetric analysis because sites below 750 m formed a cohesive group and were spread more evenly across the range of human influence.

Percent area logged measured human influence at the watershed scale; instream condition measured human influence at the riparian scale. We expected measurements at 2 different scales to roughly agree for these sites because the primary human activity (logging) affected both the watershed and the instream channel (Fig. 3). Given the many sources of variability and error in land-use data, we expected the agreement to be rough at best. Plotting land-use variables against each other provided a check for data consistency. Unexpected outliers pointed us toward data errors and new hypotheses to be test-

ed. For example, 2 sites with low harvested area but very poor instream condition made us aware of several watersheds for which timber harvest was underestimated: estimates were only for public land and did not include logging on private land within the watershed. We corrected estimates of percent area logged for watersheds with private land by estimating that the area cut on private land was the same as on federal land, probably an underestimate.

Decline in taxa richness is one of the most reliable indicators of degradation for many aquatic groups (Ford 1989, Barbour et al. 1995): periphyton (Bahls 1993); phytoplankton (Schelske 1984); riverine fish (Karr 1981, Miller et al. 1988, Ohio EPA 1988, Lyons 1992, Lyons et al. 1995); lake fish (Minns et al. 1994); estuarine fish (Deegan et al. 1993); and invertebrates (Ohio EPA 1988, Kerans and Karr 1994, DeShon 1995). Therefore we used taxa richness to test whether road density, presence or absence of roads adjacent to the channel, or percent area logged most accurately reflected the impact of human activities on the biota. Taxa richness declined significantly (Spearman's rho, $p < 0.05$) as area logged increased (Fig. 4), but did not change systematically with density or presence of roads. We were initially concerned that most

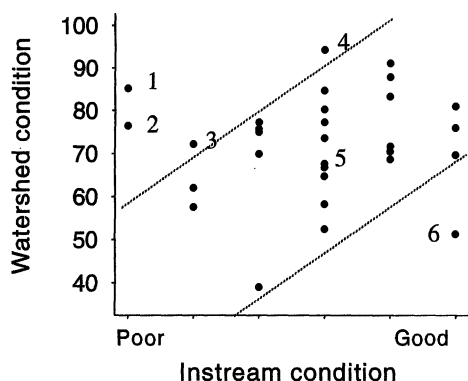


FIG. 3. Relationship between watershed condition at a site (percent area *not* logged) and instream condition (based on channel morphology, bank erosion and stability, and riparian condition). Dashed lines denote direction of expected correlation and are drawn by eye to bound groups of sites. Points that fell outside the lines were judged outliers. For outlier sites, area logged was underestimated (1, 2, 4); data were tabulated incorrectly (for 3 and 5, instream condition was reversed); and some logging had occurred long ago (6).

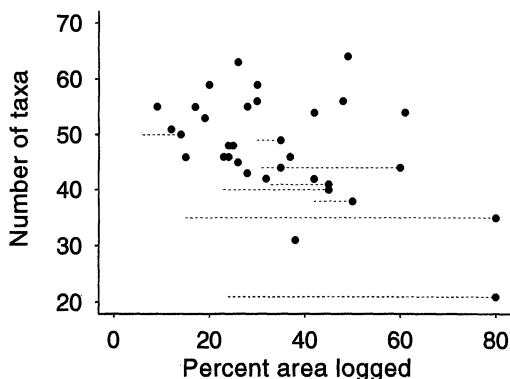


FIG. 4. Relationship between number of taxa and percent of watershed logged for the Umpqua NF. For some sites, we increased logged area (dotted lines) to reflect private-land logging within the watershed (see text for details). Before data correction, correlation (Spearman's rho) was not significant; after correction, it was significant ($p < 0.05$).

of the sites were similarly degraded and therefore might not provide a broad enough range of disturbance to test metrics. Significant correlation between area logged and taxa richness, in spite of the inaccuracy of the land-use data, gave us confidence that the sites were sampled across a gradient of human-induced change, and that best and worst sites were truly different. Because our analysis was limited to existing data sets, our sites probably did not include true extremes.

Defining and testing metrics

Among 30 biological attributes (Table 2) in 4 broad classes—taxa richness and composition, tolerant and intolerant taxa, feeding ecology, and population attributes—only 10 successfully distinguished most and least disturbed sites for 1990 and 1992. Those 10 metrics and 1 other (see discussion of abundance below) were included in B-IBI (Table 3).

Taxa richness and composition.—Total taxa richness and richnesses of Ephemeroptera, Plecoptera, and Trichoptera easily separated the best sites from the poor sites and, in general, declined as disturbance increased (See Fig. 5 for examples). Plecoptera are apparently more sensitive to human influence, because they disappeared from sites at lower levels of disturbance than did either Ephemeroptera or Trichoptera. *Pteronarcys* species richness indicated better

TABLE 3. Metrics included in B-IBI for 2nd- to 4th-order Oregon streams, their response to human disturbance, and class boundaries used as scoring criteria. A score of 5 indicates a metric value that deviates little or none from expected condition, 3 indicates moderate deviation, and 1 strong deviation from expectation.

Metric	Response to disturbance	1	3	5
Total number of taxa	Decrease	<40	40–54	>54
Ephemeroptera taxa	Decrease	<8	8–11	>11
Plecoptera taxa	Decrease	<6	6–9	>9
<i>Pteronarcys</i> species	Decrease	0	—	>0
Trichoptera taxa	Decrease	<6	6–9	>9
Intolerant taxa	Decrease	<2	2–5	>5
Sediment intolerant taxa	Decrease	0	1	>1
% Tolerant	Increase	>40	20–40	<20
% Sediment tolerant	Increase	>15	5–15	<5
% Dominance (3 taxa)	Increase	>55	40–55	<40
Abundance	Decrease	<500	500–1500	>1500

stream condition. Neither Diptera nor Chironomidae taxa richnesses changed systematically along a gradient of human influence.

Tolerance and intolerance metrics.—For tolerance and intolerance metrics, we identified the most tolerant (about 35 out of 200 total taxa) and extremely intolerant taxa (about 35 taxa). We also identified taxa that were specifically tolerant or intolerant to fine sediment, which included some of the taxa in the general tolerance/intolerance metrics and others as well.

We express tolerance metrics as a percentage of the total number of individuals in a sample that are tolerant; intolerance metrics are expressed as taxa richness (Karr et al. 1986). Tolerant organisms occur at all sites but tend to dominate in relative abundance as conditions are degraded. In contrast, intolerant taxa, when present, are typically never very abundant, making accurate estimates of their relative abundances difficult with a reasonable sampling effort; yet, the presence of these taxa sends a strong signal about local environmental quality. Our results confirmed the generality of this pattern: tolerance metrics are best expressed as percentages of total abundance and intolerance metrics should be expressed as taxa richness.

Feeding ecology.—We were surprised to find that feeding ecology metrics (Merritt and Cummins 1984), calculated as either percentage of individuals or taxa richness, failed to distinguish the most and least disturbed sites (See Fig. 5 for examples). Percentage of scrapers in the sample distinguished good from poor sites in both 1990 and 1992, but the best sites had a

smaller percentage of scrapers in 1990 yet a higher percentage in 1992. The year 1992 was unusually dry, suggesting that temporal variation in environmental conditions may affect scrapers and possibly the relative abundances of other trophic groups as well.

Population attributes.—High abundance relative to other species was tested as cumulative percentage of the 1–5 most abundant taxa. The best form for this metric calculated dominance as the cumulative abundance of the 3 most abundant taxa divided by total abundance. Total abundance did not satisfy our criteria for selection. Nonetheless, we included it because we expected very low abundances only in the most extreme situations, such as high levels of urban or industrial contaminants. Very high abundances may also indicate degradation, particularly in agricultural areas where nutrient enrichment affects invertebrates. In general, species abundances (or population size or density) are perhaps the most variable aspects of an assemblage; thus, all other metrics were calculated as a percentage of total abundance (relative abundance) or as taxa richness.

Constructing and testing B-IBI

Metrics differed in their range of values; therefore, we transformed them to a similar scale before combining them into a summary index, B-IBI. We divided the range of each metric into three parts and assigned a new score of 5 (similar to expected, or reference, condition), 3, or 1 (strong deviation from expected condi-

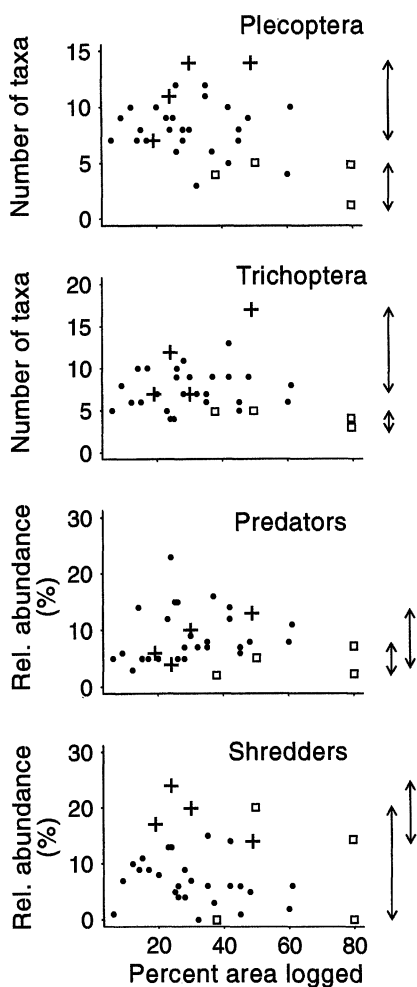


FIG. 5. Relationship between invertebrate metrics and percent area logged. Metrics that clearly distinguish least degraded (+) from most degraded stream sites (□) are used in B-IBI. Moderately degraded sites are also shown (●). Two taxa richness metrics (upper panels) distinguished the 2 groups of sites. Trophic structure metrics, such as percent of predators and shredders (lower panels), failed to distinguish the groups of sites. Arrows to the right of each plot indicate the range of metric scores for good and poor sites.

tion) (Table 3, Fig. 6). The 11 selected metrics were added to get a final B-IBI score (maximum = 55; minimum = 11). We determined the potential range of metrics based on 80 sites: 32 from the Umpqua NF (1990), 30 from the Roseburg District of the BLM (1991), and 18 from the Medford District of the BLM (1991).

Some data used to set scoring criteria (30 Roseburg sites) were subsequently used to test B-IBI. Their inclusion should not bias our test of the index because they were used only to assess the range of metrics. For most metrics, very low and very high scores were found in all 3 study areas. Roseburg sites did not have the lowest score (1) for most metrics, and neither did Umpqua sites always score high (5); rather, metric scores for Umpqua sites were on average higher than scores for Roseburg sites, which provided another check of the metrics.

We tested the hypothesis that B-IBI scores for Umpqua stream sites would be higher, reflecting higher biotic integrity, than for Roseburg sites. Although the range of B-IBI scores was similar for the 2 areas, the poorest sites in Roseburg were more degraded than the poorest sites in the Umpqua (Fig. 7). The Roseburg area includes BLM and private land in a checkerboard pattern. Private land, in general, has been logged more heavily and has been affected further by urbanization, roads, mining, and agriculture. A significant difference between index scores for the 2 areas (1-sided t -test, $p < 0.05$) demonstrated that B-IBI scores reflect the influence on instream biota of different levels of human activities in the 2 study regions.

Testing RBP

The RBP index as modified by Oregon DEQ (Table 4) failed to detect a difference between sites in Roseburg BLM and Umpqua NF (1-sided t -test, $p > 0.05$; Fig. 7). RBP metrics represent a set of working hypotheses that must be tested (like all metrics) with data from the region to be monitored. RBP metrics have not been tested in northwestern USA (and only minimally tested elsewhere). We tested the component RBP metrics with the same criteria described above and found that only 4 of the 10 metrics could distinguish between most- and least-disturbed sites (Table 4). One of those metrics, total taxa richness, was included in B-IBI. Taxa richness for Ephemeroptera, Plecoptera, and Trichoptera is included in B-IBI but as a separate metric for each order, rather than as a single metric for all three richnesses (EPT).

Interpreting principal components analysis

PCA results for the 46 Umpqua 1990 sites were similar for 3 versions of the data (normal-

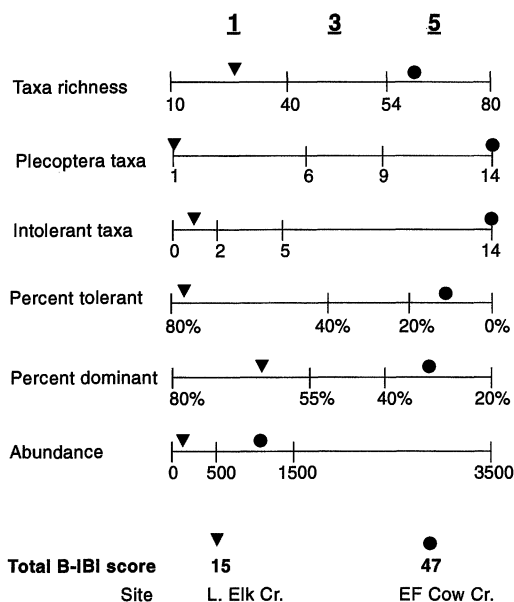


FIG. 6. Range and quantitative values for 6 B-IBI metrics at a severely disturbed site (Lower Elk Cr.) and a less disturbed site (East Fork Cow Cr.). Because metrics have different ranges, they are scored as 5 (indicating little or no deviation from expected, or reference, condition), 3, or 1 (indicating strong deviation from expected condition) to put them on the same scale. Tick marks represent the class boundaries used to assign scores of 5, 3, or 1. Total B-IBI score equals the sum of the 6 metric scores and 5 others not shown.

TABLE 4. Metrics included in Oregon DEQ's RBP (Mulvey et al. 1992), their predicted response to human influence, and whether they could distinguish least disturbed from most disturbed sites in southwest Oregon.

Metric	Predicted response	Could it distinguish?
Total number of taxa	Decrease	yes
EPT taxa	Decrease	yes
% EPT taxa	Decrease	yes
% Chironomidae	Increase	no
% Dominance (1 taxon)	Increase	no ^a
% Shredder	Increase	no
% Filterer	Increase	no
% Scraper	Decrease	no
Hilsenhoff biotic index	Increase	no
Diversity index, H'	Decrease	yes ^b

^a A better form for this metric uses 2 or 3 taxa

^b Strongly correlated with taxa richness

ized, log transformed, and no transformation). For all 3 data versions, sites identified as most disturbed, large river, and high elevation showed a slight tendency to form their own groups when plotted against the 1st and 2nd principal components. Least disturbed sites and moderately disturbed sites did not form distinct groups and none of the groups were exclusive; all showed considerable overlap (Fig. 8A). Sites that were outliers from their groups in 1 analysis were also outliers in the other 2 analyses. The second PCA excluded large river and high elevation sites and was based on the same 32 sites used in the B-IBI analysis. This PCA was even less successful in grouping sites according to a gradient of human disturbance (Fig. 8B).

We suspect that zeros in the data matrix had a strong effect on the PCA solutions for 3 reasons. First, data transformation had little effect

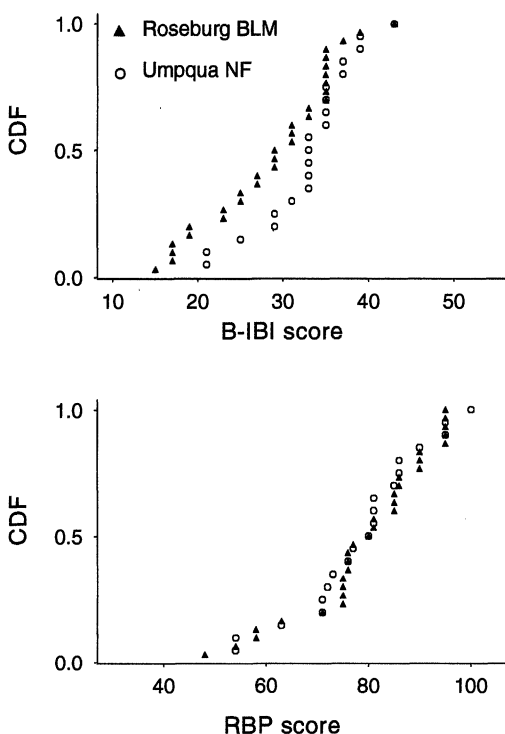


FIG. 7. The cumulative distribution function (CDF) of B-IBI scores (significant difference, 1-sided *t*-test, $p < 0.05$) and RBP scores (not significant) for sites in the Roseburg district ($n = 30$) and the Umpqua district ($n = 20$). The CDF is the percentage of sites with B-IBI scores lower than the B-IBI score on the x-axis. Distance between the 2 CDFs indicates the degree of difference between index scores for the 2 areas.

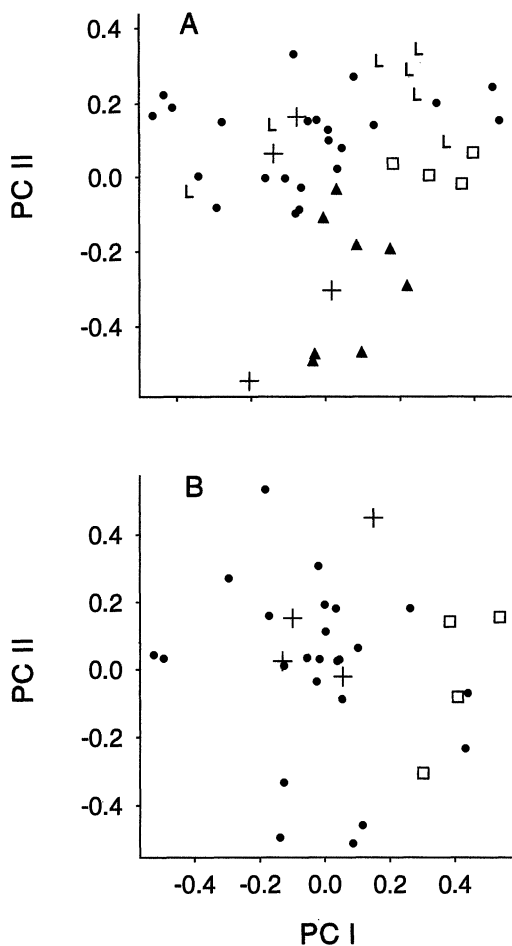


FIG. 8. Sites from Umpqua NF sampled in 1990 plotted according to principal components analysis of the "normalized" data for $n = 46$ sites (A) and $n = 32$ sites (B). Most degraded (□), high elevation (▲), and large river (L) sites showed a tendency to group together (A); least-degraded sites (+) did not. When high elevation and large river sites were excluded from the analysis (B), most and least degraded sites did not group together, nor were moderately degraded sites (●) arrayed between the extremes.

on the PCA outcome, probably because zero abundances were not changed by transformation. Second, disturbed sites grouped together best, probably because, with the fewest taxa, they contained the largest number of zeros in the data matrix. Third, PC I (for the second PCA, $n = 32$) was strongly correlated with taxa richness (Spearman's $\rho = -0.61$, $p < 0.01$).

We looked for patterns in the loadings of the 198 identified taxa on the principal components.

We found a slight tendency for Ephemeroptera, Plecoptera, and Trichoptera taxa to load more frequently on the component that separated most and least disturbed sites (PC I). We did not detect patterns in the loadings for intolerant and tolerant taxa. Other taxa grouped together on principal component axes provided little additional insight into the structure of the assemblage. Rather than trying to work through the statistical analysis and determine how and why species were grouped together, we found it easier to compare sites and search for patterns in the raw data by simply noticing which taxa were present or absent in similar sites. By focusing on biological patterns associated with specific, testable hypotheses, we were better able to rank sites according to environmental condition with B-IBI (Fig. 9A) than we could from PCA plots (Fig. 8, Fig. 9B, 9C).

Discussion

In this paper, we emphasize the process of testing hypotheses about how invertebrates respond to human influences and developing an index that can communicate this response to a broad audience. Even with a rather narrow range of disturbance, we were able to test metrics and develop an index that varied with human-induced disturbance. Component metrics of the index varied as the intensity of logging, road building, and other human activities increased. One may ask, if human influence can be determined so easily, why bother with biological monitoring? Although the condition of sites near the ends of the spectrum are easy to judge, moderately degraded sites are not. Thus we started with the extremes, tested our hypotheses, and developed an index that could rank sites with intermediate levels of human influence. Judging the condition of streams based on the instream biota allows us to identify smaller differences than those a human observer might notice, such as large woody debris, which may be less important to the resident biota than more subtle changes in river ecology.

B-IBI functions as a yardstick for measuring the condition of stream sites. Multimetric indexes formalize what any good biologist, familiar with local biota, knows about the biological condition of a stream. The effects of human-induced degradation on rivers and streams that biologists are being asked to describe are not

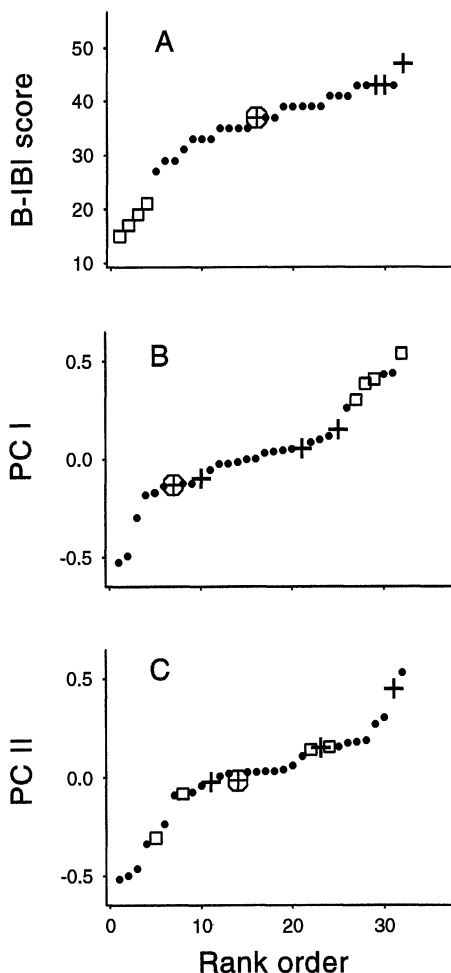


FIG. 9. Sites sampled in Umpqua NF in 1990 ($n = 32$) ranked according to B-IBI score (A), 1st principal component (B), and 2nd principal component (C). B-IBI correctly ranked the most ($\square$) and least (+) degraded sites at the extremes except for Calf Cr. ($\oplus$), the only least-disturbed site with substantial disturbance (bridges and removal of woody debris from the channel) adjacent to the sampling site. PC I (see Figure 8B) ranked the most degraded sites near one extreme but did not separate them from moderately disturbed sites ($\bullet$). Neither PC I nor PC II separated least disturbed sites from moderately disturbed sites; PC II also failed to distinguish most disturbed sites.

subtle. Multiple metrics are critical to a useful index; they are analogous to the measurements (e.g., blood pressure, pulse, and blood chemistry) a doctor might make on a single patient. Multiple measurements increase the probability of an accurate diagnosis. Narrative and numeric

indexes communicate biological condition to policy makers and concerned citizens and thus provide a scientific basis for the management decisions that affect aquatic resources.

Too often, biological patterns in complex data sets are reported in terms of complicated statistics, such as principal components analysis. In contrast, we demonstrated how graphical analysis and simple statistics can be used to interpret complex data and understand biological patterns. To many field biologists, "statistics" means "multivariate statistics." Monitoring reports have been using the same multivariate techniques since the 1960s in spite of a plethora of statistical techniques available (Potvin and Travis 1993).

Our primary concern regarding multivariate analyses is that decisions are often based on statistical properties of the data, e.g., the structure of the covariance matrix, rather than the biological knowledge and understanding of the investigator. PCA uses mathematical algorithms to extract variance from the covariance matrix of species abundances. For our data, PCA failed to clearly distinguish sites with different levels of human influence. As is typical of most multivariate analyses, PCA focuses on numbers of animals present (one of the most variable aspects of biology) rather than biological knowledge of how the animals feed, reproduce, use the habitat, or respond to human activities. In contrast, this biological information is the foundation for multimetric indexes.

Ordination was developed for data that follow a multivariate normal distribution; our data did not. Numerous zeros and frequent high abundances typical of biomonitoring data represent outliers to the analysis and can have a strong influence on the solution (Gauch 1982, Tabachnick and Fidell 1989). Transformation and editing of data are designed to address departures from normality typical of taxa lists and abundances. The recommendation to remove rare taxa from the analysis (typically those present in less than 5% of the samples) is a striking example of statistical requirements eclipsing biological common sense. The presence of rare taxa in the data matrix means lots of zeros which can degrade the statistical solution. The presence of rare taxa in the field indicates pristine conditions capable of supporting these (often) sensitive taxa. Thus rare taxa provide critical clues to biological condition.

Even if PCA, or another type of multivariate analysis, had arrayed sites perfectly according to our a priori classifications, our larger point is that the inherent statistical complexity of multivariate analyses distracts biologists from making clear, testable statements to each other and to nonscientists about how the biota responds to human influence. Multivariate analyses are most appropriate for exploratory analysis, that is, constructing hypotheses about systems for which little is known. We know enough about freshwater systems to test hypotheses rather than continue generating them; we must move forward with testing hypotheses directly rather than simply reporting presumed complexity. More important, we must use monitoring data to make difficult policy decisions; more effort must be dedicated to defining the magnitude of differences and their biological consequences.

Metric definition

Metrics were calculated as taxa richness or percentages depending on the attribute measured. A good biomonitoring index should not be sensitive to slight differences in rainfall or location; therefore we are less interested in which particular species of stonefly is found at a site because the identity of the species may depend on subtle microhabitat differences or broader biogeographic patterns. Similarly, numeric abundance of species responds dramatically to small changes in the environment. On the other hand, taxa richness of stoneflies as a whole, for example, is robust to natural variability. We count how many stonefly taxa are present because most stoneflies require large cobble and fast clean cool water.

Trophic composition metrics (e.g., shredders, scrapers, and predators), tolerance metrics, and dominance (relative abundance of most numerous taxa) metrics are scored as percentages rather than taxa richness because they evaluate processes at the level of the assemblage. Taxa in these groups may be present at all sites, but their relative proportion changes with degradation. We also tested trophic composition metrics as taxa richness metrics when they failed to discriminate sites as percentages; neither approach showed consistent variation in metric values along a gradient of human influence. No doubt human disturbance affects the trophic composition of the resident biotic assemblage.

We suspect that these metrics did not vary systematically because our classification of invertebrates according to feeding ecology is as yet imprecise. In future studies we intend to identify obligate members of each trophic group and enlarge the category of unknowns and omnivores.

Although ratio metrics have been included in other monitoring protocols, e.g., ratio of scrapers to filterers in RBP, we recommend against them for 2 reasons. First, the numerator and denominator vary together; therefore a pair of very large values for scrapers and filterers may yield the same ratio value as a pair of very small values. We expect that large values for these 2 groups may have a different biological meaning than small values. In short, if 2 attributes of an assemblage are potentially important, they should be evaluated individually. Second, when 2 variables are combined as a ratio, the ratio tends to have higher variance than either variable alone (Sokal and Rohlf 1981). Barbour et al. (1992) found ratio metrics for invertebrates to be more variable than other metrics. On the other hand, percentages, or relative abundances, are based on the binomial distribution and are much more reliable.

Our goal is to supplement the general tolerance and intolerance metrics with more specific metrics that describe to what the organisms are tolerant; for example, the metrics for sediment-tolerant and intolerant taxa that were defined for these data. As appropriate data sets become available, we suspect another more inclusive metric will include *Pteronarcys* along with other taxa, such as freshwater mussels, that are sensitive to scour because of large adult size and time needed to complete their life cycle.

"Indexes"

To most biologists, the word "index" carries with it semantic baggage that is the legacy of diversity indexes. Multimetric indexes are often condemned along with diversity indexes (Suter 1993, Norris and Georges 1993) although they differ fundamentally in what they measure and how they are constructed (Fausch et al. 1990). Semantic confusion also surrounds the terms "parameter", "metric", and "index", which are often used interchangeably. In the language of statistics, "parameter" refers to a summary statistic, such as the mean, computed for a partic-

ular variable. Often the word "variable" is more accurate when referring to taxa richness, nitrogen level, or area logged. "Metrics" refer to attributes of the biological assemblage, such as taxa richness, included in a multimetric index. The word "index" has several meanings and a "diversity index" is sometimes included as a component of a multimetric index.

Biomonitoring indexes can be divided into three types: (1) diversity and similarity indexes; (2) pollution-tolerance indexes, or "biotic indexes"; and (3) multimetric indexes. Diversity indexes and most similarity indexes combine information about abundance (or density) and species richness into a measure of "evenness". The concept of evenness was developed in the context of information theory, but has not been meaningfully interpreted in terms of ecological theory (Hurlburt 1971). Diversity indexes often vary with degradation because they are strongly correlated with taxa richness (Camargo 1992), an attribute of the assemblage that, on its own, is easy to interpret in terms of biology. The response of these indexes to systematic changes in the assemblage are often erratic, inconsistent, dependent on initial conditions, and can give misleading interpretations of biological data (Wolda 1981, Boyle et al. 1990).

Pollution-tolerance indexes, or biotic indexes, are based on assignment of a pollution tolerance value to every taxon. These biotic indexes are calculated by multiplying the taxon-specific tolerance value by the number of individuals for that taxon. These values are summed across taxa and divided by the number of individuals in the sample (Chutter 1972, Winget and Mangum 1979, Hilsenhoff 1982, Lenat 1993) to estimate an average pollution tolerance for the assemblage. In western streams as elsewhere, chemical and organic pollution resulting from agriculture and urban sewage represent only part of the problem. Grazing, timber harvest, and road building affect streams and their resident biota differently than organic pollution. A potential problem with this index format is that by defining the tolerance of every organism, a strong response by a few taxa can be missed because most assemblages are dominated by taxa that are neither sensitive nor insensitive; that is, the signal may be lost in the noise. We avoided this problem with tolerance and intolerance metrics by concentrating on the most and least tolerant organisms.

Each component metric of a multimetric index measures an attribute of the assemblage that is the product of evolutionary and biogeographic processes at a site (Karr 1996). Metrics are interpreted (and assigned a score of 5, 3, or 1) based on metric values observed in pristine, or best available, sites. The n metrics represent n measurements of the biotic integrity of a stream. We found for the fish IBI that variance was reduced and precision increased by using an average of several metrics (Fore et al. 1994). Metric values are not independent because they are calculated from a single collection of invertebrates; nonetheless, the functional form of multimetric indexes, i.e., the average (or sum) of metrics, allows us to take advantage of properties of the mean. We found that multimetric index scores approximate a normal distribution (Fore et al. 1994), probably because averages tend to be distributed normally by the Central Limit Theorem (Cassella and Berger 1990); consequently multimetric indexes can be tested with familiar statistics, such as ANOVA or regression.

Multimetric indexes are structured like familiar indexes, such as the consumer price index or the index of leading economic indicators (Mitchell and Burns 1938). These indexes summarize the current condition or state of the economy and reduce the inherent natural variability by including the prices of several items or the current levels of several indicators. Like multimetric indexes, economic indicators are combined into an overall index by comparing them to a standard year; after conversion, index components are summed. These composite indexes have survived 60 years of discussion and criticism and are still widely used by economists, policy makers, and the media to interpret economic trends (Auerbach 1982).

Making decisions

Collecting data for biological monitoring is not a goal in itself; data should be collected to answer specific questions relevant to environmental management (Yoder 1994). Statistical tests are appropriate in some situations, such as, when the goal is to regulate a polluter by assessing sites upstream and downstream of a point source. But when a biologist or statistician reports a significant difference based on a p -value, an alert audience will ask, how different? and, in what way?

In a monitoring context, we are less interested in testing for statistical significance than we are in evaluating differences, especially their biological relevance (Stewart-Oaten et al. 1992, Stewart-Oaten 1996). Multimetric indexes can be used for statistical comparisons, but they also provide a yardstick that can be used to rank sites, for example, to determine which sites are the best candidates for restoration. Index scores for similar sites within a region provide a context for interpreting scores and identifying trends. As data accumulate across a region, we can calculate the statistical power of our yardstick (Peterman 1990, Fore et al. 1994) and report confidence intervals with index scores.

In contrast, multivariate analyses, such as ordination, are not easily adapted to a simple task like ranking streams from severely to minimally degraded. Management decisions can be (and have been) based on multivariate statistical analysis of biological data (Davies et al. 1993, Reynoldson and Zarull 1993, Wright et al. 1993), but the decision process is opaque to anyone who does not understand the mathematical algorithms and the underlying assumptions of the models.

Multimetric indexes condense and summarize information and can be used to compare sites over a large geographic area. On a smaller scale, the component metrics are available to make more site-specific assessments, such as to pinpoint sources of degradation (Yoder and Rankin 1995a) or identify which attributes of the assemblage are affected by human activities. Although a single number, the index, is used to evaluate the relative condition of sites within a region, detailed information about individual sites is not lost, but is retained in the component metrics (Simon and Lyons 1995).

We recognize that biological criteria are urgently needed; but if an index is not composed of biologically meaningful and reliable metrics, the biological criteria will be useless. The RBP index failed to detect differences in southwestern Oregon because the component metrics were never adequately tested. Finally, we urge caution about the concept of "rapid" bioassessment. Our goal is to accurately convey information about biological condition at a site. Speed should only be invoked as a specific criterion *after* a method is shown to be effective.

Fine-tuning B-IBI

Some steps involved in developing a biomonitoring index may need to be reconsidered as data accumulate, but we expect many of the conclusions from our study to hold true across the Pacific Northwest. In particular, the metrics that responded reliably to the effects of logging will likely respond to other types of disturbance such as grazing, agriculture, or urbanization. We expect this to be true because many of the same metrics (8 out of 11) that responded to human influences in southwestern Oregon also responded to agriculture, mining, and damming in the Tennessee Valley (Kerans and Karr 1994). For the fish IBI, the same set of metrics or their functional equivalents responds to many types of human influence across most regions of the USA (Miller et al. 1988); a similar pattern is emerging for the benthic IBI.

The relationship of a metric to human influence—whether it goes up or down—will likely be consistent as well. As data accumulate, samples from severely degraded and relatively pristine streams will probably extend the range of metrics beyond the values reported here. Scoring criteria will shift to reflect this new information. Scoring criteria may also shift as we apply B-IBI to other regions, elevations, and stream sizes. Just as physical sciences recognize the reality that water takes different forms (solid, liquid, gas) as a function of temperature, multimetric indexes reflect the reality that biological systems vary as a consequence of physical and geographic gradients. Scoring criteria may change as we move from west to east across the Cascade Mountains, but our goal is to identify and include metrics in B-IBI that are broadly applicable to a large region and many types of human influence.

A good multimetric index includes a balance of metrics that respond across the range of degradation and to different types of degradation. Because our data sets came from moderately disturbed sites, we were unable to specifically test the sensitivity of metrics under extremely degraded conditions. Percent sediment-tolerant taxa was sensitive at the most degraded end of the spectrum, and taxa richnesses of Plecoptera and percent sediment-intolerant taxa were more sensitive for least disturbed sites. Total taxa richness has a more monotonic response across the range of disturbance. As new data are eval-

uated, we intend to assess the redundancy of metrics and the overall balance of B-IBI (Kerans and Karr 1994).

By carefully selecting component metrics, we can minimize the effects of natural physical variability and weather so that changes in the assemblage caused by human actions are revealed in sharp relief. Complete separation of the effects of natural variability from human-induced variability is not realistic because disturbance exacerbates the effects of natural events (Schlosser 1990); floods are more extreme, for example, in damaged watersheds. Multimetric index scores reflect this reality and consequently are more variable for heavily degraded sites (Karr et al. 1987, Steedman 1988, Fore et al. 1994, Yoder and Rankin 1995b).

Statistical variability of the index determines how precisely it compares sites and guides decisions. State and federal agencies in the US have not yet agreed on the number of individuals to be identified from an invertebrate sample or the number of replicates needed from a stream site. We intend to evaluate the statistical power of B-IBI for different sampling protocols to determine how many classes of integrity B-IBI can distinguish, as was done for the fish IBI (Fore et al. 1994). If we can reduce the amount of effort involved in identifying samples without sacrificing reasonable precision, the money saved for each site assessment translates into more streams visited each year.

Acknowledgements

We thank M. Jones and J. J. Dose (Umpqua Ranger District), L. Lindell (Medford BLM), and E. Rumble and S. Hofford (Roseburg BLM) for providing information on land use. M. Lagerloef (US EPA, Region 10), G. A. Hayslip (US EPA, Region 10), R. Hafele (Oregon Department of Environmental Quality), R. W. Plotnikoff (Washington Department of Ecology), and D. W. Zaroban (Idaho Department of Environmental Quality) participated in discussions that guided the preparation of this paper. D. P. Larsen, R. J. O'Connor, and R. M. Hughes reviewed an early version. The research described here has been funded wholly or in part by the U.S. Environmental Protection Agency under Cooperative Agreement #3B1053NAEX to JRK.

Literature Cited

- ALLAN, J. D., AND A. S. FLECKER. 1993. Biodiversity conservation in running waters. *BioScience* 43:32–43.
- AUERBACH, A. J. 1982. The index of leading indicators: "Measurement without theory", thirty-five years later. *Review of Economics and Statistics* 64: 589–595.
- BAHLS, L. L. 1993. Periphyton bioassessment methods for Montana streams. Water Quality Bureau, Department of Health and Environmental Sciences, Helena, Montana.
- BARBOUR, M. T., J. L. PLAFKIN, B. P. BRADLEY, C. G. GRAVES, AND R. W. WISSEMAN. 1992. Evaluation of EPA's rapid bioassessment benthic metrics: metric redundancy and variability among reference stream sites. *Environmental Toxicology and Chemistry* 11:437–449.
- BARBOUR, M. T., J. B. STRIBLING, AND J. R. KARR. 1995. Multimetric approach for establishing biocriteria and measuring biological condition. Pages 63–77 in W. S. Davis and T. P. Simon (editors). *Biological assessment and criteria: tools for water resource planning and decision making*. Lewis Publishing, Boca Raton, Florida.
- BOYLE, T. P., G. M. SMILLIE, J. C. ANDERSON, AND D. R. BEESON. 1990. A sensitivity analysis of nine diversity and seven similarity indices. *Research Journal of the Water Pollution Control Federation* 62:749–762.
- CAMARGO, J. A. 1992. New diversity index for assessing structural alterations in aquatic communities. *Bulletin of Environmental and Contaminant Toxicology* 48:428–434.
- CASELLA, G., AND R. L. BERGER. 1990. *Statistical inference*. Wadsworth, Inc., Belmont, California.
- CHUTTER, F. M. 1972. An empirical biotic index of the quality of water in South African streams and rivers. *Water Research* 6:19–30.
- CUMMINS, K. W., M. A. WILZBACH, D. M. GATES, J. B. PERRY, AND W. B. TALIAFERRO. 1989. Shredders and riparian vegetation. *BioScience* 39:24–30.
- DAVIES, S. P., L. TSOMIDES, D. L. COURTEMANCH, AND F. DRUMMOND. 1993. Maine biological monitoring and biocriteria development program. Maine Department of Environmental Protection, Bureau of Water Quality Control, Division of Environmental Evaluation and Lake Studies, Augusta, Maine.
- DAVIS, W. S., AND T. P. SIMON (EDITORS). 1995. *Biological assessment and criteria: tools for water resource planning and decision making*. Lewis Publishing, Boca Raton, Florida.
- DESHON, J. E. 1995. Development and application of the invertebrate community index (ICI). Pages 217–244 in W. S. Davis and T. P. Simon (editors). *Biological assessment and criteria: tools for water*

- resource planning and decision making. Lewis Publishing, Boca Raton, Florida.
- DEEGAN, L. A., J. T. FINN, S. G. AYVASIAN, AND C. RYDER. 1993. Feasibility and Application of the Index of Biotic Integrity to Massachusetts estuaries (EBI). Final Project Report to Massachusetts Executive Office of Environmental Affairs, Department of Environmental Protection, North Grafton, Massachusetts.
- DOSE, J. J., AND B. B. ROPER. 1994. Long-term changes in low-flow channel widths within the South Umpqua Watershed, Oregon. *Water Resources Bulletin* 30:993–1000.
- FAUSCH, K. D., J. R. KARR, AND P. R. YANT. 1984. Regional application of an index of biotic integrity based on stream fish communities. *Transactions of the American Fisheries Society* 113:39–55.
- FAUSCH, K. D., J. LYONS, J. R. KARR, AND P. L. ANGERMEIER. 1990. Fish communities as indicators of environmental degradation. *American Fisheries Society Symposium* 8:123–144.
- FORD, J. 1989. The effects of chemical stress on aquatic species composition and community structure. Pages 99–144 in S. A. Levin, M. A. Harwell, J. R. Kelly, and K. D. Kimball (editors). *Ecotoxicology: problems and approaches*. Springer Verlag, New York.
- FORE, L. S., J. R. KARR, AND L. L. CONQUEST. 1994. Statistical properties of an index of biotic integrity used to evaluate water resources. *Canadian Journal of Fisheries and Aquatic Sciences* 51:1077–1087.
- FRISSELL, C. A. 1993. Topology of extinction and endangerment of native fishes in the Pacific Northwest and California (USA). *Conservation Biology* 7:342–354.
- GAUCH, H. G. 1982. *Multivariate analysis in community ecology*. Cambridge University Press, Cambridge, UK.
- HAYSLIP, G. A. (EDITOR). 1993. EPA Region 10 in-stream biological handbook (for wadable streams in the Pacific Northwest). EPA-910-9-92-013. U.S. Environmental Protection Agency-Region 10, Environmental Services Division, Seattle, Washington.
- HILSENHOFF, W. L. 1982. Using a biotic index to evaluate water quality in streams. Department of Natural Resources, Madison, Wisconsin. Technical Bulletin Number 132.
- HUGHES, R. M., AND J. R. GAMMON. 1987. Longitudinal changes in fish assemblages and water quality in the Willamette River, Oregon. *Transactions of the American Fisheries Society* 116:196–209.
- HUGHES, R. M., AND R. F. NOSS. 1992. Biological diversity and biological integrity: current concerns for lakes and streams. *Fisheries* 17(3):11–19.
- HURLBURT, S. H. 1971. The nonconcept of species diversity: a critique and alternative parameters. *Ecology* 52:577–586.
- JAMES, F. C., AND C. E. MCCULLOCH. 1990. Multivariate analysis in ecology and systematics: panacea or Pandora's box? *Annual review of Ecology and Systematics* 21:129–166.
- KARR, J. R. 1981. Assessment of biotic integrity using fish communities. *Fisheries* 6(6):21–27.
- KARR, J. R. 1995. Clean water is not enough. *Illahee* 11:51–59.
- KARR, J. R. 1996. Ecological integrity and ecological health are not the same. Pages 100–113 in P. Schulze (editor). *Engineering within ecological constraints*. National Academy of Engineering, National Academy Press, Washington, D. C.
- KARR, J. R., K. D. FAUSCH, P. L. ANGERMEIER, P. R. YANT, AND I. J. SCHLOSSER. 1986. Assessment of biological integrity in running water: a method and its rationale. *Illinois Natural History Survey Special Publication*, Number 5, Champaign, Illinois.
- KARR, J. R., AND B. L. KERANS. 1991. Components of biological integrity: their definition and use in development of an invertebrate IBI. Pages 1–16 in W. S. Davis and T. P. Simon (editors). *Proceedings of the 1991 Midwest pollution control biologists meeting*. EPA 905-R-92-003. U.S. Environmental Protection Agency, Environmental Sciences Division, Chicago, Illinois.
- KARR, J. R., L. A. TOTH, AND D. R. DUDLEY. 1985. Fish communities of Midwestern rivers: a history of degradation. *BioScience* 35:90–95.
- KARR, J. R., P. R. YANT, AND K. D. FAUSCH. 1987. Spatial and temporal variability of the index of biotic integrity in three midwestern streams. *Transactions of the American Fisheries Society* 116:1–11.
- KENKEL, N. C., AND L. ORLÓCI. 1986. Applying metric and nonmetric multidimensional scaling to ecological studies: some new results. *Ecology* 67:919–928.
- 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.
- KERANS, B. L., J. R. KARR, AND S. A. AHLSTEDT. 1992. Aquatic invertebrate assemblages: spatial and temporal differences among sampling protocols. *Journal of the North American Benthological Society* 11:377–390.
- KOLKWITZ, R., AND M. MARSSON. 1909. *Oekologie der tierischen Saprobien. Beiträge zur Lehre von der biologischen Gewässerbeurteilung*. Internationale Revue der gesamten Hydrobiologie und Hydrogeographie 2:126–152. [Translated 1967. *Ecology of animal saprobia*. Pages 485–495 in L. E. Keup, W. M. Ingram, and K. M. Mackenthun (editors). *Biology of water pollution*. Federal Water Pollution Control Administration, Washington, DC.]

- LENAT, D. R. 1993. A biotic index for the southeastern U.S.: derivation and list of tolerance values, with criteria for assigning water-quality ratings. *Journal of the North American Benthological Society* 12:279–290.
- LUDWIG, J. A., AND J. F. REYNOLDS. 1988. *Statistical ecology*. John Wiley and Sons, New York.
- LYONS, J. 1992. Using the index of biotic integrity (IBI) to measure environmental quality in warmwater streams of Wisconsin. US Forest Service, United States Dept. of Agriculture, St. Paul, MN. General Technical Report NC-149.
- LYONS, J., S. NAVARRO-PEREZ, P. A. COCHRAN, E. SANTANA-C., AND M. GUZMAN-ARROYO. 1995. Index of biotic integrity based on fish assemblages for the conservation of streams and rivers in west-central Mexico. *Conservation Biology* 9:569–584.
- MANLY, B. F. J. 1994. *Multivariate statistical methods: a primer*. 2nd edition. Chapman and Hall, London, UK.
- MERRITT, R. W., AND K. W. CUMMINS (EDITORS). 1984. *An introduction to the aquatic insects of North America*. 2nd edition. Kendall/Hunt Publishing Company, Dubuque, Iowa.
- MILLER, K. L., AND 13 OTHER AUTHORS. 1988. Regional applications of an index of biotic integrity for use in water resource management. *Fisheries* 13(5):12–20.
- MINNS, C. K., V. W. CAIRNS, R. G. RANDALL, AND J. E. MOORE. 1994. An index of biotic integrity (IBI) for fish assemblages in the littoral zone of Great Lakes' areas of concern. *Canadian Journal of Fisheries and Aquatic Sciences* 51:1804–1822.
- MITCHELL, W. C., AND A. F. BURNS. 1938. *Statistical indicators of cyclical revivals*. National Bureau of Economic Research, New York.
- MOYLE, P. B., AND J. E. WILLIAMS. 1990. Biodiversity loss in the temperate zone: decline of the native fish fauna of California. *Conservation Biology* 4: 275–284.
- MULVEY, M., L. CATON, AND R. HAFELE. 1992. Oregon nonpoint-source monitoring protocols and stream bioassessment field manual for macroinvertebrates and habitat assessment. Oregon Department of Environmental Quality Laboratories, Portland, Oregon.
- NEHLSSEN, W., J. E. WILLIAMS, AND J. A. LICHATOWICH. 1991. Pacific salmon at the crossroads: stocks at risk from California, Oregon, Idaho, and Washington. *Fisheries* 16(2):2–21.
- NORRIS, R. H., AND A. GEORGES. 1993. Analysis and interpretation of benthic macroinvertebrate surveys. Pages 234–285 in D. M. Rosenberg and V. H. Resh (editors). *Freshwater biomonitoring and benthic macroinvertebrates*. Chapman and Hall, New York.
- OHIO EPA (ENVIRONMENTAL PROTECTION AGENCY). 1988. *Biological criteria for the protection of aquatic life*, volumes 1–3. Ecological Assessment Section, Division of Water Quality Monitoring and Assessment, Ohio EPA, Columbus, Ohio.
- PACIFIC RIVERS COUNCIL. 1995. *A call for a comprehensive watershed and wild fish conservation program in eastern Oregon and Washington*. 2nd edition. Pacific Rivers Council, Eugene, Oregon.
- PETERMAN, R. M. 1990. Statistical power analysis can improve fisheries research and management. *Canadian Journal of Fisheries and Aquatic Sciences* 47:2–15.
- PLAFKIN, J. L., M. T. BARBOUR, K. D. PORTER, S. K. GROSS, AND R. M. HUGHES. 1989. *Rapid bioassessment protocols for use in streams and rivers: benthic macroinvertebrates and fish*. Assessment and Water Protection Division, U.S. Environmental Protection Agency, Report EPA/440/4-89-001. Washington, D.C.
- POTVIN, C., AND J. TRAVIS (EDITORS). 1993. Special feature. *Statistical methods: an upgrade for biologists*. *Ecology* 74:1614–1676.
- REYNOLDSON, T. B., AND M. A. ZARULL. 1993. An approach to the development of biological sediment guidelines. Pages 177–200 in S. Woodley, J. Kay, and G. Francis (editors). *Ecological integrity and the management of ecosystems*. St. Lucie Press, Delray Beach, Florida.
- SCHELSKE, C. L. 1984. *In situ* and natural phytoplankton assemblage bioassays. Pages 15–47 in L. E. Shubert (editor). *Algae as ecological indicators*. Academic Press, London.
- SCHLOSSER, I. J. 1990. Environmental variation, life history attributes, and community structure in stream fishes: implications for environmental management and assessment. *Environmental Management* 14:621–628.
- SIMON, T. P., AND J. LYONS. 1995. Application of the index of biotic integrity to evaluate water resource integrity in freshwater ecosystems. Pages 245–262 in W. S. Davis and T. P. Simon (editors). *Biological assessment and criteria: tools for water resource planning and decision making*. Lewis Publishing, Boca Raton, Florida.
- SOKAL, R. R., AND F. J. ROHLF. 1981. *Biometry*. 2nd edition. Freeman and Company, New York.
- STEEDMAN, R. J. 1988. Modification and assessment of an index of biotic integrity to quantify stream quality in southern Ontario. *Canadian Journal of Fisheries and Aquatic Sciences* 45:492–501.
- STEWART-OATEN, A. 1996. Goals in environmental monitoring. Pages 17–28 in R. J. Schmitt and C. W. Osenberg (editors). *Detecting ecological impacts: concepts and applications in coastal habitats*. Academic Press, San Diego, California.
- STEWART-OATEN, A., J. R. BENICE, AND C. W. OSENBURG. 1992. Assessing effects of unreplicated perturbations: no simple solutions. *Ecology* 73:1396–1404.

- SUTER, G. W. 1993. A critique of ecosystem health concepts and indexes. *Environmental Toxicology and Chemistry* 12:1533–1539.
- TABACHNICK, B. G., AND L. S. FIDELL. 1989. Using multivariate statistics. 2nd edition. HarperCollins Publishers, Inc., New York.
- USFS (US FOREST SERVICE). 1992. Wild and scenic river eligibility and determination. Umpqua National Forest, Roseburg, Oregon.
- WILDERNESS SOCIETY. 1993. The living landscape. Volume 1. Wild salmon as natural capital; volume 2. Pacific salmon and federal lands. The Wilderness Society, Washington, D.C.
- WINGET, R. N., AND F. A. MANGUM. 1979. Biotic condition index: integrated biological, physical and chemical stream parameters for management. US Forest Service, Intermountain Region, Provo, Utah.
- WOLDA, H. 1981. Similarity indices, sample size, and diversity. *Oecologia* 50:296–302.
- WRIGHT, J. F., M. T. FURSE, AND P. D. ARMITAGE. 1993. RIVPACS—a technique for evaluating the biological quality of rivers in the U. K. *European Water Pollution Control* 3:15–25.
- YOCOZ, N. G. 1991. Use, overuse, and misuse of significance tests in evolutionary biology and ecology. *Bulletin of the Ecological Society of America* 71:106–111.
- YODER, C. O. 1994. Integrating basic and applied benthic science. *Journal of the North American Benthological Society* 13:391–398.
- YODER, C. O., AND E. T. RANKIN. 1995a. Biological response signatures and the area of degradation value: new tools for interpreting multimetric data. Pages 263–286 in W. S. Davis and T. P. Simon (editors). *Biological assessment and criteria: tools for water resource planning and decision making*. Lewis Publishing, Boca Raton, Florida.
- YODER, C. O., AND E. T. RANKIN. 1995b. Biological criteria program development and implementation in Ohio. Pages 109–144 in W. S. Davis and T. P. Simon (editors). *Biological assessment and criteria: tools for water resource planning and decision making*. Lewis Publishing, Boca Raton, Florida.

Received: 27 July 1995

Accepted: 7 March 1996