Development and Evaluation of Predictive Models for Measuring the Biological Integrity of Streams

Charles P. Hawkins; Richard H. Norris; James N. Hogue; Jack W. Feminella

DEVELOPMENT AND EVALUATION OF PREDICTIVE MODELS FOR MEASURING THE BIOLOGICAL INTEGRITY OF STREAMS

CHARLES P. HAWKINS,[1,3] RICHARD H. NORRIS,[2] JAMES N. HOGUE,[1,4] AND JACK W. FEMINELLA[1,5]

[1]*Department of Fisheries and Wildlife, Watershed Science Unit, and Ecology Center, Utah State University, Logan, Utah 84322-5210 USA*
[2]*Cooperative Research Centre for Freshwater Ecology, University of Canberra, ACT, 2616 Australia*

Abstract. The ratio of the number of observed taxa to that expected to occur in the absence of human-caused stress (*O/E*) is an intuitive and ecologically meaningful measure of biological integrity. We examined how *O/E* ratios derived from stream invertebrate data varied among 234 unimpaired reference sites and 254 test sites potentially impaired by past logging. Data were collected from streams in three montane ecoregions in California. Two sets of River Invertebrate Prediction and Classification System (RIVPACS) predictive models were built: one set of models was based on near-species taxonomic resolution; the other was based on family identifications. Two models were built for each level of taxonomic resolution: one calculated *O* and *E* based on all taxa with probabilities of capture (P_c) > 0; the other calculated *O* and *E* based on only those taxa with $P_c \geq 0.5$. Evaluations of the performance of each model were based on three criteria: (1) how well models predicted the taxa found at unimpaired sites, (2) the degree to which *O/E* values differed among unimpaired reference sites and potentially impaired test sites, and (3) the degree to which test site *O/E* values were correlated with independent measures of watershed alteration. Predictions of species models were more accurate than those of family models, and predictions of the $P_c \geq 0.5$ species model were more robust than predictions of the $P_c \geq 0$ model. *O/E* values derived from both species models were related to land use variables, but only assessments based on the $P_c \geq 0.5$ model were insensitive to naturally occurring differences among streams, ecoregions, and years.

Key words: bioassessment; biological integrity; California; invertebrates; logging; monitoring; predictive models; RIVPACS; species-based vs. family-based models; streams; water quality.

INTRODUCTION

During the last 20 yr, water resource scientists have developed a variety of biologically based approaches to assess the ecological effects of pollution and landscape alteration (Ohio Environmental Protection Agency [EPA] 1987, Reynoldson and Metcalfe-Smith 1992, Lenat 1993, Rosenberg and Resh 1993, Wright et al. 1993, Norris and Norris 1995, Southerland and Stribling 1995). Two major approaches have emerged from this activity: a multimetric approach based on the ideas of Karr (1981) that uses the values of several indices to measure biotic condition, and assessment schemes that compare observed faunas to those predicted by empirical models to occur in the absence of human alteration (Wright et al. 1993, Wright 1995, Norris 1996, Moss et al. 1999). Although considerable effort has been directed toward developing the multimetric approach in the United States (e.g., Ohio EPA 1987,

Plafkin et al. 1989, Kerans and Karr 1994, Barbour et al. 1995, 1996, 1997, DeShon 1995, Fore et al. 1996, Gibson et al. 1996, Karr and Chu 1999), little effort has been directed at either developing or evaluating the utility of predictive models in the United States.

Wright et al. (1993) and Norris (1996) describe the British and Australian versions of these predictive models. These models are conceptually and mechanistically similar, and they differ only in two subtle aspects, which we describe. Because of their similarities, we refer to them collectively as RIVPACS models, an acronym Wright (1995) and coworkers derived from the name of the original assessment scheme (River Invertebrate Prediction and Classification System). The analyses on which this paper is based were performed with a modified version of this method developed in Australia (Australian River Assessment System, or AUSRIVAS).

River Invertebrate Prediction and Classification System models allow an assessment of biological condition by comparing the taxa observed at sites of unknown or suspect biological condition (hereafter called "test sites") with the biota expected to occur in the absence of stress (see Wright [1995] for details). In brief, a predictive model is built from biological and environmental data collected from reference sites that are min-

imally affected by human activities. For practical reasons, reference sites are seldom pristine, but should represent the least impaired sites within the area of interest. Measurements at these reference sites provide an empirical foundation from which to judge the relative condition of other sites (Bailey et al. 1998, Reynoldson and Wright 2000), and the model predicts the biotic composition expected at a site from a suite of environmental features unlikely to be influenced by human activity (e.g., latitude, elevation, channel slope, and alkalinity).

In RIVPACS assessments, the ratio of the number of taxa observed at a test site that were expected to occur to the number of taxa expected (O/E ratio) is used as a measure of biological impairment. Low observed-to-expected ratios ($O/E \ll 1.0$) imply that test sites are adversely affected by some environmental stressor. The number of taxa expected at a site is calculated as the sum of individual probabilities of capture (P_c) of all taxa found in the region of interest. In the British models, all values of $P_c > 0$ are summed to calculate the expected number of taxa (E), and this number is compared to the total taxa observed (O) at a site. In the Australian models, only taxa with $P_c \geq 0.5$ are used to calculate E, and this number is then compared to the number of taxa with $P_c \geq 0.5$ that were collected.

Because these models predict the actual taxonomic composition of a site, they also provide information about the presence or absence of specific taxa. If the sensitivities of taxa to different stressors are known, this information can lead to derived indices and diagnoses of the stressors most likely affecting a site.

The primary goal of this paper is to describe the potential utility of RIVPACS assessments for measuring the biological condition of streams draining landscapes that have been altered by logging activities. Comparing the sensitivity of these assessments to timber harvest practices should represent a particularly rigorous test of their utility, because the composition of stream biota can vary substantially with naturally occurring local and regional conditions in the western United States (Hawkins et al. 1997), and biotic responses to logging practices are often subtle, complex, and difficult to distinguish from natural background variation (e.g., Newbold et al. 1980, Hawkins et al. 1982, 1983).

We address two main objectives: (1) Do assessments derived from species data differ from those derived from data based on family identifications? (2) Do assessments based on $P_c > 0$ differ from those based on $P_c \geq 0.5$. In addressing these objectives and evaluating model performance, we focused on seven questions: (1) was it possible to identify classes of sites that described most of the natural biotic variability among sites, (2) did the models correctly predict the condition of known reference sites, (3) did assessment scores of test sites vary in understandable ways with independent measurements of catchment condition, (4) was there any difference in assessments of the same location made in different years, (5) did the ecoregion from which samples were collected influence assessment scores, (6) did the type of habitat sampled influence assessment scores, and (7) were there specific taxa indicative of logging effects?

Methods

Study area

The study area included three montane ecoregions in California (Omernik 1987, Barbour et al. 1997) that differ in forest type and cover, geology, and precipitation (Hawkins et al. 1994, 1997). One ecoregion (Klamath Mountains; KM) was the mixed conifer forest area of the interior portion of northwestern California, including portions of the Klamath Mountains and the Shasta-Trinity Mountains. The second ecoregion was the drier Sierra Nevada (SN), consisting predominantly of pine and fir forests. The third ecoregion, Eastern Cascades Slopes and Foothills (ECSF), is drier than the others and sparsely forested because it lies within a rain shadow. Climate in all three ecoregions is warm in the summer with rain and snow at higher elevations in winter. Runoff in KM streams is heavily influenced by rainfall, whereas runoff in the SN and ECSF ecoregions is largely related to snowmelt. Study sites spanned ~6° latitude (36°–42° N) and >2200 m in elevation (330–2547 m). Bedrock geology underlying basins in these three ecoregions varies considerably and includes granitic, volcanic, and sedimentary rock (Mount 1995). The dominant land use in these basins is silviculture.

Sampling design and site selection

The sampling program was designed to assess the degree to which landscape alteration associated with timber harvesting has affected stream biota. Because lack of statistical rigor and small sample size has compromised many past studies of the effect of management on stream biota (Hall and Knight 1981, Hawkins et al. 1983), we designed a geographically extensive sampling program to quantify the physical, chemical, and biological heterogeneity that existed both among and within basins.

In 1988, we requested that 13 National Forests in California (Region 5 of the U.S. Forest Service) submit candidate basins for this study. We also requested that background information on each basin be provided to facilitate site selection. Our primary criterion for the initial submission of candidate basins was that they be 3000–6000 ha in size. We also asked that basins with a range of land use be submitted, which we initially characterized as having low, moderate, and high amounts of silviculture.

Of the 167 basins submitted in 1988, we selected 45 basins for study (Fig. 1), 15 in each of the land use
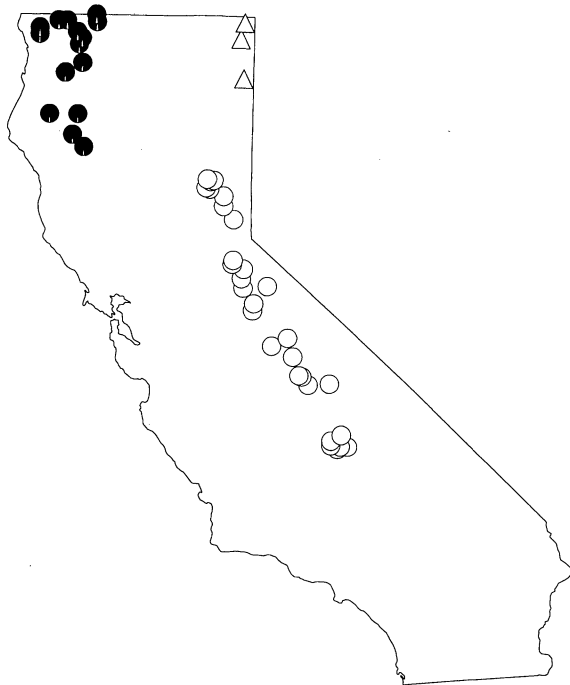
FIG. 1. Locations of 45 basins in California. Shaded circles are basins in the Klamath Mountain ecoregion, open circles are basins in the Sierra Nevada ecoregion, and triangles are basins in the East Cascade Slopes and Foothills ecoregion.

categories. These basins met the following sampling design criteria: basin size ~4000–5000 ha (third- and fourth-order streams; Strahler 1964), triplets of basins in the low, moderate, and high land use categories occurring within close proximity of each other (within ~50 km) and at similar elevation, and the number of basins selected within the three 3 ecoregions occurring in approximately the same proportion as their respective areas within California (70% SN and 30% KM, with ECSF sites combined with SN sites). Twenty-four basins were sampled during summer 1988; the remaining 21 basins were sampled during summer 1989. To examine how much benthic faunas differed between years, we sampled two streams (one low and one high land use) in each ecoregion in both years, thereby providing 49 total sets of samples.

In each year, streams were sampled from south to north, in order to minimize differences in assemblage structure associated with sampling date. This sampling design was based on the assumption that initiation of life history phenomena in northern populations would be delayed relative to more southerly populations because of a latitudinal gradient in the seasonal warming of streams (see Newbold et al. 1994). Sampling began in mid-June in the southern Sierra Nevada and ended in mid-September near the California–Oregon border.

Data collection

We collected four types of data in each basin: benthic invertebrates, channel conditions, chemical water quality, and hill slope conditions. Several fast-water habitats were sampled at ~600-m intervals along 4–11 km of the main stream channel in each basin. Fast-water habitat types included runs, step runs, low-gradient riffles, high-gradient riffles, and cascades (see Bisson et al. [1981] and Hawkins et al. [1993] for descriptions of habitat types). These habitat types represent a continuum of "riffle" type habitats from low-gradient, low-turbulence units (e.g., runs) to high-gradient, high-turbulence units (e.g., cascades). In this paper, we refer to this combination of fast-water habitats collectively as "riffles."

Benthic invertebrates, macroalgae, and moss.—Nine 0.1-m² benthic samples (modified Surber sampler with 250-μm mesh) were collected per riffle and combined in the field to provide a single sample from each riffle. In conjunction with the benthic invertebrate samples, we used a 100-cell, 30 × 30 cm transparent grid to estimate the percentage of each sample area covered by visible growths of algae and moss.

Invertebrate samples were sorted in the laboratory and individuals identified to the lowest possible taxonomic level (usually genus or species; hereafter called operational taxonomic units; OTU).

Riffle habitat conditions.—The following environmental data were collected at the time of sampling from each riffle: riffle slope (%), mean depth (m, calculated from 20 measurements taken along five to seven transects); mean width (m, calculated from five to seven measurements); mean length (m, calculated from three to five measurements); pool volume (m³, calculated from mean depth, mean width, and mean length); median substrate size (D50, calculated from measurements of 50–100 individual particles collected at 0.3-m intervals along systematically spaced transects and reported as a phi (log₂) scaled index, where 0–1 mm = 1, 1–2 mm = 2, 2–4 mm = 3, etc.; see Wolman 1954); percent embeddedness of cobble substrates estimated as the relative depth (0–20%, 20–40%, 40–60%, 60–80%, and 80–100%) that bed particles ≥64 mm were buried by fine material; percent shade over the riffle estimated with a spherical densiometer and calculated as the percentage of the riffle shaded by riparian vegetation; and percentage of the near-stream floodplain occupied by herbaceous, deciduous, and coniferous vegetation.

Stream temperatures.—In addition to these sample site measurements, we also recorded water temperature at hourly intervals while in the stream, and used the >60 hourly measurements taken over the eight days required to sample a stream to calculate mean daytime temperature at the time of sampling (see Hawkins et al. 1997).

Near-stream erosion.—Between each sampling lo-

cation, we measured potentially active areas of erosion on the stream bank and nearby off-channel hillside (<50 m from channel). Areas of potentially active erosion were defined as regions >1 m² with exposed soil. Values of both stream bank and hillslope erosion were summed between each sampling location and for the entire sampled channel upstream of a sampling location. These values were then summed to give an index of total erosion occurring immediately above a site. A standardized estimate of the cumulative amount of eroding area upstream of each site was estimated by summing all eroding areas observed upstream of each site and then dividing by the length of channel surveyed upstream of each site. Both site and cumulative values were expressed as number of square meters of eroding area per 100 m of stream channel.

Water chemistry.—We collected water chemistry samples at two locations on each stream. One location was the basin outlet (first sampling site), and the other site was located approximately midway between the first and last sampling sites. At each location, water chemistry (Na, K, Mg, Ca, NO_3, PO_4), specific conductance, and pH were sampled at 3-h intervals over a 24-h period. Because of funding constraints, water chemistry samples were collected from the 25 SN and 3 ECSF basins in the summer and fall of 1990 and from the 17 KM basins in the summer of 1991. For these analyses, we assumed that conditions at the time of chemistry sampling were similar to those when invertebrate samples were collected, an assumption we believe is reasonable because climatic conditions during 1988–1991 were generally similar because of a modest 7-yr drought. We used 24-h means in all statistical analyses.

Basin conditions.—Information describing watershed boundaries, stream networks, resource type and abundance data (e.g., soils, vegetation), management history data (e.g., roads, timber harvest, fire history), and stream sampling locations were entered into a geographical information system (GIS). All data were compiled at 1:24 000 scale. The GIS allowed us to calculate basin conditions upstream of every sampling point. For this paper, we used the following derived basin scale variables: basin area, drainage density, elevation, latitude, longitude, length of permanent stream channel above the sampling location, area of basin burned above the sampling site, and azimuth of the stream channel. Azimuth was measured as the difference in decimal degrees between the orientation of the stream channel and due south, where due south was given by 0°. Direction of flow was ignored in order that orientation of all channels could be scaled from −90° to +90° (Bartholow 1989). In our statistical analyses, we used the absolute value of azimuth, because streams with orientations of equal magnitude but different sign would generally receive the same amount of solar radiation.

Identification of reference and test sites

River Invertebrate Prediction and Classification System (RIVPACS) assessments require comparison of biota collected at potentially disturbed test sites with biota collected from environmentally similar reference sites (Norris 1994, Reynoldson et al. 1997, Bailey et al. 1998). We used sites with minimal upstream logging activity as reference sites and considered all other sites as potentially degraded. Because no data exist describing how stream invertebrate assemblages change with the amount of logging among these basins, we used an arbitrary, but conservative, criterion for defining reference sites. Reference sites were defined as sites in which timber had been removed from <5% of the basin up-gradient from the site, as measured from maps dating to ~1960. The five-percent criterion was chosen after inspecting how percent upstream logging was distributed among sites, and represents a compromise that balanced the need for minimum possible upstream human disturbance with a sample size sufficiently large to allow adequate statistical analysis.

During the field surveys, 668 separate riffles were sampled for benthic invertebrates. However, we were unable to obtain some basin-level information for some sites, and data were missing for one or more environmental variables from other sites. We also discovered that invertebrate assemblages collected from riffles dominated by large boulders or bedrock were usually statistical outliers in the analyses we conducted. We therefore eliminated all sites with missing data and those dominated by bedrock and large boulder substrate from subsequent analyses. Removal of these sites provided initial sample sizes of 261 reference and 281 test sites, although we excluded several more test sites because they did not belong to the same statistical population as the reference sites (see *Models: Model performance: Check on applicability*).

Model construction and calculation of O/E

The statistical procedures used in constructing RIVPACS models have been described in detail elsewhere (Wright et al. 1984, Moss et al. 1987, Wright 1995, Clarke et al. 1996, Norris 1996, Parsons and Norris 1996, Marchant et al. 1997), and only a brief description is given here. To construct and implement models, we used the following procedures:

1) A biological classification of reference sites was derived from cluster analysis (unweighted pair–group arithmetic averaging with β = −0.1; Belbin and McDonald 1993) following calculation of Bray–Curtis dissimilarities (Faith et al. 1987). For the site classification, but not subsequent predictions of biotic composition, we deleted all taxa that occurred at >95% and <5% of the sites, because taxa that occur everywhere will not discriminate groups and very rare taxa have a high probability of being accidentals and thus

could obscure our ability to detect biologically significant differences between sites.

2) Stepwise multiple discriminant analysis (MDA) was used to create discriminant functions models (DFM) that estimate the probabilities of group membership of test sites from environmental data.

3) Probabilities of group membership were used in conjunction with frequencies of occurrence of specific taxa within the different reference site groups to estimate taxon-specific probabilities of capture (P_c) for individual sites.

4) Probabilities of capture (P_c) were used to identify the specific taxa predicted to occur in samples for both $P_c \geq 0$ and $P_c \geq 0.5$ and estimate the number of taxa expected to occur within the sample for both P_c thresholds. These data were then used to calculate O/E values for both P_c thresholds, i.e., O/E_0 and O/E_{50}.

5) Model error was estimated by comparing the number of taxa observed at reference sites to that expected from the model and generating a distribution of reference site O/E ratios. This population of O/E ratios describes the distribution of errors in predicting taxa richness and is used to determine if the O/E ratio of a test site is different than that expected from model error alone.

6) Environmental data from test sites were then input to the models, and O/E ratios were calculated as described.

Model performance

Training and evaluation data.—To determine the accuracy and precision of each model, we used data from 80% of the reference sites to build the models (training sites), and we used data from the remaining 20% of reference sites (evaluation sites) to evaluate the ability of the models to correctly assess sites of known condition. To ensure that evaluation data represented the range of biotic conditions found among sites, and that groups were proportionately represented in these analyses, we selected every fifth site as arrayed in the cluster tree as an evaluation site.

From these analyses, we examined how well the discriminant models predicted group membership and if O/E scores for the evaluation sites differed from those predicted by the models. We used two cross-validation procedures to evaluate the DFMs. First, we used the jackknifing procedures within the SYSTAT (Version 8) discriminant functions routine to predict group membership of each site in the training data from models constructed after removing them one at a time from the data set. We then compared the jackknifed results with how well the model predicted group membership of the evaluation sites.

Effects of taxonomic resolution on assessments.—To explore how the level of taxonomic resolution affected sensitivity of assessments, we built models from two data sets that were identical, except in the level of taxonomic resolution used in classifying individuals.

The first model was built with data in which taxa were identified to the lowest level possible, generally genus or species, and the second with identifications only to family.

Check on applicability of models to test sites.—Before calculating O/E values for test sites, we determined if the environmental attributes of each test site belonged to the same statistical population as the reference sites (i.e., was the test site within the experience of the models). This test followed the procedures described by Moss et al. (1987) and Clarke et al. (1996), in which the Mahalanobis squared distances between a test site and each classification group in multivariate discriminant space are calculated. These distances are distributed as a χ^2 function, and a new site is inferred as being outside the range of the model if the smaller of these values exceeds a predefined critical χ^2 value (here $\alpha = 0.05$) in which the number of degrees of freedom equals the number of discriminant functions. We eliminated any test site that failed this test from further consideration.

O/E values and environmental conditions.—We conducted two types of analyses to determine if O/E values varied with both naturally occurring environmental conditions and channel and basin conditions likely caused by land management. ANOVA was used to determine if test site O/E values differed from those of reference sites, and if O/E values differed between site classes, year of sampling, type of fast water habitat sampled, and ecoregion. To determine if test site O/E values varied with site class, we assigned each site to its most probable class as identified by the DFM. To determine if O/E values varied with factors associated with land use above each test site, we first conducted a principal components analysis on management-related variables to describe statistically independent axes of environmental variation. We then regressed O/E values on the three principal components axes that described most of the variation in management variables.

Taxa missing from and present at test sites.—To identify those taxa that appeared to be most sensitive to logging practices, we compared how densities of each taxon varied, on average, between reference and test sites. Differences in densities were calculated in two ways: (1) as the logarithm (base 10) of the difference between the two means with positive values assigned to taxa with highest densities in reference sites, and negative values assigned to taxa with highest mean densities in test sites; and (2) as the proportional difference in the means calculated as the logarithm of reference site density divided by test site density. These values were then plotted against the rank abundance of taxa found at the reference sites. The first plot revealed which taxa differed most strongly in absolute densities between reference and test sites, whereas the second plot revealed which taxa differed the most, relative to their densities in reference sites.

Differences between reference and test sites in dom-

TABLE 1. Range of channel and basin conditions observed among the reference and test sites.

Variable	Reference sites			Test sites		
	Minimum	Mean	Maximum	Minimum	Mean	Maximum
Sampling variables						
Sampling date (Jan 1 = 0)	158	211	256	161	211	255
Basin variables						
Azimuth (°)	2	44	84	0	40	86
Elevation (m)	330	1484	2547	410	1346	2353
Basin area above site (m² × 10⁷)	0.3	3.1	10.0	0.1	3.1	6.2
Length of stream above site (km)	4.3	36.7	95.5	2.0	37.4	70.9
Drainage density (m/ha)	6.1	12.1	20.6	7.4	12.4	18.2
Calcareous geology (ha)	0	142	2424	0	265	17449
Area burned (ha)	0	227	5941	0	299	25041
Total area cut (ha)	0	55	403	16	792	3584
Percent of basin cut	0	1.5	5	5	26	77
Length of roads (km)	0	29	167	1.3	66	152
Stream-road crossings (no.)	0	18	112	0	32	97
Riparian variables						
Riparian shade (%)	0	45	96	0	58	96
Herbaceous cover (%)	0	35	100	0	33	100
Deciduous cover (%)	0	45	100	0	53	100
Coniferous cover (%)	0	21	98	0	24	100
Channel variables						
Channel gradient (%)	0	2.8	26	0	3.1	26
Riffle depth (m)	0.02	0.13	0.70	0.03	0.13	0.56
Riffle width (m)	0.7	3.5	12.8	0.6	3.3	11.4
Particle size (mm)	0.5	32	1536	0.5	31	1536
Substrate diversity (1/S)	1.0	5.7	10.4	2.1	5.8	11.0
Embeddedness (%)	10	25	64	10	25	80
Moss cover (%)	0	3	68	0	3	56
Macroalgae cover (%)	0	6	62	0	3	96
CV macroalgae cover (%)	0	1.1	3.0	0	0.9	3.0
Large woody debris (no. pieces/100 m)	0	5	49	0	5	42
Site erosion (m²/100 m)	0	218	5216	0	158	1956
Upstream erosion (m²/100 m)	0	203	2034	0	154	988
Water variables						
pH	6.3	7.6	8.3	6.8	7.7	8.1
Conductivity (μS/cm)	34	91	447	29	91	447
Sodium (1 mg/kg H₂O)	0.8	6.1	24.8	0.8	5.0	19.9
Potassium (1 mg/kg H₂O)	0.2	1.3	4.1	0.5	1.5	3.9
Magnesium (1 mg/kg H₂O)	0.4	5.1	15.1	0.4	5.4	15.1
Calcium (1 mg/kg H₂O)	2.7	15.8	48.7	2.2	15.7	48.7
NO₃/NO₂ (1 μg/kg H₂O)	3.2	21.1	74.2	3.3	24.8	76.8
PO₄ (1 μg/kg H₂O)	0.8	3.5	12.7	1.1	4.4	12.7
Temperature at site (°C)	8.4	15.4	19.8	11.6	16.0	20.3

Notes: Site erosion is the area of near-channel banks and hillslopes eroding within 600 m upstream of each site, standardized to a 100 m length of stream channel. Upstream erosion is the cumulative area of near-channel banks and hillslopes eroding upstream of a site, standardized to a 100 m length of stream channel. Substrate diversity is calculated as the inverse of Simpson's diversity index (S).

inance–diversity relationships.—To determine if differences in the performance of O/E_0 and O/E_{50} were associated with either an increase or decrease in dominance by taxa, we compared the dominance–diversity curves derived from the mean abundances observed at reference and test sites.

RESULTS

Invertebrate fauna

Two hundred ninety three operational taxonomic units (OTUs) in 83 families were recognized from the study sites, of which chironomids represented the largest taxonomic group (31% of all OTUs). Only 14 non-insect OTUs were collected, although 11 of these OTUs were only identified to family or higher levels of classification. The number of OTUs collected from individual sites ranged 9–80. Of the 222 OTUs collected from reference sites, 79 OTUs and 27 families occurred at ≤5% of the sites. Six OTUs and six families occurred at ≥95% of sites. Elimination of these taxa resulted in 143 OTUs and 50 families on which subsequent site classifications were based.

Variation in physical and chemical conditions among sites

Both naturally occurring and human-caused environmental conditions varied widely among sites (Table 1).

TABLE 2. Mean values of environmental factors for sites located in the three ecoregions: Sierra Nevada (SN), Klamath Mountains (KM), and Eastern Cascade Slopes and Foothills (ECSF).

Variable	SN (265)	KM (169)	ECSF (30)
Sampling variables			
Sampling date (Jan 1 = 0)	191	240	224
Basin variables			
Azimuth (°)	48	34	32
Elevation (m)	1749	804	1893
Basin area above site (m² × 10⁷)	3.3	2.9	3.1
Length of stream above site (km)	39	36	11.9
Drainage density (m/ha)	12.3	12.9	7.6
Calcareous geology (ha)	8.0	217	0
Area burned (ha)	126	498	329
Total area cut (ha)	537	323	26
Percent of basin cut	17	12	1
Length of roads (km)	51	41	50
Stream-road crossings (no.)	31	17	16
Riparian variables			
Riparian shade (%)	42	73	19
Herbaceous cover (%)	28	37	68
Deciduous cover (%)	41	65	28
Coniferous cover (%)	29	15	9
Channel variables			
Channel gradient (%)	2.6	3.6	2.2
Riffle depth (m)	0.14	0.12	0.12
Riffle width (m)	3.5	3.2	2.9
Particle size (mm)	31	32	27
Substrate diversity (1/S)	5.7	6.0	5.5
Embeddedness (%)	25	26	20
Moss cover (%)	3	4	0
Macroalgae cover (%)	4	5	2
CV macroalgae cover (%)	1.1	0.9	1.0
Large woody debris (no. pieces/100 m)	4.0	8.3	2.5
Site erosion (m²/100 m)	166	.205	120
Upstream erosion (m²/100 m)	146	220	94
Water variables			
pH	7.5	7.9	7.6
Conductivity (μS/cm)	74	145	46
Sodium (1 mg/kg H_2O)	7.1	3.6	3.7
Potassium (1 mg/kg H_2O)	1.4	1.4	1.3
Magnesium (1 mg/kg H_2O)	2.9	9.4	2.5
Calcium (1 mg/kg H_2O)	13.7	21.0	5.8
NO_3/NO_2 (1 μg/kg H_2O)	18.0	33.3	8.6
PO_4 (1 μg/kg H_2O)	3.8	4.2	5.0
Temperature at site (°C)	16.1	14.8	16.6

Notes: Numbers of sites are shown in parentheses. Site erosion is given by area of near-channel banks and hillslopes eroding ≤600 m upstream of each site, standardized to a 100 m length of stream channel. Upstream erosion is given by cumulative area of near-channel banks and hillslopes eroding upstream of a site, standardized to a 100 m length of stream channel. Substrate diversity is calculated as 1/Simpson's diversity index (S).

Few substantial differences existed between reference and test sites in mean values of naturally occurring factors, although the average reference site was slightly higher in elevation than the average test site, and test sites tended to occur in slightly more calcareous regions than references sites. In contrast, the mean values of several variables associated with land use varied markedly between reference and test sites. Area cut, percentage of the basin above each site that had been cut, length of road above each site, number of stream–road crossings, percentage of the channel shaded by riparian vegetation, bankside cover by deciduous vegetation, and potassium and phosphate stream water concentrations were all higher on average for test sites.

Sodium stream water concentrations were lower on average for test sites.

In contrast to the few differences among reference and test sites in naturally occurring factors, pronounced differences existed both within and among ecoregions in elevation, drainage density, calcareous geology, riparian shade, deciduous and coniferous cover surrounding stream banks, substrate particle size, stream water pH, conductivity, and ionic concentration of most cations (Table 2). Klamath sites (KM) were generally more heavily shaded, lowest in elevation, coolest, and generally had the highest amounts of woody debris and stream water concentrations of most ions. Eastern Cascade Slope and Foothill (ECSF) sites had low drainage

FIG. 2. Dendrograms showing species and family site classes. Individual sites are not shown. Linkage distances are those calculated from raw Bray–Curtis pairwise distances by the flexible β unweighted pair group method using arithmetic averages (UPGMA) clustering algorithm of Belbin and McDonald (1993) and do not therefore scale exactly from 0 to 1.
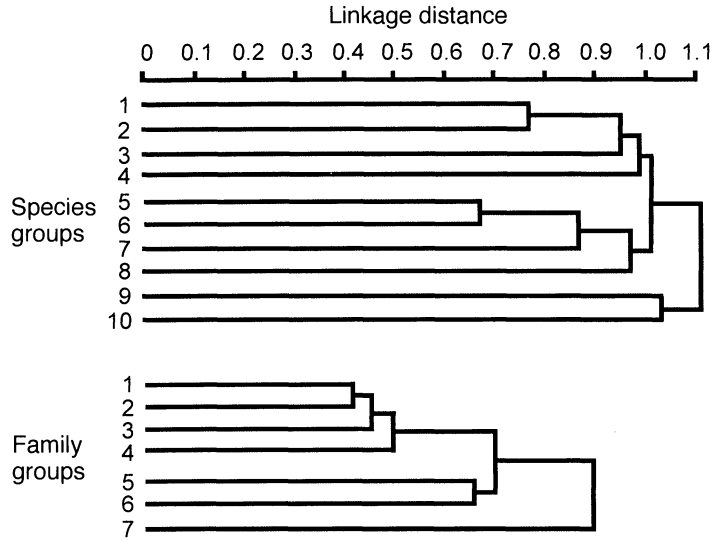
densities and were sparsely shaded. However, these sites also had the most bankside cover of herbaceous plants, had the lowest specific conductivity and stream water ionic concentrations, and had little woody debris within their channels. Sierra Nevada (SN) sites were generally intermediate to Klamath and ECSF sites.

Model development

Site classification and annual variation in group membership.—We identified 10 groups when species data were used in the cluster analysis, and seven groups when family identifications were used (Fig. 2). For the species-level classification, sites within the same basin tended to cluster within just one or two groups implying basin-level features (including proximity to potential colonists) were strongly associated with faunal composition (see the Appendix, Table A1). In contrast, for the family data, sites from the same basin were spread across as many as five groups (see the Appendix, Table A2). On average, Bray–Curtis distances between sites were greater for species data than for family data (Fig. 2, mean site distances = 0.61 for species data and 0.33 for family data).

For both the species- and family-level classifications, sites sampled from the same basin in different years tended to occur in the same groups (see the Appendix, Tables A1 and A2). Nineteen of 24 samples from Thompson Creek (KM) were assigned to species Group 2 (11 samples from 1988 and eight from 1989). The remaining samples were assigned to adjacent Group 1. Of the 16 samples from the SN basin (North Fork American River), 15 (six from 1988 and nine from 1989) occurred in Group 5. One sample from 1989 occurred in Group 8. Within groups, sites sampled the same year tended to cluster together (data not shown).

Site fidelity for family-based groups was not as strong as observed for species data (see the Appendix,

Tables A1 and A2). Thompson Creek sites occurred in four widely separated groups (1, 2, 4, and 5), and sites from the North Fork American River occurred in three of the seven groups (2, 3, and 5). However, 88% of Thompson Creek sites occurred in two contiguous groups (1 and 2), and 80% of North Fork American River sites occurred in Group 3. Unlike the species data, there was no tendency for sites from the same year to cluster together within family-defined groups (data not shown).

Discriminant functions models (DFM) classification errors.—Jackknifed evaluations of internal consistency and external tests with independent sites both showed that the species DFM was substantially better than the family DFM in predicting group membership (see the Appendix, Tables A3 and A4). The jackknifed estimate of error for the species DFM was 23%, and the estimate based on the external test was 18%. The jackknifed estimate of error was 45% for the family DFM, and the estimate based on external data was 49%. The species model was least successful in predicting membership of sites from Group 5 and misclassified a high percentage of these sites into Groups 7 and 8. Misclassification rates were generally high for all family groups except Group 7. For the species, but not the family model, the percentage of within-group misclassifications were directly related to both the number of sites ($r = 0.91$, $P < 0.001$) and the number of basins ($r = 0.93$, $P < 0.001$) within a group.

Discriminant functions models predictor variables.—Because the internal and external estimates of error were nearly identical, we combined data from both the training and evaluation reference sites to estimate the contribution of different predictor variables to the models. These analyses identified 11 and 9 predictor variables for the species and family models, respectively (Table 3). For both models, variables that

described water chemistry (conductivity), climate regime and ecoregion (elevation, latitude, longitude), and stream size (basin area, length of channels upstream of the site, and stream width and depth) were most important in discriminating groups.

Error in predicting the expected fauna: the distribution of reference site O/E values.—The distribution of *O/E* values derived from both the training and evaluation reference sites was close to being normally distributed for both the species and family models (Fig. 3). For all models, the distributions of *O/E* values derived from the evaluation reference data were statistically indistinguishable from those obtained from the training data (Table 4), implying that all models generated stable and unbiased estimates of error in predicting the number of taxa at unimpaired sites. Mean *O/E* values for both species and family models were very close to unity, which showed that both models produced unbiased estimates of the number of taxa ex-
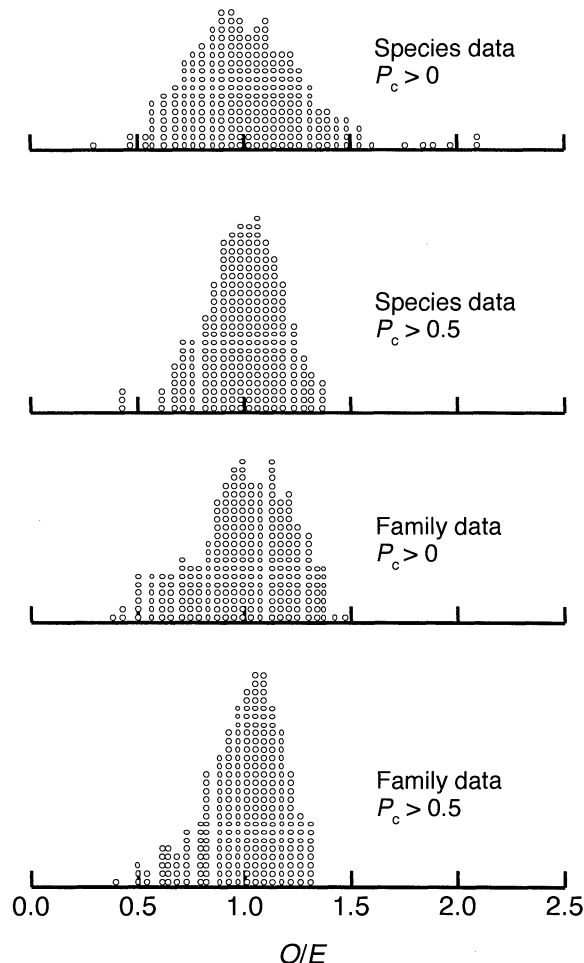
TABLE 3. Predictor variables identified by stepwise discriminant analysis for the species- and family-level models.

Variable	F
Species-level model	
log(Specific conductance)	29.6
Longitude	21.0
Basin area	13.1
Elevation	14.6
log(Depth)	13.1
Latitude	11.9
Stream length	9.5
log(Width)	7.2
Sampling date	5.2
log(Stream slope)	4.8
Azimuth	4.3
Family-level model	
log(Specific conductance)	13.9
log(Depth)	11.1
Longitude	10.0
Sampling date	8.6
Elevation	8.1
log(Stream slope)	5.6
Basin area	3.9
Stream length	3.9
log(Width)	3.4

Note: Variables are listed in order of their importance as measured by model *F* values.

FIG. 3. Distribution of observed-to-expected ratios (*O/E*) for reference sites. *O/E* values based on probabilities of capture >0 and ≥0.5 are shown for both the 10-group species and 7-group family models.

pected to occur at a site. Standard deviations of O/E_{50} (i.e., models based on $P_c \geq 0.5$) were lower than that of O/E_0, implying there was greater error predicting rarer taxa than more common ones. Although the distribution of overall error appeared to be unbiased in all models, bias occurred for some site classes in some models (Fig. 4). For example, 8 of 10 species site groups had O/E_0 means near unity, but the mean for Group 5 was significantly higher than the mean for Group 7 (Tukey's pairwise comparison, $P < 0.004$), implying that this model produced biased estimates of richness for these two groups. This bias was not noticeable among species groups for O/E_{50}, but was clearly apparent for both of the family models with Groups 5, 6, and 7 having consistently and significantly (Tukey's pairwise test, $P < 0.05$) lower mean *O/E* values than the other groups.

Differences among reference and test sites in O/E values.—Mean *O/E* values for the test sites were significantly lower than mean values for reference sites for species but not family models (Table 4). This difference in the ability of species and family models to detect differences in mean *O/E* values between reference and test sites was not related to model error; standard deviations of reference site *O/E* values were similar for the two models (Table 4).

The species model based on $P_c \geq 0.5$ was slightly more sensitive than the model based on $P_c \geq 0$. Mean test site *O/E* values were significantly lower for the O/E_{50} model than the O/E_0 model (Table 4; ANOVA, $P < 0.01$), a consequence of the smaller standard deviation for O/E_{50} values than O/E_0 values.

Observed-to-expected ratios (O/E_{50} and O/E_0 values)

TABLE 4. Means and standard deviations of O/E values for training (T) reference, evaluation (E) reference, and test sites for species- and family-level models calculated using probabilities of capture (P_c) >0 and >0.50.

| Data set | N | Species models | | | | Family models | | | |
| | | $P_c > 0$ | | $P_c > 0.50$ | | $P_c > 0$ | | $P_c > 0.50$ | |
		Mean	1 SD	Mean	1 SD	Mean	1 SD	Mean	1 SD
Combined groups									
Reference (T)	203	1.03	0.28	1.02	0.17	1.00	0.22	1.01	0.17
Reference (E)	51	1.03	0.23	0.98	0.16	1.00	0.23	1.01	0.17
Test	234	0.92†	0.26	0.87†	0.18	1.03	0.17	1.03	0.13
Combined reference sites by group									
1	2/19‡	1.04	0.24	1.02	0.15	1.04	0.15	1.02	0.09
2	9/41‡	1.00	0.15	1.04	0.10	1.11	0.14	1.08	0.11
3	2/66‡	1.02	0.18	1.00	0.14	1.01	0.16	1.05	0.12
4	5/68‡	1.02	0.19	1.02	0.11	1.16	0.13	1.04	0.11
5	7/44‡	1.15	0.37	1.02	0.17	0.75	0.16	0.84	0.14
6	10/5‡	1.01	0.16	1.00	0.11	0.79	0.25	0.81	0.20
7	6/10‡	0.92	0.28	0.96	0.22	0.65	0.16	0.64	0.12
8	11	0.98	0.28	0.99	0.20	—	—	—	—
9	11	0.96	0.28	0.97	0.25	—	—	—	—
10	11	1.01	0.26	1.02	0.22	—	—	—	—

† Indicates means for evaluation and test sites that are statistically ($P < 0.05$) different from means for the training data.
‡ Slash (/) separates N values for species- and family-level groups.
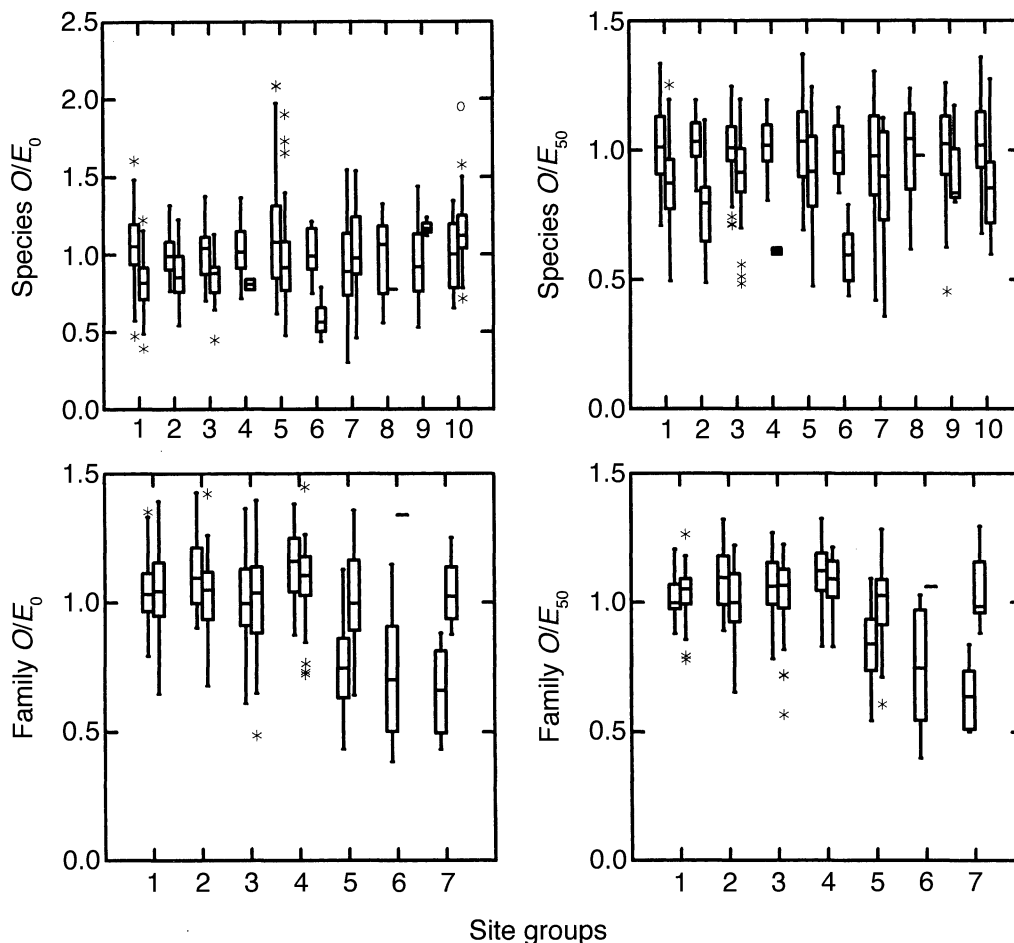

FIG. 4. Box plots showing the distribution of O/E values by site class and site condition (reference vs. test sites). For each site class, reference data are on the left, and test data are on the right. Asterisks are outliers, and the open circle is a distant outlier.
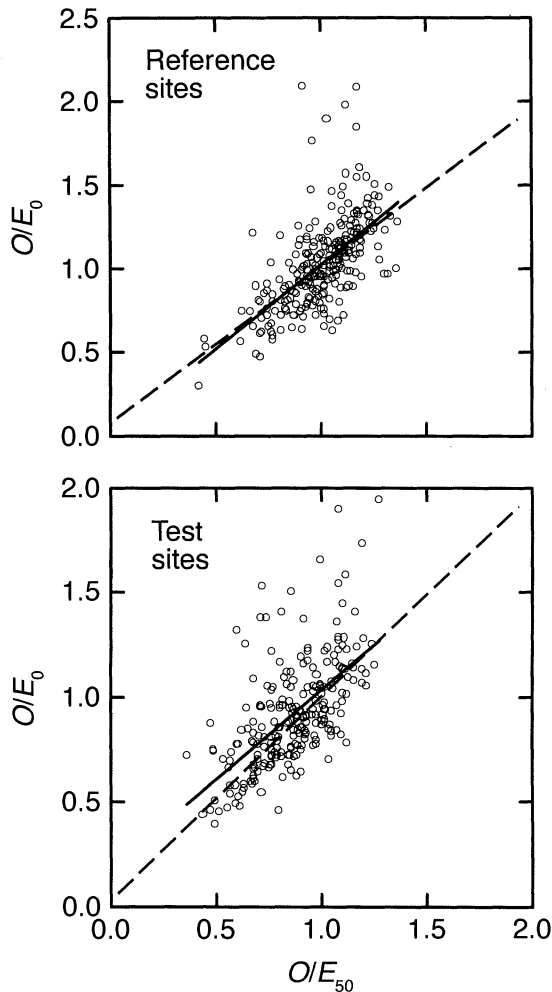
FIG. 5. Relationship between O/E_{50} and O/E_0 for the species data from reference (top panel) and test (bottom panel) sites. The diagonal lines represent the line of perfect agreement between the two measures. For reference sites, $O/E_0 = 0.014 + 1.012(O/E_{50})$, $n = 254$, $r^2 = 0.39$, $P < 0.001$ for H_0: slope $= 1$. For test sites, $O/E_0 = 0.182 + 0.853(O/E_{50})$, $n = 234$, $r^2 = 0.37$, $P < 0.001$ for H_0: slope $= 1$ and H_0: intercept $= 0$.

were linearly correlated with one another (Fig. 5), although they were not strongly related for either reference or test sites ($r^2 = 0.39$ and 0.37, respectively). Fig. 5 gives the impression that the relationship between O/E_0 and O/E_{50} was nonlinear, but residuals from the 1:1 were unrelated to O/E_{50}. Furthermore, fitting the data to a power function did not improve the overall fit. Although O/E_{50} was only a slightly biased predictor of O/E_0 for reference sites (intercept, 0.014; slope, 1.012), it was more strongly biased for the test site data (intercept, 0.182; slope, 0.853). On average, O/E_{50} overestimated O/E_0 with the amount of bias greatest at low O/E_{50} values.

The difference between reference and test sites also varied with site group (Fig. 4). For O/E_0, test sites

predicted to occur in Groups 1, 4, and 6 had substantially lower O/E values than reference sites, whereas reference and test values were similar in other site groups. For O/E_{50}, test sites predicted to occur in Groups 2, 4, and 6 had lower O/E values than reference sites. For both family models, test site O/E values were higher in family Groups 5, 6, and 7 than reference site values.

Differences among reference and test sites in faunal composition and dominance diversity relationships.— On average, the abundances (hence their probabilities of capture) of many taxa observed at test sites differed considerably from those observed at reference sites, with some taxa being less abundant and other taxa being more abundant at test sites (Fig. 6). Taxa that were most abundant at reference sites were most likely to show large numerical responses to management with the magnitude of differences generally diminishing with the rank abundance of taxa at reference sites. Furthermore, the 100 most common taxa were less abundant, on average, at test sites than reference sites, whereas rarer taxa (i.e., rank > 150) were more abundant at test sites (Fig. 6). The proportional differences in abundances between test and reference sites were generally similar for all taxa, regardless of rank abundance (Fig. 6). This analysis also showed that taxa that were rare at reference sites tended to respond positively to the conditions found at test sites.

Although test sites clearly differed from reference sites in the abundances of many individual taxa and hence assemblage composition, on average, little difference existed in total densities or in assemblage dominance diversity relationships. For example, the tendency for test sites to have lower abundances of some taxa was generally matched by higher abundances of other taxa such that there was little difference in total densities between reference and test sites (Fig. 6). Furthermore, when composition was ignored, there was essentially no difference in the overall assemblage structure of reference and test site assemblages as measured by dominance diversity relationships (Fig. 6).

Compositional differences between reference and test sites.—Several taxa were substantially less abundant at test sites than reference sites and would thus have been less likely to be captured at test sites than predicted (Table 5). These taxa would therefore have had a strong influence on O/E values of test sites. Densities of 50 taxa were 3–500× less abundant at test sites than reference sites (Table 5). Blackflies (Simuliidae), aquatic mites (Hydracarina), the riffle beetles *Optioservus quadrimaculatus* and *Zaitzevia parvula*, five chironomid midges (*Cricotopus* [*Nostococladius*], *Cladotanytarsus* [*vanderwulpi*], *Polypedilum* [Pentapedilum], *Corynoneura*, and *Orthocladius* [*Eudactylocladius*]), aquatic worms (Oligochaeta), the caddis *Micrasema*, the stoneflies *Taenionema* and Chloroperlidae, copepods, and the clam *Pisidium* appeared to be
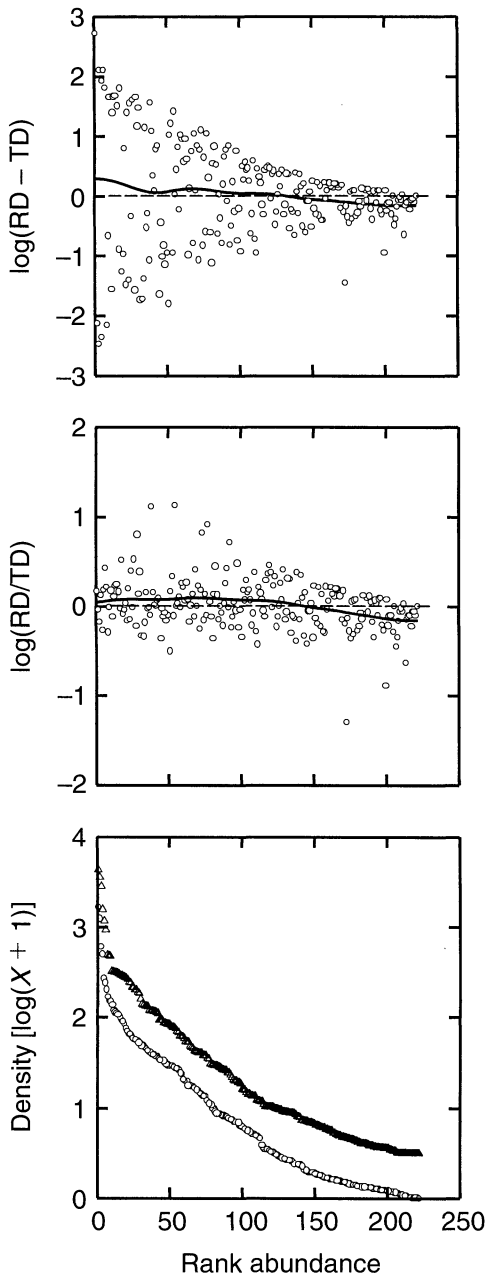
FIG. 6. Top panel: Relationship between the absolute difference in mean taxon abundances between reference and test sites ($\log_{10}$[reference site density − test site density]) and the rank order of reference taxa abundances. Taxa with values greater than 0 were more abundant at reference sites. The smoothed line describes how, on average, densities varied between reference and test sites as a function of rank abundance. The line was fitted by distance-weighted sum of squares. Middle panel: Relationship between the proportional difference in mean taxon abundances between reference and test sites ($\log_{10}$[(reference site density)/(test site density)]) and the rank order of reference taxa abundances. The smoothed line was fit as in the top panel. Bottom panel: Dominance–diversity curves derived from mean abundances of taxa found at reference (circles) and test (triangles) sites. Because the two

most sensitive (i.e., $>30\times$ difference in abundance) to conditions at test sites.

Densities of an additional 50 taxa were 3–$300\times$ more abundant at test sites (Table 5). The taxa most indicative of conditions at test sites ($>30\times$ more abundant than at reference sites) included the midges *Micropsectra/Tanytarsus, Tvetenia* [*bavarica*], *Eukieferiella* [*brevicalar/claripennis*], *E.* [*devonica*], *Pagastia, Rheocricotopus,* and *Orthocladius* (*Orthocladius*); the stoneflies *Zapada/Malenka*; the mayflies *Baetis* and *Paraleptophlebia*; and the caddis *Apatania.*

Taxa within certain higher taxonomic groups appeared to be more sensitive to conditions at test sites than those in other groups (Table 5). For example, mayflies as a group appeared to respond positively to conditions at test sites: of the 100 taxa that showed the largest difference between reference and test sites, 11 mayfly taxa were most abundant at test sites, whereas only two taxa were most abundant at reference sites. Conversely, only one beetle was most abundant at test sites, whereas eight beetles were most abundant at reference sites. No obvious trends existed for Diptera (27 taxa most abundant in reference sites and 27 in test sites), Plecoptera (3/1), or Trichoptera (6/8). Although little change occurred in overall representation at reference and test sites for these groups as a whole, the specific taxa that were most abundant within each group of taxa at these two groups of sites differed markedly.

Variation in species O/E values with year, ecoregion, and habitat type.—Effects of habitat type, year, and ecoregion were tested with ANOVA for both reference and test sites (Table 6). These analyses showed that habitat type did not influence either O/E_0 or O/E_{50} values for reference sites, but both O/E_0 and O/E_{50} values from test sites varied significantly among habitat types. Samples from high-gradient riffles and step runs had lower O/E values than samples from low-gradient riffles and runs, implying these habitat types were more sensitive to land use than the other habitats.

Values of O/E_0, but not O/E_{50}, differed between the two years of sampling, with 1988 values being higher than 1989 values. Year had a similar effect on both reference and test site O/E_0 values and was associated with ~2°C difference in mean July water temperatures between the two years (1988, 17.7°C; 1989, 15.6°C).

Ecoregion had no effect on reference site O/E values, but test sites from the Klamath ecoregion had slightly lower O/E values than sites from the Sierra Nevada for the $P_c \geq 0$ model, but not the $P_c \geq 0.5$ model. Test sites from the ECSF ecoregion were all outside the

curves almost exactly overlap each other, the data for the test sites are offset 0.5 units on the ordinate to make both curves visible. For both sets of data, $\log(X + 1)$ densities were plotted against their respective rank abundances.

TABLE 5. Mean difference between test and reference sites in abundance of different taxa.

Taxon	Operational taxonomic unit (OTU)	log(RD − TD)	log(RD/TD)
Most abundant at reference sites			
Simuliidae	Simuliidae	2.722	0.167
Hydracarina	Hydracarina	2.103	0.128
Coleoptera	*Optioservus quadrimaculatus*	2.098	0.421
Chironomidae	*Cricotopus (Nostococladius)*	1.923	0.186
Oligochaeta	Oligochaeta	1.809	0.211
Chironomidae	*Cladotanytarsus (vanderwulpi)* "A"	1.796	0.514
Chironomidae	*Polypedilum (Pentapedilum)*	1.670	0.249
Chironomidae	*Corynoneura*	1.663	0.232
Trichoptera	*Micrasema*	1.658	0.176
Plecoptera	*Taenionema*	1.649	0.798
Copepoda	Copepoda	1.607	0.525
Bivalva	*Pisidium*	1.566	1.108
Coleoptera	*Zaitzevia parvula*	1.553	0.392
Plecoptera	Chloroperlidae	1.501	0.163
Chironomidae	*Orthocladius (Eudactylocladius)* "A"	1.470	0.389
Chironomidae	*Sublettea*	1.417	1.121
Chironomidae	*Thienemanniella*	1.393	0.196
Ephemeroptera	*Diphetor hageni*	1.390	0.103
Chironomidae	*Stilocladius*	1.209	0.334
Chironomidae	*Synorthocladius*	1.161	0.136
Coleoptera	*Heterolimnius*	1.143	0.159
Chironomidae	*Stempellina*	1.096	0.818
Coleoptera	*Eubrianax edwardsii*	1.090	0.155
Chironomidae	*Rheosmittia*	1.031	0.907
Chironomidae	*Cladotanytarsus (vanderwulpi)* "B"	1.031	0.415
Turbellaria	Turbellaria	1.016	0.264
Plecoptera	*Yoroperla*	1.010	0.169
Coleoptera	*Rhizelmis nigra*	0.955	0.266
Coleoptera	*Ordobrevia nubifera*	0.946	0.218
Chironomidae	*Paratrichocladius*	0.843	0.713
Coleoptera	*Cleptelmis*	0.842	0.086
Trichoptera	*Wormaldia*	0.836	0.226
Chironomidae	*Heterotrissocladius (marcidus)*	0.821	0.447
Chironomidae	*Potthastia (gaedii)*	0.808	0.112
Trichoptera	*Agapetus*	0.789	0.034
Diptera	*Bezzia*	0.779	0.069
Trichoptera	*Brachycentrus*	0.767	0.430
Chironomidae	*Microtendipes (pedellus)*	0.759	0.183
Trichoptera	*Gumaga*	0.749	0.141
Chironomidae	*Orthocladius (Eudactylocladius)* "B"	0.748	0.479
Chironomidae	*Phaenopsectra*	0.657	0.259
Chironomidae	*Polypedilum (Tripodura)*	0.641	0.141
Chironomidae	*Pentaneura*	0.609	0.084
Trichoptera	*Rhyacophila (nevadensis)*	0.588	0.384
Chironomidae	*Cricotopus (Isocladius) (laricomalis)*	0.576	0.065
Chironomidae	*Diamesa*	0.571	0.291
Ephemeroptera	*Cloeon*	0.548	0.143
Diptera	*Hexatoma*	0.514	0.133
Coleoptera	*Hydraena*	0.496	0.457
Chironomidae	*Cardiocladius*	0.487	0.189
Most abundant at test sites			
Diptera	*Maruina*	−0.490	−0.232
Trichoptera	*Heliocopsyche borealis*	−0.494	−0.453
Ephemeroptera	*Drunella coloradensis/flavilinea*	−0.502	−0.117
Trichoptera	*Rhyacophila (hyalinata)*	−0.513	−0.325
Diptera	*Atherix*	−0.529	−0.351
Chironomidae	*Parametriocnemus*	−0.534	−0.025
Ephemeroptera	*Attenella*	−0.547	−0.302
Chironomidae	*Microtendipes rydalensis*	−0.602	−0.099
Trichoptera	*Glossosoma*	−0.613	−0.120
Chironomidae	*Nanocladius* "A"	−0.621	−0.419
Chironomidae	*Cricotopus (bicinctus)*	−0.634	−0.218
Ephemeroptera	*Drunella spinifera*	−0.642	−0.145
Chironomidae	*Virgatanytarsus*	−0.650	−0.631
Ephemeroptera	*Caudatella heterocaudata*	−0.655	−0.392
Ostracoda	Ostracoda	−0.671	−0.011
Ephemeroptera	*Serratella teresa*	−0.737	−0.282
Trichoptera	*Rhyacophila (brunnea)*	−0.751	−0.236

Table 5. Continued.

Taxon	Operational taxonomic unit (OTU)	log(RD − TD)	log(RD/TD)
Diptera	*Chelifera*	−0.792	−0.215
Nematoda	Nematoda	−0.797	−0.122
Chironomidae	*Heleniella*	−0.815	−0.244
Chironomidae	*Cricotopus (tremulus) "A"*	−0.829	−0.075
Trichoptera	*Rhyacophila (betteni)*	−0.834	−0.157
Chironomidae	*Rheotanytarsus*	−0.901	−0.032
Chironomidae	*Nilotanypus*	−0.930	−0.347
Diptera	*Pericoma*	−0.954	−0.111
Ephemeroptera	*Ameletus*	−0.956	−0.105
Ephemeroptera	*Serratella sequoia*	−0.956	−0.427
Chironomidae	*Psectrocladius (sordidellus)*	−0.961	−0.887
Ephemeroptera	*Epeorus*	−0.975	−0.047
Chironomidae	*Larsia*	−0.989	−0.177
Trichoptera	*Hydropsyche*	−1.021	−0.114
Chironomidae	*Eukiefferiella (gracei)*	−1.080	−0.108
Trichoptera	*Amiocentrus aspilus*	−1.083	−0.251
Ephemeroptera	*Heptagenia*	−1.126	−0.358
Coleoptera	*Ampumixus dispar*	−1.150	−0.160
Chironomidae	*Cricotopus (tremulus) "B"*	−1.267	−0.083
Chironomidae	*Brillia*	−1.391	−0.187
Chironomidae	*Eukiefferiella (brehmi)*	−1.403	−0.144
Ephemeroptera	*Serratella michneri*	−1.460	−1.303
Chironomidae	*Paraleptophlebia*	−1.490	−0.163
Chironomidae	*Orthocladius (Orthocladius)*	−1.571	−0.117
Chironomidae	*Rheocricotopus*	−1.582	−0.227
Chironomidae	*Pagastia*	−1.649	−0.364
Chironomidae	*Eukiefferiella (devonica)*	−1.733	−0.337
Chironomidae	*Eukiefferiella (brevicalcar)/(claripennis)*	−1.740	−0.327
Trichoptera	*Apatania*	−1.808	−0.503
Ephemeroptera	*Baetis*	−2.144	−0.046
Chironomidae	*Tvetenia (bavarica)*	−2.163	−0.294
Plecoptera	*Zapada/Malenka*	−2.364	−0.271
Chironomidae	*Micropsectra/Tanytarsus*	−2.478	−0.177

Notes: The 100 taxa showing the largest absolute difference in abundance between test and reference sites are listed in rank order, from most abundant in reference sites to most abundant in test sites. RD = reference site density, and TD = test site density; log(RD − TD) is sensitive to absolute difference in abundance between sites, whereas log(RD/TD) is sensitive to relative differences in abundance between sites.

experience of the model, so no comparisons were possible.

Variation in mean species O/E values between basins.—Mean O/E values are based on a mix of reference and test sites that occur within each basin and estimate the overall condition of the invertebrate biota within basins. These mean values varied substantially across the 49 study basins (by a factor of 1.81 for O/E_{50} and a factor of 2.79 for O/E_0; Table 7), implying that basins varied, on average, in the biological integrity of their stream invertebrate fauna. Although mean values of O/E_{50} and O/E_0 were generally similar to one another, in some cases values of O/E_0 were markedly higher than O/E_{50} (e.g., see values for basins 20, 45, 13, and 19 in Table 7). The two measures therefore behaved somewhat differently across basins and would thus provide somewhat different inferences of biological quality at this scale of observation, an observation consistent with the relationship between O/E_0 and O/E_{50} at the site scale (Fig. 5).

Relationships between O/E values and land use variables.—Principal components analysis showed that variation among test sites in the eight land use variables could be collapsed onto three independent axes of variability (Table 8). Variables describing the amount of timber harvest and length of roads above each test site all loaded strongly on Factor 1; percent riparian shading and stream water NO_3/NO_2 concentration loaded on Factor 2; and streambed particle embeddedness and stream water PO_4 concentration loaded on Factor 3. Regression analysis of O/E values showed that estimates of the amount of timber harvest and road construction were not associated with O/E values for any model (Table 9). However, O/E values for both species models were negatively related to riparian shading and concentration of NO_3/NO_2, but family model values were not. Finally, values for three of four models were negatively related to percent substrate embeddedness and PO_4 concentration, the exception being the family O/E_0 model. In combination, land use variables were associated with only 8–11% of the variation in test site O/E values, suggesting land management activities in these basins have had subtle effects on stream biota.

DISCUSSION

In large measure, our results corroborated previous work from Great Britain and Australia that show pre-

TABLE 6. Relationships between mean O/E values, habitat type, year, and ecoregion.

O/E	Site	Habitat type					Year			Ecoregion			P
		HR	LR	R	SR	P	1988	1989	P	KM	SN	ECSF†	
$P_c > 0$	R	1.02	1.06	1.00	1.06	NS	1.07	0.99	0.02	0.99	1.05	1.00	NS
		(86)	(68)	(36)	(38)		(110)	(121)		(82)	(120)	(32)	
	T	0.85	1.10	0.96	0.89	0.001	0.98	0.87	0.001	0.87	0.95	—	0.03
		(94)	(46)	(41)	(50)		(105)	(129)		(145)	(88)	—	
$P_c > 0.5$	R	1.00	1.04	0.98	1.00	NS	1.01	1.00	NS	0.99	1.01	1.00	NS
	T	0.85	0.96	0.84	0.85	0.001	0.86	0.88	NS	0.86	0.87	—	NS

Notes: Site abbreviations: R, reference; T, test. Habitat-type abbreviations: HR, high-gradient riffle; LR, low-gradient riffle; R, run; and SR, step run. Ecoregion abbreviations: KM, Klamath Mountains; SN, Sierra Nevada; and ECSF, Eastern Cascade Slopes and Foothills. Sample sizes are shown in parentheses for each set of contrasts and are the same for both O/E_0 and O/E_{50} models.

† Test sites within the ECSF ecoregion were not assessed because of missing data.

dictive models can provide a powerful means of assessing the biological condition of streams (Wright et al. 1984, 1993, Moss et al. 1987, Parsons and Norris 1996, Marchant et al. 1997). Given the similarities among these studies, the River Invertebrate Prediction and Classification System (RIVPACS) approach may offer a standard method of biological assessment that can be applied to streams throughout the world. This latter point is important in that such a standard approach would potentially allow direct comparison of the biological condition of streams at local, regional, and global scales. Further work will be necessary, however, before such comparisons can be made with confidence. For example, RIVPACS-type models have largely been developed for small, easily sampled (i.e., wadeable) streams. Applying this approach to larger rivers, which tend to be more degraded, will likely require use of reference sites located within least-impaired reaches of river. Given that the current conditions of the best river sites are clearly less pristine than that of the best small-stream sites, observed-to-expected ratios (O/E values) derived from these two types of lotic environments will not be absolutely comparable in terms of their historical potential. However, they may very well provide comparable assessments in terms of the biotic conditions that society believes to be realistically attainable in different types of systems.

Predictor variables and biotic structure

That models developed in Great Britain (e.g., Wright 1995), Australia (Parsons and Norris 1996), and California (this study) use largely the same variables to predict biotic composition has interesting implications for both our understanding of the factors influencing stream biota and our ability to assess biological condition. For example, good predictions of biotic composition in Great Britain, Australia, and California suggest stream invertebrate assemblages in different parts of the world are highly structured by environmental filters (sensu Poff 1997) and not random assemblages, as has been implied by some stream ecologists (e.g., Winterbourn et al. 1981, McCreadie et al. 1997). These results also imply that local richness in streams is not

simply proportional to regional richness, as seems to be the case for many other assemblages (see Cornell 1993, Caley and Schluter 1997). Instead, richness appears to conform best to a niche model in which species composition and richness varies with the diversity of resources available at each location, an idea supported by a large literature on stream invertebrates (see review by Vinson and Hawkins [1998]).

The predictor variables used in these models provide insight into the specific niche axes that are important in partitioning taxa among sites. For example, the importance of geographic variables such as latitude, longitude, and elevation implies temperature is a primary factor determining the composition of stream invertebrate faunas, a result consistent with the ideas of Ide (1935), Vannote and Swenney (1980), Hawkins et al. (1997), and others. The importance of alkalinity or conductivity as predictors suggests that either the ionic composition of the water or the geologic parent material from which bed materials are derived are also important determinants of biotic structure. Finally, the importance of several measures of stream size, such as width, depth, basin area, and length of stream above a site, corroborate previous observations that taxonomic composition shifts in predictable ways from headwaters to larger rivers (Vannote et al. 1980). The similarities in the predictor variables used in models from the three regions are all the more remarkable given that these regions are distinctly dissimilar in both physical setting and their biota.

Perhaps the most valuable practical aspect of these results is that most of the important predictor variables can be easily and quickly measured at low cost. Geographic variables usually can be easily measured in the office. Alkalinity or conductivity, stream depth, and stream width take only a few minutes to measure, thus adding little cost to that associated with the primary biological sampling.

Presence–absence and biotic structure

RIVPACS models use information on the presence and absence of taxa to measure biotic structure and ignore abundance information. We might therefore ex-

TABLE 7. Mean species-level O/E values, with one standard deviation, for individual study basins.

Ecoregion and basin	N	O/E_0 Mean	O/E_0 1 SD	O/E_{50} Mean	O/E_{50} 1 SD
KM ecoregion					
East Fork Beaver Creek (28)	7	1.058	0.230	0.889	0.172
Indian Creek (1988) (4)	13	0.966	0.158	1.037	0.129
Indian Creek (1989) (38)	13	0.869	0.212	0.881	0.130
Thompson Creek (1988) (9)	13	0.960	0.118	1.034	0.078
Thompson Creek (1989) (49)	11	1.036	0.152	0.975	0.138
Patricks Creek (8)	6	1.019	0.249	0.976	0.191
Walker Creek (51)	15	0.785	0.144	0.795	0.105
Jones Creek (5)	10	0.966	0.294	0.960	0.257
Thompkins Creek (50)	15	0.839	0.112	0.905	0.137
Kelsey Creek (41)	13	0.891	0.224	0.845	0.148
North Fork Russian River (46)	9	1.124	0.140	1.122	0.127
Nordheimer Creek (7)	12	0.870	0.198	0.962	0.141
Big Creek (Shasta) (2)	4	0.955	0.171	0.846	0.090
Grouse Creek (3)	10	0.769	0.190	0.640	0.118
Upper Rattlesnake Creek (11)	13	0.992	0.177	0.982	0.150
Upper East Fork of the South Fork Trinity River (10)	13	1.195	0.235	1.069	0.117
ECSF ecoregion					
Lassen Creek (42)	16	0.964	0.180	0.966	0.139
East Creek (35)	16	1.045	0.218	1.025	0.110
SN ecoregion					
Nelson Creek (43)	7	1.144	0.224	1.067	0.135
Jamison Creek (39)	15	0.967	0.207	1.060	0.173
Poplar Creek (48)	15	1.106	0.212	0.973	0.141
Canyon Creek (32)	10	0.939	0.310	0.947	0.145
Lavazolla Creek (20)	8	1.368	0.444	1.033	0.203
Haypress Creek (1988) (17)	13	1.023	0.250	0.910	0.176
Haypress Creek (1989) (37)	10	0.889	0.202	0.974	0.193
Middle Fork Yuba River (21)	11	1.187	0.278	0.967	0.161
North Fork American River (1988) (22)	9	1.141	0.340	1.124	0.279
North Fork American River (1989) (45)	12	0.735	0.153	0.982	0.171
Alder Creek (12)	11	1.299	0.382	0.883	0.221
North Fork Consumnes River (23)	11	0.848	0.167	0.920	0.170
Cole Creek (14)	11	1.012	0.262	1.019	0.218
Blue Creek (31)	12	0.779	0.170	0.872	0.173
Deadman Creek (33)	10	1.016	0.260	1.010	0.200
Beaver Creek (27)	15	0.860	0.123	0.885	0.138
Bourland Creek (13)	7	1.657	0.419	0.982	0.123
Jawbone Creek (19)	7	1.252	0.244	0.687	0.107
Jackass Creek (18)	6	1.078	0.166	0.878	0.127
Big Creek (Sierra) (30)	10	0.953	0.357	0.907	0.299
Kaiser Creek (40)	7	0.913	0.173	0.907	0.109
Dinkey Creek (34)	6	0.670	0.249	0.721	0.188
Deer Creek (15)	6	0.912	0.189	1.007	0.067
Rancharia Creek (25)	6	1.179	0.129	1.142	0.082
Ninemile Creek (24)	11	0.994	0.158	0.977	0.133
Freeman Creek (36)	6	0.837	0.156	0.800	0.138
Peppermint Creek (47)	2	1.039	0.228	0.919	0.201
Fish Creek (16)	11	0.834	0.277	0.820	0.255
Rattlesnake Creek (26)	5	0.593	0.114	0.628	0.111
Nobe Young Creek (44)	9	0.890	0.127	0.838	0.259

Notes: Basins are ordered from north to south within ecoregions. The number following the basin name is the basin code that corresponds with those given in the Appendix.

pect these models to be insensitive to many stressors, because individual populations can suffer considerable degradation before going locally extinct. However, at the assemblage level, presence–absence data appear sufficiently robust to allow detection of reasonably subtle differences among sites. Although not explored here, previous work has shown that RIVPACS can distinguish four bands of degradation with acceptable er-

rors in misclassification (Clarke et al. 1996). These bands represent classes of biological impairment ranging from equivalent to reference to poor. The robustness of presence–absence data in allowing an assessment of biological condition has important implications for how we approach bioassessment in the United States and elsewhere. If presence–absence data allow a reasonably accurate assessment of biological condition,

TABLE 8. Loadings of the original land use variables on three principal component axes.

Variable	Principal component axis		
	1	2	3
No. km^2 of basin cut	0.943	0.028	−0.062
No. km of roads	0.880	−0.026	0.252
Percent basin cut	0.861	0.104	−0.164
No. stream-road crossings	0.682	−0.090	0.371
Percent riparian shading	0.088	0.840	0.056
NO_3/NO_2 (1 µg/kg H_2O)	−0.066	0.752	0.099
Percent substrate embeddedness	−0.112	0.033	0.802
PO_4 (1 µg/kg H_2O)	0.253	0.173	0.634

Note: All variables were transformed (square root) prior to extraction of principal components, and principal axes were rotated (varimax) to reveal best separation of variables.

we may not need to find ways of detecting abundance changes in taxa that show inherently high seasonal and annual variation in population densities.

Adequacy of one-season sampling

Wright et al. (1993) recommend that models and assessments be based on samples collected from three different seasons to ensure that all taxa present at a site are sampled. However, we were able to build sensitive and accurate models with data collected from one-time sampling. Because both the misclassification errors in the discriminant functions models (DFM) we constructed and the standard deviations of reference site O/E values were generally lower than reported by Moss et al. (1987) for British models, it may be possible to relax the three-season sampling criterion in this region without significant loss of sensitivity. However, if one-season sampling is used for assessments, we concur with the recommendations of others (e.g., Gibson et al. 1996, Barbour et al. 1997) that sampling be restricted to a standard index period to minimize the influence of phenological shifts in taxonomic composition.

Taxonomic resolution and assessment sensitivity

Previous studies have reported only slight to moderate loss of sensitivity when using family identifications (Furse et al. 1984, Marchant et al. 1995, Norris 1996). However, in this study, assessments based on family data were unable to detect differences among reference and test sites. There are at least two possible reasons for the difference in sensitivity between species and family assessments. First, the degree of impairment

at test sites may be too subtle for family assessments to detect. This explanation is plausible given the slight differences detected by the species assessments. Alternatively, family assessments would also be less sensitive than species assessments if families were comprised of several species that differ considerably in their ecological requirements. We believe this second explanation may apply to the California invertebrate stream fauna. Several families show extensive adaptive radiation in diet and use of habitats (e.g., ephemerellid mayflies and limnephilid caddisflies). This type of specialization within families is not as apparent in the British or Australian (Victoria) faunas (Hawkins and Norris 2000). For regions with few genera and species per family, species and family assessments would be expected to perform similarly. For regions with many genera and species per family, important information on species-specific taxon–habitat relationships could easily be lost by adopting a coarser level of taxonomic resolution. Although this idea requires careful scrutiny, we expect family assessments may be relatively inaccurate in environmentally heterogeneous regions, such as many parts of the western United States. However, family assessments are also likely to work adequately in environmentally diverse regions that have been recently glaciated and where taxonomic richness would be expected to be low (such as the parts of the Canadian Rockies).

Quantifying site and regional biological conditions

River Invertebrate Prediction and Classification System predictive models were developed to provide site-

TABLE 9. Results of multiple regression analyses testing relationships between test site O/E values and principal component (PC) scores of the land use variables shown in Table 8.

Variable	Species O/E_0 ($R^2 = 0.08$)		Species O/E_{50} ($R^2 = 0.11$)		Family O/E_0 ($R^2 = 0.02$)		Family O/E_{50} ($R^2 = 0.11$)	
	SRC	P	SRC	P	SRC	P	SRC	P
PC-1 (timber harvest and roads)	−0.074	0.282	−0.091	0.179	0.104	0.139	0.121	0.074
PC-2 (shading and NO_3)	−0.267	<0.001	−0.209	0.002	0.120	0.086	0.063	0.344
PC-3 (embeddedness and PO_4)	−0.134	0.039	−0.261	<0.001	0.088	0.186	−0.287	<0.001

Notes: Standardized regression coefficients (SRC) and probabilities of significance (P) given for each O/E model. $N = 230$ for all analyses.

specific assessments of the biological quality of streams. In this respect, use of species models revealed distinct effects of timber harvest practices at some test sites. Previous attempts to measure effects of timber harvest on aquatic invertebrates have provided ambiguous results (cf., Newbold et al. 1980, Hawkins et al. 1982). Lack of a standard scientifically defensible method has also greatly hindered our ability to describe the condition of stream biota on basin and regional scales. Such assessments are critical, however, in developing forest management plans. We believe RIVPACS models can be a potentially powerful means of describing how conditions vary among basins and entire regions. For example, our study showed that, on average, some streams have clearly been more degraded than others (Table 7) and that throughout the region studied land use associated with timber harvest has resulted in a subtle but distinct loss of ~10% of the invertebrate fauna occurring in streams draining altered catchments (Fig. 3, Table 4). Presentation of O/E values in this way provides a means of meeting the Environmental Monitoring and Assessment Program's (EMAP; a branch of the U. S. Environmental Protection Agency) goal of describing resource status and trends on a regional basis (Paulsen and Linthurst 1994). By tracking how O/E values vary over time among randomly sampled test sites (probability-based sampling), we would be able to determine if the condition of the stream invertebrate fauna in a region was improving, declining, or stable and thus provide critically needed information to resource managers and policy makers.

Effects of rare taxa on assessments

In the British RIVPACS models, O/E values are calculated from all nonzero probabilities of capture. In the Australian River Assessment System (AUSRIVAS) models, O/E values are calculated from only those probabilities ≥ 0.5. Theoretically, it seems sensible that using all of the information available would provide a more sensitive measure of biological condition, especially given that rare taxa are typically more at risk of extinction than taxa that are more common. In fact, the sensitivity of RIVPACS in detecting impairment of British streams appears to decrease as the minimum probability of capture used to calculate O/E increases (J. Wright, *personal communication*). This behavior is implied by the data in Fig. 5, which shows the slope of the O/E_{50} vs. O/E_0 plot to be less than one. A slope that is less than one should occur among test sites if O/E_{50} underestimates faunal richness at good sites, or overestimates richness at poor sites. Cao et al. (1998) provide additional empirical support for this view by showing that the assessed difference between biologically rich and depauperate sites declined as rare taxa were excluded from the comparisons. However, across the range of O/E values observed, the magnitude of this bias was much smaller (<0.1 O/E units) than the random sources of variation observed in the relation-

ship between O/E_{50} and O/E_0 (Fig. 5). The relatively weak correspondence between O/E_{50} and O/E_0 values implies that the confidence intervals associated with individual assessments are relatively large, regardless of which O/E measure is used.

Our overall results, however, imply that more information is not always better for making assessments and appear to support the view tentatively adopted in Australia that O/E values that are based on intermediate P_c values provide more robust assessments. Our study also provided some basis for understanding why O/E_{50} may be a more robust measure than O/E_0, in spite of being based on less information.

The reason O/E_{50} was more sensitive than O/E_0 seems to be related to how common and rare taxa varied in response to land use. Although there was essentially no difference in the dominance diversity curves for reference and test sites (Fig. 6), the difference between reference and test sites in individual taxa abundances depended on their rank abundance in reference streams. For example, common taxa (i.e., those with high to intermediate P_c's) tended to be less abundant at test sites. Many of these taxa would have had $P_c > 0.5$ at reference sites but $P_c < 0.5$ at test sites, and their "absence" at test sites would have caused O/E ratios to be less than unity. However, taxa with rank abundances >150 tended to have higher abundances in test sites, implying land use either directly or indirectly benefited these taxa. This tendency for rare taxa to increase in abundance at test sites would have partly compensated for the tendency of the loss of more common taxa to lower O/E_0, which includes data on all taxa. However, this increased probability of capturing rare taxa at test sites would have been ignored in the calculation of O/E_{50}, because these taxa would have had $P_c < 0.5$ at reference sites. O/E_{50} thus appears to be more sensitive in assessing initial stages of degradation in these oligotrophic streams than does O/E_0. Furthermore, O/E_{50} was less sensitive to natural sources of variation (stream type, habitat type, year, and ecoregion) than was O/E_0.

Spatial and temporal robustness of models

To be most useful, predictive models should be robust to natural environmental variation in both space and time. The process of model construction directly addresses natural spatial variation through the creation of stream classes and the use of spatially varying environmental data to predict sites into classes. Although we had limited data from which to address the issue of temporal robustness, the initial results were encouraging. Mean O/E_{50} values for the two reference catchments sampled in 1988 were within 6% and 14% of the mean values observed in 1989 (Table 7), and species O/E_{50} values for both reference and test sites were not related to year of sampling.

Model predictions and responses of specific taxa

To fully evaluate the utility of RIVPACS models for assessing effects of timber harvest practices, we need to scrutinize the predictions of the models. We need to be particularly aware of whether the models are producing false positives (i.e., inference of impairment when none exists). False positives could potentially arise if important predictor variables were missing from the model. In the absence of being able to compare predictions with the true condition of a site, one way of assessing the overall accuracy of a model is to ask if the observed trends in biotic composition among test sites make ecological sense.

Inspection of how abundances of specific taxa varied among reference and test sites (Table 5) suggests that at least some taxa are responding to land use in ways consistent with model predictions. The tendency for some taxa to have lower abundances at test sites and others to have higher abundances is consistent with the idea that logging causes changes in sediment and nutrient loads that would adversely affect some taxa while benefiting others. In general, increased scour and accumulation of fine sediment would be expected to exclude some bare-surface and moss dwellers as would increased periphyton growth associated with increased nutrient loading. The tendency for test sites to have lower densities of blackflies, which require clean surfaces for attachment; riffle beetles and fingernail clams, which inhabit interstitial spaces; the caddisfly *Micrasema*, which inhabits moss; and several "clean-water" stoneflies, caddisflies, and midges is consistent with this idea. The trend for certain taxa to have higher densities at test sites is also interpretable. For example, Tanytarsini midges (*Micropsectra/Tanytarsus* in this study), the small nemourid stoneflies (*Malenka/Zapada*), and the ubiquitous grazing mayfly *Baetis* have been previously observed in either light- or nutrient-enriched streams (Newbold et al. 1980, Hawkins et al. 1982, Gregory et al. 1987, Campbell and Doeg 1989). Unfortunately, we know too little about the habitat and feeding requirements of most of the taxa collected during this study to draw inferences regarding why they were more or less abundant at test sites.

The most rigorous way to evaluate the accuracy of these models would be to experimentally mimic the stressors of interest and determine if the taxa that are predicted to occur, but missing from sites, were sensitive to a stressor of interest. To our knowledge, such experiments have not been conducted; furthermore, so little basic knowledge exists regarding the habitat requirements and pollution tolerances of different taxa that it is difficult to determine if model predictions are consistent with the ecology of different taxa. Experimentally testing model predictions will be a rich and rewarding research endeavor for many years to come.

It is clear from our results that not all species within an order or family respond similarly to the same stress.

In many cases, substitutions occurred within the same taxonomic group that compensated for those taxa that decreased in abundance at test sites (Fig. 6). It follows then that assessment methods, like the EPT index (mayflies [Ephemeroptera], stoneflies [Plecoptera], and caddisflies [Trichoptera]; see Barbour et al. [1997]), that count taxa but ignore composition may be insensitive to changes in composition caused by a stressor unless that stressor also alters overall taxa richness (Ford 1989, Gray 1989).

CONCLUSIONS

The ultimate goal of biological water quality assessment is to provide an accurate measure of the biotic integrity of aquatic ecosystems that can be easily understood by managers and policy makers who may have little biological training and that is cost efficient (Schindler 1987, Karr 1991). We believe the predictive model approach meets each of these criteria. One of the main advantages of the predictive-model approach to bioassessment is that it is grounded in long-established empirical generalizations regarding the importance of specific environmental factors to aquatic invertebrates and one of the most robust ecological principles known: that of niche theory. We believe these models work well because stream invertebrate taxa differ in their habitat requirements, streams vary substantially in both regional and local habitat features, and stream invertebrate habitat relationships can be described by the models to predict what taxa should occur at new sites.

Although some have argued that multivariate approaches to resource assessment are too complicated to be used by anyone other than a few highly trained professionals or that they obscure important biological information (e.g., Gerritsen 1995, Fore et al. 1996, Karr and Chu 1999), we believe this is a specious argument when applied to predictive models. Model output is intuitive, logically derived, biologically meaningful, and easily interpreted. Furthermore, the specific taxa responsible for deviations of *O/E* values from unity can be identified, thus providing a basis for inferring the mode of action of a stressor. Model development does require both sophisticated biological and statistical knowledge (building a bridge or a car requires equally sophisticated knowledge), but routine model implementation and interpretation of output does not. These predictive models are now becoming available in computer programs that are easy to access and use. For example, the models described here run within a spreadsheet environment by executing a series of macros. To use these models, managers need only collect biota and data for a few environmental variables from suspect sites and enter these data into the computer using an interactive interface. The computations are automatic and "painless," and the outputs we have described are easily interpretable. As a further example of ease and economy of use, the Australian River As-

sessment System (AUSRIVAS) program has region-specific models loaded in a central computer that is accessible to users in every state and territory of Australia.[6] All the user need do is upload data from test sites, then call up and run the appropriate model.

We recognize that there may be significant psychological and perhaps institutional reluctance by the bioassessment community in the United States and elsewhere to embrace a predictive-model approach to assessment. For example, none of the 38 chapters in two recent books edited by Loeb and Spacie (1994) and Davis and Simon (1995) examine the predictive-model approach to monitoring and assessment, nor was it referred to by Fore et al. (1996), even though one of their objectives was to make a comparison between the major approaches to bioassessment. This lack of attention in the United States occurred even though the predictive-model method was first described in 1984 in a well-established and easily accessible journal (Furse et al. 1984, Wright et al. 1984, Moss et al. 1987).

Our results for streams in California show that the predictive model approach should be seriously evaluated in the United States and elsewhere for use in streams and other environments. We have already created models for littoral wetlands associated with small high-mountain lakes, and these models work well in quantifying the effects of exotic fish on benthic invertebrate assemblages (C. P. Hawkins, *unpublished data*). Furthermore, use of the predictive-model approach does not exclude use of multimetric indices as often practiced in the United States (e.g. Karr 1991, Karr and Chu 1999). In fact, these two approaches could complement one another in much the same way as indices are used in both Great Britain and Australia (Armitage et al. 1983, Chessman 1995). Essentially the same data are used in both approaches; where predictive models require additional data, they should be easily obtained with little extra cost. We believe the potential strengths of the predictive-model approach in satisfying the main criteria for biological monitoring listed by both Schindler (1987) and Karr (1991) are too compelling to ignore.

[6] See, for example, ⟨http://ausrivas.canberra.edu.au/ausrivas⟩

LITERATURE CITED

Armitage, P. D., D. Moss, J. F. Wright, and M. T. Furse. 1983. The performance of a new biological water quality score system based on macroinvertebrates over a wide range of unpolluted running-water sites. Water Research 17:333–347.

Bailey, R. C., M. G. Kennedy, M. Z. Dervish, and R. M. Taylor. 1998. Biological assessment of freshwater ecosystems using a reference condition approach: comparing predicted and actual benthic invertebrate communities in Yokon streams. Freshwater Biology 39:765–774.

Barbour, M. T., J. Gerritsen, G. E. Griffith, R. Frydenborg, E. McCarron, J. S. White, and M. L. Bastian. 1996. A framework for biological criteria for Florida streams using benthic macroinvertebrates. Journal of the North American Benthological Society 15:185–211.

Barbour, M. T., J. Gerritsen, and B. D. Snyder. 1997. Revision to Rapid Bioassessment Protocols for Use in Streams and Rivers. EPA 841-D-97-002, Office of Water, United States Environmental Protection Agency, Washington, D.C., USA.

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. Lewis, Boca Raton, Florida, USA.

Bartholow, J. M. 1989. Stream temperature investigations: field and analytic methods. Biological Report No. 89(17). U.S. Department of Interior, Fish and Wildlife Service, Research and Development, Washington, D.C., USA.

Belbin, L., and C. McDonald. 1993. Comparing three classification strategies for use in ecology. Journal of Vegetation Science 4:341–348.

Bisson, P. A., J. L. Nielson, R. A. Palmason, and L. E. Grove. 1981. A system of naming habitat types in small streams, with examples of habitat utilization by salmonids during low streamflow. Pages 62–73 *in* N. B. Armantrout, editor. Acquisition and utilization of aquatic habitat inventory information. Proceedings of a symposium held 28–30 October, 1981, Portland, Oregon, USA. Western Division, American Fisheries Society, Bethesda, Maryland, USA.

Caley, M. J., and D. Schluter. 1997. The relationship between local and regional diversity. Ecology 78:70–80.

Campbell, I. C., and T. J. Doeg. 1989. Impact of timber harvesting and production on streams: a review. Australian Journal of Marine and Freshwater Research 40:519–539.

Cao, Y., D. D. Williams, and N. E. Williams. 1998. How important are rare species in aquatic community ecology and bioassessment? Limnology and Oceanography 43:1403–1409.

Chessman, B. C. 1995. Rapid assessment of rivers using macroinvertebrates: a procedure based on habitat-specific sampling, family level identification and a biotic index. Australian Journal of Ecology 20:122–129.

Clarke, R. T., M. T. Furse, J. F. Wright, and D. Moss. 1996. Derivation of a biological quality index for river sites: comparison of the observed with the expected fauna. Journal of Applied Statistics 23:311–332.

Cornell, H. V. 1993. Unsaturated patterns in species assemblages: the role of regional processes in setting local species richness. Pages 243–252 *in* R. E. Ricklefs and D. Schluter, editors. Species diversity in ecological communities. University of Chicago Press, Chicago, Illinois, USA.

Davis, W. S., and T. P. Simon, editors. 1995. Biological assessment and criteria. Lewis, Boca Raton, Florida, USA.

DeShon, J. E. 1995. Development and application of the Invertebrate Community Index (ICI). Pages 217–243 *in* W. S. Davis and T. P. Simon, editors. Biological assessment and criteria. Lewis, Boca Raton, Florida, USA.

Faith, D. P., P. R. Minchin, and L. Belbin. 1987. Composi-

tional dissimilarity as a robust measure of ecological distance. Vegetatio **69**:57–68.

Ford, J. 1989. The effects of chemical stress on aquatic species composition and community structure. Pages 99–144 *in* S. A. Levine, M. A. Harwell, J. R. Kelly, and K. D. Kimball, editors. Ecotoxicology: problems and approaches. Springer–Verlag, New York, New York, USA.

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.

Furse, M. T., D. Moss, J. F. Wright, and P. D. Armitage. 1984. The influence of seasonal and taxonomic factors on the ordination and classification of running-water sites in Great Britain and on the prediction of their macro-invertebrate communities. Freshwater Biology **14**:257–280.

Gerritsen, J. 1995. Additive biological indices for resource management. Journal of the North American Benthological Society **14**:451–457.

Gibson, J. R., M. T. Barbour, J. B. Stribling, J. Gerritsen, and J. R. Karr. 1996. Biological criteria—technical guidance for streams and rivers, revised edition. EPA 822-B-96-001, Office of Water, U.S. Environmental Protection Agency, Washington, D.C., USA.

Gray, J. S. 1989. Effects of environmental stress on species rich assemblages. Biological Journal of the Linnean Society **37**:19–32.

Gregory, S. V., G. A. Lamberti, D. C. Erman, K. V. Koski, M. L. Murphy, and J. R. Sedell. 1987. Influence of forest practices on aquatic production. Pages 233–256 *in* E. O. Salo and T. W. Cundy, editors. Streamside management: forestry and fishery interactions. Contribution No. **57**.College of Forest Resources, University of Washington, Seattle, Washington, USA.

Hall, J. D., and N. J. Knight. 1981. Natural variation in abundance of salmonid populations in streams and its implications for design of impact studies. **EPA-600**:1–4. Environmental Protection Agency Project Summary Environmental Research Laboratory, Corvallis, Oregon, USA.

Hawkins, C. P., J. P. Dobrowolski, J. W. Feminella, J. N. Hogue, T. Hougaard, and D. Glatter. 1994. Cumulative watershed effects: an extensive analysis of responses by stream biota to watershed management. Final report, Cooperative Agreement No. **PSW-88-0011CA**, U.S. Forest Service, Pacific Southwest Forest and Range Experiment Station, Berkeley, California, USA.

Hawkins, C. P., J. N. Hogue, L. A. Decker, and J. W. Feminella. 1997. Channel morphology, water temperature, and assemblage structure of stream insects. Journal of the North American Benthological Society **16**:728–749.

Hawkins, C. P., J. L. Kershner, P. A. Bisson, M. D. Bryant, L. M. Decker, S. V. Gregory, D. A. McCullough, C. K. Overton, G. H. Reeves, R. J. Steedman, and M. K. Young. 1993. A hierarchical approach to classifying stream habitat features. Fisheries **18**:(6)3–12.

Hawkins, C. P., M. L. Murphy, and N. H. Anderson. 1982. Effects of canopy, substrate composition, and gradient on the structure of macroinvertebrate communities in Cascade range streams of Oregon. Ecology **63**:1840–1856.

Hawkins, C. P., M. L. Murphy, N. H. Anderson, and M. A. Wilzbach. 1983. Density of fish and salamanders in relation to riparian canopy and physical habitat in streams of the northwestern United States. Canadian Journal of Fisheries and Aquatic Sciences **40**:1173–1185.

Hawkins, C. P., and R. H. Norris. 2000. Effects of taxonomic resolution and use of subsets of the fauna on the performance of RIVPACS-type models. Pages 217–228 *in* J. F. Wright, D. W. Sutcliffe, and M. T. Furse, editors, Assessing the biological quality of freshwaters. RIVPACS and other

techniques. Freshwater Biological Association, Ambleside, England, UK.

Ide, F. P. 1935. University of Toronto Studies No. 39: the effect of temperature on the distribution of the mayfly fauna of the stream. Publications of the Ontario Fisheries Research Laboratory No. **50.** University of Toronto Press, Toronto, Ontario, Canada.

Karr, J. R. 1981. Assessment of biotic integrity using fish communities. Fisheries **6**:(6)21–27.

Karr, J. R. 1991. Biological integrity: a long neglected aspect of water resources management. Ecological Applications **1**:66–84.

Karr, J. R., and E. W. Chu. 1999. Restoring life in running waters: better biological monitoring. Island, Covelo, California, USA.

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.

Lenat, D. R. 1993. A biotic index for the southeastern United States: derivation and list of tolerance values, with criteria for assigning water quality ratings. Journal of the North American Benthological Society **12**:279–290.

Loeb, S. L., and A. Spacie, editors. 1994. Biological monitoring of aquatic systems. Lewis, Boca Raton, Florida, USA.

Marchant, R., L. A. Barmuta, and B. C. Chessman. 1995. Influence of sample quantification and taxonomic resolution on the ordination of macroinvertebrate communities from running waters in Victoria, Australia. Marine and Freshwater Research **46**:501–506.

Marchant, R., A. Hirst, R. H. Norris, R. Butcher, L. Metzeling, and D. Tiller. 1997. Classification and prediction of macroinvertebrate communities from running waters in Victoria, Australia. Journal of the North American Benthological Society **16**:664–681.

McCreadie, J. W., P. H. Alder, and J. F. Burger. 1997. Species assemblages of larval black flies (Diptera: Simuliidae): random or predictable? Journal of the North American Benthological Society **16**:760–770.

Moss, D., M. T. Furse, J. F. Wright, and P. D. Armitage. 1987. The prediction of the macro-invertebrate fauna of unpolluted running-water sites in Great Britain using environmental data. Freshwater Biology **17**:41–52.

Moss, D., J. F. Wright, M. T. Furse, and R. T. Clarke. 1999. A comparison of alternative techniques for prediction of the fauna of running-water sites in Great Britain. Freshwater Biology **41**:167–181.

Mount, J. F. 1995. California rivers and streams. University of California Press, Berkeley, California, USA.

Newbold, J. D., D. C. Erman, and K. B. Roby. 1980. Effects of logging on macroinvertebrates in streams with and without buffer strips. Canadian Journal of Fisheries and Aquatic Sciences **37**:1076–1085.

Newbold, J. D., B. W. Sweeney, and R. L. Vannote. 1994. A model for seasonal synchrony in stream mayflies. Journal of the North American Benthological Society **13**:3–18.

Norris, R. H. 1994. Rapid biological assessment, natural variability and selecting reference sites. Classification of rivers and environmental health indicators. Pages 129–166 *in* M. C. Uys, editor. Water Research Commission Report No. **TT 63/94.** Proceedings of a joint South African/Australian workshop, Cape Town, South Africa.

Norris, R. H. 1996. Predicting water quality using reference conditions and associated communities. Pages 32–52 *in* R. C. Bailey, R. H. Norris, and T. B. Reynoldson, editors. Study design and data analysis in benthic macroinvertebrate assessments of freshwater ecosystems using a reference site approach. Technical Workshop, North American Benthological Society, Kalispell, Montana, USA.

Norris, R. H., and K. R. Norris. 1995. The need for biological

assessment of water quality: Australian perspective. Australian Journal of Ecology **20**:1–6.

Ohio Environmental Protection Agency (EPA). 1987. Biological criteria for the protection of aquatic life. Volume II: users manual for biological field assessment of Ohio surface waters. Ohio EPA, Division of Water Quality Monitoring and Assessment, Surface Water Section, Columbus, Ohio, USA.

Omernik, K. M. 1987. Ecoregions of the conterminous United States. Annals of the Association of American Geographers **77**:118–125.

Parsons, M., and R. H. Norris. 1996. The effect of habitat-specific sampling on biological assessment of water quality using a predictive model. Freshwater Biology **36**:419–434.

Paulsen, S. G., and R. A. Linthurst. 1994. Biological monitoring in the environmental monitoring and assessment program. Pages 297–322 *in* S. L. Loeb and A. Spacie, editors. Biological monitoring of aquatic systems. Lewis, Boca Raton, Florida, USA.

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. **EPA/444/4-89-001.** Environmental Protection Agency, Washington, D.C.,USA.

Poff, N. L. 1997. Landscape filters and species traits: towards mechanistic understanding and prediction in stream ecology. Journal of the North American Benthological Society **16**:391–409.

Reynoldson, T. B., and J. L. Metcalfe-Smith. 1992. An overview of the assessment of aquatic ecosystem health using benthic invertebrates. Journal of Aquatic Ecosystem Health **1**:295–308.

Reynoldson, T. B., R. H. Norris, V. H. Resh, K. E. Day, and D. M. Rosenberg. 1997. The reference condition: a comparison of multimetric and multivariate approaches to assess water-quality impairment using benthic macroinvertebrates. Journal of the North American Benthological Society **16**:833–852.

Reynoldson, T. B., and J. F. Wright. 2000. The reference condition: problems and solutions. Pages 303–313 *in* J. F. Wright, D. W. Sutcliffe, and M. T. Furse, editors. Assessing the biological quality of freshwaters. RIVPACS and other techniques. Freshwater Biological Association, Ambleside, UK.

Rosenberg, D. M., and V. H. Resh, editors. 1993. Freshwater biomonitoring and benthic macroinvertebrates. Chapman and Hall, New York, New York, USA.

Schindler, D. W. 1987. Detecting ecosystem responses to anthropogenic stress. Canadian Journal of Fisheries and Aquatic Sciences **44**:6–25.

Southerland, M. T., and J. B. Stribling. 1995. Status of biological criteria development and implementation. Pages 81–96 *in* W. S. Davis and T. P. Simon, editors. Biological assessment and criteria. Lewis, Boca Raton, Florida, USA.

Strahler, A. N. 1964. Quantitative geomorphology of drainage basins and channel networks. Section 4-II *in* V. T. Chow, editor. Handbook of applied hydrology. McGraw–Hill, New York, New York, USA.

Vannote, R. L., G. W. Minshall, K. W. Cummins, J. R. Sedell, and C. E. Cushing. 1980. The river continuum concept. Canadian Journal of Fisheries and Aquatic Sciences **37**: 130–137.

Vannote, R. L., and B. W. Sweeney. 1980. Geographic analysis of thermal equilibria: a conceptual model for evaluating the effect of natural and modified thermal regimes on aquatic insect communities. The American Naturalist **115**: 667–695.

Vinson, M. R., and C. P. Hawkins. 1998. Biodiversity of stream insects: variation at local, basin, and regional scales. Annual Review of Entomology **43**:271–293.

Winterbourn, M. J., J. S. Rounick, and B. Cowie. 1981. Are New Zealand stream ecosystems really different? New Zealand Journal of Marine and Freshwater Research **15**:321–328.

Wolman, M. G. 1954. A method of sampling coarse river-bed material. Transactions of the American Geophysical Union **35**:951–956.

Wright, J. F. 1995. Development and use of a system for predicting the macroinvertebrate fauna in flowing waters. Australian Journal of Ecology **20**:181–197.

Wright, J. F., M. T. Furse, and P. D. Armitage. 1993. RIVPACS: a technique for evaluating the biological water quality of rivers in the UK. European Water Pollution Control **3**:15–25.

Wright, J. F., D. Moss, P. D. Armitage, and M. T. Furse. 1984. A preliminary classification of running-water sites in Great Britain based on macro-invertebrate species and the prediction of community type using environmental data. Freshwater Biology **14**:221–256.

APPENDIX

Further data regarding site classification and discriminant functions model performance are available in ESA's Electronic Data Archive: *Ecological Archives* E081-008.