Appendix 7-B

DOI: 10.1007/s00267-004-0035-8

A Quantitative Tool for Assessing the Integrity of Southern Coastal California Streams

PETER R. ODE*
Aquatic Bioassessment Laboratory
Water Pollution Control Laboratory
Department of Fish and Game
2005 Nimbus Road
Rancho Cordova, California 95670, USA

ANDREW C. REHN
Aquatic Bioassessment Laboratory
Chico State University Research Foundation
2005 Nimbus Road
Rancho Cordova, California 95670, USA

JASON T. MAY
3421 I Street 1
Sacramento, California, 95816, USA

ABSTRACT / We developed a benthic macroinvertebrate index of biological integrity (B-IBI) for the semiarid and populous southern California coastal region. Potential reference sites were screened from a pool of 275 sites, first with quantitative GIS landscape analysis at several spatial scales and then with local condition assessments (in-stream and riparian) that quantified stressors acting on study reaches. We screened 61 candidate metrics for inclusion in the B-IBI based on three criteria: sufficient range for scoring, responsiveness to watershed and reach-scale disturbance gradients, and minimal correlation with other responsive metrics. Final metrics included: percent collector-gatherer + collector-filterer individuals, percent noninsect taxa, percent tolerant taxa, Coleoptera richness, predator richness, percent intolerant individuals, and EPT richness. Three metrics had lower scores in chaparral reference sites than in mountain reference sites and were scored on separate scales in the B-IBI. Metrics were scored and assembled into a composite B-IBI, which was then divided into five roughly equal condition categories. PCA analysis was used to demonstrate that the B-IBI was sensitive to composite stressor gradients; we also confirmed that the B-IBI scores were not correlated with elevation, season, or watershed area. Application of the B-IBI to an independent validation dataset (69 sites) produced results congruent with the development dataset and a separate repeatability study at four sites in the region confirmed that the B-IBI scoring is precise. The SoCal B-IBI is an effective tool with strong performance characteristics and provides a practical means of evaluating biotic condition of streams in southern coastal California.

Assemblages of freshwater organisms (e.g., fish, macroinvertebrates, and periphyton) are commonly used to assess the biotic condition of streams, lakes, and wetlands because the integrity of these assemblages provides a direct measure of ecological condition of these water bodies (Karr and Chu 1999). Both multimetric (Karr and others 1986; Kerans and Karr 1994; McCormick and others 2001; Klemm and others 2003) and multivariate (Wright and others 1983; Hawkins and others 2000; Reynoldson and others 2001) methods have been developed to characterize biotic condition and to establish thresholds of ecological impairment. In both approaches, the ability to recognize degradation at study sites relies on an understanding of the organismal assemblages expected in the absence of disturbance. Thus, the adoption of a consistent and quantifiable method for defining reference condition is fundamental to any biomonitoring program (Hughes 1995).

Southern California faces daunting challenges in the conservation of its freshwater resources due to its aridity, its rapidly increasing human population, and its role as one of the world's top agricultural producers. In recent years, several state and federal agencies have become increasingly involved in developing analytical tools that can be used to assess the biological and physical condition of California's streams and rivers. For example, the US Environmental Protection Agency (EPA), the US Forest Service (USFS), and California's state and regional Water Quality Control Boards (WQCBs) have collected fish, periphyton and benthic macroinvertebrates (BMIs) from California streams and rivers as a critical component of regional water

KEY WORDS: Benthic macroinvertebrates; B-IBI; Biomonitoring; Mediterranean climate

Published online May 10, 2005.

*Author to whom correspondence should be addressed; *email:* pode@ospr.dfg.ca.gov

quality assessment and management programs. Together, these agencies have sampled BMIs from thousands of sites in California, but no analysis of BMI assemblage datasets based on comprehensively defined regional reference conditions has yet been undertaken. In the only other large-scale study within the state, Hawkins and others (2000) developed a predictive model of biotic integrity for third- to fourth-order streams on USFS lands in three montane regions in northern California. This ongoing effort (Hawkins unpublished) is an important contribution to bioassessment in the state, but the emphasis of this work has been concentrated on logging impacts within USFS lands. The lack of a broadly defined context for interpretation of BMI-based bioassessment remains the single largest impediment to the development of biocriteria for the majority of California streams and rivers. This article presents a benthic index of biotic integrity (B-IBI) for wadeable streams in southern coastal California assembled from BMI data collected in the region by the USFS, EPA, and state and regional WQCBs between 2000 and 2003.

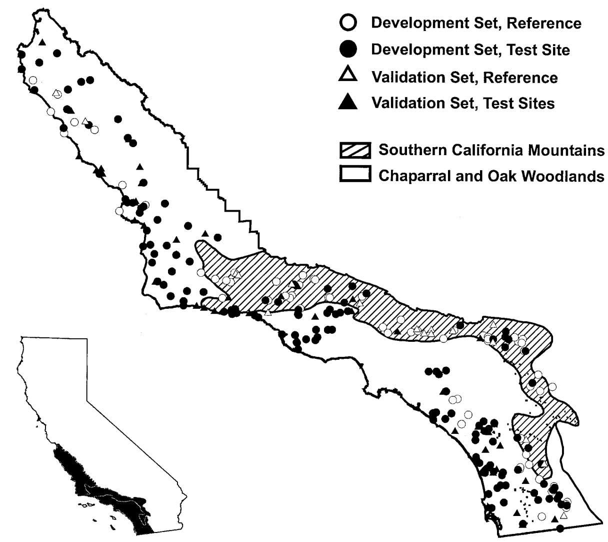

Figure 1. Map of study area showing the location of the study area within California, the distribution of test and reference sites and development and validation sites, and the boundaries of the two main ecoregions in the study area.

Methods

Study Area

The Southern Coastal California B-IBI (SoCal B-IBI) was developed for the region bounded by Monterey County in the north, the Mexican border in the south, and inland by the eastern extent of the southern Coast Ranges (Figure 1). This Mediterranean climate region comprises two Level III ecoregions (Figure 1; Omernik 1987) and shares a common geology (dominated by recently uplifted and poorly consolidated marine sediments) and hydrology (precipitation averages 10–20 in./year in the lower elevations and 20–30 in./year in upper elevations, reaching 30–40 in./year in the highest elevations and in some isolated coastal watersheds (Spatial Climate Analysis Service, Oregon State University, www.climatesource.com). The human population in the region was approximately 20 million in 2000 and is projected to exceed 28 million by 2025 (California Department of Finance, Demographic Research Unit, www.dof.ca.gov).

Field Protocols and Combining Datasets

The SoCal B-IBI is based on BMI and physical habitat data collected from 275 sites (Figure 1) using the 3 protocols described in the following subsections. Sites were sampled during base flow periods between April and October of 2000–2003.

California Stream Bioassessment Protocol (CSBP, 144 sites). Several of the regional WQCBs in southern coastal California have implemented biomonitoring programs in their respective jurisdictions and have collected BMIs according to the CSBP (Harrington 1999). At CSBP sites, three riffles within a 100-m reach were randomly selected for sampling. At each riffle, a transect was established perpendicular to the flow, from which three separate areas of 0.18 m^2 each were sampled upstream of a 0.3-m-wide D-frame net and composited by transect. A total of 1.82 m^2 of substrate was sampled per reach and 900 organisms were subsampled from this material (300 organisms were processed separately from each of 3 transects). Water chemistry data were collected in accordance with the protocols of the different regional WQCBs (Puckett 2002) and qualitative physical habitat characteristics were measured according to Barbour and others (1999) and Harrington (1999).

USFS (56 sites). The USFS sampled streams on national forest lands in southern California in 2000 and 2001 using the targeted riffle protocol of Hawkins and others (2001). All study reaches were selected nonrandomly as part of a program to develop an interpretive (reference) framework for the results of stream biomonitoring studies on national forests in California. BMIs were sampled at study reaches (containing at least four fast-water habitat units) by disturbing two separate 0.09-m^2 areas of substrate upstream of a 0.3-m-wide D-frame net in each of four separate fast-water units; a total of 0.72 m^2 was disturbed and all sample material from a reach was composited. Field crews used a combination of qualitative and quantitative measures to collect physical habitat and water chemistry data (Hawkins and others 2001). A 500-organism subsample was processed from the composite sample and identified following methods described by Vinson and Hawkins (1996).

Environmental Monitoring and Assessment Program (EMAP, 75 sites). The EPA sampled study reaches in southern coastal California from 2000 through 2003 as part of its Western EMAP pilot project. A sampling reach was defined as 40 times the average stream width at the center of the reach, with a minimum reach length of 150-m and maximum length of 500-m. A BMI sample was collected at each site using the USFS methodology described earlier (Hawkins and others 2001) in addition to a standard EMAP BMI sample (not used in this analysis). A 500-organism subsample was processed in the laboratory according to EMAP standard taxonomic effort levels (Klemm and others 1990). Water chemistry samples were collected from the midpoint of each reach and analyzed using EMAP protocols (Klemm and Lazorchak 1994). Field crews recorded physical habitat data using EPA qualitative methods (Barbour and others 1999) and quantitative methods (Kaufmann and others 1999).

As part of a methods comparison study, 77 sites were sampled between 2000 and 2001 with both the CSBP and USFS protocols. The two main differences between the methods are the area sampled and the number of organisms subsampled (discussed earlier). To determine the effect of sampling methodology on assessment of biotic condition, we compared the average difference in a biotic index score between the two methods at each site. Biotic index scores were computed with seven commonly used biotic metrics (taxonomic richness, Ephemoptera, Plecoptera, and Trichoptera (EPT) richness, percent dominant taxon, sensitive EPT individuals, Shannon diversity, percent intolerant taxa, and percent scraper individuals) according to the following equation:

$$Score = \sum (x_i - \bar{x})/sem_i$$

where x_i is the site value for the ith metric, x is the overall mean for the ith metric, and SEM_i is the standard error of the mean for the ith metric. A score of zero is the mean value.

Because USFS-style riffle samples were collected at all EMAP sites, only two field methods were combined in this study. All EMAP and CSBP samples were collected and processed by the California Department of Fish and Game's Aquatic Bioassessment Laboratory (ABL) and all USFS samples were processed by the US Bureau of Land Management's Bug Lab in Logan, Utah. Taxonomic data from both labs were combined in an MS Access© database application that standardized BMI taxonomic effort levels and metric calculations, allowing us to minimize any differences between the two labs that processed samples. Taxonomic effort followed standards defined by the California Aquatic Macroinvertebrate Laboratory Network (CAMLnet 2002; www.dfg.ca.gov/cabw/camlnetste.pdf). Sites with fewer than 450 organisms sampled were omitted from the analyses.

Screening Reference Sites

We followed an objective and quantitative reference site selection procedure in which potential reference sites were first screened with quantitative Geographical Information System (GIS) land-use analysis at several spatial scales and then local condition assessments (instream and riparian) were used to quantify stressors acting within study reaches. We calculated the proportions of different land-cover classes and other measures of human activity upstream of each site at four spatial scales that give unique information about potential stressors acting on each site: (1) within polygons delimiting the entire watershed upstream of each sampling site, (2) within polygons representing local regions (defined as the intersection of a 5-km-radius circle around each site and the primary watershed polygon), (3) within a 120-m riparian zone on each side of all streams within each watershed, and (4) within a 120-m riparian zone in the local region. We used the ArcView® (ESRI 1999) extension ATtILA (Ebert and Wade 2002) to calculate the percentage of various land-cover classes (urban, agriculture, natural, etc.) and other measures of human activity (population density, road density, etc.) in each of the four spatial areas defined for each site. Two satellite imagery datasets from the mid-1990s were combined for the land-cover analyses: California Land Cover Mapping & Monitoring Program (LCMMP) vegetation data (CalVEG) and a recent dataset produced by the Central Coast Watershed Group (Newman and Watson 2003). Population data were derived from the 2000 migrated TIGER dataset (California Department of Forestry and Fire Protection, www.cdf.ca.gov). Stream layers were obtained from the US Geological Survey (USGS) National Hydrography Dataset (NHD). The road network was obtained from the California Spatial Information Library (CaSIL, gis.ca.gov) and elevation was based on the USGS National Elevation Dataset (NED). Frequency histograms of land-use percentages for all sites were used to establish subjective thresholds for elim-

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

Stressor metric	Definition	Threshold
N_index_L	Percentage of natural land use at the local scale	≤ 95%
Purb_L	Percental of urban land use at the local scale	> 3%
Pagt_L	Percentage of total agriculture at the local scale	> 5%
Rddens_L	Road density at the local scale	> 2.0 km/km^2
PopDens_L	Population density (2000 census) at the local scale	> 150 indiv./km^2
N_index_W	Percentage of natural landuse at the watershed scale	≤ 95%
Purb_W	Percentage of urban landuse at the watershed scale	> 5%
Pagt_W	Percentage of total agriculture at the watershed scale	> 3%
Rddens_W	Road density at the watershed scale	> 2.0 km/km^2
PopDens_W	Population density (2000 census) at the watershed scale	> 150 indiv./km^2

inating sites from the potential reference pool (Table 1). Sites were further screened from the reference pool on the basis of reach-scale conditions (obvious bank instability or erosion/ sedimentation problems, evidence of mining, dams, grazing, recent fire, recent logging).

Eighty-eight sites passed all the land-use and local condition screens and were selected as reference sites, leaving 187 sites in the test group. We randomly divided the full set of sites into a development set (206 sites total: 66 reference/140 test) and a validation set (69 sites total: 22 reference/47 test). The development set was used to screen metrics and develop scoring ranges for component B-IBI metrics; the validation set was used for an independent evaluation of B-IBI performance.

Screening Metrics and Assembling the B-IBI

Sixty-one metrics were evaluated for possible use in the SoCal B-IBI (Table 2). A multistep screening process was used to evaluate each metric for (1) sufficient range to be used in scoring, (2) responsiveness to watershed-scale and reach-scale disturbance variables, and (3) lack of correlation with other responsive metrics.

Pearson correlations between all watershed-scale and reach-scale disturbance gradients were used to define the smallest suite of independent (nonredundant) disturbance variables against which to test biological metric response. Disturbance variables with correlation coefficients $|r| \geq 0.7$ were considered redundant. Responsiveness was assessed using visual inspection of biotic metric versus disturbance gradient scatterplots and linear regression coefficients. Metrics were selected as responsive if they showed either a linear or a "wedge-shaped" relationship with disturbance gradients. Biological metrics often show a "wedge-shaped" response rather than a linear response to single disturbance gradients because the single gradient only defines the upper boundary of the biological response; other independent disturbance gradients and natural limitations on species distributions might result in lower metric values than expected from response to the single gradient. Biotic metrics and disturbance gradients were log-transformed when necessary to improve normality and equalize variances. Metrics that passed the range and responsiveness tests were tested for redundancy. Pairs of metrics with product–moment correlation coefficients $|r| \geq 0.7$ were considered redundant and the least responsive metric of the pair was eliminated.

Scoring ranges were defined for each metric using techniques described in Hughes and others (1998), McCormick and others (2001), and Klemm and others (2003). Metrics were scored on a 0–10 scale using statistical properties of the raw metric values from both reference and nonreference sites to define upper and lower thresholds. For positive metrics (those that increase as disturbance decreases), any site with a metric value equal to or greater than the 80th percentile of reference sites received a score of 10; any site with a metric value equal to or less than the 10th percentile of the nonreference sites received a score of 0; these thresholds were reversed for negative metrics (20th percentile of reference and 90th percentile of nonreference). In both cases, the remaining range of intermediate metric values was divided equally and assigned scores of 1 through 9. Before assembling the B-IBI, we tested whether any of the final metrics were significantly different between chaparral and mountain reference sites in the southern California coastal region, in which case they would require separate scoring ranges in the B-IBI. Finally, an overall B-IBI score was calculated for each site by summing the constituent metric scores and adjusting the B-IBI to a 100-point scale.

Table 2. The 61 BMI metrics screened for use in the SoCal IBI

Candidate metrics	Disturbance variables: U_index_W	Pagt_W	Purb_L	RdDens_L	Channel Alteration	Bank Stability	Percent Fines	Total Dissolved Solids	Total Phosphorus	Total Nitrogen	Range Test
Taxonomic group metrics											
Coleoptera richness*	M	w	M	S	S	—	—	—	—	—	P
Crustacea + Mollusca richness	—	—	—	—	—	—	—	—	—	—	F
Diptera richness	—	—	—	—	—	—	—	—	—	—	P
Elmidae richness	w	—	w	M	S	M	S	M	—	—	F
Ephemerellidae richness					S	M	M	S	—	M	F
Ephemeroptera richness	S	S	M	S	w	M	S	S	—	S	P
EPT richness*	S	S	S	S	S	S	S	S	—	S	P
Hydropsychidae richness	—	—	w	—	S	—	—	—	—	—	F
Percent Amphipoda individuals	—	—	—	—	—	—	—	—	—	—	P
Percent Baetidae individuals	—	—	—	—	w	—	—	—	—	—	P
Percent Chironomidae individuals	—	—	—	—	—	—	—	—	M	—	P
Percent Corbicula individuals	—	—	—	—	—	—	—	—	—	—	P
Percent Crustacea individuals	—	—	—	—	—	—	—	—	—	—	P
Percent Diptera individuals	—	w	—	—	—	—	—	—	—	—	P
Percent Elmidae individuals	—	—	—	w	M	S	S	w	—	M	P
Percent Ephemeroptera individuals	—	w	w	—	M	w	—	—	—	—	P
Percent EPT individuals	—	—	M	M	M	M	—	—	—	—	P
Percent Gatropoda individuals	—	—	—	w	—	—	—	—	—	—	P
Percent Glossosomatidae individuals	—	—	—	—	w	—	—	—	—	M	F
Percent Hydropsychidae individuals	—	—	—	M	w	M	—	—	—	—	P
Percent Hydroptilidae individuals	—	—	—	M	—	w	—	—	—	—	F
Percent Mollusca individuals	—	—	—	w	w	—	—	—	—	—	P
Percent non-*Baetis/Fallceon* Ephemeroptera individuals	w	w	—	M	w	M	—	—	w	—	P
Percent non-*Hydropsyche* Hydropsychidae individuals	—	—	—	M	w	w	—	—	—	—	F
Percent non-*Hydropsyche/Cheumatopsyche* Trichoptera individuals	w	w	—	M	w	M	M	w	—	—	P
Percent non-insect Taxa*	M	w	M	M	w	—	—	—	w	M	F
Percent Oligochaeta individuals	—	—	—	—	w	—	—	—	—	—	P
Percent Perlodidae individuals	—	—	—	w	w	—	—	—	—	—	F
Percent Plecoptera individuals	—	—	—	M	M	M	M	M	w	S	P
Percent Rhyacophilidae individuals	—	—	—	w	S	S	w	—	—	M	F
Percent Simuliidae individuals	—	w	—	w	S	w	—	—		—	P
Percent Trichoptera	w	—	—	M	M	M	M	w	w	—	P
Plecoptera richness	M	S	w	M	w	w	M	S	—	S	F
Total taxa richness	M	M	w	S	w	w	w	w	w	M	P
Trichoptera richness	S	S	S	S	S	M	S	w	—	w	P

Table 2. Continued.

Candidate metrics	Disturbance variables: U_index_W	Pagt_W	Purb_L	RdDens_L	Channel Alteration	Bank Stability	Percent Fines	Total Dissolved Solids	Total Phosphorus	Total Nitrogen	Range Test
Functional feeding metrics											
Collector (filterers) richness	w	—	M	S	S	M	w	—	—	—	F
Collector (gatherers) richness	—	—	—	—	—	—	—	—	—	w	P
Percent collector (filterer) + collector (gatherer) individuals*	M	—	—	S	—	w	—	M	w	M	P
Percent collector (filterer) individuals	—	—	—	w	M	M	w	—	—	—	P
Percent collector (gatherer) individuals	—	—	—	w	M	—	—	w	M	w	P
Percent predator individuals	—	—	—	w	M	—	—	—	—	—	P
Percent scraper individuals	w	w	—	M	M	w	w	—	—	—	P
Percent scraper minus snails individuals	—	—	—	w	—	w	—	—	—	—	P
Percent shredder individuals	—	—	—	w	w	—	—	—	—	—	P
Predator richness*	S	S	w	M	w	—	—	S	—	M	P
Scraper richness	S	M	M	S	S	S	S	S	—	S	P
Shredder richness	M	M	—	M	S	—	—	—	—	M	F
Tolerance metrics											
Average tolerance value	M	w	w	S	w	—	M	—	—	w	P
Intolerant EPT richness	M	w	w	M	S	—	S	S	—	S	P
Intolerant taxa richness	M	w	w	M	S	M	S	S	—	S	P
Percent intolerant Diptera individuals	—	—	—	—	—	—	—	—	—	—	F
Percent intolerant individuals*	M	w	—	M	S	M	M	S	—	M	P
Percent intolerant scraper individuals	—	—	—	w	M	w	w	w	—	—	P
Percent of intolerant Ephemeroptera individuals	—	—	—	w	w	—	w	w	—	—	P
Percent of intolerant Trichoptera individuals	—	w	—	—	w	w	w	w	—	—	P
Percent sensitive EPT individuals	w	w	—	M	M	M	M	M	w	M	P
Percent tolerant individuals	—	—	—	—	—	—	w	w	—	—	P
Percent tolerant taxa*	w	—	w	M	—	—	—	w	—	M	P
Tolerant taxa richness	—	—	—	—	—	M	—	—	—	—	P
Others											
Percent dominant taxon	—	—	—	—	—	—	—	—	—	—	P
Shannon Diversity Index	w	w	w	M	M	w	—	w	w	w	P

Note: Each metric is indicated as having either no response (—), weak response (w), moderate response (M), or strong response (S) to each of eleven minimally correlated disturbance variables and whether each metric passed (P) or failed (F) the range test. The final seven minimally correlated metrics are indicated with an asterisk (*).

Table 3. Scoring ranges for seven component metrics in the SoCal B-IBI

Metric score	Coleoptera taxa (all sites)	EPT taxa		Predator taxa (all sites)	% Collector individuals		% Intolerant individuals		% Noninsect taxa (all sites)	% Tolerant taxa (all sites)
		6	8		6	8	6	8		
10	>5	>17	>18	>12	0–59	0–39	25–100	42–100	0–8	0–4
9		16–17	17–18	12	60–63	40–46	23–24	37–41	9–12	5–8
8	5	15	16	11	64–67	47–52	21–22	32–36	13–17	9–12
7	4	13–14	14–15	10	68–71	53–58	19–20	27–31	18–21	13–16
6		11–12	13	9	72–75	59–64	16–18	23–26	22–25	17–19
5	3	9–10	11–12	8	76–80	65–70	13–15	19–22	26–29	20–22
4	2	7–8	10	7	81–84	71–76	10–12	14–18	30–34	23–25
3		5–6	8–9	6	85–88	77–82	7–9	10–13	35–38	26–29
2	1	4	7	5	89–92	83–88	4–6	6–9	39–42	30–33
1		2–3	5–6	4	93–96	89–94	1–3	2–5	43–46	34–37
0	0	0–1	0–4	0–3	97–100	95–100	0	0–1	47–100	38–100

Note: Three metrics have separate scoring ranges for the two Omernik Level III ecoregions in southern coastal California region (6 = chaparral and oak woodlands, 8 = Southern California mountains).

Validation of B-IBI and Measurement of Performance Characteristics

To test whether the distribution of B-IBI scores in reference and test sites might have resulted from chance, we compared score distributions in the development set to those in the validation set. We also investigated a separate performance issue that ambient bioassessment studies often neglect: spatial variation at the reach scale. Although our use of a validation dataset tests whether the B-IBI scoring range is repeatable (Fore and others 1996; McCormick and others 2001), we designed a separate experiment to explicitly measure index precision. Four sites were resampled in May 2003. At each site, nine riffles were sampled following the CSBP, and material from randomly selected riffles was combined into three replicates of three riffles each. B-IBI scores were then calculated for each replicate. Variance among these replicates was used to calculate the minimum detectable difference (MDD) between two B-IBI scores based on a two-sample *t*-test model (Zar 1999). The index range can be divided by the MDD to estimate the number of stream condition categories detectable by the B-IBI (Doberstein and others 2000; Fore and others 2001).

Results

Combining Datasets

Unmodified CSBP samples (900 count) had significantly higher biotic condition scores ($t = -6.974$, $P < 0.0001$) than did USFS samples (500 count). However, there was no difference in biotic condition scores between USFS samples and CSBP samples that were randomly subsampled to reduce the 900 count to 500 ($t = -0.817$, $P = 0.416$). Thus, data from both targeted-riffle protocols were combined in B-IBI development.

Selected Metrics

Ten nonredundant stressor gradients were selected for metric screening: percent watershed unnatural, percent watershed in agriculture, percent local watershed in urban, road density in local watershed, qualitative channel alteration score, qualitative bank stability score, percent fine substrates, total dissolved solids, total nitrogen, and total phosphorous. Twenty-three biotic metrics that passed the first two screens (range and dose response) were analyzed for redundancy with Pearson product–moment correlation, and a set of seven minimally correlated metrics was selected for the B-IBI: percent collector-gatherer + collector-filterer individuals (% collectors), percent noninsect taxa, percent tolerant taxa, Coleoptera richness, predator richness, percent intolerant individuals, and EPT richness (Table 3). All metrics rejected as redundant were derived from taxa similar to those of selected metrics, but they had weaker relationships with stressor gradients. Dose–response relationships of the selected metrics to the 10 minimally correlated stressor variables are shown in Figure 2 and reasons for rejection or acceptance of all metrics are listed in Table 2. Regression coefficients were significant at the $P \leq 0.0001$ level among all seven selected metrics and at least two stressor gradients: percent watershed unnatural and road density in local watershed (Table 4). The final seven metrics included several metric types: richness, composition, tolerance measures, and func-

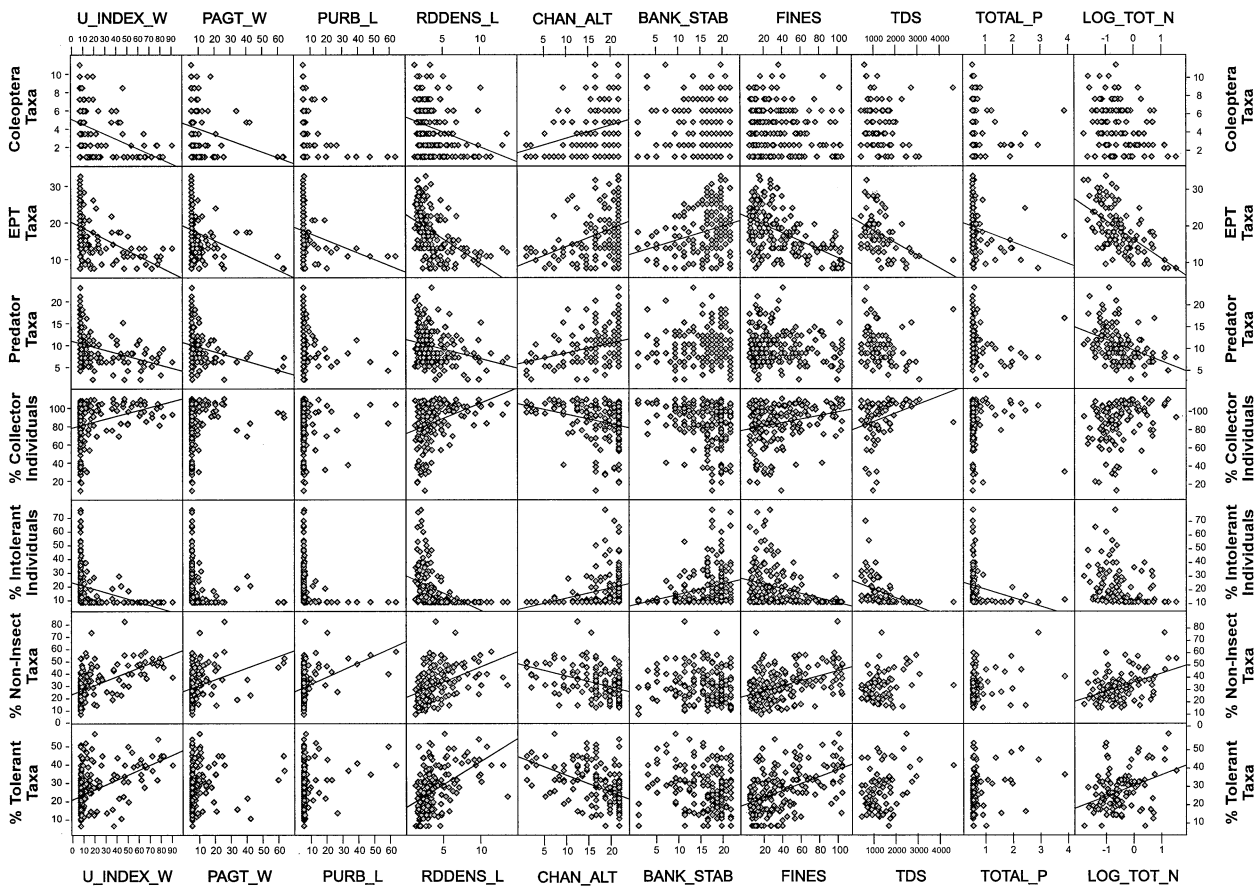

Figure 2. Scatterplots of dose–response relationships among 10 stressor gradients and 7 macroinvertebrate metrics (lines represent linear "best-fit" relationships; see text for abbreviations).

Table 4. Significance levels of linear regression relationships among 10 stressor metrics and 7 biological metrics

Metric	Coleoptera taxa	EPT taxa	Predator taxa	% Collector individuals	% Intolerant individuals	% Noninsect taxa	% Tolerant taxa
Bank Stability	0.813	**<0.0001**	0.3132	0.0009	0.0001	0.1473	0.0013
Fines	0.0017	**<0.0001**	0.0171	**0.0003**	**<0.0001**	**<0.0001**	**<0.0001**
Chan_Alt	**<0.0001**	**<0.0001**	**<0.0001**	**0.0003**	**<0.0001**	**<0.0001**	**<0.0001**
Log_U_Index_W	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**
Log_PAgT_W	**0.0007**	**<0.0001**	**0.0004**	0.0054	0.0014	**<0.0001**	0.0012
Log_PUrb_L	0.0367	**0.0007**	0.0344	0.6899	0.0045	**0.0002**	0.0215
Log RdDens_L	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**
Log_TDS	0.0094	**<0.0001**	0.0035	**0.0005**	**<0.0001**	0.0271	0.004
Log_Tot_N	0.0019	**<0.0001**	**<0.0001**	0.0078	0.0019	**<0.0001**	**<0.0001**
Log_Tot_P	0.062	**<0.0001**	0.0085	0.0162	**0.0001**	0.0018	0.0059

Note: Significant *P*-values corrected for 70 simultaneous comparisons ($P < 0.0007$) are highlighted in bold. Abbreviations are defined in Table 1 and in the text.

tional feeding groups. Because there are only seven metrics in the B-IBI, final scores calculated using this IBI are multiplied by 1.43 to adjust the scoring range to a 100-point scale.

The B-IBI scores were lower in chaparral reference sites than in mountain reference sites when calculated using unadjusted metric scores (Mann–Whitney *U*-test; $P = 0.02$). Although none of the final seven metrics

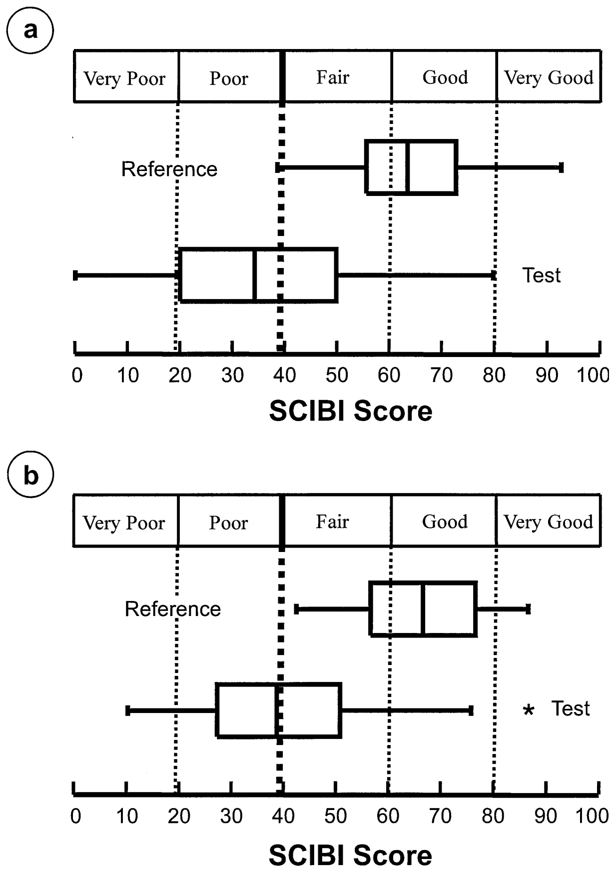

Figure 3. Box plots of B-IBI site scores for reference and test groups showing B-IBI scoring categories: (a) development sites and (b) validation sites. Dotted lines indicate condition category boundaries and heavy dotted lines indicate impairment thresholds.

were significantly different between chaparral reference sites and mountain reference sites at the $P = 0.05$ level ($P < 0.007$ after Bonferroni correction), scores for three metrics (EPT richness, percent collector-gatherer + collector-filterer individuals, and percent intolerant individuals) were substantially lower in chaparral reference sites than in mountain reference sites. We adjusted for this difference by creating separate scoring scales for the three metrics in the two ecoregions (Table 3). There was no difference in B-IBI scores between reference sites in the two ecoregions after the adjustment (Mann–Whitney *U*-test, $P = 0.364$).

Validation of B-IBI and Measurement of Performance Characteristics

The distribution of B-IBI scores at reference and nonreference sites was nearly identical between the development and validation data sets (Figure 3), indicating that our characterization of reference conditions and subsequent B-IBI scoring was repeatable and not likely due to chance. Based on a two-sample *t*-test model (setting $\alpha = 0.05$ and $\beta = 0.20$), the MDD for the SoCal IBI is 13.1. Thus, we have an 80% chance of detecting a 13.1-point difference between sites at the $P = 0.05$ level. Dividing the 100-point B-IBI scoring range by the MDD indicates that the SoCal B-IBI can detect a maximum of seven biological condition categories, a result similar to or more precise than other recent estimates of B-IBI precision (Barbour and others 1999; Fore and others 2001). We used a statistical criterion (two standard deviations below the mean reference site score) to define the boundary between "fair" and "poor" conditions, thereby setting B-IBI = 39 as an impairment threshold. The scoring range below 39 was divided into two equal condition categories, and the range above 39 was divided into three equal condition categories: 0–19 = "very poor", 20–39 = "poor", 40–59 = "fair", 60-79 = "good", and 80–100 = "very good" (Figure 3).

We ran two principle components analyses (PCAs) on the environmental stressor values used for testing metric responsiveness: 1 that included all 275 sites for which we calculated 4 watershed scale stressor values and another based on 124 sites for which we had measurements of 9 of the 10 minimally correlated stressor variables. We plotted B-IBI scores as a function of the first multivariate stressor axis from each PCA. We log-transformed percent watershed unnatural, percent watershed in agriculture, percent local watershed in urban, road density in local watershed, total nitrogen, and total phosphorous. Only PCA Axis 1 was significant in either analysis, having eigenvalues larger than those predicted from the broken-stick model (McCune and Grace 2002). In both PCAs, the B-IBI score decreased with increasing human disturbance (Figure 4) and was correlated (Spearman ρ) with PCA Axis 1 ($r = -0.652$, $P < 0.0001$ for all 275 sites; $r = -0.558$, $P \leq 0.0001$ for 124 sites). In the analysis of all 275 sites, all 4 watershed-scale stressors had high negative loadings, with percent watershed unnatural and local road density being the highest (Figure 5a). In the analysis of 124 sites, percent watershed unnatural, percent watershed in agriculture, and local road density had the highest negative loadings on the first axis, and channel alteration had the highest positive loading (Figure 4b).

Finally, we found no relationship between B-IBI scores and ecoregion (Mann–Whitney U, $P = 0.364$), Julian date ($R^2 = 0.01$, $P = 0.349$), watershed area ($R^2 = 0.002$, $P = 0.711$), or elevation ($R^2 = 0.01$, $P = 0.349$), indicating that the B-IBI scoring is robust with respect to these variables (Figure 5). Our ecoregion scoring adjustment probably corrects for the

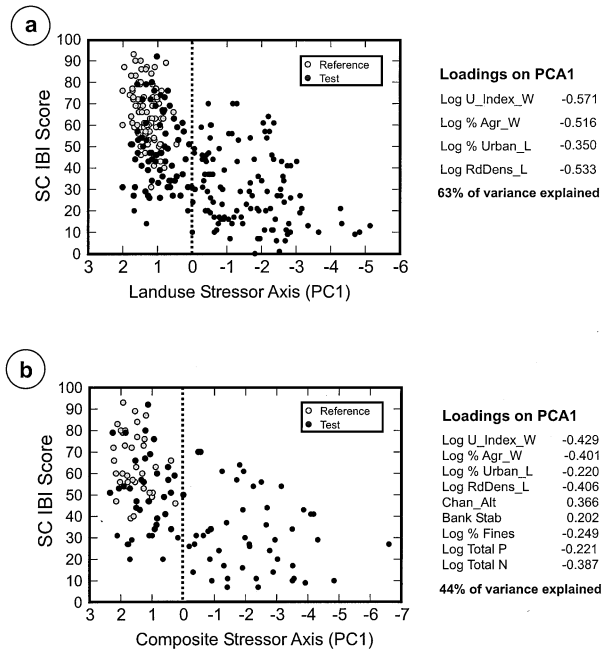

Figure 4. Scatterplots of SoCal B-IBI scores against two composite stressor axes from PCA: (a) values for all 275 sites; composite axis includes 4 land-use gradients; (b) values for 124 sites; composite axis includes 9 local and watershed scale stressor gradients.

strongest elevation effects, but there is no evidence that B-IBI scores are related to elevation differences within each ecoregion.

Discussion

The SoCal B-IBI is the most comprehensive assessment to date of freshwater biological integrity in California. As in other Mediterranean climate regions, the combination of aridity, geology, and high-amplitude cycles of seasonal flooding and drying in southern coastal California makes its streams and rivers particularly sensitive to disturbance (Gasith and Resh 1999). This sensitivity, coupled with the burgeoning human population and vast conversion of natural landscapes to agriculture and urban areas, has made it the focus of both state and federal attempts to maintain the ecological integrity of these strained aquatic resources.

Unfortunately, growing interest in biomonitoring is unmatched by financial resources available for this monitoring. Thus, combination of data among programs is very desirable, although this goal is rarely achieved in practice. We demonstrated that macroinvertebrate bioassessment data from multiple agencies could be successfully combined to produce a regional index that is useful to all agencies involved. This index is easy to apply, its fundamental assumptions are transparent, it provides precise condition assessments, and it is demonstrated to be responsive to a wide range of anthropogenic stressors. The index can also be applied throughout a long index period (mid-spring to mid-fall): Just as biotic factors tend to have more influence on assemblage structure during the summer dry period of Mediterranean climates than during the wet season when abiotic factors dominate (Cooper and others 1986; Gasith and Resh 1999), it is likely that our biotic index is more sensitive to anthropogenic stressors during the summer dry period. Because of these qualities, we expect the SoCal B-IBI to be a practical management tool for a wide range of water quality applications in the region.

This B-IBI is a regional adaptation of an approach to biotic assessment developed by Karr (1981) and subsequently extended and refined by many others (Kerans and Karr 1994; Barbour and others 1996; Fore and others 1996; Hughes and others 1998). We drew heavily upon recent refinements in multimetric index methodology that improve the objectivity and defensibility of these indices (McCormick and others 2001; Klemm and others 2003). A central goal of bioassessment is to select metrics that maximize the detection of anthropogenic stress while minimizing the noise of natural variation. One of the most important recent advances in B-IBI methods is the emphasis on quantitative screening tools for selecting appropriate metrics. We also minimized sources of redundancy in the analysis: (1) between watershed and local-scale stressor gradients for dose–response screening of biotic metrics and (2) in the final selection of metrics. The former guards against a B-IBI that is biased toward a set of highly correlated stressors and is, therefore, of limited sensitivity; the latter assures a compact B-IBI with component metrics that contribute independent information about stream condition. Combined with an assessment of responsiveness to specific regional disturbance gradients, these screening tools minimize the variability of B-IBI scores and improve its sensitivity.

The seven component metrics used in this B-IBI are similar to those selected for other B-IBIs (DeShon 1995; Barbour and others 1995, 1996; Fore and others 1996; Klemm and others 2003), but some of the metrics are either unique or are variations on other commonly used metrics. Like Klemm and others (2003), we found noninsect taxa to be responsive to human stressors, but richness was more responsive than percent of individuals. Some authors have separated the EPT metric into two or three metrics based on its component orders because the orders provided unique signals (Clements 1994; Fore and others 1996; Klemm

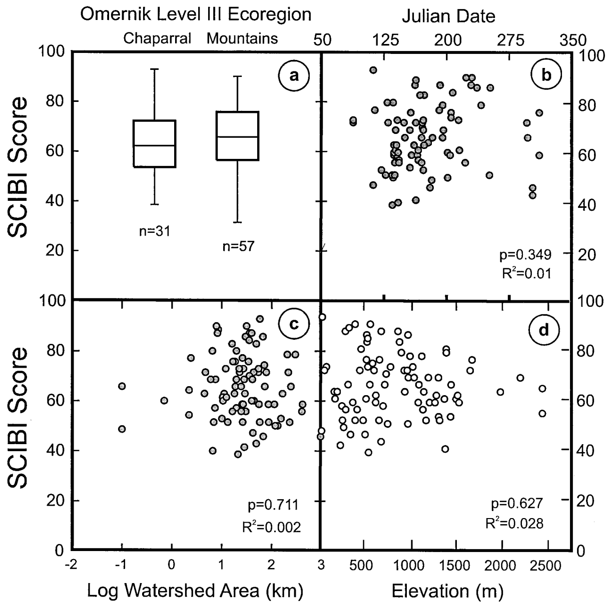

Figure 5. Relationship between B-IBI scores at 88 reference sites and (a) Omernik Level III ecoregion, (b) Julian date, (c) log watershed area, and (d) elevation.

and others 2003), but we found very similar patterns in these orders' response to various stressors we measured. To our knowledge, Coleoptera richness has not previously been included in a B-IBI, but beetle taxa might be a good indicator of the effects of fine sediments at impaired sites in this region (Brown 1973). A recent study of benthic assemblages in North Africa noted a high correspondence between EPT and EPTC (EPT + Coleoptera) (Beauchard and others 2003), but these orders were not highly correlated in our dataset. Feeding groups appear less often in B-IBIs than other metric types (Klemm and others 2003), but they were represented by two metrics in this B-IBI: predator richness and percent collectors (gatherers and filterers combined). Scraper richness was also responsive, but was rejected here because it was highly correlated with EPT richness.

The SoCal IBI should prove useful as a foundation for state and regional ambient water quality monitoring programs. Because the 75 EMAP sites were selected using a probabilistic statistical design, it will also be possible to use those samples to estimate the percentage of stream miles that are in "good", "fair", and "poor" condition in the southern California coastal region. These condition estimates, combined with stressor association techniques, have great potential to serve as a scientifically defensible basis for allocating precious monitoring resources in this region.

Acknowledgments

Major funding for this study was provided by David Gibson of the San Diego WQCB and Joseph Furnish of the USFS, with supplemental funding from other regional WQCBs in southern California. We thank James Harrington and other staff of California Department of Fish and Game's Aquatic Bioassessment Laboratory for efforts that contributed to the success of this project. Biologists Mike Dawson, Shawn McBride, and Jennifer York collected much of the field data; taxonomists Dan Pickard, Doug Post, Brady Richards, and Joe Slusark performed most of the taxonomy, except for the USFS samples, which were processed by the Bug Lab at Utah State University. Dan Heggem (US EPA, Landscape Ecology Branch), Mark Rosenberg (California Department of Forestry), and Mark Angelo (Central Coast WQCB) helped identify appropriate GIS land-use layers, and Glenn Sibbald (CDFG) assisted with watershed delineation. Mary Adams (Central Coast WQCB) collected invertebrate samples and provided local site condition data for sites in the Central Coast region. This manuscript was greatly improved by comments from Larry Brown, Joseph Furnish, Robert Hughes, Leska Fore, and three anonymous reviewers.

Literature Cited

Barbour, M. T., J. B. Stribling, and J. R Karr. 1995. The multimetric approach for establishing biocriteria and measuring biological conditions. Pages 63–77 *in* W. S. Davis, T. Simon (eds.), Biological assessment and criteria: tools for water resource planning and decision making. Lewis Publishers, Boca Raton, Florida.

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

Beauchard, O., J. Gagneur, and S. Brosse. 2003. Macroinvertebrate richness patterns in North African streams. *Journal of Biogeography* 30:1821–1833.

Brown, H. P. 1973. Survival records for elmid beetles, with notes on laboratory rearing of various dryopoids (Coleoptera). *Entomological News* 84:278–284.

Clements, W. H. 1994. Benthic invertebrate community response heavy metals in the Upper Arkansas River Basin, Colorado. *Journal of the North American Benthological Society* 13:30–44.

Cooper, S .D., T. L. Dudley, and N. Hemphill. 1986. The biology of chaparral streams in southern California. Pages 139–151 *in* Proceedings of the Chaparral Ecosystems Research Conference, J. J. DeVries (ed.). California Water

Resources Center Report 62. University of California, Davis, California, 155 pp.

DeShon, J. E. 1995. Development and application of the invertebrate community index (ICI). Pages 217–243 *in* W. S. Davis, T. P. Simon (eds.), Biological assessment and criteria: Tools for water resource planning and decision making. CRC Press, Boca Raton, Florida.

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

Ebert, D. W., and T. G. Wade. 2002. Analytical tools interface for landscape assessments (ATtILA), Version 3.0. US EPA, Office of Research and Development, Washington, DC.

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

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

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

Gasith, A., and V. H. Resh. 1999. Streams in Mediterranean climate regions: Abiotic influences and biotic responses to predictable seasonal events. *Annual Review of Ecology and Systematics 30* : 51–81.

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

Hawkins, C. P., R. H. Norris, J. N. Hogue, and J. M. Feminella. 2000. Development and evaluation of predictive models for measuring the biological integrity of streams. *Ecological Applications* 10:1456–1477.

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

Hughes, R. M. 1995. Defining acceptable biological status by comparing with reference conditions. Pages 31–47 *in* W. S. Davis, T .P. Simon (eds.), Biological assessment and criteria: Tools for water resource planning and decision making. CRC Press, Boca Raton, Florida.

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

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

Karr, J. R., K. D. Fausch, L. Angermeyer, P. R. Yant, and I. J. Schlosser. 1986. Assessment of biological integrity in running waters: A method and its rationale. Illinois Natural History Survey Special Publication No. 5, Illinois Natural History Survey, Urbana Champaign, IL.

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

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

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

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

Klemm, D. J., and J. M. Lazorchak. 1994. Environmental monitoring and assessment program, surface water and Region 3 regional monitoring and assessment program, 1994 pilot laboratory methods manual for streams. EPA/62/R-94/003. US EPA, Office of Research and Development, Washington, DC.

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

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

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

Newman, W. B., and F. R. G. Watson. 2003. Land use history and mapping in California's Central Coast region. The Watershed Institute, California State University, Monterey Bay, California.

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

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

Reynoldson, T. B., D. M. Rosenberg, and V. H. Resh. 2001. Comparison of models predicting invertebrate assemblages for biomonitoring in the Fraser River catchment, British Columbia. *Canadian Journal of Fisheries and Aquatic Sciences* 58:1395–1410.

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

Wright, J. F., M. T. Furse, and P. D. Armitage. 1983. RIVPACS: A technique for evaluating the biological quality of rivers in the U.K. *European Water Pollution Control* 3:15–25.

Zar, J. H. 1999. Biostatistical analysis, 4th ed. Prentice-Hall, Upper Saddle River, New Jersey.