Bioassessment tools in novel habitats: An evaluation of indices and sampling methods in low-gradient streams in California

Raphael D. Mazor, Kenneth C. Schiff, Kerry J. Ritter, Andy Rehn[1] and Peter Ode[1]

Abstract

Biomonitoring programs are often required to assess streams for which assessment tools have not been developed. For example, low-gradient streams (slope ≤1%) comprise 20 to 30% of all stream miles in California and are of particular interest to watershed managers, yet most sampling methods and bioassessment indices in the State were developed in high-gradient systems. This study evaluated the performance of three sampling methods: targeted riffle composite (TRC), reachwide benthos (RWB), and the margin-center-margin modification of RWB (MCM); and two indices: the Southern California Index of Biotic Integrity (SCIBI) and the ratio of observed to expected taxa (O/E) in low-gradient streams in California for application in this habitat type. Performance was evaluated in terms of efficacy (i.e., ability to collect enough individuals for index calculation), comparability (i.e., similarity of assemblages and index scores), sensitivity (i.e., responsiveness to disturbance), and precision (i.e., ability to detect small differences in index scores). The sampling methods varied in the degree to which they targeted macroinvertebrate-rich microhabitats, such as riffles and vegetated margins, which may be naturally scarce in low-gradient streams. The RWB method failed to collect sufficient individuals (i.e., ≥450) to calculate the SCIBI in 28 of 45 samples, and often collected fewer than 100 individuals, suggesting it is inappropriate for low-gradient streams in California. Failures for the other methods were less common (TRC:16 samples; MCM:11 samples). Within-site precision, measured as the minimum detectable difference (MDD), was poor but similar across methods for the SCIBI (ranging from 19 to 22). RWB had the lowest MDD for O/E scores (0.20 vs. 0.24 and 0.28 for MCM and TRC, respectively). Mantel correlations showed that assemblages were more similar within sites among methods than within methods among sites, suggesting that the sampling methods were collecting similar assemblages of organisms. Statistically significant disagreements among methods were not detected, although O/E scores were higher for RWB samples than TRC. Index scores suggested impairment at all sites in the study. Although index scores did not respond strongly to several measurements of disturbance in the watershed, % agriculture showed a significant, negative relationship with O/E scores.

Introduction

Large-scale biomonitoring programs are often confronted with the need to assess habitat types for which assessment tools have not been developed. This problem is severe in large heterogeneous regions like California (Carter and Resh 2005). Developing and maintaining unique assessment tools for multiple habitat types may be prohibitively expensive and may impede comparison of data from different regions. Therefore, assessing the applicability of tools in diverse habitat types is a critical need for large biomonitoring programs.

In southern California, biomonitoring programs use tools like the SCIBI (Ode *et al*. 2005), which were developed using reference sites that were predominantly in high-gradient (i.e., >1% slope) streams. However, low-gradient streams are a major feature in alluvial plains of this region (Carter and

[1] *California Department of Fish and Game, Aquatic Bioassessment Laboratory, Water Pollution Control Laboratory, Rancho Cordova, CA*

Resh 2005). According to the National Hydrography Dataset Plus (NHD+; USEPA and USGS 2005) approximately 20 to 30% of all stream miles in California have slopes below 1%. Because these habitats are subject to numerous impacts and alterations (SMCBWG 2007), several biomonitoring efforts in California specifically target low-gradient streams, even though the applicability of assessment tools created and validated in high-gradient streams has not been tested.

Low-gradient streams differ from high-gradient streams in many respects (Montgomery and Buffington 1997). For example, bed substrate is typically composed of fines and sands, rather than cobbles, boulders, or bedrock. In California and other semiarid climates, low-gradient channels are often complex, with ambiguous and dynamic bank structure. Frequent floods create new channels and cause streams to abandon old ones (Carter and Resh 2005). For bioassessment programs, an important distinction between high- and low-gradient streams is the scarcity of riffles and other microhabitats that are typically targeted by macroinvertebrate sampling protocols (e.g., Harrington 1999).

In this study, application of three sampling methods and two bioassessment indices for use in low-gradient streams in California were evaluated. Sampling methods were assessed for efficacy (i.e., the ability to collect sufficient numbers of benthic macroinvertebrates), comparability (i.e., community similarity and agreement among assessment indices), sensitivity (i.e., responsiveness of the indices to watershed disturbance), and precision of the assessment indices (i.e., power of assessments to detect differences among sites).

Methods

Study Areas

Twenty-one low-gradient sites were sampled in several regions across California (Table 1; Figure 1). Most sites were in heavily altered rivers, although a few were in protected watersheds. Slopes were estimated from the NHD+ (USEPA and USGS 2005), or from digital elevation models (at Jack Slough, Wadsworth Canal, and the Santa Ana River, which lacked associated data in the NHD+). All sites were on reaches defined in the NHD+ as having slopes below 1%.

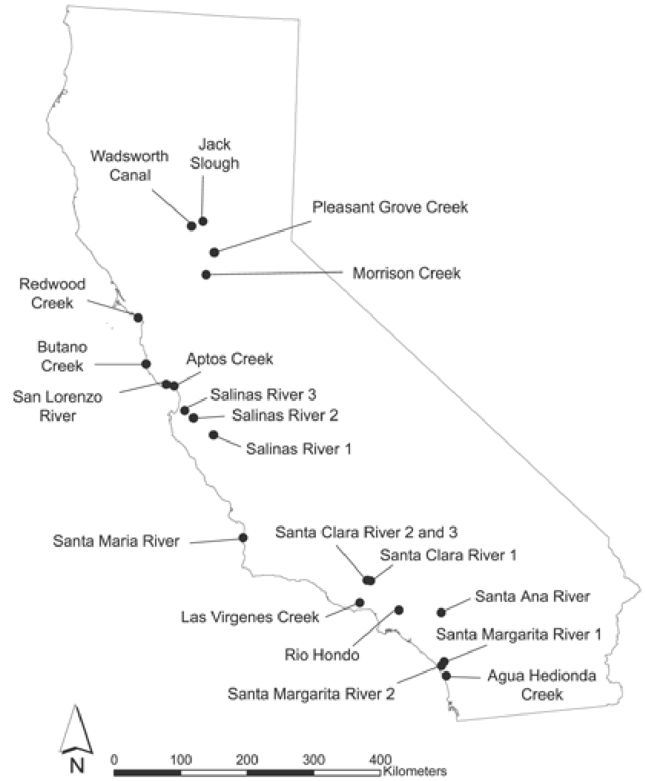

Figure 1. Location of study sites.

Sampling

At each site, TRC, RWB, and MCM sampling methods were used to collect benthic macroinvertebrates. The three sampling methods differ in the degree to which they target the richest microhabitats (e.g., riffles or vegetated margins). TRC and RWB are similar to methods used in the nationwide Environmental Monitoring and Assessment Program (EMAP; Peck *et al*. 2006), and both methods are currently used in California's bioassessment programs (Ode 2007). MCM is intended to capture marginal habitats not sampled by RWB, and has been adopted for use in low-gradient streams in California (Ode and van Buuren 2008). Samples were displaced upstream or downstream by 1 m when necessary to avoid interference among different methods. At 12 sites, triplicate samples were collected for each method (Table 1).

For the TRC method, 11 equidistant transects were established along the 150-m reach, and 3 1-ft^2 areas of streambed were sampled at three randomly selected transects. At each transect, field crews targeted the richest microhabitats and sampled a total of 9 ft^2 of streambed in three riffles. This method is

Table 1. Low-gradient sites included in the study. S = assessed using Southern California Index of Biotic Integrity; X = not assessed using an index of biotic integrity; WS = watershed; Local = within 500 m of sampling point; Ndel = ambiguous watersheds which could not be delineated; Ndet = ambiguous stream network for which stream order could not be determined; and * = triplicate samples collected.

Site	Watershed	County	Watershed Size (km^2)	Stream Order	% Developed		% Agricultural		% Open space	
					WS	Local	WS	Local	WS	Local
Within Central and Southern California										
Central Coast										
S	Aptos Creek*	Santa Cruz	200	3	18	92	0	0	82	8
S	Salinas River 1	Monterey	10940	6	14	71	0	1	86	28
S	Salinas River 2*	Monterey	10666	7	5	28	7	61	88	11
S	Salinas River 3	Monterey	9141	7	5	13	4	27	90	60
S	San Lorenzo River	Santa Cruz	378	4	5	7	6	56	88	37
S	Santa Maria River*	Santa Barbara	1844	6	4	4	6	0	91	96
South Coast										
S	Agua Hedionda Creek*	San Diego	80	3	76	77	0	0	24	23
S	Las Virgenes Creek*	Los Angeles	63	3	19	29	0	0	81	71
S	Rio Hondo*	Los Angeles	325	3	70	83	0	0	30	17
S	Santa Ana River	Riverside	1965	6	25	78	1	0	74	22
S	Santa Clara River 1	Los Angeles	817	4	14	68	0	0	86	32
S	Santa Clara River 2	Los Angeles	1107	5	16	76	0	1	84	23
S	Santa Clara River 3	Los Angeles	1107	5	16	75	0	5	84	20
S	Santa Margarita River 1*	San Diego	1856	6	13	48	3	0	84	52
S	Santa Margarita River 2*	San Diego	1888	6	14	24	3	0	83	76
Outside Central and Southern California										
Bay Area										
X	Butano Creek	San Mateo	234	3	11	34	0	0	89	66
X	Redwood Creek*	Marin	44	2	4	10	2	24	94	67
Central Valley										
X	Jack Slough	Yuba	Ndel	3		7		91		2
X	Morrison Creek*	Sacramento	114	3	40	100	4	0	56	0
X	Pleasant Grove Creek	Placer	40	3	69	34	3	16	28	50
X	Wadsworth Canal	Sutter	Ndel	Ndet		12		87		1

similar to the targeted riffle composite method used by EMAP, which sampled a total of 8 ft^2 of streambed from four to eight riffles (Peck *et al*. 2006). A second difference was the fixed reach length of 150 m, in contrast to EMAP, which had a variable reach length set at 40 times the wetted width.

In contrast to TRC, which allowed the field crew to sample the richest microhabitats within transects, the RWB method used systematically distributed sampling locations. For RWB, eleven equidistant transects were established along the 150-m reach, and one sample was collected with a D-frame kick-n*et al*ong each transect at 25, 50, or 75% of the stream width (with the position changing at each transect). A total of 11 ft^2 of streambed was sampled. This method is similar to the Reach-Wide Benthos method used by EMAP, except that EMAP used variable reach length set to 40 times the wetted width (Peck *et al*. 2006).

The MCM method was identical to RWB with minor modification. Instead of collecting samples at 25, 50 and 75% of stream width, samples were collected at 0, 50, and 100%. Unlike RWB, MCM samples were collected from the margins, which in low-gradient streams often contain the richest, most stable microhabitats (e.g., vegetated margins). As with RWB, 11 ft^2 of streambed were sampled.

Benthic macroinvertebrates were sorted and identified to the Standard Taxonomic Effort Level 1 (i.e., most taxa to genus, with Chironomidae left at family) established by the Southwestern Association of Freshwater Invertebrate Taxonomists (Richards and Rogers 2006). When possible, at least 500 individuals were identified in each sample.

Data Analysis

For each sample, bioassessment metrics and indices were calculated and analyzed to evaluate the efficacy, comparability, sensitivity, and precision of the three sampling methods.

Calculation of indices and metrics

The SCIBI was calculated for 15 sites located on coastal drainages from Santa Cruz to San Diego Counties. No IBIs were calculated for the two sites in the San Francisco Bay Area and the four sites in the Central Valley because IBIs for these regions were not available at the time of the study. Furthermore, small sample sizes in these regions and unknown comparability of IBIs for different regions would limit the utility of including these sites. In order to calculate the SCIBI, benthic macroinvertebrate data were processed according to the index requirements. For example, samples containing more than 500 individuals were randomly subsampled with replacement to obtain 500 individuals per sample.

Calculation of O/E scores

Observed-over-expected scores were calculated for all sites using a predictive model developed for the state of California (Charles P. Hawkins pers. com.; Western Center for Monitoring and Assessment. Accessed online March 30, 2007: http://129.123.10.240/wmcportal/DesktopDefault.aspx). These scores are the ratio of observed to expected taxa, and are based on only those taxa with a probability of occurrence ≥50%. The original identifications were converted to operational taxonomic unit (OTU) names used in the models, and ambiguous taxa (i.e., those that could not be assigned to an OTU and those that could not be adequately identified, such as early instars), as well as all Chironomidae larvae, were eliminated. The resulting sample counts were reduced to 300, if more than 300 individuals remained after removal of ambiguous taxa. Sites were assigned to the appropriate submodel based on climate (i.e., low mean annual precipitation, and high mean monthly temperature), which were used to predict expected taxa occurrence (E) using longitude, percent sedimentary geology in the watershed, and log mean annual precipitation. Climatic data were obtained from the Oregon Climate Center (accessed online March 30, 2007: http://www.ocs.orst.edu/prism), and geologic data were obtained from a generalized geological map of the United States (accessed online March 30, 2007: http://pubs.usgs.gov/atlas/geologic). Details of these predictive models can be found in Ode *et al*. 2008.

The two Central Valley sites were located in streams with ambiguous watersheds, and therefore required that percent sedimentary geology be estimated, rather than calculated by geographic information systems (GIS). For this study, percent sedimentary geology was estimated at 100%. Using different percent sedimentary geology values (i.e., 0, 20, 40, 60, and 80%) had negligible effect on O/E scores; coefficient of variation for scores within each sample at the two Central Valley sites was <2%, (data not shown), perhaps as a result of the low numbers of observed taxa at these sites.

Evaluation of sampling methods and indices

Efficacy

To assess the efficacy of the sampling methods, the percentage of samples was calculated for each method that collected at least 450 individuals (within 10% of the minimum number for calculating the SCIBI) or at least 270 individuals (within 10% of the minimum number for calculating O/E, counting only unambiguous taxa). In bioassessment applications, smaller samples would be rejected and represent wasted resources. In order to minimize the effects of pseudoreplication, the percentage of samples containing an adequate number of individuals was calculated for each site; then, this percentage was averaged across all 21 sites. This rate estimated the likelihood of collecting adequate samples from the population of sites in the study. McNemar's test was used to test differences between methods (paired within sites) for statistical significance (Zar 1999, Stokes *et al*. 2000). Because McNemar's test requires binary data, within-site rates were rounded to 1 or 0 at replicated sites. A Bonferroni correction was used to account for multiple tests across methods (i.e., $\alpha = 0.05/3 = 0.017$).

Comparability

To see if the different sampling methods collected similar types of organisms, community structure between sampling methods was compared using a Mantel test (Mantel 1967). Mantel tests provide a measure of correlation (Mantel's R) between two sampling methods. Sorensen distance was used as a dissimilarity measure. For sites where multiple samples were collected, mean distances were used; that is, matrices comprised mean or observed distances between pairs of sites, not samples. All samples were included in this analysis, regardless of the number of individuals collected. Significance was tested against correlation values for 999 runs with randomized data.

A Bonferroni correction was used to account for multiple tests across methods (i.e., $\alpha = 0.05/3 = 0.017$). PC-ORD [Version 5.12] was used to run Mantel tests (MJM Software Design, Glendeden Beach, OR).

To determine the relative influence of sampling method on assessment indices, a variance components analysis was used to determine how much of the variability was explained by differences among sites, sampling methods, and their interaction. Restricted maximum likelihood (REML) was used to calculate variance components because of the unbalanced design. SAS was used for all calculations (using PROC VARCOMP method=REML, SAS Institute Inc. 2004). Unlike the mean square method of estimating variance components, REML ensures that all components are greater than or equal to zero (Larsen *et al*. 2001). Because sites were a fixed factor and not a random factor, the variance component attributable to site must be considered a finite, or pseudo variance (Courbois and Urquhart 2004). Only sites where all three sampling methods were represented (after excluding samples containing inadequate numbers of organisms) were used in this analysis.

To assess agreement among the sampling methods, mean SCIBI and O/E values were calculated and regressed for each pair of methods. Slopes were tested against 1 and intercepts to 0 ($\alpha = 0.05$); Theil's test for consistency and agreement, which is based on differences between sampling methods, was used as an additional test of comparability (Theil 1958). Pairwise differences between mean SCIBI and O/E scores were regressed against log watershed area and stream order to see if these gradients contributed to the observed disagreements. A Bonferroni correction was not used for either analysis in order to increase the ability to detect disagreements. Bias was not explicitly assessed because none of the methods could be assumed to represent a true value. Only samples with adequate numbers of individuals were used in this analysis.

Sensitivity

The sensitivity of the assessment indices to watershed alteration was assessed by correlating mean SCIBI and O/E scores against land cover metrics, including percent open, developed, and agricultural land within the watershed for all sites with unambiguous watersheds (Table 1). This analysis assumed that the biology of the streams respond to these watershed alterations. Open water was excluded from all calculations. Land cover data was obtained from the National Land Cover Database (USGS 2003). Relationships were assessed by calculating the Spearman rank correlation, which is robust to non-normal distributions and extreme values in land cover metrics (Zar 1999). Only samples with the minimum number of individuals for each index were used in this analysis. Data from each sampling method were analyzed independently. A Bonferonni correction was used to account for multiple comparisons ($\alpha = 0.05/6 = 0.008$) across two indices and three land cover classes within each method.

Precision

Precision was evaluated by calculating the MDD of each sampling method for SCIBI and O/E scores (Zar 1999, Fore *et al*. 2001). The MDD was calculated using the mean squared error from a nested ANOVA (replicates within site) as an estimate for average within-site variance. Only data from site and method combinations with replication (after exclusion of samples lacking adequate numbers of individuals) were used to estimate variability. These estimated variabilities were applied to a two-sample *t*-test ($\alpha = 0.05$, $\beta = 0.10$) with three replicates in each sample. Additionally, the coefficient of variation (CV) of the indices for each method, averaged across sites, was calculated.

Results

One hundred thirty-five samples were collected at 21 sites throughout the state; 15 of these sites were located along the southern and central California coast. All three methods were used at each site, and 196 taxa were identified. For all sampling methods, SCIBI and O/E scores were low at most sites (Figure 2). For example, mean SCIBI scores were well under 39 (the impairment threshold) at all but one site (Aptos Creek). Observed-over-expected scores indicated impairment in nearly every sample, as scores were below the impairment threshold of 0.66 in all but three samples.

Efficacy

Efficacy was low for all methods, and many samples contained fewer than the required number of individuals. Ideally, each sample should have contained at least 500 individuals. However, only 46 of 135 samples met this target; 34 of the remaining 89 samples had at least 450 individuals, the minimum required for calculation of the SCIBI. For the 55 samples with fewer than 450 individuals, IBIs may

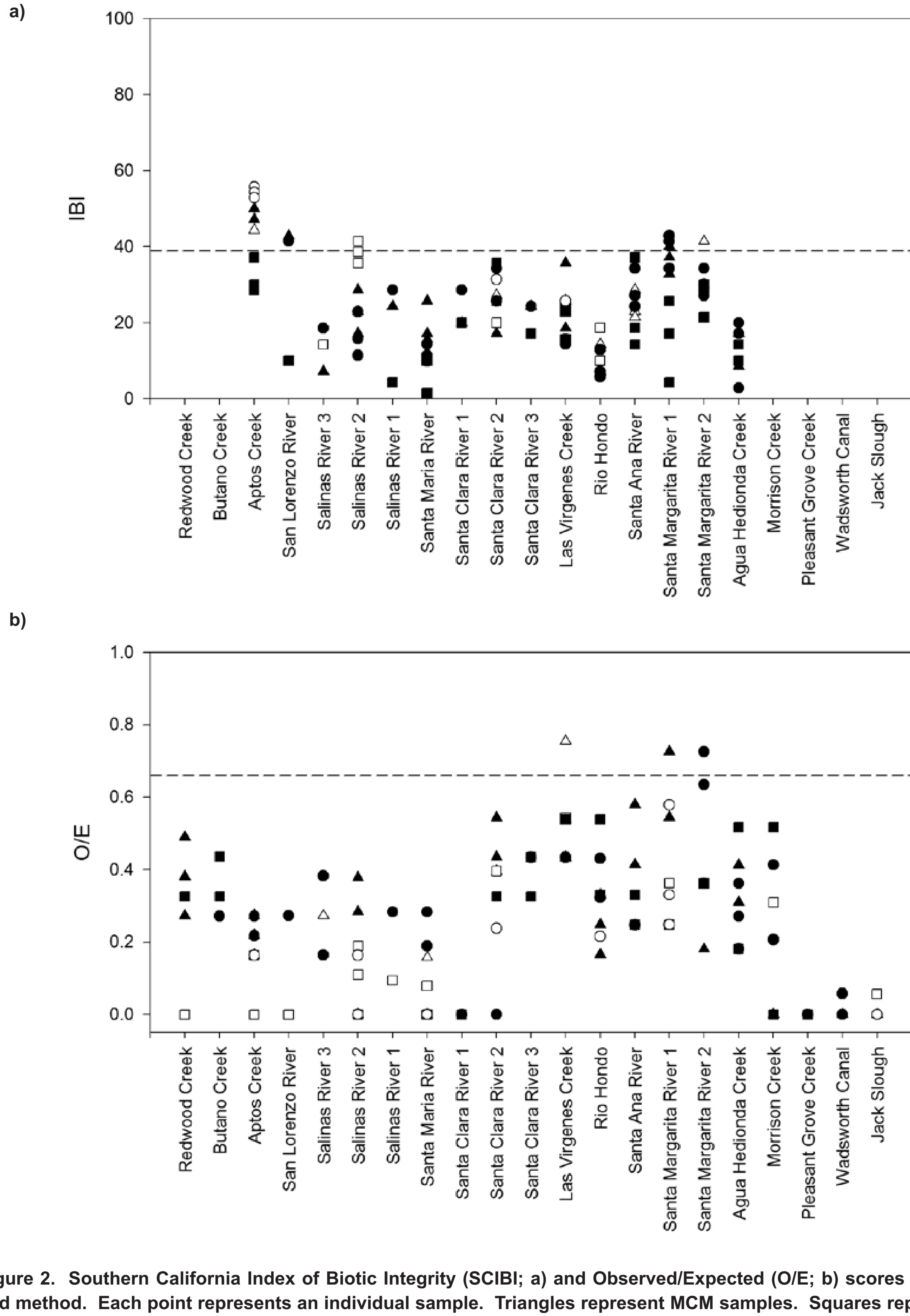

Figure 2. Southern California Index of Biotic Integrity (SCIBI; a) and Observed/Expected (O/E; b) scores by site and method. Each point represents an individual sample. Triangles represent MCM samples. Squares represent RWB samples. Circles represent TRC samples. Black symbols are samples containing sufficient individuals for index calculation, and white symbols are samples containing insufficient individuals for index calculation. Dashed lines represent the threshold for identifying impairment with each index (i.e., 39 for the SCIBI, and 0.66 for the O/E).

not be valid. Furthermore, 55 samples had fewer than 270 unambiguously identified individuals, meaning that O/E scores may not be valid for these samples.

Several samples had extremely low counts (e.g., four individuals; Table 2). Most of these samples were collected by the RWB sampling method. Nearly half (21 out of 45) of RWB samples had fewer than 450 individuals. In contrast, only 2 MCM samples and 6 TRC samples had fewer than 450 individuals. The adjusted efficacy rate, a site-adjusted estimate of sampling efficacy, for the MCM method (54%) was twice that of RWB (27%). The adjusted efficacy rate for TRC (46%) was nearly as high as that of the MCM method. However, these differences fell short of statistical significance after Bonferroni corrections were applied (i.e., $p > 0.017$). The rates were slightly higher for samples with at least 270 individuals at 67, 32, and 67% for MCM, RWB, and TRC, respectively, and these differences were statistically significant (McNemar's test $p = 0.0039$).

Comparability

Sampling methods comparability was good in terms of both multivariate community structure and index scores. Mantel's test showed significant correlations among benthic macroinvertebrate communities collected by all three sampling methods (Table 3). However, the RWB method had weaker correlations with both TRC (0.40) and MCM (0.45), compared to the higher correlation observed between TRC and MCM (0.69). In all cases, the correlations were significant ($p < 0.002$).

Variance components analysis showed that the methods were highly comparable and that site accounted for nearly all of the explained variance in both indices. The analysis of SCIBI scores included 7 sites and 26 samples; the analysis of O/E scores included 10 sites and 52 samples. Site accounted for 100% of the explained variance in SCIBI scores and 95% in O/E scores. Method and interaction between site and method explained none or negligible components of the variance in these indices (0 to 5%).

Table 3. Mantel correlations between sampling methods. Asterisk denotes statistical significance ($p < 0.017$).

Method 1	Method 2	Mantel's R	P
RWB	MCM	0.45	0.001*
RWB	TRC	0.40	0.002*
MCM	TRC	0.69	0.001*

Significant disagreements between pairs of sampling methods were not observed for either index (Table 4; Figure 3). Slopes for all three comparisons were not significantly different from 1, and no intercepts were significantly different from 0. Consistency among SCIBI scores was best (i.e., slope closest to 1) between the MCM and TRC methods (slope = 0.96) and worst for the MCM and RWB methods (slope = 0.62). In contrast, consistency among O/E scores was best between the MCM and RWB methods (slope = 0.97) and worst for the RWB and TRC methods (slope = 0.72). Theil's test confirmed the lack of significant disagreements among IBI and O/E scores between pairs of methods. No differences between sampling methods were significantly related to log watershed area or stream order (regression slope and intercept $p > 0.05$).

Sensitivity

Sensitivity of both indices to gradients in land cover was poor, although to some extent the relationships were affected by sampling method, specific cover type, and geographic scale (Table 5; Figure 4). For example, O/E scores were strongly and negatively correlated with agricultural land cover in the

Table 2. Samples, sites, and efficacy by method. Adjusted Rate = site-adjusted estimate of efficacy rate.

Method	Total		≥ 450 Organisms			≥ 270 Organisms		
	Samples	Sites	Samples		Adjusted Rate	Samples		Adjusted Rate
MCM	45	21	34	76%	54%	32	71%	67%
RWB	45	21	17	38%	27%	14	31%	32%
TRC	45	21	29	64%	46%	30	67%	67%

Table 4. Regressions of mean IBI and O/E scores for each method. Slopes were tested against 1 and intercepts were tested against 0. Methods 1 and 2 plotted on x and y axis, respectively, in Figure 3. SE = Standard error.

Index	Method 1 (x)	Method 2 (y)	n	r^2	Slope	SE	p	Intercept	SE	p
SCIBI	MCM	TRC	14	0.77	0.96	0.15	0.803	2.52	3.96	0.537
	MCM	RWB	7	0.45	0.62	0.25	0.194	6.31	5.53	0.305
	MH	TRC	7	0.74	1.18	0.28	0.540	-0.30	5.63	0.959
O/E	MCM	TRC	14	0.78	0.86	0.13	0.284	0.02	0.04	0.633
	MCM	RWB	8	0.90	0.97	0.13	0.816	0.02	0.04	0.653
	RWB	TRC	8	0.71	0.72	0.19	0.185	0.06	0.06	0.401

watershed (Spearman's ρ ranged from -0.46 to -0.89 across sampling methods). However, most relationships between index scores and land cover metrics were not statistically significant (i.e., p <0.008). Only the relationship between O/E scores from RWB samples were significantly correlated with agricultural land use in the watershed (ρ = -0.89, p = 0.003). Although the direction of correlation often met expectations (e.g., % open space in the watershed vs. SCIBI; Figure 4c), a few showed no clear relationship (e.g., % developed land in the watershed vs. O/E; Figure 4d).

Precision

Sampling method affected the precision of both the SCIBI and O/E scores (Table 6). For example, the RWB sampling method had the largest MDD for the SCIBI: 22 vs. 19 for the other two methods. However, RWB had the lowest MDD when O/E

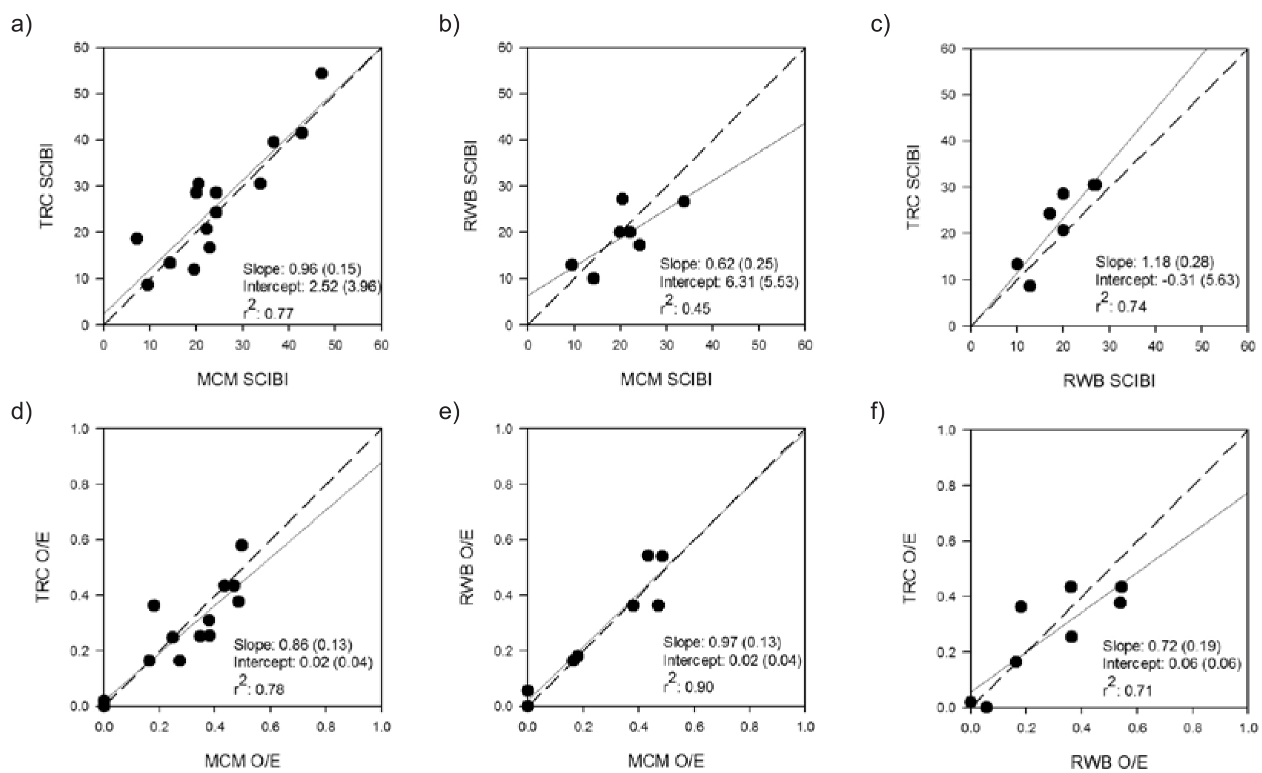

Figure 3. Agreement between the sampling methods for Southern California Index of Biotic Integrity (SCIBI; a – c) and Observed/Expected (O/E; d – f) scores Each point represents the mean index score at a site. Solid lines represent linear regressions, and dashed lines represent perfect 1:1 relationships. Numbers in parentheses are standard errors. Slopes were tested against 1, and intercepts were tested against 0.

Table 5. Spearman rank correlations (*ρ*) between bioassessment indices and landscape metrics. * = statistical significance (p <0.008).

Index	Land Cover	Method	Watershed			1 km radius		
			n	*ρ*	p	n	*ρ*	P
SCIBI	% Developed	MCM	15	-0.08	0.783	15	0.11	0.685
		RWB	7	-0.75	0.054	7	-0.59	0.159
		TRC	14	-0.32	0.914	14	0.20	0.487
	% Open	MCM	15	-0.04	0.892	15	0.09	0.742
		RWB	7	0.62	0.139	7	0.67	0.102
		TRC	14	-0.04	0.890	14	-0.08	0.782
	% Agricultural	MCM	15	0.06	0.842	15	-0.11	0.689
		RWB	7	0.12	0.799	7	0.22	0.628
		TRC	14	0.00	0.991	14	-0.02	0.954
O/E	% Developed	MCM	15	0.14	0.640	15	0.35	0.202
		RWB	8	-0.28	0.509	8	-0.07	0.866
		TRC	17	0.23	0.370	17	0.31	0.222
	% Open	MCM	15	-0.05	0.857	15	0.01	0.980
		RWB	8	0.40	0.333	8	0.17	0.693
		TRC	17	-0.24	0.355	17	0.02	0.948
	% Agricultural	MCM	15	-0.67	0.009	15	-0.24	0.388
		RWB	8	-0.89	0.003*	8	-0.15	0.719
		TRC	17	-0.46	0.064	17	-0.31	0.220

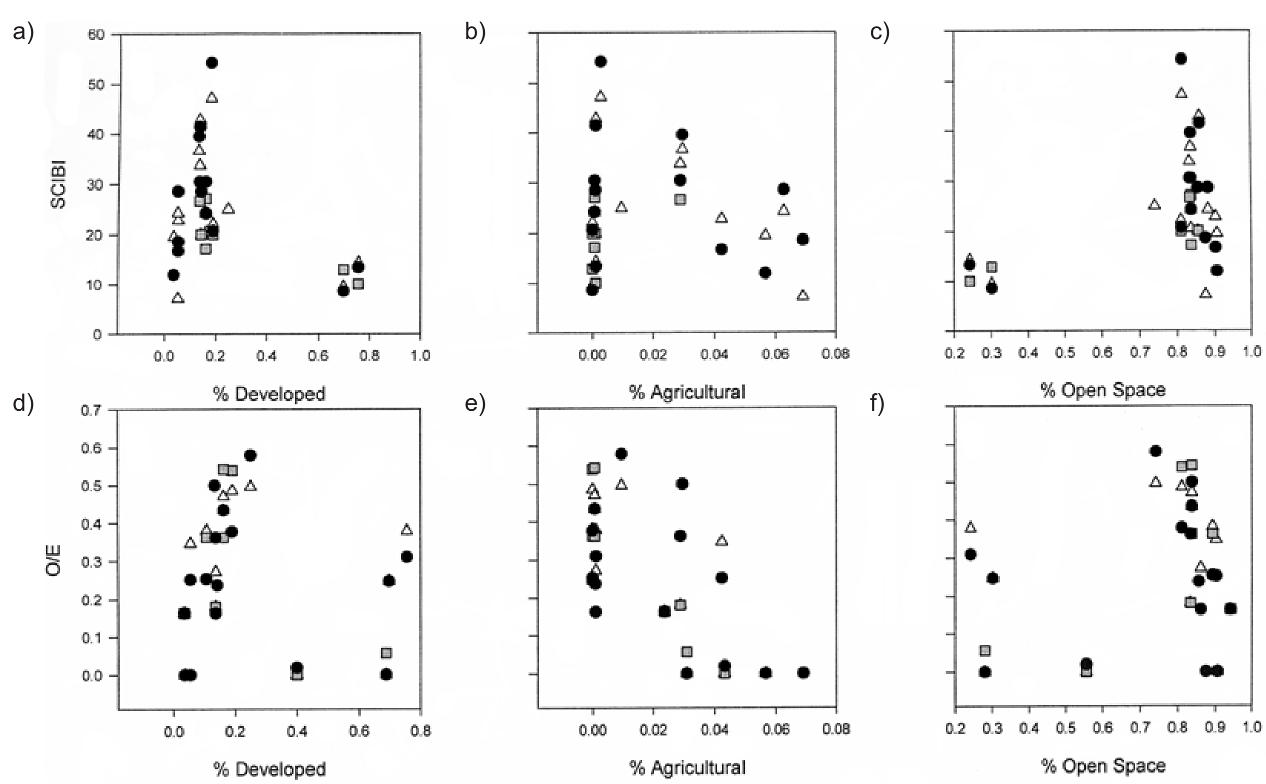

Figure 4. Index scores versus land cover metrics. Each point represents the mean of all samples collected by one method at each site. White triangles represent MCM samples. Gray squares represent RWB samples. Black circles represent TRC samples.

Table 6. Within-site variability (expressed as mean square error, MSE) and minimum detectable difference (from a two-sample, 2-tailed t-test with n = 30, α = 0.05, and β = 0.1) for each of the sampling methods. d.f.: degrees of freedom. SS: sum of squares. MSE: mean square error. MDD: mean detectable difference.

Index	Method		d.f.	SS	MSE	F	p	MDD
SCIBI	TRC	Sites	7	2507	358	12.5	>0.0001	19
		Residuals	15	430	29			
	RWB	Sites	3	403	134	3.7	0.0701	22
		Residuals	7	254	36			
	MCM	Sites	8	1745	218	8.0	0.0002	19
		Residuals	16	437	27			
O/E	TRC	Sites	8	0.625	0.078	12.7	>0.0001	0.28
		Residuals	13	0.074	0.006			
	RWB	Sites	3	0.115	0.038	14.5	0.0037	0 20
		Residuals	6	0.016	0.003			
	MCM	Sites	9	0.860	0.096	20.9	>0.0001	0.24
		Residuals	17	0.078	0.005			

scores were used: 0.20 vs. 0.28 for TRC and 0.24 for MCM. Coefficients of variation showed similar trends in variability among methods when SCIBI scores were used, (ranging from 22 to 27%), and lower CVs for RWB when O/E scores were used: 12 vs. 20% for MCM and 45% for TRC.

The low number of samples containing adequate numbers of individuals meant that estimates of within-site variance were sometimes based on very small samples. For example, only four sites in the region using the SCIBI had multiple samples with sufficient numbers of organisms collected by the RWB method. This problem was less severe for estimates based on O/E scores because fewer individuals per sample are required for index calculation, and because sites in the Central Valley and San Francisco Bay area could be included in the estimates.

Discussion

Low-gradient streams are distinct from other streams in many aspects, such as substrate material, bed morphology, and the distribution of microhabitats (Montgomery and Buffington 1997). As a consequence of these differences, traditional bioassessment approaches in California that were developed in high-gradient streams with diverse microhabitats have limited applications in low-gradient reaches. The sampling methods evaluated in this study differed in the extent to which they targeted the richest microhabitats (such as riffles, or vegetated margins). For example, the TRC method allows field crews to select the richest microhabitats specifically. In contrast, the RWB method may systematically undersample or miss these habitats entirely, as the richest areas in low-gradient streams are typically found at the margins (Montgomery and Buffington 1997). The MCM method, a modification of the RWB method, was designed so that these margins could be targeted.

Caution should be used when applying sampling methods or assessment tools that were calibrated for specific habitat types (e.g., high-gradient streams) to new habitats (e.g., low-gradient streams). The present study's evaluation of assessment tools unveiled a number of shortcomings that weaken application of these tools in low-gradient streams, including the inability to collect adequate numbers of organisms, poor sensitivity of assessments, and low precision of the sampling methods. Significant disagreements among the methods were not detected, although power was low because of the low number of samples. The inability of the RWB sampling method to collect an adequate number of individuals in nearly half of all samples makes it unsuitable for low-gradient streams, even though this method is widely used by bioassessment programs in California (Ode 2007) and across the USA (Peck *et al*. 2006). Although biomonitoring programs must assess a diverse range of habitat types with available tools, the present study indicates that these programs may be well served by evaluating tools in novel habitats where monitoring activities occur.

Variance components analysis of assessment indices showed that differences among sites explained more of the variance in index scores than differences among sampling methods, suggesting that similar types of benthic macroinvertebrates are collected by the different methods. However, analysis of disagreements among the methods indicated that some samples collected by RWB were distinct from those collected by TRC, and samples collected by MCM were intermediate between the other two. For example, samples collected by TRC had lower O/E scores than samples collected by MCM, which in turn were lower than those collected by RWB. However, differences among these methods did not reach statistical significance.

Other studies comparing single, targeted habitat sampling methods (e.g., TRC) to multi-habitat sampling methods (e.g., RWB) have shown similar results. For example, MDDs reported in other studies (or calculated from reported variabilities) were comparable to those reported here, although generally larger (Rehn *et al*. 2007, Blocksom *et al*. 2008). However, these studies found that multi-habitat sampling reduced variability in multimetric indices, whereas the present study found that variability was lower for the single habitat method (i.e., TRC; Table 7). As in Rehn *et al*. (2007), the present study found that TRC samples had higher O/E scores than RWB samples, but that the strength of disagreement was inconsistent in the largest watersheds.

The generally weak response of the indices to land cover metrics suggests that the SCIBI and O/E may not be sensitive to variability in watershed-scale disturbance in low-gradient streams. This conclusion is tempered by small sample sizes that limited power, and sensitivity to reach-scale degradation was not explored in this study for lack of data. Several studies have shown the strong impact of reach-scale factors on benthic macroinvertebrates, which may exceed the influence of watershed-scale stressors (e.g., Hickey and Doran 2004, Sandin and Johnson 2004). Furthermore, most of the watersheds in the study were highly altered, particularly those in the region of the SCIBI, and portions of the disturbance gradient to which these indices are more sensitive may not have been adequately sampled. Several studies have found that biota responds to disturbance gradients ≤10% development in a watershed, but responses above this gradient are muted (e.g., Hatt *et al*. 2004, Walsh *et al*. 2007). Agricultural land cover, which was low in most watersheds (<10%), showed strong responses with the indices, suggesting that the study was able to capture portions of this gradient to which both the SCIBI and O/E were sensitive.

The low numbers of organisms collected from the low-gradient streams in the study may reflect the naturally low population densities of benthic macroinvertebrates in these reaches. The River Continuum Concept hypothesizes that higher order streams with larger watersheds have a lower energy base because of reduced allochthonous input and depressed autochthonous productivity (Vannote *et al*. 1980). This lower energy base would be expected to support reduced biomass. However, observation of the sites in this study suggests that the lack of stable microhabitats (e.g., riffles and vegetated margins) may account for the reduced numbers of macroinvertebrates, as few species are adapted to the shifting sandy substrate found in most low-gradient streams in California. A well known, but extreme, example of the impact of shifting sandy substrates on maintaining low densities of benthic macroinvertebrates are the migrating submerged dunes in the lower

Table 7. Minimum detectable differences in multimetric indices. Southern California Index of Biotic Integrity (SCIBI); Northern California Index of Biotic Integrity (NICIBI); Virginia Stream Condition Index (VSCI); Macroinvertebrate Biotic Integrity Index (MBII); California O/E Index (O/E); and NT = not tested.

Index type	Method	Present study	Rehn *et al*. 2007	Blocksom *et al*. 2008	
Multimetric index	Single-habitat	19.2 (SCIBI)	19.7 (SCIBI + NCIBI)	19.88 (VSCI)	29.79 (MBII)
	Multi-habitat	22.6 (SCIBI)	15.5 (SCIBI + NCIBI)	17.37 (VSCI)	17.91 (MBII)
Predictive model	Single-habitat	0.28 (O/E)	0.22 (O/E)	NT	NT
	Multi-habitat	0.20 (O/E)	0.19 (O/E)	NT	NT

Amazon River (Sioli 1975, Lewis, Jr. *et al*. 2006). Although very high productivity of Chironomidae and other benthic macroinvertebrates has been observed in low-gradient sandy rivers of the southeastern United States, this productivity was attributed to snags and other stable microhabitats, more than to the shifting sandy substrate (Benke 1998). Thus, the vast majority of the macroinvertebrate activity in a large reach of river was found in small areas containing snags (Wallace and Benke 1984). Snag microhabitats are arguably less common in streams of the arid Southwest, which lack dense riparian forests to contribute snag-forming woody debris and may be less likely to be sampled using a systematic sampling method like RWB.

Bioassessment programs are often required to make do with available tools to fulfill regulatory mandates, yet they lack resources to evaluate the tools for applications in all habitats of concern. Although all sampling methods in this study suffered from poor efficiency in collecting organisms, the MCM method greatly improved efficacy and reduced the frequency of rejected samples. Furthermore, the lack of significant disagreements and inconsistencies suggests that the MCM method produced results that were comparable to the other methods already in use in California, which may facilitate integration of historical data sets (Cao *et al*. 2005, Rehn *et al*. 2007). Therefore, the present study supports the use of MCM in low-gradient streams in California as a substitute for the currently preferred RWB method. Overall, bioassessment programs can improve data quality and avoid unnecessary expenses by explicitly evaluating assessment tools when assessing novel habitat types.

Literature Cited

Benke, A.C. 1998. Production dynamics of riverine chironomids: Extremely high biomass turnover rates of primary consumers. *Ecology* 79:899-910.

Blocksom, K.A., B.C. Autrey, M. Passmore and L. Reynolds. 2008. A comparison of single and multiple habitat protocols for collecting macroinvertebrates in wadeable streams. *Journal of the American Water Resources Association* 44:577-593.

Cao, Y., C.P. Hawkins and A.W. Storey. 2005. A method for measuring the comparability of different sampling methods used in biological surveys: implications for data integration and synthesis. *Freshwater Biology* 50:1105-1115.

Carter, J.L. and V.H. Resh. 2005. Pacific Coast Rivers of the Coterminous United States. pp. 541-590 *in*: A.C. Benke and C.E. Cushing (eds.), Rivers of North America. Elsevier Academic Press. Boston, MA.

Courbois, J.-Y.P. and N.S. Urquhart. 2004. Comparison of survey estimates of the finite population variance. *Journal of Agricultural, Biological, and Environmental Statistics* 9:236-251.

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

Harrington, J. M. 1999. California Stream Bioassessment Procedures. California Department of Fish and Game, Aquatic Bioassessment Laboratory, Water Pollution Control Laboratory. Rancho Cordova, CA.

Hatt, B.E., T.D. Fletcher, C.J. Walsh and S.L. Taylor. 2004. The influence of urban density and drainage infrastructure on the concentrations and loads of pollutants in small streams. *Environmental Management* 34:112-124.

Hickey, M.B.C. and B. Doran. 2004. A review of the efficiency of buffer strips for the maintenance and enhancement of riparian ecosystems. *Water Quality Research Journal* 39:311-317.

Larsen D.P., T.M. Kincaid., S.E. Jacobs and N.S. Urquhart. 2001. Designs for evaluating local and regional scale trends. *Bioscience* 51:1069-1078.

Lewis, Jr., W.M., S.K. Hamilton and J.F. Saunders, III. 2005. Rivers of northern South America. pp. 219-256 *in*: C.E. Cushing, K.W. Cummins and G.W. Minshall (eds.), River and Stream Ecosystems of the World. University of California Press. Berkeley, CA.

Mantel, N. 1967. The detection of disease clustering and generalized regression approach. *Cancer Research* 27:209-220.

Montgomery, D. and J. Buffington. 1997. Channel-reach morphology in mountain drainage basins. *Geological Society of America Bulletin* 109:596-611.

Ode, P.R. 2007. Standard operating procedures for collecting benthic macroinvertebrate samples and associated physical and chemical data for ambient bioassessment in California. Available from http://mpsl.mlml.calstate.edu/phab_sopr6.pdf

Ode, P.R. and B.H. van Buuren. 2008. Amendment to SWAMP Interim Guidance on Quality Assurance for SWAMP Bioassessments. Surface Water Ambient Monitoring Program. Sacramento, CA.

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

Ode, P.R, C.P. Hawkins and R.D. Mazor. 2008. Comparability of biological assessments derived from predictive models and multimetric indices of increasing geographic scope. *Journal of the North American Benthological Society* 27:967-985.

Peck, D.V., A.T. Herlihy, B.H. Hill, R.M. Hughes, P.R. Kaufmann, D.J. Kemm, J.M. Lazorchek, F.H. McCormick, S.A. Peterson, S.A. Ringold, T. Magee and M. Cappaert. 2006. Environmental Monitoring and Assessment Program—Surface Waters Western Pilot study: Field operations manual for wadeable streams. EPA/620/R-06/003. United States Environmental Protection Agency, Office of Research and Development. Corvallis, OR.

Rehn, A.C., P.R. Ode and C.P. Hawkins. 2007. Comparison of targeted-riffle and reach-wide benthic macroinvertebrate samples: implications for data sharing in stream-condition assessments. *Journal of the North American Benthological Society* 26:332-348.

Richards, A.B. and D.C. Rogers. 2006. List of Freshwater Macroinvertebrate Taxa from California and Adjacent States including Standard Taxonomic Effort Levels. Southwest Association of Freshwater Invertebrate Taxonomists. Accessed online March 1, 2008: http://www.swrcb.ca.gov/swamp/docs/safit/ste_list.pdf

Sandin, L. and R.K. Johnson. 2004. Local, landscape and regional factors structuring benthic macroinvertebrate assemblages in Swedish streams. *Landscape Ecology* 19:504-514.

SAS Institute, Inc. 2004. SAS OnlineDoc 9.1.3. Cary, NC.

Sioli, H. 1975. Tropical river: The Amazon. pp. 461-488 *in*: B.A. Whitton (ed.), River Ecology. University of California Press. Berkeley, CA.

Stokes, M.E., C.S. Davis and G.G. Koch. 2000. Categorical Data Analysis Using the SAS System. 2nd edition. SAS Institute, Inc. Cary, NC.

Stormwater Monitoring Coalition Bioassessment Working Group (SMCBWG). 2007. Regional Monitoring of Southern California's Coastal Watersheds. Technical Report 539. Southern California Coastal Water Research Project. Costa Mesa, CA.

Theil, H. 1958. Economic Forecasting and Policy. North Holland Publishing Co. Amsterdam, The Netherlands.

United States Environmental Protection Agency and United States Geological Survey (USEPA and USGS). 2005. National Hydrography Dataset Plus. Edition 1.0. Available from: http://www.horizon-systems.com/nhdplus/

United States Geological Survey (USGS). 2003. National Land Cover Database. Edition 1.0. http://www.lsc.usgs.gov.

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 Science* 37:130-137.

Wallace, J.B. and A.C. Benke. 1984. Quantification of wood habitat in subtropical Coastal Plain streams. *Canadian Journal of Fisheries and Aquatic Sciences* 41:1643-1652.

Walsh, C.J., K.A. Waller, J. Gehling and R. MacNally. 2007. Riverine invertebrate assemblages are degraded more by catchment urbanisation than by riparian deforestation. *Freshwater Biology* 52:574-587.

Zar, J. 1999. Biostatistical Analysis. 4th edition. Prentice Hall. Upper Saddle River, NJ.

Acknowledgements

The authors thank the Aquatic Bioassessment Laboratory of the California Department of Fish and Game for laboratory processing and identification; Aquatic Bioassay and Consulting, Weston Solutions, Inc., and California Regional Water Quality Control Boards for field sampling; and Chuck Hawkins for assistance with predictive models. This project was partially supported by the Stormwater Monitoring Coalition of Southern California.