American Fisheries Society Symposium 47:179–191, 2005

Setting Limits: The Development and Use of Factor-Ceiling Distributions for an Urban Assessment Using Macroinvertebrates

JAMES L. CARTER* AND STEVEN V. FEND

U. S. Geological Survey, 345 Middlefield Road MS465, Menlo Park, California 94025, USA

Abstract.—Lotic habitats in urban settings are often more modified than in other anthropogenically influenced areas. The extent, degree, and permanency of these modifications compromise the use of traditional reference-based study designs to evaluate the level of lotic impairment and establish restoration goals. Directly relating biological responses to the combined effects of urbanization is further complicated by the nonlinear response often observed in common metrics (e.g., Ephemeroptera, Plecoptera, and Trichoptera [EPT] species richness) to measures of human influence (e.g., percentage urban land cover). A characteristic polygonal biological response often arises from the presence of a generalized limiting factor (i.e., urban land use) plus the influence of multiple additional stressors that are nonuniformly distributed throughout the urban environment. Benthic macroinvertebrates, on-site physical habitat and chemistry, and geographical information systems–derived land cover data for 85 sites were collected within the 1,600-km^2 Santa Clara Valley (SCV), California urban area. A biological indicator value was derived from EPT richness and percentage EPT. Partitioned regression was used to define reference conditions and estimate the degree of site impairment. We propose that an upper-boundary condition (factor-ceiling) modeled by partitioned regression using ordinary least squares represents an attainable upper limit for biological condition in the SCV area. Indicator values greater than the factor-ceiling, which is monotonically related to existing land use, are considered representative of reference conditions under the current habitat conditions imposed by existing land cover and land use.

Introduction

Identification of reference sites is fundamental to study designs that evaluate the composition and structure of benthic macroinvertebrate assemblages for the assessment of lotic systems. The concept underlying the use of reference sites in impact assessments is based on the use of controls in manipulative experimental designs. Reference and test sites are assumed to be sufficiently similar that, in the absence of impact, the chosen response variable(s) can be logically compared.

Green (1979) provided the principal rationale for the use of controls in field-based aquatic monitoring studies. However, most of the experimental designs were focused on geographically small-scale studies. In the last few decades, the geographic extent of assessment programs has increased substantially. Many programs are national in scope. Some examples include the River Invertebrate Prediction Classification system (RIVPACS; Wright 2000), Australian River Assessment System (AUSRIVAS; Davies 2000; Simpson and Norris 2000), Environmental Monitoring and Assessment Program (EMAP; USEPA 2002a), and National Water-Quality Assessment Program (NAWQA; Gilliom et al. 1995, 2001). These programs include thousands of sites located over many thousands of square kilometers. Large-scale biomonitoring programs are also integral to many state water quality programs (Carter and Resh 2001; USEPA 2002b).

Natural environmental gradients often increase as the geographic scale of a study increases (Corkum 1989; Carter et al. 1996; Fend et al. 2005, this volume). Longer environmental gradients invariably lead to higher species turnover, which can compromise the logical use of standard reference approaches. This increase in geographic scale and gradient length has influenced the methods used to identify and apply reference sites (Hughes et al. 1986; Wright 2000) and

* Corresponding author: jlcarter@usgs.gov

has also affected their perceived applicability (Polls 1994; Reash 1995).

Although numerous methods have been developed for identifying reference sites, the first step is to determine whether the populations of both reference and test sites possess similar nonimpact-associated physical and chemical characteristics. Similar sites are then presumed to have the same biological potential in the absence of impact; consequently, impact can be logically inferred by differences in macroinvertebrate assemblage composition and structure between the reference and test sites. This initial site classification step is overtly stated in some methods (Hughes 1995; Barbour et al. 1999), while in others it is less apparent (Wright et al. 1984). Following the initial identification of reference sites, a refinement to eliminate (Barbour et al. 1999) or account for (Wright 2000) further differences in either the physical habitat and/or the benthic assemblages among the chosen reference sites generally occurs.

Lotic habitats within urban settings are often more physically modified than in other anthropogenically influenced environmental settings (Paul and Meyer 2001). Factors influencing lotic systems include increases in the percentage of impervious surface, stream canalization, and loss of or highly modified riparian corridors. In most urban settings, these modifications often form a gradient of increasing anthropogenic effect from more distant rural areas to the core of the urban center. In many areas, particularly in the western United States, natural gradients in landform (e.g., altitude) often coincide with these anthropogenic gradients (Carter et al. 1996).

The extent and degree of modifications to lotic systems associated with urban settings often compromises the use of the reference-based study designs using benthic invertebrates. This is particularly true when assessments are relatively small-scale, such as when assessments are constrained by political boundaries but still encompass a variety of habitat conditions. Establishing justifiable expected biological conditions that can be used for evaluating lotic impairment or establishing restoration goals is difficult, and expected biological condition must be estimated or modeled.

Directly relating biological responses, such as metrics derived from the composition of macroinvertebrate assemblages, to a combination of effects, particularly those of urbanization, is further complicated by the nonlinear response often observed between commonly used biological metrics and urbanization (Fend et al. 2005). Bivariate plots of metrics and measures of urbanization or even natural geomorphic conditions are often polygonal in form (Fausch et al. 1984; Karr and Chu 1999). These polygonal responses often arise from the presence of a generalized limiting factor plus the influence of site-specific stressors (Thomson et al. 1996).

The gradient of urbanization that increases from the outskirts to the area of maximum urban influence creates an increasingly constrained physical and chemical environmental template that sets an upper limit on potential stream quality and consequently, the lotic community. In this paper, we use the concept of polygonal distributions and factor ceilings (Thomson et al. 1996; Scharf et al. 1998) to establish a reference condition that is linearly related to and accounts for the underlying effects of this urban gradient. The reference condition identifies a potential upper limit on the condition of the benthic assemblage.

The purposes of this study were to (1) present a conceptual framework for establishing reference conditions in urban settings, (2) objectively identify reference conditions that reflect a realistic maximum biological potential that is a function of the constraints of urbanization, and (3) develop a simple bioindicator that is inexpensive to determine, has the potential to be highly comparable among programs, and reflects impairment in an urban environmental setting.

Study Area

The study location is in the Santa Clara Valley area of the San Francisco Bay region of California (Figure 1). The area is approximately 1,600 km^2. The study area is surrounded on the west, south, and east by a topographic divide and bordered on the north by San Francisco Bay. The physical setting includes upland areas that are sparsely populated and more densely populated lowland areas. Urbanization increases from the uplands to the lowlands near San Francisco Bay.

The area has a Mediterranean climate with most precipitation falling as rain during the winter and spring; almost no precipitation occurs during the summer months. Many streams are impounded, while instream withdrawals alter the natural hydrologic cycle of others. Water management is extremely complex and water is released from most impoundments for aquifer recharge, flood control, and to support sensitive species. Approximately 50% of the water used in the basin is imported.

The San Francisco Bay area contains extremely variable landforms and climates compared to many less topographically complex urban centers in the United States. Altitude ranges from sea level to more than 1,200 m and precipitation and temperature vary on a subregional basis. These variations lead to important differences in factors that influence the distribution of the local flora and fauna. Spatial differences in temperature, runoff, channel morphology, and local potential and realized vegetation create a mosaic of gradients that directly influence the habitat and resources available to lotic invertebrates throughout the area. As a result of these differences, physical habitat responses, and consequently, biotic responses to any given stressor are likely to vary both within and among regions.

Two factors confound the use of a reference site approach for determining impairment of San Francisco Bay area streams. First, humans have been significantly influencing local streams for at least 150 years; therefore, streams that could be classified as pristine or even "least-impaired" are lacking. This is particularly true for the higher order, downstream, more urbanized reaches. Second, naturally high habitat variability throughout the Bay Area leads to a diversity of potential macroinvertebrate assemblages.

Methods

Biological

Eighty-five sites located on 14 streams within the Santa Clara Valley area were sampled during May 1997 for macroinvertebrates (Figure 1). All sites were also sampled for nutrients, dissolved trace elements, and channel and riparian structure during May and June 1997. Streams included in the study were San Francisquito Creek, Corte Madera Creek, Los Trancos Creek, Stevens Creek, Saratoga Creek, Guadalupe River, Los Gatos Creek, Ross Creek, Guadalupe Creek, Alamitos Creek, Barret Creek, Arroyo Calero, Coyote Creek, and Upper Penitencia Creek.

Sampling locations were chosen to be equidistant, with sites located at approximately 2-km intervals. The most downstream site within each subbasin was located at either the point of observed or assumed intermittent flow or where a tidal influence to river flow and/or

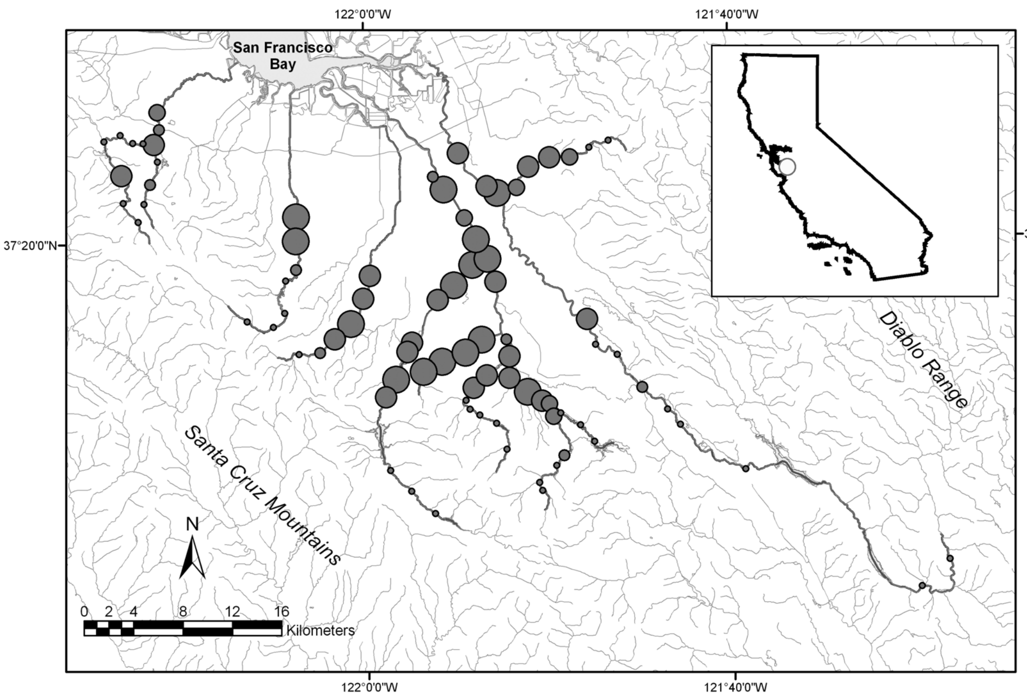

FIGURE 1. Map of study area (Santa Clara Valley, California) depicting percentage urban land cover within 200 m × 2 km buffer strips upstream of each of the 85 sites. Percentage urban land cover was derived from 30 m National Land Cover Data. Circle size increases in 10% intervals from near zero to near 100% urban land cover.

substratum type was apparent. In general, the most upstream site was at an altitude of approximately 300 m.

At each site a semiquantitative collection of macroinvertebrates was made from a single riffle. Each collection was a composite of five 0.1-m^2 kick samples taken with a 0.3-m-wide D-frame kicknet fitted with a 500-µm mesh. The five individual kick samples were taken systematically in each riffle in a downstream to upstream direction, crossing the stream twice along two perpendicular diagonal lines (i.e., a v-shape across the riffle's breadth). Two of the five samples were obtained in the thalweg, and the remaining three samples were taken between the thalweg and the margins. At sites containing extremely long riffles, sampling focused on the upstream portion of the riffle.

Samples were cleaned of large debris and preserved with 10% buffered formalin in the field. All samples were randomly subsampled in the laboratory using a gridded tray (Moulton et al. 2000). Approximately 500 organisms were sorted from each subsample using ~8× magnification. After each subsample was sorted a second person searched the just-sorted material for a maximum of 30 min to remove organisms that may have been overlooked during the initial sorting. All individuals sorted were placed in vials and stored in 80% ethanol prior to identification. All organisms were identified to the lowest practicable taxon.

Estimates of the combined number of taxa in the insect orders Ephemeroptera, Plecoptera, and Trichoptera (EPT) (EPT richness) and the relative abundance of individuals in the same three orders (% EPT) were used for the principal analyses. These estimates were based on the mean of 100 computer-generated, 300-organism subsamples. Samples were rarefied to a constant size to reduce the effect that variations in the number of organisms sorted has on comparing richness-based metrics among sites. Ephemeroptera, Plecoptera, and Trichoptera richness was further corrected by eliminating less-resolved higher-level taxa when at least one individual of a lower taxonomic designation was present in the sample.

An indicator value (EPT score) was constructed from EPT richness and percentage EPT. We restricted our analysis to these two metrics because the EPT are (1) the most often used taxa in biomonitoring and are generally considered intolerant of most stressors (Resh and Jackson 1993; Kerans and Karr 1994), (2) identified to species more often than any other taxa (Carter and Resh 2001), and (3) highly correlated to many other metrics (Lenat and Penrose 1996). Additionally, no assumptions regarding higher level ecological responses are necessary when evaluating these two metrics, as there are when evaluating functional-type metrics (e.g., functional feeding-groups) because EPT richness and percentage EPT are strictly taxonomically based.

Physical

An estimate of the degree of habitat impairment near the channel due to increasing urbanization was determined for each site. Seven factors were evaluated. The first five factors were categorized on a 1–4 ordinal scale and included (1) channel form, which ranged from natural to v-shaped concrete; (2) riparian composition (including canopy cover), which ranged from all native to absent; (3) riparian width, which ranged from greater than 30 m to absent; (4) siltation, which ranged from no obvious deposited silt to a visible silt layer over the substrate; and (5) turbidity, which ranged from clear to an inability to see the bottom.

Both canopy and embeddedness were difficult to visually categorize in the field, so ordinal values (1–4) were developed from field measurements. Riparian canopy shading was measured at three mid-channel points within the sample area, using a Solar Pathfinder (Solar Pathways, Inc.) to measure the solar arc for May; the value was expressed as a percentage of the expected insolation. Sediment embeddedness was estimated as the percentage depth (the vertical axis) to which 10 randomly chosen particles were buried in sand or finer material. The mean of all seven factors was calculated and used to represent a near-site estimation of urbanization (UHA [urban habitat assessment], see Fend et al. 2005).

Land Cover

The spatial coordinates (latitude, longitude, and altitude) were determined for each site using a Global Positioning System and topographic maps. Urban land cover for a 200 m wide × 2 km long buffer strip (100 m on either side of the stream) upstream of each site was estimated using 30 m National Land Cover Data (NLCD) (Vogelmann et al. 2001). The summation of four NLCD categories (low intensity residential, high intensity residential, commercial/industrial/transportation, quarries/strip mines/gravel pits) was used to represent urban land cover.

Statistics

The relationship between individual EPT-based metrics and the total benthic assemblage was determined by correlating the first axis of a Detrended Correspondence Analysis (McCune and Mefford 1999) derived using $\log_{10}$ transformed abundance data from the entire data set. All percentage data (% EPT and land cover) were transformed using arcsine square root transformations. Rarified richness was not further transformed.

The EPT score was formed by standardizing each metric (EPT richness and % EPT) by its maximum, thereby creating values that ranged between 0 and 1. The standardized metrics were summed and the total multiplied by 5 to provide an easily interpretable multimetric that had the potential of ranging from 0 to near 10.

Partitioned regressions were used to (1) estimate an upper boundary condition (factor-ceiling) of the bivariate distribution of the EPT score and percentage urban land cover per site, and (2) define four impairment categories. The procedure was as follows. First, a primary regression by ordinary least squares (OLS) was performed using data from all sites. This separated (partitioned) the data into those EPT scores with positive residuals and those with negative residuals. Next, a secondary regression was performed using just those data identified by the primary regression to possess positive residuals. The OLS line of this secondary regression was used as the boundary condition above which EPT scores represent the proposed reference condition for any given percentage urban land cover. The boundary represents a continuously varying reference condition that accounts for the effects of urbanized land cover on the potential magnitude of EPT scores. To complete the analysis, a final regression was fit using those EPT scores identified as having negative residuals from the primary regression. This partitioned this portion of the data into two additional groups. The final two groups, along with the two groups formed from partitioning the data with positive residuals, allowed the formation of four potential categories: one category representing the reference condition and three categories representing increasing levels of impairment.

Two single factor Analysis of Variance (ANOVA) tests were used to determine whether there were significant differences in altitude and urbanization (the UHA score) among the derived impairment categories. Among-group variances were tested using Levene's test for homogeneity of variances prior to the ANOVAs, and Newman-Keuls tests were used for post hoc testing of differences among groups when appropriate. All analyses were performed using STATISTICA (StatSoft, Inc. 2004).

Results

General Description

The spring collections yielded an abundant and diverse fauna. Total number of individuals identified from the 85 sites was 65,571. Total richness (number of different taxa) across all sites was approximately 300 taxa. Mean total richness per site based on nonrarefied samples was 44 ± 8.6 (±1 SD, n = 85) and ranged from 27 to 67. Although our collection methods were semiquantitative, estimated mean density per site was 9,590 individuals/m^2 and ranged from 1,386 to 29,581/m^2.

Random subsampling to provide at least 500 organisms yielded a mean of 653 ± 137 individuals and ranged from 470 to 992 individuals per subsample. This variation in the number of individuals sorted per sample necessitated rarefying the samples prior to comparing richness estimates among samples. Mean EPT richness per site, based on randomly sorted samples, was 10 ± 5.3 and ranged from 2 to 27. Mean EPT richness per site based on samples rarefied to 300 individuals was 8 ± 4.3 and ranged from 1.5 to 22.8.

Percentage EPT per site based on randomly sorted samples was 35 ± 19.0 and ranged from 1.2 to 78.2. As expected, rarefaction had little influence on percentage composition and mean percentage EPT per site based on samples rarefied to 300 individuals was 35 ± 19.1 and ranged from 1.3 to 78.3.

Urban land cover varied widely across the basin. Mean percentage of land cover per site classified as urban within the 200 m wide x 2 km long buffer strips was 41 ± 35.4 and ranged from 0 to 95.9.

The correlation between the benthic assemblage as represented by the first axis of the Detrended Correspondence Analysis (DCA) ordination and the EPT score (Figure 2) was relatively high (r^2 = 0.44, P < 0.001). Most scatter was related to the poor relationship between percentage EPT (r^2 = 0.09) and the first ordination axis compared to a much better relationship between EPT richness and the first axis (r^2 = 0.69). Both metrics were negatively correlated with percentage urban land cover and displayed considerable scatter that could best be described as polygonal in form (Figure 3A, B). The combined EPT score when plotted against percentage urban land cover also contained substantial scatter, which was polygonal in form (Figure 4A).

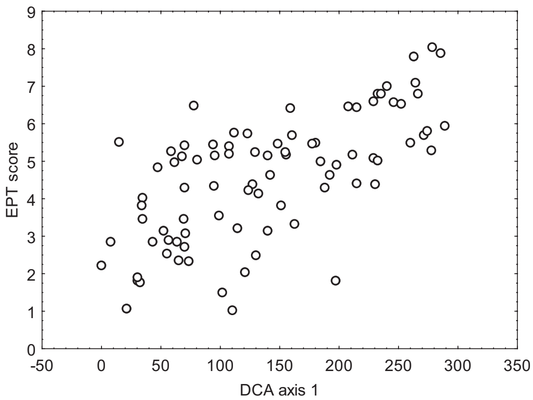

FIGURE 2. Relationship between detrended correspondence analysis (DCA) axis 1 derived from $\log_{10}$ transformed benthic data and the derived EPT score ($r^2 = 0.44$, $P < 0.001$, $n = 85$) from the Santa Clara Valley area, California.

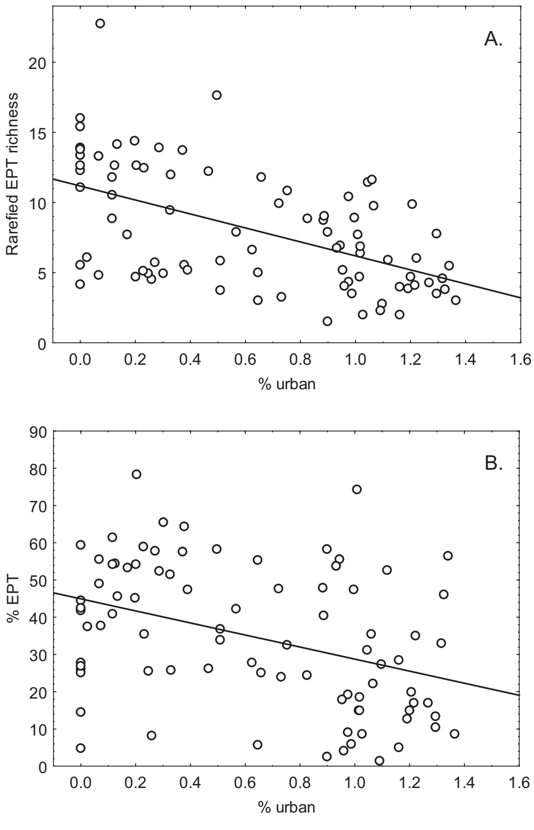

FIGURE 3. Relationship of transformed percentage urban land cover of buffer strips with (A) EPT richness rarefied to 300 individuals ($r^2 = 0.28$, $P < 0.001$, $n = 85$) and with (B) percentage EPT ($r^2 = 0.16$, $P < 0.001$, $n = 85$) for the Santa Clara Valley area, California.

Regressions

The primary OLS regression of percentage urban land cover and the EPT score partitioned the data into two near equal groups (Figure 4A). About 30% of the variation ($r^2 = 0.29$) in EPT scores was accounted for by the estimated percentage urban land cover of the site-specific buffer strips when all EPT data were used.

The OLS line formed by the secondary regression of the subset of sites with positive residuals based on the primary regression (Figure 4B) identified a conservative upper boundary condition for the EPT score (the uppermost regression line of Figure 4C, E). An EPT score above the uppermost regression line is considered to be a reference value for a given level of percentage urban land cover. The secondary regression using only those data possessing negative residuals from the primary regression (Figure 4D) provided a separation of sites into those with low EPT scores and those with even lower EPT scores over the full range of percentage urban land cover.

Assembling all three regressions results in four site-groups representing least-impaired to most-impaired conditions (n = 20, 23, 27, and 15, respectively) and provides a potential classification of impairment based solely on the relationship between EPT score and percentage urban land cover of the buffer strips (Figure 4E). A conservative potential value of the EPT score that represents the least-impaired condition as set by the factor-ceiling (the uppermost regression line of Figure 4E) varied from 6.8 at near zero urbanization to 4.0 at 100% urbanization.

Physical Characteristics of Site Groups

There was no significant difference in the variances among proposed impairment groups in altitude (Levene's test of homogeneity of variances; $F = 2.34$, $P = 0.079$). There also was no significant difference in mean altitude among groups ($F = 1.96$, $P = 0.126$). However, the site group that represented the most impaired sites had a mean altitude lower than the other three groups (Figure 5A).

There was no significant difference in the variances among proposed impairment groups in the UHA (Levene's test of homogeneity of variances; $F = 2.30$, $P = 0.083$). However, there was a significant difference among proposed impairment groups in the mean UHA ($F = 3.16$, $P = 0.029$). The most highly urbanized site group (group 4) was significantly different from the two least urbanized site groups. Although no other site groups differed, there was an apparent near-linear decrease in UHA across the four groups (Figure 5B).

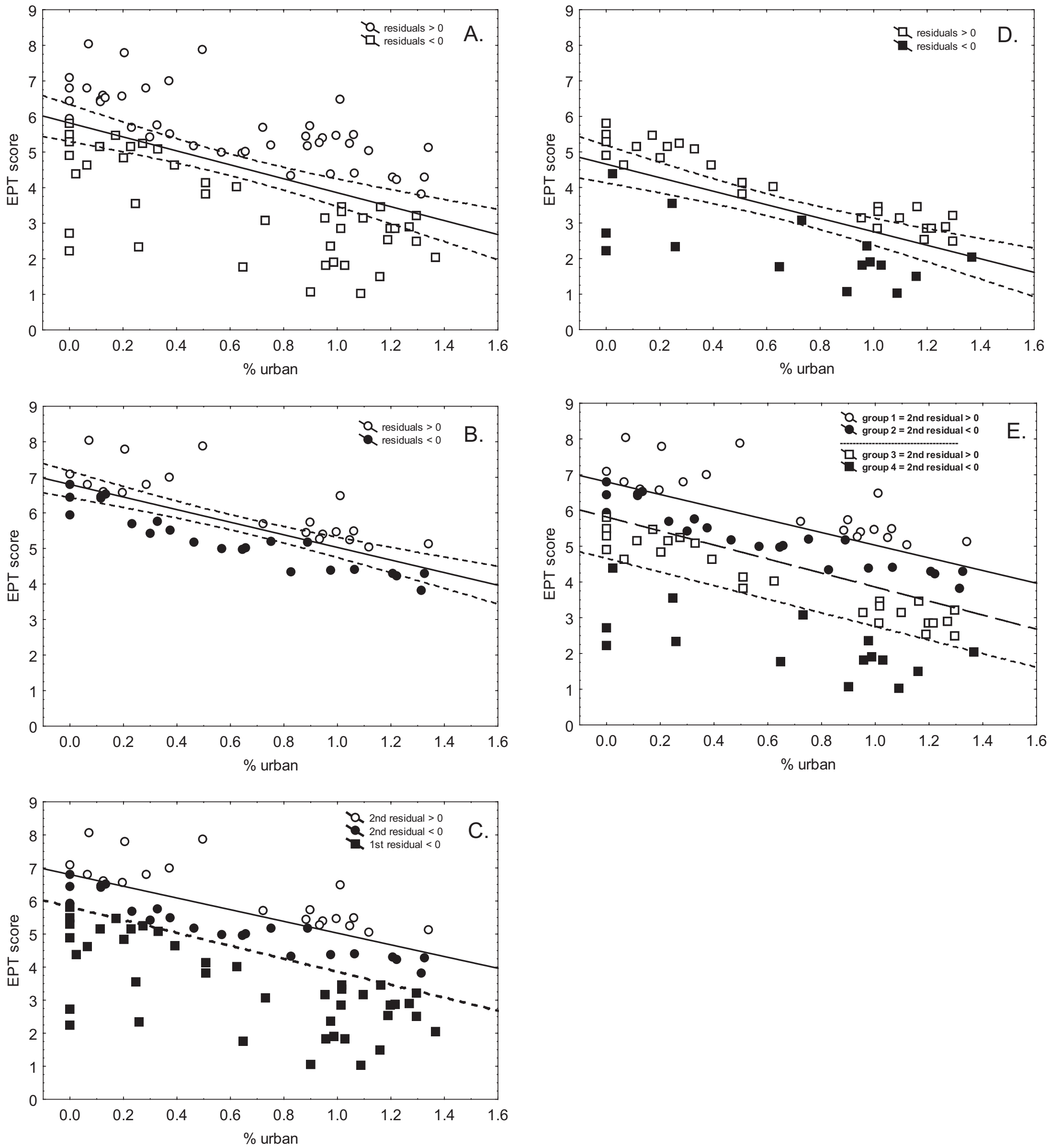

FIGURE 4. Results of primary and secondary ordinary least squares regression between transformed percentage urban land cover of buffer strips and EPT score from the Santa Clara Valley area, California: (A) primary regression ($r^2 = 0.29$, $P < 0.001$, $n = 85$), (B) secondary regression of positive residuals, (C) Solid line indicates the proposed factor-ceiling with EPT scores greater than the boundary representative of reference conditions, (D) secondary regression of negative residuals, and (E) formation of potential percentage urban-specific-impairment categories from least impaired most impaired.

Discussion

Current bioassessment methods are based on comparing test sites to reference sites or a reference condition. Whether these comparisons use the benthic assemblage directly as in RIVPACS-type models, or indirectly when metrics and multimetrics derived from the benthic assemblage (e.g., richness, percentage composition) are used, a reference condition is necessary to evaluate the biological condition of a test site. Reference conditions can be represented by a single site on a stream, as in upstream-to-down-

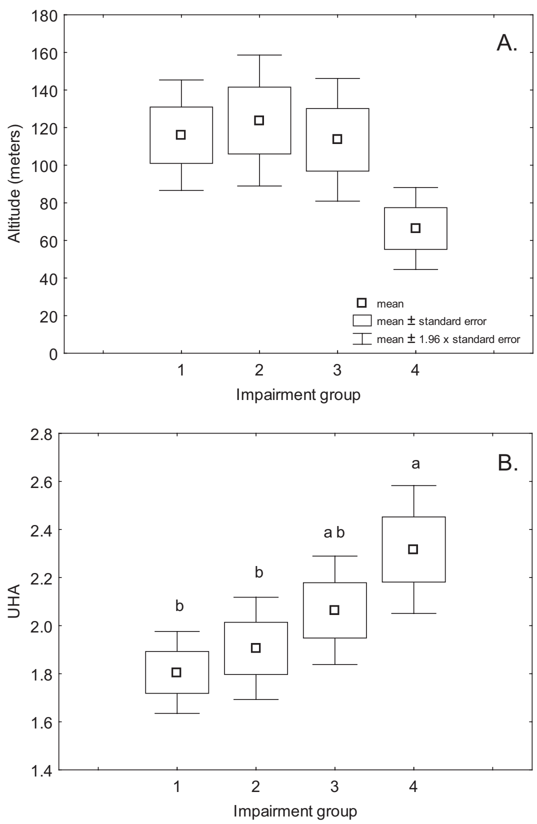

FIGURE 5. Values of (A) altitude, and (B) study-specific urban habitat assessment (UHA) for potential impairment groups as determined by single factor ANOVAs for data from the Santa Clara Valley area, California. Means with different letters are significantly different (Newman-Keuls test).

stream designs often used to identify the influence of point source impacts, although arguments have been presented against this approach (Underwood 1997; Downes et al. 2002; Bailey et al. 2004). Alternatively, reference conditions can be represented by the average biological condition among numerous sites and indicate an expected condition for a region as in larger-scale studies of impairment (Reynoldson et al. 1997; Bailey et al. 2004). Reference sites can also be used to model the expected species composition of a test site given a database of species occurrences and a limited suite of environmental variables (Wright 2000).

Large-scale studies have used various criteria for establishing reference conditions. Wright et al. (1984) initially stratified by stream size, excluding both low and high order streams. Large U.S. programs that are state specific or regional have also stratified prior to establishing reference conditions. These programs have used various methods and criteria to geographically partition study areas in an effort to control for environmental variability. Examples include the use of physiographic provinces (Lenat 1993), general landform (Yoder and Rankin 1995), and ecoregions (Barbour et al. 1999). Stratifying based on ecoregion is often a first step in the stratification process (Omernik and Bailey 1997). Whittier et al. (1988) were among the first to detail the similarities and differences in assemblage composition and structure of macroinvertebrates and other taxa within and among ecoregions of the northwestern United States.

Unfortunately, some studies showed that within-ecoregion variability in physical habitat, and consequently the potential biota, is often too high to establish logical comparisons (Hawkins et al. 2000; Fend et al. 2005). Thus, it is often necessary to subdivide ecoregions into subecoregions or stratify study areas and/or sites based on other physical characteristics such as basin size, stream order, or local conditions. Barbour et al. (1996), in an assessment of Florida streams, found that partitioning into subecoregions was necessary to adequately represent expected faunal composition and structure.

Hawkins et al. (2000) summarized numerous studies concerned with the applicability of ecoregions as a stratifying factor in water quality studies. These studies were worldwide in scope and evaluated responses of a variety of taxa. In general, ecoregions were considered too coarse a structure (species turnover was too high) to be applicable for most lotic assessments. Coincidentally, USEPA Science Advisory Board lists the "state of the science in defining ecoregions and reference areas" as one limitation in the use of biocriteria in water quality studies (SAB 1993).

A basic assumption of bioassessment designs is that reference conditions that are sufficiently comparable to test sites exist and that reference sites and test sites have a similar biological potential. The widespread degradation of surface waters (Karr and Chu 1999) has led to a dearth of pristine or pre-Columbian conditions that can be used as reference sites. Consequently, the standard of pristine conditions has been supplanted by the acceptance of least-impaired conditions for defining both reference conditions and restoration goals (Barbour et al. 1999; Karr and Chu 1999). This strategy allows assessments to be designed when least-impaired conditions can be identified. Acceptance and use of least-impaired conditions establishes a precedent for using a factor-ceiling approach to identify reference-type conditions across environmental gradients, as shown in the present study.

Urban environmental settings often display a continuum of potential stream function from near-natural potential at near-zero urbanization to extremely limited potential at the urban core (Paul and Meyer 2001). There are many well-known and obvious factors that create this gradient in potential stream function and are detailed throughout this volume. Although the fluvial hydrologic effects of some factors can be mitigated (e.g., impacts on stream chemistry), others will likely not change under most restoration scenarios. These latter factors (e.g., road corridors, canalization, high imperviousness) form a complex gradient of increasing effect that starts at the rural–urban interface and progresses to the urban core and set limits to stream function and probable restoration goals (Booth and Jackson 1997).

These limits to stream function influence the biotic potential of lotic systems along a rural–urban gradient (Allan and Flecker 1993; Morley and Karr 2002). In our study, biological condition ranged from near-zero to a higher, maximum value across the full range of urban land cover. The objectively modeled factor-ceiling represents an attainable biological potential given current urban land cover. Therefore, sites displaying an EPT score below the potential can be nominally ranked into impairment categories, while accounting for the constraints existing because of background urban land cover.

Altitude is a significant constraint on the distribution of benthic macroinvertebrates (Ward 1986) and often confounds detecting impairment in streams (Carter et al. 1996; Cuffney et al. 2005, this volume; Fend et al. 2005). Our analyses indicated that altitude was not significantly different among the impairment groups, and given the relatively large within-group sample size, similarity among groups in mean altitude was likely attributable to a large range in altitude within each impairment group. However, estimated differences in mean UHA of each group displayed an almost linear relationship among impairment groups. This indicates at least a partial decoupling of estimated urban impairment from altitude and a partial controlling of the confounding effect of altitude on interpreting impairment based on macroinvertebrate distributions in the Santa Clara Valley basin.

High species turnover along the extensive environmental gradients present in large-scale studies limits logical comparisons among sites (Hawkins and Vinson 2000). One of these gradients in lotic studies is longitudinal change (Ward 1986; Carter et al. 1996) and is predicted by the River Continuum Concept (Vannote et al. 1980). In urban studies, a second gradient in stream structure and function results from the background template of urban land use cover. This second gradient also leads to excessive species turnover, which further confounds comparing assemblages for evaluating water quality impacts. Even though this latter gradient is anthropogenic, both of these gradients are continuous in nature, which strongly argues against partitioning (stratifying) geographical areas even as small as our study basin (Hawkins et al. 2000). The coincidence of these gradients is particularly common in areas that are environmentally diverse, such as the topographically complex western United States. (see Cuffney et al. 2005).

If a continuous gradient in potential condition is not used, but a method which partitions a basin is chosen, the boundaries between partitions should be viewed as somewhat artificial. Use of rigid partitions places the onus on those who develop monitoring strategies to make the boundaries at least fuzzy if not probabilistic relative to expected response values (biological potential), such as in RIVPACS-type models (Wright et al. 2000). Conceptually, establishing a reference system by incorporating the effects of either a natural gradient or, in the case of an urban setting, an immutable anthropogenic gradient, functionally integrates, instead of ignores, the influence of environmental setting on stream structure and function. This reference condition can be represented by a modeled factor-ceiling in urban environmental settings.

We consider the urban environment part of the overall template within which bioassessments must be designed. Although some mitigation of the influence of the urban setting is possible and desirable from a cultural standpoint, there are other aspects that society probably will not change. The method we present to estimate the biological potential acknowledges many of these limitations; it seems only pragmatic to include this urban portion of the template in our assessment designs and development of potential indicators (Palmer et al. 2004).

One of our goals was to develop a simple, but effective bioindicator. The majority of bioassessment programs attempt to identify all taxa that are sorted from benthic samples collected for biomonitoring (Carter and Resh 2001). Even with this level of effort—or possibly because of it—variability exists among programs. This leads not only to poor comparability among programs (Houston et al. 2002), but also, in general, a less tractable biological response. We based our bioindicator on two metrics derived from the EPTs because we desired an indicator that was more tractable, inexpensive to determine, had a high probability of among-program comparability, and adequately responded to anthropogenic stressors. Wallace et al. (1996) showed that the EPT and the total benthic assemblage responded similarly when evaluating impairment. Our data also indicate a high correlation between the entire benthic assemblage and the derived EPT score.

Although rare in biomonitoring studies, basing the biological response on fewer taxa likely contributes to a clearer understanding of the effects of both anthropogenic and natural factors on individual metrics. Basing an indicator on fewer taxa also allows more effort to be spent on species-level identifications, which leads to a better understanding of each taxon's contribution to the chosen metric (Resh and Unzicker 1975). Most importantly, this knowledge can lead to a more mechanistic understanding of the specific indicator response (Lenat and Resh 1999). Our failure to understand these responses is currently one of the fundamental limitations to incorporating biocriteria in water quality standards (SAB 1993).

We estimated a potential indicator value using partitioned OLS regression using percentage urban land cover derived from NLCD database. Other measures of urbanization (or even other land cover types or natural gradients) as well as other bioindicators could be used in these analyses. Also, other methods of defining the upper limits of polygonal distributions such as quantile regression (Scharf et al. 1998; Cade et al. 1999) are available and may even be more appropriate in some situations. However, the method presented here can be objectively implemented with relatively little effort using readily available statistical packages.

The advantages of assessing water quality using the biota have been repeatedly stated (Rosenberg and Resh 1993; Barbour et al. 1999; Karr and Chu 1999). However, numeric criteria based on the composition and structure of lotic benthic macroinvertebrate assemblages currently exist for only a few state water quality programs (USEPA 2002b). Concerns still surround the nonimpact related effects of specific natural and anthropogenic factors, as well as the identification of reference conditions (SAB 1993; Polls 1994; Reash 1995). Clearly, a more thorough understanding of these factors is necessary. In the interim, incorporation of the effects of known limiting factors, which are often represented by continuous gradients, seems an important component of establishing practical and broadly acceptable limits for biomonitoring lotic systems. For the foreseeable future, humans will continue to impact natural systems (Palmer et al. 2004); acknowledging these influences seems a necessary component when modeling optimal ecosystem structure and function within urban systems.

Acknowledgments

This research was partially funded by the Santa Clara Valley Urban Runoff Pollution Prevention Program (project SC22.6.0) and the NAWQA program of the U.S. Geological Survey.

References

Allan, J. D., and A. S. Flecker. 1993. Biodiversity conservation in running waters: identifying the major factors that threaten destruction of riverine species and ecosystems. BioScience 43:32–43.

Bailey, R. C., R. H. Norris, and T. B. Reynoldson. 2004. Bioassessment of freshwater ecosystems: using the reference condition approach. Kluwer Academic Publishers, Boston.

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, B. D. Snyder, and J. B. Stribling. 1999. Rapid bioassessment protocols for use in streams and wadeable rivers: periphyton, benthic macroinvertebrates and fish, 2nd edition. U.S. Environmental Protection Agency, Office of Water, Washington, D.C.

Booth, D. B., and C. R. Jackson. 1997. Urbanization of aquatic systems: degradation, thresholds, storm-water detection, and the limits of mitigation. Journal of the American Water Resources Association 33:1077–1090.

Cade, B. S., J. W. Terrell, and R. L. Schroeder. 1999. Estimating effects of limiting factors with regression quantiles. Ecology 80:311–323.

Carter, J. L., S. V. Fend, and S. S. Kennelly. 1996. The relationships among three habitat scales and stream benthic invertebrate community structure. Freshwater Biology 35:109–124.

Carter, J. L., and V. H. Resh. 2001. After site selection and before data analysis: sampling, sorting, and laboratory procedures used in stream benthic macroinvertebrate monitoring programs by USA state agencies. Journal of the North American Benthological Society 20:658–682.

Corkum, L. D. 1989. Patterns of benthic invertebrate assemblages in rivers of northwestern North America. Freshwater Biology 21:191–205.

Cuffney, T., H. Zappia, E. Giddings, and J. Cole. 2005.

Effects of urbanization on benthic macroinvertebrate assemblages in contrasting environmental settings: Boston, Massachusetts; Birmingham, Alabama; and Salt Lake City, Utah. Pages 361–407 *in* L. R. Brown, R. Gray, R. M. Hughes, and M. R. Meador, editors. Effects of urbanization on stream ecosystems. American Fisheries Society, Symposium 47, Bethesda, Maryland.

Davies, P. E. 2000. Development of a national river bioassessment system (AUSRIVAS) in Australia. Pages 1–24 *in* J. F. Wright, D. W. Sutcliffe, and M. T. Furse, editors. Assessing the biological quality of fresh waters: RIVPACS and other techniques. Freshwater Biological Association, Ableside, Cumbria, UK.

Downes, B. J., L. A. Barmuta, P. G. Fairweather, D. P. Faith, M. J. Keough, P. S. Lake, B. D. Mapstone, and G. P. Quinn. 2002. Monitoring ecological impacts: concepts and practice in flowing waters. Cambridge University Press, Cambridge, UK.

Fausch, K. D., J. R. Karr, and P. R. Yant. 1984. Regional application of an index of biotic integrity based on stream fish communities. Transactions of the American Fisheries Society 113:39–55.

Fend, S. V., J. L. Carter, and F. R. Kearns. 2005. Relationships of field habitat measurements, visual habitat indices, and land cover to benthic macroinvertebrates in the urbanized streams of the Santa Clara Valley, California. Pages 193–212 *in* L. R. Brown, R. H. Gray, R. M. Hughes, and M. R. Meador, editors. Effects of urbanization on stream ecosystems. American Fisheries Society, Symposium 47, Bethesda, Maryland.

Gilliom, R. J., W. M. Alley, and M. E. Gurtz. 1995. Design of the National Water-Quality Assessment Program: occurrence and distribution of water-quality conditions. U.S. Geological Survey Circular 1112. Sacramento, California.

Gilliom, R. J., P. A. Hamilton, and T. L. Miller. 2001. The National Water-Quality Assessment Program – entering a new decade of investigations. U.S. Geological Fact Sheet 071–01, Reston, Virginia.

Green, R. H. 1979. Sampling design and statistical methods for environmental biologists. Wiley, New York.

Hawkins, C. P., R. H. Norris, J. Gerritsen, R. M. Hughes, S. K. Jackson, R. K. Johnson, and R. J. Stevenson. 2000. Evaluation of the use of landscape classifications for the prediction of freshwater biota: synthesis and recommendations. Journal of the North American Benthological Society 19:541–556.

Hawkins, C. P., and M. R. Vinson. 2000. Weak correspondence between landscape classifications and stream invertebrate assemblages: implications for bioassessment. Journal of the North American Benthological Society 19:501–517.

Houston, L., M. T. Barbour, D. Lenat, and D. Penrose. 2002. A multi-agency comparison of aquatic macroinvertebrate-based stream bioassessment methodologies. Ecological Indicators 1:279–292.

Hughes, R. M. 1995. Defining acceptable biological status by comparing with reference conditions. Pages 31–47 *in* W. S. Davis and T. P. Simon, editors. Biological assessment and criteria: tools for water resource planning and decision making. Lewis Publishers, Washington, D.C.

Hughes, R. M., D. P. Larsen, and J. M. Omernik. 1986. Regional reference sites: a method for assessing stream potentials. Environmental Management 10:629–635.

Karr, J. R. and E. W. Chu. 1999. Restoring life in running waters: better biological monitoring. Island Press, Washington, D.C.

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.

Lenat, D. R., and D. L. Penrose. 1996. History of the EPT taxa richness metric. Bulletin of the North American Benthological Society 13:305–307.

Lenat, D. R., and V. H. Resh. 1999. Taxonomy and stream ecology - the benefits of genus - and species-level identifications. Journal of the North American Benthological Society 20:287–298.

McCune, B., and M. J. Mefford. 1999. PC-ORD. Multivariate analysis of ecological data, Version 4. MjM Software Design, Gleneden Beach, Oregon.

Morley, S. A., and J. R. Karr. 2002. Assessing and restoring the health of urban streams in the Puget Sound basin. Conservation Biology 16:1498–1509.

Moulton S. R., II, J. L. Carter, S. A. Grotheer, T. F. Cuffney, and T. M. Short. 2000. Methods of analysis by the U.S. Geological Survey National Water Quality Laboratory - processing, taxonomy, and quality control of benthic macroinvertebrate samples. U.S. Geological Survey, Open-File Report 00–212, Denver.

Omernik, J. M., and R. G. Bailey. 1997. Distinguishing between watersheds and ecoregions. Journal of the American Water Resources Association 33:935–949.

Palmer, M., E. Bernhardt, E. Chornesky, S. Collins, A. Dobson, C. Duke, B. Gold, R. Jacobson, S. Kingsland, R. Kranz, M. Mappin, M. Martinez, F. Micheli, J. Morse, M. Pace, M. Pascual, S. Palumbi, O. J. Reichman, A. Townsend, and M. Turner. 2004. Ecology for a crowded planet. Ecological science and sustainability for a crowded planet: 21st century vision and action plan for the Ecological Society of America. Available: www.esa.org/ecovisions (September 2004).

Paul, M. J., and J. L. Meyer. 2001. Streams in the urban landscape. Annual Review of Ecology and Systematics 32:333–365.

Polls, I. 1994. How people in the regulated community view biological integrity. Journal of the North American Benthological Society 13:598–604.

Reash, R. J. 1995. Biocriteria: a regulated industry perspective. Pages 153–166 *in* W. S. Davis and T. P. Simon, editors. Biological assessment and criteria: tools for water resource planning and decision making. Lewis Publishers, Washington, D.C.

Resh, V. H., and J. K. Jackson. 1993. Rapid assessment approaches to biomonitoring using benthic macroinvertebrates. Pages 195–233 *in* D. M. Rosenberg and V. H. Resh, editors. Freshwater biomonitoring and benthic macroinvertebrates. Routledge Chapman and Hall Inc, New York.

Resh, V. H., and J. D. Unzicker. 1975. Water quality monitoring and aquatic organisms: the importance of species identification. Journal of the Water Pollution Control Federation 47:9–19.

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.

Rosenberg, D. M., and V. H. Resh. 1993. Introduction to freshwater biomonitoring and benthic macroinvertebrates. Pages 1–9 *in* D. M. Rosenberg and V. H. Resh, editors. Freshwater biomonitoring and benthic macroinvertebrates. Chapman and Hall, New York.

SAB (Science Advisory Board). 1993. An SAB report: evaluation of draft technical guidance on biological criteria for streams and small rivers. Biological Criteria Subcommittee of the Ecological Processes and Effects Committee. 1993. EPA-SAB-EPEC-94–003 Science Advisory Board, Washington, D.C.

Scharf, F. S., F. Juanes, and M. Sutherland. 1998. Inferring ecological relationships from the edges of scatter diagrams: comparison of regression techniques. Ecology 79:448–460.

Simpson, J. C., and R. H. Norris. 2000. Biological assessment of river quality: development of AUSRIVAS models and outputs. Pages 125–142 *in* J. F. Wright, D. W. Sutcliffe, and M. T. Furse, editors. Assessing the biological quality of fresh waters: RIVPACS and other techniques. Freshwater Biological Association, Ableside, Cumbria, UK.

StatSoft, Inc. 2004. STATISTICA (data analysis software system), Version 6. Available: www.statsoft.com (September 2004).

Thomson, J. D., G. Weiblen, B. A. Thomson, S. Alfaro, and P. Legendre. 1996. Untangling multiple factors in spatial distributions: lilies, gophers, and rocks. Ecology 77:1698–1715.

Underwood, A. J. 1997. Experiments in ecology: their logical design and interpretation using analysis of variance Cambridge University Press, Cambridge, UK.

USEPA (U.S. Environmental Protection Agency). 2002a. Research strategy: environmental monitoring and assessment program. U.S. Environmental Protection Agency, EPA-620/R-02/002, Washington, D.C.

USEPA (U.S. Environmental Protection Agency). 2002b. Summary of biological assessment programs and biocriteria development for states, tribes, territories, and interstate commissions: streams and wadeable rivers. U.S. Environmental Protection Agency, EPA-822-R-02–048, Washington, D.C.

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.

Vogelmann, J. E., S. M. Howard, L. Yang, C. R. Larson, B. K. Wylie, and N. Van Driel. 2001. Completion of the 1990s National Land Cover Data Set for the conterminous United States from Landsat Thematic Mapper data and ancillary data sources. Photogrammetric Engineering & Remote Sensing 67:650–662.

Wallace, J. B., J. W. Grubaugh, and M. R. Whiles. 1996. Biotic indices and stream ecosystem processes: results from an experimental study. Ecological Applications 6:140–151.

Ward, J. 1986. Altitudinal zonation in a Rocky Mountain stream. Archiv fur Hydrobiologie Supplement 74:133–199.

Whittier, T. T., R. M. Hughes, and D. P. Larsen. 1988. Correspondence between ecoregions and spatial patterns in stream ecosystems in Oregon. Canadian Journal of Fisheries and Aquatic Sciences 45:1264–1278.

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 macroinvertebrate species and prediction of community type using environmental data. Freshwater Biology 14:221–256.

Wright, J. F. 2000. An introduction to RIVPACS. Pages 1–24 *in* J. F. Wright, D. W. Sutcliffe, and M. T. Furse, editors. Assessing the biological quality of fresh waters: RIVPACS and other techniques. Freshwater Biological Association, Ableside, Cumbria, UK.

Wright, J. F., D. W. Sutcliffe, and M. T. Furse. 2000. Assessing the biological quality of fresh waters:

RIVPACS and other techniques. Freshwater Biological Association, Ableside, Cumbria, UK.

Yoder, C. O. and E. T. Rankin. 1995. Biological criteria program development and implementation in Ohio. Pages 109–144 *in* W. S. Davis and T. P. Simon, editors. Biological assessment and criteria: tools for water resource planning and decision making. Lewis Publishers, Washington, D.C.