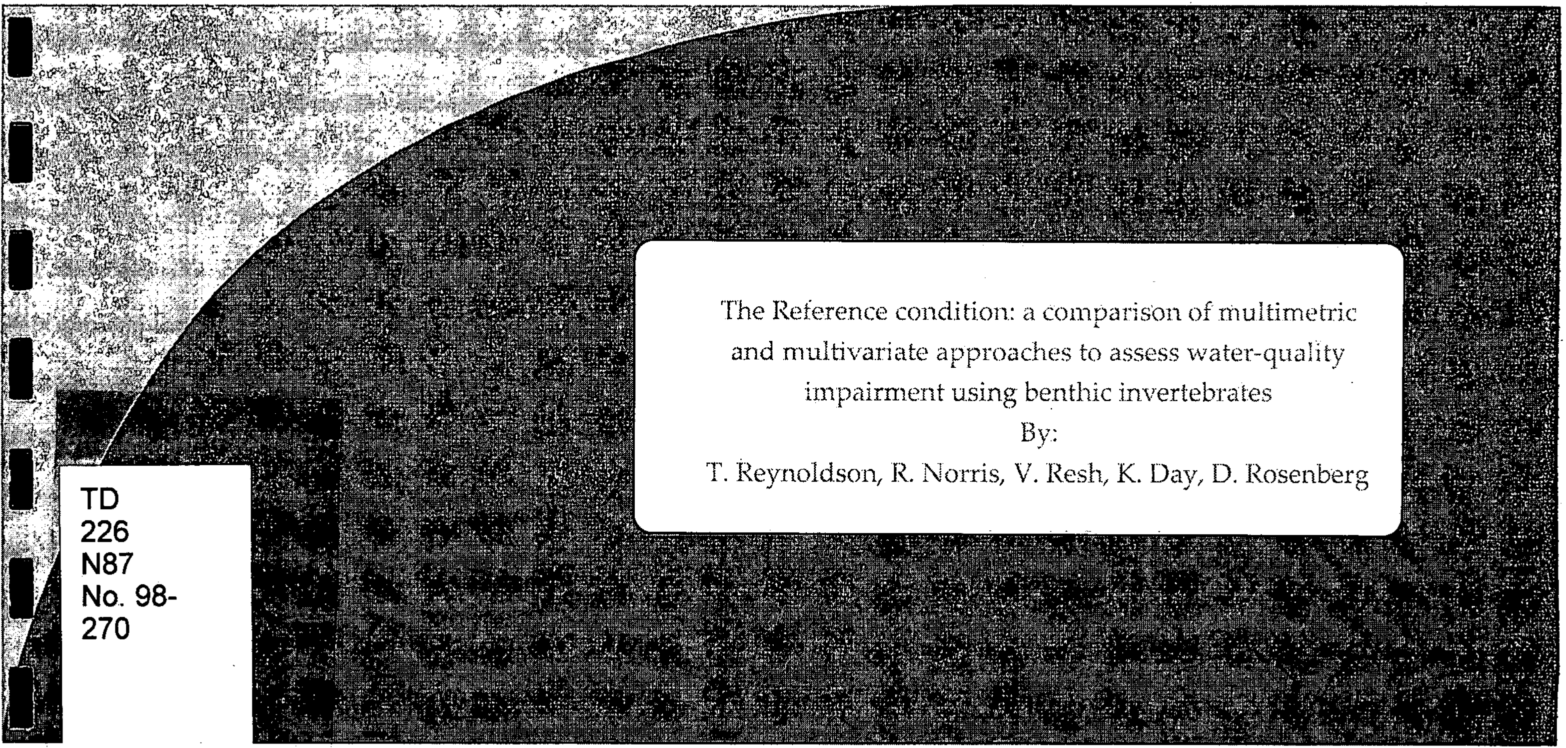
98-270

Environment Canada
Water Science and Technology Directorate

Direction générale des sciences et de la technologie, eau
Environnement Canada



The Reference condition: a comparison of multimetric and multivariate approaches to assess water-quality impairment using benthic invertebrates

By:

T. Reynoldson, R. Norris, V. Resh, K. Day, D. Rosenberg

18-270

MANAGEMENT PERSPECTIVE

Title The reference condition: a comparison of multimetric and multivariate approaches to assess water-quality impairment using benthic invertebrates

Authors T.B. Reynoldson, R.H. Norris, V.H. Resh, K.E. Day and D.M. Rosenberg

NWRI Publication No: 98-270

Citation: Journal N. American Benthological Society 16:

EC Priority/Issue: Current EEM programmes use traditional methods of establishing control sites in assessing the performance of federal regulations. The reference-condition approach offers a powerful alternative because sites serve as replicates rather than the multiple collections within sites that are the replicates in traditional designs using inferential statistics. With the reference-condition approach, an array of reference sites characterises the biological condition of a region; a test site is then compared to an appropriate subset of the reference sites, or to all the reference sites with probability weightings.

Current status: This paper compares the procedures for establishing reference conditions, and assesses the strengths and deficiencies of multimetric (as used in the USA) and multivariate methods (as used in the UK, Canada, and Australia) for establishing water quality status. A data set of environmental measurements and macroinvertebrate collections from the Fraser River, British Columbia, was used in the comparison. Precision and accuracy of 2 multivariate methods were consistently higher than for the multimetric assessment. The complementary emphases in the multivariate methods examined (presence/absence in AUSRIVAS cf. abundance in BEAST) lead us to recommend that they be used together.

Next steps NWRI is promoting the use of the reference condition approach as the basis of a national reference site data base programme that could be used in different EEM programmes and to address other site specific aquatic ecosystem helath issues.

J. N. Am. Benthol. Soc., 1997, 16(4):833–852

The reference condition: a comparison of multimetric and multivariate approaches to assess water-quality impairment using benthic macroinvertebrates

T. B. REYNOLDSON[1], R. H. NORRIS[2], V. H. RESH[3], K. E. DAY[1], AND
D. M. ROSENBERG[4]

[1]*Canada Centre for Inland Waters, 867 Lakeshore Rd., P.O. Box 5050, Burlington, Ontario,
Canada L7R 4A6*

[2]*Cooperative Research Centre for Freshwater Ecology, University of Canberra, P.O. Box 1, Belconnen,
Australian Capital Territory 2616 Australia*

[3]*Department of Environmental Science, Policy and Management, 201 Wellman Hall,
University of California, Berkeley, California 94720 USA*

[4]*Freshwater Institute, Department of Fisheries and Oceans, 501 University Crescent, Winnipeg,
Manitoba, Canada R3T 2N6*

Abstract. Traditional methods of establishing control sites in field-oriented biomonitoring studies of water quality are limited. The reference-condition approach offers a powerful alternative because sites serve as replicates rather than the multiple collections within sites that are the replicates in traditional designs using inferential statistics. With the reference-condition approach, an array of reference sites characterises the biological condition of a region; a test site is then compared to an appropriate subset of the reference sites, or to all the reference sites with probability weightings. This paper compares the procedures for establishing reference conditions, and assesses the strengths and deficiencies of multimetric (as used in the USA) and multivariate methods (as used in the UK, Canada, and Australia) for establishing water-quality status. A data set of environmental measurements and macroinvertebrate collections from the Fraser River, British Columbia, was used in the comparison. Precision and accuracy of the 2 multivariate methods tested (AUStralian RIVer Assessment Scheme: AusRivAS, BEnthic Assessment of SedimenT: BEAST) were consistently higher than for the multimetric assessment. Classification by ecoregion, stream order, and biotic group yielded precisions of 100% for the AusRivAS, 80–100% for the BEAST, and 40–80% for multimetrics; and accuracies of 100%, 100%, and 38–88%, respectively. Multimetrics are attractive because they produce a single score that is comparable to a target value and they include ecological information. However, not all information collected is used, metrics are often redundant in a combination index, errors can be compounded, and it is difficult to acquire current procedures. Multivariate methods are attractive because they require no prior assumptions either in creating groups out of reference sites or in comparing test sites with reference groups. However, potential users may be discouraged by the complexity of initial model construction. The complementary emphases in the multivariate methods examined (presence/absence in AusRivAS cf. abundance in BEAST) lead us to recommend that they be used together, and in conjunction with, multimetric studies.

Key words: water-quality assessment, reference condition, multimetrics, multivariate analysis, benthic macroinvertebrates, pollution.

Fundamental to the scientific method is the use of controls or control conditions against which results obtained under test conditions are compared. In laboratory experiments, all variables are controlled except the variable of interest, which usually is set at several levels; replicates are randomly assigned to the different treatment levels. In field experiments, all variables cannot be controlled and replicates cannot be randomly assigned to treatments, so an attempt is made to choose test and control conditions (often represented by different sites) that are as similar as possible; the variable of interest is then manipulated, and uncontrolled variables are assumed to fluctuate similarly. The actual choice of separate sites in the field that are similar in all aspects and that can be divided into control and experimental groups is difficult. Traditionally, this problem has been solved in aquatic studies by choosing adjacent sites in streams (i.e., upstream and downstream comparisons, Norris et al. 1982), by dividing lakes into halves (Schindler 1974), or by using mesocosms (Graney et al. 1984). Such approaches have several problems (Cooper and Barmuta 1993); a major one in streams is confounded designs (Eberhardt 1978), often called "pseudoreplication" (Hurlbert 1984).

How have control sites been used in past freshwater studies? Voshell et al. (1989) and Resh and McElravy (1993) examined study designs described in articles on 45 lotic and 45 lentic studies of water-pollution effects and benthic macroinvertebrates that had been recently published in scientific journals. The 2 studies revealed that: 1) 63% of lotic and 26% of lentic studies involved a spatial comparison within the same water body; 2) 15% of lotic and 28% of lentic studies used a spatial comparison in different water bodies; and 3) 22% of lotic and 46% of lentic studies did a temporal (i.e., before and after) comparison. However, current trends suggest that reliance on only 1 or a few field sites as controls is becoming less common because of: 1) limited capacity for extrapolation to other sites; 2) limited ability to calculate variance estimates; and 3) a need to address increasingly common non-point-source factors, such as species introductions and habitat alteration, rather than point-source problems (Hughes 1995).

A recent development in water-quality monitoring has been the attempt to describe **reference conditions** based on pre-established criteria that exist at a wide range of sites rather than relying on information from 1 or a few control sites. These reference conditions then serve as the control against which test-site conditions are compared. The notion of **reference condition** is really one of **best available condition** and it is represented by information from numerous sites.

The concept of a reference condition is a critical element in approaches now being developed for biomonitoring and bioassessment of aquatic resources. For example, the reference condition is central to currently accepted ideas of "biocriteria" being developed by the US Environmental Protection Agency (EPA) (Davis and Simon 1995). The same approach has been used in the UK for river classification and water-quality assessment (Wright 1995), is currently being used in Canada to develop sediment guidelines for the Great Lakes (Reynoldson et al. 1995), and is the basis for the National River Health Program in Australia (Parsons and Norris 1996).

The purpose of this paper is: 1) to formalize what is meant by the reference-condition approach; 2) to compare and contrast 2 major methods of establishing and testing reference conditions, the USA multimetric approach and multivariate classifications used in the UK, Australia, and Canada; and 3) to identify strengths and deficiencies of each method.

What is the reference condition ?

Many approaches have been used to describe reference conditions (Table 1). In this paper, we emphasise techniques for establishing **regional reference sites**. The other approaches described in Table 1 have less applicability than regional reference sites in water-quality monitoring programs and eventually may be superseded by the approaches described here.

We define **the reference condition as the condition that is representative of a group of minimally disturbed sites organized by selected physical, chemical, and biological characteristics.** The reference condition is used by comparing the biological attributes of individual test sites with a group of reference sites expected to be similar. The reference-condition approach differs fundamentally from other approaches commonly used for water quality assessments (e.g., traditional studies using Before After Control Impact designs and ANOVA) in that sites, rather than multiple collections within sites, serve as replicates. An advantage of using our definition of the reference condition is that, after reference sites have been grouped by some method (e.g., classification using biota), independent data (e.g., physical and chemical) can be used to match test sites to the most appropriate group of reference sites for bioassessment.

Establishing regional reference conditions and determining underlying assumptions

The critical feature of the reference-condition approach is that it uses an array of reference sites that characterises the potential biological conditions in a region for which assessments are to be made. A test site is subsequently compared to what is deemed either the most appropriate subset of the reference sites or to all the reference sites using probability weightings (i.e., a test site is assigned a probability of belonging to each group of reference sites). Selection of the

TABLE 1. Summary of approaches for determining reference conditions based on Hughes (1995) and Johnson et al. (1993).

Approach	Where used	Application	Limitations
Regional reference sites	UK, USA, Canada, New Zealand; for lakes, streams, wetlands	Ordination and indicator analyses are used to determine representativeness of reference sites; acceptable levels of disturbance must be determined	Ecoregions are difficult to apply to wetlands; aquatic ecoregions are applicable to whole faunal assembly but there is some difficulty in applying ecoregions to particular communities; habitat classification still needed
Historical data	Eastern and midwestern USA streams and lakes	Useful if sites have been periodically resampled, or if making general statements about conditions	Usually limited to a single invertebrate community; often, comparisons with historical data only can reflect serious deterioration; data incomplete or methods sometimes unknown; frequency of collection often masks normal variation
Paleoecological data	Lakes and large rivers throughout the world	Essentially limited to lakes, diatoms, and chironomids	Poorly suited to streams, reservoirs, and wetlands; diatoms, the most widely used group, are affected by changes in water quality but perhaps less from changes in habitat structure or introduced species
Biotic indices	World wide	For comparison with a predetermined hierarchy of values	Conditions represented by indices may not be obtainable because of habitat differences; tolerances usually developed for organic contamination
Experimental laboratory data	Not widely used	Relationships between test species and some stressors (specific toxins, temperatures, etc.) are well known so field data may be used to exclude some reference sites	Data not applicable to entire invertebrate communities and are unsuitable for systems disturbed by other stressors
Quantitative methods	Some studies in USA	Plotting metric or index values against disturbance or natural variables can establish reference conditions through curve fitting	Outliers, uneven distribution of data, and absence of data from minimally disturbed sites can distort models
Best professional judgment	Pacific Northwest of USA	Convening expert panels to determine reference conditions and peer review of data and conclusions are the usual basis for this approach	Value of judgment is a function of scientists' expertise and the quality of data supplied to them

appropriate subset can be done using a number of techniques described below.

The different types of biological condition that can exist within a region require classification of types of sites in streams or lakes to establish the expected condition at a test site, which is then compared with reference-site characteristics. It is important that classification methods place reference sites into groups with similar habitat and invertebrate community characteristics because comparisons need to be made where site attributes are expected to yield similar invertebrate communities in the absence of disturbance.

Stream classification schemes currently are based mainly on geomorphological features and the creation of ecoregions (e.g., Hughes and Larsen 1988, Plafkin et al. 1989, Hughes et al. 1990, Omernik and Griffith 1991), or multivariate analysis of biological features (e.g., Wright et al. 1984, Moss et al. 1987, Parsons and Norris 1996). Lake classification schemes have also used multivariate analysis of biological features (Johnson and Wiederholm 1989, Reynoldson et al. 1995).

The establishment of ecoregions based on geomorphological characteristics (Omernik 1987) or subecoregion groups based on professional judgement (Gerritsen 1995), within which comparisons are made, is common in the Rapid Biological Assessment approach as developed by the US EPA (Plafkin et al. 1989). This approach assumes that test-site characteristics match the chosen ecoregion reference sites exactly. However, there seems to be little evidence that invertebrate communities show high levels of similarity within such regions (Corkum 1990, 1991, Richards et al. 1993).

Individual sites are grouped according to faunal characteristics in methods used in the UK (Wright et al. 1984), Canada (Reynoldson et al. 1995), and Australia (Parsons and Norris 1996). Such an approach can provide an objective way of creating groups of reference sites with which to compare test sites having similar characteristics. It also avoids the need to determine ecoregions, which may not provide similar benthic communities. The multivariate approach does not assume that test sites exactly match reference site groups, but instead calculates the probability of belonging to each of the groups. These probabilities may be incorporated in subsequent analyses that assess test sites (Wright et al. 1984, Parsons and Norris 1996).

Analytical approaches for comparisons with reference conditions

Major types of data analysis involving reference conditions include the use of biotic indices with pre-established thresholds (Metcalfe-Smith 1994), multimetric indices (Gerritsen 1995), and taxonomic prediction using multivariate analysis (Wright 1995). Biotic indices have the longest history; they have been widely used and codified in legislation in several European countries (Metcalfe-Smith 1994). They are not considered further because they do not use reference conditions as defined here (i.e., based on particular sites).

The other 2 approaches to assessing the degree of disturbance at a test site are being widely used: multimetric indices in the USA and multivariate methods in the UK, Canada, and Australia. Gerritsen (1995) maintains that additive multimetric indices that are developed specifically for assessment and management of environmental quality are sensitive to biological degradation and function well when developed from reference data bases. In contrast, he maintains that multivariate methods are more complex, require specialized practitioners, and are difficult to convey to managers and the public. In defense of multivariate methods, Norris (1995) has argued that predictive models developed from multivariate analysis of reference data bases are effective in assessing water quality (e.g., Wright 1995), and that these methods have been incorporated into interactive computer systems for use by managers, in which the complexities of model construction are hidden.

In practice, multimetric and multivariate approaches differ considerably in determination of whether a test site is equivalent to the reference condition (Fig. 1). However, both methods begin from the same premise and require the same data. As commonly used, multimetric methods classify reference sites based on geographic and physical attributes, whereas the multivariate approaches classify sites using multivariate analysis of the macroinvertebrate fauna. For the multivariate methods, selection of the most appropriate group of reference sites to which test sites are compared (BEAST [BEnthic Assessment of SedimenT] model), or comparison of a test site to all the reference sites with probability weightings (AusRivAS [AUStralian RIVer Assessment Scheme] model) is based on a predictive model. This selection is generally based on the location of the site (e.g., the ecoregion) when using the multimetric approach. Finally, when comparing the test site with the reference condition (as described by the reference sites) the multimetric approach uses taxa counts and assumptions about the taxa to derive a set of indices, whereas the multivariate approach uses only taxa counts.

Gerritsen (1995) suggests that a lack of consensus about which multivariate approaches are the most reliable demonstrates that the use of these techniques for management of resources may be premature. However, there is also continuing debate on the use of multimetrics (e.g., Hannaford and Resh 1995); and the philosophy underlying the multivariate approach, regardless of the different methods used, has been suc-

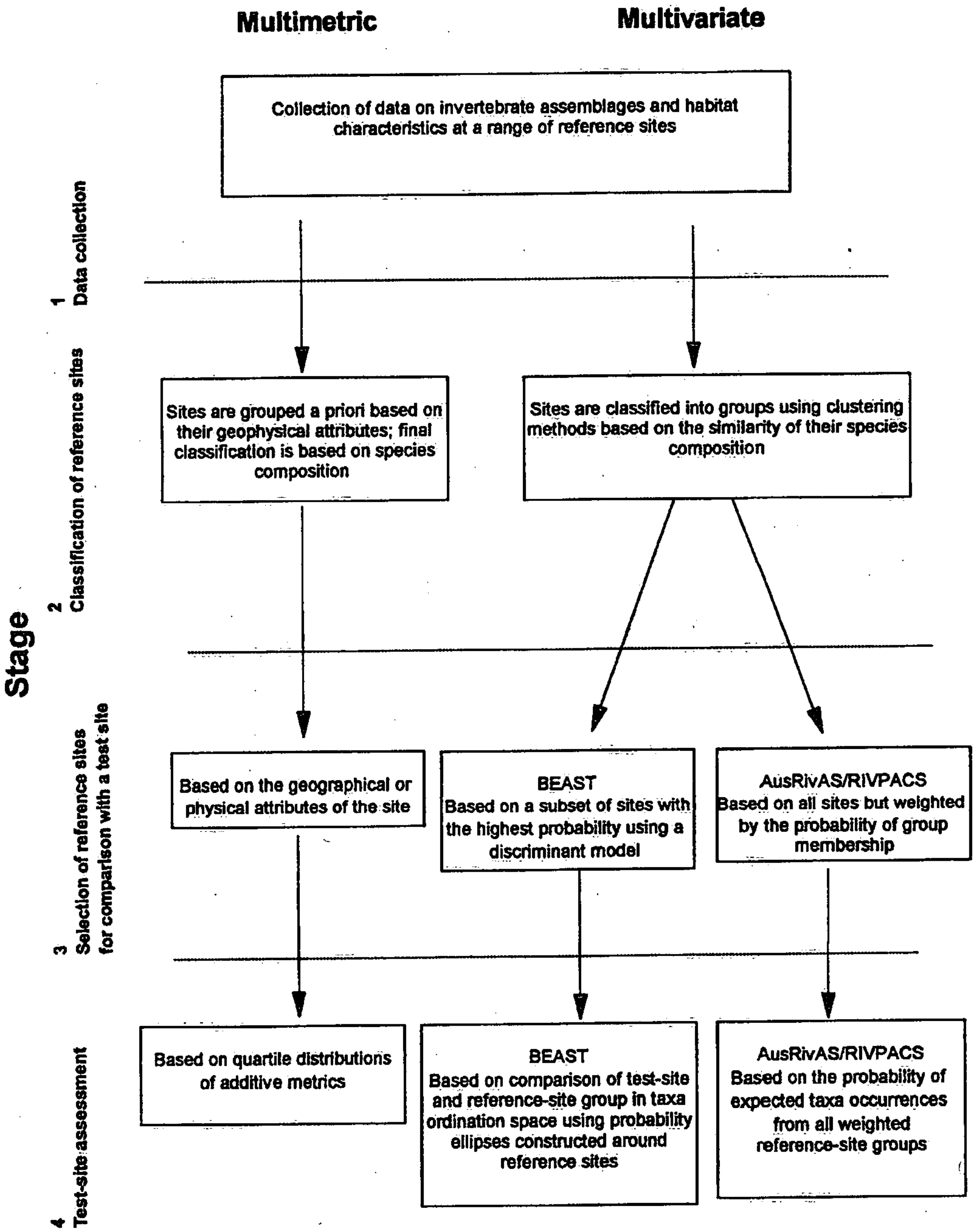


FIG. 1. Flowchart of assessment methods using multimetric and multivariate approaches. AusRivAS = AUStralian RIVer Assessment Scheme; BEAST = BEnthic Assessment of SedimenT; RIVPACS = River InVertebrate Prediction And Classification System.

cessfully applied in the UK (Wright 1995), Canada (Reynoldson et al. 1995), and Australia (Parsons and Norris 1996).

The debate on choosing multimetric analysis or multivariate analysis options is far from over. In the examples discussed below, we examine features of both approaches with the idea that a common ground may emerge.

The multimetric approach to establishing reference conditions.—Barbour et al. (1995, 1996) describe the multimetric approach to establishing reference conditions, often referred to as establishing "biocriteria". Reference sites are chosen from streams with small catchments representing, and completely within, subecoregions (Barbour et al. 1995). The ecoregions and subecoregions used in this approach are predefined largely using geomorphological characteristics such as climate, physiography, geology, soils, and vegetation (Omernik 1987). Streams are then placed into a small number of groups that account for substantial variation in the metrics (i.e., the individual measures that contribute to the additive index) used for data analysis (Barbour et al. 1995).

Two alternative classification systems for Florida were tested by Barbour et al. (1996): 1) geographic regions or ecoregions, and 2) hydrologic and dominant stream chemistry. Barbour et al. (1996) then used 3 statistical approaches to test the predetermined classification groups: 1) multiple discriminant analysis (MDA) of the metric values; 2) refinement of the classification using ordination of the species-composition data; and 3) visual examination of box-and-whisker plots of metric values. As a result, sites were aggregated into 3 groups of subecoregions (called "bioregions"), which provided the lowest error rate (8% for the Florida data) for classification of reference sites into these predetermined groups. The classification error rate using MDA was higher (29%) for an alternative allocation into 9 bioregions but, although not pointed out by Barbour et al. (1996), it was proportionally about the same as that obtained for the 3 original bioregions. Larger numbers of groups almost inevitably increase error rates simply because of an increase in possibilities of group membership, but larger numbers also result in greater sensitivity. The other 2 statistical approaches were eventually only used to verify outlying sites and to test the metrics (Barbour et al. 1996). Metrics were chosen that were consistent within a bioregion (Barbour et al. 1996), and these were then used to evaluate water quality at test sites within the same bioregion.

In summary, Barbour et al. (1995, 1996) classified sites biogeographically, where separation of the regions was adjusted to maximize the differences in the biota among the regions. Classification of test sites was then geographic; a site either was or was not in a given region. Rather than probability of class membership, this classification and assessment method explicitly considered the variability of the population of reference sites within a region. However, the level of grouping was intuitive; it was tested using MDA without cognisance of the effect of the number of groups or of the probabilities of a test site belonging to each of the reference-site groups (Gerritsen 1995). The reference-site groups were then broken down into subecoregions, which were grouped again into bioregions; the strength of these bioregions was tested with MDA. It was assumed that sites to be tested within each bioregion matched with reference sites in that bioregion. Thus, the reference condition to which a test site was matched was determined using geomorphological features in the initial classification.

The multivariate approach to establishing reference conditions.—The multivariate approach to establishing reference conditions differs from the multimetric approach in that it makes no a priori assumptions about the similarity of invertebrate communities at different sites based on either physical or chemical descriptions. Rather, faunal data are used to group sites that have similar taxonomic composition, thus providing an objective way of grouping reference sites with similar invertebrate assemblages.

A method is required that matches a test site to the appropriate reference group once reference sites have been classified into groups based on the similarity of their invertebrate fauna. Clearly, if a test site can be associated with a group of reference sites representing the reference condition, then those reference sites can be used to predict the fauna expected at the test site in the absence of disturbance.

Habitat attributes of the groups of reference sites defined by the similarity of the benthic communities are compared to identify a subset of variables for use in Discriminant Function Analysis (DFA) for the prediction of group membership. Emphasis is placed on those variables known to be little affected by most human activity (e.g., latitude, longitude, altitude, alkalinity). The variable subset may be selected in 2 ways: 1) Stepwise DFA, which exploits correlations among the predictors to maximize discrimination of the groups (Moss et al. 1987, Parsons and Norris 1996); and 2) correlation analysis between the physical/chemical data and the ordination scores of the biological data (Wright et al. 1984), and Principle Axis Correlations (PCC in the PATN analysis package; Belbin 1993), which is a multiple regression method using the environmental variables as predictor variables and ordination axes as response variables (see Faith and Norris 1989, Parsons and Norris 1996). Once the variables have been selected, a final discriminant model is developed through an iterative process using DFA. This procedure ensures the lowest possible error rate in using the selected environmental variables to predict membership of each reference site in a predetermined biotic group. We recommend using cross-validation rather than re-substitution when testing the discriminant model; the former removes each site in turn from the data set, re-constructs the model, and tests the removed site against the model. In contrast, the latter constructs the model with all sites and then tests each site in turn.

Misclassifications are less important in analyses where the probabilities of test sites belonging to all of the groups are used for predictions (such as RIVPACS [River InVertebrate Prediction And Classification System], Wright 1995 and AusRivAS, Parsons and Norris 1996). Misclassified sites are most likely to be allocated to neighboring groups with relatively strong probabilities of belonging to >1 group. All reference data contribute to the predictions at a test site because probability weightings from all groups are used, which contrasts with other methods that use only data from the most probable group (BEAST, Reynoldson et al. 1995) or the allocated region (Barbour et al. 1995).

A number of test sites (few or many) is selected, the predictor variables identified during the building of the model are measured in the field, and the invertebrate fauna is sampled. Then, the measured environmental variables are used to match the test sites with the groups of reference sites and the probabilities of site membership are calculated (using DFA). Once the probabilities of group membership of a test site have been determined from the environmental data, 2 approaches can be used to compare taxa at the test sites with taxa at the reference sites.

The first approach, the AusRivAS (http://ausrivas.canberra.edu.au/ausrivas, also briefly described in Parsons and Norris 1996) is an adaptation of the system for predicting the macroinvertebrate fauna in flowing waters in the UK (Wright 1995). AusRivAS predicts the taxa that should occur at a site at different probability levels of occurrence (usually 50% and 75%), in contrast to the standard practice in RIVPACS, which is to include all families from 100 to 0.1% probability. Because most taxa will have a <100% chance of occurrence, not all of them would be expected to be present or collected. In both AusRivAS and RIVPACS, the number of taxa expected is calculated as the sum of the probabilities of those predicted (Moss et al. 1987). The number of these taxa actually collected is then compared with the number expected.

In AusRivAS and RIVPACS, the severity of any environmental impact is assessed based on how much the number of taxa observed (O, counted from those with a probability ≥ the probability used to calculate the expected number) deviates from the number expected (E), calculated as the O/E ratio. When the O/E ratio indicates impairment (i.e., less than the mean minus 2 standard deviations for the reference-site O/E), the types of organisms predicted to occur but not collected are used in interpretation.

The 2nd approach, the BEAST (Reynoldson et al. 1995), predicts group membership for a test site using a discriminant model. The invertebrate data for only the reference sites to which the test site is predicted and the test site are merged into one data matrix. The new data matrix is re-ordinated and 90% probability ellipses are plotted for the reference sites. The community structure of the test site can then be compared with the reference sites, and divergence resulting from environmental disturbance can be assessed.

In contrast to the use of presence/absence data in RIVPACS and AusRivAS, the BEAST uses quantitative data on the taxa present to assess test sites (although methods are being developed to include log abundance categories in the RIVPACS/AusRivAS approach; J. F. Wright, Institute of Freshwater Ecology, personal communication). Also, the BEAST combines test and reference sites in a new ordination matrix; thus, the test sites will inevitably affect the distribution of reference sites in ordination space. Therefore, it is important that an appropriate ratio of reference to test sites be used in the BEAST. A specific ratio has not yet been established by testing but we suggest a minimum of 5:1 reference:test sites be used.

Methods

A field comparison of multimetric and multivariate approaches

In this section, we describe how we compared multimetric and multivariate approaches to establishing reference conditions and their ability to assess test-site condition. We examined 2 components: 1) the most accurate classification method for grouping reference sites and assigning test sites to a reference group; and 2) the most precise and accurate assessment method for comparing test sites with the reference condition. We have used a data set from the Fraser River, British Columbia, Canada, which consists of 37 reference sites and 6 test sites collected from 6 subcatchments sampled in the fall of 1994. Macroinvertebrate samples were collected from riffle areas using 400-μm-mesh kick nets; sampling duration was 3 min per kick-net sample (D. M. Rosenberg and others, Freshwater Institute, unpublished data). Five replicated kick-net samples were taken from each site. Selected environmental variables were also measured. Specimens were identified to the family level.

The reference-site data set consisted of 1 of the 5 kick-net samples taken from the 37 reference sites. We examined precision and accuracy of the classification and assessment approaches at 4 of the reference sites (Chilcotin, Clearwater, Pitt, and Stuart rivers) by using the 4 additional replicated macroinvertebrate kick-net samples, and at Salmon River test-site 3 by examining 3 replicated samples. A precise classification and assessment method would designate the additional replicates from the reference and test sites consistently as all either unimpaired or impaired. Accuracy of the methods was determined using only the 4 replicates from the reference sites, and was assessed by the number of reference-site replicates designated as unimpaired. The 3 replicate macroinvertebrate samples from Site 3 of the Salmon River, a site exposed to agricultural activities, were not used to assess accuracy because there was no a priori way of knowing whether this site was disturbed. Another 5 test sites were taken from potentially disturbed areas on the Willow River (extensive logging), Fraser River (downstream of pulp and paper, industrial, agricultural, and municipal discharges), and Sites 1, 2, and 4 on the Salmon River (varying degrees of agricultural impact, increasing in severity from Site 1 to Site 4). These test sites cannot be used to assess either precision (because replicate samples were not taken), or accuracy (because there is no a priori way to define them as disturbed or undisturbed).

Classification methods

Three classifications were used to compare the efficacy of multimetric and multivariate methods. Two of these were physical groupings (i.e., ecoregion and stream order), as recommended by Hughes (1995) and Omernik (1995), which have been adopted by a number of agencies in the USA. The 3rd was a biological grouping based on the invertebrate fauna at reference sites (e.g., Reynoldson et al. 1995, Wright 1995, Parsons and Norris 1996). As discussed above, these classifications use different methods for matching test sites to the appropriate groups of reference sites. The a priori physical approach uses simple assignment of the test site based on either the ecoregion or stream order of the test site. The multivariate faunal classification uses habitat conditions at the test site in a discriminant function model developed from the reference sites to predict the reference-site group to which the test site is most likely to belong.

Assessment methods

Each of the classification methods was then assessed using multimetric and multivariate comparisons of community structure. The multimetrics approach modified from Plafkin et al. (1989) included: 1) total abundance; 2) number of families; 3) % Ephemeroptera, Plecoptera, and Trichoptera (EPT) individuals; 4) % Chironomidae; 5) number of EPT individuals/number of EPT + Chironomidae individuals; 6) number of Hydropsychidae/number of Trichoptera; 7) % dominance; and 8) Family Biotic

Index (FBI; Bode 1988, Hilsenhoff 1988, Lenat 1993). Based on input from M. T. Barbour and J. Gerritsen (Tetra Tech Inc.) a 2nd multimetric assessment was also done. It deleted "1) total abundance", and replaced "5) number of EPT individuals/number of EPT and Chironomidae individuals" with "number of Baetidae/number of Ephemeroptera". Metrics were scored ordinally based on their similarity to the appropriate reference-condition classification. We used the quartiles and 5th or 95th percentiles, for the reference sites, to assign metric scores. All metrics scored 5 when they were between the quartiles from the reference sites. Depending on the metric (Table 2) and the expected response (i.e., increase or decrease), a site scored 3 if it was between the quartile and the 5th or 95th percentile, and 1 if it was less than the 5th or greater than the 95th percentile (Gerritsen 1995). For example, number of taxa is expected to decrease with disturbance, so a site would score 5 if it was above the lower quartile, 3 if it was between the quartile and the 5th percentile, and 1 if it was less than the 5th percentile. Thus, a score was assigned to the test site for each metric and the total score was compared with the reference score. Based on the 7 or 8 metrics used in either assessment, a maximum score of 35 or 40 was achievable; a test site was considered to be disturbed if the metric score was lower than the

TABLE 2. Values for metrics used in scoring sites in 4 ecoregions of the Fraser River catchment. (EPT = Ephemeroptera, Trichoptera, and Plecoptera).

Metric (expected response to stress)	Score	Fraser Basin	Fraser Plateau	Southern Rocky Mountain Trench	Pacific Range
Abundance (increase or decrease)[a]	5	278–1325	2918–9645	447–1973	106–595
	3	1325–3273	9645–20,920	1973–9164	595–702
	3	53–278	1511–2918	393–447	63–106
	1	>3273	>20,920	>9164	>702
	1	<53	<1511	<393	<63
No. families (decrease)	5	>12	>12	>12	>10
	3	7–12	8–12	9–12	8–10
	1	<7	<8	<9	<8
% EPT (decrease)	5	>46.8	>22.4	>53.0	>84
	3	26.5–46.8	8.3–22.4	40.3–53.0	59.8–84
	1	<26.5	<8.3	<40.3	<59.8
% Chironomidae (increase)	5	<45.8	<72.8	<41.9	<12.2
	3	45.8–70.5	72.8–89.9	41.9–56.5	12.2–29.6
	1	>70.5	>89.9	>56.5	>29.6
% EPT/EPT + Chironomidae (decrease)	5	>0.50	>0.25	>0.57	>0.87
	3	0.27–0.50	0.09–0.25	0.42–0.57	0.67–0.87
	1	<0.27	<0.09	<0.42	<0.67
% Baetidae/Ephemeroptera (increase)	5	<0.53	<0.48	<0.52	<0.55
	3	0.53–0.60	0.48–0.82	0.52–0.59	0.55–0.75
	1	>0.60	>0.82	>0.59	>0.75
% Hydropsychidae/Trichoptera (increase)	5	<0.35	<0.10	<0.23	0
	3	0.35–0.96	0.10–0.22	>0.23–0.54	0.00–0.33
	1	>0.96	>0.22	>0.54	>0.33
% dominance (increase)	5	<45.8	<72.8	<45.5	<46.1
	3	45.8–70.5	72.8–89.9	45.5–56.6	46.1–69.2
	1	>70.5	>89.9	>56.5	>69.2
Family Biotic Index (increase)	5	<5.21	<6.13	<4.69	<3.55
	3	5.21–5.73	6.13–6.67	4.69–5.37	3.55–4.44
	1	>5.73	>6.67	>5.37	>4.44

[a] Scores: The first 3 and 1 represent increases above those expected; the second 3 and 1 represent decreases below those expected (see text for explanation)

25th percentile for the appropriate group of reference sites (Barbour et al. 1995, Yoder and Rankin 1995).

The multivariate approach compared reference sites and test sites from the same classification group (ecoregion, stream order, or faunal group) in ordination space (BEAST), or compared O/E ratios of taxa occurrence from all the groups based on the probability of a test site belonging to all of the reference groups and the probability of a taxon's occurrence in the reference group (AusRivAS). The BEAST assessment constructs 90% probability ellipses around the reference sites; location of the test site outside that probability ellipse would designate it as disturbed (T. B. Reynoldson, unpublished data). The O/E ratio of the AusRivAS assessment compares the number of taxa observed at a test site with the number of taxa expected from the reference database. The number of taxa predicted at a test site at a given probability level (in this case 50%) is calculated using the probabilities of a site belonging to each of the classification groups and the frequencies with which each taxon occurs in the classification groups (Wright 1995). Thus, if a site has a 0.70 probability of belonging to 1 group and a taxon occurs at 80% of the sites in the group, and the site has a 0.30 probability of belonging to a 2nd group and the taxon occurs at 60% of the sites, then the taxon has a $(0.7 \times 80) + (0.3 \times 60) = 74\%$ chance of occurring at the test site. The number of taxa expected is simply the sum of the probabilities of those predicted to have a 50% or higher probability of occurrence at the test site (Wright 1995). Determination of whether a site is impaired relative to reference is based on the distribution of O/E ratios for the reference sites. For AusRivAS, a site is defined as equivalent to reference if the O/E ratio is within 2 standard deviations (SDs) of the mean of the O/E ratios for the reference data, and impaired if the O/E ratio is 4–2 SDs (mildly impaired), 6–4 SDs (moderately impaired), or >6 SDs below the mean (severely impaired; R. H. Norris, unpublished data).

In current benthic macroinvertebrate monitoring programs, multimetric assessments would normally use ecoregion or stream-order classifications, and multivariate approaches would normally use biotic groupings. However, these are not mutually exclusive applications in that: 1) some multimetric approaches use biotic groupings (as when multivariate analysis forms reference-site groupings); and 2) some multivariate studies are confined to a single ecoregion and others are conducted within a single stream order (e.g., upstream–downstream comparisons).

Results

Classification by ecoregion

The 37 reference sites from the Fraser River encompassed 5 ecoregions. Test sites were located in 4 of those ecoregions (Table 2). Test sites were always assessed by comparison with reference sites from the same ecoregion. Although the AusRivAS approach uses all the data and does not assume exact matches between reference-site groups and test sites, we have used the AusRivAS in a more restrictive sense here for the sake of simplification of the analysis (see comments below).

Multimetric assessment.—The calculated values for assessing site condition based on the 2 sets of metrics used are shown in Table 2; they reveal the importance of setting site-specific values. For example, for sites in the Fraser Basin and Fraser Plateau ecoregions, % Chironomidae is naturally much higher than for sites in the Pacific Range. Thus, a hypothetical test site with 30% Chironomidae would score only 1 for this metric if it were located in the Pacific Range but would score 5 if it were in the Fraser Basin or Fraser Plateau.

Scores for each metric were calculated for the reference sites and test sites, and summed to produce an overall score out of a possible maximum of 35 or 40. Scores for the reference sites in the ecoregions were developed as box plots showing the median, 25th, and 75th percentiles, and maximum and minimum scores; test sites and replicated reference sites were plotted individually. For example, using the 1st multimetric assessment, a number of sites in the Southern Rocky Mountain Trench ecoregion would be assessed as impaired (scores <25th percentile), including all the replicate samples from a single riffle at a reference site on the Clearwater River (CLR 6.1–6.4) (Fig. 2A).

The results of the analysis of precision (estimated by determining the designation of replicates as either all impaired or all unimpaired at the Salmon 3, Chilcotin, Clearwater, Pitt, and

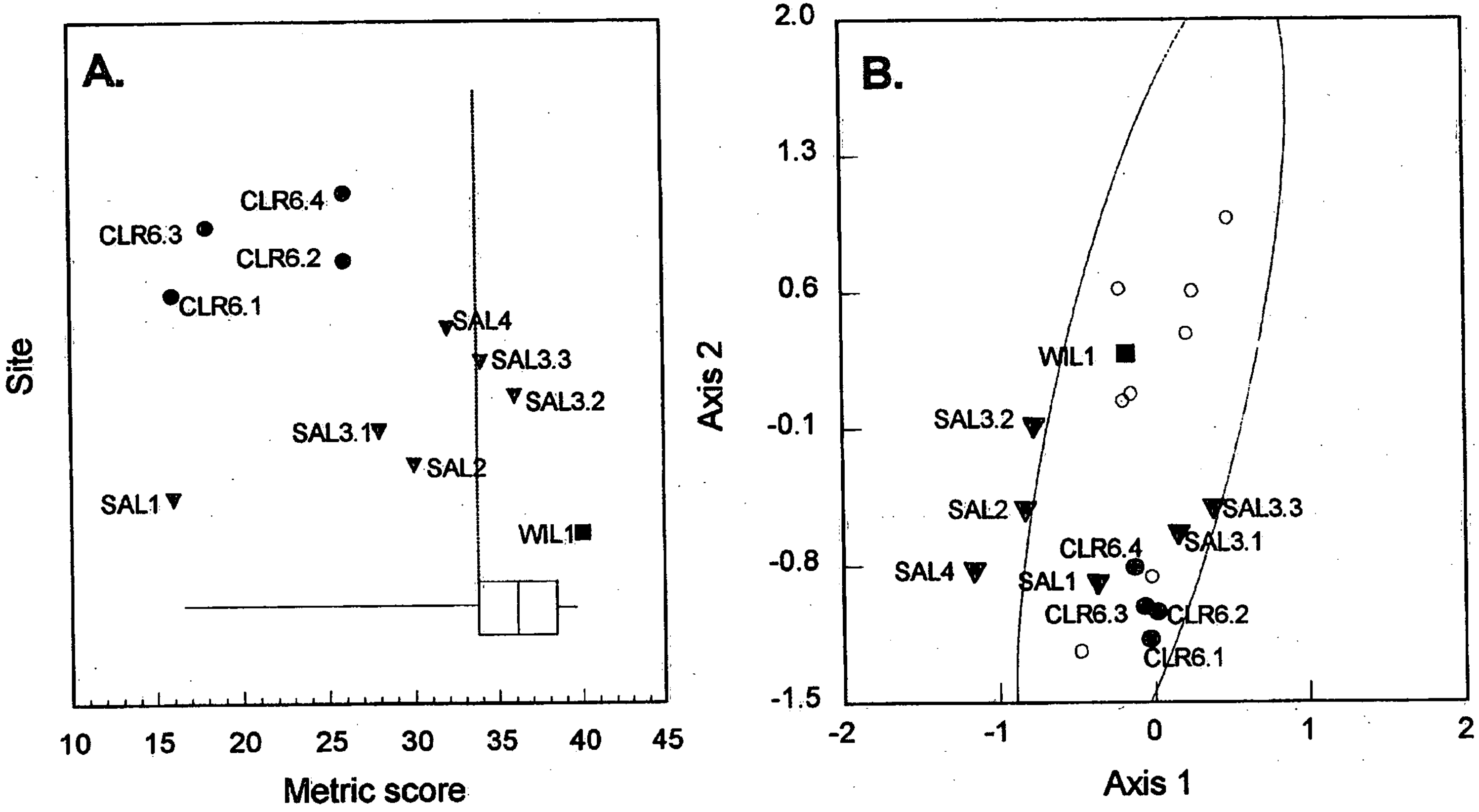


FIG. 2. Comparison of multimetric (A) and multivariate (B) assessment of disturbance based on ecoregion classification (Southern Rocky Mountain Trench). Multimetric assessment involved 8 metrics (see text). CLR = Clearwater, SAL = Salmon, WIL = Willow rivers. The number following an abbreviation is the station; the decimal is a replicate. A.—Vertical dotted line is the 25th percentile. Box plot summarizes reference-site scores: 25th percentile—vertical solid line on left; 75th percentile—vertical solid line on right; median percentile—vertical solid line in middle. Bar on left: minimum score; bar on right: maximum score. B.—Ellipse represents 90% probability. Open circles are the reference sites used for assessment. The Clearwater (CLR) sites are also reference, but they are highlighted here for comparison with (A). Two dimensions used in multidimensional scaling; stress level = 0.1334.

TABLE 3. Precision and accuracy of classification (ecoregion, stream order, biotic group) and assessment (multimetric, BEAST, AusRivAS^a) methods. (Note: Sites are **precisely** assessed when all replicates are designated as either reference [R] or impaired [I]; replicates from 4 reference sites are **accurately** assessed when designated as reference.)

Classification	Ecoregion																Stream order							
Assessment	Multimetric 1				Multimetric 2				BEAST				AusRivAS				Multimetric 1				Multimetric 2			
Replicate	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Test sites																								
Willow	R				R				R				R				R				R			
Salmon 1	I				I				R				R				I				I			
Salmon 2	I				I				I				R				R				R			
Salmon 3	I	R	R		I	R	R		R	I	R		R	R	R		R	R	R		R	R	R	
Salmon 4	I				R				I				R				R				R			
Fraser 28	I				I				R				R				R				R			
Reference sites																								
Chilcotin 5	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	I	I	I	I	I	I	I	I
Clearwater 6	I	I	I	I	I	I	I	I	R	R	R	R	R	R	R	R	I	I	I	I	I	I	I	I
Pitt 7	I	I	R	I	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
Stuart 2	R	R	I	R	R	R	I	R	R	R	R	R	R	R	R	R	R	I	I	R	R	I	I	R
No.precise	2 of 5				3 of 5				4 of 5				5 of 5				4 of 5				4 of 5			
% precision	40%				60%				80%				100%				80%				80%			
No. reference	8 of 16				11 of 16				16 of 16				16 of 16				6 of 16				6 of 16			
% accuracy	50%				69%				100%				100%				38%				38%			

Stuart river sites) indicated that only the Chilcotin and Clearwater sites had consistent replicates in the 1st multimetric assessment; this result was 2 out of 5 sites, providing a 40% precision (Table 3). The result was 60% for the 2nd multimetric assessment. In terms of accuracy (established using the 16 replicates from the reference sites), a site assessment was considered accurate if replicates of the 4 reference sites were identified as unimpaired. Eight of the 16 replicates provided the correct assessment for the 1st multimetric assessment, yielding a 50% accuracy. The result was 69% for the 2nd multimetric assessment.

Multivariate assessment.—Data from reference and test sites for each ecoregion were plotted together in ordination space using the BEAST model. A 90% probability ellipse was plotted around the reference sites (Fig. 2B); a site was deemed impaired if it was located outside the 90% ellipse. In the Southern Rocky Mountain Trench ecoregion (Fig. 2B), the Willow River, the upstream Salmon River (SAL1), and the 4 replicates from the Clearwater River were all classified as similar to reference (cf. Fig. 2A). The downstream Salmon River samples (SAL 2, 3.2, and 4) were outside the reference ellipse and would be assessed as impaired. Two of the replicate samples from Salmon River Site 3 (SAL 3.1 and 3.3) were inside the ellipse, so the result from Salmon Site 3 is imprecise. The replicated samples from the Clearwater River were in close proximity in species ordination space, indicating high similarity. Precision was 80% and accuracy was 100%.

For the AusRivAS model, numbers of families present at each reference site in the ecoregion and the average were established. An O/E value for a test site was determined from the ratio of number of families at the test site to the average for the reference sites in the same ecoregion. The normal application of AusRivAS incorporates the variable probability of a test site belonging to more than 1 reference group. This feature is lost using an ecoregion classification because the probability of a test site belonging to the matching ecoregion is clearly 1.0. Both precision and accuracy were 100%, based on O/E family ratios (Table 3).

The BEAST and AusRivAS differed in their assessment of 3 sites (Table 3). The BEAST identified Site 2 and Site 4 on the Salmon River as

TABLE 3. Extended.

Stream order								Biotic group															
BEAST				AusRivAS				Multimetric 1				Multimetric 2				BEAST				AusRivAS			
1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
R				R				R				R				I				R			
R				R				R				R				R				R			
R				I				R				R				I				I			
R	R	R		R	R	R		R	R	R		R	R	R		R	I	I		R	R	R	
R				R				R				R				I				R			
R				R				—[b]				—[b]				—[b]				I			
R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
R	R	R	R	R	R	R	R	I	R	R	I	R	R	R	R	R	R	R	R	R	R	R	R
R	R	R	R	R	R	R	R	R	I	I	R	R	I	I	R	R	R	R	R	R	R	R	R
5 of 5 100%				5 of 5 100%				3 of 5 60%				4 of 5 80%				4 of 5 80%				5 of 5 100%			
16 of 16 100%				16 of 16 100%				12 of 16 75%				14 of 16 88%				16 of 16 100%				16 of 16 100%			

[a] BEAST = BEnthic Assessment of SedimenT; AusRivAS = AUStralian RIVer Assessment Scheme

[b] No assessment possible because only 1 reference site available for comparison

impaired, whereas the AusRivAS did not. The AusRivaAS identified all 3 replicates at Site 3 as reference, whereas the BEAST called 1 replicate impaired (SAL 3.2).

Classification by stream order

The reference data set of 37 sites included representatives of 1st- to 8th-order streams, but test sites included only 2nd-, 3rd-, 4th- and one 8th-order sites (Fraser 28). Therefore, multimetric scores were only calculated for reference sites of 2nd- to 4th-order streams. There were only three 8th-order reference sites, so we composited reference sites from 5th- and higher-order streams for purposes of assessment.

Multimetric assessment.—The same approach as for the ecoregion classification was taken in calculating metric scores for reference sites classified by stream order. Individual scores for test sites were compared with reference sites from the same stream order (Table 3).

In terms of precision, the level for both multimetric assessments was 80%, which was higher than in the ecoregion classification (40% and 60%), and both multimetric assessments produced identical results. However, there were differences between the stream-order and ecoregion classifications (Table 3). In terms of accuracy, the level for both multimetric assessments was 38%, which was lower than in the ecoregion classification (50% and 69%); only 6 of the 16 replicates from the reference sites were assessed correctly (Table 3).

Multivariate assessment.—Assessment of sites using the BEAST model yielded slightly better results than in the ecoregion classification. Both precision and accuracy were 100% (Table 3).

The feature of AusRivAS that incorporates the variable probability of a test site belonging to more than 1 reference groups is also lost using a stream-order classification, as discussed above for the ecoregion classification. Nevertheless, the O/E ratio of the AusRivAS model still showed 100% precision using a stream-order classification (Table 3). The BEAST and AusRivAS models differed in their assessment only once (SAL 2) out of 24 instances.

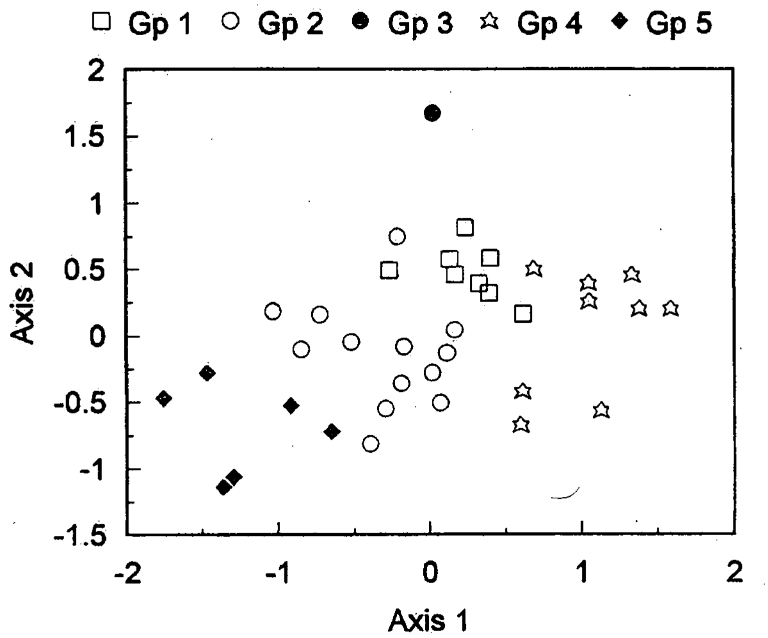


FIG. 3. Site groups (Gp) formed by multidimensional scaling (MDS) ordination for 37 reference sites in the Fraser River catchment, based on invertebrate families. Two dimensions used in MDS; stress level = 0.1776.

Biotic classification

The 2 previous classification methods assume that there is a direct relationship between physical structure and the communities that occur in riffle habitats. The biotic classification does not make the same assumption. Instead, the biological data are used to classify the sites and form biotic groupings. Selection of the appropriate biotic group in assessing any test site is determined by applying the predictive model to environmental data from the test sites.

Five groups of reference sites were discerned using cluster analysis (unweighted pair-groups using arithmetic averages), and the degree of discrimination between the groups thus formed was examined with hybrid multidimensional scaling (Belbin 1991) on 2 axes (Fig. 3). The Bray–Curtis association metric was used for these pattern analyses. Discriminant function analysis was used to relate habitat structure to the biotic groups. Sixteen potential predictor variables were tested (Table 4); those inappropriate because of human influence (e.g., phosphorus in agricultural areas) were excluded (i.e., "no" in Table 4).

The discriminant model with the lowest error rate (32%) used 6 predictor variables (latitude, channel width, maximum channel depth, maximum velocity, % silt, and alkalinity). Twenty-four sites or site replicates were assigned a probability of belonging to a reference group (Table 5) using this model, and were compared with the most probable reference condition using both multimetric and multivariate assessments.

Multimetric assessment.—The performance of the multimetric assessment was improved by the biotic classification. Precision remained at 80% for the 2nd multimetric assessment and declined to 60% for the 1st multimetric assessment (Table 3); however, accuracy increased to 75% for the 1st multimetric assessment and to 88% for the 2nd method (Table 3).

Multivariate assessment.—Precision was 80% and accuracy was 100% in the BEAST assessment (Table 3). Salmon River Sites 2, 4, and 2 of 3 replicates at Salmon Site 3 were assessed as impaired (Fig. 4). Fraser Site 28 could not be designated (Table 3) because there was only 1 reference site in that biotic group (Fig. 4).

Using the AusRivAS assessment, precision and accuracy were both 100%. The AusRivAS

TABLE 4. Environmental variables measured and considered as potential predictors for biotic groups in the Fraser River catchment.

Environmental variable	Suitable as predictor	Rationale for exclusion
Latitude	yes	
Longitude	yes	
Altitude	yes	
Stream order	yes	
Channel width	yes	
Channel depth (avg. and max.)	yes	
Velocity (avg. and max)	yes	
Substrate	yes	
% gravel	yes	
% sand	yes	
% silt	yes	
Supended nitrogen	no	responds to eutrophication
Suspended carbon	no	responds to eutrophication
Chlorophyll	no	responds to eutrophication
Slope	yes	
Water temperature	no	temporally variable
Oxygen	no	temporally variable
Conductivity	yes	
Alkalinity	yes	
pH	yes	
Total suspended solids	yes	
Nitrate-nitrite	no	responds to eutrophication
Kjeldahl nitrogen	no	responds to eutrophication
Total phosphorus	no	responds to eutrophication

method uses all the reference sites based on weighted probabilities, so it was able to assess the status of Fraser River Site 28 as impaired.

Discussion

Using the same data set, precision and accuracy were consistently better for both multivariate assessment methods than for the multimetric assessments. This result was obtained both with commonly used applications (multimetric approaches within stream-order or ecoregion classifications; multivariate approaches within biotic classifications) and with specialized applications (multivariate approaches within stream-order or ecoregion classifications; multimetric approaches within biotic classifications). The imprecision and inaccuracy observed with the multimetric methods is our greatest concern in terms of advocating their continued use. The accuracy of site-replicate identification as unimpaired was highly variable for the multimetric methods (between 38 and 88%). In contrast, the BEAST and AusRivAS methods consistently identified as unimpaired all replicates from all reference sites (Table 3). Multimetric methods were most accurate when using a biotic classification.

The ability of the methods to identify stressed sites varied. Multimetric and AusRivAS methods identified the site at which logging activity had occurred (Willow River) as unimpaired in all classifications. The BEAST method agreed under the ecoregion and stream-order classifications, but designated the Willow River site as impaired under the biotic classification. Fraser River Site 28 was inconsistently designated by all 3 methods, when a designation was possible. For example, the BEAST and AusRivAS both identified Fraser River Site 28 as equivalent to reference in ecoregion and stream-order classifications, whereas the AusRivAS identified the Fraser River site as impaired under the biotic classification. The BEAST had insufficient reference sites for comparison in the biotic classification (Table 3). In contrast, the AusRivAS used all of the reference sites, weighted by probability, which is a strength of this approach.

The variability of the 3 methods in assessing the Salmon River sites is of most interest. Site 1

TABLE 5. Probabilities of sites belonging to 1 of 5 biotic reference groups (Gp) based on prediction using latitude, channel width, maximum channel depth, maximum velocity, % silt, and alkalinity. No sites were predicted to occur in Group 1. The number following an abbreviation is the station; the decimal is a replicate. Test sites: WIL = Willow; SAL = Salmon; FRA = Fraser; reference sites: CHI = Chilcotin; CLR = Clearwater; PIT = Pitt; STU = Stuart. For the reference sites, 1 replicate was used to form the reference-site groups (see Methods) and the remaining 4 replicates were used here.

	Pre-dicted	Probability of membership				
Site	group	Gp 1	Gp 2	Gp 3	Gp 4	Gp 5
WIL1	5	0.234	0.204	0	0	0.562
SAL1	5	0.074	0.040	0	0	0.886
SAL2	5	0.272	0.314	0	0.063	0.351
SAL3.1	5	0.129	0.185	0	0	0.686
SAL3.2	5	0.150	0.248	0	0	0.603
SAL3.3	5	0.157	0.260	0	0	0.583
SAL4	5	0.202	0.229	0	0	0.569
FRA28	3	0	0	1	0	0
CHI5.1	5	0.270	0.325	0	0	0.406
CHI5.2	5	0.258	0.311	0	0	0.431
CHI5.3	5	0.270	0.325	0	0	0.406
CHI5.4	5	0.258	0.311	0	0	0.431
CLR6.1	5	0.199	0.165	0	0	0.636
CLR6.2	5	0.160	0.129	0	0	0.711
CLR6.3	5	0.162	0.130	0	0	0.708
CLR6.4	5	0.161	0.130	0	0	0.709
PIT7.1	4	0	0	0	1	0
PIT7.2	4	0.001	0	0	0.998	0
PIT7.3	4	0.001	0	0	0.998	0
PIT7.4	4	0	0	0	1	0
STU2.1	2	0.279	0.550	0	0	0.172
STU2.2	2	0.272	0.520	0	0	0.208
STU2.3	2	0.288	0.545	0	0	0.167
STU2.4	2	0.276	0.545	0	0	0.179

was the farthest upstream with a well-vegetated riparian zone and limited agricultural activity in the vicinity. Site 2 was located downstream in a similar setting. At Site 3, the banks had been infilled and rip-rapped; the site is also in an area of agricultural activity. Site 4, the farthest downstream, was within 100 m of a ranch dwelling and horse paddock. In general, the area and intensity of agricultural activity increased in a downstream direction. In the analyses, the BEAST and AusRivAS methods both consistently identified Site 1 as unimpaired but were inconsistent in their designations of downstream sites (Table 3). The AusRivAS method was more consistent in its designations than the BEAST, but detected impairment less often at the downstream sites compared to the BEAST. Weighted probabilities were not used in either of the physical classifications, which may have led to loss of power in the AusRivAS model. The BEAST appeared to be more sensitive than the AusRivAS to change at the downstream Salmon sites under the biotic classification. Conversely, the BEAST was less robust than the AusRivAS, as suggested by the imprecise designation of Salmon Site 3. The difference in the AusRivAS and BEAST methods likely relates to the fact the BEAST is responding to quantitative changes in community composition, whereas the AUSRIVAS does not respond until a taxon is absent. For example, under the biotic classification, the BEAST identified Salmon River Sites 2, 3 (2 replicates), and 4 as impaired (Table 3) because total abundance and abundances of families such as the Baetidae and Hydropsychidae differed from Salmon Site 1. The AusRivAS identified Salmon Sites 3 and 4 as reference because the same families were still present at both sites, although numbers in these families differed.

Assessment of the Salmon River sites using multimetrics was even more inconsistent (Table 3). Site 1 was defined as impaired under the ecoregion and stream-order classifications but was designated as reference under the biotic classification. The downstream sites were generally designated as reference. These results are contrary to what was expected based on the intensity of agricultural land use.

Conclusions and Recommendations

The results of the analysis and interpretation done in this study indicate that each of the assessment methods has advantages and disadvantages. Multimetrics are attractive because they produce a single score that is readily comparable to a target value; this approach is a traditional one that appeals to many managers. Multimetrics are purported to incorporate ecological information, including how animals feed, reproduce, and exploit their habitats (Fore et al. 1996), depending on the metrics used in calculating an overall score. Disadvantages of their use are that: 1) they discard information; 2) some metrics are redundant 3) some can compound error; and 4) it is difficult to acquire cur-

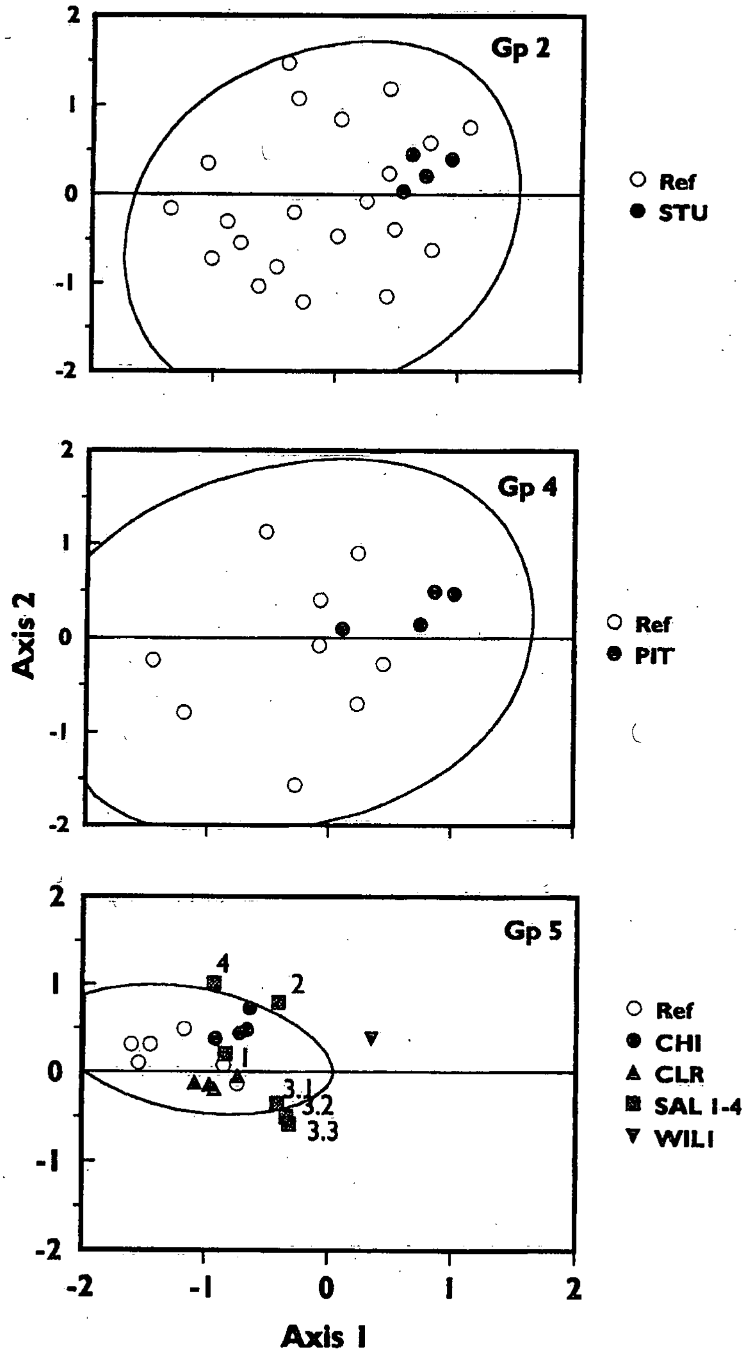


FIG. 4. Comparison of test sites (WIL1, SAL1–4) and reference sites (Ref: open symbols and STU, PIT, CLR, and CHI; see Table 5) using 90% probability ellipses (BEAST model). Abbreviations of rivers as in Table 5. Gp = group. No sites were predicted to Gp 1 and only 1 site was predicted to Gp 3, so these groups are not shown here. Two dimensions used in multidimensional scaling; stress levels: Gp 2 = 0.2326; Gp 4 = 0.1746; Gp 5 = 0.1867.

rent procedures. For example, in the 1st assessment method for multimetrics used in this study, EPT taxa are included in 3 of the 8 metrics and Chironomidae are included in 2 metrics (see above). Therefore, a high variability in abundance of any of these taxonomic groups in a sample may confound interpretation and may have caused the imprecision of assessments in multiple replicates taken from the same riffle in this study (Table 3).

The BEAST and the AusRivAS both use multivariate methods for defining reference groups. The multivariate approach is attractive because it requires no prior assumptions in either creating faunal groups out of reference sites or in comparing sites suspected of being impaired with reference faunal groups. A disadvantage is that initial model construction for the multivariate methods is complex and consequently discourages potential users. However, software development for the multivariate methods has reached the stage that programs with simple-to-interpret outputs can be run by managers or community groups (e.g., http://AusRivAS.canberra.edu.au/AusRivAS); the user does not do the actual multivariate analysis. This development should allay apprehensions about multivariate methods among biologists responsible for water-quality assessment (e.g., Gerritsen 1995, Fore et al. 1996).

An advantage of the multivariate AusRivAS approach is that it accounts for the fact that test sites are predicted to belong to a reference group with a variable probability, and uses a weighting method to predict the number of taxa that should occur. However, it uses only presence/absence data and does not incorporate quantitative changes in the community, which can also be indicative of stress. In comparison, an advantage of the BEAST approach is that it incorporates quantitative changes in the assemblage of organisms at a site. However, it does not account for the probability that a site may be placed in the wrong reference group (Table 5). Thus, we recommend that: 1) users understand how the strengths and weaknesses of both AusRivAS and BEAST affect data interpretation; and 2) these 2 methods be used together, where possible, because of their basic but complementary differences.

Given that the majority of the US environmental regulatory agencies use the multimetric approach, how can they proceed in the near future in light of the results of our study? A safe, cost-effective strategy for these agencies may be to: 1) supplement the multimetric biological collections (they are fundamentally the same as those used for the multivariate approaches) with the environmental measurements required for using the BEAST or AusRivAS methods; and 2) do multimetric and multivariate analyses side by side and base the ultimate decision of site impairment on analysis and interpretation of both approaches.

Acknowledgements

We thank M. T. Barbour and J. Gerritsen for discussions on the use of multimetric approaches. J. F. Wright has provided invaluable help in the development of Australian and Canadian multivariate approaches. R. C. Bailey provided comments that helped clarify much of this manuscript. Finally, we thank R. J. Mackay, former editor of J-NABS, for shepherding the manuscript through to the final stage of publication.

Literature Cited

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. B. STRIBLING, AND J. R. KARR. 1995. Multimetric approach for establishing biocriteria and measuring biological condition. Pages 63–77 *in* W. S. Davis and T. P. Simon (editors). Biological assessment and criteria. Tools for water resource planning and decision making. Lewis Publishers, Boca Raton, Florida.

BELBIN, L. 1991. Semi-strong hybrid scaling, a new ordination algorithm. Journal of Vegetation Science 2:491–496.

BELBIN, L. 1993. PATN, pattern analysis package. Division of Wildlife and Ecology, CSIRO, Canberra, Australia.

BODE, R. W. 1988. Quality assurance work plan for biological stream monitoring in New York State. Stream Biomonitoring Unit, Bureau of Monitoring and Assessment, Division of Water, New York State Department of Environmental Conservation, Albany, New York.

COOPER, S. D., AND L. A. BARMUTA. 1993. Field experiments in biomonitoring. Pages 399–441 *in* D. M. Rosenberg and V. H. Resh (editors). Freshwater biomonitoring and benthic macroinvertebrates. Chapman and Hall, New York.

CORKUM, L. D. 1990. Intrabiome distributional patterns of lotic macroinvertebrate assemblages. Canadian Journal of Fisheries and Aquatic Sciences 47: 2147–2157.

CORKUM, L. D. 1991. Spatial patterns of macroinvertebrate distributions along rivers in eastern deciduous forest and grassland biomes. Journal of the North American Benthological Society 10:358–371.

DAVIS, W. S., AND T. P. SIMON. 1995. Introduction. Pages 3–6 *in* W. S. Davis and T. P. Simon (editors). Biological assessment and criteria. Tools for water resource planning and decision making. Lewis Publishers, Boca Raton, Florida.

EBERHARDT, L. L. 1978. Appraising variability in population studies. Journal of Wildlife Management 42:207–238.

FAITH, D. P., AND R. H. NORRIS. 1989. Correlation of environmental variables with patterns of distribution and abundance of common and rare freshwater macroinvertebrates. Biological Conservation 50:77–98.

FORE, L. S., J. R. KARR, AND R. W. WISSEMAN. 1996. Assessing invertebrate responses to human activities: evaluating alternative approaches. Journal of the North American Benthological Society 15: 212–231.

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

GRANEY, R. L., D. S. CHERRY, AND J. CAIRNS. 1984. The influence of substrate, pH, diet and temperature upon cadmium accumulation in the Asiatic clam (*Corbicula fluminea*) in laboratory artificial streams. Water Research 18:833–842.

HANNAFORD, M. J., AND V. H. RESH. 1995. Variability in macroinvertebrate rapid-assessment surveys and habitat assessments in a northern California stream. Journal of the North American Benthological Society 14:430–439.

HILSENHOFF, W. L. 1988. Rapid field assessment of organic pollution with a family-level biotic index. Journal of the North American Benthological Society 7:65–68.

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, Boca Raton, Florida.

HUGHES, R. M., AND D. P. LARSEN. 1988. Ecoregions: an approach to surface water protection. Journal of the Water Pollution Control Federation 60:486–493.

HUGHES, R. M., T. R. WHITTIER, C. M. ROHM, AND D. P. LARSEN. 1990. A regional framework for establishing recovery criteria. Environmental Management 14:673–683.

HURLBERT, S. H. 1984. Pseudoreplication and the design of ecological field experiments. Ecological Monographs 54:187–211.

JOHNSON, R. K., AND T. WIEDERHOLM. 1989. Classification and ordination of profundal macroinvertebrate communities in nutrient poor, oligo-mesohumic lakes in relation to environmental data. Freshwater Biology 21:375–386.

JOHNSON, R. K., T. WIEDERHOLM, AND D. M. ROSENBERG. 1993. Freshwater biomonitoring using individual organisms, populations, and species assemblages of benthic macroinvertebrates. Pages 40–158 *in* D. M. Rosenberg and V. H. Resh (editors). Freshwater biomonitoring and benthic macroinvertebrates. Chapman and Hall, New York.

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.

METCALFE-SMITH, J. L. 1994. Biological water-quality assessment of rivers: use of macroinvertebrate communities. Pages 144–172 *in* P. Calow and G. E. Petts (editors). The rivers handbook. Volume 2. Blackwell Scientific, London.

MOSS, D., M. T. FURSE, J. F. WRIGHT, AND P. D. ARMITAGE. 1987. The prediction of the macro-invertebrate fauna of unpolluted running-water sites in Great Britain using environmental data. Freshwater Biology 17:41–52.

NORRIS, R. H. 1995. Biological monitoring: the dilemma of data analysis. Journal of the North American Benthological Society 14:440–450.

NORRIS, R. H., P. S. LAKE, AND R. SWAIN. 1982. Ecological effects of mine effluents on the South Esk River, north-eastern Tasmania III. Benthic macroinvertebrates. Australian Journal of Marine and Freshwater Research 33:789–809.

OMERNIK, J. M. 1987. Ecoregions of the conterminous United States. Annals of the Association of American Geographers 77:118–125.

OMERNIK, J. M. 1995. Ecoregions: a spatial framework for environmental management. Pages 49–62 *in* W. S. Davis and T. P. Simon (editors). Biological assessment and criteria. Tools for water resource planning and decision making. Lewis Publishers, Boca Raton, Florida.

OMERNIK, J. M., AND G. E. GRIFFITH. 1991. Ecological regions versus hydrologic units: framework for managing water quality. Journal of Soil and Water Conservation 46:334–340.

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

PLAFKIN, J. L., M. T. BARBOUR, K. D. PORTER, S. K. GROSS, AND R. M. HUGHES. 1989. Rapid bioassessment protocols for use in streams and rivers. Benthic macroinvertebrates and fish. EPA/444/4-89/001. Office of Water Regulations and Standards, US Environmental Protection Agency, Washington, DC.

RESH, V. H., AND E. P. MCELRAVY. 1993. Contemporary quantitative approaches to biomonitoring using benthic macroinvertebrates. Pages 159–194 *in* D. M. Rosenberg and V. H. Resh (editors). Freshwater biomonitoring and benthic macroinvertebrates. Chapman and Hall, New York.

REYNOLDSON, T. B., R. C. BAILEY, K. E. DAY, AND R. H. NORRIS. 1995. Biological guidelines for freshwater sediment based on BEnthic Assessment of SedimenT (the BEAST) using a multivariate approach for predicting biological state. Australian Journal of Ecology 20:198–219.

RICHARDS, C., G. E. HOST, AND J. W. ARTHUR. 1993. Identification of predominant environmental factors structuring stream macroinvertebrate communities within a large agricultural catchment. Freshwater Biology 29:285–294.

SCHINDLER, D. W. 1974. Eutrophication and recovery in experimental lakes: implications for lake management. Science 184:897–899.

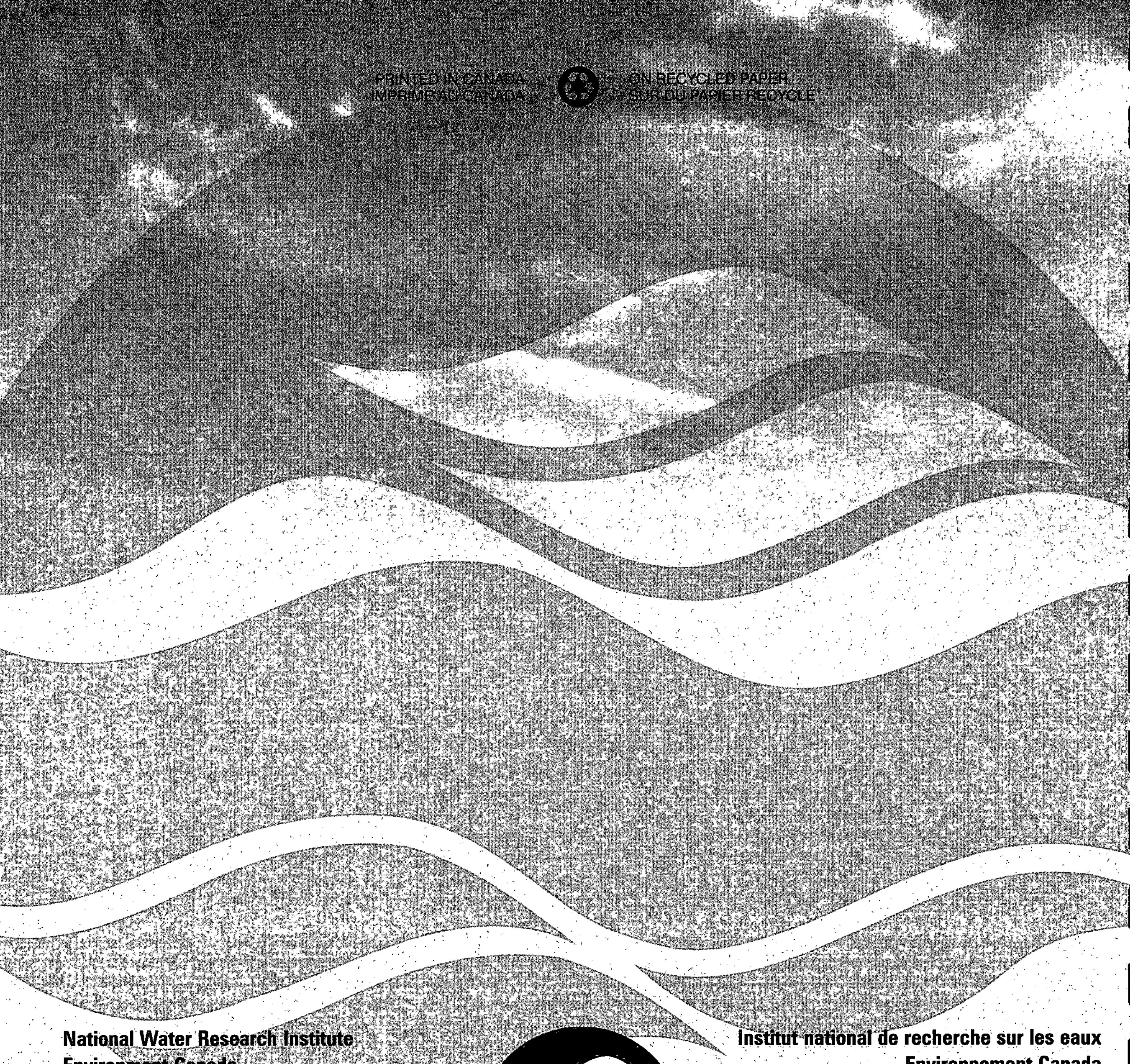
VOSHELL, J. R., R. J. LAYTON, AND S. W. HINER. 1989. Field techniques for determining the effects of toxic substances on benthic macroinvertebrates in rocky-bottomed streams. Pages 134–155 *in* U. M. Cowgill and L. R. Williams (editors). Aquatic toxicology and hazard assessment: 12th volume. American Society for Testing and Materials (ASTM) Special Technical Publication 1027. ASTM, Philadelphia.

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

WRIGHT, J. F., 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 the prediction of community type using environmental data. Freshwater Biology 14:221–256.

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, Boca Raton, Florida.

Received: 23 December 1996
Accepted: 2 August 1997



PRINTED IN CANADA
IMPRIME AU CANADA

ON RECYCLED PAPER
SUR DU PAPIER RECYCLE

National Water Research Institute
Environment Canada
Canada Centre for Inland Waters
P.O. Box 5050
867 Lakeshore Road
Burlington, Ontario
L7R 4A6 Canada

National Hydrology Research Centre
11 Innovation Boulevard
Saskatoon, Saskatchewan
S7N 3H5 Canada

NATIONAL WATER
RESEARCH INSTITUTE
INSTITUT NATIONAL DE
RECHERCHE SUR LES EAUX

Institut national de recherche sur les eaux
Environnement Canada
Centre canadien des eaux intérieures
Case postale 5050
867, chemin Lakeshore
Burlington, Ontario
L7R 4A6 Canada

Centre national de recherche en hydrologie
11, boul. Innovation
Saskatoon, Saskatchewan
S7N 3H5 Canada

Environment Canada
Environnement Canada

Canada