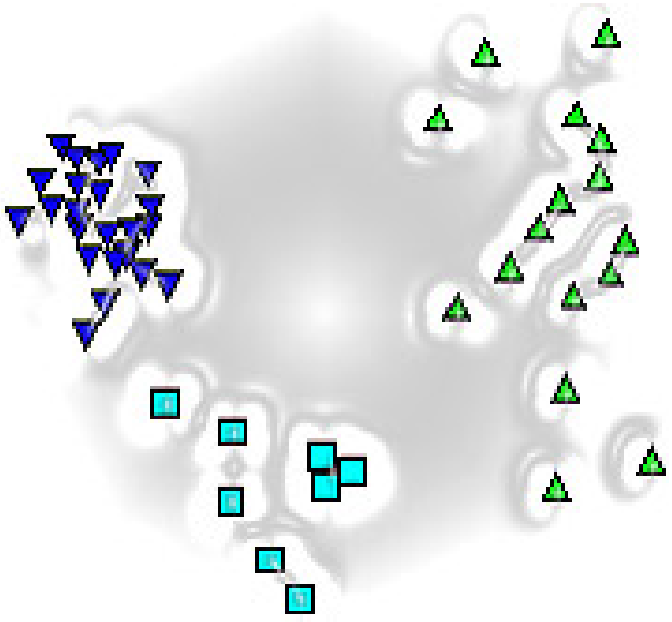
APPLICATION OF MULTIVARIATE TECHNIQUES TO EXAMINE QUALITY OF STREAMS IN THE OKANAGAN REGION

Final Report

March 31, 2006



LIMNOTEK

APPLICATION OF MULTIVARIATE TECHNIQUES TO EXAMINE QUALITY OF STREAMS IN THE OKANAGAN REGION

Final Report

Submitted to

B.C. Ministry of Environment
Penticton, B.C.

Prepared by

Chris J. Perrin, MSc. RPBio.
Limnotek Research and Development Inc.,
Vancouver, BC

March 31, 2006

Citation: Perrin, C.J. 2006. Application of multivariate techniques to examine quality of streams in the Okanagan region. Report prepared by Limnotek Research and Development Inc. for B.C. Environment. 39p.

Cover image: a stylized ordination plot of groups of sampling sites in the Okanagan region.

EXECUTIVE SUMMARY

The Reference Condition Approach (RCA) is gaining wide acceptance as an innovative and cost effective tool for water quality monitoring and site testing in British Columbia. For a particular region of interest, a multivariate model is used to explain variability of benthic invertebrate communities among reference sites using environmental attributes at those sites. The model is used to predict what biota should be present at a "test" site if it is in a reference condition. If the biota found at the test site is different from that predicted to be present, the site is rated as being stressed to varying degrees.

As part of a process to examine the potential merits of using the RCA in the Okanagan region, the assignment of stream condition based on an index of biotic integrity (IBI) was compared with a derivation of stream condition using the RCA. Biological and habitat data from selected stream sites were compiled by staff of the B.C. Ministry of Environment (MOE) and entered into a web accessed database called the Canadian Aquatic Biomonitoring Network (CABIN) that is maintained by Environment Canada. In the RCA procedures, all samples were tested for degree of stress using the Benthic Assessment of Sediment (BEAST) software that is available on-line on the CABIN website. Data from 90 site and year observations from the Okanagan region were independently tested using BEAST and compared to results from testing the same samples using an IBI that was developed by MOE staff.

The BEAST assessment was conservative in showing fewer unstressed sites and more severely stressed sites than was found by the IBI. There was complete agreement between the methods for 29% of the sites, BEAST assigned a worse condition than IBI for 55% of the sites, and BEAST assigned a better condition than IBI for 13% of the sites. Where differences in output between the methods occurred, a spread of one or two stress categories was found. A spread of more than two stress categories between methods was not found.

Both methods were good at detecting stress where a degree of disturbance was known to be present based on local knowledge of stream water quality. The variation in assigning a degree of stress between the methods was not surprising because the methods do different things. BEAST combines multivariate modeling of entire biological communities with a comparison of a test sample to a reference condition, which is the basis of the IBI. BEAST can be considered more comprehensive because it includes all taxa in the stream communities rather than using metrics of selected parts of communities, which is the focus of IBI. Error may have been introduced to the BEAST assessment by sample collection methods that were done according to IBI protocols, not the RCA procedures. BEAST assessment error may also have occurred by application of the Fraser/Georgia Basin model that does not include reference conditions in the Okanagan watershed. Despite these shortcomings, the substantial agreement between the BEAST and IBI approaches in detecting stress suggests that sampling method and geographic specificity of the BEAST model may not be critical for use in site testing. It is possible that a Surber sampler can be used for BEAST assessments although error would likely be reduced if collections were achieved by moving to multiple substrates within a sampling area, as is done using a kick net in the RCA procedures. In this respect one should not hesitate using the BEAST even when samples have not been collected with a kick net. In planning site testing wherein there is an intended application of the BEAST, it is recommended that sampling methods conform to RCA protocols to minimize prediction error.

The IBI and the RCA are complimentary. The key advantage of the RCA is that it integrates habitat attributes with information about the biota to enable an empirical definition of ecosystem health. That feature is powerful in making a case for site quality because it allows a water manager to have an understanding of habitat and biological attributes that should be present in an undisturbed state. If they are not present, data are available to "drill down" to examine potential cause of site disturbance. Clear hypotheses can then be tested using more detailed experimentation or monitoring. An advantage of IBI is that it produces a single score that is intuitively simple to understand and can be compared to target values. In adopting simplicity, however, it intentionally discards ecological information that may be important. However, data that are compiled for an IBI analysis may also be used for more in-depth analysis that could involve multivariate analysis of habitat attributes that may distinguish IBI scores and help to examine potential cause of site disturbance. Combinations of the benefits of IBI and RCA methods may be most powerful for water managers in the Okanagan. If it can be shown that sample collection methods can be integrated to a single technique, it is recommended that both approaches be applied to surface water monitoring in the Okanagan region. Once the reference condition model is in place, there is little difference in cost between using only one approach compared to using both approaches.

Supplemental multivariate analysis was run to examine habitat attributes of greatest importance in defining the quality of sampling sites in the Okanagan region. Cluster analysis and non-metric multidimensional scaling (MDS) was used to assign sample groups based on biological composition. This is the same process that is used in development of an RCA model, except in this case, the sample groups were specific to the Okanagan region. Discriminant function analysis (DFA) was used to identify habitat variables that were most important in discriminating between the sample groups. All data for the MDS and DFA analyses included data exported from CABIN and supplemental habitat information provided by the MOE.

Among all samples that were collected over several years from Okanagan streams, three sample groups were found by the combination of cluster analysis and MDS. All samples from streams that were known to be in a reference condition or influenced little by disturbance, based on local knowledge, were found to group tightly together while sites known to be affected to some degree by anthropogenic disturbance were found in the two other main groups. A list of 26 invertebrate families contributed to 70% or more of the dissimilarity between the three sample groups and 15 of these families were found to capture nearly the same multivariate pattern as the full set of 65 families that were found in all sample enumerations. The 26 families that best defined the dissimilarity between the sample groups included four mayfly families, six stonefly families, six caddisfly families, the naidid and lumbriculid worms, Elmid beetles, freshwater snails, and six true fly families including the chironomids.

The concentration of sediment PAH, sediment Mn, sediment Ni, and alkalinity were best at discriminating between the three sample groups. Known toxicity and relatively large difference in the concentration of sediment PAH compared to the other variables suggested that PAH constituents may be most important in determining site quality.

Site specific testing may be conducted to unequivocally show the effects of the contaminants. Simple ground truthing may reveal one or more sources of contaminants or diffuse sources from land use activities. The approach used in this report can be

imagined as a screening tool that focuses attention to “hot spots” of disturbance where follow-up detailed testing can be conducted to clearly define cause of disturbance.

Measurement of biological indicators in this approach is much more powerful than simple water sampling because it provides a time integrated measure of stress over the life spans of many taxa that use the habitat. Water chemistry and even sediment chemistry measurements that are compared to regulatory standards can miss detection of site contamination when it is present and thus is expected to be less sensitive and potentially have a higher error rate than the application of biological measures.

ACKNOWLEDGEMENTS

This project was completed under contract to B.C. Environment, Penticton, BC. Mr. Vic Jensen was the contract manager. Mr. Jensen is thanked for providing all data for this project. Vic is also thanked for our numerous discussions that helped to define the scope of work and potential application of multivariate analysis for the Okanagan region. I appreciate the assistance of Stephanie Sylvestre with resolving quirks with the Environment Canada CABIN web site that was used for preliminary site testing.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

1 INTRODUCTION

Benthic invertebrates are good indicators of water quality (Rosenberg and Resh 1993) and ecosystem health (Norris and Hawkins 2000), where the term "health" is useful for defining the condition of a stream that can be understood by the general public and resource managers. A stream in good condition or good health can be defined as having clean water and a functioning food web having a diversity of organisms that can support highly valued endemic fish species. A stream in poor condition might be one receiving some degree of anthropogenic disturbance that may modify the water chemistry, the physical habitat, or the biological communities that are considered representative of an undisturbed or pristine condition. Because of continuous exposure to water flow, benthic biota provide an integrated record of physical and chemical environmental quality. They are ubiquitous, largely sedentary, and there are large numbers of species that can provide an integrated measure of response to stress. Their characteristics allow effective spatial and temporal analyses of disturbance among stream reaches. The invertebrates act as a major food supply for fish, particularly salmonids, and provide an indication of food availability for fish populations through time and space. The result is that monitoring of benthic invertebrates can provide an indication of change in the chemical and physical attributes of the water they inhabit.

Some layouts and approaches to testing an effect or degree of disturbance in surface waters can involve multiple lines of evidence from a suite of univariate statistical tests. Where possible, control and impact stations can be sampled before and after start-up of a disturbance or discharge, thus facilitating a layout known as a before-after/control-impact (BACI) design (Steward-Oaten et al. 1986). In this approach, replicated differences between some measure (e.g. counts or biomass) at an upstream control site and a downstream treatment site among years before treatment are compared to the differences after treatment is applied using analysis of variance (ANOVA). In its simplest form on a single stream, years are replicated and the paired differences between an impacted and control site over years are analyzed by one way ANOVA. In a stronger case involving two or more control sites, asymmetric analysis of variance (Underwood 1994) can be used to test for an interaction between the difference between a measurement of some endpoint at an impacted location and that at control locations before compared to after a disturbance began. While there are many variations of these designs to test control versus potentially impacted sites over time, each involve univariate analytical approaches that can require sampling over years. They require the "correct" selection of an endpoint among many taxa. To avoid missing taxa of potential importance, total abundance or biomass may be analysed or a series of metrics (e.g. abundance of the combination of mayflies, stoneflies, caddisflies; abundance of chironomids; etc.) may be selected for independent analysis. Evidence from these multiple tests and other observations can then be combined to determine the effect of specific disturbances and to examine cause – effect pathways that are critical

for supporting water management decisions. Although the time required for field testing can be shortened and additional control can be applied to the tests by running experiments in mesocosm scale facilities (e.g. Perrin and Richardson 1997), most of these approaches require long time periods before a definitive description of water quality and cause – effect pathways may be found. They can be very expensive and impractical to complete on a large regional scale. In addition, basic assumptions of the statistical analyses may be violated perhaps due to insufficient funding to collect enough samples at any site, logistics that prevent repeated sampling, or other factors that constrain an ideal layout of sample collection. Hence, these approaches are most powerful at the site specific level and are best suited to definitive experimentation rather than providing evidence of water quality condition at a regional level.

Over large spatial scales, more rapid bioassessment procedures based on the use of multimetric indices of benthic invertebrate composition and abundance has gained widespread use as a screening tool for monitoring water quality, particularly in the United States (Karr 1981, Karr and Chu 1999, Barbour et al. 1999). A multimetric index is the combination of a number of individual metrics (e.g. number of mayflies, stoneflies, and caddisflies (EPT), percent chironomids, etc.) to form a single score. An assessment involves a comparison of the score found at a test site to that expected to be present in the absence of a disturbance (e.g. Kearns and Karr 1994). The Index of Biotic Integrity (IBI) that was developed by Karr (1981) and Karr and Chu (1999) is perhaps the best known and most widely used of the many multimetric bioassessment methods. Because the IBI requires development of a score for the undisturbed condition, it requires calibration throughout the region to which the IBI assessment is being applied. While best known in the United States, a multimetric IBI based on the methods developed by Karr (1981) was successfully developed for the Skeena region of British Columbia (Rysavy 2000).

Another rapid impact assessment approach that is based on multivariate statistical analysis and modeling (Bailey et al. 2004) is gaining increasing interest and widespread use, particularly in the UK (Wright et al. 2000), Australia (Parsons and Norris 1996), and Canada (Bailey et al. 2004, Sylvestre et al. 2005a, Reynoldson et al. 1997, Reynoldson et al. 2001). It is known as the Reference Condition Approach (RCA) (Bailey et al. 2004). For a particular watershed of interest, a multivariate model is developed to explain variability of benthic invertebrate communities among reference sites using environmental attributes at those sites. The model then predicts what biota should be present at a “test” site for a given set of environmental attributes found at that test site. If the biota found at the test site is similar to that predicted to be present, the site can be considered in a “reference condition”. If the biota found at the test site is different from that predicted to be present, the site is considered disturbed to varying degrees.

Both the IBI and the RCA can be considered screening tools for water quality assessment within a large region. Both approaches are based on the concept of comparison to a reference condition and can be considered complimentary. The IBI is based on the sum of a selected number of biological metrics that are found to be sensitive to gradients of water quality within a region. The RCA combines the ideas of multivariate modeling of entire biological communities (Wright et al. 2000) with the concept of comparison to a reference condition. RCA is more comprehensive because it includes complete communities rather than parts of communities in a final predictive model. The RCA differs from IBI in providing an explanation of habitat attributes that best explain variation in biological communities in the reference condition, which provides insight into site functioning and can be important in establishing cause – effect hypotheses for sites that are found to be impacted. While the RCA is more computationally complex than IBI, the computations can easily be run on a web site wherein calculations run behind the scenes, making site testing a very rapid and simple process. The website called CABIN (Canadian Aquatic Biomonitoring Network; *http://cabin.cciw.ca/cabin/asp/english/welcome.asp*) is the portal where testing of sites in Canada using the RCA can be run.

The use of the RCA is dependent on the availability of a model that describes the reference condition for a particular watershed or region of interest. Up until recently, these models have not been available for watersheds in British Columbia, but this is rapidly changing. An RCA model was first developed for the Fraser River Basin (Reynoldson et al. 2001). The Georgia Basin has now been added (Sylvestre et al. 2005a) and a model for the Skeena region with links to other northern basins is under development (Sharpe et al. 2005). Hence, the RCA approach for bioassessment can now be applied or soon will be applicable to much of the Province. With the establishment of CABIN and this increasing scope of development of the RCA models in British Columbia, there is increasing interest in use of the RCA.

Using benthic invertebrate counts from Surber samples collected in 1999 through 2004, Jensen (2006) developed an IBI for streams in the Okanagan region. Results showed a gradient of stresses ranging from sites that were found in very poor condition to those that were pristine and in excellent condition. With momentum shifting to the use of the RCA in British Columbia, there is interest by the Ministry of Environment (MOE) to further examine site condition by comparing the assignment of stream condition based on the IBI scores with a derivation of stream condition using the RCA. This report provides that comparison. Further multivariate analysis is provided to determine if groups of sites can be distinguished using measures of similarity and dissimilarity among the benthic invertebrate data and to determine if there are specific habitat variables that appear most important in determining any site grouping that may be present in the Okanagan region. Results from this report can be used as technical criteria to support mapping of zones or points of disturbance within the Okanagan region and to show what habitat attributes differentiate any zonation that may be present.

2 METHODS

2.1 Data Compilation

Biological data were compiled by staff of the B.C. Ministry of Environment, Penticton, BC.and entered into the CABIN database using procedures outlined by Sylvestre et al. (2005b). The compilation consisted of counts of benthic invertebrates per sample that were collected from 33 stream sites distributed within the Okanagan region (Table 1) as described by Jensen (2006). Those sites were sampled irregularly in 1999 through 2004 during the late August through October period, resulting in some sites being sampled annually, some being sampled in only one year, and some being sampled at an intermediate frequency. Three replicate samples were collected from each of 90 sampling episodes (site and time combinations), which resulted in 270 observations. Each sample was collected over a standard period of 1 minute using a Surber sampler that was equipped with a 210 µm mesh net. The sampling methods followed IBI protocols (Karr and Chu 1999). Sample sorting, subsampling and enumeration procedures were described by Jensen (2006).

Among all biological data compiled in CABIN for the Okanagan region, five sites were excluded from the present analyses because they were outside of the geographic area of interest. They included site codes called HAR01, HAR02, SND, STH, and YNG (V. Jensen, Ministry of Environment, Penticton, Pers. Comm.). The final compilation of biological data with the invertebrate taxonomic resolution set to the family level was used for all subsequent analyses (Appendix A).

At the time of invertebrate sampling, a suite of physical and chemical measurements and samples for analysis were collected as described by Jensen (2006). Measurements from surface water samples included alkalinity, conductivity, temperature, pH, and concentration of NO_3-N plus NO_2-N and total phosphorus (TP). Sediment samples were analyzed for concentration of polycyclic aromatic hydrocarbons (PAHs) and total Cr, Cu, Fe, Mn, Ni, Zn, As, Cd, Se, Ag, and Pb using the methods reported by Jensen (2006). The concentrations of As, Cd, and Pb could not be determined from most measurements because values were at or below the detection limit and the detection limit changed between 2001 and 2002. Concentrations for these metals were omitted from the final data compilation. Data for Se and Ag were also omitted because values were always less than the detectable limit. Data for physical habitat measurements included a categorical ranking of dominant and second dominant substrate, cover of coniferous trees, cover of deciduous trees, embeddedness, flow permanence, grass cover, macrophyte cover, presence of shrubs, slope, and surrounding material. Other measurements included percent pool, rapid, riffle, and run habitat. The average water velocity, bank full width, percent canopy closure, average

channel depth, maximum channel depth, and channel width was also measured. Methods for each of the measurements are described by Jensen (2006).

Table 1. List of stream sites in the Okanagan region where benthic invertebrate samples were collected in 1999 through 2004. The site codes and site names are from files provided by Jensen (B.C. Ministry of Environment, Penticton, BC. Pers. Comm.)

Site code	Site name	Years sampled
BEL01	Bellevue Cr at Lakeshore Rd	1999, 2000, 2001, 2004
BX01	BX creek	annually 1999 - 2004
BX02	BX Creek @ 30th	1999, 2000, 2001, 2004
BX03	BX creek2	1999 only
CHT01	Chute Cr @ Glenfir Rd E239620	1999, 2000, 2001, 2003
CLD01	Coldstream Cr upstream Municipal Intake E206374	2000 only
CLD02	Coldstream C @ Creekside park E208089	2002 only
ELL0104	Ellis Creek @ Mouth 0500027	annually 2000-2004
ELL02	Ellis C u/s Pent Diversion E224191	2003 and 2004
ENA01	Eneas C near Legion 0500684	annually 1999 - 2003
EQU01	Equesis C near Westside Rd 0500028	2002 and 2003
GRT01	Greata C near WSC stn E239618	1999 only
INK01	Inkaneep Creek upstream	2000 and 2001
INK02	Inkaneep Creek near mouth	annually 2000 - 2002
KEL0104	Kelowna Creek @ Abbott	1999, 2000, 2002, 2003, 2004
KEL0204	Kelowna Creek @ Bulman Rd	2003 and 2004
LBY01	Lambly Creek near Westside Rd 0500041	1999, 2000, 2003
MCD01	McDougall Creek @ Shannon Lake Rd E242784	2000 and 2001
NMT01	Naramata C near school 0500755	2000 and 2003
NMT02	Naramata Creek near mouth	1999 only
PCH01	Peachland C downstream Intake 0500856	annually 2000 - 2004
PCH02	Peachland Creek at Hwy 97	1999 only
PRA01	Prairie Creek near Okanagan Lake 0500325	1999 and 2000
PWR01	Powers Creek near mouth 0500059	1999, 2000, 2002
SGL01	Shingle Creek downstream Shatford Creek E242849	annually 2000 - 2003
SHR01	Shorts Creek near Westside 0500067	2000 and 2003
SHR02	Shorts Creek	1999 only
SHT01	Shuttleworth Creek @ Willow St E242896	annually 2000 - 2002
TRE01	Trepanier Creek near intake 500352	2000 only
TRT01	Trout Creek upstream Municipal Intake E22131	annually 1999 - 2003
TRT02	Trout Creek near Hwy 97 0500080	1999, 2000, 2002, 2003
VRN01	Vernon Creek @ 25th E249392	2002 only
WHT01	Whiteman Creek @ Westside br 0500099	2002 and 2003

Several other landscape variables were derived using GIS techniques that were complete by staff of the Ministry of Environment in Penticton. Measurements included ecoregion, altitude, stream order, distance from source (headwater origin or dam),

drainage area, latitude, longitude, total area, logged area, area of the forested landbase, area within the agricultural land reserve (ALR), urban area, number of road crossings, road length, area 100 m from streams, and road density. Other GIS variables were percent of total area that is logged, percent of forested landbase that is logged, percent of total area represented by ALR land, and percent of total area that is urban. Further explanation of the measured and derived variables is provided in Appendix C.

In preparation for running the RCA model, relevant habitat data were entered into the CABIN database by staff of the Ministry of Environment. The habitat variables included the surface water chemistry (alkalinity, conductivity, temperature, pH, and concentration of NO_3-N plus NO_2-N and TP) and all of the riparian and in-channel physical habitat measurements.

In addition to the standard site assessment in CABIN, other analyses were run in this project to examine groupings of sites based on similarities of biological composition and abundance between and among samples and to examine habitat attributes that best explained the dissimilarities of the sample groups, specific to the Okanagan region (Section 2.3). Raw data that were used to run the additional analyses were individually exported from CABIN and saved into Excel workbooks. The sediment chemistry and GIS landscape data were appended to the habitat data from CABIN to form a master habitat data file for statistical analyses. That habitat data is listed in Appendix B. QAQC included numerous checks to ensure that date and site coding in the appended data matched the date and site coding in the file that was exported from CABIN.

2.2 RCA analytical tools run on the CABIN website

The biological and habitat data residing on the CABIN database were examined using the Benthic Assessment of Sediment (BEAST) software that is available on-line on the CABIN website. Data from all 90 site and year observations were independently tested using BEAST. BEAST classified the habitat attributes of a test site to a reference condition that has already been defined for the Fraser River Basin (Rosenberg et al. 1999, Reynoldson et al. 2001, Sylvestre et al. 2005a). BEAST then compared the composition of the invertebrate community at the test site to that of the assigned reference condition and produced output in the form of an ordination plot. The degree of stress at the test site was related to the distance it was placed from reference sites on the ordination plot. Final output identified the level of stress from the ordination output and categorized it as unstressed, potentially stressed, stressed, or severely stressed according to methods described by Sylvestre et al. (2005b).

The model that was used by BEAST to classify a test site is the one that has been developed for a combination of reference conditions in the Fraser River Basin and the Georgia Basin (Sylvestre et al. 2005a). In that model the independent habitat variables include average channel depth, cover by coniferous trees, ecoregion, latitude,

stream order, wetted width, dominant substrate, embeddedness, pH, slope, and maximum velocity. Data were available for each of these variables for the Okanagan sites with the exception of maximum velocity (missing from all sites) and slope (missing from PCH02 in 1999 and from SGL01 in 2001, 2002, and 2003). Hence, the classification of test sites was run by excluding maximum velocity and slope from the model. This change was not considered important in affecting the error rate because both variables were relatively weak predictors in the DFA model (S. Sylvestre, Env. Can. Pers. Comm.).

Output from the BEAST analysis was used to construct a table in which each test site from the Okanagan was categorized as being unstressed, potentially stressed, stressed, or severely stressed according to the BEAST output. For each year of observation at a given site, the stress category that was assigned by BEAST was assigned a numeric code: unstressed was 1, potentially stressed was 2, stressed was 3, and severely stressed was 4. The average value of all codes that were assigned across years was the stress code that was assigned for a given site. The stress description for a site across all years was as follows: "unstressed" was any value less than 1.5, "potentially stressed" was any value between 1.5 and 2.5, "stressed" was any value between 2.5 and 3.5, and "severely stressed" was any value greater than or equal to 3.5. The stress category into which each site was placed by the IBI analyses, as reported by Jensen (2006), was added to the table to facilitate comparison of results. Some sites were also assigned a category of anthropogenic influence based on local knowledge. Before any IBI or BEAST analysis was run, Jensen (MOE, Penticton, Pers. Comm.) identified sites having a low influence from disturbance (CHT01, PCH01, SHR01, and WHT01) and other sites that were thought to be highly influenced by anthropogenic disturbance (KEL0104, PRA01, ENA01, BX01). The tabular comparison revealed similarities and differences in the assignment of site condition between the BEAST analysis, the IBI analysis, and local knowledge. In this comparison the IBI categories were changed to the same nomenclature that is used by BEAST as shown in Table 2.

Table 2. Corresponding stress categories used for comparison of RCA and IBI site classifications. Nomenclature in this report was standardized to the RCA classification.

RCA stress classification	Corresponding IBI stress classification
Unstressed	Excellent
Unstressed	Good
Potentially stressed	Fair
Stressed	Poor
Severely stressed	Very poor

2.3 Habitat attributes determining site condition

2.3.1 Classification of samples from sites of known condition

The output from BEAST was expected to reveal the degree of disturbance of each test site compared to a reference condition that was derived for the entire Fraser and Georgia Basins. Because those basins encompass a large geographic area that includes many ecoregions, there was interest by the MOE to determine whether any groups of samples that were characteristic of water quality condition could be distinguished within the Okanagan region alone. We might expect that Okanagan samples from sites found to be unstressed in the BEAST analysis might fall into one group (possibly more) while samples from sites that were found to have some degree of stress might fall into other groups, with the number of groups corresponding to different degrees of stress.

The sample groups were derived from the Okanagan biological data that presently reside in CABIN. Those data were exported from CABIN and in preparation for analysis, the average count of invertebrates, by family, among the three replicates from each combination of site and year was determined in Systat v8 (SPSS 1998). The average count for each invertebrate family across all site and year combinations was used in the subsequent multivariate analyses to examine sample groups.

Samples from sites known to be unstressed or stressed based on local knowledge (Jensen BC MOE, Penticton, Pers. Comm.) were first examined for obvious groupings. The unstressed sites were CHT01, PCH01, SHR01, and WHT01 and the stressed sites included BX01, ENA01, KEL0104, and PRA01. The multiple years of sampling at these sites resulted in 39 observations from unstressed sites and 48 observations from sites thought to be stressed to some degree. The data were opened in PRIMER multivariate analysis software (Clarke and Gorley 2001, Clarke and Warwick 2001) and the sample counts were fourth root transformed to down-weight the very abundant taxa and to allow the midrange and rarer taxa to exert some influence on the calculation of between – sample similarities. Similarities between every pair of samples was calculated using the Bray Curtis coefficient, which is defined as:

$$S_{jk} = 100\left\{1 - \frac{\sum_{i=1}^{p}\left|y_{ij} - y_{ik}\right|}{\sum_{i=1}^{p}(y_{ij} + y_{ik})}\right\}$$

where S_{jk} is the similarity between the *j*th and *k*th samples, y_{ij} is the abundance for the *i*th species in the *j*th sample, y_{ik} is the abundance for the *i*th species in the *k*th sample, and there are $i = 1,2,...,p$ species. Results formed a similarity matrix from which a dendrogram was plotted using the group average linkage in the hierarchical,

agglomerative clustering algorithm in PRIMER. The dendrogram was examined for obvious groupings of samples and group labels were assigned to each sample where groupings existed. An *a priori* decision was made that any one group should contain not less than 10% of the total number of data records at a given similarity level, which is similar to a rule applied to model development in the Reference Condition Approach for bio-assessment. The group assignment was assisted with interpretation of a non-metric multi-dimensional scaling analysis (MDS) that was run in PRIMER on the same similarity matrix that was used for the cluster analysis. MDS is a procedure for fitting a set of points in space such that the distances between points correspond as closely as possible to dissimilarities between objects. Output is displayed on two-dimensional images having no scaling units wherein space between objects on the image provides a perspective of dissimilarities. These images are called ordinations. An important computation that accompanies each ordination is something called a "stress" value. Stress increases with reducing dimensionality of the ordination and it indicates if a 2-dimensional plot is a usable summary of the sample relationships. An ordination of the Okanagan samples was only accepted if the stress value was <0.2, which reflects a useful 2-dimensional picture of sample relationships. Consistency between sample groupings on the ordination plot and the cluster dendrogram provided confidence in the group classifications. If one or more samples appeared in different sample groups between the different analyses, those samples were removed from the classification and the analyses were rerun. This process was continued until consistency of output between the cluster analysis and MDS was achieved.

A routine called SIMPER that is part of PRIMER was then used to determine which invertebrate families contributed most to the overall change in community structure between sample groups that was identified in the MDS analysis. This procedure compared the percentage composition each family made to the average dissimilarity between two sample groups and the average similarity within groups of samples. The families that cumulatively contributed to 70% or more of each of the between-groups dissimilarity were considered discriminators of sample groups. An algorithm called BVSTEP, which was also part of PRIMER, was then run to identify a subset of families whose among-sample relationships captured nearly the same multivariate pattern as the full set of families. The families in that subset were considered biological indicators of streams in the Okanagan region.

2.3.2 Classification of all samples

The analyses that are explained in Section 2.3.1 were run on only 48 of the 90 possible observations; the remaining 42 observations were from sites of unknown condition. A possible grouping of the outstanding observations was examined in another

sequence of analyses. The analytical procedures were the same as those described in Section 2.3.1 but they included data from all sites and times. Classification of the Okanagan sites in this way helped to identify zones of disturbance or "hot spots" of disturbance and it helped to reveal unstressed sites that were not initially known.

The classification of all observations proceeded in three steps:

Step 1: Cluster analysis and MDS was run on the complete set of biological data (90 observations). Any site for which a water quality condition was unknown that was found in close association with sites known to be unstressed was reassigned as an unstressed site. A test site was considered to be in close association with a group of unstressed sites if it met ***both*** of two rules:

1. The test site grouped tightly to the unstressed group in each of the clustering and MDS analyses using the grouping criterion that any one group could contain not less than 10% of the number of observations at a given similarity level.
2. The test site was shown to be either unstressed or only potentially stressed using either of the IBI analyses (Jensen 2006) or BEAST analyses (Section 2.2)

A test site was considered to be stressed if it met ***either*** of three other rules:

1. The test site grouped tightly to a stressed group in each of the clustering and MDS analyses using the grouping criterion that any one group could contain not less than 10% of the number of observations at a given similarity level.
2. The test site grouped within a new group that was not similar to a group of unstressed sites in output from the clustering and MDS analyses. Again, the grouping criterion was used that any one group could contain not less than 10% of the number of observations at a given similarity level.
3. The test site was shown to be either potentially stressed, stressed, or severely stressed using each of the IBI analyses (Jensen 2006) and BEAST analyses (Section 2.2).

Step 2: Observations that did not meet the rules that were applied in Step 1 were removed and were considered outliers. These outliers were observations that could not be clearly grouped and thus could not be used for subsequent analysis to determine the relative importance of habitat attributes in determining groupings of biological communities. Clustering and MDS analysis was repeated and the remaining outliers were removed. This process continued until clear groupings of observations without outliers were found.

Step 3: SIMPER was run in PRIMER to identify biological indicators of each new sample group using procedures described in Section 2.3.

2.3.3 Habitat attributes that discriminated between sample groups

Discriminant function analysis (DFA) was used to determine which habitat variables best discriminated between the sample groups that were defined by the clustering and MDS analyses. The DFA was first run on groups that were defined from samples of sites where stream condition was known from local knowledge. It was run a second time on the groups formed from the classification of all samples. The habitat variables were all of those included in the master habitat data file that was described in Section 2.1. All data except for percentages were $\log_{10}(x+1)$ transformed prior to analysis to standardize units of measure. The DFA was run in SYSTAT v8 (SPSS 1998) using the backward stepping procedure, resulting in the computation of what are called canonical variables. Canonical variables are linear combinations of measured predictor variables that best discriminate sample groups. The first of these canonical variables usually accounts for most (e.g. >95%) of the total dispersion of the sample groups and thus was used as the final model. Values called "F-to-remove" that were calculated for each predictor (e.g. habitat variable) in the backward stepping procedure indicated the relative importance of each predictor in contributing to the discriminant model. Relatively large F-to-remove values indicated greater importance of the associated habitat predictor variable than predictors having smaller F-to-remove values. In this way, habitat variables that were included in the model were identified and ranked as ones that were most important in discriminating sample groups grading down to ones that were least important in discriminating sample groups.

The accuracy of the discriminant model was determined using two procedures in SYSTAT v8 (SPSS 1998). A resubstitution procedure was run in which each case was reclassified to a group using the final model (e.g. the classification function or first canonical variable). This test is optimistic because the model is evaluated using all values that were used to compute it. One would expect a low rate of resubstitution error in this approach. A jackknife cross validation procedure was used as a second test in which the model was computed in the absence of one observation that was to be classified and then the model was used to classify the case that was removed. The term "jackknife" simply means that an observation is cut out for testing the predictive ability of the model by using that cut out observation.

An iterative approach was used to select the final model computed by the DFA. After a first run using the backwards stepping procedure, tolerance values that were computed with the F-to-remove values were examined. The tolerance values indicated correlation of a candidate variable with other variables in the discriminant model with a possible range of 0-1. Very low tolerance values indicated high redundancy between the candidate variable and one or more of the other model variables. Any tolerance values less than 0.1 were considered too low and indicated possible instability of the model. Any variable that had a low F-to-remove value (e.g. something less than 4) and a very

low tolerance value (e.g. something less than 0.1) was removed and the model was recomputed. This process was continued until tolerance values of selected predictors were all greater than or equal to 0.1. This process resulted in several computed models. The final accepted model was the one having acceptable tolerance values for each predictor variable (≥0.1) and the lowest reclassification error using both the resubstitution and jackknife cross validation tests. In this final model, the F-to-remove values were again examined. Variables with the highest F-to-remove values were considered most important in contributing to the discriminant model (e.g. classification of sites to groups) and variables with the lowest F-to-remove values were considered least important in contributing to the discriminant model. While it was desirable to have F-to-remove values that were >4, lower F-to-remove values were accepted in the final model if they contributed to lowering the overall model error rate as tested by the resubstitution and jackknife cross validation tests. The final model showed what physical and chemical attributes were most important in potentially contributing to the biological differences that were shown between groups of samples throughout the Okanagan region, including samples from paired sites that were upstream and downstream on single streams.

3 RESULTS

3.1 BEAST and IBI assessment of Okanagan stream sites

Similarities and differences between the IBI scoring and the BEAST output were found among all Okanagan stream sites (Table 3). The BEAST assessment was more conservative in showing fewer unstressed sites and more severely stressed sites than was found by the IBI (Figure 1). There was complete agreement between the methods for 29% of the sites, BEAST assigned a worse condition than IBI for 55% of the sites, and BEAST assigned a better condition than IBI for 13% of the sites. Where differences in output between the methods occurred, a spread of one (17 of 23 sites) or two (6 of 23 sites) stress categories was found. A spread of more than two stress categories between methods was not found.

Where some level of stress was present, Table 3 shows there was good agreement between the BEAST assessment, the IBI scores, and the *a priori* assignment of stress based on local knowledge. All sites that were thought by Jensen (MOE, Penticton, Pers. Comm.) to be highly influenced by anthropogenic stress were shown in the BEAST and IBI output to range from potentially stressed to severely stressed. There was some variation between the IBI and BEAST results in assigning a level of stress to these sites, but for any given site, any difference was only by one stress category.

Table 3. Assessment of Okanagan stream sites using IBI scoring, BEAST methods on the CABIN website, and local knowledge. The scoring nomenclature for IBI and BEAST is based on the BEAST output where U means unstressed, PS means potentially stressed, S means stressed, and SS means severely stressed.

Site Code	Site description	IBI scoring	BEAST assessment	Local knowledge*
BEL01	Bellevue Creek @ Lkshr br E241303	PS	PS	
BX01	BX creek	PS	PS	High influence
BX02	BX Creek @ 30th	SS	PS	
CHT01	Chute Cr @ Glenfir Rd E239620	U	PS	Low influence
CLD01	Coldstream Creek upstream of Municipal Intake 206374	PS	SS	
CLD02	Coldstream Creek @ Creekside park E208089	U	S	
ELL0104	Ellis Creek @ Mouth 0500027	S	SS	
ELL02	Ellis Creek u/s Pent Diversion E224191	U	PS	
ENA01	Eneas Creek near Legion 0500684	SS	S	High influence
EQU01	Equesis Creek near Westside Rd 0500028	U	S	
GRT01	Greata Creek near WSC stn E239618	PS	S	
INK01	Inkaneep Creek upstream	PS	PS	
INK02	Inkaneep Creek near mouth	PS	U	
KEL0104	Kelowna Creek @ Abbott	SS	SS	High influence
KEL0204	Kelowna Creek @ Bulman Rd	S	SS	
LBY01	Lambly Creek near Westside Rd 0500041	U	PS	
MCD01	McDougall Creek @ Shannon L Rd E242784	U	PS	
NMT01	Naramata Creek near school 0500755	Not tested	PS	
NMT02	Naramata Creek near mouth	PS	PS	
PCH01	Peachland C downstream Intake 0500856	U	U	Low influence
PCH02	Peachland Creek at Hwy 97	PS	SS	
PRA01	Prairie Creek near OK L 0500325	S	SS	High influence
PWR01	Powers Creek near mouth 0500059	PS	S	
SGL01	Shingle Creek d/s Shatford Creek E242849	PS	SS	
SHR01	Shorts Creek near Westside 0500067	U	PS	Low influence
SHR02	Shorts Creek	PS	PS	
SHT01	Shuttleworth Creek @ Willow st E242896	S	S	
TRE01	Trepanier Creek near intake 500352	PS	PS	
TRT01	Trout Creek u/s Municipal Intake E22131	PS	U	
TRT02	Trout Creek near Hwy 97 0500080	S	SS	
VRN01	Vernon Creek @ 25th E249392	S	PS	
WHT01	Whiteman Creek @ Westside br 0500099	U	PS	Low influence

*V. Jensen, BC MOE. Pers. Comm. “High influence” means the site was thought to be highly influenced by anthropogenic disturbance and “low influence” means the site was thought to be not influenced or influenced very little by anthropogenic disturbance.

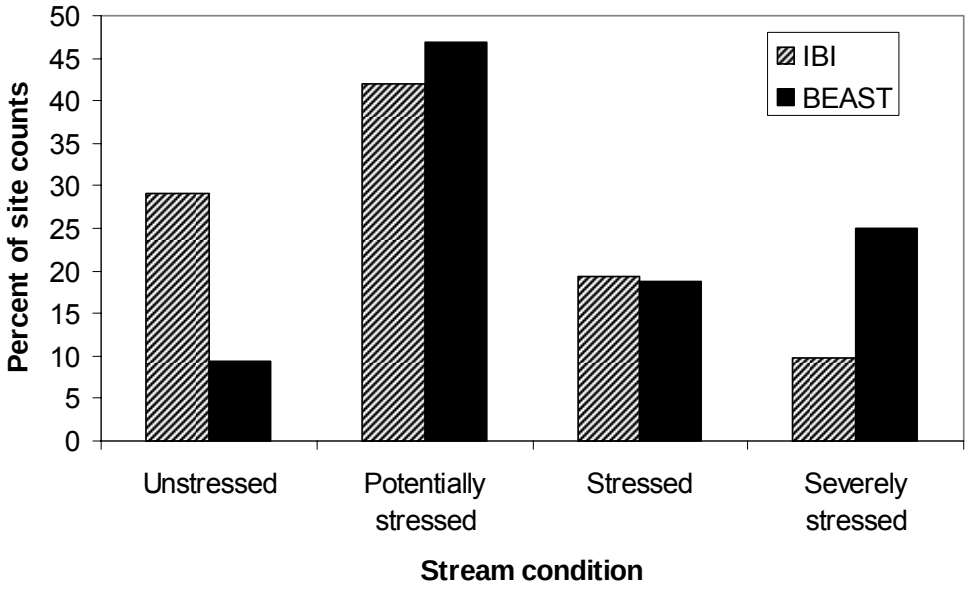


Figure 1. Relative differences between IBI and BEAST methods in assigning condition of Okanagan streams.

On each of four streams (Inkaneep Creek, Kelowna Creek, Trout Creek, and Ellis Creek), a site was sampled upstream and downstream of zones where disturbance was thought to affect stream water quality (Jensen, MOE, Penticton, Pers. Comm.). Paired site testing by either or both the IBI and BEAST would be expected to show a drop of at least two stress categories over the downstream gradient to provide confidence that a change in condition was present. This change was detected in Trout Creek and Ellis Creek but not in the other two streams (Table 4). No deterioration in condition was found by either IBI or BEAST over the downstream gradient in Inkaneep Creek. Both of the upstream and downstream sites were found to be potentially stressed or unstressed. BEAST found that Kelowna Creek was severely stressed at both of the paired sites while the IBI found only a marginal change from a stressed condition at the upstream site to a severely stressed condition downstream. BEAST detected a larger decline in condition than did IBI between upstream and downstream sites in Trout Creek while the two methods detected an equal degree of change (two stress categories) between the paired sites in Ellis Creek. In both streams, however, the BEAST found condition at the downstream site to be worse than was found by IBI.

Table 4. Comparison of stress levels determined by BEAST and IBI at paired upstream and downstream sites on selected Okanagan streams. The scoring nomenclature for IBI and BEAST is based on the BEAST output where U means unstressed, PS means potentially stressed, S means stressed, and SS means severely stressed.

Stream and site code	Method	Stream condition upstream	Stream condition downstream
Inkaneep Creek	IBI	PS	PS
Upstream is INK01 Downstream is INK02	BEAST	PS	U
Kelowna Creek	IBI	S	SS
Upstream is KEL0204 Downstream is KEL0104	BEAST	SS	SS
Trout Creek	IBI	PS	S
Upstream is TRT01 Downstream is TRT02	BEAST	U	SS
Ellis Creek	IBI	U	S
Upstream is ELL02 Downstream is ELL0104	BEAST	PS	SS

3.2 Habitat attributes determining site condition

3.2.1 Classification of samples from sites of known condition

Among the sites of known condition, three sample groups were found by the combination of cluster analysis and MDS (Figures 2 and 3). To define samples groups, a horizontal line was first drawn across the top of the cluster dendrogram. That line was moved down the dendrogram (in the direction of increasing between-sample similarity) to a point where fewest outliers and the largest number of groups was found using the 10% rule (any one group should contain not less than 10% of the total number of data records at a given similarity level). That point was approximately at the 55% similarity level. It showed three groups of samples and two outliers (BX0104 (site BX01 in 2004) and KEL010404 (site KEL0104 in 2004)). Group 1 included samples that were all from sites that were thought to be minimally impacted or had low influence from

anthropogenic stresses. Those sites were CHT01 sampled in 1999 – 2001 and in 2003, PCH01 sampled in 2000 – 2004, SHR01 sampled in 2000 and 2003, and WHT01 sampled in 2002 and 2003. Group 2 included three samples from two sites ENA01 in 1999 and PRA01 in 1999 and 2000, both of which were known to be influenced or stressed in some way. Group 3 included samples that were all from sites that were thought to be highly influenced from anthropogenic stresses. They included BX01 in 1999 – 2003, ENA01 in 2000 – 2003, and KEL0104 in 1999 – 2000 and 2002 – 2003. The MDS ordination confirmed this group allocation. The plot had a stress level of 0.15, which indicated a useful and acceptable 2-dimensional picture of sample relationships. Figure 3 shows the ordination plotted with groups from the cluster analysis distinguished from each other. The Group 1 samples that were thought to be in good condition were very tightly grouped, indicating very close similarities of benthic community structure and abundance between the samples. In contrast, the Group 3 samples were well separated from Group 1 and more spread out. This imaging suggested that Group 3 samples were not only very different from Group 1 but there was a greater range of sample similarities within Group 3 than in Group 1 with some samples indicating worse conditions that other samples. This effect was also apparent on the dendrogram; links between samples in Group 3 were spread over a greater range of the similarity scale than was found among the Group 1 samples. Group 2 samples were an intermediary group. Both the ordination plot and the dendrogram showed them to be more similar to the Group 1 samples than were the Group 3 samples but were sufficiently dissimilar from Group 1 to justify forming another group. The outliers were samples from sites that were clearly impacted because of their link to Group 3 but they were distant from Group 3 on both the ordination and dendrogram. Using the 10% rule, the two outlier samples were sufficiently remote on the similarity scale to not be considered part of Group 3.

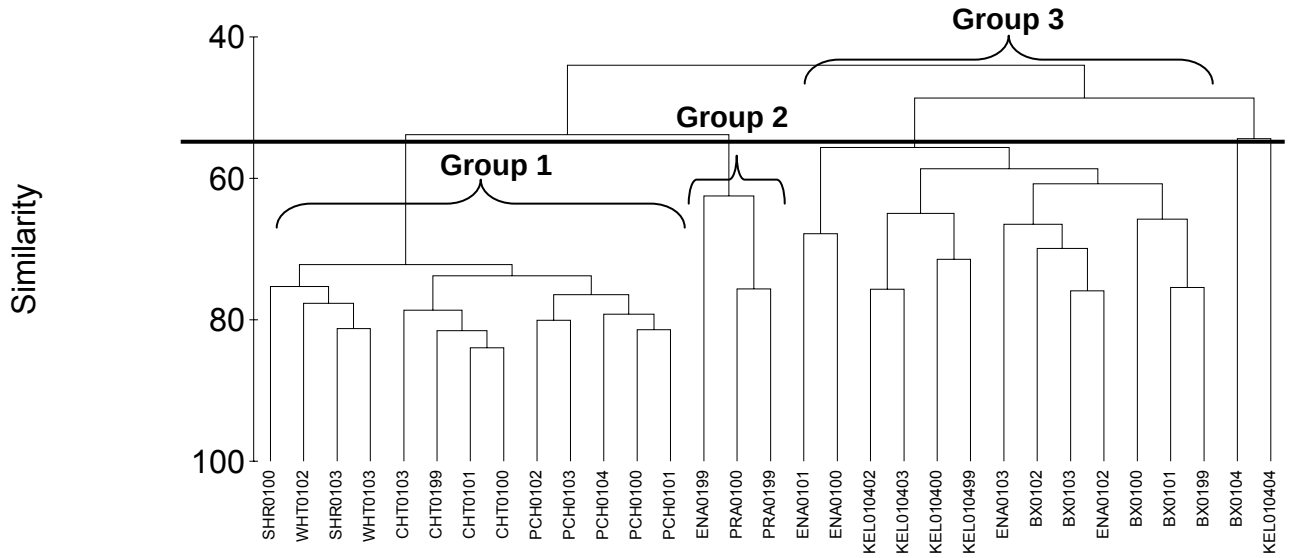


Figure 2. Assignment of groups of samples of known condition using cluster analysis with a rule that no one group could contain less than 10% of the total number of samples. Sample codes along the bottom scale follow the coding shown in Table 3 with an addition of two digits at the end showing year of sample collection (e.g. 01 means the sample was collected in 2001). By dropping the thick horizontal line down the similarity scale, the largest number of groups and fewest outliers were formed at approximately the 55% similarity level. The result formed three groups and two outliers (BX0104 (site BX01 in 2004) and KEL010404 (site KEL0104 in 2004)). Group 1 samples were from sites known from local knowledge to be in good condition while Group 2 and 3 samples were from sites known to be in poorer condition.

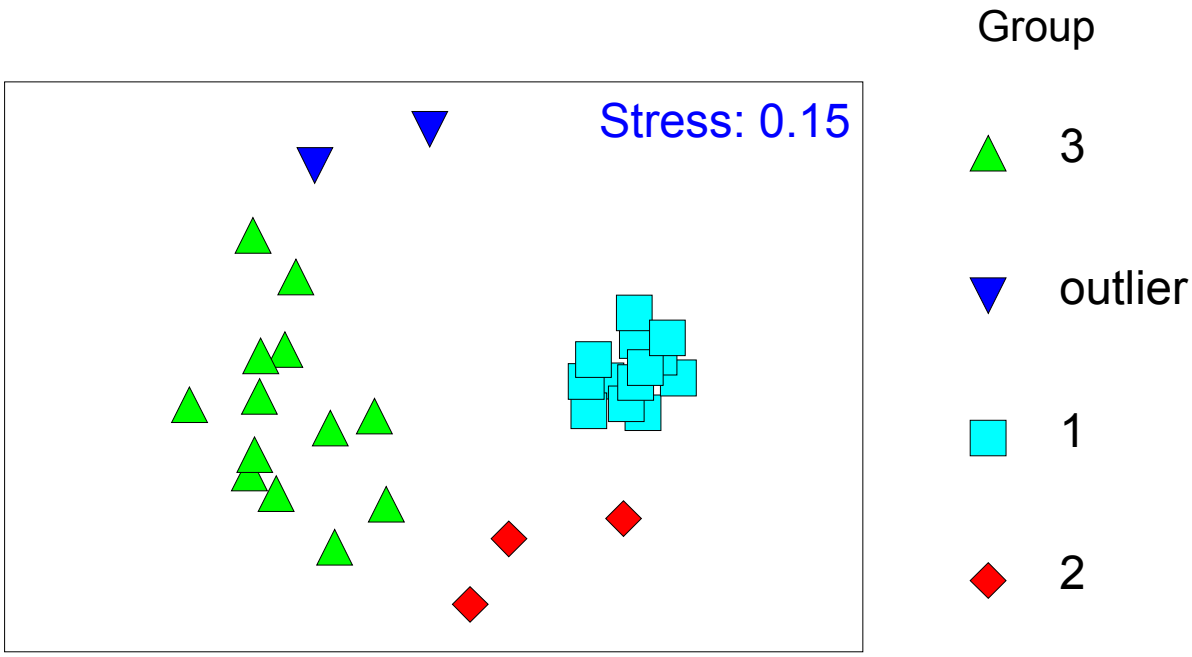


Figure 3. Ordination of samples from sites of known condition. The tightly packed light blue squares of Group 1 are samples from sites known to be in good condition. The more spread out green (Group 3) and red (Group 2) samples were from sites known to be in poorer condition. The two dark blue samples were the outliers shown in the cluster analysis. Group numbers refer to the same groups of samples that are shown on the cluster dendrogram in Figure 2.

A list of 21 families contributed to 70% or more of the dissimilarity between the three sample groups and 12 of these families were found to capture nearly the same multivariate pattern as the full set of 65 families that were found in the sample enumerations (Table 5). The families that best defined the dissimilarity between the sample groups included four mayfly families, five stonefly families, four caddisfly families, the naidid worms, Elmid beetles, freshwater snails, and five true fly families including the chironomids (midges). Abundance of the mayflies and stoneflies was greater among Group 1 samples than among samples from the other groups and of these families the Ephemereliidae, Heptageniidae, Leptophlebiidae, and Perlidae were found to be indicator taxa. With the exception of the Hydropsychidae, the abundance of the caddisflies was greatest in Group 1, next greatest in Group 2 and lowest in Group 3. The Hydropsychidae abundance was greatest in Group 3, next greatest in Group 1, and it was found in only trace numbers in Group 2. Three of the caddisfly families (Glossosomatidae, Hydropsychidae, and Rhyacophilidae) were found to be indicator taxa. With the exception of the black flies, abundance of all of the dipterans was greatest in Group 1. The chironomids represented more than 94% of the mean

abundance of this group of dipterans across all groups and as a result they were found to be an indicator family. The Simulidae (black flies) often occur in a very patchy distribution and were present in greatest abundance in Group 2 among all taxa. The simulids were an indicator family, which was likely due to this relative dominance in Group 2. Abundance of the naidid worms was highest in Group 1 and it graded down to lowest abundance in Group 3. Although the naidids are often associated with poor water quality, they are also ubiquitous and do not always indicate poor conditions. The present data shows an example where this is true given that they were found in highest abundance along with families that typically do indicate high water quality (mayflies, stoneflies and most caddisflies). The freshwater snails were absent from Group 1 samples and they were found in increasing but low abundance in Group 2 and 3 samples. The elmid beetles were unique in being in low abundance in Group 2 samples and in relatively high abundance in Group 1 and particularly in Group 3 samples. The snails and beetles were also found to be indicator families.

Five habitat variables were found to discriminate between the three groups of samples that were collected from the sites of known condition. They were sediment PAH concentration, alkalinity, sediment copper concentration, sediment chromium concentration, and sediment nickel concentration (Table 6). Both the resubstitution and cross validation model testing procedures showed 100% correct assignment of samples to groups. Based on the F-to-remove values, alkalinity was the most important discriminator, followed by PAH, sediment Cr concentration, sediment Ni concentration, and sediment Cu concentration. Table 6 shows that the Group 1 samples were exposed to relatively low sediment PAH concentration, low alkalinity, and moderate metals concentrations in the sediments. Group 2 samples were exposed to an order of magnitude higher PAH concentrations than were found in the Group 1 samples, high alkalinity, and moderate to low metals concentrations. Group 3 samples were exposed to very high PAH concentrations, high alkalinity, and relatively high metals concentrations in the sediments.

Table 5. Mean family abundance, by sample group, and percent contribution of families to sample group dissimilarity from sites of known condition. Blanks indicate the family was not important in contributing to between group dissimilarity (e.g. <1% dissimilarity between a pair of groups).

Order	Family	Common name	Mean abundance (number/sample)			Percent contribution to dissimilarity between group pairs		
			Group 1 (good condition)	Group 2 (intermediate condition)	Group 3 (poorer condition)	1 vrs 2	1 vrs 3	2 vrs 3
Basommatophora	Physidae*	freshwater snails	0	1.5	19	2	2.3	3.4
Coleoptera	Elmidae*	riffle beetles	116	27	380	2.4	2.3	6.1
Diptera	Ceratopogonidae	biting midges	13	0.7	0.6	3.3	3	1.5
Diptera	Chironomidae*	midges	601	50	416	5	2.7	4.9
Diptera	Empididae	dance flies	14	0	10	4	1.8	3.9
Diptera	Psychodidae	sand flies	14	0	0	4.1	3.4	
Diptera	Simuliidae*	black flies	13	116	61	4.3	2.9	3.9
Ephemeroptera	Baetidae	mayflies	338	99	196	3	3.1	3.6
Ephemeroptera	Ephemerellidae*	mayflies	205	0.3	0.6	7.5	6.5	
Ephemeroptera	Heptageniidae*	mayflies	215	11	0.2	5	7.1	4.5
Ephemeroptera	Leptophlebiidae*	mayflies	100	45	1.4	1.8	5.6	7.2
Haplotaxida	Naididae*	naidid worms	117	76	61	4.8	3.9	5.5
Plecoptera	Chloroperlidae	stoneflies	12	0	0.07	3.8	3	
Plecoptera	Leuctridae	stoneflies	14	0	0.1	3.3	2.7	
Plecoptera	Nemouridae	stoneflies	28	18	0.3		4.3	5.8
Plecoptera	Perlidae*	stoneflies	10	4	0	3.1	3.5	1.8
Plecoptera	Taeniopterygidae	stoneflies	28	0	0.07	3.6	2.9	
Trichoptera	Glossosomatidae*	caddisflies	24	2.7	0.3	3.7	3.8	1.8
Trichoptera	Hydropsychidae*	caddisflies	116	2.5	194	4.8	2.1	6.8
Trichoptera	Rhyacophilidae*	caddisflies	22	5	0.5	1.7	3.8	3.8
Tricoptera	Brachycentridae	caddisflies	32	12	2.2	1.9	3.8	5.1

*Included in a subset of 12 families that were found in the BVSTEP algorithm to capture nearly the same multivariate pattern as the set of all 65 families

Table 6. Mean values of discriminant variables identified with discriminant function analysis run on the three groups of samples collected from sites of known condition. The sample groups were defined by cluster analysis and MDS ordination of the biological data as shown in Figures 2 and 3. The F-to-remove values indicated relative importance of each variable as a predictor (highest values indicated greater importance as a predictor) and tolerance values indicated the correlation of a given predictor variable with other variables in the model. Tolerance above 0.1 was considered acceptable.

Variable	F-to-remove	tolerance	Mean concentration		
			Group 1 (n=18)	Group 2 (n=3)	Group 3 (n=21)
Sediment PAH	36	0.4	0.03 µg/g	0.5 µg/g	5 µg/g
Alkalinity	65	0.4	58 mg/L	324 mg/L	240 mg/L
Sediment Cu	1	0.5	12 µg/g	18 µg/g	36 µg/g
Sediment Ni	3	0.6	15 µg/g	4 µg/g	17 µg/g
Sediment Cr	14	0.3	22 µg/g	28 µg/g	28 µg/g

3.2.2 Classification of all samples

Despite the low error rate of the DFA that was run on samples from sites of known condition, small sample sizes among groups introduced some question of precision of the DFA model. Those small sample sizes were caused by low numbers of sites that were sampled over all years and by the DFA omitting entire cases when any habitat data was missing. The Group 1 data was restricted to 18 observations that were only available from CHT01 (Chute Creek) in 1999 - 2001 and from PCH01 (Peachland Creek) in 2000 – 2002. There were only 3 complete observations for Group 2, all coming from ENA01 in 1999. Group 3 was characterized by 21 observations, but even they came from only two sites (BX01 in 1999 through 2002 and ENA01 in 2000 – 2002).

To improve model precision and include a wider scope of sites for the assessment of regional water quality, the sample classification was expanded to include all samples and sites that were logged into CABIN.

Three sample groups were again found by the combination of cluster analysis and MDS (Figures 4 and 5). To define the sample groups, a horizontal line was again drawn across the top of the cluster dendrogram and moved down the scale of increasing similarity to a point where fewest outliers and the largest number of groups was found using the 10% rule (any one group should contain not less than 10% of the total number of data records at a given similarity level). That point was approximately at the 56% similarity level. Group 1 included samples from sites that were thought to be minimally impacted (Section 3.2.1; CHT01 sampled in 1999 – 2001 and in 2003, PCH01 sampled

in 2000 – 2004, SHR01 sampled in 2000 and 2003, and WHT01 sampled in 2002 and 2003). New sites added to Group 1 included LBY01 sampled in 1999 and 2000 and 2003, MCD01 sampled in 2000 and 2001, ELL02 sampled in 2003 and 2004, EQU01 sampled in 2002 and 2003, and NMT01 sampled in 2000. All of these new sites met the criterion for inclusion into an unstressed condition as outlined in Section 2.3.2. Group 2 included samples from sites that were most impacted in the analysis in Section 3.2.1 (BX01 in 1999 – 2003, ENA01 in 2000 – 2003, and KEL0104 in 1999 – 2000 and 2002 – 2003). Two new sites were added to this group; KEL0204 sampled in 2003 and 2004 and VRN01 sampled in 2002. Each of these new sites met the criterion for inclusion into a stressed condition as outlined in Section 2.3.2. Group 3 was new and did not include samples from sites that were previously grouped (Section 3.2.1). This new group included samples from SHT01 sampled in 2000 and 2001 and 2002, and from ELL0104 that was sampled each year in 2000 – 2004. Each of these sites were found to be either stressed or severely stressed in the IBI and BEAST analyses, which suggested this group was impacted in some way. All samples that were found in Group 2 from sites of known condition (Section 3.2.1) were found to be outliers in this cluster and MDS analysis, which indicated some unique attributes that did not fit with the other sample groups. Those sites (ENA01 in 1999, and PRA01 in 1999 and 2000) were not further considered here but they were examined in Section 3.2.1.

The MDS ordination (Figure 5) confirmed the group allocation. The plot had a stress level of 0.16, which indicated a useful and acceptable 2-dimensional picture of the sample relationships. The Group 1 samples that were thought to be in good condition were very tightly grouped, indicating very close similarities of benthic community structure and abundance between the samples. In contrast, the Group 2 and Group 3 samples were well separated from Group 1 and more spread out. This imaging suggested that the Group 2 and Group 3 samples were not only very different from Group 1 but there was a greater range of sample similarities within the Group 2 and the Group 3 samples than in Group 1. This effect was also apparent on the dendrogram; links between samples in Group 2 and between samples in Group 3 were spread over a greater range of the similarity scale than was found among the Group 1 samples.

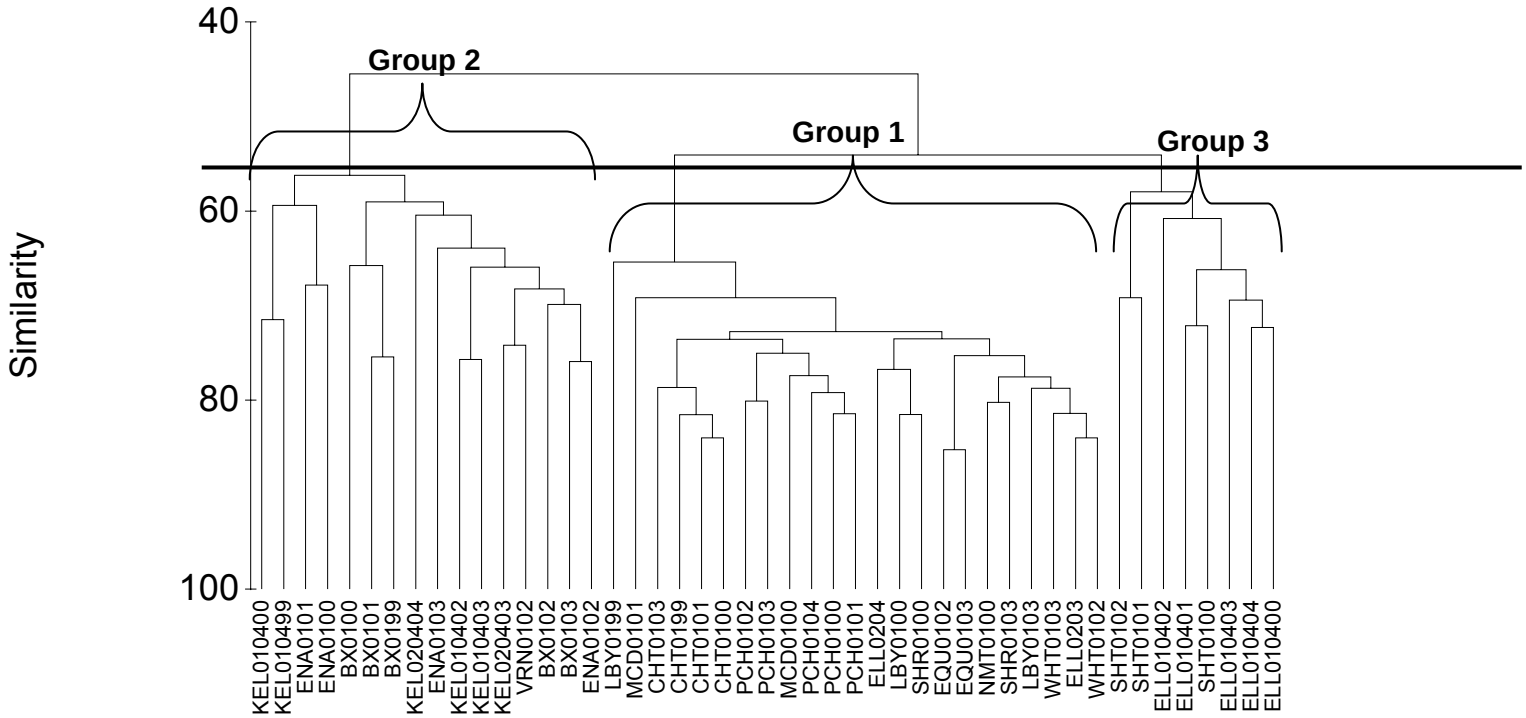


Figure 4. Assignment of groups of all samples with cluster analysis using a rule that no one group could contain less than 10% of the total number of samples. Sample codes along the bottom scale follow the coding shown in Table 3 with the addition of two digits at the end showing year of sample collection (e.g. 01 means the sample was collected in 2001). By dropping the thick horizontal line down the similarity scale, the largest number of groups and fewest outliers were formed at approximately the 56% similarity level. Note that samples from PWR, BEL, SGL, and TRE could not be clearly grouped and thus could not be used for subsequent analysis to determine the relative importance of habitat attributes in determining groupings of biological communities. These samples were omitted from this figure.

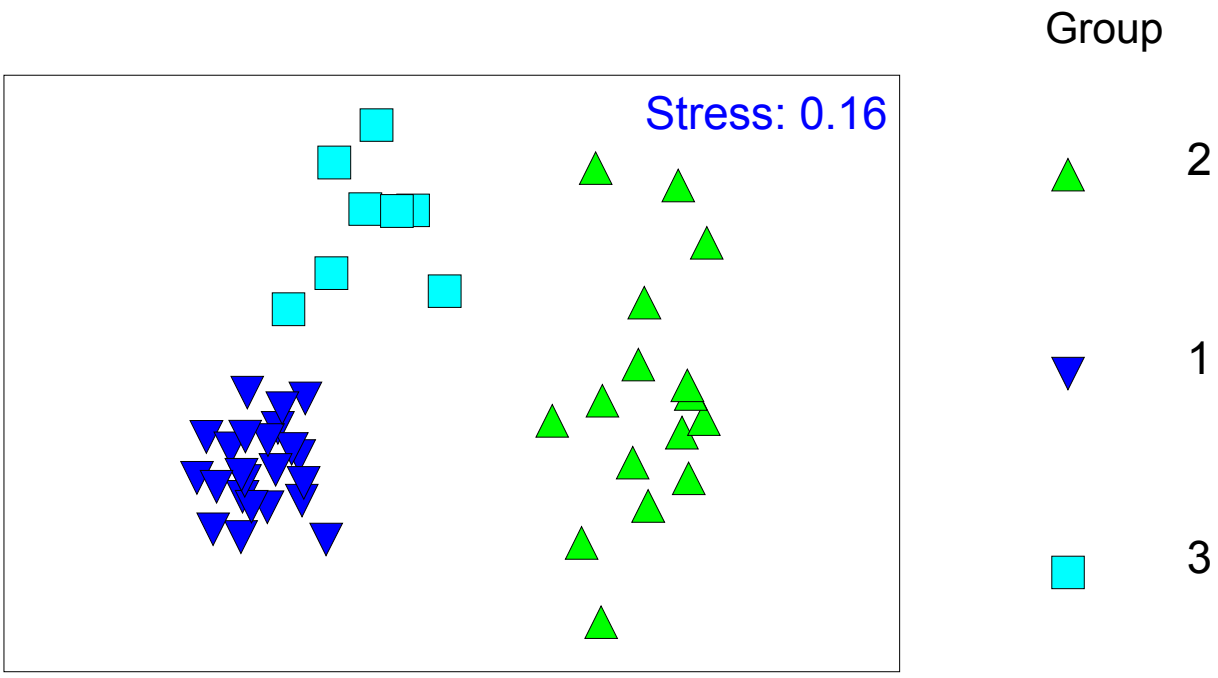


Figure 5. Ordination of all samples using MDS. Group numbers refer to the same groups of samples that are shown on the cluster dendrogram in Figure 4.

A list of 26 families contributed to 70% or more of the dissimilarity between the three sample groups and 15 of these families were found to capture nearly the same multivariate pattern as the full set of 65 families that were found in all sample enumerations (Table 7). The 26 families that best defined the dissimilarity between the sample groups included four mayfly families, six stonefly families, six caddisfly families, the naidid and lumbriculid worms, Elmid beetles, freshwater snails, and six true fly families including the chironomids (midges). All of the families that were found in the groups of samples from sites of known condition (Table 5) were again found among the three groups comprising all samples in Table 7.

More than 90% of the stoneflies were found in Group 1 samples. Counts among the six families were in trace numbers in the other two groups, indicating that conditions were relatively favourable in Group 1. Stoneflies are considered to be intolerant of disturbance and poor water quality (Reice and Wohlenberg 1993), which is evidence that water quality among Group 1 sites was relatively good.

Group 1 samples were characterized by relatively high abundance of all the mayfly families (829 individuals/sample). Excluding the Baetidae, Group 2 samples had very low abundance of the mayflies (2 individuals/sample). The Baetidae were one of the most abundant families in that group. This separation of the Baetids from the other mayflies suggests there was variation in environmental tolerances between the Baetidae and the other mayflies between Groups 1 and 2. Group 3 samples had an average count of 200 mayflies/sample with particularly high abundance of the Leptophlebiidae (125 individuals/sample).

With the exception of the Hydropsychidae, the abundance of the caddisflies was greatest in Group 1, next greatest in Group 2 and lowest in Group 3. The Hydropsychidae abundance was greatest in Group 2, next greatest in Group 3, and lower although still in high numbers in Group 1.

Overall dipteran abundance was greatest in Group 1 but there were differences in abundance among families between the sample groups. The chironomids represented 80% of total dipteran abundance among all groups with greatest abundance in Group 1 (604 individuals/sample) and lowest abundance in Group 3 (185 individuals/sample). Cumulative abundance of the other five dipteran families was greatest in Group 1 (130 individuals/sample), lower in Group 2 (110 individuals/sample), and lowest in Group 3 (68 individuals/sample). The Simuliidae abundance was notably higher in Group 2 than in the other groups.

Abundance of the naidid worms in Group 3 (170 individuals/sample) was approximately double the abundance in the other groups. Although numbers were very low, the Lumbriculidae worm abundance was at least 4 times greater in Groups 1 and 2 than it was in Group 3.

The freshwater snails were absent from Group 1 samples and they were found in increasing abundance in Group 2 and 3 samples.

The elmid beetles were unique in being in low abundance in Group 3 (21 animals/sample) and in high abundance in Group 1 (142 individuals/sample) and particularly in Group 2 (384 individuals/sample).

Table 7. Mean invertebrate abundance, by family and sample group, and percent contribution of families to sample group dissimilarity from all sites. Blanks indicate the family was not important in contributing to between group dissimilarity (e.g. <1% dissimilarity between a pair of groups).

Order	Family	Common name	Mean abundance (number/sample)			Percent contribution to dissimilarity between group pairs		
			Group 1	Group 2	Group 3	1 vrs 2	1 vrs 3	2 vrs 3
Basommatophora	Physidae*	freshwater snail	0	18	62	2.5	3.9	4.7
Coleoptera	Elmidae*	riffle beetles	142	384	21	2.7	2.8	6.2
Diptera	Ceratopogonidae*	biting midges	11	0.8	3	2.8	2.4	2.1
Diptera	Chironomidae*	midges	604	436	185	2.5	3	3.1
Diptera	Empididae	dance flies	18	11	19	1.7	1.4	2.2
Diptera	Psychodidae*	sand flies	23	0.1	0	3.4	4	
Diptera	Simuliidae*	black flies	25	90	5	3	3	5.3
Diptera	Tipulidae	crane flies	53	8	41	2.2	2.1	3.1
Ephemeroptera	Baetidae	mayflies	376	206	33	3	4.5	4.1
Ephemeroptera	Ephemerellidae*	mayflies	171	0.5	35	6.4	2.9	5.7
Ephemeroptera	Heptageniidae*	mayflies	208	0.3	7	7	5.6	3.1
Ephemeroptera	Leptophlebiidae*	mayflies	74	1.3	125	5.1	3.5	6.4
Haplotaxida	Naididae*	worms	90	70	170	3.9	5	5.9
Lumbriculida	Lumbriculidae*	worms	8	11	2	2	2.1	2.7
Plecoptera	Chloroperlidae*	stoneflies	13	0.1	0.6	3.4	3.4	
Plecoptera	Leuctridae	stoneflies	25	0.1	0	3.5	4.1	
Plecoptera	Nemouridae	stoneflies	26	0.3	2	4.3	3.9	1.6
Plecoptera	Perlidae	stoneflies	9	0	0.5	3.3	2.9	
Plecoptera	Taeniopterygidae	stoneflies	43	0.1	0.5	3.3	3.6	
Plecoptera	Perlodidae*	stoneflies	4	6	2	2.2	1.7	2.7
Trichoptera	Glossosomatidae	caddisflies	15	0.4	0	2.8	3.5	
Trichoptera	Hydropsychidae*	caddisflies	95	266	148	2.8	1.9	2.9
Trichoptera	Rhyacophilidae*	caddisflies	16	0.4	0.5	3.3	3.7	
Trichoptera	Lepidostomatidae	caddisflies	7	0.3	41	1.2	4	5.2
Trichoptera	Limnephilidae	caddisflies	3	0.3	14	1.3	2	2
Tricoptera	Brachycentridae	caddisflies	20	1.9	4	3.1	2.3	3.1

*Included in a subset of 12 families that were found in the BVSTEP algorithm to capture nearly the same multivariate pattern as the set of all 65 families

Four habitat variables were found to discriminate between the three groups of samples that were collected from all sites. They were sediment PAH concentration, alkalinity, sediment manganese concentration, and sediment nickel concentration (Table 8). Model testing by resubstitution showed 100% correct assignment of samples to groups and the cross validation test showed 95% accuracy. All errors shown by the cross validation test were in Group 3. Based on the F-to-remove values, alkalinity was the most important discriminator, followed by PAH, sediment Mn concentration, and sediment Ni concentration. Table 8 shows that the Group 1 samples were exposed to relatively low sediment PAH concentration, moderate alkalinity, high Mn concentration in sediments, and low Ni concentration in the sediments. Group 2 samples were exposed to two orders of magnitude higher PAH concentrations than were found in the Group 1 samples, high alkalinity, moderate sediment Mn concentration, and low sediment Ni concentration. Group 3 samples were exposed to PAH concentrations between those in Groups 1 and 2, low alkalinity, and relatively low metals concentrations in the sediments. Again the alkalinity is an index of concentration of bases (mainly OH^-, CO_3^{2-}, and HCO_3^-) and because it indicates acid neutralizing capacity, higher values may be associated with anomalous concentrations of cations, some of which may be metals or other contaminants.

Table 8. Mean values of discriminant variables identified with discriminant function analysis run on the three groups of samples collected from all sites. The sample groups were defined by cluster analysis and MDS ordination of the biological data as shown in Figures 4 and 5. The F-to-remove values indicate relative importance of each variable as a predictor (highest values indicated greater importance as a predictor) and tolerance values indicated the correlation of a given predictor variable with other variables in the model. Tolerance above 0.1 was considered acceptable.

Variable	F-to-remove	Tolerance	Mean concentration		
			Group 1 (n=24)	Group 2 (n=21)	Group 3 (n=18)
Sediment PAH	46	0.8	0.04 µg/g	5.0 µg/g	1.7 µg/g
Alkalinity	52	0.9	79 mg/L	240 mg/L	49 mg/L
Sediment Mn	21	0.6	615 µg/g	432 µg/g	234 µg/g
Sediment Ni	1	0.6	16 µg/g	17 µg/g	8 µg/g

4 DISCUSSION AND RECOMMENDATIONS

4.1 Site classifications by IBI and BEAST

One or sometimes two stress categories were found between the IBI and BEAST assessments, but evidence of site impairment was consistent between the two approaches. Both methods were good at detecting some level of stress where a degree of disturbance was thought to be present based on local knowledge of stream water quality by regional water management staff of the MOE. The variation in assigning a degree of stress between the methods was not surprising because they are doing different things. The RCA combines multivariate modeling of entire biological communities (Wright et al. 2000) with a comparison of a test sample to a reference condition. IBI compares scores of metrics (selected taxa considered sensitive to disturbance in the sampling area) to a reference condition (Karr 1981, Karr and Chu 1999, Barbour et al. 1999). RCA can be considered more comprehensive and more rigorous because it includes all taxa in the stream communities rather than using metrics of selected parts of communities, which is the focus of IBI. The IBI intentionally selects certain taxonomic groups that are shown to be sensitive to variation in water quality among sites within a region and so the final score is entirely based only on the presence and abundance of those taxa. Rare taxa that are not included in the list of metrics will be missed and so the IBI may be insensitive to somewhat subtle changes in community composition associated with anthropogenic stress. In contrast, the BEAST analysis includes all taxa, common and rare, and there is no preconceived notion about what taxa might respond to regional variation in water quality. Inclusion of all rare as well as common taxa that are tolerant or intolerant to disturbance intuitively makes the BEAST analysis more sensitive to detecting site disturbance. This difference between methods in what taxa are included and what taxa are not included in site testing may be one factor contributing to the BEAST finding more samples to be in a disturbed state than was found by the IBI among the Okanagan samples.

There may be some question whether the DFA model that was the basis of the BEAST assessment was relevant to sample testing in the Okanagan watershed. The model represents the reference condition among 274 stream sites that were sampled in 1994 through 2002 from locations in the Fraser River Basin (Rosenberg et al. 1999) and the Georgia Basin (Sylvestre et al. 2005a), all of which were outside of the Okanagan watershed. One might argue that biological attributes of the reference condition in Fraser and Georgia Basin streams might be different than the composition of reference sites within the drier climate of the Southern Interior Ecoregion (Perrin and Blyth 1998), where the Okanagan sampling sites were located. Regional biological differences are open to question and may have contributed to the BEAST assigning some degree of disturbance to sites that were locally known to be undisturbed or little affected by

anthropogenic activity (Table 3). This difference was not found in the comparison of IBI test results and the evidence from local knowledge, mainly because the IBI scoring was adjusted to a reference condition that came from local knowledge. Recent surveys in Yukon streams show substantial similarity of biological composition to that found in the Fraser Basin (B. Bailey, University of Western Ontario, Pers. Comm.). A similar overlap may exist in benthic communities between the Okanagan Basin and the Fraser Basin but some differences may be present that could contribute to the variation in results that were shown between the IBI and BEAST test results and between local knowledge and the BEAST test results.

Any error that was present could be resolved by development of an RCA model that is specific to the Southern Interior Ecoregion (Perrin and Blyth 1998). Given the potential for even drier conditions to develop in this ecoregion as a result of climate change (*http://www.climatechange.gc.ca/english/publications/okanagan*) and the likelihood of increasing demand for that water as human populations increase within the ecoregion (*http://www.climatechange.gc.ca/english/publications/okanagan*), there is substantial technical merit to development of a reference condition model for use in managing surface water quality. Ideally, the development of a model would be linked with a similar effort in the United States in a joint effort to classify and monitor quality of surface waters in the Okanagan watershed. This initiative will require additional sampling to increase sample size among reference condition sites. At present there are 69 samples from only 9 sites (Appendix B). The number of sites should be expanded to cover a wider spatial scope and the number of samples needs to be increased. A sample size of over 100 covering 30 or more sites is a minimum target to consider given the geographic complexity of the Okanagan watershed (Bailey et al. 2004). It is important to remember that reference site data are constantly re-used in the RCA approach and sequential sampling at new reference sites can be added as budgets allow over years, thus adding precision to the predictive model and increasing its use over time.

Another factor potentially contributing to the differences in outcome between the two methods may relate to how samples were collected. Triplicate Surber samples were collected from each site following the IBI protocols. Each replicate sample was a timed collection from a single placement of the sampler. In contrast, the RCA protocol is based on the timed collection and compositing of invertebrates using a kick net that is moved from spot to spot within a sampling site. Hence, fresh surfaces are repeatedly sampled and composited using the kick net while each triplicate Surber sample is collected from a single fixed point. The kick net sample may result in more animals and greater diversity of taxa within a sample than may be achieved in a single Surber sample. Because of this potential difference, the application of mean invertebrate counts that were calculated from the three replicate Surber samples in this study, rather than use of kick net data to the BEAST assessment may have resulted in the BEAST assigning a worse site condition to a sample than what actually may have been present.

This sampling effect may have contributed to BEAST assigning generally poorer conditions among the Okanagan sites than was found using the IBI.

The application of the BEAST analysis was opportunistic; an exercise to explore how different the outcomes of the BEAST and IBI approaches might be with recognition that the data used by the BEAST was not collected according to the RCA protocols and that the BEAST model may not have been entirely appropriate for the Okanagan streams. Given these shortcomings, it is perhaps surprising that the IBI and BEAST test results were not more different. The fact that there was substantial agreement between them suggests that sampling method and geographic specificity of the BEAST model may not be critical for use in site testing. It is possible that a Surber sampler can be used for BEAST assessments although error would likely be reduced if collections were achieved by moving to multiple substrates within a sampling area, as is done using a kick net in the RCA procedures. In this respect one should not hesitate using the BEAST even when samples have not been collected with a kick net. In planning site testing wherein there is an intended application of the BEAST, it is recommended, however, that sampling methods include kick net collections. This precaution to conform to RCA protocols will minimize potential error associated with sampling method and minimize error in site assessment using the BEAST.

Outcomes of testing the Okanagan sites suggest that either the IBI or BEAST can be used with confidence that a site that is impacted to some degree will be detected by either method. So the question is raised, why bother with the reference condition approach when the IBI appears to work just fine in the Okanagan region? Both methods are vast improvements over simple chemical analysis of water samples because the biological communities integrate effects of contaminants and other disturbance over time and give a truer picture of site condition than does monitoring that is restricted to instantaneous measurement of water chemistry. To answer to the question of “Why bother with RCA?” it is important to note that the reference condition approach is the bioassessment method of choice at the national level in many developed countries, including Canada. Within Canada, the development of a RCA model for the Fraser Basin (Rosenberg et al. 1999), the Georgia Basin (Sylvestre et al. 2005a), and most of the northern part of British Columbia (Sharpe et al. 2005) makes the Province a leader within Canada of the RCA approach (Bailey et al. 2004). With the exception of Ontario, no other Province has advanced the technique to this degree. The approach is cost effective over the long term and it is technically sound having been reviewed mulitiple times in peer reviewed international scientific journals (e.g. Bailey et al. 2004, Reynoldson et al. 1997, Reynoldson et al. 2001). Cost is initially high for the development of the reference condition model, but once that model is in place, site testing is inexpensive, it is scientifically robust, it is accepted internationally, and it can be done routinely as is done for water chemistry analysis. While it must be recognized that IBI and the RCA can be complimentary, the key advantage of the RCA is that it integrates habitat attributes with information about the biota to enable an empirical

definition of ecosystem health. That feature is powerful in making a case for site quality because it allows a water manager to have an understanding of habitat and biological attributes that should be present in an undisturbed state. If they are not present, data are available to "drill down" to examine potential cause of site disturbance as was done in this report. Clear hypotheses can then be tested using more detailed experimentation or monitoring as necessary to meet water management objectives. An advantage of IBI is that it produces a single score that is intuitively simple to understand and can be compared to target values. In adopting simplicity, however, it discards ecological information that may be important. Another important feature of IBI is that it is commonly used in the United States. Water management in the Okanagan Region of British Columbia is closely tied to that further downstream in the United States. Transboundary water management agreements may opt for the use of IBI because it is in place and working in the US and Canada. From this point of view, continued use of IBI may be warranted, but given the greater sensitivity of RCA, combinations of the two approaches may be most powerful for water managers, particularly in the Okanagan where correspondence of indicators used on both sides of the Canada US border are of great interest. If it can be shown that sample collection methods can be integrated to a single technique in the Okanagan region, it is recommended that both approaches be applied to surface water monitoring in the Okanagan region. Once the reference condition model is in place, there is little difference in cost between using only one approach compared to using both approaches.

4.2 Indicator taxa

The application of cluster analysis and MDS distinguished three groups of samples from Okanagan streams. All samples from streams that were known to be in a reference condition or influenced little by disturbance (based on local knowledge) were found to group tightly together (Group 1) while sites known to be affected to some degree were found in the two other main groups. Hence, there was excellent agreement between the statistical tests based on similarity of invertebrate composition and abundance between samples and knowledge of local water quality conditions. Local knowledge based on a long history of water management experience is seldom wrong. Hence, the grouping of low influence sites together and higher influence sites together provided confidence that unique attributes of the composition of invertebrates between the sample groups indicated relatively tolerant (Groups 2 and 3) and relatively intolerant (Group 1) taxa among the Okanagan stream sites.

Most mayflies and stoneflies were found in greatest abundance and diversity where water quality was thought to be highest. These insect orders are known to be relatively intolerant of site disturbance (Reice and Wohlenberg 1993) and are often used

independently or as a combined metric in IBI assessments to indicate a reference state (Bennett and Rysavy 2003). The relative abundance of these orders among Group 1 samples implied relatively good conditions at sites where those samples were collected (Chute Creek at CHT01, Ellis Creek at ELL02, Equesis Creek at EQU01, Lambly Creek at LBY01, McDougall Creek at MCD01, Naramata Creek at NMT01, Peachland Creek at PCH01, Shorts Creek at SHR01, and Whiteman Creek at WHT01). These sites were considered to be in a reference condition mainly because of this association between the relatively high abundance of the pollution intolerant mayflies and stoneflies and local knowledge that these sites were either undisturbed or under little influence from anthropogenic disturbance.

In contrast, the abundance of the naidid worms in Group 3 was evidence of some disturbance at those Group 3 sites. The naidid worms can be ubiquitous but they increase in relative abundance where the more pollution intolerant taxa are rare or absent, thus indicating disturbance. Group 3 was also characterized by an abundance of snails, midges, Leptophlebid mayflies, and three caddisfly families (Hydropsychidae, Lepidostomatidae, Limnephilidae) and relatively few beetles, some dipterans, stoneflies, and some caddisflies. Group 2 was characterized by a lack of stoneflies and most caddisflies and mayflies, but an abundance of beetles, midges, Baetid mayflies, and Hydropsychid caddisflies.

In all of these groupings, there was considerable overlap of similar invertebrate orders across groups (e.g. caddisflies, mayflies, dipterans) but it was variation in abundance of the families that distinguished the sample groups. While disturbance of streams is often described in terms of the abundance or relative abundance of insect orders (e.g. mayflies and stoneflies are pollution intolerant (Karr and Chu 1999)), the Okanagan sample groups show that site quality is better distinguished at the family level. Many of the samples that were clustered in Group 2 were locally known to be influenced to some degree by disturbance (BX Creek at BX01, Eneas Creek at ENA01, Kelowna Creek at KEL0104, and Prairie Creek at PRA01). Relatively poor water quality at those sites was found to be reflected in the presence of freshwater snails, low abundance of the Ephemerelid and Heptagenid mayflies, almost complete absence of all stoneflies, and low abundance of the non-Hydropsychid caddisflies.

4.3 Habitat attributes determining site condition

Output of the IBI and BEAST indicated that the biological composition of some of the samples was outside of a reference condition. The output did not indicate what factors may have caused potential site disturbance. The subsequent DFA was a valuable tool to supplement the initial site testing and contribute insight into factors

contributing to disturbance among many sites. It was a method of "drilling down" through the multivariate data to reveal attributes of the stream habitats that might be important in contributing to site impairment. DFA is the statistical technique that is used in BEAST to test whether a test site is within or outside of a reference condition. BEAST uses a model that is already built for this testing procedure. It runs in the background on the CABIN website when each site is tested.

In the present case, a new DFA model was built just using the Okanagan data. The output showed what physical and chemical variables best discriminated sample groups that were defined by the biological composition among all sites. The model building was different from that used to construct a model of the reference condition in that habitat variables that might be affected by pollution or indicated the presence of a pollutant or some form of anthropogenic activity were not excluded. All variables were considered important. Variables that were a measure of chemical or physical disturbance and were found to be important predictors in the DFA model were particularly valuable in contributing insight into why groups of samples might be distinguished by the clustering and MDS analyses.

Development of the DFA model was iterative and while rules were set for the acceptance or rejection of variables, it is important to note that there is no absolutely correct model. Iterations proceeded until a model was derived having low error, low tolerance between predictor variables (low correlation between variables) and as high F-to-remove values as was possible to achieve a low error rate. One could argue that another combination of variables would be equally viable but the ones that were finally selected were the best predictors based on the criteria that were established a priori and based on multiple model runs.

This process showed that the combination of concentration of sediment PAH, sediment Mn, sediment Ni, and alkalinity were best at discriminating between the three sample groups when all samples were included. It did not show that the discriminating variables *caused* change in biological composition of the sample groups. It only showed that PAH, alkalinity and the two metals were good discriminators between the sample groups that were distinguished by the composition and abundance of all invertebrates.

Mn is ubiquitous in the environment as a consequence of the weathering of crustal rock (Howe et al. 2004). It is present in most vegetation and thus is released to surface waters after forest fires that have been common in the Okanagan region in recent years. The major anthropogenic source of manganese that may be present at the Okanagan stream sites may be the combustion of fossil fuels. Mn can be toxic to aquatic biota (Howe et al. 2004) but sediment quality guideline concentrations are not reported for Canadian waters by CCME (2003). It is unknown if the difference in Mn concentrations in sediments between the Okanagan sample groups was sufficient to contribute either directly via toxicity or indirectly to the differences in biota between the three sample groups.

Nickel is also a crustal metal and has been studied little with respect to aquatic toxicity (Keithly et al. 2004). Very few invertebrate taxa have been tested for Ni toxicity. Modes of toxicity are not known except that toxicity decreases with increasing water hardness (Keithly et al. 2004). Hence, potential toxicity by Ni may be present in Okanagan streams but again sediment quality concentrations are not reported for Canadian waters by CCME (2003). Again, it is unknown if the relatively small difference in Ni concentrations in sediments between the Okanagan sample groups was sufficient to contribute either directly via toxicity or indirectly to the differences in biota between the three sample groups.

Alkalinity is an index of concentration of bases (mainly OH^-, CO_3^{2-} and HCO_3^-) and because it indicates acid neutralizing capacity, it may be associated with anomalous concentrations of cations, some of which may be metals or other contaminants.

PAH's are well known to be toxic to aquatic biota (CCME 2003) and are introduced to surface waters via atmospheric deposition from forest fires, volcanic activity, combustion of fossil fuels, among other sources. They persist in stream sediments because they are relatively non-volatile and poorly soluble. Hence, benthic invertebrates may be one of the best indicators of the presence of PAH contamination because of direct contact in stream and lake sediments. The difference in total PAH concentration particularly between Group 1 samples and the other groups was more than two orders of magnitude, which was much more than was found for any of the other predictor variables. Low PAH concentrations were associated with Group 1 samples that were considered to be in a reference condition while the high PAH concentrations were associated with the other two groups that were biologically different from the reference condition and were thought to be impacted to some degree by anthropogenic stress based on local knowledge. This close association between PAH concentration and biological condition suggests that one or more of the PAH constituents may have contributed to lower quality of water among the sites from which samples of Groups 2 and 3 were collected.

The preliminary DFA model that revealed habitat variables that best discriminated groups of samples that were collected only from sites of known condition showed that concentration of sediment Cu and Cr were important predictors along with alkalinity and sediment Ni and PAH. Cr and particularly Cu can be toxic to aquatic biota (CCME 2003), but the model output can be questioned because of the small sample sizes. A sample of only 3 observations, all coming from a single site to form Group 2 was considered too small and too spatially restrictive. It may have produced misleading results, including the list of predictor variables. For this reason, one should not accept the list of predictor variables in Table 6 with confidence.

4.4 Application of the multivariate analyses

A question that may arise after the multivariate analyses is "Where do we go from here?". The main advantage of the DFA is it provided a short list of candidate habitat variables that were important in discriminating groups of samples from the Okanagan region. Known toxicity and relatively large difference in the concentration of sediment PAH compared to the other variables suggested that PAH constituents may be most important in determining site quality. The sites that hosted the highest concentration of PAH can be pinpointed as potential "hot spots" in the Okanagan region. Those were sites belonging to Group 2 (BX Creek at BX01, Eneas Creek at ENA01, Kelowna Creek at KEL0104, and Vernon Creek at VRN01). Other sites hosting lower but still high concentrations of PAH's might receive a different coding (e.g. "warm spots"). They included Group 3 sites, which actually consisted only two sites from which 24 samples were collected in 2000 – 2004 (Appendix B). Those sites were Ellis Creek at ELL0104 and Shuttleworth Creek at SHT01. Other sampled sites might be coded "cool sites" because they were found to represent a reference condition. The list of predictor variables that resulted from the DFA are hypothesized to contribute to site quality. Site specific testing may be conducted, potentially by experimentation, to unequivocally show the effects of those variables. If this experimental testing or even simple observations of land use at "hot spots" clearly reveals one or more sources of contaminants or diffuse sources from land use activities, water managers may engage in actions to change land use or other activity to improve quality of water at those hot or warm spots, if warranted among water management priorities.

Application of DFA modeling to examine water quality attributes close to the source is a proactive approach to management of water quality at downstream sites that are heavily used by people. Measurement of biological indicators, whether by multivariate or multimetric approaches, is much more powerful than simple water sampling because it provides a time integrated measure of stress over the life spans of many taxa that use the habitat. Instantaneous water chemistry and even sediment chemistry measurement can miss detection of site contamination when it is present and thus is expected to have a higher error rate than the application of biological measures.

5 LIST OF REFERENCES

Bailey, R.C., R.H. Norris, and T.B. Reynoldson. 2004. Bioassessment of freshwater ecosystems using the reference condition approach. Kluwer Academic Publishers. Norwell, MA. 170p.

Barbour, M.T., J. Gerritsen, B.K. Snyder and J.B. Stribling. 1999. Rapid bioassessment protocols for use in streams and wadeable rivers: Periphyton, benthic macroinvertebrates and fish. Second Edition. Assessment and Water Protection Division, U.S. Environmental Protection Agency. Report EPA 841-B-99-002. Washington, D.C. *http://www.epa.gov/owow/monitoring/rbp* .

Bennett, S. and K. Rysavy. 2003. Guidelines for calibrating a benthic invertebrate multimetric index of biological integrity (B-IBI) for streams in British Columbia. Report prepared by Bio Logic Consulting Ltd. for B.C. Ministry of Sustainable Resource Management. Victoria, B.C. 43p.

CCME. 2003. Canadian Environmental Quality Guidelines. Canadian Council of Ministers of the Environment. Winnipeg. Manitoba.

Clarke, K.R. and R.N. Gorley. 2001. PRIMER v5: User Manual/Tutorial. PRIMER-E: Plymouth. UK.

Clarke, K.R. and R.M. Warwick. 2001. Change in marine communities: an approach to statistical analysis and interpretation, 2nd edition. PRIMER-E: Plymouth UK.

Howe, P.D., H.M. Malcolm, and S. Dobson. 2004. Manganese and its compounds: environmental aspects. Inter-Organization Programme for the Sound Management of Chemicals. World Health Organization. Geneva.

Jensen, E.V. 2006. Cumulative Effects Monitoring of Okanagan Streams using Benthic Invertebrates, 1999 to 2004. B.C. Ministry of Environment draft management report. Penticton, B.C. 32p.

Kearns, 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.

Keithly, J., J.A. Brooker, D.K. DeForest, B. K. Wu, and K.V. Brix. 2004. Acute and chronic toxicity of nickel to a cladoceran (*Ceriodaphnia Dubia*) and an amphipod (*Hyalella Azteca*). Environmental Toxicology and Chemistry. 23: 691-696.

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

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

Norris, R.H. and C.P. Hawkins. 2000. Monitoring river health. Hydrobiologia. 435: 5-17.

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.

Perrin, C.J. and C.A. Blyth. 1998. An ecozone classification for lakes and streams of British Columbia; Version 1.0. Report prepared by Limnotek Research and Development Inc. and AXYS Environmental Consulting Ltd. for Ministry of Environment, Lands and Parks. Water Quality Branch. Victoria, B.C. 95p plus map.

Perrin, C.J. and J.S. Richardson. 1997. N and P limitation of benthos abundance in the Nechako River, British Columbia. Can. J. Fish. Aquat. Sci. 54: 2574-2583.

Reice, S.R. and M. Wohlenberg. 1993. Monitoring freshwater benthic macroinvertebrates and benthic processes: measures for assessment of ecosystem health. In: Rosenberg, D.M. and V.H. Resh. (Ed). 1993. Freshwater biomonitoring and benthic macroinvertebrates. Chapman and Hall. New York.

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

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

Rosenberg, D.M. and V.H. Resh. (Ed). 1993. Freshwater biomonitoring and benthic macroinvertebrates. Chapman and Hall. New York.

Rosenberg, D.M., T.B. Reynoldson and V.H. Resh. 1999. Establishing reference conditions for benthic invertebrate monitoring in the Fraser River catchment, British Columbia, Canada. Environment Canada's Fraser River Action Plan Report DOE FRAP 1998-32. *http://www.rem.sfu.ca/FRAP/9832.pdf*.

Rysavy. S. 2000. Calibration of a multimetric benthic invertebrate index of biological integrity for the upper Bulkley River watershed: A tool for assessing and monitoring stream condition. Report prepared by Bio Logic Consulting Ltd. for Community Futures Development Corporation of Nadina. Houston, B.C. 20p plus appendices.

Sharpe, I., C.J. Perrin, S. Bennett, S. Linke, and B. Bailey. 2005. Benthic macroinvertebrate sustainability indicator development project: Summary of progress in year 1. Report prepared by B.C. Ministry of Water Land and Air Protection. Smithers, B.C. 40 pp.

SPSS. 1998. SYSTAT 8.0. SPSS Inc. Chicago Ill.

Stewart-Oaten, A. W.M. Murdoch, and K.R. Parker. 1986. Environmental impact assessment: "Pseudoreplication" in time? Ecology. 67: 929-940.

Sylvestre, S., M. Fluegel, and T. Tuominen. 2005a. Benthic invertebrate assessment of streams in the Georgia Basin using the reference condition approach: expansion of the Fraser River invertebrate monitoring program 1998-2002. Environment Canada report EC/GB/04/81.

Sylvestre, S., T. Reynoldson, T. Pascoe, I. Wong, T. Tuominen, and M. Fluegel. 2005b. CABIN data management and BEAST analysis handbook. Environment Canada. Vancouver, B.C. 38p.

Underwood, A.J. 1994. On beyond BACI: sampling designs that might reliably detect environmental disturbances. Ecological Applications 4: 3-15.

Wright, J.F., D.W. Sutcliffe and M.T. Furse. 2000. Assessing the Biological Quality of Freshwaters. RIVPACS and Other Techniques. Freshwater Biological Association, Ambleside, UK. 373pp.

6 RAW DATA APPENDICES

Raw data appendices are available on CD or via file transfer from B.C. Ministry of Environment.