

## CHAPTER 12

# Establishing reference conditions in the Fraser River catchment, British Columbia, Canada, using the BEAST (Benthic Assessment of Sediment) predictive model

DAVID M. ROSENBERG<sup>1</sup>, TREFOR B. REYNOLDS<sup>2</sup> AND VINCENT H. RESH<sup>3</sup>

<sup>1</sup>*Department of Fisheries and Oceans, Freshwater Institute, 501 University Crescent, Winnipeg, MB R3T 2N6, CANADA*

<sup>2</sup>*National Water Research Institute, Environment Canada, CCIW, 867 Lakeshore Road, Burlington, ON L7R 4A6, CANADA*

<sup>3</sup>*Department of Environmental Science, Policy & Management, 201 Wellman Hall, University of California, Berkeley, CA 94720, USA*

### Summary

A biomonitoring program based on the BEAST predictive model is described for the Fraser River, British Columbia, Canada. The BEAST model is a three-stage multivariate approach to establishing reference conditions. Reference sites (representative of the "best available condition") were chosen using ecoregions and stream orders within ecoregions. A small number of test sites (suspected of being impacted) were used for comparison with the reference sites when developing the model. Environmental variables representing map, site, channel and water-column scales were measured at each site. Single macroinvertebrate samples were collected from riffle habitats at each site using a kicknet sampler (400- $\mu$ m mesh) for 3 minutes. Sites in the Fraser River catchment were sampled in the autumn of 1994, 1995 and 1996. Results from the 1994 and 1995 field seasons used 127 reference sites and produced five reference groups based on macroinvertebrate community structure. Ten optimum predictor variables were identified for matching new sites to the appropriate reference group, and included map, site, channel and water-column measurements. The error rate for predicting test sites to particular reference groups in cross-validation studies was 37%. The model was capable of discriminating the most impacted of the test sites used.

### Introduction

Traditional methods of establishing control sites in field-oriented biomonitoring studies of water quality are largely limited to areas upstream of impacts in lotic waters, and areas near impacts in lentic waters. Such designs are often "confounded" (Eberhardt 1978; Hurlbert 1984). Moreover, reliance on only one or a few control sites is problematic because of limited capacity to extrapolate to other locations, limited ability to calculate variance estimates, and difficulties in addressing problems of a nonpoint-source nature (Hughes 1995).

The reference condition approach offers a powerful alternative to the above-described

traditional methods because the sites themselves serve as replicates; this is in contrast to the multiple collections within sites that are the replicates in designs using inferential statistics (Reynoldson, Norris *et al.* 1997). We define the "reference condition" as being ". . . representative of a group of minimally disturbed sites organized by selected physical, chemical and biological characteristics" (Reynoldson, Norris *et al.* 1997).

In this study, we use the BEAST model (Benthic Assessment of Sediment) (Reynoldson, Bailey *et al.* 1995; Reynoldson, Norris *et al.* 1997; and see Chapter 11), in which an array of reference sites characterizes the biological condition of a region, and a test site is compared with an appropriate subset of reference sites. The probability of a test site belonging to each of the subsets or classification groups is determined, and the test site is assigned to the highest probability group. In other multivariate models based on the reference condition approach, the probabilities are computed for a test site being a member of each classification group, and the prediction for the test site is weighted in relation to the probabilities for each classification group (RIVPACS (Wright 1995 and Chapters 1 to 3) and AUSRIVAS (Parsons & Norris 1996, and Chapters 9 to 11)). Another reference condition approach, multimetrics (e.g. Barbour *et al.* 1995; Barbour, Gerritsen *et al.* 1996) is considered in Chapters 13 and 19.

The reference condition approach is well-suited for large-scale biomonitoring programs because reference sites can be scattered throughout the catchment. Furthermore, local knowledge and expertise, published information, or simple reconnaissance trips, can be used to identify sites that represent "best available condition" for use in building the reference site model. The model can represent either a one-time investment (Rosenberg *et al.* 2000) or it can be continually improved (e.g. Wright 1995). The reference condition approach already forms the basis of large-scale biomonitoring programs in the UK (e.g. Wright 1995), Australia (e.g. Parsons & Norris 1996) and Canada (Reynoldson, Bailey *et al.* 1995). The objective of this contribution is to describe a Canadian program on the Fraser River, British Columbia (BC), one of Canada's largest rivers, and to report results from the first two years of field research. We include a description of the biotic groups formed from the reference sites, and an assessment of sites suspected of being impacted.

### The Fraser River catchment

The Fraser River catchment (Fig. 12.1) covers *ca* 230,000 km<sup>2</sup> or 25% of British Columbia (BC Ministry of Environment, Lands and Parks and Environment Canada 1993). The Fraser River mainstem (>1350 km long) has a mean annual discharge of 3620 m<sup>3</sup> s<sup>-1</sup>, which makes it the fourth largest river in Canada after the St Lawrence (10,800 m<sup>3</sup> s<sup>-1</sup>), the Mackenzie (9910 m<sup>3</sup> s<sup>-1</sup>) and the Yukon (6370 m<sup>3</sup> s<sup>-1</sup>) (Dynesius & Nilsson 1994). Two-thirds of British Columbia's population, or *ca* 1,700,000 people, live in the Fraser catchment, mostly in its southern regions.

The Fraser River mainstem is one of the last unregulated large rivers of North America. However, the catchment has a variety of land-use and management problems, which include forest harvest and pulp mills throughout the catchment, widespread agricultural activities (e.g. ranching, fruit growing and other crops), flow regulation for hydroelectric power generation on some of the tributaries, and urban development in the lower Fraser River, near Vancouver (Richardson & Healey 1996). Mining, fishing and wilderness recreation and tourism are additional pressures (Environment Canada 1995). All of these activities pose a threat to the health of the Fraser River ecosystem, a fact recognized by the establishment of the Fraser River Action Plan (FRAP) in 1991 by Environment Canada (Environment Canada 1995). The research described in this chapter was funded by the FRAP, to address the problem of pollution in the Fraser River by developing a biomonitoring program for water quality assessment. The

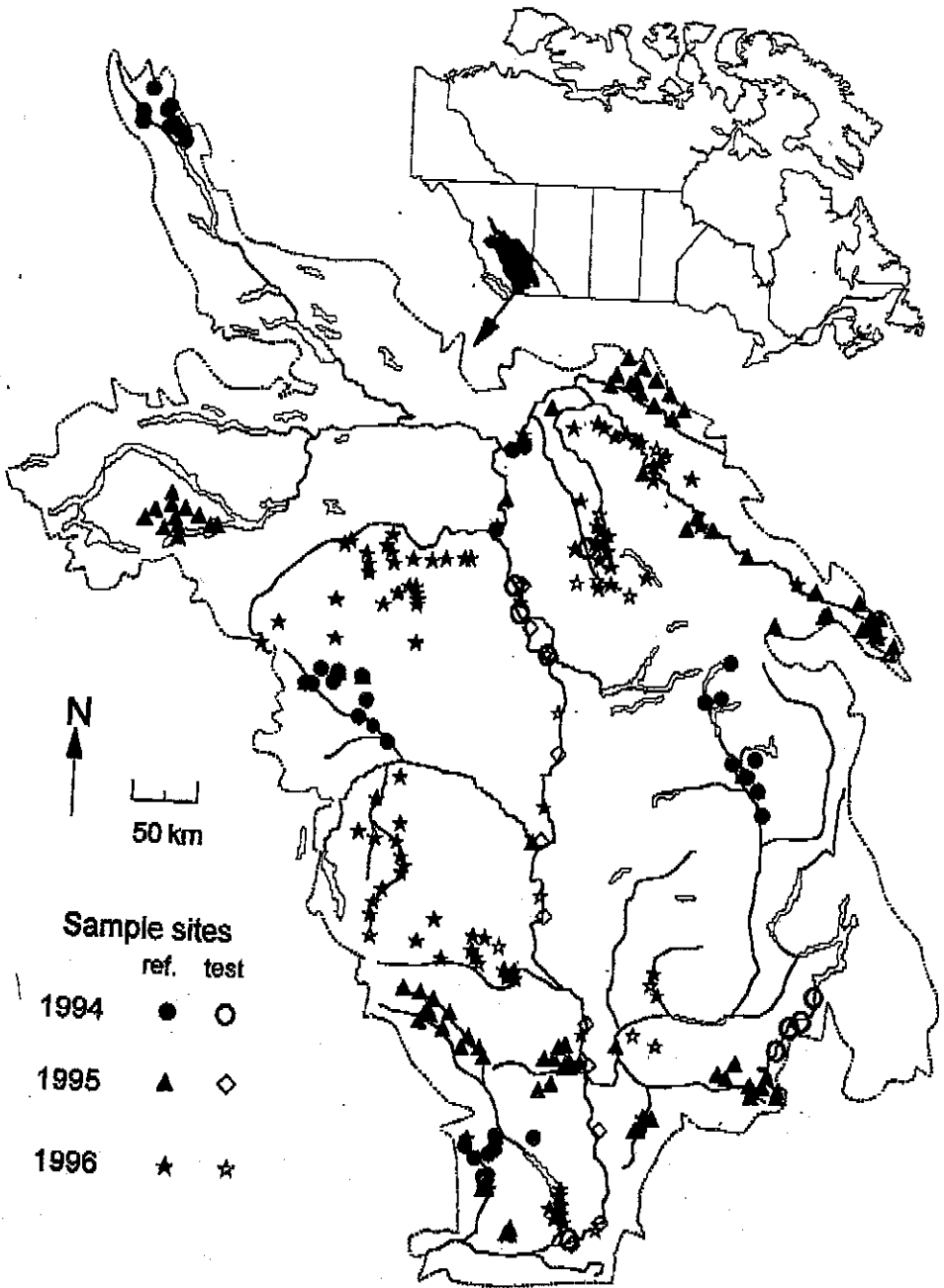


Figure 12.1. Location of the Fraser River catchment in British Columbia, Canada, showing the mainstem river, major tributaries and sampling sites (1994-1996). ref. = reference site.

Fraser River is the first Canadian river to which the reference condition approach, using benthic macroinvertebrates, has been applied.

## Methods

### *Design of the program*

The multivariate methods used to match test sites to a subset of reference sites require a substantial initial investment in sufficient sites to characterize both the different conditions in a catchment and the variability associated with these conditions. Generally, *ca* 250 sites are required to characterize variability adequately, and to build appropriate predictive models (e.g. see Reynoldson, Bailey *et al.* 1995; Wright 1995). A small set of additional (test) sites, which may or may not be impacted, is included to verify performance of the reference site model during its development. A minimum ratio of five reference sites to one test site is recommended because the BEAST combines test and reference sites in a new ordination matrix, so the test sites will affect the distribution of reference sites in ordination space (see below; Reynoldson, Norris *et al.* 1997).

Two spatial scales, ecoregion and stream order within an ecoregion, were used to maximize the diversity of sites chosen. The ecoregion scale ensured the inclusion of different climatic and landscape conditions, whereas the stream-order scale ensured that a range of hydraulic conditions was included (Rosenberg *et al.* 2000). A series of workshops with local experts served to identify subcatchments that were unimpacted and those that had different degrees of impact. Reference and test sites were randomly selected from the appropriate subcatchments (Rosenberg *et al.* 2000).

Two temporal scales, annual and seasonal, were also examined. Most of the reference sites were visited only once, in the autumn, because of the large geographic scale of the study. The autumn period was chosen because it is a low-water period in the long-term hydrograph of the Fraser River, which is important in terms of accessibility to streams in the catchment. Thus, our intent was to examine the effects of annual and seasonal variability on the accuracy of the predictive model. Annual variation was examined by re-sampling *ca* 10% of sites visited initially, in the second and third years of the study. Seasonal variation was examined by sampling six sites monthly over a 2-year period (Dymond 1998). Preliminary results of the seasonal study indicated unpredictable effects of season on the predictive models, so Dymond (1998) recommended that sampling should be restricted to the autumn or done over multiple seasons. However, the overall results of temporal investigations are not available at this time and will be reported elsewhere.

### *Site selection*

Thirty-nine subcatchments (including the Fraser River mainstem) were sampled over the 3-year field study (Rosenberg *et al.* 2000). Fifteen of these were reference (unimpacted) subcatchments. The Fraser mainstem flows through seven ecoregions; tributaries individually flow through a maximum of three ecoregions. A maximum of seven stream orders was sampled in a subcatchment.

A total of 266 sites (233 reference, 33 test) was visited over three years (Fig. 12.1): 46 in 1994 (37 reference, 9 test), 99 in 1995 (90 reference, 9 test) and 121 in 1996 (106 reference, 15 test). In addition, nine sites were re-sampled in 1995 and 1996 to measure annual variability. Thus, the sampling program approximated the target level of 250 reference sites, and it achieved the guideline of a 5:1 ratio of reference to test sites (the actual ratio was *ca* 7:1).

Test sites came from the Fraser mainstem (pulp and paper mill impacts), the Salmon River

(agricultural impacts), and the Willow River (logging impacts). For the Fraser mainstem sites, FRA13 is ca 50 km downstream of Prince George and its pulp mills; FRA14 is ca 100 km downstream of the pulp mill at Quesnel; FRA16 is ca 20 km upstream of Quesnel; FRA28 is ca 400 km further downstream, at Hope, and away from pulp mill influences. For the Salmon River sites, SAL1 is upstream of most agricultural activity, which progressively increases downstream from SAL2 to SAL4. Three replicated samples taken at SAL3, as part of an earlier sampling design, were included. A single test site came from the logged-over Willow River (WIL1).

#### Environmental variables measured

Environmental variables are used to relate habitat conditions to subsets of sites selected, based on similarities in macroinvertebrate communities, and to build the predictive model for matching new sites to the appropriate subset of reference sites (see below). An optimum set of predictor variables cannot be determined *a priori*, so a maximum number of likely variables had to be chosen beforehand. Those variables came from previously published multivariable studies that examined the relationship between environmental characteristics and community structure of lotic benthic macroinvertebrates. The list of variables was discussed, amended and supplemented at the initial Fraser River workshop (Rosenberg *et al.* 2000). The final list (Table 12.1) was measured at all sites, using methods described in Rosenberg *et al.* (2000).

Variability of environmental measurements was assessed by sampling 26 (ca 10%) of the sites in triplicate. Coefficients of variation (CV) were calculated for each site, and overall mean CVs were calculated for each variable measured.

Table 12.1. The environmental variables measured in the Fraser River biomonitoring program, listed under various scales (map, site, channel and water).

(1) Diameter (cm) of the dominant substratum: one of seven categories at each site (>0.1–0.2; 0.2–0.5; 0.5–2.5; 2.5–5.0; 5.0–10.0; 10.0–25.0; >25 cm), averaged for all sites in a group.

(2) Diameter of the next dominant substratum (see note (1) above).

(3) Degree of exposure of the dominant substratum: one of five categories ranging from completely embedded (score = 1) to unembedded (score = 5) at each site, averaged for all sites in a group.

| Map          | Site                                     | Channel                          | Water                                 |
|--------------|--|----------------------------------|---------------------------------------|
| Latitude     | Date of sampling                         | Wetted width                     | pH                                    |
| Longitude    | Flow state                               | Mean channel depth               | Dissolved oxygen                      |
| Altitude     | Macrophyte cover                         | Maximum channel depth            | Conductivity                          |
| Ecoregion    | Riparian vegetation (%)                  | Bankfull width                   | Temperature                           |
| Stream order | [grasses, shrubs, deciduous, coniferous] | Slope                            | Total phosphorus                      |
|              | Canopy cover                             | Water velocity [mean, max.]      | Nitrate-nitrite and Kjeldahl nitrogen |
|              | Extent of logging in riparian zone       | Framework <sup>(1)</sup>         | Alkalinity                            |
|              |  | Matrix <sup>(2)</sup>            | Total suspended solids                |
|              |  | Interstitial material (%)        |                                       |
|              |  | [silt/clay, sand, gravel]        |                                       |
|              |  | Embeddedness <sup>(3)</sup>      |                                       |
|              |  | Periphyton biomass               |                                       |
|              |  | Periphyton chlorophyll- <i>a</i> |                                       |

*Community structure of benthic macroinvertebrates*

Benthic macroinvertebrates were collected from the riffle areas of small streams and the cobble shoulders of large streams/ivers using a triangular kicknet (38.5 cm on each side), as described by Rosenberg, Davies *et al.* (1977). The kicknet was inexpensive to construct, durable and easy to transport and use. A series of calibration studies revealed that a single sample per site, collected for 3 minutes using a 400  $\mu\text{m}$  mesh, provided an optimal combination of taxon recovery and cost effectiveness (Rosenberg *et al.* 2000).

Samples were subsampled using a Marchant box (Marchant 1989), and the first 200 specimens, located at random in the box, were counted (see Rosenberg *et al.* 2000 for details). Barbour & Gerritsen (1996) reported that fixed-count subsampling discriminated better than other subsampling methods, including fractions, in a study of Florida lakes. A subsample containing 200 organisms required about 4 hours to sort and identify, which represented *ca* 1/30 of the time required to process three full replicates from each site.

The eventual need for non-specialists to identify benthic macroinvertebrates is a special concern of the Fraser River study. Only results from the family level are reported here, but a final calibration step will involve a comparison of the efficacy of identification at family level and lower taxa (mostly genus and species) for establishment of site groups and in model development.

*Reference condition statistics*

A three-stage procedure (Reynoldson, Bailey *et al.* 1995) was used to analyze the Fraser River data.

*Stage 1.* Pattern recognition techniques (cluster analysis and ordination) were used to describe the biological structure of reference site data. The Bray-Curtis association measure was used to describe the species matrix because it performs well under a variety of conditions (Faith *et al.* 1987). Clustering was done using an agglomerative hierarchical fusion method with unweighted pair-group mean arithmetic averaging (UPGMA). Groups were selected by examining group structure and spatial location of the groups in ordination space. Ordination (semi-strong hybrid multidimensional scaling; Belbin 1991) was used to reduce the variables required to identify structure of the data. All clustering and ordination was done using PATN (Belbin 1992).

*Stage 2.* The observed biological structure was related to environmental variables. Variables that were most likely to be influenced by anthropogenic activity (e.g. total phosphorus and chlorophyll-*a*) were excluded from the analysis. The relationship was examined in three ways.

(a). Principal axis correlation (PAC) in PATN – a multiple-linear regression method that determines how well a set of attributes (environmental variables) can be fitted to an ordination matrix space (the species). The method determines the orientation of the best-fit vectors for each environmental variable in ordination space. These vectors can be represented as axes on an ordination plot, and correlation values of axes with the ordination are provided. Monte Carlo simulations can be used to establish the statistical significance of the correlations.

(b). ANOVA in SAS – to identify the environmental variables that differed significantly among biological groups.

(c). Stepwise discriminant analysis (STEPDIS in SAS) – to identify the variables that best described the biological groups (and to minimize effects of auto-correlation among those variables).

Key environmental variables, identified by the above three analyses, were used in multiple

discriminant analysis (MDA in SAS) to relate the biological groups to environmental variables. Untransformed environmental data were used to generate discriminant scores and to predict the probability of group membership of individual sites. Cross-validation was used to verify the accuracy of predictions from the discriminant model. In this method, each of the sites is in turn removed from the dataset, a model is generated without that site and then the group to which the site belongs is allocated probabilistically. Allocated and actual groups are then compared to provide group and total error rates based on the percentage of sites allocated correctly.

The optimal set of predictor variables was chosen by iteration. Various combinations of predictor variables were selected from the stepwise discriminant analyses and principal axis correlations. The set of variables having the lowest error rate was regarded as optimal.

*Stage 3.* The final stage is devoted to assessing the biological condition (i.e. equivalent to reference or impacted) of new sites tested. The first step is to assign the new site to one of the reference groups by comparing community composition of the benthic macroinvertebrates. Values of the optimal predictor variables are used to assign the new site to one of the reference groups, using a multiple discriminant analysis (PROC DISCRIM in SAS). Reference sites are then combined with the new site to create a new data-matrix, which is ordinated using semi-strong hybrid multidimensional scaling (see Stage 1). Test sites having a similar probability of belonging to more than one group are compared to each of those groups; the current protocol involves sites having a probability of membership >30%. All sites are re-plotted in ordination space, and probability ellipses are constructed around the reference sites (SCATTER-PLOT in SYSTAT). A new site is considered to be equivalent to reference if it is located inside the 90% probability ellipse, and impacted if it falls outside the 90% probability ellipse. Other probability ellipses (i.e. 99% and 99.9%) can be used to measure the degree to which impacted sites differ from the reference condition (e.g. see Chapter 11).

## Results and discussion

### *Variability of environmental measurements*

Mean coefficients of variation for nine variables concerned with channel characteristics, and eight variables connected with the water column, revealed a range from <1% to 87%. Measures of substratum composition (CV = 10 to 87%) and biomass of benthic algae (CV ca 35%) were the most variable in the channel group, whereas water variables (e.g. pH, nutrients, total suspended solids) had CVs <25%. We were unable to separate the relative contributions of field and laboratory procedures on the variables measured. In the end, values of the available CVs for variables used as optimal predictors at the family level ranged from 5 to 87%.

### *Model building at the family level*

The three-stage process of model building using the BEAST, as described above, was applied to the 127 reference sites sampled in the 1994 and 1995 field seasons. A further 106 reference sites from 1996 remain to be completed. The model derived from the complete dataset will be reported in subsequent publications.

A total of 131 families was identified but only 39 families, each one of which accounted for  $\geq 0.5\%$  of total abundance, were used in the analysis. These 39 families included 99.7% of the organisms collected. Taxa having abundances <0.5% were deleted, because the presence of large numbers of rare taxa adds unwanted noise to the classification analysis (Reynoldson, Bailey *et al.* 1995).

Thirteen families represented more than 90% of total numbers; predominantly these were the



Figure 12.2 (on facing page). The most abundant families and most frequently occurring families of macroinvertebrates in the Fraser River study during 1994 and 1995.

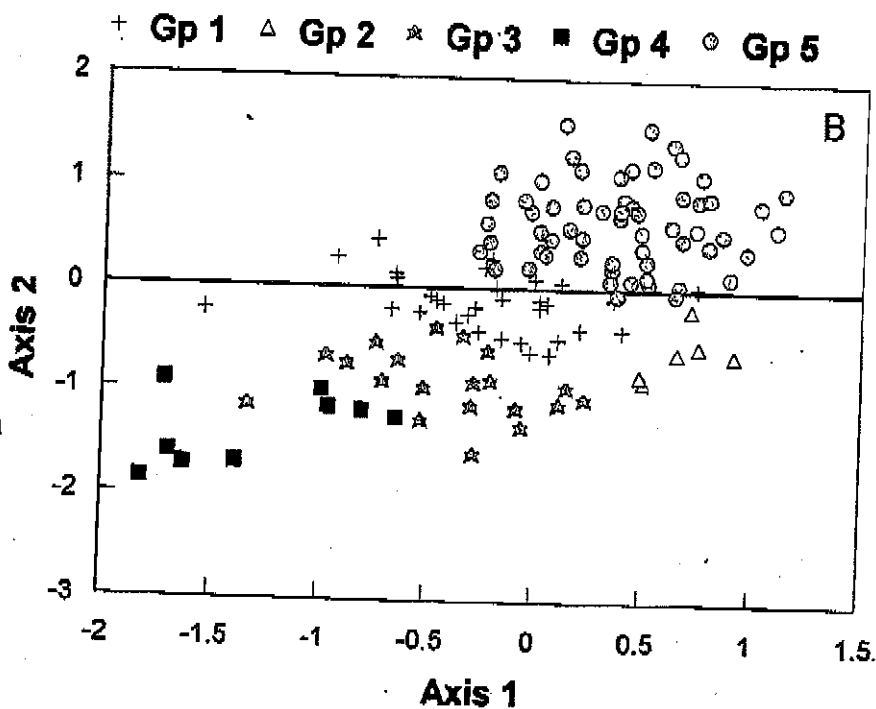
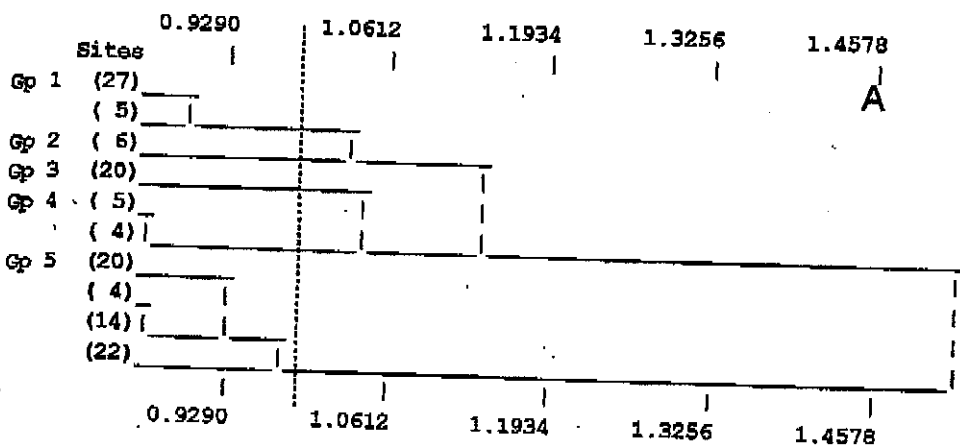


Figure 12.3. Family-level classification of 127 reference sites on the Fraser River, British Columbia. Data are from the 1994 and 1995 field seasons. Gp = Group. (A) Cluster diagram; vertical dotted line indicates the cut-off point for group formation. (B) Ordination of five macroinvertebrate groups; stress level = 0.1728.

Chironomidae and three families of Ephemeroptera (Fig. 12.2). Twelve families occurred at more than 50% of all the sites, once again led by the Chironomidae and the three families of Ephemeroptera.

*Stage 1.* We identified five groups of sites formed by the macroinvertebrate families from the cluster analysis and ordination (Fig. 12.3). Figure 12.3B presents the ordination of taxa data only; the contribution of individual taxa to the ordination will be reported elsewhere. The distribution of sites along axis 1 (Fig. 12.3B) suggests a strong gradient, although we have not examined the nature of that gradient. Twenty-eight of the 39 common families contributed significantly ( $p < 0.01$ ) to the ordination axes, according to the PAC. Each of the five site groups had a characteristic family composition, which is shown in Table 12.2 for the ten families contributing most to the structure observed in the ordination analyses. No geographic pattern was obvious (Fig. 12.4), which suggests that geographic distribution at the family level is not an important consideration in the Fraser River catchment.

Table 12.2. Relative occurrence at sites (%) and relative abundance of families (%) best correlated with five groups of sites (Gp 1 to Gp 5) formed from 127 reference sites in the Fraser River catchment. The mean numbers of individuals (per family) collected in each group of sites are given, and the total number of sites in each group is also shown.

PAC  $r$  = principal axis correlation coefficient.

| Families            | PAC<br>$r$ | Occurrence        |           | Mean numbers per kicknet sample in each group |      |      |      |      |
|---------------------|------------|-------------------|-----------|---|------|------|------|------|
|                     |            | Total<br>at sites | abundance | Gp 1  | Gp 2 | Gp 3 | Gp 4 | Gp 5 |
| Chironomidae        | 0.615      | 98.4              | 31.4      | 326   | 110  | 48   | 26   | 3692 |
| Heptageniidae       | 0.543      | 92.9              | 16.9      | 342   | 444  | 63   | 8    | 1843 |
| Bactidae            | 0.503      | 87.4              | 13.2      | 172   | 71   | 129  | 6    | 1492 |
| Nemouridae          | 0.487      | 74.8              | 5.0       | 96  | 105  | 7    | 10   | 367  |
| Leptophlebiidae     | 0.461      | 33.9              | 1.7       | 16  | 0    | 0    | 0    | 202  |
| Taeniopterygidae    | 0.461      | 53.5              | 3.9       | 73  | 1848 | 146  | 15   | 213  |
| Perlodidae          | 0.450      | 66.1              | 0.8       | 22  | 49   | 14   | 23   | 73   |
| Empididae           | 0.426      | 55.1              | 0.5       | 11  | 6    | 2    | 2    | 57   |
| Chloroperlidae      | 0.425      | 75.6              | 1.6       | 72  | 99   | 24   | 2    | 144  |
| Tipulidae           | 0.403      | 55.9              | 0.7       | 14  | 11   | 3    | 2    | 73   |
| Total no. of sites: | —          | 127               | 127       | 32  | 6    | 20   | 9    | 60   |

*Stage 2.* PAC revealed ten variables that were important to predicting macroinvertebrate community structure: (1) two map variables – altitude and longitude; (2) two site variables – maximum water velocity and percent grasses; (3) five channel variables – mean channel depth, percent gravel, percent sand, percent silt/clay and framework (dominant substratum diameter); (4) one water variable – alkalinity.

Table 12.3 shows the mean values of these optimal predictor variables for each reference site group, as well as the PAC value between the variables and the family ordination matrix. Overall similarity of variables in Table 12.3 may reflect the use of riffles as a standardized habitat for sampling. Cross-validation procedures to test the accuracy of predictions from the discriminant model, using the ten predictor variables, resulted in an overall error rate of 37% (Table 12.4).

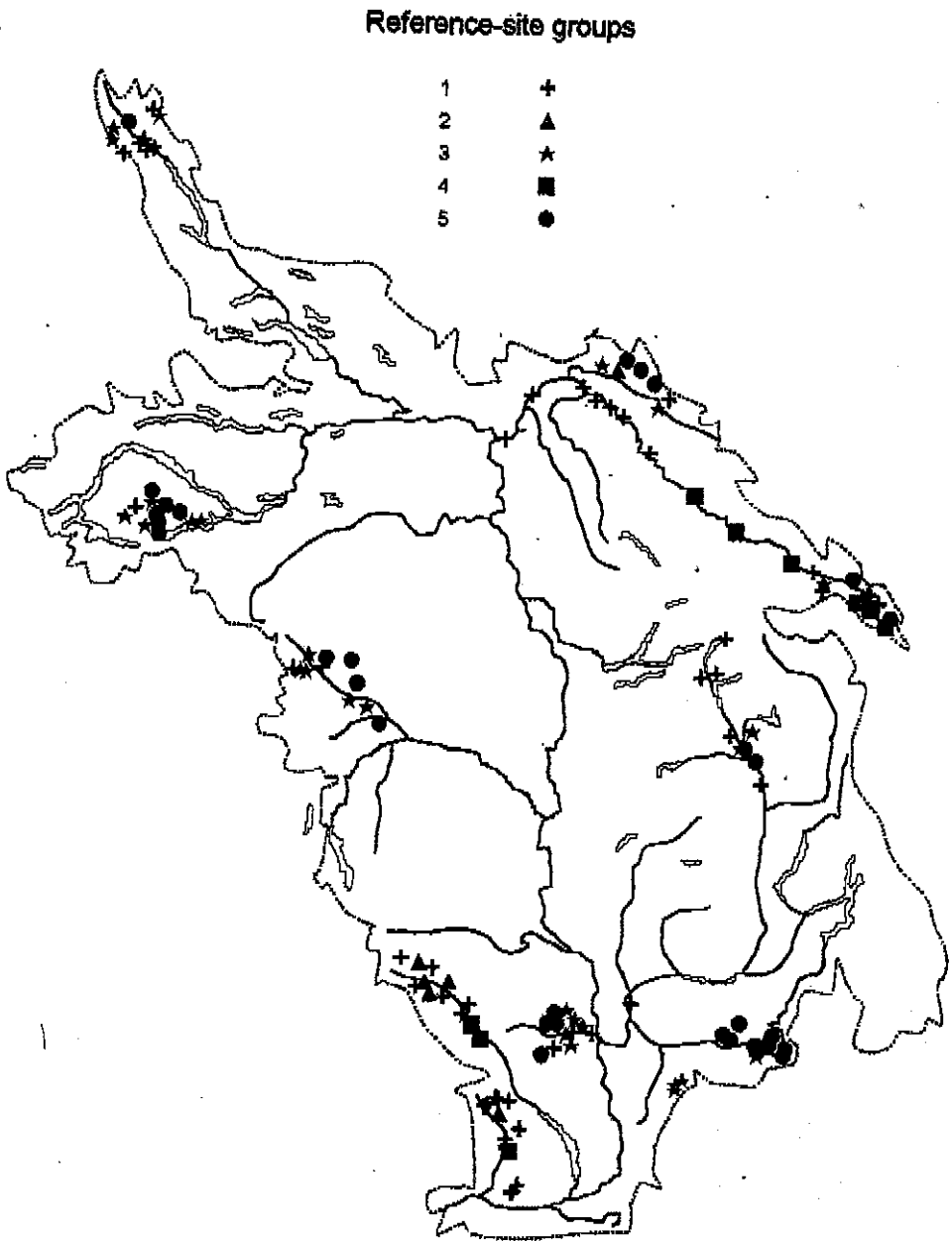


Figure 12.4. Geographic distribution of each of the five reference groups of macroinvertebrates throughout the Fraser River catchment.

Table 12.3. Mean values of predictor variables selected by stepwise discriminant analysis in five reference groups (Gp), and principal axis correlation (FAC r), with the ordination matrix rank given in parentheses.

(1) Presence or absence at each site, averaged for all sites in a group.

(2) Diameter (cm) of the dominant substratum (see note (1) in Table 12.1 for explanation).

| Predictor variables                    | FAC r      | Gp 1   | Gp 2   | Gp 3   | Gp 4   | Gp 5   |
|--|------------|--------|--------|--------|--------|--------|
| Altitude (ft amsl)                     | 0.343 (9)  | 3113   | 2688   | 2234   | 2766   | 3784   |
| Longitude ( $^{\circ}$ decimal mins)   | 0.193 (21) | 123.00 | 122.88 | 121.50 | 120.50 | 122.60 |
| Grasses (% occurrence) <sup>(1)</sup>  | 0.476 (1)  | 22     | 17     | 5      | 0      | 52     |
| Framework (cm) <sup>(2)</sup>          | 0.353 (7)  | 6.1    | 7.2    | 7.2    | 5.3    | 5.8    |
| Interstitial material (%)              |            |        |        |        |        |        |
| Silt/clay                              | 0.327 (12) | 0.17   | 4.24   | 1.32   | 1.96   | 0.26   |
| Sand                                   | 0.419 (3)  | 71.4   | 84.2   | 75.6   | 86.3   | 63.4   |
| Gravel                                 | 0.432 (2)  | 27.3   | 11.5   | 19.2   | 13.1   | 35.3   |
| Mean channel depth (cm)                | 0.409 (4)  | 31.0   | 22.3   | 35.3   | 25.0   | 20.4   |
| Maximum velocity (cm s <sup>-1</sup> ) | 0.398 (5)  | 0.58   | 0.52   | 0.74   | 0.66   | 0.51   |
| Alkalinity (mg l <sup>-1</sup> )       | 0.349 (8)  | 37.4   | 23.9   | 24.4   | 23.5   | 50.0   |

Table 12.4. Prediction of family groups (Gp) by discriminant analysis. Overall error rate = 37%.

| Error rate | To Gp 1  | To Gp 2 | To Gp 3  | To Gp 4 | To Gp 5  |
|------------|----------|---------|----------|---------|----------|
| From Gp 1  | 14 (44%) | 0       | 6        | 2       | 10       |
| From Gp 2  | 1        | 2 (33%) | 1        | 1       | 1        |
| From Gp 3  | 2        | 0       | 14 (70%) | 3       | 1        |
| From Gp 4  | 0        | 1       | 0        | 8 (89%) | 0        |
| From Gp 5  | 5        | 2       | 7        | 4       | 42 (70%) |

Stage 3. The last stage assesses test sites, suspected of being impacted, against the reference site groups, and requires reordination of the appropriate reference site group with the test sites. All of the test sites were predicted to occur in either reference Group 1 or reference Group 5 (Fig. 12.5). The Fraser mainstem sites were predicted to Group 1; FRA16 had a similar probability of belonging to either group, so it was included in Group 1. FRA13, FRA14 and FRA16 fell well outside the 90% probability ellipse, indicating probable pulp and paper mill impacts at these sites; FRA13 and FRA16 were more affected than FRA14. The downstream FRA28 site seems to be unimpacted.

The remaining test sites were predicted to occur in Group 5 (Fig. 12.5). The WIL1 site appears to have been marginally affected by past logging activities in its catchment. For the Salmon River sites, SAL1 and SAL2 are equivalent to reference. The evidence for agricultural impacts at SAL3 and SAL4 is equivocal because of the proximity of SAL4 and two of the three SAL3 replicates (SAL3.1, SAL3.2) to the 90% probability ellipse. In addition, SAL3.3 is equivalent to reference.

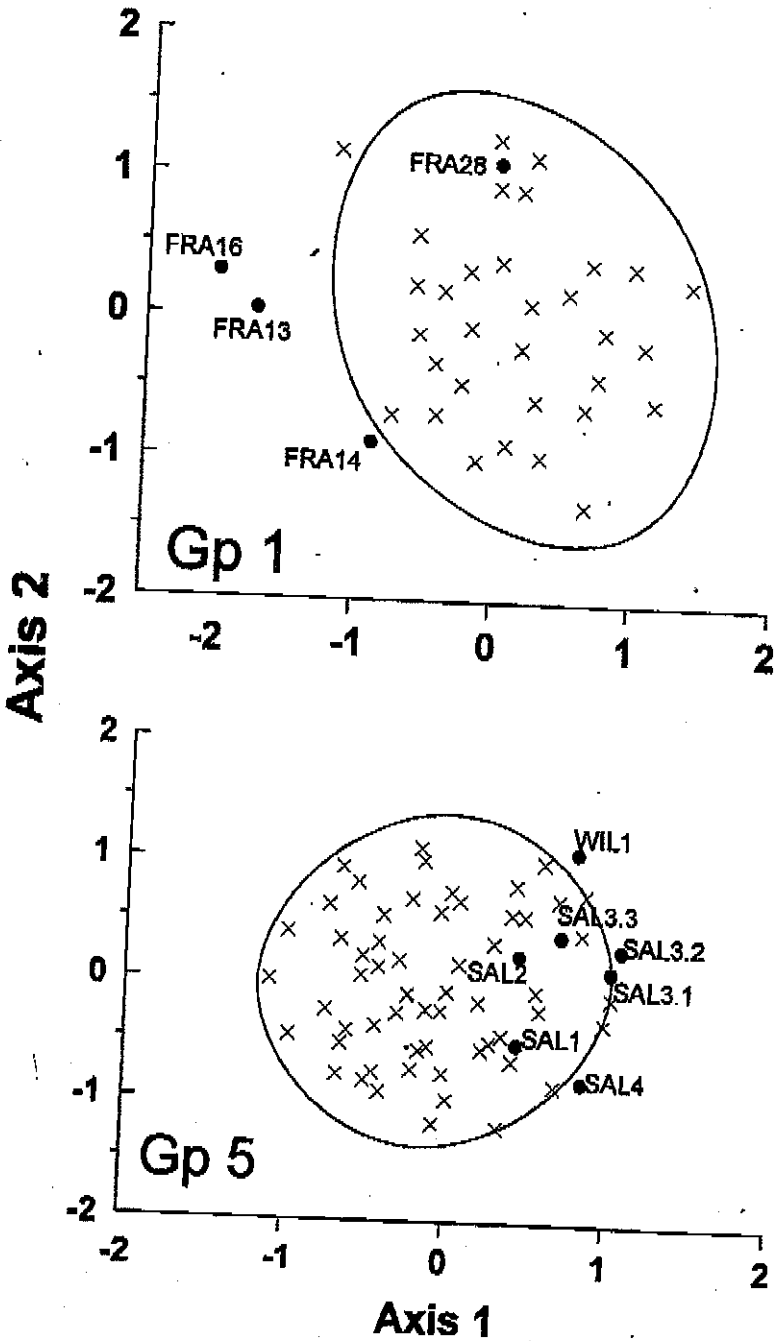


Figure 12.5. Assessment of macroinvertebrate community structure at test sites, using the BEAST model and the 1994 and 1995 data from the Fraser River, British Columbia. Groups (Gp) 1 and 5 are shown along with their 90% probability ellipses. Crosses = reference sites. Solid circles = test sites (FRA = Fraser River mainstem; SAL = Salmon River; WIL = Willow River). Three dimensions were used for Gp 1, stress level = 0.2334. Three dimensions were used for Gp 5, stress level = 0.1850; only the first two axes are shown for Gp 5.

### Conclusion

We have developed a family-level predictive model that uses five reference groups based on macroinvertebrate community structure, and requires ten predictor variables. Our error rate is 37% for predicting reference sites to a group.

Preliminary assessment of test sites using the BEAST model indicates that a variety of environmental perturbations can be evaluated. Eventual use of the full dataset should provide a valid method for setting numerical criteria for decision-making over large geographic scales. The resulting reference database can be used by itself or added to, over time, as funding becomes available. The database, models, and eventual software development, will enable routine assessment of sites suspected of being impacted in the Fraser River catchment.

### Acknowledgements

We thank J. Hilton and J. Wright for making possible the preparation of this paper. D. Cobb and R. Stewart reviewed earlier drafts of the manuscript. Environment Canada's Fraser River Action Plan provided funding for this study. We also thank T. Tuominen and field and laboratory personnel for their support throughout the project.