Development of a Northern Ontario Benthic Invertebrate Reference Condition Approach (RCA) Biomonitoring Network to Meet Metal Mining Environmental Effects Monitoring Requirements

Phase One

Unnamed Creek on South Regan Road in Hemlo

Cooperative Freshwater Ecology Unit

2005

Development of a Northern Ontario Benthic Invertebrate Reference Condition Approach (RCA) Biomonitoring Network to Meet Metal Mining Environmental Effects Monitoring Requirements

Phase One

Prepared by:

T.B. Reynoldson, C.I. Brereton, W. Keller, and C. Sarrazin-Delay

With the participation of:

N. Ali, M. Bowman, A. Faubert, D. Gelderland, C. Logan, T. Pascoe, W. Plant, W. Sencza, K. Somers, G. Watson, R. Wepruk, W. Wisniewski

May 2005

Cooperative Freshwater Ecology Unit
Department of Biology, Laurentian University
Sudbury, Ontario, P3E 2C6, Canada

List of participants

Nardia Ali
Environment Canada
Toxics Prevention Division
Environmental Protection Branch - Ontario Region
4905 Dufferin Street, Downsview, Ontario M3H 5T4
Tel: 416-739-5884; Fax: 416-739-4342
Email: nardia.ali@ec.gc.ca

Christine Brereton
Inco Limited, Environment Health and Safety
Copper Cliff, Ontario, P0M 1N0
Tel: 705-682-6182
Email: cbrereton@inco.com
Formerly with the Cooperative Freshwater Ecology Unit

David Gelderland
Placer Dome (CLA) Limited - Campbell Mine
Box 10, Mine Road, Balmertown, ON, P0V 1C0.
Tel: 807-735-2075 ext.230; Fax: 807-735-5012
Email: david_gelderland@placerdome.com

Craig Logan
Environment Canada
Aquatic Ecosystem Management Research Branch
National Water Research Institute
Canada Centre for Inland Waters
867 Lakeshore Road, P.O. Box 5050
Burlington, Ontario, L7R 4A6
Tel: 905-336-6418 or 905-336-6416; Fax: 905-336-6430
Email: craig.logan@cciw.ca

Wesley Plant
Environment Canada
Toxics Prevention Division
Environmental Protection Branch - Ontario Region
4905 Dufferin Street, Downsview, Ontario M3H 5T4
Tel: 416-739-4272; Fax: 416-739-4342
Email: wesley.plant@ec.gc.ca

Chantal Sarrazin-Delay
Cooperative Freshwater Ecology Unit
Department of Biology, Laurentian University
935 Ramsey Lake Road, Sudbury, ON P3E 2C6
Tel: 705-675-1151 ext. 4451; Fax: 705-671-3857
Email: csarrazindelay@laurentian.ca

Keith Somers
Ontario Ministry of the Environment,
Environmental Monitoring and Reporting Branch
Dorset Environmental Science Centre
Bellwood Acres Road, P.O. Box 39, Dorset, Ontario, P0A 1E0
Tel: 705-766-2408; Fax: 705-766-2254
Email: keith.somers@ene.gov.on.ca

Randy Wepruk
Goldcorp Inc. Red Lake Mine
Mine Road, General Delivery
Balmertown, Ontario, P0V 1C0
Tel: 807-735-2077 ext. 208; Fax: 807-735-3348
Email: rwepruk@goldcorp.com

Michelle Bowman
University of Toronto / Ontario Ministry of Environment
Dorset Environmental Science Centre
Bellwood Acres Road, P.O. Box 39, Dorset, Ontario, P0A 1E0
Tel. 705-766-1790; FAX: (705) 766-2254
Email: michelle.bowman@ene.gov.on.ca

Adele Faubert
Williams Operating Corporation
P.O. Bag 500
Marathon, ON, P0T 2E0
Tel: 807-238-1100 ext. 4264; Fax: 807-238-1050
Email: afaubert@hemlomines.com

Bill Keller
Ontario Ministry of the Environment
Cooperative Freshwater Ecology Unit
Department of Biology, Laurentian University
Ramsey Lake Road, Sudbury, Ontario, P3E 2C6
Tel: 705-671-3858; Fax: 705-671-3857
Email: bkeller@vianet.ca

Timothy Pascoe
Environment Canada
Canada Centre for Inland Waters
867 Lakeshore Road
Burlington, Ontario
L7R 4A6
Tel. 905-336-6239; Fax: 905-336-4699
Email: tim.pascoe@ec.gc.ca

Trefor B. Reynoldson
Acadia Centre for Estuarine Research
Box 115 Acadia University
Wolfville, Nova Scotia, B4P 2R6
Tel: 902-585-1638; Fax: 902-585-1054
Email: trefor.reynoldson@acadiau.ca

Walter Sencza
Newmont Canada Limited
Yellow Brick Road, Junction Hwy 17 & Hwy 614
P.O. Box 40, Marathon, Ontario, P0T 2E0
Tel: 807-238-2108; Fax: 807-238-2044
Email: Walter.Sencza@Newmont.com

Glen Watson
Inco Limited, Environment Health and Safety
Copper Cliff, Ontario, P0M 1N0
Tel: 705-682-8231; Fax:705-682-6976
Email: gwatson@inco.com

Wendy Wisniewski
Inco Limited, Environment Health and Safety
Copper Cliff, Ontario, P0M 1N0
Tel: 705-682-8247; Fax:705-682-6976
Email: wwisniewski@inco.com

Summary

1. The Northern Ontario Benthic Invertebrate Reference Condition Approach (RCA) Biomonitoring Network was created with the participation of Environment Canada, Ontario Ministry of the Environment, INCO Ltd., Williams Operating Corporation, Goldcorp Inc., Placer Dome Inc., Newmont Canada Inc., Laurentian University and Acadia University. It was designed to assist the metal mining industry in locating suitable reference sites to meet the Environmental Effects Monitoring (EEM) requirements of the federal Metal Mining Effluent Regulation (MMER).
2. In the fall of 2003, 214 sites were sampled across four mining areas in northern Ontario to represent different geographic regions and habitat types. Sites included 58 reference streams, 19 urban streams, 6 historically impacted streams, 49 large lakes, 70 small lakes, 8 historically impacted small lakes, and 4 historically impacted large lakes.
3. Stream sampling was done using Environment Canada's CABIN protocol while lake sampling followed the Ontario Ministry of the Environment's protocol with a few modifications to suit EEM. Aquatic invertebrates were sampled using the travelling kick-and-sweep method.
4. The physical, chemical and biological data, collected from lake and stream sites, were used to create the various models that will help to select reference sites for a given test site. One model was created using stream data only, one used lake data only, one combined lake and stream data, and one global model used data from a Moose River study as well as all the data from the EEM RCA study.
5. The stream model was based on a 3-groups classification and resulted in 70% accuracy; a similar model for 3 groups of lakes performed less well, correctly predicting only 49% of sites. The most likely explanation for the poor performance of the lake models was lower variability in the lake assemblages.
6. Examination of the invertebrate assemblages in streams and lakes showed a gradient in communities based on lake and stream habitat, but also a considerable similarity in the assemblages at many lake and stream sites.
7. Due to the similarity in the invertebrate assemblages in lakes and streams, a single model was constructed that used 149 lake and stream reference sites; 11 habitat predictors resulted in 80% correct assignment of sites into 4 reference-site groups.

8. Assemblages from 76 Moose River reference sites sampled in 1999, were sufficiently similar to those in the EEM RCA study to justify a composite northern Ontario model.
9. Classification of 225 reference sites resulted in 5 groups based on invertebrate assemblages; however, one group included only two sites and was not used in subsequent model building. The 4-group predictive model correctly predicted group membership for 79% of sites using 11 habitat variables.
10. The RIVPACS and BEAST methods of test site assessment have high concordance in describing the condition of communities, and were able to differentiate between different degrees of disturbance. Both methods suggested that urban reference sites are generally in reference condition, and can also distinguish the response of communities to urban and mine associated disturbance. RIVPACS tends to be more sensitive to taxa loss and BEAST to changes in abundance. Both methods should be used in assessments of test sites.
11. Based on repeated sampling of reference sites, the two test site assessment methods have high accuracy (91.7%) in describing the degree of disturbance.
12. To address continuing questions pertaining to the differences between urban and pristine reference sites, whether or not substantial temporal changes occur, and to fill gaps outlined during the initial model development, ten streams (8 reference and 2 historically impacted) and six lakes were re-sampled during the fall of 2004; an additional 53 new sites were sampled from the Sudbury and Hemlo areas.
13. Predictive models will be refined using the samples collected from the fall of 2004. Following sorting, taxonomic identification and appropriate QA/QC, the second round of model development is expected to commence in the spring 2005. With the increase in sample size and additional habitat variables, the development of lake and stream specific models will be explored.
14. Our partners will be able to access RCA data via internet through the use of the CABIN website (http://cabin.cciw.ca). A web portal has been developed to provide access to standard sampling protocols, data entry and management, analytical and reporting tools.

Table of contents

List of Figures

List of Tables

List of Appendices

Acknowledgements

Funding and support for this project were provided by Environment Canada, Ontario Ministry of the Environment, the National Water Research Institute (Environment Canada), Laurentian University, Acadia University, INCO Ltd., Williams Operating Corporation, Goldcorp Inc., Placer Dome Inc., Newmont Canada Inc., FedNor, and Human Resources Development Canada.

Patrick Barnes, Nicole Breau, Christine Brereton, Kim Carmichael, Jen Davidson, Kim Fram, Robert Godwin, Emily Lemieux, Chantal Sarrazin-Delay, and Dan Tremblay collected and processed samples. Taxonomic identification was done by Lynne Witty, Chantal Sarrazin-Delay and Kim Carmichael. Craig Logan (Environment Canada's National Water Research Institute) verified benthos identification.

Water samples were analyzed at the Elliot Lake Research Field Station. Tim Pascoe (Environment Canada) managed the integration of study data into the BIRC database. Model development was performed by Trefor Reynoldson at Acadia University.

Special thanks to the staff of the Cooperative Freshwater Ecology Unit for administrative and moral support.

1.0 INTRODUCTION

The "reference condition" is defined as "the condition that is representative of a group of minimally disturbed sites organized by selected physical, chemical, and biological characteristics" (Reynoldson et al. 1995; Reynoldson et al. 1997; Bailey et al. 2004). The Reference Condition Approach (RCA) to bioassessment is based on the premise that when a site is to be assessed it is compared to minimally impacted reference sites with similar habitat characteristics. The determination of an "effect" at the test site is based on the differences observed in the benthic community. The magnitude and direction of the divergence from reference condition is the measure of effect.

There are two attributes of the RCA that differ from more traditional bioassessment study designs. First, the site matching process is unique in RCA as it is done quantitatively and probabilistically from a large data set of reference sites. The test site is matched to the reference sites by using a set of habitat variables. The habitat attributes used to match test to reference sites are selected through their importance in structuring the biological community, and their resistance to anthropogenic disturbance. The selection of appropriate habitat attributes is done through establishing the relationship in undisturbed (reference) sites between the biota and the habitat. This relationship allows one to build a model that predicts a test site to its appropriate reference-site group (Benthic Assessment of SedimenT (BEAST)) (Reynoldson et al. 2000). The reference sites in the appropriate group are then used to generate a list of taxa expected to be present at the test site if it is undisturbed (in reference condition) (River InVertebrate Prediction and Classification System (RIVPACS)) (Wright 2000). The second major difference in the RCA is the spatial scale at which replication occurs. Replication and variance in the RCA are measured at the site scale rather than replicates within a site. This avoids many of the issues of pseudo-replication (e.g. Hurlbert, 1984, Stewart-Oaten et al.1986, Underwood 1997).

The Northern Ontario Benthic Invertebrate Reference Condition Approach (RCA) Biomonitoring Network (Northern Ontario RCA Study) is the first large-scale, multi-partner RCA project designed to assist the metal mining industry in locating suitable reference sites to meet the Environmental Effects Monitoring (EEM) requirements of the Federal Metal Mining Effluent Regulations (MMER). A number of mines in northern Ontario have collaborated in this study

using methods developed by the National Water Research Institute (NWRI), Environment Canada (Reynoldson et al. 2001), and the Ontario Ministry of the Environment (OMOE) (David et al. 1998). The partners involved in this study include: Goldcorp Inc., Inco Ltd., Newmont Canada Inc., Placer Dome Ltd., Williams Operating Corporation, Environment Canada - Ontario Region, National Water Research Institute, Ontario Ministry of the Environment, Acadia University and the Cooperative Freshwater Ecology Unit (Co-op Unit) of Laurentian University.

The objective of this study was to develop a large network of reference and test sites to assess and monitor mining effects on surface waters by detecting any impairment in benthic invertebrate community structure. This project has attempted to define appropriate reference sites for comparison of benthic invertebrate communities with sites potentially impacted by mining operations. A large number of reference sites were sampled to characterize normal benthic assemblages for the region. The reference sites were classified by similarity of their assemblages. Models were developed to select habitat characteristics important in classifying reference site to help select appropriate reference sites for a given test site. These reference sites represent different geographic regions and habitat types and are used to establish the type of invertebrate community expected to occur in the various habitats or regions. Once reference conditions are established, they can be used to assess environmental impacts, conservation status, or biodiversity at any new site (Rosenberg et al. 1998).

This network is similar to that proposed for a Canadian Aquatic Biomonitoring Network (CABIN) by Environment Canada (Reynoldson et al. 1999), and the Ontario Benthos Biomonitoring Network (OBBN) by the Ontario Ministry of the Environment. Our data and those of the OBBN are contributing to the national network. CABIN provides a consistent protocol for aquatic ecosystem assessment that addresses:

- The environmental problems that affect large regions and have cumulative effects on freshwater ecosystems.
- The ongoing requirements for benthic invertebrate surveys under EEM.
- The regional requirements for biological assessment.
- The need for an early warning system for ecosystem changes and identification of long term trends.

The study design followed other RCA designs developed by the National Water Research Institute (NWRI) of Environment Canada, with modifications sufficient to meet the needs of the EEM program and requirements of the participating mines. The program was designed to develop a network of reference and test (exposed) stations, and identify appropriate variables to be measured for ongoing monitoring of community structure and the detection of impairment. The reference sites can be used to define numeric targets for interpretation of the state of the ecosystem in the vicinity of mining operations.

2.0 METHODS

2.1 Selection of Reference and Test Sites

The utility of a monitoring program lies in its ability to detect effect. Historically, water quality monitoring programs have mostly used chemical indicators. However, the adoption of an ecosystem based approach has required the inclusion of methods that define biological effects. Biological assessment is difficult because of the variability of biological systems and the lack of available assessment criteria. The use of reference sites is the most practical method for establishing normal or control conditions. This requires sampling a network of reference stations that reflect acceptable, non-impacted or minimally impacted conditions. A minimum number of 50 reference sites has been suggested for building RCA predictive models (Reynoldson and Wright, 2000). In order to capture the normal range of biological variation, these reference sites need to be distributed randomly in minimally impacted habitats that are similar to those where test sites are located. In our study, reference sites were distributed equally among stream and lake orders and in different size classes of lakes.

To make data collection more cost effective and to address issues of confounding factors and too few local reference sites, a number of mines in northern Ontario have collaborated in this study. To meet the needs of the mining industry, reference sites were located in habitats that matched the environments of the receiving areas. Both stream and lake habitats were sampled and included in the reference site database because near-field sites for metal mining Environmental Effects Monitoring (EEM) can be in both habitat types (Appendix 1). A survey of the participating mines indicated two general stream types: soft bottomed slow moving, and fast flowing gravel and cobble bottomed. For lake environments, sampling was restricted to the wadeable littoral zones (<1 m depth) and focused on lake inflow and outflow areas. Lakes were stratified by size:

large (> 50 ha) and small (< 50 ha) lakes. In addition to reference sites representing unimpaired conditions, a subset of sites was sampled to represent historic and urban impacts. These are both confounding factors that require discrimination from the effects of current activities. Inclusion of these sites in the database may permit the RCA to discriminate mining effects from other confounding effects.

The four areas surveyed during this study (Figure 1), the Nelson River Basin (Red Lake), the Lake Superior Basin (Hemlo), the Lake Huron Basin (Sudbury) and the Moose River Basin (Timmins) are in the the Boreal Shield Ecozone. In addition to sampling conducted by Co-op Unit staff, data from reference sites from the southern portion of the Moose River Basin collected in 1999 by Kilgour et al. (2000) were included in analyses.

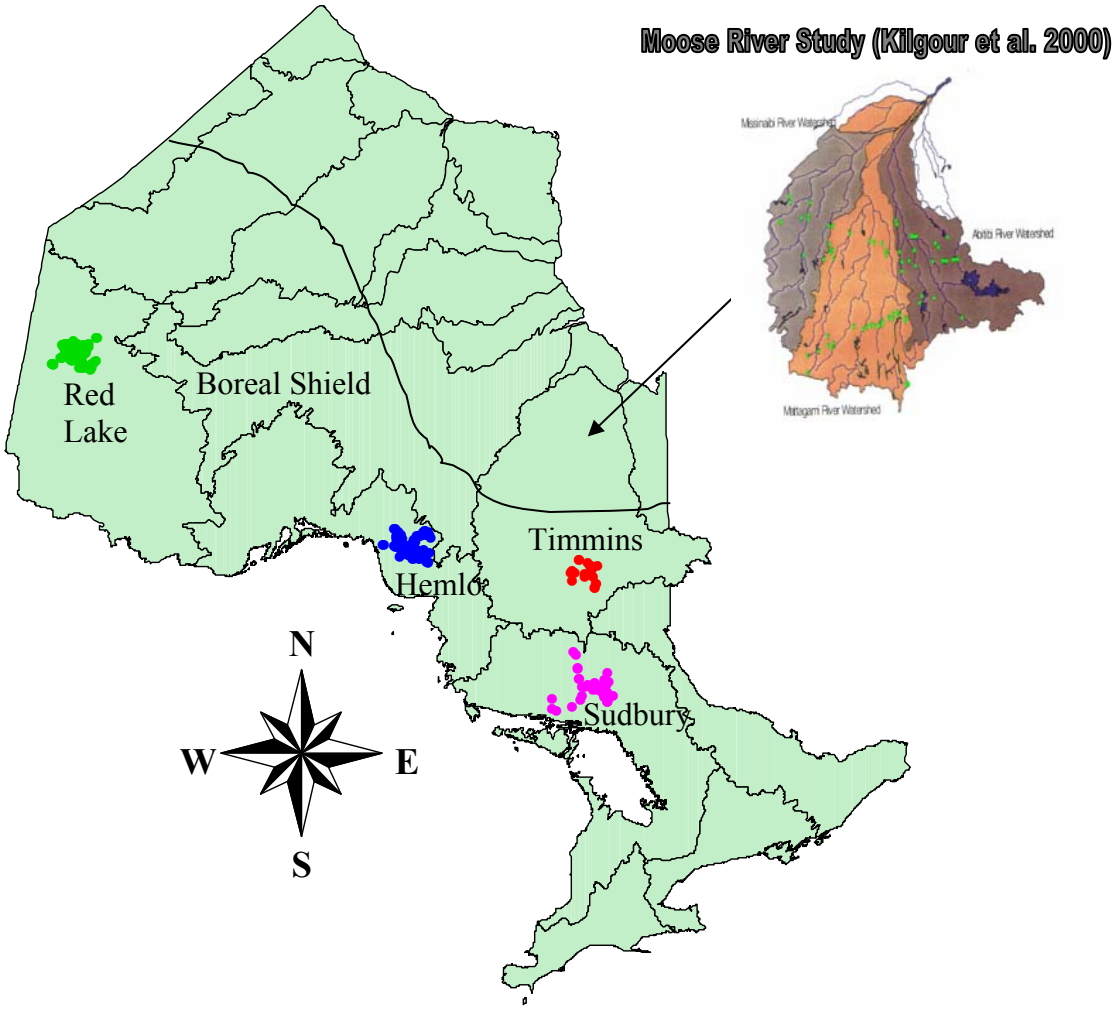

Figure 1. Sample sites (Fall 2003) included in phase one of the northern Ontario RCA study. Four mining areas were sampled: Red Lake (green circles), Hemlo (blue circles), Timmins (red circles), and Sudbury (pink circles). Data from a previous Moose River study (inset: green circles represent sample sites) was also included in some analyses.

Through a series of workshops and teleconferences with provincial and national government experts and participating mining partners the criteria for site selection were determined. Reference sites were allocated to water bodies using random numbers to select a grid squares on 1:50,000 topographic maps (Table 1). Test and historically impacted sites as well as urban streams were also sampled to test model performance. Locations were assigned until sufficient sites of each category were identified. In the event that a field visit proved a randomly selected site to be unsuitable based on accessibility issues or unsuitable habitat for sampling, the site was moved to the nearest suitable location (Appendix 1).

Table 1. Fall 2003 distribution of sites in identified habitat strata.

Location	**Reference**			**Urban**	**Historically impacted**			**Test**		
	Stream	**Lake**		**stream**	**Stream**	**Lake**		**Stream**	**Lake**	
		small	**large**			**small**	**large**		**small**	**large**
Sudbury	12	6	2	14	2	0	0	1	0	0
Red Lake	21	16	27	0	2	4	2	1	0	2
Hemlo	25	41	10	0	0	0	0	0	4	0
Timmins	0	7	10	5	0	0	0	0	0	0
Total	**58**	**70**	**49**	**19**	**4**	**4**	**2**	**2**	**4**	**2**

2.2 Sampling

At both reference and exposed (urban, historically impacted and test) sites, we sampled the benthic community, water chemistry and characterized the habitat. Habitat variables collected for each site included remotely sensed (map based) variables and field-measured variables (Table 2).

Table 2. Habitat variables collected for each site in the northern Ontario RCA study.

Landscape-scale habitat	Site-scale habitat	Channel form and flow	Chemical and physical properties of water	Degree of impact
- Latitude - Longitude - Altitude - Ecoregion - Drainage area - Land use - Stream order	- Date - Flow state - Macrophyte cover - Riparian vegetation - Canopy cover	- Channel width - Mean depth - Max. depth - Bank width - Substrate heterogeneity - Particle size - Velocity	- pH - Dissolved Oxygen - Conductivity - Temperature - Total Phosphorus - Nitrate - NH_4 - Hardness - Alkalinity - Total Suspended Solids - Cyanide - Dissolved Organic Carbon - SO_4 - Ca, Mg, Na, K - Total Metals (as per EEM): Al, Cd, Fe, Mo, As, Cu, Pb, Ni, Zn	- Evidence of historic tailings deposition

2.3 Water Chemistry

With the exception of those from TEST16, TEST17 and TEST18, water samples were shipped via Purolator every other day to the Elliot Lake Research Field Station (accredited by Standards Council of Canada in 2002) for analysis (Table 2). Two reference sites (one lake and one stream) were sampled for Ra226 in each mining area. Analysis for cyanide was performed on all samples (except TEST16, TEST17, and TEST18) by PSC Analytical Services. Eight samples (two in each mining area) were sent to the Ministry of the Environment for detection of low-level mercury.

2.4 Invertebrate Sampling

Field sampling was conducted from September 3 to November 5, 2003. Benthic invertebrates were collected using a travelling kick-and-sweep technique. The substrate was kicked to disturb and dislodge the surface sediment which was swept up with a standard D-frame net with 500 µm mesh. In areas where there was an abundance of macrophytes, detritus or organic matter, which could potentially clog the net, the contents were emptied periodically. The content of the kick net was emptied into a plastic sample jar and preserved with 10% buffered formalin. After a minimum of 72 hours, the samples were gently rinsed with tap water in a 500 µm sieve and then preserved in 70% ethanol.

2.4.1 Stream Sampling

Stream invertebrate sampling was conducted according to Environment Canada's CABIN protocols (http://cabin.cciw.ca/cabin/asp/english/cabin_online_resources.asp; Reynoldson et al.1999), to be consistent with previous reference site sampling done in the Moose River Basin (Kilgour et al. 2000). The collector zigzagged from bank to bank for 3 minutes.

2.4.2 Lake Sampling

Lake sampling methods were similar to procedures outlined in David et al. (1998) with a few modifications to accommodate the needs of the mining industry (see Appendix 3 for more detail). Benthic invertebrates were collected for approximately 10 minutes at each site along a varying number of transects (0-1 m depth) aligned perpendicular to shore and located 2 m apart. At each site (inflow and outflow), this procedure was followed for three samples located approximately 50 m apart. The resulting data was pooled prior to data analysis.

2.5 Sample Processing and Invertebrate Identification

Samples were randomly sub sampled using a Marchant box (e.g. Marchant 1989, Rosenberg et al. 1998) and sorted in the laboratory using a stereo microscope (16X magnification). Consecutive cells were sorted until the 300th animal was found and then the remainder of the material from that cell was also sorted. For lakes, a 100 organism fixed count for each lake triplicate was followed for a total of 300+ animals per lake site. Animals were identified to the family level with retention of specimens for future lowest practical level identification.

2.6 Quality Assurance / Quality Control (QA/QC) Procedures

The northern Ontario RCA study participated in Environment Canada's CABIN protocol (Reynoldson et al. 1999) sampling, sorting and taxonomic QA/QC programme. This is an ongoing process which has the goal of prevention, early detection and correction of field and analytical data collection errors (U.S. EPA 1995).

2.6.1 Field Sampling QA/QC

While in the field, technicians ensured that all data sheets (CABIN field forms modified to include more habitat variables) were accurately and completely filled in. They also determined whether the data were reasonable before leaving the field, and if not, the measurements were repeated before leaving. A partner system was used to verify substrate and habitat classification and to ensure that proper methods were used and samples were correctly labelled. To estimate the variability associated with sampling technique, 10% of the lake and stream sites were sampled in triplicate.

2.6.2 Laboratory QA/QC

To ensure acceptable sorting efficiency (>90% of animals in a sub-sample found) for sample processing, each technician was required to retain the residue (enumerated Marchant box cells only) from their first five samples and every tenth subsequent sample. These samples were re-picked in their entirety and the number of new organisms found was counted. Sorting efficiency was calculated as the average number of organisms missed divided by the total number of animals picked.

2.6.3 Invertebrate Identification QA/QC

Benthic invertebrate identification was conducted using appropriate taxonomic keys and by comparing our identifications with the reference specimens verified by Craig Logan of the taxonomy laboratory of Environment Canada's National Water Research Institute. Identification efficiency was also determined by Craig Logan. For each taxonomist (3), 10% of identified samples were verified to ensure that accuracy was greater than the desired 90%.

2.6.4 Data Management

To maintain consistency, all data were entered into Environment Canada's Benthic Information for Reference Conditions (BIRC) database. Line-by-line data verifications were conducted by a different person.

2.7 Data Analysis

Multivariate methods were used to describe and classify the invertebrate assemblages, to develop the predictive models based on the relationships between the invertebrate and habitat data sets, and to compare the test sites with appropriate reference sites. There are five steps in using reference sites to characterise the biological conditions of a region, and comparing test sites to an appropriate subset of reference sites (Reynoldson et al., 1997).

These are:

(1) The collection of data on invertebrate assemblages and habitat characteristics at a range of reference sites.

(2) The classification of sites into groups using clustering methods based on the similarity of their invertebrate composition.

(3) The development of a predictive model based on the reference-site groups using discriminant function analysis with selected habitat characteristics.

(4) Matching a test site to an appropriate reference-site group based on habitat.

(5) The comparison of a test site to one of the reference-site groups in taxonomic ordination space to determine degree of impairment.

These steps involve fairly complex analyses which are described in some detail below.

2.7.1 Classification of reference sites

Classification methods were used to describe the structure of the biological data used for four different models. The 2003 Co-op Unit data set alone was used for the development of three models:

1) A stream model with 57 reference sites and 99 taxa.

2) A lake model with 92 reference sites and 91 taxa.

3) A combined lake and stream model with 149 sites and 106 taxa.

A fourth model was:

4) A combined lake and stream model with 225 sites that included 76 reference sites sampled from the Moose River Basin in 1999 using the same collection methods (Kilgour et al. 2000) with 122 taxa.

All benthic invertebrate data were included in the analyses. The only taxa eliminated were those that were planktonic or represented meiofauna (e.g., Ostracoda, Copepoda). The optimal model developed was used to assess a number of test sites from each of the mines participating in this study.

For each model, the counts for each taxon (family) were used as descriptors of the benthic invertebrate assemblage. In the case of the lakes data set, the invertebrate data of the three replicates was combined to produce a single value comparable to a stream sample. Habitat measurements from the three lake replicates were averaged when more than one measurement was taken. The data were examined using raw counts and both 4^{th} root and log transformation to

determine which provided the strongest pattern in the assemblage. Transformation had the effect of up weighting the importance of less abundant taxa, with a 4^{th} root transformation being intermediate and a log transformation adding the most weight to less abundant taxa. The Bray-Curtis association measure was used as an association metric for the benthic invertebrate counts because it performs consistently well in a variety of tests and simulations using different types of data (Faith et al. 1987; Jackson 1993).

The reference sites were clustered using agglomerative hierarchical fusion (unweighted pair-groups method with arithmetic averages (UPGMA)) with beta set at 0.1. This has the effect of producing less "stringy" dendrograms (Belbin 1993). The appropriate number of groups was selected by examining the group structure, the spatial location of the groups in ordination space, and Analysis of Similarities (ANOSIM) and Similarity Percentage Analysis (SIMPER). ANOSIM assists in determining the differences among the groups. This method uses the similarity matrix and Monte Carlo randomization of the sites in the groups to examine the significance of differences among the groups. SIMPER looks at the contribution of the fauna to the groups, the differences among the groups, and the contribution of families to both the similarities and dissimilarities. A minimum group size of ten sites was used as the general stopping rule for the classification (Wright 1995). Groups formed with fewer than ten sites were considered to be too small to sufficiently describe normal variability and hence, not very useful for comparison with test sites.

Ordination was also used to describe and explain the variability observed among the large number of taxa with a reduced number of new variables (ordination axes). In this analysis, we used a hybrid multi-dimensional scaling (HMDS) method of ordination that incorporates both metric and non-metric scaling (Faith et al. 1987), and that is of particular value when relating ordination scores to environmental characteristics. All clustering and ordination was done using PATN, a pattern analysis software package developed by CSIRO in Australia (Belbin 1993).

2.7.2 Prediction of site groups

To evaluate the predictive value of our habitat variables, we used a variant of Discriminant Function Analysis (DFA), Multiple Discriminant Analysis (MDA) to rank variables based on their ability to discriminate between reference site groups. In addition, MDA provides insights into:

- The differences among groups.
- The most parsimonious way to distinguish groups.
- The variables which are important for group distinctions.
- The likelihood that a site belongs to one or more groups.
- The correctness of the classification by observing whether cases are classified as predicted.

Habitat variables were excluded from initial consideration as predictors if they were likely to be influenced by anthropogenic activity (e.g., nutrient enrichment, physical disturbance). Habitat variables were also excluded if temporal variability would result in a measurement being unrepresentative (e.g., temperature, dissolved oxygen).

Selection of appropriate habitat variables for use in predictive models was done in two ways. The first used principal axis correlation (PAC, procedure PCC in PATN), which determines how well the habitat data can be fitted to the benthic community data in ordination space. This method is a multiple-linear regression that takes each habitat attribute and determines the location of the vector that has the best fit in ordination space. The variables can be represented as an axis on an ordination plot, and a correlation of the variable with the ordination is provided. A randomisation model is used to establish the statistical significance of the correlation.

The second method of variable selection used stepwise DFA (SYSTAT 8.0) to establish which environmental variables best separated sites into the predefined groups formed by classification of the biological data set. Stepwise selection of variables was used, and the significance level for variable entry and retention was 0.05. Based on PAC and stepwise DFA, environmental variables were used in DFA to establish functions for the variables that best separated sites into the predefined biological groups. The SYSTAT version of DFA (SYSTAT 8.0) was used with raw environmental data to generate discriminant scores, and to predict the probability of group membership. The accuracy of predictions from the discriminant model was verified by examining

how well the reference sites were predicted to the correct group. The predicted groupings and actual groupings were then compared to provide a group and total error rate.
Final selection of the optimal predictor variables was done by iteration. Various combinations of predictor variables were selected based on the results of the stepwise DFA and PAC. The optimal set was defined as that with the lowest error rate from cross-validation in DFA. (Moss et al. 1987; Parsons and Norris 1996).

2.7.3 Assessing test sites

The objective of developing predictive models is to assess sites exposed to a stressor and to determine the degree of difference between the exposed site and the reference sites. We have used two approaches for determining if sites are equivalent to reference sites. These are RIVPACS (River Invertebrate Prediction and Classification System) and BEAST (BEnthic ASsessment of SedimenT).

2.7.3.1 RIVPACS

RIVPACS was developed in the United Kingdom by Wright et al. (1984 etc.). A similar approach is also used in Australia, (AUSRIVAS: Australian River Assessment Scheme) by Davies (2000) and in the United States by Hawkins and Norris (2000). This approach examines taxa expected (predicted) to be present at a site with those observed at the site. This allows the calculation of an Observed to Expected (O:E) ratio. It also provides a list of taxa that should be present, which can be compared to those actually present. This method is sensitive to changes in taxa richness and taxa loss, but less so to changes in abundance (Reynoldson et al. 1997, Mazor et al. (submitted)). The RIVPACS assessment method calculates the probability of any taxon occurring at a site. This prediction is a simple calculation and is the summed probability of the specific site being a member of a community group times the percent occurrence of the taxon in the group. The likelihood of a taxon occurring at a site is calculated from the product of the probability of a site being in a group multiplied by the occurrence of the taxa in the group. This provides the contribution for that group to the overall probability of a taxon occurring. The sum of the contributions for all the groups provides the summed probability of a taxon being present. This calculation is conducted for every taxon at the site to determine the predicted community at a site. Because taxon with a probability of 50% or less could be there by chance, these are not used in the calculation of the expected number of taxa. Expected taxa are based on the sum of taxa with a

more than 50% chance of occurrence: the $E_{50.}$ The actual expected number is always less than the number predicted because it is the combined probabilities. For three taxa with a P 0.75 of occurrence, the sum of 2.25 means one would expect less than three taxa. RIVPACS assessment of impairment is done by calculating the O:E ratios for all reference sites. A test site is defined as disturbed if the O:E ratio is less than 90% of the reference range. Stress Bands can be calculated by making them of equal size to the reference band (Table 3).

The RIVPACS approach has two positive attributes. First, it uses all the reference sites when developing a prediction of the expected number of taxa, and it provides a probability of each taxon occurring. The absence of specific taxon can provide some indication of the effects of disturbance. However, the disadvantage of the RIVAPCS assessment approach is that it relies on a single attribute (number of taxa) for determining disturbance. The use of presence/absence data therefore makes the RIVPACS assessment insensitive to changes in abundance.

Table 3. Stress bands for Observed to Expected (O:E) ratio scores for northern Ontario.

Mean 1.03	**O:E score**	**Assessment**
Band 0	> 1.36	A site of exceptional diversity or mild enrichment
Band 1	0.70 - 1.36	Equivalent to reference
Band 2	0.36 - 0.69	Mild disturbance
Band 3	0.02 - 0.35	Disturbed
Band 4	< 0.02	Severe disturbance

2.7.3.2 BEAST

An alternative assessment approach is the BEAST method which was originally developed for the Great Lakes (Reynoldson et al. 2000) and is now widely used in Canada. In the BEAST method, the invertebrate community of a test site is plotted in the same ordination space as a subset of reference sites (Figure 17). If the test site is within the range of variation observed for reference sites they are deemed as equivalent to reference. If they depart from normal variation then the degree of departure is an indication of the difference.

To assess a test site, it is first predicted to one of the reference-site groups formed from the original classification, using the predictive model constructed from DFA. Variation among the reference sites, in the group to which the test site is predicted, is assumed to represent the normal range of variation expected at the test site if it is unimpaired. Reference sites and test sites are plotted in ordination space. If the test site falls within the “cloud” of sites representing the

reference condition then it is considered to be equivalent to reference; if it is outside the reference-site cloud it is considered to be different than reference. Wright et al. (1991) developed a method that used a series of bands, representing grades of biological quality from good to poor, to provide a simple statement of biological quality, which allows broad comparisons in either space or time that would be useful for management purposes. Four categories or bands of biological quality based on how similar a test site is to the reference sites were established using probability ellipses. Band 1 is the inner ellipse (90% probability) formed around the reference sites, and is "equivalent to reference or unimpaired" (Figure 17). Band 2 is immediately adjacent to the reference (inner) ellipse (99% probability), and is "possibly different from reference or potentially impaired". Band 3 is between the 99% probability and the largest ellipse (99.9% probability), and is "different from reference or impaired". Band 4 is outside the 99.9% ellipse, and is "very different from reference or severely impaired". If three ordination axes were required to describe the variation in the invertebrate assemblage, then sites were plotted on all three axes and the greatest difference from reference was used to assign the status of a test site. Gaussian bivariate probability ellipses (Altman 1978; Owen and Chmielewski 1985) in SYSTAT (Procedure Scatterplot) were constructed around the reference sites to define the bounds of the reference-site cloud.

The advantages of the BEAST approach are that it is responsive to changes in both abundance and richness; it incorporates all the information about the reference community in a non-subjective manner and provides no weight to various attributes of the assemblage. The disadvantage is that it is dependent on the accuracy of the predictive model, and if the site is incorrectly predicted to a group, then it compares the test site to inappropriate reference sites. This method has shown to be more sensitive to changes in abundance than RIVPACS but less so to loss of rare species (Mazor et al. submitted). The early response to stress may appear as a loss of rare species or as change in abundance, and therefore both methods should be applied in a comprehensive assessment. Both these methods have been shown to be more precise, accurate and sensitive than traditional metric based assessment methods (Reynoldson et al. 1997, Mazor et al. submitted). It is suggested that both methods be used, as they are complementary. However, a final decision on the status of a site should probably use the more sensitive in any particular case. Using simulated and real data, the BEAST approach has been demonstrated more responsive to change in abundance and RIVPACS to change in richness (Bailey et al. 2004, Mazor et al. submitted).

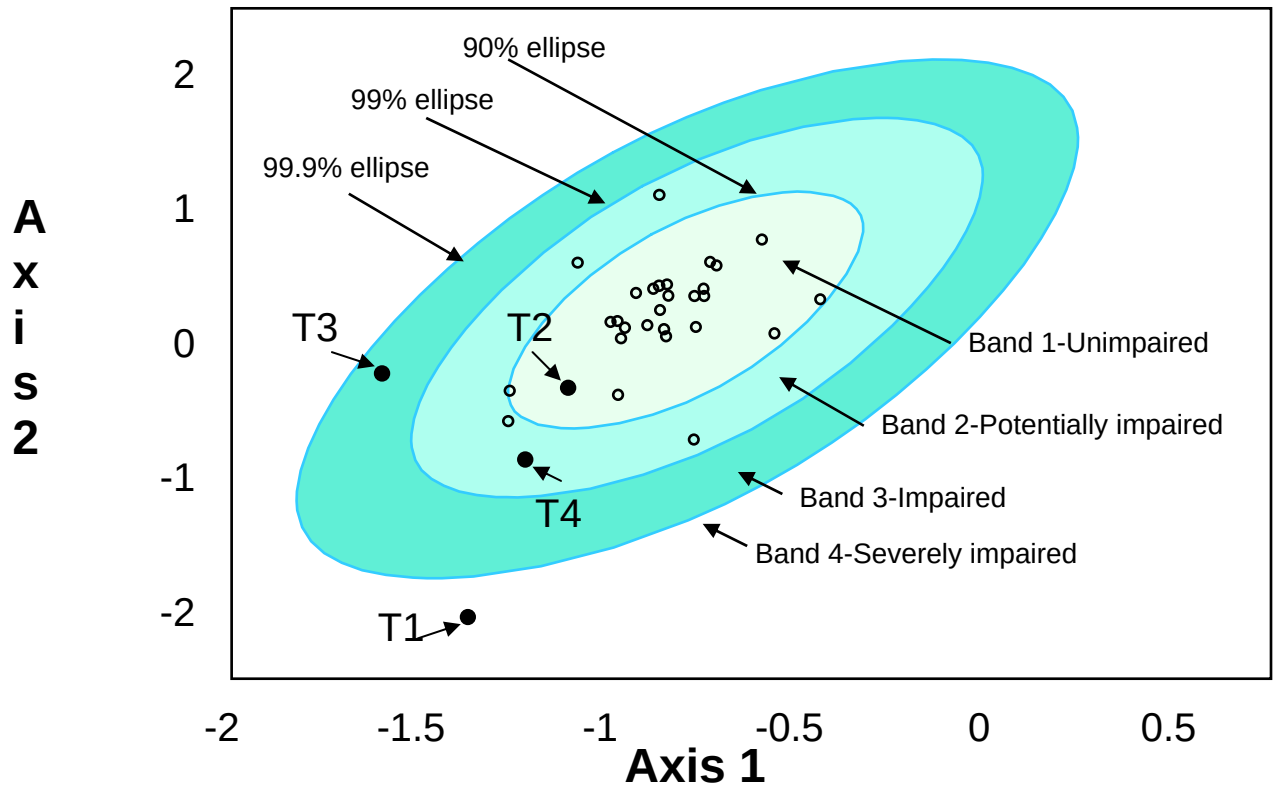

Figure 2. Theoretical BEAST assessment showing reference sites (open circles) and four test sites (solid circles) which are predicted to this set of reference sites. Site T2 is within the 90% probability ellipse constructed around reference sites only and is in Band 1, and therefore unimpaired. Site T1 is outside the 99.9% ellipse, in Band 4, and is therefore very different from reference and possibly severely impaired.

3.0 RESULTS AND DISCUSSION

3.1– Stream Sites

3.1.1 Stream Benthic Community Classification

The 57 stream sites were clustered and groups of sites were examined. A small, diverse, group of 8 sites (Group 2) was formed (Figure 3). Small groups of sites are problematic in RCA because the full range of natural variability seen in a community type is probably not represented. A minimum of 10 sites per reference group was recommended by Reynoldson and Wright (2000), therefore, a five group solution was initially considered (Table 4). The Group 2 sites have been left as a unit because the ordination (Figure 4) indicates that they are no more dispersed than the sites in the other four groups. This group represents sites with generally low abundance of Chironomidae (Table 4), but high richness and abundant Leptophlebiidae mayflies. The remaining groups represent a continuum (Figure 4) of notably Chironomidae, Sphaeridae, Ceratopogonidae and Ephemerellidae from low abundance at Group 2 to highest abundance in Group 1Bi. The latter group consists of three sites that, at this time, and until more such sites are acquired, may need to be merged with Group 1Bii.

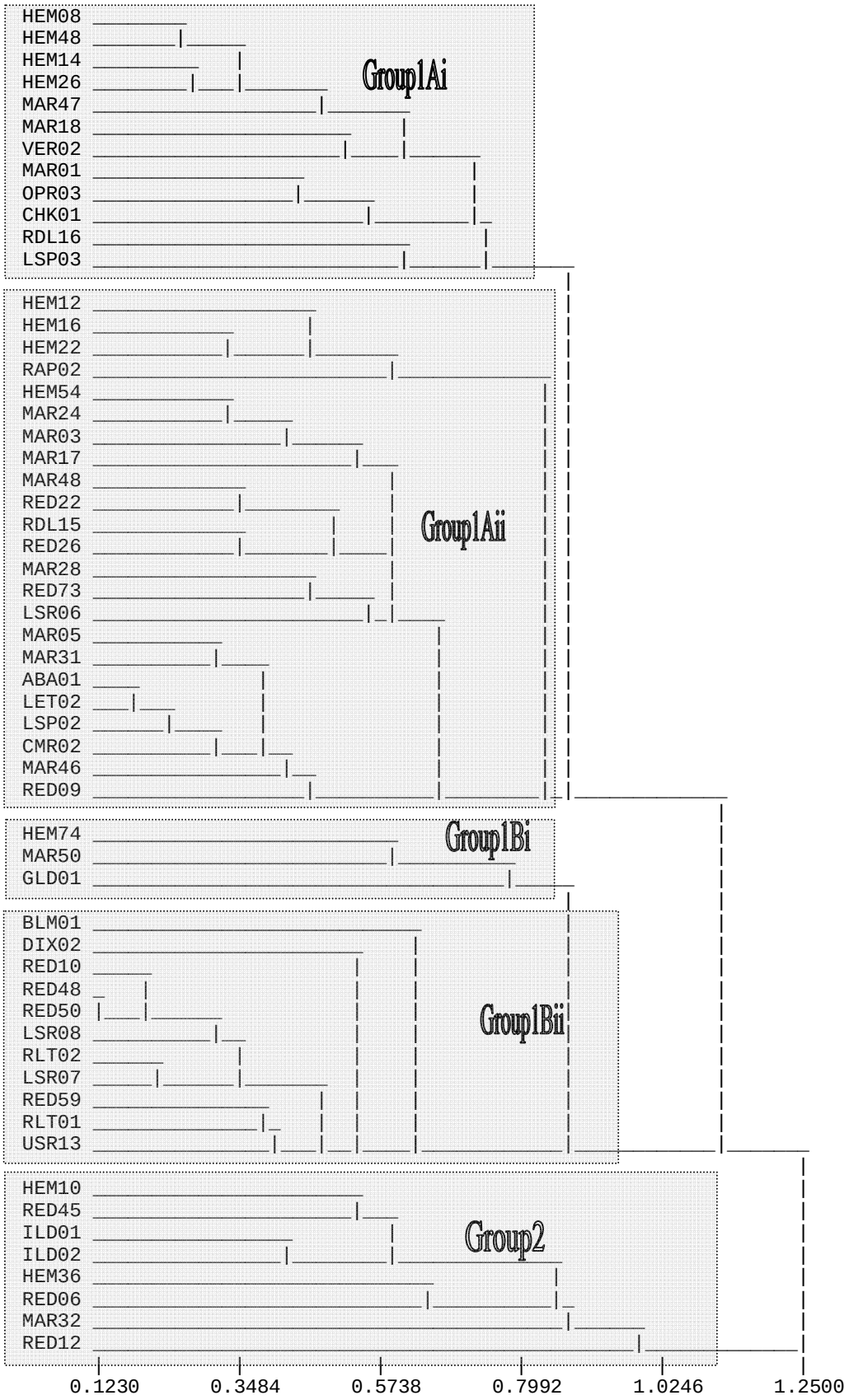

Figure 3. Dendrogram of 57 stream reference sites formed with raw counts of 99 taxa. The Y axis is categorical and represents the site names; the x axis is the coefficient of similarity.

Table 4. Major contributing taxa in 5 groups formed from 57 stream sites. Mean within-group abundance is given for each family. Richness is the mean number of families found at each site.

	Contribution of taxa to ordination	Group				
		1Ai	1Aii	1Bi	1Bii	2
No. sites per group		12	23	3	11	8
Chironomidae	0.754	422	1110	5123	3916	92
Ephemerellidae	0.597	50	139	1093	36	24
Ceratopogonidae	0.593	40	98	267	258	19
Leptophlebiidae	0.583	41	70	400	121	100
Gomphidae	0.545	47	14	177	25	11
Hyalellidae	0.505	21	150	217	370	98
Sphaeriidae	0.494	210	181	1013	171	16
Hydropsychidae	0.464	33	56	3090	7	21
Heptageniidae	0.462	26	19	680	29	40
Elmidae	0.440	29	84	1867	82	14
Abundance		1314	2790	18663	6522	640
Richness		22.1	21.6	19.0	17.6	24.1

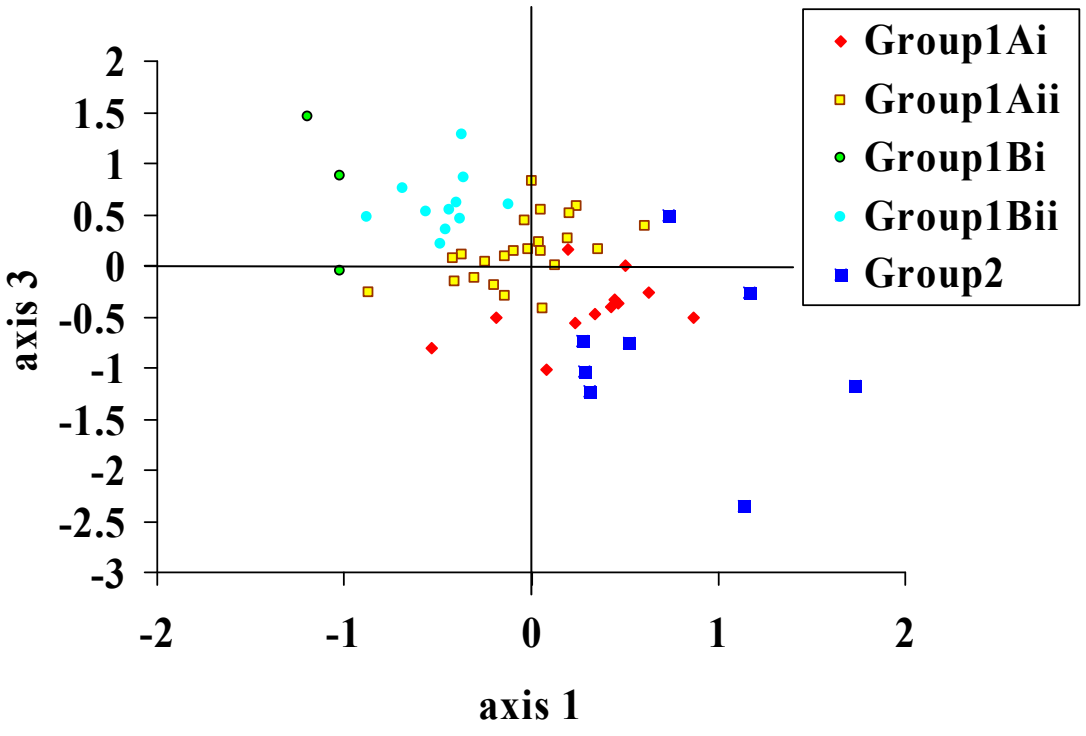

Figure 4. HMDS ordination of stream sites using untransformed data; group membership is based on clustering.

Analysis of the differences among the groups (ANOSIM), for the global test (all groups) was highly significant indicating that the groups are dissimilar (Table 5). The pairwise test indicates that Group 1Bi was very different from the other groups and Group 1Ai was the least dissimilar.

Table 5. Analysis of the differences among the groups (ANOSIM) formed from raw taxa counts.

Global R 0.697	**Group1Aii**	**Group1Bi**	**Group1Bii**	**Group2**
Group1Ai	0.502	0.934	0.957	0.575
Group1Aii		0.833	0.562	0.902
Group1Bi			0.881	0.890
Group1Bii				0.907

Transformed Stream Data

Stream data were transformed using the 4th root. The groups using transformed data are more similar in size (Table 6). The differences in abundance of the families contributing to the patterns in the community are not as large as when using the raw data. The Enchytraeidae, which are important contributors to the patterns in the transformed data only range from 6 to 50 per site in the groups. This greater similarity among the groups is confirmed from ANOSIM which shows a lower Global R value (Table 7) and less difference among the groups (lower r values in the pairwise tests) than with raw data. There is less structure in the transformed data based on the sites in ordination space (Figure 5). Therefore, only groups formed by the raw data were examined further.

Table 6. Major contributing taxa in 5 Groups formed from 4th root transformed data from 57 stream sites. Mean within-group abundance is given for each family. Richness is the mean number of families found at each site. Note: site RED 12 not used in transformed data set.

		Group				
	Contribution of taxa to ordination	**1A**	**1B**	**2A**	**2Bi**	**2Bii**
Sites per group		13	13	5	14	11
Chironomidae	0.699	620	1937	484	2234	2076
Enchytraeidae	0.623	48	6	47	25	50
Ceratopogonidae	0.561	66	88	24	204	140
Ephemerellidae	0.530	102	347	34	116	1
Caenidae	0.418	38	56	6	207	31
Abundance		1603	6181	1218	4698	3608
Richness		23	23.2	23.6	23	13.7

Table 7. Analysis of the differences among the groups (ANOSIM) formed from 4th root transformed data.

Global R 0.536	Group1B	Group2A	Group2Bi	Group2Bii
Group1A	0.283	0.018	0.411	0.437
Group1B		0.446	0.647	0.777
Group2A			0.374	0.681
GroupBi				0.546

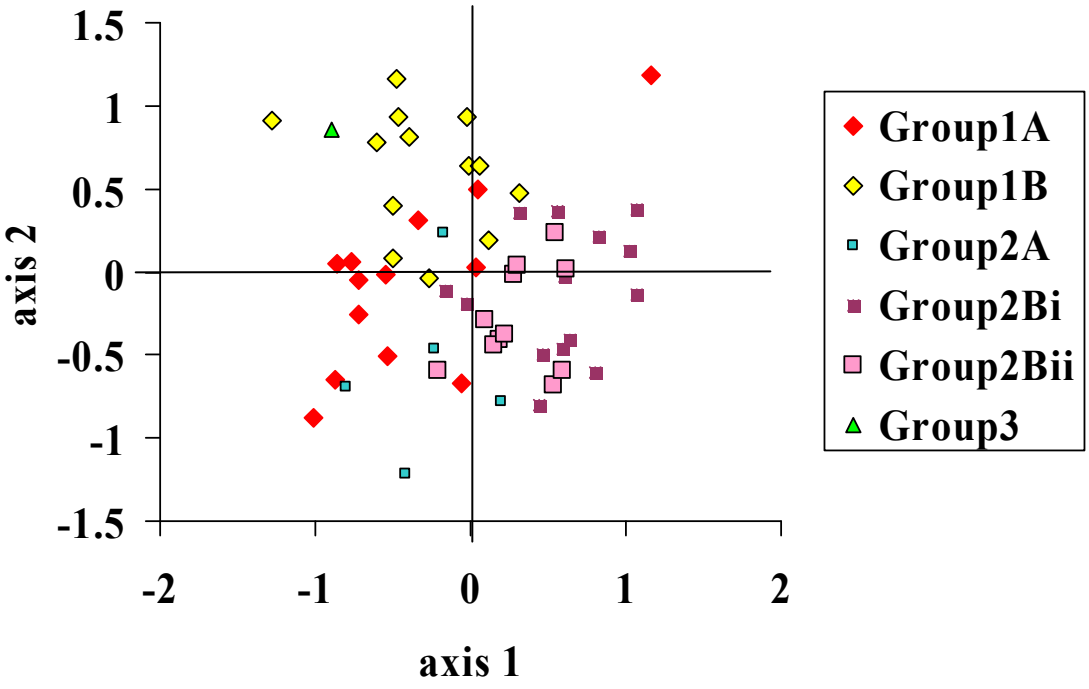

Figure 5. HMDS ordination of the 57 stream sites using transformed (4th root) data.

3.2.1 Streams Habitat Patterns

Data for 28 habitat variables were complete, a further 26 habitat variables were missing for three sites. These variables were included and the average value was used for the missing value, so as to unweight the variable at those sites. Nine variables had 30 or more missing values and were not used. All variables did not meet assumptions of normality and were $\log_{10} + 1$ transformed to meet the assumptions of PCA, ANOVA and MDA.

From PCA, regional differences in habitat attributes were evident. The sites from the three regions (Hemlo, Red Lake, and Sudbury) overlap very little in the ordination space (Figure 6). However, much of the variation was unexplained on the first two components, which together

only explained 32.7% of the variation in the data. Despite this, ANOSIM indicated that habitat differences among the three mining regions were significant (Global R 0.408) whereas habitat differences among the community groups where not significant (Global R 0.019). The first component revealed an environmental gradient based on water chemistry. In general, reference sites located in the Hemlo and Timmins area were significantly harder ($p<0.05$), more conductive and more alkaline and had higher mean concentrations of calcium and magnesium than Sudbury and Red Lake sites (Table 8). Sudbury-area sites had higher sulphate and metal levels than the other mining areas. The second gradient is spatial but is also associated with dissolved organic carbon (DOC) and sulfate.

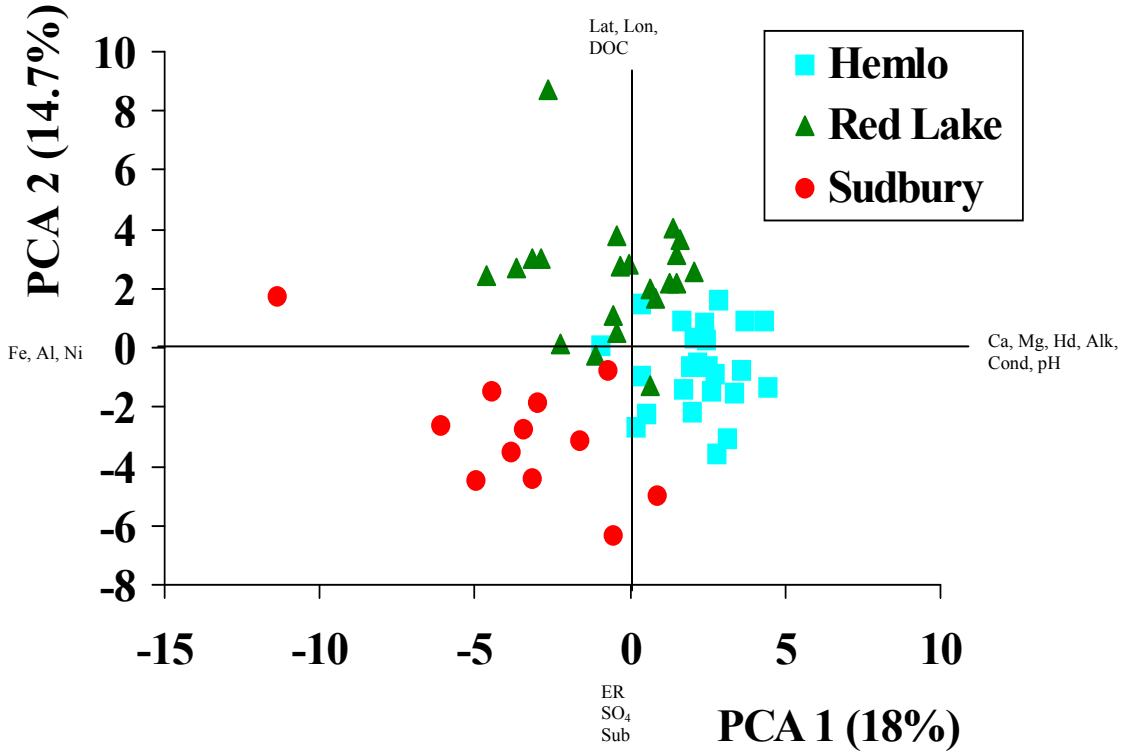

Figure 6. PCA of habitat attributes for 57 stream reference sites in Hemlo, Red Lake, and Sudbury.

Table 8. Mean values from reference sites for selected water chemistry parameters.

Location	Ca mg/L	Mg mg/L	Hardness mg/L	Conductivity µS/cm	Alkalinity mg/L	pH
Hemlo (n = 76)	21.75	4.54	72.85	172.61	66.96	7.61
Sudbury (n = 20)	4.99	1.48	18.55	66.21	13.10	6.82
Red Lake (n = 56)	7.18	1.91	25.80	59.98	20.12	6.76
Timmins (n = 17)	16.82	3.74	57.40	141.09	53.53	7.57

Fifty-one habitat variables were examined in one-way ANOVA tests, using the five groups from the raw data set. Only three variables (percent pool, percent riffle and lead) showed no significant differences ($P > 0.05$) among the groups. The other 48 variables all showed at least one significant difference ($P<0.05$). Variables with the highest F values, and therefore the most likely to discriminate among the biological groups were those associated with spatial location (ecoregion number, latitude, longitude and altitude), water chemistry (pH, DO and conductivity) and channel characteristics (channel depth and riparian vegetation).

3.1.3 Stream Models

Based on the community analyses, five groups of sites produced by untransformed counts were used in the model building. All habitat data were $\log_{10} + 1$ transformed to meet the assumptions of normality for the discriminant analysis. Only those variables not likely to be modified by disturbance were used. Metals were excluded from consideration, as were variables that were not well represented temporally by single samples, such as DO and temperature.

For the reference site classification based on raw data, the results of 10 models are summarized in Table 9. The performance of the models in classifying sites is summarized for both re-substitution and cross-validation methods. The latter is a more accurate indication of model performance as sites are sequentially removed from the model and coefficients are recalculated without that site and the classification is then reported. Not surprisingly, as the number of community groups is increased the models perform less well, ranging from a high of 89% (2 groups backward stepwise) to a low of 35% (4a groups forward stepwise) (Table 9). While five groups can be justified from the analyses of the community data, Group 1Bi only consists of three sites, and in the three-group model is merged with Group 1Bii (Figure 7). The four group model with Groups 1Ai and 1Aii (4aB and 4aF in Table 9) did not perform well with prediction accuracy only 60% and 35% for the backward and forward models and again Group 1Ai only consists of 12 sites. Therefore, at this time we would recommend using the three group backward model using 10 predictor variables (Table 9) with reclassification accuracy in cross-validation of 75%.

Table 9. Summary of predictive models for 2-5 groups formed from 57 stream sites using original counts of the invertebrate community (F are variables selected by forward and B by backward stepwise analyses). The dots represent the variables that are used to discriminate each group. The stars represent the recommended model.

	Number of Groups (stepwise direction)									
Variable in Model	**2 (F)**	**2(B)**	**3(F)**	**3(B)**	**4(F)**	**4(B)**	**4a(F)**	**4a(B)**	**5(F)**	**5(B)**
Latitude						●		●		●
Longitude		●						●		
Altitude (ft)				*		●		●		●
Ecoregion number				*		●		●		●
Wetted Width (m)	●		●		●		●		●	
Bankfull width (m)										
Slope		●		*				●		●
Average Channel Depth (cm)										
Maximum Channel Depth (cm)										
Pool (Percent)										
Rapid (Percent)										
Riffle (Percent)								●		●
Run (Percent)										
Dominant Substrate			●							
2nd Dominant Substrate										
Embeddedness								●		
Surrounding Material										
Conifer trees (presence/absence)		●			●					
Deciduous trees (presence/absence)										
Grasses and ferns (presence/absence)	●	●						●		
Macrophyte (percent)	●	●			●			●	●	●
Shrubs (presence/absence)		●								
Canopy (percent)										
Alkalinity (mg/L)		●		*		●		●		●
Hardness (mg/L)										
Conductivity (µS/cm)	●	●		*				●		●
Total Suspended Solids (mg/L)										
pH (pH)		●		*		●				
Cyanide (CN(t)) (mg/L)										
Dissolved Organic Carbon (DOC) (mg/L)						●				●
Calcium (Ca) (mg/L)		●		*		●		●		●
Potassium (K) (mg/L)		●	●	*	●		●		●	
Magnesium (Mg) (mg/L)										
Sodium (Na) (mg/L)		●						●		●
Iron (Fe) (mg/L)		●		*				●		
Ammonia (NH_3) (mg/L)										
Nitrate (NO_3) (mg/L)					●	●			●	
Sulphate (SO_4) (mg/L)		●		*		●		●		●
Total Phosphorus (TP(Wat)) (mg/L)										
Total variables	4	14	7	10	5	9	2	15	4	12
Classification success (resub)	81	95	61	86	68	79	40	88	53	75
Classification success (crosval)	79	89	58	75	60	70	35	60	39	49

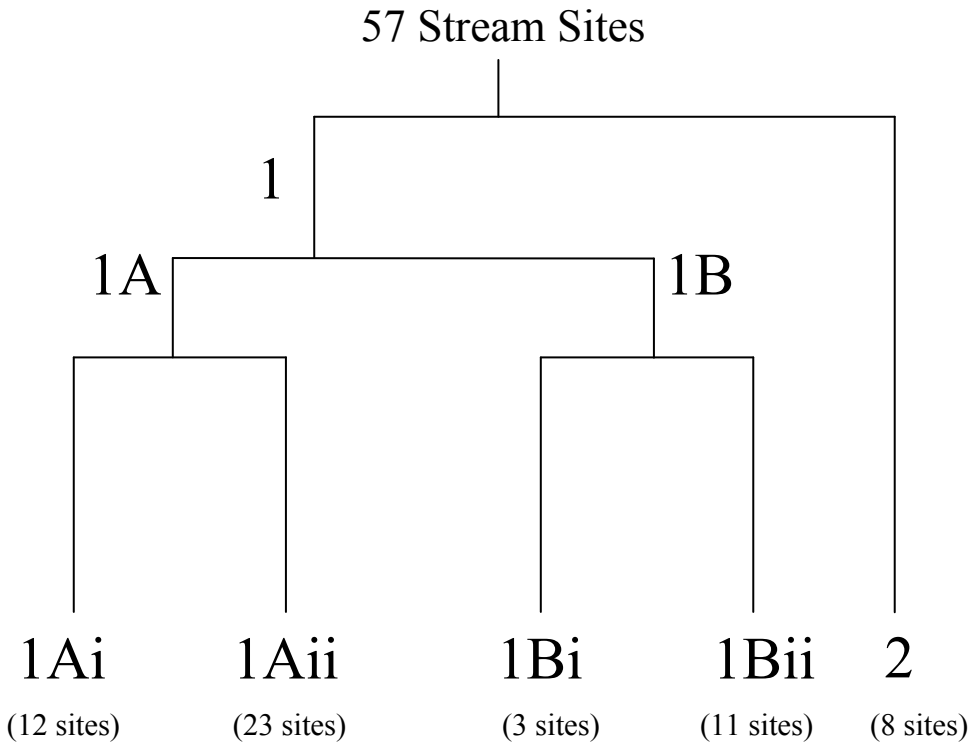

Figure 7. Stream sites - Summary of dendrogram of 57 stream reference sites for 5 community groups based on raw counts.

Details of the model using the three groups are presented in Table 10. The model is well balanced in that all three groups are equally well predicted. From examination of the groups in ordination space (Figure 4) and from the composition of the fauna in the groups (Table 4), the combination of 1Ai and 1Aii into group 1A and 1Bi and 1Bii into 1B is appropriate at this time. In fact, the sites represent a continuum of distributions and this is clear from examining the sites in ordination space (Figure 4). With the current number of available sites, it is preferable to increase model accuracy and characterize group variation than increase the number of groups.

Table 10. Performance of discriminant model using cross validation in predicting sites to groups formed from benthic invertebrate abundance using 10 habitat variables.

	Group 1A	**Group 1B**	**Group 2**
Group 1A	26 (74%)	5	4
Group 1B	2	11 (79%)	1
Group 2	2	0	6 (75%)

3.2 Lake Sites

3.2.1 Lake Benthic Community Classification

Classification of 92 lake sites using 91 taxa at the family level produced the dendrogram shown in Figure 8. The dendrogram beyond the six groups is not shown as the structure in the data set tends to be lost and groups of single sites are formed. Examination of regional differences and differences with lake size showed no pattern in community assemblages associated with either of these attributes (r 0.051 and 0.093 respectively). To guide the determination of the appropriate number of groups to examine, ANOSIM was conducted with 2 to 6 groups as well as the region and lake size (Table 11).

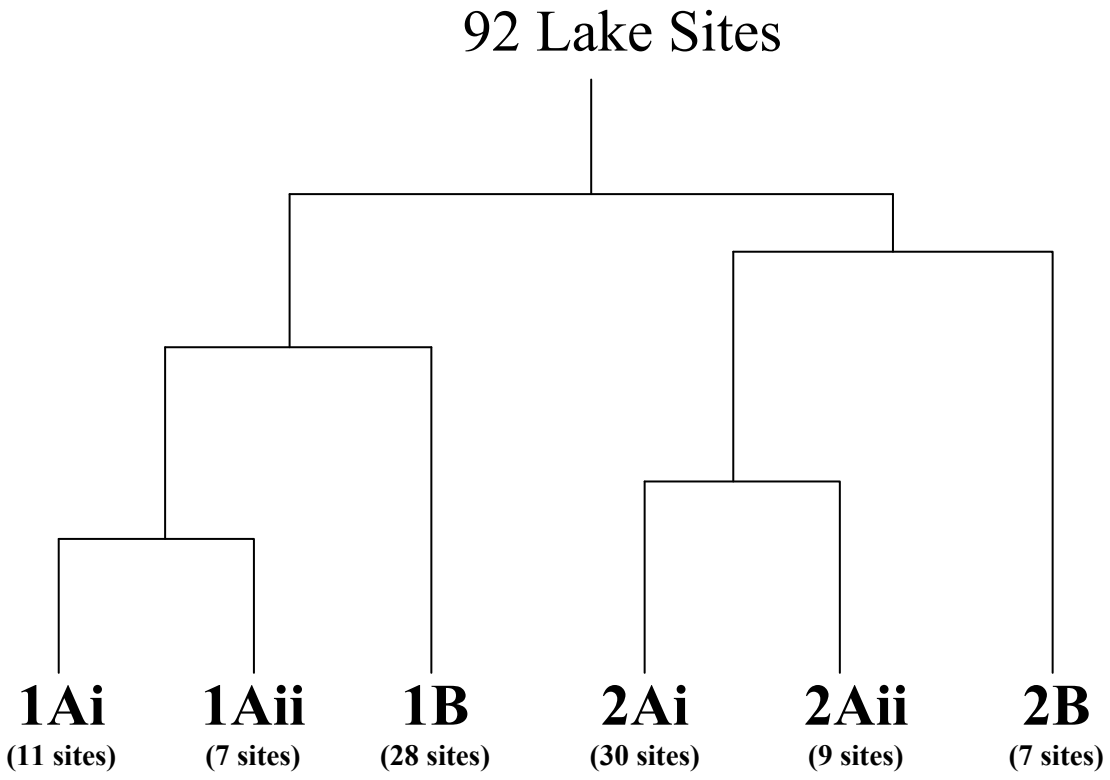

Figure 8. Lake sites - Summary of dendrogram of 92 lake reference sites for 6 community groups based on raw counts.

Table 11. Results of ANOSIM for 92 lake sites (by region, by lake size, plus 5 faunal-composition-based classifications.

	Region	**Size**	**2 groups**	**3 groups**	**4 groups**	**5 groups**	**6 groups**
R value	0.051	0.093	0.446	0.590	0.610	0.679	0.710
P	.68	.04	.01	.01	.01	.01	.01

The differences in the biological groups (Figure 9) relate primarily to abundance of the Chironomidae and Hyalellidae and in all groups except 2B these are the two numerically dominant families. In Group 2B, the Ceratopogonidae, Naididae and Enchytraeidae are numerically dominant; however, group 2B is also the most variable (Figure 9).

Ordination suggests that there are distinct groups of sites that appear to be associated with four identifiable assemblages of taxa: 1) A Chironomidae complex with which the Caenidae and Phryganeidae are associated; 2) a Hyalellidae complex associated with Leptophlebiidae, Planorbidae and Leptoceridae; 3) a Naididae complex associated with the Ephemerellidae, Empididae, Elmidae and Ephemeridae, and lastly; 4) an Enchytraeidae-Dytiscidae complex. Although these family associations can be identified, individual sites are located along a general abundance gradient of the Chironomidae and Hyalellidae.

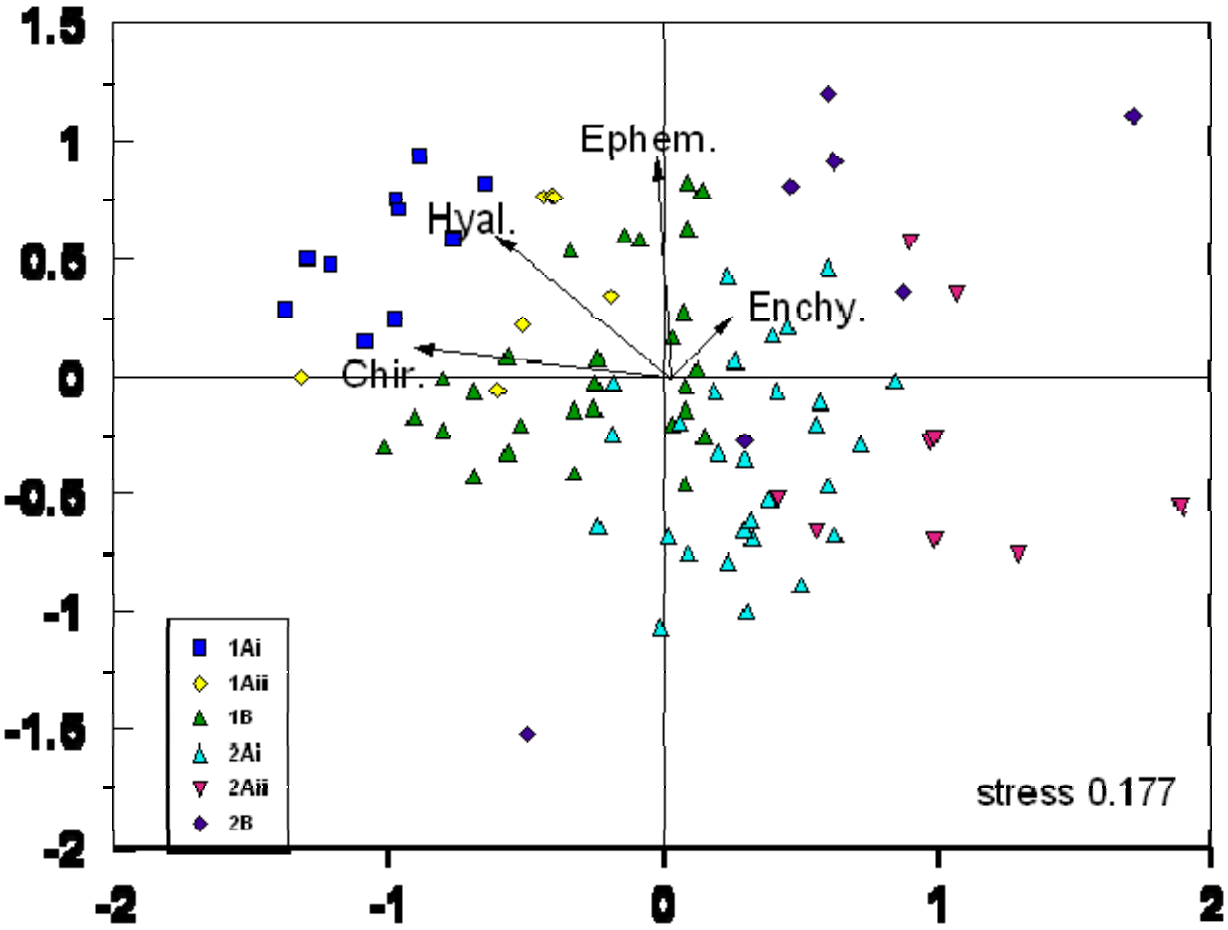

Figure 9. Lakes sites – HMDS ordination using Bray-Curtis distance. The positions of the taxa explaining most of the variation in the ordination space are also shown. (Chir – Chironomidae, Ephem – Ephemerellidae, Enchy – Enchytraeidae, Hyal – Hyalellidae).

3.2.2 Lake Models

Predictive models were built and evaluated for 2-, 3-, 4-, 5-, and 6-group reference-site classifications (Table 12). None of the models performed well in discriminating the assemblages. The best model was with only two groups and even that only predicted 65% of the sites correctly. This is poor considering that half the sites would be correct by chance alone.

There are three possible explanations for this poor performance. The first is that the variables measured are unimportant in structuring the invertebrate communities, and second, that the communities are not very variable and that stochastic processes are more important than deterministic ones, or possibly some combination of these two explanations. The third is that perhaps the data set represents a continuum and groups do not exist. It was decided to examine

the lake and stream sites as a single data set. This provides an indication of the relative variability of the two types of communities, and the degree of similarity, among lake and stream assemblages. A merged data set may provide a more robust model by increasing the number of reference sites per community group available for assessing test sites.

Table 12. Summary of lake predictive models for 2-6 groups of sites formed from original counts of the invertebrate community from 92 sites (B backward stepwise; F forward stepwise) using cross-validation. The dots represent the variables that are used to discriminate each model. The stars represent the recommended model.

	Number of Groups (stepwise direction)									
Variable in Model	**2(B)**	**2(F)**	**3(B)**	**3(F)**	**4(B)**	**4(F)**	**5(B)**	**5(F)**	**6(B)**	**6(F)**
Julian day									●	
Latitude							●		●	
Longitude						●				
Altitude (ft)										
Lake Size (Small or Large)									●	
Ecoregion number									●	
Wetted Width (m)										
Bankfull width (m)										
Slope					●	●	●	●		●
Average Channel Depth (cm)										
Maximum Channel Depth (cm)										
Pool (Percent)										
Rapid (Percent)										
Riffle (Percent)										
Run (Percent)										
Dominant Substrate					●					
2nd Dominant Substrate	*	●								
Embeddedness										
Surrounding Material	*									
Conifer trees (presence/absence)										
Deciduous trees (presence/absence)			●	●	●	●	●		●	
Grasses (presence/absence)			●	●	●	●	●			
Macrophyte (percent)										●
Shrubs (presence/absence)										
Canopy (percent)		●								
Alkalinity (mg/L)										
Hardness (mg/L)					●			●	●	
Conductivity (µS/cm)	*	●								
Total Suspended Solids (mg/L)										
pH (pH)										
Cyanide (CN(t)) (mg/L)										
Dissolved Organic Carbon (mg/L)										
Calcium (Ca) (mg/L)	*	●					●		●	
Potassium (K) (mg/L)										
Magnesium (Mg) (mg/L)	*	●								●
Sodium (Na) (mg/L)										
Ammonia (NH_3) (mg/L)	*									
Nitrate (NO_3) (mg/L)										
Sulfate (SO_4) (mg/L)							●		●	
Total Phosphorus (TP(Wat)) (mg/L)										
Total variables	6	5	2	2	5	4	6	2	8	3
Classification success (crosval)	65	61	49	49	45	38	39	27	32	25

3.3 Combined Lake and Stream Sites

The results of HMDS ordination of the combined data set is presented in Figure 10. Three dimensions were required to attain a suitable stress level (0.164). There is a gradient along a lake-stream axis, but there is considerable overlap in the lake (open circles) and stream sites (shaded squares) with the stream sites encompassing the complete gradient (Figure 10). The lake sites are less variable, being mostly located in the lower left-hand portion of the ordination. There are five taxa or taxa groups associated with this gradient. First, a Chironomidae-Ephemerellidae-Caenidae association that is the primary gradient; this complex is also associated with a Leptoceridae-Planorbidae association. Three other individual families impinge on the main gradient, the Sphaeriidae and the Hyalellidae, which are negatively associated, and the Leptoceridae that are weakly associated with the Hyalellidae and the Leptoceridae-Planorbidae association. The examination of the combined lake and stream data set indicated that constructing a lake-stream model may be appropriate primarily because there is considerable overlap and similarity among lake and stream sites. Combining the data increases the total number of sites and may increase the number of sites available, particularly for groups represented by few sites, thus providing better characterization of the reference assemblages.

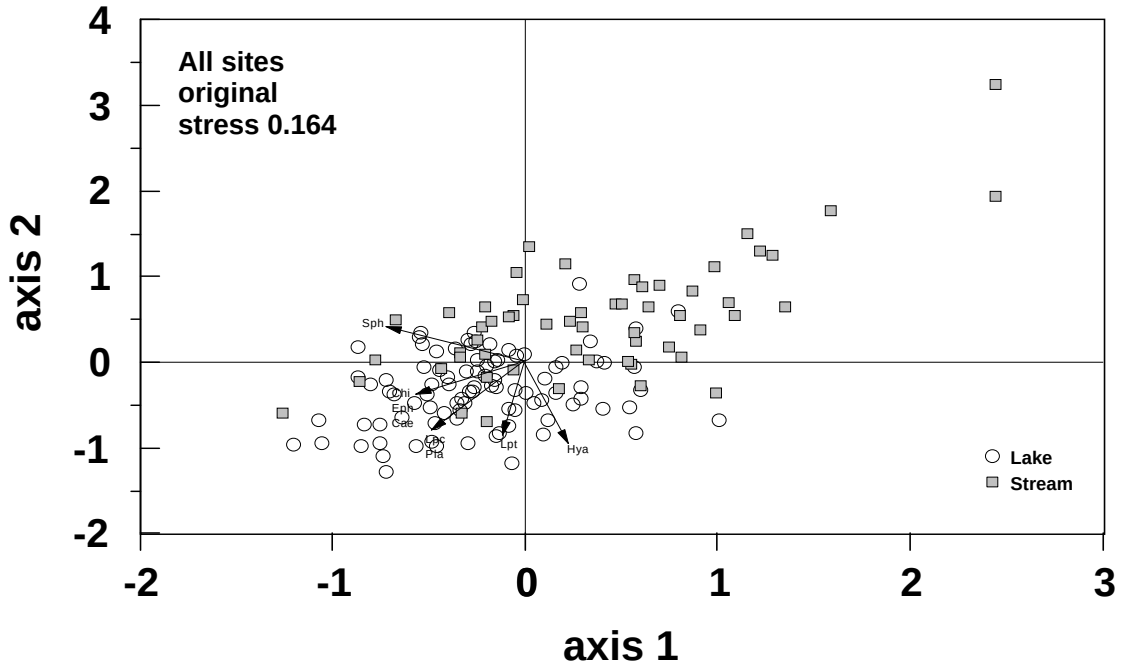

Figure 10. HMDS ordination of 149 lake and stream sites in northern Ontario from untransformed counts of invertebrate fauna. The first two dimensions (axes) are presented. Taxa most strongly associated with the ordination ($r > 0.400$; $P < 0.01$ from Monte Carlo randomization) are also shown. (Sph – Sphaeriidae, Chi – Chironomidae, Eph – Ephemerellidae, Cae – Caenidae, Lpt – Leptophlebiidae, Lpc – Leptoceridae, Pla - Planorbidae, Hya – Hyalellidae).

Although both untransformed data and transformed ($\log_{10} + 1$) data were used for classification and for model building, the performance of the models constructed from groups based on raw data was not as good as that with groups formed from transformed data. The transformed data captures some of the underlying richness in the data set and is not dominated by the most abundant taxa, primarily the Chironomidae. Therefore, only the results from the analysis of the transformed data are presented here.

3.3.1 Stream and Lakes Sites: Community Classification

The 149 lake and stream sites were classified using unweighted pair group mean averaging cluster analysis. A partial dendrogram is presented in Figure 11 that illustrates the first 12 groups formed from this analysis. Branches below this are not shown, as there is little further structure in the data.

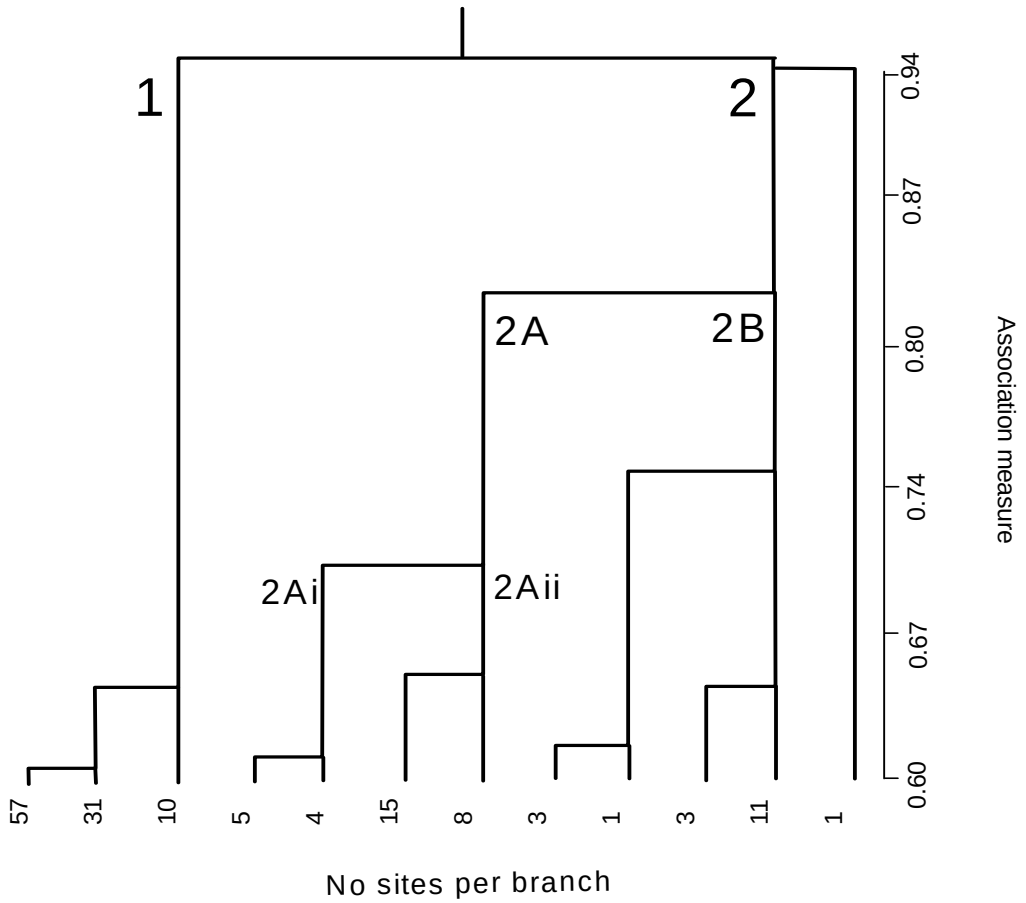

Figure 11. Dendrogram of transformed ($\log_{10} + 1$) invertebrate data showing the first 12 branches for 149 lake and stream sites for northern Ontario.

The sites first separate into two large groups of 98 (Group 1) and 51 (Group 2) sites. This is primarily a split between lake and stream sites; all but four of the lake sites are in this group, together with 10 of the 57 stream sites. Group 2 consists of 47 stream sites and four lake sites. A single stream site (RED12) separates immediately from Group 2. This single site has very low abundance (39 individuals) compared to the average of more than 7700 individuals per site. Although richness is within 1 SD of the average for the sites, 16 taxa versus the average of 22 taxa per site were found. Regardless, the fact that this single site forms a separate group means that it should not be included in model building. The next division is of Group 2 into two groups, a group of 32 sites (Group 2A) and 18 sites (Group 2B). Group 2A can be further split into a small group of nine sites (Group 2Ai) and a larger group of 23 sites (Group 2Aii). Beyond that, small groups of sites begin to form (1-5 sites) suggesting little further pattern in the data. It also clear from the dendrogram that there is far more homogeneity among the lake sites (Group 1) than the stream sites.

The validity of the groups is also illustrated in the HMDS ordination for which axes 1 and 3 are shown in Figure 12. Groups 1A and 1B are merged, but are indicated by different colours, and are clearly indistinguishable on these axes (or on axis 2 that is not shown). Group 1 is quite separate from Group 2. This separation appears to be on an axis associated with higher numbers of a faunal complex characterized by Hyalellidae and a strong association of Caenidae, Chironomidae and Naididae, and largely, but not exclusively, represents the lake sites.

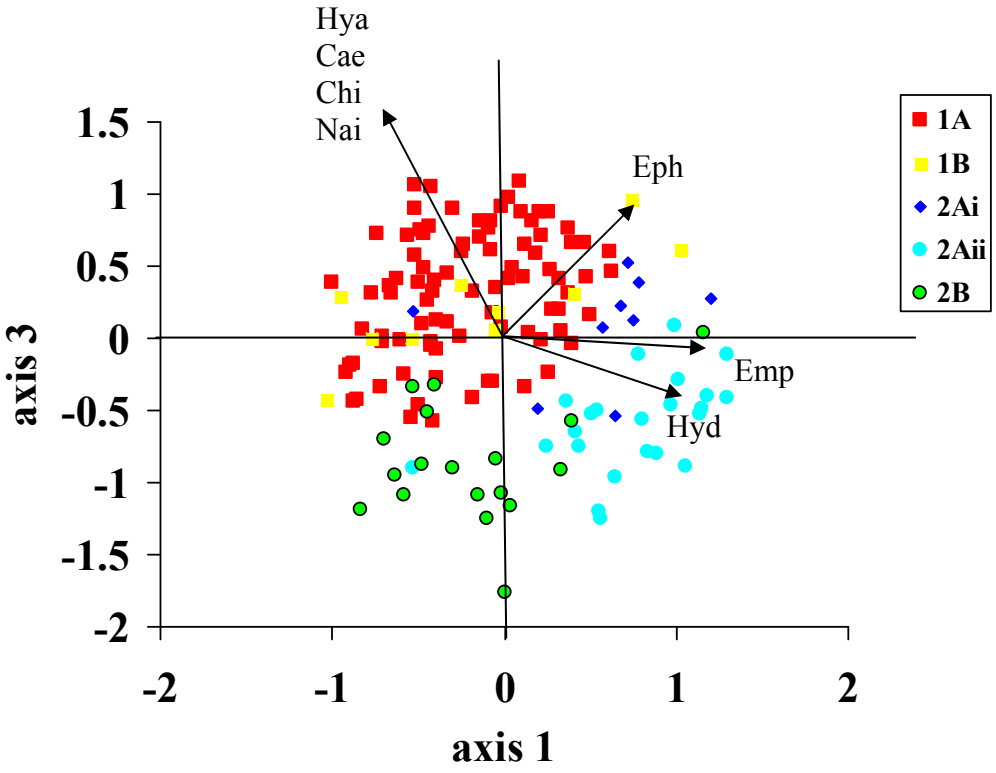

Figure 12. HMDS ordination of $\log_{10}$ +1 transformed counts of the invertebrate community from 149 lake and stream reference sites in northern Ontario. Five groups formed from classification analysis are shown. Vectors associated with taxa are also shown. (Hya – Hyalellidae, Cae – Caenidae, Chi – Chironomidae, Nai – Naididae, Eph – Ephemerellidae, Emp – Empididae, Hyd – Hydropsychidae).

Group 2B is also distinct and is characterized by intermediate abundance of Chironomidae and lower numbers of Ephemerellidae, Empididae and Hydropsychidae, as well as high numbers of Ceratopogonidae (not shown in Figure 12). The final division is of Group 2A. The two groups are primarily differentiated by lower numbers of the Hyalellidae complex, and higher numbers of caddisflies (Hydropsychidae and Hydroptilidae). As well as classifying sites based on the invertebrate fauna, classification based on regional and habitat typologies was also examined (Table 13).

Table 13. Global R values from ANOSIM of 149 lake and stream sites based on classification of region, lake vs. stream, and 2-5 groups from cluster analysis of the invertebrate fauna.

Typology	Region	Habitat	2 groups	3 groups	4 groups	5 groups
Global R	0.042	0.398	0.632	0.733	0.756	0.743

In ANOSIM, a low value for R indicates little difference among groups. In Table 13, it is evident that there are no differences among the regions. There is some difference between lake and stream habitat (R sign. $P < 0.01$), however, the classification of sites into four groups based on the invertebrate fauna provided the greatest discrimination among the groups (Table 13).

A summary of the four groups is presented in Table 14, and again it is clear that Group 1 is characterized by the Chironomidae-Hyalellidae complex. Groups 1 and 2Ai represent the most productive sites with average abundance of more than 9000 individuals per sample. Group 2Ai however is characterized by many more stream taxa. Groups 2Aii and 2B represent less productive sites, again largely streams. However Group 2Aii has the lowest overall abundance but the greatest number of taxa, and Group 2B is the least diverse group of sites.

Table 14. Summary of families characterizing four reference-site groups formed by transformed data of 149 lake and stream sites. The average abundance per group is shown together with the within group similarity (in parentheses). Those families whose abundance is in bold contributed to > 85% of the average similarity for that group (determined by SIMPER analysis).

	Group (within group percent similarity)			
	1 (43.1)	**2Ai (38.0)**	**2Aii (36.0)**	**2B (34.4)**
Abundance	9743	9329	2066	3246
Richness	22.4	23.3	25.1	15.3
Chironomidae	**4161**	**3129**	**619**	**1857**
Hyalellidae	**1471**	**411**	91	77
Caenidae	**626**	122	48	6
Naididae	**697**	76	21	62
Ceratopogonidae	**424**	**276**	**44**	**426**
Elmidae	135	**887**	**42**	16
Hydroptilidae	35	**331**	15	6
Ephemerellidae	212	**312**	**200**	6
Sphaeriidae	290	270	**144**	**217**
Hydropsychidae	0	941	**116**	3
Heptageniidae	100	239	**52**	6
Simuliidae	0	0	**134**	5
Leptophlebiidae	355	276	**47**	5
Enchytraeidae	162	29	**29**	209

3.3.2 Stream and Lake Site Models

The same process of constructing models using stepwise Multiple Discriminant Analysis was used with the combined lake and stream data set, as described above. However, a slightly different set of variables was used, as not all variables were common across the two data sets. In addition, a categorical variable for habitat was used to identify stream and lake sites. Up to five groups were examined as candidates for the model, although the division of Group 1 (Figures 11 and 12) is probably not justified. The variables considered for inclusion in the various models, the ones used by each model, and their performances are summarized in Table 15.

Table 15. Summary of Stream/Lake (Global) predictive models for various classifications (2-5 groups) of sites formed from transformed ($log_{10} + 1$) counts of the invertebrate community from 149 lake and stream sites using cross-validation (B backward stepwise; F forward stepwise). The dots represent the variables that are used to discriminate each model. The stars represent the recommended model.

	Number of Groups (stepwise direction)							
Variable in Model	**2 (B)**	**2(F)**	**3(B)**	**3(F)**	**4(B)**	**4(F)**	**5(B)**	**5(F)**
Latitude			●		●			
Longitude								
Altitude (ft)								
Lake or Stream	●	●	●	●	●	*	●	●
Ecoregion number			●		●			
Slope	●	●	●	●	●	*	●	●
Pool (Percent)	●	●	●	●	●	*	●	●
Rapid (Percent)	●	●	●	●	●	*	●	●
Riffle (Percent)	●	●	●	●	●	*	●	●
Run (Percent)								
Dominant Substrate						*		
2nd Dominant Substrate	●	●	●	●				
Embeddedness								
Surrounding Material								
Macrophyte								
Grasses								
Shrubs								
Deciduous trees								
Conifer trees								
Canopy				●			●	●
Alkalinity (mg/L)								
Hardness (mg/L)							●	
Conductivity (µS/cm)					●		●	●
Total Suspended Solids (mg/L)				●				
pH (pH)			●	●	●	*	●	●
Cyanide (CN(t)) (mg/L)								
Dissolved Organic Carbon (mg/L)								
Calcium (Ca) (mg/L)								
Potassium (K) (mg/L)	●	●	●					
Magnesium (Mg) (mg/L)			●	●	●			●
Sodium (Na) (mg/L)								
Ammonia (NH_3) (mg/L)								
Nitrate (NO_3) (mg/L)								
Sulfate (SO_4) (mg/L)							●	●
Total Phosphate (TP(Wat)) (mg/L)			●	●	●	*	●	●
Total variables	7	7	12	11	11	8	11	11
Classification success (crosval)	92	92	85	86	81	80	71	74

All the models performed well (Table 15), even the five-group model, even though the biological differences between Groups 1A and 1B, which created the fifth group are marginal. In fact, closer examination shows that although the re-classification success rate of the five group models are high (71 and 74%), the model performed poorly in discriminating Group 1B sites and none of the 10 sites in the group were predicted correctly to Group 1B. The high accuracy of the 5 group version of the model is derived from the accurate prediction of sites to other groups. Therefore, it is recommended that the four group model be used. The forward stepwise version is preferred, as it is more parsimonious, using fewer predictor variables (Table 15). It performs generally well among all groups (Table 16), particularly with the two groups with large numbers of sites. The model correctly predicts 80% of the reference sites.

Table 16. Cross-validation classification matrix for 149 sites.

Group from/to	**1**	**2Ai**	**2Aii**	**2B**	**% correct**
1	88	9	0	1	90
2Ai	2	6	1	0	67
2Aii	0	6	17	0	74
2B	3	3	5	7	39

3.4 Co-op Unit (2003) and Moose River (1999) sites

Data collected in 2003 study covered the region from Sudbury in the southeast to Timmins in the north and Red Lake to the west, and included 149 stream and lake reference sites that were used in the first round of model development. The methods used in this study were selected to be compatible with those used in a 1999 Moose River basin study in which 76 stream reference sites were sampled (Kilgour et al. 2000). The compatibility of the Moose River data with our study was examined by pooling all of the data, classifying references sites, and building models. Model suitability was then compared with the previous lake and stream models built for our study area.

3.4.1 Comparison of Moose River and Co-op Unit Data Sets

The two data sets were merged at the family level. The combined number of families recorded was 122. The Co-op Unit data set contained 105 families and the Moose River data set contained 93 families. Examination of regional differences between the two data sets using HMDS ordination (Figure 13) showed fairly conclusively that there was marked similarity between the invertebrate communities. While there is a strong gradient on the second and third ordination axes, which again partially discriminates the lake and stream sites in the Co-op Unit data, the Moose River sites are distributed along the entire gradient (Figure 13). This gradient describes changing abundance of Chironomidae, together with the strongly associated families of Planorbidae, Leptophlebiidae and Caenidae.

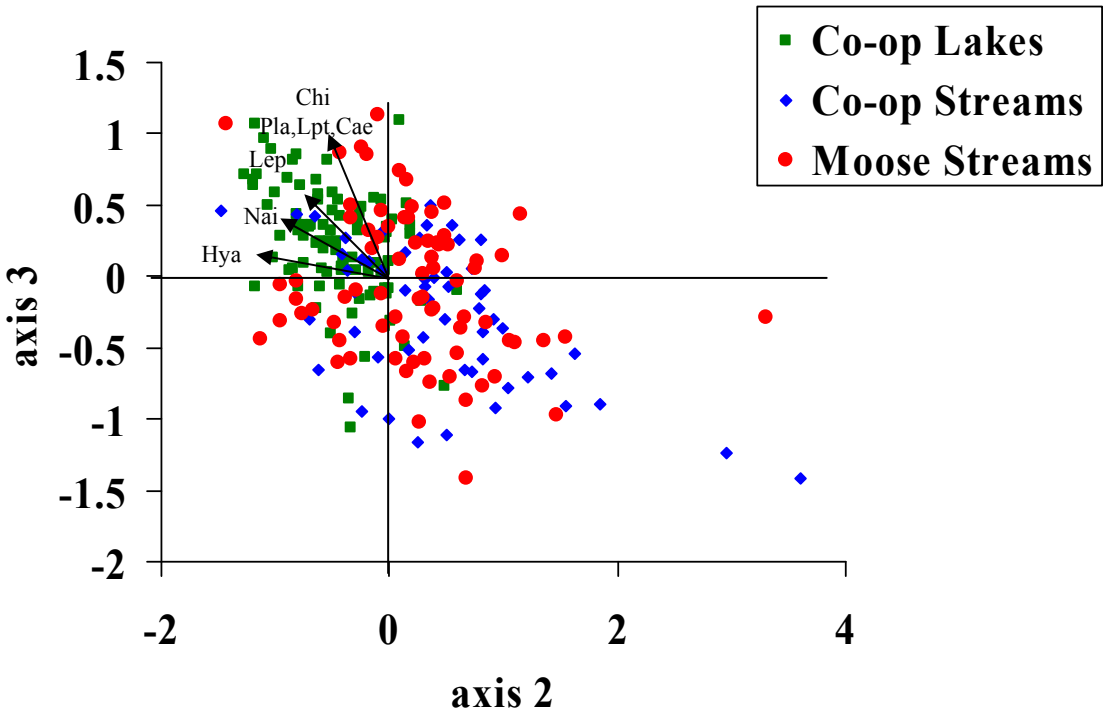

Figure 13. HMDS ordination of untransformed invertebrate community data from 225 reference sites in northern Ontario. Sites are identified as to the original data (Co-op Unit or Moose River) and by habitat (stream or lake). Vectors for the original taxa are also shown. (Chi – Chironomidae, Pla – Planorbidae. Lpt – Leptophlebiidae, Cae – Caenidae, Lep – Leptoceridae, Nai – Naididae, Hya – Hyalellidae).

The dominance of the Chironomidae in structuring the pattern in the communities is shown from the correlation of the families to the ordination axes (Table 17). The R value for the Chironomidae is almost twice that of the other families. To down weight the abundance of Chironomidae in characterizing the patterns in the region, data were transformed ($\log_{10} + 1$) to increase the contribution of other abundant and commonly occurring families. The effect of transformation is apparent in Table 17. Hydropsychidae, Ceratopogonidae and Heptageniidae are the three families now contributing to the observed patterns in the data set, and all the families listed in Table 17 contribute significantly ($P < 0.01$). The results of the transformed ordination are shown in Figure 14. Again, the Moose River basin sites are broadly distributed across the range of variation observed in the Co-op Unit stream and lake sites.

Table 17. Families contributing to HMDS ordinations of original and transformed (log_{10} + 1) data from northern Ontario invertebrate communities. Significant correlations (R) from Monte Carlo randomization with ordination axes are indicated in bold together with percent occurrence in original data sets.

Family	R value (orig.)	R value (trans.)	Co-op Unit (% occurrence)	Moose R (% occurrence)
Chironomidae	**0.707**	**0.466**	99.3	100
Hyalellidae	**0.474**	**0.590**	81.2	23.7
Caenidae	**0.473**	**0.571**	73.8	63.2
Leptophlebiidae	**0.467**	**0.396**	84.6	81.6
Planorbidae	**0.425**	**0.497**	64.4	36.8
Leptoceridae	**0.424**	**0.301**	61.7	65.8
Naididae	**0.402**	**0.492**	94.0	80.3
Hydropsychidae	0.177	**0.719**	15.4	38.2
Ceratopogonidae	**0.382**	**0.614**	95.3	44.7
Heptageniidae	**0.269**	**0.607**	47.0	63.2
Empididae	0.220	**0.566**	32.2	35.5
Lepidostomatidae	0.136	**0.548**	16.8	32.9
Baetidae	**0.334**	**0.546**	45.0	81.6
Unionicolidae	0.168	**0.524**	30.2	27.6
Enchytraeidae	0.236	**0.519**	71.8	18.4
Philopotamidae	0.178	**0.509**	11.4	21.1

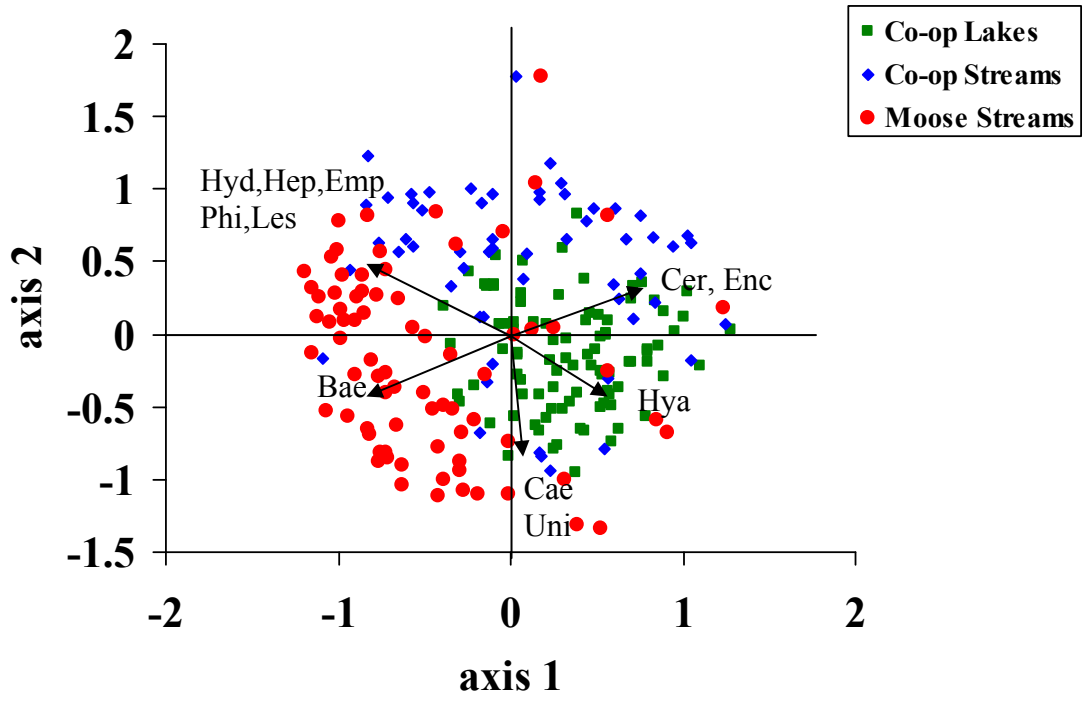

Figure 14. HMDS ordination of log_{10} +1 transformed invertebrate community counts data from 225 reference sites in northern Ontario. Vectors for the original taxa are also shown. (Hyd – Hydropsychidae, Hep – Heptageniidae, Emp – Empididae, Phil – Philopotamidae, Les – Lepidostomatidae, Bae – Baetidae, Lpt – Leptophlebiidae, Cae – Caenidae, Uni – Unionicolidae, Cer – Ceratopogonidae, Enc - Enchytraeidae, Hya – Hyalellidae).

There appear to be four distinct community types, a Hydropsychidae community with four associated families: Heptageniidae, Empididae, Philopotamidae and Lepidostomatidae. This community is negatively associated with Hyalellidae, and represents a stream to lake environmental gradient. A second orthogonal gradient is one from a Baetidae community which is negatively associated with a Ceratopogonidae-Enchytraeidae assemblage. This may represent a larger scale regional gradient, as the Baetidae community is more associated with sites from the Moose R. basin. A fifth possible assemblage is characterized by the Caenidae and Unionicolidae. This pattern, showing some large scale regional difference between the Moose River and Co-op Unit sites, provides an argument for using the larger data set as it will encompass any larger scale regional variation not included in the Co-op Unit data. To verify the significance of large scale regional variation ANOSIM analysis showed similar and substantial differences between the lake and stream sites for both the Co-op Unit and the Moose River data, however, there was little difference (low r) between the two sets of stream sites (Table 18).

Table 18. ANOSIM analysis of combined Moose River and Co-op Unit data using site habitat (lake vs. stream) and data (1999 Moose River vs. 2003 Co-op Unit) set as factors.

	R statistic
Global	0.332
Co-op lakes vs. Co-op streams	0.397
Co-op lakes vs. Moose R. streams	0.421
Co-op streams vs. Moose R. streams	0.087

From this preliminary analysis it was concluded that it was acceptable to merge the data sets and that a classification using $\log_{10} + 1$ transformed data would be more informative.

3.4.2 Moose River and Co-op data: Benthic Community Classification

The classification of the 225 sites is shown in Figure 15. The first division (Group 3) is the separation of 2 stream sites (RED12 and one from the Moose MR110), both with low abundance. A group of 49 sites is then formed (Group 2); these are all stream sites representing a combination of the Co-op Unit and Moose datasets. The group is characterized by fewer Chironomidae (Table 19), but is distinguished from the other groups of sites by the complex of families associated with lotic habitats that include the Hydropsychidae, Heptageniidae, Empididae, Philopotamidae and Lepidostomatidae (Table 19, Figure 16). The other Group (Group 1) then forms two groups,

Group 1A of 143 sites and Group 1B of 31 sites, the majority of which are streams from the Co-op Unit data set but includes 7 Moose stream sites and 3 lake sites.

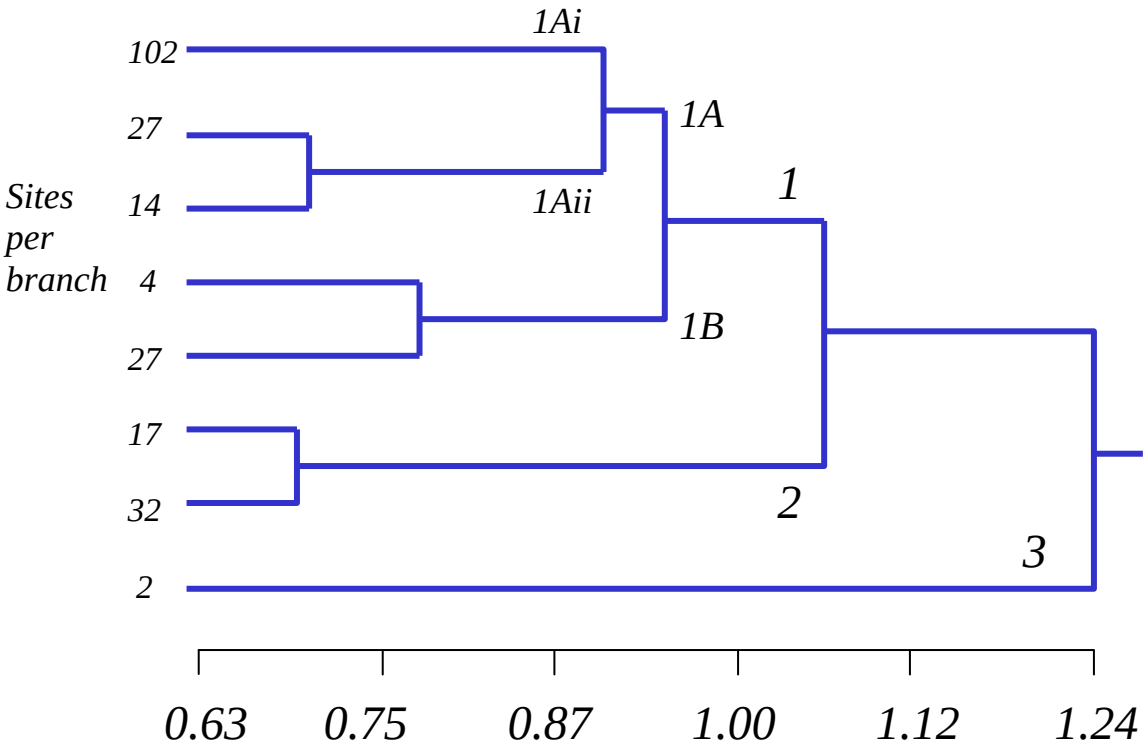

Figure 15. Dendrogram of log transformed invertebrate community counts data from 225 reference sites in northern Ontario. Eight groups are shown.

Table 19. Origin of sites and biological characteristics of 5 groups formed from 225 reference sites in northern Ontario.

	Group 1Ai	**Group 1Aii**	**Group 1B**	**Group 2**	**Group 3**
Co-op Streams (# sites)	11	0	21	24	1
Moose R Streams (# sites)	2	41	7	25	1
Co-op Lakes (# sites)	89	0	3	0	0
Abundance	9598	5765	3493	3785	87
Family Richness	22.8	17.9	16	24	12.5
Chironomidae	4040	1923	1761	1138	12
Hyalellidae	1443	155	63	58	0
Ceratopogonidae	417	29	293	44	2
Enchytraeidae	157	4	172	19	0
Caenidae	610	223	15	40	3
Unionicolidae	32	92	3	0	0
Hydropsychidae	0.2	14	5	300	2
Heptageniidae	100	70	6	131	7
Baetidae	118	244	9	197	0
Ephemerellidae	216	32	10	187	0
Valvatidae	109	88	45	12	37
Leptophlebiidae	360	359	121	134	0

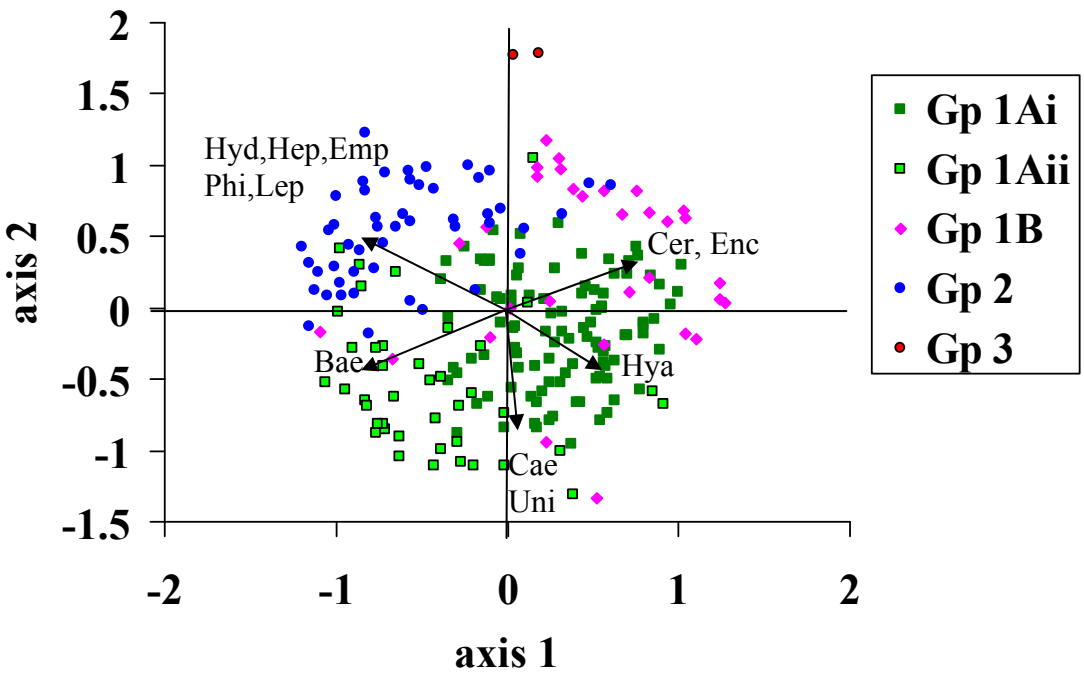

Figure 16. HMDS ordination of $\log_{10}$ +1 transformed invertebrate community counts data from 225 reference sites in northern Ontario. Sites are identified by groups formed from classification. Vectors for the original taxa are also shown. (Hyd – Hydropsychidae, Hep – Heptageniidae, Emp – Empididae, Phil – Philopotamidae, Les – Lepidostomatidae, Bae – Baetidae, Lpt – Leptophlebiidae, Cae – Caenidae, Uni – Unionicolidae, Cer – Ceratopogonidae, Enc - Enchytraeidae, Hya – Hyalellidae).

Group 1B has lower richness and is characterized by higher numbers of Ceratopogonidae and Enchytraeidae worms, both of which are associated with shallower, silty or marginal habitats. Group 1A then separates into two groups (Figure 15). Group 1Ai formed from 102 sites, 89 being lake sites, 11 are Co-op Unit streams and 2 are Moose River stream sites. This group has the highest average abundance, high diversity, and is characterized by Chironomidae and the Hyalellidae, as well as Ceratopogonidae. The latter two families are more typical of slow moving or still water habitats. Group 1Aii is made up exclusively of Moose River stream sites which have high abundance, but Chironomidae are not as numerically abundant at these sites and three mayfly families (Caenidae, Leptophlebiidae, Baetidae), typical of slow flow habitats and submergent vegetation, are found at more than 80% of these sites (Figure 16). The structure beyond this tends to be lost and smaller groups of sites are formed, therefore groups beyond this were not examined. To establish the significance of the groups ANOSIM was conducted using the five groups

(Table 20). All the pairwise comparisons were statistically significant, although r values were not very high for several tests, notably 1Aii vs. 1B and 1Aii vs. 2. Both Groups 1Aii and 1B are more variable, which may account for this lack of discrimination. However, these groups were deemed suitable to consider in model building.

Table 20. ANOSIM analysis of combined Moose and Co-op Unit data using 5 groups formed from transformed data as factors. All values are significant (using randomization) at the 0.02 level.

Global r = 0.45	**1Aii**	**1B**	**2**	**3**
1Ai	0.471	0.365	0.601	1.000
1Aii		0.161	0.213	0.964
1B			0.236	0.946
2				0.983

3.4.3 Moose River and Co-op data Models

For the complete Co-op Unit model a total of 35 habitat variables was available for consideration in the model. When combining the Co-op Unit and Moose River data sets only 24 variables were available for consideration. Others (altitude, slope, hardness, TSS, Na, Ca, K, Mg, SO_4, NO_3) were either not recorded or not sampled in the Moose River data set. Four variables had missing data with respectively, one (surrounding substrate, alkalinity), two (TP) and nine (DO) missing values. In these cases the average value for the variable was substituted. The data were transformed using a $\log_{10} + 1$ transformation to meet the assumptions of discriminant analysis. While the classification based on the transformed data captured more richness, to ensure that an optimum model was acquired, models were constructed for communities established for 2 – 5 groups formed from both original (not presented) and with transformed data.

Using from two to five groups (Table 19, Figures 15 and 16) the model accuracy ranged from 69 – 80% and with 11 variables was 75% accurate with the five groups (Table 21). Given the limited number of 24 variables, available from both data sets, of which 16 are categorical this shows remarkable robustness in the models. In fact, by removing the group of two sites (Group 3) and running a model with the remaining 4 groups the accuracy increases to 79% (Table 21), and performs well across all the groups (Table 22). Models constructed based on groups formed from original data did not perform as well. The best five group solution only predicted 45% of the sites correctly, compared to 75% with the model using groups derived from transformed counts.

Table 21. Summary of northern Ontario predictive models for 2-5 groups of sites formed from transformed counts of the invertebrate community at 225 sites using cross-validation (B backward stepwise; F forward stepwise). The dots represent the variables that are used to discriminate each model. The stars represent the best model.

	Number of Groups (stepwise direction)							
Variable in Model	2 (B/F)	3(B)	3(F)	4(B)	4(F)	5(B)	5(F)	4(B/F)*
Latitude				●		●		*
Longitude				●		●		*
Lake or Stream				●	●	●	●	*
Ecoregion number								
Pool (Percent)		●	●	●	●	●	●	*
Rapid (Percent)					●			
Riffle (Percent)		●	●	●	●	●	●	*
Run (Percent)	●	●	●	●	●	●	●	*
Dominant Substrate								
2nd Dominant Substrate								
Embeddedness				●	●	●	●	*
Surrounding Material								
Macrophyte			●	●	●	●	●	*
Canopy (presence/absence)	●		●	●	●	●	●	*
Grasses (presence/absence)								
Shrubs (presence/absence)								
Deciduous trees (presence/absence)								
Conifer trees (presence/absence)								
Canopy (percent)								
Alkalinity (mg/L)								
Conductivity (µS/cm)		●						
pH (pH)		●		●	●	●	●	*
Dissolve Oxygen (DO) (mg/L)								
Total Phosphorus (TP(Wat)) (mg/L)		●	●	●	●	●	●	*
Total variables	2	6	6	11	10	11	9	11
Classification (crossval. - % correct)	77	80	76	72	69	75	72	79
* the alternate 4B/F model is based on 4 groups with Group 3 excluded (model without TP had an accuracy of 78%)								

Table 22. DFA cross-validation classification results using four groups with 11 variables (see Table 20).

Group to/from	**1Ai**	**1Aii**	**1B**	**2**	**% correct**
1Ai	89	2	11	0	87
1Aii	0	36	2	3	88
1B	3	4	19	5	61
2	0	7	10	32	65
Overall					79

Based on these analyses showing similarities among sites at both the regional level and between habitats it is recommended that the model based on 223 reference sites, with the two Group 3 sites removed, be used in future analysis. The classification based on transformed data is preferred as it incorporates information from many more taxa, rather than being derived from changes in Chironomidae abundance, furthermore, the accuracy of the predictions from the model based on that classification is much higher.

3.3 Setting Assessment Targets and Site Evaluation

The discriminant model was run on the 223 reference sites (Group 3 excluded) using the classification and model for the transformed data (Table 21). This produced the classification described above and probabilities of group membership for each reference site. These probabilities were used to calculate an expected value for the number of taxa at every reference site. The Observed to Expected (O:E) ratio was then derived for each of the 223 reference sites. The histogram of the O:E ratios (Figure 17) has a mean O:E score for reference of 1.03 and the 10^{th} percentile is 0.33. Based on this distribution of reference site O:E ratios we can assign quality bands to the O:E ratio (Wright 1995). Sites with O:E ratios within the range of 90% of reference sites are considered to be equivalent to reference (Band 1, Table 23). Bands of equal size, in this case of 0.33 O:E units, then describe sites of various degrees of disturbance (Table 23).

3.3.1 RIVPACS

The RIVPACS assessment method calculates the probability of a taxon occurring at a site. The probability of occurrence of a given taxon is calculated as the sum of two probabilities: (1) the probability that a site belongs to the reference group(s) that the model predicted it to, and (2) the probability of occurrence of the taxon in the reference group (s) (Table 24).

The likelihood of a taxon occurring at a site is calculated from the product of the probability of a site being in a group multiplied by the occurrence of the taxa in the group. This provides the contribution for that group to the overall probability of a taxa occurring. So in this example, 0.009 x 49.0 = 0.44 is the contribution of Group 1Ai to the prediction of Baetidae (Table 24). The sum of the contributions for all the groups provides the summed probability of a taxon being present. So the overall probability of Baetidae at site MAR24 is 0.44 + 11.5 + 14.2 + 9.3 = 35.4. There is a 35.4% likelihood of Baetidae at this site. This calculation is conducted for every taxa at the site to determine the predicted community at a site. Because taxa with a probability of 50% or less could be there by chance, these are not used in the calculation of the expected number of taxa. Expected taxa are based on the sum of taxa with a more than 50% chance of occurrence: the $E_{50.}$ The actual expected number is always less than the number predicted because it is the combined probabilities. For three taxa with a P 0.75 of occurrence the sum of 2.25 means one would expect less than three taxa.

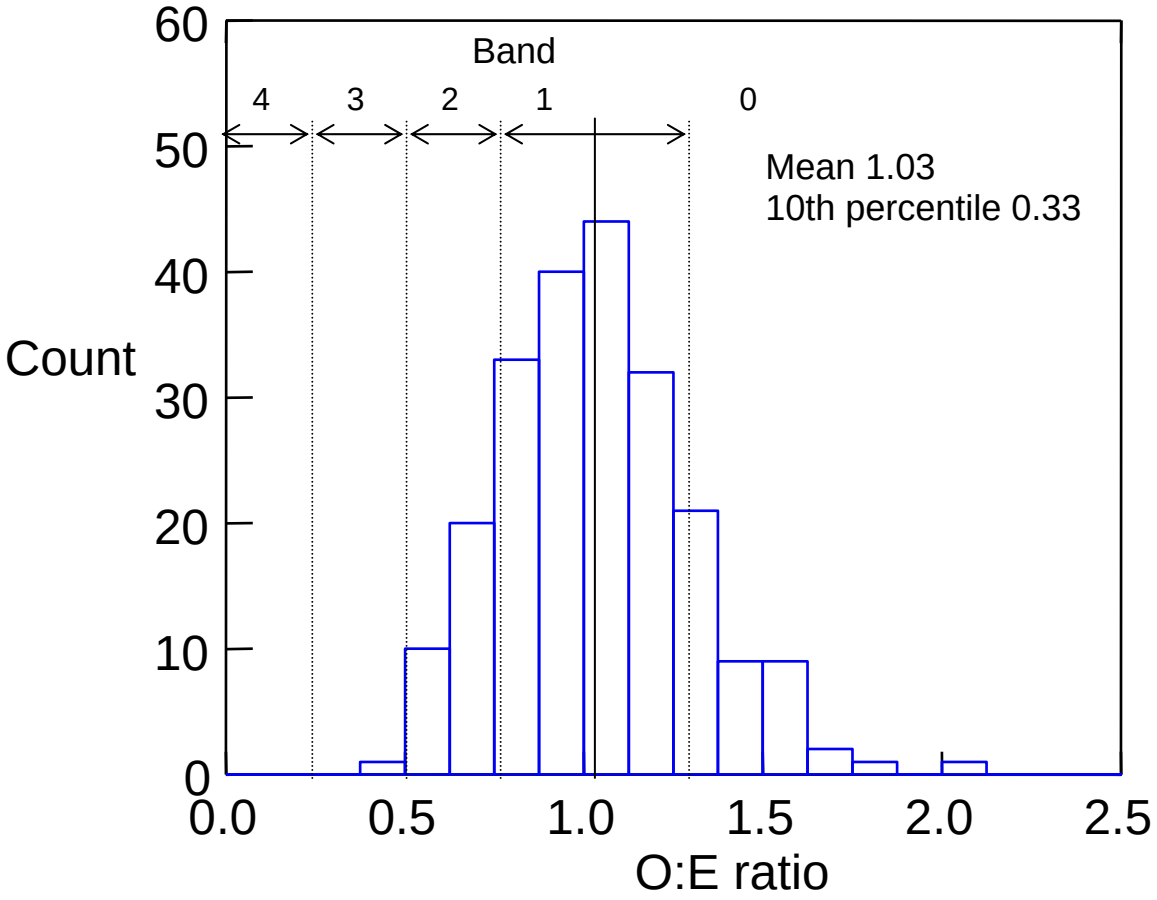

Figure 17. Histogram of Observed to Expected (O:E) ratios for 223 reference sites.

Table 23. Stress bands for Observed to Expected (O:E) ratio scores for northern Ontario.

Mean 1.03	**O:E score**	**Assessment**
Band 0	> 1.36	Either site of exceptional diversity or mild enrichment
Band 1	0.70 - 1.36	Equivalent to reference
Band 2	0.36 - 0.69	Mild disturbance
Band 3	0.02 - 0.35	Disturbed
Band 4	< 0.02	Severe disturbance

RIVPACS assessment of impairment is done by calculating the O:E ratios for all reference sites. A test site is defined as disturbed if the O:E ratio is less than 90% of the reference range. Stress Bands can be calculated by making them of equal size to the reference band, as described below.

Using this model and the O:E targets from the reference sites (Table 23) an assessment of all the test sites has been conducted (Appendix 4).

Table 24. Calculation of Baetidae being present at reference site MAR24. The probability is derived from MDA and the occurrence from the reference data set.

Site MAR24	**Group 1Ai**	**Group 1Aii**	**Group 1B**	**Group 2**	**Summed Prob. of taxa being present**
Probability	0.009	0.143	0.734	0.114	
% Occurrence in Group	49.0	80.5	19.4	81.6	
Combined probability	0.44	11.50	14.21	9.28	35.44

3.3.2 Test Site Assessment

Habitat data for the predictor models identified for the northern Ontario predictive model were assembled. Three categories of test site were assessed – a set of historically impacted sites, a set of urban sites, and a set of repeated reference sites. These latter sites provide some indication of the accuracy of the model, as they are by definition in reference condition. Three repeated reference sites (RED 57, RED42, RED 51) had missing data for pH and Total Phosphorus (TP) as water samples were lost. Average values for all reference sites were used to replace those missing values and $\log_{10} + 1$ values for each variable were used in the MDA model. The results of the discriminant analysis in predicting each of the test sites, the expected and observed number of taxa, and the RIVAPCS and BEAST assessment Bands are presented in detail in Appendix 4.

3.3.3 Historically Impacted Sites

Historically impacted sites were sampled from each of the Hemlo, Red Lake and Sudbury regions; 12 were lake sites and 6 were stream sites. The model certainty in assigning a prediction was high ($P > 0.9$) for all but one site (TEST18) which was still relatively high ($P = 0.729$). This indicates that one can be confident of the outcomes of the assessment. Particularly for the BEAST method, where the test site is compared against only those reference sites to which it is predicted as belonging.

Using RIVPACS, half the sites (9) were assessed as undisturbed, one site as either enriched or of special significance (Band 0), seven as mildly disturbed and one as disturbed (Site TEST17)

(Appendix 4, Table 25). Using BEAST ten sites were identified as undisturbed (Band 1), five as mildly disturbed (Band 2) and three as disturbed (Band 3) (Appendix 4). Concordance between the two assessment methods was high. Thirteen of the sites were assigned as having the same degree of disturbance (Appendix 4); three were assigned to a lower band (more stressed) by BEAST and two by RIVPACS. The historically impacted test lake sites ranged from a very diverse or enriched site (TEST05) with 32 taxa, to undisturbed sites Band 1 to mildly disturbed sites (Band 2) which were missing two thirds of the taxa. Finally one impacted site was the most removed from undisturbed (Band 3), with site TEST17 having only 5 taxa where almost 24 were expected (Table 25, Appendix 4). The RIVPACS assessment provides more information than simply the expected number of taxa; it also provides a list of taxa and their probability of occurrence. To illustrate this, three examples are described in more detail below in Table 26.

Table 25. Summary of RIVPACS assessment of historically impacted sites.

Habitat	Band 0	Band 1	Band 2	Band 3
Lakes	TEST05	TEST09(1) TEST09(2) TEST11 TEST12 TEST08	TEST01 TEST02 TEST03 TEST04 TEST06 TEST07	
Streams		TEST13 TEST14 TEST15 TEST18	TEST16	TEST17

Table 26. Biological assessment and attributes of three historically impacted test sites.

SITE	TEST15		TEST16		TEST17	
Predicted group	1B		1B		2	
Probability	1.000		0.939		0.982	
Expected taxa	16.0		16.4		23.9	
Observed taxa	14		6		5	
O:E	0.87		0.37		0.21	
RIVPACS band	1		2		3	
BEAST band	1		2		3	
Total abundance Ref (test) site	3493 (2627)		3493 (224)		3785 (228)	
	Predicted taxa > 0.7	P (count)	Predicted taxa > 0.7	P (count)	Predicted taxa > 0.7	P (count)
	Chironomidae	96.8 (1427)	Chironomidae	97.0 (188)	Chironomidae	99.9 (100)
	Ceratopogonidae	93.5 (73)	Ceratopogonidae	91.4 (16)	Sphaeriidae	95.8 (0)
	Sphaeriidae	90.3 (9)	Sphaeriidae	90.2 (0)	Ephemerellidae	90.7 (0)
	Naididae	87.1 (82)	Naididae	86.7 (0)	Heptageniidae	84.5 (0)
	Tubificidae	83.9 (418)	Tubificidae	82.2 (0)	Hydropsychidae	82.4 (0)
	Enchytraeidae	74.2 (0)	Enchytraeidae	71.6 (4)	Baetidae	80.5 (0)
					Naididae	79.7 (0)
					Leptophlebiidae	77.3 (0)
					Elmidae	72.7(4)
					Empididae	72.7 (52)

Site TEST15 was identified as a historically impacted stream which is fairly wide and deep. The site was predicted to Group 1B (Table 19, Figure 15) with a probability of membership of 1.000, 16 taxa were expected from the RIVPACS model and 14 were observed with an O:E ratio of 0.87 putting it in Band 1 (undisturbed). The taxa expected to occur are identified in Table 26, only those with a greater than 70% probability of occurrence are shown, as others are more affected by chance. Six families were expected to occur at greater than 70% level, and of these only one was missing the Enchytraeidae.

Using the BEAST assessment method, the Group 1B reference sites (open circles) and test sites (TEST15 and TEST16 – solid circles) are plotted in the same ordination space (Figure 18) together with probability ellipses (90, 99, and 99.9%) constructed around the reference sites only. Site TEST15 is clearly within the 90% ellipse and within the range of variation observed at reference sites (open circles). However, site TEST16 is outside the first (90%) probability ellipse but inside the 99.9% ellipse and is therefore assessed as Band 2 (the same as the RIVPACS assessment). It is also worth noting that one reference site is also in Band 2, this is because the bands are based on the distribution of the reference sites and by definition 10% of reference sites will fall outside the 90% ellipse. In fact, this describes (or defines) the Type 1 error rate, the probability of assessing a site as disturbed when it is not. Also from Figure 18, an additional attribute of the BEAST assessment method can be seen. In the site ordination space, the original taxa can be plotted (Figure 18), where the six taxa contributing most to the ordination pattern are shown. It is evident that the taxa associated with site TEST16 are reduced or has low numbers of Naididae and Chironomidae, because TEST16 is at the opposite end of the vector for those taxa.

From the RIVPACS assessment of site TEST16 only 6 taxa were observed from an expected total of 16.4 and the O:E ratio is 0.37, which puts the site just inside Band 2 (Table 23). Of the six taxa predicted to occur with a more than 70% probability, three were missing (Table 26). Chironomidae numbers (188 at TEST16) were an order of magnitude lower than at reference sites (average = 1768, Table 18). The relative strengths of the two assessment methods are captured in site TEST16. The BEAST method is responsive to the change in abundance of Chironomidae as shown in the ordination (Figure 18). The RIVPACS method picks out the loss of the Tubificidae and Sphaeriidae which have less influence in the BEAST assessment as they have relatively little influence on the ordination (r = 0.43 and 0.31 respectively). Site TEST17 was predicted as belonging to a different reference group (Group 2) and was assessed as disturbed (Band 3) by both assessment methods, only 5 of 23.9 expected taxa were collected. Of the ten taxa with a more

than 70% probability of occurring, only three (Chironomidae, Elmidae and Empididae) were present and overall abundance was low. The BEAST method also shows the site to be very different from the reference sites (Figure 18).

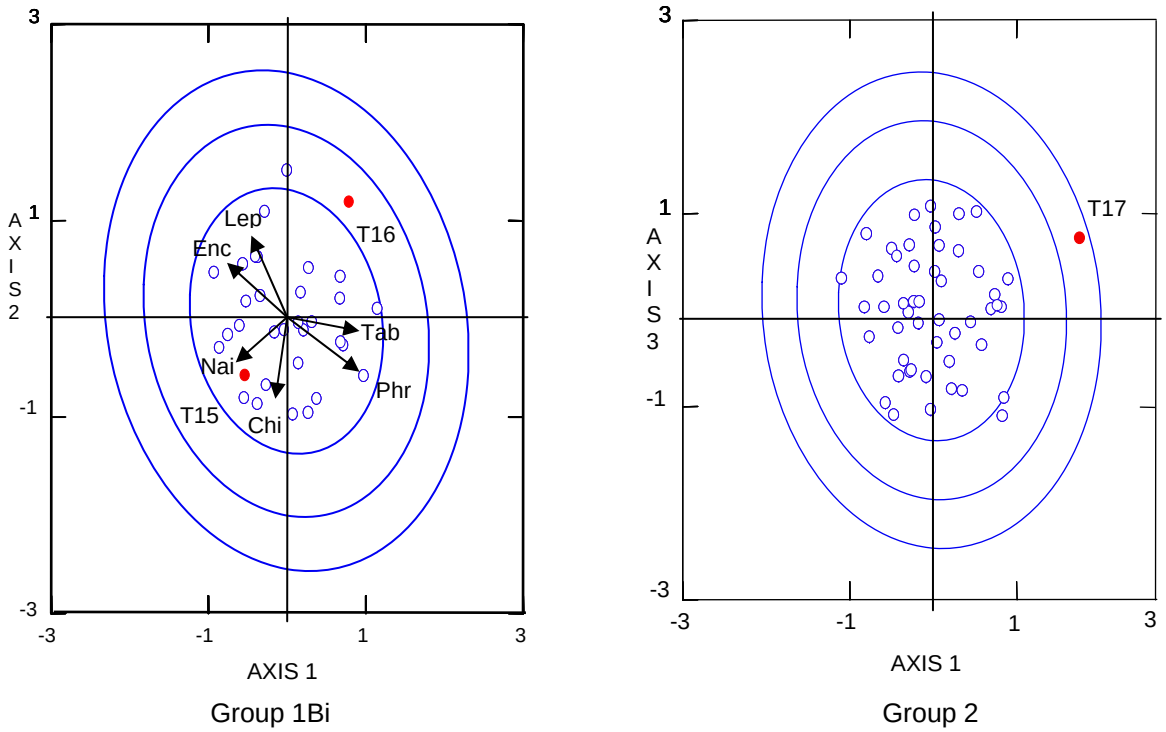

Figure 18. Ordination of reference and three test sites (T15, T16 and T17), using the BEAST assessment method. Important taxa are shown (Lep – Leptoceridae, Enc - Enchytraeidae, Nai – Naididae, Chi – Chironomidae, Tab – Tabanidae, Phr – Phryganeidae)

3.3.4 Urban Streams

Nineteen urban stream sites were sampled in Sudbury and Timmins. The rationale for sampling these sites was to assess the role of confounding influences of urban disturbance, as EEM near and far-field sites are sometimes located within urban areas. Of the 19 sites 15 were assessed as equivalent to reference by RIVAPCS, 11 were assessed by BEAST as equivalent to reference, and 10 of these were assessed as reference by both methods (Appendix 4). The two sites assigned to Band 0 by RIVPACS (Sites TEST23 and TEST29) were taxonomically richer than expected, which could be a result of mild organic enrichment; increased abundance and richness are often the first response to enrichment. In comparison, BEAST assessed Site TEST29 as undisturbed and Site TEST23 as mildly disturbed. BEAST assessed eight sites as mildly disturbed in comparison to the two sites assessed as mildly disturbed by RIVPACS, likely because of changes

in abundance to which the BEAST is more sensitive as opposed to actual loss of taxa. The fact that the majority of urban sites are assessed as equivalent to reference or mildly disturbed indicates that confounding factors are not a major concern using these assessment methods. Where urban impact and mining impact sites are assessed as disturbed the nature of the community response can differentiate between the stress factors. Two urban sites (TEST22, TEST23) assessed as Band 2 by BEAST show a very different community response (Figure 19) compared to a historically impacted site (TEST16), as indicated by their very different positions in ordination space. Site TEST23 has increased abundance of Chironomidae, Naididae and Tubificidae, a typical enrichment response, whereas site TEST22 has slightly reduced overall abundance, and reduced numbers of Chironomidae and Tubificidae but richness is reasonably high. Site TEST16 (historically impacted from mining) has low richness and abundance, and shows a general response more typical of that to toxicants.

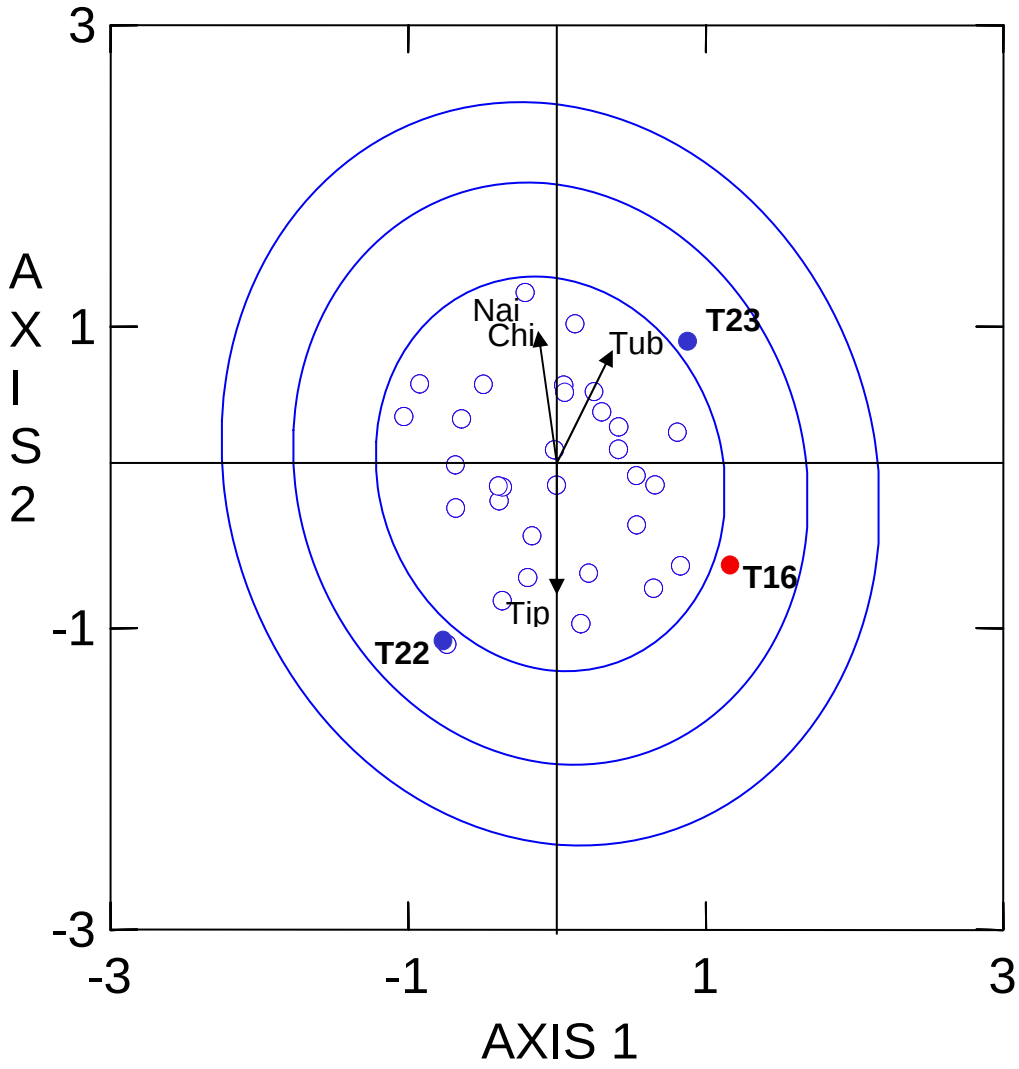

Figure 19. Differential response at mining (T16) and urban sites (T22, T23). (Nai – Naididae, Chi – Chironomidae, Tub – Tubificidae, Tip – Tipulidae)

3.3.5 Repeated Reference Sites

At six lake sites, a second set of samples was taken. The first set of samples was used in building the model. These samples were used to test the accuracy of the model in assessing sites. As the sites are all in reference condition by definition, if the model is good it will assess the second set of independent samples as being in Band 1. The results of both the RIVPACS and BEAST assessment are presented in Table 27.

Table 27. Assessment of repeated samples from reference sites using BEAST and RIVPACS assessment methods.

	Band 1	**Band 2**	**Band 3**	**Band 4**
RIVPACS	HEM52a MAR21a RED29a LSP01a TIM52a	RED08a		
BEAST	HEM52a MAR21a RED29a LSP01a TIM52a RED08a			

The BEAST assigned all sites to Band 1 (equivalent to reference), whereas RIVPACS assigned five of the six sites to Band 1. Overall, 11 of 12 assessments were correct, the one site that was assigned to Band 2 (RED08a) had an O:E score of 0.62 and is in the upper end of Band 2 (Table 24). In fact, this shows that, from these data, the site assessments are quite accurate (91.7%). Other authors have showed similarly high accuracy for these two assessment methods (Reynoldson et al. 1997, Mazor et al. submitted).

4.0 CONCLUSIONS

From these data analyses, we concluded that a combined reference data set developed from the Hemlo, Red Lake, Sudbury, Timmins, and the Moose River catchments provides a valid RCA model that can be used to conduct EEM assessments for the benthic invertebrate community. Using four invertebrate community reference groups, the model has high predictive accuracy (79 %) with easily measured habitat variables. The model can discriminate impaired sites from reference sites, and the assessments have high accuracy. While independent stream and lake models were constructed from data collected in 2003, it is suggested that the larger model that combines 2003 lake and stream sites with data from the Moose River in 1999 be used. The rationale for using the larger model is that it includes a greater number of reference sites, and therefore provides a more comprehensive description of reference conditions. It encompasses a larger geographic area and therefore has broader coverage and application for assessing test sites. Finally, it expands the time period describing reference conditions and therefore reduces concerns regarding temporal variability.

5.0 BENTHIC INFORMATION SYSTEM FOR REFERENCE CONDITIONS (BIRC) DATABASE

CABIN is an integrated national network for the collection, management, assessment, and distribution of bioassessment data. The network has been designed to help governments and community groups monitor cumulative aquatic environmental change over large parts of the country (Reynoldson, et al. 1999) using a standardized approach. Further, CABIN provides guidance, support, and long-term stability through the dissemination, and/or provision of standards, tools and information. In order to support the needs of the CABIN program, a web portal has been developed to provide access to standardized sampling protocols, data entry and management, analytical and reporting tools. The portal is currently hosted by the Ecological Monitoring and Assessment Network (EMAN) at http://cabin.cciw.ca. At the heart of the CABIN portal is the Benthic Information System for Reference Conditions (BIRC). Designed around the standard CABIN sampling protocols for benthic invertebrates, BIRC provides a secure online data management system. As a web-based interface, CABIN's data management system has several benefits over traditional software packages. Most notably, users are not required to maintain software on a local computer, meaning installation and software upgrades are not required. Further, since it operates through standard web browsers, data maintained within the system can be accessed securely from any computer connected to the internet. In addition to data management, BIRC provides a suite of analytical tools which are continually being developed and updated.

Based on multivariate analytical techniques developed by the NWRI, the primary study design is RCA. RCA permits the prediction of community structure of benthic invertebrates in reference (or minimally impacted) sites using simple habitat and water quality descriptors stored in the database. Reference data are then combined with test (potentially contaminated) site data entered by the user, and Discriminant and Principal Component Analysis is used to assess the relative health of the benthic community. Results are displayed both in tabular and graphic formats, identifying Unstressed, Potentially Stressed, Stressed, or Severely Stressed sites. As a complement to RCA analysis, 41 of the most commonly used benthic metrics can also be computed. Metrics include measures of richness, abundance, functional types, and biotic indices.

6.0 FUTURE WORK

To address questions pertaining to the differences between urban and pristine reference sites, whether or not substantial temporal changes occur, method comparability (MOE versus CABIN), and to fill gaps outlined during the initial model development, ten streams (eight reference and two historically impacted) and six lakes were re-sampled during the fall of 2004 and an additional 53 new sites were sampled from the Sudbury and Hemlo areas.

The variables used to select the new reference sites were extracted from the groups that were under-represented in the initial model development (Table 28). Site selection was prioritized according to the following groups:

- Group 2 - consisted of a major group of sites representing diverse but less abundant communities, and formed the first separation. Sites assigned to this group were larger, low slope streams with lower alkalinity, hardness and conductivity.
- Group 1Bi - consisted of only three sites which represented a community with the highest abundance. These were also larger streams with lower slopes, but higher alkalinity and hardness.
- Group 1Bii - had eleven sites with smaller, low slope streams with lower alkalinity and hardness.
- Group 1Ai - had twelve sites with medium size streams, high slope, lower conductivity and alkalinity.

Table 28. Parameters used to select 2004 stream sampling sites. The table illustrates the minimum, maximum and mean value for each parameter in each group.

	Group				
Variable	**1Ai**	**1Aii**	**1Bi**	**1Bii**	**2**
Latitude	46.233 – 51.167 48.579	46.283-51.129 49.037	48.538-51.189 49.446	46.408-51.093 49.791	46.662 –51.164 49.041
Longitude	80.957 - 94.009 85.962	81.036-94.448 87.471	85.088-93.661 87.973	81.524-93.803 90.367	81.218 – 93.735 87.503
Altitude (f.a.s.l.)	686-1280 1096.8	692-1345 1124.8	1174-1450 1317.7	850-1352 1130.0	938-1220 1112.6
ER number	90 – 98 95.3	90-98 94.4	90-96 94.0	90-98 92.091	90-97 94.0
Bankfull Width (m)	0.9 - 50 12.6	1.0-75.0 10.4	9.8-27.4 16.3	1.7-31.0 8.9	3.7 – 65 24.4
Slope (m/km)	0.64-20.33 9.06	1.9-23.8 7.1	1.2-9.8 4.7	0.1-16.4 5.7	1.00-8.27 3.93
Alkalinity (mg/L)	6-99 31.6	3-106 40.0	39-76 62.0	16-59 37.1	9.0-80.0 29.8
Conductivity (µS/cm)	31.3-302 90.6	25.9-295 112.1	96.8-185.9 155.0	78.7-142.5 104.7	43.9-185.9 80.4
Hardness (mg/L)	7.7-114 38.5	10-128 47.1	49.4-80.7 69.2	10.6-70.4 44.5	14.4-83.7 38.4

The parameters outlined in the models covered broad ranges which made site selection in 2004 difficult. All efforts were made to collect data from under-represented groups but without knowing the water chemistry of the new sites *a priori* it was difficult to ensure that the proper habitats/sites were sampled. In addition to re-sampling the inflow and outflow of six lakes, sites distant from inflows or outflows, representing three habitats types (rock, organic debris, and vegetation) were also sampled in triplicate at each lake. The purpose of sampling these other locations and habitats was to determine whether benthic communities, in lentic situations, are clearly linked to location or habitat type. The northern Ontario RCA Study is continuing and samples collected during the fall of 2004 are being used to enhance the BIRC database and will be used in the second round of model development scheduled for the spring of 2005. With the addition of more land-use and watershed data, we hope to improve the database and refine the models to better discriminate between groups. A large database of this nature, which contains physical, chemical and biological information on aquatic systems, will also be extremely useful in answering important science questions in future efforts to understand the effects of various environmental stressors on ecosystem health.

7.0 LITERATURE CITED

Altman, D.G. 1978. Plotting probability ellipses. Applied Statistics 27: 347-349.

Bailey, R.C., Norris, R.H., and Reynoldson, T.B. 2004. Bioassessment of freshwater ecosystems using the reference condition approach. Kluwer Academic Publishers, Boston. 184pp.

Belbin, L. and MacDonald, C. 1993. Comparing three classification strategies for use in ecology. Journal of Vegetation Science 4: 341-348.

David, S.M., Somers, K.M., Reid, R.A., Hall, R.J., and Girard, R.E. 1998. Sampling protocols for the rapid bioassessment of streams and lakes using benthic macroinvertebrates, Second edition. Ontario Ministry of the Environment Report, Dorset, Ontario. 29 pp. + appendices.

Davies, P.E. 2000. Development of a national river bioassessment system (AUSRIVAS) in Australia. In *Assessing the biological quality of freshwaters. RIVPACS and other techniques*, (editors J.F. Wright, D.W. Sutcliffe and M.T. Furse). Freshwater Biological Association, Ambleside, UK. Chapter 8, pp. 113-124.

Faith, D.P., Minchin, P.R.and Belbin, L. 1987. Compositional dissimilarity as a robust measure of ecological distance. Vegetatio 69: 57-68.

Hawkins, C.P. and Norris, R.H. 2000. Effects of taxonomic resolution and use of subsets of the fauna on the performance of RIVPACS-type models. In *Assessing the biological quality of freshwaters. RIVPACS and other techniques*, (editors J.F. Wright, D.W. Sutcliffe and M.T. Furse). Freshwater Biological Association, Ambleside, UK. Chapter 15, pp. 217-228.

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

Jackson, D.A. 1993. Multivariate analysis of benthic invertebrate communities: the implications of choosing particular data standardizations, measures of association, and ordination methods. Hydrobiologia 268: 9-26.

Kilgour, B.W., Dixon, D.G., Bailey, R.C., and Reynoldson, T.B. 2000. Development of a reference condition approach (RCA) to assess fish and benthic communities and in-stream habitat attributes in the Moose River Basin. Report submitted to Forest Ecosystem Science Co-operative. 37 pp. + appendices.

Marchant, R. 1989. A subsampler for samples of benthic invertebrates. Australian Society of Limnology Bulletin 12: 49-52.

Mazor, R.D., Rosenberg D. M., Reynoldson, T. B and Resh, V. H. (submitted). Effects of assemblage, classification method, and assessment method on performance of biomonitoring models. Canadian Journal of Fisheries and Aquatic Sciences.

Moss, D, Furse, M.T., Wright, J.F. and Armitage, P.D. 1987. The prediction of macro-invertebrate fauna of unpolluted running-water sites in Great Britain using environmental data. Freshwater Biology 17: 41-52.

Owen, J.G. and Chmielewski, M.A. 1985. On canonical variates ellipses and the construction of confidence ellipses in systematic studies. Systematic Zoology 34: 366-374.

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

Reynoldson, T.B., Bailey, R.C., Day, K.E., and Norris, R.H. 1995. Biological guidelines for freshwater sediment based on Benthic Assessment of SedimenT (the BEAST) using a multivariate approach for predicting biological state. Australian Journal of Ecology 20: 198-219.

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

Reynoldson, T.B., Logan, C., Pascoe, T., and Thompson, S.P. 1999. CABIN (Canadian Aquatic Biomonitoring Network) invertebrate biomonitoring field and laboratory manual. NWRI Report No. 99-211.

Reynoldson, T.B., Day, K.E. and Pascoe, T. 2000. The development of the BEAST: a predictive approach for assessing sediment quality in the North American Great Lakes. In *Assessing the biological quality of freshwaters. RIVPACS and other techniques*, (editors J.F. Wright, D.W. Sutcliffe and M.T. Furse). Freshwater Biological Association, Ambleside, UK. Chapter 11, pp. 165-180.

Reynoldson, T.B., and Wright, J.F. 2000. The reference condition: problems and solutions. *In* Assessing the biological quality of fresh waters. RIVPACS and other techniques. (editors Wright, J.F., Sutcliffe, D.W. and Furse, M.T.) Freshwater Biological Association, Ambleside, U.K. pp. 293-303.

Reynoldson, T.B., Rosenberg, D.M., and Resh, V.H. 2001. Design of a regional benthic biomonitoring programme: Development of predictive models of invertebrate community structure using multivariate and multimetric approaches in the Fraser River catchment, British Columbia. Canadian Journal of Fisheries and Aquatic Sciences 58: 1395-1410.

Rosenberg, D.M., Reynoldson, T.B., Reece, P.F., and Resh, V.H. 1998. Design of a regional biomonitoring programme for the Fraser River Basin. FRAP Report Number: DOE FRAP 1998-32.

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

Underwood, A.J. 1997. Experiments in Ecology. Their Logical Design and Interpretation Using Analysis of Variance. Cambridge University Press, Cambridge.

U.S. EPA. 1995. Volunteer stream monitoring: a methods manual. EPA 841 D95-001 April. Office of Wetlands, Oceans and Watersheds, 4503F Washington DC 20460.

Wright, J.F., Moss, D., Armitage, P.D., and Furse, M.T. 1984. A preliminary classification of running-water sites in Great Britain based on macroinvertebrate species and the prediction of community type using environmental data. Freshwater Biology 14: 221-256.

Wright, J.F., Furse, M.T., Armitage, P.D., and Moss, D. 1991. Testing and further development of RIVPACS. Institute of Freshwater Ecology interim report to the National Rivers Authority, December 1991. IFE, Wareham, Dorset, U.K.

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

Wright, J.F. 2000. An introduction to RIVPACS. In *Assessing the biological quality of freshwaters. RIVPACS and other techniques*, (editors J.F. Wright, D.W. Sutcliffe and M.T. Furse). Freshwater Biological Association, Ambleside, UK. Chapter 1, pp. 1-24.

8.0 APPENDICES

1) Appendix 1. RCA 2003 – Sample site locations

Waterbody	Site Code	Type		UTM Zone	Easting	Northing
Sudbury						
Junction Creek	JC2	imp urban stream	Test	17T	501564	5152124
Junction Creek	Nolin McNeil	imp urban stream	Test	17T	498887	5149504
Whitson River	WHR50	urban stream	Test	17T	496123	5160623
Vemillion River	VER01	ref stream		17T	503276	5185081
Onaping River	ONP01	urban stream	Test	17T	469973	5167826
Frood-Stobie 1	FST1	imp urban stream	Test	17T	501625	5152891
Robinson L. Trib.	ROB01	urban stream	Test	17T	496987	5144514
Romford Creek	RMF01	urban stream	Test	17T	511631	5147909
Rapid River	RAP01	ref stream		17T	497286	5174546
Broder Lake 23 - Outflow	PAN01	ref lake		17T	503790	5138470
Broder Lake 23 - Inflow	PAN02	ref lake		17T	502660	5137908
Lily Creek	RAM01	urban stream	Test	17T	499043	5146112
Dowes Lake - Inflow	USR01	small lake		17T	455252	5177189
Dowes Lake - Outflow	USR02	small lake		17T	454246	5176994
Vemillion River	UPV01	urban stream	Test	17T	505539	5172055
Whitson Trib. Seguin Rd.	WHR02	urban stream	Test	17T	489759	5161659
Whitson River - Cecile	WHR01	urban stream	Test	17T	499350	5163196
McKenzie Creek	WHR03	urban stream	Test	17T	479129	5157435
Whitson Trib. Errington	WHR04	urban stream	Test	17T	484780	5157901
Island Creek	ILD01	ref stream		17T	481228	5169267
Sandcherry Creek	ILD02	ref stream		17T	483307	5167582
Bull Lake - Outflow	LSP01	small lake	Test	17T	407329	5143538
Sable Trib.	LSP02	ref stream		17T	409537	5126112
Sable River - Massey	LSP03	ref stream		17T	416306	5120523
Beaudin Creek	LSR06	ref stream		17T	441766	5126996
Ministic Creek Trib.	LSR07	ref stream		17T	457789	5139512
Cameron Creek	LSR08	ref stream		17T	459743	5147909
Windy L Trib.	OPR02	ref stream		17T	463717	5161691
Sawmill L. Trib.	CMR01	ref stream		17T	459083	5160908
High Cliff Creek	OPR01	urban stream	Test	17T	469132	5164070
Whitewater L. Trib	WHR05	urban stream	Test	17T	493001	5154485
Low Water Creek	USR12	ref stream		17T	448474	5217750
Halfway Lake - Inflow	USR13	large lake		17T	451831	5194605
Halfway Lake - Outflow	USR11	large lake		17T	450784	5192320
Upper Marquette Lake	USR10	small lake		17T	445108	5223286
Whitefish Creek	PAN03	urban stream	Test	17T	504571	5141765
Unnamed Creek - Hwy 69	PAN04	urban stream	Test	17T	502589	5142266
Red Lake						
Ranger L. Trib. 1	RLT01	ref stream		15U	469284	5649089
Hist. Imp. Stream 1	HIS01	imp stream	Test	15U	442566	5651443
Ranger L. Trib. 2	RLT02	ref stream		15U	466447	5643453
Dixie Creek	DIX01	ref stream		15U	463516	5632140
Leano Lake - Inflow	LEA01	large lake		15U	398917	5629650

Appendix 1 (continued)

Waterbody	Site Code	Type		UTM Zone	Easting	Northing
Leano Lake - Outflow	LEA02	large lake		15U	397399	5628270
Leano Lake - Inflow	LEA03	large lake		15U	398878	5627573
Leano L. Trib	LET01	ref stream		15U	397987	5629201
Golden Creek	Golden Cr	ref stream		15U	453811	5671086
Chikuni River	Chikuni R.	ref stream		15U	444700	5668650
Balmer L. Trib	Balmer L. Trib	ref stream		15U	450851	5658491
Abalard Creek	Abalard Cr	ref stream		15U	449500	5662572
Red Lake - Inflow	RDL15	ref stream		15U	426924	5656558
Unnamed Creek	RDL16	ref stream		15U	429298	5657855
Stone Lake - Inflow	STO01	large lake		15U	444765	5637349
Stone Lake - Outflow	STO02	large lake		15U	448263	5638426
Red Lake - Pipestone Bay	RED01	imp large lake	Test	15U	413595	5658249
Red Lake - Pipestone Bay	RED02	large lake		15U	410722	5659474
Red Lake - St. Paul Bay	RED03	large lake		15U	435400	5652119
Red Lake - Outflow	RED04	large lake		15U	462359	5640262
Lund Lake - Outflow	RED05	large lake		15U	409195	5661881
Chikuni Trib.	RED06	ref stream		15U	466284	5641318
Unnamed Lake - Inflow	RED07	small lake		15U	410153	5662091
Tote Lake - Outflow	RED08	small lake		15U	463151	5629324
Peisk Creek	RED09	ref stream		15U	412804	5664873
Dixie Cr. Trib	RED10	ref stream		15U	444167	5628897
Unnamed Lake - Outflow	RED11	small lake		15U	416134	5666227
Dixie L. Trib	RED12	ref stream		15U	448260	5631413
Unnamed Lake - Inflow	RED13	small lake		15U	416791	5665789
Unnamed Lake - Inflow	RED14	small lake		15U	412329	5640230
Unnamed Lake - Outflow	RED16	small lake		15U	411796	5640139
Johnson Lake - Outflow	RED18	small lake		15U	416179	5649644
Unnamed Lake - Outflow	RED19	small lake		15U	429097	5659823
Johnson Lake - Inflow	RED20	small lake		15U	415255	5649405
Unnamed Lake - Outflow	RED21	small lake		15U	429728	5660279
Stupeck Outflow Trib.	RED22	ref stream		15U	417637	5649504
Unnamed Lake - Outflow	RED23	small lake		15U	432745	5662467
Unnamed Lake - Outflow	RED24	small lake		15U	409668	5641188
Unnamed Lake - Outflow	RED25	small lake		15U	410025	5661667
Hugh's Lake Trib	RED26	ref stream		15U	415508	5646687
Whiteass Lake - Inflow	RED27	large lake		15U	425612	5670971
Unnamed Lake - Outflow	RED28	small lake		15U	422750	5649796
Whiteass Lake - Outflow	RED29	large lake		15U	427292	5672595
Tack Lake - Inflow	RED30	large lake		15U	426585	5643355
Corallen Lake - Inflow	RED31	large lake		15U	431663	5666673
Tack Lake - Outflow	RED32	large lake		15U	426853	5644095
Corallen Lake - Outflow	RED33	large lake		15U	434930	5668760
Rowan Lake - Outflow	RED35	imp small lake	Test	15U	421740	5657362
Balmer Lake - Outflow	RED37	imp large lake	Test	15U	448802	5657198

Appendix 1 (continued)

Waterbody	Site Code	Type		UTM Zone	Easting	Northing
Balmer Lake - Inflow	RED39	imp large lake	Test	15U	449480	5658475
Flat Lake - Inflow	RED40	large lake		15U	432457	5646897
Sidace Lake - Outflow	RED41	large lake		15U	463642	5683155
Flat Lake - Outflow	RED42	large lake		15U	434185	5645932
Sidace Lake - Inflow	RED43	large lake		15U	465762	5685210
Snib Lake - Outflow	RED44	large lake		15U	439904	5650564
Pindar Cr. Trib.	RED45	ref stream		15U	453421	5668336
Snib Lake - Inflow	RED46	large lake		15U	438536	5649640
Unnamed Lake - Inflow	RED47	imp small lake	Test	15U	442348	5650727
Sully Creek	RED48	ref stream		15U	446113	5646557
Unnamed Lake - Outflow	RED49	imp small lake	Test	15U	442436	5651161
Killoran Creek	RED50	ref stream		15U	445142	5648262
Spiers Lake - Outflow	RED51	small lake		15U	431688	5647992
Upper Medicine Stone Lake	RED52	large lake		15U	429338	5639931
Spiers Lake - Inflow	RED53	small lake		15U	431447	5648475
Upper Medicine Stone Lake	RED54	large lake		15U	426484	5640688
Derlak Lake - Outflow	RED55	imp small lake	Test	15U	437161	5648351
Russett lake - Inflow	RED56	large lake		15U	434662	5646663
Halfway Creek	RED57	imp ref stream	Test	15U	438757	5647926
Russett lake - Outflow	RED58	large lake		15U	434006	5647271
Unnamed Creek	RED59	ref stream		15U	443776	5660406
Red Lake - Impacted site	RED61	imp large lake	Test	15U	442382	5658862
Dixie Lake - Inflow	RED69	large lake		15U	449027	5630038
Dixie Lake - Outflow	RED71	large lake		15U	452201	5631480
Bug River	RED73	ref stream		15U	446490	5641615
Balmer Creek	RED75	imp stream	Test	15U	448713	5654983
Hemlo Area						
Melgund Lake - Inflow	HEM02	small lake		16U	562505	5392981
Melgund Lake - Outflow	HEM04	small lake		16U	561367	5392781
Unnamed Lake - Outflow	HEM06	small lake		16U	604092	5391135
Unnamed Creek	HEM08	ref stream		16U	604164	5391527
Rein Creek	HEM10	ref stream		16U	591109	5374299
Binekan Creek	HEM12	ref stream		16U	592229	5377698
Triplet L. Stream	HEM14	ref stream		16U	596307	5381937
Oskabukuta River	HEM16	ref stream		16U	598005	5383222
Unnamed Lake - Inflow	HEM18	small lake		16U	599182	5383140
Unnamed Lake - Outflow	HEM20	small lake		16U	599045	5383185
East Bremner River	HEM22	ref stream		16U	617600	5375418
Animons Lake - Outflow	HEM24	small lake		16U	605400	5387483
Oskabukuta Trib	HEM26	ref stream		16U	602204	5382915
Nursery Lake - Inflow	HEM28	small lake		16U	607711	5395620
Nursery Lake - Outflow	HEM30	small lake		16U	607519	5396117
Unnamed Lake - Inflow	HEM32	small lake		16U	604136	5380148
Unnamed Lake - Outflow	HEM34	small lake		16U	604160	5379558

Appendix 1 (continued)

Waterbody	Site Code	Type		UTM Zone	Easting	Northing
White River	HEM36	ref stream		16U	602258	5394284
Deer Lake - Inflow	HEM38	small lake		16U	599261	5394945
Deer Lake - Outflow	HEM40	small lake		16U	599364	5394911
Unnamed Lake - Inflow	HEM44	small lake		16U	613901	5374060
Unnamed Lake - Outflow	HEM46	small lake		16U	613566	5373843
Unnamed Creek	HEM48	ref stream		16U	614476	5383128
Whitefish Lake - Inflow	HEM50	small lake		16U	623675	5373880
Whitefish Lake - Outflow	HEM52	small lake		16U	624422	5374748
Whitehead's Creek	HEM54	ref stream		16U	625224	5379656
North Crocker Lake - Inflow	HEM56	small lake		16U	619181	5388185
North Crocker Lake - Outflo	HEM58	small lake		16U	619176	5387885
Frank Lake - Inflow	HEM60	imp small lake	Test	16U	584775	5389604
Frank Lake - Outflow	HEM62	imp small lake	Test	16U	586057	5389325
Lim Lake - Inflow	HEM64	imp small lake	Test	16U	580853	5391375
Lim Lake - Outflow	HEM66	imp small lake	Test	16U	581604	5390698
Wabikoba Creek	HEM68	ref stream		16U	588604	5396630
Tukanee Lake - Inflow	HEM70	large lake		16U	630714	5389243
Tukanee Lake - Outflow	HEM72	large lake		16U	630677	5387719
Mink Lake Creek	HEM74	ref stream		16U	640969	5385887
Sagina Lake - Inflow	HEM80	small lake		16U	638441	5370642
Sagina Lake - Outflow	HEM82	small lake		16U	637747	5370267
Marion Lake - Outflow	HEM84	small lake		16U	639601	5366846
Dorothy Creek	MAR01	ref stream		16U	578125	5422345
Mickey Creek	MAR03	ref stream		16U	578721	5420975
Mobert Creek	MAR05	ref stream		16U	584650	5415659
Summers Lake - Inflow	MAR07	small lake		16U	584090	5409492
Summers Lake - Outflow	MAR09	small lake		16U	584003	5409348
Dead Otter Lake - Inflow	MAR11	large lake		16U	588316	5413327
Dead Otter Lake - Outflow	MAR13	large lake		16U	589702	5413264
Lunny Lake - Outflow	MAR14	small lake		16U	586403	5410325
Amwri Lake - Inflow	MAR15	small lake		16U	583115	5406029
Amwri Lake - Outflow	MAR16	small lake		16U	582291	5405021
Summers Creek	MAR17	ref stream		16U	583828	5408577
Barbara Stream	MAR18	ref stream		16U	584513	5405302
Solong Lake - Inflow	MAR20	small lake		16U	589949	5407147
Solong Lake - Outflow	MAR21	small lake		16U	591554	5406742
Wabigoon Lake - Outflow	MAR22	large lake		16U	590902	5402967
Wabigoon Lake - Inflow	MAR23	large lake		16U	592564	5405496
Philips Creek	MAR24	ref stream		16U	583290	5398562
Pakoamaga Lake - Inflow	MAR25	small lake		16U	613874	5396728
Pakoamaga Lake - Outflow	MAR26	small lake		16U	612815	5396530
Lurch Creek	MAR28	ref stream		16U	613059	5402038
Skewer Lake - Inflow	MAR29	small lake		16U	614359	5402073
Skewer Lake - Outflow	MAR30	small lake		16U	613490	5401631

Appendix 1 (continued)

Waterbody	Site Code	Type		UTM Zone	Easting	Northing
Blotter Trib	MAR31	ref stream		16U	617800	5411600
Kwinkwaga River	MAR32	ref stream		16U	624435	5413402
Bouchard Lake - Inflow	MAR33	large lake		16U	614239	5405561
Bouchard Lake - Outflow	MAR34	large lake		16U	611479	5403420
Unnamed Lake - Inflow	MAR35	small lake		16U	627312	5420262
Unnamed Lake - Outflow	MAR36	small lake		16U	627787	5420652
Kenshoe Lake - Inflow	MAR37	small lake		16U	625367	5415278
Kenshoe Lake - Outflow	MAR38	small lake		16U	624272	5414091
Mikano Lake - Outflow	MAR39	large lake		16U	621974	5413151
Mikano Lake - Inflow	MAR40	large lake		16U	621165	5412224
Unnamed Lake - Inflow	MAR41	small lake		16U	631238	5423089
Unnamed Lake - Outflow	MAR42	small lake		16U	630570	5422615
Unnamed Lake - Outflow	MAR43	small lake		16U	634718	5421474
Davis Lake - Inflow	MAR44	small lake		16U	633165	5415643
Davis Lake - Outflow	MAR45	small lake		16U	634167	5416704
Unnamed Creek	MAR46	ref stream		16U	635405	5419636
Davis Creek	MAR47	ref stream		16U	638891	5415550
Plate Creek	MAR48	ref stream		16U	639950	5411963
Depew River	MAR50	ref stream		16U	634955	5377627
Timmins						
Scott Lake - Inflow	TIM02	small lake		17U	481525	5330739
Scott Lake - Outflow	TIM04	small lake		17U	481061	5331955
McArthur Lake - Outflow	TIM06	large lake		17U	483301	5338557
McArthur Lake - Inflow	TIM08	large lake		17U	484217	5338152
Clear Lake - Outflow	TIM10	small lake		17U	478546	5350283
Mountjoy River	TIM12	urban stream	Test	17U	474192	5368486
Hillary Lake - Outflow	TIM14	small lake		17U	441729	5342229
Mattagami River	TIM16	urban stream	Test	17U	473626	5370602
South Porcupine River	TIM18	urban stream	Test	17U	484624	5369112
Reid Lake - Outflow	TIM20	small lake		17U	475656	5361555
Levalley Lake - Inflow	TIM50	small lake		17U	441412	5358749
Levalley Lake - Outflow	TIM51	small lake		17U	440337	5357942
Jowsey Lake - Outflow	TIM52	large lake		17U	445685	5357435
Jowsey Lake - Inflow	TIM53	large lake		17U	446020	5358358
Kamiskotia Lake - Outflow	TIM54	large lake		17U	454573	5380182
Kamiskotia Lake - Inflow	TIM55	large lake		17U	453657	5378455
Mattagami Tributary	TIM56	urban stream	Test	17U	469996	5375380
Town Creek	TIM57	urban stream	Test	17U	474795	5369804
Kenogamissi Lake - Outflow	TIM58	large lake		17U	463600	5354199
Kenogamissi Lake - Inflow	TIM59	large lake		17U	464143	5349360
Big Water Lake - Outflow	TIM60	large lake		17U	580230	5384900
Big Water Lake - Inflow	TIM61	large lake		17U	478272	5384330

2) Appendix 2. RCA Lake Sampling Protocol

September 2003
Revised August 2004

Prepared by
Jennifer Davidson

Documenting Discharge Lakes

Prior to sampling reference lakes, review information on the mining discharge lakes (preferably have photos and/or visit sites) so that reference sites can be sampled most effectively. Each reference lake will have two sites- one inflow and one outflow with a water sample taken at each site.

Discharge characteristics to consider:

1. How large is the receiving lake? Where possible, the size of reference lakes should be similar to receiving lakes.
2. Are the discharge points at inflows or outflows? This will help determine appropriate QA/QC sites (see below).
3. What are the physical characteristics at the inflows and outflows?
    - Substrate
    - Flow state & width- large river, small creek, marsh etc.
    - If discharge is upstream of a beaver dam (i.e. into a beaver pond), the reference site should generally be sampled *downstream* of the dam, as the pond will have steep drop-offs, unusual community etc.

Sampling Reference Lakes

Choosing the two sites:

1. Determine inflows and outflows using maps.

2. Survey the lake by boat etc. to see which inflow (if there are several) most resembles that of the discharge lake- this should be your sampling site. The outflow chosen should be the major one for the lake.

3. At each inflow and outflow site choose three, 50m segments of shore which encompass the habitats of the site area (they do not have to be physically identical segments). For outflows, it may make sense to sample both sides. In general, select sampling areas that best characterize the individual lake- consider where any effects the lake may have on invertebrate communities would most likely be detected.

4. Do not follow the inflow/outflow too far in such that you are sampling a well-defined channel (flow should be minimal). Sample the transition zone between the actual lake shore and the beginning of the channel, or, if the inflow/outflow begins abruptly, the segments will mostly be lake shore.

5. **QA/QC**: If the lake is selected for QA/QC, one of the lake sites (either inflow or outflow) is sampled in duplicate (each of the three replicates is sampled in duplicate). Choose the inflow if the relevant discharge site is at an inflow, outflow if discharge is at outflow.

6. On rare occasion reference lakes may not have appropriate inflow and outflow sites, or these sites may be impossible to sample properly. If this is the case, another nearby reference lake should be sampled instead.

Site documentation & taking the invertebrate samples:

1. Record general information (see attached field sheet): lake or stream type (small vs. large lake, urban vs. reference stream), waterbody name, site code, team, sampling date, inflow or outflow, geographic description, ecoregion, UTM zone, Easting, Northing, and elevation.

2. Take site photos:
   1 field sheet
   1 in toward inflow/outflow
   1 out toward lake
   1 across inflow/outflow to other side
   1 dry riparian edge/exposed bar (preferably with ruler)
   1 water surface with substrate visible.

3. Water chemistry: wade into 1m depth at an undisturbed area representative of the whole site, rinse 1 large water container three times then use it to carefully fill one blue-lid cyanide jar as close to top without overflowing (contains preservative). Refill large water container and seal with no air bubbles.
   If substrate is predominantly fine/organic (finer than sand), take 1 particle size sample by scraping container on the bottom, pack substrate after each scoop until container is full.
   Take 1 Mercury sample and 1 Ra226 sample **only if site is specially selected for this** (note: all mercury and radium sites are also QA/QC sites, but that not all QA/QC sites require mercury and radium samples).

4. In another undisturbed area representative of the whole site, use DO meter to record: time, air temperature, water temperature, DO, and % Saturation at 0.5 m depth.

The three 50m segments are labelled as A, B or C and are individually sampled and documented on field sheets in order to keep track of within-site variability. Use the following procedure for A, B and C as per the field sheet.

5. Estimate canopy coverage over shore to 1 metre deep.

6. Estimate macrophyte coverage from shore to 1 metre deep.

7. Record types of riparian vegetation present from shore to approximately 10m up bank.

8. Invertebrate sampling: wade out to 1m depth with sweep net and bucket. Kick and sweep along a transect perpendicular to shore, emptying net into bucket whenever it becomes full of debris. Continue kicking and sweeping additional transects down the shoreline (2 metres apart), filling the same bucket, until 10 minutes is up.

9. Transfer a random portion of the bucket sample by hand into a 1.5L white snow jar. Do not fill snow jars beyond ¾ (leave room for adding 10% buffered formalin). If you think it is possible that there are fewer than 100 animals in the sample (e.g. it's all sand and you don't see any bugs), fill a second snow jar with another random portion from the bucket and label the jars 1 and 2 of 2. It's more efficient for you and the lab to get only one snow jar per segment, however the sample is useless if it has less than 100 animals, so use your discretion (note: for reference, check with the lab about your samples to get an idea of what is actually being found in them).

10. Record the time taken for the transect(s), the # transects, the distance out, and the # of snow jars filled.

11. **If the site is QA/QC**, **take a second 10 minute kick & sweep transect(s) at each of A, B and C.** There should be 6 invertebrate samples for a QA/QC site.

12. Record the 1st & 2nd dominant substrate, the surrounding material and embeddedness (take the average of 10 rocks).

13. Record the substrate dimensions (length, width, and height) from 10 randomly selected rocks within the sampling area.

3) Appendix 3. Northern Ontario test sites, indicating predicted group, probability of Group membership, expected taxa, observed taxa and O:E ratio.

			Probability of belonging to Group								
Test site category	**Site**	**Pred Group**	**1Ai**	**1Aii**	**1B**	**2**	**Expected no. taxa**	**Observed no. taxa**	**O:E ratio**	**RIVPACS Band**	**BEAST Band**
Historically Impacted Small Lakes	Test01	1Ai	0.963	0.000	0.008	0.028	22.8	11	0.48	2	2
Historically Impacted Small Lakes	Test02	1Ai	0.861	0.001	0.018	0.120	22.8	12	0.53	2	1
Historically Impacted Small Lakes	Test03	1Ai	0.848	0.029	0.012	0.110	22.7	14	0.62	2	2
Historically Impacted Small Lakes	Test04	1Ai	0.939	0.000	0.023	0.038	22.7	11	0.49	2	3
Historically Impacted Large Lakes	Test05	1Ai	0.940	0.001	0.005	0.055	22.8	32	1.40	0	1
Historically Impacted Large Lakes	Test06	1Ai	0.937	0.000	0.063	0.000	22.4	8	0.36	2	2
Historically Impacted Large Lakes	Test07	1Ai	0.965	0.000	0.035	0.000	22.5	9	0.40	2	3
Historically Impacted Large Lakes	Test08	1Ai	0.986	0.000	0.007	0.007	22.7	18	0.79	1	1
Historically Impacted Small Lakes	Test09a	1Ai	0.989	0.000	0.003	0.008	22.8	26	1.14	1	1
Historically Impacted Small Lakes	Test09b	1Ai	0.989	0.000	0.003	0.008	22.8	28	1.23	1	1
Historically Impacted Small Lakes	Test11	1Ai	0.995	0.000	0.001	0.003	22.8	25	1.10	1	1
Historically Impacted Small Lakes	Test12	1Ai	0.990	0.000	0.010	0.000	22.7	17	0.75	1	2
Historically Impacted Streams	Test13	1B	0.004	0.003	0.967	0.025	16.3	15	0.92	1	1
Historically Impacted Streams	Test14	1B	0.010	0.013	0.910	0.067	16.7	15	0.90	1	1
Historically Impacted Streams	Test15	1B	0.000	0.000	1.000	0.000	16.0	14	0.87	1	1
Historically Impacted Streams	Test16	1B	0.002	0.018	0.939	0.041	16.4	6	0.37	2	2
Historically Impacted Streams	Test17	2	0.000	0.000	0.018	0.982	23.9	5	0.21	3	3
Historically Impacted Streams	Test18	1B	0.000	0.004	0.729	0.268	18.2	16	0.88	1	1
Urban Streams	Test19	2	0.000	0.000	0.012	0.988	23.9	17	0.71	1	2
Urban Streams	Test20	1B	0.000	0.000	0.755	0.245	18.0	14	0.78	1	1
Urban Streams	Test21	1B	0.000	0.060	0.570	0.370	19.1	18	0.94	1	1
Urban Streams	Test22	1B	0.001	0.002	0.708	0.289	18.3	11	0.60	2	2
Urban Streams	Test23	1B	0.003	0.046	0.899	0.051	16.5	26	1.57	0	2
Urban Streams	Test24	1B	0.005	0.099	0.767	0.128	17.3	20	1.16	1	2
Urban Streams	Test25	1B	0.001	0.005	0.944	0.049	16.4	16	0.97	1	1
Urban Streams	Test26	2	0.000	0.002	0.177	0.821	22.6	19	0.84	1	1
Urban Streams	Test27	1B	0.017	0.000	0.982	0.001	16.2	21	1.30	1	1

Appendix 3 cont.

			Probability of belonging to Group								
Test site category	**Site**	**Pred Group**	**1Ai**	**1Aii**	**1B**	**2**	**Expected no. taxa**	**Observed no. taxa**	**O:E ratio**	**RIVPACS Band**	BEAST Band
Urban Streams	Test28	1B	0.002	0.002	0.954	0.042	16.4	12	0.73	1	1
Urban Streams	Test29	1B	0.002	0.031	0.939	0.028	16.3	24	1.47	0	1
Urban Streams	Test30	1B	0.000	0.001	0.997	0.002	16.1	21	1.31	1	2
Urban Streams	Test31	1B	0.003	0.058	0.875	0.064	16.7	16	0.96	1	2
Urban Streams	Test32	1B	0.003	0.012	0.864	0.120	17.0	22	1.29	1	1
Urban Streams	Test33	1Aii	0.003	0.437	0.173	0.387	19.9	27	1.35	1	2
Urban Streams	Test34	1Aii	0.003	0.439	0.202	0.356	19.7	21	1.07	1	1
Urban Streams	Test35	2	0.002	0.290	0.287	0.421	19.9	17	0.85	1	1
Urban Streams	Test36	1B	0.003	0.324	0.512	0.161	17.9	22	1.23	1	1
Urban Streams	Test37	1Aii	0.003	0.445	0.114	0.437	20.3	11	0.54	2	2
Hemlo Small Reference Lakes	HEM52a	1Ai	0.959	0.000	0.007	0.033	22.8	21	0.92	1	1
Hemlo Small Reference Lakes	MAR21a	1Ai	0.974	0.000	0.004	0.023	22.8	25	1.10	1	1
Red Lake Large Reference Lakes	RED29a	1Ai	0.986	0.000	0.006	0.008	22.8	21	0.92	1	1
Red Lake Small Reference Lakes	RED08a	1Ai	0.990	0.000	0.005	0.005	22.8	14	0.62	2	1
Sudbury Small Reference Lakes	LSP01a	1Ai	0.900	0.002	0.016	0.082	22.8	26	1.14	1	1
Timmins Large Reference Lakes	TIM52a	1Ai	0.943	0.003	0.004	0.050	22.8	30	1.32	1	1