Geographic Extension of Benthic Invertebrate RCA Bioassessments: How Far Can We Go?

by

Nicole-Marie Novodvorsky

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science (MSc) in Biology

The Faculty of Graduate Studies
Laurentian University
Sudbury, Ontario, Canada

THESIS DEFENCE COMMITTEE/COMITÉ DE SOUTENANCE DE THÈSE
Laurentian Université/Université Laurentienne
Faculty of Graduate Studies/Faculté des études supérieures

Title of Thesis Titre de la thèse	Geographic extension of benthic invertebrate RCA predictive models: How far can we go?
Name of Candidate Nom du candidat	Novodvorsky, Nicole-Marie
Degree Diplôme	Master of Science
Department/Program Département/Programme	Biology
Date of Defence Date de la soutenance	January 15, 2015

APPROVED/APPROUVÉ

Thesis Examiners/Examinateurs de thèse:

Dr. John Bailey
(Co-supervisor/Co-directeur de thèse)

Dr. John Gunn
(Co-supervisor/Co-directeur de thèse)

Dr. Trefor Reynoldson
(Committee member/Membre du comité)

Dr. Robert Brua
(External Examiner/Examinateur externe)

Approved for the Faculty of Graduate Studies
Approuvé pour la Faculté des études supérieures
Dr. David Lesbarrères
M. David Lesbarrères
Acting Dean, Faculty of Graduate Studies
Doyen intérimaire, Faculté des études supérieures

ACCESSIBILITY CLAUSE AND PERMISSION TO USE

Abstract

The management of aquatic ecosystems is important to preserve the ecosystem services provided to humanity. The development of environmental assessment has allowed the management and therefore protection of these important resources. Reference Condition Approach (RCA) bioassessments using benthic macroinvertebrates as indicators are common tools that provide a means of determining impairment of stream sites by comparing exposed test sites to relatively unexposed reference sites. RCA predictive models are commonly developed at the scale of drainage basin, ecoregion, or political region (i.e. United Kingdom or Australian state), and test site assessment is restricted within the spatial boundaries of the model. If test site assessment can be applied outside the spatial scope of the model, insofar that the environmental characteristics are similar, it would reduce extensive sampling i.e. remote northern locations and time-consuming development of numerous models. The overall goal of my study was to assess whether a predictive model applied across a larger spatial extent, and therefore encompassing a greater area for test sites to be assessed, is as effective as models generally developed within smaller geographic regions such as within a basin or watershed.

Benthic invertebrates and habitat data from three areas in Canada were examined: the Attawapiskat River basin in northern Ontario, the Fraser River basin in British Columbia and the Yukon River basin. The RCA predictive model method was used in this study that determines the relationships between benthic community groups and the environmental descriptors that explain them and the Benthic Assessment of Sediment (BEAST) assessment method to compare test sites with a physical similar group of reference sites. The performance of the bioassessment was assessed using a common set of simulated impact ("simpacted") sites with known responses

of taxa to disturbance. Models for each basin and a multi-basin model were compared on prediction performance, parsimony, and sensitivity. The multi-basin model had comparable prediction performance (65% correctly classified) to single basin models (56-72%) but lacked the sensitivity that models for single basins possessed. The Attawapiskat was the most parsimonious with only 2 predictors but the Fraser and multi-basin models explained the most variance with more predictors (Wilks' $\lambda = 0.06$ and 0.1 for the Fraser and multi-basin models, respectively). The results of this study showed that sites can be assessed outside the range of their reference data insofar that the test site is within the range of environmental characteristics within the model. A test site assessed as disturbed for the multi-basin model will in fact be disturbed but disturbed sites are less frequently detected compared to single basin models. Therefore as with any bioassessment, users need to be aware of the chance of committing type 1 and 2 errors. Developing models that target a single stressor of concern to increase the pool of available candidate predictors is recommended; such a model may possess greater prediction performance and sensitivity.

Keywords

Benthic invertebrates, Bioassessment, Reference Condition Approach, Benthic Assessment of Sediment, BEAST, predictive modelling, Attawapiskat River basin, Yukon River basin, Fraser River basin, spatial application, geographic extent, Canadian Aquatic Biomonitoring Network, CABIN

Acknowledgments

There are many people that contributed to this project that I would like to thank. Thank you John Bailey and John Gunn, my supervisors, for their support and guidance throughout this project. John Bailey has given me many opportunities to gain invaluable opportunities to experience Ontario's great north, conferences, and learn from experts in other parts of the country. I would like to thank those individuals, Trefor Reynoldson, also a committee member, and Adam Yates for all their wisdom and help in multivariate predictive modelling and the art of GIS. Thank you to the many individuals for helping me sample along the great Attawapiskat River, which proved to be sometimes challenging but also rewarding: Chantal Sarrazin-Delay, your expertise and dedication did not go unnoticed. Stacey Greene for your positivity and perseverance in the field. Thank you Matthew Heerschap and Michael Ritchie for enduring and sampling the many sites in what seemed like an adventurous trip to say the least. Thanks to Isaiah Hollander and Andrew Sutherland from DeBeers, and Delaney and Junior Baxter from Marten Falls, for being amazing guides venturing down that dodgy river.

Thank you to DeBeers Victor Mine for feeding, housing, and transporting us around the area. Thanks to Brian, Rod, Terry, and Aline at DeBeers with logistics and for letting us get in their way. Thank you to Sue Salter for sorting and identifying many samples in a timely manner. Thanks to the Ministry of Environment Lab Services at Dorset Environmental Science Center and Rexdale for analyzing the many water samples. Thank you to Environment Canada and GHOST Environmental Consulting for the Fraser River and Yukon River basin datasets, respectively. Lastly, thank you for the financial support to make this project happen: NSERC CNAES Strategic grant, Ontario Ministry of the Environment and Climate Change, and Laurentian University.

Table of Contents

List of Tables

List of Figures

List of Appendices

Chapter 1

1 General Introduction

1.1 Aquatic ecosystem management and its application for the protection of ecosystem services

The impacts of human activities on the natural environment are known to result in biodiversity loss and changes to the structure and functioning of ecosystems. Biodiversity influences functions of ecosystems such as primary production (Tilman et al. 1996, Costanza et al. 2007), nutrient cycling (Altieri 1999), and decomposition (Srivastava et al. 2009, Cardinale et al. 2011). Recent studies have suggested that substantial loss in biodiversity can produce functional effects as detrimental as ozone depletion, acidification, elevated CO_2, and nutrient pollution (Tilman et al. 2012, Hooper et al. 2012). In aquatic systems, the protection of aquatic ecosystem services, such as clean water and fish resources, are often those that people can easily relate to and important in their day to day lives. Consequently, understanding the structure and functioning of the natural aquatic ecosystems that generate these essential services is important to their protection.

The protection of important ecosystem services requires management of our activities within ecosystems. To understand the repercussions of our activities, we need to understand the main drivers and processes that shape the structure and function of healthy biological communities and develop management strategies and frameworks to protect them. Bioassessments are useful in assessment and monitoring programs designed to protect the ecosystem services important to us as a society. The use of biological indicators such as measures of periphyton, fish, and benthic

invertebrates provide a means to assess the health of ecosystems and detect deviations from a healthy ecosystem that simple chemical analyses (e.g. measuring pH or metal concentration) cannot detect (Reynoldson et al. 1995). Benthic invertebrates are commonly used indicators for freshwater stream bioassessments because they are abundant, are relatively sedentary and long-lived, and they exhibit distinctive and variable levels of tolerances to many natural and anthropogenic stressors. These characteristics make the presence, absence, and relative abundance of taxa indicative of the health status of these ecosystems (Norris and Hawkins 2000).

1.1.1 History of Bioassessment

The need to assess ecosystem health stems from the early 1900's when increasing human population and industrial development significantly affected aquatic ecosystems (Bailey et al. 2004). The work of Kolkwitz and Marsson (1909) originated the concept of biological indicators for aquatic ecosystems using plankton and periphyton responses to sewage contamination. Since then biological indicators, such as benthic macroinvertebrates, have been used to develop a wide array of indices, including biodiversity indices, to assess the effects of various stressors. Although later criticized for their over-simplification, indices that reduced ecological complexity to simple numbers were widely adopted for use by many environmental agencies. Today, various indices are used most often in combination with other more statistically advanced methods.

In the 1970's bioassessments began to make use of what is called before and after, control and impact (BACI) study designs (Green 1979). The objective of BACI and other similar methods was to reduce multivariate complexity of species by site matrices into univariate data such as indices useable for univariate statistical analyses. This was mostly due to limitations of computing capacity. Although this method provides a means to assess impairment, this design proved to be difficult and expensive to implement. As a result the need for more cost-effective

bioassessments quickly became apparent and methods such as rapid bioassessment protocols were developed in the United States (Barbour et al. 1999). These multimetric methods use numerous metrics that condense the biological data into a single index revealing the effects of a disturbance. Meanwhile, multivariate predictive methods were also being developed and are now a common and the most appropriate bioassessment method for multivariate problems today.

1.1.1.1 *Predictive Modelling*

Multivariate predictive models are based on the work of Green (1971, 1979) who used discriminant analysis to quantify species' niches. However the first predictive model was developed by Field et al. (1982) using the richness of communities of nematodes for a vast number of sites in the River Exe estuary. In 1977, a national biomonitoring program in the UK at the River Laboratory of the Institute for Freshwater Ecology (IFE) began to investigate two objectives: to identify unpolluted river sites using benthic macroinvertebrate indicators and whether biota could be predicted from physical and chemical attributes (Bailey et al. 2004). This led to the development of the predictive model River Invertebrate Prediction and Classification (RIVPACS) (Wright et al. 1984). The IFE's project objectives were the founding ideas of the Reference Condition Approach (RCA). The RCA is a characterization of a number of minimally exposed (reference) sites to reveal the range of natural variation that exists among communities. A scientist uses this method to evaluate the impact of a disturbance (e.g. chemical spill) at a particular exposed (test) site. This method generates an expected range of variation in biota that exists in natural undisturbed reference sites. It has now been expanded to national assessment systems in Australia, Canada, Portugal and is being developed or considered in Sweden, South Africa and is used in many state jurisdictions in the USA.

Predictive modelling makes use of relationships between the biological community structure (or function) and the surrounding environmental characteristics. We know that biological communities are influenced by their surrounding environment, food resources, and interactions with other species. These interrelated abiotic and biotic factors work at various spatial scales to influence community composition and dynamics. Predictive modelling quantifies the relationships between the biota and their environment so as to match a site of concern to a physically similar group of least disturbed sites and thus provide an assessment of the similarity of that site to the reference sites. Predictive models are generally developed in two steps. The first step is model building which involves grouping the biological data based on the similarities among site communities, and then using Discriminant Function Analysis (DFA) to explain the biological groupings with environmental descriptors, which is important for the assessment of test sites. The second step is the assessment process that involves the matching of a test site to a physically similar group of reference sites to reveal the degree of deviation from what is expected in the reference condition. If a test site falls outside of this natural range of biological variation, then it is determined to be impacted and the degree to which it falls outside this range is the magnitude of the impact (Bailey et al. 1998).

Two common RCA bioassessment methods exist that differ in test site assessment. The first is RIVPACS and its similar counterpart the Australian River Assessment Scheme (AUSRIVAS) (Parsons and Norris 1996). This method generates an Observed (O) to Expected (E) ratio to assess the deviation of the number of taxa observed to what is expected in reference condition. The expected number of taxa of a test site is calculated using the probabilities of group membership multiplied by the proportion of each taxon within each of the groups. These values are summed to produce the expected probability of a taxon's presence and a summation across

all taxa is the expected number of taxa at a test site if it were in reference condition. The greater the deviation between O and E, the more likely a test site is not in reference condition. The second common RCA bioassessment method is the Benthic Assessment of SedimenT (BEAST) developed for Canadian streams (Reynoldson et al. 1995). This method compares a test site's biota abundances (rather than presence/absence) with one reference group that has the highest probability of group membership (Reynoldson et al. 1997). Subsequently a test site is ordinated with the reference group and confidence ellipses plotted usually at the 90%, 99, and 99.9% confidence intervals. The degree of impairment is determined by how far the test site deviated from the range of acceptable reference conditions. The methods in which bioassessment statistical analyses are conducted has been a source of debate and confounding ideas (Karr and Chu 2000, Norris and Hawkins 2000). RIVPACS and AUSRIVAS use all the reference presence/absence data to assess test sites, while the BEAST method compares a subset of reference abundance data to a test site. Both methods are still used today, with advantages and limitations to each. The development of bioassessment methods and tools have allowed us to better monitor and protect important aquatic resources, however advancements in this field come about from pushing the boundaries of what we know.

Chapter 2

2 Geographic Extension of Benthic Invertebrate RCA Bioassessments: How Far Can We Go?

2.1 Introduction

Bioassessments are conducted in many regions in the world to monitor and assess the impacts of anthropogenic activities on aquatic ecosystems. Reference Condition Approach (RCA) bioassessments are a common method to assess potential impacts to aquatic ecosystems using benthic macroinvertebrates as indicators. This method relies on reference sites sampled throughout a study region to characterize the range of natural variability of biota. Test sites exposed to human activity are compared with these relatively unexposed reference sites to assess potential deviations from the reference condition. Predictive models are increasingly used in bioassessment and are generally built using benthic macroinvertebrate abundance data and the environmental descriptors of the habitat in which the invertebrates are found. These models establish relationships between the biota and environmental descriptors to enable test sites to be matched to environmentally similar reference sites, so that benthic communities can be compared (Reynoldson et al. 1995, Wright 1995).

Predictive models are generally spatially confined to geographic boundaries as a means for controlling variability of important variables that influence community abundances and composition (Hawkins et al. 2010). Accordingly, increasing the geographic scope of a model is expected to increase the environmental variability (Corkum 1989, Vinson and Hawkins 1998) and therefore increase variability in benthic communities. An approach to control for this is by

limiting the geographic scope to regions thought to possess less variability such as within ecoregions, watersheds or even catchment (Barbour et al. 1996, Feminella 2000, Hawkins and Vinson 2000).

Ecoregions such as those in the United States defined by Omernik (1987) provide a method to partially control for environmental variability because they are based on various environmental attributes that appear common or characteristic of that region. These factors include climate, soils, geology, vegetation, and physiography. Ecoregions have therefore been found to better characterize biological variability but presumably, because of the narrowness of the environmental conditions within an ecoregion have generated imprecise predictions of the expected reference condition (see Hawkins et al. 2000 and references therein). Use of ecoregions or other similar geographic boundaries have then proven not to be necessary when using predictive modelling because biota are grouped based on their structural similarities. Although there is potentially more environmental variability to characterize, if similarities among reference sites exist beyond the bounds of an ecoregion or watershed, a predictive model could be an effective method to use over larger spatial extents.

The overall goal of my study was to assess whether a predictive model applied across a very large spatial extent, and therefore encompassing a greater area for test sites to be assessed, is as effective as models developed within smaller geographic regions such as within a basin or watershed. Previous studies showed that the sensitivity of RIVPACS models decrease when the geographic scope was expanded from 150,000km² to 2,500,000km² (Ode et al. 2008). Mykrä et al. (2008) also found poorer RIVPACS model performance (i.e. the average and variability of the root mean square error of O:E ratios) when applied across 2 rather than a single ecoregion in Finland. Their models were however based on *a priori* classifications and biological

characterization was limited by the spatial extent selected. For my study, I instead used the RCA predictive modelling and BEAST assessment approach of Reynoldson et al. (1995), a standardized method used and developed in Canada to allow for consistent interpretations of data. The Canadian Aquatic Biomonitoring Network (CABIN) implements this method to provide environmental managers a means to monitor and assess aquatic ecosystems in Canada. RCA BEAST predictive models have been developed in countries such as Canada, Portugal, and Brazil. The spatial extent of the study areas where these models were developed ranged from 4500km² with 43 reference sites in central Portugal (Feio et al. 2007) to 234,000km² with 219 reference sites in the Fraser River basin, Canada (Reynoldson et al. 2001). Prediction performance and sensitivity of these models tended to decrease with increasing spatial extent. Although the spatial extent of models has increased to encompass larger areas, test sites are still generally assessed within the geographic bounds of the sampled reference sites, and usually spatially limited within a watershed or ecoregion. In my thesis, I question whether test sites can be assessed using reference data from outside their geographic boundaries. I attempted to answer this question by creating a predictive model encompassing a large geographic region and assessing whether test sites within the bounds of this larger model produces equally sensitive test site assessments than models developed for a single region or watershed.

In this study, I developed RCA predictive models using benthic invertebrate data collected from shallow (wadeable) streams in three distinct areas of Canada: the Attawapiskat River basin in northern Ontario, the Fraser River basin in British Columbia, and the Yukon River basin. I developed models using datasets from the individual basins and on pooled data from all 3 basins, a combined area which covers approximately 570,000km². Bioassessments have been previously

conducted for the Yukon and Fraser River basins (Bailey et al. 1998, Reynoldson et al. 2001) however not in the remote and near pristine Attawapiskat River basin.

The objectives of my study were to 1) identify similarities in community composition among the 3 basins, 2) develop predictive models to compare the environmental predictors that discriminate benthic invertebrate community groups, and 3) compare prediction performance (% reference sites correctly classified) of the models and the sensitivity of assessments (type 1 and 2 error) to deviations from reference condition among models. Sensitivity was assessed using a common set of simulated impact ("simpacted") datasets for 4 levels of disturbance using known responses of benthos to stressors. I hypothesized that sites located in close proximity would tend to group together but that a common model could be developed across basins due to similar environmental descriptors influencing biological communities at the individual site level.

2.2 Methods

2.2.1 Study regions

I analyzed 3 data sets comprising macroinvertebrate abundances and environmental descriptors at reference condition wadeable streams from 3 basins in Canada: the Attawapiskat River, Yukon River, and Fraser River basins. Reference sites from the Attawapiskat basin (*n*=67) were sampled in 2013, the Yukon (*n*=293) from 2004-2012 and the Fraser (*n*=325) from 1994-2010. All 3 data sets followed the Canadian Aquatic Biomonitoring Network (CABIN) protocol, a standardized system developed and managed by Environment Canada (Environment Canada 2014).

The Attawapiskat River basin, located in northern Ontario, Canada (Figure 2.1) is approximately 50,500 km² and spans latitudes 51°N to 53°30'N and longitudes 82°W to 92°W (Table 2.1). The Attawapiskat River is 748 km long, flowing from Attawapiskat Lake through the Canadian Shield, and into the Hudson Bay lowlands and James Bay near the community of Attawapiskat. This basin is characterized by coniferous and mixed forest dominating the western Canadian Shield portion of the basin and predominantly treed and open fens and bogs as the primary vegetation in the James Bay lowland portion. The climate is colder near the coast of James Bay and warmer inland (mean annual temperatures from -2.6 to 0.5°C) with long, cold winters and cool, short-lived summers (Crins et al. 2009).

The Attawapiskat River basin is home to many First Nations communities such as Attawapiskat, Webequie, Nibinamik, Neskantaga, Eabametoong, and Marten Falls comprising approximately 3900 residents. The major components of the economy include hunting, fishing, trapping, and outdoor recreation as well as a diamond mining operation in the Lowlands area (Crins et al. 2009). Future chromite and other mineral mining (Ring of Fire Belt) is expected to affect

approximately 5000km² of the region near the transition between the Canadian Shield and Lowlands region of the Attawapiskat River. These developments are expected to contribute $5.1 billion to the province's GDP over the first 10 years, but also negatively impact the sensitive ecosystems and First Nations that reside within the basin (Chong 2014).

The second study region is the Yukon River basin, Yukon Territory, Canada. The Yukon River is the longest free flowing river in North America, which flows 3200 km from the headwaters in mountainous northern British Columbia, through diverse ecosystems, and into the Bering Sea in Alaska. The basin spans approximately 840,000 km², North America's 7th largest basin, with only 39% of the basin within Canada, 90% of which is within Yukon Territory (Figure 2.1) (Bailey 2005). Sampling sites span throughout the Yukon portion of the basin and lie between latitudes 60°N and 69°N and longitudes 130°W to 141°W (Table 2.1). This basin is characterized by mountainous terrain, plateaus, and river valleys (McKenna and Smith 2004). Land cover is characterized by open and discontinuous coniferous forests most prevalent on plateaus and discontinuous permafrost, and low shrub communities increasing northward on mountain slopes and high plateaus (McKenna et al. 2004). Common tree species are white and black spruce, while alpine fir and lodgepole pine are common in the headwaters. This basin is in the Continental climate zone represented by cool summers and very cold winters. Species richness is much lower for aquatic and terrestrial communities compared to similarly size areas in North America, however the Yukon region has higher biodiversity for this latitude compared to eastern North America (Bailey 2005). The Yukon basin is sparsely populated, with approximately 30,000 residents in the Territory, mostly concentrated in the city of Whitehorse (20,000) and the remaining mostly along the Yukon River and its tributaries. The Yukon Territory has been a source for gold since the 1897 gold rush and continues to this day with placer gold mining in the

Klondike River, Stewart River, and nearby basins. Agricultural activities are also conducted north of Whitehorse, mostly for hay production and livestock grazing (Bailey 2005).

Reference sites sampled in the Fraser River basin are located in the province of British Columbia, Canada. The Fraser River is the fifth longest river (1375 km) in Canada which flows from its headwaters in the Rocky Mountains, through the Fraser Plateau and Coast Mountains, and into the Pacific Ocean at the city of Vancouver (Reynoldson et al. 2005) (Figure 2.1). The basin is the fifth largest in Canada (234,000 km^2) and lies between latitudes 49°N and 56°N and longitudes 118°W and 125°W (Table 2.1). The basin is among the most diverse, encompassing 11 of 14 biogeoclimatic zones in British Columbia as well as 6 varied ecoregions ranging from dry sagebrush and grasslands, to coniferous forests and alpine tundra. The climate is varied with areas of mild climate in the southern valleys and humid and cold climate in the mountainous regions in the northern portions of the basin. The basin's aquatic insects are characterized by up to 50 families, with diversity decreasing from the headwaters to the downstream reaches and a varied abundance gradient, with the greatest increases of abundance found between the headwaters and Quesnel (Reynoldson et al. 2005).

The Fraser River basin comprises many urban centres such as Vancouver, the 3rd largest metropolitan area in Canada with 2.3 million people, and 3/4 of British Columbia's population (4.6 million) residing in the Lower Fraser Valley. The basin supports many natural resource operations such as a major commercial forest industry accounting for 60% of Canada's lumber exports, mining operations providing 70% and 98% of Canada's coal and copper exports, respectively, and the 3rd largest agricultural sector (salmon farming primarily) in Canada in the Fraser Valley. Recreation and tourism are another major economic sector with recreational fisheries comprising the largest component (Reynoldson et al. 2005).

Table 2.1: Summary of basin characteristics for the Attawapiskat River, Yukon River, and Fraser River basins.

	Attawapiskat River	**Yukon River**	**Fraser River**
Basin size (km²)	50,500	294,840*	234,000
Latitude range	51°N to 53°N	60°N and 69°N	49°N and 56°N
Longitude range	82°W to 92°W	130°W to 141°W	118°W and 125°W
Population	3900	30,000	2,300,000
Population density	0.077/km²	0.102/km²	9.829/km²
Topography	Flat lowlands	Mountainous terrain, plateaus and river valleys	Mountainous terrain, plateaus and river valleys
Vegetation	Coniferous and mixed wood forests, fens and bogs	Open and discontinuous coniferous forests, and low shrub communities	Dry sagebrush, grasslands, coniferous forests, and alpine tundra
Climate	Long, cold winters and cool, short-lived summers	Cool summers and very cold winters	Mild to humid and cold
Human activities	Diamond mining, hunting, fishing, trapping, future chromite mining, recreation	Mining, agriculture, urbanization	Forestry, mining, aquaculture, outdoor recreation

*** Yukon Territory, Canada portion only**

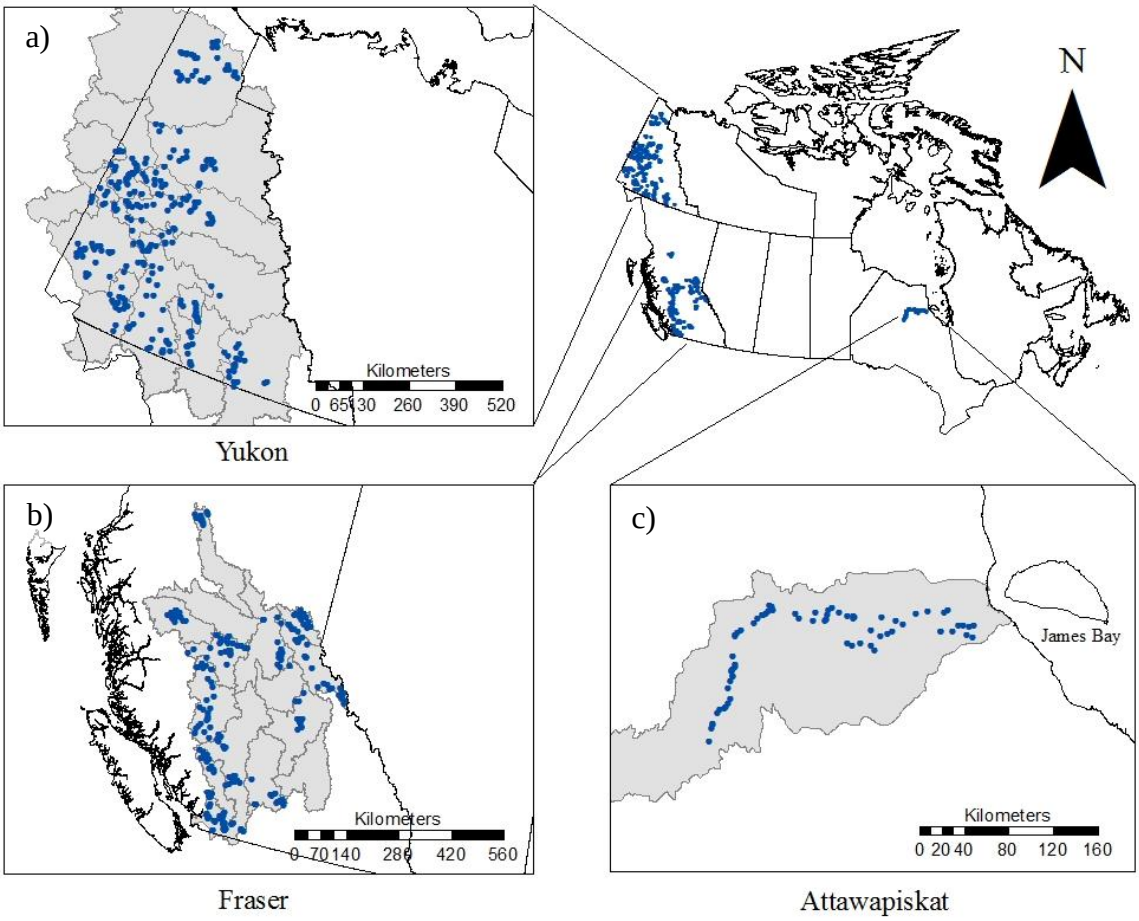


Figure 2.1: Location of sampling sites within the 3 study basins in Canada: a) Yukon River, Yukon Territory, b) Fraser River, British Columbia, and c) Attawapiskat River, Ontario.

2.2.2 Benthic invertebrate sampling procedure

For all sampling sites, invertebrates were collected using the standardized CABIN method involving a 3-minute travelling-kick technique using a 400-µm (some sites from the Yukon sampled with a 500-µm) mesh kicknet (see Environment Canada 2012a). Samples were transferred into labelled container(s) and fixed to a concentration of 10% buffered formalin.

2.2.3 Invertebrate processing and identification

In the laboratory, each sample was rinsed of formalin on a sieve, transferred back into the sample jar, and filled with 70% ethanol for better long term preservation of soft bodied invertebrates. Benthos were subsampled using a Marchant Box and a minimum of 300 organisms were sorted and identified to the lowest taxonomic level possible by an expert taxonomist (see Environment Canada 2012b).

2.2.4 Environmental descriptors

In addition to benthos collection, water samples and habitat data were also collected at each sampling site. Water samples were shipped back to the lab and analysed for various nutrients and metals, while habitat characteristics were obtained on-site (see Environment Canada 2012a).

Landscape-scale characteristics for each site were obtained using ArcGIS (10/2010-06-29, ESRI, Redlands). The bedrock and surficial geology, long term climate, land cover, and various physical catchment characteristics such as drainage area, stream length, and stream density were acquired from delineated catchments. Catchments for each site were delineated using Digital Elevation Models (ASTER GDEM V2, NASA and METI 2011) and stream networks (Geobase 2007) in the ArcHydro Tools extension (2.0/2011-10-12, ESRI, Redlands) for ArcGIS 10. Stream order was obtained from 1:50,000 scale topographic maps.

2.2.5 Statistical analyses

The objective of this study was to compare the performance of RCA predictive models developed for single basins with a model using datasets from all basins. This was accomplished by developing 3 ‘basin’ models (one for each basin) and a multi-basin model (all 3 basin datasets combined). Model development followed the RCA predictive modelling approach of Reynoldson et al. (1995). This method involves classifying reference sites into different community

assemblage groups, followed by determining the relationship between these assemblages and environmental descriptors using Discriminant Function Analysis (DFA). Models were compared for prediction performance and parsimony using these reference datasets, as well as the bioassessments' sensitivity to disturbance using simulated impact ("simpacted") sites at 4 levels of disturbance.

Accuracy (prediction performance) of models was determined by their cross validation performance in assigning the reference site to the group to which it belonged. A more rigorous method is to use a validation data subset of reference sites used in the classification but not in the model building. However, the relatively small number of sites available from the Attawapiskat precluded this approach. Parsimony was assessed by the number of predictor variables used and how much variance is explained by the environmental predictors. Fewer predictors that explain the most variance in a model is desirable. Sensitivity was determined by how well the assessments of tests sites were able to detect disturbance. For determining sensitivity I used simpacted sites because the true status of test sites is unknown. Simpacted sites are reference sites modified based on known changes in abundance and/or richness of benthic invertebrate communities to disturbance. Simpacted sites were used because known changes to community structure provide the best test of a bioassessment to deviations from reference condition. The use of simpacted sites allows us to evaluate the type 2 error rates (not detecting an effect that exists) of the bioassessment, and therefore, how it responds to known deviations from the null hypothesis/reference condition.

2.2.5.1 *Classification*

For each model, the benthic community assemblages were classified into groups using cluster analysis. Classification was conducted for each data set using raw and transformed (square root,

fourth root and log(x+1)) abundance data at the family level. I transformed the data because it down-weights the contribution of dominant taxa potentially revealing underlying trends in the benthic communities (Bennett 2011). The transformation selected was dependent on subsequent analyses in the model building phase discussed in the Model selection section (Section 2.2.6). Taxa were classified into groups through group averaging hierarchical agglomerative cluster analysis using Bray-Curtis similarity matrices of taxa abundances in PRIMER (version 6.1.13/2007, PRIMER-E, Plymouth) with beta set at -0.1. A dendrogram was produced with reference sites grouped based on their structural similarities. The decision on which solution to accept was made using the significance ($p>0.01$) of structural similarities among sites identified by the SIMPROF test in PRIMER. Reference sites forming small groups (≤10 sites) and/or individual sites were considered outliers and removed from the analysis because groups with less than 10 sites are not representative of reference group variability and not accurate for prediction (Bailey et al. 2004). Following cluster analysis, the various grouping solutions were ordinated in non-metric Multi-Dimensional Scaling (nMDS) space for visual examination of similarities and variability within and among groups. Analysis of Similarities (ANOSIM), which tests whether significant differences exist between groups of samples, was conducted to determine whether differences existed between basin benthic communities.

Following cluster analysis of the multi-basin model, the proportion of sites from each of the 3 basins was determined for each community group. Reference sites comparable in community composition will group together and reveal whether similarities are found among basins. nMDS ordination of all reference sites was also conducted to reveal the degree of overlap in Bray-Curtis similarities among basins. Concordance of communities across basins is important to the feasibility of a model applied over a large spatial extent. It will reveal potentially similar

environmental variables influencing and resulting in similar benthic communities. Similarities found among basins, and therefore a presence of all basins' reference sites within all the classified biotic groups may allow a large scale model to be as effective at assessing test sites as single basin models.

2.2.5.2 *Building Predictive Models*

Group discrimination consisted of distinguishing community groups based on environmental variables using DFA in Systat (version 13.00.05/2009, Systat Software, San Jose). The objective was to create 'multi-purpose' predictive models (used for any type of disturbance) because of the varied nature of the basins. As a result, environmental variables that would be potentially affected by disturbances such as mining, logging, agriculture, and urban development were removed (Reynoldson et al. 1997, Bailey et al. 2004). For example, deforestation causes changes to stream hydrology and erosion, therefore predictors such as stream velocity, % silt in the substrate, and total suspended solids would be inappropriate to assess the effects of logging. Environmental variables were also limited to those only shared by all three study areas for consistency and comparability purposes. Out of 150 available environmental variables, 49 candidate predictors were selected, consisting of landscape-scale variables obtainable using GIS techniques and topographic maps such as drainage characteristics, stream order, precipitation, and climate data. A list of candidate predictors is presented in Table 2.2.

Table 2.2: Candidate environmental predictors available for all RCA predictive models (49 variables).

	Mean±SE (range)		
Environmental descriptor	**Attawapiskat River (*n*=67)**	**Yukon River (*n*=293)**	**Fraser River (*n*=325)**
+Latitude (°)	52.8±0.03°N (52.2°–53.2°N)	63.1±0.1°N (60.0°-68.2°N)	51.7±0.1°N (49.1°-56.0°N)
+Longitude (°)	84.6±0.1°W (86.4°-82.5°W)	137.1±0.2°W (141.0°-128.8°W)	122.4±0.1 (126.8°-118.4°W)
+Altitude (m)	94.1±6.2 (8-176)	702.2±16.6 (244-2003)	843.1±27.1 (6-1996)
***Drainage area (km²)**	241.4±52.6 (0.2-2026.0)	142.1±17.4 (4.1-3060.3)	1738.2±305.3 (0.7-55,151.5)
***Stream length (km)**	313.7±73.7 (0.9-2495.4)	136.9±21.9 (2.0-5203.8)	3333.4±595.6 (1.0-100,360.5)
***Stream density (m/km²)**	1191±98.8 (203.4-5795.7)	948.5±23.0 (388.4-3572.1)	1976.9±36.8 (437.4-4964.7)
***Stream order**	2.1±0.1 (1-5)	3.1±0.1 (1-6)	4.1±0.1 (1-8)
^% sedimentary	96.3±2.1 (0-100)	36.8±2.5 (1-100)	26.9±2.1 (0-100)
^% volcanic	0.04±0.04 (0-2.6)	6.6±1.1 (0-100)	27.5±1.7 (0-100)
“% water	5.3±0.5 (0-16.8)	0.7±0.2 (0-29.3)	1.6±0.1 (0-11.2)
~Degree days	153.1±0.2 (151-157)	117.6±0.9 (60.5-151.4)	134.1±1.2 (62-177)
~Precip. January (mm)	25.1±0.2 (22.9-28.9)	33.1±0.9 (12.4-144.3)	117.6±4.5 (28.8-323.0)
~Precip. February	22.3±0.1 (21.5-23.9)	29.6±0.8 (11.7-124.8)	93.2±3.8 (17.8-270.0)
~Precip. March	24.6±0.3 (20.2-30.1)	27.5±0.7 (10.1-112.4)	82.5±3.2 (18.3-235.0)
~Precip. April	30.0±0.4 (24.1-36.9)	24.4±0.6 (8.9-103.9)	82.5±3.2 (18.3-235.0)
~Precip. May	38.3±0.4 (33.0-46.1)	36.2±0.8 (12.5-112.0)	65.3±1.7 (29.1-149.0)
~Precip. June	65.9±0.4 (61.8-72.1)	55.5±1.0 (17.1-117.1)	76.9±1.1 (44.0-126.0)
~Precip. July	87.6±0.4 (81.8-93.9)	69.2±1.2 (25.8-132.3)	68.1±0.9 (38.0-102.0)
~Precip. August	77.5±0.7 (68.6-85.9)	63.0±1.1 (30.5-175.1)	61.8±0.8 (34.0-87.0)
~Precip. September	77.7±0.5 (71.5-84.5)	48.3±1.4 (22.8-257.0)	60.6±1.3 (22.0-119.0)
~Precip. October	54.4±0.1 (53.8-56.3)	46.0±1.4 (21.6-258.7)	96.6±3.5 (27.0-254.0)
~Precip. November	44.3±0.5 (36.9-51.6)	36.4±1.0 (12.6-161.8)	126.4±5.1 (26.0-417.0)
~Precip. December	28.4±0.2 (25.1-32.4)	33.5±1.1 (12.1-190.0)	124.1±2.8 (32.8-360.0)
~Total annual precip. mm)	576.1±4.1 (522-643)	502.9±11.5 (199.8-1883.0)	1045.1±31.9 (370.1-2598)
~Min. temp. January (°C)	-28.4±0.01 (-28.5 to -28.1)	-29.6±0.2 (-32.2 to -12.6)	-11.3±0.3 (-15.0-0)
~Min. temp. February	-27.0±0.04 (-27.3 to -26.1)	-26.9±0.2 (-18.7 to -4.3)	-9.5±0.2 (-6.5-4)
~Min. temp March	-20.4±0.04 (-20.9 to -19.6)	-22.3±0.2 (-4.6-8.7)	-6.7±0.2 (0-9.0)
~Min. temp. April	-9.7±0.04 (-10.1 to -8.9)	-11.6±0.2 (-2.6-8.3)	-3.0±0.1 (1.0-11.0)
~Min. temp. May	-0.7±0.04 (-1.0-0.17)	-2.6±0.1 (-12.4 to -3.0)	0.3±0.1 (-5.0-5.0)
~Min. temp. June	5.2±0.1 (4.7-6.5)	3.6±0.2 (-31.4 to -12.3)	3.1±0.1 (-16.3-0)
~Min. temp. July	9.0±0.1 (8.3-10.3)	5.8±0.2 (-22.9 to -2.7)	5.3±0.1 (-5.0-7.0)
~Min. temp. August	8.6±0.1 (8.1-9.5)	3.8±0.2 (-7.2-8.1)	5.3±0.1 (0.8-13.0)
~Min. temp. September	4.0±0.01 (3.9-4.2)	-1.2±0.1 (2.7-21.8)	2.1±0.1 (8.0-19.0)
~Min. temp. October	-1.2±0.01 (-1.3 to -1.1)	-8.9±0.1 (4.0-21.1)	-1.2±0.1 (11.8-23.0)
~Min. temp. November	-10.0±0.1 (10.7 to -9.3)	-22.0±0.1 (-6.5-5.1)	-6.7±0.2 (2.0-14.0)
~Min. temp. December	-22.5±0.1 (-23.1 to -21.8)	-27.9±0.2 (-23.5 to -4.6)	-10.4±0.2 (-7.0-5.0)
~Max. temp. January (°C)	-16.9±0.05 (-17.5 to -16.1)	-19.8±0.2 (-33.1 to -14.1)	-3.3±0.2 (-17.0-0)
~Max. temp. February	-13.5±0.1 (-14.2 to -12.4)	-16.7±0.2 (-29.1 to -10.1)	-0.6±0.1 (-11.0-1)
~Max. temp. March	-5.9±0.1 (-6.7 to -4.6)	-10.1±0.3 (-8.1-2.5)	2.3±0.1 (-3.0-6)
~Max. temp. April	2.7±0.1 (2.1-4.3)	0.01±0.2 (-2.1-10.8)	6.6±0.1 (1.0-11.0)
~Max. temp. May	11.2±0.1 (10.6-13.0)	8.8±0.2 (-6.1-2.9)	11.1±0.1 (-1.0-9.0)
~Max. temp. June	17.3±0.1 (16.7-18.7)	15.6±0.2 (-25.3 to -10.2)	14.3±0.1 (-11.0-2.0)
~Max. temp. July	21.0±0.1 (20.4-22.2)	17.5±0.2 (-25.0 to -6.0)	17.6±0.1 (-8.0-5.0)
~Max. temp. August	19.8±0.04 (19.4-20.6)	15.0±0.2 (-18.1-1.4)	17.6±0.1 (-1.3-10.0)
~Max. temp. September	13.0±0.03 (12.6-13.6)	8.2±0.2 (-2.0-16.6)	13.6±0.1 (5.0-17.0)
~Max. temp. October	6.0±0.03 (5.7-6.6)	-1.8±0.2 (4.9-23.4)	7.1±0.1 (11.8-23.0)
~Max. temp. November	-2.6±0.1 (-3.3 to -1.7)	-13.4±0.2 (-0.2-15.0)	0.1±0.1 (8.0-20.0)
~Max. temp. December	-12.6±0.1 (-13.4 to -11.5)	-18.4±0.2 (-17.9 to -2.1)	-2.9±0.1 (-4.0-8.0)
~Mean annual temp. (°C)	-2.2±0.03 (-2.6 to -1.6)	-6.4±0.1 (-10.1-0.7)	2.0±0.1 (-1.5-9.6)

+ = Location descriptor, * = Catchment morphology/hydrology descriptor, ^ = Bedrock geology descriptor,
“ = Land cover descriptor, and ~ = Long term climate descriptor (1971-2000).

All datasets were screened for meeting the assumptions of DFA using R Studio (version 3.0.2/2013-09-25, R Foundation for Statistical Computing, Vienna). Grouping solutions with heterogeneous covariance matrices were not considered for final model selection. Since the best predictors for the models are unknown, forward- and backward-stepwise DFA was conducted for each grouping solution to select candidate predictors that best discriminate community groups. The tolerance for DFA was set to 0.1, which is a measure of collinearity in Systat to prevent multicollinearities and avoid redundancy of similar data (McGarigal et al. 2000). The candidate models were selected from the results of DFA including the number of variables, jackknifed (leave-one-out) cross validation (CV), F-value and Wilks' λ.

Each potential classification was examined for community group characteristics using the Similarity Percentages (SIMPER) in PRIMER. SIMPER is a method that looks at the average similarity within and average dissimilarity between community groups based on the percent contribution of each taxon to within group similarity (or between group dissimilarity) using Bray-Curtis distance measures (Clarke and Gorley 2006).

2.2.6 Model selection

Candidate models were selected from the pool of all potential models for each basin and multi-basin datasets. Potential models were removed that possessed high CV errors (%), a higher number of predictor variables than the smallest reference group size, and high Wilks' λ. High CV error is not desirable because an incorrect matching of a test site to reference group would result in an inaccurate assessment. A high number of predictor variables increases the interactions/correlations among variables and a model with the least predictors will be more robust. Lastly, a high Wilks' λ represents a weak model because predictor variables are not explaining the variance in discriminant scores. After selection of candidate models, the best

model was clearly defined using a rank-sum method based on 6 criteria considered important features of an ideal model (Table 2.3). It is most important for the ideal model to classify test sites to the correct group (higher CV) because incorrect classification results in inaccurate assessments. A similar CV across groups is 2nd for importance because assessments among groups will be more consistent and provide a well-rounded model. A lower number of predictors is 3rd for importance because it provides a more robust model (ideally less than the smallest group size). The 4th ranked criterion is evenly distributed group sizes to reduce among group variability to increase model sensitivity. A lower Wilks' λ is 5th in importance because predictor variables that explain the most total variance in discriminant scores is desirable but not as important as other criterion. Lastly, more model groups partition communities into smaller and more similar groups, and may provide a better representation of the biological communities. Within each of the 6 criteria, models were ranked best to worst. Each rank was multiplied by its weighed importance determined using the equation: $W(i) = (2(n+1-i))/(n(n+1))$, (where n = number of criteria) and all scores summed across all 6 criteria for each candidate model. The model with the lowest score was selected as the final model.

Table 2.3: Model selection criteria and their weightings for the selection of the best RCA predictive model for each basin and multi-basin dataset. RCA predictive model characteristics were used as model selection criteria. The best model is determined using a rank-sum method based on 6 criteria considered important features of an ideal model.

Model selection criteria	**Rank importance**	**Weighting**
Higher Prediction Rate	1	0.2857
Similar Errors	2	0.2381
Lower Number of Predictors	3	0.1905
Similar Group Sizes	4	0.1429
Lower Wilks' λ	5	0.0952
Higher Number of Groups	6	0.0476

2.2.7 Model evaluation (power and sensitivity)

The sensitivity and power of the assessments for each model were evaluated by determining the type 1 and type 2 errors for each community group and model. Type 1 error is inherent in the structure of the data and reveals the chance of falsely concluding that a test site is impacted. Type 2 error is the chance of falsely concluding that a test site is in reference condition. Model sensitivity to deviations from reference condition was assessed using simpacted data at 4 known levels of disturbance: undisturbed, mild, moderate, and severe. A total of 10, 30, 30, and 70 simpacted sites were assessed to determine type 1 and 2 errors for the Attawapiskat River, Yukon River, Fraser River, and multi-basin assessments, respectively. nMDS scores of a group of reference sites and a simpacted site to be assessed were plotted with 75% and 90% confidence ellipses in Systat. Type 1 errors were calculated as the proportion of undisturbed simpacted sites falling outside the reference confidence ellipse. Type 2 errors can only be estimated using simpacted sites, and was the proportion of disturbed simpacted sites that fall within the reference ellipse. The chance of committing type 1 and 2 errors was assessed using ellipses at 2 confidence intervals because the boundary between reference condition and disturbed is dependent on the decision point selected by a bioassessment user. Type 1 and 2 errors are inversely related, so for instance, using the 90% confidence interval should decrease type 1 errors, but increase type 2 errors.

Simpacted data were generated based on known sensitivities of benthic taxa to stressors using Hilsenhoff's (1987) tolerance values for organic stream pollution of the Great Lakes (used for the Attawapiskat dataset) and the Barbour et al. (1999) Idaho organic pollution tolerance index (used for the Yukon and Fraser datasets). Different indices were used for each region to match sensitivities with a published index from a region that is the most geographically similar. Taxa

were categorized into ‘tolerant’, ‘semi-sensitive’ and ‘sensitive’ using these indices (Table 2.4) and the abundances were changed and/or taxa eliminated depending on level of disturbance (Table 2.5).

Table 2.4: Classification of taxa into 3 sensitivity categories for simpact data using Hilsenhoff's tolerance values for the Attawapiskat River dataset and the Idaho pollution tolerance index for the Yukon River and Fraser River datasets.

Taxon Sensitivity Category	Hilsenhoff’s/Idaho Tolerance Value Range
Sensitive	0 – 3
Semi-Sensitive	4 – 6
Tolerant	7 - 10

Table 2.5: Reference data modifications to create simpact data at 3 levels of disturbance: mild, moderate and severe. Modifications differ between sensitive, semi-sensitive, or tolerant taxa due to differing responses to disturbance. Values represent either increasing (+) or decreasing (-) in % family richness or % abundance from the original reference data.

		Sensitive	Semi-sensitive	Tolerant
Mild	% Richness Δ	-10	0	0
	% Abundance Δ	-50	0	+200
Moderate	% Richness Δ	-50	-20	-10
	% Abundance Δ	-25	-50	-25
Severe	% Richness Δ	-100	-50	-20
	% Abundance Δ	0	-75	-50

Simpacted data were created in Microsoft Excel (version 15.0.4551.1512/2013, Microsoft, Redmond) using formulas to manipulate each sensitivity group and for each disturbance level. Abundance changes were an overall increase or decrease for all reference site abundances within each sensitivity group (and disturbance level). Richness was altered by the removal of families for each reference site within the sensitivity group experiencing the simulated disturbance. The number of families to remove within a sensitivity group depended on the total richness of the group. For example, at mild disturbance, 10% of sensitive families were removed. If a reference site has a richness of 7 sensitive families, 0.7 (rounded to 1) were removed. Families were

removed at random using the random number generator in Microsoft Excel. Each family possessed a randomly generated number and the families with the highest numbers were removed. Random removal of families for each site resulted in different families removed for each simpacted site.

2.2.8 Model comparisons

Models were compared among basins and to the multi-basin model based on their prediction performance, parsimony, and sensitivity. The correct classification of reference sites (cross validation) was used to compare the prediction performance, the number of predictors (fewer is desirable) to compare parsimony, and the type l and 2 error rates to compare power and sensitivity.

To further compare the sensitivity of the models, a subset of simpacted sites were assessed to determine if sites that failed to deviate from reference condition in the single basin models also failed to deviate in the multi-basin model. Simpacted sites were selected from each model dataset and were randomly selected within each model's biotic groups. The number of simpacted sites selected was proportional to group size. A total of 10 sites were selected from the Attawapiskat model, and 30 each from the Yukon and Fraser models. Each simpacted site was individually assessed by plotting nMDS scores of the simpacted site, its reference group, and confidence ellipses at the 75% and 90% confidence interval. A site either fell in the reference cluster (within the 75% ellipse), in the 75% band (between the 75% and 90% confidence ellipses), or outside the 90% confidence ellipse. Each simpacted site was assessed at the 4 levels of disturbance. Simpacted sites were not assigned to a group by the model because the group it belongs to is already known. The degree of concordance among model assessments was calculated as the

proportion of simpacted sites from the basin model and multi-basin model that agree within the respective confidence ellipse.

2.3 Results

2.3.1 Community characteristics

Community compositions of all three study regions are similar in richness and dominant taxa. There were 91, 86, and 95 families of benthic invertebrates found at 67, 293, and 325 reference sites in the Attawapiskat, Yukon, and Fraser River basins, respectively. The dominant taxa are midges (Chironomidae), mayflies (Baetidae, Heptageniidae, and Ephemerelidae), black flies (Simuliidae), stoneflies (Nemouridae), and riffle beetles (Elmidae) (Figure 2.2). The sampling effort, described as coverage (number of sites/ km² of basin), of the three basins was nearly equal at 0.001 (Table 2.6). Even with similar sampling effort among basins, Attawapiskat sites are the most diverse, and possess the highest: mean richness and variability, EPT taxa consisting of 36% of all taxa, and 23 families within 90% of total abundance. The Fraser and Yukon have similar EPT composition around 27%. The Fraser has the highest mean and most variable abundances while the Yukon has the lowest mean abundance and the lowest family composition within 90% of total abundances at 11 families (Table 2.6). Mean Simpson's diversity (1-λ') is similar across the basins, but the Yukon is the lowest at 0.65 and the Fraser the greatest at 0.74.

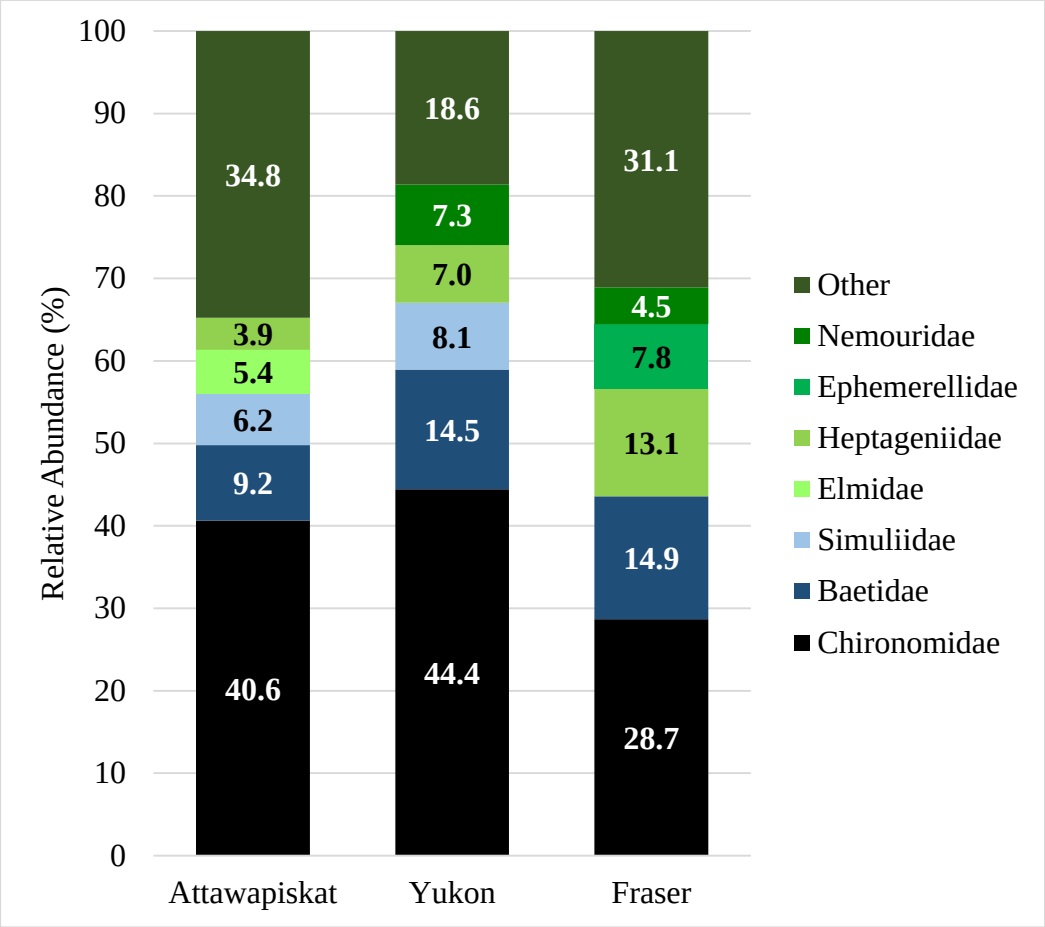


Figure 2.2: Dominant benthic invertebrate families of the Attawapiskat River (n=67), Yukon River (n=293), and Fraser River (n=325) basins.

Table 2.6: Summary of benthic invertebrate community characteristics of the Attawapiskat River, Fraser River, and Yukon River basins. EPT = Ephemeroptera, Plecoptera, and Trichoptera.

	Attawapiskat	**Yukon**	**Fraser**
Number of sites	67	293	325
Coverage (sites/km²)	0.0013	0.0010	0.0014
Total family richness	91	86	95
Mean±SE family richness	25.5±0.7	12.4±0.3	16.5±0.3
Family richness range	9-37	1-26	5-32
Mean±SE abundance	3827.2±327.2	1651.2±180.5	6006.6±414.6
Abundance range	120.7-13,580	8-25,000	24-40,000
90% most abundant families	23	11	15
Mean±SE Simpson's diversity (1-λ')	0.72±0.02	0.65±0.01	0.74±0.01
%EPT	36.3	26.7	27.4

2.3.2 Model building

After the various steps of classification, discrimination, and model comparisons, the final models for each basin and multi-basin data sets used either 4th root ($\sqrt[4]{}$) transformed (Attawapiskat, Fraser, and Multi-basin) or log(x+1) transformed (Yukon) abundances, with grouping solutions ranging from 3 to 6 (Table 2.7). The transformation selected was based on how well groups were discriminated by predictor variables, which was discerned by the model selection criteria discussed previously. The following sections describes model building results for each of the final models.

Table 2.7: Summary information of the final models selected for each basin and multi-basin datasets including the type of transformation of the abundance data, the number of biotic groups, and the number of reference sites within the models.

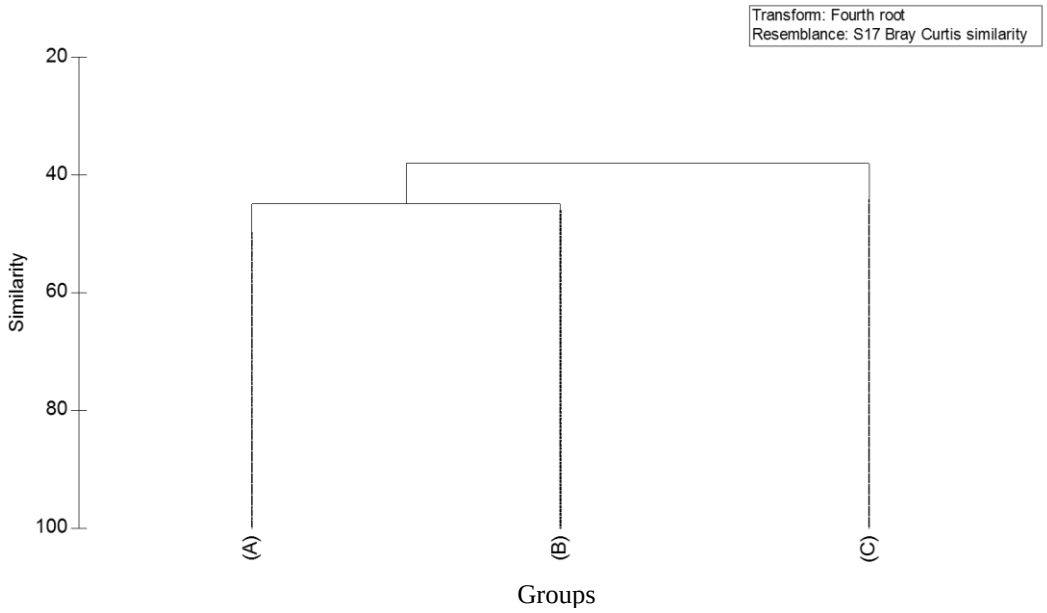
	Transformation	**No. Groups**	**No. Sites**
Attawapiskat	$\sqrt[4]{}$	3	65
Yukon	log (x+1)	4	204
Fraser	$\sqrt[4]{}$	6	270
Multi-basin	$\sqrt[4]{}$	5	617

2.3.2.1 *Classification*

Cluster analysis conducted for the models revealed grouping solutions ranging from 3 to 18 groups for all raw and transformed datasets. Further division of groups resulted in less than 10 sites per community group. The dendrograms from cluster analysis showed structural similarities among groups reaching 50% (Yukon and Fraser River basins, Figure 2.3b, Figure 2.4a) and the lowest similarities among groups at 20% (multi-basin model, Figure 2.4d). Outlier sites present at the beginning of the dendrogram (relatively low similarities) that formed groups smaller than 10 sites were removed. For the final models, total outliers removed were 2, 89, 55, and 68 for the Attawapiskat, Yukon, Fraser, and multi-basin models, respectively.

nMDS ordination of Bray-Curtis similarities showed distinct groups for each model, however some were more variable and overlapped (Figures 2.5 and 2.6). The Attawapiskat has the least among group overlap but the most variability within groups, probably a function of low number of reference sites (Figure 2.5a). nMDS ordination of Yukon sites reveals Groups A, B, and C overlap slightly while Group D is distinct from the other groups although more variable. The nMDS ordination (Figure 2.6a) of the 6 groups for the Fraser model showed significant overlap of Group C with all groups except for the substantially distinct Group F. Significant overlap is seen among groups for the multi-basin model. Groups B and C are quite variable and overlap with groups A, D, and E and may result in poor discriminations.

a)



b)

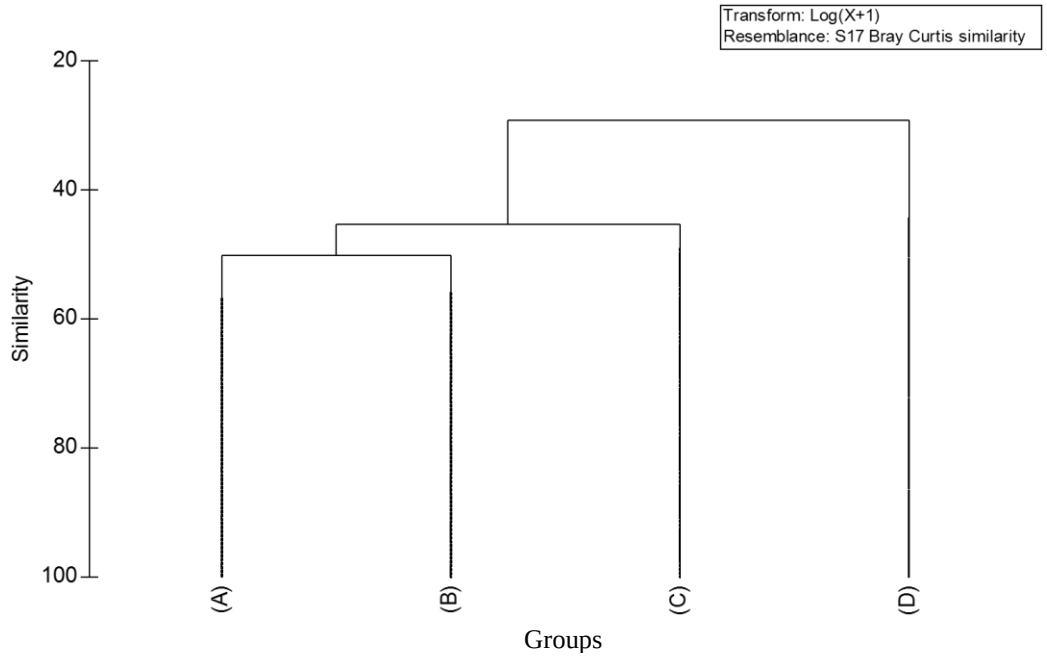


Figure 2.3: Dendrograms for hierarchical clustering (group averaging) based on Bray-Curtis similarities of (a) $\sqrt[4]{}$ transformed abundances for the Attawapiskat River basin, and (b) log(x+1) transformed abundances for the Yukon River basin. Each branch represents a significant (SIMPROF test at p=0.001) group of sites based on Bray-Curtis similarities.

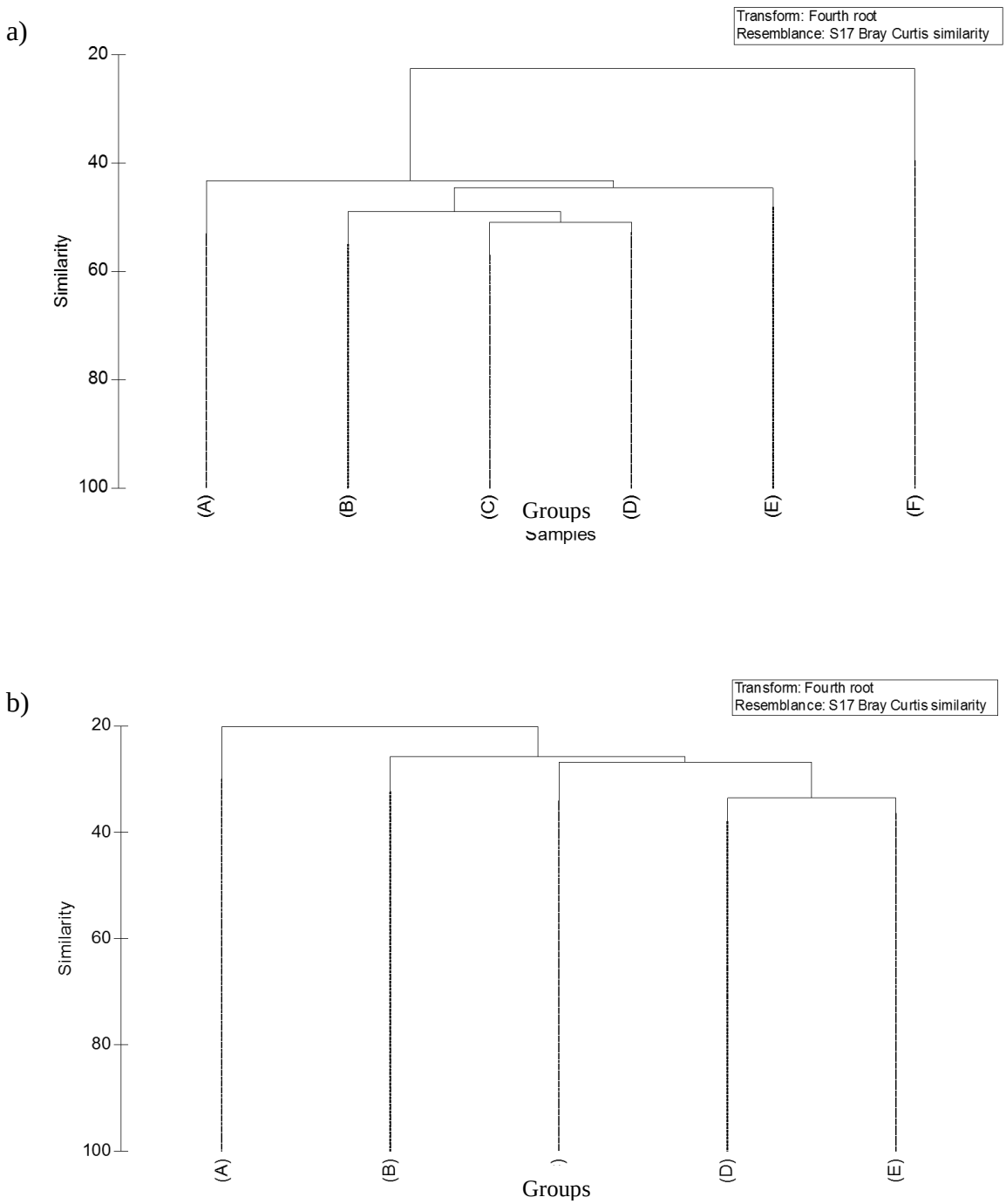
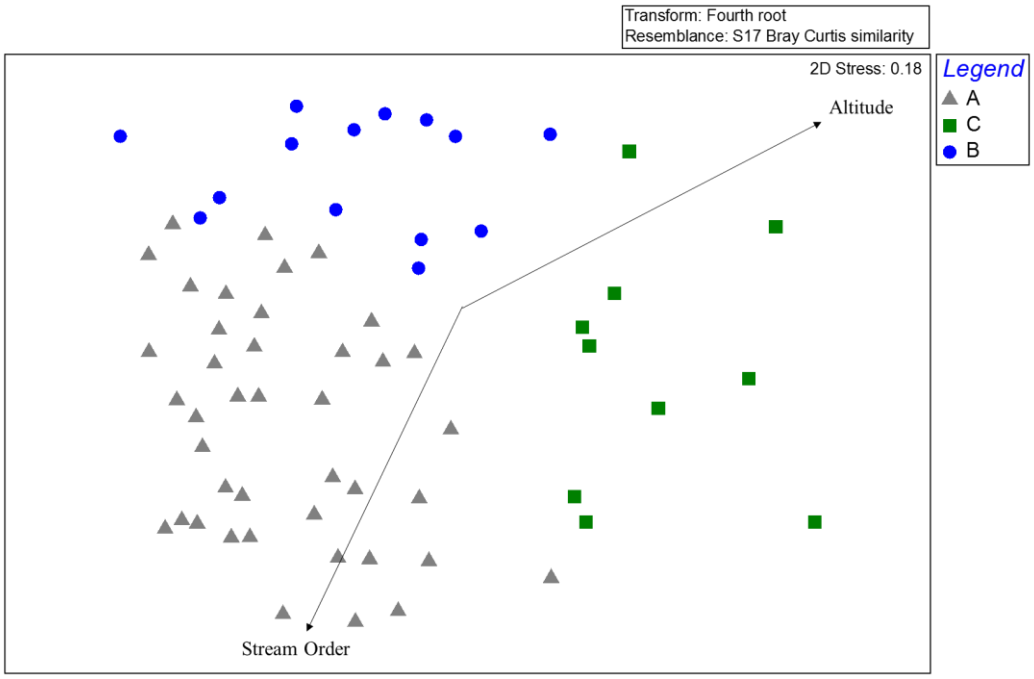


Figure 2.4: Dendrograms for hierarchical clustering (group averaging) based on Bray-Curtis similarities of (a) $\sqrt[4]{}$ transformed abundances for the Fraser River basin, and (b) $\sqrt[4]{}$ transformed abundances for the multi-basin model. Each branch represents a significant (SIMPROF test at p=0.001) group of sites based on Bray-Curtis similarities.

a)



b)

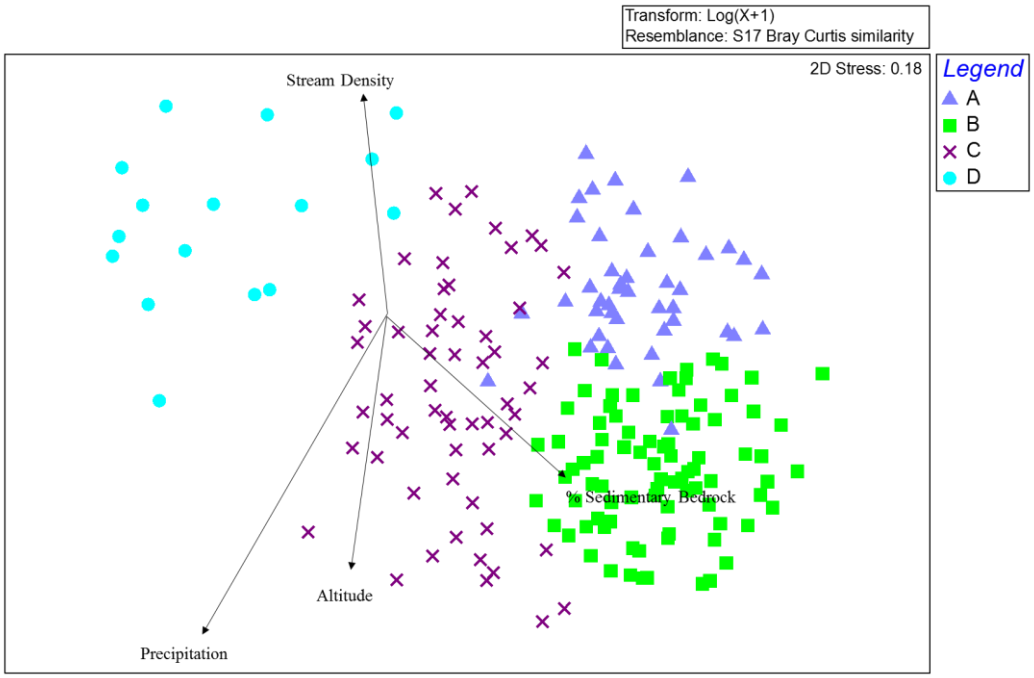
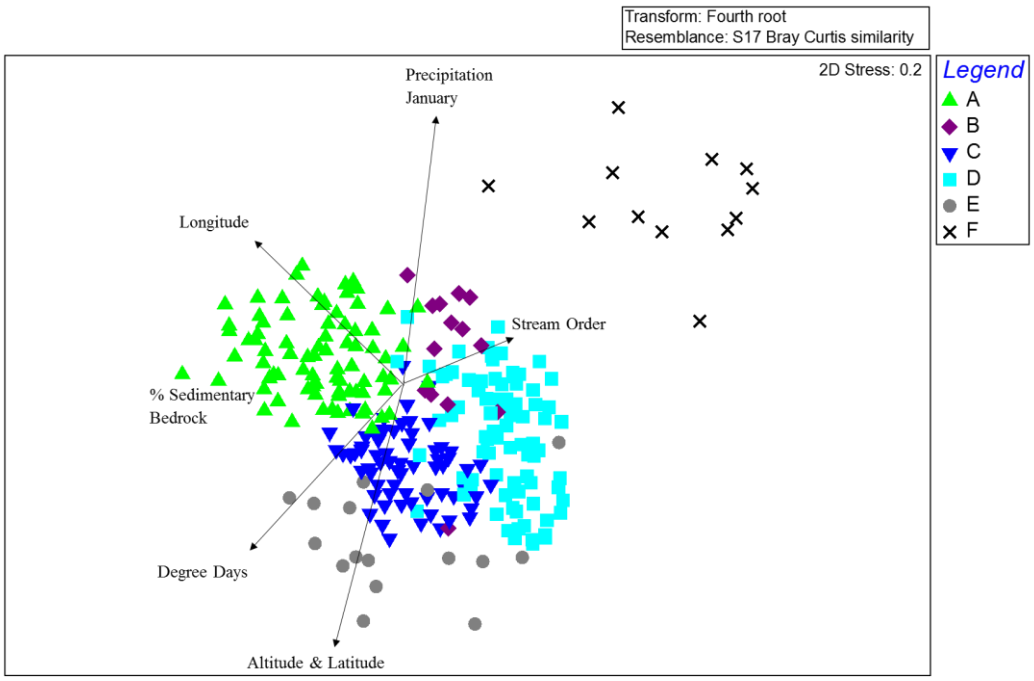


Figure 2.5: nMDS ordination of the (a) Attawapiskat River, and (b) Yukon River models' reference sites based Bray-Curtis similarities of the benthic communities (stress = 0.18). Arrows indicate the strongest Pearson correlation vectors.

a)



b)

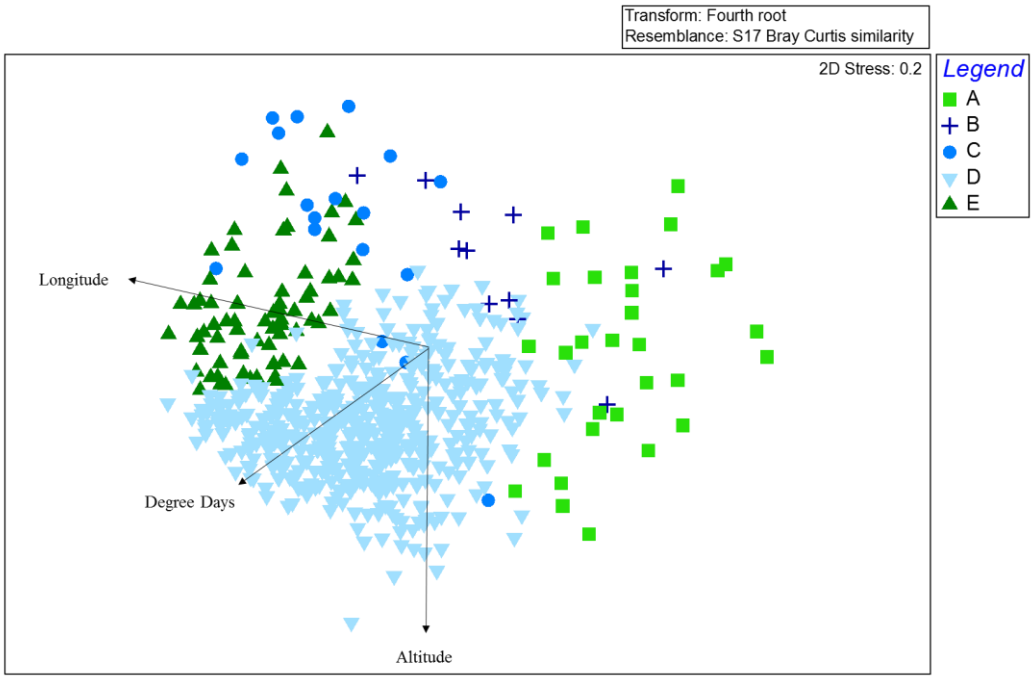


Figure 2.6: nMDS ordination of the (a) Fraser River, and (b) Multi-basin models' reference sites based Bray-Curtis similarities of the benthic communities (stress = 0.2). Arrows indicate the strongest Pearson correlation vectors.

Results of cluster analysis of the multi-basin model revealed group clustering by basin in groups B and C, whereas group similarities across basins were presented in groups A, D, and E (Figure 2.7). nMDS ordination of all the basins' reference sites reveals significant overlap in benthos community similarities, however clustering is apparent within basins. The Attawapiskat is the least variable in among-site Bray-Curtis similarities, whereas the Yukon has the most variability (Figure 2.8). Analysis of Similarities (ANOSIM) revealed there are statistically different but little differences among basin benthic communities with a global R = 0.185 (p=0.001).

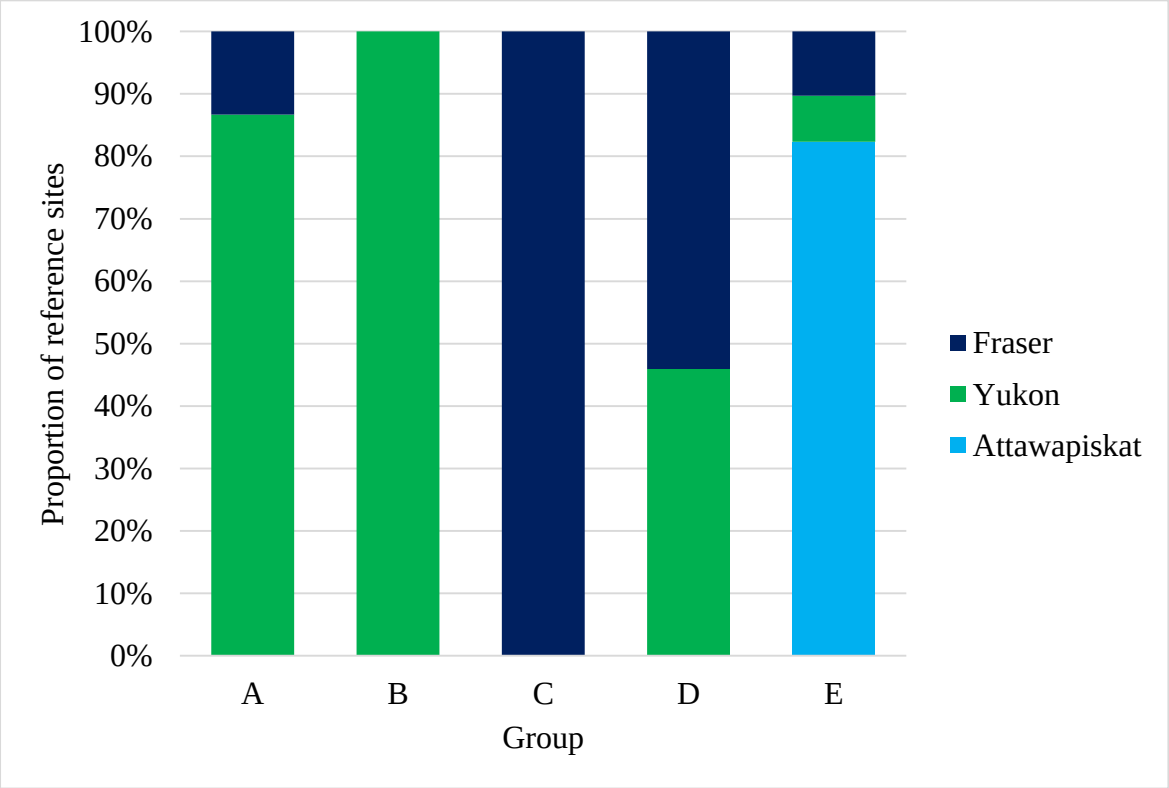


Figure 2.7: Proportion of basin sites within each community group in the multi-basin model.

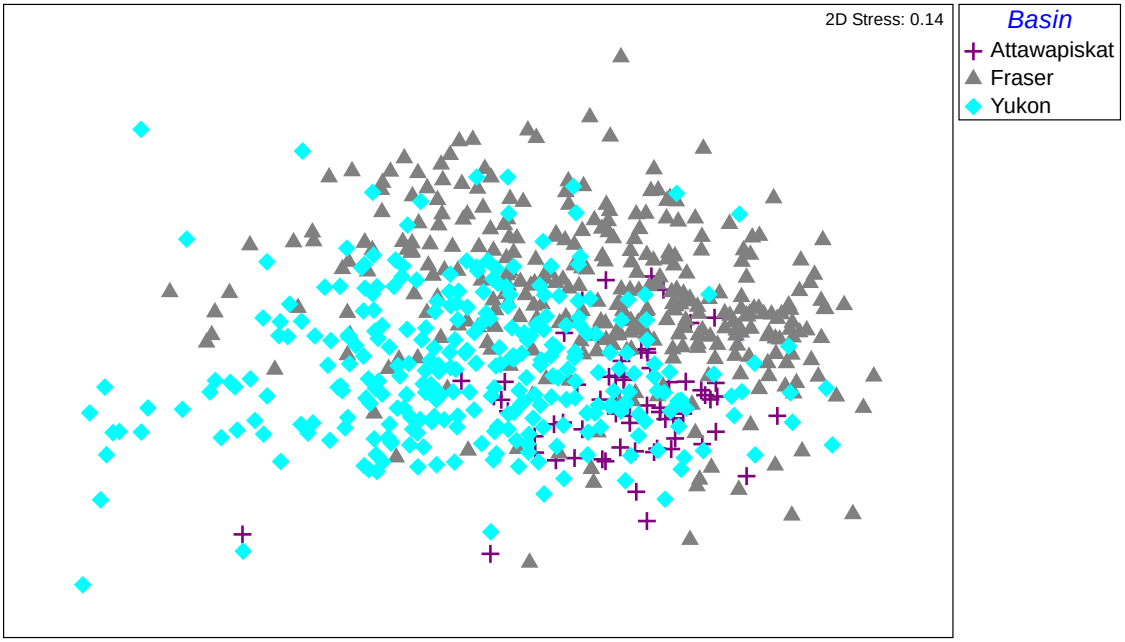


Figure 2.8: nMDS Ordination of Bray-Curtis similarities (based on raw abundances) for the Attawapiskat River, Fraser River, and Yukon River basins. Stress = 0.14. Analysis of Similarities (ANOSIM): global R = 0.185 (p=0.001).

2.3.2.2 *Discrimination*

Some potential models did not meet the homogeneous covariance matrices assumption so they were excluded as potential candidate models. DFA's were tested on all remaining potential grouping solutions. The final models selected through the rank-sum method (see Appendices D and E) have CV's ranging 56 – 72%, with Attawapiskat possessing the highest CV among models (only 1 potential model due to DFA assumptions) and the Yukon possessing the lowest CV (Table 2.8). The Attawapiskat model also possesses only 2 predictors however explains only 41% of variance in discriminant scores (Wilks' $\lambda = 0.594$). The Fraser model possesses the highest number of predictors which explains the most variance (94%) in discriminant scores of all models.

Table 2.8: Summary of DFA results for the 4 final RCA models. Only one model was possible for the Attawapiskat due to DFA assumptions.

	Cross validation (% accuracy)	**Number of predictors**	**Wilks' λ**
Attawapiskat	72	2	0.594
Yukon	56	8	0.419
Fraser	67	12	0.062
Multi-basin	65	11	0.109

The errors (CV) of groups was most balanced in the Yukon model ranging from 44 -58%, which may be attributed to the fairly balanced group sizes ranging 16 – 86 (Table 2.9). However, the Yukon model discrimination of sites from Group D with groups A and B was problematic for the model, reducing the overall model CV. The Attawapiskat model possessed the most balanced group sizes ranging 10 – 41 (highest mean:variance) whereas the multi-basin model had group sizes ranging 11 – 490 reference sites resulting in the lowest mean:variance ratio. The multi-basin model therefore has difficulty distinguishing Group B from groups A and D, and Group A from D.

Table 2.9: Jackknifed classification matrix of all 4 RCA models with cross-validation rates for each group. The percent of correctly classified reference sites is the proportion of reference sites correctly classified to the group it belongs.

			Predicted to:						
		No. Sites	**Group A**	**Group B**	**Group C**	**Group D**	**Group E**	**Group F**	**% Correct**
ATTAWAPISKAT	Group A	41	31	6	4	-	-	-	76
	Group B	14	1	12	1	-	-	-	86
	Group C	10	2	4	4	-	-	-	40
	Total	*65*	*34*	*22*	*9*	-	-	-	***72***
YUKON	Group A	44	24	4	7	9	-	-	55
	Group B	86	7	50	22	7	-	-	58
	Group C	58	5	16	33	4	-	-	57
	Group D	16	4	4	1	7	-	-	44
	Total	*204*	*40*	*74*	*63*	*27*	-	-	***56***
FRASER	Group A	81	65	0	12	1	1	2	80
	Group B	14	2	8	0	0	0	4	57
	Group C	65	11	2	32	7	13	0	49
	Group D	82	9	0	4	58	9	2	71
	Group E	16	1	0	5	1	9	0	56
	Group F	12	0	3	0	0	0	9	75
	Total	*270*	*88*	*13*	*53*	*67*	*32*	*17*	***67***
MULTI-BASIN	Group A	30	17	2	1	10	0	-	57
	Group B	11	3	5	0	3	0	-	45
	Group C	18	1	0	17	0	0	-	94
	Group D	490	146	8	27	309	0	-	63
	Group E	68	1	3	0	8	56	-	82
	Total	*617*	*168*	*18*	*45*	*330*	*56*	-	***65***

The predictors that best discriminate the biotic groupings for all models varied but each model possessed altitude and stream order as predictors. The Yukon, Fraser and multi-basin models shared predictors: stream density, % water cover, degree days and January precipitation (Table 2.10).

The Attawapiskat model revealed altitude as the lowest mean value in Group A (75.8 masl) and Group C with the highest (129.1 masl). Conversely, Group A had the highest mean stream order (2.5) and Group B had the lowest (1.1). The Pearson correlation vectors showed stream order discriminating in the negative y-axis direction and altitude in the positive x- and y-axis direction (Figure 2.5a).

Predictor variable means for the Yukon model showed that Group A is discriminated by lower than average altitude, lowest % water cover, and the lowest amounts of precipitation. Group B is characterized by the extremes: highest altitude, the lowest stream density, the highest stream order around 3, highest % sedimentary bedrock around 46%, and the highest precipitation. Group C is characterized by predictor variables that are similar to the dataset averages, with the exception of the lowest % sedimentary bedrock and the highest % water cover and degree days. Lastly, Group D is mostly characterized by lower than the data set average with the exception of having the largest stream density. Pearson correlation vectors of predictors discriminated groups with altitude, precipitation and % sedimentary bedrock in the negative y-axis direction, and stream density in the opposite direction (Figure 2.5b).

Results of predictor group means for the Fraser model showed Group A is characterized by sites located at the south western and north eastern periphery of the watershed with stream order averaging at 4, the lowest mean volcanic bedrock (7.5%), and similar to the dataset average for climatic variables. Group B sites are found in the south eastern part of the Fraser river watershed near the mouth of the Fraser river in lower-lying areas, have the lowest stream orders, have the highest sedimentary bedrock (49.4%), and higher precipitation than the overall average. Sites from Group C are at higher altitudes, have mean stream order of 3.5, sedimentary and volcanic bedrock around 30% each, and the lowest water cover (1%). Group D is characterized by sites

with the highest stream orders around 5, the lowest stream densities within site catchments, the highest water cover (2.3%) and the lowest summer precipitation. Group E sites are located in mountainous regions with the highest mean altitudes, stream orders around 3, the highest volcanic bedrock (50.9% average), the most degree days, and the lowest January precipitation. Lastly, Group F sites are also located near the mouth of the Fraser River in the lowest lying areas, have the highest catchment stream densities, the lowest sedimentary bedrock (0.3%), the least amount of degree days, and the most precipitation on average. Pearson correlations of the predictor variables (Figure 2.6a) indicate a gradient of increasing January precipitation up the y-axis, and altitude and degree days in the opposite direction. Stream order increases right along the x-axis and % sedimentary bedrock increasing left along the x-axis.

Group means of predictors for the multi-basin model showed that sites from Group A are found at higher altitudes, with the lowest mean stream density, lowest mean % water cover, and lowest mean temperatures. Group B is characterized by the most western locations in the Yukon, small site catchments, the highest mean % water cover and the lowest mean precipitation. Group C sites have the lowest mean altitudes, the largest catchments and stream density, the lowest mean degree days, and high mean precipitation and temperatures. Group D is characterized by the highest mean altitudes, and the highest mean steam orders (3.7). Lastly, Group E is characterized by the lowest mean steam orders (2.4), the most amount of degree days, and the highest precipitation in July. The strongest Pearson correlations shown in Figure 2.6b exhibit a gradient of increasing altitude down the y-axis, increasing degree days towards the plot origin and increasing longitude towards groups C and E.

Table 2.10: Summary of environmental predictor variables for each RCA model. Values represent mean values within each group. Altitude = meters above sea level, stream density = stream length (m)/catchment area (km²), precipitation = mm, temperature = °C, and bedrock = % of catchment.

Model	Predictors	Group A	Group B	Group C	Group D	Group E	Group F	Total
Attawapiskat	Altitude	75.8	122.4	129.1	-	-	-	94.0
(2 predictors)	Stream Order	2.5	1.1	1.7	-	-	-	2.1
Yukon	Altitude	661.9	860.4	619.2	727.3	-	-	738.6
(8 predictors)	Stream Density	1025	856.3	916.5	1155.1	-	-	933.2
	Stream Order	3.1	3.3	3.1	2.8	-	-	3.1
	Sedimentary Bedrock	45.4	46.7	16.4	18.6	-	-	35.6
	% Water Cover	0.05	0.3	0.9	0.07	-	-	0.4
	Degree Days	111.2	118.3	123.4	104.1	-	-	117.1
	Precipitation January	28.6	38.9	33.9	34	-	-	34.9
	Precipitation June	42.5	65	57.6	51.2	-	-	57
Fraser	Latitude	51.3	49.2	52.5	51.6	52.4	49.3	51.6
(12 predictors)	Longitude	-122.2	-122	-122.6	-122.6	-123.1	-122.4	-122.6
	Altitude	694.8	102.5	1117.8	866.3	1283.6	30.7	823.4
	Stream Order	4.1	2.6	3.5	5	3.1	2.9	4
	Stream Density	2015.3	1795.6	2062.1	1733.9	1816.7	2960.8	1960
	Sedimentary Bedrock	37	49.4	32.4	5.5	26.9	0.3	24.8
	Volcanic Bedrock	7.5	18	28.5	45.4	50.9	19.3	27.7
	% Water Cover	1.5	1	1	2.3	1.8	1.5	1.6
	Degree Days	139	97	141.9	136.7	142.1	78.8	134.3
	Precipitation January	144	258.1	78.4	88.8	60.9	301.4	119.4
	Precipitation July	75.5	72.1	68.7	56	59.6	80	67
	Precipitation September	71.3	68.7	58.2	50.6	55.1	77	61
Multi-basin	Longitude	-136.2	-137.6	-122.3	-129.1	-92.3	-	-125.3
(11 predictors)	Altitude	751.2	479.6	57.1	800.0	261.5	-	710.9
	Drainage Area	1139.3	256.1	1595.4	740	268.5	-	723.8
	Stream Density	1179.7	1226.1	3030.1	1460.8	1222.7	-	1462.5
	Stream Order	3.1	2.5	3.4	3.7	2.4	-	3.5
	% Water Cover	0.3	6.1	1.9	1	4.8	-	1.5
	Degree Days	118	118	84.3	128.3	148.9	-	128.6
	Precipitation January	42.3	24.9	289.6	75.2	33.3	-	74.3
	Precipitation July	70.1	61.9	80.5	68.4	84.7	-	70.5
	Temperature July Minimum	4.9	7.4	10	5.4	8.3	-	5.8
	Temperature May Maximum	8.5	10.2	15.2	9.9	11.1	-	10.1

2.3.2.3 *Group Community Characteristics*

The community groupings ranged in Bray-Curtis similarities within groups from 35 – 61%. Similarities for the multi-basin model groups were the lowest averaging at 42% similarity while the Yukon was the highest at 57% (Table 2.11). Analyzing group characteristics for families contributing at least 5% to within group Bray-Curtis similarity for each model, the Attawapiskat Group A is differentiated by the highest mean abundance and richness (Table 2.11), the presence of riffle beetles (Elmidae), and water mites (Hygrobatidae) (Figure 2.9a). Group B is characterized by the lowest mean abundance and richness, but possesses the highest abundances of families in Diptera. Group C is most distinct from the other groups consisting of pea clams (Pisidiidae) with the second highest relative abundance, Ceratopogonidae (biting midges), Enchytraedae (oligochaetes), and Dytiscidae (predaceous diving beetles) contributing at least 5% to group similarity.

Groups in the Yukon model are distinguished by a gradient of mean abundances, the highest in Group A with 4000 mean individuals and decreases through to 138 in Group D. The groups have similar dominant community composition however Group D is distinguished by consisting of only 3 families within 90% of within-group similarity and the highest contribution of Chironomidae with a relative abundance of 75% (Figure 2.9b). Groups A, B, and C are distinguished by the contribution of Capniidae in Group A and greater than 5% contribution to within group similarity of Perlodidae in Group B.

Groups in the Fraser model are characterized by an increasing relative abundance gradient from groups A through F of Diptera (including Chironomidae) (Figure 2.9c) and mean abundances excluding Group F (see Table 2.11). Groups A through E are similar in dominant taxa (EPT

taxa), with varied abundances within each group. Group F is differentiated by 3 families of oligochaetes (34% relative abundance) and mites (Limnesiidae) with 1.6% relative abundance.

The biotic structures of the groups for the multi-basin model are more variable in abundance and presence of taxa with within group similarities averaging around 40% (Table 2.11). Group A has the lowest mean abundance and richness, with only 2 families, Chironomidae and Nemouridae, contributing greater than 80% of relative abundance (Figure 2.9d). Group B is differentiated by the presence of blackflies (Simuliidae), snails (Valvatidae and Lymnaeidae), and pea clams (Pisidiidae). Group C consists of a high relative abundance of midges (60%), 2 families of oligochaetes, and 1.2% relative abundance of mites. Chironomids, mayflies and stoneflies contributed the most to within group similarity in Group D, ranging from 13 - 33%. Lastly, Group E has the highest abundance and richness and predominantly consists of Chironomidae and mayflies.

Table 2.11: Summary of biotic group characteristics for each final RCA model. B-C Similarity is the mean Bray-Curtis within-group similarity determined from the SIMPER test in PRIMER.

	Group	B-C Similarity ($\bar{x}$)	Abundance ($\bar{x}$) ± SE	Family Richness ($\bar{x}$) ± SE
Attawapiskat	A	54.3	4802.8±436.9	27.7±0.8
	B	52.7	2163.3±332.2	22.6±1.4
	C	47.8	2705.3±605.7	23.7±1.4
Yukon	A	60.9	4008.5±732.8	10.6±0.4
	B	59.8	1765.6±157.1	16.7±0.4
	C	53.5	343.3±38.8	10.3±0.4
	D	53.6	138±38.8	5.5±0.5
Fraser	A	57.9	764.3±75.7	15.3±0.5
	B	57.8	1451.6±285.3	21.2±1.1
	C	59.7	6946.3±664.8	17.2±0.4
	D	57	9544.3±989.2	19.2±0.4
	E	53.2	10222.6±1628.0	12.6±0.8
	F	48.6	5744.6±1512.0	11.2±0.9
Multi-basin	A	35.2	85.5±15.7	5.5±0.5
	B	41.8	346.8±146.3	12.4±0.9
	C	42.2	5682.4±1450.9	11.5±0.8
	D	44.2	3868.4±266.7	15.4±0.2
	E	44.5	5986±749.9	24.5±0.8

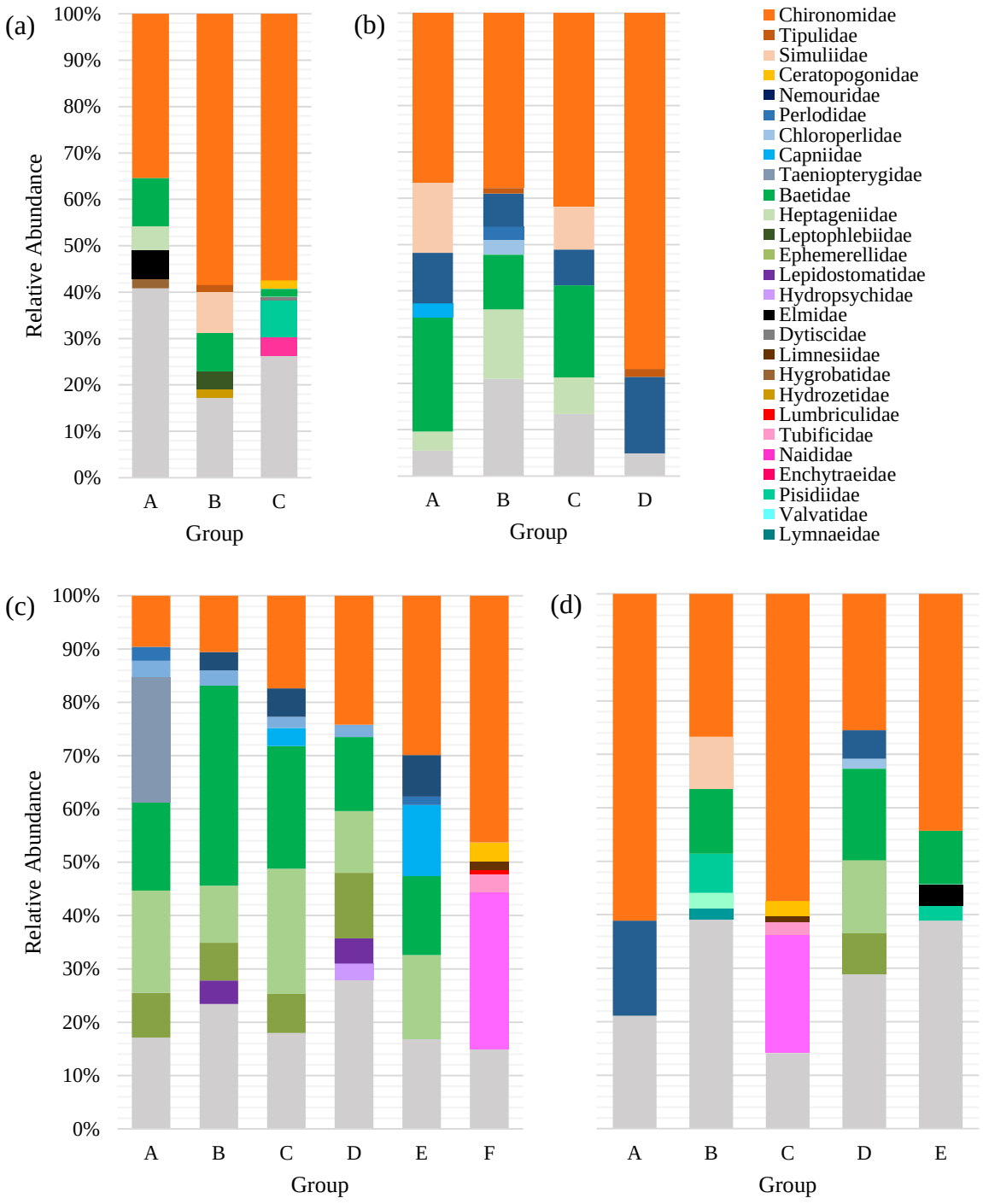


Figure 2.9: Community composition represented as relative abundance of families contributing at least 5% to Bray-Curtis within group similarities for the (a) Attawapiskat River, (b) Yukon River, (c) Fraser River, and (d) Multi-basin models.

2.3.3 Assessment evaluation

Models were evaluated and compared by determining the chance of committing type 1 and 2 errors associated with each model. Type 1 errors were determined for each group and as a total model error using undisturbed simpacted sites and are presented as a percentage in Table 2.12. Type 1 errors at the 90% and 75% confidence levels were the highest for the multi-basin model (8.6 and 18.6%, respectively). Conversely, the Attawapiskat and Fraser models had the lowest type 1 error at both 90% and 75% confidence levels.

Type 2 errors were assessed using simpacted data at each level of disturbance and presented in Table 2.13. At the 90% confidence interval, type 2 errors were highest for the multi-basin model overall, whereas the Fraser model had the lowest errors across all 3 levels of disturbance. At the 75% confidence level, the sensitivities were more variable among models, however overall the multi-basin model had the greatest type 2 errors.

Table 2.12: Type 1 errors of all 4 RCA bioassessments using undisturbed simpacted sites. Type 1 errors are represented as the proportion of simpacted sites that fall outside the reference ellipses at the 90% and 75% confidence level for each biotic group. "-" = group not present.

	90%				**75%**			
	Attawapiskat	Yukon	Fraser	Multi-basin	Attawapiskat	Yukon	Fraser	Multi-basin
Group A	0.0	0.0	0.0	0.0	16.7	14.3	11.1	0.0
Group B	0.0	0.0	0.0	0.0	0.0	15.4	0.0	100.0
Group C	0.0	25.0	0.0	0.0	0.0	25.0	28.6	0.0
Group D	-	0.0	0.0	8.9	-	0.0	11.1	17.9
Group E	-	-	0.0	12.5	-	-	0.0	25.0
Group F	-	-	0.0	-	-	-	0.0	-
Total	*0.0*	*6.7*	*0.0*	*8.6*	*10.0*	*16.7*	*13.3*	*18.6*

Table 2.13: Type 2 errors for all 4 RCA bioassessments at mild, moderate, and severe levels of disturbance. Type 2 errors are represented as the proportion of simpacted sites that fall inside the reference ellipses at the 90% and 75% confidence levels. Errors are averaged across each biotic group.

		90%			**75%**		
	n	**Mild**	**Moderate**	**Severe**	**Mild**	**Moderate**	**Severe**
Attawapiskat	10	100.0	50.0	10.0	80.0	30.0	0.0
Yukon	30	90.0	66.7	13.3	66.7	36.7	3.3
Fraser	30	90.0	30.0	0.0	70.0	10.0	0.0
Multi-basin	70	92.9	72.9	30.0	75.7	48.6	15.7

2.3.4 Model comparisons

The Attawapiskat had the highest classification performance (72% CV), was the most parsimonious (2 predictor variables), and had the lowest type 1 errors (both 90% and 75% confidence intervals). The multi-basin model does not possess the highest prediction performance however is comparable to single basin models (Table 2.14). The Attawapiskat was the best at detecting deviations from reference (lowest type 2 errors) at the 90% confidence level, but the Fraser model has the lowest type 2 errors at the 75% confidence level. The multi-basin model was the least sensitive (highest type 1 and 2 errors) overall.

Table 2.14: Model comparisons based on prediction performance, parsimony, and sensitivity. Prediction performance is the cross validation %, parsimony is the number of predictors, and sensitivity was ranked in order from lowest (1) to highest (4) for type 1 and type 2 errors at the 90% and 75% (in parentheses) confidence levels.

	Prediction performance	**Parsimony**	**Type 1 errors 90% (75%)**	**Type 2 errors 90% (75%)**
Attawapiskat	72	2	2 (3)	2 (3)
Yukon	56	8	2 (2)	3 (2)
Fraser	67	12	1 (1)	1 (1)
Multi-basin	65	11	3 (4)	4 (4)

As discussed in the methods, a subset of simpacted sites were compared between each single basin bioassessment to the multi-basin bioassessment to assess performance and sensitivity (see

example in Figure 2.10). The results are summarized in Table 2.15 as a matrix comparing the multi-basin bioassessment as rows and each single basin bioassessment as columns. Type 1 errors using undisturbed simpacted sites revealed that the Attawapiskat assessments have the same errors as the multi-basin assessments because both assessed 9 out of 10 (90%) simpacted sites as reference and possess equal errors at the 75% level with 1 simpacted site (10%) outside of reference. Type 1 errors were higher for the Yukon assessments (83.3% in reference) compared to the multi-basin assessments (86.7% in reference), and lower for the Fraser assessments (86.7% in reference) compared to the multi-basin assessments (73.3% in reference). For type 2 errors, the errors varied among bioassessments and disturbance levels. For example, comparing the multi-basin assessments to the Attawapiskat assessments at mild disturbance, both models agree that 70% of simpacted sites are within the 75% confidence ellipse, or in reference condition, denoted “Ref”. Bold text indicates where models agree, everywhere else there is a disagreement. Following the same example, a disagreement exists where 20% of mildly disturbed simpacted sites fell in the reference ellipse (“Ref”) in the multi-basin assessments but those same sites fell between the 75% and 90% confidence ellipses, denoted “75%” in the table, in the Attawapiskat assessments. This illustrates a lack of sensitivity of the multi-basin assessments for these sites compared to the Attawapiskat assessments (see also overall type 2 errors in Figure 2.11a). Comparisons of simpacted site assessments among the Attawapiskat and multi-basin indicate strong concordance (90%) among assessments to detect severe impairment, i.e. falling outside the 90% confidence ellipse (Figure 2.11c). At moderate disturbance the simpacted sites from the Attawapiskat basin are slightly better detected in the multi-basin assessment (90% of simpact sites fall outside the 75% confidence ellipse) than the Attawapiskat assessment (only 70% fall outside the ellipse) (see Table 2.15 and Figure 2.11b). Simpact sites

from the Yukon basin deviated more from reference (outside the 75% confidence interval) in the single Yukon basin at mild (33.3%), moderate (63.3%), and severe (96.7%) disturbance, than the multi-basin assessments (23.5%, 46.6%, and 86.7 for mild, moderate, and severe, respectively) (Figures 2.11d-f). Lastly, Fraser simpact sites are better detected in the Fraser basin assessment at all levels of disturbance. The type 2 errors at mild disturbance are comparable for the Fraser assessment and the multi-basin assessment (Figure 2.11g), whereas significantly more simpact sites deviated outside of the 90% confidence interval at moderate (21 simpact sites out of 30) (Figure 2.11h) and severe (30 simpact sites out of 30) disturbance (Figure 2.11i) compared with the multi-basin assessment (7 and 18 simpact sites out of 30 at moderate and severe disturbance, respectively).

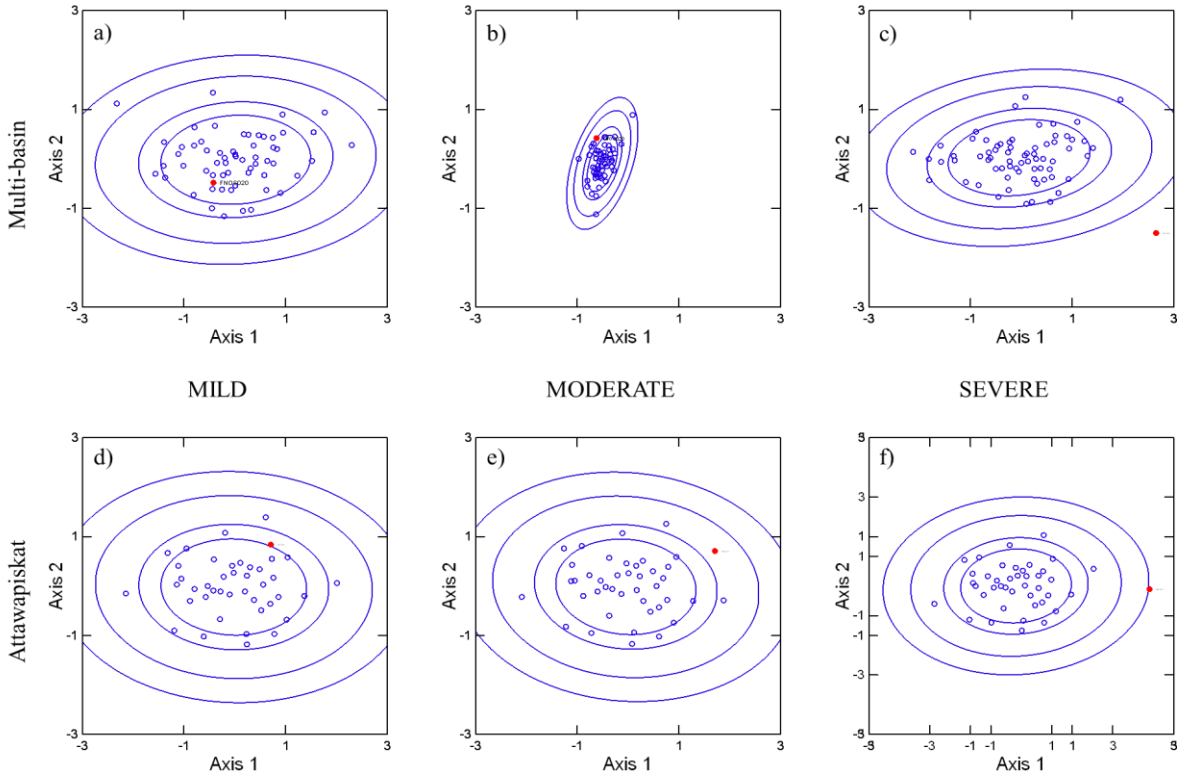


Figure 2.10: Example simpacted site assessments using the multi-basin bioassessment at a) mild, b) moderate, and c) severe disturbance, and the Attawapiskat bioassessment at d) mild, e) moderate, and f) severe levels of disturbance. nMDS scores of the reference and simpacted site are plotted with 75, 90, 99, and 99.9 confidence ellipses. Hollow blue circles represent reference sites and single solid red circle represents simpacted site.

Table 2.15: A matrix comparing the multi-basin assessments for type 1 and 2 errors to each single basins' assessments. A subset of simpact sites from each basin was used to assess type 1 errors with undisturbed simpact sites and type 2 error with mild, moderate, and severe disturbance simpact sites for each group within the models. Simpact sites either fell in the 75% reference ellipse denoted "Ref", between the 75% and 90% ellipses denoted"75%", or outside the 90% confidence ellipse denoted "90%". Values shown are summed across all biotic groups and presented as percent (the number of simpacted sites that fall outside (type 1) or within (type 2) their respective ellipse over the total number of sites assessed). Bold text indicates where simpacted sites lie within the same confidence ellipse for both models.

			Attawapiskat (*n*=10)				Yukon (*n*=30)				Fraser (*n*=30)			
			Ref	75	90	Total	Ref	75	90	Total	Ref	75	90	Total
Multi-Basin	Undisturbed	Ref	**80.0**	10.0	0.0	90.0	**76.7**	10.0	0.0	86.7	**66.7**	6.7	0.0	73.3
		75.0	10.0	**0.0**	0.0	10.0	6.7	**0.0**	0.0	6.7	10.0	**3.3**	0.0	13.3
		90.0	0.0	0.0	**0.0**	0.0	0.0	0.0	**6.7**	6.7	10.0	3.3	**0.0**	13.3
		Total	90.0	10.0	0.0	**100.0**	83.3	10.0	6.7	**100.0**	86.7	13.3	0.0	**100.0**
	Mild	Ref	**70.0**	20.0	0.0	90.0	**53.3**	20.0	3.3	76.7	**56.7**	10.0	6.7	73.3
		75.0	10.0	**0.0**	0.0	10.0	13.3	**3.3**	0.0	16.7	6.7	**6.7**	3.3	16.7
		90.0	0.0	0.0	**0.0**	0.0	0.0	0.0	**6.7**	6.7	6.7	3.3	**0.0**	10.0
		Total	80.0	20.0	0.0	**100.0**	66.7	23.3	10.0	**100.0**	70.0	20.0	10.0	**100.0**
	Moderate	Ref	**10.0**	0.0	0.0	10.0	**33.3**	13.3	6.7	53.3	**6.7**	13.3	30.0	50.0
		75.0	0.0	**0.0**	10.0	10.0	3.3	**6.7**	13.3	23.3	3.3	**6.7**	16.7	26.7
		90.0	20.0	20.0	**40.0**	80.0	0.0	10.0	**13.3**	23.3	0.0	0.0	**23.3**	23.3
		Total	30.0	20.0	50.0	**100.0**	36.7	30.0	33.3	**100.0**	10.0	20.0	70.0	**100.0**
	Severe	Ref	**0.0**	10.0	0.0	10.0	**3.3**	3.3	6.7	13.3	**0.0**	0.0	16.7	16.7
		75.0	0.0	**0.0**	0.0	0.0	0.0	**3.3**	3.3	6.7	0.0	**0.0**	26.7	26.7
		90.0	0.0	0.0	**90.0**	90.0	0.0	3.3	**76.7**	80.0	0.0	0.0	**56.7**	56.7
		Total	0.0	10.0	90.0	**100.0**	3.3	10.0	86.7	**100.0**	0.0	0.0	100.0	**100.0**

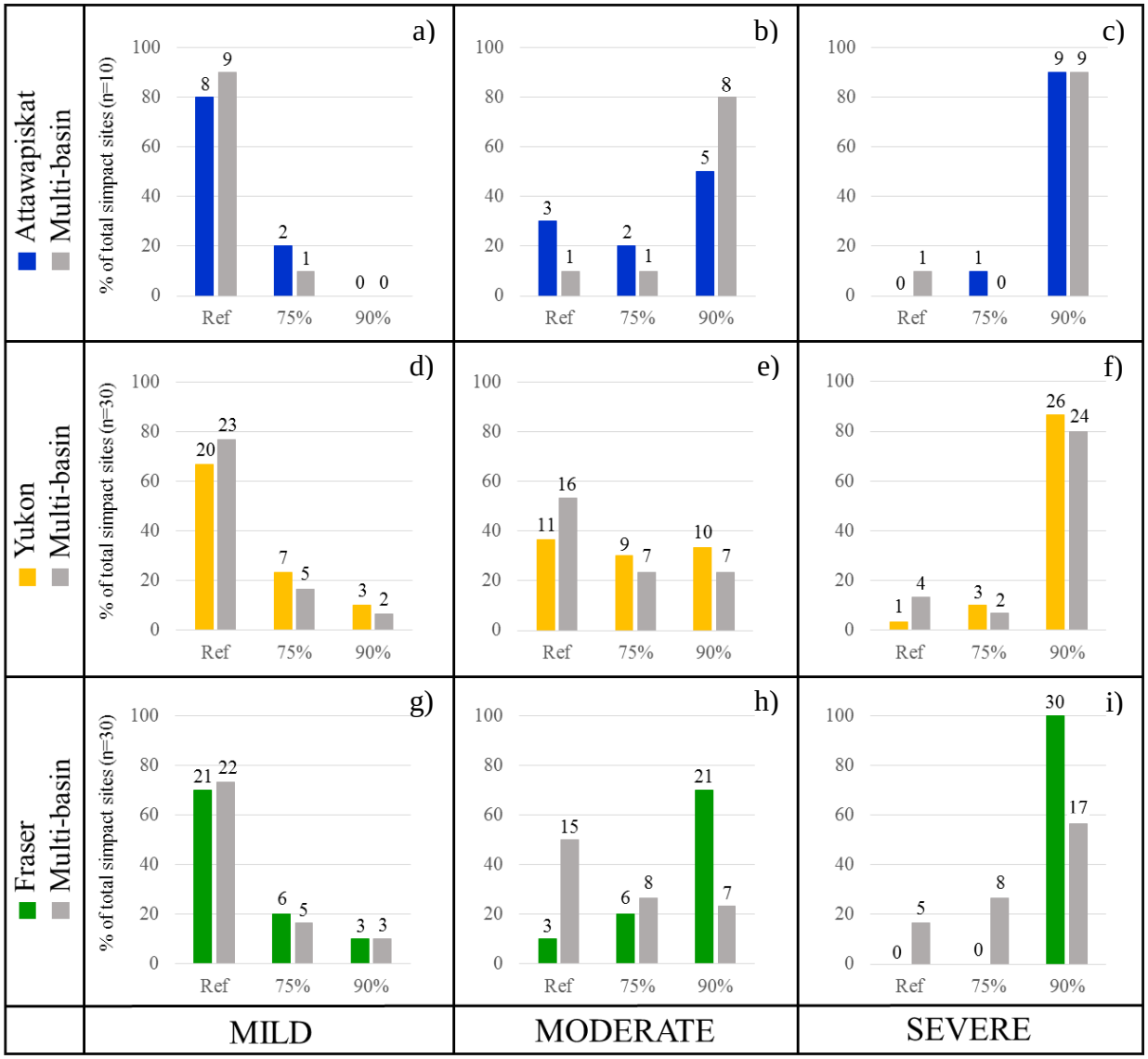


Figure 2.11: Comparisons of type 2 errors of the multi-basin model to the a-c) Attawapiskat, d-f) Yukon, and g-i) Fraser basin bioassessments. Type 2 errors are represented as the proportion of simpact sites at mild, moderate, and severe levels of disturbance that fall either within the 75% reference ellipse denoted "Ref", between the 75% and 90% ellipses denoted "75%", or outside the 90% confidence ellipse denoted "90%". The total number of simpact sites that fell within the 3 confidence intervals are presented above each bar. A total of 10, 30, and 30 simpact sites were assessed at each level of disturbance in the Attawapiskat, Yukon, and Fraser bioassessments, respectively.

2.4 Discussion

Application of the BEAST bioassessment method across basins provided a unique opportunity to observe the capabilities and limitations of this modelling approach in Canada. I examined the prediction performance, parsimony, and sensitivity of models at the basin and multi-basin level and compared my results to other studies with similar objectives.

2.4.1 Concordance of biotic assemblages

Concordance of sites among basins was assessed with the assumption that reference sites are relatively similar in structure within and adjacent to these basins if we have enough sites to capture the range of variability. In contrast to multimetric indices and some RIVPACS-type models, the BEAST method makes no *a priori* assumptions of structural patterns of biota relating to environmental features such as grouping by ecoregion (Barbour et al. 1996) or typology (Aroviita et al. 2009). Predictive modelling is concerned with characterizing the biological variability in reference sites but not whether the structure of communities are similar (Herlihy et al. 2008). In light of this, I assessed the similarities of all reference sites in nMDS ordination space and also determined the relative proportions of basin sites classified into the 5 groups of the multi-basin model. My results revealed significant overlap of basin reference sites with very little differences among basins apparent in the nMDS ordination.

The proportions of basin sites across the multi-basin model groups revealed that unique communities as well as similar communities exist across basins. As expected, some groups consisted of sites from a single basin because similarities among sites are influenced by the same environmental gradients when in close proximity (Hawkins et al. 2000). Gerritsen et al. (2000) also found strong ecoregional similarities when using clustering techniques. Hawkins et al. (2000) showed biotic variation to be partitioned best by ecoregions that differed in topography

(Feminella 2000, Marchant et al. 2000) and climate (Sandin and Johnson 2004), which are factors that indirectly influence biota. The other 3 groups in the multi-basin model comprised either two or all basins' reference sites, reflecting similarities found among basins. The concordance of similarities among basins support my objectives to uncover the relationships between the environment and biological communities that form the basis of my across-basin predictive model.

2.4.2 Prediction Performance

The prediction performance of a model is a critical measure of whether a test site is matched to the correct reference community group (Bailey et al. 2004, Ode et al. 2008, and references therein). If a site is not assigned to the correct group, this results in an inaccurate or incorrect assessment of the effects of potential stressors to the benthic community. I found that the Attawapiskat River model, developed using only 65 reference sites, to possess the best prediction performance of all models. The lower prediction performance of the other 3 models may be attributed to the larger number of reference sites within the model, larger group sizes within the model, as well as the number of biotic assemblage groups. A greater number of reference sites in a model will increase the amount of variability needing to be accounted for by the model. Wright (1995) and (Reynoldson and Wright 2000) found that prediction to the correct community group would decline with an increase in reference sites due to overlap in the biotic groups because of the increased variability. My multi-basin model exhibits this problem with uneven distribution of group sizes (11 – 490) and a large number of reference sites in the model (617). Community Group B consisting of only 11 reference sites had the lowest CV at 45%, which may be attributed to insufficient reference sites characterizing the group as well as misclassification to groups A and D (Table 2.9). This misclassification is due to larger variability of sites within

Group B and considerable overlap with other groups in the nMDS ordination (Figure 2.4d). Conversely, Group D has 490 reference sites with a CV of 63%, which is not as high as other groups in this model. The large variability to explain and the overlap with group A can be attributed to this reduction in prediction performance of this model.

The Yukon model, possessing the lowest prediction performance also contains a group of 16 sites and a CV of 44%. Misclassifications to groups A and B are occurring however there is no overlap apparent in the community groups (Figure 2.4b). Reduced prediction performance for this group may be attributed to poor predictors discriminating the groups (overlapping range of values for several predictors) and not explaining all the variance (Wilks' $\lambda = 0.419$). Reynoldson et al. (2001) points out that the lower diversity of reference sites might affect the distinction of groups because of less variability in habitat types, which is the case for the Yukon reference sites in relation to the Attawapiskat and Fraser River basins.

Similarities among groups are expected when a continuum of biological data is divided into groups (an artificial construct to partition variability); and such overlap among groups is expected in both the biological data and the predictors that discriminate these groups (Reynoldson et al. 1995). Grouping communities creates an artificial separation of a continuum, but is a requirement when using DFA. Although classification performance of DFA aims for an optimum between within group similarities and between group differences, it is dependent on the inherent characteristics of the data. I suggest that if a test site has a fairly high probability of belonging to >1 groups then to assess a test site for both groups to provide a more complete assessment. Some studies suggest the use of only common taxa which otherwise adds noise to the models (Marchant 2002, Van Sickle et al. 2007) or cause no effect to model precision (Mykrä et al. 2008), but such exclusions may reduce detection of human impact (Cao et al. 1998, 2001).

Differences in how well the models classify sites among basin models may also result from the definition of “reference condition” across basins. It is apparent that the status of sites in more remote regions such as those in the Yukon and Attawapiskat may be more in “reference” than sites found in the Fraser River basin. Although this may cause biases within and among the models (Herlihy et al. 2008, Yuan et al. 2008) I realize the implications and it is my objective to characterize the least disturbed condition possible to compare with potentially more disturbed conditions in the future.

Prediction performance is also dependent on the parsimony of a model, and whether the predictor variables explain enough of the variance to be effective. Choosing the right predictors and analyzing the relationships between environmental variables and biological structure is vital to a model with good prediction performance and parsimony.

2.4.3 Environmental Predictors

The different models applied in this study highlight how basin level and multi-basin level environmental patterns influence the biota observed at reference sites. The scope of this study limited predictor variables to landscape-scale variables, and although it would be beneficial to include reach-scale environmental patterns, landscape-scale variables provide reasonable explanations of the biological communities present (Corkum 1989, Allan et al. 1997). Landscape-scale predictor variables are also easily obtainable through currently available geographic information system (GIS) software and geospatial data. Studies show that environmental descriptors are not important at one specific scale (i.e. the use of site- and landscape-scale variables are not restricted to one scale of assessment such as local and regional) because of their interdependence and the complexity of environmental gradients across large spatial extents (Frissell et al. 1986, Sandin and Johnson 2004, Mykrä et al. 2007, Ode et al.

2008). Mykrä et al. (2007) found stream size and water acidity important at both local and regional scales however variables that best explained community structure were scale-and geographic location-dependent. Local habitat characteristics are influenced by other factors working at various regional scales (Richards et al. 1996, Allan et al. 1997), exemplifying the difficulties in pin-pointing environmental factors that are important at a specified spatial scale. My results support these findings that there are no significant differences among predictors at the basin level versus multi-basin level, with the exception of drainage area and temperature averages as predictors in the multi-basin model. Mykrä et al. (2008) also found catchment area as a significant predictor but for both ecoregion and across ecoregion models. Landscape scale variables have been found to be good predictors for large geographic regions (Corkum 1989, Mykrä et al. 2007). Catchment area and temperature predictors in my multi-basin model are unique to this model perhaps because these predictors characterize a larger region encompassing many factors that influences a site at a local scale (versus site-scale habitat variables which are more directly related and site specific). Therefore, biological patterns across large spatial extents may be best captured using landscape-scale environmental predictors.

2.4.4 Parsimony

Parsimony was measured in this study as the number of and how much variance is explained by the predictors. The highest parsimony of the Attawapiskat model with only 2 predictors is most likely due to a smaller range of variation (least amount of reference sites) captured by the predictors. However, these two predictors only explained 41% of variance explained in discriminant scores. Other candidate models for the Attawapiskat explained more of the variance with more predictors but possessed lower prediction performance (see Appendix D). Models with more reference sites and therefore more variance to capture, resulted in more predictors

used: the Fraser and multi-basin models best explained the variance ($\lambda = 0.062$ and 0.109, respectively) with 12 and 11 predictor variables, respectively. The variance not explained within all models is most likely a result of the restricted pool of candidate predictor variables. Models tailored for a specific stressor will increase the number of candidate predictors, and therefore potentially accounting for unexplained variance within the models. A number of large-scale environmental descriptors have been found to be good predictors such as forest cover (Carlisle et al. 2009), stream slope (Mykrä et al. 2008), and wetlands (% or presence) (Richards et al. 1996, Bennett 2011) and may decrease the unexplained variance within my models. I recommend developing models that target a single stressor of concern to increase the pool of available candidate predictors; such a model may possess greater parsimony and sensitivity.

2.4.5 Sensitivity

2.4.5.1 *Type 1 errors*

The type 1 errors of a bioassessment can be selected by the user as a theoretical error rate. Type 1 error is the likelihood of rejecting the null hypothesis when it is true, a false positive, and is inherent in the data. Two levels of confidence intervals were assessed because of the relationship between type 1 and 2 errors. Selecting the ellipse to use sets the theoretical type 1 error rate, and thus the potential type 2 error rate. While reducing the type 1 error may be beneficial to reduce the likelihood of a false positive, this increases the chance of type 2 errors, or false negative. Consequently a decision point is necessary to balance the willingness to commit both type 1 and 2 errors. Environmental managers concerned with the impacts on aquatic ecosystems and the risk to humans should be concerned with type 2 errors due to the risks associated with falsely accepting the alternative hypothesis. In this section, type 1 errors are discussed and compared at the 90% and 75% confidence intervals.

Analysis of the models at the 90% and 75% decision point revealed the multi-basin model to have the highest overall type 1 errors of all models. The errors across groups are not always the highest however (Table 2.12). Therefore I cannot conclude that a test site assessed as not in reference by the multi-basin model is in fact not in reference (false positive), however groups with higher errors are less likely to commit a false positive. A comparison of type 1 errors between the single basins bioassessments to the multi-basin revealed that the multi-basin assessments were comparable to the single basins and was either equal (i.e. Attawapiskat River basin), had lower type 1 errors than the Yukon River basin assessment, or relatively higher than the Fraser River assessment. These results are hypothesized to be related to the amount of variability among reference sites within each group. The number of reference sites within a group however did not attribute to this variability because groups with more reference sites did not necessarily have higher type 1 errors (data not shown, see Appendix F). This was also observed in the Attawapiskat model, but Group A (41 sites) did have higher error. Groups ordinated from the Attawapiskat model possessed tighter clusters within the confidence ellipses and therefore had the lowest type 1 errors at both the 90% and 75% confidence intervals. Tighter clusters could be due to less environmental variability and therefore less variation of biological communities among sites. Further divisions into groups at the classification stage would be beneficial (however cannot be controlled), but in many cases for this study, the prediction performance of a candidate model was sacrificed or the groups possessed heterogeneous covariance matrices. Other studies show that precision (standard deviation of O:E ratios in RIVPACS models) is decreased in models applied at larger spatial scales. For example, Yuan et al. (2008) found precision to be compromised due to a lack of reference sites to characterize a community group. This is plausible because a site with only 10 sites will not necessarily be

entirely representative of the range of variability of a community group. Sampling effort has also shown to increase the ability to detect disturbance (Cao and Hawkins 2005) and may be worth the extra effort in the field for more precise and therefore a more sensitive model.

2.4.5.2 *Type 2 errors*

The chance of committing a type 2 error has become central to hypothesis testing and statistical analyses because of the realization of the importance of erroneously accepting the null hypothesis (see Mapstone 1995, Field et al. 2004). For this study, type 2 errors are used to assess the sensitivity of my bioassessment models to deviations from the reference condition. The 3 levels of disturbance created based on known responses of benthic invertebrate communities to disturbance allowed for sensitivity assessment. Other studies compared models using test sites with an assumed level of impairment based on anthropogenic activities upstream or in the region (Reynoldson et al. 1997, Chessman 1999). These sensitivity measurements are not representative of community structure prior to disturbance or indicate the actual response of benthic communities to stressors (Cao and Hawkins 2005). The use of simpacted data allowed us to objectively assess and compare my models for sensitivity to deviations from reference.

My models revealed that type 2 errors decreased from mild to severe disturbance, illustrating the success of simpacted data to show deviation from reference conditions. Unfortunately stressors causing smaller deviations from reference (mild disturbance) will fail to be detected. To account for higher type 2 errors at low disturbance levels in comparison to severe disturbance, the decision point (confidence ellipse selected) could be changed to account for these smaller changes in the biological communities. Consequently, type 1 errors will respond oppositely, stressing the need for a compromise among the two types of error. These decisions rely on the cost and importance of committing either type of error (Mapstone 1995).

Comparison among the models shows that the errors varied across disturbances, models, and community groups within models. There are no clear patterns, however overall the multi-basin assessments had the higher type 2 errors at moderate and severe levels of disturbance. For the basin models, the highest errors were for groups possessing the smallest group sizes, whereas the highest errors for the multi-basin model were in the largest groups (data not shown, see Appendix G). These patterns may be attributed to not enough variation described for small group sizes and too much variation to describe, which is very likely with a group size of 490 reference sites. Therefore a group that is neither large nor small will likely describe enough variation to reduce errors but not possess too much variation that it cannot be explained.

Variability across groups and models may be attributed to firstly, greater variability of reference group clusters generating larger confidence intervals and therefore deviations from reference are more difficult to detect (especially for low disturbance). Second, the various changes to the 3 disturbance levels may not be affecting groups with a lower abundance and/or richness. Strachan and Reynoldson (2014) found that a data set from Australia Capital Territory with higher EPT richness is better at detecting deviations from reference than datasets with lower EPT taxa from the Great Lakes and Yukon. Groups in my models with the lowest abundance and/or richness are likely less susceptible to the various levels of disturbance incurred upon the data. Similarly, probability thresholds set for expected taxa for large scale RIVPACS models (Ode et al. 2008) resulted in underestimating impairment relative to regional models. Large-scale models based on lower richness reference sites will underestimate changes for richer areas, effectively reducing the model sensitivity to deviations from reference. Lastly, the variability in the errors may also be a function of the simpacted data selected. There were more simpacted sites assessed in models (and groups within models) that possessed more reference sites, and this may have led to a

biased assessment of sensitivity. Smaller groups had as low as 1 simpacted site (Group F in the Fraser model) to assess in a group, which may have led to a poor representation of the actual errors and therefore poorer sensitivity of the group. The variability in errors across groups suggests the importance of assessing test sites from the group from which it was predicted to. Strachan and Reynoldson (2014) also suggest using different confidence ellipses for each group within a model to account for the different sensitivities among groups.

Using the same subset of simpacted data, comparisons between each basin model and the multi-basin model revealed an overall lack of sensitivity of the multi-basin model, especially in comparison to the Yukon and Fraser models (see Table 2.13). This is most likely due to the largest group (*n*=490) in the multi-scale model that encompasses sites from both Fraser and Yukon sites. When comparing sites assessed in these 2 basin models, almost all simpacted sites were assessed in the larger group. The variability in this group is too great to detect divergence from reference. The selection of a model with more groups would reduce within group variability, but candidate models with more groups possessed lower prediction performance (see Appendix D). In this case, there is a trade-off between model prediction performance and sensitivity. A method is needed that would increase prediction performance without sacrificing sensitivity.

The lower prediction performance of the multi-basin model is also suspect to be due to the increased variability from multiple sampling years. The Attawapiskat model was built using data from a single year, whereas the other two basin models represented more years. The combination of all datasets increased the number of years accounted for and may have attributed to the increased variability in the data and therefore the decreased sensitivity of the model.

2.4.6 Considerations

There is ongoing research regarding the spatial applicability of predictive models that revolves around characterizing biotic assemblage variability and finding the appropriate predictors that explain benthic communities. The best method of explaining biological variability is still debated and whether grouping a continuum is the appropriate means of dealing with this variability. Bailey et al. (2014) summarised the results in a special issue of Freshwater Science that introduced many new and updated RCA bioassessment methods. An alternative method to grouping of biotic assemblages is the ANNA method (Linke et al. 2005) and LEDA (Chessman et al. 2008, Chessman 2014). ANNA removes the need for community groups as in RIVPACS and BEAST models by predicting biotic composition based on the nearest reference sites and comparing test sites to the reference sites based on Euclidian distance of environmental predictors. Sarrazin-Delay et al. (2014) uses a similar method to ANNA but compares test sites to its nearest neighbour reference sites using redundancy analysis modelling. The LEDA method relies on the abiotic differences among pairs of reference sites to ascertain the 'limiting environmental differences' and thereby removing the need for biotic groupings. An updated BEAST model has also been developed by Reynoldson et al. (2014) that uses multiple models in a tiered fashion to obtain better classification rates.

Likewise, environmental heterogeneity is also difficult to characterize into discrete groups. It is known that various environmental patterns, particularly landscape-scale patterns are skewed and covary with other factors that influence biotic assemblages (King et al. 2005). Problems such as these are hard to model with parametric tests and therefore non-parametric methods may provide the means to address non-linearities. Many studies use non-parametric test such as Random Forests (Cutler et al. 2007, Carlisle et al. 2009, Hawkins et al. 2010, João et al. 2014) and may

provide higher prediction rates than found in this study. Bayesian reference condition models have been found to be comparable or superior to standard RCA modelling methods (Webb et al. 2014). I therefore suggest exploring the possibilities of more sensitive assessments that may uncover underlying trends not recognized by linear models.

2.5 Conclusions

The value of applying RCA models beyond the geographic boundaries of their reference data is that it makes the models more widely applicable to environmental managers. It reduces the need for extensive sampling and reduces the number of bioassessment models necessary over large geographic regions which are costly and time-consuming to develop.

The results of my study showed that reference sites can be used outside of their spatial range to assess test sites insofar as that test site is within the range of environmental characteristics described by the model. The similarities of sites among basins show that the structure of communities and their variability is independent of geographic location itself but rather a function of similar environmental characteristics influencing them. These results demonstrate the capabilities of a model applied across a larger spatial extent.

The multi-basin model has comparable prediction performance and parsimony to the basin models but lacks some sensitivity (although not significantly) compared to the single basin models. Although community groups were not clustered by basin, the smaller clusters in the multi-basin model containing reference sites from the Attawapiskat may be under represented due to smaller sample size and therefore appear to represent poorer sensitivity. More reference sites from the Yukon and Fraser allow a greater characterization of the environmental gradients

within those basins; therefore I suggest that in the future improvements could be made with more equal representation of basin sites within a multi-basin model.

Results from Ode et al. (2008) show that precision and accuracy were sacrificed to develop a model with large spatial extent and suggests that these spatially-broad models are not ideal for use at a more regional scale, where precision and accuracy are important. Indeed, caution should be heeded with the use of a multi-basin model. Further developments of alternative RCA bioassessment methods as introduced in the Freshwater Science special issue (Volume 33, 2014) may provide greater detection of impaired sites for models applied across any spatial extent.

In a world full of human development and environmental changes, it is important to monitor and protect the structure and function of important aquatic biota. The application of a biomonitoring tool such as this will no doubt contribute substantially to our efforts to preserve biota but also to the many valuable aquatic ecosystem services that are provided to humanity.

References

Allan, D., D. Erickson, and J. Fay. 1997. The influence of catchment land use on stream integrity across multiple spatial scales. Freshwater Biology 37:149–161.

Altieri, M. A. 1999. The ecological role of biodiversity in agroecosystems. Agriculture, Ecosystems and Environment 74:19–31.

Aroviita, J., H. Mykrä, T. Muotka, and H. Hämäläinen. 2009. Influence of geographical extent on typology- and model-based assessments of taxonomic completeness of river macroinvertebrates. Freshwater Biology 54:1774–1787.

Bailey, R. C. 2005. Yukon River Basin. Pages 775–780 *in* A. C. Benke and C. E. Cushing (editors). Rivers of North America. Academic Press.

Bailey, R. C., M. G. Kennedy, M. Z. Dervish, and R. M. Taylor. 1998. Biological assessment of freshwater ecosystems using a reference condition approach: comparing predicted and actual benthic invertebrate communities in Yukon streams. Freshwater Biology 39:765–774.

Bailey, R. C., S. Linke, and A. G. Yates. 2014. Bioassessment of freshwater ecosystems using the Reference Condition Approach: comparing established and new methods with common data sets. Freshwater Science 33:1204–1211.

Bailey, R. C., R. H. Norris, and T. B. Reynoldson. 2004. Bioassessment of freshwater ecosystems using the Reference Condition Approach. Page 170. Kluwer Academic Publishers.

Barbour, M., J. Gerritsen, B. Snyder, and J. Stribling. 1999. Rapid bioassessment protocols for use in wadeable streams and rivers: periphyton, benthic macroinvertebrates, and fish, Second Edition. EPA 841-B-99-002. U.S. Environmental Protection Agency. Washington, D.C.

Barbour, M. T., J. Gerritsen, G. E. Griffith, R. Frydenborg, E. Mccarron, and M. L. Bastian. 1996. A framework for biological criteria for Florida streams using benthic macroinvertebrates. Journal of the North American Benthological Society 15:185–211.

Bennett, S. 2011. A revised predictive model for bioassessment of streams in northwest British Columbia using the reference condition approach: Skeena model. Page 41 Environmental Protection Division, Ministry of Environment, Skeena Region.

Cao, Y., and C. P. Hawkins. 2005. Simulating biological impairment to evaluate the accuracy of ecological indicators. Journal of Applied Ecology 42:954–965.

Cao, Y., D. P. Larsen, and R. S. J. Thorne. 2001. Rare species in multivariate analysis for bioassessment: some considerations. Journal of the North American Benthological Society 20:144–153.

Cao, Y., D. D. Williams, and N. E. Williams. 1998. How important are rare species in aquatic community ecology and bioassessment? Limnology and Oceanography 43:1403–1409.

Cardinale, B. J., K. L. Matulich, D. U. Hooper, J. E. Byrnes, E. Duffy, L. Gamfeldt, P. Balvanera, M. I. O'Connor, and A. Gonzalez. 2011. The functional role of producer diversity in ecosystems. American journal of botany 98:572–92.

Carlisle, D. M., J. Falcone, and M. R. Meador. 2009. Predicting the biological condition of streams: use of geospatial indicators of natural and anthropogenic characteristics of watersheds. Environmental monitoring and assessment 151:143–60.

Chessman, B. 1999. Predicting the macroinvertebrate faunas of rivers by multiple regression of biological and environmental differences. Freshwater Biology 41:747–757.

Chessman, B. C. 2014. Predicting reference assemblages for freshwater bioassessment with limiting environmental difference analysis. Freshwater Science 33:1261–1271.

Chessman, B., M. Muschal, and M. Royal. 2008. Comparing apples with apples: use of limiting environmental differences to match reference and stressor exposure sites for bioassessment of streams. River Research and Applications 24:103–117.

Chong, J. 2014. Resource Development in Canada: A Case Study on the Ring of Fire. Page 19.

Clarke, K., and R. Gorley. 2006. PRIMER v6: User Manual/Tutorial. Page 190.

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

Costanza, R., B. Fisher, K. Mulder, S. Liu, and T. Christopher. 2007. Biodiversity and ecosystem services: A multi-scale empirical study of the relationship between species richness and net primary production. Ecological Economics 61:478–491.

Crins, W., P. Gray, P. Uhlig, and M. Wester. 2009. The ecosystems of Ontario, Part 1: Ecozones and ecoregions. Page 77 Technical Report SIB TER IMA TR-01. Ontario Ministry of Natural Resources, Peterborough, Ontario.

Cutler, D., T. J. Edwards, K. Beard, A. Cutler, K. Hess, J. Gibson, and J. Lawler. 2007. Random forests for classification in ecology. Ecology 88:2783–2792.

Environment Canada. 2012a. Canadian Aquatic Biomonitoring Network Field Manual: Wadeable Streams. Page 57.

Environment Canada. 2012b. Canadian Aquatic Biomonitoring Network Laboratory Methods: Processing, Taxonomy, and Quality Control of Benthic Macroinvertebrates. Page 30.

Environment Canada. 2014. Canadian Aquatic Biomonitoring Network (CABIN). http://www.ec.gc.ca/rcba-cabin/Default.asp?lang=En&n=72AD8D96-1.

Feio, M. J., S. Almeida, S. Craveiro, and A. Calado. 2007. Diatoms and macroinvertebrates provide consistent and complementary information on environmental quality. Fundamental and Applied Limnology 169:247–258.

Feminella, J. W. 2000. Correspondence between stream macroinvertebrate assemblages and 4 ecoregions of the southeastern USA. Journal of the North American Benthological Society 19:442–461.

Field, J. G., K. R. Clarke, and R. M. Warwick. 1982. A practical strategy for analysing multispecies distribution patterns. Marine ecology progress series 8:37–52.

Field, S. A., A. J. Tyre, N. Jonzen, J. R. Rhodes, and H. P. Possingham. 2004. Minimizing the cost of environmental management decisions by optimizing statistical thresholds. Ecology Letters 7:669–675.

Frissell, C. A., W. J. Liss, C. E. Warren, and M. D. Hurley. 1986. A hierarchical framework for stream habitat classification: Viewing streams in a watershed context. Environmental Management 10:199–214.

Geobase, S. 2007. National Hydro Network, Canada. http://geobase.ca/geobase/en/data/nhn/index.html.

Gerritsen, J., M. T. Barbour, and K. King. 2000. Apples, oranges, and ecoregions: on determining pattern in aquatic assemblages. Journal of the North American Benthological Society 19:487–496.

Green, R. H. 1971. A Multivariate Statistical Approach to the Hutchinsonian Niche : Bivalve Molluscs of Central Canada. Ecology 52:544–556.

Green, R. H. 1979. Sampling design and statistical methods for environmental biologists. Page 272.

Hawkins, C. P., Y. Cao, and B. Roper. 2010. Method of predicting reference condition biota affects the performance and interpretation of ecological indices. Freshwater Biology 55:1066–1085.

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

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

Herlihy, A. T., S. G. Paulsen, J. Van Sickle, J. L. Stoddard, C. P. Hawkins, and L. L. Yuan. 2008. Striving for consistency in a national assessment: the challenges of applying a reference-condition approach at a continental scale. Journal of the North American Benthological Society 27:860–877.

Hilsenhoff, W. L. 1987. An improved biotic index of organic stream pollution. Great Lakes Entomologist 20:31–39.

Hooper, D. U., E. C. Adair, B. J. Cardinale, J. E. K. Byrnes, B. a Hungate, K. L. Matulich, A. Gonzalez, J. E. Duffy, L. Gamfeldt, and M. I. O'Connor. 2012. A global synthesis reveals biodiversity loss as a major driver of ecosystem change. Nature 486:105–8.

João, M., C. Viana-Ferreira, C. Costa, and M. J. Feio. 2014. Testing a multiple machine learning tool (HYDRA) for the bioassessment of fresh waters. Freshwater Science 33:1286–1296.

Karr, J. R., and E. W. Chu. 2000. Sustaining living rivers. Hydrobiologica 422/423:1–14.

King, R. S., M. E. Baker, and D. F. Whigham. 2005. Spatial considerations for linking watershed land cover to ecological indicators in streams. Ecological applications 15:137–153.

Kolkwitz, R., and M. Marsson. 1909. Ökologie der tierischen Saprobien. Beiträge zur Lehre von der biologischen Gewässerbeurteilung. Internationale Revue der gesamten Hydrobiologie und Hydrographie 2:126–152.

Linke, S., R. H. Norris, D. P. Faith, and D. Stockwell. 2005. ANNA: a new prediction method for bioassessment programs. Freshwater Biology 50:147–158.

Mapstone, B. D. 1995. Scalable decision rules for environmental impact studies: effect size, Type I, and Type II errors. Ecological Applications 5:401–410.

Marchant, R. 2002. Do rare species have any place in multivariate analysis for bioassessment? Journal of the North American Benthological Society 21:311–313.

Marchant, R., F. Wells, and P. Newall. 2000. Assessment of an ecoregion approach for classifying macroinvertebrate assemblages from streams in Victoria, Australia. Journal of the North American Benthological Society 19:497–500.

McGarigal, K., S. Cushman, and S. Stafford. 2000. Multivariate statistics for wildlife and ecology research. Page 283.

McKenna, K., and S. Smith. 2004. Physiography. Pages 8–10 *in* C. A. . Smith, J. C. Meikle, and C. F. Roots (editors). Ecoregions of the Yukon Territory: Biophysical properties of Yukon landscapes. Agriculture and Agri-Food Canada, PARC Technical Bulletin No. 04-01, Summerland, British Columbia.

McKenna, K., S. Smith, and B. Bennett. 2004. Vegetation. Pages 39–42 *in* C. A. . Smith, J. C. Meikle, and C. F. Roots (editors). Ecoregions of the Yukon Territory: Biophysical properties of Yukon landscapes. Agriculture and Agri-Food Canada, PARC Technical Bulletin No. 04-01, Summerland, British Columbia.

Mykrä, H., J. Aroviita, J. Kotanen, H. Hämäläinen, and T. Muotka. 2008. Predicting the stream macroinvertebrate fauna across regional scales: influence of geographical extent on model performance. Journal of the North American Benthological Society 27:705–716.

Mykrä, H., J. Heino, and T. Muotka. 2007. Scale-related patterns in the spatial and environmental components of stream macroinvertebrate assemblage variation. Global Ecology and Biogeography 16:149–159.

NASA, and METI. 2011. ASTER Advanced Spaceborne Thermal Emission and Reflection Radiometer. http://asterweb.jpl.nasa.gov/gdem.asp.

Norris, R. H., and C. P. Hawkins. 2000. Monitoring River Health. Hydrobiologia 435:5–17.

Ode, P. R., C. P. Hawkins, and R. D. Mazor. 2008. Comparability of biological assessments derived from predictive models and multimetric indices of increasing geographic scope. Journal of the North American Benthological Society 27:967–985.

Omernik, J. M. 1987. Map Supplement: Ecoregions of the Conterminous United States. Annals of the Association of American Geographers 77:118–125.

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

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

Reynoldson, T. B., J. Culp, R. Lowell, and J. S. Richardson. 2005. Fraser River Basin. Pages 697–701 *in* A. C. Benke and C. E. Cushing (editors). Rivers of North America. Academic Press.

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

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

Reynoldson, T. B., S. Strachan, and J. L. Bailey. 2014. A tiered method for discriminant function analysis models for the Reference Condition Approach: model performance and assessment. Freshwater Science 33:1238–1248.

Reynoldson, T. B., and J. F. Wright. 2000. The reference condition: problems and solutions. Pages 293–309 *in* J. F. Wright, D. W. Sutcliffe, and M. T. Furse (editors). Assessing the biological quality of fresh waters: RIVPACS and other techniques. Proceedings of an International Workshop held in Oxford, UK, on 16-18 September 1997.

Richards, C., L. B. Johnson, and G. E. Host. 1996. Landscape-scale influences on stream habitats and biota. Canadian Journal of Fisheries and Aquatic Sciences 53:295–311.

Sandin, L., and R. K. Johnson. 2004. Local, landscape and regional factors structuring benthic macroinvertebrate assemblages in Swedish streams. Landscape Ecology 19:501–515.

Sarrazin-Delay, C. L., K. M. Somers, and J. L. Bailey. 2014. Using Test Site Analysis and two nearest neighbor models, ANNA and RDA, to assess benthic communities with simulated impacts. Freshwater Science 33:1249–1260.

Van Sickle, J., D. P. Larsen, and C. P. Hawkins. 2007. Exclusion of rare taxa affects performance of the O/E index in bioassessments. Journal of the North American Benthological Society 26:319–331.

Srivastava, D. S., B. J. Cardinale, A. L. Downing, J. E. Duffy, C. Jouseau, M. Sankaran, and J. P. Wright. 2009. Diversity has stronger top-down than bottom-up effects on decomposition. Ecology 90:1073–1083.

Strachan, S. A., and T. B. Reynoldson. 2014. Performance of the standard CABIN method : comparison of BEAST models and error rates to detect simulated degradation from multiple data sets. Freshwater Science 33.

Tilman, D., P. B. Reich, and F. Isbell. 2012. Biodiversity impacts ecosystem productivity as much as resources, disturbance, or herbivory.

Tilman, D., D. Wedin, and J. Knops. 1996. Productivity and sustainability influenced by biodiversity in grassland ecosystems.

Vinson, M. R., and C. P. Hawkins. 1998. Biodiversity of stream insects: variation at local, basin, and regional scales. Annual review of entomology 43:271–293.

Webb, J. A., E. L. King, T. B. Reynoldson, and M. Padgham. 2014. Bayesian reference condition models achieve comparable or superior performance to existing standard techniques. Freshwater Science 33:1272–1285.

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

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

Yuan, L. L., C. P. Hawkins, and J. Van Sickle. 2008. Effects of regionalization decisions on an O/E index for the US national assessment. Journal of the North American Benthological Society 27:892–905.

Appendices

Appendix A: Environmental descriptors available for the Attawapiskat River basin dataset

Variable	Unit	Description
Altit	meters asl	Elevation/Altitude
Drain	km^2	Catchment Area
Latit	decimal degrees	Latitude NAD83
Longi	decimal degrees	Longitude NAD 83
StrmDen	m/km^2	Stream Density = stream length/catchment area
StrrmLng	m	Stream Length of all streams in catchment
StrnmOr	n/a	Stream Order
Perim	km	Catchment Perimeter Distance
Slope	m/m	Mean slope of catchment
Metam	percent	Metamorphic Rock
Plut	percent	Plutonic Rock
PlutVol	percent	Plutonic/Volcanic Rock
Sedm	percent	Sedimentary Rock
SedmVol	percent	Sedimentary/Volcanic Rock
Ult	percent	Ultramafic Rock
UltMet	percent	Ultramafic/Metamorphic Rock
UltVol	percent	Ultramafic/Volcanic Rock
Uncon	percent	Unconsolidated
Volca	percent	Volcanic Rock
Intru	percent	Intrusive Rock
LCBry	percent	Bryoids
LCBRDe	percent	Forest Broadleaf Dense
LCBROp	percent	Forest Broadleaf Open
LCBRSp	percent	Forest Broadleaf Sparse
LCCoDe	percent	Forest Coniferous Dense
LCCoOp	percent	Forest Coniferous Open

LCCoSp	percent	Forest Coniferous Sparse
LCExp	percent	Exposed Land
LCHrb	percent	Herb
LCMxOp	percent	Forest Mixedwood Open
LCMxSp	percent	Forest Mixedwood Sparse
LCRkRb	percent	Non-Vegetated Land: Rock/Rubble
LCShLw	percent	Shrubland: Shrub Low
LCShTa	percent	Shrubland: Shrub Tall
LCSnwI	percent	Non-Vegetated Land: Snow/Ice
LCWat	percent	Water
LCWtlHr	percent	Wetland-Herb
LCWtlSh	percent	Wetland-Shrub
LCWtlTr	percent	Wetland-Treed
LCMxDe	percent	Forest Mixedwood Dense
DegDy	days	Degree Days (number of days of the growing season above 5 deg)
PrcJa	mm	Precipitation January
PrcFb	mm	Precipitation February
PrcMr	mm	Precipitation March
PrcAp	mm	Precipitation April
PrcMy	mm	Precipitation May
PrcJn	mm	Precipitation June
PrcJl	mm	Precipitation July
PrcAu	mm	Precipitation August
PrcSp	mm	Precipitation September
PrcOc	mm	Precipitation October
PrcNv	mm	Precipitation November
PrcDc	mm	Precipitation December
PrcTtl	mm	Annual Precipitation
RnfJan	mm	Rainfall for January

RnfJne	mm	Rainfall for June
TmpJaMx	Degrees Celsius	Max Temp January (1971-2000)
TmpJaMi	Degrees Celsius	Min Temp January
TmpFbMx	Degrees Celsius	Max Temp February
TmpFbMi	Degrees Celsius	Min Temp February
TmpMrMx	Degrees Celsius	Max Temp March
TmpMrMi	Degrees Celsius	Min Temp March
TmpApMx	Degrees Celsius	Max Temp April
TmpApMi	Degrees Celsius	Min Temp April
TmpMyMx	Degrees Celsius	Max Temp May
TmpMyMi	Degrees Celsius	Min Temp May
TmpJnMx	Degrees Celsius	Max Temp June
TmpJnMi	Degrees Celsius	Min Temp June
TmpJlMx	Degrees Celsius	Max Temp July
TmpJlMi	Degrees Celsius	Min Temp July
TmpAuMx	Degrees Celsius	Max Temp August
TmpAuMi	Degrees Celsius	Min Temp August
TmpSeMx	Degrees Celsius	Max Temp September
TmpSeMi	Degrees Celsius	Min Temp September
TmpOcMx	Degrees Celsius	Max Temp October
TmpOcMi	Degrees Celsius	Min Temp October
TmpNvMx	Degrees Celsius	Max Temp November
TmpNvMi	Degrees Celsius	Min Temp November
TmpDeMx	Degrees Celsius	Max Temp December
TmpDEMi	Degrees Celsius	Min Temp December
TmpAv	Degrees Celsius	Mean Annual Temperature
DepthA	m	Average Depth
DepthM	m	Max Depth
VelocA	m/s	Average Velocity

VelocM	m/s	Max Velocity
WidthW	m	Wetted Width
WidthB	m	Bankfull Width
Domin1	Categorical	1st Dominant Substrate (0-8)
Domin2	Categorical	2nd Dominant Substrate (0-8)
Embed	Categorical	Embeddedness (1-5)
Surro	Categorical	Surrounding Material (0-8)
pH	pH	pH
Cond	mS/cm	Conductivity
Peri	Categorical	Periphyton Coverage (1-5)
Macro	Categorical	Macrophyte Coverage (1-5)
Canop	Categorical	Canopy Coverage (1-5)
Pools	Binary	Pools
Riff	Binary	Riffles
Strai	Binary	Straight Run
Conif	Binary	Coniferous Trees
Decid	Binary	Deciduous Trees
Grass	Binary	Grasses and Ferns
Shrub	Binary	Shrubs
Ag	mg/L	Ag (mg/L)
Al	mg/L	Al (mg/L)
As	mg/L	As (mg/L)
B	mg/L	B (mg/L)
Ba	mg/L	Ba (mg/L)
Be	mg/L	Be (mg/L)
Ca	mg/L	Ca (mg/L)
Cd	mg/L	Cd (mg/L)
Cl	mg/L	Chloride-Dissolved (mg/L)
Co	mg/L	Co (mg/L)

Cr	mg/L	Cr (mg/L)
Cu	mg/L	Cu (mg/L)
Fe	mg/L	Fe (mg/L)
Alkal	mg/L	Alkalinity (mg/L)
DOC	mg/L	DOC (mg/L)
Colour	n/a	Colour
Cond	µS/cm	Conductivity
DO	mg/L	DO (mg/L)
pH	pH	pH
TmpWa	Degrees Celsius	Temp. Water (Degrees Celsius)
K	mg/L	K (mg/L)
Mg	mg/L	Mg (mg/L)
Mn	mg/L	Mn (mg/L)
Mo	mg/L	Mo (mg/L)
Na	mg/L	Na (mg/L)
Ni	mg/L	Ni (mg/L)
NO2NO3	mg/L	Nitrogen-NO2+NO3 (mg/L)
TKN	mg/L	Nitrogen-TKN (mg/L)
Pb	mg/L	Pb (mg/L)
TP	mg/L	Phosphorus-TP (mg/L)
Sb	mg/L	Sb (mg/L)
Se	mg/L	Se (mg/L)
SO4	mg/L	SO4 (mg/L)
Sr	mg/L	Sr (mg/L)
Ti	mg/L	Ti (mg/L)
Tl	mg/L	Tl (mg/L)
U	mg/L	U (mg/L)
V	mg/L	V (mg/L)
Zn	mg/L	Zn (mg/L)

CoarGr	percent	Coarse grained (Glacio)Marine
FineGr	percent	Fine grained (Glacio)Marine
OrgDep	percent	Organic Deposits
TillBl	percent	Till Blanket

Appendix B: Data source information

Variable	Resolution/Scale	Source
Stream Order	1:50,000	Topographic map
Stream network	1:50,000	National Hydro Network (Geobase.ca)
Digital Elevation Model (DEM)	72m	ASTER GDEM V2, METI & NASA
Land cover	1:2,000,000	Geobase.ca
Surficial Geology	1:5,000,000	Geological Survey of Canada. Map 1880A,1995
Bedrock Geology	1:5,000,000	Geological Survey of Canada, Map D1860A, 1997
Long-term Climate (1971-2000)	7.5km	Natural Resources Canada

*Appendix C: Tolerance values for the Attawapiskat dataset (Hilsenhoff's Biotic Index), and the Fraser and Yukon datasets (Idaho Tolerance Index). * = genus level*

Family	Sensitivity	Hilsenhoff's Tolerance	Family	Sensitivity	Idaho Tolerance
Aeshnidae	Sensitive	3	**Aeshnidae**	Sensitive	3
Athericidae	Sensitive	2	**Ameletidae**	Sensitive	0
Baetiscidae	Sensitive	3	**Apataniidae**	Sensitive	1*
Blephariceridae	Sensitive	0	**Athericidae**	Sensitive	2
Brachycentridae	Sensitive	1	**Blephariceridae**	Sensitive	0
Capniidae	Sensitive	1	**Brachycentridae**	Sensitive	1
Chloroperlidae	Sensitive	1	**Capniidae**	Sensitive	1
Cordulegastridae	Sensitive	3	**Chloroperlidae**	Sensitive	1
Dixidae	Sensitive	1	**Cordulegastridae**	Sensitive	0*
Enchytraeidae	Sensitive	3	**Corduliidae**	Sensitive	2
Ephemerellidae	Sensitive	1	**Deuterophlebiidae**	Sensitive	0*
Glossosomatidae	Sensitive	0	**Dixidae**	Sensitive	1
Gomphidae	Sensitive	1	**Ephemerellidae**	Sensitive	1
Helicopsychidae	Sensitive	3	**Glossosomatidae**	Sensitive	0
Hydrodromidae	Sensitive	0	**Gomphidae**	Sensitive	1
Isonychiidae	Sensitive	3	**Helicopsychidae**	Sensitive	3
Lepidostomatidae	Sensitive	3	**Lepidostomatidae**	Sensitive	3
Leptohyphidae	Sensitive	3	**Leptophlebiidae**	Sensitive	2
Leptophlebiidae	Sensitive	2	**Leuctridae**	Sensitive	0
Leuctridae	Sensitive	0	**Nemouridae**	Sensitive	2
Metretopodidae	Sensitive	2	**Odontoceridae**	Sensitive	0*
Nemouridae	Sensitive	2	**Pelecorhynchidae**	Sensitive	3
Odontoceridae	Sensitive	0	**Peltoperlidae**	Sensitive	2
Perlidae	Sensitive	1	**Perlidae**	Sensitive	1
Perlodidae	Sensitive	2	**Perlodidae**	Sensitive	2
Philopotamidae	Sensitive	3	**Philopotamidae**	Sensitive	3

Psychomyiidae	Sensitive	2
Pteronarcyidae	Sensitive	0
Rhyacophilidae	Sensitive	0
Taeniopterygidae	Sensitive	2
Tipulidae	Sensitive	3
Uenoidae	Sensitive	0
Unionicolidae	Sensitive	0
Ancylidae	Semi-sensitive	6
Aturidae	Semi-sensitive	4
Baetidae	Semi-sensitive	4
Ceratopogonidae	Semi-sensitive	6
Chironomidae	Semi-sensitive	6
Corduliidae	Semi-sensitive	5
Crangonyctidae	Semi-sensitive	5
Dolichopodidae	Semi-sensitive	4
Dytiscidae	Semi-sensitive	5
Elmidae	Semi-sensitive	4
Empididae	Semi-sensitive	6
Ephemeridae	Semi-sensitive	4
Ephydridae	Semi-sensitive	6

Pteronarcyidae	Sensitive	0*
Rhyacophilidae	Sensitive	0
Taeniopterygidae	Sensitive	2
Tipulidae	Sensitive	3
Uenoidae	Sensitive	0
Ancylidae	Semi-sensitive	6
Baetidae	Semi-sensitive	4
Ceratopogonidae	Semi-sensitive	6
Chironomidae	Semi-sensitive	6
Dolichopodidae	Semi-sensitive	4
Dytiscidae	Semi-sensitive	5
Elmidae	Semi-sensitive	4
Empididae	Semi-sensitive	6
Ephemeridae	Semi-sensitive	4
Ephydridae	Semi-sensitive	6
Gammaridae	Semi-sensitive	4*
Gerridae	Semi-sensitive	5
Gyrinidae	Semi-sensitive	5
Heptageniidae	Semi-sensitive	4
Hydraenidae	Semi-sensitive	5

Gammaridae	Semi-sensitive	4
Gyrinidae	Semi-sensitive	5
Haliplidae	Semi-sensitive	5
Heptageniidae	Semi-sensitive	4
Hydraenidae	Semi-sensitive	5
Hydrophilidae	Semi-sensitive	5
Hydropsychidae	Semi-sensitive	4
Hydroptilidae	Semi-sensitive	4
Hydryphantidae	Semi-sensitive	4
Leptoceridae	Semi-sensitive	4
Limnephilidae	Semi-sensitive	4
Lymnaeidae	Semi-sensitive	6
Molannidae	Semi-sensitive	6
Muscidae	Semi-sensitive	6
Phryganeidae	Semi-sensitive	4
Polycentropodidae	Semi-sensitive	6
Sialidae	Semi-sensitive	4
Simuliidae	Semi-sensitive	6

Hydrophilidae	Semi-sensitive	5
Hydropsychidae	Semi-sensitive	4
Hydroptilidae	Semi-sensitive	4
Leptoceridae	Semi-sensitive	4
Limnephilidae	Semi-sensitive	4
Lymnaeidae	Semi-sensitive	6
Polycentropodidae	Semi-sensitive	6*
Sialidae	Semi-sensitive	4*
Simuliidae	Semi-sensitive	6
Tanyderidae	Semi-sensitive	5*
Asellidae	Tolerant	8*
Caenidae	Tolerant	7
Coenagrionidae	Tolerant	9
Corixidae	Tolerant	10
Culicidae	Tolerant	8
Enchytraeidae	Tolerant	10
Erpobdellidae	Tolerant	8
Glossiphoniidae	Tolerant	8

Tabanidae	Semi-sensitive	6	**Haliplidae**	Tolerant	7
Asellidae	Tolerant	8	**Hirudinidae**	Tolerant	7
Caenidae	Tolerant	7	**Hyalellidae**	Tolerant	8
Chaoboridae	Tolerant	7	**Hydrobiidae**	Tolerant	8*
Coenagrionidae	Tolerant	9	**Hygrobatidae**	Tolerant	8
Corixidae	Tolerant	9	**Lebertiidae**	Tolerant	8
Erpobdellidae	Tolerant	8	**Libellulidae**	Tolerant	9
Glossiphoniidae	Tolerant	8	**Lumbriculidae**	Tolerant	8
Hyalellidae	Tolerant	8	**Physidae**	Tolerant	8
Hydrobiidae	Tolerant	8	**Piscicolidae**	Tolerant	10*
Hygrobatidae	Tolerant	8	**Pisidiidae**	Tolerant	8
Lebertiidae	Tolerant	8	**Planorbidae**	Tolerant	7
Libellulidae	Tolerant	9	**Psychodidae**	Tolerant	10
Lumbriculidae	Tolerant	8	**Sperchontidae**	Tolerant	8
Naididae	Tolerant	10	**Staphylinidae**	Tolerant	8
Notonectidae	Tolerant	8	**Stratiomyidae**	Tolerant	8
Physidae	Tolerant	8	**Tabanidae**	Tolerant	8
Pisidiidae	Tolerant	8	**Tubificidae**	Tolerant	10
Planorbidae	Tolerant	7	**Valvatidae**	Tolerant	8
Siphlonuridae	Tolerant	7			
Sperchontidae	Tolerant	8			
Tubificidae	Tolerant	10			
Valvatidae	Tolerant	8			

Appendix D: Candidate models selected for ranking of 6 criteria to select final model.

Attawapiskat

Model	Cross Validation (%)	Error Evenness	Number of Predictors	Group Evenness	Number of Groups	Wilks' λ
4TH3	72		7		3	0.594
Not included due to assumptions:						
SQRT3	70		3		3	0.292
4TH4	65		3		4	0.226
LOG4	65		2		4	0.304
LOG3	69		2		3	0.398

Yukon

Model	Cross Validation (%)	Error Evenness	Number of Predictors	Group Evenness	Number of Groups	Wilks' λ
4TH5	47	11.94	7	3.58	5	0.507
LOG4	56	128.4	8	6	4	0.419
LOG5	51	70.51	9	7.48	5	0.348
RAW4	38	19.1	4	1.88	4	0.793
LOG6	46	9.5	8	7.64	6	0.314
SQRT5	42	7.7	7	2.01	5	0.623
SQRT6	39	11.01	7	4.62	6	0.487
4TH4	50	9.95	5	1.54	4	0.736

Fraser

Model	Cross Validation (%)	Error Evenness	Number of Predictors	Group Evenness	Number of Groups	Wilks' λ
SQRT4	66	44.8	10	1.61	4	0.279
SQRT5	64	34.1	10	1.64	5	0.193
SQRT6	58	19.9	10	3.61	6	0.114
SQRT7	52	11.3	10	8.75	7	0.103
4TH4	61	25.2	12	1.38	4	0.253
4TH5	66	82.1	11	1.52	5	0.113
4TH6	67	42.5	12	3.78	6	0.062

Multi-basin

Model	Cross Validation (%)	Error Evenness	Number of Predictors	Group Evenness	Number of Groups	Wilks' λ
SQRT6	54	6.87	11	1.88	6	0.214
SQRT7	57	5.62	11	2.46	7	0.007
4TH6	56	7.5	10	0.28	6	0.107
4TH5	65	17.64	11	0.29	5	0.109
4TH9	53	12.72	10	0.56	9	0.011
LOG6	57	12.73	10	0.46	6	0.17
RAW4	46	24.16	9	1.56	4	0.499

Appendix E: Candidate model ranks and sums across each of 6 selection criteria for each model with the exception of the Attawapiskat (no other candidate models).

Yukon

Model	Cross Validation (%)	Error Evenness	Number of Predictors	Group Evenness	Number of Groups	Wilks' λ	Rank
WEIGHT	***0.2857***	***0.2381***	***0.1905***	***0.1429***	***0.0476***	***0.0952***	
4TH5	4	4	3	4	2	5	3.81
LOG4	1	1	4	6	3	3	**2.57**
LOG5	2	2	5	2	2	2	2.57
RAW4	8	3	1	6	3	8	4.95
LOG6	5	7	4	1	1	1	4.14
SQRT5	6	8	3	5	2	6	5.57
SQRT6	7	5	3	3	1	4	4.62
4TH4	3	6	2	7	3	7	4.48

Fraser

Model	Cross Validation (%)	Error Evenness	Number of Predictors	Group Evenness	Number of Groups	Wilks' λ	Rank
WEIGHT	***0.2857***	***0.2381***	***0.1905***	***0.1429***	***0.0476***	***0.0952***	
SQRT4	2	2	1	5	4	7	2.81
SQRT5	3	4	1	4	3	5	3.19
SQRT6	5	6	1	3	2	4	3.95
SQRT7	6	7	1	1	1	2	3.95
4TH4	4	5	3	7	4	6	4.67
4TH5	2	1	2	6	3	3	2.48
4TH6	1	3	3	2	2	1	**2.05**

Multi-basin

Model	Cross Validation (%)	Error Evenness	Number of Predictors	Group Evenness	Number of Groups	Wilks' λ	Rank
WEIGHT	***0.2857***	***0.2381***	***0.1905***	***0.1429***	***0.0476***	***0.0952***	
SQRT6	4	6	3	2	3	6	4.14
SQRT7	2	7	3	1	2	1	3.14
4TH6	3	5	2	7	3	3	3.86
4TH5	1	2	3	6	4	4	**2.76**
4TH9	5	4	2	4	1	2	3.57
LOG6	2	3	2	5	3	5	3.00
RAW4	6	1	1	3	5	7	3.48

Appendix F: Type 1 errors for each model group (number of sites outside confidence ellipse/total reference sites)

	90%				75%			
	Attawapiskat	Yukon	Fraser	Mult-basin	Attawapiskat	Yukon	Fraser	Multi-basin
Group A	0/6	0/7	0/9	0/3	1/6	1/7	1/9	0/3
Group B	0/2	0/13	0/2	0/1	0/2	2/13	0/2	1/1
Group C	0/2	2/8	0/7	0/2	0/2	2/8	2/7	0/2
Group D	-	0/2	0/9	5/56	-	0/2	1/9	10/56
Group E	-	-	0/2	1/8	-	-	0/2	2/8
Group F	-	-	0/1	-	-	-	0/1	-
Total	*0/10*	*2/30*	*0/30*	*6/70*	*1/10*	*5/30*	*4/30*	*13/70*

Appendix G: Type 2 errors for each group within each model using a subset of simpacted sites at mild, moderate, and severe levels of disturbance. Total = total number of simpacted sites assessed, 90% / 75% = the number of simpacted sites that were in the reference ellipse at the 90% 75% confidence interval, % = the proportion of simpacted sites within the respective reference ellipse.

		Type 2 (%) - Mild				
		Total	**90%**	**%**	**75%**	**%**
Attawapiskat	Group A	6	6	100.0	4	66.7
	Group B	2	2	100.0	2	100.0
	Group C	2	2	100.0	2	100.0
	Total	10	10	100.0	8	80.0
Yukon	Group A	7	6	85.7	6	85.7
	Group B	13	13	100.0	7	53.8
	Group C	8	6	75.0	5	62.5
	Group D	2	2	100.0	2	100.0
	Total	30	27	90.0	20	66.7
Fraser	Group A	9	9	100.0	6	66.7
	Group B	2	2	100.0	2	100.0
	Group C	7	6	85.7	5	71.4
	Group D	9	7	77.8	5	55.6
	Group E	2	2	100.0	2	100.0
	Group F	1	1	100.0	1	100.0
	Total	30	27	90.0	21	70.0
Multi-Basin	Group A	3	3	100.0	3	100.0
	Group B	1	1	100.0	0	0.0
	Group C	2	2	100.0	2	100.0
	Group D	56	52	92.9	42	75.0
	Group E	8	7	87.5	6	75.0
	Total	70	65	92.9	53	75.7

			Type 2 (%) - Moderate			
		Total	**90%**	**%**	**75%**	**%**
Attawapiskat	Group A	6	4	66.7	2	33.3
	Group B	2	0	0.0	0	0.0
	Group C	2	1	50.0	1	50.0
	Total	10	5	50.0	3	30.0
Yukon	Group A	7	8	100.0	5	57.1
	Group B	13	3	61.5	1	38.5
	Group C	8	2	37.5	1	12.5
	Group D	2	20	100.0	11	50.0
	Total	30	3	66.7	0	36.7
Fraser	Group A	9	3	33.3	0	0.0
	Group B	2	0	50.0	0	50.0
	Group C	7	2	42.9	2	0.0
	Group D	9	0	0.0	0	0.0
	Group E	2	9	100.0	3	100.0
	Group F	1	3	0.0	3	0.0
	Total	30	1	30.0	0	10.0
Multi-Basin	Group A	3	42	100.0	27	100.0
	Group B	1	3	100.0	2	0.0
	Group C	2	51	100.0	34	100.0
	Group D	56	4	75.0	2	48.2
	Group E	8	0	37.5	0	25.0
	Total	70	1	72.9	1	48.6

			Type 2 (%) - Severe			
		Total	**90%**	**%**	**75%**	**%**
Attawapiskat	Group A	6	1	16.7	0	0.0
	Group B	2	0	0.0	0	0.0
	Group C	2	0	0.0	0	0.0
	Total	10	1	10.0	0	0.0
Yukon	Group A	7	0	42.9	0	0.0
	Group B	13	0	0.0	0	0.0
	Group C	8	1	0.0	1	0.0
	Group D	2	4	50.0	1	50.0
	Total	30	0	13.3	0	3.3
Fraser	Group A	9	0	0.0	0	0.0
	Group B	2	0	0.0	0	0.0
	Group C	7	0	0.0	0	0.0
	Group D	9	0	0.0	0	0.0
	Group E	2	0	0.0	0	0.0
	Group F	1	3	0.0	3	0.0
	Total	30	0	0.0	0	0.0
Multi-Basin	Group A	3	15	100.0	6	100.0
	Group B	1	1	0.0	1	0.0
	Group C	2	21	100.0	11	50.0
	Group D	56	1	26.8	0	10.7
	Group E	8	0	12.5	0	12.5
	Total	70	0	30.0	0	15.7