# Reference Model Supporting Documentation for CABIN Analytical Tools: Fraser Basin 2021

**MODEL NAME:** Fraser Basin 2021

**AUTHORS:** Trefor Reynoldson

**AFFILIATION:** GHOST Environmental Consulting

**DATE:** March 2021

**CONTACT(S):** Stephanie Strachan/Melissa Gledhill, CABIN Regional contacts Pacific-Athabasca-Arctic, Stephanie.Strachan@ec.gc.ca, Melissa.Gledhill@ec.gc.ca

**IMPORTANT NOTE: Reference Groups in the CABIN Analytical tools are labeled 1,2,3,4,5 and 6. These group numbers correspond to the reference groups in this document 1, 2, 3, 4A, 4B and 5, respectively.**

# 1. STUDY DESIGN AND SITE SELECTION

## 1.1 Model Purpose

The Fraser Basin 2021 model is an update to the Fraser Basin 2014 model (Strachan et al. 2014). Environment and Climate Change Canada (ECCC) started biomonitoring and model development in the Fraser Basin in the mid-1990s during the Fraser River Action Plan. Over time, more reference sites were sampled by both ECCC and BC Ministry of Environment and Climate Change Strategy (BC ENV), including repeated reference sites to track temporal variation. The model continues to provide the ability to assess streams exposed to a variety of disturbances in the basin such as forestry, mining, urban development and agriculture. This revision includes more than 190 additional reference samples and an additional 10 years of data compared to the 2014 model. This model represents established biota and habitat relationships in reference condition over a 25 year period (1994 to 2019). This model update is the first to evaluate power and sensitivity including simulated disturbance impacts (simpacts) on validation data to assess performance of the recommended model. The full analysis is documented in the detailed Technical Report (Reynoldson 2021) with the summary of the final approved model documented here.

## 1.2 Spatial and Temporal Scope

Spatial Scope: The Fraser Basin has 12 ecoregions with up to 6 stream orders (based on 1:50,000 scale). The Basin includes 13 sub-basins. Potential reference sites were distributed among all ecoregions and stream orders as much as possible based on access to sampling locations (Table 1). The geographical distribution of the reference sites in the Fraser River Basin is presented in Figure 1 with the identification of reference groups in the 2014 Fraser model and the additional reference sites considered for the 2021 update.

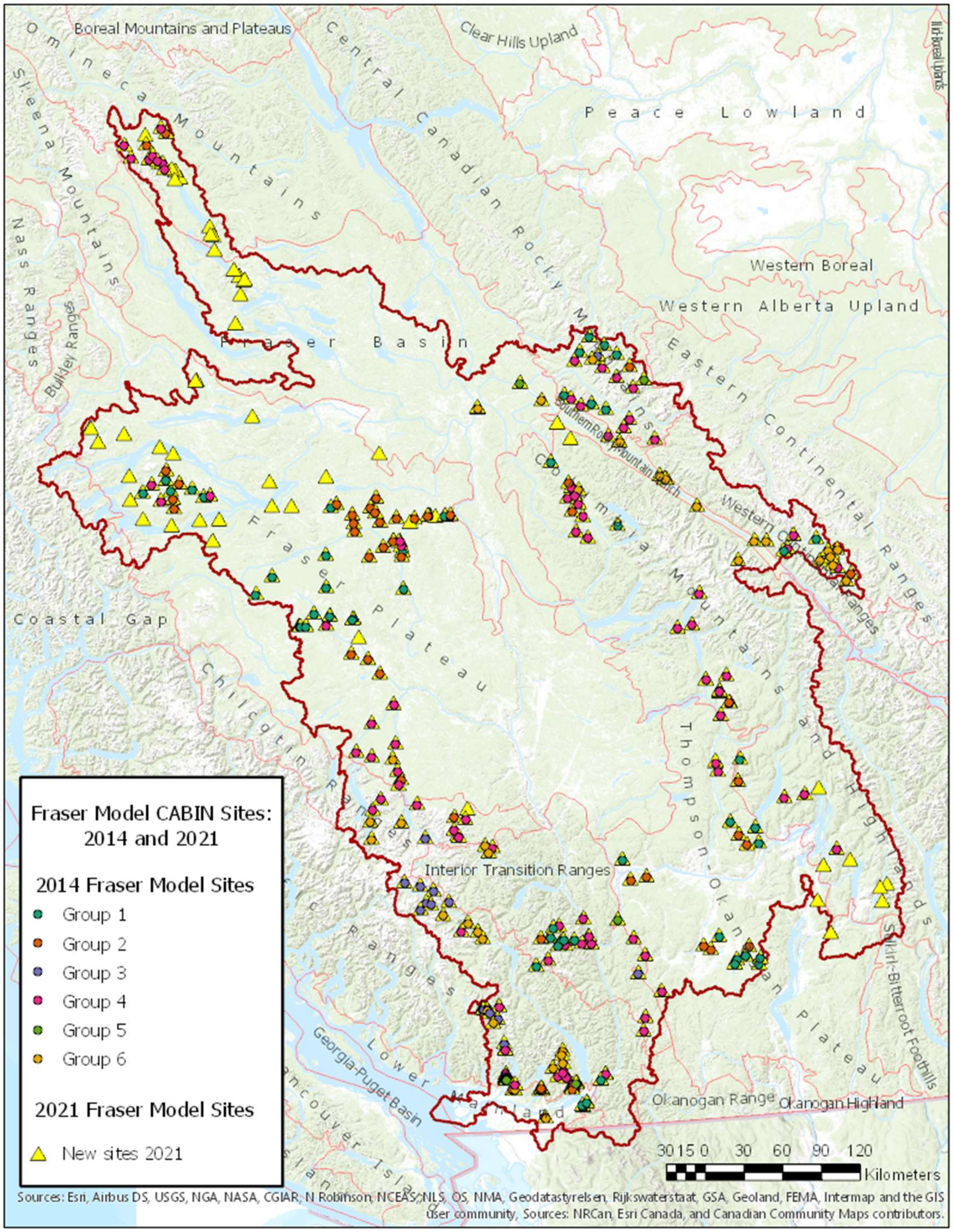


Figure 1. Distribution of reference sites across ecoregions in the Fraser River basin, showing site groupings of the 2014 Fraser model and new reference sites for the development of the 2021 Fraser model.

Table 1. Distribution of reference sites among ecoregions and stream orders (1994-2019).

| | Stream Order | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Ecoregion** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** |
| **Central Canadian Rocky Mountains** | 3 | 3 | 2 | 7 | 4 | 1 | | |
| **Chilcotin Ranges** | 1 | | 5 | 8 | 2 | | | |
| **Columbia Mountains and Highlands** | 3 | 6 | 21 | 15 | 6 | 1 | 8 | |
| **Fraser Basin** | | 2 | 6 | 2 | | | 1 | 13 |
| **Fraser Plateau** | 4 | 12 | 28 | 38 | 21 | 11 | 8 | 8 |
| **Interior Transition Ranges** | 3 | 8 | 3 | 5 | 5 | | | |
| **Lower Mainland** | 2 | 8 | 3 | | 5 | | | |
| **Omineca Mountains** | 1 | 4 | 5 | 7 | 1 | 4 | | |
| **Pacific Ranges** | 2 | 35 | 24 | 9 | 12 | | | |
| **Southern Rocky Mountain Trench** | | | | | | 2 | 14 | |
| **Thompson-Okanagan Plateau** | 2 | 12 | 14 | 4 | 56 | | 1 | |
| **Western Continental Ranges** | | 3 | 3 | 6 | | 7 | | |

Temporal Scope: The model was built using data collected in the late summer and early fall within the Fraser Basin between 1994 and 2019. Both ECCC and BC ENV resampled a subset of sites in multiple years to capture temporal variation across ecoregions (Table 2). Sampling effort has remained fairly consistent over time, after the initial large effort during the Fraser River action Plan in 1994-1996 (Figure 2).

Table 2. Temporal scope of reference samples across ecoregions and the number of years sampled to capture temporal variation.

| Ecoregion | Date Range | Total samples | Revisited sites | # years sampled for temporal variation |
|---|---|---|---|---|
| **Central Canadian Rocky Mountains** | 1995-2016 | 20 | N/A | N/A |
| **Chilcotin Ranges** | 1996-2016 | 16 | 1 | 3 |
| **Columbia Mountains and Highlands** | 1994-2018 | 60 | 8 | 14 |
| **Fraser Basin** | 1994-2018 | 24 | 4 | 9 |
| **Fraser Plateau** | 1994-2017 | 130 | 23 | 12 |
| **Interior Transition Ranges** | 1995-2018 | 24 | 4 | 4 |
| **Lower Mainland** | 1995-2005 | 18 | N/A | N/A |
| **Omineca Mountains** | 1994-2014 | 22 | N/A | N/A |
| **Pacific Ranges** | 1994-2019 | 82 | 5 | 17 |
| **Southern Rocky Mountain Trench** | 1995-2018 | 16 | 2 | 6 |
| **Thompson-Okanagan Plateau** | 1995-2019 | 89 | 11 | 18 |
| **Western Continental Ranges** | 1995-2018 | 19 | 1 | 7 |

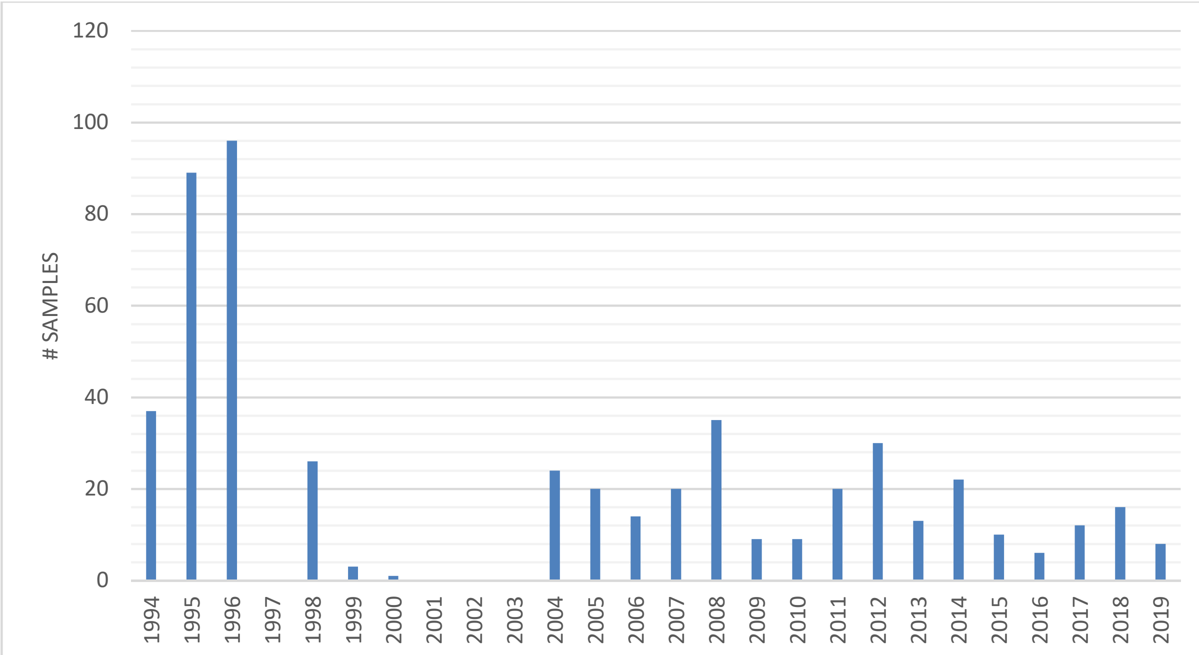


Figure 2. Fraser Basin reference samples collected in each year between 1994 and 2019.

## 1.3 Site Selection

Early reference sites (1994-2008) were selected based on expert local knowledge of the different watersheds (Rosenberg et al. 1999). Impacts such as logging, pulp mills, agriculture, ranching, recreation and urbanization as well as the degree of impact (i.e. low, moderate or high) were discussed with BC ENV staff for various sub-catchments. Final confirmation about reference site status was made during site reconnaissance based on best professional judgment. Site selection was focused on minimally disturbed sites from the wide variety of landscape types and stream sizes across the Basin, based on criteria documented elsewhere (Rosenberg et al. 1999, Davies 1994).

In the meantime, BC ENV developed a GIS-based site selection tool which defined stream level and watershed level criteria (Table 3) based on natural and stressor information to assist staff in locating potential reference sites (Norris 2012). The criteria varied among provincial regions based on resource development activity in the region. Final decision on site selection was made after site reconnaissance recognizing that the GIS based information is not always current. Reference sites selected by BC ENV after 2008 made use of the GIS-based site selection tool.

The final dataset consisted of 520 reference samples from 313 unique sites and 59 repeated sites. Refer to Section 1.2 for the distribution of reference sites among stream orders, ecoregions and years. As per the CABIN protocol, erosional/riffle habitats were targeted and a site was deemed to be 6x bankfull width. Where riffles were not present, runs were sampled.

Table 3. GIS-based reference site selection criteria to guide reconnaissance efforts for potential reference sites.

| Watershed Criteria | Reference Site Selection Tool |
|---|---|
| Urbanization | <0.1% |
| Agriculture | <5% |
| Forestry | <10% |
| Road density | <0.5 $km/km^2$ |
| Forest fire | <10% |
| Pine beetle infestation | <10% infested |
| **Stream Criteria** | |
| Downstream distance from waterbodies <5 $km^2$ | >2 km |
| Downstream distance from waterbodies >5 $km^2$ | >5 km |
| Downstream distance from flow structures | >500m |
| Upstream distance from flow structures | >50m |
| Upstream distance from road crossings | >50m upstream from any crossings |
| Downstream distance from road crossings | >500m |
| Upstream distance from current and past producing mines within 100m of stream | >500m |
| Downstream distance from Mining | No streams downstream from MINFILE |
| Riparian Areas - Natural Vegetation within 30m of stream | No human impact within 30m of stream |

# 2. Exploring Benthic Group Structure

Data from 520 reference or potential reference sites were downloaded from the CABIN website. Samples with incomplete habitat or benthic data were removed, leaving 485 reference and potential reference sites available for classification. The biological data used in the classification included 110 families.

## 2.1 Cluster Analysis

Reference groups were determined by classification and supported by ordination. While the original dataset for classification comprised 485 reference sites, a number of outlier sites were removed through preliminary investigations of the benthic data including initial classification and ordination. Two small groups of 12 and 13 sites were removed because the number of sites was insufficient for estimation of variance. A further 13 sites were outliers and also removed before final classification. Finally, four sites were removed because missing habitat data could not be acquired. Biological data from the remaining 443 reference sites were classified by Unweighted Pair Group Method with Arithmetic Mean (UPGMA) using agglomerative hierarchical fusion on a Bray-Curtis association matrix (Figure 3).

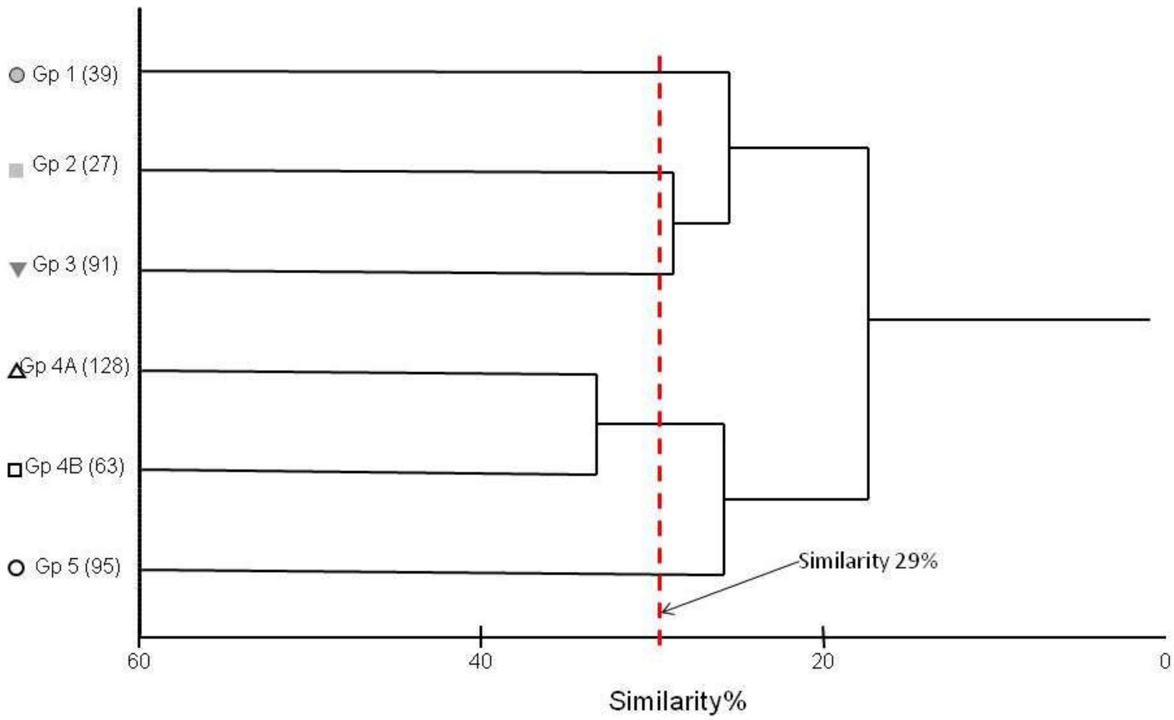


Figure 3. Classification dendrogram of 443 reference sites from the Fraser River Basin showing 6 groups based on Bray-Curtis similarity and group average linkage. Numbers in parentheses at left end of each branch indicate the number of sites on each branch.

The final grouping structure comprised six groups (Figure 3) and after various attempts using different classification methods and three to ten groups, this final solution provided the best predictive model. The geographical distribution is illustrated in Figure 4.

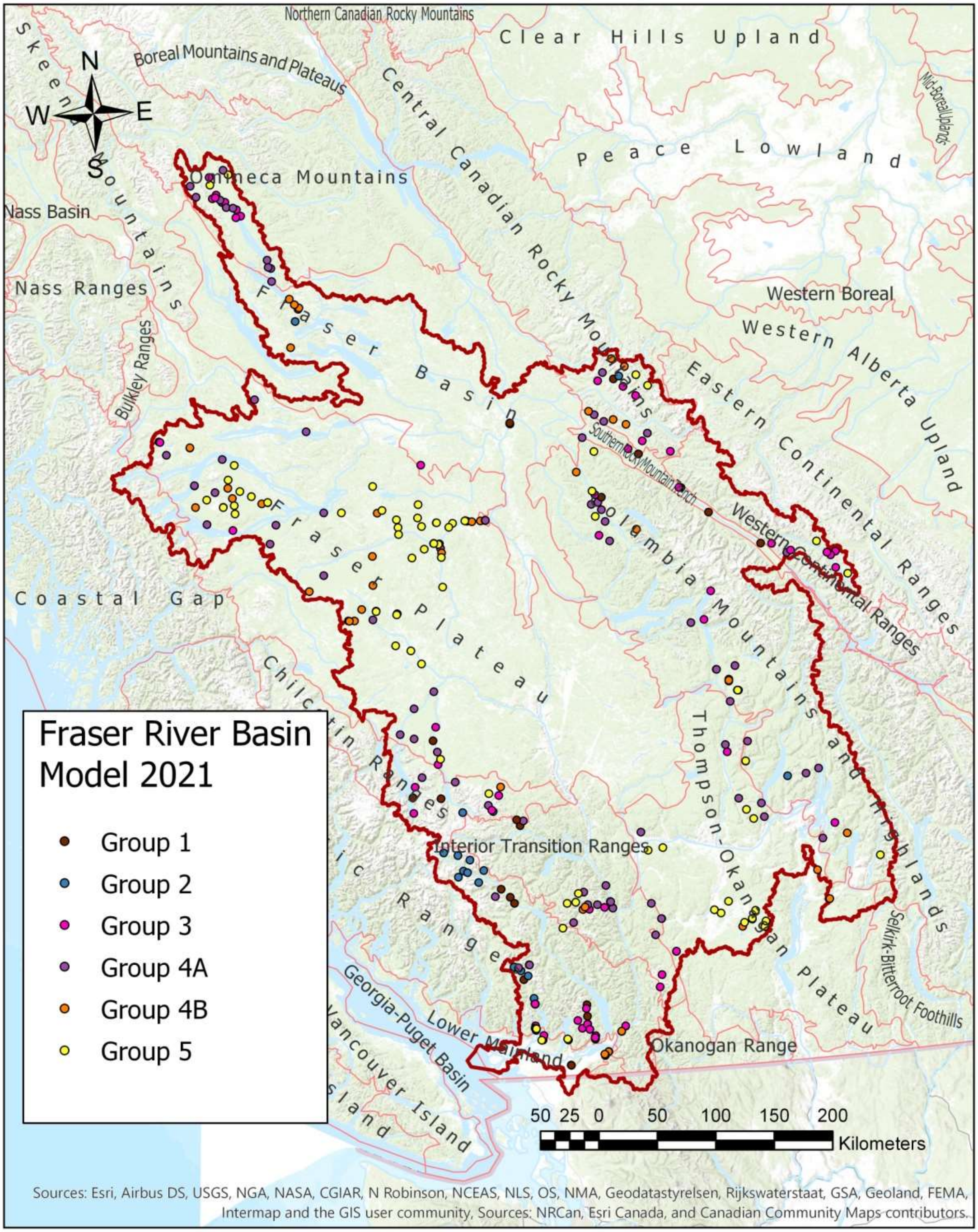


Figure 4. Geographic distribution of reference sites into six groups based on the invertebrate assemblage.

As part of the model development and evaluation procedure, 30% of reference sites (n=133) were removed for use as a validation dataset to assess model performance. The remaining sites (n=310) were used as the training dataset for model building; the following analysis and description of the reference groups is based upon the training sites only.

The biological communities of the training sites were plotted in ordination space (Figure 5) and show a gradient of increasing richness and total abundance from Group 1 to Group 5. The composition and abundance of families contributing to within group similarity are shown in Table 4.

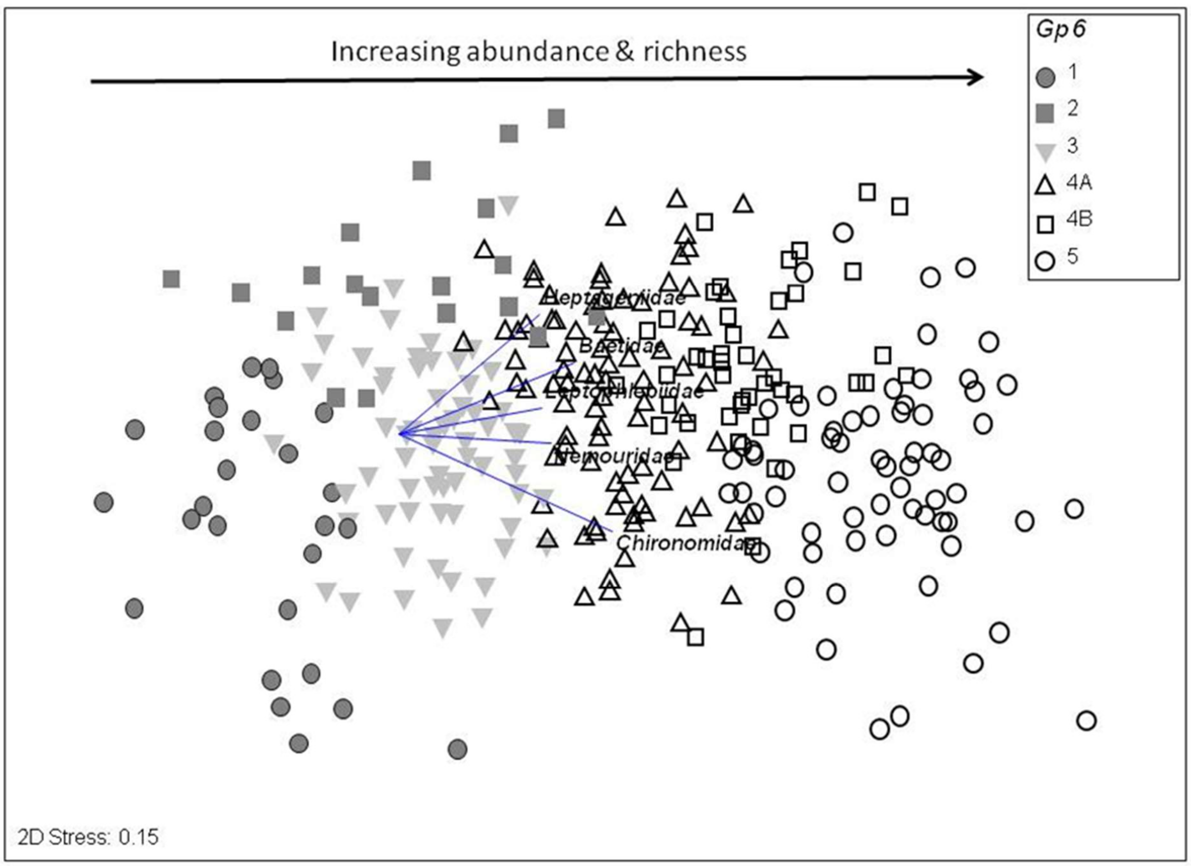


Figure 5. Ordination of 310 reference sites used for model building based on Bray-Curtis association of family level raw abundance data. The plot represents a gradient of abundance and richness with high values on the right and low values on the left.

Group 4A has the highest diversity and Groups 4A, 4B and particularly Group 5 have the highest total abundance. The gradient of increasing abundance and richness can be clearly seen in the ordination of the reference sites in Figure 5. The differences between the groups increase, as one would expect, as they move farther apart on the plot. The greatest difference (96.6%, Table 4) is between Group 1 and Group 5. The least difference (66.9%, Table 4) is between Groups 4A and 4B - and Figure 5 shows considerable overlap between these groups.

Table 4. Biological characteristics of the reference groups for the recommended six group model with 12 predictor variables.

| | Group 1 | Group 2 | Group 3 | Group 4A | Group 4B | Group 5 |
|---|---|---|---|---|---|---|
| **Biological Attributes (average)** | | | | | | |
| Richness | 15.7 | 11.4 | 18.4 | 19.1 | 18.0 | 18.0 |
| Abundance | 302.3 | 1535.5 | 955.5 | 3309.5 | 7867.8 | 14986.8 |
| Shannon's Diversity | 1.86 | 1.53 | 2.03 | 2.05 | 1.91 | 1.78 |
| Simpson's Diversity | 0.77 | 0.68 | 0.79 | 0.79 | 0.77 | 0.71 |
| Evenness | 0.68 | 0.64 | 0.70 | 0.70 | 0.67 | 0.62 |
| **SIMPER: Group Similarity** | | | | | | |
| Within Group Similarity | 37.03% | 46.47% | 42.48% | 41.56% | 49.82% | 42.99% |
| No. families contributing to 90% of similarity | 7 | 5 | 8 | 9 | 7 | 7 |
| **Mean abundance & percent contribution of common families *(NS indicates <10% contribution)*** | | | | | | |
| Chironomidae | 53 (19%) | *NS* | 116 (15%) | 480 (19%) | 924 (12%) | 5596 (60%) |
| Taeniopterygidae | 37 (15%) | 730 (60%) | *NS* | *NS* | *NS* | *NS* |
| Baetidae | 27 (12%) | *NS* | 167 (20%) | 355 (14%) | 2920 (51%) | 1856 (11%) |
| Ephemerellidae | 27 (11%) | *NS* | *NS* | 408 (13%) | *NS* | *NS* |
| **SIMPER: Dissimilarity between Groups** | | | | | | |
| Group 2 | 77.7% | | | | | |
| Group 3 | 72.2% | 69.2% | | | | |
| Group 4A | 87.0% | 74.2% | 69.6% | | | |
| Group 4B | 93.8% | 83.2% | 81.8% | 66.9% | | |
| Group 5 | 96.6% | 92.1% | 90.3% | 77.8% | 68.2% | |

The relative abundances of the major biological taxa in each reference group are illustrated in Figure 6. Figure 6 illustrates some major differences between the groups; for example, Group 2 is dominated by stonefly families (Order: Plecoptera) whereas Group 5 is dominated by midges (Order: Diptera). The families contributing to within group similarity, and their actual abundances, are presented in Figure 7.

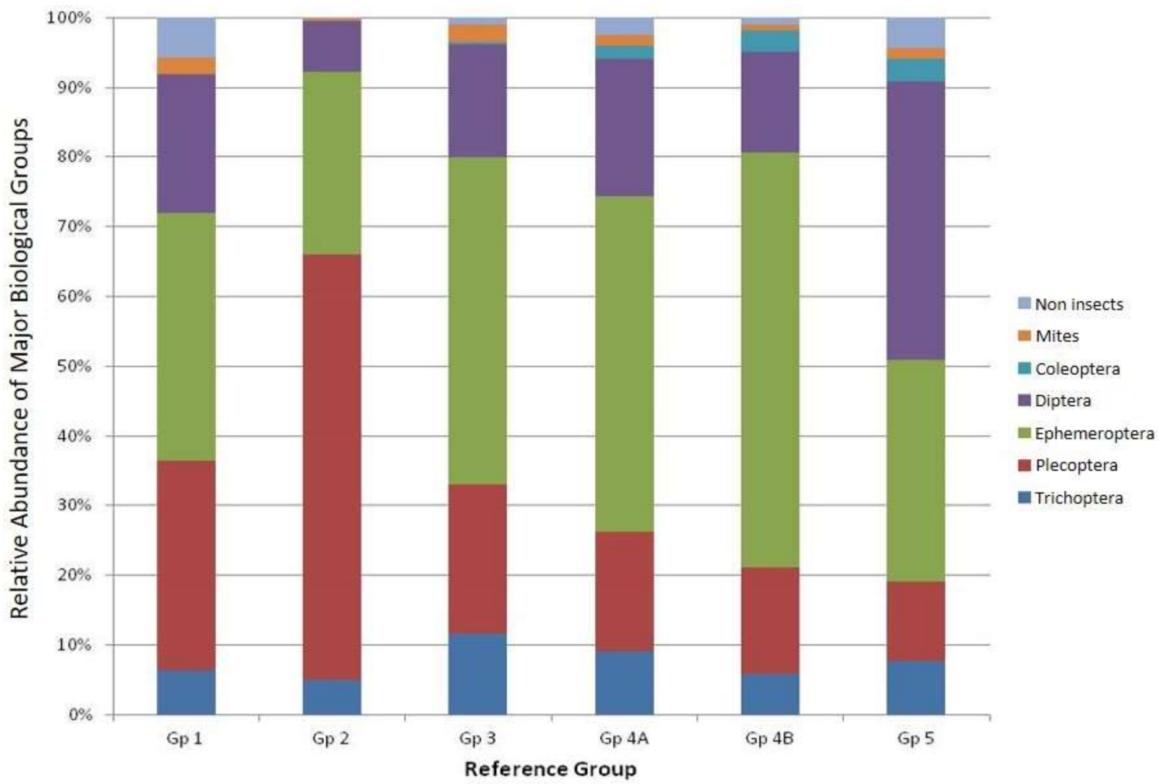


Figure 6. Relative abundance of major taxa in each reference group.

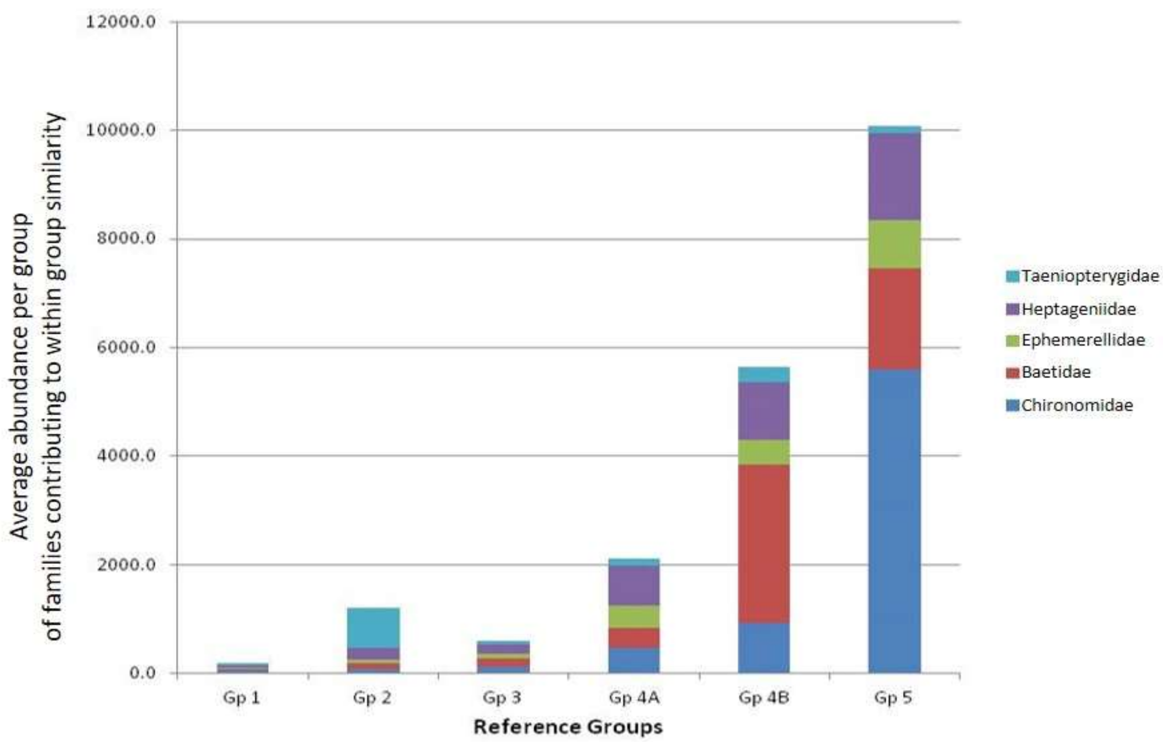


Figure 7. Average abundance of families contributing to within group similarity.

**Group 1 sites** have the lowest overall abundance and within group similarity, and are the least dominated by a single family contributing to within group similarity. Heptagenidae mayflies are the dominant family, but Chironomidae, Baetidae and Ephemerellidae also contribute. These sites tend to be at lower altitudes (Table 5) and are higher order streams (thus wider and deeper) that tend to have a higher velocity. They also have the largest drainage areas and a milder climate with higher precipitation. These 27 sites are primarily located in the Pacific Ranges (13) and Rocky Mountain Trench (9) ecoregions; and are mostly in the Headwaters (9) and Lower Fraser (13) basins.

**Group 2 sites** have the lowest biological diversity, particularly family richness, with almost half the number of families as the richest group (Group 4A, Table 4). This group is dominated by the Taeniopterygidae (winter stoneflies), a group of early emerging Plecoptera that contribute to 60% of the group similarity. Group 2 sites have the highest proportion (28%) of snow and ice cover in the drainage basin; they are also fairly wide streams. They have the highest substrate dominance category, with the average dominant cover being large cobble (Category 7: 12.8 - 25.6 cm diameter). These 19 sites are predominantly located in the Pacific Ranges ecoregion (11 sites), and the primary basins are the Harrison (6) and Lower Fraser (6).

**Group 3 sites** have the second lowest abundance but high richness and the highest diversity. Heptageneidae and Baetidae mayflies (Ephemeroptera) and Chironomidae dominate communities equally. These streams have the second highest precipitation but in most other respects are intermediate in terms of habitat attributes (Table 5). These 64 sites are distributed across 11 of the 12 ecoregions with the largest number (18) being in the Pacific Ranges. They are also widely distributed across 12 of the 13 sub-basins, with 12 sites in the Lower Fraser.

**Group 4A sites** have high total abundance and the highest richness and diversity, with the greatest number of families (9) contributing to within group similarity. Heptageneidae are again dominant, and the Chironomidae, Baetidae and Ephemerellidae also contribute substantially to within group similarity. These are higher altitude streams with the smallest drainage areas. With 90 sites, this is the largest group and sites are located across eight ecoregions, primarily the Fraser Plateau (22), the Columbia Mountains and Highlands (19), and the Thompson-Okanagan Plateau (18). The Lower Fraser (21), North Thompson (17) and Thompson (15) are the primary sub-basins of this group.

**Group 4B sites** have almost twice the total abundance of the Group 4A sites and the highest within group similarity. Communities are dominated by Baetidae mayflies and they contribute over 50% of the within group similarity. These are higher altitude steams with higher slope in the catchment, and tend to have the narrowest channel width. They also have the lowest annual precipitation and the coldest winter temperatures. Eight ecoregions are included in this group of 44 sites, with most sites (22) being located in the Fraser Plateau. Sites are widely distributed across the sub-basins, with West Road (9), Chilcotin (8) and Kemano Diversion (7) being the best represented.

**Group 5 sites** have the greatest total abundance - two orders of magnitude greater than the Group 1 sites - but they have the lowest evenness. They are dominated by Chironomidae, which on average make up a third of the community. These are the highest altitude streams, with the steepest slopes in the drainage basin and the lowest average velocity. They have a wide winter to summer temperature range. The 66 sites encompass 10 of the 12 ecoregions but five of those ecoregions only contain one or two sites; almost half (31) are located in the Fraser Plateau. Sites are located in 11 sub-basins, but the majority are in West Road (14) and the three Thompson basins (18).

Table 5. Group means for selected habitat attributes describing the six biological reference groups for the Fraser River Basin.

| | Group 1 | Group 2 | Group 3 | Group 4A | Group 4B | Group 5 |
|---|---|---|---|---|---|---|
| **Latitude (°N)** | 51.380 | 51.169 | 51.702 | 51.793 | 52.690 | 52.025 |
| **Longitude (°W)** | 121.932 | 123.148 | 122.640 | 122.536 | 123.273 | 122.869 |
| Altitude (masl) | 437 | 843 | 738 | 842 | 955 | 1029 |
| **Slope 30-50% (%)** | 28 | 29 | 27 | 22 | 11 | 10 |
| **Snow & Ice cover (%)** | 10 | 28 | 4 | 4 | 1 | 1 |
| **Water features in basin (%)** | 1.3 | 0.8 | 1.1 | 1.6 | 1.4 | 2.2 |
| Stream Order | 5.1 | 3.2 | 3.4 | 3.7 | 3.9 | 3.8 |
| Drainage Area ($km^2$) | 4646 | 963 | 846 | 635 | 1186 | 1064 |
| **Precipitation June (mm)** | 93.0 | 79.5 | 82.2 | 70.0 | 66.8 | 69.0 |
| **Precipitation August (mm)** | 76.2 | 64.9 | 65.5 | 57.8 | 53.7 | 54.7 |
| **Precipitation December (mm)** | 175.4 | 144.8 | 153.48 | 99.98 | 76.68 | 97.8 |
| Precipitation Annual (mm) | 1442 | 1181 | 1234 | 875 | 706 | 846 |
| Temp February Min (°C) | -9.6 | -10.0 | -10.1 | -11.9 | -14.0 | -12.6 |
| **Temp June Max (°C)** | 13.573 | 12.678 | 14.487 | 14.422 | 14.402 | 15.079 |
| **Temp August Max (°C)** | 16.9 | 16.4 | 17.8 | 17.7 | 17.6 | 18.4 |
| **Deciduous Streamside Veg (Proportion of sites where present)** | 0.9 | 0.5 | 0.7 | 0.7 | 0.6 | 0.3 |
| **Bankfull Width (m)** | 123 | 71 | 37 | 24 | 17 | 21 |
| Depth, Average (cm) | 41 | 25 | 27 | 23 | 20 | 20 |
| Velocity, Average (m/s) | 0.55 | 0.52 | 0.48 | 0.43 | 0.42 | 0.36 |
| Dominant Substrate (category (0-9)) | 6 | 7 | 6 | 6 | 6 | 6 |

*Note: Bold font indicates final model predictor variables.*

# 3. Relating Habitat to Benthic Group structure

## 3.1 Determining candidate predictor variables for Discriminant Function Analysis

The initial download of reference site data included 392 habitat variables. Many variables (98) had virtually no data and were excluded; a further 177 were missing data from more than 5% of the reference sites and were also excluded. Of the remaining 117 habitat variables, 31 were removed because they were no longer sampled, or not comparable through time due to methodology changes, or potentially modified by human activity and therefore could not be used to assign a test site to a reference group for assessment. Examples of this latter category include vegetation land cover variables that are modified by logging activity, substrate and other channel attributes that are modified by physical disturbance, or water quality attributes that can be modified by point source discharges. After removal of these variables, a total of 86 variables remained for potential inclusion in a predictive model.

## 3.2 Stepwise DFA results: Training Dataset

As described in section 2.1, the training dataset for final model building comprised 310 reference sites. During the analysis of these data, a total of 83 different models were examined using combinations of the 86 potential predictor variables and various strategies:

- creating forward and backward stepwise discriminant function analysis (DFA) models by removing or adding variables based on the F score for variable entry, or by adjusting the P values for variable inclusion;
- using combinations of variables identified by DFA with habitat categories, and
- backward and forward stepwise DFA using variables identified by matrix matching (PRIMER software: BEST method, BVSTEP option).

These approaches resulted in final selection of a model with six groups using 12 predictor variables. The minimum jackknifed classification rate for this model was 48% (Group 4B, Table 6) which is a 2.9x improvement over random. The details can be found in the full Technical Report (Reynoldson 2021).

Table 6. DFA classification and jackknifed cross-validation tables of reference site predictions to reference groups based on the training dataset (n=310).

| **Resubstitution** | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Group** | **1** | **2** | **3** | **4A** | **4B** | **5** | **%correct** |
| **1** | **20** | 2 | 4 | 1 | 0 | 0 | **74** |
| **2** | 4 | **10** | 3 | 1 | 0 | 1 | **53** |
| **3** | 5 | 2 | **36** | 16 | 2 | 3 | **56** |
| **4A** | 2 | 4 | 12 | **56** | 6 | 10 | **62** |
| **4B** | 0 | 0 | 5 | 5 | **25** | 9 | **57** |
| **5** | 0 | 0 | 11 | 6 | 14 | **35** | **53** |
| **Total** | 31 | 18 | 71 | 85 | 47 | 58 | **59** |
| | | | | | | | |
| **Cross-Validation (Jackknifed)** | | | | | | | |
| **Group** | **1** | **2** | **3** | **4A** | **4B** | **5** | **%correct** |
| **1** | **19** | 2 | 5 | 1 | 0 | 0 | **70** |
| **2** | 4 | **10** | 3 | 1 | 0 | 1 | **53** |
| **3** | 6 | 2 | **33** | 18 | 2 | 3 | **52** |
| **4A** | 3 | 4 | 13 | **52** | 8 | 10 | **58** |
| **4B** | 0 | 0 | 6 | 6 | **21** | 11 | **48** |
| **5** | 0 | 0 | 11 | 7 | 15 | **33** | **50** |
| **Total** | **32** | **18** | **71** | **85** | **46** | **58** | **54** |

The final set of predictors for the recommended six Group model are summarized below.

1) **Latitude** of the sampling location in decimal degrees.
2) **Longitude** of the sampling location in decimal degrees.
3) **Slope 30-50%** - the proportion of the drainage area of the site that has a gradient of 30-50%.
4) **Landcover, Snow & Ice** - percentage of upstream catchment with snow or glaciers.
5) **Landcover, Water** – percentage of upstream catchment with water, including lakes, reservoirs and rivers.
6) **Precipitation June** - 30yr monthly precipitation averages from 1971 – 2000 summarised for the upstream catchment from rasterised grids.
7) **Precipitation August** - 30yr monthly precipitation averages from 1971 – 2000 summarised for the upstream catchment from rasterised grids.
8) **Precipitation December** - 30yr monthly precipitation averages from 1971 – 2000 summarised for the upstream catchment from rasterised grids.
9) **Temperature June maximum** - 30yr average of maximum June temperatures from 1971 – 2000 summarised for the upstream catchment from rasterised grids.
10) **Temperature August maximum** - 30yr average of maximum August temperatures from 1971 – 2000 summarised for the upstream catchment from rasterised grids.
11) **Deciduous cover** - presence or absence of deciduous streamside vegetation at the sample site; see CABIN Field Manual (Environment Canada, 2012).
12) **Bankfull width** - at the sampling site, in meters.

Of these 12 variables, three (Bankfull width, Water cover, Snow & Ice cover) were included in the 2014 model, and other variables were similar - i.e. measures of summer temperatures, winter precipitation and descriptors of catchment slope. See Appendix 1 for GIS data sources for predictors 3-10.

# 4. Recommended Model

## 4.1 Evaluating Model Performance: Validation Dataset

The validation data were used to assess the model's power to detect disturbance by quantifying Type 1 and Type 2 error rates. Type 1 error is the percentage of unaltered validation sites that the model assessed as outside of reference condition (disturbance level "None" in Table 7 and Figure 8a, 8b). Type 2 error quantification used simulated disturbances to the biological communities of the validation data – simpacts – representing enrichment and sedimentation (Table 7, Figure 8a, 8b). While an array of simpacts were initially examined, as described in the full Technical Report (Reynoldson 2021), the following measures of power were based on three levels of disturbance described as mild, moderate and severe. The effects of these disturbances on abundance and richness and the associated error rates are shown in Figure 8a, 8b.

Table 7. Type 1 and 2 error rates for the proposed model and individual reference groups, for simulated enrichment and sedimentation disturbances.

| Enrichment | | Type 1 errors | Type 2 errors | | |
|---|---|---|---|---|---|
| | Level of Disturbance | **None** | **Mild** | **Moderate** | **Severe** |
| 6 Group model | No. Sites in error | 33 | 62 | 19 | 3 |
| | % error rate | 24.8% | 46.6% | 14.3% | 2.3% |
| By Group | | | | | |
| Group 1 (12 validation sites) | | 16.7% | 58.3% | 8.3% | 0.0% |
| Group 2 (8 validation sites) | | 0.0% | 75.0% | 50.0% | 25.0% |
| Group 3 (27 validation sites) | | 40.7% | 29.6% | 0.0% | 0.0% |
| Group 4A (38 validation sites) | | 13.2% | 47.4% | 0.0% | 0.0% |
| Group 4B (19 validation sites) | | 31.6% | 15.8% | 0.0% | 0.0% |
| Group 5 (29 validation sites) | | 31.0% | 69.0% | 48.3% | 0.0% |
| **Sedimentation** | | **Type 1 errors** | **Type 2 errors** | | |
| | Level of Disturbance | **None** | **Mild** | **Moderate** | **Severe** |
| 6 Group model | No. Sites in error | 28 | 40 | 27 | 2 |
| | % error rate | 21.1% | 30.1% | 20.3% | 1.5% |
| By Group | | | | | |
| Group 1 (12 validation sites) | | 8.3% | 33.3% | 8.3% | 0.0% |
| Group 2 (8 validation sites) | | 0.0% | 62.5% | 62.5% | 0.0% |
| Group 3 (27 validation sites) | | 44.4% | 11.1% | 0.0% | 0.0% |
| Group 4A (38 validation sites) | | 15.8% | 13.2% | 10.5% | 0.0% |
| Group 4B (19 validation sites) | | 21.1% | 10.5% | 0.0% | 0.0% |
| Group 5 (29 validation sites) | | 17.2% | 72.4% | 58.6% | 6.9% |

As a note of clarification to Table 7: due to the large number of validation sites, with four levels of disturbance to be assessed on each, it was deemed impractical to run a BEAST assessment for every impact for every site individually (i.e. 133 sites x 4 levels of disturbance x 2 disturbance types = 1064 ordination plots). Instead, four levels of a disturbance were assessed for each site in a single ordination plot (see example, Figure 9). Combining multiple sites with varying degrees of disturbance in a single ordination plot results in small shifts to the 90% probability ellipse used to designate reference condition, and is the reason that Type 1 errors related to enrichment and sedimentation impacts are slightly different despite the fact that no enrichment or sedimentation simpacts were made to the data used to assess Type 1 error (Table 7).

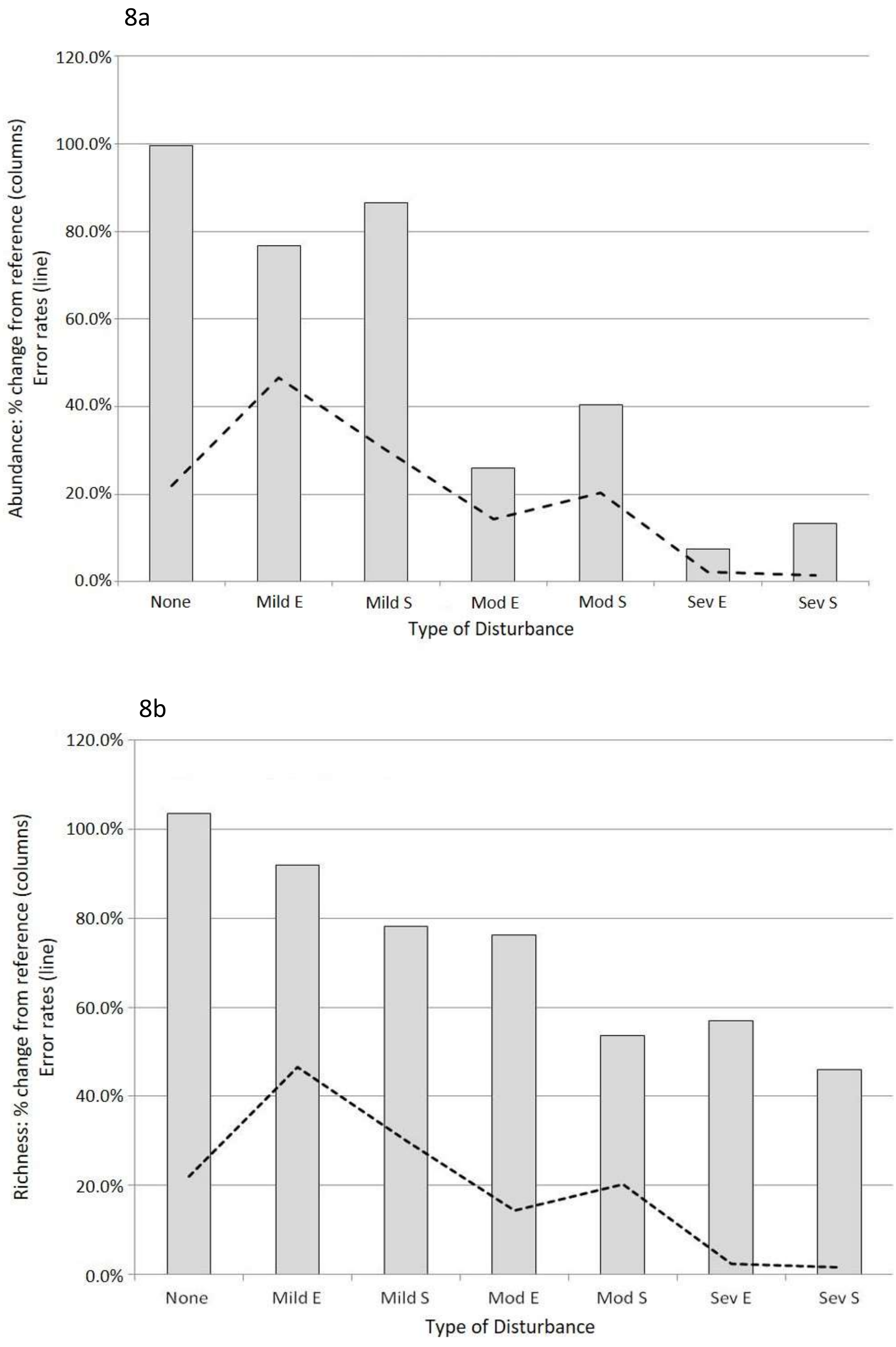


Figure 8. Percent change from reference in a) abundance (columns) and b) richness (columns) for 133 simpacted sites and associated error rates (dashed line) across different types and levels of disturbance. Type 1 error rate for the sedimentation simpact is shown on the "None" disturbance column; all others are Type 2 error rates. "E" represents an enrichment simpact and "S" represents a sedimentation simpact for each of the mild, moderate and severe levels of simulated disturbance.

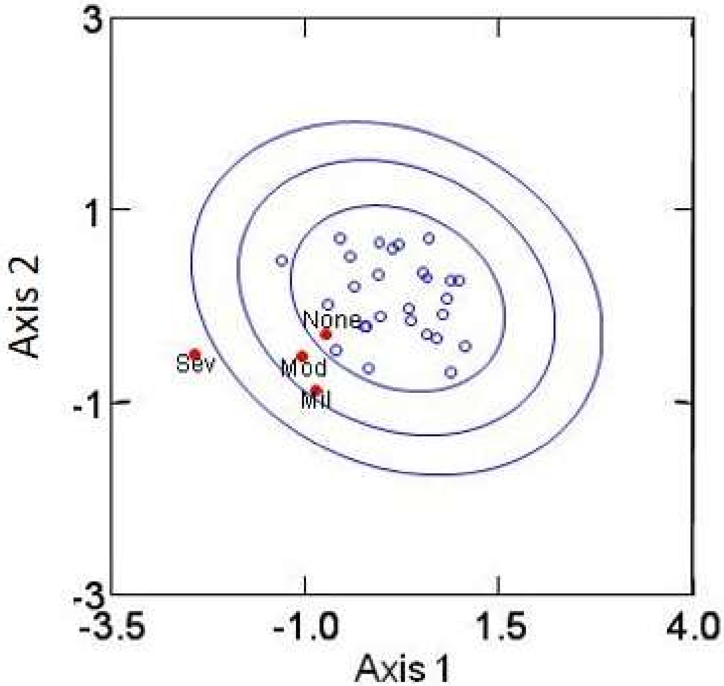


Figure 9. Example BEAST assessment plot with None, Mild, Moderate and Severe impacts plotted together for one validation site. The innermost ellipse is the 90% probability ellipse and sites within it are considered to be in Reference condition.

## 4.2 Comparison with the 2014 Fraser Basin Model and other RCA Models

The group distribution of the reference sites in the 2014 Fraser model was compared with their distribution in the proposed classification for the 2021 model. Of the 330 sites designated as reference in the 2014 model, 304 sites were classified with the 2021 model. The remaining sites were excluded as either outliers or as groups that were too small in the 2021 classification. For example, only 2 sites from Group 5 of the 2014 model were included in 2021 grouping; the other 11 sites formed a separate small group in the 2021 analysis and were excluded (Table 8).

Site classification between the 2014 and 2021 models shows high concurrence (Table 8). For example, Group 1 (2021) includes 93.5% of the Group 6 (2014) sites. Similarly, Group 2 (2021) is equivalent to Group 3 (2014). Group 4A (2021) is equivalent to Group 4 (2014), and Group 4B (2021) is Group 1 (2014).

Table 8. Concurrence in classification solutions between the 2014 and 2021 models.

| | 2014 Grp 1 | 2014 Grp 2 | 2014 Grp 3 | 2014 Grp 4 | 2014 Grp 5 | 2014 Grp 6 | Total | Concurrence |
|---|---|---|---|---|---|---|---|---|
| 2021 Group 1 | | | | 1 | 1 | 29 | 31 | 93.5% |
| 2021 Group 2 | | | 18 | | | 1 | 19 | 94.7% |
| 2021 Group 3 | | | 1 | 49 | 1 | 11 | 62 | 79.0% |
| 2021 Group 4A | 13 | 1 | | 59 | | | 73 | 80.8% |
| 2021 Group 4B | 38 | 3 | | | | | 41 | 92.6% |
| 2021 Group 5 | 18 | 57 | | 3 | | | 78 | 73.1% |
| Total | 69 | 61 | 19 | 112 | 2 | 41 | | |
| excluded | | 1 | | 4 | 11 | 10 | | |

Selection of the final recommended 2021 Fraser Basin model was based on a number of factors:

- **Number of groups** – as a general rule, a model with more groups is desirable.
- **Group size** – balanced group sizes are preferable because groups of similar size should provide more consistent assessments.
- **Number of predictor variables** - in general the number of predictor variables should not exceed the size of the smallest group; models with a smaller number of variables are preferred and seen as more robust.
- **Classification accuracy (training dataset)** – how well the model assigns sites to the correct group. For training sites, this is assessed by Discriminant Function Analysis using resubstitution and cross-validation.
- **Classification precision (training dataset)** - how variable the classification accuracy is among model groups, measured by the coefficient of variation. A lower coefficient of variation is preferred.
- **F – ratio** – a measure of the difference between the group means for the predictor variables. Higher F scores are preferred.
- **Wilks' lambda (λ)** – a measure of the null hypothesis that the groups have identical means on the discriminant equation; the larger the value of λ the better.
- **Differences among groups** – it is desirable to maximize the biological differences among the groups and minimize the differences within the groups.
- **Validation site classification accuracy** – the percentage of validation sites that are correctly predicted, by the model, to the groups they were assigned to by the classification.
- **Validation site classification precision** – the variation in accuracy among the individual model groups, measured by the coefficient of variation.
- **Power: Type 1 and 2 error rates** – using validation and simpacted validation sites, the percentage of validation sites assessed as different to reference (Type 1 error) and the percentage of simpacted reference sites assessed as in reference condition (Type 2 error).

Where these factors were also calculated in development of the 2014 Fraser model, the performance of the two models was compared (Table 9). Each of the criteria are ranked as 1 or 2 based on how the models performed relative to each other for that particular attribute. In almost every case, the new model was superior. Validation sites were not used in the 2014 model and therefore related classification, power and error rates were not reported.

Table 9. Factors used in selecting both the 2014 and 2021 Fraser Basin models, with relative ranking of results in parentheses.

| | 2021 model | 2014 model |
|---|---|---|
| ***Model Attributes: Training data*** | | |
| Number of Groups | 6 (1) | 6 (1) |
| Group sizes (number of sites/group) | 27, 19, 64, 90, 44, 66 (1) | 69, 62, 19, 117, 13, 51 (2) |
| Number of variables | 12 (1) | 12 (1) |
| Classification Accuracy (cross-validation/jackknifed) | 54% (1) | 49% (2) |
| Classification Precision (coefficient of variation) | 0.51 (1) | 0.69 (2) |

The 2021 Fraser model was compared to other RCA models for which simpacted error rates have been reported (Table 10). The Type 1 errors for the Fraser model are higher than those reported for the 2019 Peace River model (Reynoldson and Raggett 2019), but the Type 2 error rates are markedly lower. The model performed considerably better for both types of error than datasets from Australia, Yukon and the Great Lakes reported by Strachan and Reynoldson (2014).

Table 10. Comparison of error rates (%) for the proposed Fraser River model and other RCA models. "E" represents an enrichment simpact and "S" represents a sedimentation simpact.

| | Type 1 | Type 2 (mild) | Type 2 (moderate) | Type 2 (severe) |
|---|---|---|---|---|
| Fraser 2021 – 6 Groups (E) | 24.8 | 46.6 | 14.3 | 2.3 |
| Fraser 2021 – 6 Groups (S) | 21.1 | 30.1 | 20.3 | 1.5 |
| [1]Peace River 2019 – 4 Groups (E) | 8.7 | 60.9 | 52.2 | 0.0 |
| [1]Peace River 2019 – 4 Groups (S) | 8.7 | 78.3 | 17.4 | 17.4 |
| [2]Australian Capital Territory | 75.0 | 20.0 | 20.0 | 5.0 |
| [2]Yukon | 53.0 | 43.0 | 25.0 | 23.0 |
| [2]Great Lakes | 30.0 | 65.0 | 60.0 | 55.0 |

1. See Reynoldson and Raggett 2019
2. See Strachan and Reynoldson 2014

# 5. LITERATURE CITED

Davies, P. E. (Ed.) 1994. Monitoring River Health Initiative. River Bioassessment Manual. National River Processes and Management Program. (Freshwater Systems: Tasmania.)

Environment Canada. 2012. Canadian Aquatic Biomonitoring Network field manual – wadeable streams. 57pp. ISBN 978-1-100-20816-9.

Norris, S. 2012. British Columbia's Provincial Stream Biomonitoring Program Technical Documentation: GIS Tools for Reference Site Selection and Upstream Watershed Analysis. Draft report prepared by Hillcrest Geographics for Water Protection and Sustainability Branch, Ministry of Environment, Victoria, BC.

Reynoldson, T.B. 2021. A Revision of the 2014 CABIN RCA Model for the Fraser River Basin using sites collected over a 25 year period (1994-2019). GHOST Environmental Consulting. 78pp + appendices.

Reynoldson, T.B. and Raggett, J. 2019. Development of a CABIN Model for the Peace River. Report submitted to British Columbia Ministry of Environment and Climate Change Strategy by GHOST Environmental Consulting. 65pp

Rosenberg, D.M., T.B. Reynoldson and V.H. Resh. 1999. Establishing reference conditions for benthic invertebrate monitoring in the Fraser River Catchment, British Columbia, Canada. Fraser River Action Plan, Environment Canada, Vancouver BC. DOE-FRAP 1998-32.

Strachan, S.A., M. Edwards, T.B. Reynoldson, and J.L. Bailey. 2014. *Reference Model Supporting Documentation for CABIN Analytical Tools: Fraser Basin 2014.* Retrieved on Sept 16, 2020 from https://cabin-rcba.ec.gc.ca/CABIN/Beast/ModelDoc.aspx

Strachan, S.A. and Reynoldson, T.B. 2014. Performance of the standard CABIN method: Comparison of BEAST models and error rates to detect simulated degradation from multiple datasets. Freshw. Sci. 33:1225-1237.

# APPENDIX 1: DATA COLLECTION, ANALYSIS AND QUALITY ASSURANCE

## A. Field Collection

| *CABIN Study Name* | **BC MOE-Cariboo Region** | **BC MOE-FSP Skeena Region** | **BC MOE-Okanagan Kicknet** | **BC MOE-Omineca/ Peace Region** |
|---|---|---|---|---|
| *Agencies involved* | BC Ministry of Environment and Climate Change Strategy | BC Ministry of Environment and Climate Change Strategy | BC Ministry of Environment and Climate Change Strategy | BC Ministry of Environment and Climate Change Strategy |
| *Date range* | 2008-2016 | 2004-2014 | 2008-2016 | 2012-2014 |
| *Sampling season* | mid Sept to mid Oct | mid Aug to mid Sept | end of Aug to end of Sept | mid Aug to end of Sept |
| *# reference samples* | 3 | 40 | 8 | 16 |
| *Certified samplers (Y or N)* | Y | Y | Y | Y |
| *Certified team leader (Y or N)* | Y | Y | Y | Y |
| *400 um kicknet (Y or N)* | Y | Y | Y | Y |
| *Preservative used* | ethanol | ethanol | ethanol | ethanol |

| *CABIN Study Name* | **BC MOE-Thompson Region** | **EC-Fed/Prov WQ Monitoring Stations** | **EC-Fraser Large River Pilot 2018** | **EC-Fraser River** |
|---|---|---|---|---|
| *Agencies involved* | BC Ministry of Environment and Climate Change Strategy | Environment and Climate Change Canada | Environment and Climate Change Canada | Environment and Climate Change Canada |
| *Date range* | 2007-2019 | 2006-2018 | 2018 | 1994-2019 |
| *Sampling season* | mid Sept to end of Oct | end of Sept to early Oct | Early Oct | Early Sept to end of Oct |
| *# reference samples* | 45 | 6 | 1 | 401 |
| *Certified samplers (Y or N)* | Y | Y | Y | Y |
| *Certified team leader (Y or N)* | Y | Y | Y | Y |
| *400 um kicknet (Y or N)* | Y | Y | Y | Y |
| *Preservative used* | ethanol | formalin | formalin | formalin |

## B. Macroinvertebrate Identification

| CABIN Study Name | All BC MOE studies | All EC studies |
|---|---|---|
| Taxonomist | Danusia Dolecki 2006<br>Cordillera Consulting 2006-2019 | Craig Logan 1994-2005<br>EcoAnalysts 2007,-2008, 2010- 2011<br>Cordillera Consulting 2006, 2009, 2012-2019 |
| Marchant Box used (Y or N) | N, Caton Tray; Y, Marchant Box in later years | Y |
| Subsample count | 300 | 300 (except 1994-2001: 100 or 200 org count) |
| 10% of reference samples sent to National Lab for QA | Y | Y |
| Reference Collection maintained | Y | Y |

The data collection goes back to the early years of CABIN before the laboratory processing protocol and guidance for taxonomists was well established. As a result several taxa were identified and entered into CABIN that are not included in the development of a model. The excluded taxa are listed in Table A1-1 with the rationale for excluding them based on the CABIN protocol.

Table A1-1. Taxa exported from CABIN studies that were excluded from model analysis

| Taxon | Rationale |
|---|---|
| *Bosminidae* | cladoceran (pelagic) |
| *Candonidae* | ostracod (too small for 400 kicknet reliably) |
| *Chydoridae* | ostracod (too small for 400 kicknet reliably) |
| *Cyclocyprididae* | ostracod (too small for 400 kicknet reliably) |
| *Cyclopidae* | copepoda (pelagic) |
| *Cyprididae* | ostracod (too small for 400 kicknet reliably) |
| *Cypridopsidae* | ostracod (too small for 400 kicknet reliably) |
| *Cytherideidae* | ostracod (too small for 400 kicknet reliably) |
| *Daphniidae* | cladoceran (pelagic) |
| *Dugesiidae* | in protocol to not count (therefore inconsistently analysed) |
| *Hydridae* | Colonial |
| *Limnocytheridae* | ostracod (too small for 400 kicknet reliably) |
| *Lumbricidae* | terrestrial |
| *Macrobiotidae* | Tardigrada (pelagic) |
| *Macrothricidae* | cladoceran (pelagic) |
| *Planariidae* | in protocol to not count (therefore inconsistently analysed) |
| *Spongillidae* | Colonial |
| *Tetrastemmatidae* | Meiofauna (too small for 400 um kicknet reliably) |

The CABIN database includes a linkage to the Integrated Taxonomic Information System (ITIS) to ensure consistency in nomenclature. It is well known among taxonomic experts that updates to taxonomy are delayed therefore some important updates must be acknowledged that are not yet current in ITIS. Tubificidae is now recognized as a subfamily of Naididae. Due to the delay in ITIS updates, it is flagged as "unverified" in the CABIN database.

## C. GIS Analyses

GIS data were generated by Adam Yates (University of Western Ontario) for ECCC data while BC ENV (Chris Steeves, pers. comm.) generated GIS data using a slightly different procedure.

ECCC GIS data
Watersheds were delineated using ArcGIS 10 ArcHydro 2.0 (ESRI 2010). Delineations were based on 20 m resolution digital elevation models (DEM) and a 1:50,000 scale hydrological network. The DEM was subjected to pre-processing which "burned in" the stream network into the DEM and filled sinks to improve flow modeling. The corrected DEM was used to calculate flow direction and flow accumulation to carry out the terrain procession steps to model catchment areas (ArcHydro 2010). The delineated catchments were described using the GIS layers in the table below collected from publicly available sources (Table A1-2).

BC ENV GIS data
Delineations were based on 1:20,000 25m ASTER DEM using ArcGIS Spatial Analyst script to hydrologically condition the DEM. In some cases (some flat areas where there is little topographic variation, and some double line rivers are braided) this automated processing may not be successful and the bottom of the watershed must be defined manually, using the BC FWA fundamental watersheds as a guide within provincial boundaries.

Table A1-2: GIS data sources, methods and resolution used to generate watershed data for a variety of landscape level descriptors.

| ***Descriptor*** | ***Scale/ Resolution*** | ***Source and method*** |
|---|---|---|
| *Basin Morphometry* | *20m (ECCC)<br>25m (BC ENV)* | *Area and perimeter were calculated from delineated catchments as described above.* |
| *Bedrock* | *1:100,000* | *BC Ministry of Energy and Mines – BC Digital Geology Maps 2005 - http://www.empr.gov.bc.ca/Mining/Geoscience/PublicationsCatalogue/DigitalGeologyMaps/Pages/default.aspx Intersected with catchment boundaries using intersect function in ArcGIS (ESRI 2010)* |
| *Climate* | *7.5 km (ECCC)<br><br>1km (BC ENV)* | *Natural Resources Canada (contact: Dan McKenney – dan.mckenney@nrcan-rncan.gc.ca) NR Canada raster climate data is intersected with the watershed vector layer to produce an output table summarizing the mean condition for each climate variable for each watershed. Summarized using rasterized grids describing temperature normals from 1971-2000 giving long term monthly and annual averages of temperature and precipitation. Grids were used to generate average, minimum and maximum values for each catchment using Geospatial Modelling Environment v. 0.6.0.0 (Beyer 2012). Where catchments were completely contained within one grid cell, catchments were assigned the value of that cell.* |
| *Hydrology* | *1:50,000* | *www.geobase.ca – National Hydro Network Intersected with catchment boundaries using intersect function in ArcGIS (ESRI 2010)* |
| *Land Use (circa 2000)* | *1:2,000,000* | *www.geobase.ca – Land Cover Raster land cover data (circa 2000) from the North America Land Change Monitoring System project is intersected with the watershed vector layer to produce an output table summarizing the total area of land cover class, and by joining watershed area values, percent area can be calculated. Some polygons have no defined cover type due to being obscured by cloud or shadow cover.* |
| *Topography* | *20m (ECCC)<br><br>30m (BC ENV)* | *www.geobase.ca – Digital Elevation Data A zonal statistics tool is applied to the Aster DEM data to generate maximum, minimum, and mean elevation output values for each watershed. Percent slope was generated from the DEM and classified into one of four groups based on the slope value for each grid cell (i.e. <30%, 30-50%, 50-60%, >60%). Areas of each class within each catchment were then calculated.* |

### D. Laboratory Analyses

Laboratory analyses for water quality samples are stored in CABIN but are not used as predictors in the development of the model. The laboratories and methods used varied for each CABIN study.

### E. Statistical Analyses

Several software packages were used in the development of the Fraser model:
1. Excel - data manipulation and storage
2. PATN V.3.12 - classification and ordination of test sites for assessment
3. PRIMER 6 - classification, MDS ordination, ANOSIM, SIMPER, BEST
4. SYSTAT 11 - discriminant analysis and plotting BEAST assessments with probability ellipses