Reference Model Supporting Documentation for CABIN Analytical Tools

MODEL NAME: Vancouver Island Updated Model 2021

AUTHORS: Keith Somers[1], Catherine Proulx[1], Bruce Kilgour[1], Jolene Raggett[2]

AFFILIATION: [1]Kilgour & Associates Ltd, [2]BC Ministry of Environment and Climate Change Strategy

DATE: July 2021

CONTACT(S):   Jolene Raggett Jolene.Raggett@gov.bc.ca

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

1. STUDY DESIGN, SCOPE AND SITE SELECTION

The Canadian Aquatic Biomonitoring Network (CABIN) Program is a national program that provides tools for collaborative freshwater ecosystem health monitoring and assessment using benthic macroinvertebrates (e.g., protocols, online database, web-accessible analytical and reporting tools; Environment Canada, 2012, 2014). There are CABIN bioassessment models for all the major watersheds in British Columbia (B.C.).

The B.C. Ministry of Environment and Climate Change Strategy (BC-ENV) built a CABIN model for Vancouver Island and South Gwaii Haanas in 2012 (Gaber, 2012). However, the model required an update for several reasons:

1)   The 2012 model included reference data from sites on Haida Gwaii, which is now part of the North/Central Coast CABIN model. The update will ensure that the model covers a geographically distinct boundary, as recommended by the CABIN Science Team.
2)   The model as developed by BC-ENV (Gaber, 2012) was not loaded to the CABIN website due to concerns related to the predictor variables and modifications were made to the list of predictors before being uploaded to the CABIN analytical tools. The update will ensure that the model uploaded to the CABIN analytical tools is consistent with technical documentation.
3)   The 2012 model only contained data up to 2010. The update will ensure the model includes more recent data (i.e., up to 2019).

1.1 Model Purpose

This document describes an updated CABIN model with a focus on watersheds on Vancouver Island only (see Figure 1). The Reference Condition Approach (RCA) and modelling methods are based on the CABIN model builder checklist and existing BEAST model documentation (e.g., Environment Canada, 2012; Reynoldson and Bailey, 2015; Novodvorsky and Bailey, 2016; Strachan and Pappas, 2016). A full description of the development and testing of the model is provided in a separate technical report (Somers, 2021). Streams on Vancouver Island are exposed to a variety of disturbances, including extensive forest harvest, rural and urban settlements, mining, and agriculture, among others. This updated model will help assess the effects of these stressors on aquatic biota in streams from across Vancouver Island.

1.2 Spatial Scope

Vancouver Island is located off the southwest coast of B.C. It is the largest island on the Pacific Coast of North America, with an area of 32,134 km^2 and has a population of 870,297 (as of 2019). It is rich with indigenous culture and is the home of many Nations within the Nuu-chah-nult, Kwakwaka'wakw, and Coast Salish linguistic groups.

The island is divided into two ecoregions, with the sites distributed approximately evenly among both (Western Vancouver Island [45%] and Eastern Vancouver Island [55%]). The ecoregions are further separated into 6 different ecosections. This model considers data from 125 reference and potential reference samples, which are distributed among all ecozones (Figure 1) and stream orders (i.e., 1 – 11 samples, 2 – 35, 3 – 39, 4 – 34, and 5 - 6).
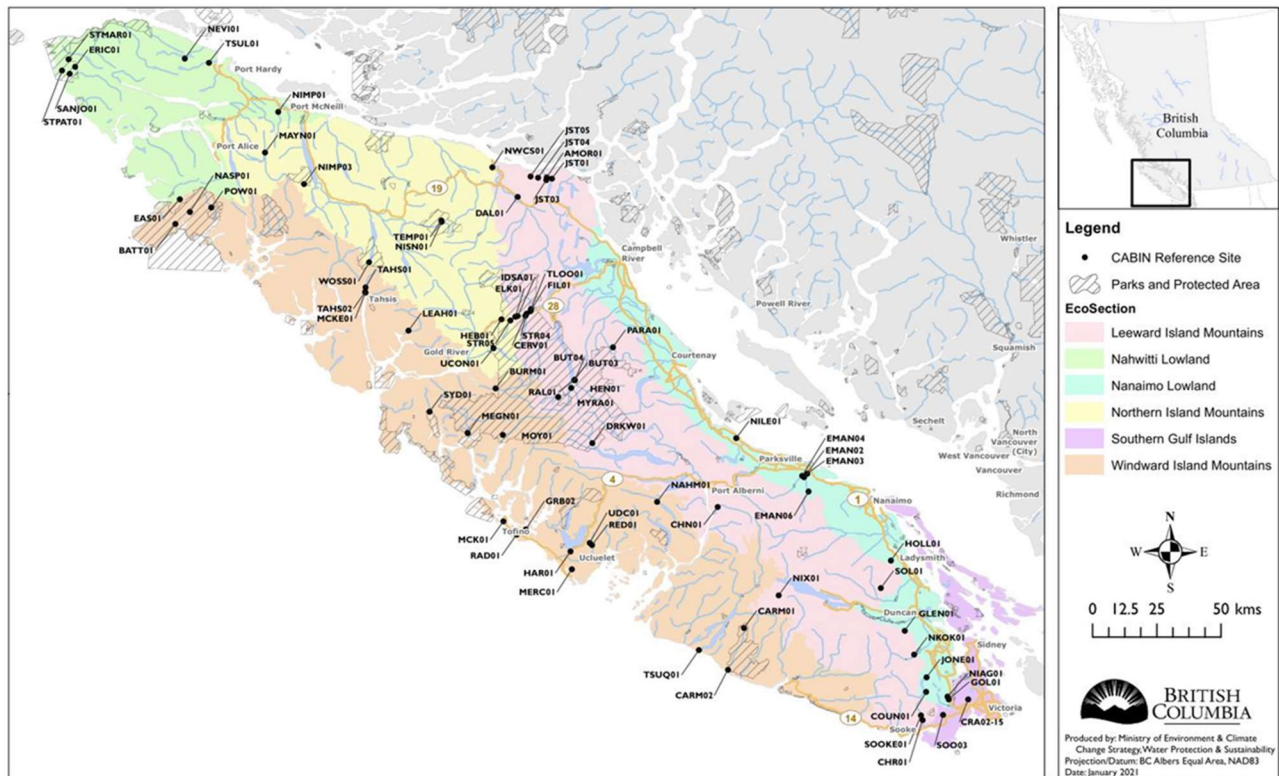

Figure 1 Map of Vancouver Island showing locations of the reference sites.

1.3 Temporal Scope

Samples were collected between 2001 and 2019 (Table 1) by CABIN certified field staff. All samples were collected in the late summer or early fall following the methods outlined in the CABIN field manual (Environment Canada, 2012). Some sites were re-visited and sampled in multiple years to capture temporal variation (Table 2). For the updated model, samples from the same site collected over multiple years were treated as if they were separate sites.

Table 1 Number of samples collected on Vancouver Island between 2001 and 2019.

Site Type	Sampling Year																		Total	
	2001	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	Sites	Samples
Reference	21	0	0	0	6	18	22	12	8	6	5	4	6	3	2	3	5	4	78	125
Test	6	6	8	9	19	11	1	9	36	22	7	13	3	5	0	1	4	0	93	160
All	27	6	8	9	25	29	23	21	44	28	12	17	9	8	2	4	9	4	166	285

Table 2 Number of years of sample collection for the different types of Vancouver Island CABIN sites.

Site Type	Number of Years Sampled												Total	
	1	2	3	4	5	6	7	8	9	10	11	12	Sites	Samples
Reference	53	19	3	0	1	0	0	1	0	0	0	1	78	125
Test	62	21	3	2	1	0	0	3	0	1	0	0	93	160
All	106	43	7	2	2	0	0	4	0	1	0	1	166	285

1.4 Reference Site Selection

Minimally disturbed sites were selected from the wide variety of landscape types and stream sizes across Vancouver Island. Samples were collected by Environment and Climate Change Canada (ECCC) and BC-ENV and uploaded to three CABIN studies: BC MOE-Vancouver Island Region (96), EC-Pacific Rim/Western Vancouver Island (8), and EC-Vancouver Island (21).

The approach to identifying reference sites evolved over the years of data collection (2001 – 2019) with new geographic information systems (GIS) techniques and availability and accessibility to landscape level data. Prior to 2008, reference sites were selected using local area knowledge and the best professional judgment of the CABIN team collecting the samples to target streams with minimal stressors in the upstream watershed. Beginning in 2008, a GIS based tool was developed and implemented to select reference sites, which is described by Norris (2012). This approach used a query to select watersheds that met specific stressor criteria for Vancouver Island (see Table 3). Sites were short-listed from this query and were ultimately selected to represent a range of different stream orders and eco-sections across Vancouver Island based on accessibility and the presence of appropriate habitat at the site.

Table 3 GIS-based reference site selection criteria for Vancouver Island CABIN sites (based on Norris, 2012).

Watershed Criteria	Reference Site Selection Tool
Urbanization	<8%
Agriculture	<30%
Forestry	<20%
Road density	<0.5 km/km^2
Forest fire	<20%
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 downstream
Upstream distance from current and past producing mines within 100m of stream	>500m
Downstream from current and past producing mines within 100m of stream	None
Riparian (30m from stream) vegetation disturbance	None

2. REFERENCE DATA AND MODEL DEVELOPMENT

The initial steps in the BEAST bioassessment approach involve a series of analyses of the benthos samples from the reference sites (Reynoldson et al., 1995). The benthos data are based on counts of 110 Families of benthic invertebrates collected between 2001 and 2019 (see protocols in Environment Canada, 2012, 2014).

2.1 Biological description

The initial step in the analysis of the benthos data involved the calculation of a Bray-Curtis distance matrix between all possible pairs of the 125 reference site samples. The resultant Bray-Curtis matrix was evaluated using cluster analysis and nonmetric multidimensional scaling (NMDS) ordination. The goal of these analyses is to identify distinct groups (or clusters) of reference site samples that are characterized by different benthos assemblages, or different biological communities.

The cluster analysis used the 'average' option that is also known as the unweighted pair-group method using averages (UPGMA). The dendrogram was inspected to identify groups or clusters of sites representing different types of biological communities. A randomization test was completed to determine whether the observed tree structure was nonrandom (Clarke et al., 2008). The solution indicated the existence of 15 nonrandom clusters of samples (see Figure 2), although 11 of the 15 clusters were represented by small numbers of sites (i.e., <10 samples per group). Reference sites belonging to each cluster are listed in Appendix C in Somers (2021). A 9-group solution at a Bray-Curtis value of 0.675 distinguished 4 main groups (with 15 [yellow – Group 2], 28 [light green – Group 1], 38 [blue – Group 6] and 28 [purple – Group 5] observations) leaving several smaller groups and outliers (i.e., with 8, 4, 2, 1 and 1 samples, see Figure 2). The smaller groups and outliers were dropped from further consideration as suggested by Reynoldson and Wright (2000).
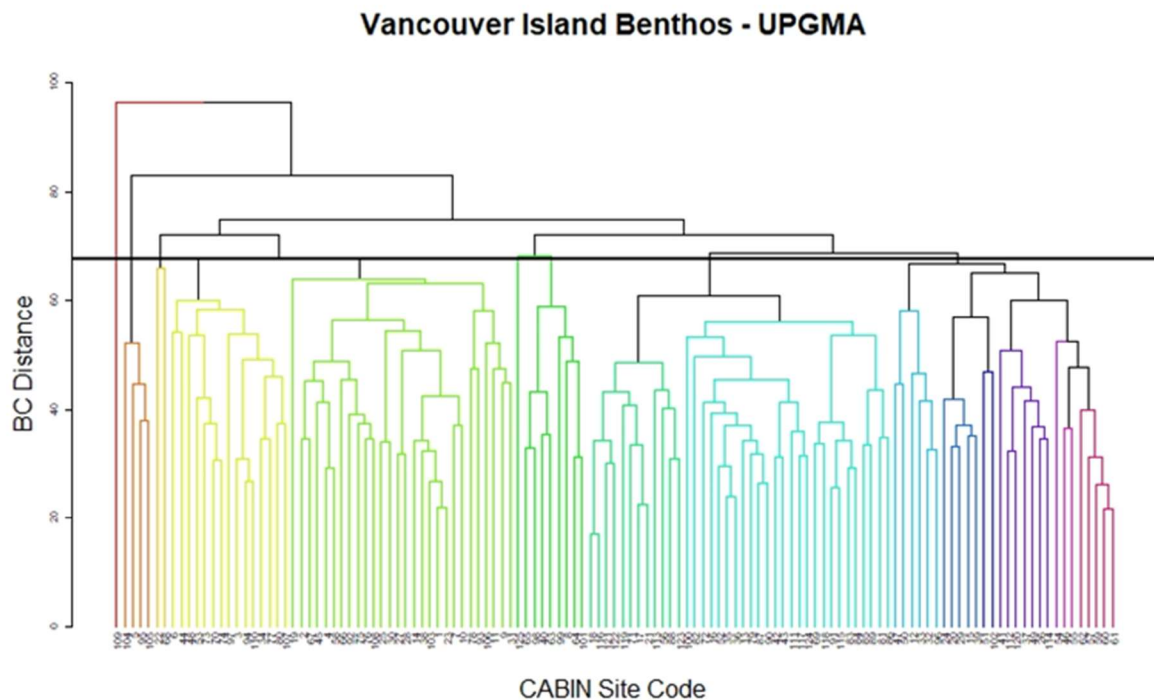
Vancouver Island Benthos - UPGMA

Figure 2 The dendrogram (or tree diagram) resulting from the cluster analysis (UPGMA) of the Bray-Curtis distance matrix for the benthos from 125 reference site samples on Vancouver Island. The dendrogram is colour-coded with the 15-group randomization test result (i.e., branches with solid black lines are nonrandom) and proposed 9-group solution at a Bray-Curtis distance of 67.5 (horizontal line).

Different benthos communities characterize the 4 major groups of reference site samples (Table 4). For example, the average abundance (or density) of the benthos in Group 1 and Group 2 is less than 1/3 of the average abundance for samples in Group 5 or Group 6. By comparison, richness (the number of benthos Families) is similar across all 4 groups with ranges that overlap considerably. Group 5 has fewer EPTs (i.e., Ephemeroptera – mayflies, Plecoptera – stoneflies, and Trichoptera – caddisflies) and more chironomids (i.e., midges) than the other 3 groups of reference site samples. Generally, the diversity (based on the Shannon or Simpson diversity index) of benthos samples collected from sites in Group 6 is lower than the diversity found at the other reference sites. Interestingly, abundance at the test sites is almost double, on average, the abundance at reference sites. Test-site richness is similar to richness at the reference sites, on average, although the minimum richness for the test sites is almost half the minimum richness at the reference sites. The EPT richness and proportion of EPTs at test sites was lower, on average, than the reference-site groups. By contrast, diversity (both measures) for the test sites was comparable to the reference sites, although the minimum value for the test sites was lower than the reference sites.

Table 4 Summary statistics for 7 common benthos indices calculated for the 4 largest groups of reference site samples in the cluster analysis of the benthos data. Values are also tabulated for the test sites to highlight differences among groups of sites. Abundance is the total number (or density) of individuals found at a site. Richness is the number of Families. The % EPT is the proportion of Ephemeroptera (E), Plecoptera (P), and Trichoptera (T) in the total abundance. Richness is also presented as EPT_Rich; the number of E, P, and T Families. The %Chir index is the proportion of Chironomidae in the total abundance. Shannon and Simpson are two common diversity measures.

Group	Statistic	Abundance	Richness	%EPT	EPT_Rich	%Chir	Shannon	Simpson
Group 1	Mean	559.0	20.8	76.3	11.9	12.4	2.174	0.828
N=28	StDev	293.58	3.66	12.72	1.92	9.40	0.263	0.055
	Min	160.0	14.0	50.1	8.0	1.9	1.598	0.696
	Max	1327.8	28.0	95.0	15.0	38.4	2.609	0.900
Group 2	Mean	796.2	20.5	66.4	10.4	8.2	2.197	0.810
N=15	StDev	288.82	3.36	16.72	2.03	5.67	0.428	0.114
	Min	288.0	13.0	32.5	5.0	1.2	1.307	0.527
	Max	1274.0	27.0	96.6	13.0	24.3	2.785	0.926
Group 5	Mean	2022.9	21.6	52.9	11.9	28.4	2.130	0.794
N=28	StDev	944.73	4.57	19.77	2.22	15.60	0.309	0.080
	Min	558.4	13.0	20.4	8.0	4.2	1.461	0.592
	Max	4457.2	30.0	85.9	16.0	62.8	2.621	0.887
Group 6	Mean	2552.2	19.2	67.4	10.9	10.0	1.992	0.781
N=38	StDev	1550.33	3.69	16.78	2.27	5.89	0.362	0.112
	Min	553.0	11.0	32.2	6.0	0.6	0.787	0.310
	Max	8125.0	26.0	96.1	15.0	22.9	2.484	0.886
Test Sites	Mean	4492.6	20.4	45.2	9.5	26.5	2.024	0.775
N=160	StDev	6419.94	4.12	24.26	3.00	19.16	0.430	0.131
	Min	144.0	7.0	0.4	1.0	0.6	0.587	0.238
	Max	57840.0	30.0	91.4	16.0	87.1	2.758	0.909

The benthos Families that most contributed to separating the 4 main groups of reference site samples were identified using the Simper approach that decomposes the Bray-Curtis distance matrix into a series of contrasts between pairs of clusters (Table 5; see Clarke, 1993). Taxa that contribute to the separation of two clusters are ranked and tabulated based on their relative importance. Several mayfly Families and chironomids were generally among the most important taxa that separated the 4 groups of reference site samples.

Table 5 Pairwise comparisons of the 4 major groups of reference site samples using the Simper approach. Each sub-table lists the top 10 benthos Families contributing to the Bray-Curtis distance (BC) separating the two groups (i.e., average [Avg.BC], standard deviation [sd.BC], ratio of mean:sd [ratio], and cumulative sum [C.Sum] for the BC contribution). The average abundance for a given benthos Family in each group is also provided (e.g., Avg(Grp1) and Avg(Grp2)).

Grp1 vs Grp2	Avg.BC	sd.BC	ratio	Avg(Grp1)	Avg(Grp2)	C.Sum
Baetidae	0.126	0.085	1.479	54.1	199.1	0.186
Heptageniidae	0.074	0.067	1.097	140.2	45.6	0.296
Chironomidae	0.047	0.041	1.144	68.0	72.0	0.365
Chloroperlidae	0.042	0.036	1.184	39.1	62.2	0.428
Ephemerellidae	0.038	0.042	0.883	53.5	17.5	0.483
Taeniopterygidae	0.036	0.052	0.687	51.3	1.4	0.536
Torrenticolidae	0.032	0.039	0.817	15.0	44.9	0.583
Glossosomatidae	0.032	0.043	0.727	14.8	44.3	0.630
Nemouridae	0.027	0.026	1.008	26.4	35.3	0.669
Elmidae	0.024	0.038	0.630	1.1	36.7	0.705

Grp5 vs Grp6	Avg.BC	sd.BC	ratio	Avg(Grp5)	Avg(Grp6)	C.Sum
Baetidae	0.156	0.117	1.331	108.7	837.0	0.227
Chironomidae	0.087	0.075	1.149	511.5	267.8	0.353
Simuliidae	0.060	0.075	0.798	17.5	279.6	0.440
Heptageniidae	0.057	0.051	1.120	207.8	295.8	0.523
Ephemerellidae	0.043	0.063	0.673	194.6	94.0	0.585
Chloroperlidae	0.032	0.031	1.033	74.0	179.9	0.632
Lepidostomatidae	0.028	0.058	0.484	128.8	16.3	0.673
Taeniopterygidae	0.028	0.060	0.466	105.9	14.4	0.714
Nemouridae	0.028	0.029	0.972	123.6	94.6	0.754
Torrenticolidae	0.025	0.021	1.194	120.9	93.7	0.791

Grp1 vs Grp5	Avg.BC	sd.BC	ratio	Avg(Grp1)	Avg(Grp5)	C.Sum
Chironomidae	0.182	0.119	1.531	68.0	511.5	0.245
Heptageniidae	0.071	0.074	0.959	140.2	207.8	0.341
Ephemerellidae	0.063	0.093	0.681	53.5	194.6	0.426
Taeniopterygidae	0.056	0.094	0.590	51.3	105.9	0.501
Lepidostomatidae	0.044	0.085	0.515	6.4	128.8	0.560
Torrenticolidae	0.041	0.031	1.320	15.0	120.9	0.615
Nemouridae	0.039	0.045	0.872	26.4	123.6	0.668
Baetidae	0.036	0.032	1.124	54.1	108.7	0.716
Chloroperlidae	0.025	0.028	0.911	39.1	74.0	0.750
Elmidae	0.013	0.020	0.619	1.1	44.8	0.767

Grp2 vs Grp5	Avg.BC	sd.BC	ratio	Avg(Grp2)	Avg(Grp5)	C.Sum
Chironomidae	0.163	0.109	1.493	72.0	511.5	0.222
Baetidae	0.061	0.051	1.205	199.1	108.7	0.305
Heptageniidae	0.060	0.071	0.838	45.6	207.8	0.387
Ephemerellidae	0.058	0.088	0.662	17.5	194.6	0.467
Lepidostomatidae	0.040	0.078	0.510	19.1	128.8	0.521
Taeniopterygidae	0.039	0.091	0.433	1.4	105.9	0.574
Nemouridae	0.038	0.041	0.920	35.3	123.6	0.626
Torrenticolidae	0.035	0.028	1.238	44.9	120.9	0.673
Chloroperlidae	0.024	0.024	0.998	62.2	74.0	0.705
Naididae	0.020	0.044	0.459	30.9	44.5	0.733

Grp1 vs Grp6	Avg.BC	sd.BC	ratio	Avg(Grp1)	Avg(Grp6)	C.Sum
Baetidae	0.233	0.150	1.552	54.1	837.0	0.313
Simuliidae	0.090	0.106	0.848	5.1	279.6	0.433
Heptageniidae	0.069	0.057	1.206	140.2	295.8	0.526
Chironomidae	0.063	0.050	1.275	68.0	267.8	0.610
Chloroperlidae	0.049	0.043	1.118	39.1	179.9	0.676
Torrenticolidae	0.033	0.033	1.006	15.0	93.7	0.719
Nemouridae	0.027	0.028	0.955	26.4	94.6	0.755
Ephemerellidae	0.027	0.038	0.692	53.5	94.0	0.791
Taeniopterygidae	0.021	0.030	0.691	51.3	14.4	0.819
Hydropsychidae	0.015	0.018	0.800	3.4	48.4	0.839

Grp2 vs Grp6	Avg.BC	sd.BC	ratio	Avg(Grp2)	Avg(Grp6)	C.Sum
Baetidae	0.180	0.131	1.371	199.1	837.0	0.263
Simuliidae	0.080	0.097	0.829	16.3	279.6	0.380
Heptageniidae	0.075	0.053	1.406	45.6	295.8	0.489
Chironomidae	0.058	0.047	1.234	72.0	267.8	0.573
Chloroperlidae	0.041	0.038	1.081	62.2	179.9	0.633
Torrenticolidae	0.030	0.028	1.061	44.9	93.7	0.676
Nemouridae	0.026	0.026	1.006	35.3	94.6	0.714
Ephemerellidae	0.023	0.034	0.668	17.5	94.0	0.748
Elmidae	0.018	0.028	0.648	36.7	39.0	0.775
Glossosomatidae	0.016	0.024	0.679	44.3	20.4	0.798

To illustrate the relative proximity of groups of reference sites identified by the cluster analysis, an NMDS ordination of the Bray-Curtis distance matrix was completed. The analysis used the Bray-Curtis distance with the 'hybrid' fitting model and a 3-D solution. Fifty random starting configurations were used to find the optimal solution with a minimum stress (i.e., lack of fit). The final solution had a stress value of 0.145 suggesting a reasonable approximation of the Bray-Curtis distance matrix. Groups of samples based on the UPGMA clustering solution were highlighted with coloured symbols in the ordination plot (Figure 3). The outlier samples were clearly separated from the other samples, although many groups overlapped to some extent.
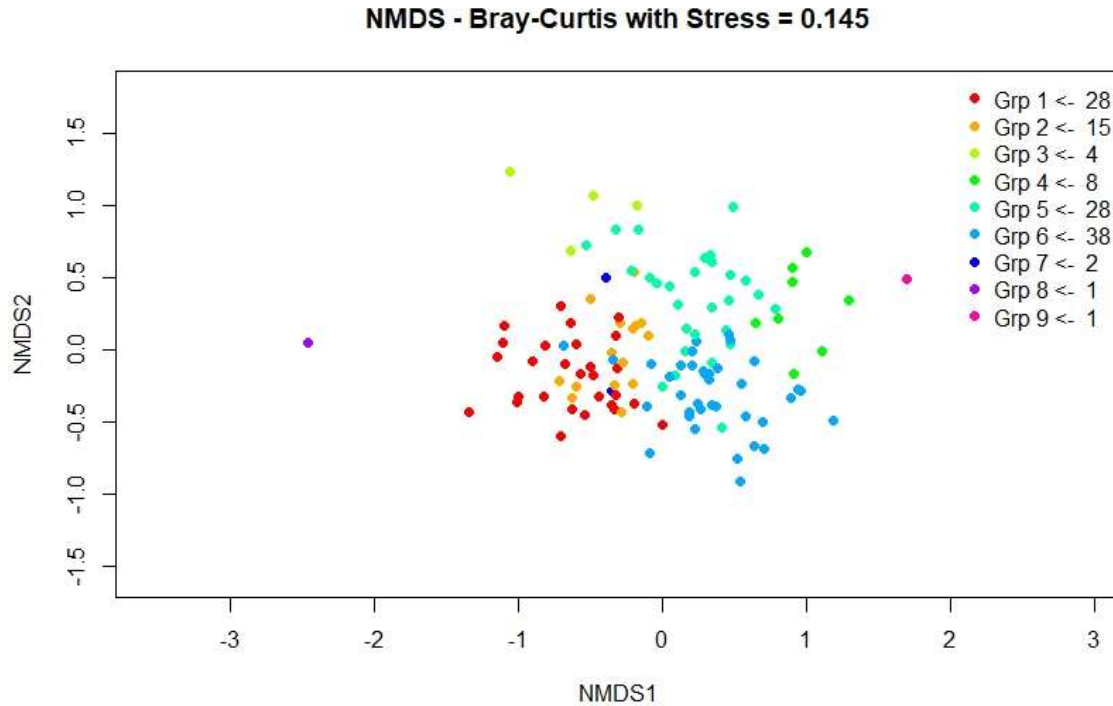

Figure 3 NMDS ordination solution for the first two axes showing the relative positions of members of the 9 groups of reference site samples identified in the UPGMA cluster analysis.

A correlation analysis of benthos abundance at the Family taxon level with the axis scores for the NMDS ordination indicated that abundances of only 21 of the 110 taxa produced correlations that were larger than +/- 0.30 (Table 6). Axis 1 (NMDS1) was correlated with the abundances of mayflies (e.g., Baetidae, Heptageniidae), caddisflies (Brachycentridae, Hydropsychidae), stoneflies (Chloroperlidae, Nemouridae), dipterans (Ceratopogonidae, Chironomidae, Tipulidae), and water mites (Lebertiidae, Sperchontidae, Torrenticolidae) suggesting a gradient in overall abundance with lower abundance at the negative end of the axis. By contrast, NMDS2 was correlated with the abundances of mayflies (negatively) and several caddisfly families (positively, e.g., Lepidostomatidae, Polycentropodidae), whereas NMDS3 was correlated with several stonefly and mayfly families (positively), and blackflies (negatively, Simuliidae).

Table 6 Pearson correlations between taxon (Family) abundances and the three NMDS ordination axes for the 125 reference site samples. Only taxa with correlations larger than |0.30| were retained for the table.

Taxon Family	NMDS1	NMDS2	NMDS3
Baetidae	0.5635	-0.4459	-0.2075
Brachycentridae	0.3617	0.1469	-0.0433
Candonidae	-0.1754	-0.1440	0.3877
Ceratopogonidae	0.3403	0.1950	0.1328
Chironomidae	0.6065	0.2879	0.0787
Chloroperlidae	0.4303	-0.2921	-0.1861
Elmidae	0.3213	0.1247	-0.1393
Ephemerellidae	0.3537	0.0711	-0.0398
Heptageniidae	0.4932	-0.2926	0.3139
Hydropsychidae	0.3869	0.0200	-0.1179
Lebertiidae	0.3416	0.0102	-0.0049
Lepidostomatidae	0.2640	0.4761	-0.1236
Leptophlebiidae	0.3029	0.1159	0.1274
Nemouridae	0.4947	0.1297	0.1091
Polycentropodidae	0.0980	0.3615	0.2481
Rhyacophilidae	0.3117	-0.0045	-0.0225
Simuliidae	0.3222	-0.2380	-0.2814
Sperchontidae	0.4059	0.1136	-0.0563
Taeniopterygidae	-0.0168	-0.1664	0.4744
Tipulidae	0.3700	0.0799	0.0066
Torrenticolidae	0.5611	0.2723	-0.0893

2.2 Habitat Description

A variety of habitat variables were measured for each site either in the field or by GIS. The habitat variables were grouped according to type: regional characteristics (e.g., latitude, longitude, altitude), bedrock geology, stream channel measurements, hydrology, topography, land cover, and climate.

The full set of habitat variables for the reference sites was screened (as described below) to find the characteristics that best described each site. The full data set included 317 candidate descriptors within the CABIN database. However, 75 contained no data and these variables were deleted. An additional 141 descriptors were dropped because they contained 10-or-more missing values. Because monthly and annual climate variables were highly correlated, a principal components analysis (PCA) of the 12 precipitation variables and a separate PCA of the 27 temperature variables were conducted. The variables that were most highly correlated (positively and negatively) with the first 3 principal components were retained; the other variables were considered redundant and were dropped. For the precipitation PCA, annual total, August, and November precipitation were retained (i.e., 9 of the 12 variables were dropped). For the temperature PCA, March max, annual max, annual min, December max, and July max were retained (i.e., 5 variables were retained and 22 were dropped). Three habitat variables were deleted because they were easily altered by anthropogenic activity (i.e., embeddedness, canopy cover, and exposed land). Additionally, 8 land cover variables were also removed because logging is an important stressor and logging can affect the proportions of various land cover types. This editing resulted in the retention of 33 habitat variables (see listing in Table 7).

For these 33 habitat variables, missing values for the reference and test sites were highlighted. One reference site was missing two values; the missing values were replaced with the mean value for the other reference sites. Two test sites were missing data that was replaced with results from a survey one year earlier in 2009 (i.e., LTSO05-10 and LTSO07-10). For the other test sites, 3 had one missing value, 2 had 2 missing values, and 1 had 3 missing values - all these missing values were replaced with the mean value for the other test sites.

To address potential concerns with model-building issues associated with multi-collinearity among the 33 habitat variables, a Pearson correlation matrix was calculated among all possible pairs of variables. Large positive and negative correlations were highlighted (i.e., abs(r)>0.6). Additionally, the variance inflation factor (VIF), tolerance, and associated redundancy (R^2) statistics were calculated for each variable relative to the other habitat variables. Large values for the VIF (i.e., > 10), tolerance (i.e., <0.1), and redundancy (i.e., R^2>0.9) were identified and compared. Based on this evaluation, no additional habitat variables were removed from the data set. The final data matrix of 33 habitat variables was populated with 125 reference and 160 test site samples (i.e., a total of 285 samples; Table 2).

To characterize the different groups of reference site samples identified by the cluster analysis, the 33 habitat characteristics for the 4 major groups of samples were summarized using the average, minimum and maximum values (Table 7). For comparison, the average and range for the 160 test site samples are also presented. Differences among the groups of samples based on the benthos assemblages (e.g., Table 4) are also evident in differences in habitat characteristics. For example, sites in Group 1 and Group 5 tend to be at higher altitude than sites in Group 2 and Group 6. Groups 1 and 6 have larger proportions of volcanic bedrock in their catchments compared to Groups 2 and 5. Stream bankfull width is greater in sites in Groups 2 and 6, relative to sites in Groups 1 and 5. Climatically, Group 1 sites have fewer degree days than the other sites, with lower average maximum air temperatures, but average annual

precipitation. Group 1 sites tend to have smaller watersheds, are located at higher altitude, and with steeper basins. Group 2 sites are generally found at lower altitude with less steep watersheds with larger amounts of intrusive bedrock, more degree days, and larger amounts of precipitation annually. Reference sites in Group 5 are at higher altitude with somewhat less annual precipitation, but they tend to be 'average' for many habitat characteristics. By contrast, Group 6 sites are at a lower altitude with watersheds with a large proportion of volcanic bedrock. The watersheds tend to be large and the streams are wider being somewhat larger in stream order.

Given the large number of test site samples (160), it is not surprising that the test sites span the range of characteristics of the reference sites. For example, the minimum depth for test-site streams is smaller than the reference sites and maximum depth is larger than reference streams. Similar patterns exist for degree days, annual total precipitation, and annual minimum temperature, although the minimum annual maximum temperature is higher for the test sites than the reference sites. The largest test-site watersheds are almost twice the size of the maximum size of the reference-site watersheds and this difference is also evident in maximum watershed perimeter and stream length. On average, the test sites tend to be at low altitude like reference sites in Group 2 and Group 6, but the highest altitude test site is higher than all the reference sites in Group 1, Group 2, and Group 6. Despite the wide range in habitat characteristics at the test sites, the 4 major groups of reference site samples generally span the range of habitats found at test sites.

Table 7 Summary of habitat characteristics for the 4 major groups of reference site samples and the test sites.

Habitat Characteristic	Group 1 (28)			Group 2 (15)			Group 5 (28)			Group 6 (38)			Test Sites (160)		
	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max
Latitude	49.7071	48.4800	50.7400	49.6127	48.4100	50.7000	49.2132	48.4100	50.3400	49.9905	48.4800	50.7600	49.2592	48.3900	50.5300
Longitude	-125.7275	-128.2900	-123.5700	-125.7900	-128.3400	-123.7100	-124.7796	-127.6200	-123.5700	-126.3232	-128.3000	-123.4600	-124.5359	-127.4600	-123.3100
Altitude (m)	232.4	18.0	610.0	89.7	10.0	290.0	234.2	8.0	1114.0	87.5	5.0	420.0	89.4	1.0	1071.0
StreamOrder	2.8	1.0	4.0	2.7	1.0	5.0	2.7	1.0	5.0	3.5	1.0	5.0	3.6	1.0	6.0
BG-Intrusive (%)	16.16	0.00	97.39	46.70	0.00	94.55	38.50	0.00	100.00	17.27	0.00	100.00	25.60	0.00	100.00
BG-Volcanic (%)	74.59	0.00	100.00	32.29	0.00	100.00	54.59	0.00	100.00	75.01	0.00	100.00	47.09	0.00	100.00
CH-Depth-Avg (cm)	19.1	3.0	39.5	18.7	2.4	36.3	15.7	2.3	42.8	20.0	3.5	60.8	25.1	0.1	78.6
CH-Depth-Max (cm)	29.9	4.0	70.0	29.4	3.5	60.0	25.6	4.0	68.0	31.3	7.0	112.0	40.4	0.2	140.0
CH-Velocity-Avg (m/s)	0.3	0.0	0.5	0.3	0.1	0.7	0.2	0.0	0.8	0.3	0.1	0.7	0.3	0.0	1.5
CH-Velocity-Max (m/s)	0.6	0.1	1.1	0.6	0.2	2.2	0.4	0.0	1.2	0.5	0.1	1.2	0.5	0.0	2.4
CH-Width-Bankfull (m)	14.2	2.0	36.3	19.6	2.7	75.0	13.9	0.9	54.0	22.2	4.1	88.0	27.0	1.0	97.8
CH-Width-Wetted (m)	8.5	1.0	28.7	12.5	1.3	55.9	8.1	0.9	33.5	11.5	2.6	34.5	17.5	0.9	76.0
CL-DegreeDays (Days)	617.6	69.0	1532.0	1123.9	109.0	1560.0	982.6	79.0	1468.0	1037.1	113.0	1639.0	877.7	37.0	1756.0
CL-Precip08_AUG (mm)	80.6	36.0	119.0	79.5	39.0	114.0	64.8	37.0	114.0	86.1	29.0	114.0	55.4	23.0	101.0
CL-Precip11_NOV (mm)	279.5	40.0	453.0	348.9	254.0	455.0	291.6	63.0	429.0	346.7	182.0	459.0	279.1	26.0	445.0
CL-PrecipTotal_ANNUAL (mm)	2219.8	1428.0	3234.0	2372.5	1516.0	3189.0	1952.9	1428.0	3052.0	2391.3	1042.0	3074.0	1811.4	750.0	3084.0
CL-Temp03_MARmax (Degrees Celsius)	5.6	2.0	9.9	7.3	4.3	10.0	6.4	3.0	8.7	6.7	4.1	9.7	7.8	3.0	10.7
CL-Temp07_JULmax (Degrees Celsius)	18.2	16.0	21.6	18.9	16.2	21.3	19.7	16.4	21.7	18.2	16.3	21.7	21.1	15.7	23.0
CL-Temp12_DECmax (Degrees Celsius)	2.4	0.0	7.7	4.0	0.9	7.8	2.7	-0.2	4.8	3.6	0.7	6.5	3.7	-0.2	7.7
CL-TempANNUALmax (Degrees Celsius)	19.1	16.9	22.5	19.7	17.1	22.1	20.6	17.0	22.5	19.0	17.1	22.3	21.7	16.6	23.3
CL-TempANNUALmin (Degrees Celsius)	-1.8	-4.5	2.4	-0.9	-3.7	2.2	-1.7	-5.1	0.0	-1.2	-4.0	2.0	-1.1	-5.1	2.2
HY-Drainage-Area (km^2)	25.0	0.8	85.5	56.4	0.9	305.8	55.7	0.8	410.2	50.6	2.2	184.6	251.9	0.8	920.7
HY-Perimeter (Km)	27.0	4.1	71.2	33.2	5.0	101.2	36.6	4.1	141.4	41.2	7.6	106.7	109.3	4.6	360.2
HY-StreamLength (m)	63698.5	594.6	226817.8	130855.0	1000.0	729483.4	121974.9	675.0	850035.6	115158.2	6422.7	364683.2	527612.9	250.4	2031185.3
TO-ElevationAvg (m)	817.4	40.0	1306.3	508.6	30.0	1046.0	705.1	315.2	1253.0	556.9	171.0	1074.0	456.7	22.0	1231.0
TO-ElevationMax (m)	1364.3	124.0	2025.0	1042.4	76.0	2025.0	1218.8	570.0	2084.0	1117.7	377.0	2049.0	1256.9	79.0	1790.0
TO-ElevationMin (m)	220.0	16.0	610.0	89.3	0.0	294.0	220.9	13.0	1087.0	96.1	6.0	426.0	71.3	1.0	1068.0
TO-Slope30-50% (%)	21.27	1.80	36.70	18.20	0.50	30.90	19.78	2.30	38.10	18.30	4.90	39.10	14.98	0.00	39.10
TO-Slope50-60% (%)	8.00	0.00	17.70	4.89	0.00	16.10	4.86	0.00	14.50	4.48	0.50	12.00	3.13	0.00	9.30
TO-SlopeAvg (%)	43.53	13.10	64.00	33.23	14.70	56.70	37.38	11.30	67.30	39.09	16.50	67.30	23.27	2.50	69.40
TO-SlopeGT60% (%)	29.50	0.00	56.90	16.88	0.00	48.20	21.43	0.00	61.50	25.75	0.20	64.90	7.74	0.00	71.30
TO-SlopeLT30% (%)	41.26	11.40	98.20	60.02	20.50	99.50	53.93	11.40	97.70	51.50	12.70	94.20	74.15	7.20	100.00
TO-SlopeMax (%)	155.69	45.00	293.90	156.51	46.90	329.10	166.97	42.50	439.80	174.22	75.50	381.40	142.87	12.90	276.10

2.3 Predictive Model and Associated Performance

With the identification of 4 biologically based groups of reference site samples, the next step in the BEAST bioassessment approach involved the development of a habitat – faunal group predictive model. The predictive model utilizes discriminant functions analysis (DFA) to maximally separate the 4 groups of sites using a subset of the habitat variables, known as the predictor variables. Because of the large number of variables relative to the number of reference site samples, a series of stepwise DFA analyses was completed using forward and backward variable selection algorithms that add or remove variables based on a variety of criteria. Because the modelling goal is to maximally separate the 4 groups of reference sites, the criterion for selecting a final model was the ability to separate groups of sites.

The number of sites correctly classified by a given model is generally higher in models with larger numbers of habitat variables. A 9-variable model was selected because it correctly classified 64% of the reference sites (i.e., 70 of 109 sites). The model includes the following predictor variables, where Tree_group identifies the 4 groups of sites from the cluster analysis:

Tree_group ~ Altitude + Stream Order + %Volcanic Bedrock + Channel Bankfull Width + Degree Days +

August Precipitation + Stream Length + Watershed Max Elevation + Watershed Min Elevation

The model included regional characteristics, bedrock geology, channel measurements, climate, and topographic variables. The DFA produced a Wilk's lambda of 0.368 with an F value of 4.31 (P<0.001). The F value can be viewed as a type of signal-to-noise ratio suggesting strong group separation. Mean values for the 4 groups of reference site samples are tabulated for the 9 habitat variables associated with the discriminant analysis model (Table 8). The magnitude of the differences between groups was evaluated with one-way ANOVA. The F value and associated P value provide an indication whether any group differed from the other groups for that variable. For example, altitude, volcanic bedrock, degree days, August precipitation, and minimum elevation produced highly significant F values with P values of <0.001. Although all 9 of these variables were selected for the model, individually they did not necessarily produce significant ANOVA results (e.g., stream length and maximum elevation).

Table 8 Mean values for the 4 groups of reference sites associated with the 9 predictor variables selected in the DFA. One-way ANOVA F values and associated P values are also presented.

Habitat Characteristic	Group 1	Group 2	Group 5	Group 6	F value	P value
Altitude	232.39	89.73	234.18	87.47	9.48	<0.001
StreamOrder	2.79	2.73	2.71	3.47	4.480	0.005
BG-Volcanic (%)	74.59	32.29	54.59	75.01	6.200	<0.001
CH-Width-Bankfull (m)	14.17	19.63	13.91	22.20	2.762	0.046
CL-DegreeDays (Days)	617.53	1123.94	982.58	1037.13	8.23	<0.001
CL-Precip08_AUG (mm)	80.59	79.43	64.78	86.00	6.96	<0.001
HY-StreamLength (m)	63699	130855	121975	115158	1.197	0.315
TO-ElevationMax (m)	1364.29	1042.40	1218.82	1117.74	1.397	0.248
TO-ElevationMin (m)	220.00	89.33	220.86	96.13	7.011	<0.001

To evaluate the performance of the DFA model, the classification results were tabulated to compare the original ('true') group membership with the predicted group membership based on the discriminant functions. The number or proportion of sites correctly classified to their respective group is a measure of model prediction accuracy. The tables provide separate results for each group of sites to illustrate that some groups may be better distinguished than others. That is, a larger proportion of sites may be incorrectly classified in some, but not all groups. Tables were generated from the original DFA classification results and a jack-knife (or leave-one-out) cross-validation approach.

The initial classification table shows that 70 of the 109 reference site samples (0.642) were correctly classified into their appropriate group (values on the diagonal of the table identify correct classifications; see Table 9). The largest group (Group 6 with 38 sites) was classified with the lowest proportional error (11 of 38 sites, or 0.289), although misclassified sites were allocated to every group. Nine of the 28 sites for Group 1 (0.321) and 13 of the 28 sites for Group 5 (0.464) were misclassified. By contrast, 6 sites from Group 2 (0.400) were misclassified and erroneously allocated to the other groups.

The jackknife table provides an estimate of the reliability of the overall classification based on the DFA model (Table 9). Although the original results indicated that 64.2% of the sites were correctly classified, the jackknife results showed that 57.8% of the reference sites were correctly classified using a leave-one-out strategy. That is, the original model misclassified 39 sites whereas the jackknife approach misclassified 46 sites. Group 6 had the lowest misclassification error (0.316) and Group 5 had the highest error (0.643).

Table 9 Classification tables (or confusion matrices) for the 4-group DFA model based on counts of observations contrasting observed 'true' group and predicted group. The lower table provides results based on the jackknife sub-sampling approach.

	Predicted Group 1	Predicted Group 2	Predicted Group 5	Predicted Group 6	% Correct
Assigned Group 1	19	3	5	1	67.9%
Assigned Group 2	1	9	2	3	60.0%
Assigned Group 5	7	3	15	3	53.6%
Assigned Group 6	2	5	4	27	71.1%
Total	29	20	26	34	64.2%

Jackknife	Predicted Group 1	Predicted Group 2	Predicted Group 5	Predicted Group 6	% Correct
Assigned Group 1	19	3	5	1	67.9%
Assigned Group 2	1	8	3	3	53.3%
Assigned Group 5	10	4	10	4	35.7%
Assigned Group 6	2	6	4	26	68.4%
Total	32	21	22	34	57.8%

The cross-validation classification efficacy of the DFA model was also evaluated by separating the original data set into two fractions. The larger fraction, the calibration set, was used to generate a DFA model using the 9 variables, and the smaller fraction, the validation set, was classified using that model. The classification results for both the calibration and validation sets were tabulated and compared to the known group membership to evaluate model prediction success. That is, the 109 reference sites were randomly partitioned in an 80:20% split in proportion to the relative sizes of the 4 groups.

Because modelling and prediction success is a function of a given random draw of sites, 20 sets of calibration and validation sites were sampled and evaluated (Table 10). The first row in the table (Run Number 0) indicates the true allocation of reference sites to the respective groups with an 80:20 split. The subsequent 20 rows identify the number of correctly allocated calibration and validation sites. The original DFA model correctly allocated 64.2% of the reference sites to their correct group (Table 9). By contrast, the partitioned data sets correctly classified from 61-72% of the calibration observations and 38-71% of the validation sites. On average, the classification success of the calibration set (0.660) approximated results based on the full data set (0.642, see Table 9). By contrast, correct classification for the validation set was always lower than the calibration set with the poorest classification associated with the smallest group (i.e., Grp2 = 0.213). This result was also observed with the calibration set where the classification success for Group 2 was 0.568. The best cross-validation rates were produced by Group 1 and Group 6. Overall classification success was generally comparable to results from other cross-validation studies (e.g., Strachan et al., 2014; Novodvorsky and Bailey, 2016; Strachan and Pappas, 2016).

Table 10 Cross-validation classification results for the 4-group DFA model based on 20 random 80:20 (calibration: validation) partitions of the 109 reference sites.

Run Number	Calibration Set - Number Correct					Validation Set - Number Correct				
	Grp1	Grp2	Grp5	Grp6	%	Grp1	Grp2	Grp5	Grp6	%
0	23	11	23	31		5	4	5	7	
1	14	5	12	27	0.6591	4	0	3	2	0.4286
2	18	6	12	22	0.6591	4	1	2	6	0.6190
3	17	5	12	24	0.6591	2	1	3	6	0.5714
4	14	5	14	24	0.6477	5	1	2	6	0.6667
5	14	4	16	25	0.6705	3	1	0	6	0.4762
6	15	9	11	25	0.6818	3	0	3	6	0.5714
7	15	6	10	24	0.6250	4	1	3	6	0.6667
8	13	8	12	26	0.6705	4	0	1	6	0.5238
9	12	7	14	23	0.6364	4	1	1	6	0.5714
10	17	7	12	27	0.7159	3	0	2	5	0.4762
11	20	5	11	25	0.6932	3	1	1	3	0.3810
12	13	8	11	25	0.6477	4	1	1	4	0.4762
13	15	6	13	25	0.6705	4	1	3	5	0.6190
14	13	8	8	25	0.6136	3	0	5	7	0.7143
15	18	3	12	24	0.6477	3	4	2	6	0.7143
16	14	6	11	26	0.6477	4	2	2	7	0.7143
17	12	8	14	23	0.6477	4	0	2	5	0.5238
18	15	6	11	23	0.6250	4	1	5	4	0.6667
19	19	6	8	28	0.6932	1	1	3	5	0.4762
20	15	7	12	27	0.6932	3	0	2	4	0.4286
Average (%)	0.6587	0.5682	0.5130	0.8032	0.6602	0.6900	0.2125	0.4600	0.7500	0.5643

2.4 Assessing Sites using the Updated Model

The Vancouver Island CABIN data set consisted of samples from 125 reference sites and 160 test sites (test sites are sites of unknown condition). The cluster and ordination analysis of the benthos data indicated that the reference site samples could be classified into 4 groups or types of benthos communities and the DFA model used 9 habitat variables to classify sites into one of the 4 groups of reference sites. Using habitat data, the 160 test sites were projected with the DFA model to classify each test site to the most likely group of reference sites. If a test site is not biologically impaired, then the benthos at that test site should be similar to the benthos found at matching reference sites with similar habitat characteristics.

The benthos data for each test site was compared to the benthos from the predicted (matching) group of reference sites with an NMDS ordination. To determine if the test-site benthos differed from the benthos at the reference sites, the position of the test site was compared to a series of confidence ellipses constructed around the reference sites in the NMDS. Three ellipses were constructed to contain 90%, 99% and 99.9% of the reference sites in each of the NMDS plots. These confidence ellipses represent 3 thresholds for assessing impairment. Sites lying within the 90% confidence ellipse are judged to be equivalent to reference and are deemed to be unimpaired. Sites lying between the 90% and 99% ellipses have low levels of impairment. Similarly, sites lying between the 99% and 99.9% ellipses have moderate impairment, whereas sites falling outside the 99.9% confidence ellipse are highly impaired.

As an example, the 109 reference site samples were individually evaluated with the 3 ellipses on plots of the 3 NMDS axes (Table 11A). None of the reference sites was assessed as displaying moderate or high levels of stress. Only 11 reference site samples were identified with a low level of impairment, whereas the majority (i.e., 98 of 109, or 0.899) were assessed as in reference condition. If all the reference sites were truly in reference condition, the 11 mis-classified reference site samples represent a false positive classification error, also known as a Type I error (i.e., 11 of 109, or 0.101). Interestingly, a 10% Type I classification error is consistent with using a 90% confidence ellipse as a threshold to characterize the reference condition.

The 109 reference site samples were also re-classified using the DFA model (Table 11B). That is, the reference sites were treated as test sites and classified to one of the four reference-site groups using the DFA model. Assessments using NMDS and 3 confidence ellipses indicated that 18 of the 31 reference sites assigned to Group 1 were in reference condition. Four of 13 reference sites assigned to Group 2 were assessed as impaired. Similarly, 8 of 24 Group 5 sites, and 14 of 41 Group 6 sites were impaired. Of the 39 reference sites that were deemed to be impaired, the majority exhibited a low level of impairment (i.e., 24 of 39, or 0.615). Regardless, 35.8% of the presumed reference sites were incorrectly judged to be impaired (i.e., 39 of 109). This result suggests that the Type I error, incorrectly evaluating a reference site as impaired, is 35.8%. If reference sites falling in the low category of impairment are treated as minimally impaired and lumped with the other unimpaired reference sites, the number of impaired reference sites falls to 15 producing a Type I error rate of 13.8% (i.e., 15 in 109).

The 160 test site samples were classified using the DFA model and assessed with NMDS and the confidence ellipses (Table 11C). The model predicted that 10 test site samples belonged to Group 1, 10 to Group 2, 103 to Group 5, and 37 to Group 6. Assessments relative to the 90%, 99% and 99.9% ellipses indicated that none of the Group 1 test sites were in reference condition and 4 had a low level of impairment. By contrast, one of the 10 test sites classified to Group 2 was in reference condition, 5 had low, none had moderate, and 4 had high levels of impairment. Overall, almost 1/3 of the test sites were unimpaired (0.313), just over 1/3 exhibited a low level of impairment (0.356), and about 1/6 of the test sites had moderate (0.175) or high (0.156) levels of impairment.

Table 11 Summary of the assessment results for the 109 reference and 160 test site samples.

| A - Ref Sites | Reference Site Status | | | | No. Impaired | | Total |
Group	Ref	Low	Mod	High	Total	Proportion	Sites
Grp1	24	4	0	0	4	0.143	28
Grp2	15	0	0	0	0	0.000	15
Grp5	24	4	0	0	4	0.143	28
Grp6	35	3	0	0	3	0.079	38
Total	98	11	0	0	11		109
Proportion	0.899	0.101	0.000	0.000		0.101	
B - Modelled Ref	Reference Site Status				No. Impaired		Total
Group	Ref	Low	Mod	High	Total	Proportion	Sites
Grp1	18	9	4	0	13	0.419	31
Grp2	9	2	1	1	4	0.308	13
Grp5	16	7	1	0	8	0.333	24
Grp6	27	6	3	5	14	0.341	41
Total	70	24	9	6	39		109
Proportion	0.642	0.220	0.083	0.055		0.358	
C - Modelled Test	Test Site Status				No. Impaired		Total
Group	Ref	Low	Mod	High	Total	Proportion	Sites
Grp1	0	4	2	4	10	1.000	10
Grp2	1	5	0	4	9	0.900	10
Grp5	44	35	14	10	59	0.573	103
Grp6	5	13	12	7	32	0.865	37
Total	50	57	28	25	110		160
Proportion	0.313	0.356	0.175	0.156		0.688	

Misclassifying truly impaired test sites as unimpaired is a false negative (or Type II) error. If all test sites were truly impaired, no test site samples should have been classified as in reference condition. However, about 1/3 of the test sites were deemed to be unimpaired producing a Type II error of 31.3% (i.e., 50 of 160; Table 11C). Because the true condition of the test sites is unknown, the relative proportions of test sites assessed as equivalent to reference is not a true estimate of the Type II error. That is, the observed result is simply a consequence of the selection of this set of test sites. A different set of test sites would likely produce a different result suggesting a different Type II error rate.

2.5 Assessing Sites using Simulated Disturbance (SIMPACT)

As noted above, attempts to estimate misclassification errors for real-world sites are confounded by the fact that the true status of a real site is generally unknown (Bailey et al., 2012; Strachan and Reynoldson, 2014). Although resampling methods such as jackknife and calibration-validation subsets can be used to evaluate classification-based DFA models (e.g., Strachan et al., 2014; Reynoldson and Bailey, 2015), quantifying assessment errors is difficult because true status is unknown. One solution involves the alteration of presumably unimpaired sites by simulation (Downie, 2011; Bailey et al., 2014). That is,

presumed reference site samples are manipulated to change the underlying benthos data to resemble hypothesized impaired benthos communities (Bailey et al., 2012). The manipulations involve increases or decreases in abundances, as well as decreases in occurrences, of certain taxa based on known sensitivities or tolerances to specific stressors (Downie, 2011). Different combinations of multipliers are used to simulate impacted (i.e., simpacted) test sites with low, moderate, or high levels of stress.

Simpacted test site samples were generated using taxon tolerances presented in Strachan and Pappas (2016) and Strachan (2020), and the weighting scheme described in Bailey et al. (2012). Ten sites from each of the 4 groups of reference sites (i.e., 40 sites) were randomly selected, with replacement. The benthos for each reference site was modified using the weights in Bailey et al. (2012) to create simpacted sites with low, moderate, and high levels of impairment. Simpacted sites in reference (no change), low, moderate, and highly impaired categories were summarized with the appropriate set of reference sites using NMDS and the Bray-Curtis distance. The resultant NMDS axis scores were plotted with 90%, 99% and 99.9% confidence ellipses. Positions of the simpacted sites relative to the confidence ellipses were evaluated and tabulated by simulated stressor level for each group of reference sites.

Results based on the simpacted sites can be used to evaluate false positive (Type I) and false negative (Type II) error rates for the NMDS ellipse assessment (Table 12). Based on the 40 simpacted sites, 3 of the randomly selected reference sites fell outside of the 90% confidence ellipse, suggesting that all but 3 of the sites were unimpaired before the simulation. Consequently, the Type I error of incorrectly judging a reference site to be impaired is 7.5% (i.e., 3/40 or 0.075). This result is somewhat larger than the 2% rate reported by Strachan and Pappas (2016) in their simpact analyses. By contrast, the Type II error rate reflects the number of impaired sites that were incorrectly classified as unimpaired. For the low impairment category, 11 simpacted sites were identified as exhibiting low impairment (or greater) indicating that 29 of the simpacted sites in this category were incorrectly classified as unimpaired (i.e., 29/40 or 72.5%). For the moderately impaired simpacted sites, 31 (77.5%) were identified as impaired (22.5% were judged unimpaired) and 14 (35%) were judged to be moderately impaired (or higher). For simpacted sites with the highest simulated stress level, 37 (92.5%) were deemed to be impaired (7.5% were unimpaired), and 26 (65%) were correctly judged to be highly impaired.

These results show that false negative, Type II errors decreased with simulations involving increasing levels of stress (i.e., low – 72.5% unimpaired, moderate – 22.5%, and high – 7.5%), although simpacted sites were often judged to be impaired at a lower threshold than the simulated level of impairment. Given these results, one might conclude that the approach lacks power to correctly identify impaired sites; however, this conclusion rests on the assumption that the simpact process correctly generates impaired sites at a specified level. Plots of the NMDS scores for simpacted sites (not shown) suggest that this assumption does not hold with some simpacted sites remaining in reference condition. Regardless, the false negative (Type II) error for simpacted sites at the highest level of impairment was 7.5%, suggesting that most sites with high levels of impairment will be detected.

Table 12 Assessment summary for 40 simpacted sites based on 10 randomly selected sites from each of the 4 groups of reference site samples. Values in each sub-table indicate the number of simpacted sites that met a particular assessment threshold (i.e., simpact status is based on 0.9, 0.99 or 0.999 confidence ellipses). The maximum number in any cell in the table is 10 (i.e., 10 simpacted sites per cluster group).

Simpact Type (N)	Cluster Group	Simpact Status				Number Impaired		Number Correct	
		Ref	Low	Mod	High	Total	Proportion	Total	Proportion
Ref (40)	Grp1	9	1	0	o	1	0.100	9	0.900
	Grp2	10	0	0	0	0	0.000	10	1.000
	Grp5	10	0	0	0	0	0.000	10	1.000
	Grp6	8	2	0	0	2	0.200	8	0.800
	Total	37	3	0	0	3		37	40
	Proportion	0.925	0.075	0.000	0.000		0.075		0.925
Low (40)	Grp1	5	4	0	1	5	0.500	4	0.400
	Grp2	8	2	0	0	2	0.200	2	0.200
	Grp5	10	0	0	0	0	0.000	0	0.000
	Grp6	6	2	2	0	4	0.400	2	0.200
	Total	29	8	2	1	11		8	40
	Proportion	0.725	0.200	0.050	0.025		0.275		0.200
Mod (40)	Grp1	0	5	3	2	10	1.000	3	0.300
	Grp2	0	4	1	5	10	1.000	1	0.100
	Grp5	7	3	0	0	3	0.300	0	0.000
	Grp6	2	5	1	2	8	0.800	1	0.100
	Total	9	17	5	9	31		5	40
	Proportion	0.225	0.425	0.125	0.225		0.775		0.125
High (40)	Grp1	0	1	0	9	10	1.000	9	0.900
	Grp2	0	0	0	10	10	1.000	10	1.000
	Grp5	3	5	1	1	7	0.700	1	0.100
	Grp6	0	1	3	6	10	1.000	6	0.600
	Total	3	7	4	26	37		26	40
	Proportion	0.075	0.175	0.100	0.650		0.925		0.650

2.6 A Comparison of the Preliminary and Updated Models

The preliminary model used 3 habitat variables (i.e., percent minimum watershed elevation, percent wetlands area, and number of National Parks in the catchment) to classify reference sites. By contrast, the updated model included 9 habitat variables: altitude, stream order, stream length, bankfull width, 2 climate variables (degree days and August precipitation), 2 topography variables (watershed maximum elevation and minimum elevation), as well as one bedrock feature (percent volcanic bedrock). The updated model correctly classified 64.2% of the reference sites with a jackknife-based classification rate of 57.8% and a cross-validation (i.e., repeated 80:20 calibration-validation subsets) success rate of 66% for the calibration sites and 56% for the validation sites. Prediction success for the updated 4-group model was lower than the results based on the preliminary model, although the preliminary model did not include the same habitat features.

To compare assessment results for the preliminary model and the updated 9-variable model, 62 Vancouver Island sites were evaluated using both models (Table 13). The preliminary model classified 19 of the 62 sites in reference condition (30.6%) and the remaining 43 sites as impaired (69.4%). By contrast, the updated model classified 13 sites in reference condition (21.0%) and 49 sites (79.0%) as impaired. Five

sites that were classified in reference condition by the preliminary model were also classified in reference condition by the updated model (i.e., 5 of 19 or 26.3% agreement). Nine of 10 sites classified as highly impaired by the preliminary model (90%) were also classified as highly impaired by the updated model. Overall, 27 of the 62 sites (43.5%) were assigned the same assessment status by both models. On average, the updated model classified sites to a higher level of impairment than the preliminary model (i.e., 26 sites or 41.9%). That is, the updated model was more sensitive than the preliminary model, classifying 85.5% of the sites to the same or higher level of impairment.

Table 13 Comparison of assessment results for 62 sites using the preliminary model and the updated 9-variable model.

| Updated Model Status | Gaber (2012) Model Status | | | | Updated Model | |
	Ref	Low	Mod	High	Sum	% Agreement
Ref	5	5	2	1	13	0.385
Low	10	10	1	0	21	0.476
Mod	2	6	3	0	11	0.273
High	2	4	2	9	17	0.529
Gaber Sum	19	25	8	10	62	
% Agreement	0.263	0.400	0.375	0.900		0.435

3. CONCLUSIONS AND RECOMMENDATIONS FOR MODEL USERS

This document describes an updated CABIN model for Vancouver Island streams. A preliminary model was developed using 106 reference site samples from Vancouver Island and south Gwaii Haanas using data collected between 2001 and 2010 (Gaber, 2012). The 106 reference site samples were clustered into 3 groups of 32, 36, and 34 sites with differing benthic invertebrate assemblages. The preliminary model used 3 habitat variables (i.e., percent minimum watershed elevation, percent wetlands area, and number of National Parks in the catchment) to classify reference sites into the 3 clusters based on different benthos communities.

The updated model was based on 125 reference site samples from Vancouver Island using habitat and benthos data collected between 2001 and 2019. A cluster analysis of the benthos community data indicated 4 groups of reference site samples with different biological assemblages (i.e., 28, 15, 28, and 38 samples). A few outliers and small groups of samples were dropped from the analysis leaving 109 reference site samples.

A predictive model contrasting the 4 groups of reference site samples was developed with DFA using 33 habitat variables including regional characteristics, bedrock geology, climatic, topographic, and stream-channel measurements. The updated predictive model included 9 habitat variables: altitude, stream order, stream length, bankfull width, 2 climate variables (degree days and August precipitation), 2 topography variables (watershed maximum elevation and minimum elevation), as well as one bedrock feature (percent volcanic bedrock).

The model correctly classified 64.2% of the reference sites with a jackknife-based classification rate of 57.8% and a cross-validation (i.e., repeated 80:20 calibration-validation subsets) success rate of 66% for the calibration sites and 56% for the validation sites. Prediction success for the updated 4-group model was lower than the results based on the preliminary Gaber (2012) model, although the Gaber model did not include the same habitat features.

To compare assessment results for the preliminary model and the updated 9-variable model, 62 Vancouver Island site samples were evaluated using both models. The preliminary model classified 19 of the 62 sites in reference condition (30.6%) and the remaining 43 sites as impaired (69.4%). By contrast, the updated model classified 13 sites in reference condition (21.0%) and 49 sites (79.0%) as impaired. Overall, 27 of the 62 sites (43.5%) were assigned the same assessment status by both models. On average, the updated model was more sensitive than the preliminary model, classifying 85.5% of the sites to the same or higher level of impairment.

The updated model was used to predict the group membership of 160 test site samples and each sample was evaluated in an NMDS ordination using 90%, 99% and 99.9% confidence ellipses around the matching group of reference sites. Fifty of the test site samples were judged to be unimpaired (31%), 57 exhibited low impairment (36%), 28 were moderately impaired (18%), and 25 were highly impaired (16%). Because the true biological condition of the test sites is unknown, the test-site assessment results cannot be used to estimate the false negative or Type II error rate of the model. An assessment of the modelled 109 reference site samples (i.e., the reference sites were classified to one of the 4 groups using the updated model) indicated that 70 were correctly assessed as unimpaired (64%), 24 were judged to have a low level of impairment (i.e., 22%), and 15 samples (13.8%) fell outside of the 99% threshold leading to an incorrect conclusion of moderate or high impairment. If the low level of impairment is assumed to be minimally impaired, the false positive, Type I error for assessing reference sites as impaired was 13.8%.

A further assessment was conducted using simulated impacted (simpacted) test sites, although the ordination results suggested that the simulation process generated test sites that were often less impaired than anticipated (i.e., the simpacted sites were often closer to the reference condition than expected). Regardless, 26 of 40 highly impaired simulated test sites were correctly identified as highly impaired (65%) and 37 of the 40 (92.5%) were deemed to be impaired. These results suggest that this type of assessment has a high probability of detecting highly impaired test sites.

3.1 Recommendations

1. Several errors were discovered in the Vancouver Island data extracted from the CABIN database. Some errors could be prevented with appropriate data quality objectives (e.g., site altitude values of 0m are not possible, but were found; so were duplicate names for the same site). New data should be checked for errors. Protocols for screening CABIN data could be considered to ensure data quality as new data are added to the database.

2. Resampled sites are important for evaluating changes in site status over time. A number of resampled sites (i.e., sites resampled in multiple years) were used to build the updated model. The use of resampled sites can be problematic. For example, a single group could be formed from repeated samples for a single site. Most of the habitat data for that group would have no variation – all values would be the same for many of the habitat variables. This lack of variation affects the 'degrees of freedom' for evaluating the habitat model. A strategy for dealing with repeated

measures for the same site should be developed to avoid this 'degrees of freedom' issue at the model-building stage.

3. The approach of using the DFA model to predict group membership for reference sites followed by a BEAST assessment provides an estimate of the false positive, Type I error, associated with incorrectly evaluating a reference site as impaired. Methods for evaluating Type II errors (i.e., the false negative result of erroneously judging an impaired site as unimpaired) need further consideration. The current approach modifies the benthos of a subset of reference sites using a 'simpact' strategy and evaluates the simpacted sites for impairment, but the habitat model is not involved. Alternative approaches such as re-shuffling habitat and benthos sub-matrices would create manipulated reference sites with mis-matching benthos and habitat data. Classifying these sites with the habitat model should lead to a BEAST assessment that highlights the mismatch. Failure to detect the mismatch could be used to provide better estimates of the Type II, false negative, error associated with the DFA model.

4. REFERENCES

Bailey, R.C., G. Scrimgeour, D. Cote, D. Kehler, S. Linke and Y. Cao. 2012. Bioassessment of stream ecosystems enduring a decade of simulated degradation: lessons for the real world. Can. J. Fish. Aquat. Sci. 69: 784-796.

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(4): 1204-1211.

Clarke, K.R. 1993. Non-parametric multivariate analyses of changes in community structure. Australian Journal of Ecology 18: 117–143.

Clarke, K.R., P.J. Somerfield and R.N. Gorley. 2008. Testing of null hypothesis in exploratory community analyses: similarity profiles and biota-environment linkage. J. Exp. Mar. Biol. Ecol. 366: 56-69.

Downie, A.J. 2011. Using impact simulations to evaluate the power of Skeena reference condition approach stream bioassessments. MSc Thesis, Royal Roads University, Victoria, BC.

Environment Canada. 2012. Canadian Aquatic Biomonitoring Network (CABIN) Field Manual for wadeable streams. 49pp.

Environment Canada. 2014. CABIN Processing, Taxonomy, and Quality Control of Benthic Macroinvertebrate Samples. May 2014.

Fox, J. and others. 2018. Package 'car', Companion to Applied Regression; Functions to accompany J. Fox and S. Weisberg. An R Companion to Applied Regression. 3rd Edn., Sage Publ. Ver. 3.0-0.

Gaber, L. 2012. A predictive model for bioassessment of streams in coastal British Columbia using the reference condition approach: 2011. Report prepared by Water Protection and Sustainability Branch, B.C. Ministry of Environment, Victoria, B.C. 28p plus appendices.

Hammer, Ø., D.A.T. Harper and P.D. Ryan. 2001. PAST: Paleontological Statistics Software Package for Education and Data Analysis. Palaeontologia Electronica 4(1): 9pp.

Kuhn, M. and others. 2020. Package 'caret', classification and regression training. Miscellaneous functions for training and plotting classification and regression models. Ver. 6.0-86.

Norris, S. 2012. British Columbia's Provincial Stream Biomonitoring Program. Technical Documentation. GIS Tools for Reference Site Selection and Upstream Watershed Analysis. Prepared for B.C. Ministry of Environment. March 2012.

Novodvorsky, N. and J.L. Bailey. 2016. Reference Model Supporting Documentation for CABIN Analytical Tools - Near North Ontario 2016. Ghost Consulting Report.

Oksanen, J. and others. 2019. Package 'vegan', Community Ecology Package. Ordination methods, diversity analysis and other functions for community and vegetation ecologists. Ver. 2.5-6.

Pokotylo, O., P. Mozharovskyi, R. Dyckerhoff and S. Nagy. 2020. Package 'ddalpha', Depth-based classification and calculation of data depth. Ver. 1.3.11.

R Core Team. 2016. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

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. and J.L. Bailey. 2015. Reference Model Supporting Documentation for CABIN Analytical Tools - BC Central/North Coast 2015. Ghost Consulting Report.

Reynoldson, T.B. and J.F. Wright. 2000. Chapter 20 The reference condition: problems and solutions. Pages 293-304 in D.W. Sutcliffe, M.T. Furse and J.F. Wright, editors. Assessing the biological quality of fresh waters: RIVAPCS and other techniques. Freshwater Biological Association, Ambleside, Cumbria, UK.

Ripley, B., W. Venables, D.M. Bates, K. Hornik, A. Gebhardt and D. Firth. 2019. Package 'MASS', Support functions and datasets for Venables and Ripley's Modern Applied Statistics with S. Ver. 7.3-51.4.

Roever, C., N. Raabe, K. Luebke, U. Ligges, G. Szepannek, M. Zentgraf and D. Meyer. 2020. Package 'klaR', Classification and visualization. Ver. 0.6-15.

Somers, K.M. 2021. Updated CABIN Model for Vancouver Island. Report submitted to British Columbia Ministry of Environment and Climate Change Strategy by Kilgour & Associates Ltd. 122 pp.

Strachan, S.A. 2020. Resource development tolerance scores for BC (developed based on CABIN database extraction Spring 2020). Unpublished Draft Technical Report. 3 pp.

Strachan, S.A., M. Edwards, T.R. 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., 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: 1225-1238.

Strachan, S.A., and S. Pappas. 2016. Reference Model Supporting Documentation for CABIN Analytical Tools: Northeast BC Preliminary Model 2016. Retrieved on Sept 16, 2020 from https://cabin-rcba.ec.gc.ca/CABIN/Beast/ModelDoc.aspx

Whitaker, D. and M. Christman. 2014. Package 'clustsig', Significant cluster analysis. Ver. 1.1.

Wickham, H., R. Francois, L. Henry and M. Kirill. 2020. Package 'dplyr', A grammar of data manipulation. Ver. 1.0.2.

APPENDIX: DATA COLLECTION, ANALYSIS AND QUALITY ASSURANCE

1. Field Collection

CABIN Study Name	EC-Vancouver Island	EC-Pacific Rim/Western Vancouver Island	BC MOE-Vancouver Island Region
Agencies	Environment and Climate Change Canada	Environment and Climate Change Canada	BC Ministry of Environment and Climate Change Strategy
Date range	2001	2008 - 2009	2006 - 2019
Sampling season	Late August/ Sept	Late August/Sept	Late August/early Oct
# of reference samples	21	8	96
Certified samplers (Y or N)	Y	Y	Y
Certified team leader (Y or N)	Y	Y	Y
400 um kick net (Y or N)	Y	Y	Y
Preservative	Formalin	Formalin	Ethanol/Formalin

2. Macroinvertebrate Identification

CABIN Study Name	EC-Vancouver Island	EC-Pacific Rim/Western Vancouver Island	BC MOE-Vancouver Island Region
Taxonomist	Environment Canada	Cordillera Consulting	Cordillera Consulting
Marchant Box used (Y or N)	Y	Y	Y
Subsample count	300	300	300
10% of reference samples sent to National Lab for QA	Y	Y	N
Reference Collection maintained	Y	Y	N

3. GIS Analyses

GIS analyses were done by Chris Steeves, GeoSpatial Services, Thompson Okanagan Region, B.C. Ministry of Forests, Lands, Natural Resource Operations and Rural Development. Watersheds were delineated using ArcGIS 10. 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. The delineated catchments were described using the GIS layers in the table below collected from publicly available sources.

Description	Scale/ Resolution	Source
Basin Morphometry	20 m	Area (km^2) and perimeter (km) were calculated from delineated catchments, as described above
Hydrology	1:50,000	www.geobase.ca – National Hydro Network Intersected with catchment boundaries using intersect function in ArcGIS (Variables: stream order based on 1:50,000, stream length in m)
Bedrock	1:50,000 to 1:250,000	BC Ministry of Energy and Mines – BC Digital Geology Maps 2005 - http://www.empr.gov.bc.ca/Mining/Geoscience/PublicationsCatalogue/Digital GeologyMaps/Pages/default.aspx Intersected with catchment boundaries using intersect function in ArcGIS (ESRI 2010) (Variables: Intrusive, Volcanic, Metamorphic, Sedimentary, Ultramafic, Alluvium as % of upstream watershed area).
Geology	1:5,000,000	Geoscape Canada - A Map of Canada's Earth Materials - Surficial and bedrock geology. http://geogratis.gc.ca/api/en/nrcan-rncan/ess-sst/9636bf0e-aba3-59c3-9736-1ac66bab4ac0.html?pk_campaign=recentItem Using the ArcGIS 10.1 intersect function, all vector layers were intersected with the delineated upstream basins to derive attributes within each catchment.
Climate	560 m	Natural Resources Canada (contact: Dan McKenney – dan.mckenney@nrcan-rncan.gc.ca) . Summarized using rasterized grids describing temperatures normal from 1971-2001 giving long term monthly and annual averages of temperature and precipitations. 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 (Variables: Min & max temp for each month, precip for each month, annual precip, annual min, max and mean temp).
Topography	20 m	www.geobase.ca – Digital Elevation Data Described using 20 m DEM and the Geospatial Modeling Environment v. 0.6.0.0 (Beyer 2012) to describe the maximum and minimum elevation in each catchment. Percent slope was generated from the DEM using the slope function in ArcGIS (ESRI 2010) and classified into one of four groups based on the slope value for each grid cell (i.e. 60%) (Variables: Areas of each class within each catchment; Elevation min, max, mean; and Slope min, max, mean).
Land Use	30 m	www.geobase.ca – Land Cover Intersected with catchment boundaries using intersect function in ArcGIS (ESRI 2010) (Variables: all national landcover variables as % of upstream watershed area).

4. Laboratory Analyses

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

5. Statistical Analyses

Several software packages were used in the development of the model.

1. Excel - data manipulation, data editing and calculation of simple summary statistics

2. PAST: Paleontological Statistics Software Package for Education and Data Analysis (Hammer et al., 2001) - ordination for selecting subsets of climate variables

3. R programming language (R Core Team, 2016) and associated libraries – data manipulation: 'dplyr' – Wickham et al., 2020; clustering classification: 'clustsig' - Whitaker and Christman, 2014; 'vegan' – Oksanen et al., 2019; ordination: 'vegan' – Oksanen et al., 2019; discriminant function analysis: 'caret' – Kuhn et al., 2020; 'klaR' – Roever et al., 2020; 'MASS' – Ripley et al., 2019; confidence ellipses and associated tests: 'car' – Fox et al., 2018; 'ddalpha' – Pokotylo et al., 2020.

4. SYSTAT 13 – confirmation of discriminant analysis results

The updated Vancouver Island model technical report and associated CABIN model was reviewed and approved by the CABIN Science Team: June 2021.