Reference Model Supporting Documentation for CABIN Analytical Tools

MODEL NAME: Okanagan 2017

AUTHORS: Stephanie Strachan and Shauna Bennett

AFFILIATION: Environment and Climate Change Canada, Limnotek Research and Development Inc.

DATE: February 2018

CONTACT(S): Stephanie Strachan 604-664-4099, stephanie.strachan@canada.ca

1. BACKGROUND

A preliminary predictive bioassessment model for the Okanagan and Columbia basins was developed based on data collected from 2006-2010 (Gaber 2012). The preliminary model demonstrated limited sensitivity to detect human disturbance in the Okanagan. An effort was made to fill data gaps to improve the sensitivity to detect disturbance in the Okanagan. Land use in the Okanagan is different than the Columbia which also warrants a separate model given that reference sites criteria may also be different in this area). The Okanagan is dominated by orchards, vineyards, rangeland, pasture, logging, urban and recreational development while the Columbia is primarily influenced by dams/reservoirs, industry, forestry as well as agriculture, urban and recreational development. Watershed concerns in the Okanagan include nutrient enrichment, erosion, sedimentation, stormwater runoff, invasive species and water quantity.

The model developed for the revised Okanagan model described here follows the standard model building procedures for CABIN as described in the online CABIN training modules (Canadian Rivers Institute http://canadianriversinstitute.com/training/cabin/) and originally described in Rosenberg et al. (1999).

2. STUDY DESIGN AND SITE SELECTION

2.1 Model Purpose

The intention of the revised Okanagan model is to enable users of CABIN in the Okanagan to compare test site data to a reference model for which the characteristics of Okanagan biological communities and habitat types have been specifically modeled. This will allow CABIN users in the Okanagan to evaluate stream health and will support informed decision-making and management actions. Results generated from this tool by CABIN users will expand the science and knowledge of the health, productivity and threats to aquatic ecosystems in the Okanagan Basin.

To update the preliminary model and improve the sensitivity of bioassessment in the Okanagan, additional reference samples were collected to fill data gaps and improve the characterization of natural variation. Samples were collected by Environment and Climate Change Canada (ECCC) and British Columbia Ministry of Environment (ENV) between 2010 and 2016. The spatial and temporal scope (see Section 2.2) was defined followed by a process to objectively select additional reference sites (see Section 2.3).

2.2 Spatial and Temporal Scope

The Okanagan model includes three major watersheds: Similkameen, Okanagan, and Kettle watersheds. It also includes the small Vernon Diversion drainage area. From west to east, this area extends from Manning Provincial Park to Granby and Gladstone Provincial Parks and from north to south it extends from Armstrong to Osoyoos. This area includes five ecoregions: Interior Transition Ranges, Okanagan Highlands, Okanagan Range, Selkirk-Bitterroot Foothills, and the Thompson-Okanagan Plateau. The Thompson-Okanagan Plateau comprises the largest area in this region. A range of stream orders was sampled to ensure a variety of stream sizes was included. Refer to Table 1 for the distribution of stream order and ecoregion.

Table 1. Number of reference sites sampled by stream order and ecoregion.

	Stream Order					
Ecoregion	**1**	**2**	**3**	**4**	**5**	**Total**
Interior Transition Ranges		1	1			2
Okanagan Highlands			1			1
Okanagan Range		3	16	13	1	33
Selkirk-Bitterroot Foothills			6	1		7
Thompson-Okanagan Plateau	18	32	19	7	2	78
Total	**18**	**36**	**43**	**21**	**3**	**121**

All data were collected in the late summer and early fall (late-August to mid-October) by CABIN certified field technicians. The complete reference dataset includes a total of 121 samples from 69 unique sites from the Silmilkameen, Okanagan and Kettle watersheds. Refer to Table 2 for the distribution of samples by watershed.

Table 2. Number of reference sites samples collected each year by watershed. Unique sites are indicated in parentheses.

	Watershed			
Year	**Kettle**	**Okanagan**	**Similkameen**	**Total**
2005		4 (4)		**4 (4)**
2006	1 (1)	14 (13)	3 (3)	**18 (17)**
2007	4 (4)	4 (1)	7 (4)	**15 (9)**
2008		2 (0)		**2 (0)**
2009	3 (0)	5 (2)	1 (0)	**9 (2)**
2010		5 (2)	2 (0)	**7 (2)**
2011		5 (2)		**5 (2)**
2012	2 (0)	4 (2)	1 (0)	**7 (2)**
2013		2 (0)	5 (1)	**7 (1)**
2014	9 (9)	2 (0)	5 (3)	**16 (12)**
2015		5 (2)	16 (14)	**21 (16)**
2016	2 (0)	5 (0)	3 (2)	**10 (2)**
Total	**21**	**57**	**43**	**121 (69)**

Fifteen sites were sampled in multiple years to capture temporal variation in local climate or hydrology affecting the reference communities resulting in 67 repeated samples collected between 2005 and 2016 by ENV and ECCC. Refer to Table 3 for the list of repeated reference sites. This dataset provides a balance of spatial variation from all three watersheds, over three ecoregions and temporal variation over 12 years.

Table 3. Repeated reference sites sampled between 2005 and 2016 within each ecoregion and basin

Site	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	Total
Okanagan Range (Similkameen basin)													
ASH01		1	1			1				1			**4**
ASH02		1	1		1				1		1		**5**
BUC01			1					1	1	1	1	1	**6**
GDF01			1						1				**2**
Thompson Okanagan Plateau (all 3 basins)													
BX01		1	1	1	1	1	1	1	1	1	1	1	**11**
CHT01		1	1									1	**3**
ELL02	1	1				1						1	**4**
EQU01	1						1						**2**
EQU02		1	1	1	1	1	1	1	1	1	1	1	**11**
KER05		1	1			1			1				**4**
MCL01						1						1	**2**
PEA01		1			1								**2**
WHI02		1					1						**2**
WKET01			1		1								**2**
Selkirk-Bitterroot Foothills (Kettle basin)													
GRA01			1		1			1				1	**4**
GRA02			1		1			1				1	**4**
Total	**2**	**9**	**11**	**2**	**7**	**6**	**4**	**5**	**6**	**4**	**4**	**7**	

2.3 Reference Site Selection

The reference sites included the 2010 preliminary Columbia-Okanagan model were selected by best professional judgement. The data gaps were identified and filled for the revised Okanagan model after applying the Human Activity Gradient (HAG, Yates and Bailey 2010) to the Okanagan, Similkameen and Kettle watersheds and retrospectively confirming the status and suitability of reference sites identified in the preliminary model. The HAG process is described in detail in Yates and Bailey (2010). The following paragraphs describe some of the detail of the HAG approach applied here.

The first step of the HAG required that the model study area was subdivided into microbasins. Microbasins of 3rd order and higher were identified using 1:50K BC Freshwater Atlas delineations from BCMOE's DataBC (https://www2.gov.bc.ca/gov/content/data/about-data-management/databc). These microbasins were grouped into similar landscape types based on natural habitat characteristics such as bedrock geology, landcover and climate. Classification analysis (K means clustering) was performed to identify possible natural landscape groups. Discriminant function analysis (DFA) was used to evaluate the ability to distinguish various landscape groups. More than 400 microbasins were identified in the study area and classified into two to nine natural landscape groups. Evaluation of the natural landscape groups found that five natural landscape groups best describe the landscape types in the Okanagan study area (Figure 1). The natural habitat characteristics of these landscape groups are described in Table 4.

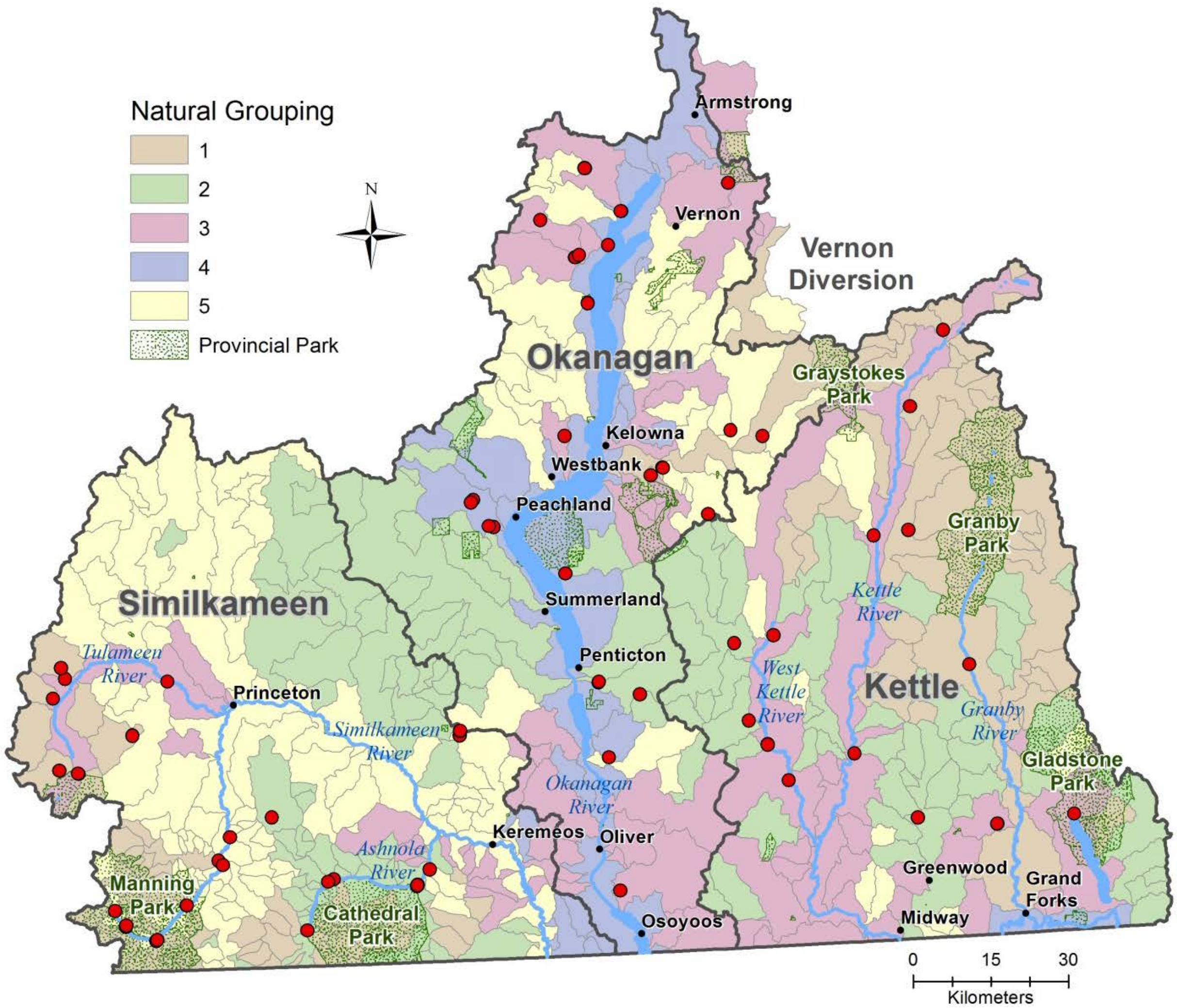


Figure 1. Geographical distribution of reference sites in the Okanagan, Similkameen, and Kettle watersheds distributed across five natural landscape groups.

Table 4. Median of natural habitat and stressor characteristics of micro-basins in each of the five natural landscape groups derived from National Land Cover data (refer to Natural Resources Canada 2009 for stressor definitions).

Natural Landscape Groups	1 (n=109)	2 (n=113)	3 (n=71)	4 (n=24)	5 (n=111)
Natural habitat characteristics					
Basin size	18.0 km^2	37.2 km^2	29.4 km^2	17.0 km^2	26.7 km^2
Low Shrub landcover	2.8%	7.4%	6.9%	7.6%	9.1%
Herb landcover	6.1%	7.7%	11.3%	18.1%	7.7%
Grassland landcover	0%	0%	1.4%	6.2%	0%
Intrusive bedrock	52.5%	94.2%	13.0%	23.5%	4.8%
Metamorphic bedrock	0%	0%	26.0%	5.6%	1.1%
Volcanic bedrock	0%	0.4%	4.1%	8.9%	47.1%
Annual mean precipitation	956 mm	638 mm	525 mm	379 mm	666 mm
Annual mean temperature	1 ^oC	2 ^oC	4 ^oC	7 ^oC	2 ^oC
Stressor characteristics					
Road Density	0 km/km^2	0.003 km/km^2	0.249 km/km^2	0.569 km/km^2	0.0001km/km^2
Developed	0%	0%	0.01%	0.12%	0%
Urban	0%	0%	0%	0%	0%
Agricultural	0%	0%	1.45%	6.08%	0%
Rangeland	0%	0%	6.6%	35.88%	0%
Young Forest	30.87%	56.65%	46.83%	33.88%	44.86%
Recently Logged	0%	10.05%	3.49%	0%	12.32%
Active cut block	0.61%	5.35%	0.94%	0%	7.14%

Reference sites should be distributed among all natural landscape groups and stratified among all stream orders to capture the full range of natural variation. These natural landscape groups were retroactively applied to previous reference sites to examine the distribution of reference sites over the different landscapes and identify data gaps. One natural landscape type, Group 4, was deficient in reference sites. This group was associated with microbasins in populated areas along the Okanagan Lake corridor. Additional data gaps were identified in the Similkameen and Kettle watersheds corresponding to natural landscape groups 1, 2 and 5.

In the second step of the HAG, potential reference microbasins were objectively identified by establishing a stressor gradient for each natural grouping and defining the best available reference sites within each natural grouping (Yates and Bailey 2010). Stressors in the Okanagan study area are described in Table 4 and were described for each natural landscape group by principal components analysis (PCA). Natural landscape groups 3 and 4 tended to be warmer and drier areas and were associated with small communities and agricultural areas. Natural groups 1, 2 and 5 tended to be colder, wetter areas differentiated by bedrock and were associated with logging activity (Table 4). As a result of the different natural features in each group, the land use in the different areas was different (i.e. logging vs agriculture vs rangeland vs development) and was the reason for requiring unique stressor gradients for each group.

The stressor gradient for each natural landscape group was examined individually. The best possible (least disturbed) reference microbasins were selected by identifying the best 25% reference microbasins in each natural landscape group. The reference site criteria for each stressor are different for each group and are provided in Table 5.

The most significant stressor variables for this area that eliminated microbasins from consideration as potential reference sites included forestry related landcover stressors (i.e. active cutblock) and developed or urban landcover. In many cases, the following potential stressor variables were considered to be not significant in the desktop exercise either because information must be verified on site or it was dependent on the date of the GIS data layer: recent burn area, recent logging, exposed land (may not be human induced), recreational (may be protected park), and young forest.

Mining was not a stressor of concern in the Okanagan basin and rangeland could not be avoided using 25% as the threshold for selection of the best reference microbasins (Table 5). Rangeland occurs throughout most of the Okanagan and was not considered a significant stressor variable in the selection of reference microbasins (Table 5). Due to the small size of Group 4 with its unique natural features (i.e. dry landscape, herb landcover), the reference site criterion for rangeland was set relatively high in order to capture the best available microbasins in this landscape type. Similarly, Group 3 landscape type tends to be where communities are centered and the criteria for road density, developed and urban landcover are relatively high in order to capture the best available microbasins of this landscape type (Table 5).

Table5. Criteria for potential reference microbasins in each natural landscape group (NL Gp) for the Okanagan area.

	NL Gp 1	NL Gp 2	NL Gp 3	NL Gp 4	NL Gp 5
Road Density	0 km/km^2	0 km/km^2	0.030 km/km^2	0.168 km/km^2	0 km/km^2
Developed	0%	0.1%	1.5%	2%	0.1%
Urban	0%	0%	2%	0%	0%
Agricultural	0.2%	0.3%	15%	22%	0.5%
Rangeland	7%	0.6%	38%	62%	8.7%
Mining	0%	0%	0%	0%	0%
Recently Logged	13%	32%	13%	0.1%	19%
Active cut block	5.5%	23%	13%	0%	19%

A subset of potential reference microbasins identified through this desktop site selection exercise was verified through field reconnaissance by ENV and ECCC. Final reference site selection was determined by habitat suitability (i.e., presence of a riffle), safe access, and visual confirmation that surrounding land use was not affected by human disturbance.

3. REFERENCE DATA

3.1 Biological description of the Okanagan dataset

The completed dataset of 121 samples collected from 69 unique sites in the Okanagan, Similkameen and Kettle watersheds collected over a 12-year period (section 2.2) were analysed. The dataset was comprised of 65 families of invertebrates to be used for subsequent analyses. Three families were removed from further analysis as they were either not strictly aquatic fauna, or were excluded according to laboratory and sample processing requirements of CABIN (Environment Canada 2014) which included Hydridae, Planariidae, and Sparganophilidae. Refer to Environment Canada 2014 for further details.

Classification of family level reference data was performed using the Bray-Curtis similarity measure and group average linkage in PRIMER 6. The SIMPROF procedure was also used to test for structure in the data. The cluster analysis produced a hierarchical structure with several groups (Figure 2). Some of the small groups were rolled up to the next most similar group or removed as outliers to reduce the number of groups with fewer than 10 sites. The final classification result was four biological groups as indicated in Figure 2 by the symbol and colour along the base of the dendrogram. SIMPER analysis in Primer was performed to identify the taxa which contribute to the similarity within groups (Table 6) and dissimilarity among groups (Table 7). The within group similarity was comparable among all groups with an average and standard deviation of 53.45% ± 1.56%. The average dissimilarity between groups was 61.84% ± 6.63%.

Within the reference dataset, biological communities from 15 reference sites were identified as outliers and removed from further analysis to ensure relatively high within site similarity and to maximize among site dissimilarity. The outliers tended to be among the smallest or the largest streams in this database and their dissimilarity from other reference sites are illustrated in Figure 3. More streams of these sizes should be sampled to further understand the biological communities found in these stream types.

Ordination of the biological data based on the Bray-Curtis similarities showed some overlap of the reference sites within the four groups (Figure 3) and the outliers were distributed on the outer edges of the ordination. The biological description of each group is summarized in Table 8.

106+

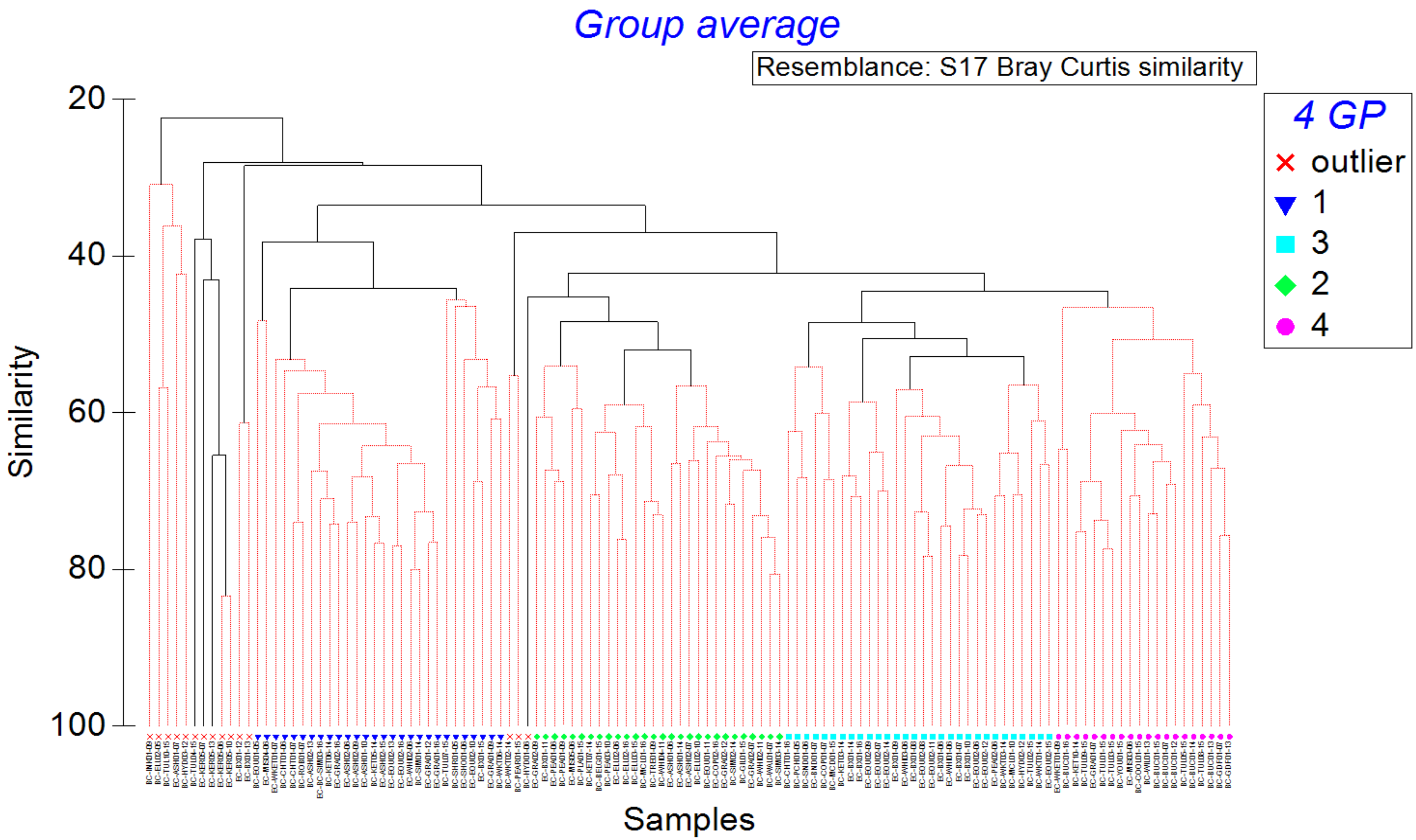


Figure 2. Cluster analysis based on Bray-Curtis similarity and group average linkage followed by a SIMPROF test for structure of 121 reference samples. Red dotted lines indicate no difference between sites connected at the node in the dendrogram. A solid black line indicates a significant difference between sites connected at the node. Final group structure and outliers for the model are indicated by the coloured symbols.

Table 6. Primer SIMPER analysis of biological group similarity and the top 5 contributing taxa.

Group	Species	Av.Abund	Contrib%	Cum.%
Group 1 within group similarity = 51.49%	Baetidae	339.36	39.73	39.73
	Heptageniidae	159.67	17.71	57.44
	Nemouridae	83.14	6.29	63.73
	Chironomidae	84.07	6.16	69.88
	Ephemerellidae	67.61	5.85	75.73
Group 2 within group similarity = 54.17%	Baetidae	1781.56	45.14	45.14
	Heptageniidae	589.73	11.37	56.51
	Nemouridae	551.05	9.09	65.6
	Chironomidae	539.52	8.53	74.13
	Ephemerellidae	396.34	7.14	81.27
Group 3 within group similarity = 53.04%	Baetidae	468.13	19.2	19.2
	Nemouridae	433.33	17.39	36.59
	Heptageniidae	411.61	15.48	52.07
	Ephemerellidae	251.84	8.79	60.86
	Elmidae	227.52	7.37	68.23
Group 4 within group similarity = 55.11%	Chironomidae	997.49	33.6	33.6
	Nemouridae	436.59	14.45	48.05
	Heptageniidae	429.91	13.42	61.48
	Baetidae	289.4	7.23	68.7
	Ephemerellidae	226.35	6.25	74.95

Table 7. Primer SIMPER analysis of biological group dissimilarities and the top 5 contributing taxa.

	Group 1	Group 2	Group 3
Group 2	**70.86%**		
	Baetidae Heptageniidae Nemouridae Chironomidae Ephemerellidae		
Group 3	**59.69%**	**56.23%**	
	Nemouridae Heptageniidae Baetidae Elmidae Ephemerellidae	Baetidae Heptageniidae Nemouridae Chironomidae Elmidae	
Group 4	**69.36%**	**59.29%**	**55.59%**
	Chironomidae Nemouridae Heptageniidae Baetidae Taeniopterygidae	Baetidae Chironomidae Heptageniidae Nemouridae Ephemerellidae	Chironomidae Baetidae Heptageniidae Nemouridae Taeniopterygidae

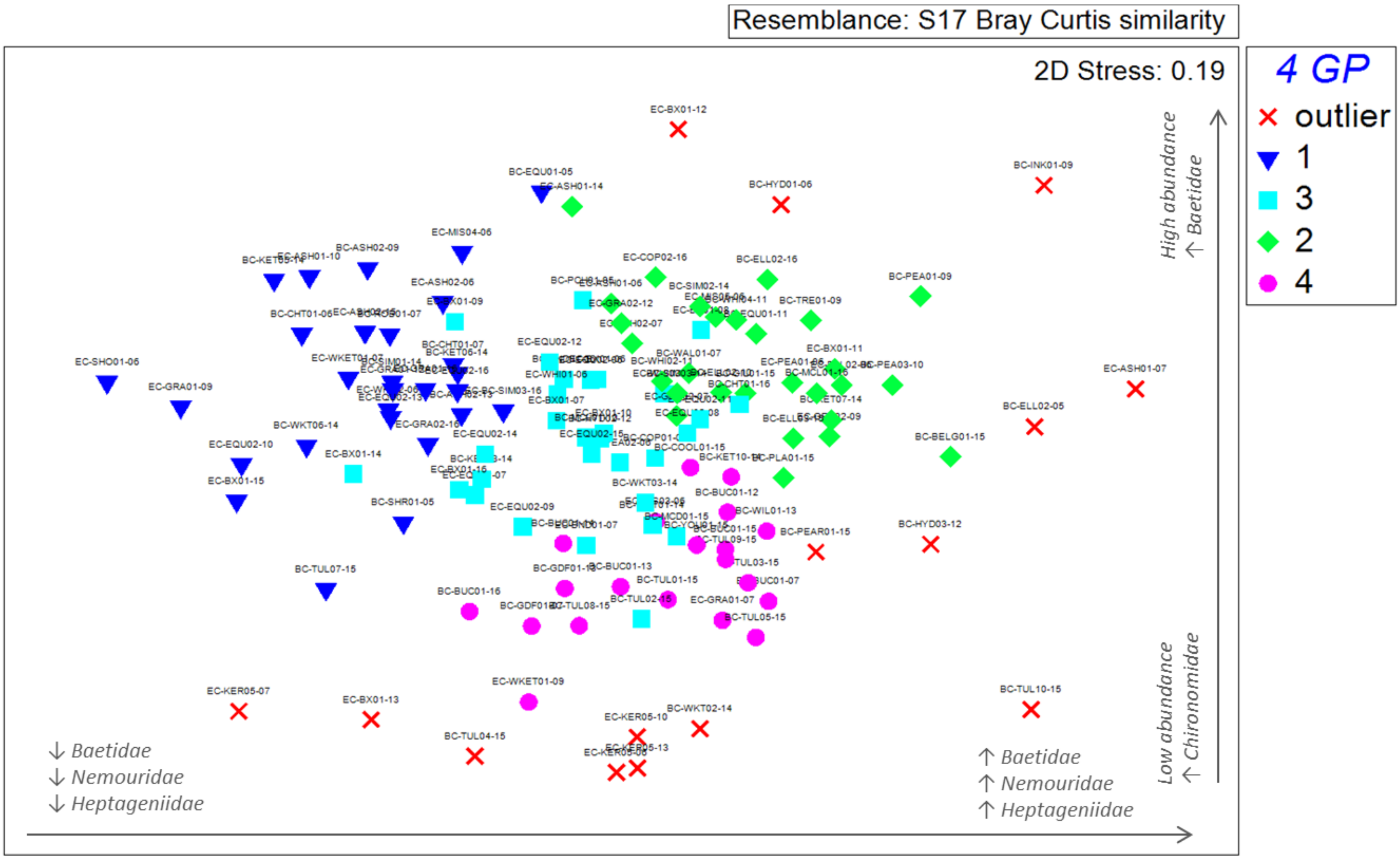


Figure 3. Multidimensional scaling (MDS) ordination plot of Okanagan reference communities (n-=121) based on Bray-Curtis similarity. The group classification of each site is indicated by the coloured symbols. The primary biological drivers for the location of sites in ordination space are indicated by arrows and identified in italics.

Table 8. Summary of biological community descriptions (mean, standard deviation, range) for each of the four reference groups for the complete Okanagan dataset, excluding outliers (n=106).

	Group 1 (n=28)		Group 2 (n=28)		Group 3 (n=30)		Group 4 (n=20)	
	Average ±SD	**Range**	**Average ±SD**	**Range**	**Average ±SD**	**Range**	**Average ±SD**	**Range**
Total Abundance	1253.74 ±525.1	(490-2972)	5789.83 ±2410.78	(1894.05-11400)	3153.81 ±1043.45	(1167.84-5499.74)	3808.67 ±1348.24	(1629.2-5800.01)
Total Richness	23 ±3.5	(17-33)	23.9 ±3.4	(17-31)	24.2 ±3	(15-30)	24.1 ±3.6	(14-30)
EPT* richness	14.8 ±2.2	(11-20)	14.6 ±2	(9-19)	15.3 ±2.5	(9-19)	15.4 ±2	(10-18)
% EPT*	78.3 ±13.3	(39.2-96.5)	78.3 ±7.2	(63.9-93.4)	78.8 ±6.8	(64.4-94.8)	64.6 ±12.9	(33.8-83.8)
# E*	4.7 ±0.4	(4-5)	4.1 ±0.6	(3-5)	4.2 ±0.6	(3-5)	4.6 ±0.5	(3-5)
# P*	4.7 ±1.3	(3-7)	5.4 ±1.5	(2-9)	6.2 ±1.4	(3-8)	6.3 ±1.1	(3-8)
# T*	5.3 ±1.3	(3-8)	5 ±1.1	(3-8)	4.9 ±1.5	(1-7)	4.4 ±1.3	(2-7)
# C*	1 ±0.6	(0-3)	0.8 ±0.5	(0-2)	1 ±0.3	(0-2)	0.3 ±0.6	(0-2)
% E*	50.9 ±16.2	(19.1-80.9)	52.5 ±11.9	(29.5-75.2)	38.2 ±10.6	(21.2-61.1)	26.3 ±9.3	(11-45.3)
% P*	12.3 ±7.7	(2.2-26.8)	15.8 ±8.6	(3.2-34.1)	26.4 ±9.6	(9.2-43)	27.3 ±12.3	(3.7-46.9)
% T*	15.1 ±6.4	(2.6-28.8)	10.3 ±5.6	(2.0-23.6)	14.1 ±6.8	(1.2-30.6)	10.9 ±6.3	(2.6-26.9)
% C*	6.6 ±9.9	(0-45.4)	4.4 ±5.6	(0-21.7)	7.2 ±5.2	(0-21.2)	0.6 ±1.8	(0-7.7)
% Chironomidae	6.5 ±5.4	(0-17.7)	8.8 ±6.1	(0-26.4)	5.3 ±3.5	(0.5-16.5)	27.0 ±12.5	(2.9-51.1)
% Dipteran + other non-insects	15.1 ±7.3	(2.4-28.2)	17.3 ±6.5	(6.3-31.5)	14.0 ±5.8	(4.6-29.7)	34.0 ±10.8	(16.2-54.0)
Simpson's Diversity	6.73 ±2.55	(2.77-12.72)	6.06 ±2.12	(2.4-10.42)	8.69 ±1.83	(5.65-12.17)	7.07 ±1.94	(3.3-10.16)
Simpson's Evenness	0.29 ±0.1	(0.13-0.58)	0.25 ±0.09	(0.09-0.5)	0.36 ±0.07	(0.21-0.51)	0.29 ±0.07	(0.16-0.43)

**E=Ephemeroptera, P=Plecoptera, T=Trichoptera, C=Coleoptera*

3.2 Habitat Description of the Okanagan dataset

The habitat data include location descriptors, catchment morphology/hydrology, topography, reach descriptors, bedrock geology, land cover, long term climate, and water chemistry. The list of potential predictor variables was reduced by excluding variables if they were missing data, were expected to vary with anthropogenic disturbance (i.e. vegetation landcover, water chemistry), were highly correlated with another variable (i.e. climate variables), or had no known relationship with invertebrate communities (i.e. statistical variables such as "Elevation standard deviation in the upstream watershed"). The reduced list consisted of 78 possible predictor variables. This list was further reduced to 26 possible predictors after examining those that explained the most variation in the reference dataset through a principal components analysis (PCA) and those that showed significant relationships with the biological data through previous analyses of the Okanagan data. The final list of potential predictor variables is provided in Table 9 with the mean and standard deviation of each variable by reference group.

Table 9. Mean (±SD) of selected potential predictor variables for the complete Okanagan reference dataset (n=106).

Potential Predictor Variables	Group 1 (n=28)	Group 2 (n=28)	Group 3 (n=30)	Group4 (n=20)
Latitude (dec. degrees)	49.642 +/-0.4614	49.5127 +/-0.4267	49.9725 +/-0.4536	49.2971 +/-0.2576
Longitude (dec. degrees)	-119.5085 +/-0.6778	-119.6999 +/-0.6581	-119.4039 +/-0.4909	-120.3488 +/-0.7971
Altitude (fasl)	2450.6 +/-721.2	2632.9 +/-842.2	2533.2 +/-689.2	3771.3 +/-637.7
Stream order (1:50,000)	2.6 +/-0.9	2.8 +/-1	2.1 +/-0.8	3 +/-0.9
%Intrusive Bedrock	44.5 +/-37.52	47.13 +/-37.88	22.87 +/-34.47	38.53 +/-36.46
%Metamorphic Bedrock	20.02 +/-30.52	13.03 +/-24.8	21.28 +/-30.1	11.19 +/-25.89
%Sedimentary Bedrock	17.13 +/-26.83	15.08 +/-27.97	26.18 +/-26.81	42.47 +/-44.76
Channel Depth, avg (cm)	21.67 +/-11.72	17.18 +/-9.35	11.78 +/-4.89	14.49 +/-7.55
Channel Depth, max (cm)	30.45 +/-15.65	23.98 +/-12.45	16.21 +/-6.6	21.06 +/-10.24
Presence of pools in reach (0/1)	0.78 +/-0.41	0.85 +/-0.35	0.73 +/-0.44	0.85 +/-0.36
Presence of runs in reach (0/1)	0.64 +/-0.48	0.92 +/-0.26	0.56 +/-0.5	0.85 +/-0.36
Channel velocity, avg (m/s)	0.39 +/-0.18	0.35 +/-0.15	0.3 +/-0.14	0.29 +/-0.18
Channel velocity, max (m/s)	0.59 +/-0.26	0.56 +/-0.18	0.45 +/-0.2	0.51 +/-0.27
Channel width, bankfull (m)	12.05 +/-6.62	10.05 +/-6.73	6.76 +/-3.52	9.06 +/-4.62
Channel width, wetted (m)	8.05 +/-5.29	5.72 +/-3.18	3.99 +/-2.06	4.84 +/-3.46
Precipitation, APR 30 yr avg (mm)	65.77 +/-21.57	60.43 +/-20.06	52.98 +/-15.23	84.02 +/-37.56
Precipitation, ANN 30 yr avg (mm)	715.57 +/-183.36	670.69 +/-144.95	602.34 +/-123.43	869.24 +/-158.24
Drainage Area (km^2)	187.03 +/-209.74	151.11 +/-233.46	53.93 +/-45.93	57.56 +/-111.09
Drainage Perimeter (km)	96.08 +/-59.09	80.02 +/-49.77	54.61 +/-26.82	47.96 +/-43.81
Stream Density (m/km^2)	1964.45 +/-683.49	1736.9 +/-604.78	1684.02 +/-314.11	1638.43 +/-227.69
Stream Length (m)	404510.61 +/-449633.4	277903.81 +/-459595.12	95247.84 +/-83967.68	102929.72 +/-217065.08
%Snow/Ice Landcover	0.02 +/-0.05	0.01 +/-0.04	0 +/-0	0 +/-0
%Water Landcover	0.18 +/-0.34	0.64 +/-1.05	0.32 +/-0.77	0 +/-0.02
%Wetland Treed Landcover	0.11 +/-0.18	0.11 +/-0.2	0.15 +/-0.19	0 +/-0.03
%Boulder substrate	11.64 +/-10.62	9.03 +/-7.88	4.1 +/-3.71	8.25 +/-7.58
%Pebble substrate	23.64 +/-12.81	24.6 +/-11.96	34.86 +/-9.85	29.65 +/-7.64
Elevation max	2158.21 +/-306.16	2103.32 +/-289.97	1885.93 +/-196.32	2067.44 +/-156.59
Elevation min	754.78 +/-204.02	809.74 +/-249.02	802.33 +/-225.19	1143.49 +/-178.32
%Slope <30%	57.68 +/-12.63	59.05 +/-18.29	63.25 +/-9.66	59.07 +/-9.82
%Slope, max	208.05 +/-100.3	168.28 +/-42.95	126.91 +/-28.98	155.05 +/-70.69

4. THE MODEL

A stepwise discriminant function analysis (DFA) was performed on various clustering options, - 2 groups, 3 groups or 4 groups. A variety of indicators from the DFA were examined to determine the best clustering option and the best model. All stepwise options with the various clusters were examined to look for the following:

a) High overall classification rate and jackknifed classification rate
b) Balanced individual group rates of the jackknifed classification and consistency in the classification of each group.
c) Better than random prediction of the group classification rate from the jackknifed results.
d) Low Wilks Lambda.
e) Tolerance values of individual predictors greater than 0.20 to avoid multicollinearity.

A 4-group model was determined to provide the best solution with the complete dataset of 106 sites.

In order to test model performance validation data are required to test the prediction success and Type 1 error rates. Using the 4 group classification of the full dataset (n=106), 20% of the reference sites were randomly selected to divide the dataset into a training (n=85) and a validation (n=21) dataset from which the final model was developed. The training data were used to derive the final model. These data were run through a stepwise DFA to further refine the optimal predictor variables. The validation data were used to test the model. These data were treated like test sites to determine if the optimal model correctly predicted the sites' membership in the original classification of the complete dataset. A summary of the division of the reference sites is provided in Table 10.

Table 10. The distribution of reference sites from the complete dataset (n=106, excluding outliers) into training and validation datasets to build and test the optimal model.

# of sites	Complete dataset	Training dataset	Validation dataset
Group 1	28	23	5
Group 2	28	22	6
Group 3	30	23	7
Group 4	20	17	3
Total	**106**	**85**	**21**

4.1 Discriminant Function Analysis Results

The forward and backward stepwise DFA results of the 4-group model provided good classification success. However, the models resulted in numerous variables with low tolerance scores indicating collinearity. The stepwise results were further refined through an iterative process of removing variables with low tolerance and adding other significant variables identified in the reduced list of potential predictors in an attempt to improve or maintain classification success while reducing collinearity. The final model DFA results are provided in Table 11 and 12. Eleven predictor variables resulted in a Wilks Lambda of 0.152, an overall classification rate of 74% and an overall jackknifed classification rate of 65%. The individual group classification rates were on average 2.4 times better than a random prediction (random prediction of 4 group model is 25%) with good consistency for all groups (average rate was 65.5% ± 7.14%). The tolerance scores were all >0.25 suggesting limited collinearity among predictor variables. The classification of the validation data indicates a 90% success rate in predicting reference sites' membership in the correct reference group.

Table 11. Optimal model variables with F scores and tolerance values from the iterative DFA process.

Variable	F score	Tolerance
LATITUDE	4.454	0.462
Bedrock Geology-%Intrusive in u/s watershed	6.506	0.449
Bedrock Geology-%Metamorphic in u/s watershed	1.803	0.691
Channel Depth Avg (cm)	1.595	0.512
Climate-Precip April (30 yr avg) (mm)	8.213	0.522
Landcover-%Water in u/s watershed	15.244	0.293
Landcover-%Wetland Treed in u/s watershed	4.961	0.266
Substrate-Pebble	3.236	0.579
Topography-Elevation in u/s watershed, min (m)	12.905	0.610
Topography-Slope in u/s watershed, max (%)	3.291	0.420
Hydrology-Drainage area (km^2)	5.371	0.326

Table 12. Classification tables from the DFA for the optimal model (n=85) identified in Table 10 including the classification of validation sites (n=21).

	Group 1	Group 2	Group 3	Group 4	%correct
Classification Matrix					
Group 1	**15**	2	5	1	65
Group 2	2	**16**	4	0	73
Group 3	1	1	**18**	3	78
Group 4	1	0	2	**14**	82
Total	19	19	29	18	**74**
Jackknifed Classification Matrix					
Group 1	**14**	2	6	1	61
Group 2	2	**14**	4	2	64
Group 3	2	3	**14**	4	61
Group 4	2	0	2	**13**	76
Total	20	19	26	20	**65**
Classification of Validation Sites					
Group 1	**4**	0	1	0	80
Group 2	0	**6**	0	0	100
Group 3	1	0	**6**	0	86
Group 4	0	0	0	**3**	100
Total	5	6	7	3	**90**

4.2 Model performance

Validation Sites: Type 1 error rate

An additional test of the model using the validation data includes an assessment of the Type 1 error rate. As the validation data were reference sites that were randomly selected to test the model, all validation data should be similar to the reference data used to build the model. CABIN uses a 90% confidence ellipse of the reference sites used in the model to identify test or validation sites that are mildly divergent from the predicted reference condition. By definition, 10% of the validation sites (i.e. 2 sites) should be expected to be outside the 90% ellipse. The assessment of the validation sites resulted in four sites falling outside the 90% ellipse suggesting a Type 1 error rate of 19%. This model would be considered very protective of the environment by these results as these validation sites known to be in good condition as four were found outside of the reference cloud while we expected only two sites to be outside of the reference cloud.

Test Sites: Comparison with 2010 preliminary model results

The model was further assessed by comparing the test site assessments with assessments from the 2010 preliminary Columbia-Okanagan model (Gaber 2012) to see if the detection of divergence from reference was similar (Table 13). Seventeen sites that were assessed with the 2010 preliminary model were also assessed with this model. Two of the 17 sites were found to be in reference condition and 15 sites were found to be out of reference condition (i.e. outside the 90% ellipse). This revised model was able to detect sites different from reference similar to the 2010 preliminary model. However, the revised model tended to find greater difference between the divergent test site and the reference condition than the 2010 model. Eight sites were assessed similarly by the two models while six sites were shown to be more different from reference. Only one site was found to be more similar to reference with this revised model than the 2010 preliminary model.

As assessment of Type 2 error rates can only be done when the extent of the impact on the benthic community is known. The real impact on a benthic community is never known therefore simulated data are needed for this assessment. When a consistent simulation procedure can be recommended as a standard approach for assessing CABIN models by the CABIN Science Team, the Type 2 error rate performance should be evaluated and amended to this template.

Table 13. Summary of test site assessment results for Okanagan area sites using the preliminary 2010 Columbia Okanagan Model (Gaber 2012) and the revised Okanagan 2017 model. Assessment results are indicated as follows: R=reference condition, MD=mildly divergent from reference condition, D=divergent from reference condition, and HD=highly divergent from reference condition.

Site code	Site name	Basin	Year	Latitude	Longitude	2010 Columbia-Okanagan	Okanagan 2017
KET04	Arthurs Creek	Kettle	2014	49.688	-118.704	R	R
SIM03	Similkameen River	Similkameen	2006	49.107	-120.841	MD	MD
COL02	Coldstream Ck 1.8 km us of WSC Station	Okanagan	2009	50.265	-119.093	MD	D
KEL0204	Kelowna Creek @ Bulman Rd	Okanagan	2006	49.938	-119.386	D	D
MIS06	Mission Creek at East Kelowna	Okanagan	2008	49.864	-119.390	R	R
MIS01	Mission Creek u/s Gordon; E268083	Okanagan	2007	49.843	-119.476	MD	D
BX01	BX creek near PV Road	Okanagan	2006	50.277	-119.247	MD	D
OKA01	Okanagan River	Okanagan	2010	49.114	-119.566	MD	MD
SIM01	Similkameen River	Similkameen	2010	49.078	-119.711	MD	D
SIM02	Similkameen River	Similkameen	2010	49.461	-120.504	MD	MD
KET01	Kettle River	Kettle	2010	49.005	-118.776	MD	MD
KEL0104	Kelowna Creek @ Abbott	Okanagan	2004	49.879	-119.489	HD	HD
BX02	BX Creek@30th	Okanagan	2007	50.264	-119.280	D	D
SHU01	Shuttleworth Creek at Willow St.	Okanagan	2010	49.341	-119.573	HD	D
KET02	Kettle River	Kettle	2010	49.020	-118.476	HD	HD
COL01	Coldstream Creek @ Kirkland Drive	Okanagan	2009	50.224	-119.262	MD	D
ELL0104	Ellis Creek @ Mouth 0500027	Okanagan	2005	49.479	-119.595	MD	D

4.3 Final Model Summary

The final model was built on the training dataset (n=85) and is described below. Not surprisingly, the training dataset closely resembles the complete dataset described in section 2.1. Reference Groups 1 and 2 are very similar in diversity and richness (Figure 4) and overall composition (Figure 5) but differ in total abundance. Groups 3 and 4 are similar in family richness, Simpson's diversity and total abundance characteristics but differ in their composition of Coleoptera, Diptera and non-insects and generally have fewer Ephemeroptera, Plecoptera, and Trichoptera (EPT) organisms.

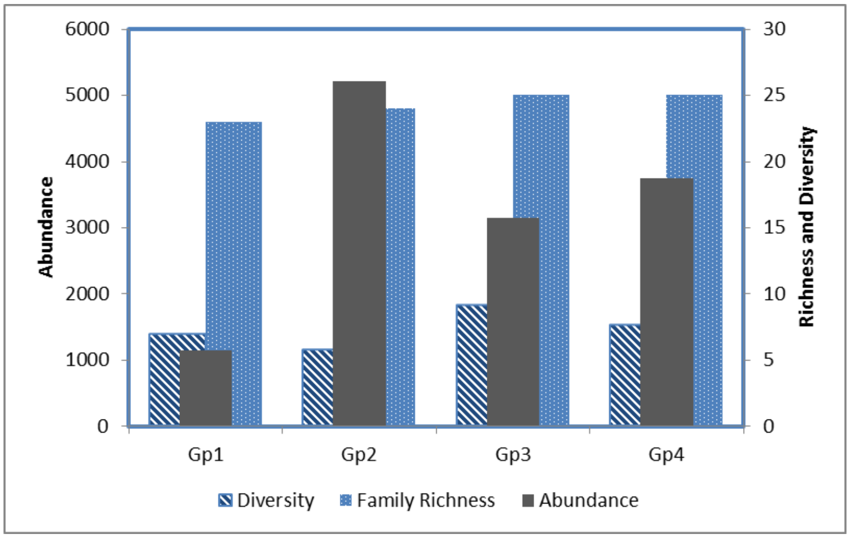


Figure 4. Summary Simpson's diversity, family richness and total abundance for each of the four reference groups from the training dataset (n=85).

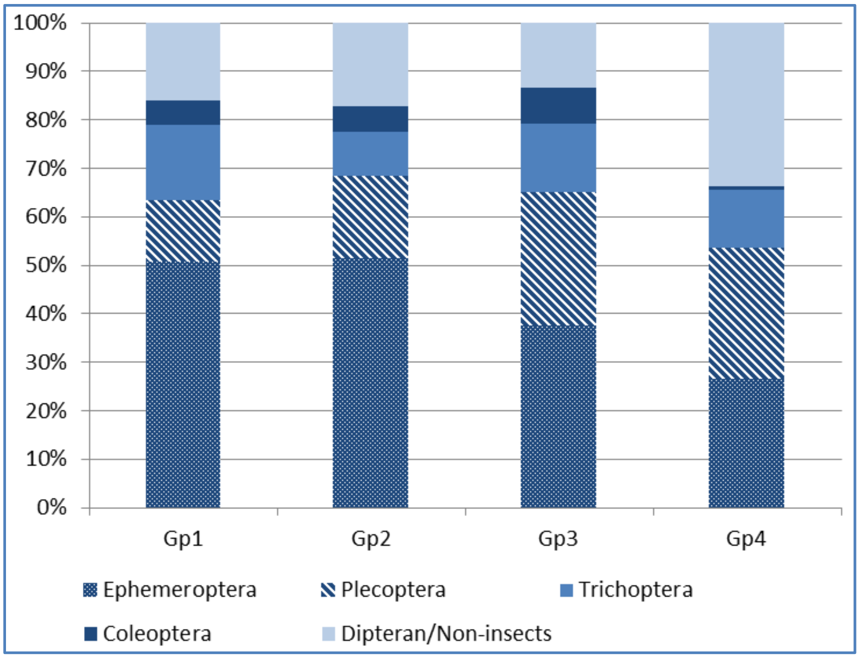
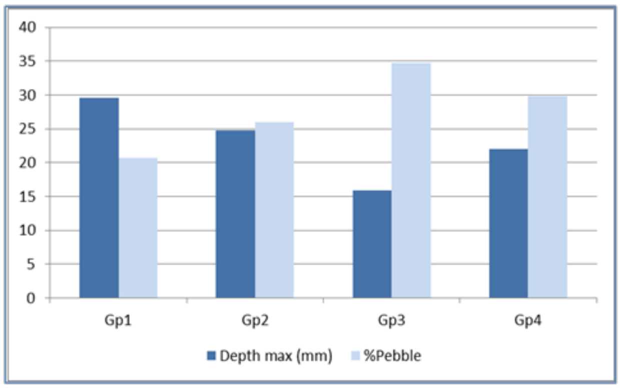
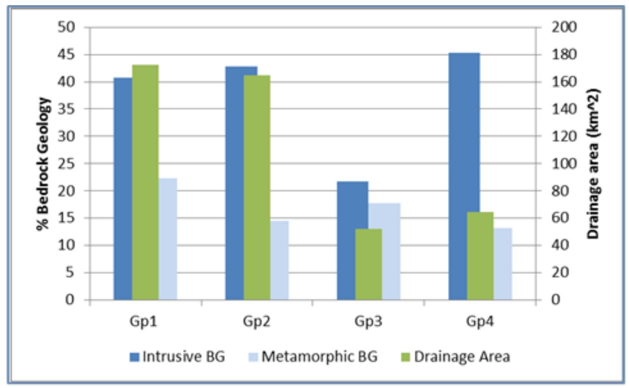


Figure 5. Distribution of taxonomic order level composition for each of the four reference groups from the training dataset (n=85).

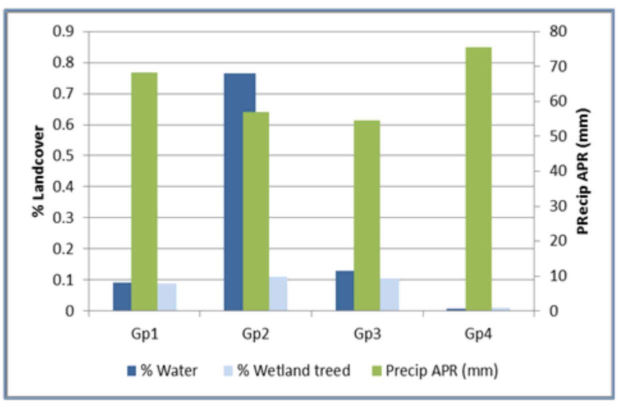
The habitat characteristics of the reference groups are summarized in Figure 6a to Figure 6d. Groups 1 and 2 tend to have large drainage areas. Group 2 is distinguished from Group 1 by the large proportion of landcover consisting of water. Group 3 tend to have the least amount of intrusive and metamorphic bedrock geology while the substrate tends to have a greater proportion of pebble with shallower streams. Group 4 tends to have the highest precipitation in April and tends to be at lower elevations with very little of the watershed landcover having water or treed wetlands.



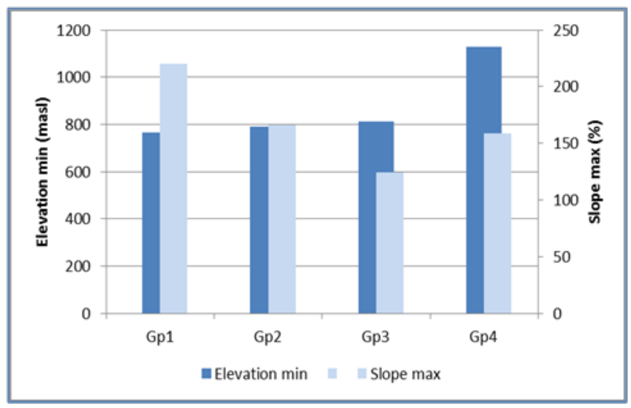
6a



6b



6c



6d

Figure 6. Graphical summary of average predictor variable characteristics of the 4 reference groups from the training dataset (n=85): a) average channel depth (cm) and % pebble substrate, b) % bedrock geology and drainage area, c) % landcover and 30 year average precipitation in April, and d) minimum elevation in the upstream watershed and maximum slope in the upstream watershed.

5. RECOMMENDATIONS

The results of this model were very good with classification rates of other models currently uploaded available in the CABIN analytical tools. The Type 1 error rate was very good and the comparison to the preliminary model indicated that this revised model was able to detect greater difference from the reference condition for test sites previously assessed suggesting increased sensitivity.

In the future as more data are collected and the mode is revised or updated, finer resolution climate variables (more recent nationally consistent records, as opposed to 30 year climate averages) would provide more meaningful relationships between the benthic community and the local environment. Changing weather patterns as a result of climate change will be particularly important in this normally dry area. Also, substrate was a really important predictor for this model. While it is possible that pebble substrates may be affected by erosion or sedimentation from human disturbances, it was assumed that most of the influences would occur from substrates smaller than pebbles. This variable demonstrated significant discriminatory power for the final model, without which the classification success would have been much less.

Evaluation of Type 2 error rates would provide more information about the performance of the model and its ability to detect known differences from the expected condition. CABIN Science Team guidance is needed to establish an appropriate means of evaluating Type 2 error rates with simulated data.

6. OTHER RELEVANT LITERATURE

Environment Canada. 2014. CABIN Laboratory Methods: Processing, Taxonomy and Quality Control of Benthic Macroinvertebrate Samples. May 2014. 36 p. http://publications.gc.ca/site/eng/476513/publication.html

Gaber, L. 2012. A predictive model for bioassessment of streams in the Columbia-Okanagan area of British Columbia using the reference condition approach: 2010. Report prepared by Water Protection and Sustainability Branch, B.C. Ministry of Environment, Victoria, B.C. 32p. plus appendices https://www2.gov.bc.ca/assets/gov/environment/air-land-water/water/water-quality/monitoring-water-quality/biomonitoring/wq_bio_columbia_okanagan_rca_2012.pdf

Natural Resources Canada, 2009. GeoBase – Landcover, circa 2000 http://geogratis.gc.ca/api/en/nrcan-rncan/ess-sst/643c4911-475b-4765-b730-2dde9be50d5b.html)

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. http://publications.gc.ca/site/eng/464069/publication.html

Yates, A. and R.C. Bailey. 2010. Selecting objectively defined reference sites for stream bioassessment programs. Environmental Monitoring and Assessment 170:129-140.

APPENDIX: DATA COLLECTION, ANALYSIS AND QUALITY ASSURANCE

1. Field Collection

CABIN Study Name	BC MOE-Okanagan Kicknet	EC-Okanagan Basin
Agencies involved	BC Ministry of Environment	Environment and Climate Change Canada
Date range	2005-2016	2006-2016
Sampling season	Late Aug-Mid Oct	Late Aug-Mid Oct
# reference samples	65	56
Certified samplers (Y or N)	Y	Y
Certified team leader (Y or N)	Y	Y
400 um kicknet (Y or N)	Y	Y
Other sampling device	--	--
Field Preservative	Ethanol	Formalin

2. Macroinvertebrate Identification

CABIN Study Name	BC MOE-Okanagan Kicknet	EC-Okanagan Basin
Taxonomist	Fraser Environmental/ Cordillera Consulting	Cordillera Consulting/ EcoAnlaysts
Marchant Box used	N	Y
Other subsampling device	Caton Tray	--
Subsample count	300	300
10% of reference samples sent to National Lab for QA	N	Y
Reference Collection maintained	N	Y

3. GIS analyses

GIS analyses for all studies were generated by Adam Yates, Department of Geography, University of Western Ontario. Watersheds were delineated using ArcGIS 10 ArcHdyro 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 (AcrHydro 2010). The delineated catchments were described using the GIS layers in the table below collected from publicly available sources

Description of data layers and sources:

Descriptor	Scale/ Resolution	Source and method
Basin Morphometry	20 m	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	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.
Hydrology	1:50,000	www.geobase.ca – National Hydro Network Intersected with catchment boundaries using intersect function in ArcGIS (ESRI 2010)
Land Use	1:2,000,000	www.geobase.ca – Land Cover Intersected with catchment boundaries using intersect function in ArcGIS (ESRI 2010)

4. Statistical Analyses

Statistical programs used:

- Excel – data manipulation and storage
- PRIMER 6 – clustering, SIMPROF, MDS ordination, ANOSIM, SIMPER for biological data
- SYSTAT 11 – discriminant analysis and test site analysis (plotting BEAST assessments with probability ellipses)

Model was reviewed by the CABIN Science Team February 2018. Comments provided and recommendations applied and final copy submitted February 2018.