

Reference Model Supporting Documentation for CABIN Analytical Tools

MODEL NAME: BC Central/North Coast 2015
AUTHORS: Trefor Reynoldson and John Bailey
AFFILIATION: GHOST Environmental Consulting
DATE: July 2015
CONTACT(S): John Bailey (705) 671-3860, ibailey@laurentian.ca;
Trefor Reynoldson (902) 538-3389 trefor.reynoldson@gmail.com

1. STUDY DESIGN AND SITE SELECTION

Reference Condition Approach (RCA) models and stream bioassessments based on benthic invertebrate communities were first developed for the Skeena River Basin in 2007 and most recently in 2010 based on 145 sites (Skeena BEAST09, Bennett 2010). Subsequent sampling up to 2013 has expanded the potential reference site database to 273 potential reference sites and accordingly it was appropriate to consider a re-build of the Skeena model for the BC Central/North Coast. Furthermore, GIS collected predictor data using national data layers were also acquired to provide a more consistent and comparable set of potential habitat predictors.

Table 1. List of different studies and sub-basins included in data set for model development.

	CABIN study name	Reference and potential reference Sites (sample period)	Basins (no. sites/basin)
1	BC MOE-FSP Skeena Region	138 (2004-2013)	Clarence Strait (1) Portland Canal (2) Stikine (15) Hecate Strait (37) Nass (13) Queen Charlotte Sound (3) Skeena (66) Taku (1)
2	BC-Skeena BC Timber Sales	22 (2007-2008)	Hecate Strait (6) Nass (7) Skeena (9)
3	BC-0887- Rescan	9 (2011-2013)	Stikine (9)
4	National Parks-Gwaii Haanas	64 (2006-2012)	Moresby Island (64)
5	BC-Kitsault - AMEC/AZIMUTH	15 (2010-2013)	Portland Canal (15)
6	BC-Long Lake IPP -Triton	4 (2011-2012)	Portland Canal (4)
7	BC-Dome Mtn Baseline	1 (2009)	Skeena (1)
8	Aquatic Studies – Rescan (BC-IM Red Chris Baseline-Golder	7 (2012-2013)	Stikine (7)
9	BC-123210182-Stantec	11 (2012)	Stikine (11)
10	BC-Eskay Creek Mine EEM – Golder	2 (2010)	Clarence Strait (2)



The distribution of reference sites considered for the model is shown by year and watershed in Table 1 and Figure 1. Data for the habitat and invertebrate assemblage were collated from 10 studies (Table 1) from reference sites for the period 2004-2013. Nine different sub-basins are included in the data set of 273 possible reference sites. It is this data set that was used for building the revised model.

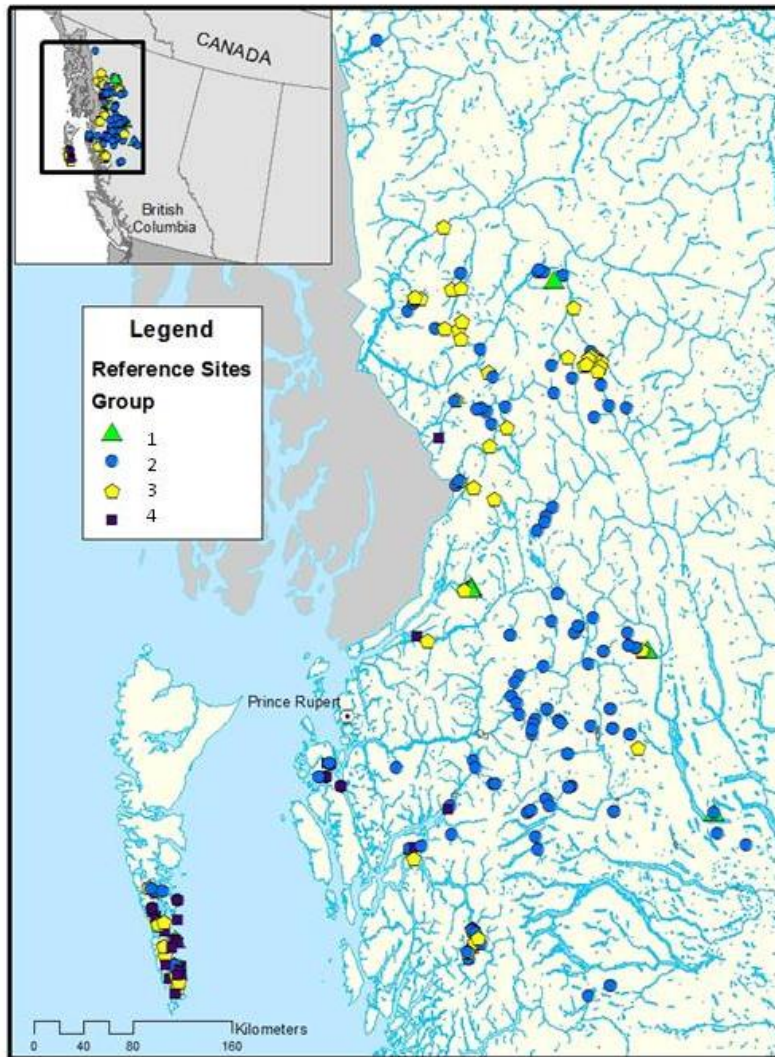


Figure 1. Geographical distribution of reference sites used in the BC Central/North Coast 2015 model.

2.0 REFERENCE DATA AND FINAL MODEL

2.1 Biological description

A total of 93 families were recorded from the 273 reference sites. Four families were removed from further analysis: Daphniidae, Oreoleptidae, Hydridae and Planariidae, as they are either meiofauna or do not meet the verification requirements of CABIN. The remaining 89 families were used to classify the sites into reference groups, where the groups represent the different invertebrate assemblages naturally occurring in the study area.

Thirteen families represent more than 90% of the organisms found at the reference sites (Table 2) and occur at between 18-98% of the reference sites. All four of the most abundant taxa are found at more than 90% of the reference sites.

Table 2. Summary of the abundant and frequently occurring families at 273 reference sites.

Phylum/Order	Family	% Total count	Cumulative total (%)	% Occurrence
Diptera	Chironomidae	18.53	18.53	98.17
Ephemeroptera	Heptageniidae	18.06	36.60	95.60
Ephemeroptera	Baetidae	14.35	50.94	97.80
Plecoptera	Nemouridae	11.82	62.76	93.04
Plecoptera	Taeniopterygidae	6.49	69.25	52.38
Mollusca	Pisidiidae	4.92	74.17	18.32
Ephemeroptera	Ephemerellidae	4.70	78.87	83.88
Plecoptera	Chloroperlidae	3.50	82.38	93.04
Trichoptera	Rhyacophilidae	2.01	84.38	84.98
Ephemeroptera	Leptophlebiidae	1.92	86.30	46.52
Diptera	Tipulidae	1.13	87.43	64.84
Plecoptera	Capniidae	0.96	88.39	47.62
Diptera	Simuliidae	0.94	89.33	53.48
Trichoptera	Hydropsychidae	0.84	90.17	45.05

Classification of family level data (Figure. 2) sites produced 4 distinct reference groups from 273 reference sites. One large group of 146 sites, two groups of 50-60 sites, one of which (Gp 4) largely occurs on Haida Gwaii, and one small group of 14 widely distributed sites (Table 3).

Table 3. Number of sites within each reference group.

GROUP Number	1	2	3	4
# of Reference Sites	14	146	59	54

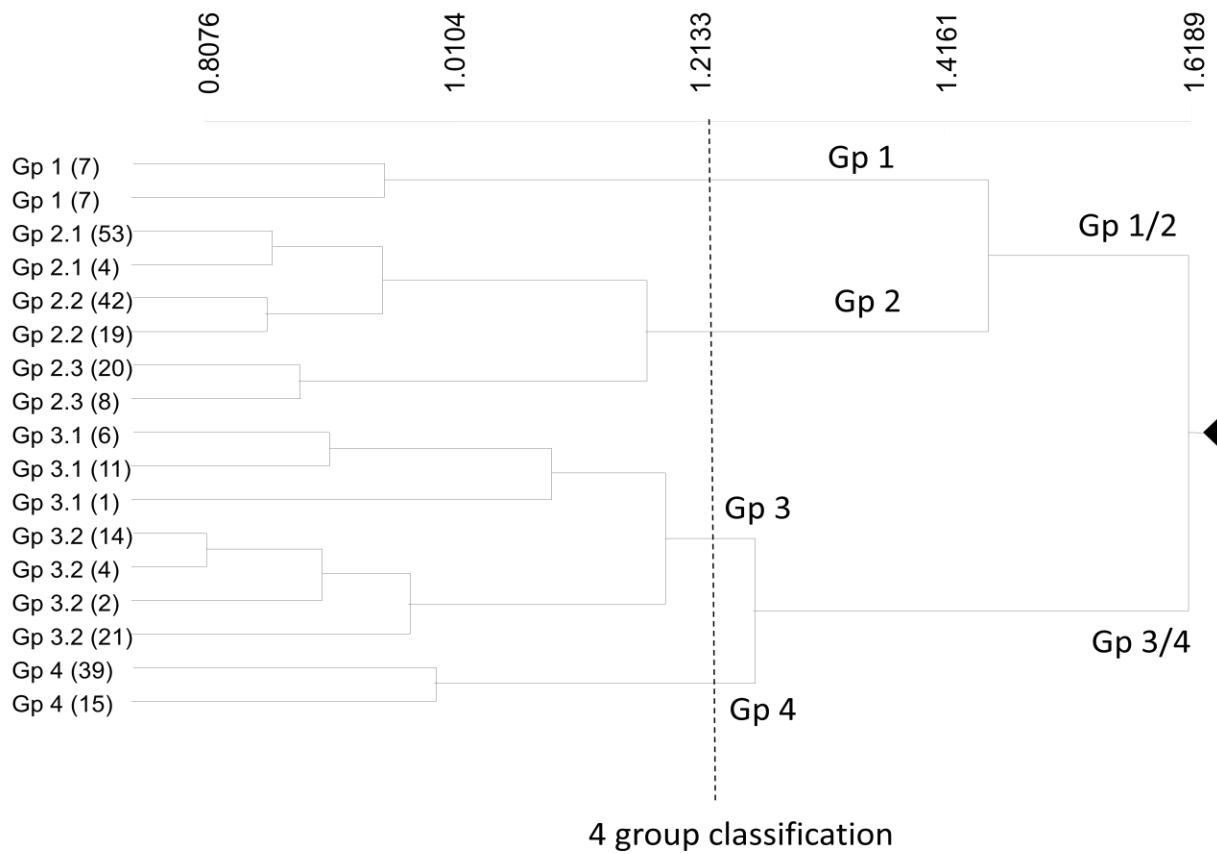


Figure. 2. Classification of 273 reference sites showing 15 branches of the dendrogram the four group classification and the seven groups examined (the number of sites in each branch in parentheses).

Table 4. Biological characteristics of four reference assemblages described for the BC Central/North Coast Basin. Habitat variables used as predictors in bold. Families contributing most to similarity within a group in bold.

Variable	Group 1	Group 2	Group 3	Group 4
Abundance	11065 (7171)	2331 (1541)	324 (274)	762 (762)
Richness	17 (3)	18 (4)	15 (4)	19 (4)
Chironomidae	2693 (3075)	404 (476)	48 (49)	64 (65)
Nemouridae	1961 (2056)	230 (266)	38 (64)	45 (45)
Pisidiidae	1901(2324)	3 (20)	0 (0)	5 (15)
Leptophlebiidae	460 (453)	23 (72)	1 (4)	14 (25)
Baetidae	1367 (3086)	268 (355)	29 (49)	364 (479)
Heptageniidae	841 (1876)	558 (615)	82 (93)	41 (38)
Ephemerellidae	133 (194)	158 (240)	8 (13)	12 (23)
Chloroperlidae	258 (709)	92 (103)	14 (15)	30 (44)
Taeniopterygidae	270 (774)	211 (556)	23 (55)	2 (8)
Rhyacophilidae	56 (54)	66 (88)	8 (9)	6 (16)
Tipulidae	212 (745)	14 (25)	4 (7)	20 (22)
Sperchontidae	112 (232)	15 (29)	2 (4)	10 (13)

2.2 Habitat Description

A total of 84 variables were available from 273 sites as potential predictors in the model. These include 6 categories of variables, almost entirely obtained through GIS. A further 18 variables were possible candidates but were missing from between 8 and 64 sites, after analysis these variables did not contribute substantially more information and were excluded from final analysis.

Several approaches were used to select the final habitat predictors:

1. Forward and Backward Stepwise DFA and then optimization by variable removal and entry based on individual F scores.
2. Variable selection based on similarity matrix matching using BVSTEP in PRIMER with raw and normalized habitat data.
3. Optimal model from the above approaches and adding variables that best predicted individual reference groups in DFA.
4. Backward and Forward Stepwise DFA but adjusting the P value for entry and removal of variables.

Selection of a final model was based on a number of factors:

1. The number of groups- where a model with more groups of similar size is preferred. Such a model will tend to show less within group variability and therefore have more sensitivity in detecting change from reference.
2. A lower lambda (λ) statistic - as this indicates greater difference among the means values for the predictor variables.
3. A higher F score - as this indicates greater among group to within group variation in the predictor variables.
4. High accuracy - as indicated by the cross validation (CV) % sites correctly assigned.
5. Smaller number of predictor variables - as these models are more robust.

The habitat characteristics of the groups for both the predictor variables and those best correlated with the biological data are summarized in Table 5.



Table 5. Environmental characteristics of four reference assemblages described for the BC Central/North Coast. Habitat variables used as predictors in bold. Families contributing most to similarity within a group in bold.

Variable	Group 1	Group 2	Group 3	Group 4
Latitude	55.331	55.018	55.210	52.827
Longitude	128.891	128.563	129.7	130.971
Altitude	2344	2062	1712	157
Stream order	2.6	3.0	3.0	1.6
Drainage area	18.5	79.1	96.9	11.9
Stream density	1782	2516	2374	2088
Shrub low (%)	27.4	7.7	12.0	6.0
Water (%)	3.2	0.6	0.3	0.9
Degree days	131	132	132	88
Precip Total	1305	1287	1658	2579
Precip. Jan (mm)	148.1	144.4	185.9	294.4
Precip. Feb	106.5	108.5	142.7	229.3
Precip. Mar	82.4	89.2	125.8	223.7
T. min Jan (°C)	-10.2	-10.9	-9.9	-0.1

Model Group Summary

Four families (Chironomidae, Nemouridae, Baetidae and Hetageniidae) are characteristic of each group, however the relative and actual abundance varies considerably (Table 4). Groups 2 and 3 are very similar in terms of the families that contribute to the group similarity however the actual counts in Group 2 are an order of magnitude higher.

Group 1 is a small group of 14 sites with very high abundance dominated by 4 taxa Chironomids, Nemourid (stoneflies), Pisidiidae, (fingernail clams) and Baetid mayflies. The land cover is dominated by low shrubs and a higher proportion of water in the stream catchments. These are higher altitude and smaller drainage area streams and are widely dispersed in the catchment area.

Group 2 is the largest group (146 sites) and these represent the most commonly occurring assemblage in the region, dominated by Heptageniids and Chironomids. Total abundance is also high at these sites. These tend to be larger streams at higher altitudes with the highest stream density, but total precipitation is lower, particularly in January and February and they also have the lowest January minimum temperatures.

Group 3 has the lowest overall abundance and richness. The characteristic taxa are the same as Group 2, but abundance is an order of magnitude lower. The environmental attributes at these sites are generally similar to the Group 2 sites; they tend to be a little further west and lower altitude with larger drainage areas and higher precipitation.

Group 4 also has low total abundance, but these sites are dominated by Baetid mayflies and Chironomids are not as abundant in this assemblage. These are the lowest altitude sites and are located in Haida Gwaii. They have the highest total precipitation and the warmest Januarys with temperatures often above zero, although based on the total degree days the summers are cooler.



Model Performance

Table 6. Classification matrices for model based on resubstitution and cross-validation.

Resubstitution Matrix (Cases in row categories classified into columns)

Group	1	2	3	4	%correct
1	9	3	1	1	64
2	6	99	31	10	68
3	1	10	35	13	59
4	0	4	6	44	81
Total	16	116	73	68	68

Jackknifed (Cross-validation) Classification Matrix

Group	1	2	3	4	%correct
1	9	3	1	1	64
2	6	98	31	11	67
3	1	11	34	13	58
4	0	6	6	42	78
Total	16	118	72	67	67

This model performs well across the groups (Table 6), Group 3 is the most variable of the groups (Figure 3) and has the lowest classification accuracy , but still exceeds 58% correct classifications. The model is very robust only requiring 7 predictor variables.

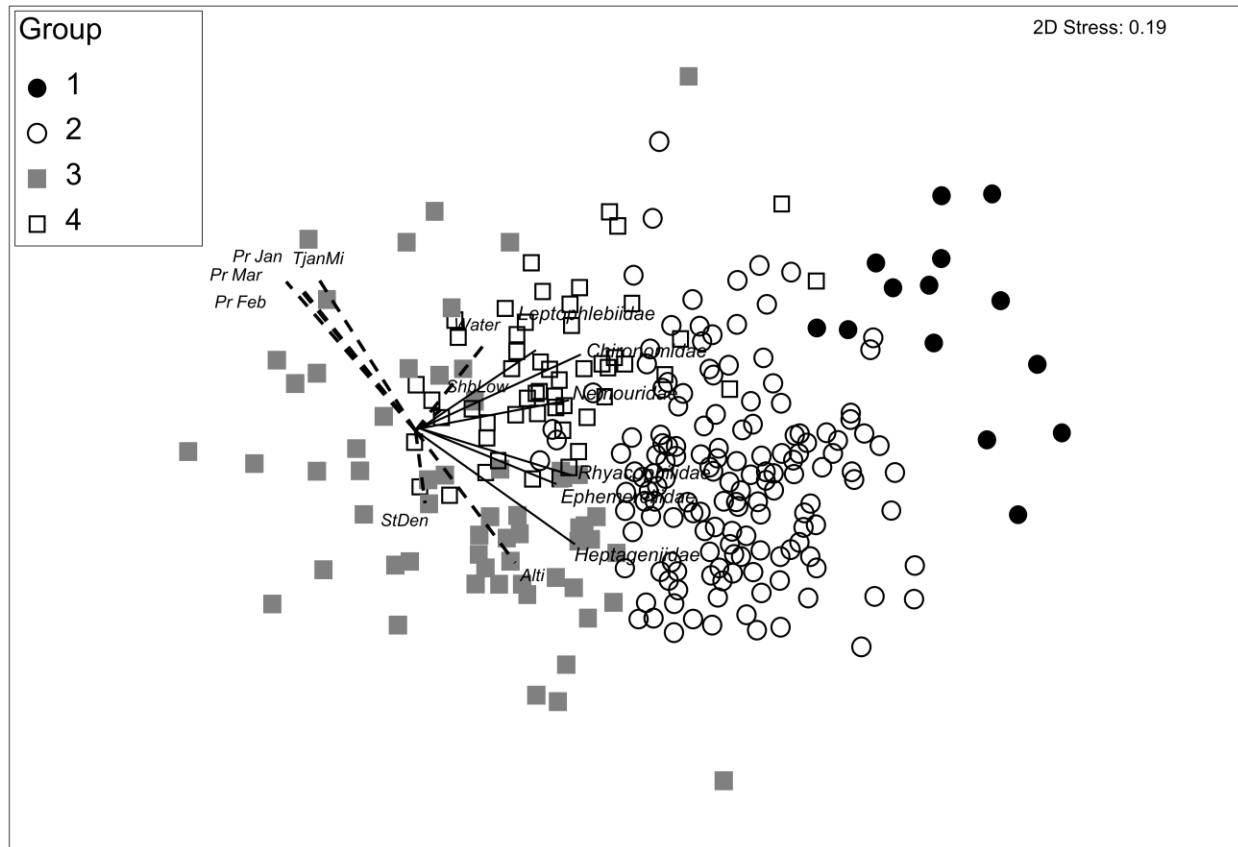


Figure 3. MDS ordination of 273 training reference sites used to build the DFA model illustrating 4 community assemblages.

To test model performance 55 sites (approx. 20%) were removed from the data set and treated as validation data (Table 7). The model was reconstructed using the training data (218 sites) and used to classify the 55 validation sites. The model derived from training has similar accuracy with cross validation, although the precision is a little lower. The validation sites are correctly classified at almost the same rate as the full model which therefore provides strong evidence that the accuracy reported by cross-validation (Table 7) is correct.

In addition the validation data were used to examine Type 1 and Type 2 error rates were examined using simulated data, referred to as simpacted data. For each of the 55 validation sites, five levels of intensity of disturbance were created for two types of disturbance, enrichment and placer mining. In both cases the number of individuals in each taxon was adjusted by a factor representing one of five levels of intensity of the disturbance (SIN = 0.2, 0.8, 1.0, 2.0, and 3.0) multiplied by the tolerance of the taxon to the type of disturbance (see Appendix 14 in Reynoldson and Bailey 2015) and the number of individuals in the validation sample. The tolerance of the taxon to disturbance was either derived from literature reported tolerances of taxa to enrichment (Barbour et al. 1999), or based on the correlation between the distribution of the taxon to habitat variables anticipated to be affected by placer mining activity (viz. TSS, % bedrock and boulder in the substrate, % sand, silt clay in the substrate and embeddedness) as reported by Reynoldson et al. (in prep.). Additionally the simpact intensity (SIN) level 0.0 was of course the unadjusted validation sites.

Table 7. Performance of model (273 sites) and training model (218 sites) and number of validation sites correctly classified.

	CABIN Model	Training data only	Validation sites correctly classified
Gp 1	64%	64%	2 of 3 (67%)
Gp 2	67%	65%	14 of 25 (56%)
Gp 3	58%	53%	10 of 16 (62%)
Gp 4	78%	84%	10 of 11 (91%)
Accuracy	67%	67%	36 of 55 (65%)
Precision (100-CV%)	87.4%	80.6%	77.8%

In order to determine the degree to which each of these disturbance levels modified the average community for the 55 sites we examined changes to richness and total abundance (Figure 4). The two simpacts show different types of response. The placer shows little effect at a SIN level of 0.2, richness is not affected till a SIN of 1.0. Enrichment simpacts show less effect on abundance and as tolerant taxa increase in response to the disturbance numbers actually increase at the highest SIN level. This is in fact a known response to enrichment. Based on these average changes simpacts at SIN levels of 1.0, 2.0 and 3.0 were assessed.

Table 8. Error rates (percent) for simpacted (P= placer, E=enrichment) sites for all sites and for each reference group at P=0.90. Type 1 error indicated by errors for SIN 0, and Type 2 errors for SIN 1.0, 2.0, and 3.0.

	SIN 0		SIN 1.0		SIN 2.0		SIN 3.0	
	P	E	P	E	P	E	P	E
All sites	12.7%	21.6%	29.1%	72.5%	5.5%	58.8%	5.5%	52.9%
Gp 1	0.0%	0.0%	100.0%	66.7%	33.3%	66.7%	33.3%	66.7%
Gp 2	8.0%	12.0%	24.0%	64.0%	4.0%	52.0%	4.0%	44.0%
Gp 5	18.8%	25.0%	12.5%	68.8%	0.0%	62.5%	0.0%	62.5%
Gp 7	18.2%	36.4%	45.5%	72.7%	9.1%	45.5%	9.1%	36.4%

Across all the sites at P = 0.90 the Type 1 error rates range from 12.7 to 21.6% and the Type 2 error rates range from 5.5-72.5% (Table 8). The Type 2 error rates in particular are substantially higher with the enrichment simpact (highest error = 72.5%) than the placer simpact (highest 29.1%). Not surprising as the community change in the placer mining simpact is substantially greater than in the enrichment simpact (Figure 4). There is considerable variation among groups, with Group 5 sites being more sensitive to detecting disturbance than Group 1 sites. This is largely a reflection of the natural variation within groups.



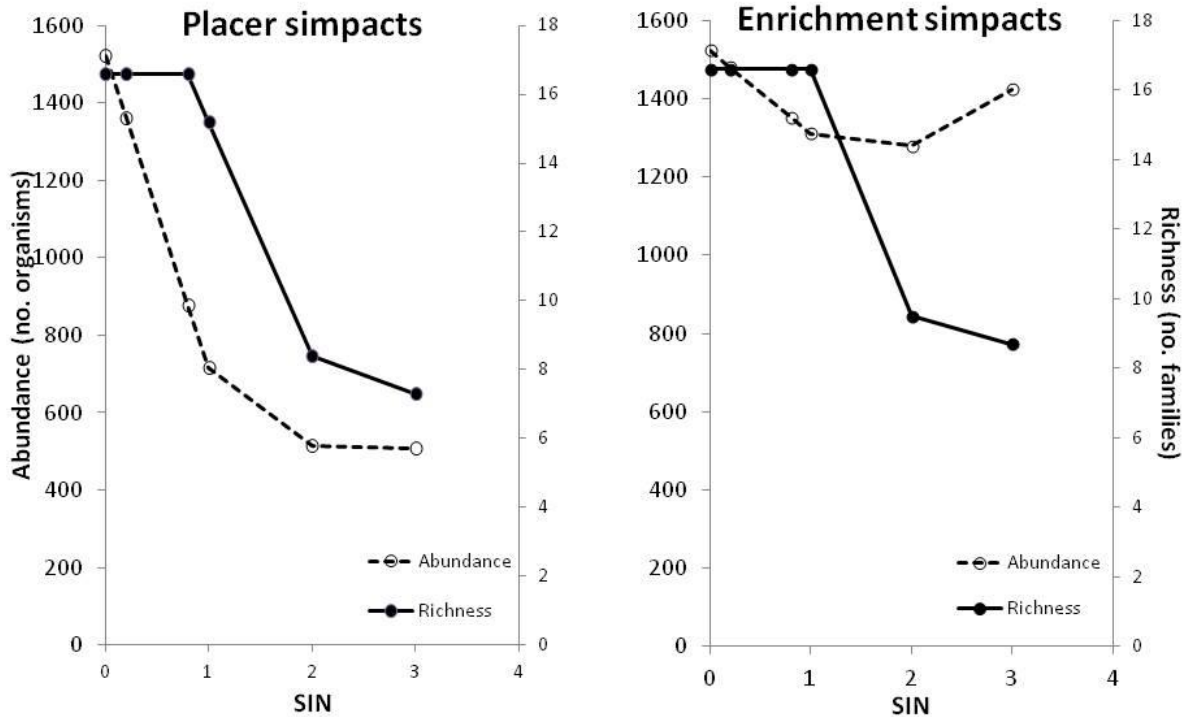


Figure 4. Summary of placer mining and enrichment simpacts on richness and abundance.

3.0 OTHER RELEVANT LITERATURE

Reynoldson, T.B., R.C. Bailey, and J.L. Bailey. (in prep). A review of the Yukon Placer Mining Aquatic Health Monitoring Protocol. Report prepared for the Yukon Territorial Government by GHOST Environmental Consulting. 46pp.

Bennett, S. (2010) Skeena Beast09: Supporting documentation for the Skeena Beast09 RCA predictive model. Report prepared for BC Ministry of Environment by BioLogic Consulting. 19pp.

Reynoldson, T.B. and J.L. Bailey (June 2015) A reference condition model for the Skeena Basin and North Coast of British Columbia, based on sites sampled from 2004-2013. Report prepared for BC Ministry of Environment by GHOST Environmental Consulting. 54pp.

APPENDIX: DATA COLLECTION, ANALYSIS AND QUALITY ASSURANCE

Field Collection

CABIN Study Name	BC MOE- FSP Skeena Region	National Parks- Gwaii Haanas	BC-Kitsault	BC- Skeena BC Timber Sales	BC-0887; BC-Dome Mtn Baseline Aesthetics	BC-Eskay Creek Mine EEM; BC-IM Red	BC-Long Lake IPP	BC- 123210182
Agencies involved	BCMOE	Parks Canada	AMEC /Azimuth	BC Timber Sales	ERM	Golder	Triton	Stantec
Date range	2004-2008	2006-2012	2010-2013	2007-2008	2009-2013	2010-2013	2011-2012	2012
Sampling season	Mid-Aug- Late Sept	Late Aug- Early Sept	Aug	Late Aug- Early Sept	Late Aug- Early Sept	Early Sept- Early Oct	Late Aug	Late Aug
# reference samples	138	63	15	22	10	9	4	11
Certified samplers (Y or N)	N	N	Y	Y	Y	Y	Y	Y
Certified team leader (Y or N)	N	N	Y	Y	Y	Y	Y	Y
400 um kicknet (Y or N)	Y	Y	Y	Y	Y	Y	Y	Y
Other sampling device	--	--	--	--	--	--	--	--
Field Preservative	Ethanol	Formalin	Formalin	Ethanol	Formalin	Formalin	Formalin	Formalin

3.2 Macroinvertebrate Identification

CABIN Study Name	BC MOE-FSP Skeena Region	National Parks-Gwaii Haanas	BC-Kitsault	BC-Skeena BC Timber Sales	BC-0887; BC-Dome Mtn Baseline Aquatic	BC-Eskay Creek Mine EEM; BC-IM Red	BC-Long Lake IPP	BC-123210182
Taxonomist	Fraser Environmenta l/ Cordillera Consulting	Cordillera Consulting/ EcoAnalysts	Cordillera Consulting	Fraser Environmenta l/ Cordillera Consulting	Biologica; Cordillera Consulting; EcoAnalysts	Cordillera Consulting	Triton (Jim Trask)	Stantec
Marchant Box used (Y or N)	N	Y	Y	N	Y	Y	N	N
Other subsampling device	Caton Tray	--	--	Caton Tray	--	--	Folsom Splitter	Sieve Sorted
Subsample count	300 fine fraction (<1mm) Total count large fraction (>1mm)	300	300	300 fine fraction (<1mm) Total count large fraction (>1mm)	300	300	200 fine fraction (<2 mm), 200 large fraction (>2 mm)	300
10% of reference samples sent to National Lab for QA	N	Y	N	N	N	N	N	N
Reference Collection maintained	N	Y	N	N	Y	N	N	Y

3.3 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 ArcHydro 2.0 (ESRI 2010). Delineations were based on 20 m resolution digital elevation models (DEM) and a 1:50,000 scale hydrological network. The DEM was subjected to pre-processing which “burned in” the stream network into the DEM and filled sinks to improve flow modeling. The corrected DEM was used to calculate flow direction and flow accumulation to carry out the terrain procession steps to model catchment areas (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-mcan.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)

3.5 Statistical Analyses

Statistical Programs used:

- Excel - data manipulation and storage
- PATN V.3.12 - classification and ordination of test sites for assessment
- PRIMER 6 - classification, MDS ordination, ANOSIM, SIMPER, PCA for habitat variables and BEST for matching invertebrate and habitat resemblance matrices
- SYSTAT 13 - discriminant analysis, multiple regression and plotting BEAST assessments with probability ellipses

Model was reviewed by the CABIN Science Team July 2015. Comments provided and recommendations applied August 2015, final copy submitted October 2015.

