

Benthic evaluation of ORegon rIverS (BORIS - A RIVPACS type model)

In this approach, sites which are considered to be in a reference condition are used as a benchmark against which test sites are compared. The reference sites are classified into distinct groups, based upon both their biological and physico/chemical attributes. A test site is compared to a reference site group with similar environmental characteristics. Taxa found at the test site is compared with that typifying the reference group and a determination of whether a test site is adequately supportive of aquatic life is made.

Essentially, the procedure involves the following steps:

1. Reference Site Classification

A set of minimally disturbed reference condition sites are sampled for aquatic macroinvertebrates and a suite of environmental parameters are measured. The base of reference sites should cover the range of environmental conditions expected at future test sites. For example, it would not be valid to assess a low elevation test site if all reference groups in the data set were composed of only high elevation sites. The macroinvertebrate assemblage sub groups that exist within the pool of reference sites are then generated using a multivariate clustering technique.

The similarity measure used to classify reference sites in the BORIS model is the Bray-Curtis or Czekanowski coefficient, defined as:

$$S_c = \frac{2 \min(X_{ij}, X_{ik})}{X_{ij} + X_{ik}},$$

where $\Sigma \min(X_{ij}, X_{ik})$ is the sum of the lesser abundance of each species i when it occurs in both sites j and k . Similarity ranges between 1 (identical species and abundance) to 0 (no common taxa). In Euclidean space, each site is thus defined in terms of its similarity to all other sites. Similar sites will cluster together in the multi-dimensional space and these clusters reflect the different biological community groups in the reference data set.

NOTE: The above returns the **similarity value** between sites. The Bray-Curtis dissimilarity measure necessary for ordination is actually $1 - S_c$. Alternatively, the following formulae calculates the dissimilarity directly:

$$D = \frac{|X_{ij} - X_{ik}|}{(X_{ij} + X_{ik})}$$

The linking strategy used to connect similar sites is the agglomerative fusion technique UPGMA (Unweighted Pair Group Averaging), provided by the software package PATN (Belbin, 1995). The RIVPACS model in Britain still uses TWINSpan as the clustering technique.

Once Reference Sub groups have been generated then the fidelity of each taxon for each reference group is then determined. For example, if *Cinygmula* was collected at 3 sites within a Reference Group A which contained a total of 10 sites then its fidelity for this reference type is simply $3/10 = 0.3$.

It should be noted that the designation of these reference community groups is really an abstraction as the aquatic invertebrate community is actually a continuum both spatially and temporally.

2. Reference Group Discriminant Function Analysis

Once reference groups have been established based upon their biological community, these groups are then described in terms of their environmental features. A Stepwise Multiple Discriminant Function Analysis is used to undertake multiple analyses of variance for environmental parameters against group membership. The parameters which best account for the variability in the environmental data set are selected one at a time and incorporated into a relatively simple model. The ability of the model to correctly discriminate between each reference group can be assessed progressively as each variable is added. Usually only three or four parameters are required to generate an acceptable model. The significant parameters are combined into arithmetic equations termed discriminant function roots. These roots define the Euclidean space in which the sites are located based upon their environmental characteristics. Sites situated closely together will have similar environmental features.

What is sought here is a model of reference site environmental characteristics which allocates sites in a manner consistent with those derived from the multivariate analysis of macroinvertebrate assemblages.

3. Applying a Test Site

The model is then used to predict the most likely reference group membership of a test site (assuming unimpaired conditions exist). The probability of a test site belonging to each of the reference groups is derived from the discriminant function roots in step 2.

4. Determination of Impairment

The probability of finding a taxon (from the reference site data set), at a test site can now be determined. The proportion of taxa observed at a test site (i.e. collected at a sampling event) vs. that predicted by the model is used as the determinant of biological impairment. In the complete absence of impairment, the ratio should be close to one (i.e. all expected taxa were actually collected). A simple pass/fail test might require collection of >50% of the expected taxa. In BORIS the standard deviation of the reference site scores is used to determine impairment classes.

BORIS in the Oregon Coast Range

Figure 1 presents a UPGMA generated dendrogram for 34 reference sites within the Oregon Coast Range. From this classification analysis three groups are extracted. These groups essentially fall out as North, Mid and South Coast sites, shown in Figure 2.

Table 1 presents a sample of the taxa found at the reference sites and their fidelity for each of the three groups shown below.

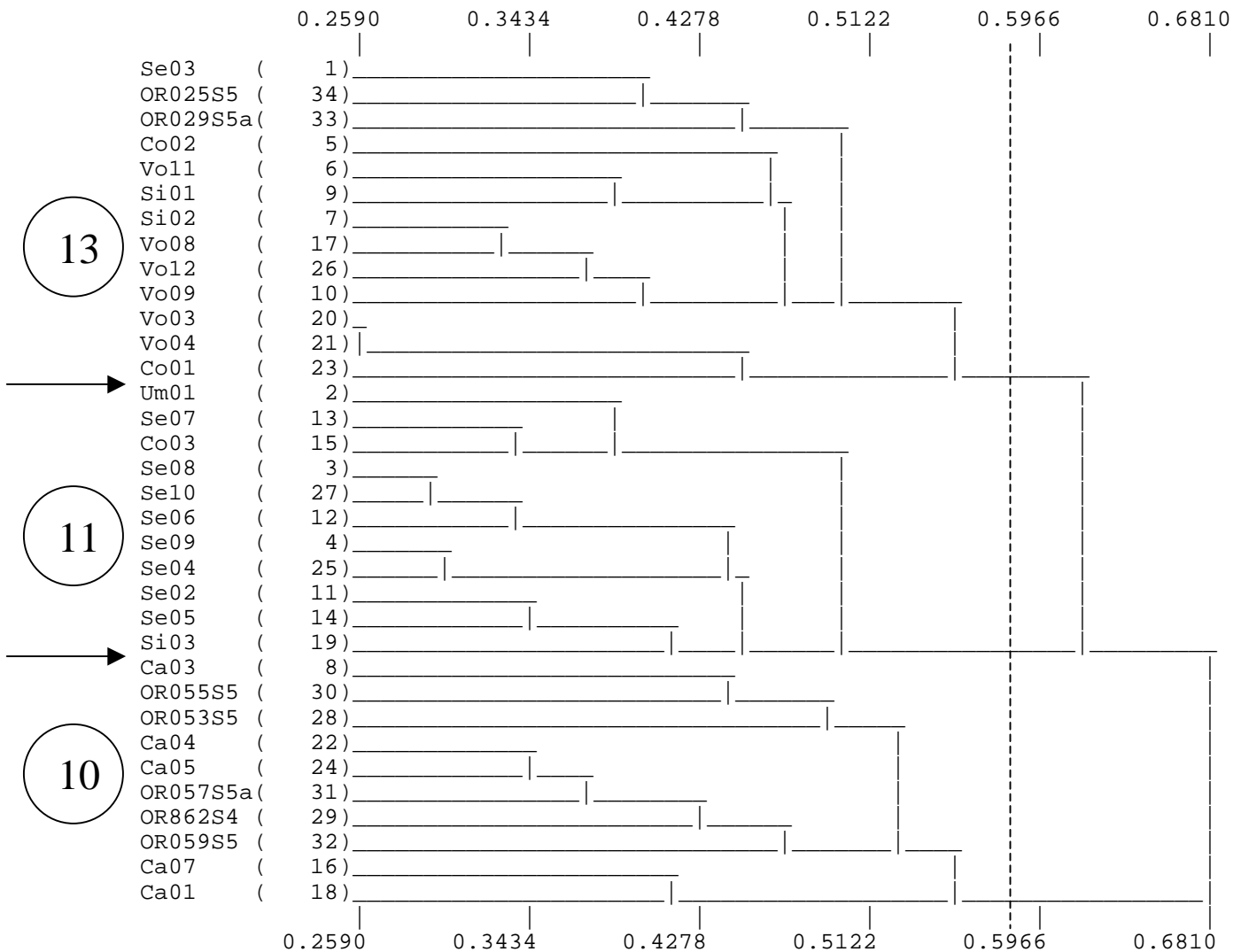


Figure 1. Reference sites in the Oregon Coast Range. Dendrogram of UPGMA analysis.

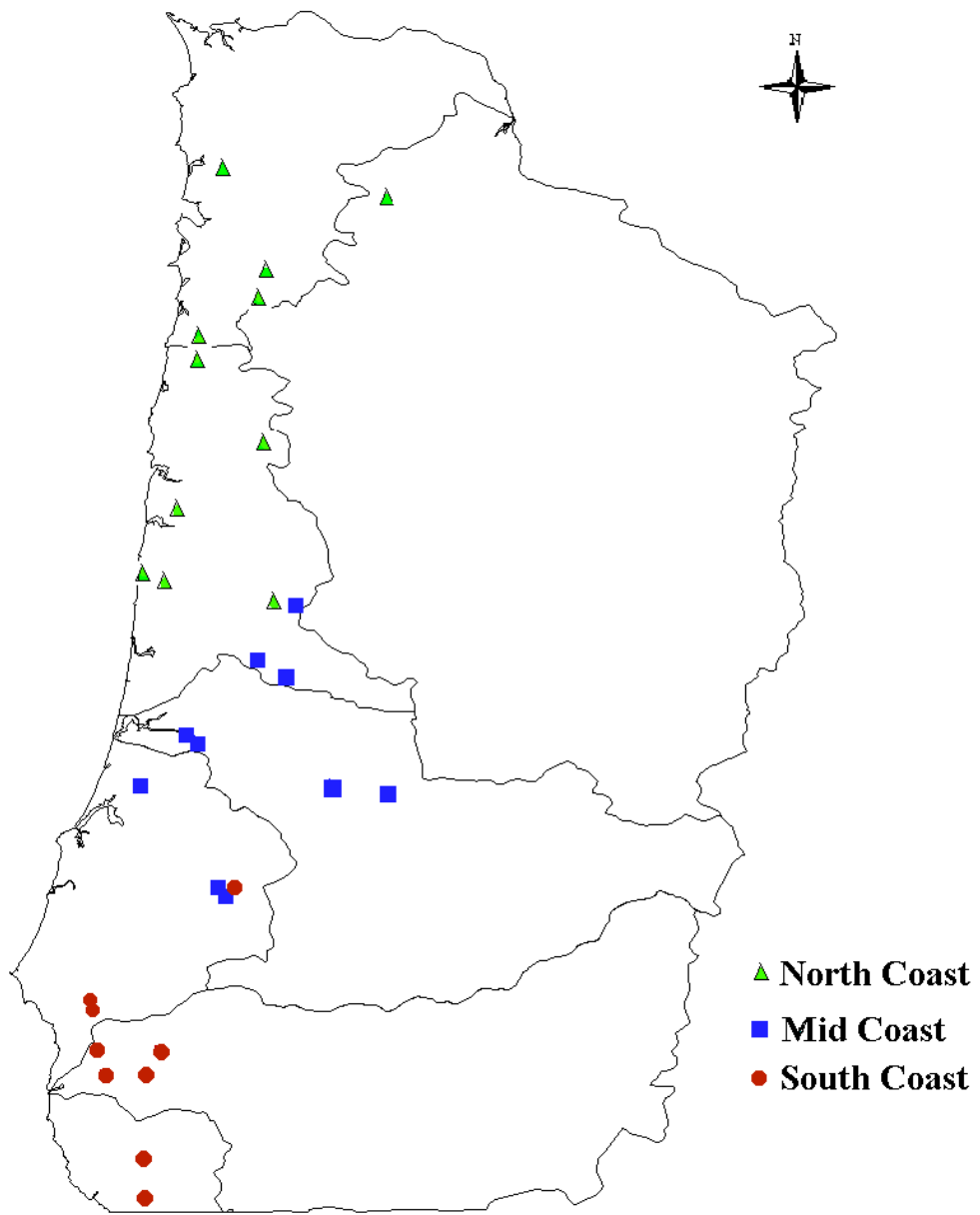


Table 1. Reference Site Taxa Fidelity

Taxon	Grp1	Grp2	Grp3	Taxon	Grp1	Grp2	Grp3	Taxon	Grp1	Grp2	Grp3
Ephemeroptera			Plecoptera			Trichoptera					
Baetis bicaudatus	0.00	0.09	0.00	Capniidae	0.23	0.00	0.00	Heteroplectron	0.08	0.09	0.00
Baetis tricaudatus	1.00	1.00	1.00	Chloroperlidae	0.69	0.09	0.80	Lepidostoma	0.31	0.55	0.30
Centroptilum	0.00	0.00	0.20	Kathroperla perdita	0.08	0.00	0.30	Amiocentrus	0.00	0.00	0.20
Dipheter hageni	0.77	1.00	0.80	Paraperla	0.31	0.00	0.20	Rhyacophila Angelita Gr.	0.08	0.09	0.60
Acentrella	0.15	0.00	0.20	Sweltsa	0.92	0.82	0.30	Rhyacophila Betteni Gr.	0.69	0.91	1.00
Attenella	0.23	0.09	0.10	Leuctridae	0.38	0.27	0.00	Rhyacophila Brunnea Gr.	0.77	0.45	0.20
Caudatella	0.08	0.00	0.10	Despaxia augusta	0.08	0.00	0.00	Rhyacophila Coloradensis Gr.	0.00	0.00	0.10
Drunella doddsi	0.77	0.27	0.60	Moselia infuscata	0.46	0.00	0.00	Rhyacophila Hyalinata Gr.	0.62	0.18	0.20
Drunella grandis	0.00	0.09	0.10	Amphinemura	0.00	0.09	0.00	Rhyacophila arnaudi	0.08	0.18	0.00
Drunella pelosa	0.00	0.00	0.10	Malenka	0.31	0.73	0.90	Rhyacophila Rotunda Gr.	0.00	0.09	0.00
Drunella coloradensis/flavilinea	0.00	0.00	0.40	Soyedina	0.08	0.00	0.00	Rhyacophila Sibirica Gr.	0.85	0.55	0.00
Ephemerella	0.31	0.00	0.20	Zapada cinctipes	0.69	1.00	0.10	Micrasema	0.46	0.27	0.40
Serratella	0.46	0.00	0.40	Zapada columbiana	0.08	0.00	0.00	Apatania	0.00	0.36	0.10
Timpanoga	0.23	0.09	0.50	Zapada frigida	0.08	0.00	0.10	Agapetus	0.00	0.00	0.60
Cinygma	0.15	0.00	0.00	Zapada Oregonensis Gr.	0.46	0.55	0.20	Anagapetus	0.08	0.00	0.00
Cinygmula	0.92	0.73	0.50	Calineuria californica	0.85	1.00	1.00	Glossosoma	1.00	0.91	0.90
Epeorus albertae	0.23	0.36	0.80	Doroneuria	0.38	0.00	0.10	Arctopsyche grandis	0.54	0.09	0.40
Epeorus longimanus	0.15	0.00	0.40	Hesperoperla pacifica	0.62	0.45	0.70	Cheumatopsyche	0.00	0.09	0.00
Epeorus grandis	0.23	0.00	0.00	Diura	0.08	0.00	0.10	Hydropsyche	1.00	0.64	0.80
Heptagenia	0.08	0.00	0.00	Isoperla	0.23	0.09	0.10	Hydroptila	0.08	0.00	0.10
Nixe/Leucocruta	0.15	0.45	0.70	Perlinodes	0.15	0.00	0.10	Neotrichia	0.00	0.09	0.30
Ironodes	0.62	0.45	0.10	Skwala	0.85	0.45	0.10	Ochrotrichia	0.00	0.00	0.40
Rhithrogena	0.92	1.00	0.60	Yoraperla brevis	0.15	0.45	0.40	Allocosmoecus	0.00	0.27	0.00
Paraleptophlebia	0.92	0.73	0.50	Pteronarcella	0.00	0.09	0.00	Dicosmoecus gilvipes	0.08	0.00	0.10
Ameletus	0.23	0.00	0.40	Pteronarcys	0.15	0.00	0.10	Ecclisocosmoecus	0.08	0.00	0.00
Tricorythodes	0.00	0.00	0.10					Ecclisomyia	0.08	0.00	0.00
								Goera archaon	0.00	0.09	0.10
								Hydatophylax	0.08	0.00	0.10
								Dolophilodes	0.08	0.00	0.00
								Wormaldia	0.69	0.82	0.70
								Polycentropus	0.08	0.00	0.10
								Psychomyia	0.08	0.00	0.00
								Neophylax	0.31	0.91	0.80
								Oligophlebodes	0.08	0.00	0.10
								Farula	0.08	0.00	0.00
								Gumaqa	0.00	0.09	0.20

Having divided reference sites into specific community groups based upon their macroinvertebrate assemblages, the next phase is to model those groups in terms of their environmental conditions. As mentioned above, only parameters that are essentially not influenced by human activities should be considered. The following parameters were used in the Discriminant Function Analysis:

- Latitude
- Longitude
- Elevation
- Wetted width
- Stream Order
- Drainage (watershed area, upstream of sample site)
- Discharge (measured at time of sampling)
- Slope (measured from 7.5 minute quad map)
- Alkalinity

Analysis was performed using the software program STATISTICA and the program yielded two significant parameters: Latitude and Longitude which combined into two discriminant root functions, shown below. These correctly assigned 94% of the reference sites. The two incorrect assignments fell into "adjacent" groups.

$$\text{Root 1 value} = 212.1686 - 2.5069 * \text{Latitude} - 0.8265 * \text{Longitude}$$

$$\text{Root 2 value} = 563.5425 - 0.5896 * \text{Latitude} - 4.3427 * \text{Longitude}$$

Figure 3 displays the location of each reference site as determined by the above root functions. Also shown are the centroids of the three groups.

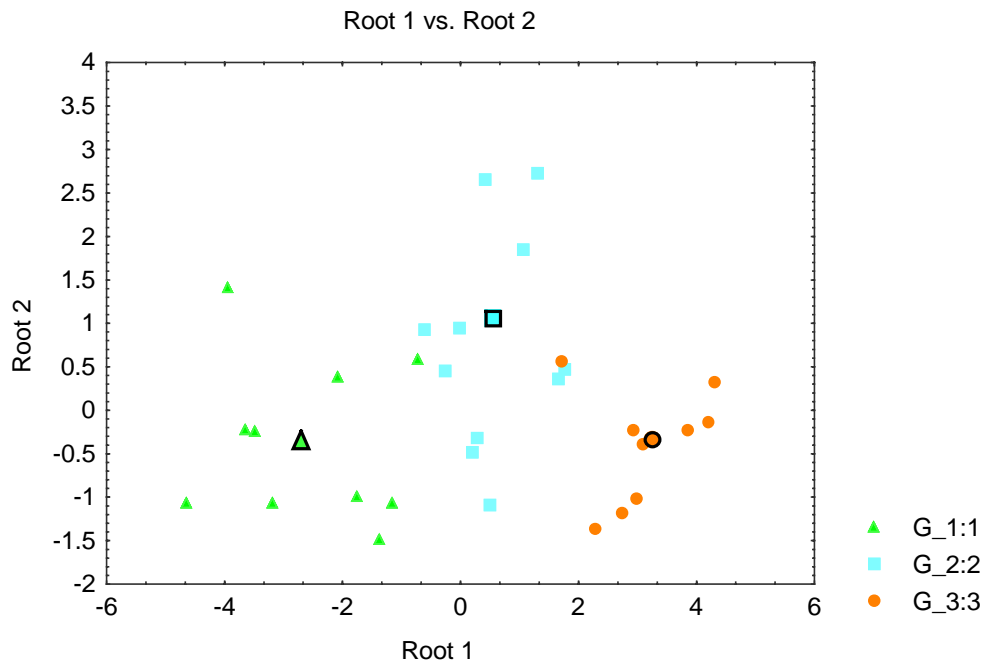


Figure 3. Plot of canonical discriminant root scores for reference sites. Highlighted symbols are mean positions of each Reference Group.

Figure 4 shows how a test site X can be located on a set of axes according to its solution to the two Discriminant Function Roots above. This site lies a certain "distance" d from the centroids of each reference group. The closer the test site is to a particular centroid, the more similar in terms of its environmental characteristics is it to the members of that cluster. This Euclidean distance is calculated by:

$$d_j^2 = \sum_{i=1} (x_i - m_{i,j})^2$$

where x_i = the discriminant root i score for test site x and d_j^2 = square of distance from test site x to group j represented by its centroid $m_{i,j}$ (mean score of discriminant root i for reference group j).

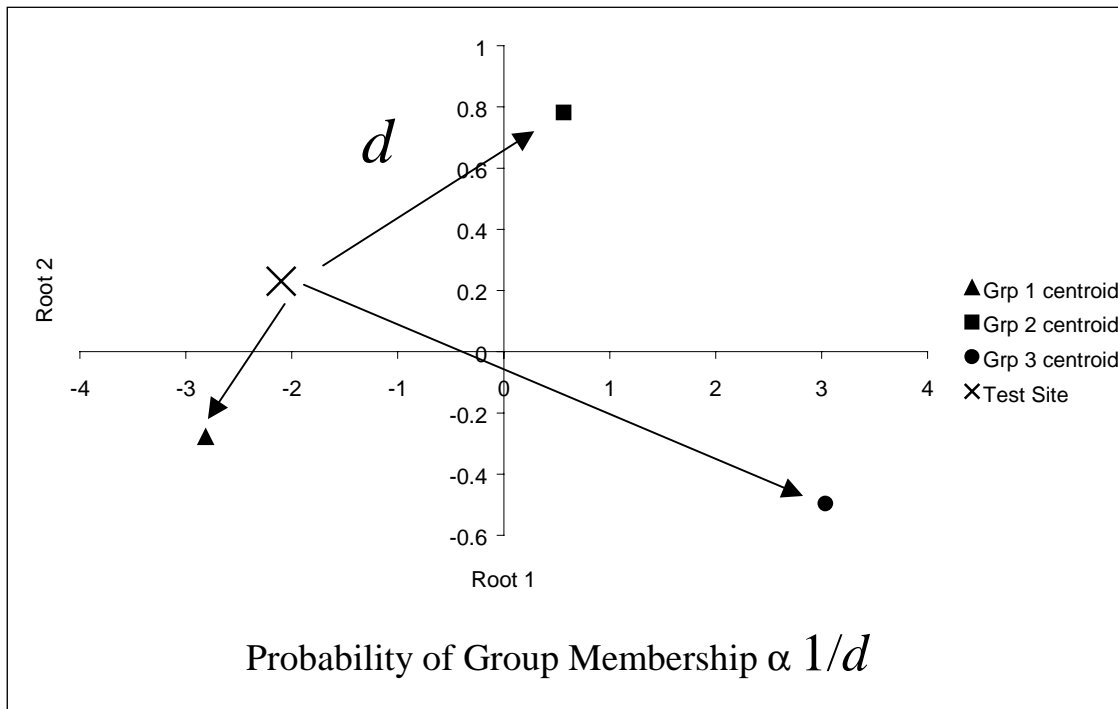


Figure 4. Location of a test site on Root Function Axes

The probability that a test site is a "member" of each reference group is proportional to the number of members of that group multiplied by the exponential of minus one half the distance from the group centroid. Defined by the formula:

$$p_j = \frac{q_j}{\sum_{j=1}^n q_j}$$

and,

$$q_j = n_j * \exp\left(-\frac{d_j^2}{2}\right),$$

where p_j = probability of membership of a test site to group j and n_j = number of reference sites in group j .

Thus, a site might have a high probability of being "like" a single reference type or it might display a character from two or more reference groups depending upon its location in relation to centroids of the reference groups. Table 2 presents a short list of coastal sites and their probabilities of reference group membership. In the absence of impairment, Agency Creek should display a macroinvertebrate assemblage which is "100%" similar to sites from Reference group 1 while Cullen Creek is predicted to display an assemblage make up which is 83% and 17% similar to that of Reference groups 1 and 2 respectively.

Table 2. Test Site Probabilities of Reference Site Group Membership

Site Name	Date	PROB_GRP1	PROB_GRP2	PROB_GRP3
AGENCY CREEK AT RM 0.2	14-Aug-96	1.00	0.00	0.00
BENSON CREEK AT RM 5.0	17-Jul-96	0.02	0.98	0.00
CABIN CREEK AT RM 5.6	23-Jul-96	0.00	1.00	0.00
CULLEN CREEK AT RM 0.3	6-Aug-96	0.83	0.17	0.00
ELK RIVER AT RM 24.0	20-Aug-96	0.00	0.30	0.71
EMILY CREEK AT RM 7.1	18-Aug-96	0.00	0.01	0.99

For a given test site, the probability (r_k) of occurrence of each taxon k found at the reference sites is the test sites probability of reference group membership (p_j) multiplied by the proportion of members of the group j at which the taxon actually occurred. This is calculated and then summed across all reference groups. What results is a weighted probability of occurrence with respect to the fidelity the taxon shows towards a reference community type and with respect to the likelihood that the test site belongs to a particular reference group. This is defined by:

$$r_k = \sum_{j=1}^n p_j * g_{j,k},$$

where $g_{i,j}$ = proportion of members of group j at which taxon k was found.

Taking the case of Cullen Cr. and the taxon Cinygmula whose fidelity is shown below it can be seen that the overall probability of finding Cinygmula at this site is:

$$0.92 \times 0.83 + 0.73 \times 0.17$$

= 0.89

Taxon fidelity	Grp1	Grp2	Grp3
Rhithrogena	0.92	1.00	0.60
Cinygmula	0.92	0.73	0.50
Epeorus albertae	0.23	0.36	0.80
Epeorus longimanus	0.15	0.00	0.40

The complete BORIS taxonomic prediction for the coastal site Brush Creek is given in Table 2. A total of 58 taxa are predicted by the model at the 20% probability level. A probability limit can be placed upon the predictive calculations, in this case only taxa with an overall probability of at least 20% are used to assess the site - in AUSRIVAS a level of 50% has been adopted while RIVPACS itself I believe does not set a lower limit; the lowering of this limit is attenuated by the fact that only progressively smaller probabilities are added to the predicted list of taxa. The total sum of probabilities for Brush Cr. comes to 32.95. Thus, any one single collection effort needs to record any 33 of the predicted 58 taxa for the site to score 1.0 (the overall reference model mean score). The final assessment for Brush Cr. would be:

	BrushC94
No. taxa predicted (at 20% probability level)	58
No. taxa expected (sum of individual probabilities) (E Value)	32.95
No. taxa actually captured (O Value)	31
O/E	0.94

Table 2: BORIS predictions for Brush Cr.

Site Code	Taxon_Code_B	SCIENTIFIC_NAME	TaxaProb	Site Code	Taxon_Code_B	SCIENTIFIC_NAME	TaxaProb		
1	BrushC94	CHI010	Chironominae	0.92	30	BrushC94	EPH750	Ameletus	0.23
2	BrushC94	CHI020	Diamesinae	0.30	31	BrushC94	GAS610	Juga	0.55
3	BrushC94	CHI030	Orthocladiinae	1.00	32	BrushC94	HYD000	Hydracarina	0.46
4	BrushC94	CHI060	Tanypodinae	0.47	33	BrushC94	OLI000	Oligochaeta	0.69
5	BrushC94	COL225	Heterlimnius	0.83	34	BrushC94	PLE100	Capniidae	0.23
6	BrushC94	COL240	Narpus	0.46	35	BrushC94	PLE200	Chloroperlidae	0.68
7	BrushC94	COL250	Optioservus	0.32	36	BrushC94	PLE220	Paraperla	0.30
8	BrushC94	COL290	Zaitzevia	0.70	37	BrushC94	PLE250	Sweltsa	0.92
9	BrushC94	DIP051	Glutops	0.53	38	BrushC94	PLE300	Leuctridae	0.38
10	BrushC94	DIP210	Ceratopogoninae	0.31	39	BrushC94	PLE331	Moselia infuscata	0.45
11	BrushC94	DIP310	Dixa	0.45	40	BrushC94	PLE430	Malenka	0.32
12	BrushC94	DIP400	Empididae	0.30	41	BrushC94	PLE491	Zapada cinctipes	0.70
13	BrushC94	DIP520	Pericoma	0.38	42	BrushC94	PLE494	Zapada Oregonensis Gr.	0.46
14	BrushC94	DIP620	Simulium	1.00	43	BrushC94	PLE511	Calineuria californica	0.85
15	BrushC94	DIP910	Antocha	0.45	44	BrushC94	PLE530	Doroneuria	0.38
16	BrushC94	DIP940	Hexatoma	0.46	45	BrushC94	PLE541	Hesperoperla pacifica	0.61
17	BrushC94	EPH112	Baetis tricaudatus	1.00	46	BrushC94	PLE620	Isoperla	0.23
18	BrushC94	EPH140	Dipheter hageni	0.77	47	BrushC94	PLE670	Skwala	0.84
19	BrushC94	EPH310	Attenella	0.23	48	BrushC94	TRI035	Lepidostoma	0.31
20	BrushC94	EPH332	Drunella doddsi	0.76	49	BrushC94	TRI1130	Rhyacophila Betteni Gr.	0.70
21	BrushC94	EPH340	Ephemerella	0.30	50	BrushC94	TRI1140	Rhyacophila Brunnea Gr.	0.76
22	BrushC94	EPH350	Serratella	0.45	51	BrushC94	TRI1170	Rhyacophila Hyalinata Gr.	0.61
23	BrushC94	EPH360	Timpanoga	0.23	52	BrushC94	TRI1210	Rhyacophila Sibirica Gr.	0.84
24	BrushC94	EPH520	Cinygmula	0.92	53	BrushC94	TRI130	Micrasema	0.46
25	BrushC94	EPH531	Epeorus albertae	0.23	54	BrushC94	TRI230	Glossosoma	1.00
26	BrushC94	EPH534	Epeorus grandis	0.23	55	BrushC94	TRI311	Arctopsyche grandis	0.53
27	BrushC94	EPH550	Ironodes	0.61	56	BrushC94	TRI340	Hydropsyche	0.99
28	BrushC94	EPH560	Rhithrogena	0.92	57	BrushC94	TRI780	Wormaldia	0.70
29	BrushC94	EPH620	Paraleptophlebia	0.92	58	BrushC94	TRI967	Neophylax rickeri	0.32

Sum of Probabilities: 32.95

A site can be placed into an impairment category based upon the distribution of reference site scores. Theory predicts a normal distribution of reference site scores around a mean of 1.0. Figure 5 shows the frequency distribution of reference scores for the current model. A confidence interval of 95% (± 1.96 X Stand. Dev.) is applied to the data set, in this case resulting in the criteria listed below.

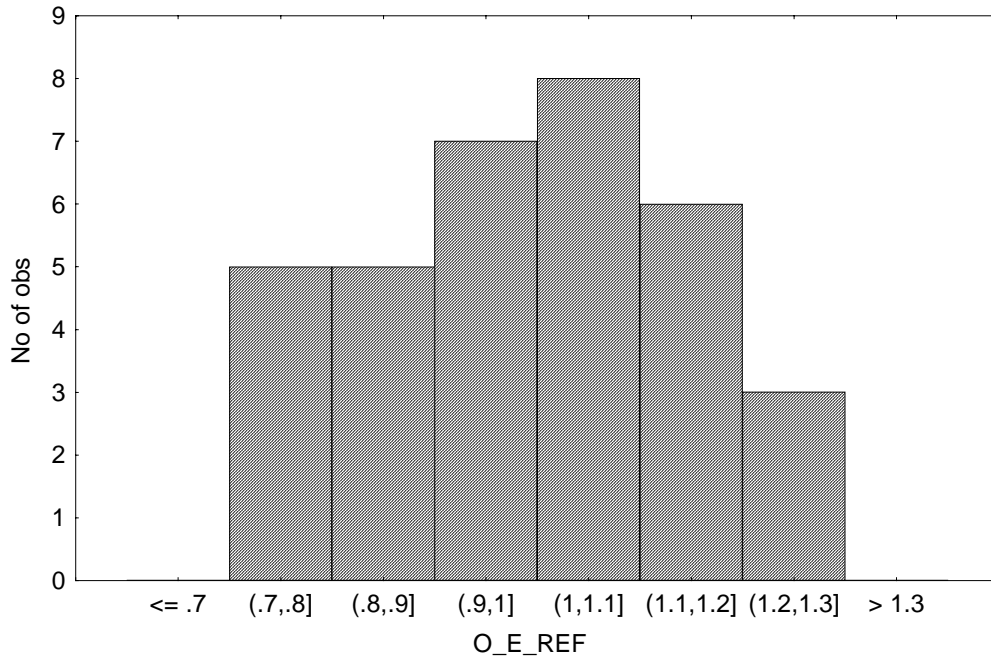
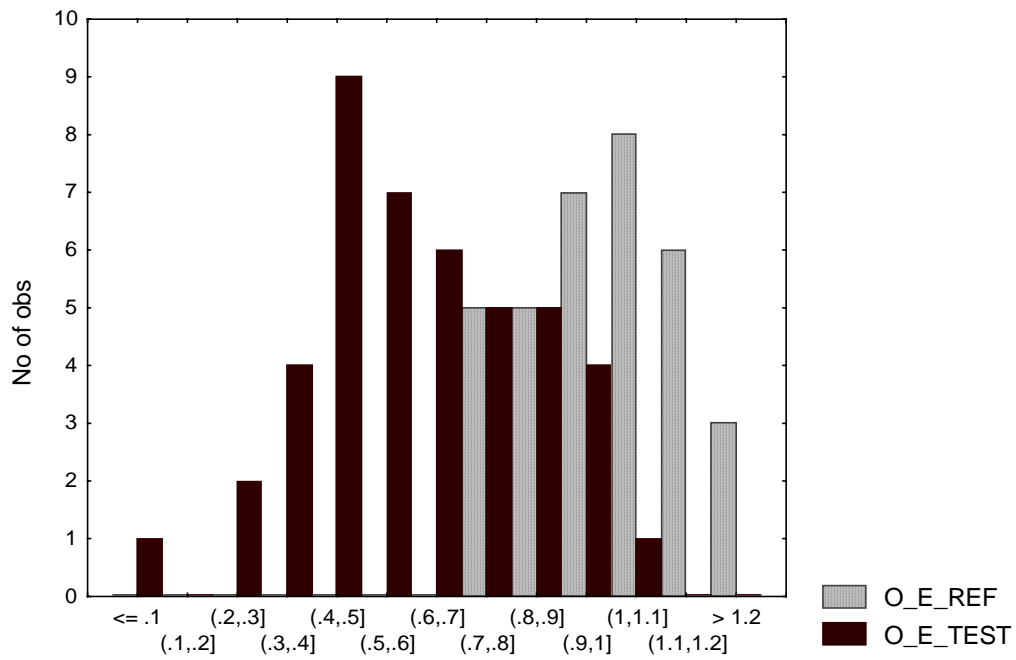


Figure 5. Frequency distribution of reference site scores

Category	Score	Score
No impairment detected	>0.72	(< 1.96 X Std. dev. ref. site scores)
Moderate impairment	$0.44 - 0.72$	($1.96 - 3.92$ X Std. dev. ref. site scores)
Severe impairment	<0.44	(> 3.92 X Std. dev. ref. site scores)



Frequency distribution of BORIS reference sites and REMAP Coast Range sites

The following list of papers present information on the RIVPACS modeling approach. Papers in bold type represent publications by the original authors of the method and should be considered required reading for developing a RIVPACS type model.

Marchant, R., A. Hirst, R. H. Norris, R. Butcher, L. Metzeling, and D. Tiller. 1997. Classification and prediction of macroinvertebrate assemblages from running water in Victoria, Australia. *Journal North American Benthological Society*, 16(3): 664-681.

Moss, D., M. T. Furse, J. F. Wright, and P. D. Armitage. 1987. The prediction of the macroinvertebrate fauna of unpolluted running-water sites in Great Britain using environmental data. *Freshwater Biology*, 17: 41-52.

Norris, R. H., and A. Georges. 1993. Analysis and interpretation of benthic macroinvertebrate surveys. Pages 234-286 in D. M. Rosenberg and V. H. Resh, editors. *Freshwater biomonitoring and benthic macroinvertebrates*. Chapman & Hall, New York.

Wright, J. F. 1996. Development and use of a system for predicting the macroinvertebrate fauna in flowing waters. *Australian Journal of Ecology*, 20: 181-197.

Wright, J. F., J. H. Blackburn, R. J. M. Gunn, M. T. Furse, P. D. Armitage, J. M. Winder, and K. L. Symes. 1996. Macroinvertebrate frequency data for RIVPACS III sites in Great Britain and their use in conservation evaluation. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 6: 141-167.

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