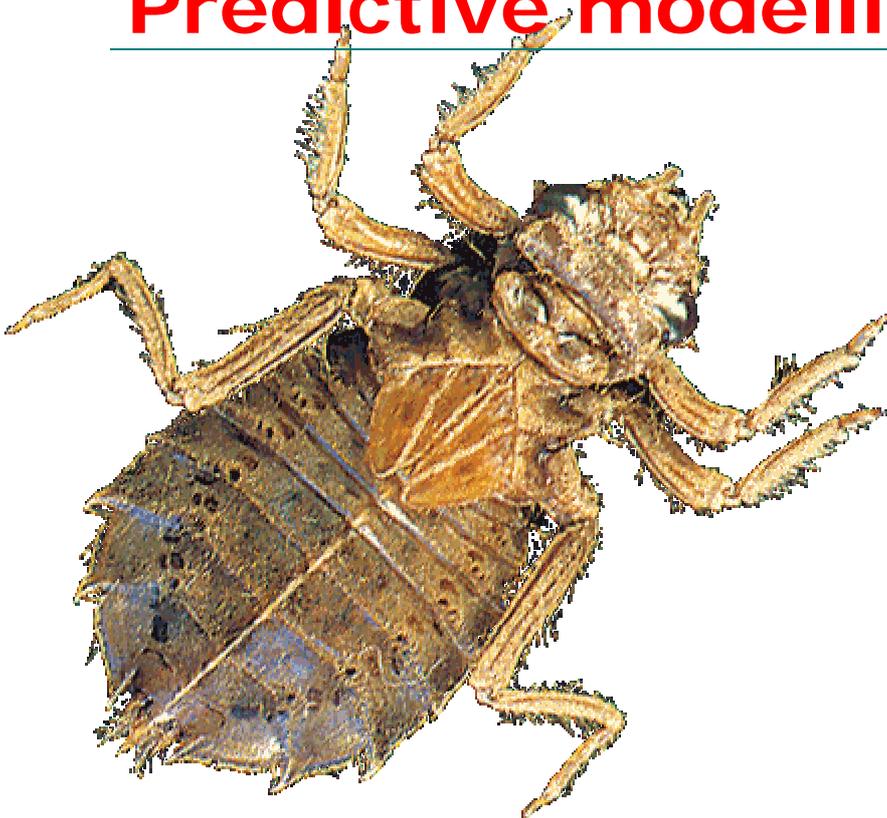# AUSRIVAS
Macroinvertebrate bioassessment

Predictive modelling manual

AUSRIVAS

Julie Coysh, Sue Nichols, Gail Ransom, Justen Simpson, Richard Norris[1],
Leon Barmuta[2] and
Bruce Chessman[3], 2000

[1]CRC for Freshwater Ecology, Building 15, University of Canberra, ACT 2601
[2]University of Tasmania, GPO Box 252-05 Hobart 7001
[3]NSW Department of Land and Water Conservation, PO Box 3720, Parramatta NSW 2124

AUSRIVAS

Contents

1.   Introduction

In response to growing concern in Australia for maintaining ecological values, the <u>National River Health Program</u> (NRHP) was formed. A major component of the NRHP was the development of the computer program, Australian River Assessment System (AUSRIVAS), for use in assessing river health. AUSRIVAS was developed at the <u>CRC for Freshwater Ecology</u> (CRCFE) in Canberra and is based on the British RIVPACS (River InVertebrate Prediction And Classification System) II program (<u>Wright et al. 1993</u>). AUSRIVAS consists of mathematical models that can be tailor made for use in different aquatic habitats and for different times of the year. These models predict the aquatic macroinvertebrate fauna expected to occur at a site in the absence of environmental stress, such as pollution or habitat degradation. The AUSRIVAS predictive system and associated sampling methods offer a number of advantages over traditional assessment techniques. The sampling methods are standardized, easy to perform and require minimal equipment. Rapid turn around of results is possible and the range of outputs from the AUSRIVAS models are tailored for a range of users including community groups, managers and ecologists.

AUSRIVAS models have been developed for each state and territory, for the main habitat types that can be found in Australian river systems. These habitats can include the edge/backwater, main channel, riffle, pool and macrophyte stands. The models can be constructed for a single season or data from several seasons may be combined to provide more robust predictions. To date, the AUSRIVAS predictive system has been developed primarily for lotic environments. Future research and development of the AUSRIVAS system is aimed at widening its scope for use in estuarine and wetland environments.

To complement the bioassessment outputs of AUSRIVAS, a mapping component will be added to AUSRIVAS in 2003. The AUSRIVAS mapping and reference site screening module will enable map based outputs for both AUSRIVAS bioassessment results and catchment impact indices, for use at a range of mapping scales. The <u>Wild Rivers</u> impact database will also be incorporated, enabling a standard framework for reporting and also aid AUSRIVAS reference site screening.

The AUSRIVAS NRHP predictive model manual has four main sections. The first introductory section provides some background on AUSRIVAS. The second section discusses the construction of predictive models and the third section looks specifically at the AUSRIVAS models and how they work. The fourth section covers the use of AUSRIVAS outputs for reporting purposes, how to interpret the outputs and some rules on how to combine outputs from different models to obtain the most accurate site assessment. Detailed model information can be found online in the Appendices section (http://ausrivas.canberra.edu.au/Bioassessment/Macroinvertebrates/Man/Pred/). A description of each State and Territory sampling protocol is available on the AUSRIVAS Website at <u>http://ausrivas.canberra.edu.au/Bioassessment/Macroinvertebrates/</u> under the section on <u>Macroinvertebrates Bioassessment: Manuals & Datasheets</u>.

2. Building predictive models

## 2.1 Collecting the data to build predictive models

'Reference sites' are sites considered to be minimally affected by most human activity and are used to construct a predictive model. 'Test sites' refer to the sites being tested by the model for biological impairment. The test sites may be sites with known or suspected impacts, sites selected for a regional assessment or reference sites revisited and sampled for periodic testing of a model. It is important that all stream and river types, which may be encountered at test sites, have been sampled at sites considered to be equivalent to reference condition. This ensures test sites will be compared against reference conditions that the test sites could be expected to have in the absence of impact. Creating a predictive model involves the selection of a large number of reference sites considered to be minimally impaired (Davies 1994). The macroinvertebrates are sampled and a wide range of environmental characteristics are measured at each site.

Davies (1994) lists the standard set of habitat variables that should be measured at each site. The standard set of habitat variables should now be incorporated into each of the State/Territory AUSRIVAS sampling protocol manuals. To build a predictive model all the variables identified as the potential predictor variables (i.e. those least influenced by human activities) **MUST** be collected at **ALL** sites because the multivariate analysis used in the AUSRIVAS models will not allow missing data. If a habitat variable is missed at a site, either that variable must be deleted from the entire data set or that site must be excluded from the analysis. Alternatively, extrapolations using data from similar sites, means from previous years or a return to the site may be used to fill in the missing data. Missing data are an expensive and wasteful problem easily avoided by taking extra care when sampling. It is important that once sampling and sub-sampling techniques are adopted that they are used consistently. Changing sampling and/or sub-sampling techniques between sampling occasions may result in data unsuitable for comparison with previously collected reference data. Model predictions are only valid for samples collected in the same way as those used to build the model.

The collected invertebrates are identified to family level with the exception of Oligochaeta (Class), Acarina (Order), Collembola (Order), Turbellaria (Order), and Chironomidae (Sub-family). Invertebrate identifications may be taken past family level if desired but at present the AUSRIVAS models only use family level data. A coding system, which incorporates all levels of taxonomic identification, is used for the invertebrates and must be used to allow the models to sort and match data in the various data sets. The complete list of invertebrate codes is available from the Victorian Environment Protection Authority (EPA) and an abbreviated list of taxa codes is also available at http://ausrivas.canberra.edu.au/Bioassessment/Macroinvertebrates/ on the AUSRIVAS Website (select 'Taxonomy' from the left-hand menu). The predictive models on the AUSRIVAS Website only use presence/absence data for predictions, thus, abundance data will be converted to presence/absence data by AUSRIVAS when used in any calculations.

2.2 Invertebrate classification

The first step of the model building process is classifying reference sites into groups, which have similar invertebrate composition, based on family level presence/absence invertebrate data. A number of classification methods can be used to form the reference site groupings. The agglomerative clustering technique, flexible Unweighted Pair-Group arithMetic Averaging (UPGMA) recommended by Belbin (1994) is the most commonly used technique. The flexible component refers to the ability to distort classification space to optimize clustering. The Bray-Curtis association measure is used on the recommendation of Faith et al. (1987) as a robust measure of association for cluster analysis and ordination. The classifications are viewed as dendrograms (Figure 1) allowing the fusion level which divides sites into groups to be selected. If adequate site discrimination is not achieved using UPGMA then either the divisive classification technique TWINSPAN or the non-hierarchical technique ALOC can be used as alternative classification techniques (Belbin 1994). Based on the recommendations of Wright et al. (1993), groups should contain not less than 5 sites. Small classification groups are either deleted from further analysis or those sites are amalgamated with another group of appropriate reference sites. Groups containing less than 5 sites can result from poor representation of a particular type of reference site in the initial sampling, problems with the initial sampling or degradation of sites in some manner resulting in loss of taxa indicative of reference conditions.
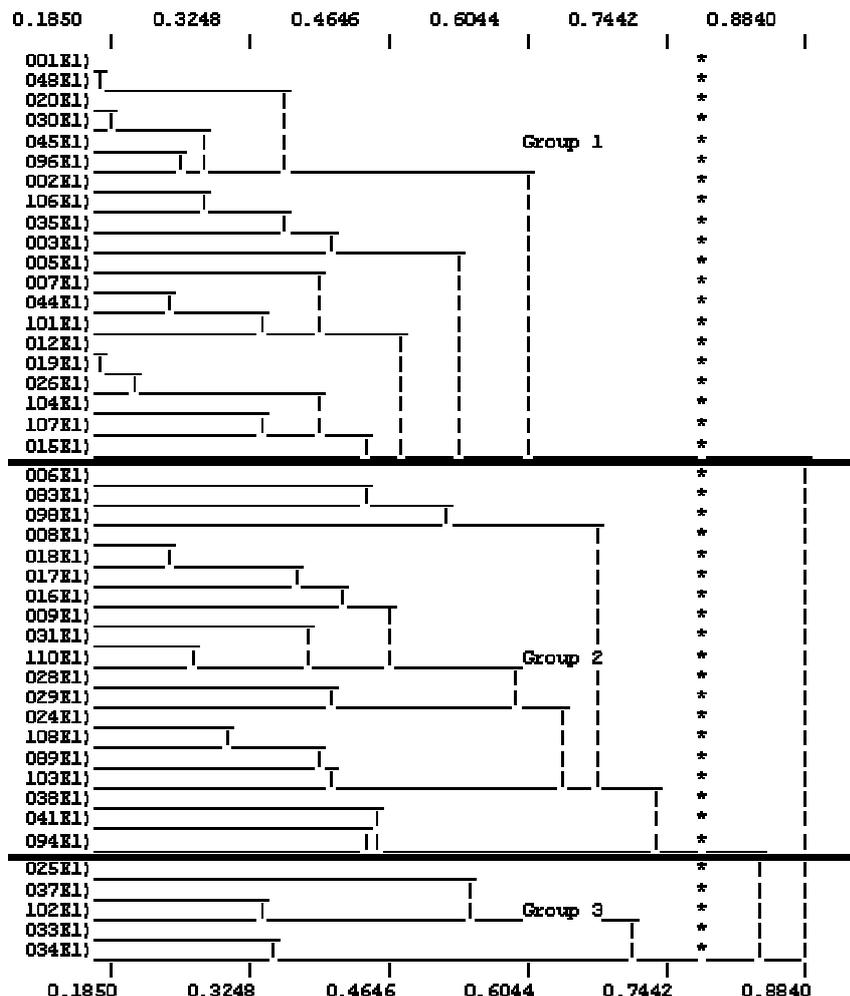
Figure 1. A typical dendrogram formed from a UPGMA classification of reference site invertebrate data.

2.3   Choosing habitat predictor variables

The AUSRIVAS models use habitat features (predictor variables) from a site to predict which taxa should occur at that site in the absence of environmental stress. Habitat variables that may be affected by human activity cannot be used as predictor variables. Variables such as turbidity, dissolved oxygen and phosphorus concentrations are often affected by anthropogenic impacts and would provide spurious predictions if used to predict the membership of test sites to the reference site groups. In contrast, habitat features such as altitude, distance from source and latitude/longitude often make good predictor variables because they are rarely affected by impacts.

The reference site groups from the classification step are entered into the reference habitat data set and a stepwise Multiple Discriminant Function Analysis (MDFA) is used to select the predictor variables used in a model. This procedure selects a subset of habitat variables that best discriminate between the groups of sites formed from the faunal classifications. The stepwise procedure includes habitat variables one at a time, selecting at each step the variables that give the best group discrimination. At each step of the analysis the significance of variables already included is checked and variables that are no longer significant are removed. The significance level for variables to enter and be retained by the Stepwise MDFA are both set at 0.05.

Alternatively, the relationship between the structure of the habitat and invertebrate data can be determined by performing an ordination followed by a Principal Axis Correlation. Semi-strong hybrid multidimensional scaling (SSH, Belbin, 1992) is performed on the invertebrate data with dimensions in ordination space added until an acceptable stress level (< 0.2) is achieved (Belbin 1992). A Monte Carlo simulation (MCSS with 100 permutations, Belbin, 1992) is then performed on the invertebrate ordination to determine the probability of the observed solution having occurred by chance alone (Faith, 1990) The subset of habitat variables, which best describe the position of reference sites in invertebrate ordination space, is then determined using a Principal Axis Correlation (PCC option, Belbin, 1992). Principal Axis Correlation is a multiple-linear regression that generates a correlation value with invertebrate ordination space for each habitat variable (Faith and Norris, 1989). The significance of these correlation values is determined using a Monte Carlo significance test (MCAO with 100 permutations, Belbin, 1992). Only those habitat variables with a significance of 0.01 or better are considered for use a model. This procedure is equivalent to the correlations of the habitat variables with ordination scores used by Wright et al. (1984) in the British RIVPACS model.

Either or both subsets of habitat variables, those from the PCC and those from the Stepwise MDFA, are then tested in a Multiple Discriminant Function Analysis to predict the probabilities of group membership for a reference site. In practice, the variables chosen by the Stepwise MDFA give the best group discriminations. Biased discriminations are avoided by using the cross-validation option, which predicts group membership of each site separately. A subset of habitat variables, which produce the lowest error in predicting the group membership of reference sites, is obtained from this procedure. However, the actual value of the misclassification error is not critical because the probabilities of a site belonging to each group are used for site predictions

rather than the allocation to a single group used by the cross-validation procedure. Thus, sites with an affinity for two or more groups can be misclassified but still provide adequate predictions for a model.

The subset of habitat variables obtained from the stepwise MDFA are used as predictor variables for the AUSRIVAS model under construction. The predictor variables and the reference site invertebrate classification form the foundation of AUSRIVAS, allowing the prediction of taxa at new test sites.

2.4   Checklist of data requirements for the construction of new models

All agencies should check data for the following before sending it to the CRCFE for early model development stages AND if agencies are developing classifications/DFA themselves.

1. All sites must have an entry for each variable (**NO MISSING VALUES**).

2. Check that all data are entered in the same units for a variable.

3. Two data sets are required; 1. the invertebrate file and 2. the habitat data file. Site numbers and codes in the invertebrate file should match exactly the site numbers and codes in the habitat data file i.e. THERE SHOULD BE THE SAME SITES AND THE SAME NUMBER OF SITES IN BOTH FILES.

4. Check for duplicate sites and remove/change code of duplicates (there can only be one entry per site code).

5. AUSRIVAS Macroinvertebrate Predictive Modelling version 3.0 allows alpha-numeric site codes. Earlier versions must have only numeric site codes.

6. Files should be saved as Excel spreadsheets, with sites as rows and invertebrates/habitat variables as columns.

7. Macroinvertebrate data should be entered with their Victorian EPA <u>taxacodes</u>, with only one column of entry per code i.e. Larvae and Adults are to be combined in one total column (NOT SEPARATE). Please supply a list describing the habitat/environmental variables. This MUST include: the UNITS OF MEASUREMENT used and the SCALE/HABITAT at which they were measured.

8. Also, supply a DESCRIPTION OF THE MODEL to be constructed (e.g. NSW SPRING-RIFFLE) and the YEARS over which the reference-site data were collected.

The rules that should be used for combining seasonal data into one dataset are as follows:

1. Both data sets (habitat and invertebrate) need to have been sampled in both seasons. If not, delete that site for that year.

2. Measurements over seasons for all habitat variables should be *averaged* to combine the single season datasets

3. Data for each invertebrate collected over seasons should be *summed* to combine datasets

If the classification stage has been completed, the data required for construction of a model is as follows:

1. Invertebrate data file.

2. Habitat data file.

3. Rare taxa to be removed according to AUSRIVAS protocols prior to classification, that is: taxa occurring at less than 10% of sites if there are less than 100 sites are considered rare, alternatively, if there are more than 100 sites, taxa occurring at less than 10 sites are considered rare.

4. Data should be transformed to presence/absence data prior to classification and the Bray Curtis dissimilarity measure should be used. Classification groups should be formed using flexible UPGMA.

5. Classification groups of less than 5 sites can result from poor representation of a particular type of reference site in the initial sampling or degradation of sites in some manner resulting in loss of taxa indicative of reference conditions. Therefore small classification groups with less than 5 sites should be deleted from further analysis if warranted, or amalgamated with another group of appropriate reference sites.

6. Group membership for each site entered as a column in the habitat file.

7. A copy of the dendrogram (preferably saved in word) with the dissimilarity level at which groups were separated marked on the dendrogram.

8. A file listing the invertebrates used in the classification. To construct this file transpose the column labels from the invertebrate file and put a 1 next to those taxa used in the classification and a 0 next to those taxa not used in the classification.

If the discriminant function analysis has been completed, the data required for construction of a model is as follows:

1. Invertebrate data file.

2. A file listing the invertebrates used in the classification. Transpose the column labels from the invertebrate file and put a 1 next to those taxa used in the classification and a 0 next to those taxa not used in the classification.

3. The habitat data file, including a labeled column, which contains the reference site group-membership of sites. The habitat data file may include only the predictor variables used, or alternatively the complete habitat data file may be sent along with the list of predictor variables you have selected to create the model.

4. The units used and the scale/habitat at which predictor variables were measured.

3. The AUSRIVAS models

3.1 How the AUSRIVAS models work

To use the AUSRIVAS models to assess test sites, the sampling must have been performed in strict accordance with the specific state or territory protocol outlined in the appropriate sampling manual. The test sites must also have been sampled in a season, or seasons, that correspond with the reference site sampling used to construct the model. For example, if a riffle habitat combined-seasons (spring+autumn) model is used, riffle samples from the test site must also have been collected in both spring and autumn. Each of the predictor variables measured at a site has a standardized code, which must be used for data entry to allow the model to recognize and order that variable. The habitat variable codes required for each model can be viewed in the AUSRIVAS program, as described in the "Preparing Data" section of the Predictive modelling software user manual at http://ausrivas.canberra.edu.au/Bioassessment/Macroinvertebrates/Man/User/.

The AUSRIVAS models predict an invertebrate assemblage that is expected to occur at test sites in the absence of impact. A comparison of the invertebrates predicted to occur at the test sites with those actually collected, provides a measure of biological impairment at the tested sites. The predicted taxa list also provides a 'target' invertebrate community to measure the success of any remediation measures taken to rectify identified impacts. The type of taxa predicted by the AUSRIVAS models may also provide clues as to the type of impact a test site is experiencing. This information can be used to facilitate further investigations e.g. the absence of predicted Leptophlebiidae may indicate an impact on a stream from trace metal input.

To obtain a site assessment, the appropriate biological and habitat data from the test site under investigation are entered and preliminary analyses performed to determine whether the test site falls within the experience of that model (see explanation of the Chi2 test in Clarke et al. 1996). Any test sites with no appropriate reference group for comparison are identified at this stage as 'outside the experience of the model'. If the test site passes the validation procedure the probability of that test site belonging to each of the reference site groupings from the faunal classification is calculated, using the habitat predictor variables for that model. The probabilities of group membership are a function of the proximity of a site to each group centroid in multivariate space and the size of that classification group.

The frequency of occurrence of each taxon in each of the reference site classification groups is then calculated. Multiplying a taxon's frequency of occurrence in a classification group by the probability of a test site belonging to that group and summing the results for all of the groups in the classification, gives the probability of a taxon occurring at that test site (Table 1).

Table 1. Calculation of the probability of a taxon occurring at a test site. Combined probability that taxon X will occur at Site Y = 76.5%.

Classification group	Probability that test site Y belongs to group	Frequency of taxon X in group (%)	Contribution to probability that taxon X will occur at Site Y (%)

AUSRIVAS

A	0.59	90	45.00
B	0.30	70	21.00
C	0.15	60	9.00
D	0.05	30	1.50
Total			**76.50**

The AUSRIVAS predictive system only considers taxa that were calculated to have a probability of 50 % or greater of occurring at a test site. These are the actual taxa 'predicted' to occur at a test site with the probability of finding them at any one sampling occasion. The sum of the probabilities of occurrence for these taxa provides the number of taxa 'expected' (E) to be found at a test site.

Taxa with a > 50% chance of occurring at a test site are considered the most useful for detecting a decline in the number of taxa at a test site. Using taxa with a < 0.5 probability of occurrence produces a larger list of predicted taxa than using taxa with a > 0.5 probability of occurrence, however, the number of expected taxa (the sum of the probabilities of those predicted) remains similar (Figure 2). This is because taxa with low probabilities of occurrence contribute little to the total number of expected taxa (Figure 2). Therefore, taxa with a > 0.5 probability of occurrence provide the best compromise between predicting taxa which are unlikely to occur at a site regardless of aquatic health and predicting taxa which will occur even at moderately impacted sites.
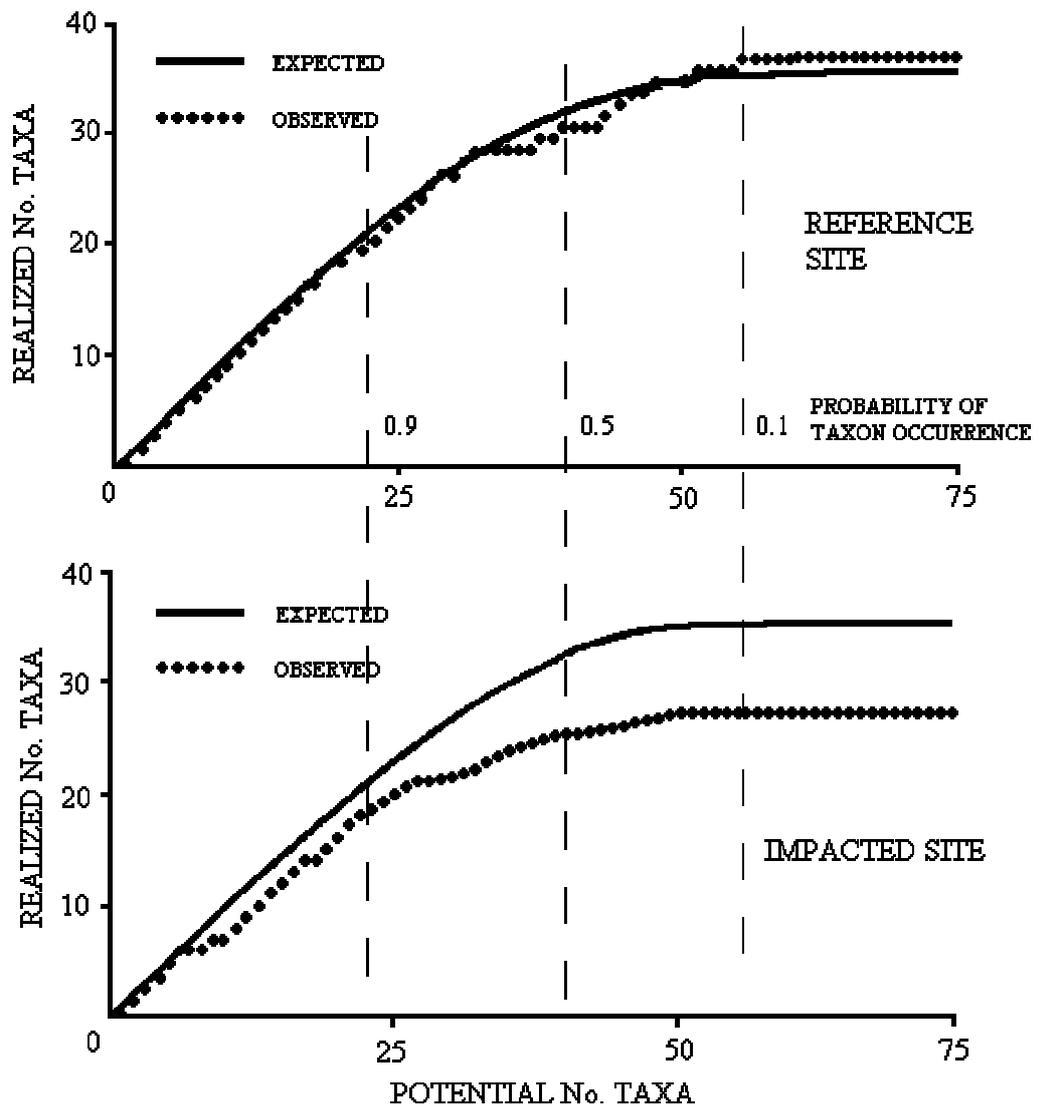
Figure 2. Taxa accretion curves showing expected taxa versus observed taxa for a reference and impacted site and the probabilities of occurrence for these taxa (from R.E.Clark, 1995 unpublished data).

3.2   Signal scores

In addition to calculating the expected number of taxa at a test site, AUSRIVAS also calculates the expected *SIGNAL score* for a site. Calculation of SIGNAL scores uses *SIGNAL grades* (Chessman, 1995), a system that assigns a value to each invertebrate family based on its sensitivity to pollution. A grade of 10 represents high sensitivity to pollution, while a grade of 1 represents high tolerance to pollution.

Each of the calculated SIGNAL scores is explained below:

E50Signal

E50Signal is the expected signal score for taxa that have a probability of occurrence of greater than or equal to 50%. It is calculated by weighting the probability of occurrence of each **predicted** taxa (those taxa that have a probability of occurrence of greater than or equal to 50%) by the taxon's SIGNAL Grade, summing these and then dividing the total by the sum of the (unweighted) probabilities of occurrence.

$$E_{50}Signal = \frac{\sum_{i=1}^{N_{P(50)}} P(Taxa_i) \times SignalGrade_i}{\sum_{i=1}^{N_{P(50)}} P(Taxa_i)}$$

O50Signal

O50Signal is the observed signal score for taxa that have a probability of occurrence of greater than or equal to 50%. It is calculated by averaging the SIGNAL Grade's for all observed taxa with P(Taxa) >= 0.5.

$$O_{50}Signal = \frac{\sum_{j=1}^{N_{O(50)}} SignalGrade_j}{N_O(50)}$$

OE50Signal

The observed to expected SIGNAL ratio, OE50Signal, is the ratio of E50Signal to O50Signal.

E0Signal

E0Signal is calculated the same way as E50Signal, except all taxa that have a probability of occurrence of greater than 0% are included in the calculation.

$$E_0 Signal = \frac{\sum_{i=1}^{N_{P(0)}} P(Taxa_i) \times SignalGrade_i}{\sum_{i=1}^{N_{P(0)}} P(Taxa_i)}$$

O0Signal

O0Signal is the observed signal score for taxa that have a probability of occurrence (*P(Taxa)*) of greater than 0%. It is calculated by averaging the SIGNAL Grade's for all observed taxa. O0Signal is equivalent to the 'raw' SIGNAL score (Chessman, 1995).

$$O_0Signal = \frac{\sum_{j=1}^{N_0} SignalGrade_j}{N_0}$$

OE0Signal

The observed to expected SIGNAL ratio, OE0Signal, is the ratio of E0Signal to O0Signal.

An example of calculating the expected signal score E50Signal is shown in Table 2.

Table 2. Calculation of expected number of taxa and expected SIGNAL score.

Family	SIGNAL grade	Taxa probability of occurrence	Weighted SIGNAL grade
Aeshnidae	6	0.4	taxon probability <0.5
Baetidae	5	0.8	4
Hydridae	4	0.1	taxon probability <0.5
Isostictidae	7	0.9	6.3
Kokiriidae	10	0.8	8
Leptoceridae	8	0.9	7.2
Notonectidae	3	0.7	2.1
Scirtidae	9	0.5	4.5
Expected no. of taxa:		4.6	
Expected SIGNAL score (E50Signal):			6.98

3.3 Bioassessment indices

AUSRIVAS compares both the expected (E) number of taxa and the expected SIGNAL score against what taxa were actually observed (O) at a test site. This provides two indices, which provide a measure of biological impairment at a test site. These are:

O/E Taxa	This is the ratio of the number of invertebrate families observed at a site to the number of families expected at that site.

O/E SIGNAL	This is the ratio of the observed SIGNAL score for a site to the expected SIGNAL score. The calculation of the observed and expected SIGNAL scores is shown in Section 3.2.

The values of both indices can range from a minimum of 0 (indicating that none of the families expected at a site were actually found at that site) to a theoretical maximum of 1.0, indicating a perfect match between the families expected and those that were found. In practice, this maximum can exceed 1.0 indicating that more families were found at that site than were predicted by the model. This can indicate an unusually biologically diverse site, but could also indicate mild enrichment by organic pollution where the added nutrients have allowed families not normally found in that site to establish. Conversely, an unimpacted, high-quality site may score an index value less than 1.0 because of chance exclusions of families during sampling.

3.4   The banding scheme

To simplify interpretation and to aid management decisions, O/E taxa can be divided into bands representing different levels of biological condition. (At present, no bands are calculated for O/E SIGNAL values). The width of the bands is based on the distribution of O/E taxa values for the reference sites in each particular model. The width of the reference band, labeled A in Table 3, is centered on the O/E taxa value of 1.0 and includes the central 80% of the reference site O/E taxa values for each particular model. A test site whose index value exceeds the upper bound of these values (i.e. the index value is greater than the 90th percentile of the reference sites) is judged to be richer than the reference condition and is allocated to "band X". A test site whose index value falls below the lower bound (i.e. the index value is smaller than the lower 10th percentile of the reference site O/E taxa values) is judged to have fewer families than expected and is allocated to one of the lower bands according to its value. The width of bands B and C are the same as for band A, the reference band. Band D may be narrower than these bands depending on the variability in the O/E taxa values of the reference sites in the particular model. In most cases, sites falling in band D will be severely impaired and have few of the families expected at the site.

Table 3. Division of O/E taxa into bands or categories for reporting. The names of the bands refer to the relationship of the index value to the reference condition (band A). Under comments for each index, an explanation of the band is stated first, followed by possible interpretations.

Band	Description	O/E taxa	O/E taxa interpretations
X	MORE BIOLOGICALLY DIVERSE THAN REFERENCE	O/E greater than 90th percentile of reference sites used to create the model.	More families found than expected. Potential biodiversity "hot-spot" or mild organic enrichment. Continuous irrigation flow in a normally intermittent stream.
A	SIMILAR TO REFERENCE	O/E within range of central 80% of reference sites used to create the model.	Expected number of families within the range found at 80% of the reference sites.
B	SIGNIFICANTLY IMPAIRED	O/E below 10th percentile of reference sites used to create the model. Same width as band A.	Fewer families than expected. Potential impact either on water and/or habitat quality resulting in a loss of families.
C	SEVERELY IMPAIRED	O/E below band B. Same width as band A.	Many fewer families than expected. Loss of families from substantial impairment of expected biota caused by water and/or habitat quality.
D	EXTREMELY IMPAIRED	O/E below band C down to zero.	Few of the expected families and only the hardy, pollution tolerant families remain. Severe impairment.

4. Use of AUSRIVAS for reporting

4.1 Introduction

For a given test site, the most comprehensive assessment of its biological condition will be based on data collected from two habitats in two seasons. This should yield the most comprehensive list of families found at the site and, in general, the fairest assessment of the status of the site relative to the reference conditions. In this situation there are potentially eight index values to be synthesized for final reporting. The following procedure is to be used for reporting such data.

4.2 Choosing between single-season or combined-seasons model

The procedure first requires a choice of the most appropriate seasonal model for computing the indices. Usually, data from both seasons for a particular habitat will be used in the combined seasons model for that habitat. If resources or time only allow one sampling occasion, then the seasonally more appropriate model will be used. In some circumstances, it may be of interest whether the status of a site has changed between sampling occasions. For example, a human disturbance such as a waste spillage or river restoration may have occurred between the sampling occasions. In that case the appropriate single-season model will be applied to each sampling occasion.

4.3 Combining assessments from different habitats

In some circumstances only one habitat will be assessed, either because other habitats are not present or because the investigation is targeted at only one habitat (such as the habitat deemed most susceptible to the disturbance of concern). In such cases, the fact that only one habitat has been employed in the assessment should be made explicit.

In many cases two bandings will be available for a given index and test site: one for each habitat. Where the bandings from both habitats allocate the site to the same band, then that is the final band allocation for the site. Where there is a mis-match in the band allocation from the two habitats, then allocate the site to the band that is farther from band A. In the rare event that the alternative bands are band B and band X, allocate to band B, because this is the most precautionary approach. Allocation to Band X should result in further assessment to determine whether the site is richer than reference because of naturally high biodiversity or an impact such as mild nutrient enrichment.

When combining assessments from different habitats, the above rules should be used to decide the final band allocation. However, in interpreting the results, the following factors should also be considered before accepting that the site is impaired:

1. Were there **unusual circumstances,** which may have affected the results? For example, severe weather conditions or accidents when sampling, sorting or typographic errors, which may have biased the data collected.

2. **Were there unusual habitat conditions**, such as a lack of vegetation in the edge habitat or domination of one habitat by bedrock, which may reduce the availability of macroinvertebrate habitat?

3. **Any periods of unusually high or low flow** before sampling, which may have affected the number and type of taxa collected. Such disturbances may impact some habitats more than others and this impact may differ between sites. For example, a river with plentiful macrophytes in the edge habitat may be unaffected by a period of low flows, however the riffle habitat may be quite reduced and consequently be assessed as impacted. Alternatively, at a different site low flows may lead to a severe reduction in edge habitat, but a substantial increase in available riffle habitat.

If the preceding steps indicate problematic data, then there should be **no post hoc alteration or 'correction' of the faunal or environmental data beyond typographic, data-entry errors**. The integrity of the data is paramount and alterations such as deletions of families whose presence is 'explained away' cannot be tolerated.

The options for further action in order of preference are as follows.

1. **Re-sampling and re-assessment**. This is the obvious choice if time and resources permit re-sampling. If this is not possible, then

2. **Draw a conclusion of 'no reliable assessment possible'**. This is the most conservative approach. Diagnostic information can still be presented, explained and qualified, but no allocation should be made to a band. The reasons that no reliable assessment could be made should be made explicit.

4.4   Assessments of multiple sites

Assessments of several sites simultaneously should use a consistent basis for comparison; mixing assessments based on different seasonal models or mixtures of single and two-habitat data should be discouraged.

Further information

Construction of a new model

Contact details

◆ Richard Norris (norris@lake.canberra.edu.au),

◆ Sue Nichols (nichols@lake.canberra.edu.au).

For information on coverage of models

Sources of information

◆ Models lists at:
http://ausrivas.canberra.edu.au/Bioassessment/Macroinvertebrates/index_ModelSelect.html

Contact details

◆ Contact your State or Territory NRHP Representative
(http://ausrivas.canberra.edu.au/nrhp_contacts.html).

Manuals

On-line or downloadable manuals

◆ Software user manual
http://ausrivas.canberra.edu.au/Bioassessment/Macroinvertebrates/Man/User/

◆ This Predictive modelling manual is also available online at
http://ausrivas.canberra.edu.au/Bioassessment/Macroinvertebrates/Man/Pred/

◆ Sampling manuals available at
http://ausrivas.canberra.edu.au/Bioassessment/Macroinvertebrates/

Contact details

◆ Sue Nichols (nichols@lake.canberra.edu.au),

◆ Gail Ransom (gail@lake.canberra.edu.au).

References

Belbin, L. and McDonald, C. (1993). Comparing three classification strategies for use in ecology. *Journal of Vegetation Science* **4**, 341-348.

Belbin, L. (1994). PATN, Technical Reference. Commonwealth Scientific and Industrial Research Organisation, Division of Wildlife and Ecology: Canberra, Australia.

Chessman, B.C. (1995). Rapid river assessment using macroinvertebrates: a procedure based on habitat specific sampling, family level identification and a biotic index. *Australian Journal of Ecology* **20**, 122-129.

Clarke, R.T., Furse, M.T., Wright, J.F. and Moss, D. (1996). Derivation of a biological quality index for river sites: comparison of the observed with the expected fauna. *Journal of Applied Statistics* **23**, 311-332.

Davies, P.E. (ed) (1994). Monitoring river health initiative: River bioassessment manual. National river processes and management program, Freshwater Systems (Tasmania: Australia.)

Faith, D.P. (1990). Benthic macroinvertebrates in biological surveillance: Monte Carlo significance tests on functional groups' responses to environmental gradients. *Environmental Monitoring and Assessment* **14**, 247-264.

Faith, D.P., Minchin, P.R. and Belbin, L. (1987). Compositional dissimilarity as a robust measure of ecological distance. *Vegetatio* **69**, 57-68.

Faith, D.P. and Norris, R.H. (1989). Correlation of environmental variables with patterns of distribution and abundance of common and rare freshwater macroinvertebrates. *Biological Conservation* **50**, 77-98.

Institute of Freshwater Ecology (1991). Testing and further development of RIVPACS. Interim Report No. R&D Project 243. (National Rivers Authority: Bristol, U.K.)

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

Stein, J.L., Stein, J.A. and Nix, H.A. (1998). The Identification of Wild Rivers: Methodology and database development. (Centre for Resource and Environmental Studies: Australian National University, Canberra.)

Wright, J.F., Furse, M.T., Armitage, P.D., and Moss, D. (1993). New procedures for identifying running-water sites subject to environmental stress and for evaluating sites for conservation, based on the macroinvertebrate fauna. *Archiv für Hydrobiologi* **127**, 319-326.

Wright, J.F., Furse, M.T. and Armitage, P.D. (1994). Use of macroinvertebrate communities to detect environmental stress in running waters. In 'Water quality and stress indicators in marine and freshwater ecosystems: linking levels of organisation - individuals, populations, communities.' (Eds D.W. Sutcliffe.) pp. 15-34. (Freshwater Biological Association: Ambleside, U.K.)