

## APPENDIX A: Data requirement checklist for the construction of AUSRIVAS models

JULIE COYSH AND SUE NICHOLS  
CRC FOR FRESHWATER ECOLOGY, CANBERRA

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

- All sites must have an entry for each variable (NO MISSING VALUES)
- Check that all data are entered in the same units for a variable
- Site numbers in the bug file should match exactly the site numbers in the chem file i.e. THERE SHOULD BE THE SAME SITES AND THE SAME NUMBER OF SITES IN BOTH FILES.
- Check for duplicate sites and remove/change code of duplicates (there can only be one entry per site code)
- The format of site codes must be numeric at present, if you have alpha numeric site codes please also give each a numeric site code.
- Files should be saved as Excel spreadsheets, with sites as rows and bugs/habitat variables as columns
- Macroinvertebrate data should be entered with their Victorian EPA taxacodes, with only one column of entry per code i.e. Larvae and Adults are to be combined in one total column (NOT SEPARATE)
- The rules that should be used for combining seasonal data into one dataset are as follows:
  - both data sets (chem and bugs) need to have been sampled in both seasons. If not, delete that site for that year.
  - Measurements over seasons for all chem/habitat variables should be averaged to combine datasets
  - All bugs collected over seasons should be summed to combine datasets

**When sending us the results of the classification, we require:**

- Bug data file
- Chem data file
- A copy of the dendrogram (preferably saved in word) with the dissimilarity level at which groups were separated marked on the dendrogram
- Group membership for each site entered as a column in the chem file.
- Rare taxa to be removed according to AUSRIVAS protocols prior to classification, that is:  
taxa occurring at less 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.
- 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.
- 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.
- A file listing the bugs used in the classification. To construct this file transpose the column labels from the bug 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.

**When sending us the results of the discriminant function analysis we require:**

- Bug data file
- A file of bugs used in the classification. Transpose the column labels from the bug 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.
- The chem file, including a labelled column which containing the reference site group membership of sites
- The chem file which includes only the predictor variables used may be sent, or alternatively the complete chem file may be sent along with a list of predictor variables you would like us to use to create the model.
- For predictor variables used, we require the units used and the scale/habitat at which they were measured.

**After constructing a preliminary model and running the reference site data through it we will identify those sites that have a low O/E taxa ratio (<0.75) and will ask you to review these sites.**

**To review reference sites with observed/expected values less than 0.75 the following should be considered:**

- the sampler and sampling effort ( inclement weather conditions on the day of sampling of untrained samplers can affect sampling effort).
- Flood or drought conditions near the time of sampling
- Potential sources of impacts