

**National River Health Program**  
**Support and Completion of Australia Wide Assessment of River  
Health Models- Phase 2**

---

**FINAL REPORT**

**November 2002**

**CRC for Freshwater Ecology**

**Julie Coysh  
Amara Barlow  
Sue Nichols  
Richard Norris  
Peter Liston**

**Building 15 University of Canberra ACT 2601**



## TABLE OF CONTENTS

Project Details	3
Introduction	4
Data formatting	5
Analyses	5
Rare Taxa	5
Classification	5
Discriminant Function Analysis	6
Validating reference sites	11
Achievement of Objectives	13
Summary of work completed	29
Problems and delays	30
Communication	30
Acknowledgments	30
References	31
APPENDIX A: Data requirement checklist for the construction of AUSRIVAS models	32
APPENDIX B: Sites used in the model creation and their group membership as determined by the invertebrate classifications	35
APPENDIX C: Numbers of sites rejected as reference sites for model development	122
APPENDIX D: Queensland sites rejected as reference sites for model development and the rationale for excluding them	123

## **Project Details**

### **Project Title:**

Support and Completion of Australia Wide Assessment of River Health Models Phase 2

### **Project Staff:**

Leader: A/Prof. Richard Norris

Research Assistants: Julie Coysh, Amara Barlow, Sue Nichols and Peter Liston

### **Project Duration:**

March 2001 to October 2002.

### **Milestones:**

#### ***Milestone 1:***

- Project Commencement

#### ***Milestone 2:***

- Completion of all alpha level models for NSW, QLD, SA, and WA including: undertaking checking and matching of data files, model validation, model error checks, construction of model bands, AUSRIVAS coding of variables and taxa
- Review and endorsement of all models by Steering Committee
- Mounting final alpha level models on WWW site and software platform, with supporting documentation
- Provide assistance and training in statistical analysis as required by lead agencies.
- Progress report to be lodged with EA

#### ***Milestone 3:***

- Final Report to be lodged with EA within 30 days of project completion

The due date for Milestones was extended by EA after delays to the completion of the Queensland models.

All models completed in the second stage of AUSRIVAS model development have been approved and endorsed by appropriate state agencies and the national steering committee. The role of that steering committee was to oversee the completion of the project including each milestone to the satisfaction of the commonwealth, and endorse the project outputs. The steering committee members during the course of the project were: Natural Heritage Trust Program Manager for NRHP (Chair): Peter Komidar, Environment Australia; Senior Project Officer for NRHP: Bruce Gray, Environment Australia; Project Officer: Lisa Evans, Environment Australia; NRHP Science Coordinator: Peter Davies, Freshwater Systems Pty Ltd; CRCFE Principle Investigator: Richard Norris; CRCFE Research Assistant: Julie Coysh; CRCFE Research Assistant: Amara Barlow; State / Territory Agency representative : Satish Choy, Qld DNRM; State / Territory Agency representative : Eren Turak, NSW EPA; State / Territory Agency representative: Leon Metzeling, EPA Victoria; State / Territory Agency representative: Stuart Halse, WA CALM; and Independent Science Representative (external to NRHP): Richard Marchant, Museum of Victoria.

### **Project Objectives:**

- 1) To complete the second iteration of the AUSRIVAS models for all states and territories, taking into account:
  - a) requirements for replacement or removal of poor quality biological and environmental data
  - b) further additions and/or removals of reference sites, in consultation with the lead agencies
  - c) further considerations with regard to site classification such as the numbers of groups selected
- 2) All models are to be revised and approved at an alpha level
- 3) To mount all revised alpha level models on the AUSRIVAS WWW site at the University of Canberra, with all necessary supporting documentation including taxonomic codes
- 4) To provide statistical advice to National River Health Program lead agencies on matters associated with data entry, site classification, discriminant function analysis and model construction.

### **Introduction**

Experience had led to a much better knowledge of the effects of data quality on the AUSRIVAS models. This project was designed to enable a second iteration of all the state/territory AUSRIVAS models to allow the incorporation of improvements in data quality and reference site selection. In addition, state and territory agencies have now had an opportunity to review existing reference sites and sample new reference sites, requiring the construction of new models. All states and territories were contacted when the project commenced and advised that they had the opportunity to have their models recreated. The first states/territories to supply the Cooperative Research Centre for Freshwater Ecology (CRCFE) a complete set of six data sets were the Australian Capital Territory, Victoria and Tasmania, therefore these states and territories were

included in Phase 1 of this project. In this phase (Phase 2) alpha models have been created for New South Wales, Queensland, South Australia, and Western Australia.

### **Data formatting**

Agencies requiring the CRCFE to complete all components of the analysis were requested to provide all relevant physical/chemical data and biological data for reference sites to be included in the model in Excel spreadsheet format.

Habitat data for combined season models were averaged for two sampling seasons in a year. If a site was not sampled in both seasons it was unsuitable for inclusion in the dataset for a combined season model. Invertebrate data for two sampling seasons per year were summed, increasing the taxon list for a site and generally providing more robust model predictions. Sites sampled in consecutive years were treated as separate sites to incorporate the year to year variation in the models. Invertebrate and habitat data files were checked for missing data and inappropriate variables (e.g. larvae and adult entries for the one taxon). Invertebrate data were then transformed to presence/absence data before analysis because the models only use presence/absence data.

QLD had all data checking and analysis performed by the CRCFE, SA provided initial classifications, NSW and WA built their own models and performed all data checking and analysis themselves. All states were provided with the guidelines in Appendix A for data checking and analysis. Thus, while some states performed parts of their own analyses, the methods were standardized to the following procedures.

### **Analyses**

#### **Rare Taxa**

Rare taxa were removed from invertebrate data sets before analysis. For AUSRIVAS, rare taxa are considered those that occur at less than 10% of sites for data sets of 100 sites or less, or taxa that occur at less than 10 sites for data sets of more than 100 sites. In some cases more taxa were removed from the Queensland regional models, to account for the patchy distribution of many taxa. The only other exception was the Western Australian models, where taxa at less than 5 sites were removed.

#### **Classification**

The first step in the model building process was to classify reference sites that had similar invertebrate compositions, based on family level presence/absence data, using the PATN multivariate analysis package (Belbin 1994). The Bray-Curtis association measure was used on the recommendation of Faith *et al.* (1987) as a robust measure of association for cluster analysis and ordination. Groups were selected based on the dissimilarity level of the clusters, the tightness of the reference site clusters and how well agency staff thought the groups represented a 'type' of reference site. 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 and McDonald (1993) was the most commonly used technique. The classifications were viewed as dendrograms allowing the fusion level, which divides sites into groups, to be selected. Based on the recommendations of Wright *et al.* (1993) groups should contain not less than 5 sites. Small classification groups were either deleted from further analysis or these sites were amalgamated with another group of appropriate reference sites. The list of sites used in all the invertebrate classifications and their final group membership for these sites are listed in Appendix B.

### **Discriminant Function Analysis**

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 damage. Habitat variables that are commonly affected by human induced impacts should not be used as predictor variables. Variables such as turbidity, dissolved oxygen and phosphorus concentrations are often affected by human impacts and might 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, latitude and longitude often make good predictor variables because they are rarely affected by human impacts.

The reference site groups from the UPGMA classification were entered as a variable into the reference habitat data set and a Stepwise Multiple Discriminant Function Analysis (MDFA) performed to select the predictor variables used in an AUSRIVAS model. The Stepwise MDFA was performed in the SAS statistical package version 6.12 (SAS Institute 1995). This procedure selected a subset of habitat variables that best discriminated among the groups of sites formed from the faunal classifications. The stepwise procedure included habitat variables one at a time, selecting at each step the variables that gave the best group discrimination. At each step of the analysis the significance of variables already included were checked and variables that were no longer significant were removed. The significance levels for variables to enter and be retained by the Stepwise MDFA both were set at 0.05.

The subsets of habitat variables from the Stepwise MDFA were then tested in a MDFA to predict the probabilities of group membership for a reference site. Biased discriminations were avoided by using the cross-validation option that predicts group membership of each site separately. A subset of habitat variables that produced the lowest error in predicting the group membership of reference sites was obtained from this procedure. However, the actual value of the misclassification error is not critical because AUSRIVAS uses all the probabilities of a site belonging to each group for site predictions, rather than the allocation to a single group as performed 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 subsets of habitat variables obtained from the stepwise MDFA were 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 predictions of the taxa expected at new “test” sites. The list of predictor variables, their descriptions and the misclassification errors for each of the agency models constructed under this contract are provided in Tables 1 and 2.

**Table 1: Predictor variables and misclassification errors for state and territory models.**

Model	Cross-validation Misclassification Rate (%)	Resubstitution Misclassification Rate (%)	Predictor Variables
NSW Autumn Edge	13.71	10.93	ALKALINITY, ALTITUDE, BEDROCK, BOULDER, COBBLE, LATITUDE, LOGDFSM, LOGSLOPE1KUS, LONGITUDE, RAINFALL
NSW Autumn Riffle	17.15	14.03	ALTITUDE, LATITUDE, LOGDFSM, LOGSLOPE1KUS, LONGITUDE, RAINFALL
NSW Combined Eastern Edge	22.85	15.42	ALKALINITY, ALTITUDE, BEDROCK, COBBLE, LATITUDE, LOGDFSM, LOGMODEWIDTH, LOGSLOPE1KUS, LONGITUDE, RAINFALL
NSW Combined Riffle	9.92	8.42	ALTITUDE, LATITUDE, LOGDFSM, LOGMODEDEPTH, LOGMODEWIDTH, LOGSLOPE1KUS, LONGITUDE, RAINFALL
NSW Combined Western Edge	19.78	17.40	ALTITUDE, LATITUDE, LONGITUDE, SLOPE1KUS
NSW Spring Edge	12.97	10.21	ALKALINITY, ALTITUDE, BEDROCK, BOULDER, COBBLE, LATITUDE, LOGDFSM, LOGMODEWIDTH, LOGSLOPE1KUS, LONGITUDE, RAINFALL
NSW Spring Riffle	25.36	24.35	ALTITUDE, LATITUDE, LOGDFSM, LOGSLOPE1KUS, LONGITUDE, RAINFALL
QLD Coast Autumn Edge	37.32	36.79	ALTITUDE, BEDSLOPE, LONGITUDE, MDMR, RAINRANGE, RAWD, WETPERCENT
QLD Coast Autumn Pool	39.62	37.10	COBBLE, DFS, MDMR, MINTEMP, RAINFALL
QLD Coast Autumn Riffle/Run	45.05	43.83	BEDROCK, COBBLE, DFS, LONGITUDE, MAXTEMP, MWMR, PEBBLE
QLD Coast Combined Edge	34.85	33.50	BEDROCK, BEDSLOPE, LONGITUDE, MAXTEMP, MDMR, PROCESSZONE, RAINFALL, RAWD
QLD Coast Combined Pool	33.90	30.39	COBBLE, MDMR, MINTEMP, MWMR, PROCESSZONE, RAINFALL, STORDER
QLD Coast Combined Riffle/Run	32.33	30.03	ALTITUDE, BEDROCK, COBBLE, DFS, DRYRANGE, LONGITUDE, MWMR, RAINRANGE
QLD Coast Spring Edge	41.55	40.56	ALTITUDE, LONGITUDE, MDMR, MINTEMP, RAWD, WETR
QLD Coast Spring Pool	27.54	25.39	COBBLE, MDMR, PROCESSZONE, RAINRANGE, RANGETEMP
QLD Coast Spring Riffle/Run	40.47	39.26	ALTITUDE, COBBLE, LONGITUDE, MAXTEMP, MWMR, PROCESSZONE, WETPERCENT
QLD West Autumn Edge	38.96	36.21	DRYRANGE, LATITUDE, MINTEMP, MWMR, STORDER, WETPERCENT
QLD West Autumn Pool	37.15	35.31	PROCESSZONE, STORDER, WETR
QLD West Autumn Riffle/Run	35.70	27.24	ALTITUDE, BEDROCK, BEDSLOPE, BOULDER, COBBLE, DRYRANGE, LATITUDE, LONGITUDE, RAINFALL
QLD West Combined Edge	27.74	25.42	COBBLE, DRYRANGE, LATITUDE, LONGITUDE, PROCESSZONE, RAWD, WETPERCENT, WETR
QLD West Combined Pool	21.29	20.91	BEDSLOPE, COBBLE, DRYRANGE, LATITUDE, MDMR, MINTEMP, MWMR, RAINFALL, RANGETEMP, RAWD, WETPERCENT

Table 1 Continued.

Model	Cross-validation Misclassification Rate (%)	Resubstitution Misclassification Rate (%)	Predictor Variables
QLD West Combined Riffle/Run	16.67	9.88	ALTITUDE, COBBLE, DFS, DRYRANGE, LATITUDE, LONGITUDE, MAXTEMP, MDMR, PEBBLE, PROCESSZONE, RAWD, STORDER, WETPERCENT, WETR
QLD West Spring Edge	40.19	38.75	ALKALINITY, DFS, LATITUDE, LONGITUDE, PROCESSZONE
QLD West Spring Pool	42.53	42.53	BOULDER, COBBLE, WETR
QLD West Spring Riffle/Run	28.35	21.74	BEDROCK, COBBLE, DFS, DRYRANGE, LONGITUDE, MAXTEMP, MDMR, MINTEMP, PROCESSZONE, RAINFALL, WETR
SA Autumn Edge	30.87	25.93	ALKALINITY, BICARBPOW0.4, EDGEAREA, LATITUDE, LOGCOND, LOGDEPTHCMCAT, LONGITUDE, SHADINGPERCENT
SA Autumn Riffle	28.14	23.36	BICARBONATES, CATCHAREAPOW-0.1, LATITUDE, LOG10BEDROCK, LOG10BOULDER, LOG10COND, LOG10STORDER, LONGITUDE
SA Combined Edge	28.08	25.20	ALTITUDEPOW0.5, CATCHAREAPOW-0.2, DFSPOW-0.4, EDGEAREA, LATITUDE, LOG10COND, LOG10DEPTHCMCAT, LOG10SLOPE1KUS, LOG10STORDER, LONGITUDE, SHADPERCENTPOW0.8
SA Combined Riffle	10.54	9.15	ALTITUDEPOW0.5, CATCHAREAPOW-0.2, DFSPOW-0.3, LATITUDE, LOG10COND, LOG10SLOPE1KUS, LOG10STORDER, LONGITUDE, PEBBLEPOW0.5, RIFFAREAPOW0.5, SHADINGPERCENTPOW0.5
SA Spring Edge	26.99	24.48	ALTITUDEPOW0.4, CATCHAREAPOW-0.2, LATITUDE, LOG10BEDROCK, LOG10BOULDER, LOG10COND, LOG10DEPTHCMCAT, LOG10DFS, LOG10GRAVEL, LOG10STORDER, LONGITUDE
SA Spring Riffle	16.31	13.39	ALTITUDEPOW0.5, CATCHAREAPOW-0.2, DFSPOW-0.3, LATITUDE, LOG10COBBLE, LOG10COND, LOG10STORDER, LONGITUDE, MEDSEDP0.7, RIFFAREAPOW0.8, SHADPERCENTPOW0.5, SLOPE1KUSPOW0.2, STWIDTHPOW-0.3
WA Autumn Channel	21.71	16.07	LATITUDE, LOGSTWIDTH, LONGITUDE, PARTSIZE, RAINFALL
WA Combined Channel	29.08	23.52	LATITUDE, LOGALTITUDE, LOGMAXVELOCITYCM, LOGRAINFALL, LONGITUDE
WA Spring Channel	23.85	18.91	DISCHARGECA, LATITUDE, LOGMAXVELOCITYCM, LONGITUDE, RAINFALL

**Table 2: AUSRIVAS codes, description and units of predictor variables used for new models created under this contract.**

<b>AUSRIVAS Code</b>	<b>Description and Units</b>
ALKALINITY	Total carbonates (mg/l)
ALTITUDE	Height above sea level (m)
ALTITUDEPOW0.4	Altitude (m) to the power 0.4
ALTITUDEPOW0.5	Altitude (m) to the power 0.5
BEDROCK	Percent bedrock in habitat
BEDSLOPE	Slope (m/m) of bed
BICARBONATES	Bicarbonates (mg/l)
BICARBPOW0.4	Bicarbonates (mg/l) to the power 0.4
BOULDER	Percent boulder (>256mm) in habitat
CATCHAREAPOW-0.1	Catchment area upstream of the sampled site (KM2) to the power -0.1
CATCHAREAPOW-0.2	Catchment area upstream of the sampled site (KM2) to the power -0.2
COBBLE	Percent cobble (64-256mm) in habitat
DFS	Distance from source (km)
DFSPOW-0.3	Distance From Source (km) to the power -0.3
DFSPOW-0.4	Distance from Source (km) to the power -0.4
DISCHARGE CAT	Mean annual discharge, logarithmic categories 1-6 (WA)
DRYRANGE	Range in dry season monthly rainfall means (mm)
EDGEAREA	Percent edge habitat in reach
LATITUDE	Latitude - Decimal Degrees
LOG10BOULDER	Log 10 (x+1) % boulder in habitat
LOG10COBBLE	Log 10 (x+1) % cobble in habitat
LOG10COND	Log 10 (x+1) Conductivity
LOG10DEPTHCMCAT	Log 10 (x+1) Mean depth category of habitat (cm)
LOG10DFS	Log 10 (x+1) Distance from source (km)
LOG10GRAVEL	Log 10 (x+1) % gravel in habitat
LOG10SLOPE1KUS	Log 10 (x+1) Slope - Diff in elevation (m) b/w the site and 1km u/s
LOG10STORDER	Log 10 (x+1) Stream Order
LOGALTITUDE	Log 10 - Altitude (m)
LOGBEDROCK	Log 10 (x+1) - Percentage bedrock in reach
LOGDFSM	Log 10 - Distance from source (m)
LOGMAXVELOCITYCM	Log 10 - Maximum flow velocity in habitat (cm/sec)
LOGMODEWIDTH	Log 10 - Mode stream width (m)
LOGRAINFALL	Log 10 - Mean annual rainfall (mm)
LOGSLOPE1KUS	Log 10 - Diff between the site and a point 1km upstream
LOGSTWIDTH	Log 10 - Mean width of water (m)
LONGITUDE	Longitude - Decimal Degrees
MAXTEMP	Mean daily max temp (°C)
MDMR	Mean dry season monthly rainfall (mm)
MEDSEDPOW0.7	% medium sediment (% pebble + % gravel) to the power 0.7
MINTEMP	Mean daily min temp (°C)
MWMR	Mean wet season monthly rainfall (mm)
PARTSIZE	Substrate particle size index (WA)
PEBBLE	Percent pebble (16-64mm) in habitat
PEBBLEPOW0.5	% Pebble in habitat to the power 0.5
PROCESSZONE	Process zone: erosional=2; transport=1; depositional=0 (category)

**Table 2 Continued.**

<b>AUSRIVAS Code</b>	<b>Description and Units</b>
RAINFALL	Mean annual rainfall (mm)
RAINRANGE	Annual range in mean monthly rainfall (mm)
RANGETEMP	Mean daily temp range ( <sup>0</sup> C)
RAWD	Ratio mean wet season / mean dry season rainfall
RIFFAREAPOW0.5	% area of riffle habitat at the site to the power 0.5
RIFFAREAPOW0.8	% area of riffle habitat at the site to the power 0.8
SHADINGPERCENT	% shading (estimated % stream shaded during daylight hours)
SHADPERCENTPOW0.5	% shading to the power 0.5 (% stream shaded during daylight hours)
SHADPERCENTPOW0.8	% shading to the power 0.8 (% stream shaded during daylight hours)
SLOPE1KUS	Elevation difference between the middle of the site and a point 1km upstream
SLOPE1KUSPOW0.2	Slope to the power 0.2 - Diff in elevation (m) b/w the site and 1km u/s
STORDER	Stream order
STWIDTHPOW-0.3	Mean width of water (m) to the power -0.3
WETPERCENT	Percentage rainfall in wet season (%)
WETR	Range in wet season monthly rainfall means (mm)

### **Validating reference sites**

The full reference invertebrate and habitat data set, minus sites deleted at the classification stage, were used to create the first version of the model. The same reference data set was then entered into the model as if they were test sites to validate reference site condition. Reference sites with observed to expected taxa ratios (O/E ratios) below 0.75 were reviewed by state/territory agencies and deleted from the model if found to be unsuitable as reference sites. Appendix C lists the number of sites rejected as reference sites for model development. The models were then reconstructed with the failed reference sites removed. The reduced reference site data sets were then run through the revised model. The sites actually used in the final AUSRIVAS models are listed in Appendix B. The O/E ratio output at this stage is considered to represent the distribution of ecological health for the population of reference sites.

### ***O/E Taxa Bands Representing Biological Health***

To simplify interpretation, AUSRIVAS presents the O/E taxa ratios as bands that represent different levels of biological condition. The widths of the bands are based on the distribution of the O/E taxa values for the reference sites of each particular model. Test sites that fall between the 10<sup>th</sup> and 90<sup>th</sup> percentiles are considered equivalent to reference condition (band A). Impaired test sites will fall into a band equivalent to the severity of impact that the site is experiencing. The next two bands (B and C), which represent increasing levels of impairment, are the same width as band A. Band D will vary in width depending on the variability of the reference O/E taxa values used to create the model. Test sites with biological condition richer than reference condition (i.e. > 90<sup>th</sup> percentile) are placed in Band X. The bands used for each of the new or revised AUSRIVAS models are listed in Table 3.

**Table 3: O/E taxa bands for the new AUSRIVAS models created under this contract.**

Model	O/E Taxa Bands				
	X	A	B	C	D
NSW Autumn Edge	> 1.17	0.82 - 1.17	0.47 - 0.81	0.12 - 0.46	0 - 0.11
NSW Autumn Riffle	> 1.13	0.87 - 1.13	0.61 - 0.86	0.35 - 0.60	0 - 0.34
NSW Combined Eastern Edge	>1.17	0.83 - 1.17	0.49 - 0.82	0.15 - 0.48	0 - 0.14
NSW Combined Riffle	>1.14	0.86 - 1.14	0.58 - 0.85	0.30 - 0.57	0 - 0.29
NSW Combined Western Edge	>1.14	0.86 - 1.14	0.58 - 0.85	0.30 - 0.57	0 - 0.29
NSW Spring Edge	> 1.16	0.84 - 1.16	0.52 - 0.83	0.20 - 0.51	0 - 0.19
NSW Spring Riffle	>1.18	0.81 - 1.18	0.44 - 0.80	0.07 - 0.43	0 - 0.06
QLD Coast Autumn Edge	>1.20	0.80 - 1.20	0.38 - 0.79	0 - 0.37	N/A
QLD Coast Autumn Pool	>1.21	0.79 - 1.21	0.37 - 0.78	0 - 0.36	N/A
QLD Coast Autumn Riffle-Run	>1.16	0.84 - 1.16	0.51 - 0.83	0.18 - 0.50	0 - 0.17
QLD Coast Combined Edge	>1.18	0.82 - 1.18	0.45 - 0.81	0.08 - 0.44	0 - 0.07
QLD Coast Combined Pool	>1.17	0.83 - 1.17	0.47 - 0.82	0.11 - 0.46	0 - 0.10
QLD Coast Combined Riffle-Run	>1.20	0.80 - 1.20	0.40 - 0.79	0.01 - 0.39	0.00
QLD Coast Spring Edge	>1.22	0.78 - 1.22	0.33 - 0.77	0 - 0.32	N/A
QLD Coast Spring Pool	>1.24	0.76 - 1.24	0.28 - 0.75	0 - 0.27	N/A
QLD Coast Spring Riffle-Run	>1.22	0.78 - 1.22	0.33 - 0.77	0 - 0.32	N/A
QLD West Autumn Edge	>1.19	0.81 - 1.19	0.43 - 0.80	0.05 - 0.42	0 - 0.04
QLD West Autumn Pool	>1.21	0.79 - 1.21	0.37 - 0.78	0 - 0.36	N/A
QLD West Autumn Riffle-Run	>1.21	0.79 - 1.21	0.36 - 0.78	0 - 0.35	N/A
QLD West Combined Edge	>1.15	0.85 - 1.15	0.55 - 0.84	0.25 - 0.54	0 - 0.24
QLD West Combined Pool	>1.15	0.85 - 1.15	0.54 - 0.84	0.23 - 0.53	0 - 0.22
QLD West Combined Riffle-Run	>1.09	0.91 - 1.09	0.73 - 0.90	0.55 - 0.72	0 - 0.54
QLD West Spring Edge	>1.22	0.78 - 1.22	0.34 - 0.77	0 - 0.33	N/A
QLD West Spring Pool	>1.30	0.70 - 1.30	0.09 - 0.69	0 - 0.08	N/A
QLD West Spring Riffle-Run	>1.18	0.82 - 1.18	0.45 - 0.81	0.08 - 0.44	0 - 0.07
SA Autumn Edge	>1.19	0.81 - 1.19	0.43 - 0.80	0.05 - 0.42	0 - 0.04
SA Autumn Riffle	> 1.19	0.82 - 1.19	0.45 - 0.81	0.08 - 0.44	0 - 0.07
SA Combined Edge	>1.16	0.84 - 1.16	0.52 - 0.83	0.20 - 0.51	0 - 0.19
SA Combined Riffle	>1.18	0.82 - 1.18	0.46 - 0.81	0.10 - 0.45	0 - 0.09
SA Spring Edge	>1.19	0.80 - 1.19	0.41 - 0.79	0.02 - 0.40	0 - 0.01
SA Spring Riffle	>1.14	0.86 - 1.14	0.58 - 0.85	0.3 - 0.57	0 - 0.29
WA Autumn Channel	>1.15	0.85 - 1.15	0.55 - 0.84	0.25 - 0.54	0 - 0.24
WA Combined Channel	>1.15	0.85 - 1.15	0.55 - 0.84	0.25 - 0.54	0 - 0.24
WA Spring Channel	>1.15	0.85 - 1.15	0.55 - 0.84	0.25 - 0.54	0 - 0.24

## Achievement of Objectives

### **1 To complete the second iteration of the AUSRIVAS models for all States and Territories, taking into account;**

- **Requirements for replacement or removal of poor quality biological and environmental data**
- **Further additions and/or removals of reference site samples, in consultation with the lead agencies**
- **Further considerations with regard to site classification such as the numbers of groups selected**

### **New South Wales**

NSW EPA undertook the development and testing of all 7 of their own models (refer to Turak *et al.* 2000 for details) in Phase 1 of the project.

The CRCFE then rebuilt these models according to standard AUSRIVAS protocols, for checking purposes and to create files necessary for mounting on the AUSRIVAS website. This was undertaken after the completion of Phase 1 and prior to commencement of Phase 2. NSW EPA performed an extensive review of reference sites to be included in models (see below).

### **Data**

The following steps were taken to prepare data sets for model development:

- Revision of reference/test status to all sites (777) sampled in NSW from 1994 to 1998.
- Classification of sites assigned “reference” status into three classes A, B and C. The classes A and B indicates near pristine and slightly modified reference sites respectively, and C indicates moderately disturbed sites which were nominated as reference sites because more appropriate reference sites were not available for that type of river.
- Revision of all biological records from reference sites against a set of criteria (explained in the section on Quality Control below) and assignment of “fail” to samples that do not meet these criteria. Removal of all failed samples from the process of model development and performance assessment.
- Revision of all environmental records against a set of criteria (explained in the quality control section) and replacement of erroneous data with plausible substitutes.
- For the single season models, compilation of data sets using two samples selected randomly from sites that had more than two years data. Only records that passed the quality control tests were included in this process.
- For combined season models, samples from consecutive seasons were combined. From this set of combined records, two were randomly selected from each site (where possible).
- Removal of families that occur at less than 10 sites and records that have less than 10 families.

- Removal of fine substratum (silt and clay) from potential predictor variables. This follows the finding that the recording of fine substratum was highly variable and this variability contributed significantly to inconsistencies in AUSRIVAS outputs for some sites.
- Examination of the distribution of the values for all potential predictor variables using a number of different tests for normality and kurtosis. Transformations of the data were used to improve normality and reduce kurtosis.

### **Models**

For the assessment of central NSW lowland sites, a separate western combined edge model was created, an improvement for the assessment of lowland sites on the 6 existing statewide models. Only a combined season edge model was developed for the western rivers as edge was the only available habitat and neither the Spring nor the Autumn collections from reference sites had enough taxa to allow the development of reliable single-season models in that area.

Previous models relied heavily on the subjective estimates of silt and clay for assessing the sites on the floodplains of the Murray-Darling system in Western NSW. NSW EPA found that silt and clay percentages recorded in different years or season were very different not only among recorders but also among different observations of the same recorder, causing error in predictions and large fluctuations in O/E values. The new combined edge model for Western NSW only uses permanent site attributes, resulting in a major improvement in site assessments using AUSRIVAS in this region.

### ***Definition of biological groups***

NSW used the following procedure to define biological groups:

- Classification of all data using UPGMA to produce dendrograms from which preliminary biological groups were defined.
- Generation of 3 dimensional ordination plots (SSH MDS).
- Principal Axis Correlation (PCC) analysis of families on the ordination to determine the contribution of each family to the patterns of ordination. This provides some indication of the macroinvertebrate families that were characteristic of different stream types and hence biological groups. Monte-Carlo simulations were used to determine the significance of these relationships.
- PCC analysis of environmental data on the biological ordination to determine the relationship between the environmental variables and patterns in the ordination. Of these variables the most likely predictor variables were those that yielded a high correlation coefficient. Monte-Carlo simulations were used to determine the values of correlation coefficient cut off at  $\alpha = 0.01$  significance.

- Generation of 3 dimensional rotating plots of ordinations that include all taxa that had an  $R^2$  value greater than 0.4 and all environmental variables that had an  $R^2$  value greater than 0.5.
- Identification of the preliminary biological groups (determined from the dendrograms) on the rotating plot with different colours or symbols.
- Examination of the patterns on the 3-dimensional ordination plot and modification of group definitions by seeking gaps in the ordination patterns, the environmental gradients and the characteristic taxa. In making these modification the experience gained during the construction of the previous models as well as the six years of sampling and analysis was utilised so that the biological groups generated made “ecological sense”. If this process is carried out carefully the groups should represent an identifiable “stream type”.
- After each modification was made constancy and fidelity tables were generated to examine the tightness of these groups. These tables show the percentage of sites within each group at which a taxon was present and hence the probability of finding a taxon within sites from a particular group. These tables constitute the biological data that is used in AusRivAS to assess site condition and were used to make final adjustments to groups by moving sites from one group to another, or occasionally removing a site altogether in order to achieve greater constancy and fidelity.

### ***Selection of predictor variables***

Predictor variables were chosen that were highly correlated with ordination patterns. Predictor variables were chosen by NSW EPA based on the following guidelines:

- minimisation of error rates in the cross validation procedure,
- minimisation of the use of predictor variables that are effected by human activities e.g. substratum, alkalinity, and avoid those that are particularly indicative of disturbance for that habitat (e.g. sand in the riffle habitat),
- preference for variables that are not subjective estimates,
- inclusion of climatic variables,
- the ease at which model users may obtain the values for test sites.

### **Queensland**

QLD NRM provided the CRCFE with a revised set of reference sites, with many additional reference sites added or removed since the previous models were constructed, with some models now having over 400 sites. Habitat types were defined by QLD using a habitat decision matrix. Based on velocity and depth, samples are divided into riffle/run and pool habitats. Models for three habitat types were built for QLD, riffle/run, pool/bed and edge habitats.

## Data

The Queensland model development has gone through 3 phases since model building began in 1997. The first generation models were built on a group of sites that were perceived to be reference based on the selection criteria outlined in Davies (1994). Queensland had 196 reference sites and 20 monitoring (equivalent to test) sites in all. Most of the environmental variables collected (phys/chem, habitat, physical features) were included in the dataset for model development. The models were Autumn and Spring, edge and bed (included both pools and riffles). These models performed poorly because reference sites were subject to human impact.

The second generation models dataset included sites where an attempt had been made to quantify reference condition. A set of nine criteria was developed to assess the sites. Each site was given a pass or fail against each criterion. To be considered reference, a site could fail no more than one criterion. If that one failure was against the 'dam or weir within 20 km' criterion, the site also failed. In addition to the sites used in the first generation models, sites sampled in the first year of the FNARH were also included in these models. A total of 323 reference sites and 199 test sites were used in model construction. The environmental variables were limited to those that were not subject to human impact. Only two of the second generation models were developed and used: Autumn Edge and Autumn Pool. As the CRCFE began development of the rest of the models it became obvious that the difficulties inherent in trying to model such a large area were making the models very unreliable.

The decision to develop regional models was a result of an inability to build a good spring model. It appeared that there was too much variability across the State at that time of year and no clear trends were apparent. The western sites were removed from the model, and it performed much better. At this point, QLD was offered the choice of persisting with state-wide models or building a new suite of regional models. QLD NR&M decided to go with the regional models as it gave the team scope to improve the datasets used to develop the models. The following changes were implemented:

1. The method used to define reference condition was refined and all sites were re-assessed. Each of ten criteria attracts a score between 1 (severely impacted) and 5 (no impact); scores were summed and sites scoring  $\geq 40$  were considered reference for the purpose of model development (some sites scoring between 40 and 44 were subsequently removed as borderline references if they were outliers during model development).
2. Habitats were reclassified using the ratio between velocity and depth to assist in separating pool, run and riffle habitats.
3. Coverage of the state by sites was limited in the first and second generation models. Since then, more sites have been sampled from the west and Cape, as

well as newer sites from previously sampled areas. These sites were included to give a better coverage of the state and thus more relevant models.

4. Some of the environmental variables used in the first and second generation models were either too difficult for others (consultants, students) to collect e.g. soil class required a GIS overlay of the sites or were subject to human influence (e.g. velocity, depth). These were removed from the dataset.

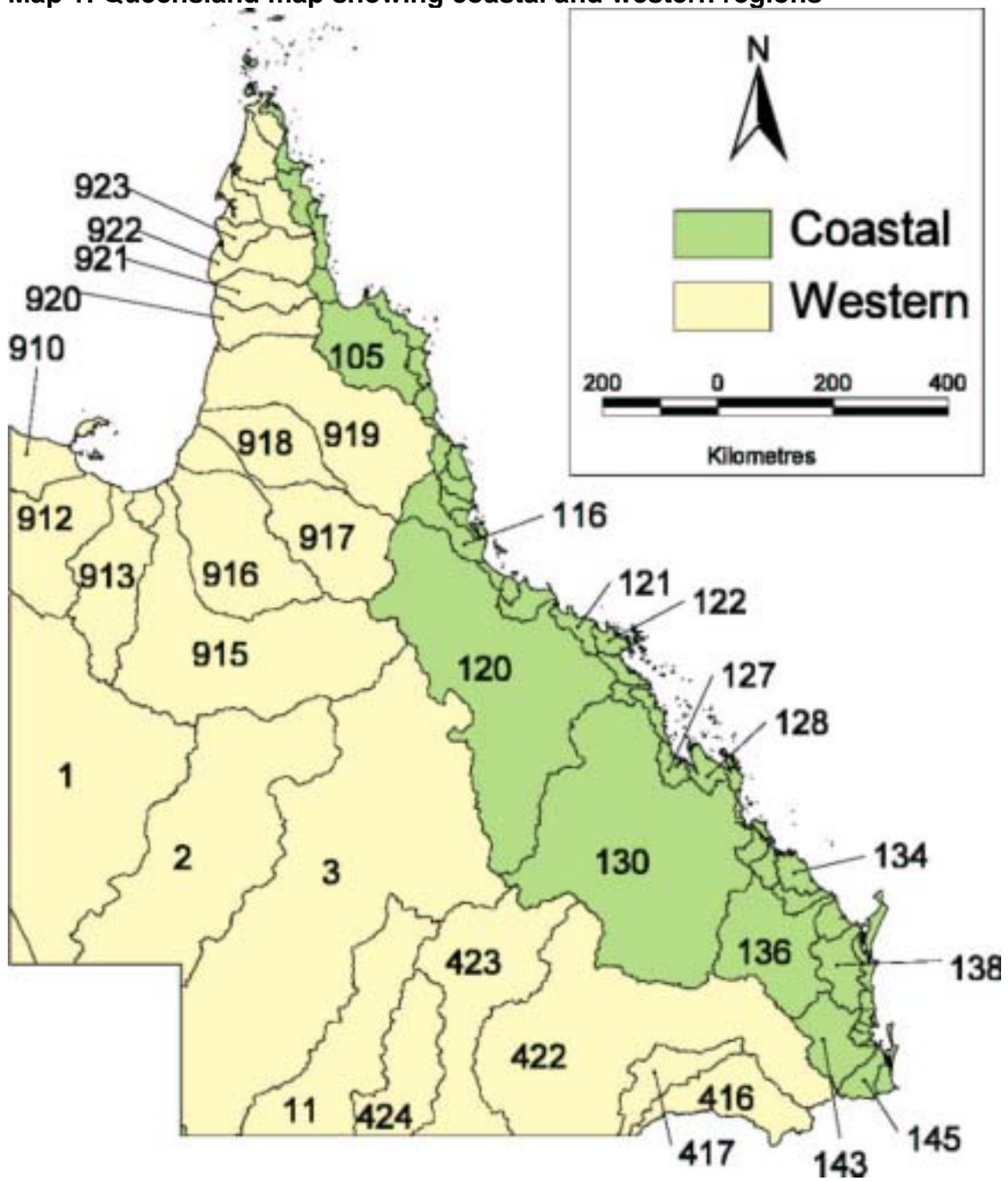
### **Models**

Initially state-wide models were constructed, however a number of problems were encountered. The CRCFE undertook development of a first iteration of 4 out of the 7 state-wide models proposed, however, considerable difficulties were encountered in obtaining useful classifications and low model errors. In addition, these models had a large number of sites with high O/E values and low expected number of taxa. The CRCFE suggested the following in the Progress Report in February 2002 to EA:

- The expected number of taxa is very low because of the low probabilities of occurrence of many of the reference site taxa. However observed numbers of taxa seem to be normal.
- Group frequencies show probabilities of occurrence are often spread between the groups. This means that very few taxa will actually be calculated to have a 0.5 chance of occurrence. Group membership is also not strong in some cases.
- This tends to suggest that there are many taxa that are generally widely distributed and patchy, i.e., their occurrence is not predictable. Such taxa may be the micro-crustacea that QLD collects, that may be locally abundant dependent more on suitable habitat than river health.
- Predictions are also confounded by the huge spatial variation in the aquatic fauna, therefore it is likely there is a need for regional models, a higher rare taxa cutoff and comparison with fewer, more closely matched reference sites (trial of another prediction method such as Assessment by Nearest Neighbour Analysis (ANNA) may be required).

EA endorsed the suggestion of the CRCFE to build regional models and QLD split all the data sets into two regions, coastal and western (see Map 1: Queensland map showing coastal and western regions), and re-sent the data to the CRCFE in March 2002. Western and coastal models were built for all habitats and seasons by the CRCFE. Sites in catchments that drain to the Pacific Ocean coast were used in the 'Coastal' models (101 – 114, 116 – 122, 124 – 127, 129 – 130, 132 – 146) and sites that drain to the Murray Darling, Lake Eyre or Gulf of Carpentaria were used in the 'Western' models (001-003, 011, 416 – 417, 422 – 424, 912 – 913, 915 – 919, 921 – 923, 925 - 927). A total of 487 reference sites and 335 test sites were used to develop the regional models.

Map 1: Queensland map showing coastal and western regions



## **South Australia Data**

A number of refinements had been made to the dataset since the last round of modelling, including:

- Review of all reference sites and exclusion of some reference sites identified as having potential or known impacts
- Checking of environmental predictor variables revealed some incorrect catchment measurements. These have since been recalculated in a GIS package and corrected in the dataset.
- Unreliable variables such as sand, silt and clay were removed as potential predictor variables
- Additional Murray River naturally salty reference sites and other sites sampled in 1997 have been added to the dataset of 94/95 sites.
- Predictor variables have been transformed to normalise distributions

## **Models**

South Australia performed initial site classifications and supplied the CRCFE with site groups formed from these. Earlier versions of the models were not very effective at predicting condition of other sites in the State (i.e. Sites not in the MDB), when River Murray sites and naturally salty sites were included in the model. Therefore some testing of models was required to ensure sites would be predicted into the appropriate group.

Classifications were analysed by SA to ensure reference sites that were naturally salty formed a classification group. Several families usually excluded as rare taxa were retained in the classifications because they were particularly distinctive of classification groups, i.e. were only found at naturally salty sites.

The classification was then redone with test sites included to ensure that naturally salty sites with other impacts were likely to be predicted to the salty group. Other sites in the state, not naturally salty but impacted by salt were also classified and in most cases grouped next to sites most geographically close to them, not into the salty group. This suggested that a model with salty reference sites included was likely to work well, providing appropriate predictor variables were chosen. Predictors such as latitude and longitude (indicators of geographic position) and a measure of 'saltiness' (conductivity/alkalinity) were considered important. When reviewing reference sites falling below 0.75, some naturally salty reference sites fell below 0.75, but were retained. As this applied to only a few sites, there was a minimal effects on band widths.

## Western Australia

CALM in WA undertook the development of three state-wide channel models (refer to Halse *et al.* 2001 for details). Pseudo-abundance models were also investigated by CALM, however the idea was abandoned because of problems with estimating abundance and a lack of differentiation between reference and test sites. Therefore, only three state-wide models have been created for WA.

## Data

- The sub-sampling procedure was refined over the course of FNARH, with a change from 30 minute to 60 minute live picking, as well as some other changes.
- Channel habitat was adopted as the main sampling habitat, as other habitats occurred too infrequently at reference and test sites to build a model with wide applicability.
- Several interim models have been developed between the first beta models that are on the website and the models in 2001 proposed as alpha models.
- Initial AUSRIVAS models were based on wet/dry seasons, however, problems were encountered assigning samples from the central part of WA to the appropriate model. Therefore, revised models are based on spring and autumn assignment of samples.
- The reference data set was scanned for sites with depauperate fauna and low O/E values using a third generation model and these sites were removed prior to construction of the autumn and spring models.
- Predictor variables were log transformed where appropriate to normalise distributions.

## Models

Taxa occurring at less than 5 sites were considered rare and masked out of all further analyses.

### *Spring Model*

Only 60 minute live pick data was used. Reference sites were screened for low O/E values based on a third generation model and 34 reference sites were removed prior to model development.

A further 11 reference sites were removed after examining the classification. "Reference sites containing families belonging to groups that did not occur at other sites of the same reference site group were discarded if geological setting or other environmental parameters suggested the site was atypical in terms of site characteristics. In a parallel analysis, reference sites were ordinated using semi-strong hybrid multi-dimensional scaling and sites that were outliers in ordination space from their classification group were subjected to the same examination as unusual sites in the two-way table." (pg.10 Halse *et al.* 2001)

Predictor variables were tested for significance. After the cross-validation procedure, four sites classified to obviously erroneous groups were removed from the dataset, the stepwise analysis repeated and the discriminant function recalculated, without improvement to the cross-validation error.

After model construction, 7 sites with low O/E values were discarded.

#### *Autumn Model*

Mostly 60 minute live pick data was used, as well as some 30 minute live pick data from the south west where geographic coverage of 60 minute live pick data was incomplete. Reference sites were screened for low O/E values based on a third generation model and 38 reference sites were removed prior to model development.

Again, examination of the classification, ordinations and a two way table resulted in removal of a further 11 outlying reference sites.

Only 4 of the canonical vectors in the discriminant function were significant, so the 5<sup>th</sup> was discarded in model construction.

After model construction, 8 reference sites with low O/E values were discarded.

#### ***Combined Model***

Both 60 minute live pick data from the north west and 30 minute live pick data from the south-west was used. A lack of sites sampled in both spring and autumn meant that no screening of reference sites based on O/E scores prior to model construction was performed.

Examination of the classification, ordinations and a two way table resulted in 5 outlying sites being removed, another two sites were removed because they had fewer than 5 families.

After the DFA, one site that was obviously misclassified was discarded and the discriminant function was redone, producing little improvement in cross-validation error.

## **2 All models are to be revised and approved at an alpha level**

### ***New South Wales***

NSW EPA extensively evaluated all models created internally, selecting the best performing iterations to become alpha models (see Turak *et al.* 2000).

In assessing the performance of the models in NSW the following steps were taken:

1. The O/E values for individual reference sites, the overall distribution of O/E values and the regression relationship between O and E values for reference sites were examined and comparisons were made between reference site data included in the model and those left out of the models.
2. The model outputs for all available test sites were examined, using results from sites with known types and degrees of disturbance to judge whether the models were providing suitable outputs.

After AUSRIVAS and EPA outputs were matched, the models were mounted on the AUSRIVAS website. Main discrepancies found by the CRCFE were related to the probability level used and the method used to calculate band widths. Standard AUSRIVAS methods were used in the case of two different methods.

### ***Queensland***

The CRCFE and QLD NR&M worked closely to assess regional model performance at a number of stages. The structure of the dendrograms and the number of classification groups was examined. Investigation of the sites in the Autumn Edge Coastal dendrogram was undertaken when no clear groups were defined. Several sites had repeated sampling occasions and the similarity of these sites was examined. It was expected that for a single site, with habitat characteristics that remain the same over repeated sampling occasions, the fauna should be similar and therefore sites classified into the same group. It was found for a large number of sites however, that very different fauna was collected. These sites were reviewed by QLD NR&M and sites showing large differences in macroinvertebrate assemblage between sampling occasions were removed (Appendix D). Many of these sites were borderline reference, based on the QLD system of scoring sites using a set of criteria to define reference condition. A trade-off was made regarding the number of groups in the QLD regional models and the misclassification error. Generally the fewer the groups, the lower the misclassification error. In the case of the QLD models, even though structure was weak it was considered important to have a greater number of groups, therefore misclassification error rates of up to 0.49 were accepted to maximize the number of groups.

The O/E distribution was examined and sites with particularly low values (<0.70) were reviewed. A number of potential impacts were identified by QLD staff and some of these sites were removed from the models (Appendix C). Sites with

very low numbers of collected taxa (3 or 5 taxa) show that a wide range of conditions are deemed acceptable reference condition. Some of these sites have been reviewed and considered to be naturally depauperate, however, there may also be sites that have been potentially impacted by natural disturbances such as droughts and floods. Sites with low Expected values ( $NTE_{50} < 6$ ) were mainly a problem in the Pool models. These sites did not form distinct classification groups in the Autumn Pool Coastal, Autumn Pool Western and Spring Pool Coastal models. Sites with low Expected values in the Spring Pool Western model are within the same group, but other sites are also included in this group that do not have low Expected values. Where low numbers of taxa are predicted and expected, this can have a large impact on the O/E values because of the chance variability of losing any one taxon at a site. For example if 5 taxa are expected and only 2 taxa are collected, an O/E value of 0.40 results. Queensland Pool habitats are extremely variable, ranging from relatively shallow, rocky beds to deeper silty beds. The variability (that can't be accounted for by environmental variables available for use as predictors) and low numbers of taxa collected from some pools makes it difficult to predict taxa in this habitat. It may be better not to use pool habitat, and perhaps look at building a composite (multiple habitat) model or a more restrictive model (e.g. sandy bed for western streams, rocky bed for coastal streams) in the future.

Problems of low O/E values in the models were also caused by sites that have reasonable numbers of observed taxa ( $NTC_{50}$ ) but much higher numbers of expected taxa (For example 11 taxa collected, but 18 taxa expected). This could be a sign of unidentified impacts in the site data, such as drought. Queensland rivers are often in either progressive drought or flood, therefore there may be problems with seasonal variability in the models that has not been accounted for by the available variables used to describe variation in the reference condition and subsequently predict the composition at test sites.

QLD samples are collected pre- and recession-wet (early May and early October). This ensures that the wet season has ceased properly, although the positive effects of the wet on the fauna and habitat is still evident during the autumn sampling, and that the storm season has not begun in earnest by the time the spring season begins. Sampling in the wet season is avoided because rivers generally have elevated flows, and the fauna has often been denuded by scouring flows in the pre-wet storm season and subsequent wet season high flows. The recession-wet sampling is conducted at least 6 weeks after flood events, so that the fauna has had time to recuperate.

Even so there have been a few instances when sample timing was off, particularly with the late-year storm season in the north of the state. Inclusion of these samples could contribute to the discrepancies between numbers of observed and expected taxa. Further review of the data is recommended to identify and remove sites with those problems.

Some sites had low numbers of observed taxa (NTC50), even though the number of taxa actually collected (total taxa richness) was not correspondingly low. This is attributed to the variable and patchy distribution of taxa, or possibly less than necessary sampling rigour. The reasons for the poor structure of the Queensland models are not clear, however, further regional divisions may be needed. This is especially so given that these models cover a distance of close to 2000km north to south and range from true tropical, through wet tropics, subtropical, temperate and arid country conditions. A similar problem was recognized with the Western Australian models.

Some taxa that are usually excluded from macroinvertebrate sampling (Copepoda, Cladocera, Neritidae, Ozobranchidae, Hymenoptera and Rotifera) were used in the Queensland models. It may be difficult for these taxa to be reliably sampled, and future work should consider excluding them from the taxa count, or perhaps modifying the sampling protocols to ensure reliable collection of these representatives of the fauna.

The broad width of Band A will mean that sites assessed as Band B and C should be treated as severely impaired. Unusually high O/E values were a problem for all of the QLD regional models, and will result in drawing a conclusion of 'no reliable assessment possible' or 'fail' for sites allocated to Band X. It will be difficult to determine whether test sites in Band X are actually richer than reference, or are a result of poor model performance. If it is concluded to be poor model performance, the conclusion will be 'no reliable assessment'. If it can be ascertained that the models has predicted well at such sites and they suffer from mild nutrient enrichment or perhaps continued irrigation flows etc., then such sites should be accorded a 'fail' assessment. Further evaluation of the various aspects of the models discussed above will be needed to improve confidence in the assessments provided by them. However, in the meantime the QLD NR&M should be encouraged to adopt formal 'weight of evidence' and 'multiple lines of evidence' approaches (see Downes et al. 2002 Chapter 9) to corroborate and strengthen their conclusions.

Validation of AUSRIVAS models is usually done by assessing the outputs of reference sites that were not used to create the model. This was the approach used for the development of the QLD models. The QLD lead agency also assessed the QLD models using test sites and concluded that there were no problems with the outputs of the models. However, the use of test sites for this purpose can be problematical as the magnitudes of potential impacts are not known, making it difficult to assess the results of the models.

With the recent development of a new piece of software, an alternative to evaluating models using test sites is recommended. This software, termed the "Impairator", builds simulated data sets with known levels of impairment. With these data sets as a test bed, a more sensitive assessment of models can be

made. Details on this software can be found on the website <http://lake.canberra.edu.au/~impairator>.

Combined season models performed better than the seasonal models with lower misclassification errors and band widths. The average number of taxa collected in the combined season models is higher than for the seasonal models, therefore reducing the variability of O/E values if any taxa are missing.

### ***South Australia***

An important improvement required from Alpha models is that they would be able to discriminate naturally salty sites from those impacted by salt. In addition, Murray River sites previously falling outside the experience of the Beta models should be able to be assessed with the addition of extra Murray River reference sites. Both of these objectives were achieved by the new models.

The South Australian models were considered by the sub-meeting of the steering committee as Alpha models and mounted on the AUSRIVAS website . The only issue raised with the models was the use of excessive data transformations, including power transformations. While some variables obviously benefit from transformations, others may predict just as well being untransformed and transformations can sometimes be 'overdone'. For the user, transformations can create a source of error, if users forget to transform their data, or use the wrong transformation, especially with more obscure transformations. At this stage, it is very much up to state agencies as part of the data preparation step whether variables are transformed and which transformations are used. It was recommended by the CRCFE that perhaps there is a role for the steering committee to provide some guidance to States/Territories on transformations.

### ***Western Australia***

Upon review of the document "Milestone Report 4" from CALM, reporting on the development of the final WA AUSRIVAS models, a number of issues were raised by the steering committee:

1. ".....there is a major practical difficulty in establishing models (the autumn and combined models) which require sites to be sampled with either 30 mins or 60 mins of live picking. Which of the two picking periods are to be used at a test site? I cannot see how these models can actually be used without first determining whether a test site is in a region where 30min or 60 min picking was carried out. If this is the case then separate models should be constructed for the two picking periods. Without being able to see whether there are any real differences in invertebrate recovery between the two periods, I can only conclude that the autumn and combined models are not likely to be useful."  
(R.M.)
2. Reference sites with O/E values falling below 0.75 seem to have been automatically removed from the models without review. More information was

requested from CALM on the review procedure used to assess sites with  $O/E < 0.75$ .

3. Another matter of concern is that reference sites seem to have been removed at inappropriate stages in the model building process, for example, after the discriminant function analysis cross validation step:

Pg. 15 "Four sites classified to obviously erroneous groups and were removed from the dataset, the step-wise discriminant analysis was repeated and the discriminant function re-calculated without improvement in cross validation error (30%)."

While sites very different from other sites identified in the classification step are sometimes removed, it is not usual to remove sites after the cross validation step. Some sites will nearly always be misclassified to another group, that is why the probability of group membership of a site to all groups in the model is used in the model calculations. Strictly speaking, removal of reference sites should be for the reason that they are not representative of reference condition, and this should be determined prior to model construction.

4. The AUSRIVAS protocol used to build all AUSRIVAS models thus far has been to classify taxa as rare and remove those that occur at less than 10% of sites if there are less than 100 sites. For more than 100 sites, those that occur at less than 10 sites are considered rare. For WA models, a lower cutoff of 5% has been used for all models (with greater than 100 sites). However, since some rare taxa removal occurred, this is not likely to cause a major problem.

5. Description of the spring model states that a number of different cut-off values were used to review reference sites, depending on their geographic location:

Pg. 10 "A total of 199 reference sites were sampled and processed with 60 mins live-pick between spring 1997 and spring 1999, with good geographical coverage of all river systems in WA and, after screening, these sites were the basis for building a revised spring model. In the initial screening process, the 74 sites that had been sampled in 1994, 1995 and at least once subsequently with 60 mins live-pick were put through a third iteration spring (wet season) model. The 9G model was used. Nine sites with consistently low  $O/E$  scores in all three rounds were discarded and then the distribution of scores at other sites was examined to determine two thresholds. The first was a general cut-off value for 60-mins live-pick scores, below which sites should automatically be omitted from model development, and the second was a score below which they should be discarded if evidence of sampling difficulties or habitat degradation existed. The threshold  $O/E$ s were 0.80 and 0.85, respectively, for north-west sites, 0.72 and 0.77 for wheatbelt sites, and 0.83 and 0.88 for south-west forest sites. This screening reduced the number of reference sites to 165."

Other AUSRIVAS models have always used O/E <0.75 as the cutoff. However, more important than the O/E value is the reason for removing the reference sites, on which no details have been provided.

6. The CRCFE was not supplied with the full data set of all reference sites, making it very difficult to check the models for consistency and possible errors. No details have been supplied to the CRCFE or in the report on the calculation of bands. Band widths are identical for all three models.

Although a number of discrepancies in model construction procedures have been highlighted, it was concluded by the steering committee that most of these discrepancies are not critical to model performance of the WA models. However, if models are to be comparable nationwide, they must be constructed according to consistent methods. Therefore, where States and Territories are involved in the process of model construction, methods consistent to those established must be followed and all data supplied for checking purposes. Thus, it was recommended that as the models represent an improvement to the beta models currently on the AUSRIVAS website, they should be mounted on the AUSRIVAS website as Alpha models, with the proviso that they were unlikely to have an extended "shelf life" and would require rebuilding in 1.5 – 2 years. As only 3 models were created for WA there is an opportunity for recreating these models in consultation with the CRCFE, addressing the following issues:

- 1) The effect of differences between 30 minute and 60 minute samples on O/E values needs to be quantified
- 2) The effect on the final model of removing sites after the classification and DFA analysis steps needs to be further investigated.

**3 To mount all revised alpha level models on the AUSRIVAS WWW site at the University of Canberra, with all necessary supporting documentation, including taxonomic codes.**

Supporting documentation for NSW, SA, WA and QLD is complete and available on the AUSRIVAS website (<http://ausriv.as.canberra.edu.au/>). Material in the manuals is specific to each state and covers aspects of habitat sampling, macroinvertebrate sampling and processing relevant to the use of the models.

Models for all states have been mounted on the AUSRIVAS website complete with new variable descriptions where required (see Table 2).

**4 To provide statistical advice to National River Health Program lead agencies on matters associated with data entry, site classification, discriminant function analysis and model construction.**

***New South Wales***

NSW consulted with the CRCFE during the model development process and sent the completed models to be checked and then mounted on the AUSRIVAS website.

***Queensland***

There was close communication between CRCFE staff and lead agency staff throughout the duration of the model building process. Communication with QLD NR&M was essential to identifying and resolving the problems encountered with the AUSRIVAS models. QLD were able to review sites that were outliers or had low O/E values and provide direction on whether or not to remove these sites from the models. The CRCFE was able to provide QLD with advice on interpreting AUSRIVAS results and on the model building process.

***South Australia***

SA visited the CRCFE in the initial stages of the contract to provide interaction between the CRCFE and SA in the classification and grouping of sites. Following this visit it was decided that it would be most appropriate for SA to perform their own classifications and form sensible groups based on their knowledge of the sites to provide to the CRCFE. Communication continued throughout the rest of the process, with SA providing useful feedback at each stage.

***Western Australia***

Less communication was had with WA during the model building process, and was the cause of some discrepancies in construction methods. These problems were mostly resolved in the checking and rebuilding phase of the models by the CRCFE, which involved considerable dialogue between WA and the CRCFE. A number of discrepancies were found in the method used to build the models by CALM when compared to the standard AUSRIVAS methods used by the CRCFE. Major issues raised were the combining of 30 and 60 minute live pick data in the same model, number of rare taxa removed and extra steps of reference site removal. A stronger dialogue with the CRCFE would have put us in a better position to provide technical advice to CALM on particular aspects of model construction, including rare taxa removal, reference site removal and model validation and checking.

## Summary of work completed

<b>Date</b>	<b>Achieved</b>
Prior to March 2001	Construction of NSW models and mounting on AUSRIVAS website
26 March 2001	Contract signed
May 2001	Alpha data supplied by WA, SA and QLD
18 May 2001	1 <sup>st</sup> Milestone payment
23 May 2001	Missing data identified in SA data after modelling commenced
June – September 2001	Model construction for WA and SA models
June 2001	QLD model construction (June–December; 1 <sup>st</sup> iterations of 4 models developed)
26 July 2001	Steering committee meeting – extension for QLD models given until 5th November 2001 because of delays
12 September 2001	Progress report submitted to Environment Australia Meeting with EA on SA and WA models
September-October 2001	Steering committee approval of SA and WA models
October 2001	SA and WA models mounted on AUSRIVAS website
October 2001	Mismatches between AUSRIVAS models and CALM models identified
14 November 2001	WA models remounted on AUSRIVAS website after identification and resolution of mismatches
January 2002	QLD contacted about problems with models
8 <sup>th</sup> February 2002	Progress report submitted to EA
March 2002	QLD regional data received
March 2002	QLD regional models built
April – May 2002	Further delays <ul style="list-style-type: none"> <li>- Duplicate samples discovered in data files requiring removal and re-construction of some models already built.</li> <li>- QLD staff away in field for 5 weeks</li> </ul>
June – July 2002	<ul style="list-style-type: none"> <li>- Sites with low O/E values reviewed by QLD NR&amp;M</li> <li>- Sites with potential impacts removed from models</li> <li>- Reference sites and test sites run through the new models; Assessments checked and models endorsed by QLD NR&amp;M</li> </ul>
August 2002	QLD Progress report sent to steering committee members
24-25 <sup>th</sup> September 2002	QLD AUSRIVAS models mounted on the web, reference sites run through the model and assessments checked against SAS outputs
September 2002	Final QLD models report submitted to EA
October 2002	Final Report Submitted to EA

## **Problems and delays**

Unfortunately a number of delays were experienced in the project and several extensions were required to compensate for this. However these delays were able to be contained within the costs of the project. WA chose to build their own models, so funds were available in the project to cover the additional costs.

Missing data was identified in some SA models after modelling commenced, resulting in the need to rebuild some models. QLD models had to be rebuilt many times, once as statewide models, again as 18 regional models, and some redone again when duplicate samples were identified in models.

## **Communication**

There was close communication between CRCFE staff and lead agency staff in most states throughout the duration of the model building process. Considerable communication occurred with SA and QLD, and although there was less interaction with NSW and WA during the model construction process, most problems were resolved in the checking and rebuilding phase of the models by the CRCFE. Communication between the CRCFE and the States was essential to identifying and resolving the problems encountered during the model building process. Communication with states that are competent in model building tasks is equally important, even if just to convey the need for consistency in approach. The CRCFE can also benefit from interaction with lead agencies leading to a better understanding of the mechanics of AUSRIVAS, suggested model improvements and new ways of thinking about and assessing model performance.

Both the AUSRIVAS models and supporting documentation information are available on the AUSRIVAS www site (<http://ausrivas.canberra.edu.au/ausrivas>).

## **Acknowledgments**

This project was funded by the National Heritage Trust under the National River Health Program and supported by the CRC for Freshwater Ecology at the University of Canberra. This project built on prior work of the CRCFE.

## 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.

Coysh, J.L., Nichols, S.J., Simpson, J.C., Norris, R.H., Barmuta, L., Chessman, B. and Blackman, P. (2000). AUSTRALIAN RIVER Assessment System (AUSRIVAS) National River Health Program Predictive Model Manual. Cooperative Research Centre for Freshwater Ecology, Building 15, University of Canberra, ACT, 2601 <http://ausrivas.canberra.edu.au/ausrivas>

Downes, B.J., Barmuta, L.A., Fairweather, P.G., Faith, D.P., Keough, M.J., Lake, P.S., Mapstone, B.D. and Quinn, G.P. (2002). *Monitoring Ecological Impacts: Concepts and Practice in Flowing Waters*. (Cambridge University Press.) pp.446

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

Halse, S.A., Scanlon, M.D. and J.S. Hocking (2001). First National Assessment of River Health: Western Australia Program. Milestone Report 5 and Final Report. Department of Conservation and Land Management, Western Australia.

S.A.S Institute. (1995) *SAS/STAT Guide for Personal Computers, Version 6.11*, SAS Institute, Carey, North Carolina.

Turak, E., Waddell, N. and G. Hose (2000). First National Assessment of River Health: New South Wales Program Milestone Report 3. NSW Environment Protection Authority.

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. *Arch. Hydrobiologia*, **127(3)**: 319-326.