



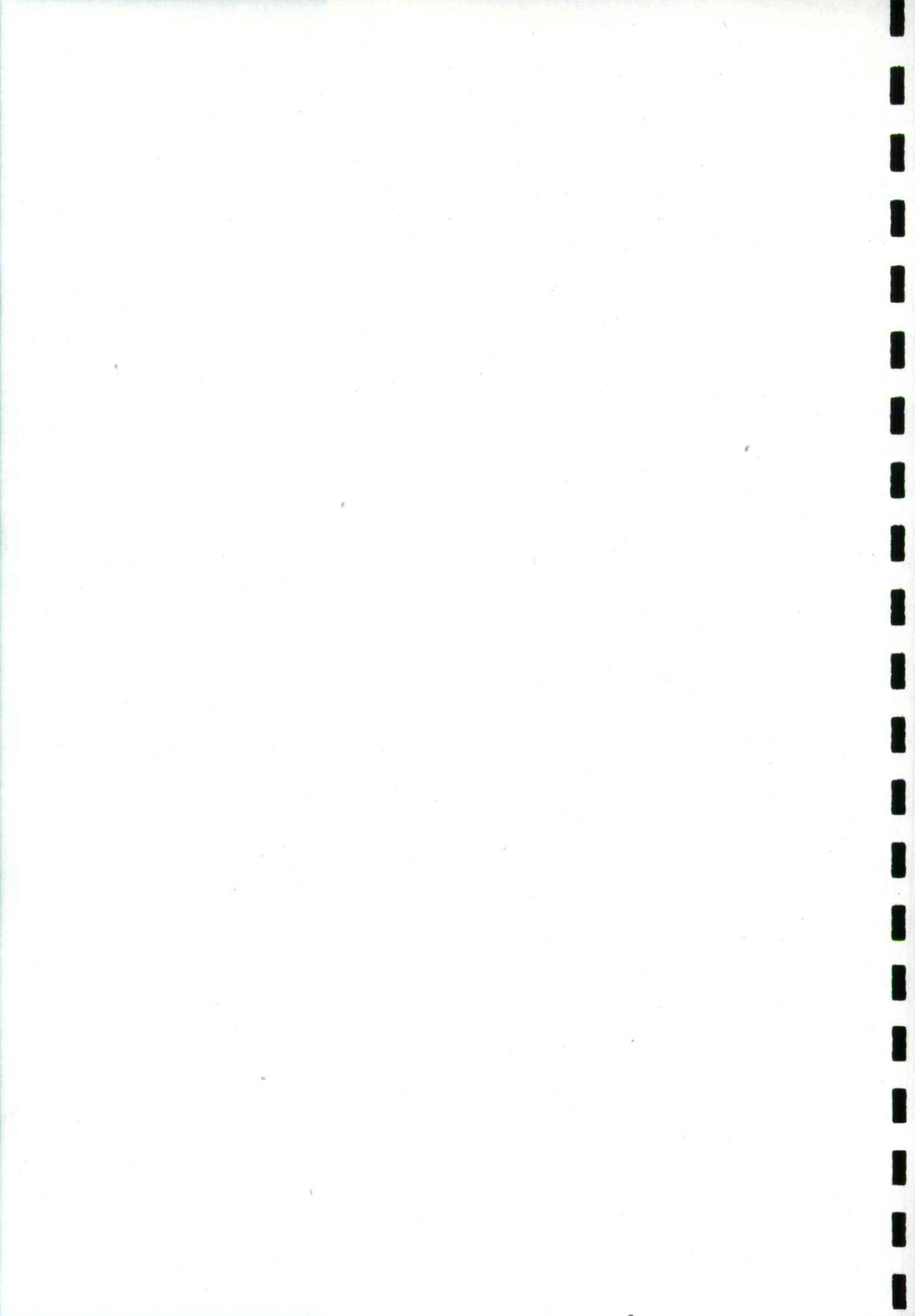
**ANALYSIS OF 1995
SURVEY DATA AND
RIVPACS UPDATE**

**R&D Progress Report E1/EMA 008/1
for the period 1st June to 31st August**

**M T Furse
R T Clarke**

**Institute of Freshwater Ecology
River Laboratory
East Stoke
WAREHAM BH20 6BB**

September 1996



1 TECHNICAL PROGRESS

The commencement date of the project was 3rd June 1996. This progress report covers the first three months of the work programme.

1.1 Objectives

The overall objective of the full research programme (Phases 1 and 2) is to:

- conduct a post-survey appraisal of the 1995 GQA biological survey data, both in terms of its assessment of biological quality, and as a tool for refining the methodology for future surveys.

The overall objective of the current phase, Phase 1 is to:

- undertake a scoping study for Phase 2 and prepare the principal tool to be used in the data analysis in order that Phase 2, comprising the data analysis and appraisal, will be undertaken most efficiently.

The specific objectives of the current phase are to:

- produce an enhanced version of RIVPACS III and its associated user manual incorporating the error terms detailed in R&D Note 412, for use in the Phase 2 data analysis and for Agency Operational purposes.
- identify and rank the options for further analysis of the 1995 GQA biological survey data and to select those most likely to meet business needs, in consultation with the Project Board and other specialists within and outside the Agency.
- produce a detailed PID and work specification for Phase 2 describing the analyses to be undertaken and the resulting products.

1.2 Work programme for the reporting period

The timetable of work for the reporting period was a compressed version of the original version, to compensate for the delayed project start date.

The work programme outlined here is based on the "Month Completed" column of the "Target and Timescales" table in the Project Initiation Document of 12th June 1996.

1.2.1 **Incorporation of error terms into RIVPACS III**

The following items were scheduled to be completed during the reporting period:

- Produce a detailed plan for the software enhancements showing the available new options and interface design.
- Finalise *[the previous item]* in consultation with the Project Board and Regional Biologists.

1.2.2 Scoping study for Phase 2

The following items were scheduled to be completed during the reporting period:

- Identify the extent, structure and accessibility of 1995 GQA Survey biological and environmental data.
- Prepare a discussion document describing the major options for analysis of the 1995 GQA biological data (the principal stage for input from the contractors).
- Agree circulation list for discussion document with Project Board (list to include Project Board members, Regional Biologists and other relevant persons both within and outside the Agency).
- Circulate discussion document requesting comments and alternative options and a ranking of all options in order of their relevance to Agency business needs.

1.3 Outputs produced

1.3.1 Incorporation of error terms into RIVPACS III

A plan for software enhancements to RIVPACS III to incorporate errors for ecological quality assessments has been completed and forwarded to the Agency Project Leader for circulation (Appendix I).

1.3.2 Scoping study for Phase 2

A discussion document listing options for additional uses of the 1995 GQA biological survey data has been completed and forwarded to the Agency Project Leader for comment (Appendix II).

2 INTERIM RESULTS

2.1 Incorporation of error terms into RIVPACS III

The first draft of the software development plan was forwarded to the Agency Project Leader, Dr R A Dines, by post in mid-July. This document formed the main basis for discussion at a meeting between Dr Dines and IFE project staff at a meeting held at the River Laboratory in August.

Following that meeting a revised a series of revisions were made to the software plan. Some involved improvements to the presentation of the menu sequence to make the document more comprehensible and others involved changes in the functionality of the new module.

The revised document (Appendix I) sets out two separate sets of changes. The first set is changes to the main RIVPACS III program to incorporate the new procedures for estimating errors, including the optional facility to use bias terms when calculating Environmental Quality Index values. Other new facilities include the option to band sites according to their ecological quality assessment and new formats for outputting data, including files in the correct format for subsequent use in the new errors module.

The second set of changes, to the RIVPACS software package as a whole, is the development of the errors module. This module allows for statistical comparison of the ecological quality of sites through detection of significant differences in Ecological Quality Index (EQI) values. It also enables sites to be assigned to ecological quality bands in a probabilistic manner and permits assessment of the probability that any two separate samples belong to the same or different bands.

The new module allows considerable flexibility in the form of data entry, spatial and temporal comparisons which can be made, the use or otherwise of bias terms and the value ranges of the quality bands which can be used. It accepts data from existing files or allows for manual data entry on screen.

The revised document was supplied to Dr Dines for circulation.

2.2 Scoping study for Phase 2

2.2.1 Review of the extent, structure and accessibility of 1995 GQA

The 1995 GQA biological and environmental data is held by Thames Region in a central data-base at Reading.

The extent and structure of the 1995 GQA data have been discussed with Paul Logan and John Steel (Thames Region) during a visit to the River Laboratory and with Julie Jefferies (Thames Region) by telephone. The accuracy and availability of the data were discussed with Dr R A Sweeing and Dr J Murray-Bligh (Thames Region) and Dr W Walley (University of Stafford).

The data-base holds information on in excess of 11,000 sites, the vast majority of which were sampled twice. In simplistic terms the data-base comprises a site file and a species file.

The site file comprises a site code, which is a sequential acquisition number, a watercourse name, which is a text string, a location name, which is also a text string and a National Grid Reference (ten character alphanumeric). It also holds the time invariant site data (eg altitude) required to run RIVPACS. The site code contains no encoded spatial information such as hydrometric area, catchment or watercourse code.

The sample file contains a linking site reference field, an eight or nine digit sample code, of which the first two characters identify the Agency region containing the site, the next two characters contain the year of sampling (i.e. 1995 = 95) and the remainder are a numeric sample code. The sample code contains no encoded spatial information other than region number. The sample file also contains the list of macro-invertebrate families known to have been present in the sample, encoded as eight digit Furse/Maitland numbers, and the time variant data required to run RIVPACS (eg water width).

The accuracy of the information held in the data-base is currently being meticulously scrutinised by Dr Walley who requires reliable data for commissions he is undertaking for the Agency. It is clear from discussions with him that the data-set still contains a substantial number of errors in both the biological and environmental data. It is difficult to identify a point when the data validation will be effectively completed. The likely controlling factors are the funding available to Dr Walley to allow him to continue his scrutiny and the speed with which regional staff provide amended information to correct the errors he detects.

A key requirement of the second Phase of the current study is that each site and sample are correctly spatially referenced. Unique identifiers are required for each sample and, ideally, these identifiers should contain encoded spatial information, linking the sample to one or more administrative regions (eg Environment Agency region, Hydrometric Area etc) and to the site, reach and river system in which they were collected. They should also cross-reference to the equivalent chemical data-sets.

It is not the purpose of the current R&D programme to develop data-bases and GIS but attaining its objectives would be helped greatly by the availability of both reliable data and appropriate mechanisms for its storage, extraction and manipulation.

No difficulties are envisaged in making the data-sets available to the organisation undertaking Phase 2 of the current study, providing that organisation has a suitable data-base system to accept, store and manipulate it.

2.2.2 Preparation of a discussion document

The first draft of the discussion document was discussed at the meeting between Dr Dines and IFE in mid-August.

As a result a broad range of changes and additions were made to the document and a second draft was produced (Appendix II) containing fifteen preliminary options for utilising the data. The reader is referred to Appendix II for full details.

2.2.3 Production of a circulation list for the discussion document

A circulation list for the discussion document has been agreed with Dr Dines. It comprises the following:

All members of the project board:

Dr R A Sweeting	Chairman, Topic Leader
Dr R A Dines	Project Leader
Dr A J D Ferguson	Project Executive, Head Office rep.
Dr J Murray-Bligh	Technical User
Mr B Hemsley-Flint	RIVPACS Project Manager
Mr D Lowson	SEPA/NI representative
[Mr M T Furse	IFE]

All Environment Agency Regional Biologists (excluding those above)

G P Green	South West
F Jones	Welsh
J Steel	Thames
S Chadd	Anglian
S Howard	Midlands
E Fisher	North West

Dr E A Chalk	(Environment Agency, North East Region)
Mrs A Lewis	(Environment Agency, North East Region)
Dr G Rutt	(Environment Agency, Welsh Region)
Mr A Warne	(Environment Agency, Anglian Region)
Professor R Edwards	(Environment Agency Board Member, Welsh Region)
Mr P Hale	(Department of the Environment, Northern Ireland)
Dr W Walley	(Stafford University)

2.2.4 Circulation of the discussion document

The document will be circulated as soon as both the Project Leader and the IFE have fully agreed its contents. It is anticipated that this will be in the latter half of September.

3 PLANS FOR THE NEXT REPORTING PERIOD

The next reporting period is from 1st September 1996 to 30th November 1996.

3.1 Incorporation of error terms into RIVPACS III

The following agreed tasks will be initiated:

- Develop and test software to derive confidence limits for EQIs using the outputs from R&D Note 412 in accordance with plan.
- Develop and test software to place sites in quality bands with attached probabilities of band membership, using the GQA banding scheme as default but with option for alternative banding schemes.
- Develop and test software to test for significance and magnitude of change in quality between sites or over time
- Modify the RIVPACS III user manual to incorporate to integrate the error modules developed in the previous tasks

These tasks are scheduled for completion by the 15th December 1996. The other task scheduled for completion by that same date is:

- Revise the RIVPACS III user manual to incorporate the new procedures

3.2 Scoping study for Phase 2

During the next reporting period all outstanding tasks from the previous quarter will be completed.

The following agreed tasks will also be completed:

- Collate returns from discussion document and prepare final list of priority items in conjunction with the three following steps:
- Identify availability of other data-sets required by priority options, and costs involved in their acquisition, manipulation and use.
- Examine compatibility between 1990 and 1995 data if consultation indicates that this is a priority option
- Ensure that the selected options will meet Agency business needs by discussion with Project Board and other relevant Agency staff and that the work required is feasible within the projected timescale and budget.

4 FACTORS WHICH MAY AFFECT THE ATTAINMENT OF ANY TARGETS OR TIMESCALES.

The software development work is currently on schedule but work on the Scoping study is approximately three weeks behind schedule.

Delays have arisen because of the late start to both the current project and other Agency projects involving the same IFE staff. These delays lead to inevitable difficulties in re-timetabling staff input when their time has been allocated to other non-Agency contracts in the interregnum. This is particularly problematic when attempts are made to compensate for the late start to projects by compressing their timetable for completion.

It is currently hoped to make-up lost time and meet the agreed completion times for each of the two major elements of the work programme.

5 FINANCE

The work conducted to date has been within the agreed budget. A financial summary for the reporting period and end-of-year out-turn may be obtained from the IFE Finance Office approximately two months after the end of the period/financial year in question.

6 REASONS FOR ANY LIKELY UNDER OR OVERSPEND OF BUDGET

No under or overspend of the budget is currently anticipated.

The most realistic risk of an overspend is likely to arise from problems in software programming. Estimating the time needed to get a fully tested new piece of software to the operational stage is notoriously difficult. In the current instance the need to use the complex software package PANEL to make the new error module compatible with RIVPACS III exacerbates this risk.

7 OTHER MATTERS

No other issues have arisen which require reporting upon here.

APPENDIX I

Plan for software enhancements to RIVPACS III to incorporate errors in biological quality assessments

SUMMARY

RIVPACS III will be enhanced to incorporate error terms for the O/E ratios (also termed Environmental Quality Indices (EQI)), based on the observed (O) and expected (E) occurrence of BMWP families. It will include options to assign individual samples probabilistically to defined biological quality bands and assess the difference or change in O/E and quality band between pairs of samples.

The error terms and computational procedures used will be those derived by the IFE as summarised in NRA R&D Note 412 'Biological Assessment Methods : Controlling the quality of biological data. Package 1 The variability of data used for assessing the biological condition of rivers ' (April 1995).

These enhancements to RIVPACS III will all be done using same style of screen question/answer interface panels as in the current version of the software.

Two case types of error assessment will be allowed for:

Case type A : Error assessments for individual sample O/E ratios

and Case type B : The comparison of paired sample O/E ratios

1. Case type A : Error assessments for individual sample O/E ratios

This is for the situations where the user needs to assess the EQI, its confidence limits and the quality banding for an individual single or combined season sample, or for a whole series or file of such individual samples, one at a time. The new output will include standard errors and (95%) confidence limits for each sample's O/E ratio and optionally its probability of belonging to previously-defined biological quality bands. There will be an option to read existing RIVPACS II and RIVPACS III output files containing O, E and O/E values (format as in Table 5.9.1 of RIVPACS III manual) and add error terms and banding probabilities

For such case type A assessments, a modified form of the RIVPACS main menu 'Prediction' option is run. The main changes occur between old screen Panels P5 and P6 by the addition of new screen panels to allow the optional specification of biases (Panels P5a - P5c) and quality band limits (Panels P5Xa-P5Xd) These new Panels need to be read in conjunction with the screen Panels on pages 134 - 142 in the RIVPACS III manual. Neither new option is compulsory and each can be bypassed by answering "No". If either are wanted they can each be entered manually or from new forms of input files. The EA's 1995 GQA biological banding limits will form the default banding scheme.

Two of the previous RIVPACS 'Prediction' output files are altered. The original output files requested in former Panels P20 and P21 are merged to form a new output file ('Prediction' option Output file Type 1). The first part of each output line in this file has been kept identical to the earlier form of P20 (specified in Table 5.9.1 of the RIVPACS III manual) so that the new outputs can be used with any old programs developed by the NRA/EA for data manipulation and, equally, old output files of this form from either RIVPACS II or III can still be used in many of the new routines developed in the errors module.

'Prediction' option Output file Type 2 contains information on the confidence limits of EQI estimates, number of quality bands (up to 10) set by the user, and the probabilities that a site with that EQI belongs to each of the those quality bands. Two blocks of statistics are produced when error terms are available, one where observed values, and hence EQIs, are uncorrected for bias and one where they are corrected. Each line ends with the bands limits used.

Examples of the new output files are enclosed together with changes that have occurred in the Panel sequence P15-P22 to accommodate the new forms of output.

We also enclose some examples of what parts of the new 'Prediction' option screen output listing file will look like.

2. Case type B : The comparison of paired sample O/E ratios

Assessments of a significant change in O/E and quality band between two samples will be made using a new errors module accessed by a new option called 'COMPARE' in the RIVPACS main menu.

The two samples being compared can be from the same site at different times, different sites at the same time or a variety of other options.

A proposed new set of screen Panels are laid out in what we hope is a fairly comprehensible format (see attached Panels B1-B13). The O and E values, the season(s) involved, the optional biases and any optional quality band limits can each either be read from files, or input manually from the screen. In particular, the O and E values and seasons involved (code 1-7) can be read from old RIVPACS II or III 'Prediction' Panel P20 output files, or these parameters and the bias values can all be read from the proposed new Output Type 1 files of the 'Prediction' option.

Comparisons of samples can be made by reading their details from either one or a pair of files. The comparisons are made as follows:

- (a) One File used : compare sample 1 with sample 2, sample 3 with 4, 5 with 6, etc
- (b) Two Files used. compare sample 1 in file 1 with sample 1 in file 2
 compare sample 2 in file 1 with sample 2 in file 2, etc

The user must ensure that their samples are ordered within the file(s) to conform to one of these two options. If the O, E and sample details are entered manually within the 'Compare' option (Panels B3-B4), then one type (a) file is created.

If old-style P20 files are used for the comparisons, Panels B7-B9 allow a additional bias term to be specified, if required. The only stipulation is that the bias must be common to all samples in the files. If sample-specific biases are required then these must first be added to the end of each line of these old output files, outside of RIVPACS.

Panels B10-B12 allow biological quality bands to be set, using band limits of the user's choice (up to ten). The bands limits can be entered manually or from files. If bands limits are provided, then the 'Compare' module will give a probabilistic assessment of the band of each of the two samples being compared and the probability they are from the same or different bands.

We have not yet devised the precise form of the output but provide you with a list of the type of contents we envisage (Panel B13).

The expected approach for making comparisons between EA GQA sites in say 1990 and 1995 would be to first run the RIVPACS 'Prediction' option to derive O, E, and O/E values for the 1990 samples in one file and the 1995 samples in another file. The new RIVPACS 'Compare' option would then be used to read the matched pairs from the 1990 and 1995 output files and assess the change in quality. The user would have to ensure that the sites were stored in the same order in each output file. Alternatively, two existing P20 output files could be used as input to the 'Compare' module, providing any optional biases were specified separately and the sites were in the same order in each of the two files.

3. Bias

The user will be allowed to specify the estimated average size of any bias (under-estimation) in the number of taxa recorded due to sample processing errors (new Panels P5a - P5c). This bias could be set to the working error limit of the EA's internal AQC (usually 2.0 we think). Alternatively, it could be set to the actual average difference between the number of taxa recorded as present in a sample by an external quality audit (QA) scheme and the original number of taxa recorded for the sample (ie the net difference between QA 'gains' minus 'losses'). R&D Note 412, found that 'losses' (ie taxa recorded as present by the Environment Agency but not found by the IFE QA) were relatively small and, as recommended, they are not allowed for separately in the assessments of errors. The bias value is used to simulate the number of missed taxa (assuming a Poisson statistical distribution).

In both the 'Prediction' option (Case A) and 'Compare' option (Case B) the bias can be either a constant for all samples or must be specified for each individual sample (This allows groups of samples with differing biases to be stored and analysed within the same file, which could be very convenient to the EA)

Assessments of site quality will be provided which both ignore and incorporate the effects of these biases

4. Quality bands

Quality band limits for EQI for each of BMWP Score, number of taxa and ASPT can optionally be set and used to classify the samples probabilistically into quality bands (eg bands a,b,c,d,e,f). The EA's 1995 GQA biological banding limits will form the default banding scheme. In addition to banding assessments based on each of ASPT, number of taxa and perhaps BMWP score, an overall quality band for a sample, based on the lower of its two bandings based on ASPT and number of BMWP taxa will be provided. This is the overall banding scheme recommended by the EA and which we will refer to as the MINTA banding scheme (ie MINimum of band based on EQI for Taxa and ASPT).

5. Batch file and manual data entry

The batch file (automated) mode of using RIVPACS will be enhanced to include the new options for both the 'Prediction' and new 'Compare' module. At this stage, it is anticipated that it may not be possible to maintain total compatibility with the old format of Batch files, which may have to be edited to use them in the new version of RIVPACS.

An alternative input method for assessing change will allow manual data entry of observed and expected BMWP index values for individual sample and for pairs of samples.

THE PLAN

This plan should be read in conjunction with the RIVPACS III manual (NRA R&D Note 454 - November 1995). In particular, repeated reference is made to the RIVPACS software screen panels (P1-P28) for the prediction option, which are shown in Appendix 4 of that manual.

Extra prediction menu panels are denoted by P5a, P5Xa, etc.

---> P6 means proceed to panel P6, etc.

The RIVPACS Main Menu panel (M1) will now consist of the following options :

M1 : Main menu	
1. Prediction	---> P1
2. Classification	---> C1
3. Compare	---> B1
4. Defaults setup	
5. Batch setup	
6. Help	
7. Exit	

Panel P4 : Biological data source

No change

Note: If the biological data are not held on file then, as only the expected (E) BMWP index values can be calculated, these will be output without an error term.

Panel P5 : Biological Filename for each season No change

NEW P5a : Do you have bias estimates for the number of missed taxa and wish to use then in calculating EQI values and their errors ?

- 1. No ---> P5Xa
- 2. Yes, input from screen ---> P5b
- 3. Yes, input from file ---> P5c

NEW P5b : Average underestimation of number of taxa for each season involved:

- Spring :) Highlight relevant season(s)
- Summer :) User enters fixed bias for each relevant season
- Autumn :) ---> P5Xa

NEW P5c : Enter name of File containing biases for each season:

Note : Bias file will contain

either 1. Fixed bias

Spring bias, Summer bias, Autumn bias (-9 if unknown or not needed)

or 2 Sample-specific bias

Sample code (20 chars) Spring bias, Summer bias, Autumn bias

---> P5Xa

Option 2: a sample-specific bias allows the user to have batches of samples with different average biases all in one file.

NEW P5Xa : Do you wish to assign sites to biological quality bands ?

- 1 No ---> P6
- 2 Yes, use defaults (EA's GQA) ---> P6
- 3 Yes, input band limits from screen ---> P5Xb
- 4. Yes, provide band limits from file ---> P5Xd

NEW P5Xb : Enter number of biological quality bands (2-10) :

---> P5Xc

NEW P5Xc : Enter lower limits of EQI of each quality band for each of BMWP Score, Number of Taxa and ASPT :

Band	a	Score	Taxa	ASPT
	b	{cells where limits are required will be highlighted}		
	c			
	d			
	e			
	f			
	...			

Note : Missing limits (or -9) treated as indication of not requiring bands for that EQI

--> P6

NEW P5Xd : Enter name of File containing band limits: ----> P6

Panels P6-P19 of RIVPACS III No change

Panel P1422 : Panels P14-P22 as shown in the RIVPACS III manual are all actually covered in one screen panel in the RIVPACS III software. This panel will hereafter be referred to as Panel P1422.

The questions in former Panels P20 and P21 will be replaced by the following two new parts of Panel P1422 :

NEW Panel P1422

Do you want the screen output to go to a disk file ? (Yes or No) : If Yes, enter Filename

Are the predicted taxa to be listed on the screen ? (Yes or No) :

Do you want create an ASCII file of the raw environmental data with the predicted BMWP score, number of taxa and ASPT appended ? (Yes or No) : If Yes, enter Filename

Do you want create an ASCII file of the BMWP statistics containing the observed and expected BMWP index values together with the derived O/E ratios and the standard errors and confidence limits of those ratios ? (Yes or No) : If Yes, enter Filename

(Creates output of form shown in Example 'Prediction' Output file Type 1 below)

Do you want create an ASCII file of the site(s) biological bandings, together with the BMWP index statistics ? (Yes or No) : If Yes, enter Filename

(Creates output of form shown in Example 'Prediction' Output file Type 2 below)

Panel P23 (last) : Predicted probability at which to stop listing taxa : No change

Return to panel M1 : Main menu

Example of New Screen Output on the Biological quality banding for a site :

****** THE OLD FORM OF OUTPUT IN RIVPACS III :**

(a) with biological data :

BMWP score, number of taxa and ASPT.

	mean	sd	lcl(95%)	ucl(95%)	obs	obs/exp
BMWP score	220.2	19.163	182.7	257.8	212.	0.96
No. taxa	33.0	2.627	27.8	38.1	32.	0.97
ASPT	6.67	0.2135	6.26	7.09	6.63	0.99

(b) without biological data :

BMWP score, number of taxa and ASPT.

	mean	sd	lcl(95%)	ucl(95%)
BMWP score	220.2	19.163	182.7	257.8
No. taxa	33.0	2.627	27.8	38.1
ASPT	6.67	0.2135	6.26	7.09

This will be replaced by :

****** NEW FORM OF OUTPUT OF O, E, O/E AND O/E ERRORS :**

SD, lcl and ucl denote Standard deviation and lower and upper 95% confidence limits

(a) with biological data and estimated bias :

BMWP Statistics :

Estimated BIAS (Mean underestimation of Number of Taxa) : 0.74

	Exp(E)	----Bias Uncorrected----					-----Bias Corrected-----				
		Obs(O)	O/E	SD	lcl	ucl	O	O/E	SD	lcl	ucl
BMWP score	220.2	212	0.96	0.12	0.71	1.25	219	0.99	0.13	0.72	1.28
No. tax	33.0	32	0.97	0.09	0.79	1.18	33	1.00	0.10	0.81	1.21
ASPT	6.67	6.63	0.99	0.05	0.89	1.09	6.64	0.99	0.06	0.88	1.10

(b) with biological data but bias unknown :

BMWP Statistics :

Estimated BIAS (Mean underestimation of Number of Taxa) : UNKNOWN

	Exp(E)	----Bias Uncorrected----				
		Obs(O)	O/E	SD	lcl	ucl
BMWP score	220.2	212	0.96	0.12	0.71	1.25
No. tax	33.0	32	0.97	0.09	0.79	1.18
ASPT	6.67	6.63	0.99	0.05	0.89	1.09

c without biological data :

BMWP Statistics :

	Exp(E)
BMWP score	220.2
No. taxa	33.0
ASPT	6.67

**** NEW SCREEN OUTPUT INFORMATION ON BANDING :

Biological Banding of the site (Bias Uncorrected):

Band	-- SCORE--		-- TAXA --		-- ASPT -		MINTA
	Lower Limit	Prob %	Lower Limit	Prob %	Lower Limit	Prob %	Prob %
a	0.77	88	0.85	91	0.91	83	79
b	0.57	12	0.70	9	0.82	17	21
c	0.40	0	0.55	0	0.73	0	0
d	0.00	0	0.00	0	0.00	0	0
e	0.00	0	0.00	0	0.00	0	0
f	0.00	0	0.00	0	0.00	0	0

Most Prob A A A A

Biological Banding of the site (Bias Corrected (WHEN GIVEN)):

Band	-- SCORE--		-- TAXA --		-- ASPT -		MINTA
	Lower Limit	Prob %	Lower Limit	Prob %	Lower Limit	Prob %	Prob %
a	0.77	89	0.85	93	0.91	84	81
b	0.57	11	0.70	7	0.82	16	19
c	0.40	0	0.55	0	0.73	0	0
d	0.00	0	0.00	0	0.00	0	0
e	0.00	0	0.00	0	0.00	0	0
f	0.00	0	0.00	0	0.00	0	0

Most Prob A A A A

Lower Limit is the lower limit of the O/E ratio for that band.

Prob% for a band is the percentage probability that the site is in that band's quality range.

MINTA is the overall site banding based on the lower of the bandings based on number of taxa and ASPT.

**** OBVIOUSLY THE NUMBER OF BANDS AND THEIR RANGES SHOWN ON THE PRINTOUT
 **** WOULD BE THE SAME AS THOSE CHOSEN BY THE USER DURING THE MENU SEQUENCE

Example 'Prediction' option Output file Type 1

Camel Pencarrow Br.	212	32	6.63	220.2	33.0	6.67	0.96	0.97	0.99	7	1
(Line 1 continuation)	0.74	0.12	0.71	1.25	0.09	0.79	1.18	0.05	0.89	1.09	
(Line 1 continuation)	0.99	0.13	0.72	1.28	1.00	0.10	0.81	1.21	0.99	0.06	0.88 1.10
Annan Williamwath Br	177	29	6.10	184.7	29.0	6.37	0.96	1.00	0.96	7	1 ...
Prome Frampton	210	37	5.68	212.4	35.8	5.92	0.99	1.03	0.96	7	1 ...

**** SEQUENCE OF ABOVE COLUMNS :

- **** First the following, EXACTLY as for existing type of output file specified in Panel P20 and Table 5.9.1 of RIVPACS III manual :
- **** Sample Code(20 digit) Observed Score, Taxa, ASPT; Expected Score, Taxa, ASPT;
- **** Observed/Expected(O/E) for Score, Taxa, ASPT; Season Code(1-7), Suitability code(1-5,7-9);
- ****
- **** Then the following new columns :
- **** Estimate of average bias, if any, for the season(s) combination requested
- **** Standard Deviation, lower and upper 95% Confidence limits of the uncorrected
- **** O/E values for Score, then Taxa and finally ASPT
- **** and finally the bias-corrected O/E value with its Standard Deviation,
- **** lower and upper 95% Confidence limits for Score, then Taxa and finally ASPT.

As the first part of each line of this output file is exactly the same as the previous RIVPACS II and RIVPACS III output file for O, E and O/E values for each site, this new file can still (almost certainly) be read by old programs used by the former NRA and others to read these output files and produce summaries of the qualities of a batch of sites.

Example 'Prediction' option Output file Type 2

There are 8 output lines per site.

Output shown for case of 4 bands (a,b,c,d) for compactness

Camel	Pencarrow	Br.	7	1	S	BU	212	220.2	0.96	0.12	0.71	1.25	4	a	88	12	0	0	0.77	0.57	0.40	0.00
Camel	Pencarrow	Br.	7	1	T	BU	32	33.0	0.97	0.09	0.79	1.18	4	a	91	9	0	0	0.85	0.70	0.55	0.00
Camel	Pencarrow	Br.	7	1	A	BU	6.63	6.67	0.99	0.05	0.89	1.09	4	a	83	17	0	0	0.91	0.82	0.73	0.00
Camel	Pencarrow	Br.	7	1	M	BU							4	a	79	21	0	0				
Camel	Pencarrow	Br.	7	1	S	BC	219	220.2	0.99	0.13	0.72	1.28	4	a	89	11	0	0	0.77	0.57	0.40	0.00
Camel	Pencarrow	Br.	7	1	T	BC	33	33.0	1.00	0.10	0.91	1.21	4	a	93	7	0	0	0.85	0.70	0.55	0.00
Camel	Pencarrow	Br.	7	1	A	BC	6.64	6.67	0.99	0.06	0.88	1.10	4	a	84	16	0	0	0.91	0.82	0.73	0.00
Camel	Pencarrow	Br.	7	1	M	BC							4	a	81	19	0	0				

***** SEQUENCE OF ABOVE COLUMNS :

***** Results are presented uncorrected for bias (BU) in lines 1-4 and corrected for bias (BC), if known, in lines 5-8

***** First line for a site refers to BMWP Score, second line to Taxa and third line to ASPT, each with the following
 ***** Sample Code(20 digit), Season Code(1-7), Suitability code(1-5,7-9), Initial Letter (S,T or A) of BMWP Index
 ***** Column denoting whether the Observed and O/E value are uncorrected (BU) or corrected (BC) for the estimated bias (due to missed taxa)
 ***** Observed value, Expected value, O/E, Standard Deviation, lower and upper 95% Confidence limits of O/E
 ***** Number of bands (K), Letter denoting most probable band, Percentage probability of belonging to each of the K bands
 ***** Lower limit of O/E for each of the K bands

***** The fourth line for a site refers to an overall assessment of site quality, called MINTA, based on using the lower of the bandings based
 ***** on number of taxa and ASPT. It contains, in the appropriate character positions, only :

***** Sample Code(20 digit), Season Code(1-7), Suitability code(1-5,7-9), Initial letter 'M' signifying overall assessment MINTA,
 ***** Number of bands (K), Letter denoting most probable band based on MINTA, Percentage probability of belonging to each of the K bands based on MINTA

***** Lines 5-8 replicate lines 1-4 but corrected for bias

***** Line 6 for corrected O/E for Taxa ends with the derived estimate of the bias or underestimation of the number of taxa present in this
 ***** seasons combination of samples.

Case B : Comparison of paired sample O/E ratios

From RIVPACS main menu M1 select : Option 3 'Compare'.

NEW Panel B1 : Compare Title page
with very brief introduction to method for comparing two samples

NEW B2 : Do you hold the data needed for sample comparison on file(s) ?
(These files must in the format of RIVPACS 'Prediction' option Output file Type 1, containing, at least, the information as in Table 5.9.1 of the RIVPACS III manual)

1. No B3
2. Yes B5

NEW B3 : Enter name for file to hold sample pair details :
(which will be entered manually) —> B4

NEW B4 : Enter details for the next pair of samples to be compared :
(which will be entered manually)

Sample Code (20 Chars)	Sample A xxx	Sample B xxx
Observed BMWP Score	x	x
Observed Number of Taxa	x	x
Observed ASPT	x	x
Expected BMWP Score	x	x
Expected Number of Taxa	x	x
Expected ASPT	x	x
Season(s) samples contributing to the observed values (Yes/No) :		
Spring	Y/N	Y/N
Summer	Y/N	Y/N
Autumn	Y/N	Y/N
Bias for each season used :		
(-9 if bias unknown or to be ignored)		
Spring	x	x
Summer	x	x
Autumn	x	x
	 B6

There will be a facility to edit data on the screen,

to add data on further pairs of samples (A and B)

and to exit this manual data input screen.

This process will make a single file of consecutive pairs of samples to be compared.

NEW B5 : How many files are to be used in these comparisons ? 1 or 2

Enter Name of File 1. x

Enter Name of File 2 x (If required) ----> B6

NEW B6 : Are, or were, the same values for the environmental variables used in deriving the predictions and expected BMWP values for both samples in each pair (as may occur if the paired samples are for the same or very close sites) ?

Y/N

---> B7

NEW B7 : Do you wish to include a bias term in the comparison of O/E ratios ?

No ----> B10

Yes ----> B8

NEW B8 : Bias term to be used ?

1. Use biases already in the input files ----> B10

2. Enter a common bias for all samples in one season of one file ----> B9

NEW B9 : Enter bias term to be used for all the samples, separately for each season and file?

	File 1 (Filename)	File 2 (Filename)	
Spring	X	X	(cells where biases are required will be highlighted)
Summer	X	X	
Autumn	X	X	

----> B10

NEW B10 : Do you wish to assign samples to biological quality bands and assess the probability of a change in band for each pair of samples ?

1. No ----> B13

2. Yes, use default limits (as for EA GQA) ----> B13

3. Yes, Input bands limits from screen ----> B11

4. Yes, Read bands limits from file ----> B12

NEW B11 : Set biological quality bands

Enter number of bands : x

Enter lower limits of EQI of each quality band for each of
BMWP Score, Number of Taxa and ASPT :

Band	a	Score	Taxa	ASPT
b		(cells where limits are required will be highlighted)		
c				
d				
e				
f				

Note : Missing limits (or -9) treated as indication of not requiring bands for that EQI

Enter name of file to contain these band limits . xxxxxxxx xxx

---- B13

NEW B12 : Enter name of file containing the required quality band lower limits :

xxxxxxx.xxx

---- B13

NEW B13 : OUTPUTS USING REQUIRED OPTIONS AND FORMATS :

The output for the comparison of each pair of samples (A and B) will contain

- (1) O, E and O/E; standard errors and confidence limits for O/E
This will be for both uncorrected and corrected for the assumed bias (if applicable)
- (2) Probability of being in each quality band separately for both the samples in the pair being compared
- (3) Estimates, standard errors and confidence limits for the difference in O/E between the two samples in each pair
(If the confidence limits include zero, then conclude there is no statistically significant difference in biological quality between the pair of samples)
- (4) Two-way table indicating the probability that
Sample A was in band j (j = a,b,c,d,e,f etc)
while sample B was in band k (k = a,b,c,d,e,f etc)
and hence an estimate of the probability that the two samples are from the same quality band
- (5) Details of the biases, band limits and data files used

---- Return to RIVPACS IV Main Menu M1

APPENDIX II

ANALYSIS OF 1995 BIOLOGICAL SURVEY DATA - LISTING OF OPTIONS FOR ADDITIONAL USES OF THE DATA-SET

INTRODUCTION

During the 1995 General Quality Assessment (GQA) macro-invertebrate samples were collected from a substantial number of running water sites throughout Great Britain. The exact number is not known but it is likely to be equal to or greater than the 8,600 sites, including 7,633 in England and Wales, reported to have been sampled in the 1990 River Quality Survey (RQS). Many sites were common between the two surveys.

Supporting environmental data were collected from each biological sampling point, including National Grid Reference, distance from source, altitude, slope, discharge category, width, depth, surface velocity and substratum composition.

During both surveys substantial chemical sampling also took place. Chemical and biological sampling sites were often not at the same location although attempts have been made to relate pairs of chemical and biological sites to defined sections of river, or "reaches". Some chemical sites have been matched to more than one biological site and vice versa.

The macro-invertebrate data collected from each site were used to evaluate the biological condition (= ecological quality) of the reach. The software package, RIVPACS (River Invertebrate Prediction and Classification System) was used to make evaluations. These were based on the ratios of the observed to expected (ie RIVPACS predicted) Biological Monitoring Working Party (BMWP) index values. Separate ratios were calculated for BMWP score, number of scoring taxa and Average Score Per Taxa (ASPT). Each ratio was termed an Environmental Quality Index or EQI. In this process expected index values were derived through use of the environmental data collected for each site, including measured or derived values of total alkalinity.

EQIs were sub-divided into value ranges or bands of ecological quality. EQI bands for individual BMWP indices can be integrated into an overall band of ecological quality for a site. Different band widths and procedures for their amalgamation were used for the 1990 RQS and the 1995 GQA.

Between these two national surveys the National Rivers Authority (NRA) commissioned the Institute of Freshwater Ecology (IFE) to undertake research on the errors, variation and biases associated with collecting, identifying and interpreting the biological material and measuring the environmental data used for assessing the condition of reaches.

This research is now complete and has provided mechanisms for attaching variance terms to EQIs, for assigning sites to bands of ecological quality in a probabilistic manner and for assessing whether there has been a statistically significant change of ecological quality and banding between sites or at the same site over time.

RIVPACS III, the version used in conjunction with the 1995 GQA, is currently being modified by the IFE to incorporate a module for calculating error terms and for making statistical comparisons between sites.

The data collected during the 1990 and 1995 survey are stored in a central data-base held by the Thames Region of the Environment Agency. Beyond the use of the biological data for evaluating the condition of sites, no other systematic national use of the extensive data holding, which includes family occurrences and, often, abundance values, has yet been made.

This document includes a preliminary listing of the potential further uses which can be made of the data and forms the initial contribution to a scoping study on the subject. The following list is not considered to be definitive. It is intended as a discussion document for circulation within the Environment Agency and Agency staff are invited to comment upon the options presented within it. They are also invited to submit outlines of alternative suggestions and new lines of research which will enable the Agency to maximise the value of the data in support of their core functions.

This scoping study and the upgrading of RIVPACS III to incorporate the "errors module" form the two parts of an Agency R&D project with the IFE entitled "Analysis of 1995 Biological Survey Data and RIVPACS Upgrade".

OPTIONS FOR FURTHER USE OF THE 1995 GQA MACRO-INVERTEBRATE DATA

Distribution of taxa in relation to other factors

The use of the 1995 GQA macro-invertebrate sampling programme only to index and band the ecological quality of sites fails to take advantage of the substantial information held on the distribution and relative abundance of the full range of aquatic macro-invertebrate families

OPTION 1 To obtain a better understanding of the environmental factors which govern the distribution of taxa.

Knowledge of the environmental range and tolerances of individual taxa is fundamental to interpretation of the results, not only of GQAs but of a wide range of environmental stresses and pollution incidents investigated by the Agency. This is demonstrated by the development of specialised algorithms to detect the impact of specific stresses such as acidification and diffuse agricultural pollution. Yet no clear documentation exists which draws together the known ranges and tolerances of in a coherent and concise fashion.

The 1995 GQA data provides the ideal data-set for developing the basic framework of such a document at the BMWP family level which can later be amplified with specific species level information from other sources.

The most relevant factors for each species are likely to be:

- RIVPACS predictor variables
- Pairs and other combinations of RIVPACS variables
- Chemical determinands (from the chemical survey and other routine Agency analytical programmes)
- Site/catchment geology, soil type and land cover
- Season

Impact of low flows

One environmental factor which has assumed particular concern in recent years is low flow.

OPTION 2 To examine the impact of low flows on the distribution, frequency and abundance of individual taxa and on the ecological quality of individual sites.

The programme would examine spatial differences between rivers differentially impacted by the drought in 1995 and temporal differences between sites sampled in both 1990 and 1995.

Selection of rivers would be based on directly gauged discharge, wherever possible, and comparisons between the 1995 annual mean flows and long term averages for the same gauge sites. It would take account of differences in analytical quality control and regional audit results in different regions and between surveys. It would also need to take account of any compounding, independent environmental stress in the selection of sites for comparison.

Evaluation of temporal and spatial changes in the biological condition of sites

The main purposes of national surveys are to periodically evaluate the condition of watercourses on a national basis and to assess changes in condition between surveys. The development of firstly the BMWP score system and secondly RIVPACS has provided far greater credibility to biological assessments of watercourse condition than had been achieved prior to 1990.

OPTION 3 Evaluation of the distribution of the ecological quality of sites in the 1995 GQA in relation to a range of environmental factors.

Option 3 provides a means of making spatial comparisons between sites sampled within the same year. Now the development of the errors module within RIVPACS allows more meaningful temporal comparison of biological samples than has been possible hitherto.

OPTION 4 Comparison of samples collected at the same sites in the 1990 RQS and the 1995 GQA in order to detect and quantify significant changes in the ecological quality of sites.

Some of the error terms developed for the NRA/Environment Agency by IFE have already been adopted for use in the report on the 1995 GQA, and have been used to present changes in the ecological quality of sites between 1990 and 1995. However the current project provides scope to examine trends and changes in far greater detail than is possible in the GQA report.

Comparisons can be made between sites sampled in different seasons or combination of seasons providing sampling has been undertaken using standard RIVPACS methodology. If required, comparisons can take account of different known levels of sorting and identification efficiency, as assessed by internal and/or external auditing of performance. The extent of changes can be expressed at different levels of probability. The analysis of change will provide a more thorough assessment of local and national trends than was previously possible.

Once meaningful spatial and temporal comparisons are available on a site-by-site basis then any regional trends or temporal changes, including seasonal changes of quality within a given survey year, can be examined in relation to a number of external factors. This will provide a sounder basis for use of the data for other purposes, such as the development of Local Environment Agency Plans (LEAPs).

Amongst the many background variables against which the ecological quality of sites can be assessed, the following are prime candidates:

- River type

Are problems concentrated in particular types of river such as low alkalinity moorland streams, small lowland watercourses, chalkstreams, individual RJVPACS groups, etc.?

- Distance from source.

Is there a tendency for the most significant changes to be taking place near to or far from source? Are there particular problems with headwaters, middle reaches or large slow flowing rivers?

- Geology and soil type

Is the poorest ecological quality and are the greatest temporal changes occurring in catchments of particular geological or soil types? If so, what are the underlying causes?

- Land cover

Diffuse and point source pollution from agriculture can have important repercussions for the biological condition of watercourses. Can recent trends in changing ecological quality of streams be associated with particular types of agriculture and changing patterns of land cover? If so, what are the implications for targeting pollution control and for managing sustainable agriculture within an ecologically acceptable framework?

- River Habitat Survey class

In addition to national surveys of the ecological and chemical quality of rivers, the Environment Agency have made substantial investment in a new form of national survey, the River Habitat Survey. RHS is a more holistic appraisal of the condition of the entire river corridor, including the water course and its riparian zones. It is strongly conservation-centred. In order to maximise the return from the investment in both the GQA and River Habitat Surveys it is suggested that changes in ecological quality of the watercourse are analyzed in relation to the results of the RHS programme in order to examine the links between the two.

Trends in family loss/gain with changes in biological condition

Closely allied to changes in indices of overall biological condition are changes in the occurrence of individual families.

The data-sets for the 1990 RQS and the 1995 GQA provide information on changes in assemblage composition as well as overall biological condition.

OPTION 5 An examination of the relationship between temporal changes in ecological quality and the losses, gains and changes in abundance of individual families.

Which families are lost and gained as the ecological and/or chemical quality of watercourses change? Are the gains as conditions improve mirrored by identical changes as conditions worsen or are the rates of deterioration different from the rates of recovery? Are there different regional patterns of taxon losses and gains for the same degree of change in ecological quality? Which taxa appear to be declining or increasing in frequency of occurrence and can these changes be linked to quantifiable changes in features of their immediate habitat or of the site catchment?

How can these changes be used to predict changes in the composition or relative abundance of faunal assemblages in response to an anticipated change in environmental conditions (eg the improvement of effluent quality from a known discharge, reduction in flow due to abstraction)?

Incorporation of GQA data in the Countryside Information System (CIS)

In many of the previous options reference has been made, directly or indirectly, to geology, soil type and land cover. The examination of the role of these factors is best achieved by use of a Geographic Information System (GIS). The cost of both acquiring and holding geological, soils and land cover data and for developing the GIS would be relatively high. An alternative is to make use of the Institute of Terrestrial Ecology (ITE) Land Classification and the CIS software package.

OPTION 6 To incorporate the results of the 1995 GQA in the Countryside Information System (CIS)

The CIS is a software package developed largely through Department of the Environment (DoE) funding and is likely to be influential in their policy forming procedures. It was originally designed to carry, display and analyze the results of the Countryside Survey 1990. Since then its remit as a data platform has considerably widened.

The system is based on each 1km square in Britain being allocated to one of 32 land classes devised by the ITE. Specific survey and census data can be held for each square or each square can be assigned the Land Class mean value for an attribute (eg average percentage cover of wheat or average frequency of occurrence of a given animal).

In addition to carrying summary statistics on the land cover of each class, the CIS can also be used as a mechanism for carrying a substantial range of other land class mean, survey or census statistics, including geographical, ecological, sociological and economic factors. It can also carry the 1990 Land Cover Map of Great Britain, developed by ITE from satellite imagery, and any other information that can be expressed on a 1km square basis. The distribution of taxa and the location of sites and their ecological quality could be mapped for individual 1km squares, or expressed as land class means. Either form of data could be displayed and interpreted against a back-drop of the other forms of data the system can hold.

CIS thus acts as a more accessible and less expensive form of GIS which can not only provide a vehicle for carrying the results of the 1995 GQA but also as a mechanism for interpreting their results in relation to other factors. The inclusion of the results of national river surveys will increase the likelihood that these will be taken into consideration in DoE policy developments.

OPTION 7 Development of theoretical taxon distribution maps within the Countryside Information System (CIS)

The CIS allows data to be held on a given attribute for each 1km square in Great Britain. Amongst the land class mean data which could be held for each square are the probabilities of capture of each family of aquatic invertebrates in any watercourse in that square. The probabilities of capture of taxa in streams of different size categories (eg headwaters, upper reaches, middle reaches, lower reaches) could be held separately. National probability of capture maps could be developed for each principal taxon. These could be compared and contrasted with the observed mean frequency of capture of each taxon in each category of river size in each land class. Areas where particular taxa are under the most severe pressure can hence be mapped.

Given knowledge of the soils, geology, altitude etc of each square, habitat suitability models could be developed which allow more detailed maps of probability of capture under unstressed conditions to be developed, in a manner akin to graphic RIVPACS predictions.

In addition to helping interpret survey data these forms of output would be useful in the broader aspects of Agency work, such as developing LEAPs.

Distribution of taxa in the urban environment

The Countryside Survey series provides a substantial body of information on the state of the British Countryside. However, the surveys paid relatively little attention to large urban areas for which a specific classification system had not been developed.

In an attempt to rectify this omission and to develop a better understanding of the urban environment, the Natural Environment Research Council (NERC) is funding a new study entitled "Environmental Characterisation of Urban Environments". The programme will involve three component institutes of the Centre for Ecology and Hydrology, IFE, which will be responsible for studies of urban waterbodies, ITE and the Institute of Hydrology (IoH).

The main aim of the study is to:

"develop a stratification of urban areas based on geographical, socio-economic and environmental characteristics which takes account of pattern and scale and which will provide a framework and stimulus for urban ecosystem process studies and for the management of urban areas in an ecologically sustainable manner"

The research programme includes the recognition that:

"developing a stratification of urban environments based on an improved understanding of the relationship between occurrence and pattern of particular land and water cover types and their associated floras and faunas is an appropriate first step in the development of a comprehensive urban ecosystem study"

An improved understanding of the processes governing sustainability of the urban environment has practical operational benefits for the Environment Agency. The aims of the NERC research programme would, in turn benefit greatly from the availability of a consistent, quality-controlled data-set of macro-invertebrate information from a wide variety of urban watercourses. The 1995 GQA data can meet that need.

OPTION 8 To develop a sub-set of the 1995 GQA macro-invertebrate survey containing sites in the urban environment and to apply those data to the objectives of the NERC "Environmental Characterisation of Urban Environments Programme" in order "to develop and extend the interdisciplinary knowledge base required to plan and achieve more sustainable urban environments".

The impact of loss of quality of headwaters upon their receiver streams

Recent findings of the "Faunal Richness of Headwater Streams" project have shown that these small watercourses, within 2.5km of source, are in generally poorer biological condition than the downstream reaches that they feed. In reports emanating from that project it has been postulated that the water quality of headwater sites will have a detrimental impact on their receiver streams. The extent to which this is true and the magnitude and nature of that impact upon faunal assemblages can be examined in detail using the results of the 1995 GQA.

OPTION 9 An examination of the relationship between the chemical and ecological quality of headwaters and that of the downstream reaches that they feed

Evaluation of the biological banding of sites

Considerable effort and inter-change of ideas and viewpoints went into the development of bands of ecological quality of sites based upon EQI value ranges for ASPT and number of scoring taxa. Similar attention was given to the integration of the two separate EQI bands into an overall biological banding for the site. A text description was developed for each of the overall bands based on presumed features of the macro-invertebrate assemblages at each band level.

The mathematical band ranges and text definitions devised for the 1995 GQA have not been subjected to an *a posteriori* evaluation of their adequacy for the purposes of the survey.

OPTION 10 To evaluate the performance of the biological banding system devised for the 1995 GQA as a means of assessing the ecological quality of sites and for representing definable changes in the structure of macro-invertebrate assemblages.

This option would also include an analysis of the separate distributions of taxa by chemical and biological bands and also the distribution of taxa by chemical band within each biological band to better understand the relationship between the two banding systems and the mismatches that arise between them.

Comparison will also be made between the biological bands based upon EQI values and the values of the abundance index Q14, included in RIVPACS III, and any other appropriate abundance-based banding system.

Assemblage structure

The 1995 data-set would enable features of assemblage structure other than individual taxon distribution and ecological quality to be examined in relation to environmental factors. Only sites of the best ecological quality would be used in analyses.

OPTION 11 The relationship between environmental factors and family richness

Earlier analyses of the RIVPACS data-set and of the macro-invertebrate data collected as part of the "Faunal Richness of Headwater Streams" project highlighted the fact that several sites had exceptionally high taxon richness, even in relationship to the RIVPACS predictions for sites of their own environmental type. The causes of exceptional species richness are not well understood but are significant in the light of the Agency's duty to "further conservation".

The 1995 GQA data-set, together with supporting environmental and chemical data, provide good opportunities to examine the distinctive features of family-rich sites and to start to develop predictive models.

This programme overlaps with possible research being planned for the development of RIVPACS and may be better undertaken under that heading. Overlap of effort should be avoided and if species richness studies are undertaken under both programmes then they must be carefully planned to be complementary to each other.

OPTION 12 Substrate/habitat diversity in relation to species richness

This is a variation on the previous theme in which the data collected on the relative abundances of four substratum particle categories, as used in RIVPACS, would be examined in relation to the family richness at the site.

OPTION 13 Identification of national reference sites of particularly high taxon richness

Examination of family richness at individual sites could be used to identify national reference sites of high bio-diversity. Representative sites could be selected for all the major RIVPACS groups. The fauna of 1990 RQS samples from these sites, held in store at IFE Wareham could be further examined at species level. Sites could be recommended for notification as SSSIs or for special status under the European Habitats directive or UK Biodiversity Action Plan. They could form the nucleus of regular monitoring programmes akin to that adopted by the UK Acid Waters Monitoring Group or added to the Environmental Change Network (ECN) suite of sites.

OPTION 14 Longitudinal patterns of zonation/community structure

The 1995 survey data could be used to examine the patterns of change in aquatic communities along watercourses and to examine the relevance of current ecological theories to the business needs of the Environment Agency.

OPTION 15 The definition of the environmental niche of individual taxa and faunal assemblages

Multi-variate techniques could be used to determine the environmental niche size/shape of individual taxa and discrete faunal assemblages in a manner akin to the determination of habitat suitability curves or the application of PHABSIM to individual taxa. The extent of overlap between taxa or faunal assemblages, however determined, could be examined by this procedure. The research programme would be targeted at a fundamental understanding of the nature of faunal assemblages rather than at any specific operational requirements of the Agency. However, ultimately, it is through this form of fundamental understanding that the problems faced by the water industry can be best understood and acted upon.

FOOTNOTE

Each of the listed options depends upon the availability of a validated and reliable data-set of biological data. Many also require an equally reliable environmental data-set.

A key requirement is that each site and sample are correctly spatially referenced. Unique identifiers are required for each sample and, ideally, these identifiers should contain encoded spatial information, linking the sample to one or more administrative regions (eg Environment Agency region, Hydrometric Area etc) and to the site, reach and river system in which they were collected. They should also cross-reference to the equivalent chemical data-sets.

As stated above, the development of a multi-functional GIS system would improve the accessibility of the data and options for its analysis.

It is not the purpose of the current R&D programme to develop data-bases and GIS but attaining its objectives would be helped greatly by the availability of both reliable data and appropriate mechanisms for its storage, extraction and manipulation.