

RL/T04053r5/2

4/91

INSTITUTE OF FRESHWATER ECOLOGY
River Laboratory, East Stoke, Wareham, Dorset BH20 6BB

Testing and further development of RIVPACS
A progress report for the period
1st January - 31st March 1991

J.F. Wright, BSc, PhD
M.T. Furse, BSc
D. Moss, BSc, PhD

Project leader: J.F. Wright
Report date: April 1991
Report to: National Rivers Authority
IFE Report Ref: RL/T04053r5/2
TFS Project No: T04053r5

This is an unpublished report and should not be cited without permission, which should be sought through the Director of the Institute of Freshwater Ecology in the first instance.

The Institute of Freshwater Ecology is part of the Terrestrial and Freshwater Sciences Directorate of the Natural Environment Research Council.

INSTITUTE OF FRESHWATER ECOLOGY

TESTING AND FURTHER DEVELOPMENT OF RIVPACS

PROGRESS REPORT FOR THE PERIOD 1st JANUARY - 31st MARCH 1991

1. INTRODUCTION

This research project is in two phases. In Phase 1 there are two specific objectives.

- a) To undertake a comprehensive testing exercise of RIVPACS II to check its robustness and find the best possible combination of environmental features for use in assessing the 1990 River Quality Survey data.
- b) To formulate a series of bands to express river quality in biological terms based upon the observed to expected (predicted) ratio (O/E) obtained from RIVPACS. The bands will be used to summarise the results of the 1990 River Quality Survey and future biological surveys.

This work was planned for the period October 1990 to March 1991. However, some elements of the work, most notably the development of the banding system for observed/expected ratios (objective b) require access to the results of the 1990 River Quality Survey. We hope to receive the first results for a small number of NRA regions in mid April 1991. This part of the programme can then proceed. Details of the progress made in other areas of Phase 1 and of a limited amount of additional work undertaken to compensate, in part, for the inevitable delay in the completion of Phase 1, are given in later sections of this report.

Phase 2, which will commence upon completion of Phase 1, has two further objectives.

- c) to undertake a comprehensive examination of the 1990 River Quality Survey data, and
- d) to undertake future development of RIVPACS.

The details of the research programme for Phase 2 are currently under discussion with the NRA Project Leader.

2. TECHNICAL PROGRESS

2.1 Testing of RIVPACS II

The testing of RIVPACS II includes two components. The first of these is to undertake an assessment of the comparative performance of the six different environmental options for prediction of the fauna at BMWP family level. The objective is to provide guidance on the most appropriate set of environmental features available in RIVPACS II for use in the 1990 River Quality Survey. A full list of the physico-chemical features used in each of the six environmental options is given in the previous progress report.

The assessment involves prediction of the fauna and various biological indices at each of the 438 sites used to construct the system, using the six environmental options. The observed fauna and biological indices at a given site can then be compared with the predictions. This offers a method of determining which set of environmental variables give the most accurate predictions on this internal test.

The comparison between the observed and predicted fauna (and biological indices) at BMWP family level (3 seasons combined) was carried out for each of the 438 sites early in the contract and described in the first progress report. The relative merits of the six environmental options have now been examined in detail and a conclusion has been reached on the most appropriate option for use in the 1990 survey. See Section 3 for details.

The second component of the testing exercise is an examination of the robustness of the prediction system. The BMWP family level (3 seasons) predictions for the 438 sites with RIVPACS II have also been used for this purpose.

Printouts which list the observed/expected ratios for number of BMWP families, BMWP score and ASPT for all six environmental options and all 438 sites were prepared in order to highlight sites which exhibit either high or low observed/expected ratios. The occurrence of extreme ratios was also assessed across each of the 25 TWINSPAN groups.

Reasons for high or low ratios within the 438 site data set were then sought. Three areas were investigated.

- a) Limitations of the current prediction system
- b) Variation in the biological quality of the 438 sites
- c) Variation in sampling effort at the 438 sites

Again, preliminary findings were given in the first progress report but the results have now been examined in more detail and some conclusions are listed in section 3.

Further internal tests using the 438 sites and environmental option 1 have been undertaken in the past three months. They include:-

1. A comparison between the observed and predicted number of taxa at species level (3 seasons combined).
2. A comparison between the observed and predicted number of BMWP families, BMWP score and Average Score per Taxon (ASPT) for spring, summer and autumn separately.
3. A more detailed assessment of the progress of each prediction at both BMWP family level and species level (3 seasons combined). This involves a comparison of the number and type of taxa predicted with those observed at 10% probability intervals (i.e. 100%-90%, 90%-80% etc.) during the progress of each prediction. Goodness of fit between the observed and expected taxon lists was then assessed using the chi-squared test.

Interim results will be reported in Section 3.

So far, these tests have been confined to internal assessments, using the 438 site data set, but it is also important to see how the system performs when predictions are made for new sites. To minimize complications in the interpretation of the results, the sites should be sampled using the standard procedures devised by the IFE, they should be of good biological quality and the data should be amenable to assessment at different taxonomic levels. Sixty-five unpolluted sites on 20 river systems throughout Great Britain will be used for the external test of RIVPACS II. All the physical data required for the predictions are available but alkalinity values for the sites are still required. Most sites are close to, or coincident with, 1990 RQS sites and NRA Thames Region have agreed to provide the relevant data for the list of sites supplied by IFE. Alkalinity values have been acquired for seven of the nine NRA/RPB regions and when the list is complete, the values will be forwarded to IFE and analyses can commence.

2.2 Banding scheme for the 1990 River Quality Survey Results

Development of a banding scheme requires access to the RIVPACS predictions (observed/expected ratios) from the 1990 survey sites. As indicated earlier, we anticipate that the first results will become available, as computer files, from mid April onwards. Examination of results from several geographical regions, encompassing sites from good to poor biological quality, will be necessary in order to set a tentative banding scheme for appraisal by selected NRA/RPB regions. Depending on the feedback received, further changes in the bands may then be necessary before the formal adoption of an agreed scheme for the 1990 River Quality Survey.

Printouts of the first RIVPACS predictions for two NRA regions (Thames and Yorkshire) were made available to IFE in late March and these are receiving urgent attention as a prelude to the banding exercise.

2.3 Additional work

Due to the unavoidable delays in completing the external testing of RIVPACS II and in undertaking the banding exercise, some of the available time has been used to start work which is essentially part of the Phase 2 programme. This includes the following:-

- a) Completion of the cataloguing of the 1990 River Quality Survey samples held at the IFE River Laboratory. This will ensure efficient retrieval of samples chosen for further processing to species level, for inclusion in a future enlarged version of RIVPACS.
- b) Listing, by individual NRA and RPB region, of the sites currently in RIVPACS II, the additional sites examined for the Nature Conservancy Council and sites chosen for the NRA project on the faunal richness of headwaters. These lists, when forwarded to each region, will form a starting point for discussion of the additional rivers which should be added to make a more comprehensive version of RIVPACS.
- c) The 65 sites already sampled for the NCC under a separate contract have had their macroinvertebrate fauna identified to the taxonomic level required by RIVPACS, except for the Oligochaeta and Chironomidae. By undertaking the further identification of these two taxonomic groups the 65 sites, which are of good quality, can be added to an enlarged version of RIVPACS. Mounting of oligochaetes and chironomids from these samples onto microscope slides has started as a prelude to their identification.
- d) At the Advisory Group meeting to the project held in January 1991, the idea of an alternative procedure for generating acceptable ranges around predicted BMWP scores, number of taxa and ASPT values was raised. Essentially, it would make use of the observed range of values exhibited by the sites within the TWINSPAN group(s) to which a new site was predicted, on the assumption that all sites in RIVPACS II are unstressed. This approach is being investigated and an indication of whether it is worth developing further in Phase 2 should be known by the end of April.

3. INTERIM RESULTS

3.1 Comparative performance of the six environmental options

The approach was to determine which set of environmental features gave BMWP level predictions for the 438 RIVPACS sites which best replicated the observed values (i.e. O/E values close to unity). Hence, for each option, the similarity between the observed and expected number of BMWP families, BMWP score and ASPT was examined at each site and the occurrence of extreme ratios (both high and low) was assessed.

The initial set of variables used in RIVPACS (Option 6) was, in general, the best, but this was discounted because it included chloride as a variable. Of the remaining five sets of variables, Option 1 was found to be the best. Eight core variables are used in option 1:

LATITUDE, LONGITUDE, WATER WIDTH, WATER DEPTH, ALTITUDE, DISTANCE FROM SOURCE, SUBSTRATUM COMPOSITION, DISCHARGE CATEGORY plus ALKALINITY, SLOPE, MEAN AIR TEMPERATURE and ANNUAL AIR TEMPERATURE RANGE.

At the Advisory Group meeting which took place on 24 January 1991 it was accepted that option 1 was the most appropriate option for use in the 1990 River Quality Survey. However, it should be noted that option 4 must be used for sites in Northern Ireland since the air temperature data are only available for Great Britain. Option 4 is a variant of option 1 which lacks MEAN AIR TEMPERATURE and ANNUAL AIR TEMPERATURE RANGE.

3.2 Robustness of the prediction system

Preliminary comments made in the previous report on reasons for high and low observed/expected ratios amongst the 438 RIVPACS II sites have been reinforced. It is also apparent that the three mechanisms listed in section 2 as likely contributory factors to extreme ratios cannot be disentangled from each other in order to determine the relative contribution from each source.

Hence, the way ahead should be to recognise the current strengths and limitations of RIVPACS II and seek to improve the present classification/prediction procedures in Phase 2 of the project. Further comments on high and low ratios are given below under the appropriate headings.

a) Limitations of the current prediction system

The system uses information from sites in one, two or more TWINSpan groups to produce a prediction of the fauna to be expected at a site with given environmental features. Since the prediction technique depends upon averaging, the output will be of the fauna to be expected at an average site. Hence, if the site has an unusually rich fauna, then the observed/expected ratio may be well above unity. This can apply both on an internal test, since the taxon richness within any one TWINSpan group varies from site to site, and also when an unclassified site is having the fauna predicted. Hence, some high O/E ratios are an inevitable feature of the present prediction system. However, if the predictions themselves are valid, then high O/E ratios can be used in a positive way to flag taxon rich sites which may be of high conservation value.

Some O/E ratios below unity should also be expected, but the system of acceptable ranges has been developed to offer a limit below which a site is regarded as being stressed.

Ideally, the mean observed/expected ratio for BMWP score, number of taxa and ASPT for each TWINSPAN group should be close to unity. In the case of ASPT all the mean ratios but one are within $\pm 5\%$ of unity, but wider variation is found with BMWP and number of taxa. This was anticipated, since ability to predict ASPT is greater than ability to predict the other indices. Nevertheless a future version of RIVPACS would benefit from a more even distribution of sites between TWINSPAN groups and O/E ratios which were close to unity for each TWINSPAN group.

b) Variation in the biological quality of the 438 sites

Here the concern is the possible inclusion of stressed sites in the data-set used for prediction of target communities. There are no sites with a BMWP score of less than 79 or an ASPT below 4.27. O/E ratios for BMWP score show that just 5 of the 438 sites had a ratio low enough to place them into class 2 (i.e. $O/E < 0.56$) of the provisional banding system. All other sites were provisional class 1A or 1B. Nevertheless, a more stringent assessment of site acceptability may be used in a future version of RIVPACS if evidence suggests that the current system is setting unacceptably low standards in some cases. Inclusion of high quality sites with high O/E ratios is not a problem in itself unless it is an artifact due to oversampling (see below).

c) Variation in sampling effort

Low O/E and high O/E ratios can result from non-standard sampling effort (i.e. under/oversampling). Alternatively, the same result can be obtained if the site in question is stressed (low O/E) or of high biological quality (high O/E). The problem of variation in sampling effort is one which NRA/RPB biologists are conscious of, following the showing of the sampling video prior to the 1990 RQS. However, the possibility of good quality, but undersampled, sites being represented within the 438 site data-set remains, despite clear instructions being issued on the procedures to be adopted at all sampling sites. Some NRA/RPB regions have an excess of low O/E ratios and few high O/E ratios but it is difficult to make a distinction between sites with low ratios because their quality is close to the lower acceptable limit and those where the ratio is due to undersampling.

The results of further internal tests carried out during the reporting period can be summarized as follows:-

d) Observed/expected number of taxa at species level (3 seasons combined)

The range and distribution of O/E ratios for the 438 sites at species level is close to that observed at BMWP family level, with only a small number of additional extreme ratios (both high and low) noted at species level. The correlation coefficient between the O/E ratios at BMWP family and species level is 0.8337.

e) Observed/expected number of BMWP families, BMWP score and ASPT for spring, summer and autumn separately

As anticipated, the range of O/E ratios is wider for individual seasons than for 3 seasons combined at BMWP family level.

- f) Detailed assessment of the progress of predictions at BMWP family level and also at species level

It is theoretically possible for a prediction to generate an O/E ratio close to unity if few of the taxa predicted with high probability are present, but many taxa predicted with low probability are present. To determine whether this ever happens in practice, the progress of predictions for the 438 sites was checked against the observed values (at both BMWP family and species level) at each 10% probability level. The goodness of fit between the observed and expected number and type of taxa was then assessed using the chi-squared test. Relatively few of the 438 sites showed a statistically significant difference, at the 5% level, between the number and type of taxa observed and those expected (22 sites at BMWP family level and 51 sites at species level). In all BMWP family level and almost all species level predictions, the significant differences were due to either high or low O/E ratios, as would be expected. In only a handful of the species level predictions were the number of observed taxa close to the number expected, but the types of taxa observed were not those with a high expected probability of capture.

4. FINANCIAL STATEMENTS

This information, which normally becomes available approximately six weeks after the completion of the period being reported on, will be made available by the IFE Finance Officer in due course.

5. FACTORS LIKELY TO AFFECT THE SATISFACTORY COMPLETION OF THE WORK

As indicated in the previous progress report, the availability of the information and biological results being collated by the NRA as part of the 1990 River Quality Survey is the chief factor which will determine the timescale required for completion of Phase 1 of the project.

We still await the information listed in the previous report. Nevertheless, we are confident that the first RIVPACS results will be available shortly and then the banding exercise can commence.

Dr D. Moss, who has been responsible for many of the computer analyses, is unavailable for a period of approximately one month early in this new financial year. However, Mr R.T. Clarke, who is based at the IFE River Laboratory, will be available for the analyses on banding.

J.F. Wright
19 April 1991

DISTRIBUTION SHEET

To be completed by all Project Leaders completing commissioned research project reports. Please bind a copy of this distribution sheet as the final page in all internal (IFE) copies of the report.

| | | | |
|-----------------------------|---|-----------|---------|
| 1. | AUTHORS: J.F. Wright, M.T. Furse, D. Moss TITLE: Testing and further development of RIVPACS. A progress report for the period 1st January - 31 March 1991 REPORT REF.: RL/T04053r5/2 MASTER COPY HELD BY: D.M. Morton REPORT ACCESS CODE (please assign a suitable code from list below): C | | |
| 2. | DISTRIBUTION LIST (A-G standard distribution; H other) | No.copies | Date |
| A) | Contract Customer: National Rivers Authority | 5 | 26.4.91 |
| B) | J.G. Jones (title page and abstract only) | | 26.4.91 |
| C) | A.D. Pickering | 1 | 26.4.91 |
| D) | A.D. Berrie (Internal Coordinator for Commissioned Research) | 1 | 26.4.91 |
| E) | Project Leader: J.F. Wright | 1 | 26.4.91 |
| F) | FBA Library, Windermere | 1 | 26.4.91 |
| G) | FBA Library, River Laboratory | 1 | 26.4.91 |
| H) | Other (please list below and indicate number of copies in right hand column) | | |
| 1. | M.T. Furse | 1 | 26.4.91 |
| 2. | D. Moss | 1 | 26.4.91 |
| 3. | | | |
| 4. | | | |
| 5. | | | |
| 6. | | | |
| 7. | | | |
| 8. | | | |
| 9. | | | |
| 10. | | | |
| Total number of copies made | | 12 | |

REPORT ACCESS CODES

- S In strict confidence - Access to named customer(s) - (could be named restricted access individuals), IFE Directorate, Project Leader and all authors.
- C In confidence - Access to customer, IFE Directorate, Project Leader, all restricted access authors, and IFE staff with permission of Project Leader.
- N 'Normal' access Access to customer and all IFE staff. Access to visitors and general public with permission of Project Leader.
- G General access General access to anyone as required.