Report

Furse, M.T.; Clarke, R.T.; Winder, J.M.; Symes, K.L.; Blackburn, J.H.; Grieve, N.J.; Gunn, R.J.M.. 1995 *Biological assessment methods: controlling the quality of biological data. Package 1: The variability of data used for assessing the biological condition of rivers.* NERC/Institute of Freshwater Ecology, 139pp. (IFE Report Ref. No: RL/T04071g1/1)

Institute of Freshwater Ecology

River Laboratory

East Stoke

Wareham

Dorset

BH20 6BB

BIOLOGICAL ASSESSMENT METHODS:
CONTROLLING THE QUALITY OF BIOLOGICAL DATA.

PACKAGE 1 THE VARIABILITY OF DATA USED FOR ASSESSING THE BIOLOGICAL CONDITION OF RIVERS

M T Furse
R T Clarke
J M Winder
K L Symes
J H Blackburn
N J Grieve
R J M Gunn

R&D Note 412

CONTENTS

R&D Note 412

LIST OF TABLES

LIST OF FIGURES (at end of document!)

FOREWORD

The National Rivers Authority's business needs require biological data of a consistently high standard for the assessment of water and environmental quality. Of particular significance are the NRA Water Quality Classification (GQA) Scheme and the European Council directive on the Ecological Quality. These needs were recognised in the 16th report of the Royal Commission on Environmental Pollution and in the NRA's response to that report.

The current document therefore forms part of a linked programme of research designed to improve the statistically reliability of the use of aquatic macro-invertebrate assemblage data for assessing the environmental quality of running waters in England and Wales.

The principal objective of the overall study, as set out by the NRA, is:-

• *To quantify and, wherever possible, control sources of variability in macro-invertebrate data for a range of river types and biological quality bands in order to increase the value of the NRA data in water quality management.*

The work programme is divided into two inter-linking packages.

Package 1 has been undertaken by the Institute of Freshwater Ecology and is reported upon here.

Package 2 has been undertaken by the Water Research Centre and forms the contents of a separate report (van Dijk 1994).

Further details of the background of the research programme for Package 1 are given in the Project Investment Appraisal (PIA) which is Schedule 1 of the Memorandum of Agreement for Research Contract between the National Rivers Authority and the Institute of Freshwater Ecology.

ACKNOWLEDGEMENTS

R&D Note 412

EXECUTIVE SUMMARY

Macro-invertebrate sampling is widely used for assessing the biological condition of rivers. Sampling programmes can be used to investigate pollution incidents or for routine monitoring or surveillance at local, regional or at national levels.

Commonly, lists of families captured during a standardised pond-net sample are used to calculate simple indices summarising the condition of a site. The most widely used indices are those comprising the Biological Monitoring Working Party (BMWP) score system. Each family present is assigned a score according to its perceived tolerance to organic pollution. The site index may be expressed either as the sum of the scores of the families present, the number of families present or the average score of those families present.

The indices observed may be compared with those predicted to occur at a site in order to provide Ecological Quality Indices (EQI's) for that site. These indices are a measure of the degree to which that site conforms to expectations and provide standard measures which allow direct spatial and temporal comparisons to be made between samples taken at the same site at different times or different sites at any moment in time.

The effective value range for EQI's is from zero, when no taxa are captured at a site, through unity, when the observed indices exactly meet expectation, to values of up to 1.25 when the sites are exceptionally taxon rich. A very small minority of sites may exceed this value. The value ranges of EQI's are often divided into bands which allow the biological condition of sites to be classified into a small number of categories from bad (no or very few taxa, index close to zero) to good (a diverse fauna of the type expected, index close to unity).

The predicted targets are derived from the River In-Vertebrate Prediction And Classification System (RIVPACS). This system includes a substantial data-base of the biological and environmental information for sites considered to be of sufficiently high biological condition to act as target standards which other sites should comply with. The prediction process relies upon derived mathematical relationships between the composition of the macro-invertebrate assemblages of sites and the environmental characteristics of the same sites.

The calculation of EQI's for a site requires the collection of both macro-invertebrate and environmental data. Each of these types of data, by nature of the collection processes, are acquired with error. For the biological data, variation and error arise from the sampling method, the efficiency with which samples are sorted and the accuracy of the taxonomic identifications. Errors and variation also occur during the collection of the environmental data.

Integrated errors in both the biological data and environmental data contribute to an overall error in the EQI's and mis-classifications when index values are divided into quality bands.

Historically, the errors and variation in all stages of the acquisition of the macro-invertebrate data have been poorly known. This has been a contributory factor to the exclusion of biological monitoring from national river quality surveys and has hampered the interpretation of the data on those occasions when macro-invertebrate sampling has been included.

Lack of knowledge of the precision with which biological indices have been calculated has also restricted the water industry's ability to determine whether spatial and temporal changes in macro-invertebrate assemblages are statistically significant.

The National Rivers Authority therefore commissioned a research programme with the objectives of determining the errors and variation in the collection of the macro-invertebrate data necessary to assess the biological condition of rivers and to devise procedures for controlling the quality of data acquisition during the sample sorting and identification phase.

Recommended analytical quality control procedures have been devised and reported upon separately in NRA R&D Note 331. The current document provides information on the errors and variation in collecting the requisite macro-invertebrate and environmental data.

It also includes procedures for quantifying these errors in terms of the precision of EQI's, methods for assessing the probability of mis-classifying the biological condition of sites using any form of quality banding and a simple statistical test for determining whether there are statistically significant differences between the EQI values of samples collected from different places or at different times.

Analytical data were derived from two sources. Errors and variation in biological sampling and environmental measurements in the field and from maps were based on sixteen specially selected sites in a four by four matrix of site type and biological condition (quality band). Each site was sampled by four different people in each of three different seasons.

Errors in the sample processing and identification stages were determined from 420 samples collected, sorted and identified by the NRA and then audited for sorting efficiency and accuracy of identification by the Institute of Freshwater Ecology.

There was no evidence that variation and errors in sampling varied consistently with site type or season and the additional variation due to inter-operator differences was very small. The absolute sampling variation in observed number of taxa and BMWP is greater at sites with more taxa but the biological condition of a site does not appear to influences the variation due to sampling in any other way.

For operational purposes the square roots of the number of taxa captured and BMWP score of a site should be assumed to have constant sampling variances dependant upon the number of season's data being used to represent the site. Similarly, the sampling variance of the observed ASPT should be best estimated by a series of constants which are also dependant on the number of season's data being used.

The sample processing and identification errors achieved by the NRA when working at about the level of performance being recommended for the 1995 River Quality Survey led to an under-estimation of taxon numbers of about 1.5 per sample unless five or fewer taxa were listed as present when on average only one taxon was missed.

Using the best available data, it was found that about 50% of taxa missed in a single sample are not subsequently found in a second sample from that site in another season of the same year. Furthermore, 37% of taxa missed in a single sample were not found in either of the samples taken the two other sampling seasons.

R&D Note 412

Based on these data, procedures were provided for adjusting for bias in each of number of taxa, BMWP score and ASPT for samples. The choice of whether to apply these correction factors is left to the NRA.

The variation between individuals in their estimation of map and field derived data generally fell within the acceptable limits for each variable set out in NRA Interim R&D Report 243/7/Y. It is recommended that the errors in expected values of BMWP indices due to variation in environmental measurements are expressed by constant standard deviations, as given here, irrespective of the number of samples used to derive the index but dependant on whether the index used is number of taxa, score or ASPT.

The variation in environmental measurements considered in this document relate exclusively to inter-operator variation in a single year. They take no account of temporal variation, particularly as that applies to setting fixed long-term mean values of each predictor variable.

Based on these identified sources of error, variation and bias, it is recommended that statistical simulation procedures are used to estimate the precision of EQI values. Detailed suggestions are included.

The simulation techniques are dependant on the constant error terms and expressions for each stage of the data collection process. They take no account of the errors in prediction generated by the prediction system, RIVPACS, failing to make best use of the environmental variables measured or available for measurement. These cannot be estimated and the RIVPACS predictions are assumed to be part of the definition of quality and hence correct except for the errors in measuring the environmental variables.

The simulation techniques may be used to estimate the likelihood of mis-classifying the biological condition (quality) of a site and for the detection of statistically significant spatial or temporal differences between the macro-invertebrate assemblages of sites, as expressed by their derived EQI's.

Further funding is recommended in order to fully integrate the findings of the current study into RIVPACS, to investigate errors and variation in the estimation of long-term fixed values of environmental variables and to investigate higher than average sampling variation in certain site types.

Keywords: Macro-invertebrate assemblages, sampling, sorting, identification, BMWP score system, RIVPACS, prediction, assessment, ecological quality, variation, errors, precision, mis-classification rates, statistical differences.

R&D Note 412

R&D Note 412

1.   INTRODUCTION

1.1   Background

1.1.1 The need for biological monitoring

NRA business needs require biological data of a consistently high standard for the assessment of water and environmental quality. These needs were recognised in the 16th report of the Royal Commission on Environmental Pollution (1992) which was exclusively devoted to freshwater quality. The commissioners made many recommendations of which the first and most fundamental was:-

"the regulatory authorities should endeavour to develop a general classification scheme based on biological assessment for use throughout the UK in the 1995 and subsequent river quality surveys"

However, to be effective, the techniques applied should be underpinned by sound biological principles and the instruments for quality determination should be amenable to rigorous statistical evaluation in order to determine their reliability in both class allocation and the detection of change. In the Project Investment Appraisal (PIA) which initiated the current research programme, it is recognised that:-

"The variability inherent in biological sampling and analysis must be quantified and controlled (as is done, in part, for water chemistry) allowing confidence limits to be placed around scores and classifications. This is particularly important for the assessment of change. As replicate sampling is only feasible for detailed studies aimed at detecting small scale impact or change, a practical assessment of variability is essential for all other work".

1.1.2 The early history of biological monitoring

The use of biological data for the evaluation of the environmental quality of waterbodies has a long history stretching back to the beginning of the twentieth century (Hellawell 1986). Much of the early development originated in eastern Europe and centred on a series of similar indexing procedures leading to what became known as the Saprobien system (Sladecek 1973). In this system a wide range of plants and animals are allocated numerical values indicative of their tolerance to organic pollution (saprobic valency), abundance and indicator value. The quality of the site is then represented by a simple index which integrates the numerical values of the taxa present (Hellawell 1986, Furse et al 1990).

R&D Note 412

In eastern Europe and Germany attention focused on refining and improving the Saprobien system which is still widely applied in various modified forms (Metcalfe-Smith 1995). However, in most western european countries macroinvertebrates became the most commonly used group for biological monitoring and surveillance.

The first regularly-used biological index which depended exclusively on invertebrates was the British Trent Biotic Index or TBI (Woodiwiss 1964). A proliferation of macro-invertebrate indices followed (Metcalfe-Smith 1995) of which the most notable early examples were those of Graham (1965) and Chandler (1970). The TBI was more or less extensively revised for use in several other European countries (Furse et al 1990).

1.1.3 National River Pollution Surveys

Many of the indices developed and used in Britain in the 1960's and early 1970's were devised to meet local circumstances and there were no consistently-used procedures which could form the basis for a national river quality reporting scheme. In contrast a common national scheme had evolved for the collection and reporting of chemical water quality which was first used in the inaugural national River Pollution Survey of 1958 (Ministry of Housing and Local Government, unpublished).

The first attempt to produce a national biological classification scheme for evaluating the quality of rivers was the scheme devised for the second national River Quality Survey of 1970 (Department of the Environment & The Welsh Office 1971). It can fairly be described as simplistic. Just four quality classes were recognised and the allocation of sites and, by extrapolation, river reaches to one of the classes largely depended on the relative frequency of three orders of insects; Ephemeroptera, Plecoptera and Trichoptera and the freshwater shrimp ("Amphipoda"). The nature of the fishery was also taken into consideration but no quantitative rules or guidelines were provided.

Allocation to quality class was inevitably a subjective process depending on how the practitioner distinguished between phrases such as *"an appreciable proportion of Plecoptera and/or Ephemeroptera, Trichoptera and Amphipoda"*, which partly defined Class A, and *"Plecoptera and Ephemeroptera populations may be restricted. Trichoptera and Amphipoda usually present in reasonable numbers"*, which partly defined Class B.

Biological classifications were excluded from the 1972 update survey (Department of the Environment and The Welsh Office 1972) but incorporated again in the 1973 survey (Department of the Environment & The Welsh Office 1975). The 1973 scheme was modified in order to downweight the contribution made by the status of the fishery. However the residual short-comings of the scheme were such that the use of biological data was once again omitted from the major 1975 survey (Department of the Environment & The Welsh Office 1978) where it was stated that:-

"the derivation of an entirely satisfactory biological classification for use in a nationwide survey is difficult at present. This is because so little is known about

the interdependence of animal and plant requirements in relation to flow, the nature of the river bed and water quality...when a more satisfactory biological classification of water quality has been developed it may be reinstated in future surveys. There is no doubt that in situ assessments of water quality in biological terms are extremely valuable"

R&D Note 412

1.1.4The Biological Monitoring Working Party

In order to devise that "more satisfactory" biological classification a working party, the Biological Monitoring Working Party (BMWP) was commissioned in 1976. Its terms of reference were to:-

•recommend a biological classification of river water quality for use in the River Pollution Survey

•consider ways and means of implementing that classification

•consider relationships, if any, between chemical and biological classifications

Significantly, in making its final report, the working party felt unable to recommend a system of biological classification of *"river quality"* (Biological Monitoring Working Party 1978). Instead it recommended a system for assessing the *"biological condition"* of a river. The system they devised was the eponymous BMWP score, which underwent a series of revisions (Chesters 1980, National Water Council 1981) before being used in the assessment of data collected during the renamed River Quality Survey of 1980.

In the BMWP score system families of animals were accorded individual scores according to the perceived tolerance to organic pollution of their most sensitive component species. Pollution intolerant taxa were given high scores (maximum 10) and pollution tolerant taxa low scores (minimum 1). The sample score, representing the biological condition of the site, was the sum of the individual scores of the families captured. Two other functions of the sample used to assess the site condition were the number of scoring taxa (not individuals) present and the average BMWP score of those taxa (Average Score Per Taxon or ASPT).

In general terms the higher the total sample score, numbers of scoring taxa or ASPT the better the biological condition of the site was taken to be.

The working party also made another important observation (Biological Monitoring Working Party 1978). This was that chemical and biological data to provide different but complementary measures of the condition of a site. They felt that the biological assessment was of greatest value when it failed to match that interpreted from chemical analyses and state that:-

"it does not serve any purpose to attempt to correlate the results of the chemical and biological assessments. If correlations were established there would be no justification to carry out both forms of assessment"

1.1.5The development of RIVPACS

The use of the BMWP system in the 1980 survey provided an empirical scoring system for recording the biological condition of sites from the macro-invertebrate samples collected from them. These results were presented as a series of maps in which the sample scores were recorded alongside the sites from which they were collected.

However, although the new scoring system was more empirical and apparently less subjective than its predecessor, the scores still needed to be interpreted in quality terms and this exercise was conspicuously not undertaken. The intrinsic worth of a score of, say, 150 was left for the map-reader to determine as was the significance of the difference between, say, scores of 130 and 160. Furthermore, in the report of the survey (National Water Council 1981), a similar anxiety to that expressed for the 1975 survey re-surfaced. The National Water Council (1981) wrote that:-

"Interpretation of biological scores is a matter for professional experts as the diversity of families present at a site depends not only on the degree of any pollution, but also on the nature of that pollution and, more particularly, on what would be present without any pollution. There are substantial natural differences between upland and lowland streams. The present biological assessment is primarily to provide a basis for future comparison".

By this time the concept of differing intrinsic biological potential of sites with contrasting character had gained sufficient credence for a research project to be initiated at the, then Freshwater Biological Association. In summary, the aim of the project was to quantify the links between the environmental characteristics of sites and the and the macro-invertebrate assemblages that will occur at them when unstressed by physical or chemical perturbations.

This research programme is still ongoing and has been responsible for the software package RIVPACS (Wright et al 1993) which is in widespread use throughout the United Kingdom and is serving as a model for the development of similar procedures in places as diverse as Spain, Canada and Australia.

The current version, RIVPACS II contains information on the macro-invertebrate assemblages and environmental characteristics of 438 sites throughout Great Britain. A revised version with 684 sites throughout the United Kingdom is nearing completion (Wright et al in preparation).

Essentially RIVPACS is a system of prediction by analogy. Through the use of multivariate statistical procedures the system provides a prediction of the fauna which should be captured at a site, using standard sampling methods, if that site was not significantly stressed (Wright et al 1993). On this basis, each site could be provided with a specific biological target against which its observed (ie sampled) fauna could be judged. The degree of compliance between the expected (ie RIVPACS-predicted) fauna and that observed has been quantified in the form of the Ecological Quality Index or EQI (Sweeting et al 1992).

The EQI of a site is the ratio of its observed BMWP index value and that predicted by RIVPACS. It can take three forms depending upon whether the function used is the BMWP score, number of scoring taxa or ASPT. In each

case the EQI of a site is unity if the observed index values fully meet expectations but zero if no taxa are present. Most sites lie within this value range but EQI's of some sites have a sufficiently diverse and high-scoring fauna that their EQI values can exceed one.

Unlike previous indices, the EQI now provides the opportunity to make direct and meaningful comparisons between the fauna of sites of entirely different character or geographic location. This is because the EQI is a measure of the extent to which each individual site meets its biological potential and this is a common factor by which all sites may be judged.

The EQI's can also be used as a numerical basis for site banding. The principles and suggested practice of banding were outlined to the NRA in a series of IFE reports (Wright et al 1991, Clarke et al 1992, Clarke et al 1994). In all cases the highest quality band width was set at the level attained or exceeded by a set percentage of the sites in RIVPACS II, which in turn were perceived to be sites with the best achievable biological condition for their environmental type. Different band widths were suggested for different functions of the BMWP system and different seasonal combinations of samples. Initially a 95%ile attainment rate was suggested to set the lower limit of the highest ASPT band whereas a 90%ile was suggested for number of scoring taxa and ASPT. These suggestions are subject to reconsideration at present but it is assumed that the ultimate responsibility for setting the number and range of biological quality bands rests with the NRA.

Unfortunately, RIVPACS was not in operational use at the time of the 1985 River Quality Survey and once again biological sampling was excluded from the evaluation process (Department of the Environment & The Welsh Office 1986). However, by 1990 a tested and fully operational version of RIVPACS was available and provided a spur for the re-inclusion of biology in the 1990 River Survey. A total of 23,083 biological samples from 8796 sites, throughout the United Kingdom, were collected for survey purposes (Sweeting et al 1992).

1.1.6 The chemical classification

Over the duration of the surveys, from 1958 to 1990 the chemical evaluations were largely based on the three determinands; dissolved oxygen, biological oxygen demand (BOD) and ammonia together with additional information on factors such as toxicity to fish, the presence of substances other than those used in the classification and even the biota actually present. Over this time the methodology of collection and interpretation of the data had gradually been modified and improved.

However the most significant improvements to the scheme were those recommended in 1990 when the NRA took over the direct task of organising the survey in place of the Department of The Environment and The Welsh Office. Whilst, for reasons of continuity and comparison, the main report retained the chemical classification scheme used in 1985 (National Rivers Authority 1991a), the NRA also recommended a revised version for use as part of a Statutory Water Quality Objectives scheme (National Rivers Authority 1991b).

The NRA's recommendations differed from the previous system in several respects. Firstly the classification was entirely restricted to dissolved oxygen, BOD and ammonia. Secondly standard non-parametric or parametric statistical procedures for interpreting the data were proscribed, depending on

the number of samples available, thirdly a three year run of results was set as a sampling requirement, fourthly the exclusion of "outlier" samples was forbidden and finally standard procedures for dealing with "less than" values were stipulated.

One result of these changes was to virtually eliminate inter-regional differences in the methods of data interpretation and classification. A second consequence of the more rigorous procedures was that the theoretical error rates in assigning sites to class and in reporting a change in class from sampling period to sampling period could be estimated (National Rivers Authority 1991a, 1991b).

1.1.7 The need to improve the collection and interpretation of biological data

The introduction of RIVPACS and the use of EQI's provided a much more rigorous method of banding sites and for making comparison between sites on different types of river. Nevertheless biological banding still lagged behind chemical in terms of the ability to estimate error rates in class allocation and the detection of temporal change. For a variety of undocumented reasons, probably including the inability to apply error rates to the classification process and doubts over the effectiveness of RIVPACS in some types of watercourse, the publication of the results and interpretation of biological data from the 1990 survey was relegated to the status of an appendix of the report (National Rivers Authority 1994).

A legacy of intermittent inclusion and exclusion of biological monitoring in the River Pollution and River Quality Surveys is a persistent perception that biology supplements rather than complements chemical monitoring. The arrival of RIVPACS with its more empirical and nationally consistent procedures certainly helped to raise the status of biology within the NRA and the 1990 survey included a more intensive and extensive biological sampling programme than any of its predecessors. However, the potential of biological monitoring has yet to be fully realised. The report of the 1990 survey (National Rivers Authority 1991a) clearly makes this point with the statement that:-

"the traditional means of assessing river water quality ... could be substantially improved by drawing upon information on the biological state of the river"

The objectives of the quinquennial River Pollution/Quality Surveys, although rarely stated must be to provide an overview and summary of the condition of British watercourses and to provide an indication of the temporal trends in change of quality in order to best formulate river management strategies.

This requires more consistent and reliable methods of data capture and interpretation than used in the past. The more structured and statistically rigorous use of chemical data provide a move in the right direction. However there still remains a 20-30% chance that an individual stretch of water may be declared to have changed chemical class when the actual chemical quality may not have changed at all (National Rivers Authority 1991a).

Preliminary indications are that the error rate associated with biological detection of change may be better than that achieved by the current level of chemical monitoring (National Rivers Authority 1994), although the situation is complicated by the different number of biological and chemical classes.

If biology is to fully recognise the intrinsic potential that both the Royal Commission and the NRA recognised it to possess then it is imperative that a

broad understanding is acquired of the errors and variation associated with all stages of data collection, processing and interpretation of macro-invertebrate assemblage data. Only then can the reliability of biological classifications and the significance of temporal change be assessed and biology complement the chemistry to the extent of that potential.

R&D Note 412

In the view of the Royal Commission (1992) the RIVPACS approach provides a sensible approach to biological monitoring and they advocate further research into similar systems. This report presents one element of that research and the significant findings of this study must be built into the further developments of the RIVPACS software.

1.1.8 Sources of biological variation

If better use is to be made of biological data then it is imperative that a fuller understanding is gained of the stages in the data collection, recording and analysis process when variation could occur and its likely magnitude.

In this report errors and variation in four separate processes are considered.

• variation in the collection of biological samples (Chapter 2)

• errors and omissions in the sorting of biological samples (Chapter 3)

• errors in the identification of macro-invertebrate specimens (Chapter 3)

Each of these sources effects the observed BMWP indices calculated from the samples and used to calculate their Ecological Quality Indices as part of the RIVPACS process. The first and third sources of error and/or variation may represent genuine variation around the mean. The second source may represent a bias in some functions of the BMWP score system. Thus failure to remove all the different BMWP families present in a sample can only lead to an underestimate of the EQI's for BMWP score and number of taxa. However this source of error may produce unbiased variation in the derived ASPT's, depending on the individual scores of the families not removed from the sample.

• errors and variation in the collection and recording of environmental variables
        (Chapter 4)

This final source leads to variation in the RIVPACS-predicted, expected BMWP index values for the site and also contributes to error and/or variation in the derivation of the site EQI as part of the RIVPACS process.

The effects of the various sources of error are integrated in Chapter 5 and methods for detecting significant spatial and temporal differences in biological assemblages are proposed. The relevance of the findings to the use of biological monitoring for river quality estimation is discussed in chapter 6.

A fifth source of variation in the derivation of EQI's is that intrinsic to the mathematical procedures employed to derive the RIVPACS model. The

significance of this is also considered in Chapter 6.

1.2 **Objectives**

The overall project objective, as set-out in the PIA, is:-

•To quantify and, where possible, control sources of variability in freshwater macro-invertebrate data for a range of river types and biological quality bands in order to increase the value of NRA data in water quality management

There are three specific objectives:-

•To assess the variability of single and combined season observed data (number of taxa, BMWP score and ASPT) due to the sampling process and analytical error.

•To assess the effect on RIVPACS predictions of errors in recording environmental variables by replicated field measurement.

•To assess the overall variability of observed and RIVPACS-predicted data due to the combined effects of the above factors.

The work programme required to meet these objectives is detailed in the project PIA and results from a series of pre-contract discussions between NRA and IFE staff. The associated Package 2 study (van Dijk 1994) provides recommended analytical quality control procedure in order to meet the standard of sample processing performance required by the NRA in order to set acceptable variation in EQI values based of the findings of the current study.

2 VARIATION IN BIOLOGICAL DATA RESULTING FROM SAMPLING

This element of the research programme concerns part of the first specific objective which is *to assess the variability of single and combined season observed data (number of taxa, BMWP score and ASPT) due to the sampling process.*

2.1 Methods

2.1.1 Experimental Design

An essential requirement of the experimental design was that the sampling programme and procedures had the potential to replicate any of the options likely to be selected for the 1995 River Quality Survey or those used in previous surveys. The relative error rates and variation associated with single and multiple-season sampling could then be compared.

The 1990 River Quality Survey involved sampling in each of the three RIVPACS "seasons", spring (February - May), summer (June - August) and autumn (September - January). No greater frequency of annual sampling seemed probable for future surveys and the three-season sampling strategy was therefore adopted for the current study.

In addition, the sampling programme needed to allow both between and within operator variation to be evaluated in both single and multiple seasons. This required that at least two people sample at each site and at least one of those take more than one sample.

The extent of the sampling programme was also regulated by the length of the research contract and its financial value.

In the context of these constraints the following experimental design was implemented with the prior agreement of the NRA Project Leader (Dr R.A. Dines) and Topic Leader (Dr R.A. Sweeting):-

• four macro-invertebrate samples were collected in each of the three seasons from each of sixteen sites.

• standard RIVPACS sampling methodology was used (National Rivers Authority in preparation).

• three of the four samples from each site were sorted and identified.

• the fourth samples were stored in case the results of statistical analyses

indicated that it would be beneficial for them to be processed and included in the analytical data-base.

•at each site in each season two samples were collected by operator A and one each by operators B and C.

•all efforts were made to standardise the identity of the operators A, B and C at each site but different operators were allowed to fulfil these three roles at each separate site.

•for purposes of continuity of experience and efficiency the same two IFE staff members were part of the sampling team at each site and alternately took the roles of operators A and C at the different sites

•an NRA staff member, whose identity varied from site to site, completed the sampling team and took the role of operator B.

The programme also needed to be suitable for comparing sampling variation in rivers of different biological condition and environmental type. The availability of sixteen sampling sites allowed a matrix of four different biological conditions to be compared for four different site types.

The chosen states of biological condition (Table 2.1) were the four quality bands associated with the "5M" system of biological grading used for reporting on the 1994 River Quality Survey (National Rivers Authority 1994). These were A (best condition) - D (worst condition),

Table 2.1   The matrix used to select sites for examining biological variation due to sampling.

BIOLOGICAL CONDITION	TWINSPAN GROUP (sensu Cox et al 1991)			
A ("good")	3a	5b	8a	9b
B ("fair")	3a	5b	8a	9b
C ("poor")	3a	5b	8a	9b
D ("bad")	3a	5b	8a	9b

The four river types needed to be as diverse as possible to ensure that the findings of the study were widely applicable throughout the NRA regions. The RIVPACS classification of sites used in conjunction with the 1990 survey (RIVPACS II) provided a suitable framework for meeting this criterion. The RIVPACS sites were classified using TWINSPAN (Hill 1979) which is a dichotomous, divisive technique in which the full set of sites are first split into two daughter groups based on the relative similarities and dissimilarities of their fauna. In the next stage of the classification each daughter group is itself divided into two sub-groups.

The four groups formed in this way represent the four predominant types of biological assemblage found at the RIVPACS sites and, because of the proven

R&D Note 412

links between assemblage composition and the physical and chemical character of sites which underpin RIVPACS, these four groups may be taken to represent the four major site types in Great Britain. These four groups were used as primary basis for the four environmental classes in the site selection matrix (Table 2.1).

R&D Note 412

However, because each of the four major biological groupings still represented a diverse range of site type, the environmental classes were refined to be the largest (most populous) sub-group of each of the major groups once the TWINSPAN division was allowed to proceed to the 25 group level at which RIVPACS II operates (Cox et al 1991).

Within RIVPACS II, Group 3a (Table 2.1) comprises a group of 24 relatively small sites at an average of 15.3km from source with average widths and depths of 7.5m and 19.8cm (Cox et al 1991). They are mainly at moderate heights (mean 74.5m) and also have moderate alkalinity values (80.8 mg l^{-1} $CaCO_3$). Their dominant substratum type is predominantly cobbles and pebbles and most are situated in the South-West and North-East of England and in Wales.

Group 5b contains 36, mainly calcareous sites with a mean alkalinity of 153.1 mg l^{-1} $CaCO_3$. On average they are smaller than sites in Group 3a; distance from source 8.2 km, width 4.8m and depth 21.7cm. They tend to be at lower altitudes than 3a (39.8m) and are principally situated in central southern England and the midlands. The predominant substratum is gravel.

Group 8a sites are mainly to be found in a central belt stretching from east Wales through the midlands to East Anglia but southern chalkstreams such as the Lee, Avon and Ed are also represented. The 26 sites have the highest mean alkalinity (228.6 mg l^{-1} $CaCO_3$) of the four types considered. Sites are similarly close to source (11.3km) and other mean values are; altitude 40.0m, width 4.8m and depth 32.5m. Their predominant substratum is gravel/sand.

The fourth group, 9b, comprises ten deep (77.5cm), low lying sites in South-East England and East Anglia with a mean altitude of 5.4m. They are, on average 33.0km from source, 13.1m wide and rather alkaline (170.5 mg l^{-1} $CaCO_3$). The sites tend to be on slow flowing, depositing reaches where the predominant substratum type is silt.

2.1.2 Site Selection

Sites were selected from those sampled biologically during the 1990 river quality survey. Selection was confined to sites within the NRA regions of England and Wales which had been sampled in each of the three RIVPACS seasons (Sweeting et al 1992) and which were held in usable form on the IFE computer data-base.

RIVPACS II was used to predict the probabilities of each of these 5006 sites (Clarke et al 1992) belonging to each of the 25 groups in the biological classification. Predictions were based on the "option 1" variable combination in RIVPACS II (Cox et al 1991). The environmental values used were those

compiled by the NRA for the 1990 survey. Sites were then assigned to the groups to which they had the highest predicted probability of membership. Only sites allocated to the four classification groups 3a, 5b, 8a and 9b were retained for further consideration.

The RIVPACS predictions for sites in these groups were used in order to calculate the individual Ecological Quality Index (EQI) values (Sweeting et al 1992) for each of BMWP score, number of scoring taxa and ASPT. EQI's were calculated from three seasons' combined observed and expected BMWP index values.

R&D Note 412

A series of successive, criteria were then applied to reduce the full list of sites in each group to a manageable set of 16 short-lists, representing each of the cells of the selection matrix (Table 2.1).

Sites excluded from each list were those:-

•sites not having the required quality banding for all three EQI's, BMWP score, number of scoring taxa and ASPT

•sites not lying fully within the scope of RIVPACS II (i.e. all sites other than those in suitability class 1 - (Cox et al 1991))

•sites with a probability of belonging to their allocated TWINSPAN groups of $p \leq 0.5$.

These procedures failed to reduce all short-lists to manageable numbers. Therefore an additional, specific criterion was then adopted for retention of sites in each cell of Table 2.1. EQI values were now required to fall within the centre of their range for the biological quality band under consideration. In some cases higher minimum acceptable probabilities of TWINSPAN group membership were also required. The net effect of these additional criteria was to reduce the short-list to those sites whose environmental character made them especially typical of the river type and quality band.

The additional criteria for retention on the short lists of each site type for each biological quality band were:-

Band A sites

•EQI values must fall within the following ranges, centred on unity:-

BMWP score0.91 to 1.09
No. taxa0.94 to 1.06
ASPT0.97 to 1.03

•The minimum acceptable probability of the relevant group membership must be $p \geq 0.6$

Band B sites

•EQI values must fall within the following ranges:-

BMWP score0.52 to 0.62
No. taxa0.64 to 0.72
ASPT0.80 to 0.85

R&D Note 412

Band C sites

•EQI values must fall within the following ranges:-

BMWP score0.29 to 0.39
No. taxa0.41 to 0.53
ASPT0.68 to 0.74

Band D sites

•EQI values must fall below the following values:-

BMWP score<0.18
No. taxa<0.30
ASPT<0.60

From the resultant short-list the five sites with the highest probability of appropriate TWINSPAN group membership in each cell were retained for the ultimate selection process.

Each was marked on a map of England and Wales and the final selections (Table 2.2) were based on the dual requirements of a single site from each of the sixteen cells and the need to maximise efficiency by limiting between-site travel on each sampling day.

Of the original sixteen sites selected, one at Storforton Land, Chesterfield (TWINSPAN group 8a, biological quality band D) proved not possible to sample on health and safety grounds and was replaced by the site with the next highest probability of correct group membership (Table 2.2)

Table 2.2   The full listing, by TWINSPAN group and biological quality band, of the sixteen sites chosen for replicate sampling

TWINSPAN GROUP	QUALITY BAND	RIVER NAME	SITE NAME	NGR	NRA REGION
3a	A	River Okement	South Dornaford	SS 600 000	South Western (SW)
3a	B	River Darracott	Tanton's Plain	SS 494 198	South Western (SW)
3a	C	River Croxdale	Croxdale House	NZ 272 379	Northumbria & Yorkshire (N)
3a	D	Twyzell Burn	B6313 Bridge	NZ 257 517	Northumbria & Yorkshire (N)
5b	A	Petworth Brook	Haslingbourne Bridge	SU 982 204	Southern
5b	B	Sheppey River	Woodford	ST 537 441	South Western (Wx)
5b	C	Sheppey River	Bowlish	ST 613 440	South Western (Wx)
5b	D	Moss Brook	PTC Bedford Brook	SJ 676 983	North West
8a	A	Summerham Brook	Seend Bridge	ST 945 595	South Western
8a	B	Cuttle Brook	Swarkestone	SK 375 288	Severn Trent
8a	C	Poulshot Stream	Jenny Mill	ST 979 592	South Western (Wx)
8a	D	Spen Beck	Dewsbury	SE 225 208	Northumbria & Yorkshire (Y)
9b	A	Old River Ancholme	Brigg	TA 001 065	Anglian
9b	B	Broad Rife	Ferry Sluice	SZ 854 963	Southern
9b	C	Skellingthorpe Main Drain	U/S Skellingthorpe	SK 937 727	Anglian
9b	D	Keyingham Drain	Cherry Cob	TA 219 224	Northumbria & Yorkshire

R&D Note 412

Key to abbreviations to former NRA regions (as at the time of the 1990 River Quality Survey)

(N)=Northumbria(SW)=South West
(Wx)=Wessex(Y)=Yorkshire

2.1.3 Sampling Procedure

All samples were collected using the standard pond-netting techniques recommended for use with the 1995 River Quality Survey (National Rivers Authority in preparation). This involved three minutes of active sampling with the objective of capturing the fullest range of taxa present at the site.

Where feasible, i.e. where suitable large particles and/or sticks occurred, samples were supplemented by a minute's continuous searching in order to find and remove individual specimens of families which may not have been captured by pond-netting. When the first operator, A, undertook a search as part of the sampling process, then all subsequent samples from that site in that season also incorporated a similar search.

Wherever possible the full width of the watercourse was sampled. The length of the sampling reach was always within the range 10-30m but varied with the size and accessibility of the watercourse.

The precise locations of the sites were determined by the need for the character of river sampled to be as consistent and spatially compact as possible in order to allow valid, within-site comparisons to be made between both samples and operators.

At all sites the sequence of sample collection was entirely consistent. The first and third samples were taken by operator A and the second by operator B. Operator C collected the fourth, reserve sample.

The sampling sequence was determined by the possibility that faunal depletion could occur with some of the collection strategies adopted. Primacy was given to the need for reliable comparisons between the first sample collected by A and the second, single sample collected by B. The most vulnerable sample to faunal depletion was the fourth, reserve sample collected by C.

In order to meet the conflicting aims of minimising the effects of faunal depletion and maximising the homogeneity of the four sampling paths, a set of sampling strategies were adopted which varied according to the physical characteristics of the site.

• In wide streams (generally >3m), where the full width of the river could be safely accessed, the three minutes of active sampling was carried out diagonally upstream from one bank to another. Samples one and two started on opposite banks at the lowermost limit of the sampling area and progressed diagonally upstream to the alternate bank. Sample three started slightly upstream (ca 2m) of sample one and followed an upstream diagonal parallel to the path of sample one. The path of sample four was parallel to and slightly upstream of that of sample two.

•In narrow streams (generally ≤3m), where the full width of stream could be safely accessed, the samples were taken successively upstream along the full width of the river. The Twyzell Burn and Petworth Brook sites, each with a width of 3.5m were also sampled in this way because the alternative form of diagonal sampling (above) would not have allowed all samples to be collected over the same range of habitat types.

•Where one margin of the site was easily and safely accessible, but the centre of the stream was not, then the samples were taken successively upstream along that river margin.

Whichever sampling strategy was used, each sample involved collecting from the accessible individual habitats in proportion to their occurrence. However, where one habitat type existed only as a small patch, which would have been totally disturbed before all four samples could have been taken, then this patch was excluded from all samples.

Sampling was undertaken in each of three seasons, "spring", "summer" and "autumn". At any given site the identity of operator A was constant throughout the three seasons. However the identity of A varied between sites, with two separate IFE staff members fulfilling this role during the sampling programme as a whole.

A total of twelve different NRA staff members fulfilled the role of operator B. In all cases except the Twyzell Burn and Croxdale River sites, the same person was operator B at any given site. At the two exceptional sites a substitute operator took the role of operator B in the summer only.

All samples were fixed at bankside, immediately after collection, using 4% formaldehyde solution. Fixed samples were held in labelled, heavy duty plastic bags placed within labelled 1.25l storage jars. Samples were taken to the IFE River Laboratory for sorting and identification.

2.1.4 Sorting and identification

Preparatory to sorting, samples were washed in a brass sieve of 500µm mesh size. This had the effect of removing the fixative and cleansing the sample of silt, clay and fine sand. Some very small animals, such as first instar chironomid larvae may have been lost during the process. However, procedures were consistent between samples and also with those adopted by the NRA.

Samples were sorted and animals removed from flat-bottomed white trays of approximate dimensions 24 x 32cm. Samples were sorted in a series of small aliquots in order to maximise the visibility of macro-invertebrate specimens amongst the other organic and inorganic material present.

The bottom of the sampling trays were sub-divided, by inked lines, into a series of sixteen cells of identical area. These cells were used to sort a sub-section of the full sample. The normal fraction sorted was a quarter of the full sample. All specimens were removed from the cell selected for sub-sampling up to a maximum of 50 from what appeared to be the same family. Further specimens of that family were recorded on a tally counter.

All specimens of families not observed and removed from the sub-sample cell but present in the remaining portion of the full sample were removed and retained for inclusion on the faunal list for that sample. Where the family was subsequently found in the sub-sample fraction these additional specimens were discarded and their numbers not included in the subsequent counts.

All specimens removed from the sample were preserved in industrial methylated spirit (IMS) with 10% by volume of glycerol added to prevent desiccation, and stored in a single labelled vial. The rest of the sample material and remaining specimens were tipped into a collecting sieve of 500 mesh after each aliquot had been fully searched and the appropriate specimens removed. This material was re-preserved and stored in labelled polythene bags and 1.25l storage jars in case quality audits needed to be undertaken on them.

Most picked specimens were identified to family level, using the most recent taxonomic keys (National Rivers Authority in preparation). The single exception was the Oligochaeta which were recorded as such because that is the level of identification required in the BMWP system (Chesters 1980). Specimens were retained in the labelled vials after identification. All microscope preparations used during the identification process were also labelled and retained. Mounts were semi-permanent and the mountant used was polyvinyl lactophenol.

For families found in the sub-sample fraction, total sample abundances were estimated by dividing the number of specimens in that fraction by the proportion of the total sample that fraction represented. For taxa found only in the non-sub-sample fraction, the number of specimens removed was taken to be the number present in the sample.

2.1.5 Quality Control

Sample collection

IFE requested that the NRA staff members attending each site should preferably include at least one person experienced in biological sampling and environmental data-collection for RIVPACS. Where it was necessary to use inexperienced staff members IFE requested that they should view the NRA/IFE training video (Furse & Gunn 1990) prior to participating in the study.

Prior to the spring sampling each participating NRA region/area was also sent an abstract from the initial RIVPACS manual (Furse et al 1990) detailing the biological sampling procedure.

At each site, in spring, the leader of the IFE sampling group re-iterated the common sampling procedure to be used by all personnel taking a macro-invertebrate sample. Where necessary these instructions were repeated prior to the summer and autumn collections. The sampling methodology used was identical to that laid down for the 1995 River Quality Survey (National Rivers Authority in preparation).

Sample sorting

R&D Note 412

All samples were sorted by three proficient IFE staff members each with wide experience of quality auditing the proficiency of NRA's sample sorting. Sorters were made fully aware of the need for extreme care and accuracy in the sorting process.

A constant, very low rate of error has been assumed for IFE's sorting but the financial constraints of the project budget have prevented this efficiency being audited internally or externally. All samples have been reconstituted and are available for audit if funding allows.

Sample identification

All specimens were identified by two experienced IFE staff members who have both attained the pass mark for family level identification on the British Museum IdQ course. The senior identifier, who checked the identification of all difficult specimens has also passed the IdQ examination in species level identification.

Once again, the financial constraints of the project budget have prevented the accuracy of IFE's identification being audited internally or externally. However all samples and microscope preparations have been retained and are available for audit if funding allows.

2.1.6 Data analysis

Data storage

All biological data have been stored, as standard format ASCII files, on a Microvax II mainframe computer situated at the IFE River Laboratory. Data have been transferred to MINITAB Version 10 (MINITAB 1994) when required for analytical purposes.

Database structure and purpose

The aim of the analyses was to estimate and summarise the variation in number of taxa (TAXA), BMWP score (SCORE) and Average Score per Taxon (ASPT) that occurs through sampling. This needed to be done for single season samples, for two seasons combined samples and for three seasons combined samples.

The single season sample variation was estimated for each of the 16 sites in each of the three seasons using the three replicate samples (2 from one IFE person, 1 from an NRA person).

For each pair of seasons (spring/summer, spring/autumn and summer/autumn), the three replicate samples in each season for a site were used to make nine examples of the possible two season combined samples for the site. Five of these nine combined season samples each involved samples taken by only one person (Table 2.3).

The three replicates samples from each season for a site were combined in all possible combinations to form 27 examples of three season combined samples from the site. Nine of these 27 combined season samples involved samples taken by only one person (8 from IFE, one from the NRA person) .

Table 2.3 Number of actual and derived samples available for analysis for

each site in each season or seasons combination.

Actual and derived samples per site	Single season samples	2 seasons combined samples	3 seasons combined samples
Total	9	27	27
Samples involving only one person	9 (3 per season)	15 (5 per pair of seasons)	9
Samples involving different people	none	12 (4 per pair of seasons)	18

Calculation of variance and mean values

The variation in the values of TAXA , SCORE and ASPT for all single and combined season samples for each site was represented by the variance (denoted by VAR) of the values.

If m_x is the mean, or average, of the n values x_1 , x_2 ,..., x_n , then:-

$$variance \ = \ VAR \ = \ \sum_{i=1}^{n} (x_i - m_x)^2 \ / \ (n-1)$$

The standard deviation (hereafter referred to as SD) is the square root of the variance.

Though multiple seasons combinations of samples for a any particular site will not be completely independent of each other, their mean and SD will be unbiased estimates of the true mean and true standard deviation for that site.

Relationship between the variance and the mean

The variance of a set of values is often larger when the mean, m_x, of the values is larger. The relationship between VAR and mean m_x for each of TAXA, SCORE and ASPT was assessed using Taylor's power law (Taylor, 1961). This assumes that the variance of the replicate values is proportional to a power b of the mean m_x of the values, namely:

$$VAR \ = \ K. (mean)^b \tag{eq 1}$$

This relationship is estimated by fitting a linear regression between the logarithm of SD^2 and the logarithm of m_x , as :

$$Log \ VAR \ = \ a + b \ Log \ (mean) , \ where \ a \ = \ Log \ (K). \tag{eq 2}$$

In each regression, between 1-4 of the observed VAR values were zero. To overcome the problem of taking logs of zeroes, the log-log regressions were done in two ways, first excluding these observations, then treating the zero VAR values as equal to the minimum observed non-zero value. The true

R&D Note 412

relationships between VAR and mean should be revealed by both approaches.

If the slope is b, a transformation of the values X to $Y=X^{1-b/2}$ makes the variance in Y independent of its mean. A slope of b = zero indicates that the variance of replicate values does not increase with their mean value.

A slope of b = 1 suggests that the variance is proportional to the mean and hence that the SD is proportional to the square root of the mean. In such cases, the variance of the square root of individual replicate values will be independent of the mean values. Therefore, in the absence of other factors influencing the replicate variance, the variation of each site is best estimated using a single variance estimate.

The mean, V_{sqr}, of the variances of the square root of replicate values for each of the sites and season combinations provides the best overall estimate of this assumed constant variance of the square root of values. V_{sqr} is estimated separately for samples based on single seasons, two seasons and three seasons combined samples.

For any observed value X, the variance estimate, V_{sqr}, based on the same number of seasons samples, can be used to derive approximate confidence limits. If Z_α is the α two-sided percentage point of a standard normal distribution (eg Z_{95} = 1.96), then an α% confidence interval for X is:

$$(\sqrt{X} - Z_\alpha SD_{sqr})^2 \quad \text{to} \quad (\sqrt{X} + Z_\alpha SD_{sqr})^2 \qquad \text{(eq 3)}$$

where $SD_{sqr} = \sqrt{V_{sqr}}$ = the best estimate of the replicate standard deviation.

A slope of b = 2 suggests that the variance varies as a constant proportion, K, of the square of the mean m_x, which is equivalent to saying that SD varies as a constant proportion, $\sqrt{K}$, of the mean. In this situation, the coefficient of variation (hereafter referred to as CV) defined as the SD divided by the mean (CV = SD / mean) will be equal to the constant $\sqrt{K}$.

Green (1979, p46-47) points out that for biological field data the slope b is often in the range 1 to 2, and strongly suggests that, for simplicity, it is assumed that SD is either independent of the mean (b=0) or that SD is proportional to the square root of the mean (b=1) or that SD is proportional to the mean (b=2). As the NRA need a simple summary of any relationship between SD of the observed values and the observed values themselves, one of these three approaches will be used here for each of TAXA, SCORE and ASPT.

The SD of ASPT is also likely to be higher for taxon-poor sites where the ASPT value has been based on averaging the BMWP scores of just a small number of taxa. Therefore, any relationship between the SD of ASPT and the mean number of taxa at a site was also assessed.

In all situations where the SD of the untransformed replicate values showed no systematic relationship with the mean value, then the best estimate, V_{unt}, was taken to be the average of the observed estimates of the sampling variance of the untransformed replicate value. V_{unt} is estimated separately for single, two and three seasons combined samples. In such cases, an α% confidence interval for an observed sample value X is:

$$(X - Z_\alpha SD_{unt}) \quad \text{to} \quad (X + Z_\alpha SD_{unt}) \quad , \quad \text{where } SD_{unt} = \sqrt{V_{unt}} . \qquad \text{(eq 4)}$$

Testing for other sources of variation

Site type

The sixteen sites were selected from four of the twenty-five site classification groups in RIVPACS II. Systematic differences in the size of the within-group variance in observed index values due to the different environmental characteristics of their component sites were therefore assessed by analyses of variance.

R&D Note 412

Where the initial log variance versus log mean regressions indicated that the variance increased with the mean value for a site, the original values were first transformed by taking square roots (or logarithms) and the within-site variances recalculated on the transformed values. This ensured any apparent differences in variability with site type were independent of the average value for the site type.

Non-parametric Kruskal-Wallis analysis of variance by ranks (Siegel 1956) was also used to test for site group differences and this gives the same test statistic whether analyzed on the variance, SD, log variance or log SD values.

If there were no apparent effects on the transformed values due to differences in the character of the sites, then site type can be ignored providing the correct transformation is used in deriving confidence limits for the observed value (ie equations (3) or (4) above).

Site quality

Another aim of the research programme was to assess whether the observed sampling variation differed according to the biological quality of the site, as represented by the four quality bands A,B,C and D (see section 2.1.1). One possibility might be that the same few taxa consistently occur in samples from a poor quality sites such that its SD is both absolutely and relatively low.

Biological quality for a site is assessed from its EQI's for TAXA, SCORE and ASPT. Therefore, it made sense to assess any relationship between SD and site quality simply by the relationship between the SD and the mean value for each site and this approached has been adopted through the log-log regressions.

Subsidiary differences in variance between quality bands were assessed by analysis of covariance. This means that the residual variances about the log-log regression lines were tested by analysis of variance for differences between the four quality bands.

Seasonal differences

Systematic seasonal differences in within-site variance were also assessed in analyses of covariance by simultaneously testing for differences between spring, summer and autumn (for single seasons values of TAXA, SCORE and ASPT) and between spring/summer, spring/autumn and summer/autumn (for two seasons combined values).

Order of sampling

A check was made as to whether there was an significant tendency for the repeated sampling to at least temporarily deplete or disturb the fauna, so that increasingly fewer taxa tended to be caught in the second and third samples.

The three replicate samples were ranked 1 (least taxa), 2 and 3 (most taxa), separately for each single season of each site. Tied values were given the average rank (eg if lowest two values were the same they were given rank 1.5).

A Friedman non-parametric two-way ANOVA of ranks (Siegel, 1956) on site(1-16) and sampling order (1-3) was used to test whether the second and/or third

sample taken tended to have fewer taxa, lower scores or lower ASPT values.

Comparisons between single and multi-person sampling

The aim here was to assess whether the sampling variance is influenced by using different personnel to sample each site. The total variance in replicate values at a site consists of that component due to basic spatial variability at the site and hence in the fauna sampled and an additional inter-person component due to variability in the precise sampling technique of different individuals. At each site one IFE person took the first and third sample while the second sample was taken by a local NRA person.

The basic sampling variance is estimated by V_{13} from the differences ($d_{13} = x_{is1} - x_{is3}$) between the values x_{is1} and x_{is3} for samples 1 and 3 taken at each site, i, in each season, s, as follows:

$$V_{13} = \sum_{i=1}^{16} \sum_{s=1}^{3} (x_{is1} - x_{is3})^2 / (2 \times 47)$$

The variance between samples from different personnel is estimated by V_{12}, where :

$$V_{12} = \sum_{i=1}^{16} \sum_{s=1}^{3} (x_{is1} - x_{is2})^2 / (2 \times 47)$$

The two SD are $SD_{13} = \sqrt{V_{13}}$, $SD_{12} = \sqrt{V_{12}}$. The difference, $SD_{12} - SD_{13}$, is used to estimate the increase in sampling SD due to differences in sampling performance between personnel.

The ratio $F_{pers} = (SD_{12} - SD_{13}) / SD_{12}$ is used to estimate the fraction of the total sampling SD which is due to using different people. If this fraction is small then most of sampling variation is due to intrinsic variability in the precise meso-habitats sampled. This would mean that variation in observed values between years would not be strongly dependent on whether the same person took the sample(s) in both years. This would be a highly desirable conclusion given the obvious difficulties associated with maintaining continuity of staffing over several years sampling. However, this approach assumes the previous analysis of order of sampling showed no general tendency for the replicate values to depend on the order the samples were taken.

Overall estimate of the sampling standard deviation

Where a single variance due to sampling variation in TAXA, SCORE or ASPT is considered to apply to all sites, then it is estimated as the mean of the sampling variance estimates for the individual sites. The precision of this estimate of the common variance is itself estimated as the standard error of the mean of the variance values for the 16 sites.

For ASPT the common variance is denoted by VAR_A. For TAXA and SCORE the

R&D Note 412

common variance is estimated for the square root of the observed values, to give estimates VAR_S and VAR_T. The common standard deviation of sampling variation for all sites is then estimated by $SD_A = VAR_A$, $SD_{\sqrt{T}} = \sqrt{VAR_{\sqrt{T}}}$ and $SD_{\sqrt{S}} = \sqrt{VAR_{\sqrt{S}}}$.

In mathematical terms, for if V_i = estimate of sampling variance for ASPT for site i, when i = 1 to 16, then:

$$VAR_A = \sum_{i=1}^{16} V_i / 16$$

and the standard error of VAR_A is estimated to be

$$SE(VAR_A) = \sqrt{\left[\sum_{i=1}^{16} (V_i - VAR_A)^2 / (15*16)\right]}$$

An identical approach is used for the square root of number of taxa or BMWP score.

2.2 Sampling variation in number of taxa

2.2.1 Variation in relation to observed number of taxa

Analyses

The mean and range of number of taxa observed in each single season's samples and in two and three season combined samples from each of the 16 sites (Table 2.4) shows that the site selection strategy has provided the requisite range of site qualities within each of the four chosen RIVPACS groups. Overall, these ranged from Cherry Cob (site 16), with at most five taxa found in any three season combined sample, to Haslingbourne Bridge (site 5) with an average of 23 taxa in any single sample.

Table 2.4 The mean (minimum to maximum) values of number of TAXA observed at each site in single season, paired and three seasons combined samples. For the two and three seasons samples the statistics are based on all n possible combinations of the appropriate single season samples (n=3, 9 & 27 for 1,2 and 3 seasons combined samples).

Site	Spring	Summer	Autumn	Spring/ Summer	Spring/ Autumn	Summer/ Autumn	All 3 Seasons
1	21.7 (21-23)	20.7 (20-21)	17.7 (14-20)	28.7 (27-30)	25.6 (23-28)	25.3 (22-28)	30.9 (28-33)
2	13.3 (11-17)	17.0 (16-18)	16.0 (15-17)	20.7 (17-24)	18.1 (16-21)	21.6 (20-23)	23.1 (21-26)
3	10.3 (9-11)	13.3 (11-16)	15.3 (14-17)	15.8 (13-18)	16.4 (15-18)	19.0 (17-21)	19.7 (18-22)
4	6.7 (4- 8)	9.3 (9-10)	11.0 (10-12)	9.8 (9-11)	12.0 (10-14)	12.6 (11-14)	12.9 (11-15)
5	12.0 (10-13)	18.7 (17-21)	23.0 (20-27)	21.6 (20-24)	24.1 (20-28)	27.8 (24-31)	28.2 (24-31)
6	17.7 (16-19)	20.0 (19-21)	18.7 (17-20)	21.8 (20-24)	21.1 (19-23)	23.0 (21-24)	23.3 (21-25)
7	11.0 (9-13)	13.3 (12-15)	11.7 (11-12)	15.0 (12-18)	14.6 (13-16)	16.3 (14-18)	17.0 (14-19)
8	6.0 (6- 6)	4.7 (4- 5)	6.7 (6- 8)	7.0 (6- 8)	8.7 (8-10)	7.6 (6-10)	9.6 (8-12)
9	18.7 (17-22)	21.7 (19-25)	18.7 (17-20)	25.0 (23-28)	22.3 (20-24)	24.6 (21-27)	26.0 (23-28)
10	11.7 (10-13)	14.3 (13-15)	9.7 (8-11)	15.7 (14-17)	13.4 (12-15)	15.4 (15-16)	16.4 (15-17)
11	11.3 (10-13)	14.0 (13-15)	13.0 (12-14)	16.0 (15-18)	16.3 (14-18)	16.4 (15-18)	18.2 (16-21)
12	5.0 (3- 7)	7.0 (7- 7)	5.7 (5- 6)	7.0 (7- 7)	6.2 (5- 7)	7.0 (7- 7)	7.0 (7- 7)
13	14.3 (13-16)	20.3 (18-22)	21.0 (20-22)	23.0 (20-25)	23.1 (22-24)	25.6 (23-28)	26.6 (24-29)

14	6.7 (6- 8)	10.7 (10-11)	8.3 (7- 9)	10.8 (10-11)	9.4 (8-10)	11.6 (10-12)	11.6 (10-12)
15	13.3 (12-15)	14.7 (13-16)	17.0 (16-18)	17.7 (17-19)	18.8 (18-20)	19.2 (17-21)	20.3 (18-22)
16	3.3 (3- 4)	3.0 (3- 3)	3.0 (2- 4)	3.3 (3- 4)	3.7 (3- 5)	3.3 (3- 4)	3.7 (3- 5)

The mean SD in the number of taxa observed in samples from any one single season (S1), or from combined seasons samples from any one pair of seasons (S2) or from three seasons combined samples (S3) tends, as expected, to be greater on sites with more taxa (Table 2.5).

Table 2.5 The mean and standard deviation (SD) of the number of TAXA observed in single (S1), two seasons (S2) and three seasons (S3) combined samples for each study site. The mean and SD are estimated from the replicate samples and derived combined season samples separately for each single season or season combination, and then averaged across seasons or combinations of seasons.

Site No	Site / River	Mean TAXA			SD TAXA		
		S1	S2	S3	S1	S2	S3
1	South Dornaford / River Okement	20.0	26.5	30.9	1.65	1.63	1.49
2	Tantons Plain / River Darracott	15.4	20.1	23.1	1.74	1.89	1.54
3	Croxdale House / Croxdale River	13.0	17.1	19.7	1.73	1.35	1.14
4	B6313 / Twyzell Burn	9.0	11.4	12.9	1.30	0.94	1.04
5	Haslingbourne Bridge / Petworth Brook	17.9	24.5	28.2	2.47	2.24	2.24
6	Woodford Bridge / Sheppey River	18.8	22.0	23.3	1.35	1.29	1.20
7	Bowlish / Sheppey River	12.0	15.3	17.0	1.37	1.43	1.30
8	ptc Bedford Brook / Moss Brook	5.8	7.7	9.6	0.58	1.07	1.28
9	Seend Bridge / Summerham Brook	19.7	24.0	26.0	2.49	1.90	1.49
10	Swarkestone / Cuttle Brook	11.9	14.9	16.4	1.40	0.92	0.64
11	Jenny Mill / Poulshot Stream	12.8	16.3	18.2	1.18	1.19	1.42
12	Dewsbury / Spen Brook	5.9	6.7	7.0	0.86	0.22	0.00

13	Brigg / Old River Ancholme	18.6	23.9	26.6	1.54	1.39	1.39
14	Ferry Sluice / Broad Rife	8.6	10.6	11.6	0.96	0.68	0.64
15	U/S Skellingthorpe /Skellingthorpe Main Drain	15.0	18.6	20.3	1.35	0.91	1.03
16	Cherry Cob / Keyingham Drain	3.1	3.4	3.7	0.53	0.57	0.68

This tendency for variance and hence SD to increase with increasing number of taxa is shown in a series of scatter plots of the log of the variance (SD^2) against the mean number of taxa observed (Figure 2.1, left-hand-side).

There are separate plots for single season, two seasons combined and three seasons combined data. For example, in Figure 2.1(b) for two seasons combined samples, there is a separate point for each site for the spring/summer, spring/autumn and summer/autumn samples. Having the same axes scales and limits for all three plots immediately shows that for a given average number of taxa in replicate single or combined season samples, the variance tends to be highest for single season values and lowest for three seasons combined samples.

Taylor's power law regressions of log(SD^2) against log(average TAXA) showed significant correlations ($p<0.01$) for each of 1,2 and 3 seasons combined samples indicating that the within-site variance does tend to increase with the mean number of taxa observed on a site (Table 2.6). The regression lines are superimposed on figure 2.1. Moreover, in all cases the regression slope b (Figure 2.1) was not statistically significantly different from unity, indicating that the square roots of the number of taxa ($\sqrt{TAXA}$) will have within-site variances independent of the number of taxa.

Table 2.6 Regression of Log variance against Log mean separately for 1,2 and 3 seasons combined samples for each of number of TAXA, SCORE and ASPT. Log Variance = a + b Log Mean, r^2 = % of variation in Log variance explained.

Index	No.	a ± SE(b)	b± SE(b)	r^2
TAXA	1	-1.79 ±	0.92 ±	22%
	2	-3.03 ±	1.21 ±	47%
	3	-2.40 ±	0.94 ±	52%
SCORE	1	-1.01 ±	1.23 ±	42%
	2	-1.46 ±	1.22 ±	66%
	3	-0.52 ±	0.96 ±	74%
ASPT	1	-3.80 ±	0.26 ±	0%
	2	-4.51 ±	0.28 ±	0%

R&D Note 412

	3	-3.84 ±	-0.44 ±	1%

Conclusion

The variance of numbers of taxa per sample or sample combination increases with numbers of taxa in the sample.

However, in the absence of other influences, the variance of the square roots of the replicate values of TAXA will be the same for all sites and is best estimated by a constant.

In these circumstances, this provides a common rule for estimating variation in observed index values due to sampling.

The next section assesses the influence of other sources of variation in √TAXA values and determines whether, in practical terms, they have a large effect on the size of the sampling variance and hence precision of observed TAXA values, such that the NRA should allow for these factors in their errors assessment for observed TAXA at sites.

R&D Note 412

2.2.2 Analysis of sources of sampling variation in number of taxa

Site quality band

Analyses

The dependence of the size of the within-site variance on river quality has already been determined by establishing that the variance, VAR, tends to increase with the average number of taxa in the actual and derived samples. Analysis of variance on the variance of square roots of replicate values of TAXA showed no systematic subsidiary influences of river quality, in terms of 5M quality bands (A,B,C,D), on the size of the variance of observed number of taxa.

Conclusion

All the influence of river quality on sampling variability can be determined by the observed number of taxa.

Site type

Analyses

The main aim in this section is to assess whether the variability in TAXA tends to differ between river types. For this reason the SD of the square root of the observed number of taxa is plotted (Figure 2.2) for each single season, pair of seasons and all three seasons combined samples for each of the 16 sites, grouped into their four RIVPACS II classification groups (3a, 5b, 8a and 9b).

There was no detectable general tendency for the observed TAXA of any one type of site to be more or less variable than the others. Analyses of variance (ANOVA) and Kruskal-Wallis non-parametric ANOVA were used to test for differences in sampling variance of the square root of the number of taxa between RIVPACS groups. These analyses were done separately for single season, two and three seasons combined samples and in no cases was there any significant differences (p>0.05) (Table 2.7(a)).

Conclusion

There was no statistically detectable difference between the variation in observed numbers of taxa due to between site differences in quality.

Season

Analyses

Any systematic differences in mean variability between the three single

R&D Note 412

seasons were also assessed by simultaneously including a factor denoting season in the ANOVA. There was some visual suggestion that variation in TAXA (and hence the highly correlated SCORE) was slightly higher in the spring (figure 2.2(a)). However, this was mostly due to high spring variation at site 2 (12, 11 and 17 taxa observed in samples 1, 2 and 3) and at site 4 (4, 8 and 8 taxa observed) and there were no statistically significant seasonal patterns to the size of sampling variation ($p > 0.05$).

Conclusion

There were no statistically detectable seasonal differences between the variation in observed numbers of taxa.

R&D Note 412

Table 2.7 Analysis of differences between site types in the mean sampling variance of (a) the square root of TAXA, (b) the square root of SCORE and (c) the ASPT (untransformed). Means are calculated separately for single and two and three seasons combined samples. Site type is based on four sites from each of the RIVPACS groups 3a, 5b, 8a and 9b.

p, p$_k$ respectively denote the significance levels of a one-way ANOVA and Kruskal-Wallis non-parametric ANOVA of ranks testing for differences in sampling variance between site types.

(a) Mean variance of replicates of √ TAXA

No. of seasons combined in sample	RIVPACS group / site type						
	3a	5b	8a	9b	Overall mean = VAR$_{\sqrt{T}}$	p	p$_k$
1	0.0698	0.0455	0.0579	0.0344	0.0519	0.43	0.86
2	0.0320	0.0377	0.0208	0.0173	0.0269	0.06	0.13
3	0.0202	0.0322	0.0140	0.0179	0.0211	0.15	0.42

(b) Mean variance of replicates of √ SCORE

No. of seasons combined in sample	RIVPACS group / site type						
	3a	5b	8a	9b	Overall mean = VAR$_{\sqrt{S}}$	p	p$_k$
1	0.584	0.230	0.331	0.238	0.346	0.11	0.42
2	0.259	0.164	0.148	0.130	0.175	0.12	0.11
3	0.161	0.136	0.093	0.130	0.130	0.53	0.83

R&D Note 412

(c) Mean variance of replicates of ASPT

No. of seasons combined in sample	RIVPACS group / site type						
	3a	5b	8a	9b	Overall mean = VAR_A	p	p_k
1	0.1066	0.0308	0.0458	0.0641	0.0618	0.13	0.44
2	0.0386	0.0131	0.0139	0.0379	0.0259	0.04	0.01
3	0.0256	0.0071	0.0061	0.0388	0.0194	0.33	0.08

R&D Note 412

Order of sampling

Analyses

A Friedman non-parametric two-way ANOVA of ranks (Siegel, 1956) on sites (1-16) and sampling order (1-3) found no statistically significant (p>0.05) overall trends or differences in the number of taxa caught according to order of sampling (Table 2.8). On detailed inspection and analysis, this was found to be generally true for each site type. There was some suggestion that less taxa and lower BMWP scores were obtained in the second and third samples from sites in RIVPACS group 5b, but this was only statistically significant for SCORE (p=0.02) (Table 2.8).

Conclusions

On the small river sites, one might expect the first sampling to remove a significant fraction of the fauna, but the careful on-site sampling procedures will eliminate most potential problems.

Few statistically detectable differences existed in the current database between the index values of samples collected at different stages of the sampling sequence of individual sites in individual seasons.

This is important as it increases the validity of the three samples to provide estimates of the standard deviation of single samples for the future and also for current comparisons of differences in replicate samples from the same person (the first and third samples of this study) with those between different people.

Variation between samples taken by different people relative to samples taken by the same person

Analyses

The overall SD of replicate values of √TAXA for a single season, averaged across all seasons and sites, is estimated by SD_{13} when based on two samples taken by the same person, and by SD_{12} when based on samples taken by different people (Table 2.9). In the same table an estimate of the overall SD based on all three replicate samples in each season, denoted by SD_O, is given for comparison. In effect, SD_O gives a single estimate of the average SD in √TAXA values irrespective of whether the same or different people took the samples on the separate occasions.

As might be expected, SD_{13} is slightly higher than SD_{12}, but the difference estimates that only about 12% of the overall sampling SD (F_{pers} in Table 2.9) is due to differences between personnel in sampling.

Conclusion

Given the imprecision in estimating the various standard deviations themselves, the estimated 12% extra source of variation due to inter-operator differences is not sufficient to justify any complicated allowance for whether the same or

R&D Note 412

different people took the one or more samples to be compared from different years for site quality assessment.

Summary of conclusions

We recommend that the overall estimate, $VAR_{\sqrt{T}}$, of the variance of $\sqrt{TAXA}$, in Table 2.7(a), be used to estimate the precision of the value for the observed number of taxa for all sites, with a separate value for single seasons, two seasons combined samples and three season combined samples.

The same variance estimates can be used irrespective of site type, season combination and whether the same of different people took the samples on different occasions.

The conclusion that inter-operator differences are trivial is dependant on the assumption that all the people must have been trained in field sampling procedures.

Table 2.8 Mean of the ranks of (a) TAXA, (b) SCORE and (c) ASPT (lowest rank = 1, highest rank = 3) in the first, second and third samples taken in each season at each site, averaged over all sites and separately for each RIVPACS site group (3a, 5b, 8a and 9b).

p_k = significance levels of a Kruskal-Wallis or Friedman non-parametric ANOVA of ranks testing for differences due to sampling order. N_1, N_2, N_3 denote the number of times the first, second and third sample taken had the highest value (joint highest counts excluded).

(a) Mean rank of number of taxa

Order of sample	RIVPACS group / site type				Overall mean
	3a	5b	8a	9b	
First	1.71	2.29	2.17	1.75	1.98
Second	2.00	2.04	1.95	2.37	2.09
Third	2.29	1.67	1.88	1.88	1.93
p_k	0.19	0.12	0.63	0.10	0.66
N_1	2	4	5	1	12
N_2	3	3	3	5	14
N_3	4	1	1	3	9

(b) Mean rank of BMWP score

Order of sample	RIVPACS group / site type				Overall mean
	3a	5b	8a	9b	
First	1.75	2.50	2.17	1.83	2.06
Second	2.00	1.75	2.00	2.25	2.00
Third	2.25	1.75	1.83	1.92	1.94
p_k	0.34	0.02	0.60	0.37	0.81
N_1	2	8	5	2	17
N_2	5	3	4	4	16
N_3	5	0	2	4	11

(c) Mean rank of ASPT

Order of sample	RIVPACS group / site type				Overall mean
	3a	5b	8a	9b	
First	1.71	2.29	2.17	1.75	2.09

Second	2.00	2.04	1.96	2.37	1.88
Third	2.29	1.67	1.87	1.88	2.03
p_k	0.89	0.19	0.18	0.26	0.51
N_1	4	4	6	2	16
N_2	4	2	3	2	11
N_3	4	4	2	6	16

R&D Note 412

Table 2.9 Inter-operator sampling variation. Estimates of:

the overall sampling standard deviation, SD_O, based on all three single season replicate samples;

the basic spatial sampling standard deviation, SD_{13} , based on the first and third samples taken by the same person and

the standard deviation, SD_{13}, based on the first and second sample taken by two different people.

$F_{pers} = 100(SD_{12} - SD_{13}) / SD_{12}$ = percentage of overall sampling SD due to inter-person variability.

N_{more} , N_{less} = number of cases (out of 16 sites by 3 seasons = 48) where the difference between replicate values for different people was more, and less, respectively than the difference in the two samples values from the same person.

	$\sqrt{}$ TAXA	$\sqrt{}$ SCORE	ASPT
SD_O	0.228	0.588	0.249
SD_{13}	0.217	0.559	0.249
SD_{12}	0.247	0.612	0.259
F_{pers}	12%	9%	4%
N_{more}	20	25	20
N_{less}	19	20	24

2.3 Sampling variation in BMWP score

2.3.1 Variation in relation to observed value of bmwp score

Analyses

Clarke et al (1994) showed that BMWP score is highly correlated with the number of BMWP taxa at a site and hence in site quality assessments it is mostly redundant if EQI's based on number of taxa and ASPT are used. However, BMWP score, or more precisely the EQI for SCORE, based on RIVPACS predictions, still has merit as a single overall quality index.

The mean and range of observed BMWP scores in each single season's samples and in two and three season combined samples from each of the 16 sites (Table 2.10) provides further verification that the sixteen selected sites covered the major range of quality conditions. The total score for a single season's sample ranged from 6 (site 16) to 162 (site 1).

R&D Note 412

The mean SD's of observed SCORE in samples from any one season (S1) or from paired seasons samples (S2) or from three seasons combined samples (S3) are given in Table 2.10.

Table 2.10 The mean (minimum and maximum) values of SCORE observed at each site in single season, paired and three seasons combined samples. For the two and three seasons samples the statistics are based on all n possible combinations of the appropriate single season samples (n=3, 9 & 27 for 1,2 and 3 seasons combined samples).

Site	Spring	Summer	Autumn	Spring/ summer	Spring/ autumn	Summer/ autumn	All 3 seasons
1	150(143-16	123(115-12	98(	190(178-20	168(153-18	151(131-17	201(181-22
2	63 (46-	85(80- 93)	78(66- 90)	109(	92(	116(103-13	127(109-14
3	41(34- 45)	53(45- 65)	63(56- 72)	66(54- 76)	68(59- 77)	80(70- 90)	83(73- 95)
4	26(12- 34)	41(36- 48)	44(41- 47)	43(36- 52)	49(41- 58)	54(46- 64)	56(46- 68)
5	60(50- 66)	94(	120(108-13	116(104-12	125(108-13	151(134-16	153(134-16
6	83(75- 88)	97(	86(76- 92)	105(	104(	114(102-12	115(102-12
7	46(37- 55)	56(51- 64)	46(44- 48)	62(51- 75)	62(57- 68)	69(60- 77)	71(60- 80)
8	21(21- 21)	12(9- 14)	21(18- 26)	25(21- 29)	33(30- 38)	24(18- 34)	36(30- 46)
9	90(	97(	84(72- 95)	121(106-13	108(	117(	127(110-13
10	41(32- 47)	54(51- 57)	31(27- 34)	62(54- 69)	50(42- 57)	59(55- 64)	66(57- 70)
11	42(38- 49)	51(46- 55)	49(44- 53)	62(56- 73)	66(56- 73)	63(55- 70)	73(62- 88)
12	13(6- 20)	20(20- 20)	15(14- 17)	20(20- 20)	17(14- 20)	20(20- 20)	20(20- 20)
13	66(52- 79)	89(	94(	108(	107(	121(102-13	127(114-14
14	20(18- 24)	34(32- 35)	23(18- 27)	34(32- 35)	30(24- 33)	38(32- 41)	38(32- 41)
15	46(42- 53)	54(50- 61)	62(60- 65)	67(59- 73)	70(65- 76)	74(62- 82)	79(65- 89)
16	7(6- 9)	6(6- 6)	7(5- 11)	7(6- 9)	9(6- 14)	8(6- 11)	9(6- 14)

The left-hand-side of Figure 2.3 comprises scatter plots the logarithm of the sampling variance of SCORE against the mean observed SCORE, with a separate point for each site by season(s) combination. The relationship between variance and mean SCORE is very similar to that for TAXA.

Taylor's power law regressions relating variance to mean (Table 2.6; regression lines superimposed on Figure 2.3), suggest that the square roots of the values of SCORE will have a variance roughly independent of the value of SCORE. This is verified in the plots on the right-hand side of figure 2.3, where the deviations of the replicate values of √SCORE from the mean for the site and season combination show no dependence on the value of SCORE.

Conclusion

The SD of √SCORE can therefore be estimated by a constant, which may depend on other sources of sampling variation.

R&D Note 412

Other possible sources of variation in SCORE due to sampling are assessed in the next section.

2.3.2 ANALYSIS OF SOURCES OF SAMPLING VARIATION IN BMWP SCORE

Analyses

The sampling variation in SCORE at the 16 study sites is summarised in Tables 2.10 and 2.11.

Table 2.11 The mean and standard deviation (SD) of observed SCORE in single(S1), two seasons(S2) and three seasons(S3) combined samples for each study site. The mean and SD are estimated from the replicate samples and derived combined season samples separately for each single season or season combination, and then averaged across seasons or combinations of seasons.

Site No	Site / River	Mean SCORE			SD SCORE		
		S1	S2	S3	S1	S2	S3
1	South Dornaford / River Okement	124	170	201	13.2	11.7	10.0
2	Tantons Plain / River Darracott	75	106	127	13.7	13.8	11.5
3	Croxdale House / Croxdale River	52	71	83	8.3	6.7	6.2
4	B6313 / Twyzell Burn	37	49	55	7.1	5.4	5.7
5	Haslingbourne Bridge / Petworth Brook	91	131	153	12.1	10.1	10.0
6	Woodford Bridge / Sheppey River	89	108	115	7.4	7.9	7.3
7	Bowlish / Sheppey River	49	65	71	6.1	5.7	5.2
8	ptc Bedford Brook / Moss Brook	18	27	36	2.3	4.1	4.9
9	Seend Bridge / Summerham Brook	90	115	127	15.0	11.9	9.2
10	Swarkestone / Cuttle Brook	42	57	66	4.8	5.0	3.7
11	Jenny Mill / Poulshot Stream	47	64	73	5.1	5.4	6.6
12	Dewsbury / Spen Brook	16	19	20	2.9	0.8	0.0
13	Brigg / Old River Ancholme	83	112	127	12.8	8.6	7.3
14	Ferry Sluice / Broad Rife	26	34	38	3.3	2.8	2.9

R&D Note 412

| 15 | U/S Skellingthorpe /Skellingthorpe Main Drain | 54 | 70 | 79 | 5.0 | 5.2 | 6.5 |
| 16 | Cherry Cob / Keyingham Drain | 7 | 8 | 9 | 1.6 | 2.3 | 2.8 |

All the tests for effects of river type, season combination, order of sampling and differences in personnel on variation in √TAXA were repeated for the square root of BMWP Score.

The were no statistically significant influences of site type, as represented by RIVPACS site group, on the size of the sampling SD of √SCORE (Figure 2.4 and Table 2.7(b)). There is perhaps some suggestion that the variance might greater on sites of type 3a, which are coarse-bottomed hill sites (see Section 2.1.1), but this was not consistent enough to merit accepting.

Although there was some (statistical) suggestion (p<0.02) that the value of SCORE tended to be lower in the second and third samples from sites of type 5b (Table 2.8(b)), there was no overall detectable effect of order of sampling on the value of SCORE obtained in the replicate samples (p=0.81). For site types 3a and 9b, the average rank of the values of SCORE was actually lowest for the samples taken first (Table 2.8(b)).

There were no subsidiary differences in the sampling SD of √SCORE according to either site quality, (A,B,C,D), or season.

The SD of √SCORE based on replicate samples taken by different people ($SD_{12} = 0.612$ in Table 2.9) was only marginally higher than that based on replicate samples taken by the same person ($V_{13} = 0.559$), such that the percentage of overall sampling variation estimated to be due to inter-operator sampling effects was only 9%. The difference in SCORE between the first and third samples (both taken by the same person) was actually greater in 20 of the 48 cases (Table 2.9). This suggests that using different personnel in different seasons has little influence on the variability, assuming, as in this sampling programme, that only properly trained people involved.

Conclusions

For a given sample or combination of two or three seasons samples a common value for SD of √SCORE can be used to estimate the precision in the observed BMWP score for all sites, irrespective of site type, biological condition of the watercourse or (trained) person taking the sample.

This common value will vary according to whether the SCORE is based on one, two or three seasons samples but will be constant for each type of seasonal combination.

2.4 Sampling variation in ASPT

2.4.1 Variation in relation to observed value of ASPT and number of taxa

Analyses

The mean and range of observed ASPT values in each single season's samples and in two and three season combined samples from each of the 16 sites are shown in Table 2.12.

The highest observed ASPT was 7.04 at site 1 in spring. At Cherry Cob (site 16), the worst quality site sampled, most samples only found three taxa (Oligochaeta, BMWP score 1), Chironomidae (score 2) and Valvatidae (score 3), giving an ASPT of 2.0.

The average and SD of the ASPT observed for samples from any one single season (S1), or from combined season samples from any one pair of seasons (S2), or from three seasons combined samples (S3) are given in Table 2.13.

A feature of Table 2.13 is that the average value of ASPT for a site tends to increase slightly with the number of seasons combined (ie S1 to S2 to S3). For every site, except site 7, the average observed ASPT for two season combined samples is always higher than that for one season, and the average ASPT for three seasons combined samples is always slightly higher than for two.

This is thought to be because the lower scoring taxa tend not to have aerial stages in their life history and, when present at a site, occur in all seasons. Conversely the higher scoring taxa are

R&D Note 412

insects with aerial stages which preclude their presence in the water at certain times of year. There may also be a tendency for lower scoring animals to be present in greater numbers than many insect families at sites were they each occur.

Table 2.12 The mean (minimum and maximum) values of observed ASPT at each site in single season, paired and three seasons combined samples. For the two and three seasons samples the statistics are based on all n possible combinations of the appropriate single season samples (n=3, 9 & 27 for 1,2 and 3 seasons combined samples).

Site	Spring	Summer	Autumn	Spring/ summer	Spring/ autumn	Summer/ autumn	All 3 seasons
1	6.90(6.81-7.04)	5.90(5.75-6.14)	5.50(5.29-5.85)	6.60(6.45-6.82)	6.60(6.23-6.84)	6.00(5.70-6.25)	6.50(6.23-6.75)
2	4.60(4.18-5.18)	5.00(4.82-5.17)	4.90(4.40-5.29)	5.20(4.82-5.54)	5.10(4.75-5.38)	5.40(5.15-5.65)	5.50(5.19-5.67)
3	4.00(3.78-4.09)	4.00(3.77-4.09)	4.10(4.00-4.24)	4.20(3.86-4.40)	4.10(3.93-4.28)	4.20(4.06-4.35)	4.20(4.06-4.38)
4	3.70(3.00-4.25)	4.30(4.00-4.80)	4.00(3.92-4.10)	4.40(4.00-4.80)	4.10(3.92-4.18)	4.30(4.00-4.67)	4.30(4.00-4.67)
5	5.00(4.85-5.08)	5.00(4.89-5.29)	5.20(4.96-5.40)	5.40(5.19-5.52)	5.20(4.96-5.40)	5.40(5.29-5.58)	5.40(5.29-5.58)
6	4.70(4.58-4.89)	4.90(4.60-5.05)	4.60(4.47-4.74)	4.80(4.60-5.04)	4.90(4.86-5.09)	5.00(4.83-5.13)	4.90(4.83-5.13)
7	4.20(4.11-4.27)	4.20(3.92-4.33)	4.00(3.92-4.00)	4.20(3.92-4.33)	4.30(4.20-4.43)	4.20(4.17-4.33)	4.20(4.11-4.33)
8	3.50(3.50-3.50)	2.50(2.25-2.80)	3.10(3.00-3.25)	3.50(3.43-3.63)	3.80(3.75-3.88)	3.20(3.00-3.40)	3.80(3.67-3.88)
9	4.80(4.77-4.94)	4.40(4.11-4.76)	4.50(4.24-4.75)	4.90(4.61-4.96)	4.80(4.67-4.92)	4.70(4.36-4.96)	4.90(4.63-5.04)
10	3.50(3.20-3.67)	3.80(3.67-3.92)	3.20(3.09-3.38)	4.00(3.80-4.06)	3.70(3.50-3.86)	3.80(3.67-4.00)	4.00(3.80-4.12)
11	3.70(3.55-3.80)	3.70(3.54-3.79)	3.80(3.67-3.85)	3.80(3.73-4.06)	4.00(3.93-4.06)	3.90(3.67-4.00)	4.00(3.88-4.19)
12	2.40(2.00-2.86)	2.90(2.86-2.86)	2.70(2.50-2.83)	2.90(2.86-2.86)	2.80(2.50-2.86)	2.90(2.86-2.86)	2.90(2.86-2.86)
13	4.60(4.00-4.94)	4.30(4.06-4.59)	4.50(4.10-4.81)	4.70(4.30-4.84)	4.60(4.32-4.86)	4.70(4.43-4.88)	4.80(4.60-4.93)
14	3.00(3.00-3.00)	3.20(3.18-3.20)	2.80(2.57-3.00)	3.20(3.18-3.20)	3.10(3.00-3.30)	3.30(3.18-3.42)	3.30(3.18-3.42)
15	3.40(3.31-3.53)	3.70(3.33-3.85)	3.70(3.61-3.75)	3.80(3.47-4.00)	3.70(3.61-4.00)	3.90(3.65-4.05)	3.90(3.61-4.19)
16	2.10(2.00-2.25)	2.00(2.00-2.00)	2.40(2.00-2.75)	2.10(2.00-2.25)	2.30(2.00-2.80)	2.30(2.00-2.75)	2.30(2.00-2.80)

The Taylor's power law regressions of log variance versus log mean value of ASPT did not yield any statistically significant relationships (slopes b for ASPT in Table 2.6 not significantly different from zero). In contrast to TAXA and SCORE, the variance in ASPT therefore shows no tendency to be higher in situations where the average value of ASPT is higher. This is seen in plots of the sampling SD of ASPT against the average observed value of ASPT, with a separate point for each

R&D Note 412

site by season(s) combination (left-hand-side of Figure 2.5).

R&D Note 412

Table 2.13 The mean and standard deviation (SD) of the observed ASPT in single(S1), two seasons(S2) and three seasons(S3) combined samples for each study site. The mean and SD are estimated from the replicate samples and derived combined season samples separately for each single season or combination of seasons averaged across seasons or seasonal combinations.

Site	Site / River	Mean ASPT			SD ASPT		
		S1	S2	S3	S1	S2	S3
1	South Dornaford / River Okement	6.1	6.4	6.50	0.2	0.1	0.14
2	Tantons Plain / River Darracott	4.8	5.2	5.47	0.3	0.2	0.16
3	Croxdale House / Croxdale River	4.0	4.1	4.23	0.1	0.1	0.10
4	B6313 / Twyzell Burn	4.0	4.2	4.29	0.3	0.2	0.22
5	Haslingbourne Bridge / Petworth Brook	5.0	5.3	5.44	0.1	0.1	0.11
6	Woodford Bridge / Sheppey River	4.7	4.9	4.95	0.1	0.1	0.10
7	Bowlish / Sheppey River	4.1	4.2	4.20	0.1	0.0	0.06
8	ptc Bedford Brook / Moss Brook	3.0	3.5	3.78	0.1	0.0	0.06
9	Seend Bridge / Summerham Brook	4.5	4.8	4.89	0.2	0.1	0.10
10	Swarkestone / Cuttle Brook	3.4	3.8	4.02	0.1	0.1	0.09
11	Jenny Mill / Poulshot Stream	3.7	3.9	4.00	0.1	0.0	0.08
12	Dewsbury / Spen Brook	2.6	2.8	2.86	0.2	0.0	0.00
13	Brigg / Old River Ancholme	4.4	4.6	4.77	0.3	0.1	0.07
14	Ferry Sluice / Broad Rife	2.9	3.2	3.28	0.0	0.0	0.10
15	U/S Skellingthorpe /Skellingthorpe Main	3.5	3.7	3.91	0.1	0.1	0.15
16	Cherry Cob / Keyingham Drain	2.1	2.2	2.31	0.1	0.2	0.34

It might be thought that the value of ASPT observed for a site would be more variable when the ASPT was based on few taxa. The right-hand side of figure 2.5 comprises scatter plots of observed SD of ASPT against the average number of taxa on which the ASPT values used to estimate the SD were based.

On average, the SD does not tend to decease systematically with the number of taxa on which it is based. However, there is a tendency for the estimates of the SD for ASPT to be much more variable when based on fewer taxa and average ASPT is low. This is especially true for single season estimates of SD which are based on only three replicate values. This tendency is investigated further in Section 2.5.

Conclusion

n the absence of other factors affecting sampling variation, the SD for ASPT based

on single, two or three seasons combined data may be best estimated by three constants (shown as horizontal lines in figure 2.5), regardless of the value of ASPT or the number of taxa present.

2.4.2 Analysis of sources of sampling variation in ASPT

Analyses

All the tests for effects of river type, season combination and order of sampling on variation in √TAXA were repeated for the sampling variance of the untransformed values of ASPT.

There was some suggestion that the sampling variance of ASPT was greater for sites from RIVPACS types 3a and 9b (Figure 2.6 and Table 2.7(c)). These site differences were statistically significant for two season combined samples ($p < 0.01$) but not for either single season or three season combined values of ASPT.

For each or one, two and three seasons data, the SD of ASPT was highest for type 3a sites, which also had the highest average value of observed ASPT. The group with the next most variable replicate values of ASPT was site type 9b, which had the lowest mean value of ASPT. Within the 3a group of sites, two sites (site 2 at Tantons Plain and site 4 on the Twyzell Burn) had higher replicate SD of ASPT in each combination of seasons (Table 2.10), but another site (site 1 on at South Dornaford) had highest average ASPT.

Overall, there does not appear to be a consistent interpretable pattern to these potential differences between site types. This and the wide variation in estimates of the SD of sites from each type leads us to recommended that the overall mean sampling variances for ASPT given in Table 2.7(c) are used to represent the sampling precision of ASPT, irrespective of the type of site. However, the influence of site type on ASPT precision may merit further study.

There was no evidence that the order of sampling had any influence on the value of ASPT obtained (Table 2.8(c). The third sample taken had the highest of the three replicate values of ASPT in as many situations as the first sample.

There were no subsidiary differences in the sampling variance of ASPT according to site quality (A,B,C,D) or season.

The SD of ASPT based on replicate samples taken by different people ($SD_{12} = 0.259$ in Table 2.9) was only marginally higher than that based on replicate samples taken by the same person ($V_{13} = 0.249$), such that the estimate of the percentage (F_{pers}) of overall sampling variation estimated to be due to inter-person sampling effects is only 4%. The difference in ASPT between the first and third samples (both taken by the same person) was actually greater than the difference between samples taken by different people in half the cases. (Table 2.9). This suggests that using different personnel in different seasons or years has no influence on the value and precision of estimates of ASPT.

Scatter plots of the within-site sampling variation of ASPT as residuals about the site mean ASPT value (Figure 2.7) re-enforce the conclusion that the size of the

sampling variation in ASPT does not generally depend on either the ASPT or the number of taxa involved.

Frequency distribution histograms of the within-site sampling variation in ASPT, the square of TAXA and the square root of SCORE (Figure 2.8) show that, in each case, the distribution is not grossly skewed. Instead it is roughly symmetrical and can be approximated by a normal distribution. The same figure also graphically displays how the effect of sampling variation is less for combined season samples than for single season samples.

R&D Note 412

R&D Note 412

R&D Note 412

Conclusions

The variation in ASPT due to sampling is best estimated by a set constant SD values irrespective of the type of site.

A tendency for greater variation in observed ASPT values in certain site types may, however warrant further investigation.

A separate constant applies to each of single, paired and three seasons values.

These constants are not dependant upon site quality, season or sampling personnel provided the latter are adequately trained.

Summary

We recommend that the sampling variation of ASPT for single or combined season samples is assumed to follow a normal distribution with a constant SD, where the constant depends on the number of seasons involved.

2.5 Variation at very poor quality sites

Although the average value of SD of ASPT may not vary consistently in relation to the mean values of TAXA or ASPT, the observed variation in the estimates of SD appears to be greater for taxon-poor sites (Figure 2.5). This is because when few taxa are present, the presence or absence of each taxon in the single or combined sample can have a large influence on ASPT. Such an observed pattern is therefore likely to be due to the individual estimates of the SD themselves having large errors with a skewed distribution. This is best illustrated by a worked example.

At Cherry Cob on Keyingham Drain (site 16 in Table 2.2) only 5 different taxa were found in total in the nine samples taken from the site over the three seasons sampled (Table 2.14).

The samples and their taxon lists from this site can be used as a realistic example to provide estimates of the probability of observing each taxon in any single sample. For simplicity of presentation two minor adjustments have been made to the real data and its treatment. Firstly, Valvatidae were taken to be present in all nine samples rather than the actual number of eight. Secondly probabilities of capture have been averaged across all seasons.

Accepting these estimates as correct for the site, the observed values in any one sample range between 3-5 for TAXA, 6-14 for SCORE and 2-2.8 for ASPT. Moreover, if three replicate samples are taken in one season, there is a 42% chance that all three samples will only contain the same three taxa (Valvatidae, Oligochaeta and Chironomidae and Valvatidae, as in sample type A of Table 2.14). If this happened the estimate of the SD for a site of this type and quality would be zero for each of TAXA, SCORE and ASPT.

R&D Note 412

At the other extreme, it is possible that the three samples could be of type A, C, and D. Then the estimates of the SD for TAXA, SCORE and ASPT would be 1.00, 4.07 and 0.45 respectively. The true SD for the site are 0.44, 1.83 and 0.24 respectively, so an estimate for a taxon-poor site based on three samples could estimate the SD to be twice as high as it really is.

Table 2.14 Simplified observed data for the taxon-poor site 16, Cherry Cob, showing the probability (P) of observing each taxon, estimated from the proportion of times they were caught and observed in nine samples. Treating the observed taxa as the only taxa truly present at the site and the P values as correct for the site, only four different combinations of taxa in a sample, (A,B,C,D) are possible. These would occur with probabilities Q_A, Q_B, Q_C and Q_D. Knowing the values of Q, the true mean value of TAXA, SCORE and ASPT for single samples from the site can be calculated, together with the true SD of such samples.

			All possible samples for this					
			A	B	C	D		
Oligochaeta	1	1	1	1	1	1		
Chironomida	2	1	1	1	1	1		
Valvatidae	3	1	1	1	1	1		
Glossiphoniid	3	1/9	0	1	0	1		
Corixidae	5	1/9	0	0	1	1		
							True	True
Number of TAXA in sample			3	4	4	5	3.22	0.44
Sample BMWP SCORE			6	9	11	14	6.89	1.83
Sample ASPT			2	2.25	2.7	2.8	2.11	0.24

Summary

It is recommended that it would be inappropriate to use the observed variation in the estimates of SD or variance as a measure of uncertainty in the true variance at any one site. The observed values of variance, especially for poor quality sites, would almost certainly grossly overestimate the true variation in sampling variance between sites.

It is probably more accurate just to use the average variance together with the 95% confidence limits of the average variance.

2.6 Sampling variation in the average of two or three single season's observed values

All previous assessments of site quality from multiple season's data have been based on amalgamating the individual single season samples into a combined season sample. It is then from this combined sample that the observed values of

TAXA, SCORE and ASPT have been compared with the RIVPACS predictions, for the same combined seasons, in order to derive EQI values.
However, it has been suggested that using the average of the EQI values for the individual seasons may give a more accurate index of quality.

It must be remembered that the combined season EQI and the average of the single season EQI are defining site quality in different ways and hence are not measuring the same thing. It is important to bear in mind that the question addressed in this section is not whether one index is a better representation of quality at a site over one year but whether one index can be estimated more precisely than the other.

The error in any EQI will depend on the error in estimating both the observed and expected values of number of taxa, BMWP score or ASPT. In this section, the magnitude of the sampling variation in the observed values are compared for combined and averaged index values.

The overall estimates of the sampling standard deviation in the observed values of ASPT and the square root of each of TAXA and SCORE for single season samples, for two season combined samples and for three season combined samples are given in Table 2.15.

Table 2.15 Overall estimates of sampling variation in the observed values of TAXA, SCORE and ASPT for single and combined seasons samples.

The observed values of ASPT are assumed to have a constant variance, VAR_A. The variance of the square root of the observed TAXA is assumed to be a constant $VAR_{\sqrt{T}}$. The variance of the square root of the observed SORE is assumed to be a constant $VAR_{\sqrt{s}}$. Standard errors (SE) for each estimate of VAR are given in brackets. $SD_A = \sqrt{VAR_A}$, $SD_{\sqrt{T}} = \sqrt{VAR_{\sqrt{T}}}$ and $SD_{\sqrt{s}} = \sqrt{VAR_{\sqrt{s}}}$.

Number of seasons in combined sample	ASPT		TAXA		SCORE	
	VAR_A	SD_A	$VAR_{\sqrt{T}}$,	$SD_{\sqrt{T}}$	$VAR_{\sqrt{S}}$	$SD_{\sqrt{S}}$
1	0.0618 (0.0120)	0.249	0.0519 (0.0078)	0.228	0.346 (0.059)	0.588
2	0.0259 (0.0043)	0.161	0.0269 (0.0030)	0.164	0.175 (0.021)	0.418
3	0.0194 (0.0072)	0.139	0.0211 (0.0030)	0.145	0.130 (0.016)	0.361

If SD_{A1} is the sampling standard deviation for observed values of ASPT in single seasons, then the standard deviation for the average of three single season observed values for ASPT can be estimated by ($SD_{A1} / \sqrt{3}$). Similarly the SD of

R&D Note 412

the average of two single season observed values was estimated by $(SD_{A1} / \sqrt{2})$.

The estimated sampling standard deviation for the average of either two or three single seasons values of observed ASPT is slightly higher that for the combined season values of ASPT using the same two or three samples, (Table 2.16), but the differences are not statistically significant ($p>0.05$). However, the average expected value of ASPT tends to be slightly higher for two and three season combined sample than for single, or the average of single, season samples, so, relatively, there is a greater sampling SD for the average of the observed values of single season's samples.

R&D Note 412

Table 2.16 Comparison of the estimates of sampling standard deviations in the observed values of TAXA, SCORE and ASPT derived from combined season samples (SD$_{Com}$) and those from the average of single season samples (SD$_{Av}$). The sampling standard deviations have been standardised by dividing by the appropriate mean expected value (Exp$_{Av}$) for ASPT, or √(Exp$_{Av}$) for TAXA and SCORE.

Number of seasons	ASPT			TAXA			SCORE		
	Exp$_{Av}$	SD$_{Av}$	SD$_{Com}$	Exp$_{Av}$	SD$_{Av}$	SD$_{Com}$	Exp$_{Av}$	SD$_{Av}$	SD$_{Com}$
1	5.37			22.1			119		
2	5.57	3.2%	2.9%	27.4	3.4%	3.1%	154	3.8%	3.4%
3	5.66	2.7%	2.5%	30.2	2.8%	2.6%	172	3.1%	2.8%

It is more difficult to make comparisons for TAXA and SCORE because the expected value is inevitably much higher for combined season samples than for the average of single seasons.

As a simple overall comparison, the sampling standard deviations of the square roots of number of taxa and BMWP score for two and three seasons combined sample values (Table 2.15) have been standardised by dividing by the square roots of the mean expected values for number of taxa and BMWP score for two and three season combined samples for the 5006 NRA sites from the 1990 River Quality Survey (Clarke et al 1992). The sampling SD for the average of two or three single season observed number of taxa or BMWP score have been standardised by dividing by the corresponding square root of the average expected value for single season samples for the 5006 sites (Table 2.16). This scales the sampling SD as a percentage of the expected value and hence in terms of change in EQI (albeit perhaps on the square root scale).

Following these procedures, the sampling SD as a percentage of the average expected value was shown to be slightly higher using the average of two or three single season observed values than using the corresponding combined season sample value for each of observed number of taxa, BMWP score and ASPT (Table 2.16).

The analyses in this section were used solely to examine the effect of variation between samples taken on the same day in each of the three seasons, spring, summer and autumn. Another source of variation, especially for the average of

R&D Note 412

single season sample values, is the within-season, temporal variation in TAXA, SCORE and ASPT. This may not be large relative to between-operator variation on any one day. However, we expect that its inclusion in SD estimates, if possible, would increase the overall sampling variation in index values calculated from averaged single season values to a greater extent than when combined season observed values were used.

Conclusions

In terms of minimising the effects of sampling variation, there is no apparent advantage in using the average of single season values to increase precision of index estimates in compared to the use of combined season samples.

R&D Note 412

2.7 **Summary**

1. Sixteen running-water sites, covering a wide and balanced range of environmental characteristics and biological condition (quality), were sampled in each of three seasons. On each occasion, two replicates sample were taken by one IFE person and a third by a local NRA person. From this information, estimates of sampling variation in number of taxa, BMWP score and ASPT were obtained.

2. There was no strong evidence that the sampling variance for each of number of taxa, BMWP score and ASPT in single, two or three seasons combined samples varied consistently with site type or season(s).

3. Using different people to takes the samples only increase the sampling variance by a very small amount. Using the same person over time is not important, but all samplers must be trained to a consistently acceptable standard. A single variance estimate based on a mixture of replicate samples from the same and different people was therefore derived for each index in each season and each number of seasons combined.

4. The overall sampling variance of the square root of the observed number of taxa in a single or combined seasons sample is best estimated by a constant, $VAR_{\sqrt{T}}$, which depends on the number of seasons' samples which have been combined, as detailed in Table 2.15.

5. The overall sampling variance of the square root of the observed BMWP score in a single or combined seasons sample is best estimated by a constant, $VAR_{\sqrt{S}}$, which depends on the number of seasons' samples which have been combined, as detailed in Table 2.15.

6. The overall sampling variance of the observed ASPT in a single or combined seasons sample is best estimated by a constant, VAR_A, which is independent of the value of ASPT or the number of taxa in the sample, but depends on the number of seasons' samples which have been combined, as detailed in Table 2.15.

7. The standard errors for the sampling variance estimates in Table 2.15 quantify the degree of uncertainty in estimating the sampling variances and could themselves be used by the NRA towards providing a range of values for the confidence limits of an EQI or quality band assessment.

8. The sampling standard deviation for the average of the observed value of number of taxa, BMWP score and ASPT in two or three single season samples is greater than the equivalent sampling standard deviation for the observed values of combined seasons samples.

R&D Note 412

3 VARIATION DUE TO SORTING AND IDENTIFICATION ERRORS

3.1 Introduction

Between 1990 and 1994 IFE have undertaken quality audits of the performance of NRA staff members at sorting and removing all the different families present in macro-invertebrate samples and identifying, to BMWP family level, the specimens that were removed. Emphasis was placed on those families in the BMWP system (Chesters 1980).

Over that period a total of 2,892 samples have been audited (Table 3.1), including representatives of each of the three "seasons" of collection.

Table 3.1 The number of NRA macro-invertebrate samples, by NRA region, audited by IFE in between 1990 and 1994. Regions given in this table are as at the end of 1994.

REGION	1990	1991	1992	1993	1994
Anglian	76	60	60	60	60
Northumbria & Yorkshire	95	120	120	101	80
North West	61	60	60	60	60
Severn-Trent	56	60	62	60	60
Southern	63	60	60	60	60
South Western	98	120	108	120	120
Thames	35	60	60	60	60
Welsh	79	60	60	60	60
TOTAL	563	600	600	581	560

The error rates pertaining in the 1990 data-set were analyzed by the Water Research Centre (Kinley and Ellis 1991) in relation to their significance to the design of an internal analytical control procedure for NRA use. Their study showed major inter-regional differences in performance. However the analyses concentrated on the number of taxa missed and no detailed study was made of the effect of missing taxa on the derivation of site EQI's and, by implication the assessment of the biological condition of sites. Neither did the report include information on the type of families missed or wrongly identified.

R&D Note 412

All these issues are now considered here.

3.2 Quality audit procedures

3.2.1 Sample selection

Samples for audit were selected internally by each of the agencies being monitored. The biologists processing these samples had no prior knowledge of the samples to be audited.

The manner of sample selection, which biologists would be monitored and the number of audit samples from each season, were left to the discretion of the agency, within the limits of the total number of samples that IFE was contracted to audit.

3.2.2 Sample processing

The normal protocol for NRA biologists was to sort their samples within the laboratory and to select examples of each scoring taxon within the BMWP system. In most cases, the macro-invertebrates were placed in a vial of preservative (4% formaldehyde solution or 70% industrial alcohol) and the BMWP taxa were listed on a data sheet. The vial of animals and the sorted material were then returned to the sample container and preservative added. Thus, each sample available to IFE for audit should have included:

- a list of the BMWP families found in the sample

- a vial containing representatives from each family

- the preserved sample

When these three elements were present, the sequence of operations at IFE was as follows:

- The remainder of the sample was sorted and the BMWP families listed

- The families contained within the vial were identified and listed

- A comparison was made between the NRA listing of families and those identified from the vial by IFE

- A comparison was made between the NRA listing of families and those found in the sample by IFE

- "Losses" or "gains" from the NRA listing of families were noted

For a number of different reasons, some samples did not include a vial containing representative examples of the families listed on the data sheet.

R&D Note 412

Others arrived with the vial damaged in transit such that the representative examples were no longer separated. For these samples NRA's accuracy of identification could not be checked.

3.2.3 Reporting

The results of each sample audit were recorded on a standard report form. Each form had separate boxes, A and B, for recording different types of error. Each box was subdivided into "gains" and "losses" sections. Gains were families found by IFE but not recorded as present by the NRA whilst losses were NRA listed taxa not found by IFE.

R&D Note 412

For audit samples where a vial of animals was included, the comparison between the NRA listing and the taxa found in the vial by IFE was shown in box A of the report form. Discrepancies could be due to carelessness, mis-identifications or errors in completing the NRA data sheet. Families not on the NRA listing but found by IFE in the remainder of the sample were entered in box B of the report form under "additional families".

When the families listed as "losses" in section A of the report form were compared with the full list of families recorded in the sample by IFE, some apparent losses from the vial were offset by the presence of those families in the remainder of the sample. These taxa were therefore listed in the "losses" box of section A and the "gains" box of section B and were neither a net loss nor a net gain. Such errors were categorised as "omissions".

Where the NRA data sheet indicated that a family was noted and released at the site, this was recorded in the notes section but not included as a "loss", even though the family was not found in the vial.

For those samples in which the vial of animals was damaged or missing many of these procedures could not be implemented and this limited the value of the audit exercise.

The results of the audit exercise were notified to the NRA as a series of annual regional reports of which Gunn et al (1991) is a typical example. Further details of the audit procedures were listed in each such report.

3.2.4 Selection of audited samples for data analysis

From the full data-base of available samples (Table 3.1) a manageable sub-set of approximately 400 was targeted for analysis in the current study.

These were chosen in approximately equal quantities from the samples audited in 1990 and 1992. These years were chosen because, they represented, respectively, the worst and best overall performances for the period 1990 - 1993. The 1994 audits had not been completed at the time of selection. This strategy ensured that the full range of performance was considered.

Selection was limited to sites that had been sampled in each of the three RIVPACS seasons during the year of audit. This enabled assessment of the proportion of taxa missed by the NRA in one season's sample that were found by the NRA in a second or third season's sample to be undertaken. Only taxa missed in one season and not found in the other season's sample contribute to any bias in the observed taxa list for the combined season sample. The few suitable sites (6 in each of 1990 and 1992) which were audited in more than one season within a year were all selected.

Only those samples which were accompanied by a separate vial of specimens for identification audit were included in the data-base used for selection.

A three-dimensional stratification matrix (Table 3.2) was used for random site selection. Its dimensions comprised NRA regions by the NRA's estimated number of taxa (1-10, 11-20, 21-30, 31+), by the three RIVPACS sampling seasons.

Table 3.2 The number of samples selected, at random, from each category. A = 1-10, B = 11-20, C = 21-30 and D = 31-41 taxa recorded by NRA in a sample. NRA regions are as in 1990.

NRA REGION	SEASON	1990					1992				
		A	B	C	D	All	A	B	C	D	All
	SPRING	4	11	2	0	17	1	3	0	0	4
	SUMMER	1	8	1	0	10	1	3	4	0	8
	AUTUMN	2	2	1	0	5	1	5	6	0	12
	SPRING	1	5	2	0	8	1	4	3	0	8
	SUMMER	2	3	0	0	5	1	5	2	0	8
	AUTUMN	1	0	0	0	1	1	3	3	1	8
	SPRING	6	5	0	0	11	2	0	3	0	5
	SUMMER	2	2	0	0	4	0	1	0	0	1
	AUTUMN	2	0	0	0	2	1	2	0	0	3
	SPRING	1	7	0	0	8	0	2	2	0	4
	SUMMER	2	2	1	0	5	0	4	3	0	7
	AUTUMN	4	5	2	0	11	0	4	2	1	7
	SPRING	0	8	5	1	14	1	5	3	0	8
	SUMMER	0	3	2	0	5	3	2	3	1	9
	AUTUMN	1	3	2	0	6	1	1	4	1	7
	SPRING	0	3	6	5	14	0	1	4	1	6
	SUMMER	2	0	7	0	9	0	2	4	0	6
	AUTUMN	0	1	0	0	1	1	3	2	0	6
	SPRING	1	0	3	1	5	2	1	5	1	9
	SUMMER	0	5	2	1	8	1	2	3	2	8
	AUTUMN	0	1	2	0	3	0	4	3	0	7
	SPRING	2	4	4	0	10	3	6	2	1	12
	SUMMER	2	4	4	0	10	0	5	4	0	9
	AUTUMN	0	0	0	0	0	0	0	0	0	0
	SPRING	2	3	2	0	7	1	0	4	3	8
	SUMMER	1	2	2	0	5	1	4	2	1	8
	AUTUMN	0	0	2	0	2	1	4	3	0	8
	SPRING	7	5	2	1	15	0	4	5	0	9
	SUMMER	1	3	2	0	6	1	3	5	0	9
	AUTUMN	0	0	2	0	2	0	1	5	0	6
	Total	46	95	59	9	209	25	84	89	13	211

R&D Note 412

The aim was to obtain a wide coverage and representative balance within the constraints of the audited samples available. Not all cells had samples available for selection. The initial computer-generated stratified random sample selection was further modified by hand to ensure a full coverage of taxa richness within each region and year.

On this basis, a final total of 420 sites were selected for final analysis (Table 3.2), divided into 209 from 1990 and 211 from 1992.

3.2.5 Recording of sample taxa lists

Listings of the biological data (faunal lists) of all samples collected and processed as part of the 1990 River Quality Survey were previously obtained from John Steel (Thames NRA) and held on the IFE Microvax II computer as standard-format ASCII files. Within these files the samples subject to quality audit were not amended to incorporate the errors detected during the checking process. These data files are referred to here as Type I NRA sample data.

As part of this analysis, we re-coded all the original sample information provided by the NRA on photocopies of their original data recording sheets to provide a second list of the taxa that the data sheets say the NRA found in the sample, referred to as Type II NRA sample data. The Type I and II data lists for each sample from 1990 only were compared. NRA's 1992 biological data-files were not held by IFE.

Any recording errors in the Type I NRA data lists, detected by reference to both the Type II files and NRA's original data-sheets were the corrected to produce amended Type I files. Next any errors detected by the IFE audit were then used to create a "true" BMWP family lists for each site. This was achieved by editing the amended Type I NRA data-files. The resultant, revised data-files were called the IFE audit lists. This ensured that differences between the NRA's taxa list and the audited sample taxa list were not due to coding and typing errors. It also provided some information on the size of recording and typing errors for taxonomic data. This time-consuming additional analysis was not specifically part of the contract, but IFE considered it important to assess this additional source of error in the observed sample values.

Each BMWP family in each sample in the new IFE audit lists was flagged using a detailed coding scheme. Flags were summarised to the four point scheme of Table 3.3.

Table 3.3 Summary IFE data coding of taxa in the IFE audit lists

Recorded as present in whole sample (vial + residual sample) by	Code	Terminology
Neither NRA nor IFE audit	0	
Both NRA and IFE audit	1	
IFE audit only	2	"gain"
NRA but not found in IFE audit	3	"loss"

R&D Note 412

The observed numbers of taxa, BMWP scores and ASPT values for each audited sample (Table 3.2) were calculated twice, first using the NRA's record of taxa present (amended Type I NRA files) and secondly using the IFE audit list of the taxa actually present in the sample. The differences (IFE minus NRA) was the degree of under-estimation, or bias, in each of TAXA, SCORE and ASPT. The average size of the under-estimation and any factors on which this may depend were assessed.

3.2.6 Derivation of combined season audited samples

The effects of sample processing errors on the observed, single season's values of number of taxa, BMWP score and ASPT are examined in Sections 3.3 - 3.7 respectively. The single season information from the audited samples is the basic data of the effect of processing errors. However, the NRA also needs to know the effect of sample processing errors on the BMWP index values for combined season samples at a site.

Unfortunately, of the several hundred NRA samples audited by IFE each year (Table 3.1) very few sites were audited in more than one season in any one year. In each of 1990 and 1992, there were only six sites whose samples which were audited in two seasons and no site was audited in all three seasons. Thus there are too few combined season audited samples to estimate directly the size of processing errors on combined season sample index values.

However, the effects of processing errors on two season combined samples can be assessed by combining estimates of the under-estimation of number of taxa in any one single season sample with estimates of the proportion of taxa missed in one sample which are recorded by the NRA as present in a second sample. If most taxa which are missed in one sample are found and recorded as present in another season's sample from the same site, then the effect of sample processing errors will be much smaller in combined season samples.

The proportion of missed taxa recorded as present in a second sample will be assessed in two ways:

Firstly using just the 12 sites for which two audited samples are available. This is the only information we have to judge the extent to which the same taxa tend to be missed in each season at a site.

Secondly, for each of the 203 sites audited from 1990 (Table 3.2 - six sites were audited in two seasons), the original Type I NRA data files were used to extract the NRA lists of the taxa they believed to be present in of the three single season samples from the same site. This enabled estimations to be made of the proportion of taxa missed by the NRA in one seasons's sample which they find and record as present in either two seasons combined samples or three seasons combined samples.

R&D Note 412

3.3 COMPARISON OF NRA AND IFE AUDITED SAMPLE TAXA LISTS

These analyses are only based on the subset of IFE audited samples (Table 3.1) that were re-analyzed for this project, namely 209 of the 563 available samples for 1990 and 211 of the 600 samples audited in 1992.

3.3.1 Data logging and typing errors

Analyses

The original NRA data files (Type I NRA data - section 3.2.5) were compared with this projects re-assessment of what the NRA recorded on their data sheets as being present in each audited sample in 1990. Within the 209 samples involving a total of 3543 taxa data entries, there were a total of 67 discrepancies spread over 47 samples. On re-checking the original data sheets, 32 of the discrepancies were due to encoding mistakes by IFE and 35 were mistakes by NRA staff. Thirty samples had only one error, but one sample had four errors made by an NRA person.

The total number of transcription errors made by NRA staff in each NRA region was never large enough to highlight inter-region differences, but it is merely noted that no such errors were recorded for North-West, Wessex or Yorkshire NRA regions in 1990.

The transcription and coding errors made by IFE staff were corrected before continuing with the analyses in this project.

Conclusion

There appears to be about a 1% error rate in transcribing taxonomic information from paper sheets onto computer file.

This error rate, if maintained or improved, is negligible compared with the other variations involved in the sampling, sorting and identification processes.

3.3.2 Frequency of missing individual taxa

The audited samples provide estimates of the proportion of times that individual taxa are missed by the NRA in their sample sorting and identification procedures. The data for 1990 and 1992 have been analyzed separately because it is known (IFE unpublished) that the sample processing error rates in most NRA regions were markedly lower in 1992 than in 1990.

R&D Note 412

The overall proportions of times that the NRA missed each BMWP taxa present in samples in 1990 and 1992 for all NRA regions combined are given in Table 3.4.

As an example, *Planariidae* were present in 83 of the IFE audited samples in 1990, but missed in 16.9% by the NRA. In the 1992 samples, the NRA missed *Planariidae* in only 9.3% of the 97 samples in which the IFE audit found it.

The taxa which the NRA as a whole seemed to have had the most difficulty picking out and identifying in 1990 are marked with an asterisk in Table 3.4 (***** actually denotes taxa missed at least 4 times and in at least 20% of the samples in which it was present). Those taxa missed by the NRA in over 25% of the samples in which they were present were *Dendrocoelidae, Valvatidae, Physidae, Planorbidae, Hydrophilidae, Scirtidae, Psychomyiidae, Hydroptilidae, Goeridae, Lepidostomatidae* and *Brachycentridae*. The most frequently missed taxa (ie over 20 times in 1990) were *Hydrobiidae, Lymnaeidae, Planorbidae, Sphaeriidae, Hydrophilidae, Elmidae* and *Hydroptilidae*, but this is partly because they are common.

As noted in section 3.3.1, by 1992 the number of taxa missed by the NRA had clearly decreased. This improvement is considered in more detail in section 3.4. The taxa which the NRA, as a whole, still seemed to have a problem picking out and/or identifying are marked with a + in Table 3.4. These are *Dendrocoelidae, Valvatidae, Caenidae, Taeniopterygidae, Haliplidae, Hydrophilidae, Scirtidae, Psychomyiidae, Hydroptilidae, Beraeidae* and *Goeridae.*

The NRA needs to know which of their regions were most frequently missing particular taxa. Table 3.5 gives a list of the number of times any particular taxon was missed by the NRA in audited samples from each 1990 NRA region. Only taxa that were missed in more than one sample in one year from at least one region are shown.

The sample sizes and especially the number of cases individual taxa were missed within each region are often small, but remembering this limitation, the information may be informative to the regional NRA biologists in showing where they may still have sample sorting or identification problems.

Conclusions

Overall the NRA missed 15.3% of all taxa occurrences in the 209 samples audited in 1990, but only 8.3% of those in samples audited in 1992.

NRA biologists competence in the removal and identification of those taxa flagged by a * and/or a + in Table 3.4 or shown to be a particular regional problem should be improved by further internal training and testing and by individual awareness of the problem areas.

Table 3.4 Overall proportion of IFE audited samples in 1990 and 1992 for which each family was present but not recorded by the NRA, or recorded as present by the NRA but not found when audit by IFE.

I90, I92 = number of times a family was present in samples audited by IFE in 1990 and 1992

G90, G92 = percentage of I90, I92 samples respectively when family was found by IFE but not recorded by the NRA.

L90, L92 = percentage of N90, N92 samples in 1990 and 1992 respectively when family was recorded as present by the NRA, but not found in the IFE audit. (Taxa not recorded as present in more than one sample in one year have been excluded).

*, + denote taxa commonly missed in 1990 and 1992 respectively.

Family	I90	I92	G90	G92	N90	N92	L90	L92
Planariidae (incl. Dugesiidae)	83	97	17	9	72	90	4	2
Dendrocoelidae	17	26	* 53	+ 19	9	24	11	13
Neritidae	12	16	17	6	11	15	9	0
Viviparidae	3	3	33	0	2	4	0	25
Valvatidae	42	51	* 43	+ 23	28	40	14	3
Hydrobiidae (incl. Bithyniidae)	161	149	14	11	140	135	1	2
Lymnaeidae	117	125	18	12	98	111	2	1
Physidae	48	46	* 29	9	35	48	0	12
Planorbidae	78	90	* 29	10	58	86	5	6
Ancylidae (incl. Acroloxidae)	110	115	16	10	95	105	1	1
Unionidae	14	10	7	10	13	9	0	0
Sphaeriidae	150	169	* 22	8	118	156	1	0
Oligochaeta	204	206	7	1	189	204	0	0
Piscicolidae	32	42	9	9	29	38	0	0
Glossiphoniidae	136	140	10	5	123	133	0	0
Erpobdellidae	110	127	6	3	106	124	2	1
Asellidae	126	133	5	7	122	125	2	2
Corophiidae	5	3	20	0	4	3	0	0
Gammaridae (incl. Crangonyctidae & Niphargidae)	171	169	7	1	160	167	0	0
Astacidae	3	5	0	0	3	5	0	0
Baetidae	163	177	10	4	147	170	0	0
Heptageniidae	74	81	1	2	72	79	0	0
Leptophlebiidae	55	43	* 25	9	41	40	0	2
Ephemerellidae	78	86	19	7	65	81	0	1
Ephemeridae	50	57	4	4	49	56	0	2
Caenidae	89	85	* 22	+ 15	71	73	1	1
Taeniopterygidae	26	20	15	+ 20	22	16	0	0
Nemouridae	55	48	* 22	4	43	46	0	0
Leuctridae	59	74	14	4	56	71	4	0
Perlodidae	47	46	6	7	47	46	4	6
Perlidae	12	12	17	8	11	11	9	0
Chloroperlidae	32	25	* 22	8	25	23	0	0
Platycnemididae	0	3	0	0	1	3	100	0
Coenagriidae	29	28	3	14	28	24	0	0
Calopterygidae	25	28	16	7	22	26	4	0

R&D Note 412

Cordulegasteridae	5	6	20	0	4	6	0	0
Aeshnidae	1	2	0	0	1	2	0	0
Mesovelidae	1	1	100	0	4	3	100	67
Hydrometridae	2	6	50	17	2	5	50	0
Gerridae	11	9	9	0	13	10	23	10
Nepidae	2	3	0	33	2	2	0	0
Naucoridae	3	1	33	0	2	1	0	0
Aphelocheiridae	4	10	0	10	4	9	0	0
Notonectidae	14	10	14	10	12	9	0	0
Corixidae	49	52	4	6	48	49	0	0
Haliplidae	68	71	19	+ 20	57	59	2	3
Dytiscidae (incl. Noteridae)	125	112	11	4	113	109	2	2
Gyrinidae	44	44	18	9	36	40	0	0
Hydrophilidae (incl. Hydraenidae)	82	67	* 42	+ 28	49	51	2	6
Scirtidae (=Helodidae)	6	21	* 50	+ 29	3	15	0	0
Dryopidae	1	3	100	67	0	1	0	0
Elmidae	154	160	17	9	127	147	0	1

Table 3.4 (continued)

Family	I90	I92	G90	G92	N90	N92	L90	L92
Sialidae	45	58	9	10	42	52	2	0
Rhyacophilidae	83	87	13	3	73	86	1	2
(incl. Glossosomatidae)								
Philopotamidae	6	5	17	20	5	4	0	0
Polycentropodidae	47	66	8	9	44	61	2	2
Psychomyiidae	25	41	* 40	+ 34	16	27	6	0
(incl. Ecnomidae)								
Hydropsychidae	115	138	6	6	108	131	0	1
Hydroptilidae	48	67	* 50	+ 21	25	53	4	0
Phryganeidae	3	6	0	17	4	5	25	0
Limnephilidae	74	108	12	6	66	103	2	1
Molannidae	9	6	33	17	8	5	25	0
Beraeidae	4	5	50	+ 80	2	1	0	0
Odontoceridae	7	13	14	8	6	12	0	0
Leptoceridae	76	88	* 24	8	60	83	3	2
Goeridae	30	41	* 40	+ 27	19	30	5	0
Lepidostomatidae	47	48	* 32	15	36	44	11	7
Brachycentridae	15	23	* 33	4	11	22	0	0
Sericostomatidae	62	79	11	6	57	74	4	0
Tipulidae	109	135	16	13	93	120	1	2
Simuliidae	106	134	19	3	86	130	0	0
Chironomidae	204	211	2	1	202	208	1	0
TOTAL	**4103**	**4475**	**15.3**	**8.3**	**3543**	**4159**	**1.9**	**1.3**

Table 3.5 Number of audited samples from each NRA region in 1990 and 1992 in which each taxa was missed by the NRA. Only taxa missed in more than two samples from at least one region in one year are included.

The NRA regions are as in 1990, namely:

1=Anglian, 2=Northumbrian, 3=North-West, 4=Severn-trent, 5=Southern, 6=South-West, 7=Thames, 8=Welsh, 9=Wessex, 10=Yorkshire

	------------ 1990 -----------										----------- 1992 ------------									
NRA Region	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10
No. of samples audited	32	14	17	24	25	24	16	20	14	23	24	24	9	18	25	18	24	21	24	24
Planariidae	5	2	1	1	0	3	0	0	1	1	3	0	0	3	0	0	1	1	0	1
Valvatidae	7	0	0	2	3	0	1	1	4	0	3	0	0	1	3	0	3	1	1	0
Hydrobiidae	4	1	6	1	3	1	1	2	1	3	2	1	5	1	1	0	4	0	0	3
Lymnaeidae	3	2	5	4	2	1	2	1	0	1	0	0	2	3	2	1	4	0	1	2
Physidae	5	0	1	2	2	0	2	1	1	0	0	0	0	0	0	0	3	0	1	0
Planorbidae	7	1	0	5	1	2	2	1	4	0	3	0	0	0	1	0	4	0	1	0
Ancylidae	2	1	4	5	0	1	1	1	2	0	0	0	1	1	2	0	4	1	2	0
Sphaeriidae	8	3	7	4	3	0	1	4	3	0	1	0	2	2	0	0	2	3	1	2
Oligochaeta	4	2	3	1	0	1	0	0	3	1	0	0	0	0	0	1	1	0	0	0
Piscicolidae	0	0	0	1	0	0	0	1	0	1	3	0	0	0	1	0	0	0	0	0
Glossiphoniidae	3	2	1	4	1	0	1	1	1	0	2	0	0	0	0	2	2	1	0	0
Asellidae	2	0	1	1	0	0	0	1	1	0	2	1	0	2	1	0	3	1	0	0

R&D Note 412

```
Gammaridae         1 0 2 0 0 2 3 2 2 0   1 0 0 0 0 1 0 0 0 0
Baetidae           3 0 1 5 1 0 2 0 3 1   2 0 0 0 1 1 2 0 0 1
Leptophlebiidae    1 1 0 2 2 5 0 1 2 0   1 0 0 1 0 0 0 1 0 1
Ephemerellidae     1 1 1 2 3 1 0 2 2 2   2 0 0 1 1 0 0 1 1 0
Caenidae           4 4 3 2 1 1 2 1 2 0   1 0 0 1 0 2 1 3 3 2
Leuctridae         0 2 2 0 0 1 0 0 3 0   1 0 0 0 0 1 0 0 1 0
Coenagriidae       1 0 0 0 0 0 0 0 0 0   3 0 0 0 1 0 0 0 0 0
Haliplidae         1 1 1 2 3 0 2 2 1 0   2 1 0 1 1 0 3 1 3 2
Dytiscidae         1 3 0 4 1 3 0 2 0 0   1 0 0 0 1 0 0 1 1 1
Hydrophilidae      7 4 0 5 3 2 4 1 6 2   2 1 3 0 1 2 3 3 1 3
Elmidae            8 0 3 7 1 0 2 1 5 0   1 1 1 1 1 0 6 2 1 0
Rhyacophilidae     1 1 3 2 1 1 0 1 1 0   0 0 0 1 0 0 0 0 0 2
Psychomyiidae      3 1 1 2 0 1 0 2 0 0   0 1 0 3 1 1 7 1 0 0
Hydropsychidae     1 0 1 0 1 1 0 1 2 0   0 0 0 1 0 0 2 2 3 0
Hydroptilidae      6 1 2 4 3 1 4 1 1 1   3 0 0 2 0 4 1 1 2 1
Leptoceridae       3 2 2 3 1 3 1 2 1 0   1 1 0 1 1 1 1 1 0 0
Goeridae           1 0 0 1 1 3 2 1 2 1   0 0 0 1 0 0 3 0 2 5
Lepidostomatidae   0 2 3 3 0 4 0 2 1 0   0 2 0 1 0 0 0 1 3 0
Tipulidae          3 1 3 4 0 0 1 1 4 0   2 0 1 0 1 1 6 2 2 2
Simuliidae         3 2 3 5 1 2 0 2 2 0   2 0 0 0 0 0 1 0 0 1
```

3.3.3 Taxa incorrectly recorded as present

Analyses

All cases where the NRA recorded a taxa as being present in a sample but it was absent from IFE audit of the sample were identified separately. The right-hand-side of Table 3.4 is a summary of the percentage of times the NRA, as a whole, mistakenly recorded a taxon as being present when it was not in the sample or vial provided to IFE nor listed on the accompanying NRA data sheet as captured but released on site. These errors could be due to mis-identifications or mis-coding of taxa once identified but it is not possible to separate the two causes.

The most obvious observation is that taxa were incorrectly assumed to be present much less frequently than they were missed. In 1992 the overall proportion of taxa incorrectly recorded as present was 1.3% (56/4159), which is, on average, 0.26 per sample, or about one taxa every four samples.

The taxa most commonly incorrectly assumed present were *Dendrocoelidae, Mesovelidae, Gerridae* and *Lepidostomatidae* (in both 1990 and 1992), *Valvatidae* (1990 only), and *Physidae* (1992 only) (Table 3.4). However, the number of cases of each was usually no more than one per region per year. The only noteworthy case was that *Physidae* were incorrectly recorded as present in four audited samples from Anglian NRA in 1992.

Conclusions

The number of taxa incorrectly assumed to be present in samples is negligible compared to the number of taxa missed and will not be examined further.

However, in analyses of the errors in the observed number of taxa, BMWP score and ASPT, both these sources of added and missed taxa are incorporated.

3.4 Effect of sample processing errors on observed number of taxa

3.4.1 Comparison of 1990 and 1992

As seen in section 3.3, the NRA have a much greater tendency to miss taxa than to incorrectly record them as present. This means that the overall effect of sample processing errors is to under-estimate the number of taxa in the sample. The IFE audited value for number of taxa minus the amended NRA Type I value measures the degree of under-estimation or bias in the NRA's sample data. This bias includes the combined effect of missing taxa and incorrectly recording taxa as present. The bias determines the extent to which the NRA may under-estimate the true biological quality of the site. This section

examines the extent to which the bias has been reduced in each NRA region between 1990 and 1992. More importantly, it assesses the relationship between the bias and the number of taxa the NRA estimated as being present in a sample.

R&D Note 412

Statistics on the average under-estimation of the number of taxa in each season of 1990 and 1992 for each NRA region (as they existed in 1990) are provided in Table 3.6, while Table 3.7 gives the averages by region in each year. As is already known from elsewhere (Kinley and Ellis 1991) many NRA regions were missing, on average 3 or 4 taxa from each sample in 1990. This was especially true during the spring 1990, the first season to be audited (Table 3.6). In 1990 samples up to 8 or 9 taxa were missed and even 15 taxa from one sample from spring.

After nationwide improvements in the NRA's sample processing procedures, the average under-estimation of the number of taxa in 1992 samples was reduced to 2.0 or less in all regions, except for Thames (where a lapse in quality of sample processing in autumn 1992 led to on average 4 taxa being missed per autumn sample).

In a completely separate 'package 2' of the NRA R&D project, the Water Research Centre (van Dijk 1994) have drawn up a quality control scheme for sample processing and auditing for an agreed tolerable under-estimation rate of an average of two taxa per sample. The results here show that, in general, this target was achieved by the NRA regions in 1992, but not 1990. Therefore, the next stage of this analysis will concentrate on the 1992 audited samples, assuming they represent the sample processing quality that will be achieved in the 1995 and future surveys.

3.4.2 Number of taxa missed in relation to number recorded as present

Analyses

Although the average number of taxa missed varied between 1.5 (1992) and 2.7 (1990), there may be a tendency for more taxa to be missed in samples with more taxa present. However, the NRA do not know how many taxa there really are in a sample, they only have their own estimate. Therefore, to be of use to the NRA, analyses were undertaken to assess whether there is a relationship between the under-estimation in the number of the taxa and the NRA's own estimate of the number of taxa.

The average under-estimation of number of taxa in samples grouped according to the NRA estimate of the number of taxa in each sample in 1992 is given in Table 3.8. For all classes of NRA estimated number of taxa (except the class 21-25 taxa), the NRA under-estimated the number of taxa by no more than one taxa in at least 50% of the samples (ie the median in Table 3.8 is one). This is encouraging. However, because several taxa are missed in a few samples, the statistical mean number missed is higher than one (range 1.0 - 1.9 in Table 3.8).

There was no firm evidence that the average under-estimation of number of taxa was strongly correlated with NRA's estimate of the total number present. Even where the NRA found 1-5 taxa (n=4) the average under-estimation was

R&D Note 412

still 1.0 compared to 1.5 in samples where the NRA recorded over 25 taxa (n=40 samples).

The under-estimation may be slightly higher than elsewhere in samples where the NRA recorded intermediate taxonomic richness (21-25 taxa). Where the NRA recorded over 30 taxa, the number missed was never more than five taxa. This pattern has some logic in that the NRA are likely to have recorded their very highest values for number of taxa in samples where they did not miss many.

R&D Note 412

Table 3.6 Mean, median (med) and maximum under-estimation of the number of taxa in each of the (n) audited samples for each season of each NRA region in 1990 and 1992. (Under-estimation = Number confirmed as present by the IFE audit minus the number recorded as present on the NRA data-sheet, i.e amended NRA Type I value)

		1990				1992			
		n	Mean	Med	Max	n	Mean	Med	Max
-	SPRING	17	3.4	4.0	6	4	0.5	0.0	2
	SUMMER	10	3.2	4.0	5	8	2.8	2.0	8
	AUTUMN	5	3.6	4.0	6	12	2.1	1.5	8
-	SPRING	8	4.4	5.0	9	8	0.5	0.5	1
	SUMMER	5	1.6	1.0	5	8	1.0	1.0	3
	AUTUMN	1	2.0	2.0	2	8	0.5	0.0	2
-	SPRING	11	4.6	4.0	9	5	3.2	3.0	5
	SUMMER	4	3.3	2.5	8	1	-1.0	-1	-11
	AUTUMN	2	1.5	1.5	2	3	0.7	0.0	2
.	SPRING	8	3.0	3.0	5	4	1.5	1.0	4
	SUMMER	5	4.0	6.0	7	7	2.1	2.0	4
	AUTUMN	11	4.3	3.0	8	7	1.6	1.0	7
—	SPRING	14	1.9	2.0	5	9	0.4	1.0	2
	SUMMER	5	1.6	2.0	5	9	1.7	1.0	4
	AUTUMN	6	1.2	0.5	5	7	1.4	1.0	4
-	SPRING	14	2.4	2.0	6	6	0.5	0.5	2
	SUMMER	9	1.0	1.0	3	6	1.8	1.5	5
	AUTUMN	1	0.0	0.0	0	6	0.8	1.0	2
—	SPRING	5	1.8	2.0	4	9	2.2	2.0	5
	SUMMER	8	2.4	3.0	4	8	2.0	1.5	4
	AUTUMN	3	1.7	1.0	4	7	3.9	2.0	8
-	SPRING	10	1.1	1.0	4	12	1.3	1.0	4
	SUMMER	10	2.8	3.0	7	9	1.3	1.0	4
	AUTUMN	0	*	*	*	0	*	*	*
-	SPRING	7	7.9	9.0	15	8	1.3	1.0	5
	SUMMER	5	3.4	3.0	8	8	1.3	1.0	4
	AUTUMN	2	0.5	0.5	2	8	1.1	1.0	4
. -	SPRING	15	0.7	1.0	3	9	0.9	1.0	3
	SUMMER	6	1.2	1.5	2	9	2.3	2.0	5
	AUTUMN	2	1.5	1.5	2	6	0.8	0.5	2

R&D Note 412

Table 3.7 Mean and standard deviation (SD) of the under-estimation of the number of taxa in each of the (n) single season audited samples of each NRA region in 1990 and 1992.

	1990			1992		
	n	Mean	SD	n	Mean	SD
1. ANGLIAN	32	3.4	1.8	24	2.0	2.2
2. NORTHUMBRIAN	14	3.2	3.1	24	0.7	0.8
3. NORTH-WEST	17	3.9	3.1	9	1.9	2.0
4. SEVERN-TRENT	24	3.8	2.1	18	1.8	1.8
5. SOUTHERN	25	1.6	1.8	25	1.2	1.7
6. SOUTH-WEST	24	1.8	1.8	18	1.1	1.3
7. THAMES	16	2.9	1.4	24	2.8	2.2
8. WELSH	20	2.0	2.2	21	1.3	1.4
9. WESSEX	14	5.2	4.7	24	1.2	1.7
10. YORKSHIRE	23	0.9	0.8	24	1.4	1.5
Overall	209	2.7	2.6	211	1.5	1.7

Table 3.8 Relationship between the under-estimation of the number of taxa in a sample and the NRA's estimate of the number of taxa for samples audited in 1992. Samples for all NRA region have been analyzed together, but grouped into classes of estimated number of taxa.

NRA estimate		Under-estimation of number of taxa			
		Mean	SD	Media	Maximu
1-5	4	1.0	1.2	1	2
6-10	21	1.4	1.6	1	5
11-15	32	1.2	2.1	1	7
16-20	52	1.2	1.4	1	5
21-25	62	1.9	2.0	1.5	8
26-30	27	1.5	1.4	1	4
31-38	13	1.5	1.8	1	5

The under-estimation of taxa (U_T) in each sample plotted against the NRA's estimated number of taxa (N_T) is demonstrated in Figure 3.1a. The right-hand plot is designed to show any general pattern in the 'average' size of the under-estimation. The dotted line is the result of using the novel smoothing method of Cleveland (1979) which was readily available within the MINITAB data analysis package (MINITAB 1994). The solid line gives the best fitting quadratic regression relationship:

$$U_T = 0.79 + 0.0612\ N_T - 0.00114\ (N_T)^2 \tag{3.1}$$

R&D Note 412

but the relationship is not statistically significant ($p>0.05$) (nor is a linear relationship).

Conclusion

Based on 1992 audit data, the OVERALL mean under-estimation of the number of taxa in single season samples, regardless of the number recorded, was 1.5.

R&D Note 412

Figure 3.1 Plot of the under-estimation of (a) number of taxa and (b) BMWP score against the value recorded by the NRA for 211 audited samples from 1992 (all NRA regions combined)

Left-hand-side : individual single season sample values;

Right hand side: two type of smoothed fitted lines for the average under-estimation, derived from a quadratic regression (solid line) and from the locally-weighted regression method of Cleveland (1979) (dashed line).

Observed underestimation — Smoothed estimates

(a) TAXA — NRA recorded TAXA

(b) SCORE — NRA recorded SCORE
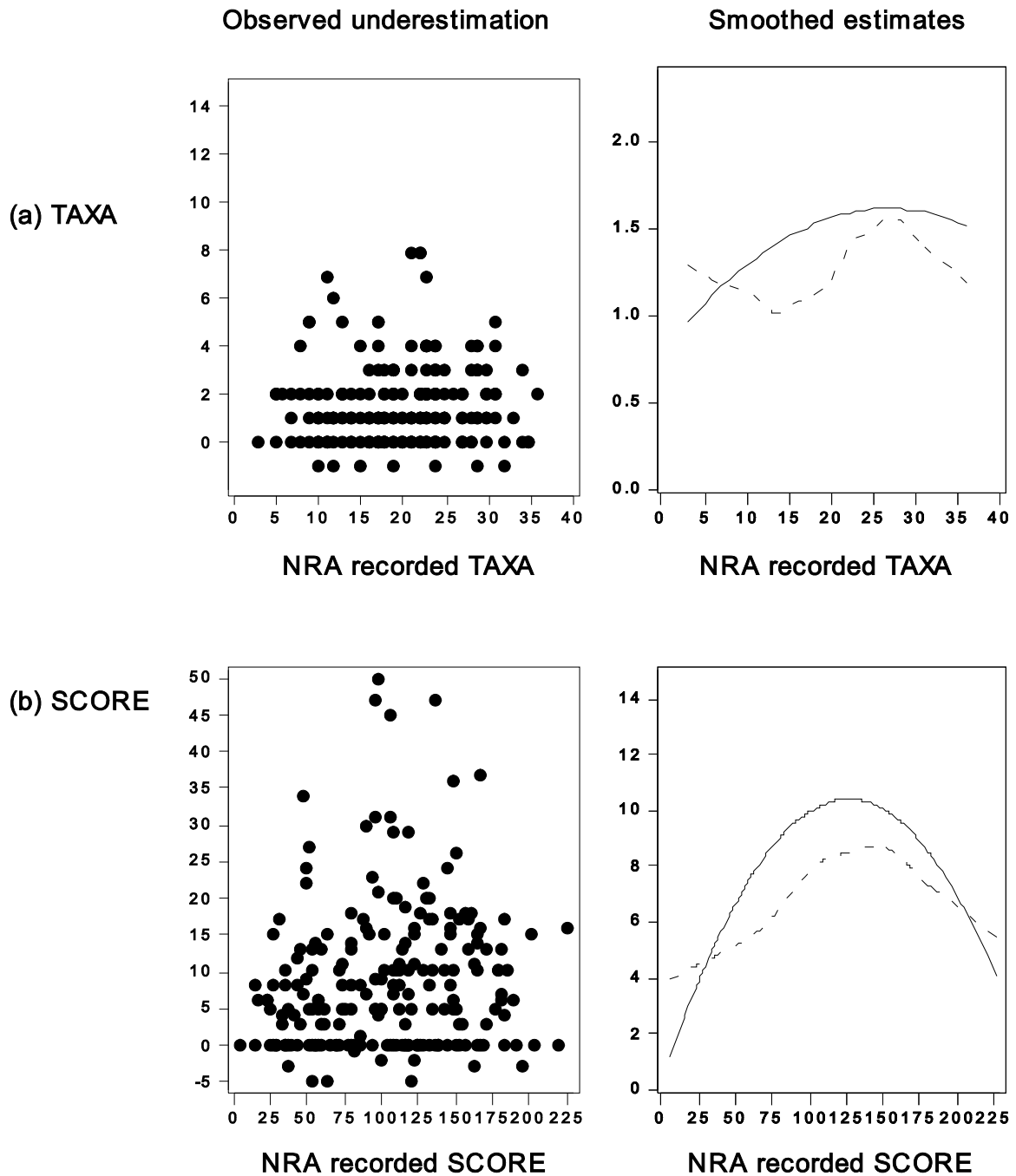
3.4.3 Procedure to estimate the under-estimation of the number of taxa in single season samples

In the preceding section it was concluded that it is reasonable to assume that the **average** under-estimation, and hence bias, of the number of taxa in single season samples is 1.5. An exception was samples with five or fewer taxa when the average should be assumed to be 1.0. However, this average does not indicate the variation in the under-estimation between individual samples.

The WRc quality control scheme (van Dijk 1994) is based on a statistical Poisson distribution for the number of missed taxa. The actual frequency distribution of under-estimation of missed taxa for all the 1992 audited samples analyzed, together with the number expected according to a Poisson distribution with a mean of 1.5 is shown in Table 3.9. Also shown is the distribution for a mean under-estimation of 2.0 which may be the target set for the 1995 River Quality Survey.

Table 3.9 Frequency (N) and percentage probability (%P) distribution of the degree of under-estimation of the number of taxa in single season's audited NRA samples from 1992; for the actual data and that predicted by a Poisson distribution with a mean of 1.5 or 2.0.

| | | | Poisson distribution with mean m | | | |
| | | | m = 1.5 | | m = 2.0 | |
Under-	N	%P	N	%P	N	%P
-1	8	3.79	(8)	(3.79)	(8)	(3.79)
0	59	27.96	45	21.47	28	13.02
1	57	27.01	68	32.20	55	26.04
2	43	20.38	51	24.15	55	26.04
3	18	8.53	25	12.10	37	17.36
4	14	6.64	10	4.53	18	8.68
5	6	2.84	3	1.36	7	3.47
6	2	0.95	1	0.34	2	1.16
7	2	0.95	0	0.08	1	0.33
8	2	0.95	0	0.01	0	0.08

In 1992 the NRA **over**-estimated the number of taxa present by one taxa in nearly 4% of the samples. This feature has been assumed to apply before fitting the Poisson distributions to the remaining samples.

A Poisson distribution with a mean of 2.0 seemed to grossly under-estimate the proportion of samples for which there is no error in the recorded number of taxa in the samples audited in 1992 (Chi-squared goodness of fit statistic = 50.80 with 7df, p<0.001).

A Poisson distribution with a mean of 1.5 is a better overall fit (Chi-squared = 26.95 with 5df), but this tends to under-estimate the small chance missing six or more taxa.

Conclusions

It is recommended that, when the NRA recorded more than five taxa as being present in a single season sample, the under-estimation of the number of taxa is assumed to follow a Poisson distribution with a mean of 1.5 taxa.

When the recorded number of taxa is five or less, it is recommended (from Table 3.8) that the number of taxa missed is assumed to follow a Poisson distribution with a mean of 1.0.

The only exception is that when no taxa are recorded as present, it is recommended that none are assumed to be present.

The NRA may choose to use a Poisson mean of 2.0 taxa missed per sample for all single season samples, as this is their target within their quality control procedures (van Dijk 1994).

3.5 Effect of sample processing errors on observed BMWP score

Analyses

Clarke et al (1994) showed that number of taxa and BMWP score are highly correlated, whether for single, two or three seasons combined samples or whether based on observed sample values or EQI index quality values. Therefore, it is likely that the sample processing under-estimation bias of both will be similarly correlated.

For the 211 audited samples from 1992, the overall correlation between under-estimation in number of taxa and under-estimation of score was 0.94. However, as observed BMWP score may still be used in some form to assess site quality, it is important to derive estimates of the size of the score under-estimation in various situations. It is believed that these are the first analyses which attempt to examine the effect of missing taxa on the under-estimation of the observed values of BMWP score.

R&D Note 412

Statistics on the average under-estimation of BMWP score for single season samples from each NRA region in each season of 1990 and 1992 are given in Table 3.9, while Table 3.10 gives the averages by region in each year.

As with number of taxa, the degree of under-estimation of BMWP score was considerably higher in 1990, especially for Wessex, but also Northumbrian, North-West and Severn-Trent (Table 3.10). In spring 1990, Wessex were often under-estimating single season BMWP score by as much as 50 (Table 3.9).

The average annual bias had been reduced in all these four regions by 1992. In 1990 the mean under-estimation of BMWP score, averaged across all NRA regions was 15. By 1992 this had been reduced to only nine.

The relationship between the degree of under-estimation of BMWP score (U_S) for a sample and the NRA's estimated value of the sample BMWP score (N_S) is set out in Table 3.11 and Figure 3.1(b).

The under-estimation is very variable, as one would might expect, with many samples having no processing error and a few even had scores over-estimated by up to five. At the other extreme, a few samples had scores under-estimated by as much as 50.

The greatest degree of under-estimation has some slight tendency to occur at sites where the NRA recorded intermediate values of BMWP score, as was found for under-estimation of number of taxa. This is supported by the best fitting quadratic regression relationship shown as the solid line in the right-hand side plot of figure 3.1(b), and given by :

$$U_S = 0.24 + 0.161\ N_S - 0.000636\ (N_S)^2 \tag{3.2}$$

This relationship is statistically significant ($p<0.02$), but only explains 4% of the variation in the degree of under-estimation.

Conclusions

Based on samples audited in 1992, the OVERALL mean under-estimation of the BMWP score in single season samples, regardless of the number of taxa recorded was nine.

The median under-estimation of score was only six, implying that in at least half the samples the NRA processing under-estimated the true sample BMWP score by no more than six.

Detailed procedures to estimate the true BMWP score and ASPT for a sample are developed in section 3.7.

R&D Note 412

Table 3.9 Mean, median and maximum under-estimation of BMWP scores in each of the (n) audited samples for each season and each NRA region in 1990 and 1992.

NRA REGION	SEASON	1990				1992			
		n	Mean	Med	Max	n	Mean	Med	Max
	SPRING	17	16	16	32	4	4	0	15
	SUMMER	10	17	21	29	8	14	13	45
	AUTUMN	5	15	13	26	12	11	5	50
	SPRING	8	28	30	59	8	5	4	10
	SUMMER	5	12	3	33	8	6	5	20
	AUTUMN	1	4	4	4	8	5	2	14
	SPRING	11	24	19	62	5	16	16	24
	SUMMER	4	21	17	49	1	−5	−5	−5
	AUTUMN	2	5	5	6	3	3	0	8
	SPRING	8	17	14	31	4	6	3	31
	SUMMER	5	22	30	41	7	13	13	30
	AUTUMN	11	23	22	58	7	10	5	47
	SPRING	14	12	15	38	9	4	5	20
	SUMMER	5	8	9	25	9	8	9	22
	AUTUMN	6	7	3	29	7	7	0	19
	SPRING	14	18	15	44	6	4	3	10
	SUMMER	9	6	7	23	6	9	10	21
	AUTUMN	1	0	0	0	6	4	4	10
	SPRING	5	9	6	21	9	12	13	23
	SUMMER	8	12	15	22	8	10	9	18
	AUTUMN	3	11	3	28	7	19	11	47
	SPRING	10	7	5	29	12	9	7	29
	SUMMER	10	15	18	34	9	7	5	18
	AUTUMN	0	*	*	*	0	*	*	*
	SPRING	7	43	50	76	8	10	5	37
	SUMMER	5	18	21	38	8	8	4	26
	AUTUMN	2	6	6	17	8	5	4	24
	SPRING	15	4	3	13	9	6	5	15
	SUMMER	6	8	7	17	9	15	10	36
	AUTUMN	2	13	13	15	6	5	2	17

R&D Note 412

Table 3.10 Mean and standard deviation (SD) of the under-estimation of BMWP scores in each of the (n) single season samples audited for each NRA region in 1990 and 1992.

	1990			1992		
	n	Mean	SD	n	Mean	SD
1. ANGLIAN	32	16	11	24	11	13
2. NORTHUMBRIAN	14	21	22	24	5	6
3. NORTH-WEST	17	21	20	9	9	10
4. SEVERN-TRENT	24	21	14	18	10	14
5. SOUTHERN	25	10	12	25	7	8
6. SOUTH-WEST	24	13	13	18	6	6
7. THAMES	16	11	9	24	13	11
8. WELSH	20	11	12	21	8	9
9. WESSEX	14	29	26	24	7	11
10. YORKSHIRE	23	5	6	24	9	10
Overall	209	15	15	211	9	10

Table 3.11 Relationship between the degree of under-estimation of BMWP scores of samples audited in 1992 and the NRA's estimate of scores for the same samples.

Samples from all NRA regions have been analyzed together, but grouped into classes of estimated BMWP score.

NRA estimate of BMWP score	Samples	Under-estimation of BMWP Score			
		Mean	SD	Median	Maximum
1-25	7	3.7	3.6	6	8
26-50	28	7.4	8.7	5	34
51-75	32	4.5	7.3	4	27
76-100	30	11.3	14.0	7.5	50
101-125	41	9.6	10.4	8	45
126-150	31	11.6	11.1	10	47
151-175	25	8.9	10.8	10	37
176-230	17	6.8	6.2	6	17

R&D Note 412

3.6 Effect of sample processing errors on observed ASPT

Analyses

Although a tendency for the NRA biologists to miss some taxa in a sample will lead to some under-estimation of the observed BMWP score, it may not lead to any general bias in the estimates of the observed ASPT.

The differences between the ASPT values for a sample based on the IFE audit taxa list and those derived from the taxa recorded as present in the sample by the NRA were calculated for each sample. Though consistently referred as the under-estimation of ASPT throughout this section, the difference could be positive (IFE>NRA) or negative (NRA>IFE) for any individual sample.

The average under-estimation of ASPT for each NRA region in each of 1990 and 1992 is given in Table 3.12. The tabulated data shows that the mean under-estimation of ASPT was positive for every region in 1990 and positive for eight of the 10 regions in 1992.

The median under-estimation was also positive (or zero) for every region in 1990 when the number of taxa missed was higher for most regions (see table 3.7). In 1992, when the number of missed taxa was generally lower, then the median error in the ASPT values due to NRA sample processing was zero in six of the 10 regions and only more than 0.02 in Thames region where it has already been established that there was lapse in the accuracy of sample processing in the autumn of 1992 (Table 3.6). The variation between samples in the estimation bias for ASPT, as represented by the standard deviation (SD), was also less for every region in 1992.

Table 3.12 Average (mean), standard deviation (SD) and median (Med) of the under-estimation of ASPT in each of the (n) single season audited samples of each NRA region in 1990 and 1992.

	1990				1992			
	n	Mean	SD	Med	n	Mean	SD	Med
1. ANGLIAN	32	0.05	0.22	0.07	24	0.08	0.16	0.01
2. NORTHUMBRIAN	14	0.03	0.34	0.05	24	0.07	0.20	0.00
3. NORTH-WEST	17	0.07	0.36	0.02	9	-0.07	0.22	0.00
4. SEVERN-TRENT	24	0.27	0.39	0.24	18	0.04	0.23	0.02
5. SOUTHERN	25	0.06	0.15	0.05	25	0.05	0.13	0.00
6. SOUTH-WEST	24	0.08	0.15	0.08	18	-0.04	0.15	0.00
7. THAMES	16	0.09	0.21	0.03	24	0.06	0.14	0.04
8. WELSH	20	0.08	0.26	0.01	21	0.02	0.11	0.00
9. WESSEX	14	0.05	0.37	0.04	24	0.04	0.13	0.01
10. YORKSHIRE	23	0.03	0.15	0.00	24	0.06	0.15	0.00
Overall	209	0.08	0.27	0.06	211	0.04	0.16	0.00

R&D Note 412

Tables 3.13 and 3.14 provide information on whether the under-estimations of sample ASPT's are related to the NRA's estimated values of either ASPT or number of taxa for the same samples. There appears to be no obvious relationship with either, except that when the NRA recorded taxa list gives a value of observed ASPT of over 7.0, it is usually a slight over-estimate, with the true sample value being on average 0.17 less.

Table 3.13 Relationship between the degree of under-estimation of ASPT's for samples audited in 1992 and the NRA's estimate of the ASPT's of the same samples.

Samples for all NRA regions have been analyzed together, but grouped into classes of estimated ASPT.

NRA estimate of ASPT	Samples	Under-estimation of ASPT				
		Mean	SD	Median	Min.	Max.
≤ 3.0	4	0.07	0.17	0.00	0.00	0.29
3.01 - 4.00	40	0.07	0.14	0.00	-0.10	0.58
4.01 - 5.00	55	0.07	0.19	0.00	-0.51	0.48
5.01 - 6.00	59	0.03	0.15	0.00	-0.37	0.46
6.01 - 7.00	48	0.02	0.15	0.00	-0.58	0.33
> 7.00	5	-0.172	0.19	-0.10	-0.59	0.00

Table 3.14 Relationship between the degree of under-estimation of ASPT's for samples audited in 1992 and the NRA's estimate of the number of taxa present in the same samples.

Samples for all NRA regions have been analyzed together, but grouped into classes of estimated number of taxa.

NRA estimate of number of taxa in sample	Samples	Under-estimation of ASPT				
		Mean	SD	Median	Min.	Max.
1-5	4	-0.05	0.33	0.00	-0.51	0.29
6-10	21	0.05	0.16	0.00	-0.23	0.58
11-15	32	0.00	0.14	0.06	-0.21	0.43
16-20	52	0.04	0.18	0.00	-0.37	0.30
21-25	62	0.06	0.13	0.00	-0.58	0.48
26-30	27	0.07	0.11	0.05	-0.12	0.46
31-38	13	0.04	0.11	0.01	-0.16	0.28

Combining across all regions for 1992, the missing of taxa in the sample processing appears to

lead, on average, to under-estimating ASPT by 0.00-0.04, but the actual effect varies considerably between samples with a standard deviation of 0.16 (Figure 3.2).

Figure 3.2 Histogram showing the distribution of values for the degree of under-estimation of ASPT in 211 audited samples from 1992 (all NRA regions combined).

Under-estimation for a sample equals the IFE audited sample value minus the NRA recorded value for the same sample.
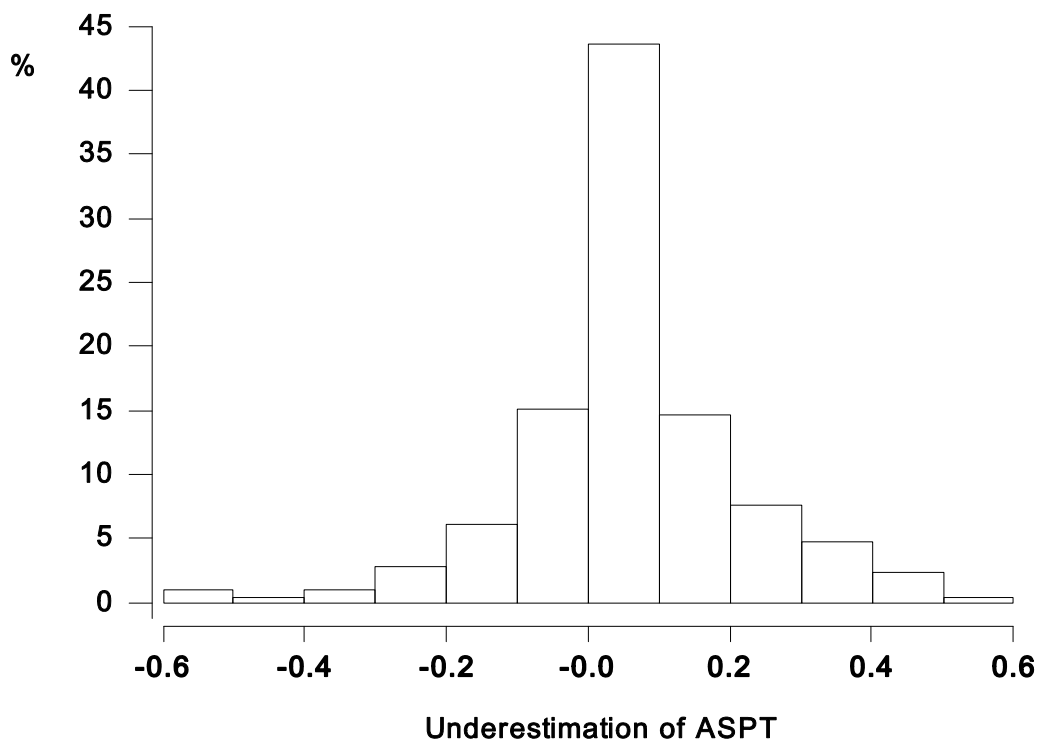
Conclusion

The effect of missing taxa during sample processing is to lead to a small but generally consistent under-estimation of the true sample ASPT and quality estimates derived from it.

Detailed procedures to estimate the under-estimation and hence the true BMWP score and ASPT for a sample are developed in section 3.7

R&D Note 412

3.7 Procedures to correct for the bias, due to sample processing errors, in the estimated number of taxa, BMWP score and ASPT for a single season sample

A very complicated way to correct for bias would be to take the site-specific RIVPACS expected probabilities of each taxa occurring and select the missed taxa using these probabilities. However, this would only be appropriate for high quality sites. For poor quality sites, the taxa missed are much more likely to be low BMWP scoring taxa rather than simply the taxa which were most expected to be present at the site (if it was unstressed). Therefore a simpler practical solution is suggested.

3.7.1 Estimation of bias for a particular sample

Analyses

In the 1992 samples the under-estimation (U_S) of BMWP score by the NRA was, on average, about nine. The corresponding estimate for the under-estimation (U_T) of the number of taxa was, on average, 1.5 (see section 3.4), implying that the overall average BMWP score of missed taxa is about six. However, if the ASPT value of the missed taxa, namely $ASPT_{miss} = U_S/U_T$, is plotted against the number of taxa (N_T) recorded as being present by the NRA, then the ASPT of the missed taxa tends to be less when few taxa are recorded (Figure 3.3).

The relationship is adequately described by the best fit linear regression (standard errors of coefficients in brackets):

$$ASPT_{miss} = \quad 4.29 \quad + \quad 0.077 \; N_T \tag{3.3}$$
$$(0.50) \qquad (0.024)$$

From equation 3.3, the mean ASPT of the missed taxa in a sample is estimated to range from around 4.5 when about five taxa are recorded as present to over 6.5 when over 30 taxa are recorded.

Figure 3.3 The ASPT of the missed taxa (ie under-estimation of BMWP score divided by under-estimation of number of taxa) plotted against the number of taxa recorded as present by the NRA for 1992 samples with missed taxa (n=154). Best fit regression line

ASPT of missed taxa

NRA recorded TAXA

R&D Note ...
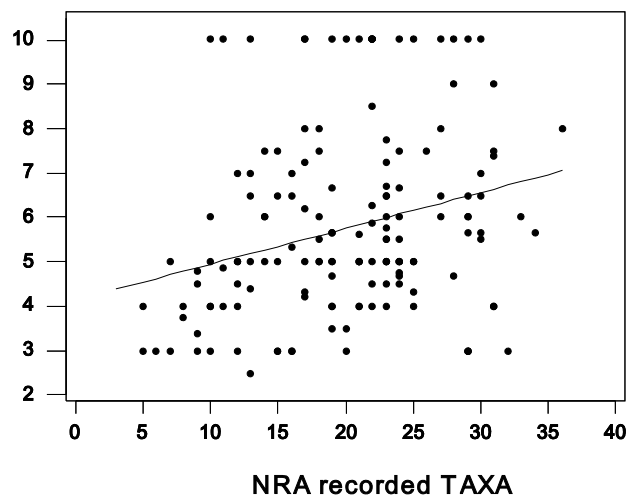
This relationship provides a simple method of estimating the under-estimation of the observed BMWP score and ASPT for a sample processed by the NRA. The number of missed taxa can be estimated by the methods recommended in section 3.4.3 based on the recorded number of taxa. The ASPT of these missed taxa is then estimated from equation 3.3 to be $ASPT_{miss}$.

Let N_T and N_S denote the NRA recorded values of observed number of taxa and BMWP score for a sample. Similarly let U_T denote the estimate of the under-estimation of the number of taxa.

The under-estimation (U_S) of the BMWP score for the sample is then estimated by :

$$U_S = U_T \cdot ASPT_{miss} \tag{3.4}$$

Then the true number of taxa (I_T), true BMWP score (I_S) and true ASPT (I_A) for the sample, are estimated as :

$$I_T = N_T + U_T \tag{3.5}$$

$$I_S = N_S + U_S \tag{3.6}$$

$$\text{and } I_A = I_S / I_T \tag{3.7}$$

For an example, consider a sample where the NRA record nine taxa as being present with a total BMWP score of 36 and hence an estimated ASPT of 4.00. Suppose the number of taxa missed from this sample is estimated to be two (from section 3.4.3). The ASPT of the two missed taxa is estimated from equation 3.3 to be 4.29 + 9 x 0.077 = 4.98. The under-estimation of total BMWP score is then estimated to be 10 (nearest integer to 2 x 4.98), such that the correct BMWP score for the sample is estimated to be 36 + 10 = 46, with an ASPT of 46/11 = 4.18. The under-estimation of observed ASPT is therefore estimated to be 4.18 - 4.00 = 0.18.

3.7.2 Variance due to bias correction of number of taxa, BMWP score and ASPT

Section 3.7.1 includes a method of obtaining a 'best' estimate of the bias in the NRA's recorded value of number of taxa, BMWP score and ASPT. Correction for this bias will add an extra error variance to the estimate of the observed value. This can most easily be incorporated into the overall variance and errors in the observed value and hence the EQI estimate for site by computer simulation, as will be explained in detail in section 5.1.

At this stage it is only necessary to record from Table 3.15 that, when only one taxon is missed, the standard deviation (SD) of the BMWP value of the missed taxa, ranges from 1.4 when few taxa were recorded as present to about 2.5 when many taxa were present.

R&D Note 412

Table 3.15 Mean and standard deviation (SD) of the ASPT of the taxa missed by the NRA in single season audited samples in 1992.

Taxa recorded as present	Mean ASPT of M missed taxa (n samples)			SD of ASPT of M missed taxa		
	M=1	M=2	M>2	M=1	M=2	M>2
1-10	4.00 (2)	3.58 (6)	3.98 (3)	1.41	0.66	0.73
11-20	6.04 (26)	5.36 (14)	5.52 (14)	2.39	1.65	1.09
21-30	6.74 (23)	6.19 (18)	5.80 (23)	2.78	2.04	1.18
31-40	6.75 (4)	7.30 (5)	5.68 (4)	2.50	1.35	1.39

Conclusions (Section 3.7)

For simplicity, it is recommended that in the error simulations, the ASPT of the U_T missed taxa is treated as a random deviate from a normal distribution with mean $ASPT_{miss}$ (equation 3.3) and SD equal to $2.0/\sqrt{U_T}$.

The simulated value of ASPT for the missed taxa should be constrained within the true limits (1-10) and taken to the nearest integer if for one missed taxa or the nearest 0.5 if for two missed taxa, etc. This will still give some impossible values, in that there are no nine-scoring taxa, etc, but it will be adequate.

The simulated value of the under-estimation of the BMWP score for the sample is then estimated by the product of the simulated values of the under-estimation of ASPT and the number of taxa.

3.8 Effect of sample processing errors on combined season index values

3.8.1 Analysis of sites audited in two seasons of one year

The effect of sample processing errors on BMWP index values for two season combined samples for the 12 sites which were audited in two seasons of any one year are summarised in Table 3.16. Though based on only a few sites, the following observations are merited:

(i)In two of the four cases where the NRA incorrectly recorded a taxa as present in a sample, IFE audit did not find the taxa in either sample from the site. Therefore, as likely as not, taxa incorrectly recorded as present at a site may be taxa which tend not to actually occur at the site. Hence they are not present in other seasons samples from the site and tend to remain as minor errors even in combined season samples .

(ii)IFE recorded 482 taxa occurrences in the 24 single season samples, an average of 20 taxa per single season sample. Of these there were a total of 47 occurrences of taxa missed by the NRA or 10% of those present. The overall rate of missing taxa in single season samples from 1990, as estimated from all 209 audited samples was actually 15.3% (see section 3.3.2 and Table 3.4).

(iii)Of the 47 cases of taxa missed by the NRA in one sample, 30 were present in the other season's sample, of which four were missed again by the NRA. This suggests that an estimated 13% (4/30) of the taxa missed in one season, are also missed, when present, in a second season's sample.

This percentage is about the same, and more importantly, not significantly greater than the overall rate of missing taxa in single season samples in 1990 (15.3%). This suggests that it is reasonable to assume that the overall probability of missing taxa, when present in a second season's sample, is independent of whether they were missed in the first season's sample.

(iv)Of the 47 cases of taxa missed by the NRA in one season's sample, 21 were not recorded as present by the NRA in the second season. Therefore the overall proportion of taxa which were missed in one season and not recorded as present in a second season was 45%. This means that about half the errors arising from missing taxa in single season samples have no effect on the index values for the two seasons combined samples.

(v)For two season combined samples, the under-estimation varied as follows:

	Mean	Min	Max
Taxa	1.25	-1	6
BMWP Score	5.67	-10	35
ASPT	-0.03	-0.26	0.21

(vi)In the 12 two seasons combined samples, the average number of taxa present (IFE   value) was 22.3, for which the mean under-estimation of the number of taxa was 1.25 This is equivalent to a 6% rate of missing taxa, compared to a 10% rate for the 24 single season samples.

R&D Note 412

Table 3.16 Summary of the NRA recorded value, the IFE audited value and the difference for each of number of taxa (T), BMWP score (S) and ASPT (A) for single and two season combined samples for the 12 sites for which samples were audited in two seasons in either 1990 or 1992.

Gain = number of taxa not recorded by NRA but found in sample by IFE.
Loss = number of taxa recorded as present by NRA but not found by IFE.

NRA Region (Site Code)	Sampling date	Season	Gain	Loss	-NRA value- T	S	A	-IFE value- T	S	A	Difference T	S	A
Northumbrian	18/04/90	Spring	5	0	23	130	5.65	28	164	5.86	5	34	.20
(1890)	01/08/90	Summer	5	0	20	111	5.55	25	144	5.76	5	33	.21
		Combined	2	0	27	161	5.96	29	167	5.76	2	6	-.20
Northumbrian	21/03/90	Spring	0	1	16	77	4.81	15	74	4.93	-1	-3	.12
(3400)	09/10/90	Autumn	2	0	8	38	4.75	10	42	4.20	2	4	-.55
		Combined	1	0	19	90	4.74	20	93	4.65	1	3	-.09
Southern	03/04/90	Spring	2	0	20	136	6.80	22	153	6.95	2	17	.15
(1016)	02/11/90	Autumn	1	0	26	162	6.23	27	167	6.19	1	5	-.05
		Combined	0	0	31	201	6.48	31	201	6.48	0	0	.00
Southern	14/03/90	Spring	3	0	17	77	4.53	20	93	4.65	3	16	.12
(3021)	23/08/90	Summer	5	0	18	85	4.72	23	110	4.78	5	25	.06
		Combined	6	0	23	111	4.83	29	146	5.03	6	35	.21
Thames	05/03/90	Spring	4	0	21	109	5.19	25	130	5.20	4	21	.01
(0109)	14/09/90	Autumn	4	0	25	119	4.76	29	147	5.07	4	28	.31
		Combined	2	0	30	157	5.23	32	169	5.28	2	12	.05
Yorkshire	26/03/90	Spring	0	0	10	37	3.70	10	37	3.70	0	0	.00
(0308)	16/08/90	Summer	0	0	9	30	3.33	9	30	3.33	0	0	.00
		Combined	0	0	11	40	3.64	11	40	3.64	0	0	.00
Anglian	09/06/92	Summer	2	0	7	23	3.29	9	29	3.22	2	6	-.06
(0306)	28/10/92	Autumn	2	1	7	26	3.71	8	31	3.88	1	5	.16
		Combined	1	0	10	40	4.00	11	43	3.91	1	3	-.09
Northumbrian	24/06/92	Summer	2	0	5	24	4.80	7	30	4.29	2	6	-.51
(4100)	20/10/92	Autumn	0	0	11	38	3.45	11	38	3.45	0	0	.00
		Combined	0	0	13	53	4.08	13	53	4.08	0	0	.00
South-West	25/03/92	Spring	0	1	32	196	6.13	31	193	6.23	-1	-3	.10
(2706)	19/06/92	Summer	5	0	17	99	5.82	22	120	5.45	5	21	-.37
		Combined	1	1	34	213	6.26	34	214	6.29	0	1	.03
South-West	29/06/92	Summer	1	0	21	149	7.10	22	154	7.00	1	5	-.10
(2932)	02/10/92	Autumn	1	0	17	113	6.65	18	123	6.83	1	10	.19

R&D Note 412

```
                Combined   1    0   22 154 7.00  23 159 6.91   1    5 -.09

Wessex     01/06/92 Summer   0    0   16  70 4.38  16  70 4.38   0    0  .00
(T137)     12/10/92 Autumn   0    1   19  87 4.58  18  77 4.28  -1  -10 -.30
                Combined   0    1   21 102 4.86  20  92 4.60  -1  -10 -.26

Wessex     09/06/92 Summer   2    0    8  27 3.38  10  35 3.50   2    8  .13
(T233)     22/10/92 Autumn   1    0   10  37 3.70  11  42 3.82   1    5  .12
                Combined   3    0   12  45 3.75  15  58 3.87   3   13  .12
```

R&D Note 412

3.8.2 Proportion of taxa missed in one season that are found in a second or third season's sample

Analyses

The aim of the following analyses was to estimate of the proportion of taxa which are present but missed by the NRA in any one season's sample that are then recorded by the NRA as present in another season's sample from the same site. This was best achieved by comparing the taxa lists for the 209 audited sample sites from 1990 with their taxa lists for each single season as used recorded by the NRA and used in making RIVPACS predictions and site quality assessments for the 1990 River Quality Survey (National Rivers Authority 1994).

For each audited sample, the missed taxa were listed, together with whether they were found by the NRA in one or both of the other two seasons samples for the same site in 1990. Summing over all audited samples and sites, led to estimates of the percentage of times each taxa when missed in one sample by the NRA would be found in a second season's sample (P2%) or in at least one of the two other seasons' samples (P3%) (Table 3.17).

The percentages P2% and P3% therefore estimate respectively the percentage of times that missing a particular taxa present in one seasons sample has no effect on the two and three seasons combined taxa list and hence observed BMWP index values.

For example, *Planariidae* occurred in 82 of the 209 audited samples. Of the 14 sites from which it was missed in one single season sample, it was recorded as present in 13 of the 28 samples from the other two seasons (ie 46%) for these sites. *Planariidae* were found in the three seasons combined sample for 10 of these 14 sites (ie 71%).

If P2% and P3% were 100% then obviously, there would be no effect of missing taxa in only one season on the BMWP index values for combined season samples. The following types of missed taxa are least likely to 'recovered' in combined season samples:

  (i) Taxa of low local abundance which are therefore unlikely to be captured in more than one season at the site.

  (ii) Taxa which, by virtue of their life cycle, are most seasonal in their availability for capture in pond-net samples and hence not likely to be caught in all seasons.

  (iii) Taxa which the NRA have most trouble in identifying within a sample and hence tend to miss in any sample

Hydrophilidae are a commonly missed taxa, of type (iii) above, which seems to have low probability of being recovered in a two season combined (P2=33%) or three season combined sample (P3=47%).

Table 3.17 Estimation of the percentage of times that each taxa is missed by the NRA in one season's sample but found and recorded in other seasons samples from the same site and year.

P2% and P3% respectively denote the percentage of times taxa missed by the NRA in one season would be found in a two and three seasons combined sample, based on 209 audited samples from 1990.

Family	Times Present	Times Missed	P2%	P3%
Planariidae (incl. Dugesiidae)	82	14	46	71
Dendrocoelidae	16	8	25	37
Neritidae	12	2	25	50
Viviparidae	3	1	100	100
Valvatidae	43	18	55	77
Hydrobiidae (incl. Bithyniidae)	160	22	63	81
Lymnaeidae	116	21	73	90
Physidae	48	13	46	61
Planorbidae	78	24	52	75
Ancylidae (incl. Acroloxidae)	110	17	41	52
Unionidae	13	1	50	100
Sphaeriidae	150	33	56	69
Oligochaeta	203	14	89	100
Piscicolidae	32	3	50	66
Glossiphoniidae	136	14	46	57
Erpobdellidae	111	7	50	85
Asellidae	125	6	66	83
Corophiidae	4	1	100	100
Gammaridae (incl. Crangonyctidae & Niphargidae)	170	12	66	91
Astacidae	3	0	0	0
Baetidae	161	14	32	50
Heptageniidae	72	1	100	100
Leptophlebiidae	54	14	50	85
Ephemerellidae	79	14	39	71
Potamanthidae	1	1	0	0
Ephemeridae	49	2	50	100
Caenidae	88	20	27	50
Taeniopterygidae	26	4	12	25
Nemouridae	54	12	33	58
Leuctridae	60	7	35	57
Capniidae	1	1	0	0
Perlodidae	46	2	0	0
Perlidae	12	2	50	50
Chloroperlidae	32	7	14	28
Coenagriidae	27	2	25	50
Calopterygidae	25	4	50	100
Cordulegasteridae	6	1	0	0
Aeshnidae	1	0	0	0
Mesovelidae	1	0	0	0
Hydrometridae	2	1	0	0
Gerridae	9	2	25	50
Nepidae	2	0	0	0
Naucoridae	3	1	0	0

R&D Note 412

Aphelocheiridae	4	0	0	0
Notonectidae	14	2	50	100
Corixidae	49	2	0	0
Haliplidae	68	13	38	61
Dytiscidae (incl. Noteridae)	124	15	40	53
Gyrinidae	45	8	43	75
Hydrophilidae (incl. Hydraenidae)	82	34	33	47

Table 3.17 (continued)

Family	Times Present	Times Missed	P2%	P3%
Scirtidae (=Helodidae)	6	3	0	0
Dryopidae	2	1	0	0
Elmidae	151	26	42	61
Curculionidae	1	0	0	0
Sialidae	43	4	75	75
Rhyacophilidae (incl. Glossosomatidae)	82	11	54	72
Philopotamidae	6	1	0	0
Polycentropodidae	45	3	33	33
Psychomyiidae (incl. Ecnomidae)	23	9	16	22
Hydropsychidae	115	6	58	66
Hydroptilidae	49	24	20	33
Phryganeidae	3	0	0	0
Limnephilidae	74	9	50	66
Molannidae	8	2	25	50
Beraeidae	4	2	0	0
Odontoceridae	7	1	0	0
Leptoceridae	75	18	50	72
Goeridae	30	11	54	72
Lepidostomatidae	49	15	43	73
Brachycentridae	16	5	50	60
Sericostomatidae	62	7	57	85
Tipulidae	110	16	40	50
Simuliidae	107	20	57	80
Chironomidae	204	3	83	100
TOTAL	**4084**	**614**	**45**	**63**

Overall an estimated 45% of the taxa missed by the NRA in single season samples were found and recorded as present in a second season's sample ('Total' line in Table 3.17). Hence in the two seasons combined sample, only 55% of taxa missed in one of the two seasons would contribute to any errors or bias in estimating the observed number of taxa, BMWP score or ASPT value for the two season combined sample.

An estimated 63% of taxa missed in any single season sample would be found or recorded in at least one of the other two seasons samples for the same site (Total P3% in Table 3.17). Therefore, only 37% of taxa missed in individual season samples contribute to the bias and under-estimation of observed BMWP index values for three seasons combined samples.

The observed effects of missing taxa on the under-estimation of (a) observed number of taxa, (b) BMWP score and (c) the bias in ASPT, for two and three season combined samples are highlighted in Table 3.18. In this analysis the NRA's taxa list for the second and third seasons were treated as correct. Ideally all three seasons samples for numerous sites would have been audited. This limitation tends to under-estimate the true effect of missed taxa on combined season sample errors in index values because it does not include taxa missed in the two non-audited seasons' samples which would increase the true observed index values for the combined season samples. However,

the table is useful to give a first-order estimate of the scale of the effects.

It is important to remember that all the estimates in this section, 3.8.2, had to be based on NRA data from 1990. The estimates of the chances of finding missed taxa in other season samples depend to some extent on the general level with which taxa are missed in any single season sample. The more taxa missed per sample, the lower the chances of missed taxa being captured in another season In section 3.4 and table 3.7 it has already been shown that the NRA missed more 15.3% of taxa in 1990 but only 8.3% in 1992.

Table 3.18 Effect of NRA sample processing errors in one season's samples on the under-estimation of observed BMWP index values in one, two and three seasons combined samples. The table shows the percentage of 209 sites audited in one season of 1990 within various degrees of under-estimation.

Note : This analysis under-estimates the true bias for combined season samples because not all seasons' samples at each site were audited.

(a) number of taxa

Under-estimation of taxa	Single season samples	Combined season samples	
		2 seasons	3 seasons
< 0 (ie over-estimated)	6.7%	6.7%	6.7%
≤ 0	22.0%	34.0%	43.5%
≤ 1	40.7%	58.9%	78.0%
≤2	55.5%	78.5%	88.5%
≤ 3	69.4%	85.7%	93.3%
≤ 4	79.4%	94.3%	96.2%
≤ 5	88.0%	97.1%	100%
≤ 7	94.7%	100%	
≤ 10	99.5%		
Max error	15	7	5

(b) BMWP score

Under-estimation of BMWP score	Single season samples	Combined season samples	
		2 seasons	3 seasons
< 0 (ie over-estimated)	9.1%	8.1%	8.1%
≤ 0	21.5%	32.1%	40.7%
≤ 5	34.5%	47.4%	61.2%
≤10	47.9%	67.5%	82.8%
≤ 15	58.9%	78.0%	88.0%
≤ 20	72.7%	87.6%	92.3%

R&D Note 412

≤ 25	80.9%	91.4%	97.6%
≤ 30	87.1%	96.2%	99.0%
≤ 40	93.8%	99.5%	100%
Max error	59	41	35

Table 3.18 (continued)

(c)ASPT

Bias in ASPT	Single season samples	Combined season samples	
		2 seasons	3 seasons
none	12.9%	24.4%	34.9%
≤ 0.05	27.3%	41.6%	56.5%
≤ 0.10	41.2%	62.2%	72.7%
≤ 0.15	53.6%	77.5%	85.2%
≤ 0.20	64.1%	84.7%	90.0%
≤ 0.30	82.3%	92.3%	94.7%
≤ 0.40	87.1%	95.1%	96.7%
≤ 0.50	91.9%	96.2%	99.0%
≤ 0.60	96.2%	97.3%	99.5%
≤ 0.80	97.6%	99.0%	99.5%
≤ 1.00	98.6%	99.5%	100%
Max error	+ 1.29	+ 1.18	+ 0.85

Assuming that the NRA have improved their sample processing and quality control procedures since 1990, the rates quoted above of finding missed taxa in other seasons will almost certainly be under-estimates of the 'recovery' rate for combined season samples since 1990 and, in particular, for the NRA's 1995 survey. **Therefore the above estimates suggest an upper limit to the size of the effects of sample processing errors and missed taxa.**

The probability of a taxon missed from one sample in 1992 being found in a second season's sample from the same site can be estimated as follows :

Probability of taxa missed in any one season in 1990 = 0.153

Proportion of missed taxa in one season being found in a second in 1990 = 0.45

If Q = overall proportion of taxa occurring in both seasons of a two season combined sample, then:

0.45 = (1 - 0.153) Q , and hence Q = 0.53.

The proportion of taxa missed in any one season in 1992 = 0.083. Therefore the probability of taxa missed in one season's sample in 1992 being recorded as present in a second season's sample is:
(1 - 0.083) Q = (1-0.083)0.53 = 0.49

R&D Note 412

Conclusions

It is recommended that, for 1992 and subsequent years, only 51% of the taxa estimated, or simulated to be missed from any single season's sample are assumed not have been recorded as present in a second season's sample and hence to still influence the combined season sample observed values of number of taxa, BMWP score and ASPT.

If the 1990 and 1995 survey results for a site are to be compared by simulating bias-corrected samples for each year , then the corresponding figure for two season samples from 1990 is 55%).

It is not feasible to extend this logic to three seasons combined samples. The simplest option is to assume that the proportion of taxa missed in one season that are found in at least one of the other two seasons samples is the same in 1992 and subsequent years as was estimated from the 1990 data (ie as 100-67% = 37%, Table 3.17).

It is therefore recommended that, for any year, the percentage of taxa missed in any single season sample which are not recorded as present in the three season combined sample for the same site is assumed to be 37%.

3.8.3 Procedure to correct for bias in combined season sample BMWP index values

The procedures detailed in section 3.7 can be extended to combined season samples as follows :

If the expected under-estimation of missed taxa in two single season samples is M_1 and M_2, where both M_1 and M_2 are usually recommended to be 1.5 (section 3.4.3), then the expected under-estimation of taxa (U_{T2}) in the corresponding two season combined sample is estimated to be:

$$U_{T2} = 0.51(M_1 + M_2) \qquad (3.8)$$

The equivalent under-estimation of number of taxa (U_{T3}) for the three seasons combined sample is:
$$U_{T3} = 0.37(M_1 + M_2 + M_3) \qquad (3.9)$$

where obviously M_3 is the expected under-estimation for the third season's sample.

The under-estimation of BMWP score is best estimated by using equations (3.3) and (3.4) to calculate the ASPT of the missed taxa and the subsequent under-estimation of BMWP score for each season separately. If S_1, S_2, S_3 are the expected or simulated values for the under-estimation of BMWP score in each of the three single seasons, then the under-estimation of BMWP score in two and three seasons combined is estimated to be, respectively:

$$U_{S2} = 0.51 (S_1 + S_2) \qquad (3.10)$$

and $U_{S3} = 0.37 (S_1 + S_2 + S_3)$ $\qquad$ (3.11)

If N_{T2}, N_{T3}, N_{S2}, and N_{S3} are the NRA recorded values of number of taxa and BMWP score for two and three season combined samples, then the true values are estimated by :

$$I_{T2} = N_{T2} + U_{T2} \;\; ; \quad I_{T3} = N_{T3} + U_{T3} \;\; ; \quad I_{S2} = N_{S2} + U_{S2} \;\; ; \; \text{and} \quad I_{S3} = N_{S3} + U_{S3} \tag{3.12}$$

The true values of ASPT for two and three season combined samples are estimated as in equation (3.7) by :

$$I_{A2} = I_{S2} / I_{N2} \quad \text{and} \quad I_{A3} = I_{S3} / I_{N3} \tag{3.13}$$

If the individual single season samples are no longer available, assume the under-estimation of taxa in a two and three season combined sample is, a Poisson variate with a mean of :

0.51 (2 x 1.5) = 1.53   for two seasons combined

and 0.37 (3 x 1.5) = 1.67   for three seasons combined.

Whatever the expected degree of under-estimation of number of taxa, it can be used as the mean of a Poisson distribution to simulate the under-estimation and hence true observed value of number of taxa in combined season samples, as part of simulations to estimate overall variation and errors in observed BMWP index values and hence EQI values the detection of change in site quality over time.

The average ASPT of the missed taxa is best assumed to be 6.0, from a normal distribution with standard deviation equal to $2.0/ \sqrt{U_{T2}}$ or $2.0/ \sqrt{U_{T3}}$, as appropriate, as in section 3.7.2. This assumption can also be used in simulations to estimate the total errors in observed index values.

3.9 Summary

1. The aim of this section was to assess the variation in observed number of taxa, BMWP score and ASPT due to biological sample sorting and processing errors and taxonomic data recording errors.

2. A representative cross-section of the IFE audit recording sheets for samples submitted to IFE for audit were re-assessed. Just over 200 audited samples were selected from both 1990 (assumed to represent the least accurate year) and 1992 (assumed to represent the improved current quality of sample processing), each covering all NRA regions, seasons and ranges of taxonomic richness. Taxa missed and incorrectly recorded as present were assessed together with net changes in number of taxa, BMWP score and ASPT.

3. There is about a 1% error rate in transcribing taxonomic information from paper sheets onto computer file. Double typing of such information would eliminate most such errors, but this may not be cost-effective.

R&D Note 412

4.In 1990 many NRA regions under-estimated the number of taxa present in a sample by, on average, 3-4 taxa. By 1992, with the exception of one region in one season, the average under-estimation was reduced to 1.5 taxa per sample, which is less than the probable NRA quality control target of two taxa.

5.Other than for very taxon-poor sites, there was no overall tendency for the number of taxa missed, to increase with the number of taxa recorded as present.

6.It is recommended that the under-estimation of the number of taxa is assumed to be, on average, 1.5 taxa per sample when the NRA recorded more than five taxa as being present. When the recorded number of taxa is five or less, it is recommended (from Table 3.8) that the average number of taxa missed is assumed to be 1.0. The only exception is that when no taxa are recorded as present, it is recommended that none are assumed to be present.

(The NRA may choose to use an estimate of an average of 2.0 taxa missed per sample for all single season samples, as this is their target within their quality control procedures (van Dijk 1994)).

7.Details are given of the taxa which are most often by the NRA in general and by particular regions (section 3.3). NRA biologists should be made aware of the problem taxa and, where necessary, further training should be given in their removal and identification. has not already been taken.

8.Detailed procedures are suggested to correct for the bias in estimating the observed values of number of taxa, BMWP score and ASPT for single season samples (section 3.7.2)

9.For 1992 and subsequent years, it is estimated that of the taxa missed in single season samples, only 51% and 37% respectively are not subsequently found in the two and three season combined sample. It is recommended that these estimates be used to correct for bias in combined season sample observed values of BMWP index values.

10.The average under-estimation of the number of taxa by the NRA in 1992 was about 1.5 taxa for single season samples and also for two or three season combined samples. The bias in estimated EQI due to NRA sample processing errors is therefore less using two and three season combined sample because they have higher expected numbers of taxa.

11.Procedures are given for estimating and correcting for the NRA bias in the observed number of taxa, BMWP score and ASPT for two and three season combined samples (section 3.8.3). These include details of how to simulate the bias-corrected observed index values as part of simulations to estimate the precision of EQI values for a site.

R&D Note 412

4 VARIATION IN RECORDING OF ENVIRONMENTAL DATA

4.1 Introduction

Errors and variation in the acquisition and processing of macro-invertebrate samples affect the observed values used in assessing Environmental Quality Indices.

Those indices also require expected values for their computation. These are derived from RIVPACS II (Wright et al 1993) and are generated by entering site-specific field and map-measured values of selected environmental variables into the predictive model (Cox et al 1991).

Variation in the measurement and recording of these environmental variables will lead to subsequent variation in the expected (RIVPACS predicted) BMWP index values for the site and hence the values of its EQI's.

In an earlier report to the NRA (Clarke et al 1994), procedures for standardising the site-specific expected BMWP index values were considered. These involved repeated annual environmental data measurements until the standard error of the mean values of individual errors fell within pre-defined limits. These, in turn, were dependant on the variation in expected values of EQI which were deemed acceptable by the NRA.

By way of illustration (Clarke et al 1994), the acceptable standard errors in the measurement of each environmental variable were presented in order to achieve 90% or 95% compliance with a pre-requisite that the EQI for number of taxa should not vary by 0.02 and that for ASPT by more than 0.01 (Table 4.1). Each of these two values was 10% of their respective three seasons band width in the 5M system used by the NRA in conjunction with the 1990 survey (Sweeting et al 1992, National Rivers Authority 1994).

In the following sections the complementary aspect of assessing the typical errors in measuring these variables is considered for a range of actual sites.

4.2 Methods

4.2.1 Site selection

The same sixteen sites were used as for the assessment of biological errors and variation.

4.2.2 Recording procedures

R&D Note 412

Eight environmental variables were measured or recorded for each site. Of these, three were obtained at the site and five from maps (Table 4.2).

Table 4.1 Estimates of the tolerable standard errors (%SE, also known as %CV) in estimating the "true" long-term mean, for any site, of each of the environmental variables used in RIVPACS II predictions (reproduced from Clarke et al 1994, Table 4.10).

Results are based on treating 0.02 for number of taxa and 0.01 for ASPT as the acceptable maximum errors in EQI for at least 90% or 95% of all sites.

Variable	Range of Site values	90% Site Compliance		95% Site Compliance	
		Taxa O/E ≤0.02	ASPT O/E ≤0.01	Taxa O/E ≤0.02	ASPT O/E ≤ 0.01
Stream Width (m)	0.3 - 2.0	30% SE	30% SE	20% SE	20% SE
	2 - 4	30% SE	30% SE	25% SE	25% SE
	4 - 20	40% SE	40% SE	25%SE	30% SE
	20 - 120	40% SE	40% SE	30% SE	35% SE
Stream Depth (cm)	4 - 10	20% SE	25% SE	20% SE	25% SE
	10 - 20	35% SE	40% SE	30% SE	35% SE
	20 - 50	40% SE	40% SE	30% SE	35% SE
	50 - 120	40% SE	40% SE	30% SE	35% SE
Stream Slope (m km^{-1})	0.2 - 1.0	40% SE	40% SE	30% SE	35% SE
	1 - 5	40% SE	40% SE	30% SE	35% SE
	5 - 75	40% SE	40% SE	25% SE	30% SE
Distance from Source (km)	0.2 - 8.0	30% SE	40% SE	20% SE	30% SE
	8 - 40	40% SE	40% SE	20% SE	30% SE
	40 - 203	40% SE	40% SE	30% SE	35% SE
Alkalinity (mg l^{-1} CaCO$_3$)	2 - 30	10% SE	30% SE	10% SE	20% SE
	30 - 150	15% SE	20% SE	15% SE	15% SE
	150 - 250	7.5% SE	7.5% SE	5% SE	5% SE
	250 - 314	5% SE	7.5% SE	5% SE	5% SE
Mean Substratum (phi units)	-7.75 : -6	SE=2.5	SE=2	SE=1.5	SE=1.5
	-6 : -3	SE=2.5	SE=2	SE=1.5	SE=1
	-3 : 3	SE=2.5	SE=1.5	SE=1.5	SE=1
	3 : 8	SE=2.5	SE=1.5	SE=1.5	SE=1

R&D Note 412

Discharge category (1-10)	1 - 2	no error allowed			
	3 - 10	none	± 1 category	none	± 1 category

Table 4.2 The eight environmental variables measured or recorded for each site

FIELD VARIABLE	UNITS	MAP VARIABLE	UNITS
Water width	m	Altitude	m
Mean water depth	cm	Distance from source	km
Mean substratum particle size	phi	Slope	$m\,km^{-1}$
		Annual mean flow	Categories
		National grid reference	Eight character alphanumeric

Mean substratum particle size was determined from the estimated percentage of cover of four different categories of particles, each with its own discrete size range; boulders/cobbles, pebbles/gravel, sand and silt/clay. Percentages were converted to phi values and integrated into a single mean particle size within RIVPACS. One of the map-measured variables, the National Grid Reference, was converted to latitude, longitude, mean air temperature and air temperature range using algorithms held within RIVPACS.

The data were recorded on a standard field pro-forma which included space for drawing a schematic site map. No quantitative data were abstracted from the maps for prediction purposes but the sketches did allow the site position and sample area, from the first spring visit, to be relocated in subsequent seasons.

The procedures for acquiring variable measurements in the field and from maps were those laid down in the initial RIVPACS manual (Furse et al 1986) and repeated in the sampling manual prepared for the 1995 River Quality Survey (National Rivers Authority in preparation).

Data collection procedures were thoroughly scrutinised by the IFE field sampling team in advance of the spring sampling. Prior to sampling, all NRA regional and area offices participating in the sampling were supplied with written details of the data-required and brief notes on its collection together with copies of the sections of the RIVPACS manual (Furse et al 1986) giving detailed instructions on how data were to be measured and recorded and the precise units of measurement.

These instructions were re-iterated verbally in individual briefing sessions immediately before sampling. Where queries arose during the briefing sessions NRA biologists were not given precise verbal instructions in case this prejudiced their interpretation of the written instructions. Instead they were referred to these instructions and requested to read the appropriate sections. They were also instructed not to confer at any stage over any element of the field or laboratory recording.

In the written and/or verbal pre-sampling instructions to NRA personnel they were informed that the area over which field measurements should be recorded was the full width of the river for the full length of the biological sample area from the lowermost point of the first biological sample to the uppermost point of the last sample.

R&D Note 412

Data recorders were also instructed as to the scale and, in the case of annual mean flow (discharge) categories, precise type of map to use.

At each site four separate people made complete, independent assessments of field environmental variables needed to make predictions. These were people who fulfilled the roles of biological operators A, B and C at the site together with a second NRA staff member who was operator D.

With the two exceptions noted for biological sampling, operators A, B and C were constant for any given site but varied between sites. There was more variation in the identity of operator D. At only eight sites did the same person fulfil this role in all three seasons.

Field data recording took place in each individual sampling season. All four operators collected their data simultaneously, without conference, immediately after the fourth biological sample had been collected. The NRA operators were asked to make their measurements using the normal methods used by their laboratory, as long as these conformed to the accepted RIVPACS methodology (National Rivers Authority in preparation). They were also asked to bring any items of equipment they normally used to record width, depth and substratum composition.

The proformas were collected by IFE immediately after completion and held in such a manner that they were not consulted by any of the operators prior to any subsequent sampling occasion. Neither were corrections allowed to be made to any recorded measurement once the proformas had been collected in, with the single exception of one width measurement known to have been recorded in the wrong units (see section 4.3.1, "Width").

During the spring sampling visit NRA operators B and D at each site were given a second, clean copy of the environmental data proforma. These were to be taken back to their workplace in order to measure or record the environmental data which needed to be abstracted from maps (Table 4.2).

The NRA staff were asked to each complete their forms entirely independently and to return them to the IFE by post. The two IFE operators also made separate estimates of the values of the same variables.

4.2.3 Data analysis

All the environmental data were initially stored on the IFE MicroVAX computer as standard-format ASCII files. The data were then extracted and stored in a PC spreadsheet in the MINITAB 10 for Windows (MINITAB 1994), ready for analysis. These spreadsheets can easily be exported as Lotus or Microsoft Excel spreadsheets, if required.

In the linear discriminant equations used in RIVPACS II predictions, altitude, distance from source, slope, stream width and stream depth are all expressed as their logarithms. Thus a constant proportional change in the value of one of the variables will have a constant influence on the values of the predictive equations. This means that it is the coefficients of variation (CV = Standard Deviation / Mean) of the estimates at a site which is most useful in assessing the importance of recorder variation.

R&D Note 412

RIVPACS predictions involve the mean values of water width, water depth and mean substrate particle size at a site as variables. In the past these means have generally been estimated from the average of spring, summer and autumn field estimates. Three season (or longer) averages are needed even when predictions are for the expected fauna in a single or pair of seasons. Therefore, the analysis here of the variables estimate in the field will concentrate on the variation between the four recorders in their estimates of the three seasons average values.

Eventually the NRA's aim is to derive fixed expectations for each site. This requires estimates of the long-term mean water width and depth and mean substrate at a site, based on several years field data. The sampling variation guidelines given by Clarke et al (1994) and reproduced in Table 4.1 apply to these estimates of the long-term means. In the present study, only the fraction of that variation which is due to inter-operator differences has been assessed. Therefore, it was hoped that the CV in recorder estimates for these field variables will be less than the tolerable CV's given in Table 4.1, as there is a potential extra source of variation, namely the between year variation in annual mean.

RIVPACS II predictions only involve substrate composition through the use of the three seasons average value of mean substrate particle size. Therefore, the analysis of variation in perceived substrate composition has concentrated on recorder variation in this parameter of annual mean substrate.

Separate RIVPACS II predictions of the expected value of each of number of taxa, BMWP score and ASPT were made from the environmental data of each of the four recorders for each of the 16 sites. In each case, predictions were made for each single season and each combination of seasons.

4.3 RESULTS

4.3.1 Variation in the recorded values

Variables derived from maps

Altitude

The estimated altitudes for each of the 16 study sites, as interpolated from 1:50,000 OS maps by each of the four recorders, are given in Table 4.3. There were no obvious gross mistakes in recording altitude. At eleven of the sites, all four recorders estimated altitude to within 5m of each other. For each of the nine sites with average estimated altitude over 25m recorder CV was less than 13%. The largest difference in estimates occurred at site 8, on Moss Brook, where person IFE1 estimated the altitude to be 35m while each of the other three recorders estimated it to be in the range 15-17m.

R&D Note 412

The estimated altitude for lowland sites 13 to 16 never more than 5m but, because of the low mean of the altitude estimates, these site could have a high CV. RIVPACS II resets values of zero altitude to 1m, so site 14 would use the same altitude for all four recorders. However at site 16, two people recorded altitude as zero metres, one at 1m and a fourth (NRA2) at 5m.

Table 4.3 Values of altitude (m) for each site, independently estimated from maps by four people (IFE1, IFE2, NRA1, NRA2), together with their site means, standard deviations (SD) and percentage coefficients of variation (%CV = 100 SD/Mean)

Site	IFE1	IFE2	NRA1	NRA2	Mean	SD	%CV
1	89	98	100	100	96.75	5.25	5.4
2	53	55	50	50	52.00	2.45	4.7
3	45	45	45	50	46.25	2.50	5.4
4	35	35	45	40	38.75	4.79	12.4
5	15	15	15	10	13.75	2.50	18.2
6	17	28	15	20	20.00	5.72	28.6
7	122	110	105	120	114.25	8.10	7.1
8	35	15	17	16	20.75	9.54	46.0
9	45	43	44	42	43.50	1.29	3.0
10	35	35	34	35	34.75	0.50	1.4
11	56	58	58	55	56.75	1.50	2.6
12	42	45	40	42	42.25	2.06	4.9
13	5	2	3	2	3.00	1.41	47.1
14	0	1	0	0	0.25	0.50	200.0
15	5	4	5	5	4.75	0.50	10.5
16	0	1	0	5	1.50	2.38	158.7

Distance from source

At the four sites nearest their stream source (sites 2, 7, 10 and 11) there was some variation in the estimates of distance, but within each site the range of values was always less than 2km (Table 4.4). Differences at sites further downstream were more marked in absolute terms and often in terms of their CV.

At both sites 6 and 12, IFE2 recorded distance as 21-22km, while the other three recorders were all in close agreement with values in the range 11.2-13km. In these cases it is believed that the distance was read from the wrong map scale on the map wheel. At site 16 recorder IFE1 estimated the distance from source to be 19.5km whereas the others all had estimates of 11-12km.

Distance from source needs to be measured with an accuracy of at most 20-30% CV for distance from source less than 40km (Table 4.1). The accuracy for several sites is around these limits, indicating that care is needed in all aspects of measuring distance from source.

Table 4.4 Values of distance from Source (km) for each site, independently estimated from maps by four people (IFE1, IFE2, NRA1, NRA2), together with their site means, standard deviations (SD) and percentage coefficients of variation (%CV = 100 SD/Mean)

Site	IFE1	IFE2	NRA1	NRA2	Mean	SD	%CV
1	21.0	20.0	20.0	21.0	20.50	0.58	2.8
2	2.6	2.8	3.0	3.0	2.85	0.19	14.9
3	11.0	11.0	11.0	9.5	10.63	0.75	7.1
4	9.0	12.0	7.0	10.0	9.50	2.08	21.9
5	6.0	6.0	6.0	5.5	5.88	0.25	4.3
6	12.0	22.0	12.3	11.5	14.45	5.04	34.9
7	3.6	3.5	3.8	3.5	3.60	0.14	3.9
8	7.0	8.0	8.8	8.9	8.18	0.88	10.8
9	8.5	8.5	7.3	8.0	8.08	0.57	7.0
10	4.5	6.0	4.8	5.0	5.07	0.65	12.8
11	4.0	3.5	2.1	4.0	3.40	0.90	26.4
12	11.2	21.0	11.5	13.0	14.18	4.62	32.6
13	28.5	40.0	35.3	35.0	34.70	4.73	13.6
14	8.0	8.0	7.0	7.8	7.70	0.48	6.2
15	8.0	8.0	8.0	8.2	8.05	0.10	1.2
16	19.5	12.0	11.0	11.0	13.38	4.11	30.7

R&D Note 412

Discharge category

Clarke et al (1994) showed that for sites with discharge category 1 or 2, the discharge needs to be recorded with no error, or only occasionally be out by one category. At eight of the 16 sites, all four recorders did estimate the discharge category to be one when read off the discharge maps (Table 4.5).

For sites with higher discharge, the estimated discharge can be out by one category (Table 4.1). With one exception, all the estimates of discharge category at each of the other sites were within one of the average of the estimates for the site. At site 13, however, the two IFE recorders interpreted discharge as category one, while the two NRA recorders interpreted it as category 4 or 5. This large discrepancy suggested that the IFE and NRA recorders treated the site as being at different points on the discharge maps. This was confirmed to be the case by the leader of the IFE team who maintained that the IFE interpretation was correct. A similar but less extreme disparity between the two organisations occurred at site 15.

Table 4.5 Values of discharge category (1-10) for each site, independently estimated from maps by four people (IFE1, IFE2, NRA1, NRA2), together with their site means, standard deviations (SD) and percentage coefficients of variation (%CV = 100 SD/Mean)

Site	IFE1	IFE2	NRA1	NRA2	Mean	SD	%CV
1	4	5	5	5	4.75	0.50	10.5
2	1	1	1	1	1.00	0.00	0.0
3	2	3	2	2	2.25	0.50	22.2
4	1	1	1	2	1.25	0.50	40.0
5	1	1	1	1	1.00	0.00	0.0
6	1	1	1	1	1.00	0.00	0.0
7	1	1	1	1	1.00	0.00	0.0
8	2	1	3	2	2.00	0.82	40.8
9	1	1	1	1	1.00	0.00	0.0
10	1	1	1	1	1.00	0.00	0.0
11	1	1	1	1	1.00	0.00	0.0
12	2	2	2	2	2.00	0.00	0.0
13	1	1	4	5	2.75	2.06	75.0
14	1	1	1	1	1.00	0.00	0.0
15	1	1	3	3	2.00	1.15	57.7

R&D Note 412

| 16 | 1 | 1 | 2 | 2 | 1.50 | 0.58 | 38.5 |

R&D Note 412

Slope

The slope of the river bed at a site is estimated from the map distance, in metres, between the upstream and downstream height contours crossing the river. This can be difficult to do, as seen from the data presented in Table 4.6. Although no obvious gross errors in measuring or recording were made at any sites, the recorder variation is quite high at most sites (all CV's greater than 17%).

Clarke et al (1994) suggested that the tolerable error is 25-35% (Table 4.1). Although this value is exceeded at several sites, most cases involve lowland sites (13-16) with hardly any slope. These sites were always recorded as having little or no slope (range 0-0.7m/km) and any small differences in estimates within this value range would be expected to have negligible influence on the RIVPACS group(s) to which the sites were predicted to belong (see section 4.3.2).

In RIVPACS II predictions, estimated slopes of zero are reset to 0.1m/km.

Table 4.6 Values of slope (m km^{-1})' for each site, independently estimated from maps by four people (IFE1, IFE2, NRA1, NRA2), together with their site means, standard deviations (SD) and percentage coefficients of variation (%CV = 100 SD/Mean)

Site	IFE1	IFE2	NRA1	NRA2	Mean	SD	%CV
1	6.7	9.1	10.0	10.0	8.95	1.56	17.4
2	33.3	28.5	20.0	20.0	25.45	6.59	25.9
3	1.0	10.0	5.7	7.4	6.03	3.79	62.9
4	25.0	33.3	17.0	18.0	23.33	7.54	32.3
5	50.0	9.1	20.0	7.0	21.53	19.82	92.1
6	3.7	4.0	6.7	4.0	4.60	1.41	30.6
7	33.3	10.0	16.0	16.7	19.00	10.00	52.6
8	0.7	0.8	0.6	1.0	0.78	0.17	22.0
9	2.0	2.0	0.8	2.0	1.70	0.60	35.3
10	1.0	0.7	1.6	1.7	1.25	0.48	38.4
11	3.3	3.3	1.9	3.3	2.95	0.70	23.7
12	3.6	4.8	2.7	2.7	3.45	0.99	28.8
13	0.0	0.3	0.2	0.0	0.13	0.15	120.0
14	0.0	0.5	0.0	0.0	0.13	0.25	200.0
15	0.0	0.2	0.6	0.6	0.35	0.30	85.7
16	0.0	0.5	0.7	0.0	0.30	0.36	118.6

R&D Note 412

Variables measured in the field

Stream width

There appears to be no practical problem in estimating the stream width at a site to within the accuracy needed for RIVPACS predictions (Tables 4.1 & 4.7). For stream widths less than 3.5m (sites 2, 4, 5, 7, 9, 10 and 11) all four estimates of the annual average at a site were within 0.73m of each other. At the two widest sites (1 and 13), both with estimated average widths of about 12m, the range of estimates was 11-15.3m. At all sites, the recording variation CV was less than the 20% limit acceptable in RIVPACS predictions (Table 4.1).

There was one important lesson to learn here. Recorder IFE1 initially had a "moment of summer madness" and recorded the summer stream width of site 7 as 6.2 FEET instead of 1.9 METRES. This was later spotted and corrected but it does highlight the constant need to remember the units in which each variable is to be recorded.

Table 4.7 Values of stream width (m) for each site, independently estimated at the sites by four people (IFE1, IFE2, NRA1, NRA2), together with their site means, standard deviations (SD) and percentage coefficients of variation (%CV = 100 SD/Mean)

Site	IFE1	IFE2	NRA1	NRA2	Mean	SD	%CV
1	12.83	11.77	12.00	12.43	12.26	0.47	3.9
2	2.33	2.60	2.57	2.13	2.41	0.22	9.1
3	6.67	6.77	7.17	8.00	7.15	0.61	8.5
4	3.17	3.33	2.67	2.60	2.94	0.36	12.4
5	3.40	2.97	3.20	2.87	3.11	0.24	7.7
6	5.07	5.83	4.83	4.60	5.08	0.54	10.5
7	2.00	2.03	1.93	2.00	1.99	0.04	2.1
8	7.00	6.33	6.67	7.00	6.75	0.32	4.7
9	2.93	2.87	3.10	3.00	2.98	0.10	3.3
10	3.47	3.60	3.10	3.03	3.30	0.28	8.4
11	1.50	1.43	1.33	1.27	1.38	0.10	7.5
12	6.93	6.87	6.17	6.67	6.66	0.35	5.2
13	11.67	15.33	11.00	13.00	12.75	1.91	15.0
14	8.20	8.77	7.77	6.97	7.93	0.76	9.6
15	6.17	5.63	6.07	5.77	5.91	0.25	4.2

R&D Note 412

16	7.00	8.93	6.33	6.00	7.07	1.31	18.6

R&D Note 412

Stream depth

For sites over 10cm deep, the recorder CV was always less than 20% (Table 4.8), considerably less than the 30-40% acceptable for RIVPACS predictions (Table 4.1). Of the shallow sites, only site 4 had a CV of over 20%, which was due to person IFE2 estimating three seasons average depth to be 4.7cm while the other three recorders had estimates of 7.5, 8.5 and 8.9cm.

Overall variation between recorders in their estimates of annual average stream depth does not appear to a major problem in RIVPACS predictions.

Table 4.8 Values of stream depth (cm) for each site, independently estimated at the sites by four people (IFE1, IFE2, NRA1, NRA2), together with their site means, standard deviations (SD) and percentage coefficients of variation (%CV = 100 SD/Mean)

Site	IFE1	IFE2	NRA1	NRA2	Mean	SD	%CV
1	14.9	13.8	16.3	15.3	15.03	1.03	6.9
2	2.7	2.2	3.5	3.1	2.87	0.56	19.4
3	10.2	10.3	8.0	11.5	10.00	1.45	14.5
4	8.9	4.7	7.5	8.5	7.39	1.91	25.8
5	5.5	5.4	4.6	5.9	5.36	0.56	10.5
6	8.9	8.5	10.4	9.5	9.33	0.83	8.9
7	6.5	6.7	7.2	6.8	6.79	0.28	4.1
8	25.8	27.8	31.2	28.2	28.27	2.24	7.9
9	14.4	12.3	14.8	13.2	13.68	1.15	8.4
10	27.5	27.8	27.2	26.4	27.22	0.62	2.3
11	24.8	22.7	24.0	26.3	24.44	1.53	6.3
12	13.1	13.5	14.3	9.5	12.60	2.12	16.8
13	65.6	78.6	72.6	76.7	73.37	5.75	7.8
14	17.8	21.1	20.4	20.8	20.02	1.49	7.5
15	35.3	38.2	35.3	36.1	36.23	1.38	3.8
16	68.8	46.6	49.9	48.2	53.36	10.37	19.4

R&D Note 412

Mean substratum particle size

Mean substratum can range from -7.75 to 8.00 in phi units. The tolerable standard error (SE) for the estimate of mean substratum at a site is 1-1.5 phi units (Table 4.1). For most of the 16 study sites, the standard deviation between the four recorders in three seasons average mean substratum was less than one phi unit (Table 4.9) and it was always less than 1.5 phi units, as required.

The greatest frequency of SD≥1.0 was at sites such as 10,11, 14 and 15 which were covered by the finer substrate types, sand and silt/clay, with positive phi unit values.

The most variably assessed site was site 14 (Ferry Sluice on Broad Rife near Pagham harbour) (Table 4.10).

Table 4.9 Value of 'Mean substratum particle size (phi units)' for each site, independently estimated at the sites by four people (IFE1, IFE2, NRA1, NRA2), together with their mean and standard deviation (SD).

Site	IFE1	IFE2	NRA1	NRA2	Mean	SD
1	-4.96	-5.92	-5.37	-6.26	-5.63	0.58
2	-4.30	-4.84	-4.88	-4.96	-4.75	0.30
3	-3.31	-2.66	-4.04	-3.76	-3.44	0.60
4	-4.53	-3.70	-4.23	-2.70	-3.79	0.80
5	-4.81	-3.86	-5.44	-5.22	-4.83	0.70
6	-2.74	-2.01	-1.31	-3.59	-2.41	0.98
7	-4.19	-4.16	-3.50	-4.02	-3.97	0.32
8	5.71	6.27	7.71	7.00	6.67	0.87
9	-2.15	-1.59	-0.80	-2.03	-1.64	0.61
10	4.51	6.46	3.74	3.95	4.67	1.24
11	4.34	6.04	6.38	6.83	5.89	1.09
12	-3.96	-4.04	-3.85	-3.00	-3.71	0.48
13	6.76	7.47	7.56	7.31	7.28	0.36
14	2.81	5.67	2.71	2.94	3.53	1.43
15	7.41	7.73	5.19	6.05	6.59	1.19
16	7.95	8.00	7.74	8.00	7.92	0.12

Table 4.10 Variation in the assessment of the substratum composition at Ferry Sluice on Broad Rife near Pagham harbour, which was the most inconsistently assessed study site.

Season	Recorder	%Boulders	%Gravel	%Sand	%Silt/Clay	Mean Substrate
Spring	IFE1	5	15	20	60	4.32
	IFE2	1	14	0	85	6.27
	NRA1	8	2	80	10	1.71
	NRA2	3	12	70	15	1.98
Summer	IFE1	1	29	10	60	3.98
	IFE2	0	10	50	40	3.87
	NRA1	10	10	40	40	2.90
	NRA2	1	15	34	50	4.11
Autumn	IFE1	20	10	60	10	0.12
	IFE2	0	10	0	90	6.87
	NRA1	10	10	30	50	3.50
	NRA2	3	15	52	30	2.72
Three seasons average	IFE1	9	18	30	43	2.81
	IFE2	0	11	17	72	5.67
	NRA1	9	7	50	34	2.71
	NRA2	2	14	52	32	2.94

Conclusions

The variation between people in recording data from maps generally fell within tolerable limits of variation in order that, on the basis of individual variables alone, the overall errors in EQI's for number of taxa and ASPT did not exceed 0.02 and 0.01 respectively (clarke et al 1994).

The acceptable limits were exceeded at sites whose location was misplaced when estimating discharge and at sites with negligible slope.

Great care needs to be exercised in map-reading and double recording, by separate people, is desirable to reduce errors.

The variation between people in their recording of the values for stream width, stream depth

R&D Note 412

and mean substratum particle size at a site are all within the tolerance limits set by Clarke et al (1994).

The effect of temporal variation was outside the scope of this study and was not tested.

R&D Note 412

4.3.2 Effect on variation in expected biotic index values

Independent RIVPACS predictions were made at each site for each of the four recorders. Predictions of expected BMWP index values were made for each single season (spring, summer and autumn) and each combination of seasons (spring/summer, spring/autumn, summer/autumn and spring/summer/autumn).

Both the National Grid reference and an estimate of the average water alkalinity are needed for a site in order to derive RIVPACS II predictions using the standard, and preferred, variables option 1. Variation in estimating and recording these two variables was not assessed in this study. The values for both variables for each site were taken from their values used by the NRA in the 1990 River Quality Survey, as supplied to IFE and given within Table 2.2.

Although the expected value of ASPT and BMWP score (SCORE) differed according to the seasonal combination involved, the correlations between the expected values for any pair of the seven possible sample combinations was always at least 0.993 for ASPT and 0.986 for SCORE. For number of taxa (TAXA), the correlations were all over 0.977 except for cases involving the expected values for summer samples, when correlations with other seasonal combinations were only 0.831 - 0.875.

This high level of correlation meant that it was not necessary to analyse the variation for every possible seasonal combination in detail. Results will be given for spring and also for spring and autumn combined samples. The latter is the intended sampling scheme for the 1995 River Quality Survey.

The maximum possible range of expected values for single seasons using RIVPACS II predictions at any site, not just these 16 study sites, is 15.9 - 29.8 for TAXA and 87.6 - 190.7 for SCORE. At the other extreme, for three seasons combined, the ranges are 21.1 - 37.6 for TAXA and 124.2 - 235.8 for SCORE. For ASPT the absolute range of expected values is roughly the same (± 0.2) for single, two and three seasons combined, namely 4.4-6.8.

The expected values of TAXA, SCORE and ASPT respectively, in spring samples from each site, as estimated from the environmental data recorded by each of the four people (IFE1, IFE2, NRA1, NRA2) are shown in Tables 4.11 to 4.13. Comparison of these values with the maximum possible ranges confirms that the 16 sites cover a good proportion of these ranges but exclude the extremes, especially those sites with the highest expected taxon richness and BMWP scores.

The equivalent expected values for TAXA and SCORE for spring and autumn combined season samples are also listed (Tables 4.14 and 4.15). The values and variation in expected ASPT are similar for each season or combination of seasons and the spring range provides a close approximation to that for each of them.

The SD in expected TAXA between recorders at a site is usually less than 0.6, which is equivalent to a 2-3% CV. This applies to both single and combined season samples, although the SD for combined seasons may be marginally higher in some cases (Tables 4.11, 4.14). For expected SCORE, most recorder SD's are less than 6, with %CV's less than 4-5% (Tables 4.12, 4.14).

R&D Note 412

Table 4.11 Variation in the RIVPACS II expected values of number of taxa in spring samples, based on independent estimates of the environmental variables for each of four people (IFE1, IFE2, NRA1, NRA2), together with the means, standard deviations (SD) and percentage coefficients of variation (%CV = 100 SD/Mean) of the expected values

Site	IFE1	IFE2	NRA1	NRA2	Mean	SD	%CV
1	24.9	24.0	24.1	24.0	24.3	0.4	2
2	24.8	24.6	25.0	25.0	24.9	0.2	1
3	22.9	23.4	23.7	24.1	23.5	0.5	2
4	24.3	24.3	24.4	24.3	24.3	0.1	0
5	24.9	25.1	25.0	25.1	25.0	0.1	0
6	24.1	22.7	24.4	24.2	23.9	0.8	3
7	24.0	22.8	23.2	23.3	23.3	0.5	2
8	24.5	22.7	23.0	22.9	23.3	0.8	4
9	20.4	20.3	20.2	20.4	20.3	0.1	0
10	20.6	20.2	20.9	20.8	20.6	0.3	2
11	19.8	19.6	19.8	19.7	19.7	0.1	0
12	20.2	20.3	20.2	20.1	20.2	0.1	0
13	19.6	19.5	20.3	21.8	20.3	1.1	5
14	22.6	20.0	22.8	22.8	22.1	1.4	6
15	20.3	19.9	21.4	21.2	20.7	0.7	3
16	19.6	19.5	19.6	20.4	19.8	0.4	2

Table 4.12 As for Table 4.11 but for predicted spring BMWP scores.

Site	IFE1	IFE2	NRA1	NRA2	Mean	SD	%CV
1	158	152	153	152	153.5	2.9	2
2	157	156	159	159	157.8	1.4	1
3	132	145	146	150	143.1	7.7	5
4	156	154	155	153	154.1	1.3	1
5	150	146	148	146	147.5	1.7	1
6	134	123	135	135	131.5	5.9	4
7	135	124	126	128	128.3	4.9	4
8	135	118	119	120	123.0	8.3	7
9	99	98	97	99	98.5	1.0	1
10	100	97	103	103	100.6	2.8	3
11	95	93	91	94	93.0	1.6	2
12	98	98	97	97	97.4	0.6	1
13	88	87	92	101	91.7	6.3	7
14	105	90	106	106	101.8	8.1	8
15	92	89	100	98	94.6	4.9	5

R&D Note 412

| 16 | 88 | 87 | 88 | 93 | 88.8 | 2.8 | 3 |

Table 4.13 As for Table 4.11 but for predicted spring ASPT values.

Site	IFE1	IFE2	NRA1	NRA2	Mean	SD	%CV
1	6.3	6.3	6.3	6.3	6.3	0.00	0
2	6.3	6.3	6.4	6.3	6.3	0.05	1
3	5.8	6.2	6.1	6.2	6.1	0.19	3
4	6.4	6.3	6.3	6.3	6.3	0.05	1
5	6.0	5.8	5.9	5.8	5.9	0.10	2
6	5.5	5.4	5.5	5.5	5.5	0.05	1
7	5.6	5.4	5.4	5.5	5.5	0.10	2
8	5.5	5.2	5.2	5.2	5.3	0.15	3
9	4.8	4.8	4.8	4.8	4.8	0.00	0
10	4.8	4.8	4.9	4.9	4.9	0.06	1
11	4.7	4.7	4.6	4.7	4.7	0.05	1
12	4.8	4.8	4.8	4.8	4.8	0.00	0
13	4.4	4.4	4.5	4.6	4.5	0.10	2
14	4.6	4.5	4.7	4.6	4.6	0.08	2
15	4.5	4.5	4.6	4.6	4.6	0.06	1
16	4.4	4.4	4.4	4.5	4.4	0.05	1

Table 4.14 As for Table 4.11 but for predicted number of taxa for spring and autumn combined samples.

Site	IFE1	IFE2	NRA1	NRA2	Mean	SD	%CV
1	30.3	29.5	29.5	29.4	29.7	0.4	1
2	30.5	30.2	30.7	30.7	30.5	0.2	1
3	29.0	28.9	29.3	29.5	29.2	0.3	1
4	29.7	29.9	29.7	29.8	29.8	0.1	0
5	30.3	30.5	30.5	30.6	30.5	0.1	0
6	29.4	28.4	29.6	29.6	29.3	0.6	2
7	29.2	28.1	28.3	28.5	28.5	0.5	2
8	29.7	27.6	28.1	27.9	28.3	0.9	3
9	25.3	25.2	25.0	25.3	25.2	0.1	1
10	25.3	24.9	25.7	25.6	25.4	0.4	1
11	24.6	24.4	24.1	24.5	24.4	0.2	1
12	25.2	25.4	25.2	25.1	25.2	0.1	0
13	24.2	24.1	25.1	26.8	25.1	1.3	5
14	27.6	24.7	27.9	27.9	27.0	1.6	6

R&D Note 412

| 15 | 25.0 | 24.6 | 26.3 | 26.0 | 25.5 | 0.8 | 3 |
| 16 | 24.2 | 24.1 | 24.3 | 25.1 | 24.4 | 0.5 | 2 |

R&D Note 412

Table 4.15 As for Table 4.11 but for predicted BMWP scores for spring and autumn combined samples.

Site	IFE1	IFE2	NRA1	NRA2	Mean	SD	%CV
1	192	186	187	186	188.0	2.9	2
2	193	192	194	194	192.9	1.0	1
3	169	179	180	183	177.8	6.1	3
4	189	189	188	187	188.3	0.8	0
5	182	178	181	179	180.0	1.7	1
6	165	155	166	166	163.0	5.2	3
7	165	153	156	158	158.0	5.2	3
8	166	145	148	147	151.7	9.6	6
9	124	123	121	124	123.2	1.3	1
10	125	120	128	127	124.9	3.3	3
11	118	117	115	117	116.9	1.4	1
12	123	124	122	122	122.7	0.8	1
13	110	110	115	126	115.1	7.4	6
14	131	113	132	132	126.9	9.4	7
15	115	113	125	122	118.6	5.8	5
16	110	110	110	117	111.7	3.3	3

The largest range of spring sample estimates of expected number of taxa was for site 14 (range 20.0 - 22.8), which is not surprising as this was the site with the most inconsistent estimation of its substratum composition (Table 4.10). For BMWP score, site 8 had slightly more variable expected values than site 14, with a range of 118-135 for spring samples.

For 10 of the 16 sites the four recorders estimates of the RIVPACS predictor variables all led to expected ASPT values for spring varying by no more than 0.1 at any one site (Table 4.13). The largest range of expected values was only 5.8 - 6.2 (at site 3), such that the coefficient of variation was at most 3% at any site.

The standard deviation of all expected values between recorders does not tend be related to the mean expected value for a site. This seems true for TAXA, SCORE and ASPT predictions for single and multiple season samples such spring and autumn combined (Tables 4.14-4.15). However, there is some suggestion that the recorder SD of expected TAXA and SCORE is higher for some types of site than others.

In particular, SD appears higher for sites of type 5b and 9b (Figure 4.2 (a) and (b)). Sites in group 5b have intermediate expected taxonomic richness. In terms of the detrended correspondence analysis used to form the RIVPACS sites groupings, these sites have intermediate ordination scores on the first axis. It may be that their central position in ordination and environmental discrimination space means variation in estimating the values of the environmental variables for such sites has more influence on changing their probability of belonging to each biological group and hence on their expected biotic index scores. However, the between recorder variation in estimating the expected

R&D Note 412

value of ASPT shows no dependence on site type (Figure 4.2(c)).

Figure 4.2 Within-site standard deviation (SD$_E$) versus site mean of the expected values of BMWP index values for spring and autumn combined samples, based on four recorders who independently derived estimates of the environmental variables used in the RIVPACS II predictions of expected values. (a) number of BMWP Taxa, (b) BMWP score, (c) ASPT.

O,●,□,■ denote RIVPACS group type (3a, 5b, 8a, 9b respectively) of the 16 study sites.

(a) SD Expected TAXA

(b) SD Expected SCORE

(c) SD Expected ASPT

R&D Note 412

R&D Note 412
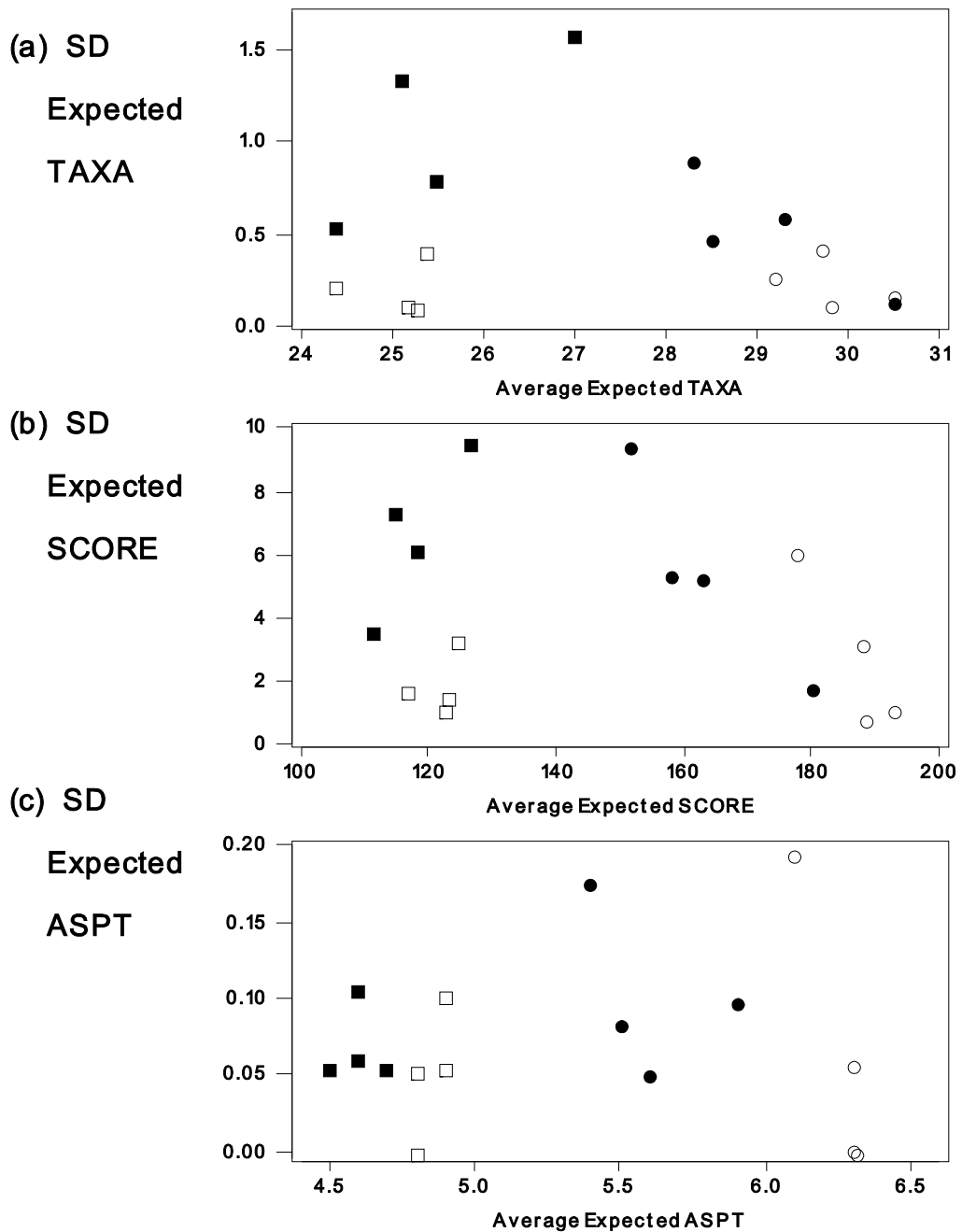
Overall, and for simplicity, it seems reasonable to estimate the recorder variation in the expected value of any one biotic index as a constant. For any season(s) samples, the overall SD is best estimated by :

$$\text{Overall SD} = SD_E = \sqrt{\sum_{i=1}^{16}(SD_i)^2/16}$$

where: SD_i is the estimated recorder SD for site i

Moreover, the overall recorder SD in expected values does not seem to change consistently according to the number of seasons samples combined (Table 4.16).

Table 4.16 The mean within-site standard deviation (SD_E) in expected values of numbers of taxa, BMWP scores and ASPT's, based on the four recorders who independently derived estimates of the environmental variables used in the RIVPACS II predictions of expected values.

Seasons combined	TAXA SD_E	SCORE SD_E	ASPT SD_E
Spring	0.60	4.7	0.083
Summer	0.32	2.8	0.077
Autumn	0.43	3.4	0.080
Spring+Summer	0.56	4.7	0.079
Spring+Autumn	0.65	5.0	0.086
Summer+Autumn	0.50	4.1	0.075
Three Seasons	0.60	5.0	0.084
Overall mean	**0.53**	**4.3**	**0.081**

This consistency suggests that variation and errors in the expected index values due to differences between individuals in the estimation and recording of the environmental variables for a site can be treated as the same constant for all types of site and irrespective of whether the expected values are for single, two or three seasons combined samples. These constant recorder SD are estimated to be :

$SD_E = 0.53$ for expected number of BMWP taxa

$SD_E = 4.3$ for expected BMWP score

$SD_E = 0.081$ for expected ASPT

R&D Note 412

4.3.3 <u>Summary</u>

Errors in map-derived time-invariant variables

1. Different people seem to record altitude and stream slope at a site within adequately consistent limits to prevent errors in recording having a large influence on RIVPACS predictions.

2. The variation between recorders in measuring the distance from source of sites from OS maps can be as much as 30%CV. At some sites, one estimate differed substantially from the other three, suggesting care is needed in measuring and converting from map to actual distances.

3. It is recommended that all map-derived variables are measured and recorded completely independently by two people. Where two estimates are in close agreement, simply use their average. Any large differences need to be explained and only the most appropriate value used. This includes latitude and longitude and air temperature means and ranges which are each dependant on the reading of the site National Grid reference, a procedure not examined here.

Errors in field-derived time-variant variables

4. The variation between people in their recording of the values for stream width and stream depth at a site is almost always less than 20%CV, such that differences have negligible effects on RIVPACS predictions.

5. The variation between recorders in their estimation of the substratum at a site in each season is such that the variation in the values they get for the annual mean substratum particle size always has a standard deviation less than 1.5 phi units, the general error permitted without significantly affecting RIVPACS predictions. Sites with finer substrata tend to be the most inconsistently estimated. Field recorders need to be consistent in their understanding and interpretation of the terms "sand" and "silt/clay".

6. Although the time-variant variables seem to be estimated consistently enough by different observers making estimates on the same day, there are other unexamined sources of error in estimating either the annual mean or the long-term average at the site for each variable. By estimating a value in each season, the average should encompass a major part of the seasonal variation, but obviously the recorders would have obtained different values for say stream depth in spring if one had gone out just before a storm and another say a few days after.

7. It is not possible within this research programme to estimate the accuracy and effects on RIVPACS predictions of using a single year's three seasons average value to estimate the long-term average value of each environmental variable for a site.

R&D Note 412

8. The %CV's of the expected values of the BMWP indices derived in this study from different observers' variable measurements are, therefore, probably reasonable estimates for the errors for year-specific predictions of the expected values of the indices. However they may be under-estimates of the errors if these one-year means are used as estimates of the long-term averages in fixed predictions for each site. More research is needed into temporal variation in the data used to provide these long-term, fixed values of each environmental variable.

Variation in expected BMWP index values

9. Variation between people in estimating, measuring and recording the RIVPACS environmental variables leads to variation between individuals in the expected BMWP index values for a site.

10. There is some suggestion that variation in expected number of taxa and BMWP score is greater for some types of site than others but there is no such pattern to the variation in expected ASPT.

11. The between-operator variation in expected numbers of taxa, BMWP scores and ASPT's does not seem to vary according to whether single or combined seasons sample predictions are involved.

12. It is recommended that for year-specific predictions of the BMWP index values for a site, the errors in the expected values due to measuring the environmental variables are assumed to have the following standard deviations, irrespective of whether predictions are for single season or two or three seasons combined samples:

Error SD of Expected : Number of taxa = 0.53

$$\text{BMWP score} = 4.3$$

$$\text{ASPT} = 0.081$$

R&D Note 412

5VARIATION IN ECOLOGICAL QUALITY INDEX VALUES

The Ecological Quality Indices (EQI) of a site are defined for each of number of taxa, BMWP score and ASPT as the ratio of the observed value (O) to the expected value (E). The expected value is that predicted by RIVPACS for unstressed sites of that environmental type.

In this chapter the results from section 2 on sampling variation, section 3 on sample processing errors and section 4 on the effects of variation in the measurement of environmental variables are integrated in order to provide methods for analyzing variation and errors in the EQI value for a site.

The observed single or combined seasons sample index value (O) divided by the expected index value (E) can be used as the best estimate of the EQI for a site. However, because several sources of error in O and E are involved, it is not feasible to calculate the overall errors in an EQI estimate mathematically. It is therefore recommended that Monte Carlo numerical simulation is used to generate many (say 1000) potential sample EQI values for the site using the results of sections 2-4 of this report on the statistical distribution of the various errors in O and E.

The statistical distribution of the simulated EQI values can then be used in site quality assessments. Most obviously, the limits within which p% (typically 90 or 95%) of the simulated O/E ratios fall can be used as confidence limits for the true EQI for the site (section 5.3.2).

The same philosophy can also be extended to assessing whether the EQI has significantly changed between two surveys at a site (see section 5.3.3).

The proportion of simulated EQI values for each index in each quality band can be used to estimate the probability each site belongs to each biological quality band and hence to estimate quality band misclassification rates (see section 5.4).

The variation and errors in the observed values are independent of those in the expected values. The total variation and errors in the observed and expected values will be summarised separately in sections 5.1 and 5.2 respectively.

5.1Total variation in the observed values

The total variation in the observed values of number of taxa, BMWP score and ASPT is due to two components :

• sampling variation

• sample processing (= sorting and identification) errors

The Monte Carlo simulated observed (O) values should first add a random sampling variation term, as detailed in section 5.1.1.

If the NRA chooses to correct for the bias due to their sample processing errors, then the underestimation should then be estimated as in section 5.1.2 and added to the simulated estimate of

R&D Note 412

the observed value of number of taxa and BMWP score.

5.1.1 Sampling variation

On the basis of the analyses of the replicated sampling programme detailed in section 2, and summarised in Table 2.15, the variation in the observed values due to sampling effects are best estimated by assuming:

(i) The sampling variation of the observed number of taxa, O_T, is such that the square root of the observed number of taxa, $\sqrt{O_T}$, has a normal sampling distribution with mean equal to $\sqrt{O_T}$ and standard deviation equal to $SD_{\sqrt{T}}$. The standard deviation $SD_{\sqrt{T}}$ equals 0.228, 0.164 or 0.145, according to whether the observed number of taxa is based on a single season sample, a two seasons combined sample or a three seasons combined sample respectively.

(ii) The sampling variation of the observed BMWP score, O_S, is such that the square root of the observed BMWP score, $\sqrt{O_S}$, has a normal sampling distribution with mean equal to $\sqrt{O_S}$ and standard deviation equal to $SD_{\sqrt{S}}$. The standard deviation $SD_{\sqrt{S}}$ equals 0.588, 0.418 or 0.361, according to whether the observed number of taxa is based on a single season sample, a two seasons combined sample or a three seasons combined sample respectively.

(iii) The sampling variation of the observed ASPT, O_A, is such that O_A has a normal distribution with mean O_A and standard deviation, SD_A, where SD_A equals 0.249, 0.161 or 0.139, according to whether the observed value of ASPT is based on a single season sample, a two seasons combined sample or a three seasons combined sample respectively.

If O_T, O_S and O_A are the observed values of number of taxa, BMWP score and ASPT and if Z_α is the α two-sided percentage point of a standard normal distribution (eg $Z_{95} = 1.96$), then $\alpha\%$ confidence intervals due to sampling variation for the observed values are:

$$(\sqrt{O_T} - Z_\alpha SD_{\sqrt{T}})^2 \ \text{to} \ (\sqrt{O_T} + Z_\alpha SD_{\sqrt{T}})^2 \qquad \text{for number of taxa} \qquad (5.1)$$

$$(\sqrt{O_S} - Z_\alpha SD_{\sqrt{S}})^2 \ \text{to} \ (\sqrt{O_S} + Z_\alpha SD_{\sqrt{S}})^2 \qquad \text{for BMWP score} \qquad (5.2)$$

$$\text{and} \ (O_A - Z_\alpha SD_A) \ \text{to} \ (O_A + Z_\alpha SD_A) \qquad \text{for ASPT.} \qquad (5.3)$$

The width of a 95% confidence interval for any observed value, O_A, of ASPT is O_A plus and minus W, where W equals 0.49, 0.32 or 0.27, according to whether O_A is based on a single season sample, a two seasons combined sample or a three seasons combined sample.

The width of the confidence intervals for the observed value of number of taxa and BMWP score depend on the observed values. Table 5.1 gives some illustrative examples of sampling variation confidence limits for observed values for both indices. The limits are slightly asymmetrical, especially for BMWP score, extending further above the observed value than below it.

R&D Note 412

Table 5.1 Examples of 95% sampling confidence intervals (95%CL) for observed values of number of taxa and BMWP score based on single or combined seasons samples

	No. seasons in the combined sample	Observed value	95% CL	
			Lower	Upper
Number of	1	5	3	7
	2	5	4	7
	3	5	4	6
	1	15	12	19
	2	15	13	18
	3	15	13	17
	1	30	25	35
	2	30	27	34
	3	30	27	33
BMWP	1	15	7	25
	2	15	9	22
	3	15	10	21
	1	50	35	68
	2	50	39	62
	3	50	40	61
	1	150	123	180
	2	150	131	171
	3	150	133	168

Usually this sampling variation will need to be incorporated into an error term for the observed values as part of the overall errors in the EQI's. This should be done by adding on a random error term, to generate simulated values O_{Tr}, O_{Sr} and O_{Ar}, as follows:

$$O_{Tr} = (\sqrt{O_T} + R_T.SD_{\sqrt{T}})^2 \qquad \text{for observed number of taxa ;} \qquad (5.4)$$

$$O_{Sr} = (\sqrt{O_S} + R_S.SD_{\sqrt{S}})^2 \qquad \text{for observed BMWP score;} \qquad (5.5)$$

$$\text{and } O_{Ar} = O_A + R_A.SD_A \qquad \text{for observed ASPT.} \qquad (5.6)$$

R&D Note 412

where R_T, R_S and R_A are all random deviates from a standard normal distribution with zero mean and unity variance.

However Table 5.2 shows that for the 16 sites in the sampling variation study, the sampling variation in values of number of taxa and BMWP score at a site are very highly correlated ($r \geq 0.95$). The correlations between the sampling variation in number of taxa and ASPT are much lower (≤ 0.5) and can be ignored.

Table 5.2 Correlations between the observed values of the square root of number of taxa ($\sqrt{\text{TAXA}}$), square root of BMWP score ($\sqrt{\text{SCORE}}$) and ASPT at a site. Differences in the mean values between sites have been eliminated, so the correlations between the residual sample values measure the true sampling correlations.

No. seasons in the combined sample	$\sqrt{\text{TAXA}}$ vs $\sqrt{\text{SCORE}}$	$\sqrt{\text{TAXA}}$ vs ASPT	$\sqrt{\text{SCORE}}$ vs ASPT
1	0.95	0.50	0.74
2	0.95	0.37	0.62
3	0.95	0.34	0.58

In order to ensure that the simulated value of observed ASPT equals the simulated value of BMWP score divided by the simulated value of number of taxa, as it should, one of two simulation methods should be used:

(i) Use two independent random numbers R_T and R_A to make simulated observed values, O_{Tr} and O_{Ar} of the number of taxa and ASPT of the site. Derive the corresponding simulated value of the site's BMWP score as $O_{Sr} = O_{Tr}.O_{Ar}$.

or

(ii) Assume $R_S = R_T$ and generate one random number to make simulated values, O_{Tr} and O_{Sr} for TAXA and SCORE for the site. Derive the corresponding simulated value of ASPT as $O_{Ar} = O_{Sr} / O_{Tr}$.

It is not clear which option is best. These approaches are especially appropriate when trying to evaluate the errors and misclassification rates for quality banding systems based on all three EQI's (TAXA, SCORE and ASPT), as in method 5M used for the 1990 River Quality Survey. Simulation method (ii) is probably most appropriate for quality assessments based on the EQI's for number of taxa and ASPT, which ignore EQI for BMWP score.

5.1.2 Sample processing errors

In section 3 the effect of sample processing errors by the NRA was examined using samples audited by IFE from 1990 and 1992. The general effect of the errors was to under-estimate the 'true' observed sample value of number of taxa and BMWP score, and also to slightly increase the error variance in the observed value of number of taxa, BMWP score and ASPT.

Procedures for correcting for the bias in the NRA estimates of the observed number of taxa in single season samples are given in section 3.4.3 whilst those devised to allow for single season sample processing errors on the value and error variance in of observed number of taxa, BMWP score and ASPT are set out in section 3.7. The latter includes ways of correcting for the bias (section 3.7.1) and also allow for the consequential increase in variance of the estimate of the 'true' observed value (section 3.7.2).

The approach is to simulate the under-estimation of the number of taxa using a Poisson distribution with mean under-estimation appropriate for the year and NRA region, then estimate the ASPT of the missed taxa as a function of the number of taxa present. These estimates are then used to derive simulated values of the 'true' observed number of taxa and BMWP score, from which the 'true' observed ASPT is calculated by division.

The ideas in section 3.7 are extended in section 3.8.3 to give procedures to simulate the 'true' observed values of number of taxa, BMWP score and hence ASPT for two and three season combined samples. It also consequentially provides a means of including the extra variance in the estimate of the observed value due to sample processing errors.

In 1992, the average underestimation of the observed number of taxa was about 1.5 for single, two season combined and three season combined samples. This value seems appropriate to use for future years, including for the 1995 NRA survey, assuming the NRA have successfully implemented the WRc quality control procedures (WRc, 1994).

5.2 Total errors in the expected values (E)

For reasons which will be discussed in the next chapter, the only relevant source of error in the expected value is assumed here to be in estimating the values of the environmental variables to make RIVPACS predictions for any particular site.

Clarke et al (1994) gave the tolerable standard errors in the estimation of each environmental variable to ensure that the effect in 90% or 95% of sites these errors would lead to errors of at most 0.01 for EQI of ASPT and at most 0.02 for EQI of number of taxa (Table 4.1 of the current report).

Section 4 of this report includes analyses of the variation between four recorders in the recording of most of the predictor variables for 16 sites of varying site types. From this study, it is recommended that, for year-specific predictions of the expected values of BMWP indices for a site, the errors in the expected values due to measuring the environmental variables are assumed to follow a normal distribution with a mean of zero and the following standard deviations (SD) :

Error SD of the expected :number of taxa= 0.53

BMWP score= 4.3

ASPT= 0.081

R&D Note 412

The same SD of expected values apply for single seasons, two season combined and three season combined samples.

In the Monte Carlo simulations to assess the overall effect of errors in the observed and expected values for a site, the expected value for each simulation is estimated by the RIVPACS predicted value plus a random error term from appropriate normal distribution given above. For example, if the predicted expected number of taxa is E_T, then the simulated value, E_{Tr}, is:

$$E_{Tr} = E_T + 0.53 \, Z \quad ; \quad \text{where Z is a standard normal deviate} \tag{5.7}$$

5.3 Variation in the observed to expected ratios (EQI)

As explained earlier, it is suggested that the variation and errors in estimating the EQI for a site should be assessed using Monte Carlo simulations which involve all the sources of error. The procedures to simulate error-included values of the observed and expected value of each BMWP index are detailed in sections 5.1 and 5.2 respectively. In each of m (say 1000) simulations, the simulated observed value is divided by the simulated expected value to get a simulated value of the EQI for the site. This process gives a simulated frequency distribution of m possible values of the EQI for the site.

5.3.1 Precision in the EQI value for a site

The lower p/2 and the upper p/2 percentile values of this simulated frequency distribution for EQI give (100-p)% confidence limits for the EQI for a site. For example the lower and upper 5 percentile values give a 90% confidence interval for the EQI.

5.3.2 Precision in detecting change in EQI at a site

The same simulated process and ideas can easily be extended to test whether a real change in EQI has occurred at a site between two surveys or years.

Case (i) Same seasons involved in each year , common estimate of expected value

If the EQI at a site in two different years (referred to as years 1 and 2) are both based on the same season or combination of seasons and use the same values of the environmental variables, then the change in EQI can be assessed by Monte Carlo simulations, as follows:

In each of the m simulations, section 5.1 should be used to derive <u>independent</u> simulation estimates, O_{1r} and O_{2r}, for the observed values in year 1 and 2, section 5.2 used to simulate one common expected value E_r, and then the simulated change in EQI estimated by :

$$D_r = (O_{2r} - O_{1r}) / E_r \tag{5.8}$$

R&D Note 412

The m simulated values of D_r provide a frequency distribution for the increase in EQI at the site between years 1 and 2. The lower p/2 and upper p/2 percentile values of the frequency distribution of D_r give (1-p)% confidence limits for the increase in EQI at the site.

•If the confidence limits include zero then the difference is not statistically significant and there is insufficient evidence to conclude that the site quality has necessarily changed at all.

•If the (1-p)% confidence limits are both positive, then this indicates that a statistically significant (at the p% probability level) increase in EQI and hence quality has occurred at the site.

•If the (1-p)% confidence limits are both negative, then a statistically significant (at the p% probability level) decrease in EQI has occurred at the site.

These procedures provide an appropriate statistical test for a change in EQI and are common to all cases which follow.

Case (ii) Same seasons involved in each year but separate estimates of expected value in each year

If the EQI at a site in two different years (referred to as years 1 and 2) are both based on the same season or combination of seasons, but use separate values of the environmental variables in each year, then the change in EQI can be assessed by Monte Carlo simulations, as follows:

RIVPACS should already have been used to derive separate expected values for each year.
In each of the m simulations, section 5.1 should be used to derive <u>independent</u> simulation estimates, O_{1r} and O_{2r}, for the observed values in year 1 and 2, after which section 5.2 is used to simulate expected values E_{1r} and E_{2r} for the expected values in year 1 and 2. The latter involves adding an appropriate independent random term to each of the RIVPACS expected values for each year.

Then the simulated change in EQI for simulation m is estimated by :

$$D_r = (O_{2r} / E_{2r}) - (O_{1r} / E_{1r}) \tag{5.9}$$

Case (iii) Different seasons, different combinations of seasons or different number of seasons involved in each year

If the EQI at a site in two different years are based on different seasons, different combinations or even different numbers of seasons, then a different approach is needed in the Monte Carlo simulations.

In each simulation, the procedures of sections 5.1-5.3 should be used to derive two independent simulated values for the observed values, expected values and hence EQI values, using errors standard deviations for the appropriate number of seasons for each year.

For each simulation the change in EQI is estimated by :

$$D_r = (O_{1r} / E_{1r}) - (O_{2r} / E_{2r}) = EQI_{1r} - EQI_{2r} \tag{5.10}$$

R&D Note 412

If the same values of the environmental variables were used to derive the expected value for each of the two years, even though the expected values were for different season combinations, then the same random deviate Z in equation (5.7) should be used to simulate the two expected values, E_{1r} and E_{2r}, even though the predicted expected values will obviously be different in the two years if they are based on different combinations of seasons.

5.3.3 Detecting differences in EQI between any two sites

Detecting such spatial differences in quality is identical to case (ii) above, assuming that separate expected values are estimated for each site. Included in this category, for example, are samples collected at the same time from sites upstream and downstream of a possible source of environmental impact.

5.4 <u>Implications for biological quality banding</u>

The Monte Carlo simulation procedures detailed in the previous section can readily be used to estimate the probability that a site belongs to each quality band.

For instance, to estimate the probability a site belongs to a band within the 5M band system devised for the 1990 River Quality Survey (National River Authority 1994), each of the m simulations would give an EQI value for each of number of taxa, BMWP score and ASPT, which would lead to the site being allocated to a band (A,B,C,D) based on each of the three EQI values and hence to an overall band using the 5M algorithm.

The proportion of all the m simulated triplets of EQI values allocated to each 5M band, estimates the probability the site belongs to each 5M quality band.

The probability of mis-banding each site can also be estimated by assuming a "true" site EQI value based on each of the observed number of taxa, BMWP score and ASPT, from which the banding rule is used to classify the site into its true quality band.

The simulation process can then be used to generate a simulate a possible value of EQI for each of the three indices, which leads to the site being classified by the same banding rule into one of the bands. The proportion of the m simulations which do not classify the site into its "true" band estimates the probability of misclassifying sites with those "true" EQI values.

This approach can also be readily extended to assess the probability that a site has erroneously changed quality band between two years by classifying the site into quality bands in each year for each simulation and calculating the proportion of the simulations in which the site was classified into different bands in the two years.

R&D Note 412

6DISCUSSION

6.1The merit of biological monitoring and justification for its use.

In the introductory chapter of this report the background to the use of macro-invertebrate sampling for assessing the biological condition of rivers was set out in some detail.

The use of macro-invertebrates in this way has become a widespread global practice over the last three decades (Furse et al 1990, Metcalfe-Smith 1995). Within Britain it has become a core activity of the NRA as well as the River Purification Boards and the Department of the Environment (Northern Ireland). Millions of pounds are invested in the activity each year.

Despite this, there has remained an under-current of uncertainty about the reliability of macro-invertebrate data, in comparison with chemical water analysis. This has led to a chequered record of the independent use of biological information in the, now, quinquennial River Pollution/Quality Surveys although chemistry has been a constant, indeed primary, part of these surveys.

The belief has undoubtedly been re-inforced by the knowledge the chemical analyses can be carried out with accuracy and reproducibilty if a single sample is analyzed and that this accuracy can be confirmed by independent audits/calibrations by other laboratories. In contrast biology is a behavioural science in which samples are inevitably collected and processed with error and, until the last five years rarely subject to formal audit.

In recent years there has been a growing awareness of the sources of variation and error in the chemical data due to, for example, temporal variation in the data, frequency of sampling, computer-logging errors and inter-regional differences in the summarisation and banding of the available data into quality classes. In response to this clear and consistent guidelines have been put in place for the collection and use of chemical data for banding purposes (National Rivers Authority 1991b, Appendix 3). With these procedures in place it then became possible to assess confidence limits to the data used in classification and hence the probability that a site was the placed in the wrong class in the banding system (National Rivers Authority 1991b, Appendix 4). The use of these statistics to summarise the errors in the chemical classification of samples was operationalised through the **CL**ass **A**llocation **M**odel, CLAM (Warn 1990).

The conjunctive, independent use of biological and chemical data in national surveys is a complementary process in which the chemical analyses gives relatively precise information on the levels of determinant concentrations at the time of sampling whilst biological data integrates environmental conditions over recent months at the same site (Furse et al 1990). Thus, although the use of macro-invertebrate data may offer no more than clues as to the source of environmental stress, it may be able to detect stresses which are either not persistent at the time of sampling or are persistent but due to factors other than those chemical determinands currently being analyzed for.

The complementary nature of biological and chemical data is recognised by the NRA (1991b) and the Royal Commission on Environmental Pollution (1992) and the need for biological monitoring has been strengthened by the proposed Council of the European Union Directive on the Ecological Quality of Water (Council of the European Union 1994).

R&D Note 412

Clearly, however, if biology in general, and macro-invertebrate sampling in particular is to fulfil its role then it must be accompanied by the same rigorous and formal procedures for the quantification of error terms in the data and the determination of error rates in the classification of the resultant site data into "quality" classes. This need has been recognised in the current contract which includes a complementary study of the design and implementation of analytical control procedures for the sorting and identification of macro-invertebrate samples.

6.2 The role of RIVPACS in the definition of the biological condition (quality) of sites

The starting point in the quantification of errors has been an assumption that the method of choice for collecting, indexing and classifying macro-invertebrate samples is the use of RIVPACS (Cox et al 1991, Wright et al 1993) and its associated standard sampling techniques (Furse and Gunn 1990, National Rivers Authority in preparation). This assumption is in line with the recommendations of the Royal Commission on Environmental Pollution (1992).

Furthermore the principal underlying RIVPACS, that the observed fauna should be compared with that to be expected at the site, has been implicitly endorsed in the Ecological Quality of Water Directive (Council of the European Union 1994) where the specification of good ecological quality includes the condition that:-

" the diversity of invertebrate communities (planktonic and bottom dwelling) should resemble that of similar water bodies with insignificant anthropogenic disturbance. Key species/taxa normally associated with the undisturbed condition of the ecosystem should be present."

Within RIVPACS the multivariate relationships between macro-invertebrate assemblage and environmental data from a reference set of sites, of perceived good biological condition, are used to predict the fauna to be expected at each monitored site in the absence of any pollution or environmental stress.

The environmental variables used to make faunal predictions were chosen, as much as was possible, to be time-invariant and of fixed valued at any one site. They were also selected on the grounds that they were not to be influenced by, nor to influence, a site's biological condition. This meant that certain other site characteristics which do influence water quality and hence biological condition, such as nitrate, chloride and potassium concentrations were purposely excluded from the derivation of RIVPACS predictions. This was because they are often a cause of the change in condition that RIVPACS-derived Ecological Quality Indices are used to detect.

Under these circumstances, there is no obvious way, other than "trial and error" of determining whether the RIVPACS model would make more accurate predictions of the expected fauna and derived index values if other combinations of environmental variables were used as predictors. Nor is it easy to calculate the efficiency of the RIVPACS model in utilising the appropriate, available environmental data for making predictions.

Through extensive testing of alternative methods of multivariate prediction, IFE believe that the system used in RIVPACS III makes near optimum use the environmental variables and hence the conditional prediction system errors are negligible (Wright et al in preparation). However, this cannot be proven.

R&D Note 412

Another reason why it is impossible to define the predictions errors is because the reference sites, though selected to all be of "high" quality or in some hypothetical "top quality band", are still of varying quality, however quality is defined. This means that it is impossible to differentiate between prediction errors and real variation in biological quality. The prediction errors must be conditional on the sub-set of environmental features used to derive the predictions and the environmental and biological condition of each site in RIVPACS at each time of sampling.

This leads us to take the viewpoint that the RIVPACS prediction system is part of the definition of biological condition and as such has little or no "true" error and <u>no</u> assumed error in assessing the precision of EQI's derived from the system.

Given the assumption that RIVPACS predictions are made without system errors the remaining sources of variation in the macro-invertebrate data were taken to be due to sampling variation and error and inefficiency and error, particularly bias, in the sorting and identification of captured specimens. Variation in environmental data was taken to be variation in field measurements and the abstraction of cartographic data from the maps.

6.3 <u>Sources of variation and error in biological sampling</u>

The findings of the analyses of sampling variation were almost entirely encouraging. All sources of variation and error were analyzed with respect to their impact upon the derived BMWP index values. For each form of the index, number of taxa, BMWP score and Average Score per Taxon (ASPT) the experimental data were examined for the independent contribution of site type, biological condition, seasons and the people taking the sample to the total variation of the estimates of index value.

No conclusive evidence emerged from the study that variance in the data varied consistently or was statistically correlated with the type of site, its biological condition or season of sampling. The only suggestion otherwise was that there was more sampling variance in the estimated ASPT values of sites selected from RIVPACS groups 3a (small sites of about 7.5m width, at mean altitudes of 75m, with a cobble and pebbly substrata and a moderate mean alkalinity of 80.8mg l^{-1} $CaCO_3$) and 9b (larger, lowland sites about 13m wide and at about 5m altitude and with a mean alkalinity of 170.5mg l^{-1} $CaCO_3$).

Even at these site types, however differences were only significant for two seasons combined samples. Given that the number of sites analyzed in this study was limited by the size of the project budget and was close to the minimum needed to obtain the necessary statistical data, this single example of statistical differences in the variance of the estimate of this index alone, for just a single seasonal combination, may be due to chance factors alone. Too much significance should not be attached to the result but further investigations of this type of site may be warranted.

Furthermore, although the sixteen sampling locations covered a wide range of site types included in RIVPACS, not every type of site could be covered and, for example headwater streams and deep, slow flowing rivers such as the lower reaches of the Thames were not included in the data-base. It is possible that these extreme site types may also have exceptional patterns of variance in estimated BMWP index values due to sampling variation and errors.

R&D Note 412

Within the sites actually samples, one of the most important findings was that the inter-operator contribution to the total sampling variation contributed only a very small percentage of extra variance in the estimates of BMWP indices. Differences, relative to the sum of the other sources of variation were, in fact, so small as to be regarded as negligible. This has major implications for the NRA because the cost implications and the sheer practicality of having to use the same people to sample each site in each year to eliminate inter-operator effects are impossible to meet. However, this conclusion is entirely based on the assumption that the people collecting the sample are fully trained and competent. Sample collection is a skilled process and the performance of an untrained operator are likely to be of a lower standard than a trained one.

Equally important to the NRA is that the sampling variances due to all sources of variation and error can be summarised by simple constant values. Thus for number of taxa and BMWP the square root of the index value can be assumed to have a constant sampling variances which are independent of site type, biological condition, the sample collector or the season of collection. However the value of the constants are dependant on the number, but identity, of seasons used alone (single seasons) or in combination (paired or triplicate) to provide the index values and also the type of index (number of taxa or BMWP score).

Similarly, the sampling variance in the observed values of ASPT can best be estimated by a constant variance term which is again only dependant in almost all cases on the number of single samples used alone or in combination to estimate the ASPT for the site.

The ability to represent sampling variation by a set of constant variance terms greatly simplifies the procedures required to estimate the precision of EQI estimates and the probabilities that a site is properly classified into one of a series of quality bands.

As a ancillary investigation, the relative magnitude of variation in the annual BMWP index values based on mean values of individual season's samples and on combined faunal lists from all seasons samples was investigated. It was shown that averaging single values led to a higher standard deviation, due to sampling, in the estimate of the mean observed value than obtained by combing samples. Although this is not conclusive evidence for preferring the latter approach, it does lend more weight to the recommendation that samples should be combined, rather than average, for reporting annual biological condition in national and regional surveys (Clarke et al 1994).

6.4 Sources of error and bias due to sample sorting and the identification of specimens

Every year since 1990 the NRA have been sending a randomly selected set of macro-invertebrate samples to IFE for auditing of the Authority's efficiency in sorting samples and identifying the specimens removed. Results of the audits have been reported back to the NRA (eg Gunn et al 1991) and this has led to a marked improvement in standards of performance. This has partly been facilitated by the setting-up or improving of internal NRA audit procedures. However the existence and nature of these audit procedures has varied from region to region and area to area.

As part of the current research programme new analytical control procedures have been recommended which include recommendations of audit procedures and acceptable levels of performance (van Dijk 1994). It is probable that these procedures will be implemented in every

R&D Note 412

NRA laboratory and that the target level of performance for each NRA region will be, on average, no more than two missed taxa per season.

Whereas the mean level of performance across the NRA as a whole was 2.7 missed taxa in the 1990 samples audited by IFE, this had improved to just 1.5 in the 1992 audited samples. As these two years represented the worst and best levels of performance of the four (1990-1993 inclusive) available for analysis in the current research programme, a stratified, random sub-set of about 200 samples per year were selected from each of 1990 and 1992 to investigate the effects of missing (including mis-identifying) taxa on estimated values of the BMWP indices for the sites in single and multiple-seasons samples. The three-dimensional stratification matrix (Table 3.2) comprised NRA regions by season by number of taxa estimated to be present by the NRA.

In order to undertake analyses all the NRA and IFE taxon lists for the audited samples had to be entered onto computer and this provided an opportunity to check the accuracy of both the NRA's and IFE's initial data-entry. In both cases a very low error rate of 1% of entries or non-entries was shown to exist for both organisations. Although this figure is relatively small it is advantageous to eliminate it by double typing of all data-entries.

The existence of the audit data-base enabled a range of relevant queries to be made concerning including the frequency of losses and gains of each BMWP family, rates of recovery of missed taxa in other samples from the same site, the total numbers of taxa and the effects of missing taxa on the index scores for sites.

The types of taxa most frequently missed fell into several categories. Firstly, several families of small black or dark beetles were commonly overlooked, including Hydrophilidae, Elmidae, Scirtidae and Haliplidae. Secondly and similarly, other inconspicuous taxa, in relation to background sample material, such as Hydrobiidae, Sphaeriidae, Caenidae and Hydroptilidae were commonly missed. Thirdly were a series of taxa which could be confused with each other or similar taxa were also missed. The latter included Dendrocoelidae (confused with Planariidae?), Lymnaeidae and Physidae (confused with each other?), Valvatidae and Planorbidae (confused with each other when small?), Taeniopterygidae (confused with Nemouridae?), Psychomyiidae (confused with Philopotamidae?) and Goeridae, Lepidostomatidae, Beraeidae and Brachycentridae (confused with each other and other cased caddis families?).

Having recognised these problem taxa, which may be regional in their inefficiency of recognition, then NRA should take steps to ensure that the additional necessary training is provided to improve overall sorting and identification efficiency.

Similarly there are a group of taxa which NRA tend to erroneously record as present. Four of these. Dendrocoelidae, Physidae, Valvatidae and Lepidostomatidae have already been considered to be subject to the confusion errors. The same applies to a fifth family, Mesovelidae which is relatively rarely present in samples and is commonly confused with immature stages of Veliidae which is curiously not included in the BMWP system. The sixth taxa commonly recorded by the NRA but not found by IFE is Gerridae. It is likely that these taxa are seen but not captured at site, recorded as present but no specimen provided to IFE and no explanation of the sighting but lack of capture.

The rates of recovery of missed taxa could only be ascertained for 1990 when NRA's overall sorting and identification efficiency was at its poorest of the years audited. This was the only year for which the IFE held data for all survey sites for all seasons of sampling. The rates of recovery of missed taxa in the other seasons was surprisingly low, with only 50% of missed taxa recovered in a

R&D Note 412

second sample and only 63% in either of two other seasonal samples. Theoretical estimates for 1992, when the NRA processed samples more efficiently and missed fewer taxa were a 49% recovery in a second sample and 63% in two others.

It is postulated that the taxa most often not recovered are high scoring insect families which have seasonal aquatic stages or occur in relatively small numbers even when present. This is corroborated by the average BMWP score of those taxa missed which is consistently over five and often over six when ten or more taxa were originally recorded as present by the NRA.

The higher rate of missing taxa present in a sample than erroneously claiming taxa to be present which were not actually there led to an inevitable under-estimation and hence bias in the number of taxa and BMWP scores of single samples or combinations of multiple samples. There was also a slight but consistent tendency for there to be an under-estimation bias in ASPT's resulting from the missed taxa tending to have a higher average score than those successfully removed from the sample and correctly identified.

Knowledge of the overall rate of missing taxa and the subsequent rates of recovery enable a set of generalised correction factors to be made in order to compensate for poor processing ability. The algorithm for adjusting observed number of taxa in the annual combined seasons sample contains two principal functions, a value for the rate of recovery of missing taxa which is separately fixed for either two or three seasons sampling and a value for the average number of missing taxa per sample being achieved by the NRA nationally or, where different from the national average, regionally.

Knowledge of the relationship between the number of taxa and their mean ASPT (Figure 3.3) allows correction factors to then be calculated for the under-estimation of annual BMWP scores and ASPT's calculated by combining multiple (two or three) seasons samples.

Although the procedures for estimating for bias in BMWP indices of combined season samples are relatively straight forward the decision about whether to apply them is less simple.

The principle underlying RIVPACS is that, in estimating site quality, like should always be compared to like. This means that the observed values for a site should be compared with the expected values for the same site based on the same season or seasons of sampling and the same standardised sampling techniques. It is also important that the efficiency of sorting and identification which generates the observed values should be the same as that which applied to the generation of the expected values.

In under-taking routine audits for the NRA, the IFE staff re-sorting and identifying the NRA samples are the same people applying the same sorting procedures and using the same keys as sorted and identified the majority of sites in the RIVPACS system. Experience of the audits over the first four years is that IFE tended to find, on average two more taxa per sample than the NRA for the same sample. This figure fell to approximately 1.5 extra taxa per sample in NRA's best year, 1992. The NRA target likely to be set for the 1995 River Quality Survey is no more, on average, than two missed taxa per sample. This means that in calculating the annual site EQI's by comparing NRA observed values with IFE (ie RIVPACS) expected values the former is likely to represent a greater degree of under-estimation than the latter.

On the continued assumption that IFE maintained the same efficiency in processing the samples used in RIVPACS as it did in auditing the NRA samples then this tendency for observed values to be subjected to greater under-estimation than the expected is independent of IFE's own rate of missing taxa, which has not been tested. In other words, the principle that RIVPACS predictions

R&D Note 412

are taken as the absolute standard of good biological conditions still applies.

What is important, therefore, is not whether there are errors in the RIVPACS data or model but how the performance of the NRA or other agency compares with that used to construct RIVPACS. On that basis there are obvious grounds for adjusting NRA's annual BMWP index values to compensate for the differential in their sorting and identification efficiency and that applying to the samples on which RIVPACS is based.

Set against this are three disadvantages in applying correction factors. The first is the impression given by adding bias-correcting values may be perceived to be on a par with the use of "fudge factors" to "massage" the average biological condition of rivers upwards. Whatever the justification for doing so this may be perceived to be politically unwise. The second is that a belief amongst NRA biologists that their inefficiency in sorting and identifying taxa can be compensated for by applying correction factors may be counter-productive to a general drive towards raising standards of common efficiency within the organisation. The third and perhaps most important disadvantage is that, whatever the overall rate of missing taxa may be there could be widely differing rates for individual samples and applying a common correction factor may lead to the condition of some sites being over-estimated whilst failing to provide adequate compensation for the most poorly processed samples.

The decision is therefore a complex one which must ultimately be taken by the NRA in consultation and common practice with the other organisations involved in the 1995 survey, the River Purification Boards and the Department of the Environment (Northern Ireland).

6.5 Sources of error in the measurement of environmental variables

In an earlier report to the NRA, Clarke et al (1994) recommended that environmental variables should be measured with such precision that the variance in the estimate of their mean value should not lead to a variation of more than 0.02 in the EQI for number of taxa or 0.01 for the EQI for ASPT for any season or seasons when the value of that variable alone was allowed to vary in the RIVPACS predictive equation.

On this basis they set out targets for the permitted variation, expressed as percentage standard errors (%SE) or percentage co-efficients of variation (%CV), in the estimate of the mean values of each variable. Target values set out in that report are reproduced as Table 4.1 of the current document and have been used to judge whether inter-operator variation in environmental measurements in a single calender year fall within the recommended range.

Generally this was found to be the case although there were a number of exceptions. In particular the estimation of percentage cover of fine substratum, sand, silt and clay appeared to be problematic and lead to greater inter-operator variation. The definition of these terms may not be fully understood and the NRA should consider steps to improve their staff's ability to recognise these different substrata correctly and consistently.

Other sources of inter-operator variation indicated the types of error that even the most experienced recorders can make. Particularly common appeared to be recording both field and map measurements in the wrong units or using the wrong scale and mis-locating sites on the wrong channel or the wrong position in relation to confluences when reading discharge values. In order to

R&D Note 412

reduce many of these errors in future it is recommended that all map-derived variables be measured completely independently by two people. When estimates are close then the average value should be used but large differences should be investigated jointly and a common decision arrived at.

Variation between people in measuring environmental variables led to variation in between the expected EQI values derived from their environmental measurements. However the between operator variation in the expected numbers of taxa, BMWP scores and ASPT's did not seem to vary according to whether single or combined seasons predictions were made. On this basis it is recommended that for year-specific predictions of BMWP index values the errors in those values due to the measurement of environmental variables are assumed to have constant standard deviations irrespective of the number of seasons involved in predictions. These constant values are provided for each index.

The key phrase in the preceding paragraph is year-specific predictions. The analyses undertaken here have concerned themselves exclusively with inter-operator differences over a range of site types with all operators recording time-variant factors like width, depth and substratum effectively simultaneously in each sampling season.

Temporal differences have not been considered but variables such as width, depth and even substratum, together with the only chemical predictor variable in RIVPACS, alkalinity, can all vary considerably both within and between years. It is for this reason that Clarke et al (1994) recommended that the ultimate objective of the use of RIVPACs should be to provide fixed predictions of the BMWP index values of each site based on their average prevailing environmental characteristics and independent of the conditions prevailing at the time of biological sampling. In that way the effects of other stresses, such as atypically low-flows, may also be evaluated. Fixed predicted values of BMWP index values depend on fixed estimates of the "true" mean of each predictor variable.

Further research is required into the effects of temporal variation on the precision of estimation of these fixed mean variable values and the amount and quality of data required in order to derive them with a tolerable degree of variance about the mean.

6.6 Application of variance terms for the assessment of the biological condition of sites

The derivation of variance terms for errors in the acquisition of the macro-invertebrate and environmental data needed to assess the biological condition of sites enables the probabilities of mis-classification of sites to be implemented in a comparable manner to the application of CLAM (Warn 1990) to chemical data. The algorithm involved may or may not, at the discretion of the NRA, include a set of correction factors to compensate for sorting and identification errors.

The recommended procedure is to use Monte Carlo techniques to generate a set of simulated Ecological Quality Index values which are dependant not only upon the NRA's observed BMWP index values and the expected index values derived from the NRA's measured environmental values but also upon the variance (and bias) functions associated with the acquisition of each.

In this way an EQI frequency distribution will be created for each BMWP index for a site. According to whatever scheme of classification is in place, based upon EQI values, the probability that each site belongs to each band of the classification can be derived from the number of simulated EQI values which fall within each band range. A thousand simulations per site are recommended.

R&D Note 412

This form of simulation technique is applicable to any division of EQI values into bands and any combination or integration of the EQI's for number of taxa, BMWP score and ASPT. They are however dependant, at present, on the use of EQI's as the basis of the banding system.

As an ancillary product of this study, the procedures derived here may also be used to detect whether there are statistically significant differences between the EQI values of samples collected at different times from the same site or at different sites at the same or different times. Separate algorithms are proposed according to the combination of temporal and spatial comparisons being made.

Whatever form of comparison is being made the principal under-lying the test is the same. Monte Carlo techniques, incorporating the appropriate variance terms, are used to create a set of simulated observed and, where appropriate, expected index values. A standard number of paired simulations (say 1000) are made for each sample being compared and a value calculated for the change in EQI for each pair of simulations.

From this a frequency distribution can be calculated for the simulated differences in EQI's between sites and this, in turn, can be used to derive confidence limits for the differences in EQI between sites. If the confidence limits range includes zero then the difference in EQI's between the samples is not statistically significant and no change or difference in site quality can be assumed. If the confidence limits have the same sign then the sample EQI's can be considered to be significantly different at the probability level used to calculate the confidence limits. The sign of the confidence limits indicates the direction of change between the two samples.

6.7 The significance of this study to the future of biological monitoring

The results of this study have potentially very important implications for the future of biological monitoring within the NRA.

For the first time realistic error terms can be applied to all aspects of the collection of the biological and environmental data used to assess the biological condition of rivers. If promoted properly within and outside the water industry, this should inevitably lead to a heightened perception of the reliability of biological data for national and regional quality assessment, pollution studies and, perhaps most importantly of all, for setting Statutory Water Quality Objectives (National River Authority 1991b) if or when these become mandatory.

In terms of national River Quality Surveys the biological and chemical data can now be subjected to comparable forms of analysis to detect their reliability for quality class allocation. Only the practical examination of real data will tell how well the reliability of the two sources of data compare.

For the investigation of pollution incidents, the appraisal of the impact of effluent discharges and the evaluation of the effectiveness of anti-pollution measures the statistical techniques proposed here offer a genuine method of comparing whether the biological condition of the case sites, as expressed by their EQI's, are statistically significantly different.

R&D Note 412

Finally, for Statutory Water Quality Objectives, the simulation techniques may be used to establish whether the biological condition of a site, as represented by its macro-invertebrate fauna, genuinely exceeds the minimum acceptable standard set for a site.

However, in conclusion it should be noted that the procedures outlined here are dependant on the use of RIVPACS to generate Ecological Quality Indices to represent the biological condition of sites and that quality classification of sites is based upon EQI value ranges. In the process the necessary assumption is made that the RIVPACS predictions are intrinsic to the definition of that site quality and are assumed to be generated without error.

The RIVPACS software is thus an obvious medium for incorporating the full set of procedures recommended in this report. At present it is recognised that the statistical analyses presented in this report may not be easily assimilated or applied by all readers. There is a clear need to develop the reported findings as an operational system supported by a clear descriptive which incorporates well presented worked examples of the procedures involved in all stages of error estimation, site classification and statistical comparison of the biological condition of sites.

R&D Note 412

7 CONCLUSIONS AND RECOMMENDATIONS

7.1 Conclusions

1. Sixteen river sites covering a wide and balanced range of environmental conditions and biological condition were sampled in each of three seasons. On each occasion two replicate samples were taken by one IFE person and a third by a local NRA person. From this information, estimates of sampling variation in the observed numbers of BMWP taxa, BMWP scores and ASPT values were obtained.

Sampling variation

2. There was no strong evidence that the sampling variance for each of number of taxa, BMWP score and ASPT in single, two or three seasons combined samples varied consistently with site type or season(s).

3. Using different people to sample does not increase the sampling variance by more than a few percent. Using the same person over time is thus not important but all samplers must be trained to the same acceptable level. A single variance estimate based on a mixture of replicate samples from the same and different people was therefore derived for each situation.

Sample processing errors

4. The NRA sample processing errors in 1992 led to an average under-estimation of the number of taxa present in single season samples of about 1.5 taxa or 1.0 taxa if five or less taxa were recorded as present (see section 3).

5. Based on 1990 data, only about 50% of cases of taxa missed in one season's samples were not found in a second season's sample from the same site and only 37% are not recorded in at least one of the single samples from the other two seasons. This means that the effect of sample processing errors is to under-estimate the number of taxa in a sample by about the same number, irrespective of whether the sample is for a single season or comprises two or three seasons combined faunal lists. The bias in estimated EQI due to NRA sample processing errors is therefore less when using two and three seasons combined lists because combined samples have higher expected numbers of taxa.

Errors in environmental predictor variables

6. The variation between recording personnel in their estimates of both the map and field derived environmental variables is generally small enough to

R&D Note 412

ensure that the resulting error variation in the RIVPACS expected values is within tolerable limits.

7.Most care is needed in estimating the distance from source of sites and in estimating the mean substratum particle size for sites with predominantly fine substratum particle sizes.

8.The between-operator variance in expected numbers of taxa, BMWP score and ASPT values is about the same for single and combined seasons expected values.

9.It is not possible, within this research programme, to estimate the accuracy and effects on RIVPACS predictions of using a single year's three seasons mean values to estimate the long-term mean values of each environmental variable for a site.

7.2 **Recommendations**

1.The combined effect of the various sources of error in the observed and expected number of taxa, BMWP score and ASPT on the precision of the EQI value for a site should be assessed by Monte Carlo simulation. Detailed simulation procedures are given in section 5.

2.The square root of the observed number of taxa in a sample should be assumed to have a constant sampling variance, $VAR_{\sqrt{T}}$, which depends on the number of seasons samples combined, as detailed in Table 2.15.

3.The square root of the observed BMWP score for a sample should be assumed to have a constant sampling variance, $VAR_{\sqrt{S}}$, which depends on the number of seasons samples combined, as detailed in Table 2.15.

4.The sampling variance of the observed ASPT in a single or combined season sample should be estimated by a constant, VAR_A, which is independent of the value of ASPT or the number of taxa in the sample, but depends on the number of seasons samples combined, as detailed in Table 2.15.

5.The effect of NRA sample processing errors in 1995 and subsequent years is assumed to be similar to that estimated for 1992. Therefore the under-estimation of the number of number of taxa should be corrected, if required by the NRA, using the procedures described in sections 3.7 and 3.8.3.

6.All map-derived variables, including National Grid references, should be measured and recorded completely independently by two people. Any large differences need to be explained and only the most appropriate value used. Where two estimates are in close agreement their average value should be used.

7.It is recommended that, for year-specific predictions of the BMWP index values for a site, the errors in the expected values due to measuring the environmental variables should be assumed to have the following standard deviations, irrespective of whether predictions are for single season or two or three seasons combined samples :

Error SD of Expected :number of taxa= 0.53

$$\text{BMWP score} = 4.3$$

$$\text{ASPT} = 0.081$$

However these values may not accurately represent the errors if these one-year means are used as estimates of the variation in predictions derived from long-term, fixed, mean values of each environmental variable for each site.

8.The simulation procedures given in section 5.3 should be used to estimate the precision of any EQI value and, hence, the probability that site quality has changed between two surveys. Suggestions for using the simulations to estimate the rate of mis-classifying sites into quality bands are also given.

7.3 Recommendations for further research

1.The findings of the current study are of major importance to the way in which the NRA collect and interpret macro-invertebrate data. For the first time the major errors, variation and bias in all stages of the collection of data used to assess the biological condition of sites have been quantified. This knowledge has allowed procedures to be developed for determining the rate of misclassification of sites in local, regional and national surveys and for identifying statistically significant differences in the condition of sites at different times and/or in different places.

However the analyses and procedures provided in this document are academic in their presentation. In order to be of value to the NRA they needed to be developed as practical operational techniques and their application to be tested and, where necessary, improved. The most appropriate medium for operational use is RIVPACS because this system not only requires macro-invertebrate and environmental data for its current purposes but also outputs data in the form of Environmental Quality Indices. It is these indices which form the basis of the statistical tests of significant differences that are recommended here.

If the techniques are to be implemented in RIVPACS then they need to be accompanied by a comprehensive but comprehensible user manual outlining the application and interpretation of the significance test and illustrating this with worked examples that provide sufficient information on the way in which biological and environmental error terms are calculated and integrated for the estimation of the precision of EQI's and their subsequent use in the statistical tests.

Further NRA funding is recommended in order to undertake this development of the RIVPACS software and to maximise the practical benefits of the findings of the current study.

2. The only major source of quantifiable variation not examined in this document is temporal variation in the estimation of the "true" mean value of each field measured environmental variable, width, depth and substratum composition.

Whilst acceptable levels of inter-operator variation were achieved in the experimental studies reported upon here, very much greater variation might occur if each recorder obtained their environmental information on different days of the same seasons. The incidence of droughts, spates, reservoir releases and even normal seasonal hydrological cycles could have major impacts on the values of these variables at different stages of the same season. Inter-annual differences are likely to be as great or greater than intra-seasonal ones.

In the Interim NRA R&D Report 243/7/Y it was recommended that the expected (RIVPACS predicted) values of numbers of taxa, BMWP scores and ASPT's should be fixed for a site according to the season or seasonal combination of samples being considered. These require fixed values of the "true" means for each environmental, predictor variable for each site. General, but not precise recommendations on the acquisition of these fixed means were set out in Report 243/7/Y.

It is recommended that research be undertaken in order to identify the best procedures for acquiring long-term mean values of each environmental variable used in RIVPACS to provide fixed predictions of the expected BMWP index values of sites in any season or seasonal combination.

3. In the current study there were indications that the variation due to sampling in the observed values of ASPT was greater at some sites than others (site types 3a and 9b, section 2.4.2).

It is recommended that further research is undertaken to determine whether certain sorts of sites and within-site habitat variation are intrinsically more variable than others and what are the specific

R&D Note 412

circumstances that induce this variability.

8REFERENCES

Biological Monitoring Working Party (1978) Final report: Assessment and presentation of the biological quality of rivers in Great Britain. December 1978.

Chandler, J.R. (1970) A biological approach to water quality management, Water Pollution Control, London, 69, 415-422.

Chesters, R.K. (1980) Biological Monitoring Working Party. The 1978 national testing exercise, Department of the Environment, Water Data Unit Technical Memorandum, 19, 1-37.

Clarke, R.T., Furse, M.T. and Wright, J.F. (1992) Testing and further development of RIVPACS. A comparison of single, paired and 3 season combined macro-invertebrate samples for the biological banding of river quality. A preliminary R&D report (243/2/Y) to the National Rivers Authority.

Clarke, R.T., Furse, M.T. and Wright, J.F. (1994) Testing and further development of RIVPACS. Phase II: Aspects of robustness. An interim R&D report (243/7/Y) to the National Rivers Authority.

Cleveland, W.S. (1979) Robust locally weighted regression and smoothing scatterplots, Journal of the American Statistical Association, 74, 829-836.

Council of the European Union (1994) Proposal for a Council Directive on the Ecological Quality of Water (94/C 222/06), Official Journal of the European Communities, No. C 222/6

Cox, R., Furse, M.T., Wright, J.F. and Moss, D. (1991) RIVPACS II. A User Manual. A report to the National Rivers Authority, the River Purification Boards and the Department of the Environment (Northern Ireland).

Department of the Environment and The Welsh Office (1971) Report of a River Pollution Survey of England and Wales 1970. Volume 1, London: HMSO.

Department of the Environment and The Welsh Office (1972) River Pollution Survey of England and Wales. Updated 1972. River Quality, London: HMSO.

Department of the Environment and The Welsh Office (1975) River Pollution Survey of England and Wales. Updated 1973. River Quality and Discharges of Sewage and Industrial Effluents, London: HMSO.

Department of the Environment and The Welsh Office (1978) River Pollution Survey of England and Wales. Updated 1975. River Quality and Discharges of Sewage and Industrial Effluents, London: HMSO.

Department of the Environment and The Welsh Office (1986) River Quality in

England and Wales 1985. A Report of the 1985 Survey, London: HMSO.

Furse, M.T, Fernandes, J.N., Wright, J.F. & Armitage, P.D. (1989) The use of aquatic macro-invertebrate assemblages for assessment of the environmental quality of rivers. In River Water Quality, Assessment and Management. Proceedings of an International Seminar, Lisbon, 17-19 May 1989, 171-181.

Furse, M.T. and Gunn, R.J.M. (1990) Pond-net techniques for sampling aquatic macro-invertebrates - standard methods for use by the National Rivers Authority. A training film for the National Rivers Authority, 45 minutes. London: Spectrum Communications.

Furse, M.T., Moss, D., Wright, J.F., Armitage, P.D. and Gunn, R.J.M. (1986) A practical manual for the classification and prediction of macro-invertebrate communities in running water in Great Britain. Preliminary version.

Graham, T.I. (1965) Biological surveys. Biological index of pollution. In Annual Report of the Lothians River Purification Board for the year ended 15th May 1965.

Green, R.H. (1979) Sampling Design and Statistical Methods for Environmental Biologists, New York: John Wiley & Sons.

Gunn, R.J.M., Wright, J.F, Blackburn, J.H. and Furse, M.T. (1991) Quality Audit of Samples for the 1990 River quality Survey,, NRA Southern Region. A report to the National Rivers Authority, Southern Region, 72pp.

Hellawell, J.M. (1986) Biological Indicators of Freshwater Pollution and Environmental Management, Pollution Monitoring Series, London: Elsevier Applied Science.

Hill, M.O. (1979) TWINSPAN - A **FORTRAN** program for Detrended Correspondence Analysis and Reciprocal Averaging, New York: Ecology and Systematics, Cornell University, Ithaca, N.Y.

Kinley, R.D. and Ellis, J.C. (1991) The Application of Statistical Quality Control Methods to Macroinvertebrate Data, National Rivers Authority Report № CO 2905

Metcalfe-Smith, J.L. (1994) Biological water-quality assessment of rivers: Use of macroinvertebrate communities. In The Rivers Handbook. Hydrological and Ecological Principles, Volume 2, edited by P. Calow and G.E. Petts, 144-170, London: Blackwell.

MINITAB (1994) MINITAB Version 10 for Windows reference manual (MINITAB package and manual available from UK distributors: CLECOM Ltd., The Research Park, Vincent Drive, Edgbaston, Birmingham).

R&D Note 412

National Rivers Authority (1991a) <u>The Quality of Rivers, Canals and Estuaries in England and Wales. Report of the 1990 Survey</u>, Bristol: National Rivers Authority

National Rivers Authority (1991b) <u>Proposals for Statutory Water Quality Objectives</u>, Bristol: National Rivers Authority

National Rivers Authority (1994) <u>The Quality of Rivers and Canals in England and Wales (1990 to 1992)</u>, Bristol: National Rivers Authority

National Rivers Authority (in preparation) A procedure for collecting and analysing river macro-invertebrate samples for RIVPACS application.

National Water Council (1981) <u>River Quality: The 1980 Survey and Future Outlook</u>, London: National Water Council.

Royal Commission on Environmental Pollution (1992) <u>Royal Commission on Environmental Pollution. Sixteenth Report. Freshwater Quality</u>, London: HMSO.

Siegel, S. (1956) <u>Non-parametric Statistics for the Behavioral Sciences</u>, New York: McGraw-Hill.

Sladecek, V. (1973) System of water quality from the biological point of view, <u>Archiv fur Hidrobiologie Beiheft. Ergebnisse der Limnologie</u>, 7, 1-218.

Sweeting, R.A., Lowson, D., Hale,P. and Wright, J.F. (1992) 1990 biological assessment of rivers in the UK. In <u>River Water Quality - Ecological Assessment and Control</u>, edited by P.J. Newman, M.A. Piavaux and R.A. Sweeting, 319-326, Luxembourg: Commission of the European Communities, publication number EUR 14606 EN-FR.

Taylor, L.R. (1961) Aggregation, variance and the mean, <u>Nature</u>, 189, 732-735.

van Dijk, P. (1994) <u>Analytical Quality Control for Macroinvertebrate Enumeration</u>, National Rivers Authority R&D Note 331.

Warn, A.E. (1990) Class Allocation Model. National Rivers Authority Report, Anglian Region.

Woodiwiss, F. (1964) The biological system of stream classification used by the River Trent Board, <u>Chemistry and Industry</u>, 14, 443-447.

Wright, J.F., Furse, M.T. and Armitage, P.D. (1993) RIVPACS - a technique for evaluating the biological quality of rivers in the UK, European Water Pollution Control, 3, 15-25.

R&D Note 412

Wright,J.F., Furse, M.T., Clarke, R.T. and Moss, D. (1991) Testing and further development of RIVPACS. An interim R&D report (243/1/Y) to the National Rivers Authority.

Wright,J.F., Furse, M.T., Clarke, R.T., Moss, D., Gunn, R.J.M., Blackburn, J.H., Symes, K.L., Winder, J.M., Grieve, N.J. and Bass, J.A.B. (in preparation) Testing and development of RIVPACS. A draft final R&D report (243/9/Y) to the National Rivers   Authority.
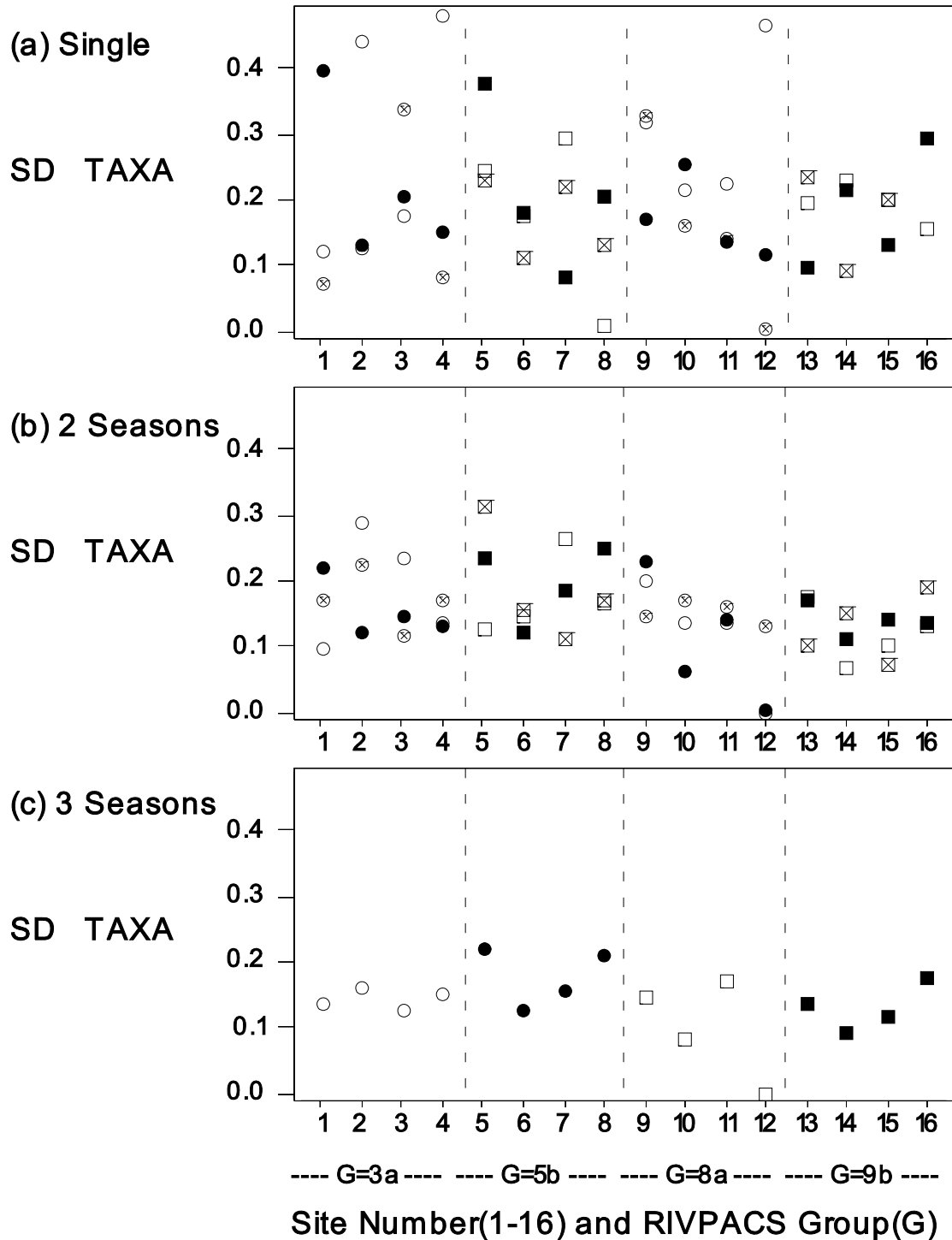
Figure 2.1 Left-hand side : Plot of the natural logarithm of the variance of the observed number of taxa versus the logarithm of the mean number of taxa observed in replicate samples for each of 16 sites. (a) Single season samples (each season's values for each site plotted separately), (b) Two seasons combined samples (each pair of seasons for each site plotted separately) and (c) Three seasons combined. Best fit regression lines of Log(Variance) versus Log(Average) are superimposed. Outlier values of zero variance which had been set to the minimum observed non-zero variance were excluded from the log-log regressions. Right-hand side : Plot of Res√TAXA (the deviation of the square root of the number of taxa in a particular single or combined seasons sample from the mean square root of the number of taxa in replicate samples from that site in that seasonal combination) versus the mean (=Average) number of taxa.
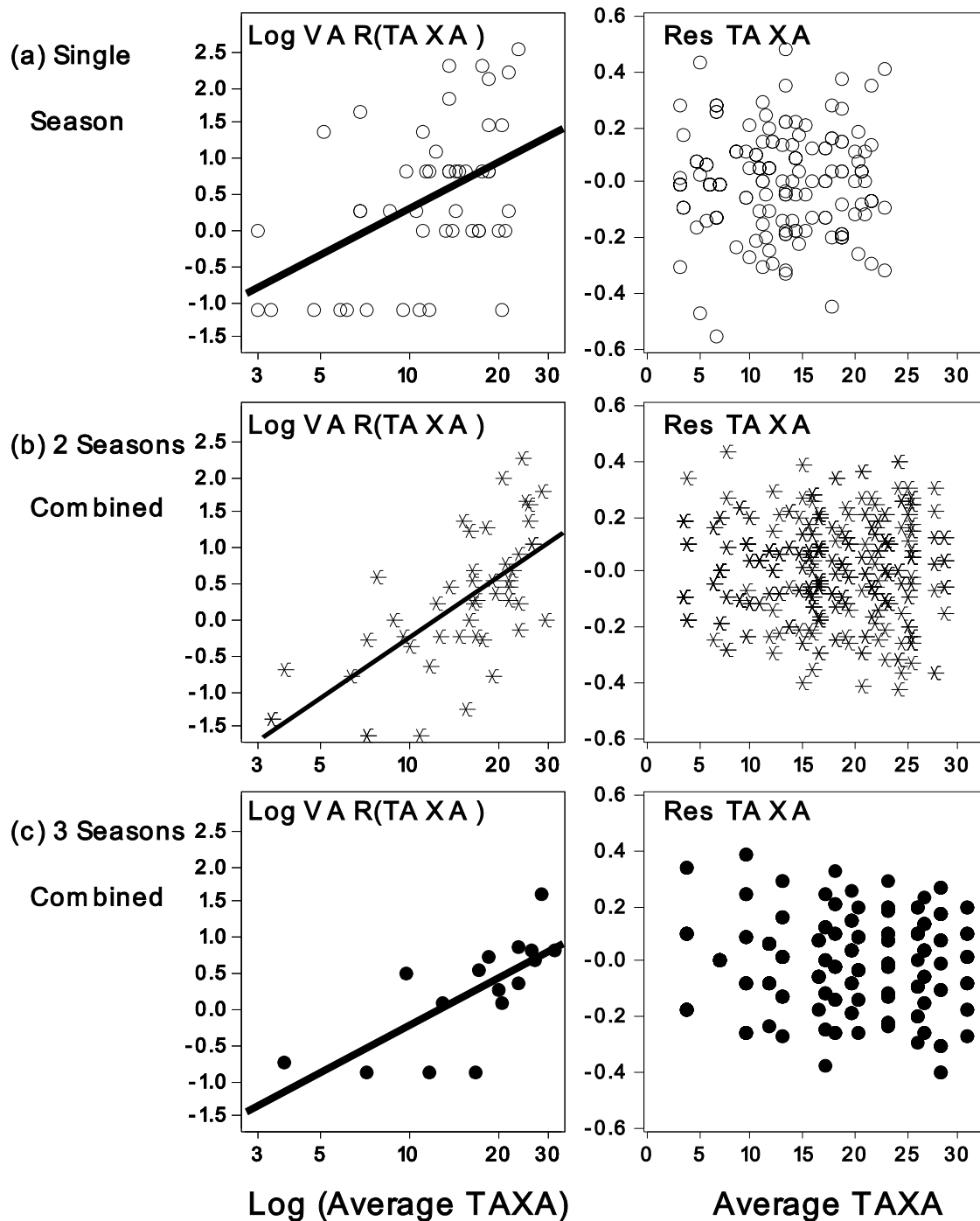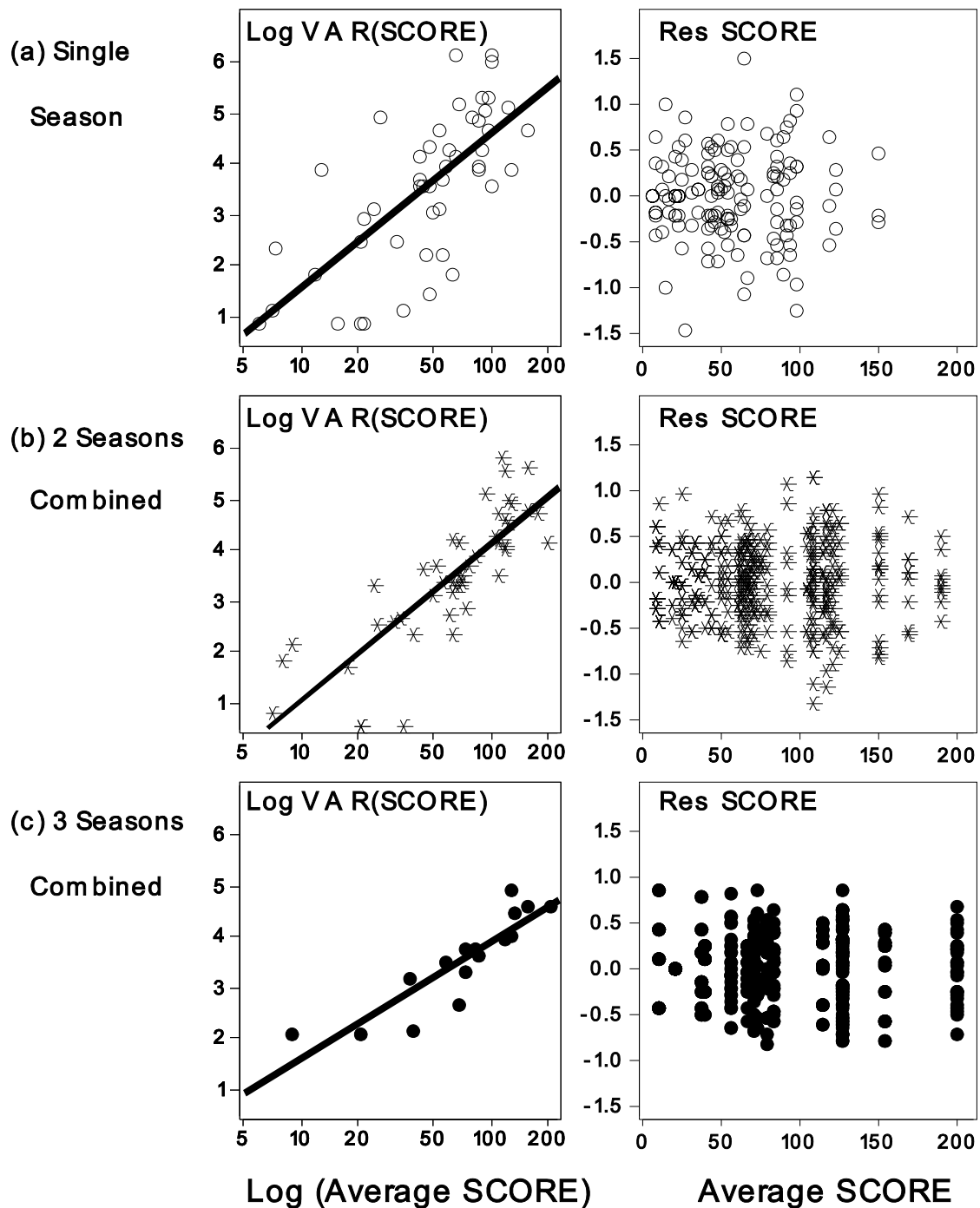

R&

Figure 2.2 Plot of the standard deviation (SD) of the square root of the observed number of taxa estimated from replicate samples for each of 16 sites (1-16). Vertical dotted lines divide the sites into their four RIVPACS site types (3a, 5b, 8a, 9b). (a) Single season samples (O▯=spring, O▯=summer and ●■=autumn values) for each site plotted separately, (b) Two seasons combined samples (O▯=spring/summer, O▯=spring/autumn, ●■=summer/autumn) plotted separately, (c) Three seasons combined values.

Figure 2.3 Left-hand side : Plot of the natural logarithm of the variance of the observed BMWP score versus the logarithm of the mean score observed in replicate samples for each of 16 sites. (a) Single season samples (each season's values for each site plotted separately), (b) Two seasons combined samples (each pair of seasons for each site plotted separately) and (c) Three seasons combined. Best fit regression lines of Log(Variance) versus Log(Average) are superimposed. Outlier values of zero variance which had been set to the minimum observed non-zero variance were excluded from the log-log regressions. Right-hand side : Plot of Res√SCORE (the deviation of the square root of the BMWP score in a particular single or combined seasons sample from the mean square root of the score in replicate samples from that site in that seasonal combination) versus the mean (=Average) BMWP score.

R&

Figure 2.4 Plot of the standard deviation (SD) of the square root of the observed BMWP scores estimated from replicate samples for each of 16 sites (1-16). Vertical dotted lines divide the sites into their four RIVPACS site types (3a, 5b, 8a, 9b). (a) Single season samples (O□=spring, O□=summer and ●■=autumn values) for each site plotted separately, (b) Two seasons combined samples (O□=spring/summer, O□=spring/autumn, ●■=summer/autumn) plotted separately, (c) Three seasons combined values.
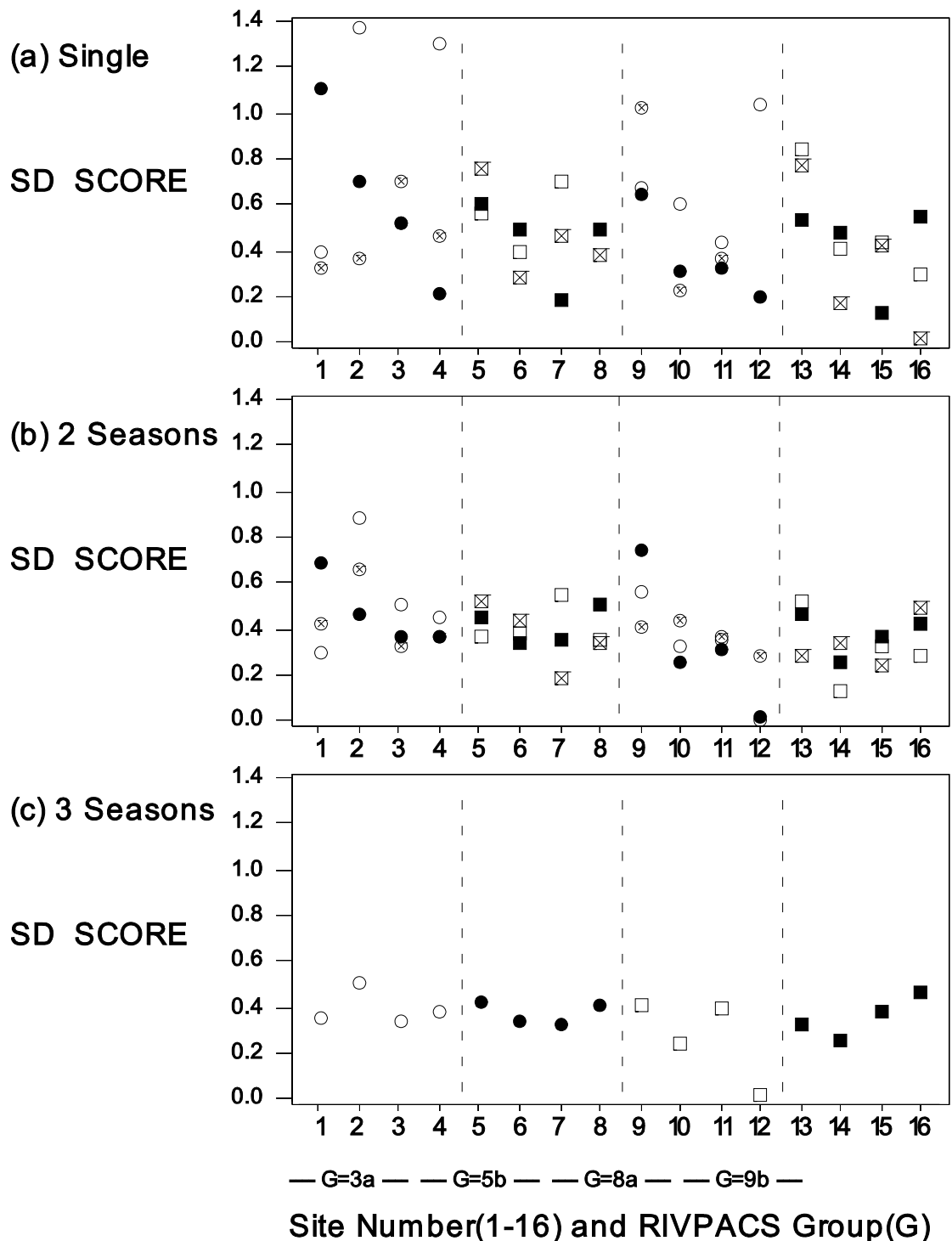
Figure 2.5 Plot of standard deviation (SD) of ASPT values versus the mean ASPT (left-hand-side) and number of taxa (right-hand-side) observed in replicate samples from each of 16 sites. (a) Single season samples (each season's values for each site plotted separately), (b) Two seasons combined samples (each pair of seasons for each site plotted separately), (c) Three seasons combined. Horizontal lines denote the best estimate overall SD (SD$_m$). For 1,2 and 3 seasons combined, SD$_m$ = 0.249, 0.161 and 0.139 respectively.

ASPT

TAXA

Average for Site/Season(s)
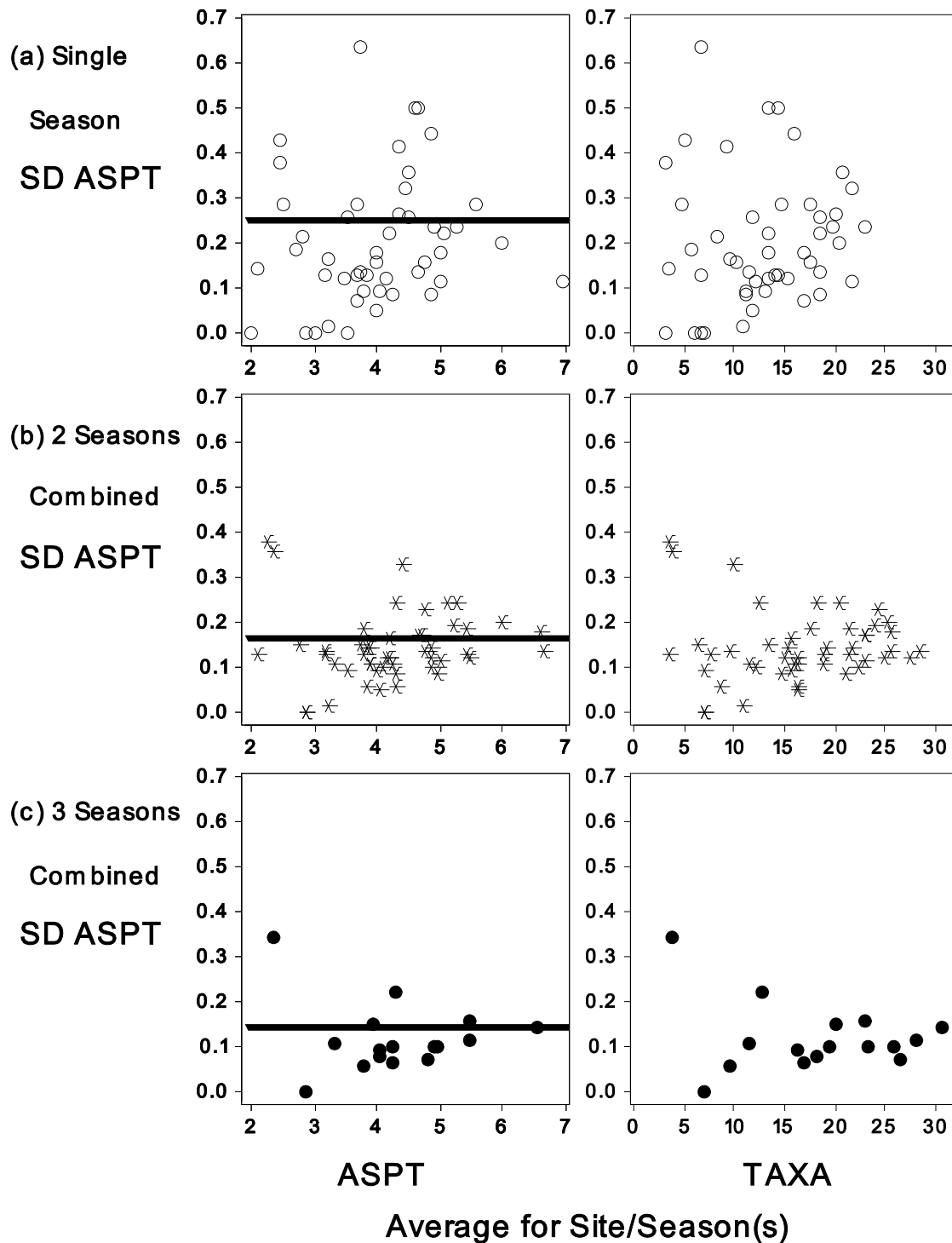
Figure 2.6 Plot of the standard deviation (SD) of the observed ASPT estimated from replicate samples for each of 16 sites (1-16). Vertical dotted lines divide the sites into their four RIVPACS site types (3a, 5b, 8a, 9b). (a) Single season samples (O☐=spring, O☐=summer and ●■=autumn values) for each site plotted separately, (b) Two seasons combined samples (O☐=spring/summer, O☐=spring/autumn, ●■=summer/autumn) plotted separately, (c) Three seasons combined values.
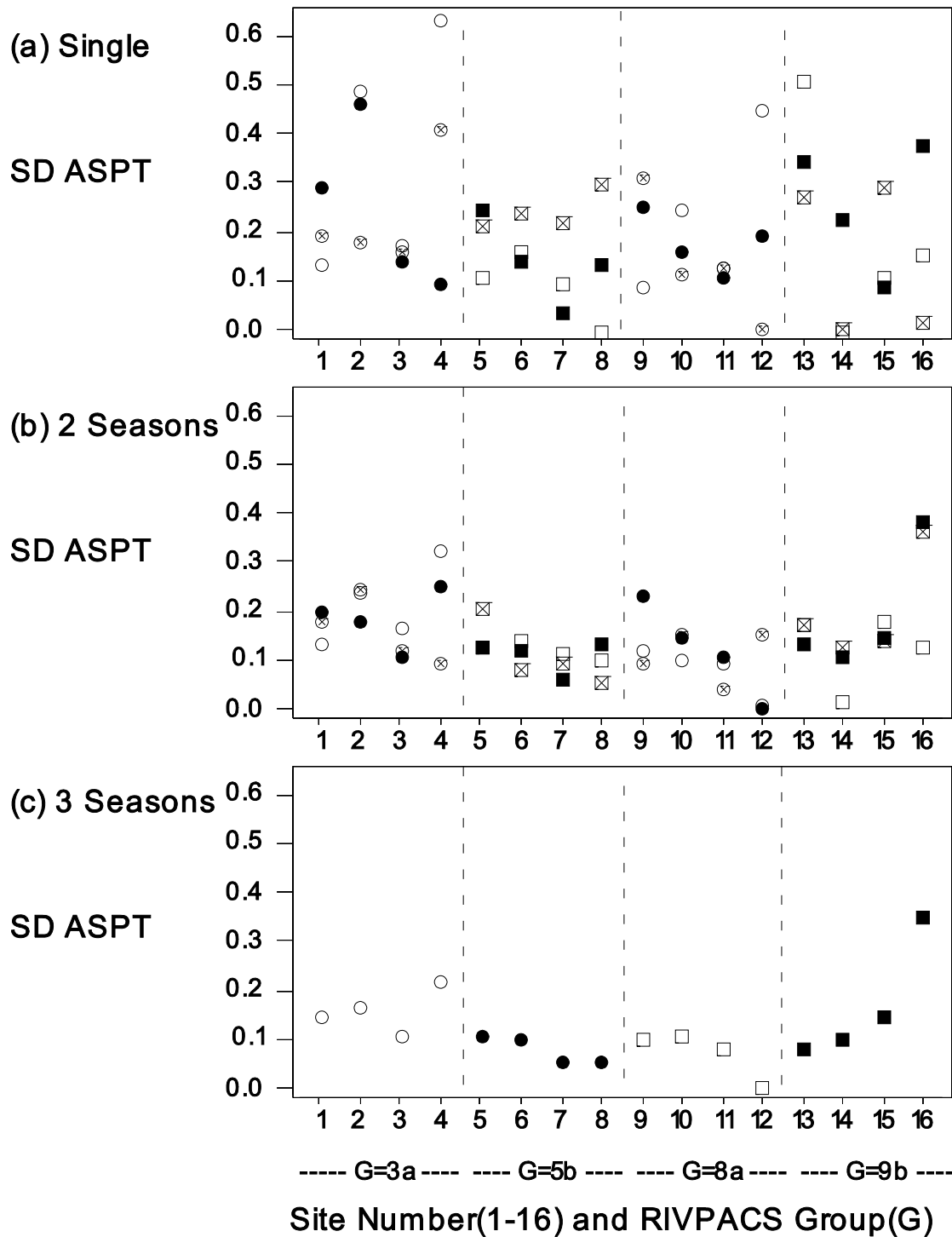
Figure 2.7 Plot of Res ASPT, the deviation of the value of ASPT for a particular single or combined seasons sample from the mean ASPT in replicate samples from that site in that seasonal combination, against the mean ASPT (left-hand-side) and the mean number of taxa on which the ASPT values for that site and season, or seasons, were based (right-hand-side).
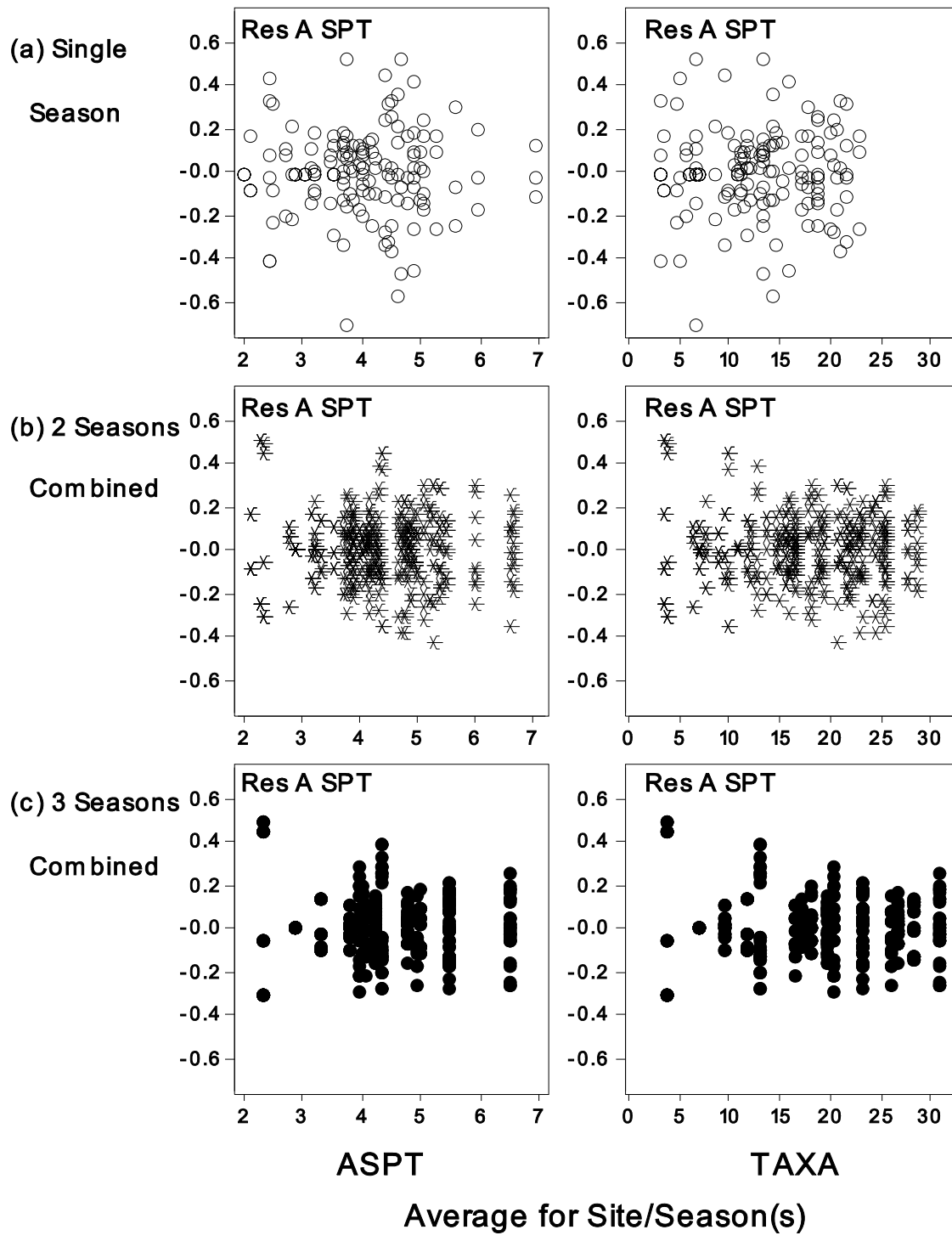
Figure 2.8 Frequency histograms showing overall statistical distribution of values of the residuals for (a) √TAXA, (b) √SCORE and (c) ASPT for single season samples (left) and three seasons combined samples (right). The residual for a sample measures the deviation of its value from the mean sample value for that site and seasonal combination.
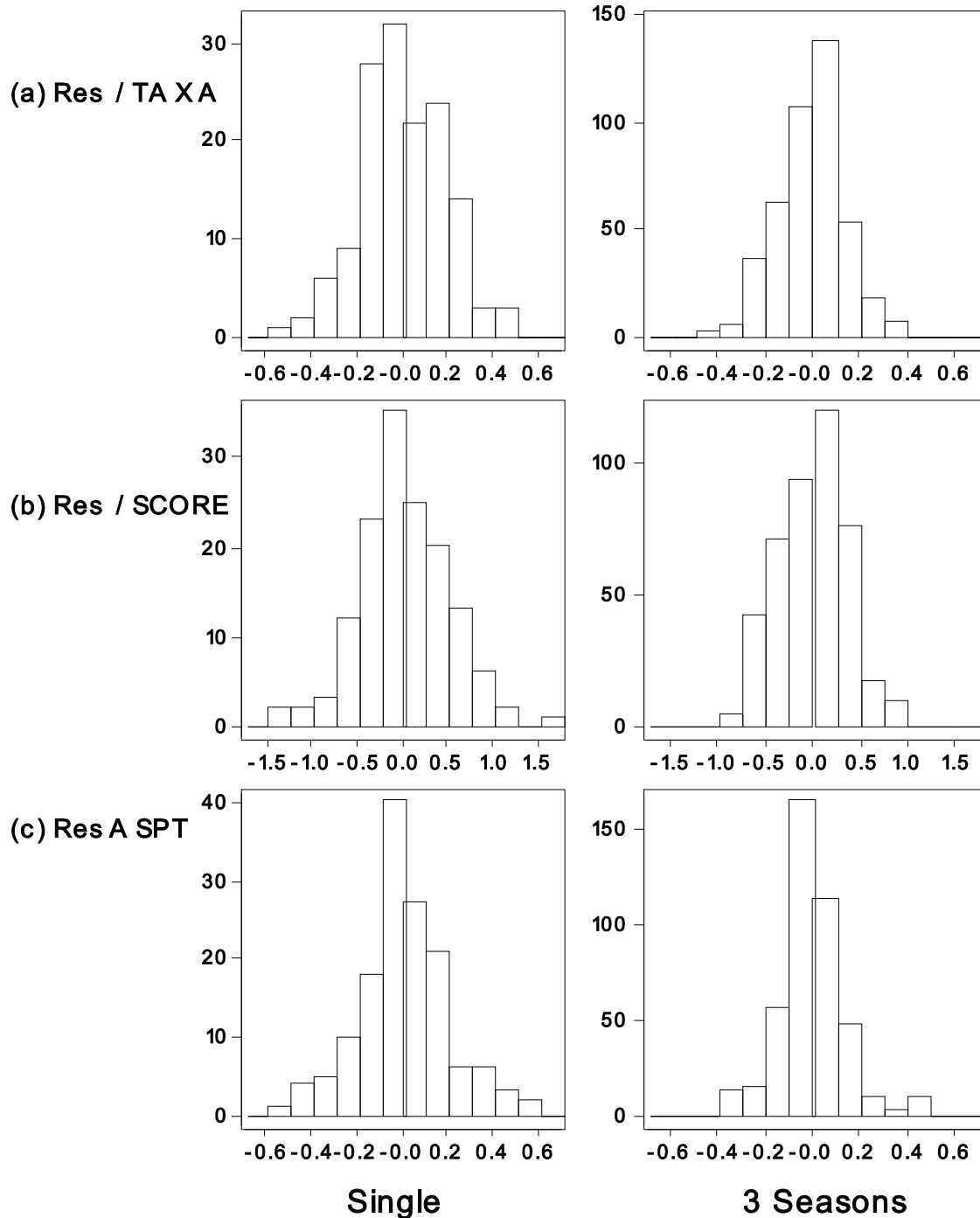
Figure 2.9 Correlation between replicate sample values of ASPT and number of taxa. Plot of Res ASPT against Res √TAXA. Res ASPT is the deviation of the value of ASPT for a particular single or combined seasons sample from the mean ASPT in replicate samples from that site in that seasonal combination. Res√TAXA is the equivalent deviation for the square root of the number of taxa observed in each sample. Separate plots for (a) single season samples (all seasons plotted together) (correlation r = 0.50), and (b) three seasons combined samples (r = 0.34).