Environment Agency

using science to create a better place

Lake benthic macroinvertebrates II: quantifying uncertainty in sampling methodology

Science Report: SC030294/SR2

SCHO0307BMHE-E-P

ea/br/e/sci/v1

The Environment Agency is the leading public body protecting and improving the environment in England and Wales.

It's our job to make sure that air, land and water are looked after by everyone in today's society, so that tomorrow's generations inherit a cleaner, healthier world.

Our work includes tackling flooding and pollution incidents, reducing industry's impacts on the environment, cleaning up rivers, coastal waters and contaminated land, and improving wildlife habitats.

This report is the result of research commissioned and funded by the Environment Agency's Science Programme.

Published by:
Environment Agency, Rio House, Waterside Drive, Aztec West, Almondsbury, Bristol, BS32 4UD
Tel: 01454 624400 Fax: 01454 624409
www.environment-agency.gov.uk

ISBN: 978-1-84432-704-1

This report is printed on Cyclus Print, a 100% recycled stock, which is 100% post consumer waste and is totally chlorine free. Water used is treated and in most cases returned to source in better condition than removed.

Further copies of this report are available from:
The Environment Agency's National Customer Contact Centre by emailing enquiries@environment-agency.gov.uk or by telephoning 08708 506506.

Authors:
I.J. Jones, R.T. Clarke, J.H. Blackburn, R.J.M. Gunn, N.T.Kneebone, & M.W. Neale

Dissemination Status:
Publicly available

Keywords:
Benthic invertebrates, littoral, sampling method, replicate, uncertainty, RIVPACS, Water Framework Directive, lake, lough, loch

Research Contractor:
CEH Dorset
Winfrith Technology Centre
Winfrith Newburgh
DORCHESTER
Dorset DT2 8ZD
Tel : 01305 213500 **Fax :** 01305 213600

Environment Agency's Project Manager:
Ben McFarland, Exeter Office

Science Project Number:
SC030294

Product Code:
SCHO0307BMHE-E-P

Science at the Environment Agency

Science underpins the work of the Environment Agency, by providing an up to date understanding of the world about us, and helping us to develop monitoring tools and techniques to manage our environment as efficiently as possible.

The work of the Science Group is a key ingredient in the partnership between research, policy and operations that enables the Agency to protect and restore our environment.

The Environment Agency's Science Group focuses on five main areas of activity:

- **Setting the agenda**: To identify the strategic science needs of the Agency to inform its advisory and regulatory roles.
- **Sponsoring science**: To fund people and projects in response to the needs identified by the agenda setting.
- **Managing science**: To ensure that each project we fund is fit for purpose and that it is executed according to international scientific standards.
- **Carrying out science**: To undertake the research itself, by those best placed to do it - either by in-house Agency scientists, or by contracting it out to universities, research institutes or consultancies.
- **Providing advice**: To ensure that the knowledge, tools and techniques generated by the science programme are taken up by relevant decision-makers, policy makers and operational staff.

Steve Killeen Head of Science

Acknowledgements

We are grateful for the input and guidance of the Environment Agency managers of this project, namely Angus Tree, Lindsey Powell and Ben McFarland. We also acknowledge the Environment Agency staff involved in the collection of the samples, and the ECUS Ltd staff involved in the sorting and identification of the samples.

Executive summary

The Water Framework Directive (WFD) requires that each Member State must assess and monitor the ecological status of its water bodies, including rivers and lakes – and that monitoring must use of a series of Ecological Quality Ratios (EQRs) based on a comparison of the observed to expected values of stress-sensitive biotic indices or metrics. Furthermore, the WFD requires that these EQRs are subdivided into five classes representing ecological status classes ('high', 'good', 'moderate', 'poor' or 'bad'). In light of this, the Environment Agency commited to examining its existing methods for assessing water bodies, leading to this report – the second reporting on this work.

While the first report in this series (CEH, 2006) reviewed existing sampling methodology and resulted in a series of recommendations, this second report specifically assesses the causes and levels of uncertainty inherent within the revised method. This work is important since the methods used for assigning an individual water body to an ecological status class are based on sample information and therefore subject to uncertainty. Specifically, the method relies on field sampling to estimate the values of individual EQRs and then using agreed rules to combine the information from different metrics into an overall assessment for the water body. Given the nature of sampling, this means that other replicate samples from the same water body at the same time could give rise to different values of one or more Ecological Quality Ratios (EQRs) and hence, possibly, even different estimates of the water body's status.

The aim of this study was, therefore, to assess the variation and uncertainty associated with a selected lake benthic macro-invertebrate sampling method, the ultimate goal being to assist in those tasked with developing tools that provide a sufficiently robust classification for assessing the ecological status of lakes.

Data provided by the Environment Agency – from samples collected using a balanced hierarchical replicated sampling design modified from the recommendations made by CEH (2006) – were examined for both inherent within- and between-site spatial heterogeneity that may change with season, and operator differences and subsequent laboratory sample processing procedures and errors. The sampling programme itself focused on acid-sensitive lakes and was designed to distinguish and quantify macroinvertebrate community variation due to each of the following sources:

i. among operators
ii. within station
iii. among stations within a lake
iv. across replicate lakes at the same level of pressure
v. across two levels of pressure, high (3 lakes) and low (3 lakes)

In addition, data from species-level macroinvertebrate samples that had been sorted and identified by ECUS were audited by CEH and analysed to assess the uncertainty associated with errors in sorting and identification.

Clearly, in order to be effective in discriminating the ecological status classes of lakes, the overall sampling error for estimates of metric values for any one lake needs to be small relative to the differences in metric values between the lakes – and especially between lakes of different pressure classes. A small 'percentage within-lake sampling variance' indicates high statistical precision and repeatability of results.

This study demonstrated that community composition of stations within a lake were, on average, always more similar to other stations from the same lake, than to stations from a different lake. It also revealed significant differences in metric values were found between

lakes **within** a pressure class, contributing 83%, 74%, 29%, 82%, 85% and 74% of the total variance of BMWP families, BMWP score, ASPT, AWIC families, AWIC total score and AWIC, respectively.

However, no consistent differences were found between lakes at high and low perceived risks of pressure from acidification. One reason for this could be that although considered to be highly susceptible to the effects of acidification, the lakes sampled may not have been affected at the time of sampling. Another possible explanation is that it may not have been possible to separate 'natural' acid lakes from those suffering from anthropogenic acidification.

The study does reveal that a considerable proportion of the variation within a lake is at the small-spatial, within-station scale and/or due to inter-operator effects. On average, the overall variance in ASPT within a lake contributed over two-thirds (71%) of the total variance in individual sample ASPT values. Meanwhile, inter-operator effects were responsible for broadly similar percentages of within-station sampling variability, ranging from 0% for ASPT and AWIC to 8% for BMWP score and 19% for number of BMWP families. These differences are small compared to pure replicate sampling variability. Inter-operator differences have no apparent systematic effect on either ASPT or AWIC.

The study also reveals that estimates of the biological status of a site generally become more precise the greater the sampling effort, and that for a given number total number of samples, the variance of the mean is always minimised by taking one sample from each station.

Finally, this study was able to demonstrate the effect of sorting or identification errors at family level on the variation. Sorting errors were responsible for 6-11% of the total variance, sample-processing errors causing an overall tendency to under-estimate the 'true' value of AWIC for a sample, but not ASPT. However, the effects of variation due to field sampling, sorting and identification errors can be incorporated into estimates of the uncertainty in bioassessments using the software package STARBUGS (STAR Bioassessment Uncertainty Guidance Software).

Contents

1 Introduction

This study addressing the requirement of the Water Framework Directive (WFD) (European Commission, 2000) that standing water quality should be assessed using biological elements, specifically benthic invertebrates. This report brings to completion the second phase of this project, which looks at the use of littoral and profundal benthic invertebrates for the classification of standing waters, and dicusses the development of tools for estimating water quality on the basis of benthic invertebrates.

- Phase I – reviews methods for sampling invertebrates and evaluates the suitability of the agencies' existing methods for producing data suitable for tool development (CEH, 2006);
- Phase II – looks at sample processing and statistical analysis in order to quantify uncertainty associated with the methods;
- Phase III – provides advice to SEPA/EA and those tasked with the development of tools.

The agencies' existing sampling methods – already in place at the start of the project – have been revised during Phase I to ensure that they are sufficient to generate data suitable for the development of diagnostic tools and predictive models (CEH, 2006). This second phase describes the measured uncertainty associated with sampling lake littoral macroinvertebrates, and hence with any classification or prediction based on these methods. The variation and uncertainty associated with sampling lake *profundal* benthic macroinvertebrates was not assessed.

1.1 Task

The task of the current study was to assess the variation and uncertainty associated with lake benthic (specifically littoral) macroinvertebrate sampling methods. For while the final sampling method has been selected to be cost-effective, consistent with existing or developing European Committee for Standardisation (CEN) standards, and compliant with existing and known future health and safety standards, a further key criterion is that it must comprise techniques that introduce minimal sampling uncertainty.

1.2 Littoral method

The existing method is based on a 3-minute kick sample with a 1-minute search for animals, with three replicates taken. It follows closely the technique used by agency staff for sampling rivers, (Environment Agency, 1997), and is semi-quantitative. Dependent upon the substrate present in the lake littoral zone to be sampled, two variations of the sampling method have been proposed (CEH, 2006) – a 'hard substrate method' for sampling stations dominated by pebbly, gravel and/or cobble substrate, and a 'vegetated area method' for areas dominated by submerged, emergent or floating macrophytes. Both methods are to be used at all sites if possible. Field workers are to sample habitats present at a site in proportion to their presence, with the exception that the substrate under macrophytes is not to be sampled. The proposed methods have the same attributes as the method used to sample rivers, being cost effective and requiring no new investment in equipment and little in training. The same sensitive taxonomic groups will be collected and, in addition, key indicator groups will now be collected from macrophytes.

1.3 Variation and uncertainty

While CEH (2006) proposed various small modifications to the existing agency sampling methods, the key issue of variation and uncertainty was flagged, but not resolved, and a sampling programme for assessing uncertainty in lake littoral sampling methodology was proposed (see Annex I). The current study carries out this assessment of uncertainty.

Sampling littoral and profundal invertebrate communities must provide a sufficiently robust classification with which to assess ecological status; as defined in Annex V of the WFD as a departure from high status. In order to achieve meaningful assessment, information obtained from sampling must be sufficient to allow evaluation of the reliability of data relative to all sources of variation. Variation between samples can be attributable to a) inherent within- and between-site differences that may change with season, and b) operator-sampling and, subsequent, laboratory or field effort. Statistical confidence is affected by sample size (including sub-sampling procedures), within site replication, habitat sampled, taxonomic resolution and statistical treatment of data.

These realities provide a particular challenge when monitoring invertebrate communities which are spatially and seasonally variable. The purpose, and goals, of monitoring surface waters, as required under Article 8 of the WFD, are outlined in Annex V and elaborated by ECOSTAT (2003). The information obtained from monitoring must be responsive to defined environmental pressures in order to assess impact.

While some of the potential sources of variation and uncertainty have been quantified for river invertebrate sampling (Clarke *et al.* 2002), they have not been quantified for lake profundal or littoral invertebrate sampling. With a view to evaluating these, a sampling regime was proposed in Phase I of this project (CEH, 2006), which would allow for the quantification of variation and uncertainty in sampling lake littoral benthic invertebrates using the 'hard substrate method' (see Annex I). This proposed sampling regime was later modified so that each operator collected two samples, in order to improve the assessment of uncertainty associated with differences between both operators and replicate samples. The sampling variation and uncertainty associated with the 'vegetated area method' were not assessed.

2 Methods

2.1 Implementing Phase I recommendations

The sampling programme used here to assess sampling variation and uncertainty is based on the recommendations made in Phase I of this project (CEH, 2006) – specifically, that the sampling programme should focus on acid-sensitive lakes and be designed to distinguish and quantify macroinvertebrate community variation arising due to each of the following:

- between operators
- within station
- between stations within a lake
- across two levels of pressure – 'high' (3 lakes) and 'low' (3 lakes)
- across replicate lakes

Phase I recommended that the study lakes should be selected to cover two perceived levels of environmental pressure (specifically acidity), with three lakes at low and three lakes at high levels of pressure, giving six study lakes in total.

Furthermore, CEH (2006) recommends that the sampling method and protocol should be the 'hard substrate method', and therefore used at sampling stations dominated by pebbly, gravel and/or cobble substrate. In addition sampling should be carried out in a single season, either in spring or autumn, and within each lake there should be a random (albeit constrained by logistics and access) selection of three 'hard substrate method' stations. Meanwhile, estimating the effect of differences between sampling personnel should follow the ideas of Clarke *et al.* (2002), such that within each station two samples should be collected by operator A and one by operator B.

2.2 Sampling study design

The final sampling study design was as follows:

2 replicate 'hard substrate method' samples
by each of 2 operators
at each of 3 'hard substrate' stations
within each of 3 lakes
at each of 2 perceived levels of risk from acidification pressure ('low' and 'high')

The study lakes and stations were selected by the Environment Agency, and were sampled by Environment Agency staff. The samples were processed by an external consultancy, ECUS Ltd. Three lakes at low perceived risk of pressure from acidification, and three at high perceived risk of pressure from acidification were identified.

This design yielded four samples at each station, 12 samples in total per lake and 72 samples in total. The six lakes were selected by the Environment Agency within England and Wales to cover a range of lake sizes within each pressure class (Table 2.1). The initial aim of this was to roughly match pairs of sites in the 'low' and 'high' pressure levels according to their surface areas.

While CEH (2006) notes that operator identity should not affect results, for logistical and operational reasons it transpired that sampling operator, Operator 2 was the same individual

involved in sampling each of the six lakes; the other sampling operator (Operator 1) varied between lakes.

Table 2.1 Location and characteristics of the six sampling study lakes

Lake name	Water body ID	EA Region/Area	NGR Easting	NGR Northing	Altitude (m)	Surface Area (ha)
Low pressure						
Llyn Caer-Euni	34701	Wales/Northern	298231	340613	317	7
Llyn y Fan Fawr	40297	Wales/South West	283086	221708	608	16
Burnmoor Tarn	29215	North West/Northern	318382	504378	253	24
High pressure						
Llyn Llagi	34319	Wales/Northern	264902	348240	375	5
Llyn Alwen	33962	Wales/Northern	289799	356534	384	26
Devoke Water	29338	North West/Northern	315778	496966	236	34

2.3 Statistical methods for assessing components of variance

Analysis of variance (ANOVA) and hierarchical nested ANOVA techniques (calculated using Minitab Release 14 statistics package (http://www.minitab.com)) were used to test for, and estimate, the various sources of variation and variance components contributing to the total variance in values of a given metric within each lake and across all lakes. This is similar to the approach used by Clarke *et al*. (in press) to assess the sampling variability of a wide range of freshwater macroinvertebrate metrics for numerous 'national' sampling methods and protocols across Europe[1].

Specifically, if Y_{ijkqr} is the value of the metric for replicate sample r taken by operator q at station k on lake j of pressure level i (ie low/high risk), then Y_{ijkqr} can be expressed in terms of the sum of the components contributing towards the overall variation in its values, namely:

$$Y_{ijkqr} = \mu + a_i + b_{ij} + c_{ijk} + d_{ijkq} + e_{ijkqr}$$

where:

- μ = overall mean value of Y within the lake type
- a_i = deviation of mean value for pressure (level) i from the overall mean value μ
- b_{ij} = deviation of mean value for lake j with pressure i from the mean for pressure i
- c_{ijk} = deviation of station k on lake j with pressure i from the mean for lake j with pressure i
- d_{ijkq} = deviation of operator q at station k on lake j with pressure i from the mean for station k on lake j with pressure i
- e_{ijkqr} = deviation of replicate r by operator q at station k on lake j with pressure i from the mean for operator q at station k on lake j with pressure

[1] This work forms part of the recently completed European Union 5th Framework research project STAR (STAndardisation of River classifications) (see www.eu-star.at for further details).

and where:

- σ_I^2 = variance of the a_i = variance due to differences between pressure classes in mean value
- σ_J^2 = variance of the b_{ij} = variance due to differences between lakes within a pressure level
- σ_K^2 = variance of the c_{ijk} = variance due to inter-station differences within a lake
- σ_Q^2 = variance of the d_{ijkq} = variance due to differences between operators within a station within a lake
- σ_R^2 = variance of the e_{ijkqr} = variance due to differences between replicate samples taken by the same operator at the same station and lake.

This approach correctly distinguishes and estimates that part of the overall variance of metric values at a station which is due to systematic differences between individual operators in the way they take the sample (namely σ_Q^2) from that part which is due to pure replicate sampling variability arising from small-scale spatial heterogeneity in fauna at the sampling station (namely σ_R^2).

The overall variance (σ_E^2) in metric values at a station is the sum of the two components, namely:

$$\sigma_E^2 = \sigma_Q^2 + \sigma_R^2$$

The percentage of the overall variance (σ_E^2) in metric values at a station which is due specifically to systematic inter-operator differences is therefore estimated by:

$$P_{Q/E} = 100\sigma_Q^2 / \sigma_E^2$$

The average variance (σ_{WL}^2) in metric values within any one lake is estimated by:

$$\sigma_{WL}^2 = \sigma_K^2 + \sigma_E^2$$

The percentage ($P_{K/WL}$) of the overall variance (σ_{WL}^2) in metric values within a lake which is due to real spatial variation in macroinvertebrate fauna between stations within a lake is estimated by:

$$P_{K/WL} = 100\sigma_K^2 / \sigma_{WL}^2$$

If a particular metric and sampling method are to be effective in discriminating the ecological status classes of lakes of differing quality, then the overall sampling error for estimates of metric values for any one lake needs to be small relative to the differences in metric values between the lakes – and especially between lakes of different pressure classes. If only one sample is taken by one person at one place to assess the condition of a lake, then the sampling variance for the estimate is the total within-lake variance (σ_{WL}^2).

The total variance (σ_T^2) in metric values across all lakes is estimated by:

$$\sigma_T^2 = \sigma_{WL}^2 + \sigma_I^2 + \sigma_J^2$$

The percentage ($P_{E/T}$) of the overall total variance (σ_T^2) in metric values across all lakes which is due specifically to small-scale sampling variation within a station is estimated by:

$$P_{E/T} = 100\sigma_E^2 / \sigma_T^2$$

The 'percentage within-lake sampling variance' ($P_{WL/T}$) is estimated by:

$$P_{WL/T} = 100\sigma_{WL}^2 / \sigma_T^2$$

If $P_{WL/T}$ is large, then the sampling process and metric jointly give results which are imprecise and cannot reliably be used to detect differences between lakes and, thus, different status classes of lakes. Conversely a small 'percentage within-lake sampling variance' indicates high statistical precision and repeatability of results. It is a separate consideration whether the metric provides a meaningful, and hence accurate (rather than merely precise), ecological indicator of lake condition.

The percentage ($P_{J/T}$) of the overall total variance (σ_T^2) in metric values across all lakes which is due specifically to differences between lakes perceived to be subject to the same pressure level, is estimated by:

$$P_{J/T} = 100\sigma_J^2 / \sigma_T^2$$

Finally, the percentage ($P_{I/T}$) of the overall total variance (σ_T^2) in metric values across all lakes which is due specifically to differences between lakes in their perceived pressure levels (ie 'low' versus 'high') is estimated by:

$$P_{I/T} = 100\sigma_I^2 / \sigma_T^2$$

If $P_{I/T}$ is substantial, it means that there are practicaldifferences in metric values between the lakes perceived to be at high pressure level and those lakes perceived to be subject to low levels of pressure.

The variance components are often quoted in their standard deviation (SD) form (eg $SD_E = \sqrt{\sigma_E^2}$ denotes the overall sampling SD within a station).

3 Results

The analyses of macroinvertebrate community variation have all been based on the data at a consistent family level, and specifically BMWP family level.

Although the main analyses concentrate on variation in commonly used macroinvertebrate metrics, we have also included summaries of the distribution of occurrence of individual taxa, as well as of multivariate ordinations displaying the overall similarity of macroinvertebrate community composition within and between lakes.

3.1 Patterns of occurrence of individual macroinvertebrate taxa

Table 3.1 gives a concise summary of the frequency and pattern of occurrence of each family within each of the three sampling stations of each lake. Several families are recorded at all three stations, and in most samples from some sites, but were completely absent from all samples from other sites (eg Planorbidae, Glossiphoniidae, Erpobdellidae, Caenidae). This suggests that in many cases, when such a family is present within a lake, it is often widespread and found in most samples. There are no obvious differences between the lakes perceived to be at high and low levels of pressure (acidity) in terms of the frequency of occurrence of any individual families.

Table 3.1 Number of samples in each of three stations of each lake with each family present

Family code	Family name	Pressure level --> low Llyn Caer-Euni	low Llyn y Fanfawr	low Burnmoor Tarn	high Llyn Llagi	high Llyn Alwen	high Devoke Water	% samples with family present	Lakes with family present
051Z0000	Planariidae	443[1]	000	000	000	111	000	19	2
16130000	Valvatidae	100	000	000	000	000	000	1	1
16210000	Physidae	230	000	000	000	000	000	7	1
16220000	Lymnaeidae	000	000	010	000	000	003	6	2
16230000	Planorbidae	444	000	000	000	000	322	26	2
162Z0000	Ancylidae	000	234	000	000	000	210	17	2
17130000	Sphaeriidae	312	202	110	444	000	444	50	5
20000000	Oligochaeta	444	010	240	344	444	444	75	6
22120000	Glossiphoniidae	444	000	120	000	000	212	28	3
22210000	Hirudinidae	100	000	000	000	000	000	1	1
22310000	Erpobdellidae	444	001	434	443	000	444	65	5
371Z0000	Gammaridae	444	000	444	000	000	444	50	3
40120000	Baetidae	000	001	000	000	000	011	4	2
40130000	Heptageniidae	000	000	000	000	000	010	1	1
40210000	Leptophlebiidae	444	000	443	444	103	232	64	5
40510000	Caenidae	444	000	442	000	000	444	47	3
41120000	Nemouridae	434	000	000	344	310	000	36	3
41130000	Leuctridae	000	001	002	021	002	100	13	5
41210000	Perlodidae	000	423	000	000	000	210	17	2
41230000	Chloroperlidae	000	111	000	344	122	000	26	3
42120000	Coenagrionidae	010	000	000	000	000	010	3	2
43610000	Corixidae	444	000	000	000	000	000	17	1
45110000	Haliplidae	101	000	100	000	000	000	4	2
451Z0000	Dytiscidae	433	004	000	444	313	000	46	4
45630000	Elmidae	444	444	321	444	433	444	89	6
46110000	Sialidae	010	000	000	000	000	000	1	1
48130000	Hydroptilidae	000	000	000	302	000	444	24	2
48240000	Polycentropodidae	444	443	221	444	444	444	89	6
482Z0000	Psychomyiidae	444	000	300	443	444	110	56	5
48310000	Phryganeidae	444	000	211	002	001	010	28	5
48330000	Lepidostomatidae	000	000	001	000	000	204	10	2
48340000	Limnephilidae	233	243	210	442	224	031	58	6
48350000	Goeridae	100	000	300	000	000	000	6	2
48370000	Sericostomatidae	101	344	200	000	100	203	29	5
48390000	Molannidae	000	201	000	000	000	000	4	1
48410000	Leptoceridae	343	000	121	433	322	433	57	5
50100000	Tipulidae	001	133	000	110	314	214	35	5
50400000	Chironomidae	443	422	231	334	334	333	75	6

[1] For example, '423' indicates that a given family is present in 4, 2 and 3 of the total 4 samples from each station (1-3) respectively of a given lake.

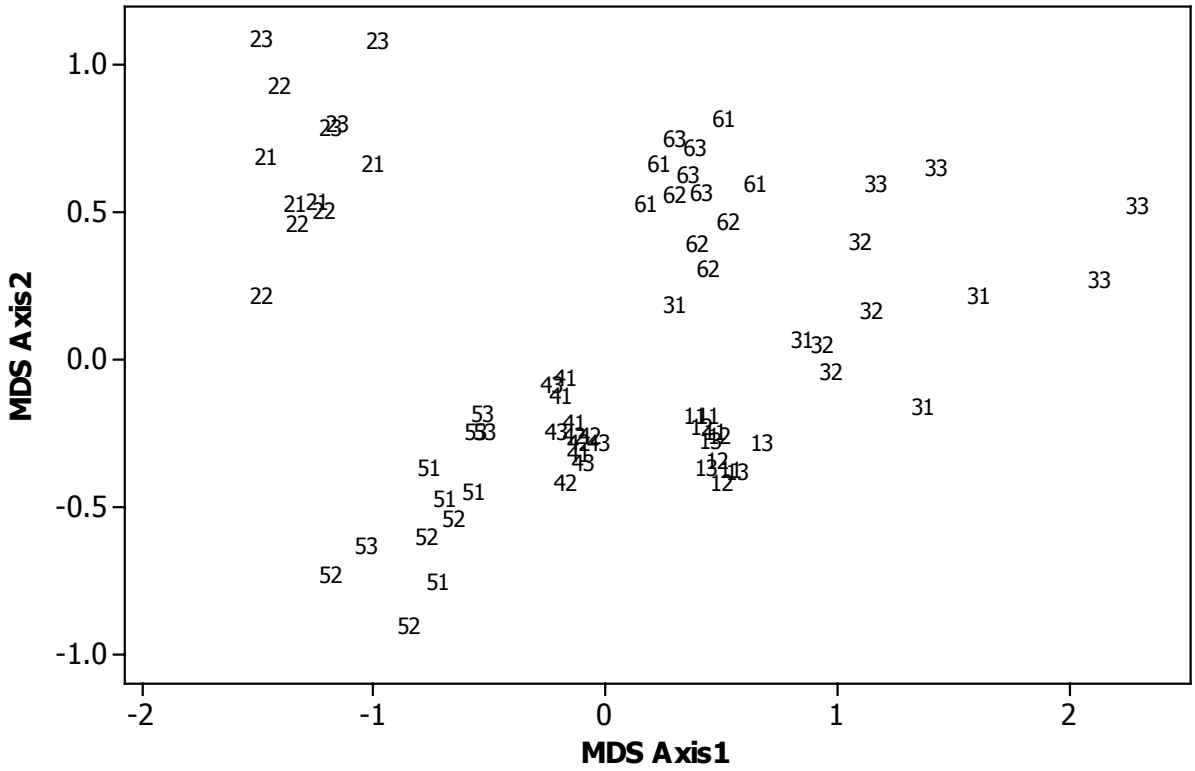
3.2 Assessing macroinvertebrate community similarity

The dissimilarity of the community composition between each pair of the 72 samples was calculated using the Bray-Curtis similarity index, which is commonly used in aquatic community studies (Clarke & Warwick, 1994). Similarities were based on double square root transformed abundances, which gives some weight to taxon abundances, but prevents the over-influence of one or two very common taxa.

The pattern of variation in the pair-wise dissimilarities of the 72 samples was represented and summarised using a multivariate statistical ordination technique called non-metric multidimensional scaling (MDS). In essence, MDS places the sample points in a two dimensional ordination plot so as to maximise the agreement between the ranks of the distances A_{ij} between each pair of samples *i* and *j* in the ordination plot and the ranks of the observed Bray-Curtis dissimilarity distances, D_{ij}. The lack of agreement is measured by a statistic called the STRESS; the lower the STRESS the better the MDS plot represents the original Bray-Curtis dissimilarities – zero indicating complete agreement (Krzanowski, 1987). The Bray-Curtis similarity calculations and the MDS ordination plot were made using the PRIMER package (Clarke & Gorley, 2001).

In this study study, the MDS two-dimensional ordination STRESS was 0.16, indicating that the plot is a reasonable representation of the inter-sample dissimilarities. Moreover, the MDS ordination plot shows that there are some distinct differences between all six lakes in terms of their taxonomic composition (Figure 3.1). This was supported by 2-way nested Analysis of Similarities (ANOSIM) randomisation tests of stations (the true level of replication) within lakes, made using PRIMER, which showed that there were statistically significant overall differences ($p<0.001$) between the six lakes.Furthermore, stations within a lake were, on average, always more similar to other stations from the same lake, than to any stations from a different lake (Figure 3.1).

Within a lake, the grouping of the samples from each station was less distinct, suggesting that a considerable part of the variation within a lake is at the small-spatial within-station scale and/or due to inter-operator effects, but with some systematic differences between stations in community composition (Figure 3.1).



Samples are codec by lake and station number. For example: 52 = sample from station 2 on lake 5.
Lakes codes are as follows: low pressure: 1=Llyn Caer-Euni, 2 = Llyn y Fan Fawr, 3= Burnmoor Tarn;
high pressure: 4 = Llyn Llagi, 5 = Llyn Alwen, 6 = Devoke Water

Figure 3.1 MDS ordination plot showing the macroinvertebrate community similarity of samples

3.3 Assessing variability in metric values due to spatial, inter-operator and replicate sampling variability

Mean and simple standard deviation of key metric values for each sampling station within each lake are presented in Table 3.2.

Table 3.2 Mean ± standard deviation of key metric values for each sampling station

	Sampling station			
(a) Number of BMWP families	**1**	**2**	**3**	**Overall**
Llyn Caer-Euni	19.8 ± 1.3	18.8 ± 1.1	18.0 ± 2.8	18.8 ± 1.9
Llyn y Fanfawr	7.3 ± 1.7	7.4 ± 1.5	9.3 ± 1.3	7.9 ± 1.7
Burnmoor Tarn	10.3 ± 3.3	8.0 ± 2.1	5.3 ± 2.2	7.8 ± 3.1
Llyn Llagi	13.0 ± 0.8	13.3 ± 1.0	13.8 ± 2.2	13.4 ± 1.4
Llyn Alwen	9.3 ± 2.2	6.6 ± 2.1	10.0 ± 2.0	8.6 ± 2.5
Devoke Water	14.5 ± 2.5	13.8 ± 1.3	15.2 ± 0.8	14.5 ± 1.6
(b) ASPT	1	2	3	Overall
Llyn Caer-Euni	5.54 ± 0.19	5.57 ± 0.09	5.73 ± 0.13	5.61 ± 0.15
Llyn y Fanfawr	6.65 ± 0.34	6.48 ± 0.70	6.43 ± 0.30	6.52 ± 0.47
Burnmoor Tarn	6.64 ± 0.61	5.33 ± 0.20	6.79 ± 0.53	6.18 ± 0.82
Llyn Llagi	6.05 ± 0.37	6.13 ± 0.19	6.17 ± 0.33	6.12 ± 0.29
Llyn Alwen	5.86 ± 0.75	5.57 ± 0.54	6.09 ± 0.32	5.84 ± 0.55
Devoke Water	5.69 ± 0.43	5.57 ± 0.45	5.41 ± 0.37	5.55 ± 0.40
(c) AWIC	1	2	3	Overall
Llyn Caer-Euni	5.14 ± 0.04	5.24 ± 0.11	5.10 ± 0.13	5.17 ± 0.11
Llyn y Fanfawr	3.71 ± 0.22	3.92 ± 0.35	4.36 ± 0.45	3.99 ± 0.42
Burnmoor Tarn	5.27 ± 0.28	5.52 ± 0.24	5.03 ± 0.73	5.29 ± 0.47
Llyn Llagi	4.75 ± 0.35	4.39 ± 0.12	4.38 ± 0.21	4.49 ± 0.28
Llyn Alwen	4.34 ± 0.25	4.22 ± 0.15	4.28 ± 0.34	4.27 ± 0.24
Devoke Water	5.08 ± 0.30	5.30 ± 0.16	5.06 ± 0.08	5.14 ± 0.21

Legend: (a) Number of BMWP families; (b) ASPT; (c) AWIC for each sampling station (1-3) of each lake (n = 4 samples) and overall for each lake (n = 12 samples)

Figures 3.2-3.4 show the individual sample metric values and range of values for each sampling station in each of the six lakes studied. The vertical lines indicate the range of the four sample values for each station and different symbols are used for the two operators to indicate clearly how values varied according to who took the sample. (Although the six audit-corrected sample values are also shown on the plots for information, they are discussed further in Chapter 4 on the results of the CEH audit of sampling processing and taxonomic identification errors).

Table 3.3 summarises the results from the hierarchical statistical ANOVA in terms of the estimates of the variance in biological metric values due to each factor described in Section 2.3 – namely replicate sampling, and differences due to operators, stations, lakes and perceived pressure levels (low/high) of lakes. In ANOVA statistical tests, only the between-lake within-pressure level variance components were usually detected as statistically significant. This does not mean the other hierarchical sources of variance are zero, or even negligible, but merely that, with these necessarily relatively small levels of replication, the estimates of variance are themselves subject to considerable uncertainty. In this study we have assumed that all factors contribute to the overall variability in metric values, and used the ANOVA estimates as the best available for comparing the relative importance of different factors.

For example, the total variance across all 72 samples in the 'number of BMBP families' recorded as present in a sample was 28.15, equivalent to a total standard deviation of 5.31. This was broken down into within- and between-lake sources of variance. Sample differences between replicates taken by the same person at the same station caused, on average, a

variance of 3.14 in the number of BMWP families, while systematic differences between operators caused an estimated additional variance of 0.72. Thus operator differences were responsible for an estimated 19% of the average total variance (3.86) within any one station in recorded number of BMWP families (Table 3.3).

Differences between stations in BMWP family richness were, on average, relatively small, adding only a further variance of 0.84, which is only 18% of the estimated total within-lake variance of 4.70. This within-lake variability was responsible for only 17% of the total variance in number of BMWP families per sample across all 72 samples and six lakes, indicating major detectable differences between lakes. However, all the between-lake variability (83% of the total variance in all 72 values) was due to differences between lakes perceived to be at the same pressure level. Amongst the six study lakes, there were no systematic differences in BMWP family richness according to whether lakes were pre-classified as being subject to low or high levels of pressure (Figure 3.2(a)).

Broadly similar results are obtained for BMWP score (Table 3.3), which has previously been shown to be highly correlated with the number of BMWP families present (Furse *et al*., 1995).

Table 3.3 Estimates of the variance in biological metric values due to replicate sampling and differences due to operators, stations, lakes and perceived pressure levels ('low'/'high') of lakes

Variance component	Symbol	BMWP Families	BMWP Score	ASPT	AWIC Families	AWIC Total Score	AWIC
Replicates	σ_R^2	3.14	167.7	0.199	3.32	76.8	0.099
Operator differences	σ_Q^2	0.72	15.5	0.000	0.64	11.4	0.000
Between Stations	σ_K^2	0.84	28.4	**0.082	0.69	30.1	0.019
Between Lakes within pressure	σ_J^2	***23.45	***610.3	*0.113	***20.72	***672.7	***0.336
Between pressure levels	σ_I^2	0.00	0.0	0.000	0.00	0.0	0.000
Overall Total	σ_T^2	28.15	821.9	0.392	25.37	791.0	0.454
Total within-station	σ_E^2	3.86	183.2	0.199	3.96	88.2	0.099
Total within-lake	σ_{WL}^2	4.70	211.6	0.279	4.65	118.3	0.118
% within-station due to operator	$P_{Q/E}$	19%	8%	0%	16%	13%	0%
% within-lake due to stations	$P_{K/WL}$	18%	13%	29%	15%	25%	16%
% Total due to within-station	$P_{E/T}$	14%	22%	51%	16%	11%	22%
%Total due to within-lake	$P_{WL/T}$	17%	26%	71%	18%	15%	26%
%Total due to between lakes within pressure	P_J	83%	74%	29%	82%	85%	74%
%Total due to pressure level	P_I	0%	0%	0%	0%	0%	0%

Note: *, **, *** indicate the ANOVA test of the variance component was statistically significant at the $p < 0.05$, $p < 0.01$ and $p < 0.001$ probability levels respectively.

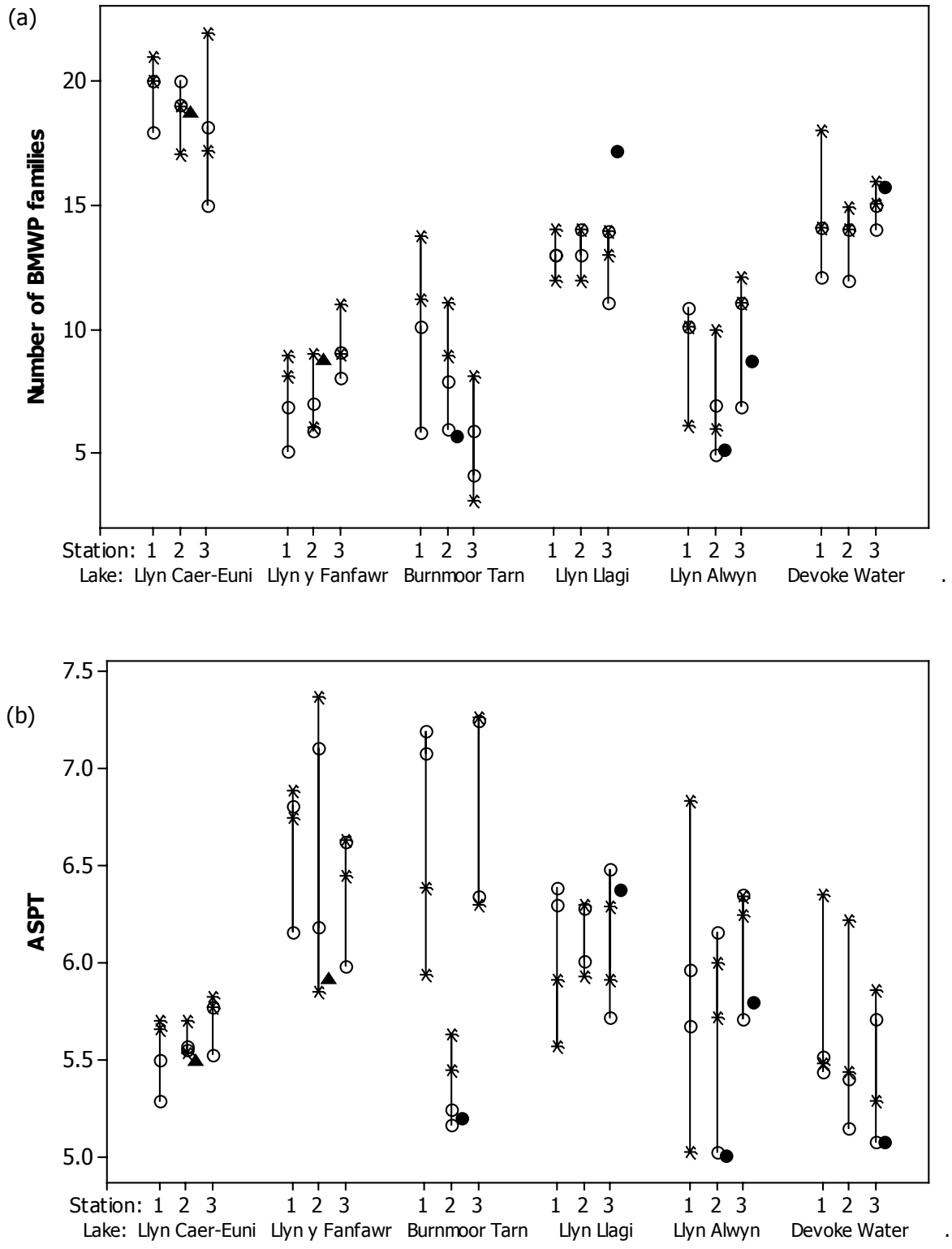
The breakdown of variance components for ASPT was quite different, with relatively more variability occurring within individual lakes. On average, the overall variance within a lake was 0.279, equivalent to a SD of 0.528, and contributing over two-thirds (71%) of the total variance (0.392) in individual sample ASPT values over the six study lakes (Table 3.3). However, the overall breakdown of variance within and between lakes is strongly influenced by the large systematic spatial differences in ASPT between the three sampling stations on Burnmoor Tarn (Figure 3.2(b), Table 3.2). All four sample ASPT values for station 2 are considerably lower than those for samples from the other two stations – this merits further checking and investigation. This increases the estimate of average variance between stations within a lake to 0.082, which is statistically significant and contributes nearly one-third (29%) of the estimated total within-lake variance of 0.279.

Within a given sampling station, differences between operators seem to have no effect on ASPT levels and the total within-station variance (0.199) is due solely to replicate sampling

variation. With, on average, 29% of total variability being due to differences between lakes it seems that it is more difficult to detect differences between lakes in ASPT than in BMWP family richness.

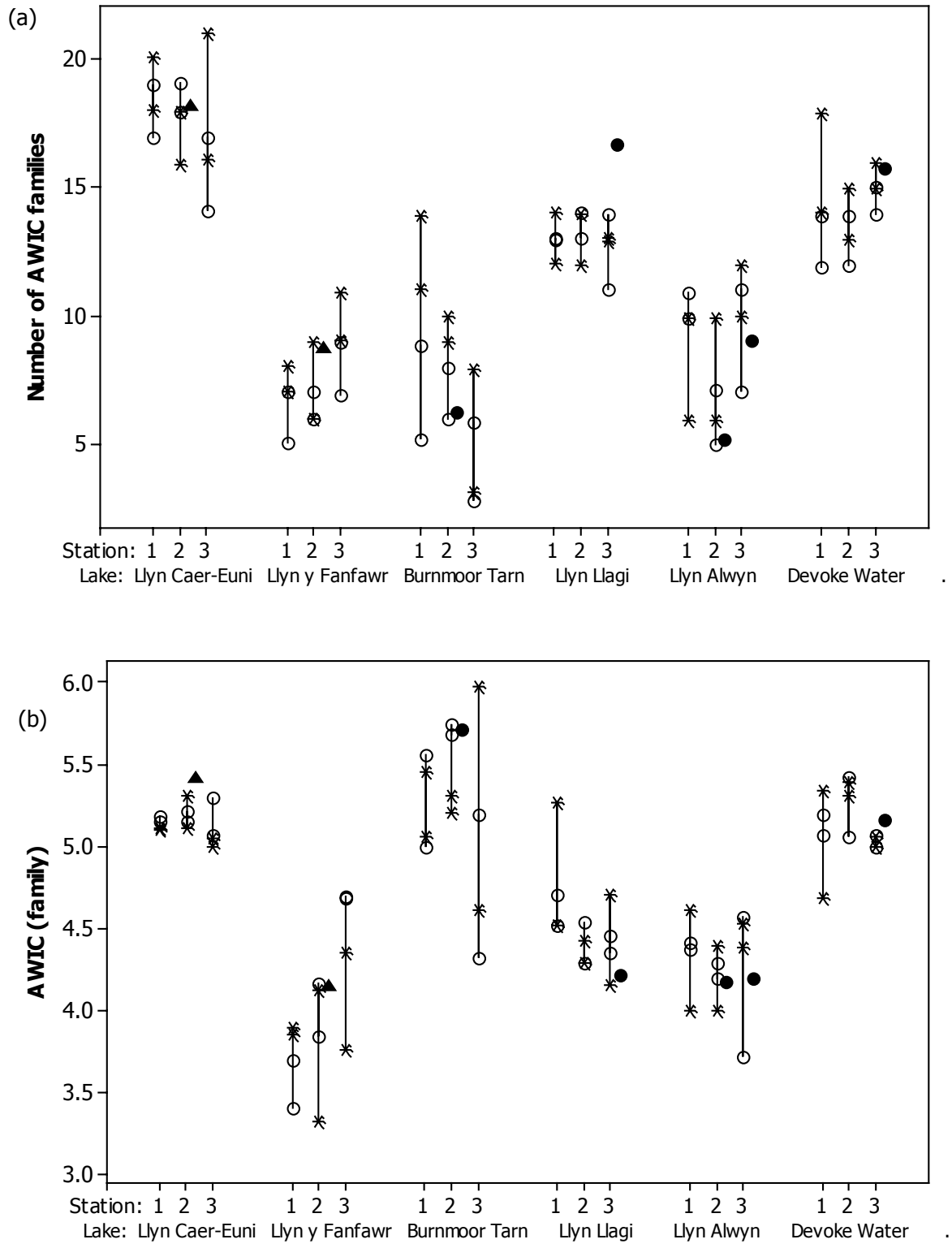
For the AWIC metric, intended to indicate acid conditions, there are systematic statistically significant differences betweenthe six lakes (Figure 3.3). Between-lake differences contribute an estimated 74% of the total variance in AWIC values, but none of this appeared to be due to differences in the perceived pressure level (high/low) of the lakes. Within a lake, large-scale spatial variation between stations was responsible for a relatively small part (16%) of the average overall within-lake variance of 0.118 (Table 3.3, Figure 3.3(b)).

Within a given sampling station, differences between operators taking the samples had no apparent effect on AWIC values, and all of the average within-station variance of 0.099 was due to pure replicate sampling variability, equivalent to a replicate sampling SD of 0.315. However, within-station SD in AWIC values varied between the six study lakes, being least for Llyn Caer-Euni (range 0.04 – 0.13) and greatest for Burnmoor Tarn (range 0.24 – 0.73) (Table 3.2 and Figure 3.3(b)).



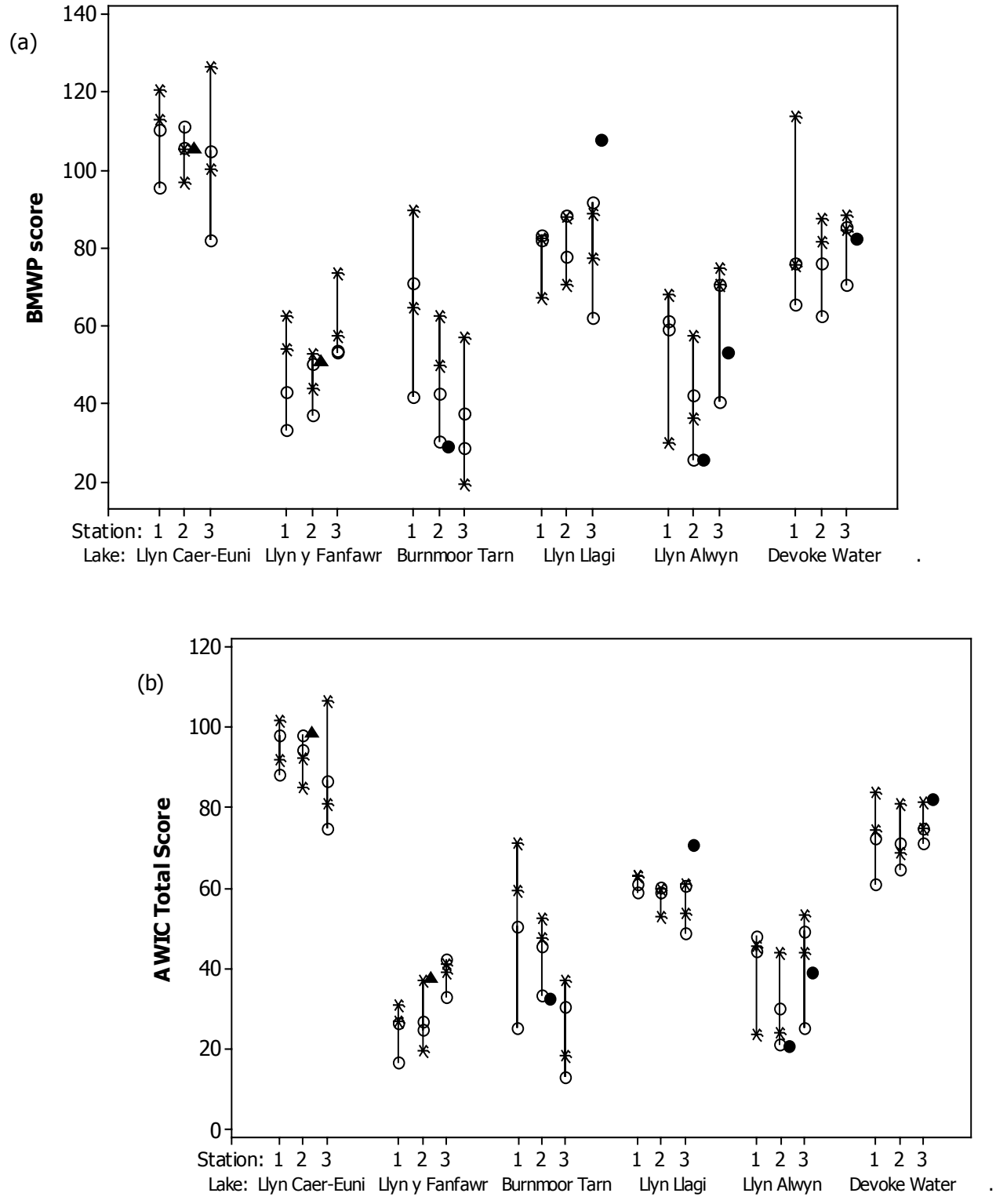
Legend: Samples identified by operator 1 (○) and 2 (*); ● and ▲ denote audit-corrected values of operators 1 and 2 respectively

Figure 3.2 Plot of all individual sample values of (a) number of BMWP families and (b) ASPT for each sampling station at each lake



Legend: Samples identified by operator 1 (○) and 2 (*); ● and ▲ denote audit-corrected values of operators 1 and 2 respectively

Figure 3.3 Plot of all individual sample values of (a) number of AWIC families and (b) AWIC (family) for each sampling station of each lake



Legend: Samples identified by operator 1 (○) and 2 (*); ● and ▲ denote audit-corrected values of operators 1 and 2 respectively

Figure 3.4. Plot of all individual sample values of (a) BMWP score and (b) AWIC total score for each sampling station of each lake

4 Audit of lake invertebrate samples

4.1 Methods of auditing samples

Twenty species-level macroinvertebrate samples that had been sorted and identified by ECUS were audited by CEH. Of these, seven of the samples were from the six lakes studied in this replicated sampling project to assess uncertainty in sampling lake benthos (see Chapter 3).

4.2 Assessment of missed or mis-identified taxa

No sorting or identification errors at family level were detected in six of the 20 samples; but nine errors (including both losses and gains) were made in one sample (Table 4.1).

Table 4.1 Number of family-level errors made within any individual sample (n = 20 samples)

Number of family-level errors	0	1	2	3	4	5	6	8	9
Number of samples	6	3	5	4	1	0	0	0	1

Within just the 20 samples audited, six BMWP families were found in the CEH audit in at least two samples where they had not been recorded by ECUS in the original sample processing; these were Sphaeriidae, Erpobdellidae, Baetidae, Caenidae, Planariidae and Hydroptilidae (Table 4.2). Seventeen other families were found in one sample by CEH where they had not been recorded by ECUS, and a total of seven families (including Erpobdellidae and Baetidae) had been recorded by ECUS in error where they were not present. There did not appear to be any consistent bias in the taxa missed or misidentified. However, it appeared that taxa were more likely to be missed in samples that contained a larger numbers of families (Figure 4.1). It was noted that a large proportion of the missed families (60%) had an AWIC score of 6 (possible range 1-6) with most of the remainder not scoring, whereas the BMWP score of the missing families was more evenly distributed, with a modal value of 5 (possible range 1-10).

The effect of sorting or identification errors at family level on the variation within the data has been estimated from the 7 audit samples from the lakes used to assess uncertainty in sampling lake benthos (see Section 4.3). Part of the variation in the data is likely to be due to inconsistency in sorting, identification and quantification (where appropriate) between samples.

Here we report only errors at family level, even though the samples were supposed to have been identified to species level by ECUS, and the number of individuals of each taxa quantified. The species-level errors in identification (as missed, misidentified or insufficiently resolved taxa) and quantification (as incorrect number of individuals) have been reported within the CEH audit. At species-level resolution, family-level errors should be minimal. Inaccuracy in sorting and identification will compromise the data produced, and reduce the value of the work. The frequency of error at family level in these data (55% of samples with 2 or more errors, 10% of samples with 4 or more errors) are sufficiently high that it is likely they would have been brought to the attention of supervisors in the Environment Agency.

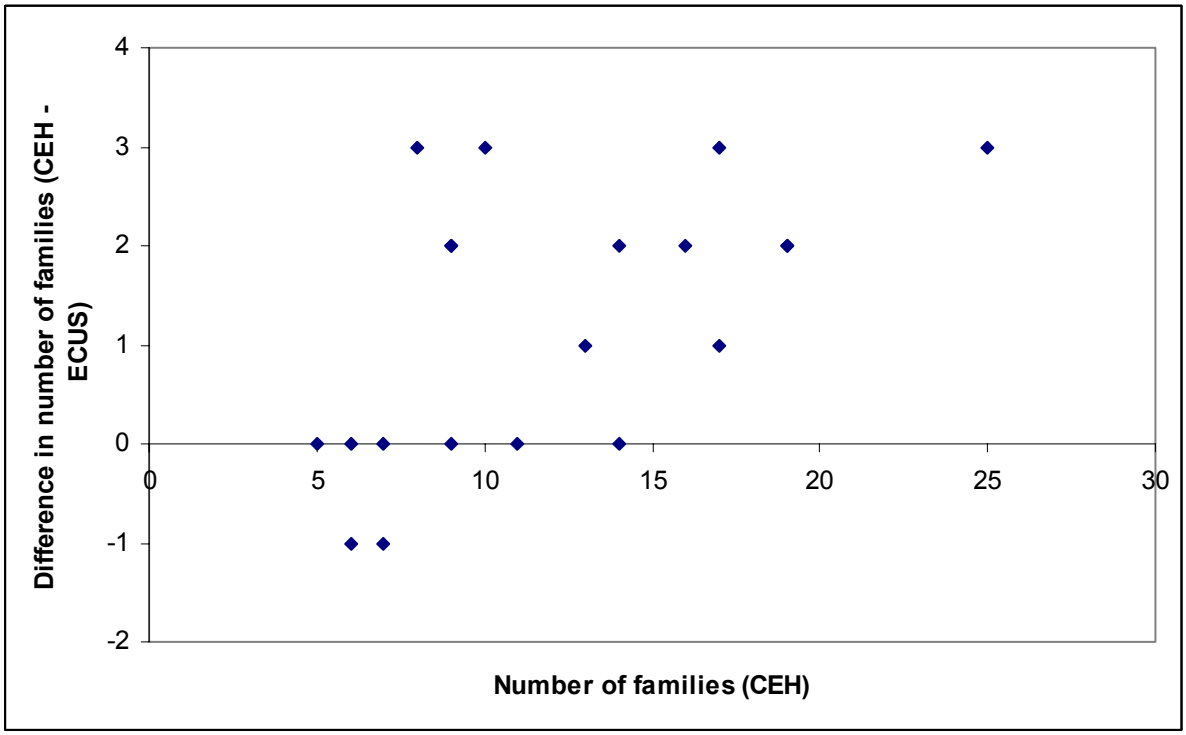


Figure 4.1 **Difference in recorded number of taxa (CEH minus ECUS) in relation to the CEH audit-corrected number of taxa present in a sample (*n* = 20 samples)**

Table 4.2 Number of audited samples in which individual families were recorded as present by ECUS and by the CEH auditors

BMWP Family	BMWP score of family	AWIC score of family	Samples with family recorded present by:		
			ECUS	CEH	Errors
Sphaeriidae	3	6	6	9	3
Erpobdellidae	3	6	8	9	3
Baetidae	4	6	4	5	3
Caenidae	7	6	10	13	3
Planariidae	5	4	3	5	2
Hydroptilidae	5	6	4	6	2
Dendrocoelidae	5		0	1	1
Ancylidae	6	6	3	4	1
Glossiphoniidae	3	6	4	3	1
Gammaridae	6	6	7	8	1
Leptophlebiidae	10	6	7	8	1
Leuctridae	10	1	2	3	1
Lestidae	8		0	1	1
Corduliidae	8		1	0	1
Libellulidae	8		0	1	1
Gerridae	5		1	0	1
Nepidae	5		2	1	1
Notonectidae	5		0	1	1
Corixidae	5	6	7	8	1
Haliplidae	5	6	1	2	1
Hygrobiidae	5		0	1	1
Dytiscidae	5	6	7	8	1
Dryopidae	5		0	1	1
Elmidae	5	6	12	13	1
Psychomyiidae	8	6	9	10	1
Polycentropodidae	7	1	10	9	1
Limnephilidae	7	4	13	14	1
Leptoceridae	10	6	10	11	1
Hydrobiidae	3	6	1	1	0
Physidae	3	6	3	3	0
Lymnaeidae	3	6	2	2	0
Planorbidae	3	6	3	3	0
Oligochaeta	1	6	20	20	0
Hirudinidae	3		1	1	0
Asellidae	3	6	5	5	0
Siphlonuridae	10		1	1	0
Heptageniidae	10	6	1	1	0
Ephemeridae	10	6	1	1	0
Nemouridae	7	1	4	4	0
Perlodidae	10	2	1	1	0
Chloroperlidae	10	1	2	2	0
Coenagrionidae	6	6	4	4	0
Hydrometridae	5		1	1	0
Naucoridae	5		1	1	0
Pleidae	5		1	1	0
Gyrinidae	5	3	1	1	0
Sialidae	4	6	3	3	0
Phryganeidae	10		4	4	0
Lepidostomatidae	10	2	4	4	0
Sericostomatidae	10	4	1	1	0
Tipulidae	5	4	4	4	0
Chironomidae	2	4	17	17	0

4.3 Effect of sample sorting and identification errors on biological metric values

This audit (Sections 4.1 and 4.2) allowed the identification of families that were either incorrectly identified as present or were present in a sample but not recorded, and the sample taxa list corrected. Values of each metric were then re-calculated for each of the 20 audited samples in turn.

Figures 4.2-4.4 show the original metric values and the audit-corrected values for the 20 audited samples. The effect of ECUS' sample processing errors on estimated ASPT values was variable. Obviously in the 30% of samples where no family-level errors were made, recorded ASPT values were the same for ECUS and CEH (samples 1, 2, 7, 8, 10 and 18 in Figure 4.2(b)). However, in 20% of samples, sample processing errors alone changed ASPT values by more than 0.3 (i.e. by +0.23 (sample 6), -0.44 (13), +0.31 (14), -0.31 (20) and by a massive -0.82 in sample 19).

Sample processing errors were the cause of similar levels of error in estimated values of the AWIC metric designed to indicate community response to the acidity of the water body (Figure 4.3(b)). Although the CEH audit revealed no errors in estimates of AWIC values for 40% (8/20) of samples, sample processing errors caused errors in AWIC values of up to -0.51, -0.78 and -0.7.

The audit reveals that, on average, the original sample analyses underestimate the number of BMWP families present by 1.2 families. Indeed, of the 20 samples examined in the CEH audit, two were underestimated by one family, six by two families and three by three families (Table 4.3, Figure 4.2(a)). Across the audited samples, a Wilcoxon signed-rank non-parametric test indicates statistically significant higher median recorded number of BMWP families for the audit-corrected values (p = 0.003). The audit also reveals an under-estimation in BMWP score (median difference = 6.5).

A Wilcoxon test also indicated that the ECUS sample processing errors caused an overall tendency to underestimate the true value of AWIC for a sample (test p = 0.038). Of the 12 audited samples where ECUS and CEH differed in their values of AWIC, the audit-corrected AWIC value was higher in 10 (83%) of cases (Table 4.3).

Table 4.3 Mean and median values of biotic metrics for the original (O) and audited-corrected (A) samples

	Mean			Median				W
	O	A	A-O	O	A	A-O	A : O	test *p*
Number of BMWP families	10.8	12.0	1.2	10.0	10.5	1.5	12 : 2	0.003
BMWP score	58.6	65.8	7.2	56.0	64.5	6.5	13 : 1	0.003
ASPT	5.42	5.49	0.07	5.39	5.47	0.00	9 : 5	0.315
Number of AWIC families	10.2	11.2	1.0	10.0	10.0	1.0	11 : 1	0.005
AWIC total score	51.7	57.7	6.0	44.5	49.5	5.0	11 : 1	0.003
AWIC	4.93	5.06	0.13	5.14	5.21	0.01	10 : 2	0.038

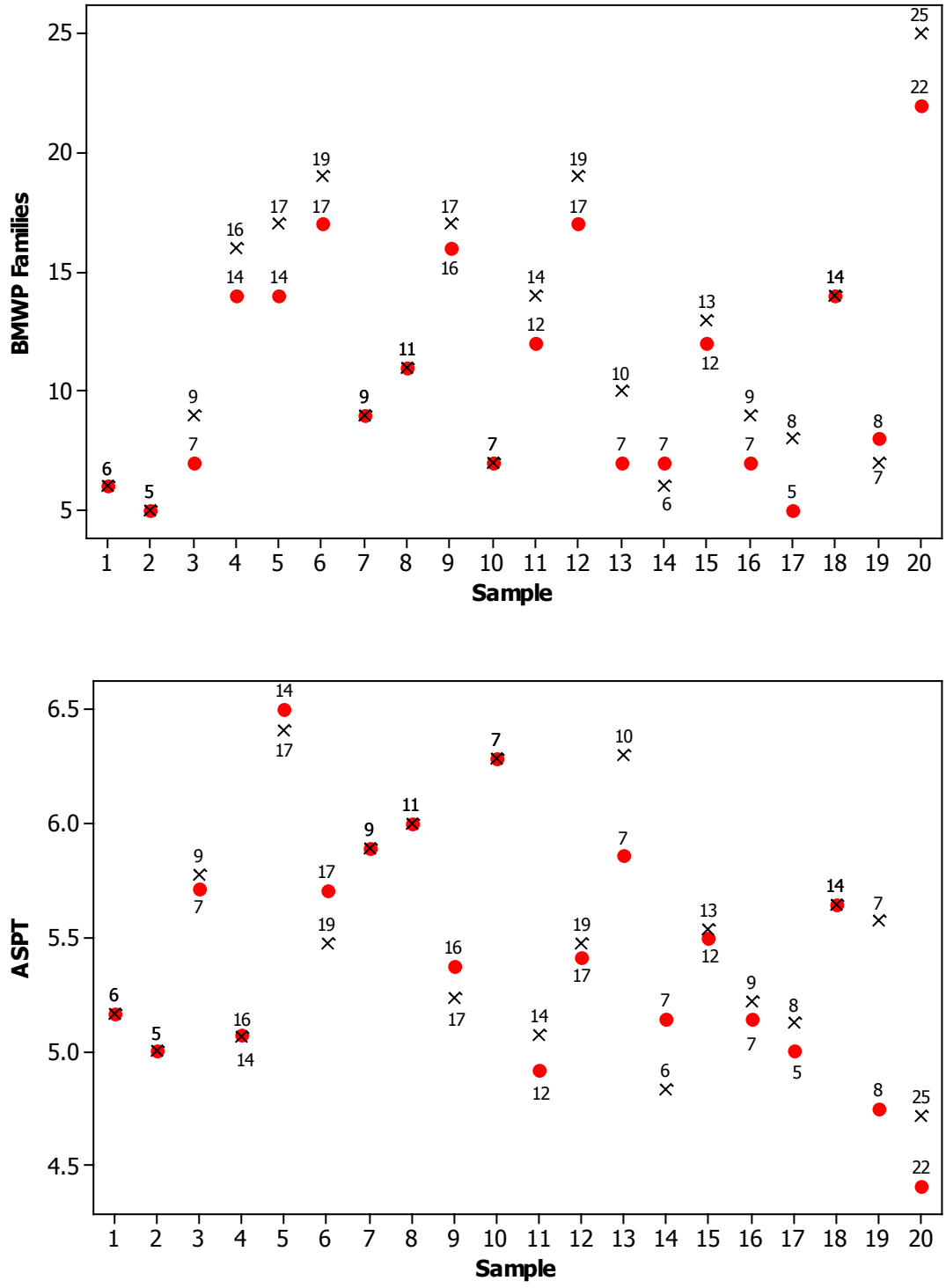
Data include the test probability value (p) of a Wilcoxon signed rank non parametric test for differences (A-O) in medians between original and audited samples; A:O denotes the ratio of the number cases where A>O to those where O>A.

Analysis of variance components techniques were used to estimate the average variance between the original and audited values for each sample and express this sample processing error variance (σ_E^2) as a percentage (P_E) of the total variance (σ_T^2) amongst the 40 metric values (20 audited samples x 2 values) (Table 4.4). Sample processing errors were responsible for 6-11% of the total variability in metric values amongst the 20 samples which

were from 19 different lakes. This error is quite separate from real field sampling variation. It is an extra source of laboratory error introduced to the bioassessment of the ecological status of lakes.

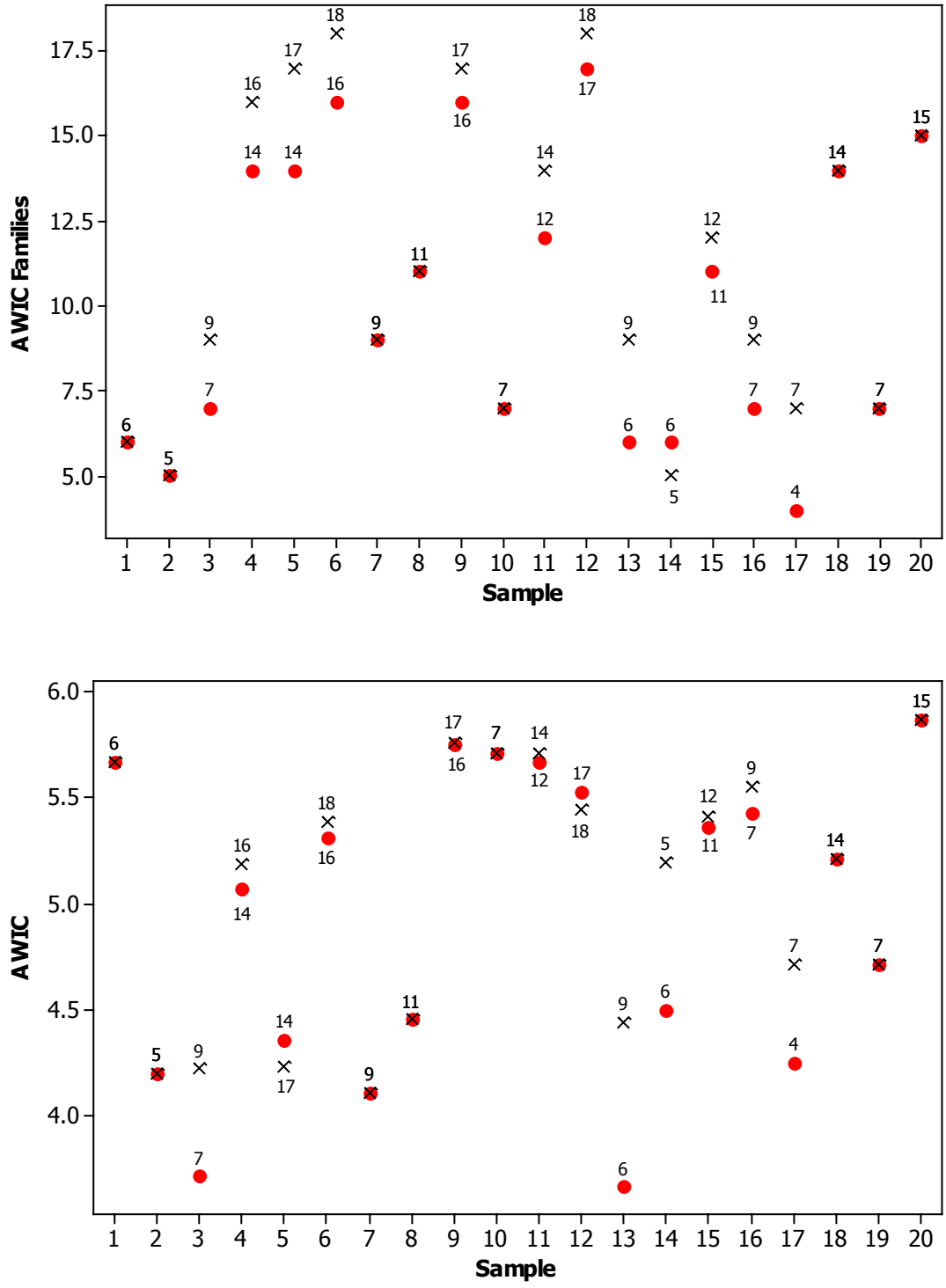
Table 4.4 Sample processing error variance (σ_E^2) as a percentage (P_E) of the total variance (σ_T^2) in metric values for 20 audited samples (2 values per sample (original and audited))

	σ_E^2	σ_T^2	P_E
Number of BMWP families	1.6	25.0	6%
BMWP score	58	670	8%
ASPT	0.030	0.233	11%
Number of AWIC families	1.3	18.3	7%
AWIC total score	38	645	6%
AWIC	0.041	0.407	9%



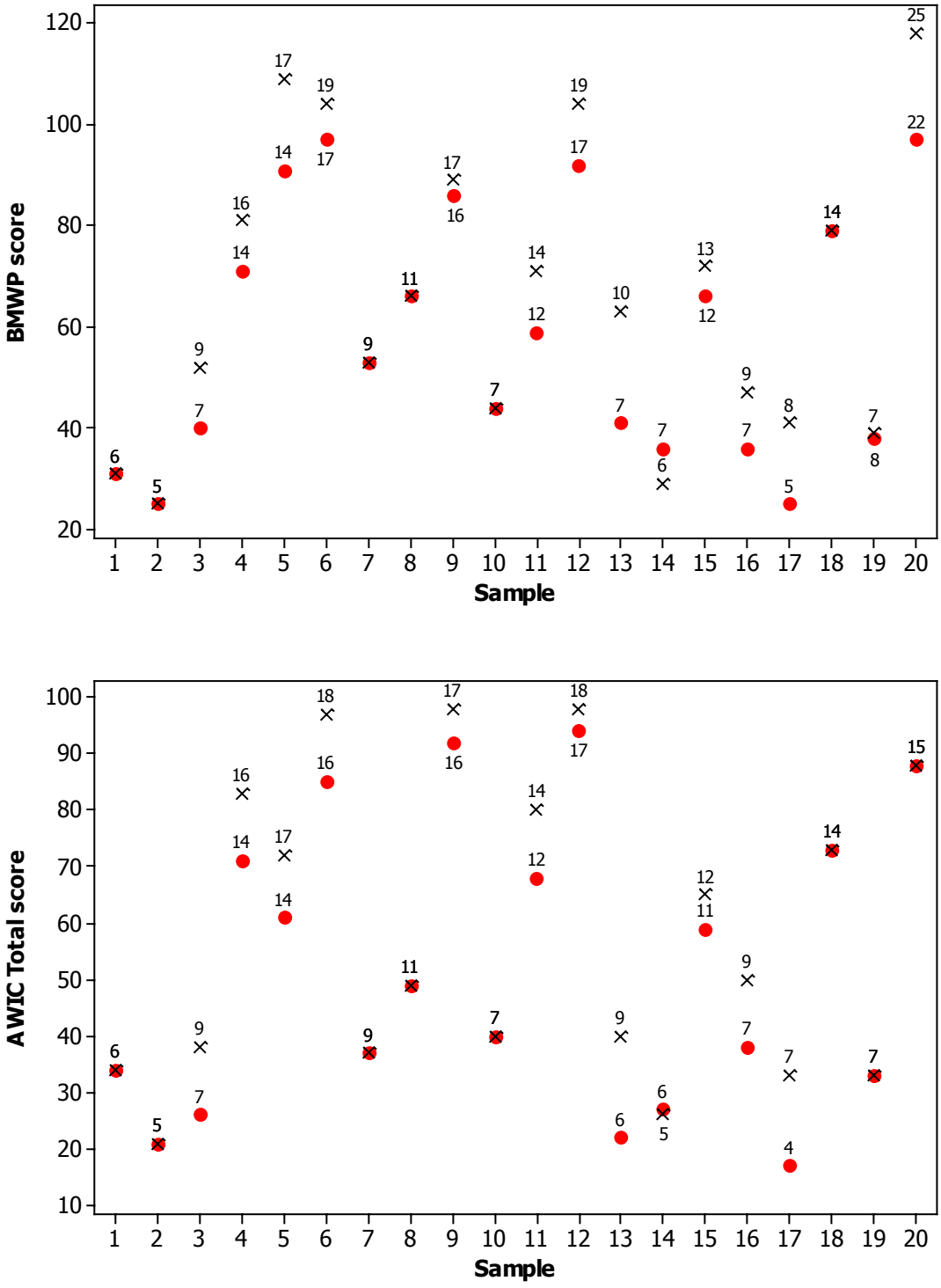
Labels indicate number of BMWP families on which ASPT is based.

Figure 4.2 **Values of (a) number of BMWP families and (b) ASPT for the original ECUS-processed sample (●) and the audit-corrected sample (X) for each of the 20 samples audited by CEH**



Labels indicate number of AWIC-scoring families on which AWIC is based.

Figure 4.3 **Values of (a) number of AWIC families and (b) AWIC for the original ECUS-processed sample (●) and the audit-corrected sample (X) for each of the 20 samples audited by CEH**



Labels indicate number of scoring families on which total scores are based.

Figure 4.4 **Values of (a) AWIC total score and (b) BMWP score for the original ECUS-processed sample (●) and the audit-corrected sample (X) for each of the 20 samples audited by CEH**

5 Discussion

5.1 Inter-operator effects

In a similar study, Clarke *et al.* (2002) assessed replicate sampling variability in the number of BMWP families, BMWP score and ASPT based on standard 3-minute RIVPACS samples across a wide range of types and biological qualities of river site. At each site in each season, two samples were taken by one person and a third by a second local biologist. In this way inter-operator effects were found to be small and contributed only 12%, 9% and 4% to the total sampling variability of number of BMWP families, BMWP score and ASPT respectively. In the current lakes study, inter-operator effects were responsible for broadly similar percentages of within-station sampling variability, ranging from 0% for ASPT and AWIC to 8% for BMWP score and 19% for number of BMWP families.

Thus different operators do produce some systematic differences in the number of taxa present in their samples, but these differences are small compared to pure replicate sampling variability. Inter-operator differences have no apparent systematic effect on either ASPT or AWIC, whose values are based on the average of the indicator scores (organic or acid) of the taxa present. These results are important because, in any practical bioassessment or biological monitoring scheme, samples from a site at different times are likely to be taken by different personnel. If inter-operator effects were very large, this could lead to false apparent changes over time (Clarke, 2000).

5.2 Number of samples and stations required per lake

Estimates of the biological status of a site generally becomes more precise the greater the sampling effort. Prior to sampling other lakes, the estimates of average variance in metric values due to variation within a sampling station (σ_E^2) and between sampling stations (σ_K^2) within a given lake can be used to estimate the likely precision of an estimate of the mean value of a metric for a lake based on *M* samples from each of *S* stations. The variance V_{SM} of such an estimate is given by:

$$V_{sm} = \sigma_K^2 / S + \sigma_E^2 / SM$$

For example, the metric AWIC has an estimated within-station variance (σ_E^2) of 0.099 and a between-station variance (σ_K^2) of 0.019, giving an estimate variance of the mean based on *M* samples from each of *S* stations of:

$$V_{sm} = 0.019 / S + 0.099 / SM$$

Thus with two samples from each of three stations the variance is predicted to be:

$$V_{SM} = 0.019 / 3 + 0.099 / 6 = 0.023$$

equivalent to a standard error of 0.152 for the estimate of average AWIC score for the lake.

For a given number total number of samples, the variance of the mean is always minimised by taking one sample from each station. However, this may incur extra costs associated with collecting the samples. In the example above most of the variation occurs at the small spatial scale within a station, so AWIC scores do not appear to show dramatic changes in different ‘hard substrate’ parts of lakes – at least for the six lakes involved in this study.

If confidence limits are for the estimate of the lake mean are needed, then the degrees of freedom of the estimate of the standard error (SE_{SM} = $\sqrt{}$ V_{SM}) of the estimator of the mean depend on the degrees of freedom of the two variance components involved, namely σ_E^2 and σ_K^2. With three stations in each of six lakes, the overall between-station (within-lake) variance σ_K^2 has 12 degrees of freedom. Treating the four samples within each station as replicates, the overall within-station variance has 54 degrees of freedom. Satterwaite's (1946) approximation is then used to get the approximate degrees of freedom of V_{SM} (and SE_{SM}) as follows:

$$df_{SM} = (\sigma_K^2 / S + \sigma_E^2 / SM)^2 / ((\sigma_K^2 / S)^2 / 12 + (\sigma_E^2 / SM)^2 / 54)$$

95% confidence limits for the estimate $\bar{x}$ of the lake mean based on *M* samples from each of *S* stations is then given by :

$$\bar{x} \pm t.SE_{SM}$$

where *t* is the two-sided 95% Student's t with df_{SM} degrees of freedom. Table 5.1 gives examples of the half width of the confidence limits for each metric for a range of values of *S* and *M*. In these cases, for these metrics, the within-station variance is always much greater than the between-station variance so the overall degrees of freedom are most determined by the DF of the within-station variance which was large (ie 54) in this study.

Table 5.1 Half width (w) of the 95% confidence intervals for a lake mean ($\bar{x}$) for a metric based on *M* samples from each of *S* stations within the lake

Variance component	Symbol	BMWP Families	BMWP Score	ASPT	AWIC Families	AWIC Total Score	AWIC
Between Stations		0.84	28.40	0.082	0.69	30.10	0.019
Total within-station		3.86	183.20	0.199	3.96	88.20	0.099
		Half width (*w*) of confidence interval of estimate $\bar{x}$ of mean **confidence interval = $\bar{x}$ ± *w***					
Number of stations (S)	**Number of samples per station (M)**	**BMWP Families**	**BMWP Score**	**ASPT**	**AWIC Families**	**AWIC Total Score**	**AWIC**
1	1	4.33	29.05	1.060	4.31	21.73	0.686
1	2	3.33	21.88	0.858	3.26	17.31	0.523
1	3	2.93	18.93	0.782	2.84	15.61	0.457
2	1	3.06	20.54	0.750	3.04	15.37	0.485
2	2	2.35	15.47	0.607	2.31	12.24	0.370
2	3	2.07	13.38	0.553	2.01	11.04	0.323
3	1	2.50	16.77	0.612	2.49	12.55	0.396
3	2	1.92	12.63	0.496	1.88	9.99	0.302
3	3	1.69	10.93	0.452	1.64	9.01	0.264
4	1	2.16	14.53	0.530	2.15	10.87	0.343
4	2	1.66	10.94	0.429	1.63	8.65	0.262
4	3	1.46	9.46	0.391	1.42	7.80	0.229
4	4	1.35	8.64	0.371	1.30	7.35	0.210

An alternative approach with multiple samples per lake is to pool the samples to produce a composite sample with a more comprehensive taxa list. Assessing the sampling variability of such composite samples is beyond the resources of this current study, but would merit further subsequent analyses.

5.3 Problems in detecting differences between lakes at 'high' and 'low' risk of acidification

In this study, no consistent differences were found between lakes at 'high' and 'low' perceived risks of pressure from acidification. There are several possible explanations. First, although a lake may be considered to be highly susceptible to the effects of acidification, it may not have been affected at the time of sampling. Secondly, the 'natural' acidity of the lakes in each group may vary and this will influence the 'natural' (reference condition) macroinvertebrate community to be expected at each lake. This may mask any effect of differences in community structure due to anthropogenic acidification. In other words, it is difficult to separate 'natural' acid lakes from those suffering from anthropogenic acidification. This can only be resolved, in future, by setting site- or type-specific reference conditions. Thirdly, the biological metrics we have used may not be sensitive enough to identify acid stress-related impacts. The AWIC has previously been shown to be a good indicator of pH levels for rivers (Davy-Bowker *et al.,* in press), although whether the ratio (WFD EQR) of observed to expected (ie reference condition) values of AWIC are a good indicator of anthropogenic acidification still needs to be tested.

5.4 Sample sorting and identification

Together with the uncertainty associated with the natural variation in communities and sample collection, an additional bias is introduced through the processing of samples. This can be minimised through quality assurance and training. Any errors in sorting and identification will have an impact on metrics derived from the data. Using 20 samples audited for ECUS (including seven samples from the data used to assess uncertainty) from 19 different lakes, processing errors were identified as responsible for 6-11% of the total variability in metric values. This error is quite separate from real field sampling variation. It is an extra source of laboratory error introduced to the bioassessment of the ecological status of lakes, and was assessed separately here to allow those tasked with tool development to remove this bias from estimates of uncertainty.

Use of ASPT reduces the impact of processing errors, and although 11% of the variance in ASPT was attributable to processing error, there was no significant difference between the original and audited ASPT. This is a consequence of there being no consistent bias, either high or low (modal score 5 out of 10), in the BMWP of taxa missed or recorded in error. There was bias in the acid tolerance of the taxa missed, however, with acid-sensitive taxa (modal score 6 out of 6) most likely to be overlooked, such that AWIC was significantly higher in the audited samples.

5.5 Using estimates of sampling variation to assess uncertainty in ecological status class assignment

The WFD requires each Member State to assess and monitor the ecological status of its water bodies, including rivers and lakes. This must be through the use of a series of Ecological Quality Ratios (EQRs) based on a comparison of the observed to expected values of stress-sensitive biotic indices or metrics. The expected of reference condition values should either be site-specific, as in the case of RIVPACS for river sites (Clarke *et al*., 2003), or at least water-body type-specific. Furthermore, the WFD requires that these EQRs be subdivided into five classes representing ecological status classes (‘high’, ‘good’, ‘moderate’, ‘poor’ or ‘bad’). The method for assigning an individual water body to an ecological status class involves using field sampling to estimate the values of these individual EQRs, and using agreed rules for combining the information from different metrics into an overall assessment for the water body. Because the assessments are based on sample information, they will be subject to uncertainty, in that other replicate samples from the same water body at the same period are likely to give different values of one or more EQRs – and hence possibly different estimates of the water body’s current status class. A thorough discussion of the varied sources of error and uncertainty in ecological quality and status class assessments is given in Clarke *et al* (1996).

More specifically, for lakes, if the ‘hard substrate method’ of sampling macroinvertebrates used in this study is used in subsequent assessments of lakes, then the estimates of sampling variation obtained here could be used to help assess uncertainty in estimates of status class. This study has derived estimates of sampling variability both within a sampling station and for a lake as a whole for some commonly used biotic indices. However, if other metrics are to be used in lake assessments, this same dataset could be used to calculate the values for the new metrics for each of the 72 samples and analysed to derive estimates of their susceptibility to sampling variation.

Clarke and Hering (in press) discuss how to incorporate the effects of field sampling variation, sample sorting and identification errors into estimates of the uncertainty in bioassessments. They also discuss and refer to a software package called STARBUGS (STAR Bioassessment Uncertainty Guidance Software), which helps to assess the

consequence on sampling and other errors in estimating the observed (and expected) values of biotic metrics on the uncertainty and probability of mis-classifying the ecological status of water bodies. This software is based on an extension of the philosophy used in the uncertainty simulation module incorporated into the RIVPACS III+ software system for bioassessment of rivers (Clarke 2000; Clarke *et al*., 2002). Further details of STARBUGS are also available from the STAR web-site (www.eu-star.at).

Developing procedures for using this sampling method to assess both lake condition and any uncertainty in resulting methods should be part of the next phase of this project.

Abbreviations

ANOSM	analysis of similarities
ANOVA	analysis of variance
ASPT	Average Score Per taxa
AWIC	Acid Water Indicator Community
BMWP	Biological Monitoring Working Party
CEH	Centre for Ecology and Hydrology
CEN	European Committee for Standardisation
ECUS	Environmental Consultancy
EQR	Ecological Quality Ratio
MDS	multi-dimensional scaling
RIVPACS	River Invertebrate Prediction and Classification System
SD	standard deviation
STARBUGS	STAR Bioassessment Uncertainty Guidance Software
STAR	Standardisation of River Classifications
WFD	Water Framework Directive

References

CEH (2006) *Lake benthic macroinvertebrates and water quality I: improving methodology.* Environment Agency Science Report SC030294/SR1, Environment Agency, Bristol

Clarke, R.T. 2000. Uncertainty in estimates of river quality based on RIVPACS. In: Assessing the biological quality of freshwaters: RIVPACS and similar techniques. Wright, J.F., D.W. Sutcliffe and Furse, M..T. (eds). *pp 39-54. Freshwater Biological Association*, Ambleside.

Clarke, R.T., Cox, R., Furse, M.T., Wright, J.F. & Moss, D. 2002RIVPACS III+. River InVertebrate Prediction And Classification System with error assessments. Release 2.1 (March 2002). Software and User Manual. 64pp & 8 Appendices.

Clarke, R.T., Davy-Bowker, J., Sandin, L., Friberg, N., Johnson, R.K. & Bis, B. “in press”. Estimates and comparisons of the effects of sampling variation using European ‘national’ macroinvertebrate sampling protocols on the precision of metrics used to assess ecological status. *Hydrobiologia*.

Clarke, R.T., Furse, M.T., Gunn, R.J.M., Winder, J.M. & Wright, J.F. 2002 Sampling variation in macroinvertebrate data and implications for river quality indices. *Freshwater Biology*, **47**, 1735-1751.

Clarke, R.T., Furse, M.T., Wright J.F. & Moss, D. 1996. Derivation of a biological quality index for river sites : comparison of the observed with the expected fauna. *Journal of Applied Statistics*, **23,** 311-332.

Clarke, K.R. & Gorley, R.N. (2001) PRIMER v5 User Manual/Tutorial. Primer-E, Plymouth Marine Laboratory, England

Clarke, R.T. & D. Hering, "in press". Overview of uncertainty in bioassessment methods and assessing the implications using STARBUGS. *Hydrobiologia*.

Clarke, K.R. & Warwick, R.M. 1994. Changes in marine communities: an approach to statistical analysis and interpretation. Plymouth marine Laboratory, England

Clarke, R.T., Wright, J.F. & Furse, M.T. 2003. RIVPACS models for predicting the expected macroinvertebrate fauna and assessing the ecological quality of rivers. *Ecological Modelling*, **160**, 219-233.

Davy-Bowker, J., Clarke, R.T., Johnson, R.K., Kokes, J., Murphy, J.F. & Zahrádková, S. "in press". A comparison of the European Water Framework Directive physical typology and RIVPACS-type models as alternative methods of establishing reference conditions for benthic macroinvertebrates. *Hydrobiologia*.

Environment Agency. 1997. Procedures for collecting and analysing macroinvertebrate samples. Quality Management Systems for Environmental Monitoring: Biological Techniques, BT001. Environment Agency, Bristol.

European Commission, 2000. Directive 2000/60/EC. Establishing a framework for community action in the field of water policy. European Commission PE-CONS 3639/1/100 Rev 1, Luxemburg.

Furse, M.T., Clarke, R.T., Winder, J.M., Symes, K.L., Blackburn, J.H., Grieve, N.J. and 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.* R&D Note 412, National Rivers Authority, Bristol.

Krzanowski, W.J. 1987. *Principles of multivariate analysis: a user=s perspective*. Clarendon Press, Oxford.

Annex I: Letter to Project Manager

Centre for Ecology & Hydrology
NATURAL ENVIRONMENT RESEARCH COUNCIL

CEH Dorset

Winfrith Technology Centre
Winfrith Newburgh
Dorchester
Dorset DT2 8ZD

Telephone (01305) 213559 (direct)
Main Fax (01305 213600)

7th April 2006

Dear Ben,

Below is our response to the queries raised over R & D Technical Report 13765: Biological quality of lakes: Phase II: Quantifying uncertainty associated with macroinvertebrate sampling methods for lake benthos.

- Even if the sorting/identification error doesn't have a large impact on the error, I still find it difficult to work with a dataset which has a large number of identification errors (at family level) VA

We agree and we hope that this will influence the future choice of contractors for such taxonomic work. In such critical work quality rather than economy should influence choice.

However, we do not think that this does not invalidate the broad conclusions of the work.

- ANOVA method seems to be applied in same way as done as applied in the 'risk of misclassification' protocol. Thought it was all very well explained in the report, as well as the statistical significance of the estimates.VA

Thanks

- A bit strange there was no significant differences detected between lakes at high and low risk. Not sure if this is due to: (1) quality of the dataset or (2) lakes not characterized in the right way, or (3) sampling in lake not representative or (4) ASPT, BMWP and AWIC not able to seperate between high and low risk. VA Specifically for AWIC, natural acidity in some lakes probably covers up anthropogenic acidification. BM

We also found this surprising, that neither the community analysis nor the metrics differentiated high and low risk lakes, but agree with (2) above. We did not have any environmental data to determine if the lakes were stressed or just sensitive to stress.
As you know the lakes were chosen by agency staff, presumably with the best information available to them.
CEH's own data for some of the lakes indicated that the high risk group did not differ in general levels of pH from the low risk group.

- It would be really interesting to see what the effect would be of more samples in space/time. I mean: can you work out how Confidence of Class would change with the applied monitoring programme. However, due to time constraints, I would only start doing this if you're sure about metrics/classification tool used for WFD. VA

The sampling of different stations has given estimates of spatial variability within lakes. We agree that it would be useful to assess temporal variability, but this was beyond the resources of the current contract. Perhaps future work could involve an assessment of natural inter-annual variability or seasonal differences and the optimum season(s) to use.

'The variation and uncertainty associated with sampling lake profundal benthic macroinvertebrtaes was not assessed' Concern why this was not the case. It states in Work Package 6 'Establish the potential degree of method uncertainty associated with both sampling and analytical (specifically the sub-sampling of profundal benthic macroinvertebrate samples) techniques' Did lack of data prevent this from being done? LAKE TAG, BM

We were not provided with any data from profundal samples to analyse.

MDS (supported by the ANOSIM tests) show distinct differences between all the lakes and that stations within a lake were, on average, more similar to other stations in the same lake than to stations from a different lake. However, groupings of samples from each station were less distinct = seems to be mainly caused by small scale within station differences (seen in replicate sample variability) and not variance caused by operator differences. In summary is this conclusion correct? BM

Yes, the percentage within station variance due to operators was always low and never more than 19% (Table 3.3). Operator differences were not assessed on the MDS community ordination.

Inter operator differences of 0% for ASPT and AWICS is to be expected because they are averages and hence more robust with regard this type of potential error. Good to see that differneces in number of families between operators is small compared to the total variance. However, this could suggest the total variance is very large and not that sampler differences should not be of any concern. BM

From Table 3.3 the inter-operator variance of number of families was 0.72, equivalent to a standard deviation of 0.85 and thus 95% confidence intervals of ±1.7 families. This contributes to a relatively high coefficient of variation (i.e. percentage error) due to inter-operator difference for taxon poor sites but a low percentage error for taxon rich sites.

BUT - No differences in BMWP family richness according to whether they were at high or low levels of pressure. To me this suggests a tool based (or one that includes a metric) on diversity of families, is unlikely to work. Would you agree? GP, BM

No, it could just be due to site risk not being classified correctly. See point above.
Diversity metrics will not distinguish between natural and anthropogenic acidification for lakes of similar pH. This highlights the value of a RIVPACS type approach which compares the observed with the expected value, and thus attempt to allow for natural variation between sites in the biota.

The suggestion that the biological status of a site becomes more precise the greater the sampling effort is perhaps not surprising.

However, limited resources, as ever, will dictate a limit on the number of samples for each lake. Clearly small scale within site differences, caused by 'patchy' habitat types is crucial. At the same time, site differences for the BMWP analysis were responsible for only 18% of the estimated total within-lake variance. Does this suggest, given limited resources, it would be better to concentrate on samples replicates at the same sites, rather than single samples from a number of sites? In WP6 it states 'Determine the degree of confidence that may be ascribed to the proposed sampling methods, implentation strategy and analytical techniques' eg what would be better, 3 samples from 3 stations or 3 samples from 1 station?
LAKE TAG, BM

For a given number of samples the most precise estimates of a mean for a lake will be obtained by taking 1 sample from as many different sites as possible. The greater the between site variance the more advantageous it is to spread the samples across as many sites as possible. So in your example, 3 samples from 3 stations are better than 3 samples from 1 station.
This ignores additional costs of travel and time associated specifically with sampling different sites around a lake. If these additional costs are high then the optimal, i.e. most cost effective, sampling scheme could be to take several replicates from fewer sites. It is a balance between these extra costs and incorporating the spatial variability between sites in maximising the precision of the lake mean, as discussed in section 5.2 page 24.

Pooled samples - what would be the advantages of this approach? Would this be through pooling replicates or pooling samples from different sites? LAKE TAG

The sampling variance of a pooled sample metric value could be less than the sampling variance of the average of the metric values of each individual sample, as was found marginally in Clarke et al. 2002 (Freshwater Biology). Obviously metrics such as the number of taxa would be higher for the pooled sample. Also you would still need at least two composite samples per lake to have any estimate of variability.

Bias in acid tolerant taxa missed, so AWIC was significantly higher in the audited samples. I would think this is because such a high percentage of the families scores 6. Therefore, if say a primary sorter missed 4 families, the chances are high that at least 3 of these would be 6 scorers. Some of the common and abundant 'families' score less than 6 (eg chironomidae and Oligochaeta), and therefore these lower scoring families are less likely to be missed in sorting, raising the chances that those famiiies that are missed, score 6. In diverse samples this would not have a sigificant impact, but in less diverse samples it may. Another factor, relevant to these data, are that many of the lakes do have low pH. Therefore, one would expect that the 6 scoring families to have lower abundances than the low scoring taxa, due to stress. Again, this would mean those families missed in sorting are more likely to be 6 scorers. I realise this is no different for BMWP families at organically polluted sites. BM

We agree with this interpretation, with the exception that more errors tended to be made in more taxon rich samples Figure 4.1.
We also point out that acidic sites tend to have lower diversity, so the impact of missing any low abundance sensitive taxa (scoring 6) will be more pronounced.

To look at the effect of sorting and ID errors on values Wilcoxon was used. Why did you use a non-parametric test? I am not a statistician,

however after having discussions with people who are (Reading Uni Statistical Services), they seem to think you should never need to use non-parametric techniques. I realise that biological data are often more tricky and have strange distributions, but they were still of the opinion parametric techniques will work with pretty much every type of data, through logging the data or assigning numbers to the data. Since the data were already in number 'form' ie number of families, then would a parametric technique have been suitable? I should point out, Reading were talking generally and were not commenting on this report. I would appreciate a comment on this and as a generally approach too. BM

Student's t tests would probably have given a similar result but we favoured non-parametric tests based on the ranks of values because they have an intuitive interpretation of how frequently the audited samples (A) give higher value than the original samples (O), hence we give the column "A : O" telling us the ratio of number of case where A is greater than O to cases where O is greater than A.

I cannot remember whether Bray-Curtis is parametric or not. I do not have my stats book to hand. Could you let me know on this? BM

Bray-Curtis is a quantitative similarity index measuring the community similarity of each pair of samples. It is non-euclidean but the non-metric MDS attempts to represent these similarities between all pairs of samples as closely as possible by plotting them in 2 or more ordination axes (Figure 3.1).

To conclude, would it be useful to include these valid queries and our response to an official appendix at the end of the report?

Yours sincerely

Dr John Iwan Jones

We are The Environment Agency. It's our job to look after your environment and make it **a better place** – for you, and for future generations.

Your environment is the air you breathe, the water you drink and the ground you walk on. Working with business, Government and society as a whole, we are making your environment cleaner and healthier.

The Environment Agency. Out there, making your environment a better place.

Published by:

Environment Agency
Rio House
Waterside Drive, Aztec West
Almondsbury, Bristol BS32 4UD
Tel: 0870 8506506
Email: enquiries@environment-agency.gov.uk
www.environment-agency.gov.uk