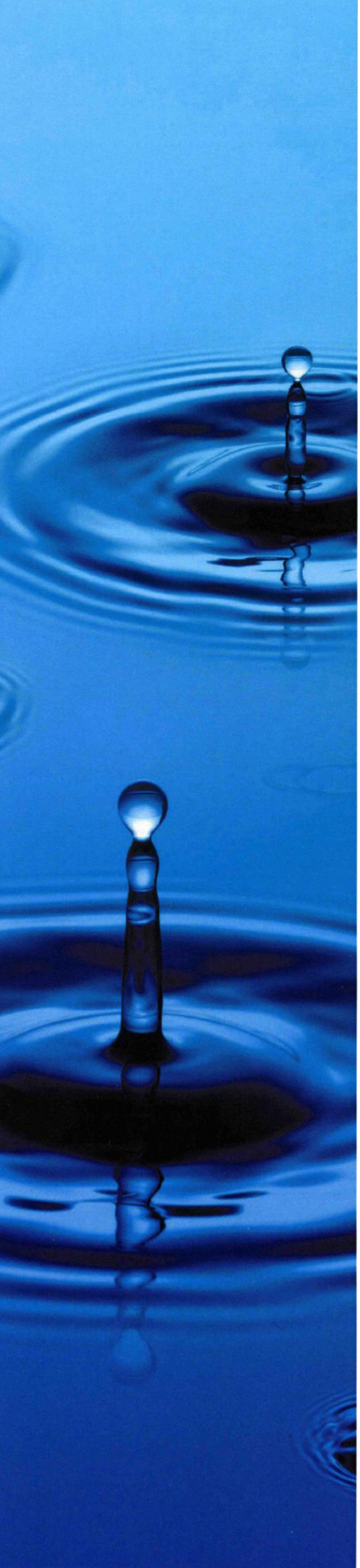


SEPA
Scottish Environment
Protection Agency

Freshwater Biological
Association

River Invertebrate Classification Tool Science Development Project: Modifications for WHPT and other Abundance-Weighted Indices

A Report to the
Scottish Environment Protection Agency

R.T. Clarke
J. Davy-Bowker

March 2014

Research Contractor

This document was produced by the Freshwater Biological Association:
Ralph Clarke, John Davy-Bowker[†]

[†]The Freshwater Biological Association, River Laboratory, East Stoke, Wareham, Dorset,
BH20 6BB, United Kingdom.

Project Funders

This project was funded by the Scottish Environment Protection Agency.

Disclaimer

Whilst this document is considered to represent the best available scientific information and expert opinion available at the stage of completion of the report, it does not necessarily represent the final or policy positions of the project funders or contractors.

Dissemination status

Unrestricted.

SEPA Project Manager

SEPA’s project manager for this contract was:
David Rendall

FBA Project Manager

FBA’s project manager for this contract was:
John Davy-Bowker

FBA Project Code

S/0008/R

The Freshwater Biological Association

The Freshwater Biological Association
The Ferry Landing
Far Sawrey, Ambleside
Cumbria, LA22 0LP, United Kingdom

The Freshwater Biological Association
River Laboratory
East Stoke, Wareham
Dorset, BH20 6BB, United Kingdom

Web site: www.fba.org.uk
Email: info@fba.org.uk

Registered Charity No. 214440
Company Limited by Guarantee No. 263162, England
UKPRN No. 10018314

Registered Office: The Ferry Landing, Far Sawrey, Ambleside, Cumbria, LA22 0LP, United Kingdom

EXECUTIVE SUMMARY

River Invertebrate Classification Tool (RICT) Science Development: Workstream 1

Project funders: SEPA

Background to research

The environment agencies in the UK (the Environment Agency; Scottish Environment Protection Agency; Natural Resources Wales and the Northern Ireland Environment Agency) use the River Invertebrate Classification Tool (RICT) to classify the ecological quality of rivers for Water Framework Directive compliance monitoring. The current system is based on using RIVPACS observed (O) to expected (E) ratios (EQIs) of two macroinvertebrate indices BMWP NTAXA and BMWP ASPT.

The UK environment agencies need to update this river assessment system to take account of new indices both for the WFD main classification and to widen assessments to take account of other pressures. These new indices, namely, WHPT, LIFE and PSI need to be introduced to improve the assessment of general degradation, and report on hydro-morphological impacts and sediment stress. All three of these indices will also for the first time incorporate abundance weighting to better reflect the requirements of the Water Framework Directive to take into account not only the structure, but also the abundance of biological quality elements in water quality classifications.

The capacity to predict and classify using abundance–weighted indices is an entirely new development for RIVPACS/RICT and these new indices will need to be incorporated into all of the existing steps that enable a classification to be performed. Specifically, the new indices will require (i) methods to base site assessments for a single year or a three year period on the average of the single season sample estimates of index EQI values, (ii) methods and estimates to correct for bias arising from laboratory sample processing errors, (iii) EQR factors to adjust EQI values to a standard WFD reference state, (iv) banding systems to permit classification of EQRs into water quality classes, and (v) estimates of sampling uncertainties to allow the calculation of confidences of class.

The priority in this current project, Workstream 1, is to develop methods, algorithms and parameter estimates to enable the abundance-based WHPT indices, WHPT NTAXA and WHPT ASPT to be incorporated into revised RICT software to provide improved river site assesments.

Objectives of research

- Develop algorithms and uncertainty parameter estimates for the incorporation of abundance-weighted classification indices, especially WHPT, into RICT
- Develop algorithms and parameter estimates for the incorporation of sample biases of abundance-weighted indices, especially WHPT, into RICT

Key findings and recommendations

Algorithms were derived to estimate river site ecological status on the average of the single season sample observed (O) to expected (E) ratios of WHPT, LIFE and/or PSI indices.

Sampling uncertainty components in the abundance-weighted WHPT, LIFE and PSI indices were estimated using a combination of existing datasets from the Environment Agency,

Scottish Environment Protection Agency and Northern Ireland which had a mixture of sites with combinations of different samples from the same site on the same day (replicates), different days and months in the same season, different seasons, and different years within and between three-year periods.

Estimates are provided of the sampling variance components (replicate and temporal) for WHPT NTAXA, WHPT ASPT, LIFE and PSI, together with derived detailed algorithms for incorporating the simulation of this sampling uncertainty into confidence of status class assessments.

A dataset of 427 externally-audited RIVPACS samples encompassing all EA regions were analysed to determine the biases (i.e. differences) between the observed (pre-audit) sample value and the audit-corrected sample value of each index.

Detailed statistical analyses provided algorithms to simulate the estimated sample processing biases in the abundance-weighted WHPT indices (WHPT NTAXA and WHPT ASPT) from the observed index values and/or the audit-based estimates of the bias in BMWP NTAXA.

A detailed algorithms section is provided to enable the RICT software programmers to encode these new methods and uncertainty parameter estimates for the abundance-weighted WHPT indices into the next version of RICT for their use in river site status classifications.

Key words: River Invertebrate Classification Tool, RIVPACS, RICT, WHPT, LIFE, PSI, abundance-weighted, Water Framework Directive.

Table of Contents

1. INTRODUCTION

The current RIVPACS Reference site model (RIVPACS IV) used in the current version of the RICT software to derive predictions of the RIVPACS expected values of biotic indies for all river sites in GB was developed in the SNIFFER WFD72C project in 2007-08 (Davy-Bowker *et al* 2008). The development involved removing some 40 sites used in the previous RIVPACS III model, developing a method to adjust predictions for variation in the perceived quality of the remaining reference sites and building a single model for the whole of GB.

The final RIVPACS model (RIVPACS IV) for GB involved the classification of 685 reference sites was into 43 end groups as shown in Figure 1. The number of reference sites in each end group varies from 6 (End group 7) to 32 (End groups 32 and 41), with a median size of 15 sites.

It is useful for mapping and descriptive summary purposes to combine the 43 groups into higher level groupings based on the hierarchical TWINSPAN classification in Figure 1. This has been done to form 7 super-groups (Table 1 and Figure 2).

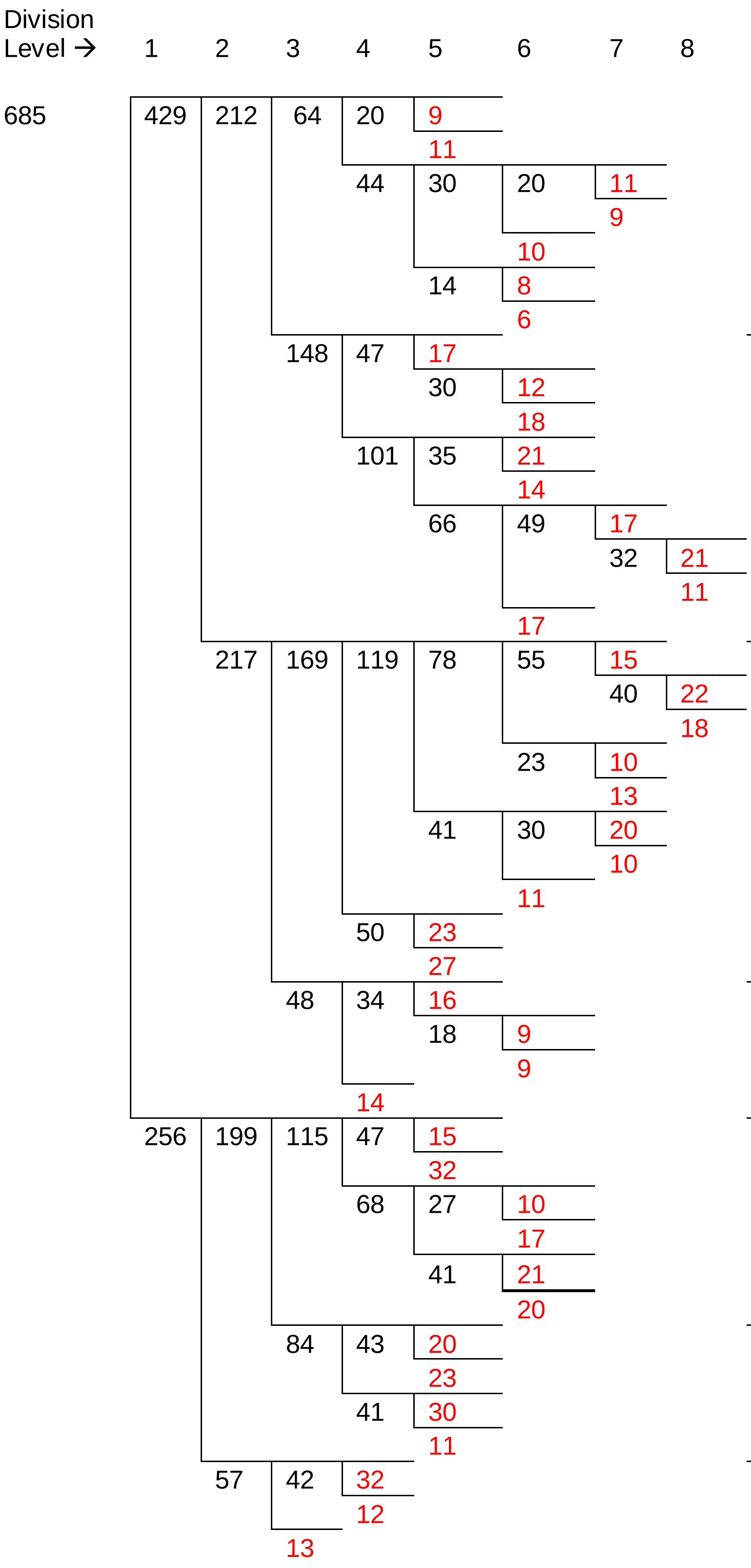
Table 1 Seven super-group level of classification of the 43 end groups of the 685 reference sites

Super-group	N sites	Mean TAXA	Mean ASPT	Dominant characteristics
1-7	64	23.0	6.27	All in Scotland mostly islands
8-16	148	25.2	6.79	Upland streams, mainly in Scotland and N England
17-26	169	31.7	6.42	Intermediate rivers, SE Scotland, Wales, N & SW England
27-30	48	27.1	6.25	Small steeper streams, with 13km of source, discharge1/2
31-36	115	34.8	5.84	Intermediate size lowland streams, including chalk, SE
37-40	84	32.7	5.58	Small lowland streams, including chalk, SE Britain
41-43	57	32.7	5.14	Lowland streams, SE England, larger, fine sediments

The distribution of the observed (O) values and O/E values for BMWP NTAXA and ASPT for the GB references sites, grouped by end-group, are shown in Figure 3 and Figure 4. The overall frequency distribution of the O/E values for these indices across all GB reference sites is shown in Figure 5.

There is a separate RIVPACS IV predictive model in RICT for river sites in Northern Ireland (NI) based on a classification of 108 reference sites into 11 end groups (two previous RIVPACS III reference sites were retrospectively judged to be inadequate and excluded from the RICT model for NI.

Figure 1. Dendrogram showing the hierarchical classification of the 685 New GB model Reference sites into 43 End-groups (1-43). End-group code indicates the TWINSPAN binary code for the hierarchical splitting for groups



End Group Id	End group Code	N sites
1	00000	9
2	00001	11
3	0001000	11
4	0001001	9
5	000101	10
6	000110	8
7	000111	6
8	00100	17
9	001010	12
10	001011	18
11	001100	21
12	001101	14
13	0011100	17
14	00111010	21
15	00111011	11
16	001111	17
17	0100000	15
18	01000010	22
19	01000011	18
20	0100010	10
21	0100011	13
22	0100100	20
23	0100101	10
24	010011	11
25	01010	23
26	01011	27
27	01100	16
28	011010	9
29	011011	9
30	0111	14
31	10000	15
32	10001	32
33	100100	10
34	100101	17
35	100110	21
36	100111	20
37	10100	20
38	10101	23
39	10110	30
40	10111	11
41	1100	32
42	1101	12
43	1111	13

Figure 2 Geographic location of the reference sites in each of the seven major groupings of the 43 end groups for the 685 reference sites in the new all GB-inclusive RIVPACS model.

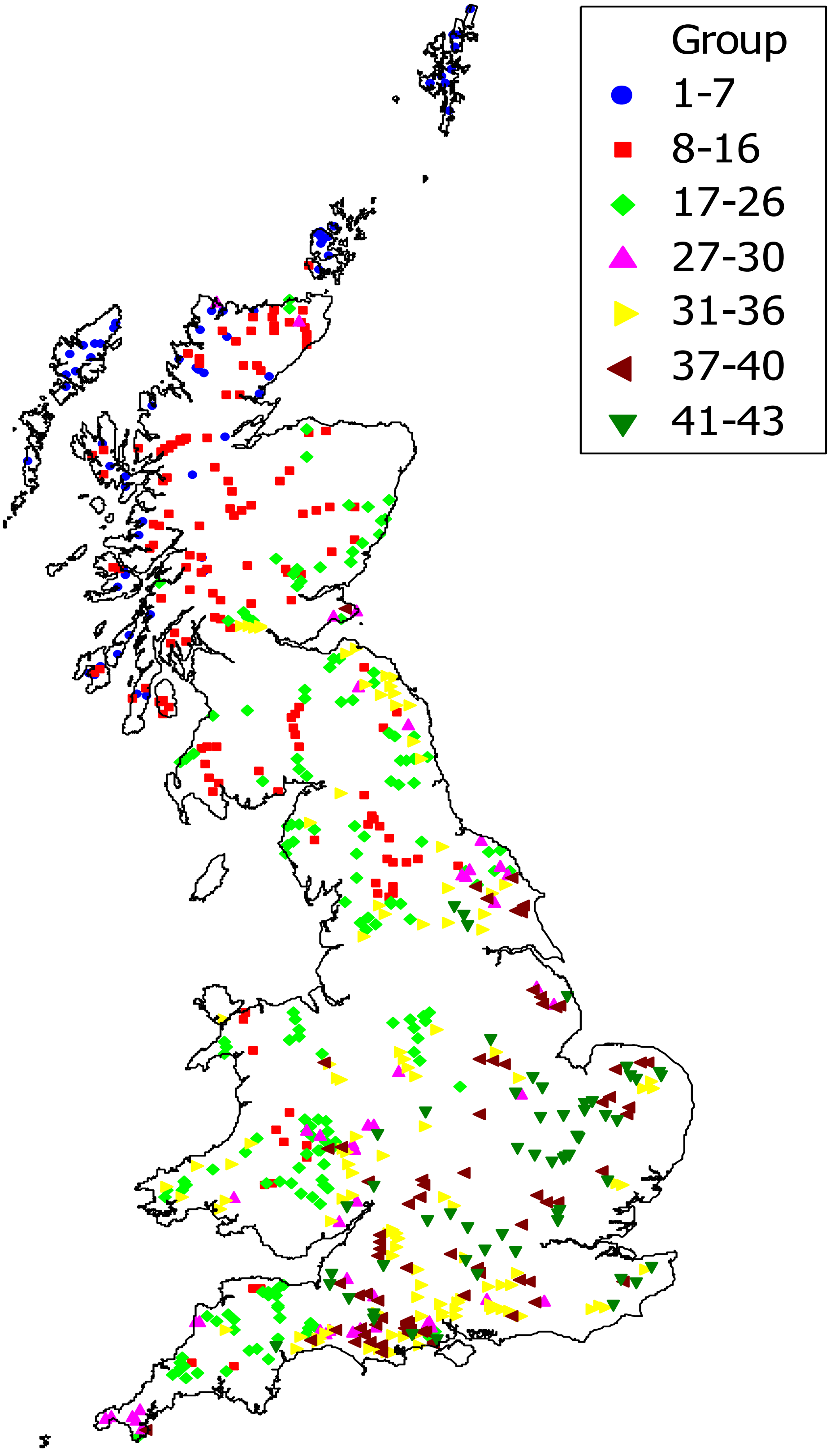


Figure 3 (a) observed number of BMWP taxa (TAXA) and (b) observed ASPT for the 685 RICT reference sites, grouped by their end group (1-43). Vertical dashed lines separate the seven super-groups described in Table 1.

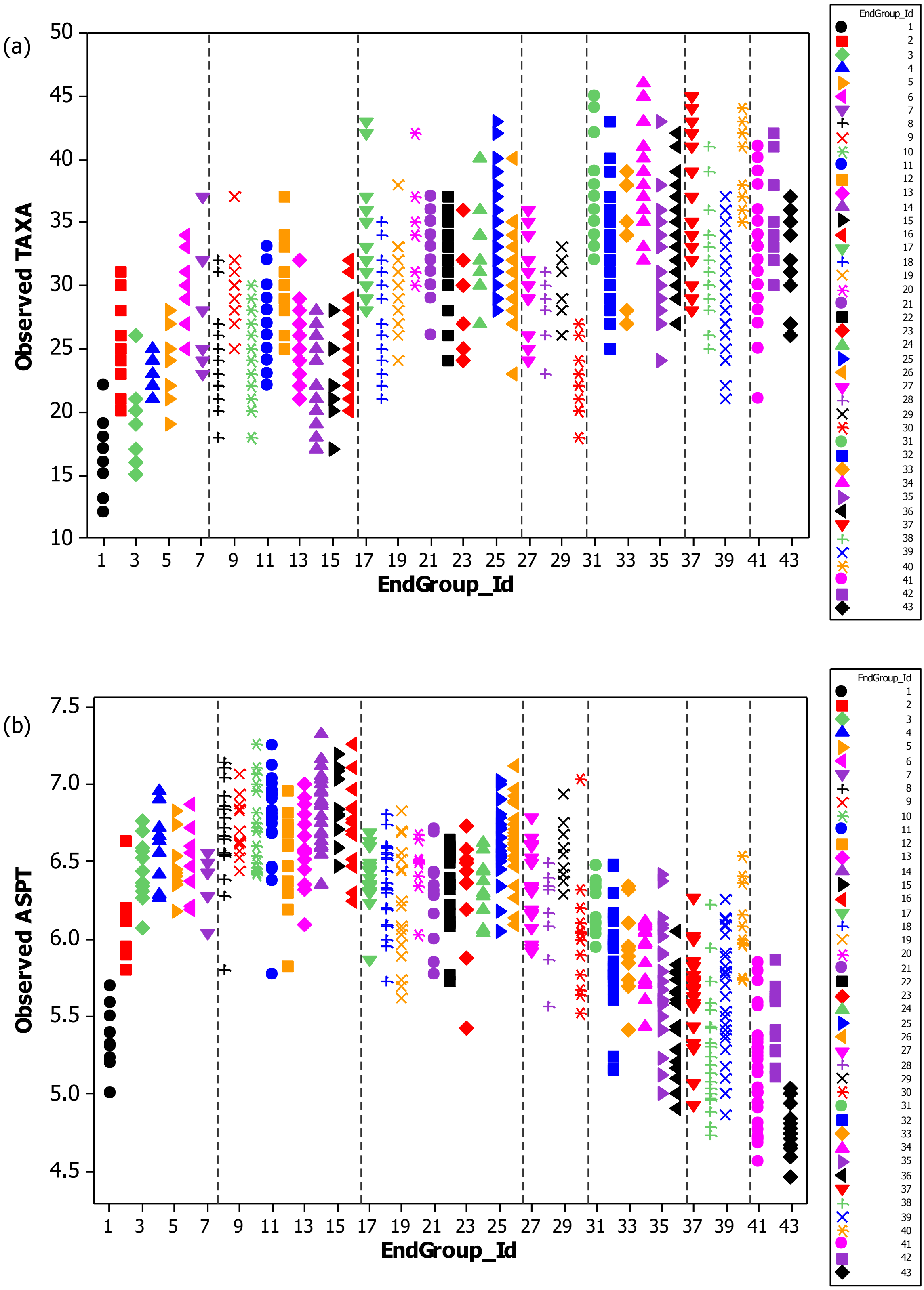


Figure 4 (a) O/E number of BMWP taxa (NTAXA) and (b) O/E ASPT for the 685 RICT reference sites grouped by their end group (1-43). Dashed lines indicate O/E values of unity.

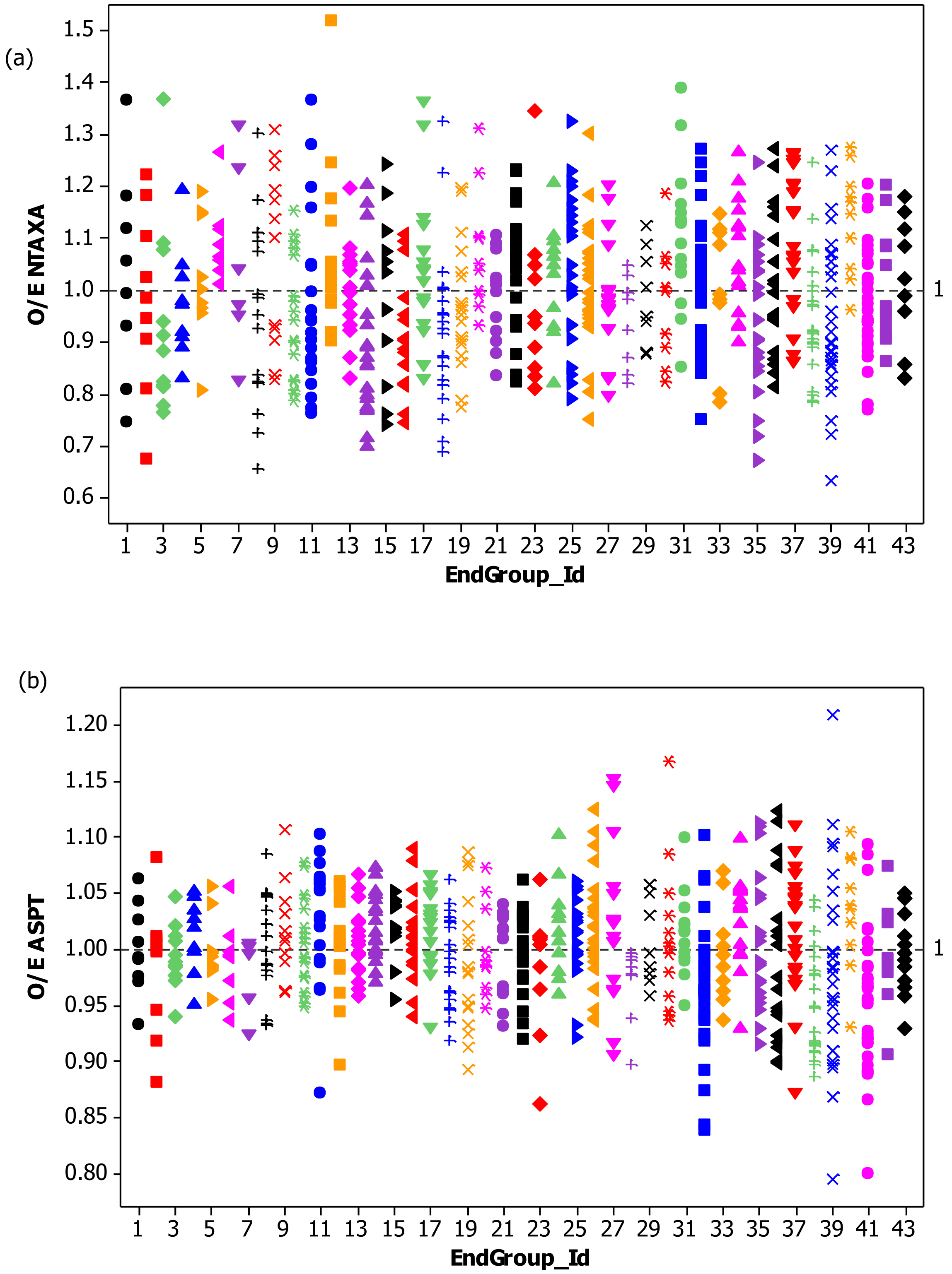
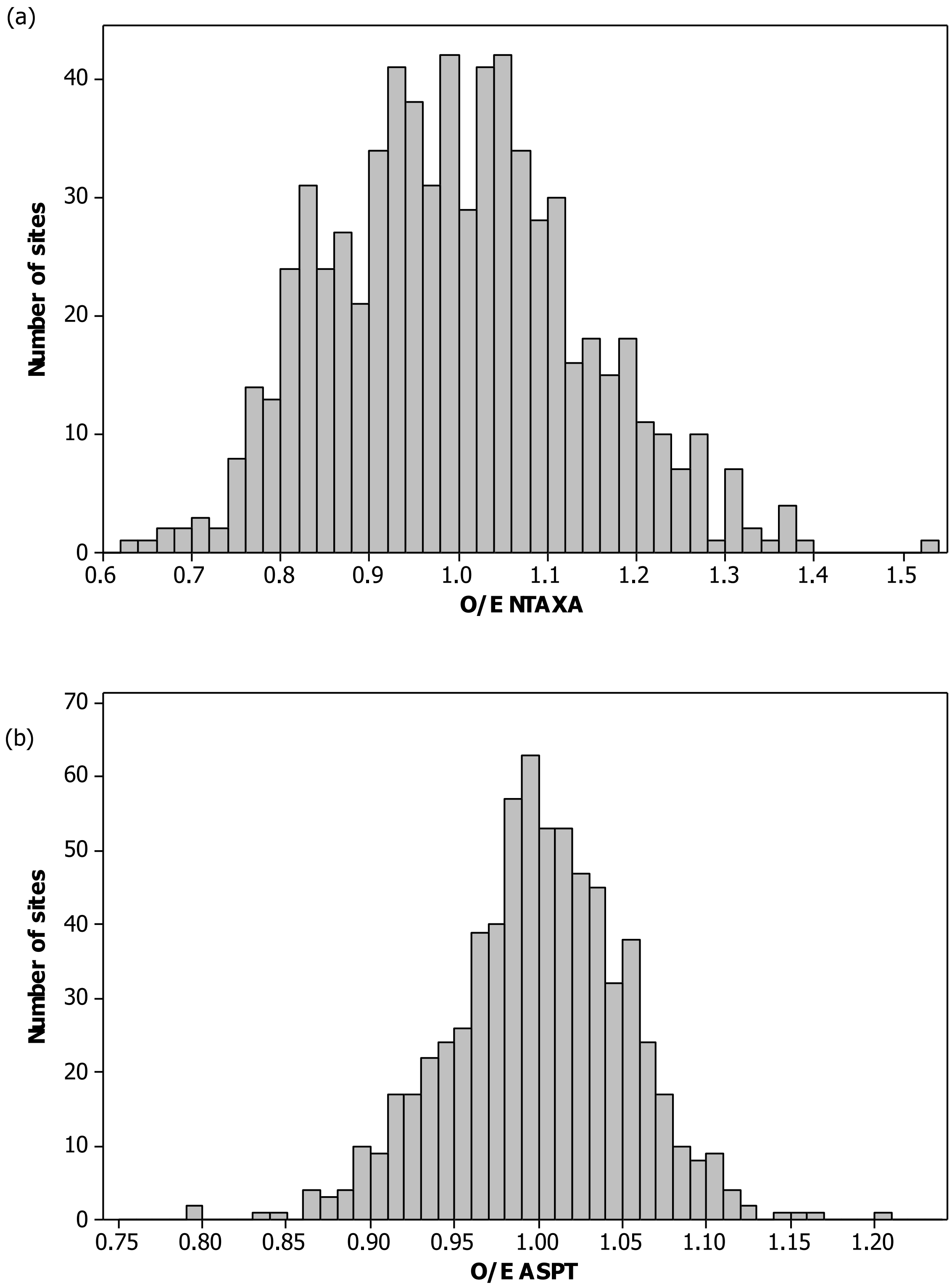


Figure 5 Frequency histogram showing the statistical distribution of the O/E values for (a) BMWP NTAXA and (b) BMWP ASPT for the 685 reference sites in the new GB-inclusive RIVPACS model with 43 end-groups.



2. Datasets used to estimate sampling error variances

RIVPACS predictive models produce predictions of the fauna and biotic index values to be expected at test sites in reference state. In RIVPACS and RICT software, comparison of RIVPACS predicted (E) biotic index values with the observed (O) values at a test site through the use of O/E ratios (EQI) provide an assessment of the biological quality of the site. As with any bio-assessment methodology it is vital to be able to quantify the sampling and other uncertainty associated with these assessments.

In this current project, estimates of the size of the various components of sampling variability between single season samples are required for each of a range of abundance-weighted indices, namely:

Abundance Weighted WHPT Score
Abundance Weighted WHPT NTAXA
Abundance Weighted WHPT ASPT
LIFE (family level)
PSI (family level)

Estimates are required of the sampling uncertainty in the average O/E value for a site for either a single season, a single year, or a three year period. This requires estimates of sampling variability due to:

Replicate sampling variability
Within-season temporal variability
Between-year (within-period) temporal variability

These estimates were derived using the same best-available combination of datasets that was used in SNIFFER project WFD72c to provide estimates of sampling variability and assessment uncertainty for the BMWP indices NTAXA and ASPT used in the initial development of RICT in 2008. The four datasets are:

28-site dataset from the Tay River Purification Board
416-site dataset from East and North-East Scotland
12-site Community Change dataset from Northern Ireland
16-site Biological Assessment Methods (BAMS) dataset

The datasets and the standardisation of their taxonomic resolution are described below.

2.1 Dataset 1: 28 Tay RPB sites

This dataset was generated by biologists from the Tay River Purification Board (RPB) that is now part of SEPA. The then Tay RPB had a network of “primary sites”, mainly on larger rivers in the Tay catchment (including the River Earn) and various other rivers between the Tay and the North Esk catchment in Angus (Figure 6).
The biological quality of the sites was generally high or good, but 4 of the 28 stand out as having impacted invertebrate faunas. These are:

8538 DEAN W. AT BRIDGEND
7989 DIGHTY W. AT BALMOSSIE MILL
7844 LUNAN W. AT KIRKTON MILL
7672 LUTHER W. AT LUTHER BRIDGE

Of the remainder, site 8688 (RIVER TUMMEL AT ALDOUR RD BDG PITLOCHRY), occasionally had poor biotic index scores due to the difficulty of sampling this river which has quite marked variation in water levels due to hydro-electric influences. The size of site 8322 (RIVER TAY AT RAIL BR. PERTH) has also made sampling difficult on occasion. The remaining sites were generally of high or good quality although sporadic sheep dip problems in the mid-1990s affected many of the rivers.

The sites were sampled between 1988 and 1997 and four replicate samples were taken at each site on each sampling occasion in spring and autumn. Although not all sites were sampled in all years, many sites have concurrent runs of data, especially in the 5-year period 1990 to 1994.

Of the 4 replicate samples taken on each sampling occasion, three were analysed to BMWP family level and one was analysed to a mixed taxonomic level.

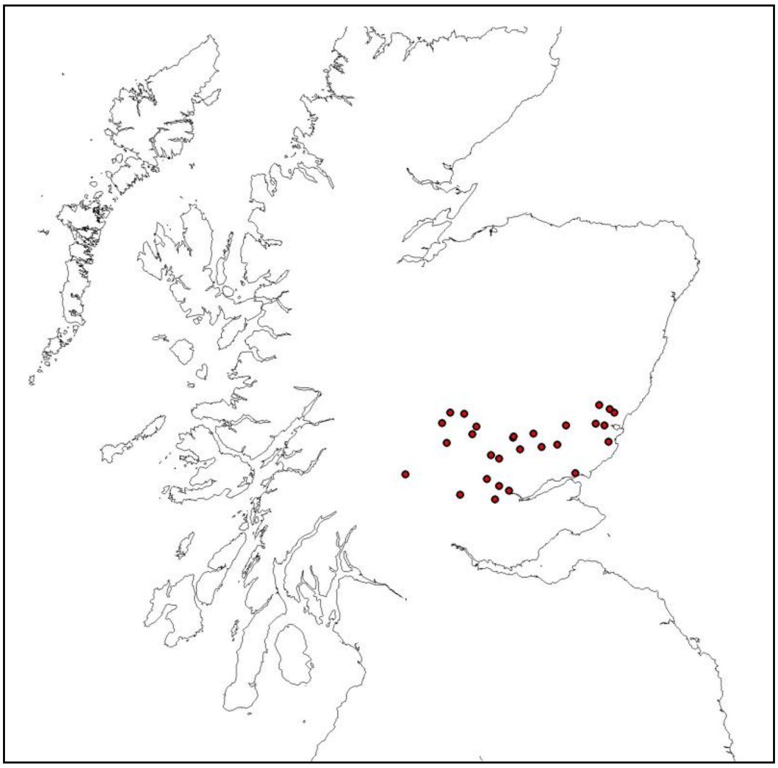


Figure 6 Map of the 28 sites in the Tay River Purification Board

From 1990 onwards the sites were sampled and processed following the now standard SEPA methodology with the same AQC/audit scheme as used in the 1990 GQA survey. Pre 1990 the exact methods are uncertain. The taxonomic data was stored on paper until approximately 2004, when the data was compiled into electronic format. The data has been checked extensively by Robin Guthrie (RG) of SEPA and was considered free of any systematic errors.

The mixed taxonomic level samples varied in the extent to which taxa were resolved, particularly for earlier samples. Generally, most taxa were taken to species or genus where possible with the exception of Oligochaeta (which were often but not consistently taken to family level and sometimes to species level), Chironomidae (sometimes sub-family), Sphaeriidae (Genus usually), Simuliidae often were only recorded as Simuliidae. Non-scoring Diptera were frequently left at family, as were other non-scoring taxa.

In the family level replicates the 8 artificial BMWP composite families were not distinguished. The families more recently regarded as composites, namely Siphlonuridae (including Ameletidae), Heptageniidae (including Arthropleidae) and Limnephilidae (including Apataniidae), were not split:

Taxonomic Resolution

The taxonomic resolution of the samples was then addressed. The original database contained 5 replicates from each site, in each year, in both spring and autumn:

Rep 1 - a Species level sample
Rep 2 - a Family level sample (including some non-BMWP families)
Rep 3 - a Family level sample (including some non-BMWP families)
Rep 4 - a Family level sample (including some non-BMWP families)
Rep 5 - a BMWP Family level sample generated by RG from Replicate 1

The original Replicate 1 (Species level sample) was retained (now called 1S) and a new Family level Replicate 1 was generated (now called 1F) based on the species sample giving the following replicates:

Rep 1S - Species level sample (processed at Species level in the first instance)
Rep 1F - Family level sample including some non-BMWP families (derived from replicate 1S)
Rep 2F - Family level sample including some non-BMWP families (processed at Family level)
Rep 3F - Family level sample including some non-BMWP families (processed at Family level)
Rep 4F - Family level sample including some non-BMWP families (processed at Family level)

Replicates 1F, 2F, 3F and 4F can be regarded as fully equivalent family level replicates. Replicate 1S provides further information on the prevalence of species within the dataset.

The next step was to convert the raw data TAXA table (containing the replicates above) into 5 different datasets, each adjusted to the exact taxonomic resolution for the calculation of the required biotic indices:

1) BMWP Family Level Taxa
2) AWIC Family Level Taxa
3) LIFE Family Level Taxa*
4) WHPT Family Level Taxa*
5) All Separated Family Level Taxa^

^Although the 5th dataset was not used for any specific indices it was created for future use with other types of indices that may include families not used in the above indices.

*For the LIFE and WHPT data, the BMWP artificial taxon groups from the raw data were split as follows:

Hydrobiidae (including Bithyniidae) are all taken to be Hydrobiidae
(as there were no species level records for Bithynia)

Planariidae (including Dugesiidae) are all taken to be Planariidae
(as there were only two species level records for Dugesia)

Ancylidae (including Acroloxidae) are all taken to be Ancylidae
(as there was only one species level record for Acroloxus)

Psychomyiidae (including Ecnomidae) are all taken to be Psychomyiidae
(as there were no species level records for Ecnomidae)

Dytiscidae (including Noteridae) are all taken to be Dytiscidae
(as there were no species level records for Noteridae)

Hydrophilidae (including Hydraenidae) are all taken to be Hydraenidae
(as there were only two species level non-Hydraenidae records)

Gammaridae (including Crangonyctidae & Niphargidae) are all taken to be Gammaridae
(as there were only 6 species level Crangonyx records and no Niphargidae)

Splitting the BMWP composite family Rhyacophilidae (including Glossosomatidae) proved to be more problematic as both of the families Rhyacophilidae and Glossosomatidae occurred frequently in the species level replicates. To derive separate Rhyacophilidae and Glossosomatidae records from the composite families, the species level replicate 1S was used to find out if either or both families were present in that replicate. The Rhyacophilidae (including Glossosomatidae) records at the other replicates at the same site on the same day were then split into separate Rhyacophilidae and Glossosomatidae records based on the species level sample. The abundances of the species in the species level sample were also used to calculate the proportion of Rhyacophilidae and Glossosomatidae across the overall $\log_{10}$ abundance of the combined Rhyacophilidae (including Glossosomatidae) records. This was then used to distribute the recorded $\log_{10}$ abundance of Rhyacophilidae (including Glossosomatidae) across the $\log_{10}$ abundance categories in the separate families Rhyacophilidae and Glossosomatidae in the same proportions.

Other families also regarded as composites were treated as follows:

Siphlonuridae (including Ameletidae) – were all regarded as Siphlonuridae
Heptageniidae (including Arthropleidae) – were all regarded as Heptageniidae
Limnephilidae (including Apataniidae) – were all regarded as Limnephilidae

Biotic Index Calculation
Biotic indices were calculated using the taxonomic levels 1) to 4) above. For LIFE and both the non-abundance weighted WHPT and abundance weighted WHPT indices, distinct families were used for scoring rather than BMWP composites.

2.2 Dataset 2: 416 East and North-East Scotland SEPA sites

This dataset generated by SEPA, and provided by Robin Guthrie (SEPA) comprised 416 sites predominantly from the East and North-East of Scotland and covered a wide range of Scottish river types from very large, oligotrophic rivers such as the Spey through to small, lowland streams in arable areas and rivers in predominantly urban settings (Figure 7).

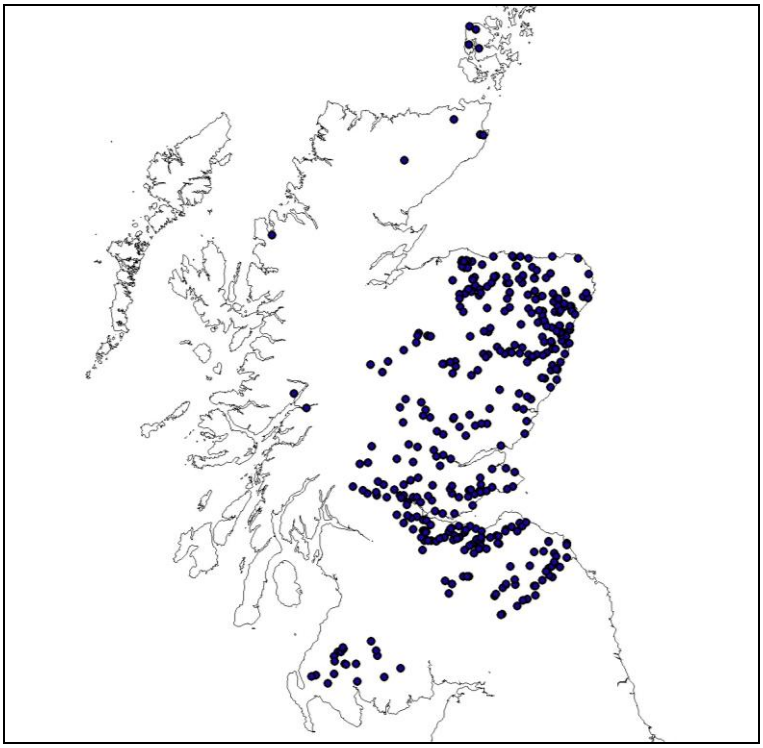


Figure 7 Map of the 416 sites in the East and North-East of Scotland

The sites ranged in quality from nearly pristine to very severely degrade. The range of impacts included organic pressures, hydro-morphological pressures, various toxic pressures, nutrient pressures and acidification.

The dataset has been compiled by RG from a range of databases held by the former River Purification Boards and from SEPA's current corporate systems. RG has checked the data extensively and was satisfied with the quality. The scores for all samples were consistent with RG's expectations for these sites (many of which RG was very familiar with). Additionally, the scores for samples from any given site were generally consistent with each other (low scoring sites tended to have consistently low scores while and high scoring sites tend to be consistently high). RG checked the unexpectedly low scores with local biologists and found them to be correct (i.e. not due to a data problem).

The sites were sampled between 1990 and 2004 and included samples from spring, summer and autumn in each year (although summer samples were fewer in number as monitoring over the later part of this period tended to be based primarily on spring and autumn samples alone). From 1990 onwards the samples were sampled and processed following the now standard SEPA methodology with the same AQC/audit scheme as used in the 1990 GQA survey. Should bias correction be required, RG has estimated that a figure of 1.7 net gains per sample would be appropriate as this was consistent with the overall SEPA performance at the time.

The samples were originally processed to a mixture of species and family levels with most sites typically having several family level samples and one species sample in a given year. RG has converted the taxonomic resolution of all of the samples to BMWP family level.

Where there were no abundances (numerical values) recorded, RG has generated synthetic abundances by allocating the average numerical abundance of that taxon at the recorded log abundance category, derived from that portion of the dataset where numerical abundances were recorded. Thus, to obtain an overall abundance for a family when deriving this from a species sample where the abundances were only recorded as log abundance RG allocated the mean numerical abundance value for that species at that log abundance category derived from those species samples where there were abundance counts. The synthetic abundance for each species within a family was then summed to give the overall family abundance.

RG also allocated synthetic abundances to all other family level records where there were no numerical abundances by deriving the mean numerical abundance of each family at each abundance category from data where the actual counts existed. In some instances (particularly for log abundance category E or for rare taxa) there weren't values available for each abundance category for that taxon so the mean numerical abundance of all taxa in that abundance category was used (e.g. the numerical abundance value 11000 was allocated to log abundance category E).

Taxonomic Resolution

The taxonomic resolution of the samples was then addressed. The taxonomic records as supplied by RG had been converted to BMWP family level. The next step was to convert the TAXA table into 5 different datasets, each adjusted to the exact taxonomic resolution for the calculation of the required biotic indices:

1) BMWP Family Level Taxa
2) AWIC Family Level Taxa
3) LIFE Family Level Taxa*
4) WHPT Family Level Taxa*
5) All Separated Family Level Taxa^

^Although the 5th dataset was not used for any specific indices it was created for future use with other types of indices that may include families not used in the above indices.

*For the LIFE and WHPT data, the BMWP artificial taxon groups from the raw data were split as follows:

Hydrobiidae (including Bithyniidae) are all taken to be Hydrobiidae
(as there was only one species level record for Bithynia)

Planariidae (including Dugesiidae) are all taken to be Planariidae
(as only 1.5% of the original species level records were for Dugesia)

Ancylidae (including Acroloxidae) are all taken to be Ancylidae
(as only 1.2% of the original species level records were for Acroloxus)

Psychomyiidae (including Ecnomidae) are all taken to be Psychomyiidae
(as there were no species level records for Ecnomidae)

Dytiscidae (including Noteridae) are all taken to be Dytiscidae
(as there were no species level records for Noteridae)

Hydrophilidae (including Hydraenidae) are all taken to be Hydraenidae
(as only 5% of the original species level records were non-Hydraenidae)

Gammaridae (including Crangonyctidae & Niphargidae) are all taken to be Gammaridae
(as only 3.5% of the original species level records were for Crangonyx
and there were no records for Niphargidae)

Rhyacophilidae (including Glossosomatidae) - both families occur commonly in Scotland so each record of Rhyacophilidae (including Glossosomatidae) was split into a record of Rhyacophilidae and a record of Glossosomatidae. Both were allocated the $\log_{10}$ abundance category of the composite family.

Other families also regarded as composites were treated as follows:
Siphlonuridae (including Ameletidae) – were all regarded as Siphlonuridae
Heptageniidae (including Arthropleidae) – were all regarded as Heptageniidae
Limnephilidae (including Apataniidae) – were all regarded as Limnephilidae

Within Seasons Replicates
A further element of variability that needs to be quantified is the extent to which samples taken on a different day vary (within the same season in the same year at the same site). Examination of the 416-site dataset has shown that there are 180 occasions where the same site was sampled in the same year and in the same season but on a different day. Typically there were either 2 spring samples taken or 2 autumn samples taken (replicate summer samples were much rarer). In a subset of 4 of these, 2 samples were taken in spring and 2 samples were taken in autumn, on different days, in the same year.

Biotic Index Calculation
Biotic indices were calculated using the taxonomic levels 1) to 4) above. For LIFE and both the non-abundance weighted WHPT and abundance weighted WHPT indices, distinct families were used for scoring rather than BMWP composites.

2.3 Dataset 3: 16 Biological Assessment Methods (BAMS) sites

The Environment Agency 16-site Biological Assessment Methods (BAMS) dataset was already used to quantify uncertainty in BMWP NTAXA and ASPT within RIVPACS III+ (Furse *et al*, 1995).The BAMS dataset comprised 16 sites covering a range of physical stream types within England and Wales and a range of biological qualities (Figure 8). Each site was sampled in the three RIVPACS sampling seasons spring, summer and autumn in 1994. In each season, 3 replicate samples were taken (two samples were taken by one operator and one sample was taken by a different operator).

Taxonomic Resolution
The samples were identified to BWMP family level and abundances were recorded as $\log_{10}$ categories. Although the BAMS dataset had only been identified to BMWP family level, the wide geographical coverage of the 16 sites made it difficult to develop rules to split the artificial BMWP composite groups into their constituent families and this was not attempted. All of the indices were therefore calculated using BMWP family level data.

Biotic Index Calculation
Biotic indices were calculated, although this time all using BMWP family level data. For LIFE and both the non-abundance weighted WHPT and abundance weighted WHPT indices, BMWP composite families were used for scoring as opposed to distinct families.

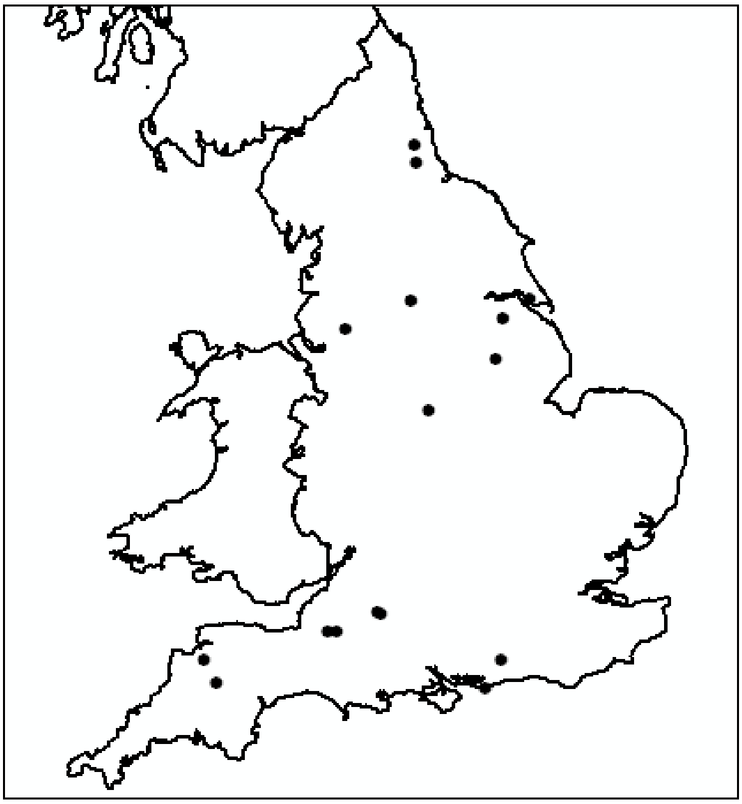


Figure 8. Map of the 16 BAMS sites in England and Wales

2.4 Dataset 4: 12 Northern Ireland Community Change Study sites

This Community Change Study dataset comprising samples from 12 sites in Northern Ireland (see map below) was supplied by Tommy McDermott (then of the Environment and Heritage Service, Lisburn (Figure 9). Each site was sampled in February, March, April, May, June, July, August, September, October and November/December 2006 and in January 2007. There were therefore 132 samples in all.

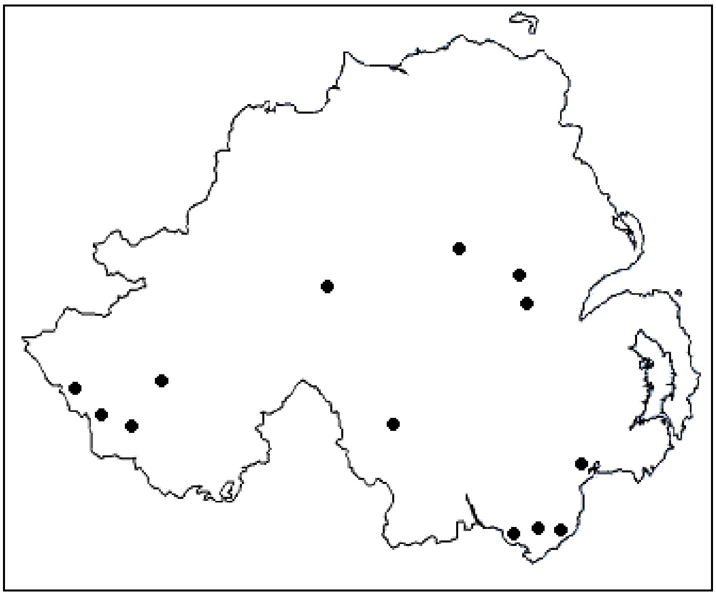


Figure 9 Map of the 12 Northern Ireland Community Change Study sites

The Community Change Study dataset is of particular interest because unlike the other datasets it contains replicate samples within each of the spring, summer and autumn RIVPACS sampling seasons with samples collected on different days. This dataset potentially enables estimates to be made of within-season sampling variability.

Taxonomic Resolution

All of the samples were identified to BMWP family level and most taxa were enumerated with $\log_{10}$ abundance categories. Taxa in some samples were only enumerated as presence/absence records and in these cases all taxa were simply regarded as having occurred at $\log_{10}$ abundance category 1. Information on the nature of the abundance data (true $\log_{10}$ abundance categories versus presence/absence based $\log_{10}$ category 1 assigned abundance) has been preserved in the collated dataset.

Replicate Data

The dataset was collated to obtain all valid combinations of single season samples from each of the 12 sites. To achieve this, the original 11 samples from each site were filtered down to include only those 9 months in the RIVPACS sampling seasons Spring (March, April and May), Summer (June, July and August) and Autumn (September, October and November/December). December samples were regarded as being suitable for inclusion as autumn samples.

Biotic Index Calculation

For LIFE and both the non-abundance weighted WHPT and abundance weighted WHPT indices, BMWP composite families were used for scoring as opposed to distinct families.

Particular attention is drawn to the fact that $\log_{10}$ abundance category data for some of the Community Change Study dataset was not available and that this will affect the abundance weighted WHPT and LIFE scores.

3. Estimation of sampling variance components for each index

3.1 Background requirements

This section focuses on developing estimation approaches, estimates and resulting algorithms for the sampling variance for a range of abundance-weighted macroinvertebrate indices required by the UK environment agencies for WFD river classification and/or regulatory purposes. This will include the WHPT abundance-weighted revised BMWP indices, the LIFE (family level) index and the relatively new PSI index. The indices and their abundance-weighted taxonomic scoring systems are described in Appendices 1-3.

The environment agencies are also moving from the use of single year's combined-season sample data towards the use of multiple years' macroinvertebrate data for stream WFD ecological status assessment. Estimates of sampling uncertainty and resulting confidence of status class are needed for these new multi-year average quality site assessments. Specifically, the agencies wish to base their site WFD status classifications on up to three years' worth of sample data in order to reflect the longer term underlying condition of the biology. For each metric, the agencies will use the average of the EQR values for each of the individual years available over the three year period of interest. Thus, class is defined for a three year period but does not necessarily require three separate years' data. If only one year's spring and autumn combined sample data were used it would still give an estimate of the three year mean condition. Three years' data would, however, give a more precise estimate.

Estimates of WFD class for sites are also still required for individual years.

The focus of this current project (workstream 1) is to develop sampling uncertainty methods and estimates to help allow the WHPT, LIFE and PSI indices to be used and incorporated into assessments of site ecological status. Unlike the previously-used BMWP indices which made use of only the presence-absence of taxa, these newer indices are abundance-based metrics where the abundance-dependent scores (weights) for each taxa have been derived either from prior statistical modelling (in the case of WHPT) or by agreement amongst a group of freshwater taxonomic experts (in the case of LIFE and PSI). The scores are only dependent on the RIVPACS log ten categories of abundance (1 = 1-9, 2 =10-99, 3 = 100-999, 4 = 1000+ ; see Appendices 1-3 for further details).

The problem is that the scores are based on either analysis or expert judgement for single season samples. Because the expected (and observed) abundances of individual taxa for combined season samples will generally be higher, it was perceived by the UK environment agencies that this would invalidate the use of these indices in combined season (e.g spring and autumn) sample assessments. However, in two-season combined samples, the expected log abundance category of a taxa usually only increases by about one category. Also, as two-season combined observed sample index values (O) are compared with the appropriate two-season combined RIVPACS expected index values (E) as O/E ratios and EQR, like is still compared with like, and the assessment would still have been valid, albeit with maybe slight sub-optimal abundance-dependent weights for some taxa.

However, a separate reason for using single season samples, rather than combining them into two- or three-season lumped samples, is that there is merit and additional information in calculating the O/E (and EQR) for each single sampled season and then defining the 'overall' WFD site quality for either a single year or a three-year period as the average of the single season O/E (EQR) values available for that year or that three-year assessment period.

3.2 Approach and data limitations on variance component estimation

Estimates of the various variance components were obtained by analysis of variance and hierarchical mixed modelling techniques using a blend of four previously available datasets described in Section 2, namely:

- BAMS dataset: 16 sites throughout England (4 types by 4 quality classes), each with 3 replicates in each of 3 seasons in one year

- TAY dataset: 28 Tay River Purification Board sites, including 18 sites with 4 replicate samples in each of spring and autumn for each of five years (1990-1994)

- SEPA dataset: 416 sites covering a much wider area of Scotland and with c.200 instances of samples being taken on more than one day within the same RIVPACS season of a year.

- NI dataset: 12 sites in Northern Ireland sampled once each month in one year

Estimation of within-season temporal variability in each index requires cases where RIVPACS samples have been taken on different dates within the same RIVPACS season (spring, summer or autumn) and more than one replicate sample on at least some days. No such datset is available. Data on sites sampled more than once in the same season are only available for the 12 site Northern Ireland dataset and the 416 site SEPA dataset, where there are 181 situations where two (and in three cases three) sample were taken on different days (and usually months) within the same season of the same year.

Ideally, at these same sites and seasons, there would also be replicate samples taken on the same day, so that we could easily ‘subtract’ away the variability between samples on different days which was due to the fact that any two replicate samples vary. However, neither the NI dataset nor the 416 site SEPA dataset has any same-day sample replication. Therefore, to estimate the variance due to real within-season temporal variability, we needed to analyse the 416 sites SEPA dataset combined with the other datasets. We could have just combined the SEPA dataset with the 28 sites Tay dataset, as both datasets are sites in Scotland, which might be expected to make the sampling variability more similar. However, the 416 SEPA sites cover a much wider geographical and environmental range than the 28 Tay sites. For this reason, it was considered best to also include the 16 BAMS sites dataset, even though, as mentioned before, the 28 Tay sites were sampled in more years and thus carrying far greater weight in determining the overall estimate and replicate sampling SD. These analyses of replicate sampling variation also assessed the best transformation (square root, logarithim or Arcsine square root for proportions or percentages) of index values to mimimise dependency of the size of the replicate sampling variablility to vary with the index values.

A further requirement is to allow the agencies to make assessments of site ecological status based on average quality over a three-year period. The uncertainty in these estimates when all three years are not sampled will depend on inter-year variance in index values due to differences between years in the (unknown) average index values for each year. Therefore, we need to derive an estimate of the inter-year variance parameter for three-year periods rather than over all years sampled at each site within the datasets. This was done by coding the years into three-year periods as follows: (1987-89, 1990-92, 1993-1995, 1996-98, 1999-2001, 2002-04).

The statistical estimation of parameters was carried out using a hierarchical model with the following variance components (Standard Deviation (SD) is the square root of the Variance(Var)) :

$\mathrm{Var}_{\mathrm{Rep}} = (\mathrm{SD}_{\mathrm{Rep}})^2$ = Replicate sampling variance
$\mathrm{Var}_{\mathrm{TSeas}} = (\mathrm{SD}_{\mathrm{TSeas}})^2$ = Within-season temporal variance
$\mathrm{Var}_{\mathrm{TYear}} = (\mathrm{SD}_{\mathrm{TYear}})^2$ = Inter-year within 3-year period temporal variance
+
$\mathrm{Var}_{\mathrm{TPeriod}} = (\mathrm{SD}_{\mathrm{TPeriod}})^2$ = Inter-period variance
$\mathrm{Var}_{\mathrm{Site.Seas}} = (\mathrm{SD}_{\mathrm{Site.Seas}})^2$ = Variance due to differences between all site x season combinations

The last two parameters are of less interest and the ‘all Site by Season combinations’ component could perhaps have been considered as a fixed effect factor, but their effects needed to be allowed for in order to estimate the important three lower-level parameters appropriately.

The above SD component parameters which can be estimated directly, or partially, using information from each dataset are indicated in Table 2.

Table 2 Components of variability which can be estimated, or for which there is information, within each dataset (indicated by ticks).

Variability component	SD	28 TAY sites	416 SEPA sites	16 BAMS sites	12 NI sites
Replicate sampling	SD_{Rep}	✓		✓	
Within-season Temporal	SD_{TSeas}		✓		✓
Inter-year Temporal	SD_{TYear}	✓	✓		

Strictly speaking the temporal SD parameters, SD_{TSeas} and SD_{TYear}, can only be estimated from analyses of variance based on the 416 SEPA sites dataset when that data set is combined with the other two datasets in order to enable us to 'subtract' the variance due to replicate sampling effects from the overall observed temporal variation in index values.

Strictly speaking, the 12 site NI dataset, which has a single replicate in each RIVPACS sampling month (Mar-Nov) in 2006, can only be used to estimate the combined effect of replicate variability (SD_{Rep}) and within-season temporal variability (SD_{TSeas}).

However, by assuming the average replicate sampling variance, and typical within-season temporal variability in index values is constant across all river sites, these datasets were then analysed in appropriate combinations to derive estimates of the various variance components for each index.

Initially, the BAMS and TAY datasets were analysed separately, then together and finally as weighted averages (based on the number of sampled sites) to derive a range of estimates of the replicate sampling standard deviations (SD_{Rep}) for each index for single season samples.

Then the temporal variance components (within-season and inter-year) were estimated by integrated analysis of the combined BAMS, TAY and SEPA datasets. The NI dataset was used as a check on the within-season temporal variance estimates.

Simultaneous estimation of the replicate and temporal sampling variances for each index was done by fitting statistical mixed (random and fixed) effect models to the combined sample data from these three mainland UK datasets described. The mixed models were fitted to the sample values of each index, where needed on the transformation scale which we had previously determined would make the replicate sampling variance between sites least heterogeneous. The mixed models were fitted using the 'lmer' function in the freely-available 'R' software package (version 3.02).

3.3 Estimates of replicate sampling variability and transformation scale

As a first stage, preliminary analyses and plots were used to assess whether the variation between replicate samples for a particular index tended to be greater for sites with either larger average values of the index or perhaps with fewer index-scoring taxa present. If sampling variance increased with site mean index value, then a transformation (such as square root or logarithm) of the raw observed index values may help make the sampling variability more constant between sites (when re-analysed on the transformed scale). This constancy of variance is a desirable property when trying to apply variance estimates to other sites for which no replicate samples exist. From an analysis of the BAMS dataset, Furse *et al* (1995) and Clarke *et al* (2002) found that both BMWP score and BMWP NTAXA replicate sampling variance increased with site mean replicate value and that by working with the square root transformed values, the replicate sampling variance was roughly constant and independent of both the quality and physical type (RIVPACS end-group) of river site.

The first step in assessing the variability in biotic index values due to replicate sampling effects was therefore to plot the sampling standard deviation of replicate samples from the same site, year and season against the mean value of those replicate samples to help assess whether, and how, the variability between replicate samples varies with the general level of each index. This was done for the two datasets with replicate samples, namely the 16 BAMS and 28 Tay sites.

The approach in the previous version of RIVPACS to simulating index uncertainty was to find the best transformation of observed values of an index to make the replicate sampling variability as homogeneous as possible, or at least not varying systematically with the replicate mean value. An established method of determining an appropriate transformation with these types of metric is to regress the logarithm of variance of replicate samples against the logarithm of mean of the replicate samples, as: Log (Replicate variance) = *a* + *b* Log (Replicate mean) (Clarke *et al* 2002). The regression slope *b* indicates the power with which the variance amongst replicate increases with their mean value. Moreover, values of *b* of around 1.0 and 2.0 indicate, respectively, that a square root transformation and a logarithmic transformation of index values will make the replicate variance (of transformed values) more independent of the replicate mean value. Values of *b* around zero indicate that no simple power transformation will either be needed or be effective in making the replicate variance for the index more homogeneous across all sites. Negative values of *b* would suggest that variation in the index values between replicates tends to decrease as their value increases.

3.3.1 Abundance-weighted WHPT indices replicate sampling SD

Figure 10 highlights the differences between the two replicated sample datasets in the typical WHPT index values. The BAMS sites were originally carefully selected to cover a wide range of biological qualities and thus encompass a wide range of values of BMWP score, NTAXA and ASPT and also WHPT indices. In contrast, the vast majority of samples from the Tay dataset are from high or reasonable good quality relatively taxon-rich sites with generally high values of WHPT NTAXA and WHPT ASPT, although the high values may be partly due to the natural physical characteristics of these sites. However, the plots of replicate SD against replicate mean for the two datasets overlap with no major systematic differences in the SD to mean relationship (Figure 10), which is encouraging; but this is assessed in more detail below. Together the two datasets complement each other by covering most of the practical range of values recorded for these indices in any field samples, which means that variance parameter estimates based on their combined analysis should provide reasonable estimates for general use in uncertainty simulation software.

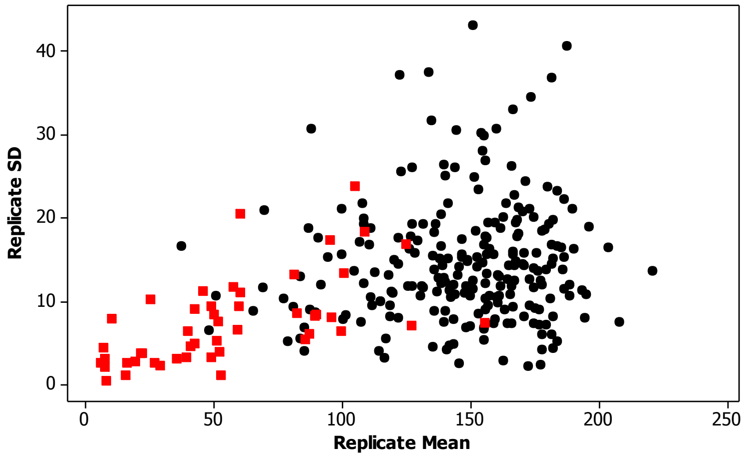
<u>WHPT Score</u>
The log variance to log mean regression slopes for the unweighted WHPT score and abundance-weighted WHPT score indices are all greater than one (1-07 – 1.21) but not statistically significantly different from 1.0 (95% confidence limits for slope (b +/- 2SE(b)) encompass 1.0). This indicates that the square root transformation is optimal (Table 3). This is the same as previously found by Clarke *et al* (2002) for the original BMWP score and as used in RIVPACS III+ and RIVPACS IV in the current version of RICT.

Table 3 Taylor's power law regressions of log replicate variance again log replicate mean for the single season samples for each biotic index based on (a) 16 BAMS sites and (b) BAMS + Tay datasets combined; b = regression slope, SE(b) = standard error of b, r^2 = % variation explained

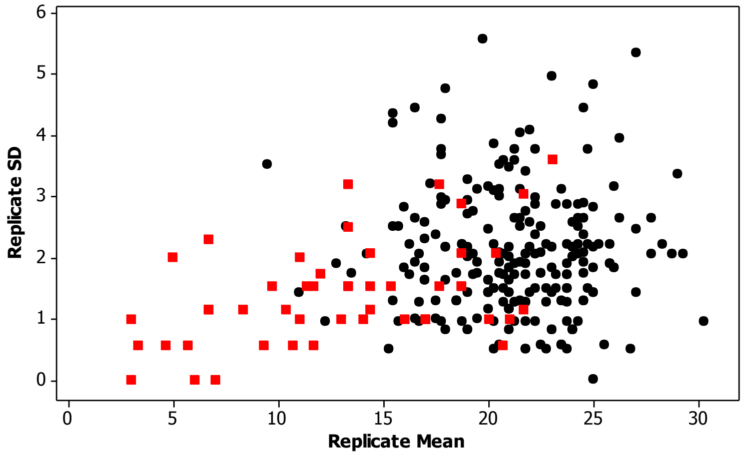
		(a) BAMS sites			(b) BAMS + Tay sites		
		b	SE(b)	r^2	b	SE(b)	r^2
Original BMWP	BMWP Score	1.23	0.22	42%	1.07	0.12	21%
	NTAXA	0.92	0.26	22%	0.81	0.18	7%
	ASPT	0.26	0.88	0%	0.03	0.30	0%
WHPT Non-weighted	WHPT Score	1.20	0.23	39%	1.07	0.13	20%
	WHPT NTAXA	0.77	0.29	14%	0.81	0.18	7%
	WHPT ASPT	0.61	0.84	1%	0.47	0.29	1%
WHPT Abundance-weighted	WHPT Score	1.21	0.21	42%	1.14	0.10	29%
	WHPT NTAXA	0.77	0.29	14%	0.81	0.18	7%
	WHPT ASPT	-1.40	0.54	13%	-0.36	0.20	1%
LIFE (family level)		-2.94	1.75	6%	-1.41	0.66	2%
PSI (family level)		0.38	0.23	6%	-0.02	0.14	0%

Figure 10 Plot of the relationship between replicate sampling SD and mean of the replicate single season sample values for all available combinations of sites and seasons with replicate sampling for the 16 BAMS sites (■) and the 28 Tay sites (●) for the abundance-weighted (a) WHPT score, (b) WHPT NTAXA and (c) WHPT ASPT.

(a) Abundance-weighted WHPT score



(b) Abundance-weighted WHPT NTAXA



(c) Abundance weighted WHPT ASPT

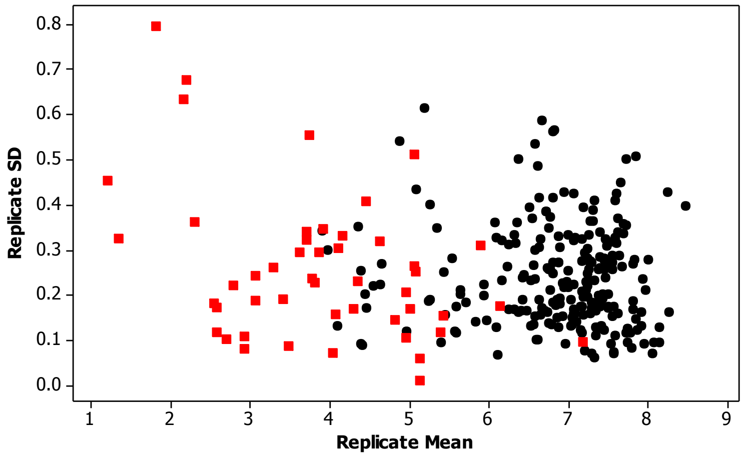
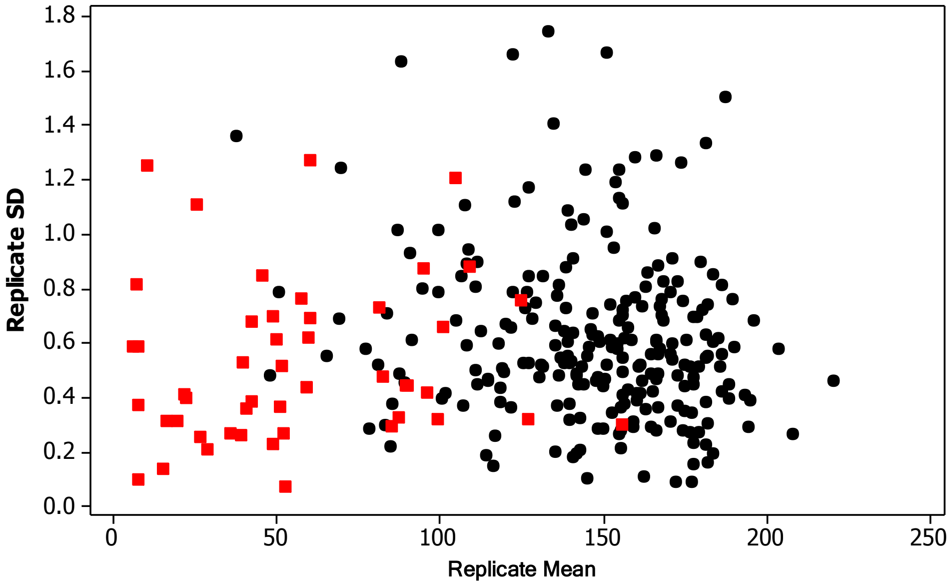
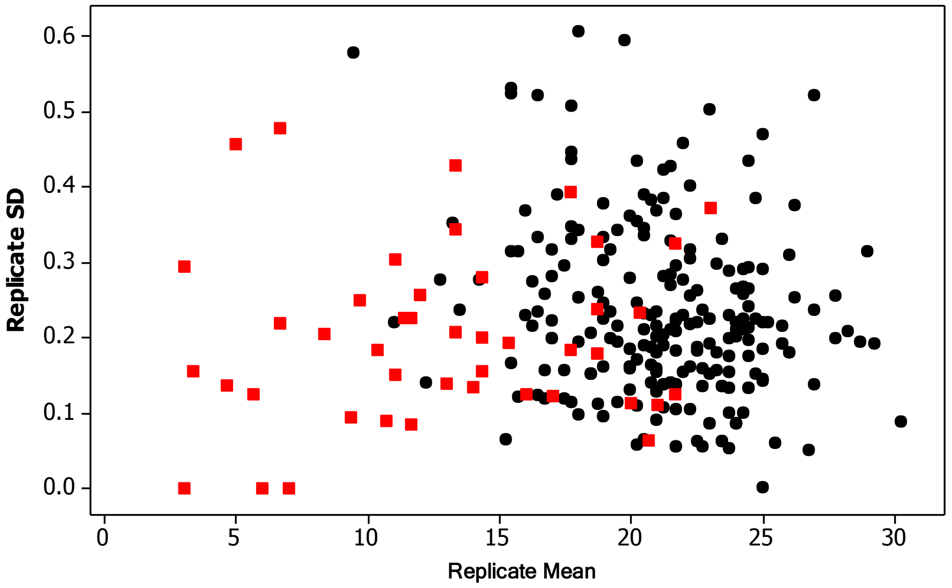


Figure 11 Plot of the relationship between the replicate sampling SD and mean of the replicate single season sample values for all available combinations of sites and seasons with replicate sampling for the 16 BAMS sites (■) and the 28 Tay sites (●) for (a) the square root of abundance-weighted WHPT score and (b) the square root of WHPT NTAXA.


(a) Square root of abundance-weighted WHPT score


(b) Square root of abundance-weighted WHPT NTAXA

<u>WHPT NTAXA</u>
Similarly, for the WHPT revised form of NTAXA, (unweighted WHPT NTAXA and abundance-weighted WHPT NTAXA are the same), the log-variance versus log mean regression slopes range from 0.77 – 0.81, but they are not statistically significantly different from 1.0 (Table 3).

This indicates that analysing variation in the WHPT NTAXA index on the square root scale is also an appropriate way of reducing systematic between-site heterogeneity in the replicate sampling standard deviation (SD_{Rep}).

The implication, by assumption, is that the square root transformation will also make the temporal variability terms, SD_{TSeas} and SD_{TYear}, of the transformed values of WHPT score and WHPT NTAXA index values less heterogeneous between sites.

By working with the square root of the sample values of abundance-weighted WHPT score or WHPT NTAXA, the replicate sampling SD no longer varies in any systematic way with the average size of the replicate values (Figure 11). This indicates that the square root transformation helps us to make the inter-sample variability of such indices more consistent across sites, increasing the justification for the use of single overall SD parameter estimates for variability in BMWP score and NTAXA (on this transformed scale).

WHPT ASPT

There appears to be no major systematic variation between the SD of replicate sample values of WHPT ASPT and the replicate mean value of WHPT ASPT, whether unweighted or weighted for taxon abundance (Figure 10(c)). However, there is some suggestion that replicate variation in the revised abundance-weighted form of ASPT may be greater for some very poor quality sites (i.e. two BAMS sites) with WHPT ASPT values less than 2.5 (and few taxa) (Figure 10(c)).

The original BMWP scoring system scored BMWP families on the scale 1-10, so that ASPT must lie between 1 and 10. However, in practice ASPT values greater than 8 are very rare (only 7 of 17011 samples amongst the three study datasets) because 10-scoring sensitive taxa always occur with many lower-scoring less sensitive BMWP families.

With the revised WHPT scoring system, the abundance-weighted scores vary from -1.6 for 1000+ Asellidae to 13.0 for 10+ Perlidae in a sample. Amongst the samples in the four datasets analysed here, the sample abundance-weighted WHPT ASPT values varied from 0.7 to 9.3. The taxa present in taxon-rich sites tend to have a wider range of WHPT scores than those few taxa occurring at poorer quality sites with lower WHPT ASPT values. However because the WHPT ASPT values of taxon-rich sites are based on the (weighted) average of the WPHT scores of many more taxa, they tend to be no more variable between replicate samples than between replicate values at poorer quality taxon-poor sites with low WHPT ASPT values (Figure 10(c)).

Therefore, overall across all sites, no transformation of WHPT ASPT values is needed to remove any systematic differences in sampling variability.

Summary

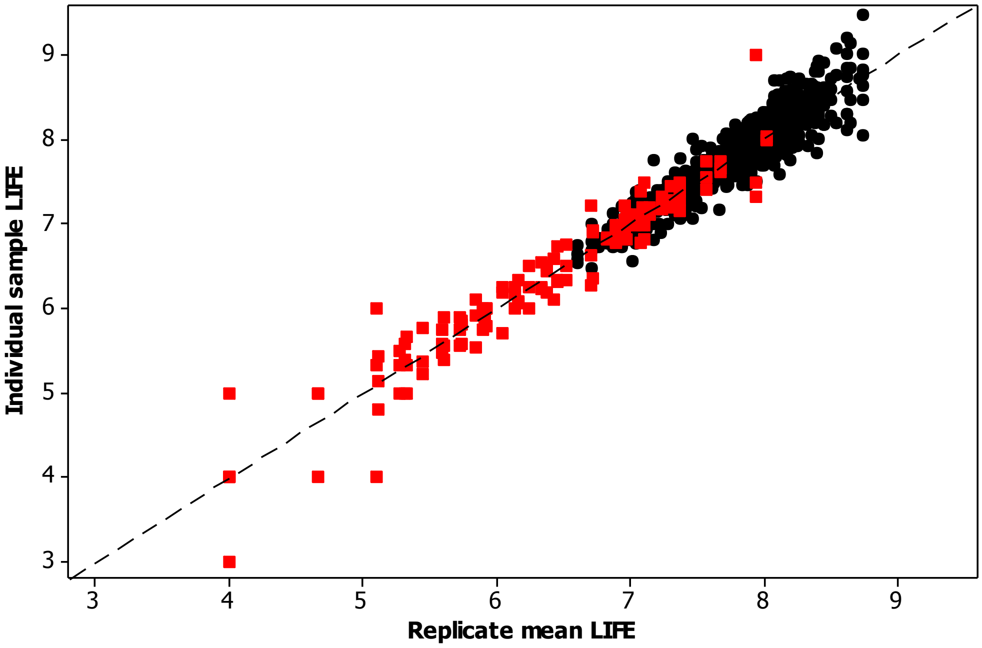
The square roots of the observed sample values of either the abundance-weighted WHPT score or the WHPT NTAXA indices have approximately constant size replicate sampling variance which does not depend on the index value. Therefore RICT simulation of overall sampling variation in these indices for any particular site should be based on square root transforming observed sample index values, then repeatedly adding a random sampling variation term with the best estimate constant variance and back-transforming (i.e. squaring) to get a simulated distribution of possible observed index values for that site.

Replicate sampling variance of abundance-weighted WHPT ASPT does not appear to vary systematically across its range and across sites; therefore no transformation is needed and single estimates of the replicate sampling variances (and by implication the other temporal sampling variance components) can be used for all sites.

3.3.2 LIFE (family) replicate sampling SD

The LIFE index (Extence *et al* 1999) is an abundance-weighted 'average-score-per-taxon' index, with taxa scores varying from 1 to 12 (see Appendix 1 for further details of taxon abundance-weighted flow-group scores). However, in practice most sites' family-level LIFE values vary between 5 and 9 (Figure 12).

Figure 12 Plot of individual sample LIFE values in relation to the replicate mean LIFE value for each site by season combination of the 16 BAMS sites (■) and the 28 Tay sites (●)

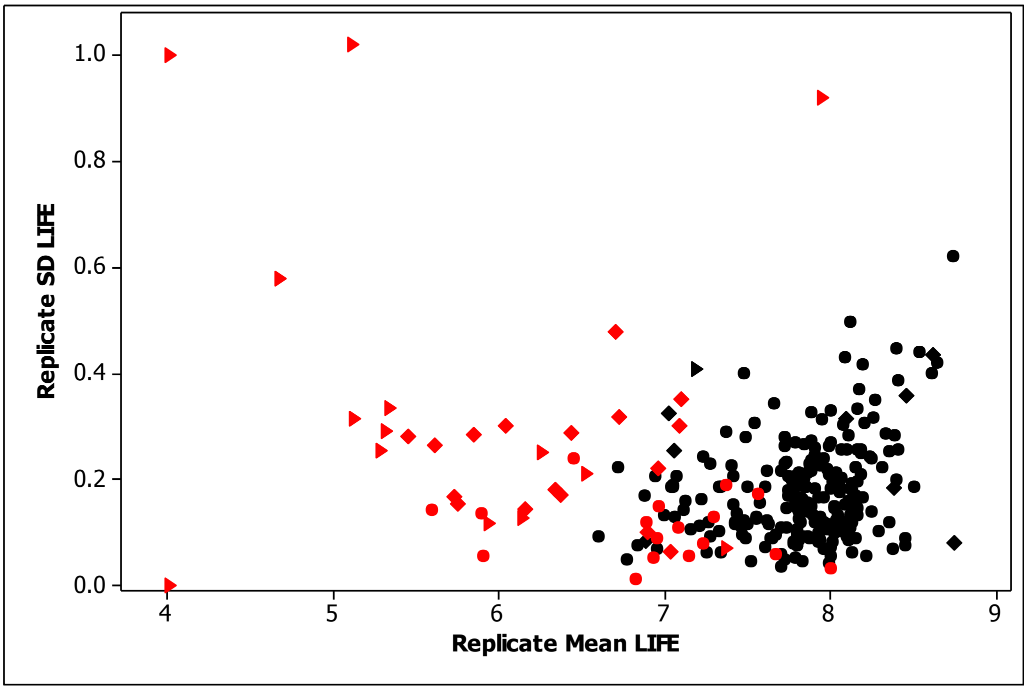


Amongst the 16 BAMS sites, the replicate sampling SD of LIFE values varied from 0.000 (discussed below) to 1.018, with a mean and median of the SD equal to 0.236 and 0.170 respectively. Amongst the 28 TAY sites, the SD in LIFE values between replicate samples from the same site, year and season, varied from 0.033 to 0.619, with a mean and median of the SD equal to 0.178 and 0.161 respectively.

There was no apparent general relationship between the replicate SD of single season sample LIFE values and the mean of the replicate sample LIFE values; the log variance – log mean regression relationship did have negative slopes, suggesting some tendency for replicate sampling variance to decrease with site replicate mean LIFE value, but the relationship was not significant or did not explain much variation ($r^2 \leq 6\%$) (Table 3, Figure 13). This was also the conclusion of Clarke *et al* (2003) in their corresponding analyses based on just the BAMS dataset. Clarke *et al* (2003) also concluded that the sampling SD of LIFE does not vary systematically between different types of site or between seasons.

However, in an earlier investigation of LIFE sampling variability, Clarke *et al* (2003) found that although the sampling SD does not appear to vary with the mean of the replicate values of LIFE, some pattern emerges when the SD of replicate samples from a BAMS site was plotted against the mean number of LIFE-scoring families involved in calculating the replicate values of LIFE for that site.

Figure 13 Relationship between replicate sampling SD family-level LIFE values and mean of the replicate single season sample LIFE values for all available combinations of sites and seasons with replicate sampling for the 16 BAMS sites (▶,♦,●) and the 28 Tay sites (▶,♦●) with symbol based on the replicate mean NTAXA (≤ 10 (▶), 10-15 (♦), 15-32 (●)).



To be of practical use within RICT, we need a predictive relationship to estimate and simulate the expected sampling uncertainty in LIFE values from readily available information about any particular river site. The value of the index BMWP NTAXA is input and therefore known for any sample presented for RICT classification, whereas the number of LIFE-scoring families present in a sample is not. The number (N_{LIFE}) of LIFE-scoring families present in a sample is very highly correlated with the number (NTAXA) of BMWP families present. Amongst the BAMS samples, the numbers (N_{LIFE} and NTAXA) never differed by more than two taxa for the same sample and their correlation *r* was >0.999.

There is some suggestion, especially amongst the BAMS sites that replicate SD of LIFE tends to be higher when there are fewer taxa present. However most cases with low replicate mean LIFE (i.e. <6) tend to have few taxa present but a wide range of SD; so the pattern is not entirely clear (Figure 13).

To investigate this further in this study, we aimed to relate the LIFE replicate sampling SD for a site to the mean number of BMWP taxa present in the replicate samples from the site (Figure 14). The highest values of SD (i.e. >0.7) all occurred when the replicate values of LIFE were based on few families, as indicated by having any average of less than seven BMWP families. At the other extreme, when the average number of BMWP families found in replicate samples was at least 20, the estimated sampling SD was nearly always relatively small (i.e. <0.3).

This potential for increased sampling variability at sites with few families present is illustrated by BAMS site 4 in spring, which has a very high average LIFE score, but it is still very variable between replicate samples. The second and third replicate samples had similar values of LIFE (7.33 and 7.50) both based on six families, but sample 1 only had two LIFE-scoring families present, Baetidae at log abundance category 3 and Simuliidae at log abundance category 1, both in LIFE flow group II, giving a value of LIFE of 9.00. This gave a SD between the three replicates of 0.92 (pointed marked Y in Figure 14(b).

When few LIFE-scoring families are present at a site, the sampling variance of LIFE is more volatile and potentially more difficult to predict. As an example of one extreme, all three replicate samples at Site 16 in summer contained only Hydrobiidae at log abundance category 3 (plus the ubiquitous Oligochaeta and Chironomidae, which are ignored in the LIFE system). All three samples therefore had values of LIFE of 4.00 and hence an estimated sampling SD of zero. Finding just one more family in one sample could have given a quite different value for LIFE and hence estimated SD and therefore this site (shown as point Z in Figure 14(b)) was excluded from the subsequent analyses (Figure 14(a)).

We concluded that the sampling SD of LIFE does tend to decline systematically with an increase in the number of families present (as represented by BMWP NTAXA) (Figure 14(a)). The relationship is best estimated by a linear regression relationship between log SD and *NTAXA*, which is statistically significant ($r = -0.37$; $p < 0.001$), explains 14% of the variation and is given by (standard errors of regression coefficients given underneath in brackets):

$$\log_e \text{SD LIFE} = \underset{(0.147)}{-0.891} - \underset{(0.0074)}{0.0499}\,NTAXA \qquad \text{(Eqn3.1a)}$$

When back-transformed (by taking exponential of both sides of the equation), the predicted relationship is:

$$\text{sampling SD LIFE} = 0.410(0.951)^{NTAXA} \qquad \text{(Eqn3.1b)}$$

which is superimposed as the solid line in Figure 14(b). [Note: Because $\log_e$ Variance LIFE = 2 $\log_e$ SD LIFE, the same relationship would effectively be derived by initially fitting $\log_e$ Variance LIFE against NTAXA].

To overcome any concerns that this relationship may be wholly or partly due to differences between the Tay and BAMS datasets in both general level of NTAXA and of replicate SD, we re-fitted the relationship allowing for dataset differences (as represented by the variable *DSET*: where $DSET = 1$ for BAMS sites and $DSET = 0$ for Tay sites) and their interaction with NTAXA relationships to give:

$$\log_e \text{SD LIFE} = \underset{(0.144)}{-0.844} + \underset{(0.111)}{0.424}\,DSET - \underset{(0.0091)}{0.0707}\,NTAXA$$

$$\log_e \text{SD LIFE} = \underset{(0.210)}{-0.312} + \underset{(0.312)}{0.581}\,DSET - \underset{(0.0148)}{0.1111}\,NTAXA + \underset{(0.0185)}{0.0635}\,NTAXA.DSET$$

Although the estimates of regression slope naturally changed to some extent, the relationship with NTAXA was always highly statistically significant (all test $p < 0.001$), supporting our conclusion of a real effect of NTAXA on sampling SD of LIFE. As a dataset-dependent relationship is of no practical use for the general prediction and simulation in RICT of the sampling variance of LIFE for other river sites, it is best to use the single relationship of equation (Eqn3.1).

Equation (Eqn3.1) can be used to provide an estimate for the unknown replicate sampling SD of single season samples for any site using just the observed number (NTAXA) of BMWP families present in a sample; examples are given in Table 4. In reality, very few samples have fewer than four BMWP families present, so the usual range of estimates of replicate SD of LIFE across the spectrum of sites is from about 0.34 down to about 0.09.

Figure 14 Relationship between replicate sampling SD of family-level LIFE values and mean of the replicate sample BMWP NTAXA values for all available combinations of sites and seasons with single season replicate sampling for the 16 BAMS sites (■) and 28 Tay sites (●); (a) shows fitted line regression line (solid) and fitted non-linear LOWESS line (dashed) to log SD, (b) shows back-transformed linear regression predictions for LIFE replicate SD for a mean value of NTAXA for a site; points Y and Z are discussed in text.

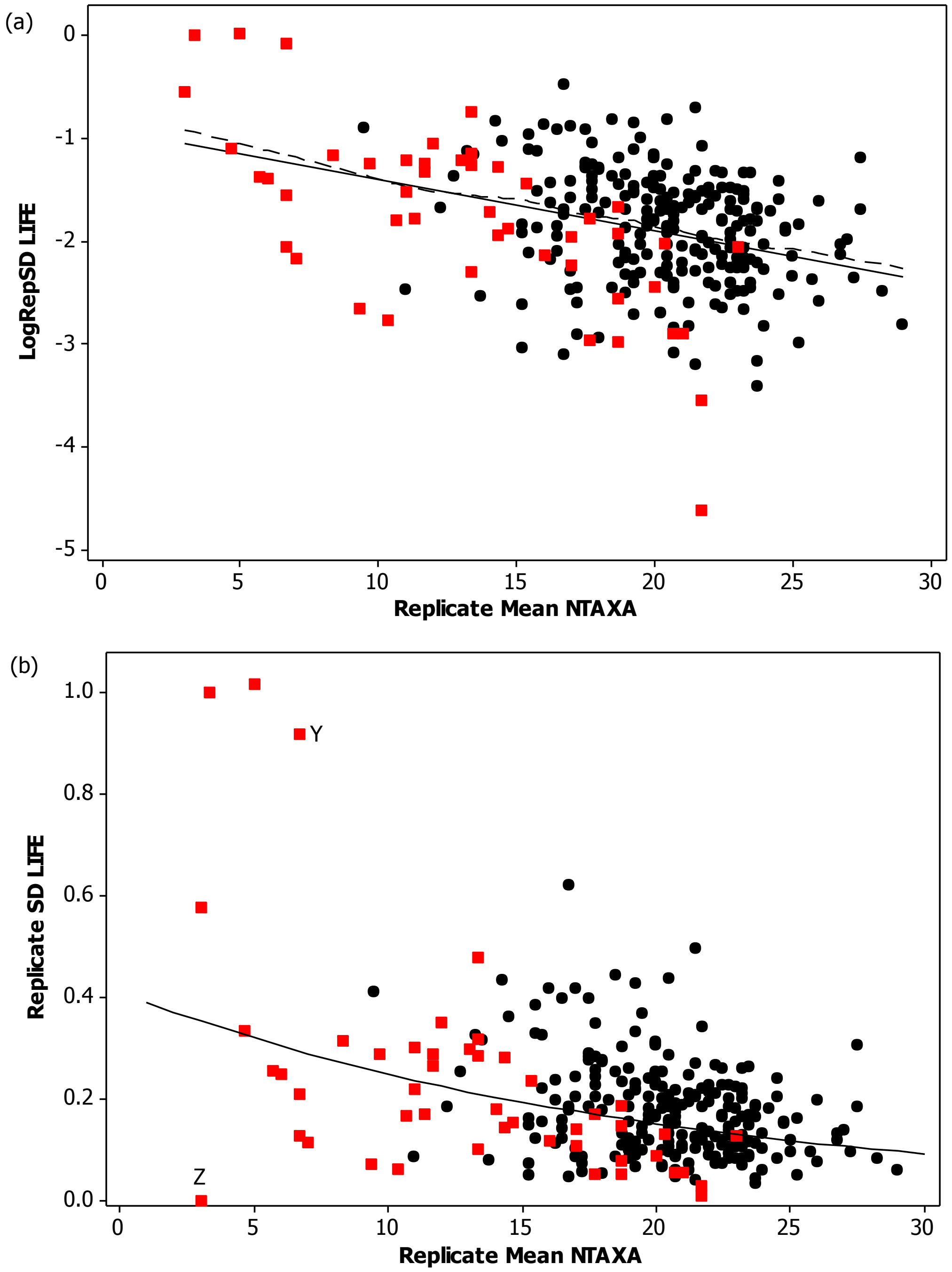


Table 4 Estimates of replicate sampling standard deviation (SD) of observed LIFE for sites based on the average number (*NTAXA*) of BMWP families present in a sample from that site (estimates based on equation (Eqn 3.1))

Number of BMWP families present (*NTAXA*)	LIFE Sampling SD
1	0.390
2	0.371
3	0.353
4	0.336
5	0.320
6	0.304
7	0.289
8	0.275
9	0.262
10	0.249
12	0.225
15	0.194
20	0.151
25	0.118
30	0.092

The implication is that when very few taxa are present, variation between possible replicate samples in LIFE value tends to be greater, so the uncertainty variance and the confidence limits for LIFE EQR are greater and a larger change in value of LIFE between years would be needed to have any confidence that the difference was not just due to chance sampling variation.

Summary:
We recommend assuming sampling SD of LIFE is constant across the range of LIFE values, but varies with the number of BMWP taxa present (NTAXA) in the mathematical form provided by equation (Eqn 3.1), as presented in Table 4 and Figure 14.

3.3.3 PSI replicate sampling SD

The PSI (Proportion of Sediment-sensitive Invertebrates) index is a new index developed by Chris Extence and colleagues (Extence *et al* 2011) which measures the abundance-weighted percentage frequency of taxa which are sensitive to fine sediment deposition. In this study, we are developing methods and estimates of sampling uncertainty in PSI derived from family-level taxonomic sample data ready for its inclusion in a future version of RICT. Table 5 and Appendix 3 contains further details of the PSI index definition in terms of the individual families involved and their ascribed sediment sensitivity grouping.
The PSI index is defined as:

$$\text{PSI} = \frac{\text{Sum of Ss Scores for observed taxa in Sediment Sensitivity Groups A \& B}}{\text{Sum of Ss Scores for observed taxa in all Sediment Sensitivity Groups A-D}} \times 100$$

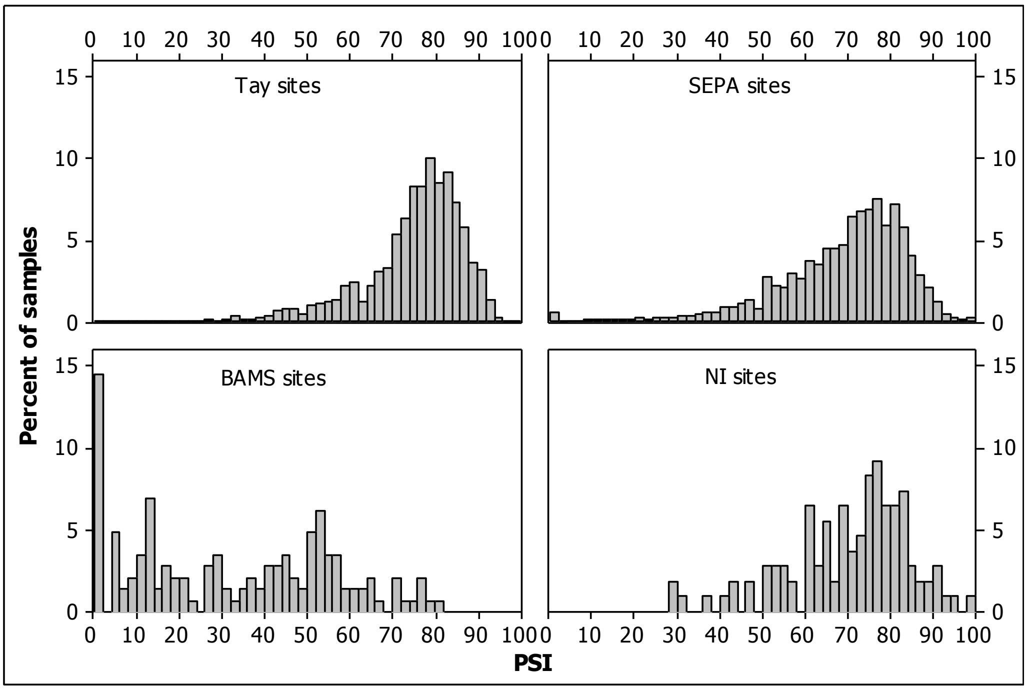
Table 5 Number of families in each PSI sediment sensitivity group

Sensitivity Group	**Sensitivity Group Description**	**Number of families**	**Log$_{10}$ Abundance Category (individuals)**			
			1 (1-9)	**2 (10-99)**	**3 (100-999)**	**4+ (1000+)**
A	Highly Sensitive	20	2	3	4	5
B	Moderately Sensitive	21	1	2	3	4
C	Moderately Insensitive	13	1	2	3	4
D	Highly Insensitive	41	2	3	4	5

Sampling variation in PSI values calculated at the family level was assessed using the same four datsets defined in Section 2, as used to assess sampling variation in the WHPT and LIFE indices.

The distributon of the individual sample values of PSI for the four datasets are compared in Figure 15.

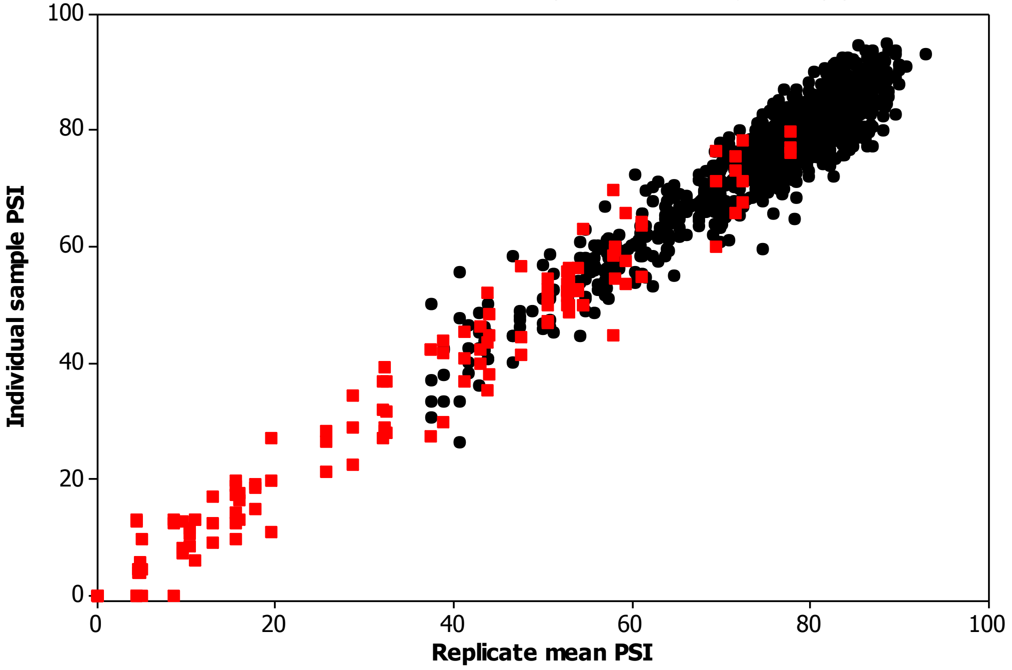
Figure 15 Histogram of the individual sample values of PSI for the four datasets



The 16 BAMS sites include both high and poor quality sites whose PSI values range from 0 to 80 with a median value of 32, whilst the 28 Tay sites, mostly of good/high quality, had sample PSI values varying from 28 to 95 with a median of 78. Thus together these two datasets should provide estimates of sampling variability in PSI for sites throughout most of its realised range (Figure 16). The SEPA

dataset PSI values have a range of 0 - 100, with a median of 77, whilst those of the Northern Ireland (NI) sites have a range from 29 – 100 with a median of 74.

Figure 16 Plot of individual sample PSI values in relation to the replicate mean PSI value for each site by season combination of the 16 BAMS sites (■) and the 28 Tay sites (●)



The relationship between the replicate sampling SD in PSI values and the replicate mean PSI value for the same site-season combination seems to be more complex. For BAMS sites which varied in both physical type and quality, there were seven site-season combinations with replicate mean PSI less than 10 but they had SD broadly encompassing the full range of SD observed within the BAMS and Tay datasets (Figure 17).The Scottish Tay sites tended to have higher PSI values and the average replicate SD may be slightly less for such high-PSI sites.

However, unlike the BMWP, WHPT and LIFE indices, the PSI index is a form of percentage, which may make its sampling variability behaviour different. When the mean replicate value for a particular site and season is near 100, then the individual replicate values must also be close to 100 and the replicate SD might be expected to be less than for site-season combinations with more intermediate values of PSI. Similarly sites with replicate sample PSI values consistently close to zero, the mean will be close to zero and the replicate sampling SD might be expected to be lower than more intermediate sites.

Amongst the Tay and BAMS sites, there were five site-season occasions where all three or four replicates had PSI values of zero, but there were a further six cases with a mixture of zero and non-zero PSI values amongst replicates. Therefore a sample PSI value of zero does not necessarily indicate that the true site-season meam value is zero and the sampling SD is therefore also greater than zero when some sample PSI values are zero.

To check the overall relationship, we fitted linear, quadratic regression and a non-parametric locally-weighted ‘lowess’ regression relationship (Figure 18). The fitted quadratic relationship (coefficient SE in brackets):

$$\begin{array}{l}\mathrm{SD(PSI)} = \underset{(0.59)}{2.55} + \underset{(0.0246)}{0.0962}\ \mathrm{MeanPSI} - \underset{(0.000239)}{0.000965}\ (\mathrm{MeanPSI})^2 \qquad \text{(Eqn 3.2)}\end{array}$$

was statistically significant ($p < 0.001$) because of the large number of site-season combinations available, but only explained 5% of the total variation in replicate sample SD in PSI values. The quadratic regression predicted replicate SD of PSI values ranged from 3.0 when the mean PSI value is around 5 or 95 up to 5.0 when the mean value is around 50 (Figure 18).

Figure 17 Relationship between replicate sampling SD family-level PSI values and mean of the replicate single season sample PSI values for all available combinations of sites and seasons with replicate sampling for the 16 BAMS sites (▶,♦,●) and the 28 Tay sites (▶,♦●) with symbol based on the replicate mean NTAXA (≤ 10 (▶), 10-15 (♦), 15-32 (●)).

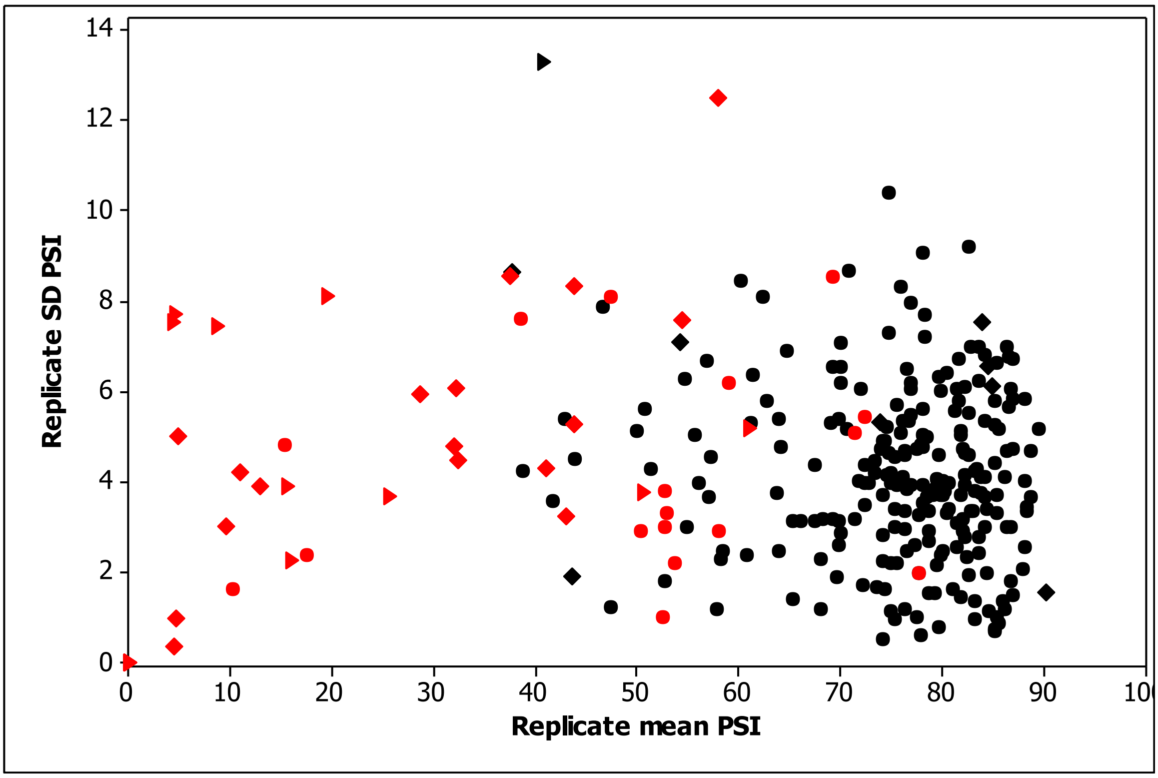
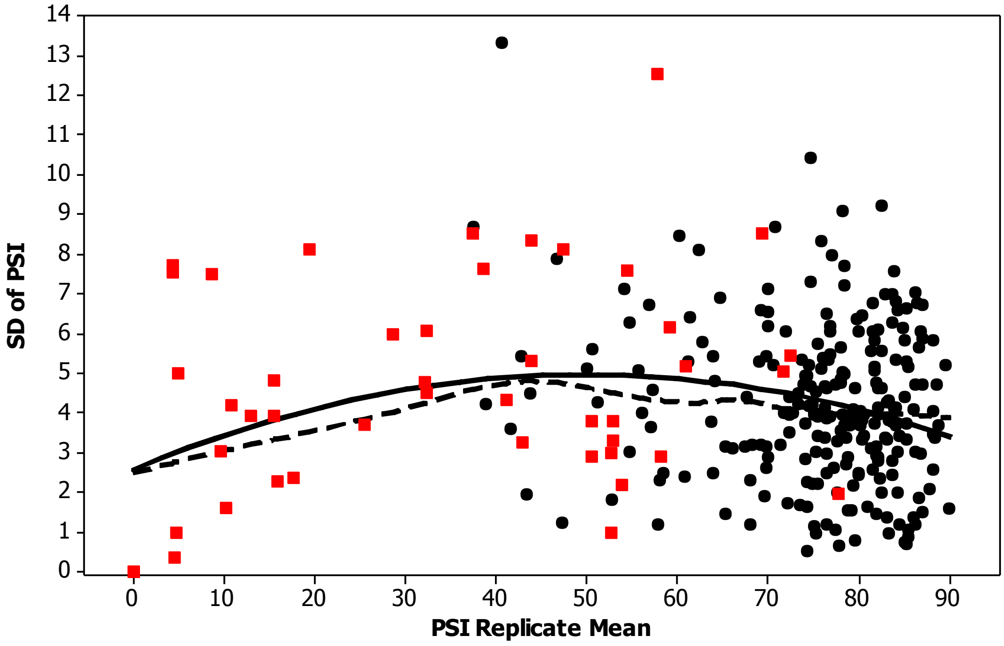


Figure 18 Plot of the relationship between replicate sampling SD of PSI and replicate mean PSI sample value for all available combinations of sites and seasons with replicate sampling for the 16 BAMS sites (■) and the 28 Tay sites (●). Lines denote quadratic (solid) and locally-weighted lowess (dashed) fitted regression relationships



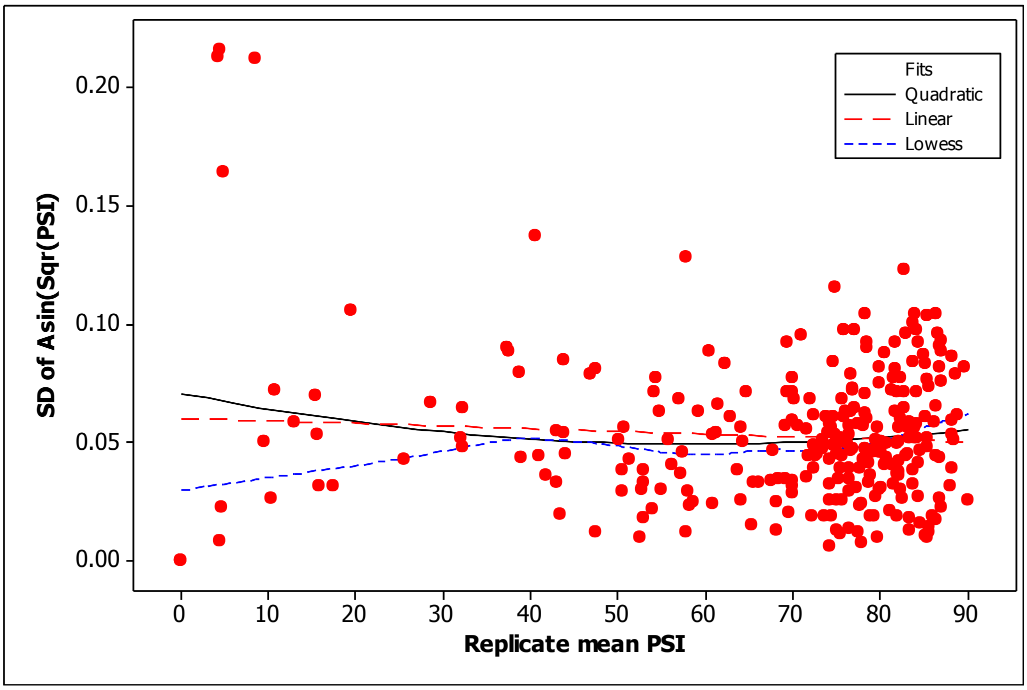
In general statistics, for indices which are proportions or percentages, it has been shown that using the arcsine (Asin) transformation of the square root of the index values (or index values divided by 100 for percentages), can make sampling variability in values independent of sampling mean value. We applied this transformation to the individual sample PSI values as follows:

Arcsine(PSI) = Asin(Sqrt(PSI/100) (Eqn 3.3)

where the transformed values are angles in radians.

We found that the replicate SD of the Arcsine transformed sample PSI values becomes independent of the replicate mean PSI value; neither a linear or quadratic regression relationship is statistically significant (all $p > 0.20$) (Figure 19). [Amongst the 286 site-season combinations with more than one replicate sample, the overall median value of replicate sampling SD of Arcsine transformed PSI values was 0.0494, with inter-quartile (i.e. mid- 50%) range of 0.0322 to 0.0682.]

Figure 19 Relationship between replicate sampling SD of the Arcsine square root transformed values of PSI and the replicate mean PSI values for the combined Tay and BAMs datsets



However, as was found with the LIFE index, the replicate sampling SD of PSI, even on the Arcsine square root transformed scale, appears to decrease with number of taxa present in the sample. The following linear regression relationship between replicate SD of arcsine transformed PSI values and the replicate mean number of BMWP NTAXA was statistically significant relationship ($r = -0.32$, $p < 0.001$) and explained about 10% (r^2) of the variation :

$$\log_e \text{SD Arcsine PSI} = \underset{(0.162)}{-2.195} - \underset{(0.0081)}{0.0460}\, NTAXA \qquad \text{(Eqn 3.4a)}$$

When back-transformed (by taking exponential of both sides of the equation), the predicted relationship is:

$$\text{sampling SD of Arcsine PSI} = 0.111(0.955)^{NTAXA} \qquad \text{(Eqn 3.4b)}$$

which is superimposed as the solid line in Figure 20. This approach provides the estimate of the replicate sampling SD of the Arcsine transformed PSI values for any observed sample value of BMWP NTAXA, as summarised in Table 6.

To overcome any concerns that this relationship may be wholly or partly due to differences between the Tay and BAMS datasets in both general level of NTAXA and of replicate SD, we re-fitted the relationship allowing for dataset differences (*DSET*: BAMS=1, Tay=0) but neither the regression slope or intercept showed any significant difference between datasets (test p = 0.854 and 0.075 respectively), supporting the use of equation (Eqn 3.4).

In RICT, to make use of any estimates of the replicate sampling SD of Arcsine transformed values of PSI back on the original PSI scale (0-100), we will need to Arcsine transform the observed sample PSI values, then create many (10000) simulated other possible sample values by repeatedly adding on a random (normal) term with the appropriate estimate of sampling SD to the transformed observed sample PSI value and then back-transform these simulated transformed values ($PSI_{AsinSim}$) using the trigonometric ‘Sine’ function, as follows:

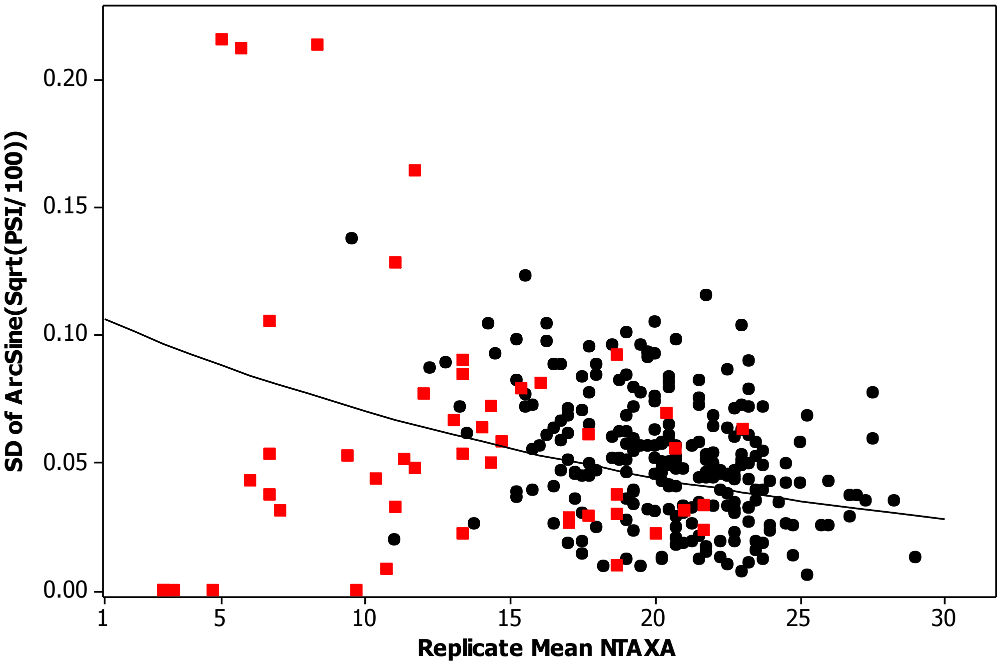
$$PSI_{sim} = 100\ (Sine(PSI_{ArcSim}))^2$$

Further details are given in Section 6 which specifies the new algorithms for the future version of the RICT software.

By calculating the SD of the (10000) simulated sample values for each possible PSI sample mean value over the full range 0-100 and for each value of NTAXA, we can obtained a prediction of the typical replicate sampling SD for any sample value of PSI based on a sample with any number of BMWP taxa present. Using this simulation approach with the NTAXA-dependent estimates of sampling SD of Arcsine PSI values in Equation (3.4) and Table 6 gives the pattern of estimates shown in Figure 21 and Table 7.

Figure 21 and Table 7 also show, for comparison, the estimates of replicate sampling SD in relation to PSI sample mean value based on using the best estimate of a single (assumed constant) value of SD of PSI on the Arcsine scale regardless of the number of taxa present (namely SD = 0.0596 from the mixed modelling of all sampling variance components in Section 3.4 and Table 10). Also shown is the fitted quadratic relationship between SD and PSI sample mean value given previously by equation (3.2).

Figure 20 Relationship between replicate sampling SD of the Arcsine square root transformed values of PSI and the replicate mean BMWP NTAXA for the combined BAMS (■) and Tay (●) sites datasets; included fitted regression line equation 3.4(b).



The estimates of replicate SD of PSI based on the quadratic vary the least from just over 2.5 when PSI is either nearly zero or 100, up to around 5.0 when PSI is around 50 (Table 7). The back-transformed estimate of assumed constant Arcsine-transformed PSI are more variable ranging from less than two at the extremes up around six at mid-range PSI sample values of 50 The greatest range of predicted SD for PSI are obtained using the fitted regression relationship (Eqn 3.4) where SD of PSI for site with average sample PSI values around 50 are predicted to range from around 8-10 when based on five or fewer BMWP taxa down to around 2.8 for samples containing 30 or so BMWP taxa. Using equation (3.4), predicted SD of PSI is less for very high or very low sample PSI values, but still decreases with the number of BMWP taxa present.

Table 6 Estimates of replicate sampling standard deviation (SD) of Arcsine transformed PSI values for sites based on the average number (*NTAXA*) of BMWP families present in a sample from that site (estimates based on equation (Eqn 3.4))

Number of BMWP families present (*NTAXA*)	PSI Sampling SD
1	0.106
2	0.102
3	0.097
4	0.093
5	0.088
6	0.084
7	0.081
8	0.077
9	0.074
10	0.070
12	0.064
15	0.056
20	0.044
25	0.035
30	0.028

Figure 21 Predictions of replicate sampling SD of PSI in relation to replicate mean PSI value based on estimates of SD on Arcsine transformed scale as either contant or in relation to BMWP NTAXA (1,5,10,20,30; eqn(3.4)), or as quadratic between SD and mean on untransformed scale (eqn(3.2)).

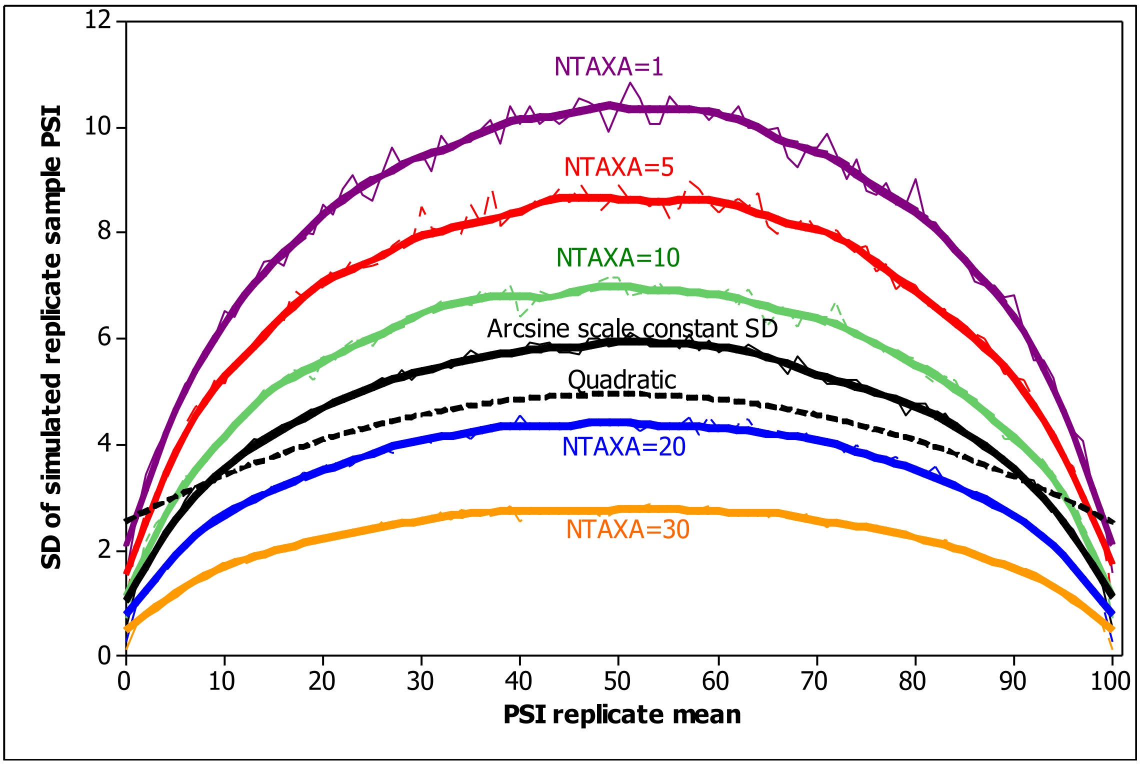


Table 7 Estimates of replicate sampling SD of PSI in relation to sample PSI values (0-100): (a) based of best estimate of constant SD (0.0596) on Arcsine scale (b) based on relationship with NTAXA on Arcsine scale (Eqn 3.4) (c) based on quadaratic relationship with sample mean untransformed PSI (Eqn 3.2)

Sample PSI	(a) Constant SD On Arcsine scale	(b) SD on Arcsine scale depends on NTAXA (Eqn(3.4))					(c) SD v mean Quadratic
		1	5	10	20	30	Eqn(3.2)
0	0.52	1.54	1.07	0.71	0.27	0.11	2.55
2	1.77	3.45	2.67	2.13	1.26	0.83	2.74
4	2.32	4.24	3.32	2.63	1.63	1.04	2.92
6	2.96	5.13	4.32	3.39	2.15	1.31	3.10
8	3.24	5.73	4.81	3.75	2.47	1.56	3.26
10	3.53	6.52	5.14	4.17	2.68	1.74	3.42
12	3.90	6.78	5.65	4.62	2.86	1.83	3.57
14	4.07	7.44	6.06	5.02	3.02	1.93	3.71
16	4.25	7.40	6.34	5.25	3.21	2.04	3.85
18	4.51	8.00	6.79	5.35	3.45	2.11	3.97
20	4.69	8.55	6.97	5.67	3.53	2.26	4.09
25	5.13	8.60	7.36	5.86	3.87	2.36	4.36
30	5.41	9.51	8.49	6.48	4.14	2.58	4.57
35	5.76	9.71	8.53	6.62	4.10	2.73	4.74
40	5.72	10.07	8.40	6.43	4.55	2.55	4.86
45	5.74	10.23	8.56	6.82	4.32	2.75	4.93
50	5.89	10.35	8.91	7.15	4.48	2.77	4.95
55	6.00	10.59	8.29	6.83	4.32	2.80	4.93
60	5.75	10.33	8.38	6.78	4.28	2.64	4.85
65	5.65	9.96	8.12	6.22	4.15	2.61	4.73
70	5.33	9.61	8.02	6.40	4.12	2.56	4.56
75	5.03	8.78	7.76	6.04	3.78	2.43	4.34
80	4.77	9.02	7.00	5.60	3.51	2.22	4.08
82	4.48	8.21	6.57	5.19	3.57	2.16	3.96
84	4.34	7.69	6.22	5.22	3.21	2.02	3.83
86	4.12	7.09	6.12	4.71	3.03	1.88	3.69
88	3.86	7.00	5.67	4.55	2.81	1.79	3.55
90	3.55	6.83	5.27	4.09	2.67	1.63	3.40
92	3.31	5.60	4.81	3.79	2.44	1.56	3.24
94	2.83	5.25	4.24	3.43	2.10	1.28	3.08
96	2.38	4.24	3.57	2.89	1.76	1.12	2.90
98	1.73	3.17	2.62	1.99	1.27	0.80	2.72
100	0.52	1.57	1.04	0.70	0.27	0.12	2.53

Summary:
We recommend assuming sampling SD of PSI is constant across the Arcsine transformed scale, but varies with the number of BMWP taxa present in the mathematical form provided by equation (eqn 3.4), as presented in Table 7 and Figure 21

3.4 Estimates of replicate sampling SD for abundance-weighted indices

The estimates of replicate sampling SD for each index (transformed as appropriate) were obtained using one-way analysis of variance (ANOVA) on the BAMS and Tay datasets where each combination of site, year and season was treated as a separate factor level. The residual mean square then equates to overall estimate of the average replicate sampling variance across the whole dataset; the estimate of ‘average’ replicate SD is then the square root of this residual mean square. Separate estimates of replicate SD for single season samples were derived for the 16 BAMS sites dataset, the 28 TAY sites dataset and for both datsets combined (Table 8).

If the ANOVA are repeated on the combined datasets, the estimates of Replicate SD are very similar to those for the Tay dataset. This is because the majority (18) of the 28 Tay sites have four replicate samples in each of spring and autumn for each of five years (1990-1994), which means they provide the vast majority of the degrees of freedom and weight in the combined dataset estimates of replicate SD. Because the overall estimates of replicate SD are intended to be used in the new RICT software to assess uncertainty across all river sites throughout the UK, we think it is probably best not to weight the estimates unduly towards just the sites in the relatively small Tay region. Therefore, the overall estimates of replicate SD for each index were obtained as weighted averages of the estimates for the two datasets given in Table 8 columns (a) and (b), where the two estimates were weighted by the number of sites sampled, namely 16 for BAMS and 28 for Tay; the weighted average estimates are given in Table 8 column (d).

Table 8. Estimates of the replicate sampling standard deviation (SDRep) of indices (transformed where appropriate) for single season samples based on (a) 16 sites BAMS dataset, (b) 28 SEPA Tay sites dataset, (c) both datasets combined and (d) weighted average (i.e. weighted by number of sites in dataset); highest of individual dataset estimates highlighted in bold.

Index		Transform scale	(a) BAMS	(b) Tay	(c) BAMS + Tay	(d) Weighted average
Original BMWP	BMWP Score	√	0.588	**0.668**	0.659	0.639
	NTAXA	√	0.228	**0.244**	0.242	0.238
	ASPT	none	0.249	**0.250**	0.250	0.250
Non-weighted WHPT	BMWP Score	√	0.576	**0.685**	0.673	0.645
	NTAXA	√	0.230	**0.250**	0.247	0.243
	ASPT		0.207	**0.251**	0.246	0.235
Abundance-weighted WHPT	BMWP Score	√	0.600	**0.680**	0.671	0.651
	NTAXA	√	0.230	**0.250**	0.247	0.243
	ASPT	none	**0.305**	0.262	0.268	0.278
LIFE (family level)		none	**0.326**	0.202	0.220	0.247
PSI (family level)	Arcsine(Sqr(PSI/100))		**0.0789**	0.0569	0.0599	0.0649

3.5 Correlation among indices: overall and amongst replicate samples

The WHPT score index is highly correlated with the original BMWP score index, whether in revised non-weighted form (correlation *r* = 0.984) or abundance-weighted form (*r* = 0.981) (Table 9 (a)).

Although the WHPT values for NTAXA were up to six taxa higher than the original BMWP NTAXA index, the two forms of NTAXA are highly correlated (*r* = 0.985) (Table 9). Note, that the values of NTAXA for the non-weighted and abundance-weighted form of the revised BMWP indices are, by definition, always the same.

Table 9 Pearson correlations between the biotic indices (transformed where appropriate) based on single season samples for (a) raw values for all three datasets combined and (b) residual variation in index values among replicate samples after allowing for all site and season combination differences for the BAMS dataset. Correlation >0.9 highlighted in bold.

		Original			WHPT Non-Weighted		WHPT Abundance-weighted		
(a) raw values		Score	NTAXA	ASPT	Score	ASPT	Score	NTAXA	ASPT
Original	NTAXA	**0.939**							
	ASPT	0.775	0.522						
WHPT non-weighted	Score	**0.984**	**0.917**	0.779					
	ASPT	0.703	0.440	**0.970**	0.737				
WHPT Abundance-weighted	Score	**0.981**	**0.910**	0.783	**0.997**	0.743			
	NTAXA	**0.926**	**0.985**	0.521	**0.930**	0.448	**0.923**		
	ASPT	**0.986**	0.776	0.477	**0.959**	0.485	**0.986**	0.776	0.485
(a) replicate residuals		Score	NTAXA	ASPT	Score	ASPT	Score	TAXA	ASPT
Original	NTAXA	**0.947**							
	ASPT	0.738	0.502						
WHPT non-weighted	Score	**0.974**	**0.958**	0.639					
	NTAXA	**0.939**	**0.999**	0.494	**0.958**				
	ASPT	0.673	0.504	0.800	0.708				
WHPT Abundance-weighted	Score	**0.931**	**0.920**	0.625	**0.951**	0.670			
	NTAXA	**0.939**	**0.999**	0.494	**0.958**	0.503	**0.929**		
	ASPT	0.504	0.415	0.602	0.514	0.672	0.708	0.414	

Indices which respond in the same way to physical, environmental and/or anthropogenic variation will tend have high correlations across a wide range of sites, for example, as found in this study for BMWP score and NTAXA. The independence, or lack of it, amongst indices, as indicators of general and specific stresses, is obviously of great importance in general bio-assessment.

However, within this project on assessing uncertainty, we restricted our interest to assessing the extent to which the sampling variability amongst these indices is correlated. Specifically, Table 9(b) gives the correlations between the residual values of each pair of indices after removing all differences in values due to site, year and season combination differences. To avoid involving the many sites/year/season combinations with no replication which would all have had zero residuals and inappropriately increased the apparent correlation amongst residuals, the correlations were based on the replicate residuals for just the 16 BAMS sites.

The correlations among replicate residuals was very high between original BMWP score and TAXA (r = 0.947), but relatively low between TAXA and ASPT (*r* = 0.502). These residual correlations, first obtained by Furse *et al* (1995 - their Table 5.2), were used to justify simulating independent random error terms for sampling variation in NTAXA and ASPT in the development of the uncertainty simulation algorithms used in RIVPACS III+, RPBATCH and the current RIVPACS IV models in RICT.

The residual (and raw) values of BMWP score and TAXA in their revised forms are all highly correlated; they are also highly correlated with the residual values for the original BMWP score and NTAXA (all *r* > 0.92) (Table 9). These high correlations between indices suggest that the revised forms of BMWP score and NTAXA are likely to give broadly similar EQI values to those obtained using the original BMWP scoring system for the vast majority of sites; however the changes in values may lead to some small improvements in the river status assessment system while also satisfying the WFD requirement to involve taxonomic abundances.

3.6 Mixed modelling estimates of temporal and other variance parameters

3.6.1 Overall estimates based on SEPA, Tay and BAMS datasets combined

The estimates of each variance component parameter were obtained from the REML fits to the appropriate mixed model for each index (BioIndex), as specified and fitted by the following 'lmer' function within the 'R' software:

Model1 <- lmer(BioIndex~1+(1|SiteSeason/Period/Year/DateDiff))) (Eqn 3.5)

Where

'SiteSeason' is a variable representing each separate combination of site and season
'Period' identifies each separate 3-year period of data
'Year' identifies each separate year
'DateDiff' identifies each separate sampled date with any one season of the same year

The variance parameter estimates from the fitted models are given in Table 10(a); they are equal to the square of the equivalent SD parameter estimates given in Table 10(c). The same mixed model was fitted to each index, where necessary on the best transformed scale. However, these initial variance component mixed models ignored the previously detected dependence of the variance of LIFE and Arcsine of PSI on the number of BMWP tax present in a sample. This is investigated further in Section 3.6.2.

To assess the relative size of the three variance components which determine the total variance of index values in a typical three-year period, the components for replicate variance, within-seasonal temporal and inter-year-within-period variance are expressed as a percentage of their sum in Table 10(b). Replicate sampling variance generally contributes just under half of the total variance within a 3-year period, ranging from 38% for (square root of) abundance-weighted WHPT score, to 55% for the family-level LIFE index.

It is useful to calculate the following parameter:

%TempSeas = estimate of percentage of total within-period temporal variance which is due to within-season temporal variability
= $100\ Var_{TSeas}\ /\ (Var_{TSeas} + Var_{TYear})$

Estimates of this parameter, given in Table 10(b), highlight that the variance estimates for short term within-season temporal variability are, rather surprisingly, about the same or higher than the longer-term inter-year-within-period temporal variance estimates for all indices except LIFE. This raises the concern held by us prior to any data analysis that any additional samples taken on a later date within the same season may be more likely to have been taken from a site if it was suspected, or known, that there was either some recent problem at the site, or the previous sample in that season was suspect. Thus the available data to estimate within-season temporal variance may not be completely typical, but moreover may tend to over-estimate the typical/average within-season temporal variance, which in turn would lead to some under-estimation of the true inter-year variance components. However, with that caveat, these estimates are the best available.

For the new PSI index, based on variance estimates from these three datsets, all of the temporal within-period variance appeared to be due to shorter term within-season variation.

Table 10(c) gives the estimates of the SD parameters (obtained as the square roots of the equivalent variance terms). Estimates of SD_{Rep}, SD_{TSeas} and SD_{TYear} can be used to estimate the overall uncertainty SD (SD_{Obs}) associated with the estimate of either the single year or three-year average observed index value used in estimate the uncertainty associated with the average of the single season EQR values and resulting ecological status class of the site for that year 3-year period (see Section 6 for further algorithm details).

Table 10 Estimates of index values for (a) variance and (c) SD (√Variance) parameters for within-season temporal variability (SD_{TSeas}), inter-year variability (SD_{TYear}), replicate sampling (SD_{Rep}) and other variance components based on all data from the BAMS, Tay and SEPA datasets combined; (b) gives variance components as a percentage of the average total variance (VarRep + VarTSeas + Var_{TYear}) within three-year periods; %TempSeas = 100Var_{TSeas} / (Var_{TSeas}+Var_{TYear}).

(a) Variance	Index	Var Rep	Var TSeas	Var TYear	Var TPeriod	Var Site.Seas
Original BMWP	√ Score	0.4320	0.2957	0.2746	0.2615	2.5196
	√ NTAXA	0.0578	0.0350	0.0365	0.0291	0.2154
	ASPT	0.0654	0.0596	0.0209	0.0359	0.7859
Revised BMWP Non-weighted	√ Score	0.4519	0.3626	0.3124	0.3517	2.8809
	√ NTAXA	0.0607	0.0444	0.0393	0.0391	0.2294
	ASPT	0.0616	0.0657	0.0173	0.0396	0.9748
Revised BMWP Abundance-weighted	√ Score	0.4496	0.4217	0.3141	0.3917	3.2214
	√ NTAXA	0.0607	0.0444	0.0393	0.0391	0.2294
	ASPT	0.0722	0.0776	0.0308	0.0589	1.2042
LIFE *		0.0446	0.0139	0.0221	0.0132	0.2462
PSI *	ArcsineSqr	0.00355	0.00421	0.0000	0.00114	0.03059
(b) % Variance		%Var Rep	%Var TSeas	%Var TYear	%Temp Seas	
Original BMWP	√ Score	43	30	27	52	
	√ NTAXA	45	27	28	49	
	ASPT	45	41	14	74	
Revised BMWP Non-weighted	√ Score	40	32	28	54	
	√ NTAXA	42	31	27	53	
	ASPT	43	45	12	79	
Revised BMWP Abundance-weighted	√ Score	38	36	26	57	
	√ NTAXA	42	31	27	53	
	ASPT	40	43	17	72	
LIFE *		55	17	27	39	
PSI *	ArcsineSqr	46	54	0	100	
(c) SD		SD Rep	SD TSeas	SD TYear	SD TPeriod	SD Site.Seas
Original BMWP	√ Score	0.657	0.544	0.524	0.511	1.587
	√ NTAXA	0.240	0.187	0.191	0.171	0.464
	ASPT	0.256	0.244	0.144	0.189	0.886
WHPT Non-weighted	√ Score	0.672	0.602	0.559	0.593	1.697
	√ NTAXA	0.247	0.211	0.198	0.198	0.479
	ASPT	0.248	0.256	0.132	0.199	0.987
WHPT Abundance-weighted	√ Score	0.670	0.649	0.560	0.626	1.795
	√ NTAXA	0.246	0.211	0.198	0.198	0.479
	ASPT	0.269	0.279	0.176	0.243	1.097
LIFE *		0.211	0.118	0.149	0.115	0.496
PSI *	ArcsineSqr	0.0596	0.0649	0.0000	0.0338	0.1749

* Variance and SD above for LIFE and PSI ignore any relationship with NTAXA

3.6.2 Comparison with estimates from Northern Ireland sites dataset

In the previous sub-section 3.6.1, we highlighted our concern about potential over-estimation of the within-season temporal SD (SD_{TSeas}) derived from the SEPA dataset (when combined with the Tay and BAMS replicated sample datasets). Our concern was that the occasions when samples were taken on more than one day in the same season at a site may often have been because of some perceived recent problem at the site and hence tend to over-estimate typical within-season variability.

A separate Northern Ireland (NI) River Community Change study dataset (dataset 4) contained monthly samples over a period of one year (Feb-Jan) at each of 12 sites in Northern Ireland. From this NI dataset, we extracted a sample in each of the three months in each of the three RIVPACS seasons, spring (Mar-May), summer (June-Aug), and autumn (Sep-Nov) at each of the 12 sites.

We then analysed this dataset using variance components analysis removing all site x season effects to provide an independent estimate of the average overall within-season variance (Var_{WSeas}) or its SD equivalent (SD_{WSeas}).

The overall variance within a season (Var_{WSeas}) is the sum of the replicate sampling variance (Var_{Rep}) and the within-season temporal variance (Var_{TSeas}), and therefore:

$SD_{WSeas} = \surd(SD_{Rep}^2 + SD_{TSeas}^2)$.

However, for the NI dataset, only a single sample was taken at each site in each month, so there is no information to separate replicate variance from within-season temporal variance; we can only estimate their combined effect (SD_{WSeas}).

The estimates of SD_{WSeas} for each index based on the NI dataset are compared with those given in Table 10(c) based on the combined Tay+SEPA+BAMS datasets. The estimates of SD_{WSeas} were actually higher for the NI dataset for each of the indices (Table 11). However, the NI within-season sample cases are all spread evenly (one per month) and thus maximally across the three months in each RIVPACS season, whereas those taken a sites on different days in the same season in the other datasets are taken at varying times, including only a few days apart. This may at least partly explain the differences.

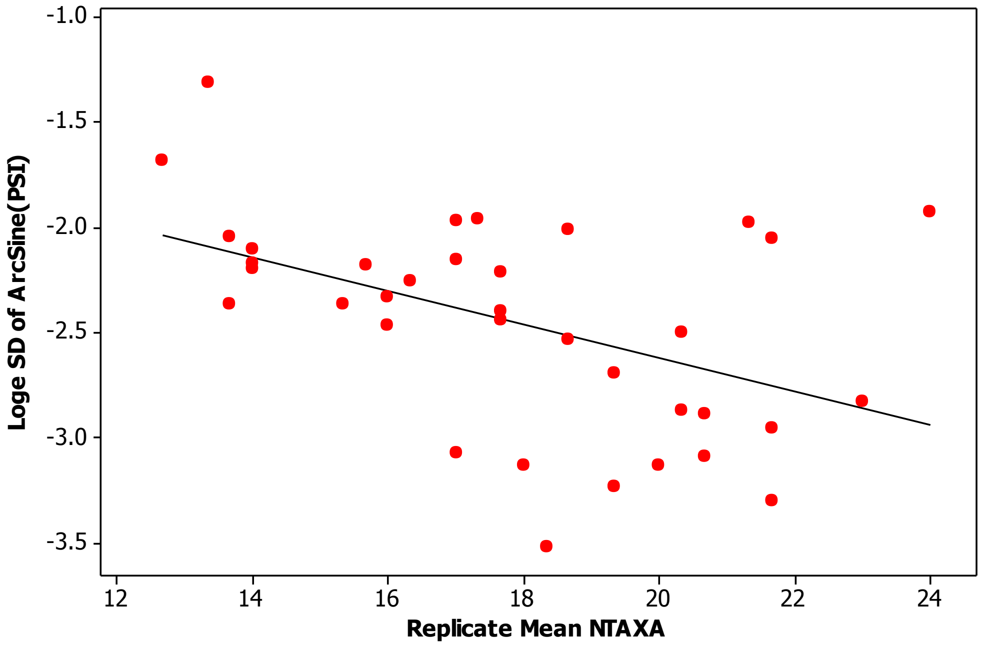
Table 11 Estimates of single season sample values for overall SD within a season (SD_{WSeas}), based on the combined effect of replicate and within-season temporal variability for (a) SEPA, Tay and BAMS datasets combined, and (b) NI monthly-sampled dataset

		(a) SEPA+Tay+BAMS			(b) NI
Index	Index form	SD Rep	SD TSeas	SD WSeas	SD WSeas
Original BMWP	√ Score	0.657	0.544	0.853	1.127
	√ NTAXA	0.240	0.187	0.304	0.392
	ASPT	0.256	0.244	0.354	0.467
WHPT Non-weighted	√ Score	0.672	0.602	0.906	1.125
	√ NTAXA	0.247	0.211	0.325	0.392
	ASPT	0.248	0.256	0.357	0.480
WHPT Abundance-weighted	√ Score	0.670	0.649	0.936	1.172
	√ NTAXA	0.246	0.211	0.325	0.392
	ASPT	0.269	0.279	0.388	0.510
LIFE *		0.211	0.118	0.242	0.267
PSI *	ArcsineSqr	0.0596	0.0649	0.0881	0.1079

* SD above for LIFE and PSI ignore any relationship with NTAXA

The regression relationship between $\log_e SD_{WSeas}$ for a NI site and season and the mean number of BMWP taxa (NTAXA) for that NI site and season was statistically significant ($r^2 = 20\%$, $p = 0.004$) (Figure 22), adding further support to the use of this form of relationship to estimate sampling uncertainty of PSI values for UK river sites.

Figure 22 Relationship between within-season sampling SD (SD_{WSeas}) of Arcsine transformed PSI values and mean of the replicate sample BMWP NTAXA values for all available combinations of sites and seasons for the Northern Ireland dataset



<u>Summary</u>
We conclude from our corroborative analyses of Northern Ireland sites, that our previous estimates of within-season temporal SD based on the SEPA, Tay and BAMS combined datasets are reasonable and appropriate for use in assessing uncertainty and that SD of PSI and LIFE both decrease with the number of taxa present (i.e. BMWP NTAXA).

3.6.3 Variance of LIFE and PSI in relation to sample NTAXA

In our analyses of sampling variation amongst replicate samples we found that the replicate sampling variance of LIFE and of the Arcsine transformed values of PSI both declined with the number of BMWP taxa (NTAXA) present in samples (section 3.3). It is difficult to incorporate this feature into the mixed modelling structure, including using the 'lmer' function in the R software. We tried to fit the following model (where RecipNTAXA = 1/NTAXA):

Model1 <- lmer(BioIndex~1+(RecipNTAXA|SiteSeason/Period/Year/DateDiff)))

which would allow each variance component to vary as a multiple of 1/NTAXA; however the model fitting did not converge, probably because of lack of sufficient information within the datasets on the relation of every variance component with NTAXA.

As a way forward, we fitted the mixed model of equation (Eqn 3.5) to subsets of the individual samples with BMWP NTAXA values with particular ranges, namely 1-10, 11-20, >20 BMWP taxa.

For the LIFE index, the mixed model estimate of each of the variance components decreases with the number of taxa present (Table 12). Although obviously based on the same datasets, this adds to our confidence that sampling variance of LIFE does decrease with NTAXA.

Table 12 Estimates of single season sample values for SD (√Variance) parameters for within-season temporal variability (SD_{TSeas}), inter-year variability (SD_{TYear}), replicate sampling (SD_{Rep}) and other variance components based on mixed model (Eqn 3.5) fitted to all data from the BAMS, Tay and SEPA datasets combined and also for subsets restricted to samples within a range of BMWP NTAXA (1-10, 11-20 or >20 taxa)

Index	NTAXA range	SD_{Rep}	SD_{TSeas}	SD_{TYear}	$SD_{TPeriod}$	$SD_{Site.Seas}$
LIFE	≤10	0.442	0.164	0.204	0.145	0.729
	11-20	0.198	0.129	0.137	0.117	0.441
	>20	0.147	0.088	0.107	0.063	0.286
	All	0.211	0.118	0.149	0.115	0.496
ArcsineSqr PSI	≤10	0.1068	0.1756	0.0000	0.0000	0.2721
	11-20	0.0588	0.0443	0.0288	0.0347	0.1433
	>20	0.0439	0.0112	0.0354	0.0250	0.0942
	All	0.0596	0.0649	0.0000	0.0338	0.1749

For the Arcsine transformed PSI values, the estimates of replicate variance (Var_{Rep}) and within-season temporal variance (Var_{TSeas}) decrease with the number of BMWP taxa present (Table 12). The longer term between-year-within-period and between-period variances terms do not seem to vary in a consistent manner with the number of taxa present in individual samples from a site. However, the between-year component SD (SD_{TYear}) estimate is zero when based on all non-NI data, but non-zero for two of the three subsets of the data and the mid value when based on samples within intermediate NTAXA range of 11-20 is 0.0288 (Table 12).

We recommend that the mid-value estimate of SD_{TYear} of 0.0228 is used for all sites and samples

Overall, our conclusion is that sampling variation, both replicate and within-period temporal variation in both the LIFE and PSI indices is greater when fewer taxa are present in the samples. The remaining problem is how to express this in a practical quantitative way to use within RICT.

In order to make combined use of the derived relationships of SD with NTAXA for LIFE and PSI given by equations (Eqn 3.1) and (Eqn 3.4) and the estimates of average temporal variance (Table 10), we need to decide the average sample NTAXA to which the estimates of average replicate and temporal variance apply.

Both the mean and median number of BMWP taxa per taxa were similar across the Tay, SEPA and NI datasets, varying between 17.7 to 21, but for the BAMS dataset of sites, specifically chosen to encompass a wide range of site qualities including very poor sites, the mean and median NTAXA was only 13 (Table 13). The variance component mixed models were fitted to the samples from the combined Tay, BAMs and SEPA datasets, for which the mean and median NTAXA was 18.

Table 13 Summary statistics for number of BMWP taxa (NTAXA) per sample in each dataset

Dataset(s)	Samples	Mean	Median	Min	25%	75%	Max
Tay	1058	20.7	21	6	18	23	32
BAMS	144	13.0	13	2	9	17	27
SEPA	7549	17.7	18	1	15	21	39
NI	108	17.9	18	8	16	21	29
Tay+BAMS	1202	19.8	20	2	18	23	32
Tay+BAMS+SEPA	8751	18.0	18	1	15	21	39
All	8859	18.0	18	1	15	21	39

Therefore we recommend that the estimates of SD_{Rep}, SD_{TSeas} and SD_{TYear} for LIFE and Arcsine PSI are assumed to apply to samples with an average NTAXA of 18 and that samples with an average NTAXA above or below 18 are adjusted by factors derived from equations (Eqn 3.1) and (Eqn 3.4).

Specifically, the adjustment factor (K) for each sampling component SD in Table 10(c) for a site with average number of sample BMWP taxa equal to NTAXA is:

$K = 0.951^{(NTAXA-18)}$ for LIFE (Eqn 3.5a)

and $K = 0.955^{(NTAXA-18)}$ for Arcsine PSI (Eqn 3.5b)

3.7 Recommended sampling SD estimates for WHPT, LIFE and PSI for use in RICT

The recommended set of sampling standard devation (SD) parameter estimates for each form of WHPT index, weighted and unweighted, for use in the RICT software are given in Table 14.

Table 14 Recommended estimates of SD parameters of each form of WHPT index due to replicate sampling (SD_{Rep}), within-season temporal variability (SD_{TSeas}) and inter-year variability (SD_{TYear}) based on single season samples

Index	Index form	SD_{Rep}	SD_{TSeas}	SD_{TYear}
Original BMWP	√ Score	0.657	0.544	0.524
	√ NTAXA	0.240	0.187	0.191
	ASPT	0.256	0.244	0.144
WHPT Non-weighted	√ Score	0.672	0.607	0.554
	√ NTAXA	0.247	0.211	0.198
	ASPT	0.248	0.257	0.131
WHPT Abundance-weighted	√ Score	0.670	0.654	0.556
	√ NTAXA	0.247	0.211	0.198
	ASPT	0.269	0.279	0.174

The recommended set of sampling standard devation (SD) parameters estimates for the LIFE and PSI indices for use in the RICT software are given in Table 15. Specifically, the multiplicative adjustment factor (K) for each sampling component SD (SD_{Rep}, SD_{TSeas} and SD_{TYear}) in Table 15 for a site with average number of sample BMWP taxa equal to NTAXA is:

$K_{LIFE} = 0.951^{(NTAXA-18)}$ for LIFE (Eqn 3.5a)

and $K_{PSI} = 0.955^{(NTAXA-18)}$ for Arcsine PSI (Eqn 3.5b)

Table 15 Recommended estimates of sampling component SD for LIFE and Arcsine PSI (Arcsine(Sqr(PSI/100))), together with the multiplicative adjustment factors (K_{LIFE}, K_{PSI}) for each component SD based on the average observed number of BMWP taxa (NTAXA) per sample from the site to be assessed (Eqn(3.5)).

Index	LIFE	Arcsine PSI
SD_{Rep}	0.211	0.0596
SD_{TSeas}	0.118	0.0649
SD_{TYear}	0.149	0.0288
NTAXA	K_{LIFE}	K_{PSI} for Arcsine PSI
1	2.349	2.187
2	2.234	2.089
3	2.125	1.995
4	2.021	1.905
6	1.827	1.738
8	1.653	1.585
10	1.495	1.445
15	1.163	1.148
18	1.000	1.000
20	0.904	0.912
25	0.703	0.724
30	0.547	0.575
35	0.426	0.457

As an example, for a site with an average sample NTAXA of 25 over the period to be assessed, then for in assessing the uncertainty in the average LIFE (or LIFE EQI) values for the site, we use:

$SD_{Rep} = 0.703 \times 0.211 = 0.148$
$SD_{TSeas} = 0.703 \times 0.118 = 0.083$
$SD_{Tyear} = 0.703 \times 0.149 = 0.105$

The algorithms using these parameters to simulate the sampling uncertainty in observed values, EQI and thus EQRs and confidence of status class are specified in Section 6.

4. Datasets used to estimate biases in abundance-weighted indices

Our approach to estimating bias values for WHPT, LIFE and PSI in RICT was to explore whether relationships might exist between BMWP NTAXA bias and biases in these new indices. These relationships could then be used as a basis to derive algorithms that could calculate biases for these new metrics using, crucially, a simple user-supplied NTAXA (of BMWP) bias. It is important to note that biases will vary with the extent of sample processing errors, which may vary over time.

From our unique experience in CEH via the BAMS project (Furse *et al* 1995) and the EU STAR project Haase *et al* 2006), we know that it is very difficult to assess and quantify the effect of sample processing errors on potential biases in metrics. Each metric (e.g. WHPT, LIFE, PSI) should be assessed in its own right as they will almost certainly behave differently in response to the same errors. This can only be done by reference to sample audit data (i.e. CEH/QMUL audit of agency samples) and subsequent re-construction of corresponding audit-corrected “true” samples. From this, a comparison of the paired original and audit-corrected sample index values for a wide range of samples can be used to assess the effect of sample processing errors on biases in observed index values. CEH/QMUL audit data were available but these only recorded those families involved in an error. Changes in score could be calculated, but calculation of absolute scores was impossible. A new paired dataset of pre- and post-audit samples with complete taxa lists was needed so that calculation of pre- and post-audit index scores for BMWP, WHPT, LIFE and PSI would be possible.

A number of compromises in the choice of a new audit dataset had to be addressed. Environment Agency audit data were chosen because these comprised audits of primary samples as opposed to audits of samples that had only been supplied for auditing after internal analytical quality control (AQC). The 2010 dataset was also chosen because at this point in time, 20 samples were being audited from each Environment Agency laboratory. After 2010, the number of samples externally audited began to fall. Going further back than 2010, the taxonomic coverage of the audit was more restricted, focussing more exclusively on BMWP families. By 2010 however the audit was reporting sample processing errors for LIFE and WHPT families, which also include most PSI families. The exclusive use of Environment Agency data had the drawback of restricting the geographical coverage of the dataset to England and Wales (still part of the EA at this time), but the overriding importance of the number of samples that could be obtained and the taxonomic coverage required to perform analyses for WHPT, LIFE and PSI meant that this was the best single dataset to choose.

During late 2013, a total of 427 audit samples, representing all 2010 Environment Agency samples that had been externally audited by QMUL in that year, were entered into a database from first principles (Table 16). Samples were entered as primary taxa lists together with records of sample losses, gains and omissions, and vial losses, gains and omissions. Taxonomic coverage of data entry included all BMWP, WHPT, LIFE and PSI families. Other non-scoring families (non-scoring in any index) were also entered where available.

This new database of 427 audit samples from 2010 was used to construct a matched set of 427 pre- and post-audit family lists including $\log_{10}$ abundance data where available. A total of 21,316 family records existed after creating separate pre- and post-audit samples.

The QMUL audit reported all types of losses, gains and omissions of families in each sample audited but QMUL were not contracted by the Environment Agency to audit the abundances of families per se. Pre- audit Log_{10} abundances from the primary sample analysis by the Environment Agency were assumed to also be correct post-audit. These were therefore also ascribed to the post-audit samples. Where a gain was recorded by the audit, the additional family was given a $\log_{10}$ abundance category of 1 (1-9 individuals) since it seemed likely that any missed family would have only be present at this lowest Log_{10} abundance category. Had it been present in a primary sample at a higher Log_{10} abundance category it is more likely that it would not have been missed by the primary analyst. Losses did not require any such post-audit estimation of Log_{10} abundances since these families disappeared from the dataset.

Pre- and post audit biotic indices were then calculated for all 427 samples.

For all abundance weighted indices, abundance weighting was used for of index calculation, despite in some cases there being an option to calculate the index without abundance weighting. For all indices where an option existed to calculate that index using either BMWP composite taxa (e.g. Planariidae including Dugesiidae) or with separate taxon scores for the individual families, separate (distinct) families were always used for index calculation. These two approaches were considered to most closely match the current and future laboratory practices of the Environment Agency, Scottish Environment Protection Agency, Natural Resources Wales and Northern Ireland Environment Agency, and therefore make the bias relationships derived in the chapter that follows most relevant to the data being used for classification by the UK agencies.

Table 16 Summary of the 427 Environment Agency audit samples from 2010

EA Region	Area Lab	Number of audit samples entered
Anglian	Northern	20
	Central	20
	Eastern	20
North East	North	20
	Yorkshire	20
North West	Northern	20
	Southern	20
Midlands	Western	20
	Central	20
	Eastern	20
Southern	Kent & Sussex	20
	Solent & S. Downs	20
Thames	North East	20
	South East	20
	West	20
Wales	Northern	20
	South Western	20
	South Eastern	20
South West	Cornwall	13
	Devon	14
	Wessex (Bridgwater)	20
	Wessex (Blandford)	20

The following pre- and post audit biotic indices were calculated for all 427 samples:

Pre audit BMWP Score	Post audit BMWP Score
Pre audit BMWP NTAXA	Post audit BMWP NTAXA
Pre audit BMWP ASPT	Post audit BMWP ASPT
Pre audit WHPT Score	Post audit WHPT Score
Pre audit WHPT NTAXA	Post audit WHPT NTAXA
Pre audit WHPT ASPT	Post audit WHPT ASPT
Pre audit LIFE	Post audit LIFE
Pre audit PSI	Post audit PSI

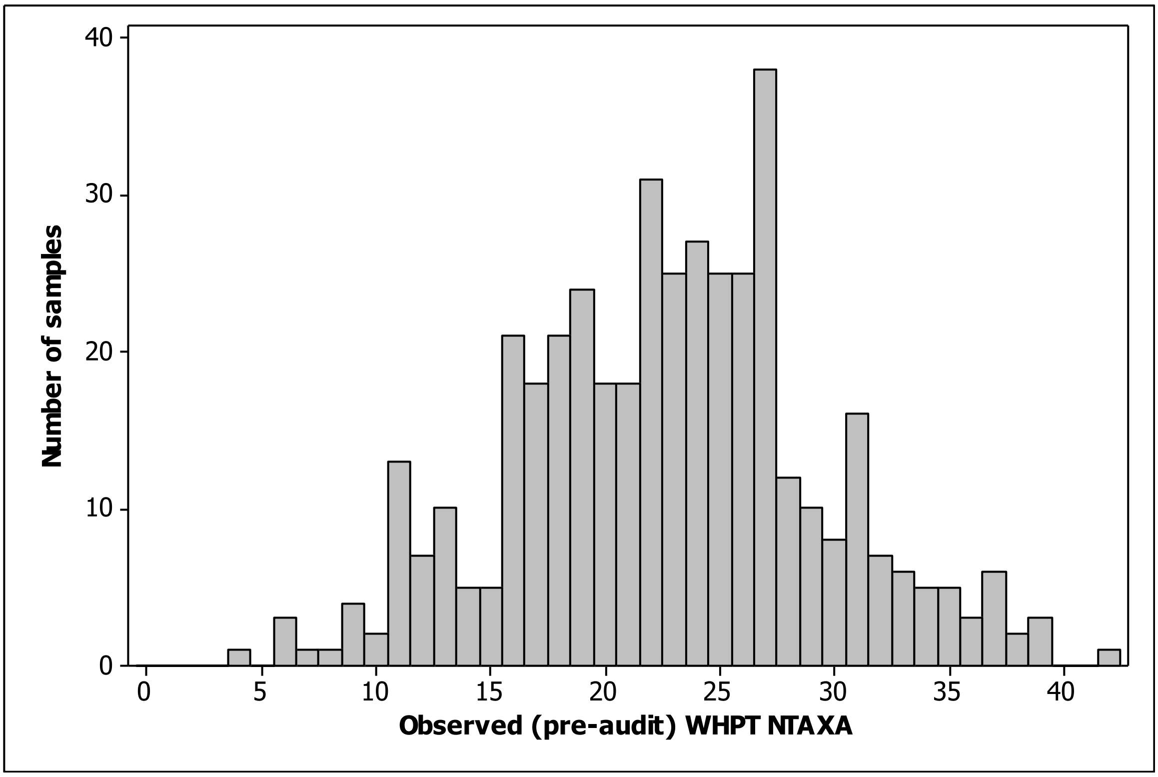
Section 5 of this report describes analyses of relationships between biases in the WHPT NTAXA and WHPT ASPT indices. The algorithms and parameter estimates for their implementation in RICT are reported in section 6 of this report. Corresponding analyses and algorithms for biases in LIFE and PSI will follow under a separate contract and report to the Scottish Executive.

5. Estimation of biases in abundance-weighted indices

5.1 WHPT NTAXA bias

The observed number of WHPT taxa (NTAXA) amongst the 429 audited sample dataset varied from 4 to 42, encompassing a wide range in sample richness, but with an inter-quartile range (i.e. middle 50%) of 18-27 with a mean and median of around 23 (Figure 23).

Figure 23 Frequency histogram of the observed (pre-audited) values of WHPT NTAXA amongst the 429 audited samples dataset



5.1.1 Relationship between WHPT NTAXA bias and BMWP NTAXA bias

The ‘Bias’ for any particular sample and index is defined to be the post-audit taxonomically-corrected value of an index minus the pre-audit original ‘observed’ sample (O) value for the index.

For the BMWP NTAXA index, the maximum bias (i.e. net under-estimation of the number of BMWP taxa present) observed in our dataset of 427 audit-corrected samples was 5 , but the median bias was one taxa and the average bias was 1.14 (Table 17). Three samples had two more taxa recorded than actually existed in the audited sample and 12 samples had one more BMWP taxa recorded than existed in the audited sample. Thus 3.5% of samples actually had higher observed sample values of BMWP NTAXA than they should have. A further 30% of samples had the correct values of BMWP NTAXA. However, on average 66%, or for two out of three samples, the recorded number of BMWP taxa present was less than the audit-corrected number (Table 7).

For the WHPT NTAXA index, the maximum bias (i.e. net under-estimation of the number of WHPT taxa present) was 8, but the median was one taxa and the average bias amongst the 427 samples analysed was 1.68 (Table 17). Three samples had two more WHPT taxa recorded than actually existed in the audited sample and 18 samples had one more WHPT taxa recorded than existed in the audited sample. Thus nearly 5% of samples actually had higher observed sample values of WHPT NTAXA than they should have. A further 20% of samples had the correct values of WHPT NTAXA. However, for the vast majority (75%) of samples, namely on average for three out of four samples, the recorded number of WHPT taxa present was less than the audit-corrected number (Table 7).

Table 17 Two-way table summarisng number of samples with each level of bias in BMWP NTAXA and WHPT NTAXA (equal biases are highlighted)

		Bias in WHPT NTAXA												
		-2	-1	0	1	2	3	4	5	6	7	8	Total	%
Bias in BMWP NTAXA	-2	3	0	0	0	0	0	0	0	0	0	0	3	0.7
	-1	0	7	5	0	0	0	0	0	0	0	0	12	2.8
	0	0	11	78	30	11	0	0	0	0	0	0	130	30.4
	1	0	0	3	76	47	12	2	0	0	0	0	140	32.8
	2	0	0	0	5	39	29	5	5	1	0	1	85	19.9
	3	0	0	0	0	1	13	12	9	1	1	0	37	8.7
	4	0	0	0	0	0	1	6	6	1	0	0	14	3.3
	5	0	0	0	0	0	0	0	0	4	2	0	6	1.4
	Total	3	18	86	111	98	55	25	20	7	3	1	427	
	%	0.7	4.2	20.1	26.0	23.0	12.9	5.9	4.7	1.6	0.7	0.2		100

The correlation between the individual sample bias in WHPT NTAXA and individual sample bias in BMWP NTAXA was high at 0.86 amongst all 427 samples.

In nearly half (48%) of all samples the bias in WHPT NTAXA was the same as in BMWP NTAXA. In 42% of samples the (positive) under-estimation of WHPT was greater than the under-estimation of BMWP NTAXA, whereas there were only 10 samples (2%) where the (positive) under-estimation of BMWP NTAXA was greater than the bias for WHPT NTAXA. Thus there is clear evidence that the average bias for WHPT NTAXA is greater than the bias for the original BMWP NTAXA.

With the exception of those few samples with more taxa recorded in the observed sample than found in the audited sample (i.e the 15 samples with negative BMWP NTAXA bias), at each level of BMWP NTAXA sample bias (0, 1, 2, 3, 4 or 5) the mean bias in WHPT NTAXA was greater than the BMWP NTAXA bias, the mean increase varied between 0.32 when BMWP NTAXA bias was zero up to a maximum of 1.33 when BMWP NTAXA bias was 5 (Table 8). This suggests that WHPT NTAXA bias might tend to vary as a multiple of BMWP NTAXA bias. The average ratio of WHPT NTAXA bias to BMWP NTAXA bias varied from 1.53 when BMWP NTAXA bias was 1 to 1.12 when BMWP NTAXA bias was 4 (Table 18).

Table 18 Average levels of bias in WHPT NTAXA for each level of BMWP NTAXA bias

BMWP NTAXA sample bias	No. of samples	Mean WHPT NTAXA : BMWP NTAXA bias		Bias in WHPT NTAXA			
		Difference	Ratio	Mean	Min	Median	Max
0	130	0.32	---	0.32	-1	0	2
1	140	0.53	1.53	1.53	0	1	4
2	85	0.69	1.35	2.69	1	2	8
3	37	0.97	1.32	3.97	2	4	7
4	14	0.50	1.12	4.50	3	4.5	6
5	6	1.33	1.27	6.33	6	6	7

Amongst the 282 audited samples with some under-estimation of BMWP NTAXA, the average ratio of WHPT to BMWP NTAXA bias was 1.42, suggesting WHPT NTAXA bias is on average 42% higher. However, for statistical reasons (Snedecor and Cochran 1980), the most precise estimate from these samples is probably the ratio of the mean WHPT NTAXA bias of 2.45 to the mean BMWP NTAXA bias of 1.80, namely a ratio of 1.36.

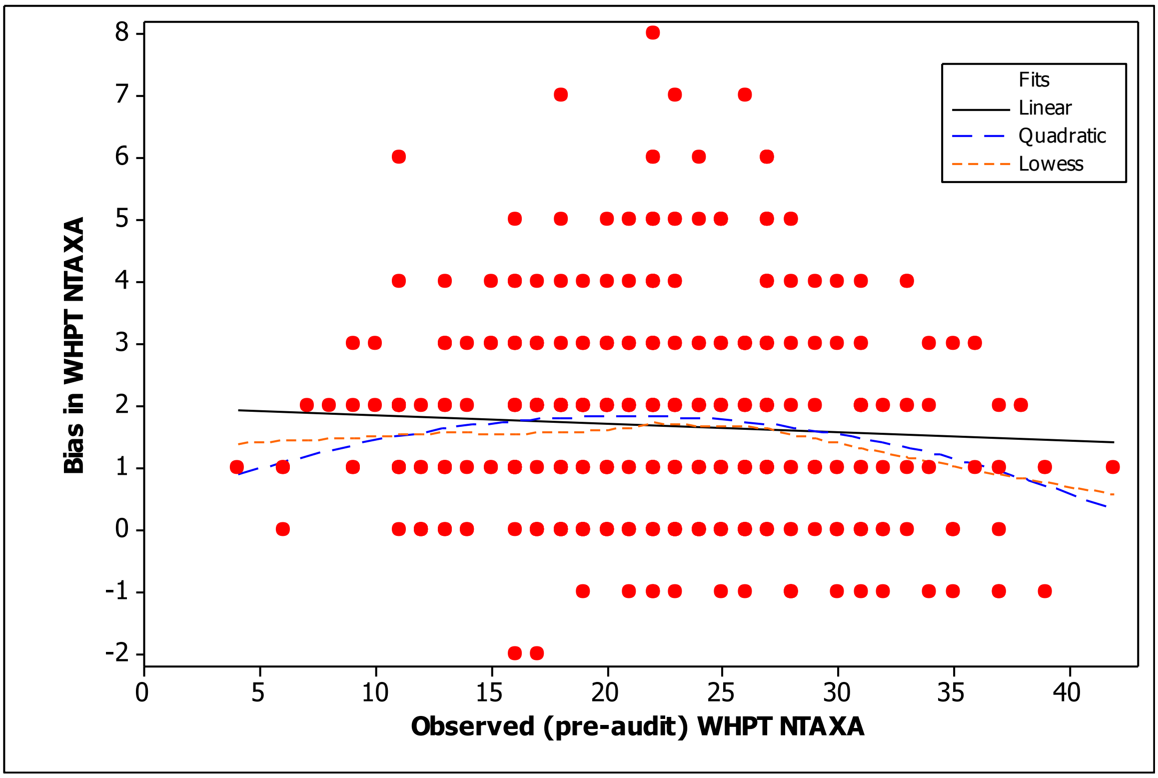
This close relationship between the two biases is fortunate as a user-supplied estimate of BMWP NTAXA bias is used in the current RICT software (and previous RIVPACS III+) to correct for sample processing errors in the estimate of site EQR values and status class and to simulate the extra uncertainty in river site status class due to these sample processing errors. In particular, the RICT user supplies an estimate of the audit-derived average BMWP NTAXA bias for the UK environment agency lab, region or agency sub-contractor lab appropriate to a river site being assessed.

5.1.2 WHPT NTAXA Bias in relation to observed (pre-audit) WHPT NTAXA

It might be thought that the more taxa there are at a site and in a sample, the more taxa are likely to be missed. However, there was no obvious relationship between the degree of under-estimation of true sample WHPT NTAXA and the number of WHPT taxa recorded as present in a sample (Figure 24). Although a quadratic relationship amoungst the large number of audited samples was statistically significant ($p = 0.021$), the relationship only explained a trivial amount of the variation ($r^2 = 1.8\%$, Figure 24). Unsurprisingly, there was also no linear or quadratic relationship between bias in WHPT NTAXA and the observed BMWP NTAXA sample values ($r^2 = 1.0\%$). This is the same result that was found in the orginal analysis of bias in BMWP NTAXA in the original BAMS study by Furse *et al* (1995).

Therefore there is no evidence to suggest that the average under-estimation in the number of WHPT taxa present in a sample increases (or decreases) with the recorded observed number of WHPT taxa present. The bias in WHPT NTAXA just seems to be, on average, about 36% larger than the BMWP NTAXA bias, roughly independent of the number of taxa recorded as present.

Figure 24 Bias in WHPT NTAXA in relation to the observed sample value of WHPT NTAXA (n = 427 audited samples)



Recommendation

The mean bias in WHPT NTAXA should be estimated as 1.36 times the appropriate user-supplied estimate of BMWP NTAXA mean bias for a RICT site (as input in the current RICT).

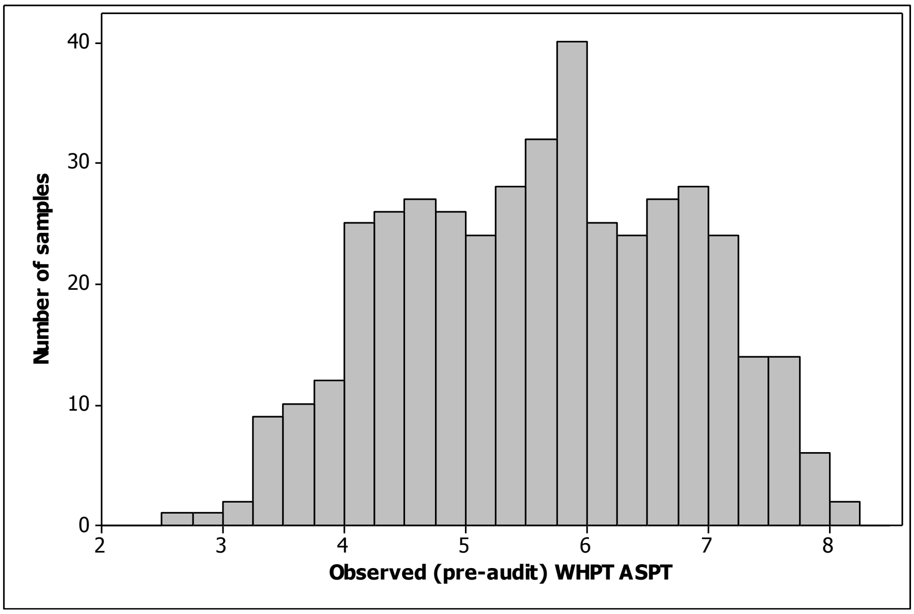
The RICT simulated biases for individual simulated samples should still be assumed to follow a statistical Poisson distribution, but with this higher Poisson mean.

5.2 WHPT ASPT bias

These analyses are all based on the abundance-weighted form of WHPT ASPT, as this is what is intended for use within the new RICT software. It is hereafter refered to simply as WHPT ASPT. WHPT ASPT bias for a particular samples equals the audit-corrected (post-audit) WHPT sample WHPT ASPT value minus the pre-audit 'observed' WHPT ASPT value for the sample.

The observed value of WHPT ASPT amongst the 429 audited sample dataset varied from 2.56 to 8.14 (which encompasses a major part of the range of WHPT ASPT observable in practice), with an inter-quartile range (i.e. middle 50%) of 4.69-6.58 and with a mean and median of around 5.7 (Figure 25).

Figure 25 Frequency histogram of the observed (pre-audited) values of WHPT NTAXA amongst the 429 audited samples dataset



5.2.1 Factors affecting WHPT ASPT bias values

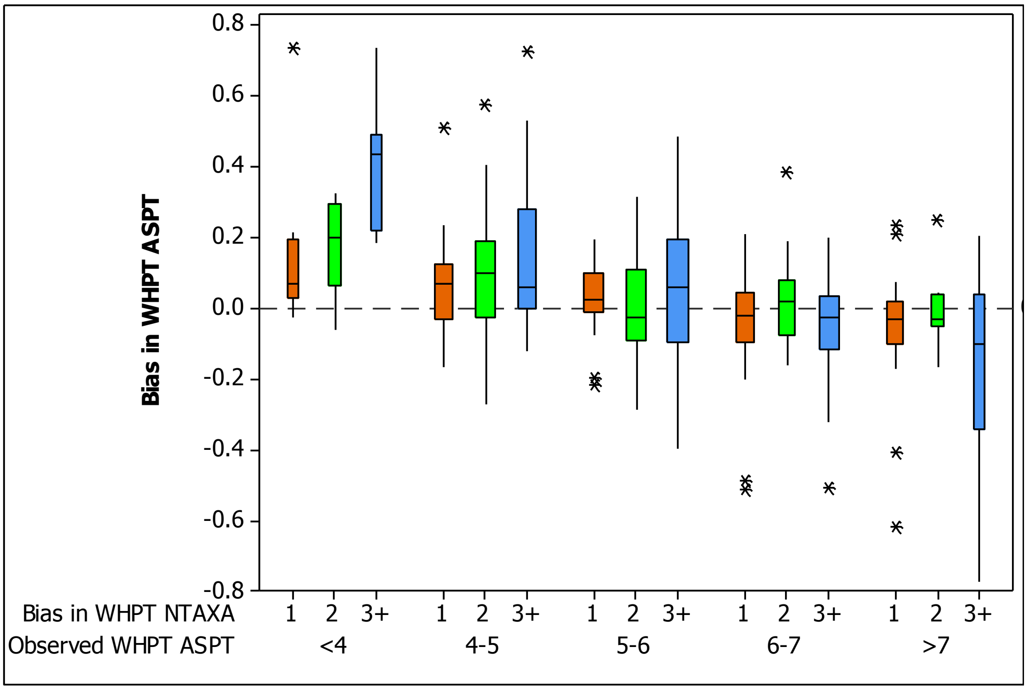
The sample bias in observed WHPT ASPT seems to vary systematically with the observed (pre-audit) value of WHPT ASPT (Table 19). When the observed WHPT ASPT was less than 4.0 the bias was positive in 71% of samples with a mean bias of +0.17. However, the average bias and the percentage of positive biases decreases with increasing observed WHPT ASPT, such that for the 57 samples with high (>7) observed WHPT ASPT, only 25% of WHPT ASPT biases were positive and the average bias was -0.11 (Table 19 Bias in WHPT ASPT in relation to observed (pre-audited) sample WHPT ASPT).

Table 19 Bias in WHPT ASPT in relation to observed (pre-audited) sample WHPT ASPT

Observed (pre-audit) WHPT ASPT	Samples	Bias in WHPT ASPT Mean	Min	Median	Max	% positive
2.5 - 4	34	0.17	-0.36	0.15	0.74	71%
4 - 5	104	0.09	-0.27	0.04	0.73	60%
5 - 6	116	0.02	-0.40	0.00	0.48	48%
6 - 7	95	-0.03	-0.51	-0.00	0.38	38%
7 - 8.2	57	-0.11	-0.80	-0.04	0.25	25%

Thus for sites with low recorded WHPT ASPT, the average WHPT score of the missed taxa is slightly higher than the average abundance-weighted WHPT score of the observed (i.e. recorded) taxa, whereas for sites with high recorded WHPT ASPT, the average WHPT score of the missed taxa is usually slightly lower than that of those recorded as present in the sample. This is shown graphically in Figure 26, which also aims to show whether the bias in WHPT ASPT varies with the degree of under-estimation (i.e. bias) of the number of WHPT taxa present. For samples with very low observed WHPT ASPT values (<4), the bias is greater when the under-estimation of WHPT NTAXA is 3 or more, but there were only four such samples in our audit dataset of 429 samples. For the rest of the range of observed WHPT ASPT values, the bias does not seem to vary systematically with the extent of under-estimation of the number of WHPT taxa present (Figure 26).

Figure 26 Boxplot of bias in WHPT ASPT in relation to the observed sample WHPT ASPT value and the bias in WHPT NTAXA (1, 2 or 3-8); boxplot width proportional to number of samples



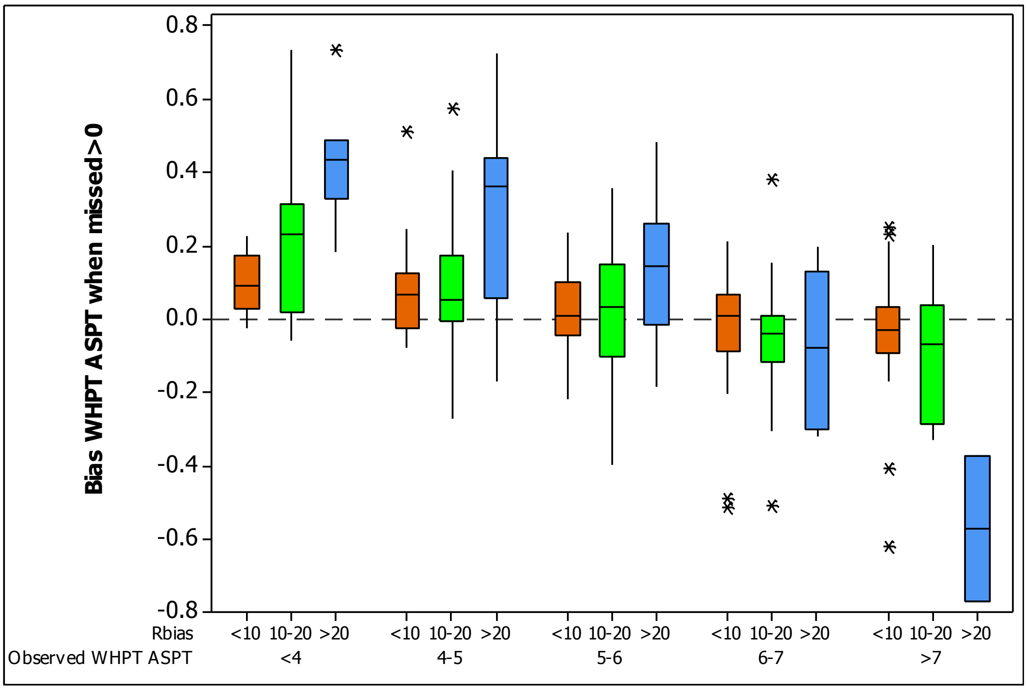
The best of these approaches is judged to be to predict the bias in WHPT ASPT for a sample from its observed WHPT ASPT using the following regression relationship (coefficient standard errors in brackets):

Bias in WHPT ASPT = 0.401 - 0.0681 Observed WHPT ASPT (Eqn 5.1)
(0.041) (0.0072)

Equation 5.1 explains 18% of the total variation in WHPT ASPT bias values (amongst samples with under-estimation of WHPT NTAXA) and has a residual SD of 0.170 which can be used to estimate the uncertainty in the predictive estimates of WHPT ASPT bias for individual samples.

The influence of a certain number of missed taxa on WHPT ASPT bias may tend to be less when the observed WHPT ASPT values were based on many observed WHPT taxa. However, the percentage ratio (R_{Bias}) of bias in WHPT NTAXA to observed WHPT NTAXA did not significantly improve (test p =0.166) the relationship of equation 5.1. The effect of having relatively larger numbers of missed taxa varies with the observed WHPT ASPT, but tends to increase the variability (and thus uncertainty) in values for the bias in WHPT ASPT (Figure 27).

Figure 27 Bias in WHPT ASPT in relation to observed WHPT ASPT and the percentage ratio (R_{Bias}) of bias to observed WHPT NTAXA (<10%, 10-20%, >20%)



5.2.2 Factors affecting WHPT ASPT of missed taxa

The WHPT ASPT of the missed WHPT taxa in a sample is calculated by the increase in WHPT score when audit-corrected divided by the sample bias in WHPT NTAXA, but this is defined only when WHPT NTAXA bias is positive.

The WHPT ASPT of the missed taxa has a tendency to increase with the observed (pre-audit) sample WHPT ASPT, from an average of 5.31 up to an average of 6.41 as observed WHPT ASPT increases from less than 4.0 up to 7.0 or more (Table 20). (Note: It is possible for the WHPT ASPT of the missed taxa to be negative or greater than the largest WHPT for any individual taxa (which is 13.0 for high abundance of Perlidae, Appendix 2) because the audit correction can includes losses as well as gains which are all absorbed into the net change in number of WHPT taxa and WHPT score).

Thus the WHPT ASPT of the missed taxa tends to be higher than the observed recorded WHPT when the observed value is low and lower than the observed value when the latter is high; this re-enforces our earlier finding from Table 9 and Figure 26.

Table 20 WHPT ASPT of missed taxa in relation to observed (pre-audited) sample WHPT ASPT

Observed (pre-audit) WHPT ASPT		WHPT ASPT of missed taxa			
	Samples	Mean	Min	Median	Max
2.5 - 4	26	5.31	3.0	5.26	8.5
4 - 5	81	5.21	2.7	5.30	11.6
5 - 6	94	5.90	-0.5	5.90	11.1
6 - 7	77	6.06	-5.1	6.27	13.6
7 - 8.2	42	6.41	-8.8	6.70	14.0

The WHPT ASPT of the missed taxa also has a tendency to be lower when the observed (pre-audit) number of WHPT taxa recorded present is low, increasing from an average of 5.13 for samples with 10 or less WHPT taxa observed up to averages of 6.2 or more when more than 20 WHPT taxa are recored as present (Table 21). A similar pattern is observed between the WHPT ASPT of the missed taxa and the number of BMWP taxa observed in the sample pre-audit (Table 22).

Table 21 WHPT ASPT of missed taxa in relation to observed (pre-audited) sample WHPT NTAXA

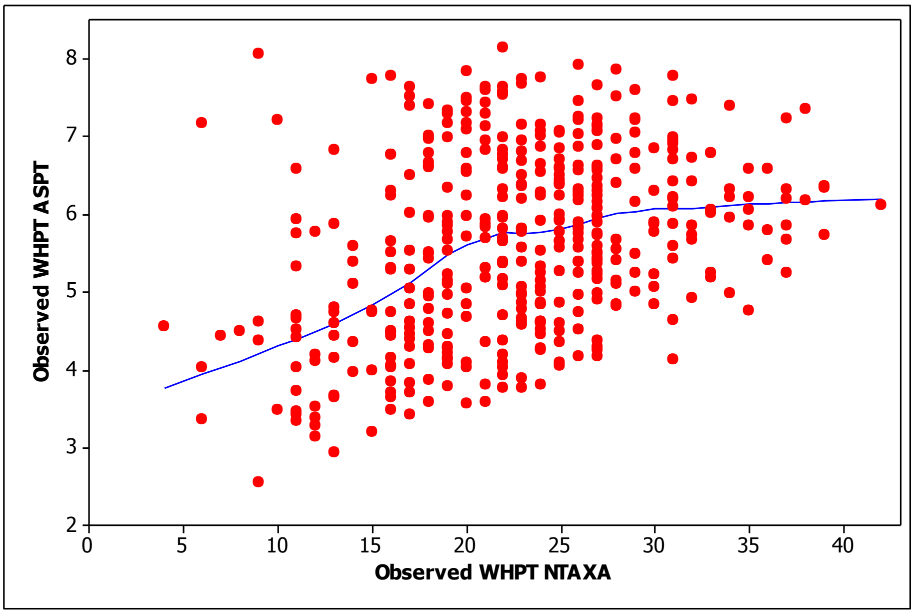
Observed (pre-audit)		WHPT ASPT of missed taxa			
WHPT NTAXA	Samples	Mean	Min	Median	Max
≤10	11	5.13	2.4	5.13	8.5
11-15	31	5.44	1.2	5.40	11.6
16-20	79	6.00	-4.0	6.00	11.4
21-25	100	6.83	-5.1	5.57	14.5
26-30	78	6.21	-8.8	6.30	14.0
31-42	42	6.66	-0.2	6.52	17.1

Table 22 WHPT ASPT of missed taxa in relation to observed (pre-audited) sample BMWP NTAXA

Observed (pre-audit)		WHPT ASPT of missed taxa			
BMWP NTAXA	Samples	Mean	Min	Median	Max
≤10	24	5.19	2.4	5.30	8.5
11-15	44	5.61	1.2	5.42	11.6
16-20	82	6.08	-4.0	6.07	11.4
21-25	118	6.78	-8.8	5.95	14.0
26-30	35	6.11	-0.2	6.30	13.6
31-36	16	6.43	3.2	6.62	9.4

The observed number of taxa (WHPT or BMWP) and observed WHPT ASPT are not completely independent in practice in that when many taxa are recorded present (>30), observed WHPT ASPT values are never low (i.e less than 4) (Figure 28). Thus the WHPT ASPT of the missed taxa in samples with many observed taxa will also tend to have relative high observed WHPT ASPT values.

Figure 28 Relationship between observed WHPT ASPT and observed WHPT NTAXA amongst the 429 audit samples (correlation r = 0.36); line indicates fitted Lowess regression smoother.



The best approach is judged to be to predict the WHPT ASPT of the missed taxa by the observed (pre-audit) value of WHPT ASPT using the following fitted regression relationship (coefficient standard errors in brackets):

WHPT ASPT of missed taxa = 4.35 + 0.271 Observed WHPT ASPT (Eqn 5.2)
(0.62) (0.107)

Although only explaining a small proportion of the total variation in WHPT ASPT of the missed taxa in individual samples, the average of the predicted values from Equation 5.2 closely matches the actual mean for samples in each band of observed WHPT ASPT values, indicating useful broad predictive power across a range of quality of sites (Table 23).

Table 23 Mean of actual and predicted (from Eqn 5.1) values of the WHPT ASPT of the missed taxa in relation to observed (pre-audited) sample WHPT ASPT

Observed (pre-audit) WHPT ASPT		Mean WHPT ASPT of missed taxa	
	Samples	Actual	Predicted from Equation 5.1
2.5 - 4	26	5.31	5.32
4 - 5	81	5.21	5.56
5 - 6	94	5.90	5.85
6 - 7	77	6.06	6.11
7 - 8.2	42	6.41	6.35

To assess whether the uncertainty in the WHPT ASPT of the missed taxa is less when there are more missed taxa (i.e WHPT NTAXA sample bias is greater), we grouped the audited samples by both their observed WHPT value and the under-estimated number (M) of WHPT taxa and calculated the mean and SD of WHPT ASPT of the missed taxa for the samples in each category (Table 24). The best single estimate of SD for each value of M was obtained from a one-way analysis of variance of WHPT ASPT of missed taxa on class of observed WHPT to gives estimates of SD of 3.28, 1.53 and 1.24 for M equal to 1, 2 and more than 2 respectively (Table 24).

Table 24 WHPT ASPT of missed taxa in relation to observed (pre-audited) sample WHPT ASPT and bias (M) in WHPT NTAXA (samples per category in brackets)

Observed (pre-audit) WHPT ASPT	Mean of WHPT ASPT for M under-estimated WHPT taxa			SD of WHPT ASPT for M under-estimated WHPT taxa		
	M=1	M=2	M>2	M=1	M=2	M>2
2.5 - 4	5.36 (9)	5.12 (10)	5.51 (7)	1.55	1.28	0.95
4 - 5	5.87 (27)	5.42 (28)	5.27 (26)	2.07	1.38	1.20
5 - 6	6.37 (29)	5.58 (27)	5.77 (38)	2.56	1.61	1.27
6 - 7	5.34 (26)	6.83 (21)	6.13 (30)	4.26	1.52	1.17
7 - 8.2	6.08 (20)	6.88 (12)	6.50 (10)	4.41	1.87	1.54
Overall average SD from ANOVA				3.28	1.53	1.24
Predicted SD using SD=(2.5/√ M)				2.5	1.76	1.44 (M=3) 1.25 (M=4)

Thus the SD of the WHPT ASPT of the missed taxa declines with the number of missed taxa (or more precisely, with the bias in WHPT NTAXA). This is a similar type of result to that found for the variability in the ASPT of the missed BMWP taxa in the original BAMS study of Furse *et al* (1995) and used in the current RICT for bias-correction of BMWP ASPT. One difference is that the SD are slightly larger for WHPT ASPT than BMWP ASPT and also the WHPT ASPT of missed taxa is most closely related to observed WHPT ASPT rather than the observed number of BMWP or WHPT taxa.

To enable prediction of the SD of WHPT ASPT for any value of M, we suggest using a simple statistical formula to represent the main pattern of observed SD, namely:

$$\text{SD of WHPT ASPT when WHPT NTAXA Bias is } M = 2.5 / \sqrt{M} \qquad \text{(Eqn 5.3)}$$

where √ M indicate square root of M. The effect of using equation 5.3 is shown in Table 24.

Recommendation

In RICT, the WHPT ASPT of a sample with a simulated estimated bias of WHPT NTAXA of M should be estimated as random deviate from a statistical Normal distribution with a mean given by Equation 5.2 (i.e a linear increasing function of observed WHPT ASPT) and a standard deviation equal to 2.5 divided by the square root of M).

6. Algorithms to simulate uncertainty for WHPT, LIFE and PSI in RICT

6.1 General simulation approach to assessing uncertainty in RICT

The RICT (i.e. RIVPACS) approach to assessing the ecological status of UK river sites compares the observed (O) sample values of the biota with the RIVPACS predictive model site- and season-specific expected (E) biota, currently through the use of the ratio (O/E) of observed to expected values of derived biotic indices. Past RIVPACS and current RICT site assessments are based on classifiying O/E values (or more specifically adjusted EQR values) for each of BMWP NTAXA and BMWP ASPT into pre-set WFD ecological status classes and then taken the lower of the two index classes as the overall status class of the site. In theory, both the observed and expected index values will be subject to a combination of sampling variation, error and uncertainty.

In RICT, uncertainties in observed (O) and expected (E) values of one or more biological indices and thus in their Ecological Quality Index (EQI = O/E) values, adjusted Ecological Quality Ratios (EQR) and the resulting confidence of assigning river sites to ecological status classes are all assessed using statistical Monte Carlo simulation procedures. Based on previously-derived estimates of variance parameters for the effects of typical replicate sampling variation, within-season temporal variation and between-year-within-period variation (for multi-year assessments) in biotic index values, Monte Carlo simulations are used to generate a large number of simulated potential values of the observed (O) and the expected (E) values of each index and hence of the O/E ratios. The lower and upper 2.5% values of the frequency distribution of the simulated O/E values are used to estimate 95% confidence limits for the true average O/E ratio for the site during that time period. Each simulated O/E value can also be classified to ecological status class based on O/E class limits for each index. The proportions of simulated values falling in each class can be used to estimate the probabilities of each class being the true (average) quality of the test site during that time period.

At least 9999 simulations should be used, and the same arbitrary but fixed random number seed is used in each run of RICT so that exactly the same uncertainty results and probabilities of class are obtained in repeated runs on the same data.

The general statistical estimation and simulation approaches were first summarised in detail in the RIVPACS III+ Release 1.2 User manual (Clarke *et al* 1997) and in Clarke (2000).

The precise detailed equations and algorithms used by the IT programmer in the initial and current version of the RICT software were provided by us (Ralph Clarke) on pages 41-60 of the SNIFFER project WFD72C Final Report (Davy-Bowker *et al* 2008).

Nearly all of those equation and algorithms will still be applicable and unchanged for the new version of RICT which will include the ability to base river site assessments on the new taxonomic abundance-based WHPT indices (WHPT NTAXA and WHPT ASPT) instead of the previous BMWP system indices (BMWP NTAXA and BMWP ASPT).

These same general uncertainty simulation equations and algorithms in the current RICT will mostly also apply if site assessments incorporate use of one or both of the newer LIFE and PSI indices. One difference is that in our recent assessment and estimation of sampling variance component (Section 3 of this report), we found that the size of sampling variation in both LIFE and PSI varied negatively with the number of BMWP taxa (NTAXA) present; the required modifications to the sampling simulation algorithms are given below in Section 6.3.

For ease of use by the IT programmers enhancing the RICT software to include the new WHPT, LIFE and PSI uncertainty assessments, we repeat all of the relevant previous parameter definitions and simulation algorithms given on pages 43-60 of the WFD72C Final Report, but highlight any additions for these new indices in green shaded text.

6.2 Statistical Methods used for single- and 3-year site assessments

The RICT software includes the ability to make river site assessments of ecological status over either single years or 3-year periods.

Single year assessments are assumed to be based on one observed sample obtained in each of one, two or three of the RIVPACS seasons (Spring (Mar-May), Summer (June-Aug), Autumn (Sep-Nov)) in the year.

Note: Variance = the square of the Standard Deviation. Below the statistical equations are given in variance form, as is usual and more succinct to write. However, in the algorithms sections, the equations are given using the SD form of parameters, as this is the form input into the RICT software and as given in the orginal algorithms sections of the SNIFFER WFD72C report.

6.2.1 Statistical methods for assessments based on BMWP indices

In the past RIVPACS software and in the current RICT, assessments were based on O/E ratios for the BMWP indices using either a single season observed sample or the combined taxonomic sample information obtained from two or three seasons (spring, summer autumn) combined sample. Thus there is one observed value for each index per year. This is divided by the single expected value of the index for that site and season or combined season to give a single O/E value for that particular index for the site for for the year.

In the current RICT, multi-year assessments using the BMWP indices are also available; these are based on an estimate of the average quality over a three year period. Observed (O) index values (one per year) will be available for one, two or all three years in the period. The same season or combination of seasons are assumed to be sampled each year with any sampling.The same field-based environmental predictor variables may or may not have been measured in each year. If a single expected (E) index value is used in each of the up to three sampled years, then:

average of the (O/E) $= \{O/E\}_{Av}$ = average of the (O) / E = O_{Av} / E

If different E values are available for each year then:

average of the (O/E) $= \{O/E\}_{Av}$ ≠ average of the (O) / average of the (E) = O_{Av} / E_{Av}

However, for any single site, the expected (E) is relatively constant and insensitive/robust to the typical degree of within-site changes in the environmental variables that occur (Armitage, 2000).

To assess uncertainty in site assessments, we have available and have analysed a mixture of datasets containing varying degress of replicated samples and time series of observed RIVPACS sample data (see Section 2 of this report). These historical datasets did not have the RIVPACS environmental predictor variables or RIVPACS expected (E) index values available for each site. However, we were able to fit statistical variance component models to estimate (i) the components of uncertainty (replicate, and within-season temporal) in the single observed (O) combined season sample value for a single year assessment and (ii) the variance components (replicate, within-season temporal and between-year-within-period) involved in the uncertainty of average (across years) observed O index value for a 3-year assessment period.

For a combination of the above reasons, the estimate of the average quality of a site for a three year period that is to be used in RICT 3-year assessments is defined as:

Estimate of average quality = average of the (O) / average of the (E) = O_{Av} / E_{Av}

where the individual O value for each year is the observed index value either for the sampled single season or for the combined season sample (assumed to be the same season(s) sampled each year).

The uncertainty variance associated with this estimator of average quality can then be estimated from the uncertainty variance (Var_{ObsAv}) associated with estimate average O value and the uncertainty associated with estimating the average E value for the three year period at that site.

For single year assessments:

Estimated variance of the O value = $Var_{Obs} = Var_{Rep} + Var_{TSeas}$

and for 3-year assessments:

Estimated variance of the average of the O value = $Var(O_{Av})$ =
$Var_{ObsAv} = (Var_{Rep} + Var_{TSeas} + Var_{Tyear}\ (1 - N_{ObsYear}\ /3)\ )\ /\ N_{ObsYear}$

where Var_{Rep} = Variance between replicate samples at a site on the same day(s)
Var_{Tseas} = Variance due to typical temporal within-season variability (i.e different days/months)
Var_{TYear} = Variance due to temporal between-year within-period variability
and $N_{ObsYear}$ = Number of years (1, 2 or 3) for which a sample was involved in the estimate of the average O value

Note: The above variance terms may apply to single season samples or to two- or three-season combined samples. These equations are the same as proposed in WFD72C and as used in the original RICT

Notes on assumptions for all assessments (using BMWP or abundance-weighted indices):

(i) The RICT simulations of the stochastic variability assume that the observed index values are based on a single sample from that 'season' in any one year (where 'single sample' for a spring and autumn combined season sample analysis means one sample in spring and one sample in autumn in the same year have been combined). If some future assessments were based on more than one sample in each 'season' of each year then the average of the observed sample values would have smaller random uncertainty estimates; but such costly more intensive sampling strategies are not expected to occur and thus are not catered for here.

(ii) The RICT (and RIVPACS III+) uncertainty assessments are based on the assumption that the same single sampling site with a WFD water body has been sampled on each occasion, and that furthermore, this site is representative of the ecological status of the WFD water body as a whole. Therefore the uncertainty estimates are actually for ecological quality at that river site; there is potentially additional uncatered for uncertainty due to un-quantified spatial variability between possible sampling sites with the water body.

6.2.2 Statistical methods for assessments based on average of single season sample EQR for the abundance-weighted WHPT, LIFE and PSI indices

The next upgraded version of RICT is to include the option of site assessments which are based on using the abundance-weighted WHPT indices, perhaps optionally also involving the LIFE and/or PSI indices. For any of these abundance-weighted indices, the use of the individual season sample information is different to that for the BMWP indices. Rather than combining the sample taxonomic information across two- or three sampled seasons to form a single 'combined-season' observed taxonomic compostion sample for the year, for these newer indices, the assessment will be based on the average of the individual seasons' O/E ratios (or derived EQR), where the E values are as always site- and season-specific. This is a major change to the assessment process, involving not only taxonomic abundances but also allowing poorer site quality (i.e much lower O/E) in any one season to have a greater impact on the estimate of site quality for the year than using the combined season sample approach.

For single year site assessments:

Average site O/E = $\{O/E\}_{Av}$ = average of the sampled single season O/E values

For 3-year period site assessments:

Average site O/E = $\{O/E\}_{Av}$ = average of individual years' average O/E values for the period

which with equal sampling per sampled year will just be the overall average of the O/E.

Note: It is a requirement and assumption within RICT for 3-year period site assessments that the same number of seasons have been sampled in each year during the period for which samples are involved in 3-year period assessment. Usually the same season (e.g. autumn) or same pair of seasons (e.g. spring and autumn) will be sampled in each year. However, the sampling uncertainty for an estimate of average site O/E over the period is only dependent on the assumption that the same numbers of seasons ($N_{ObsSeas}$) were sampled in each of the $N_{ObsYear}$ years (1, 2 or 3) sampled during the period. This is because the individual season sample EQI/EQR values have standardised for differencs between seasons and the replicate and within-season temporal variances in index values do not appear to vary between seasons. Therefore, it is permitted for say years 1 and 3 average quality to be based on the average of the standard spring and autumn sample EQIs/EQRs, whilst year 2 average quality is based on average of spring and summer sample EQIs/EQRs.

Although the estimate of average site quality, $\{O/E\}_{Av,}$ over a single year or over a 3-year period will be estimated by the average of the individual season sample O/E values, the uncertainty variance associated with this average O/E will need to be approximated. This is because we do not have direct estimates of the replicate and temporal variance components associate with these sample O/E ratios, but rather only with the observed sample (O) index values.

We assume for variance estimation purposes only, that the variation in the E values of an index for a particular site is relatively small (i.e. has a low coefficient of varation between seasons). The UK environment agencies tend to use the same single set of measured RIVPACS environmental predictor variables and thus the same season-specific E values for a site for every year, so these will generally not change between years over the 3-year period anyway. This assumption allows us to use the average of the season-specific E values for an index in our estimation of the variance of the average O/E for the site.

Specifically, the uncertainty variance, $Var(\{O/E\}_{Av})$, of the average O/E ($\{O/E\}_{Av}$) is estimated by:

$$Var(\{O/E)\}_{Av}) = Var(O_{Av}) / E_{Av}$$

where O_{Av} = average of the observed (O) values over the assessment period (1 or 3 years)
E_{Av} = average of the expected (E) values over the assessment period

Therefore we need an estimate of the variance, $Var(O_{Av})$, of the average observed values O_{Av}.

For single year assessments:

$$Var(O_{Av}) = (Var_{Rep} + Var_{TSeas}) / N_{ObsSeas}$$

and for 3-year assessments:

$$Var(O_{Av}) = ((Var_{Rep} + Var_{TSeas}) / N_{ObsSeas} + Var_{Tyear} (1 - N_{ObsYear} / 3)) / N_{ObsYear}$$

where Var_{Rep} = Variance between replicate samples at a site on the same day
Var_{Tseas} = Variance due to temporal within-season variability (i.e. different days/months)
Var_{TYear} = Variance due to temporal between-year within-period variability
$N_{ObsSeas}$ = Number of seasons (1, 2 or 3) sampled each year
$N_{ObsYear}$ = Number of years (1, 2 or 3) for which samples were involved in the estimate of the average O value

Errors in expected (E) values:

There is also, in theory, an error for the RIVPACS predicted E values for each site due to RIVPACS modelling inefficiencies and other errors. For the original RIVPACS III+ software, it was possible to estimate the size of the errors in predicting the E values of the BMWP indices due to inter-personnel variability in measurement of the RIVPACS environmental predictor variables, but this has not been possible for the newer indices (WHPT, LIFE and PSI).

For the WHPT indices, it should be reasonable to approximate the error SD for the expected (E) values of the WHPT NTAXA and WHPT ASPT indices with the estimates derived for BMWP NTAXA and BMWP ASPT respectively.

From experience with the BMWP indices, these errors were relatively small compared to those in the observed (O) values due to sampling variation. Moreover, this source of errors in the E values would also be much lower if the field-based meaurments of stream width, depth and substrum composition were not involved in site predictions of E values, as recommended when using the flow and fine sediment stress-related indices of LIFE and PSI.

Estimates of sampling component SD values

The newly-derived recommended estimates of the variance components for each abundance-weighted index to be involved in future developments of RICT are based on the best information currently available and the standard devations (square root of variance estimates) for each component are given in Section 3.7 of this report. These estimates should be used in the updated RICT tool, as detailed in the Algorithms section 6.3 below.

Assessments of uncertainty of change between two estimates of O/E or two estimates of average O/E in the RICT algorithms and software are based on a similar simulation logic to that developed for RIVPACS software procedure COMPARE).The two cases of (average) O/E values being compared could, in theory, be from the same site in different individual years or different 3-year periods, or from different sites in the same or different individual years or 3-year periods. In any one simulation in the RICT ‘Compare’ procedure, the appropriate variance of the observed and O/E values for each of the two cases (which could involve different seasons or number of seasons) is used to derive independent stochastic simulations of the potential O/E values and thus of their difference that could have been obtained. The frequency distribution of these simulated differences in O/E and changes in status class is then used to assess the likelihood of change in O/E and likelihood of change in class.

6.3 Algorithms for estimates and simulation of their uncertainty

6.3.1 Definitions

For any particular site:

i = id of current biological index
y = year id within assessment period (1, 2 or 3)
s = id of selected season(s) combination of seasons (referred to as 'season');
(1 = spring, 2 = summer, 3 = autumn, 4 = spring + summer,
5 = spring + autumn, 6 = summer + autumn, 7 = all three seasons)

The term "sample" refers to the total sample for which the index values were calculated; this could be a single season sample, or a two or three season combined sample.

Obs_{isy} = Observed sample value of index i in season s of year y for current site
Exp_{isy} = Expected value of index i in season s of year y for current site

O/E_{isy} = Obs_{isy}/Exp_{isy} = O/E ratio value of index i in season s of year y for current site

$Obs_{Av(i)}$ = Average of the observed sample values (Obs_{isy}) of index i over the assessment period (either 1 or 3 years) for the current site

$ExpIDX_i$ = Average of the Expected sample values (Exp_{isy}) of index i over the assessment period (either 1 or 3 years) for the current site

$\{O/E\}_{Av(i)}$ = Average of the individual sample O/E values of index i over the assessment period (either 1 or 3 years) for the current site

$ObsIDX_i$ = value of index i for current site around which to centre sampling uncertainty simulations

$ObsIDX_{ir}$ = r^{th} simulated value of average observed sample value of index i for current test site

Index id i	Index Name	Transformation of index prior to adding simulated "error" terms
1	BMWP Score	derived from indices 2 & 3 [+]
2	NTAXA	Square root
3	ASPT	none
4	WHPT Score (non-abundance weighted)	derived from indices 5 & 6 [+]
5	WHPT NTAXA (non-abundance weighted)	Square root
6	WHPT ASPT (non-abundance weighted)	none
7	WHPT Score (abundance weighted)	derived from indices 8 & 9 [+]
8	WHPT NTAXA (abundance weighted)	Square root
9	WHPT ASPT (abundance weighted)	none
10	LIFE (family level)	None, but depends on NTAXA
11	PSI (family level)	Arcsine(Sqr(PSI/100)) and depends on NTAXA

[+] Because WHPT ASPT = WHPT Score / WHPT NTAXA, then $ObsIDX_{7sy} = ObsIDX_{8sy} * ObsIDX_{9sy}$ (apart from rounding errors) so, for consistency, we derive simulated values for $ObsIDX_{7r}$ from simulated values for $ObsIDX_{8r}$ and $ObsIDX_{9r}$ (the latter two have statistically uncorrelated sampling variation) - this is as equivalently done previously for BMWP Score, and potentially for unweighted WHPT indices

Note: Because any single software run only involves one selected season or combination of seasons (s = 1-7) the subscript s is dropped for the observed and expected index values to make it easier to include a subscript for the r^{th} simulated value of the observed and expected values of each index.

SD denotes standard deviation = square root of the equivalent variance component
(where appropriate on the transformed index scale)

$SDRep_i$ = Replicate sampling SD of transformed observed values of index *i*

$SDTSeas_i$ = Within-season temporal variability SD of transformed observed values of index *i*

$SDTYear_i$ = Between-year within-period variability SD of transformed observed values of index *i*

$SDObs_i$ = Overall uncertainty SD for average observed value of index *i* over the assessment period

(All of the required estimates of SD were derived in Section 3 and summarised in Section 3.7)

$N_{ObsYear}$ = number of years (1, 2 or 3) for which a sample was involved in the estimate of the average O value

$N_{ExpYear}$ = 1 for single-year assessment period
= number of years (1, 2 or 3) for which a separate estimate of the E value for a season was involved in the estimate of average E value (for 3-year assessment period)
If a single estimate of E was used for all three years, then $N_{ExpYear}$ = 1)

6.3.2 Algorithms for simulating sampling variation in (average) observed index values

A simulated potential average observed index value for a site and period (1 or 3 years), denoted $ObsIDX_{ir}$ for each simulation is generated as follows:

Transform (if necessary) the average observed sample value (O_{Av}) to the appropriate scale
Adding the appropriate random error term ($ZObs_{ir}$)
Back-transform (if necessary) to the original index scale.

For example, with index 8 (abundance-weighted WHPT NTAXA), we square root (√) the average observed sample value, add a random error term based on normal distribution with zero mean and SD of $SDObs_i$, and then square the result to get a simulation of the potential value of average observed sample WHPT NTAXA we could have obtained by chance if different single replicate samples had been taken at the site in the same seasons in each sampled year of the assessment period (1 or 3 years).

6.3.2.1 Algorithm Equations applicable to every index:

For each simulation *r*, for both single- and multi-year runs, derive:

$ZNorm_{ir}$ = Random number deviate from a standard Normal distribution with a mean of 0.0 and SD of 1.0 for index *i* in simulation *r*

$ZObs_{ir}$ = Random deviate for potential average observed value of index *i* in simulation *r*

$= ZNorm_{ir} * SDObs_i$

where $SDObs_i$ is specific to index i

Note: RICT software already includes code to generate random numbers from a standard normal distribution.

6.3.2.2 Algorithm Equations and SD estimates specific to each index:

The precise equation to generate simulated average observed values differs between indices according to:

(i) the transformation required to make the variability independent of index value
(ii) whether or not the average of single season EQR (O/E) are being used
(iii) whether the variation depends on the number of taxa present.

The algorithm for each index group and index is given below:

(a) BMWP indices (NTAXA, ASPT and BMWP score):

Estimates (copied from Table 14) of the replicate sampling SD ($SDRep_i$), within-season temporal SD ($SDTSeas_i$) and between-year-within-period SD ($SDTYear_i$), for use in the above equations to estimate $SDObs_i$ are:

Index	Index Id	Index form	SD_{Rep}	SD_{TSeas}	SD_{TYear}
Original BMWP	1	√ Score	0.657	0.544	0.524
	2	√ NTAXA	0.240	0.187	0.191
	3	ASPT	0.256	0.244	0.144

$ObsIDX_2 = Obs_{Av(2)}$
$ObsIDX_3 = Obs_{Av(3)}$
$ObsIDX_1 = ObsIDX_2 * ObsIDX_3$

For single-year runs:

$$SDObs_i = \sqrt{(SDRep_i)^2 + (SDTSeas_i)^2}$$

For multi-year (3-year period) runs:

$$SDObs_i = \sqrt{((SDRep_i)^2 + (SDTSeas_i)^2 + (SDTYear_i)^2 * (1 - N_{ObsYear} / 3)) / N_{ObsYear}}$$

$ObsIDX_{2r} = (\sqrt{ObsIDX_2} + ZObs_{2r})^2$ = r^{th} simulated value for observed BMWP NTAXA
$ObsIDX_{3r} = ObsIDX_3 + ZObs_{3r}$ = r^{th} simulated value for observed BMWP ASPT
$ObsIDX_{1r} = ObsIDX_{2r} * ObsIDX_{3r}$ = r^{th} simulated value for observed BMWP Score

(b) All WHPT indices:

Estimates (copied from Table 14) of the replicate sampling SD ($SDRep_i$), within-season temporal SD ($SDTSeas_i$) and between-year-within-period SD ($SDTYear_i$), for use in the above equations to estimate $SDObs_i$ are:

Index	Index Id	Index form	SD_{Rep}	SD_{TSeas}	SD_{TYear}
WHPT Non-weighted	4	√ Score	0.672	0.607	0.554
	5	√ NTAXA	0.247	0.211	0.198
	6	ASPT	0.248	0.257	0.131
WHPT Abundance-weighted	7	√ Score	0.670	0.654	0.556
	8	√ NTAXA	0.247	0.211	0.198
	9	ASPT	0.269	0.279	0.174

For the WHPT, LIFE and PSI indices, the site assessment is based on the average of the individual sample O/E values for the assessment period. Because the average single season sample O/E value is not equal to the average of the O values divided by the average of the E values, we need to centre the simulated sampling uncertainty around the appropriate O value to ensure that the simulated uncertainty in the average O/E is centred on the estimate of average O/E. This is done using the following:

$ObsIDX_5 = \{O/E\}_{Av(5)} * ExpIDX_5$
$ObsIDX_6 = \{O/E\}_{Av(6)} * ExpIDX_6$
$ObsIDX_4 = ObsIDX_5 * ObsIDX_6$

$ObsIDX_8 = \{O/E\}_{Av(8)} * ExpIDX_8$
$ObsIDX_9 = \{O/E\}_{Av(9)} * ExpIDX_9$
$ObsIDX_7 = ObsIDX_8 * ObsIDX_9$

For single-year runs:

$$SDObs_i = \sqrt{((SDRep_i)^2 + (SDTSeas_i)^2) / N_{ObsSeas} }$$

For multi-year (3-year period) runs:

$$SDObs_i = \sqrt{(((SDRep_i)^2 + (SDTSeas_i)^2) / N_{ObsSeas} + (SDTYear_i)^2 * (1 - N_{ObsYear} / 3)) / N_{ObsYear}}$$

For the non-abundance-weighted WHPT indices:

$ObsIDX_{5r} = (\sqrt{(ObsIDX_5)} + ZObs_{5r})^2$ = r^{th} simulated value for observed unweighted WHPT NTAXA
$ObsIDX_{6r} = ObsIDX_6 + ZObs_{6r}$ = r^{th} simulated value for observed unweighted WHPT ASPT
$ObsIDX_{4r} = ObsIDX_{5r} * ObsIDX_{6r}$ = r^{th} simulated value for observed unweighted WHPT Score

For the abundance-weighted WHPT indices:

$ObsIDX_{8r} = (\sqrt{(ObsIDX_8)} + ZObs_{8r})^2$ = r^{th} simulated value for observed weighted WHPT NTAXA
$ObsIDX_{9r} = ObsIDX_9 + ZObs_{9r}$ = r^{th} simulated value for observed weighted WHPT ASPT
$ObsIDX_{7r} = ObsIDX_{8r} * ObsIDX_{9r}$ = r^{th} simulated value for observed weighted WHPT Score

(c) LIFE index (index id 10):

For the LIFE index, the estimates of the sampling component SD for a site depend on:

$NTAXA_{Av}$ = the average number of BMWP taxa present in the observed samples over the assessment period.
= $Obs_{Av2)}$ = $ObsIDX_2$ in the RICT software.

Based on variance component statistical analyses of existing datasets with suitable replication and time series of samples (see Section 3, sections 3.3.2 and 3.7 and Table 15), the sampling components SD for the LIFE index, obtained without allowing for NTAXA, were estimated to be:

SD Component (average)	$SDRep_{Av}$	$SDTSeas_{Av}$	$SDTYear_{Av}$
Estimate for NTAXA = 18	0.211	0.118	0.149

However, the component SD for a site were found to vary with the number of BMWP taxa present in the samples from that site and period.

In RICT, the various component SD for a particular site with an average BMWP NTAXA of $NTAXA_{Av}$ over the assessment period (1 or 3 years) are obtained by multiplying each of these average component SD by the same constant K_{LIFE}, where

$K_{LIFE} = 0.951^{(NTAXAdiff)}$ = 0.951 to the power $NTAXA_{diff}$, where $NTAXA_{diff} = NTAXA_{Av} - 18$

Thus for LIFE (index =10):

$SDRep_{10} = K_{LIFE} . SDRep_{AV}$
$SDTSeas_{10} = K_{LIFE} . SDTSeas_{AV}$
$SDTYear_{10} = K_{LIFE} . SDTYear_{AV}$

See Section 3 for further details and Table 15 for examples of K_{LIFE} for a range of NTAXA values.

$ObsIDX_{10} = \{O/E\}_{Av(10)} * ExpIDX_{10}$

For single-year runs:

$SDObs_{10} = \sqrt{(\ ((SDRep_{10})^2 + (SDTSeas_{10})^2) / N_{ObsSeas}\ )}$

For multi-year (3-year period) runs:

$SDObs_{10} = \sqrt{((((SDRep_{10})^2 + (SDTSeas_{10})^2) / N_{ObsSeas} + (SDTYear_{10})^2 * (1 – N_{ObsYear} / 3)) / N_{ObsYear})}$

Then for the rth uncertainty simulation for LIFE (index 10):

$ObsIDX_{10r} = ObsIDX_{10} + ZObs_{10r}$ = rth simulated value for average observed LIFE

(d) PSI index (index id 11):

For the PSI index, the estimates of the sampling component SD for a site depend on:

$NTAXA_{Av}$ = the average number of BMWP taxa present in the observed samples over the assessment period.
= $Obs_{Av(2)}$ = $ObsIDX_2$ in the RICT software.

Based on variance component statistical analyses of existing datasets with suitable replication and time series of samples, the sampling components SD for the PSI index, obtained without allowing for NTAXA, were found to be less for values near to the extremes of the index's potential range near zero and 100. The variance was made independent of index value by transforming PSI to the 'Arcsine' scale using the trigonometric asin function as:

$$\text{Arcsine(PSI)} = \text{asin}(\sqrt{(PSI/100)}) \qquad \text{expressed in angular radians (not degrees)}$$

On the Arcsine scale, the various component SD (not allowing for NTAXA) were estimated to be:

SD Component (average)	$SDRep_{Av}$	$SDTSeas_{Av}$	$SDTYear_{Av}$
Estimate for NTAXA = 18	0.0596	0.0649	0.0288

However, the component SD for a site were found to vary with the number of BMWP taxa present in the samples from that site and period.

In RICT, the various component SD for any particular site with an average BMWP NTAXA of $NTAXA_{Av}$ over the assessment period (1 or 3 years) are obtained by multiplying each of these average component SD by the same constant K_{PSI}, where

$$K_{PSI} = 0.955^{(NTAXAdiff)} \quad = 0.955 \text{ to the power } NTAXA_{diff}\ , \text{ where } NTAXA_{diff} = NTAXA_{Av} - 18$$

Thus for PSI (index 11):

$$SDRep_{11} = K_{LIFE} \,.\, SDRep_{AV}$$
$$SDTSeas_{11} = K_{LIFE} \,.\, SDTSeas_{AV}$$
$$SDTYear_{11} = K_{LIFE} \,.\, SDTYear_{AV}$$

See Section 3, sections 3.3.3 and 3.7 for further details and Table 15 for examples of K_{LIFE} for a range of NTAXA values.

$ObsIDX_{11} = \{O/E\}_{Av(11)} * ExpIDX_{11}$

For single-year runs:

$$SDObs_{11} = \sqrt{(\ ((SDRep_{11})^2 + (SDTSeas_{11})^2\ ) / N_{ObsSeas}\ )}$$

For multi-year (3-year) runs:

$$SDObs_{11} = \sqrt{((\ ((SDRep_{11})^2 + (SDTSeas_{11})^2\ ) / N_{ObsSeas} + (SDTYear_{11})^2 * (1 - N_{ObsYear} / 3)) / N_{ObsYear})}$$

Then for the r^{th} uncertainty simulation for PSI (index 11):

$ObsIDX_{11r}$ = $100\ (\sin(\text{asin}(\sqrt{(ObsIDX_{11} / 100\ )}) + ZObs_{11r})^2$
= r^{th} simulated value for average observed PSI
where 'sin' is standard trigonometric sine function (the reverse of asin)

As a check for the RICT software programmers in selecting the correct asine and sine functions, they should find that:

asin(0.5) =0.5236 , asin(1.0) = 1.5708
sin(0.5) = 0.4794 , sin(1.0) = 0.8415, sin(1.5708) =1.0

6.3.3 Algorithms for correcting for sample processing biases in index values

6.3.3.1 Algorithms for simulating bias corrections in the BMWP indices

The quantitative effects of sample processing errors can only be assessed by a detailed analysis of the data provided by an external audit and re-examination of the sorting and taxonomic identification accuracy for macroinvertebrate samples. Prior to this study and report, this has currently only been done for the existing BMWP indices (BMWP Score, NTAXA and ASPT). Below, we detail the precise algorithms used to simulate the biases for these three BMWP indices (1-3).

Definitions :

$ObsIDX_{ir}$ = Simulation *r* Observed sample value of index *i* for current test site (uncorrected for bias)

$Ubias_2$ = Estimate of average net under-estimation of NTAXA (index 2) for selected 'season' sample

$Ubias_{ir}$ = Estimate of bias (net under-estimation) of index *i* for simulation *r*

$ObsIDX_{irB}$ = Bias-corrected observed value of index *i* for simulation r

$Bias_{2s}$ = User-supplied estimates of the average under-estimation of NTAXA (index 2) due to sample sorting and identification errors for single season samples taken in season *s* (1 = spring, 2 = summer, 3 = autumn)

$Kseas_s$ = 1 if the overall sample of interest involves season *s*
= 0 otherwise

$Ubias = (Bias_{21} * Kseas_1) + (Bias_{22} * Kseas_2) + (Bias_{23} * Kseas_3)$

For single season samples : $Ubias_2 = Ubias$
For two season combined samples : $Ubias_2 = 0.51 * Ubias$
For three season combined samples : $Ubias_2 = 0.37 * Ubias$

Special case : when no BMWP taxa were recorded in the sample (i.e. $ObsIDX_2 = 0$), assume none were missed (i.e, set $Ubias_2 = 0$)

$Ubias_{2r}$ = bias (net under-estimate of number of BMWP taxa) for simulated sample r, estimated as a random deviate from a Poisson distribution with a mean of $Ubias_2$

Note: RICT software already has existing code to generate such random numbers from a Poisson distribution)

$Zbias_{3r}$ = Random number deviate from a standard Normal distribution with a mean of 0.0 and SD of 1.0

$Ubias_{3r}$ = ASPT of the $Ubias_{2r}$ missed taxa for simulated sample r

$$= u_{3a} + u_{3b} * ObsIDX_2 + Zbias_{3r} * (u_{3c} / \sqrt{Ubias_{2r}})$$

where $u_{3a} = 4.29$, $u_{3b} = 0.077$, $u_{3c} = 2.0$

Note: It is unlikely, but mathematically feasible with this algorithm to derive values of $Ubias_{3r} > 10$ (maximum real value), such values of $Ubias_{3r}$ should be reset to 10; similarly, any values of $Ubias_{3r} < 1$ should be reset to 1).

$Ubias_{1r} = Ubias_{2r} * Ubias_{3r}$ = under-estimate of BMWP score for simulated sample r

$ObsIDX_{1rB} = ObsIDX_{1r} + Ubias_{1r}$ = bias-corrected observed BMWP Score for simulation r

$ObsIDX_{2rB} = ObsIDX_{2r} + Ubias_{2r}$ = bias-corrected observed NTAXA for simulation r

$ObsIDX_{3rB} = ObsIDX_{1rB} / ObsIDX_{2rB}$ = bias-corrected observed ASPT for simulation r

6.3.3.2 Algorithms for simulating bias-corrections for the unweighted WHPT indices (4-6)

It is likely that the bias-correction algorithms for the (unweighted) BMWP indces (indices 1-3) will also be broadly appropriate for their unweighted counterpart WHPT (indices 4-6), but these analyses have yet to be completed.

Not under current development from new audit data analyses as the unweighted form of the WHPT indices are not expected to be used in the new RICT

If required, then for simplicity, it is assumed that the bias for the unweighted WHPT NTAXA (index 5) is the same as the bias for the number of BMWP taxa (index 2)

Ubias_{6r} = Unweighted WHPT ASPT of the Ubias_{2r} missed taxa for simulated sample r
= $u_{6a} + u_{6b} * \text{ObsIDX}_2 + \text{Zbias}_{6r} * (u_{6c} / \sqrt{\text{Ubias}_{2r}})$

where estimates of u_{6a} , u_{6b} and u_{6c} are currently unavailable, but it may be adequate to use the equivalent values to those derived for the original ASPT, namely:

$u_{6a} = 4.29$, $u_{6b} = 0.077$, $u_{6c} = 2.0$

Then for simulation *r* :

$\text{Ubias}_{4r} = \text{Ubias}_{5r} * \text{Ubias}_{6r}$ = under-estimate of BMWP score for simulated sample r

$\text{ObsIDX}_{4rB} = \text{ObsIDX}_{4r} + \text{Ubias}_{4r}$ = bias-corrected observed unweighted WHPT Score (index 4)
$\text{ObsIDX}_{5rB} = \text{ObsIDX}_{5r} + \text{Ubias}_{5r}$ = bias-corrected observed unweighted WHPT NTAXA (index 5)
$\text{ObsIDX}_{6rB} = \text{ObsIDX}_{4rB} / \text{ObsIDX}_{5rB}$ = bias-corrected observed unweighted WHPT ASPT (index 6)

Note: It is unlikely, but mathematically feasible with this algorithm to derive values of Ubias_{6r} (the unweighted WHPT ASPT of missed taxa) which are outside the mathematically possible bounds of this index (i.e. -0.8 to 12.7, see Appendix 2). Values less than the min possible should be reset to min and values >max possible should be reset to max.

6.3.3.3 Algorithms for simulating bias-corrections for the abundance-weighted WHPT indices (7-9)

Definitions :

$ObsIDX_{ir}$ = Simulation *r* Observed sample value of index *i* for current test site (uncorrected for bias)

$Ubias_8$ = estimate of average net under-estimation of WHPT NTAXA for the observed sample

$Ubias_8$ is either:
- (i) input by the user of the RICT software
- (ii) estimated as 36% higher than the user-input bias ($Ubias_2$) for number of BMWP taxa i.e. $Ubias_8 = 1.36\ Ubias_2$

$Ubias_{ir}$ = Estimate of bias (net under-estimation) of index *i* for simulation *r*

$ObsIDX_{irB}$ = Bias-corrected observed value of index *i* for simulation r

Special case : when no WHPT taxa were recorded in the sample (i.e. $ObsIDX_8 = 0$), assume none were missed (i.e, set $Ubias_8 = 0$)

$Ubias_{8r}$ = bias (net under-estimate of number of WHPT taxa) for simulated sample r, estimated as a random deviate from a Poisson distribution with a mean of $Ubias_8$

$Zbias_{9r}$ = Random number deviate from a standard Normal distribution with a mean of 0.0 and SD of 1.0

$Ubias_{9r}$ = abundance-weighted WHPT ASPT of the $Ubias_{8r}$ missed WHPT taxa for simulated sample r
= $u_{9a} + u_{9b} * ObsIDX_9 + Zbias_{9r} * (u_{9c} / \sqrt{Ubias_{8r}})$

where $u_{9a} = 4.35$, $u_{3b} = 0.271$, $u_{9c} = 2.5$

Then:

$Ubias_{7r} = Ubias_{8r} * Ubias_{9r}$ = bias of abundance-weighted WHPT score for simulated sample r

$ObsIDX_{7rB} = ObsIDX_{7r} + Ubias_{7r}$ = bias-corrected observed abundance-weighted WHPT Score for simulation r

$ObsIDX_{8rB} = ObsIDX_{8r} + Ubias_{8r}$ = bias-corrected observed abundance-weighted WHPT NTAXA for simulation r

$ObsIDX_{9rB} = ObsIDX_{7rB} / ObsIDX_{8rB}$ = bias-corrected observed abundance-weighted WHPT ASPT for simulation r

6.3.3.4 Algorithms for simulating bias-corrections for LIFE (index 10)

Currently under development from new audit data analyses – to be be reported later in 2014 in separate on-going contract on assessing the effect of LIFE and PSI in relation to fine sediment and flow-related pressures

6.3.3.5 Algorithms for simulating bias-corrections for PSI (index 11)

Currently under development from new audit data analyses – to be be reported later in 2014 in separate on-going contract on assessing the effect of LIFE and PSI in relation to fine sediment and flow-related pressures

6.4 Algorithms for simulating uncertainty in expected values

The estimates of the site- and season-specific expected values of the indices are also assumed to have uncertainty associated with them. It is very difficult, or almost impossible, to measure the unknown (or unknowable) true errors in estimating the macroinvertebrate fauna expected at a site in the absence of any environmental stress because the expected depends on which, and how, environmental variables are measured and used to set the "target" expected fauna.

In RIVPACS III+, only the uncertainty in the original BMWP indices due to variation and errors in different people measuring and deriving the values of the environmental predictor variables for a site were assessed and incorporated into the uncertainty assessments. This was done within the same BAMS replicated sampling study. At each BAMS site, four people measured each of the following variables completely independently:

From maps	National Grid Ref., distance from source, altitude, slope, discharge category
In the field	Stream width, stream depth, mean substratum composition (each measured in spring, summer and autumn and then averaged for use as RIVPACS predictor variables)

Then each person's values for the environmental variables were run through RIVPACS III to derive four independent RIVPACS estimates of the expected fauna and expected index values for each site. The typical SD in these estimates of expected index values for a site were then included in the RIVPACS III+ uncertainty assessments.

The same approach and estimates were included in the original RICT software for BMWP indices 1-3.

Equivalent SD parameters and simulation of errors in expected values should be included in the RICT software for the unweighted and abundance-weighted WHPT indices. As no independent estimates are currently available and the WHPT indices are on broadly similar scales to their BMWP counterparts, it is reasonable to use the same estimates of the consequences of measurement errors in the RIVPACS enviromental predictor variables on the uncertainty of estimates of the expected values for the WHPT indices. This is the current recommendation, as detailed below.

Definitions :

$ExpIDX_i$ = Expected value of index *i* (for selected season *s*) for the current test site
(this is the average site-specific E value in multi-year assessments)

$ExpIDX_{ir}$ = Expected value of index *i* in simulation *r* for the current test site

$SDExp_i$ = Error SD for expected value of index *i*

$NExp_{year}$ = 1 for single-year run
= number of years (1, 2 or 3) for which a separate estimate of the E value was involved in the estimate of average E value (for multi-year run)

Algorithms for expected (E) values :

$ZExp_{ir}$ = Random number deviate from a standard Normal distribution with a mean of 0.0 and SD of 1.0, for use in simulation *r* for index *i*

$ExpIDX_{ir}$ = $ExpIDX_i + ZExp_{ir} * SDExp_i / \sqrt{(NExp_{year})}$

Note: RICT software programmers can either use an existing function code to generate such random numbers from a standard normal distribution or we can provide the FORTRAN code used in RIVPACS III+)

Parameter Estimates :

$SDExp_1$ = 4.3 = Measurement error SD of Expected values of BMWP Score (index 1)

$SDExp_2$ = 0.53 = Measurement error SD of Expected values of BMWP NTAXA (index 2)

$SDExp_3$ = 0.081 = Measurement error SD of Expected values of BMWP ASPT (index 3)

$SDExp_4$ = 4.3 = Measurement error SD of Expected values of unweighted WHPT Score (index 4)

$SDExp_5$ = 0.53 = Measurement error SD of Expected values of unweighted WHPT NTAXA (index 5)

$SDExp_6$ = 0.081 = Measurement error SD of Expected values of unweighted WHPT ASPT (index 6)

$SDExp_7$ = 4.3 = Measurement error SD of Expected values of weighted WHPT Score (index 7)

$SDExp_8$ = 0.53 = Measurement error SD of Expected values of weighted WHPT NTAXA (index 8)

$SDExp_9$ = 0.081 = Measurement error SD of Expected values of weighted WHPT ASPT (index 9)

6.5 Algorithms for simulating O/E ratios (EQI), confidence limits and confidence of class

6.5.1 Simulating O/E ratios (EQI)

For each simulated sample r, the simulated observed ($ObsIDX_{ir}$) and expected ($ExpIDX_{ir}$) values of each index are used to derive the O/E ratio (EQI_{ir}) for any index *i*, ignoring or uncorrected for any biases due to sample sorting and identification errors. These are known as the "face value" O/E ratios, in the sense that these would be the quoted values in the absence of any knowledge of sample processing errors.

Similarly, the O/E ratio (EQI_{irB}) for simulation *r* of index *i*, corrected for bias, is based on the simulation *r* of the observed value of index *i*, corrected for bias, namely $ObsIDX_{irB}$, and $ExpIDX_{ir}$.

Definitions:

EQI_{ir} = EQI value for index *i* in simulation *r* (uncorrected for bias) = 'face value' EQI

EQI_{irB} = EQI value for index *i* in simulation *r* (corrected for bias)

Algorithms:

$EQI_{ir} = ObsIDX_{ir} / ExpIDX_{ir}$

$EQI_{irB} = ObsIDX_{irB} / ExpIDX_{ir}$

6.5.2 Uncertainty SD and Confidence limits for an EQI

The frequency distribution of all the *rN* simulated EQI_{ir} values for a particular index *i* represents the degree of uncertainty in the true EQI value for that index for the site at that time period (either uncorrected or corrected for bias). This uncertainty can be summarised by the standard deviation (SD) of the simulated values. Additionally, 95% confidence limits for the true value are estimated as the lower and upper 2.5 percentiles of this frequency distribution (ie 2.5% of simulated O/E values were less than the lower limit and 2.5% were higher than the upper limit).

When corrected for bias, the confidence limits for EQI will tend to be wider because of the extra degree of uncertainty introduced by estimating the bias for the sample. However, bias-corrected EQI values should, by definition, be correct, on average, and no longer have a tendency to under- (or over-) estimate the true quality at the site during that time period.

Definitions:

$SDEQI_i$ = SD of the *rN* simulated values of EQI_i (uncorrected for bias) for index i for the current test site

$SDEQI_{iB}$ = SD of the *rN* simulated values of EQI_i (corrected for bias) for index i for the current test site

$L95EQI_i$ = Lower 2.5 percentile of the *rN* simulated values (EQI_{ir}) of EQI (uncorrected for bias) for index *i* for the current test site

$U95EQI_i$ = Upper 2.5 percentile of the *rN* simulated values (EQI_{ir}) of EQI (uncorrected for bias) for index *i* for the current test site

$L95EQI_{iB}$ = Lower 2.5 percentile of the *rN* simulated values (EQI_{irB}) of EQI (corrected for bias) for index *i* for the current test site

$U95EQI_{iB}$ = Upper 2.5 percentile of the *rN* simulated values (EQI_{irB}) of EQI (corrected for bias) for index *i* for the current test site

Algorithms:

Separately for simulated values (a) uncorrected for bias and (b) corrected for bias:

Calculate the SD, ($SDEQI_i$ and $SDEQI_{iB}$) of the *rN* values of EQI_{ir} and EQI_{irB} respectively, in the usual way for calculating any SD.

Determine the lower and upper 2.5 percentiles and thus 95% confidence limits by sorting all of the *rN* simulated EQI values into order from smallest to largest. Then the lower and upper percentiles are given by the mL and mU smallest values, where:

mL = nearest integer to 0.025 * (*rN* + 1)
mU = nearest integer to 0.975 * (*rN* + 1)

For the recommended *rN* = 9999, mL = 250 and mU = 9750.

The average ($AvEQI_{iB}$) of the bias-corrected EQR_{irB} values can be used in the output as the best bias-corrected estimate of EQI_i for the sample/site.

6.5.3 Index class limits and confidence of class

Assignment to ecological status class and confidence of class

The WFD requires that all water bodies, including rivers sites, are classified into one of five ecological status class. For the purposes of software coding, it is recommended that they are coded as classes 1-5 with 1= 'high', 2='good', 3='moderate', 4='poor' and 5='bad'.

Individual index class limits and classifications

The ecological status class of a test site is based on the EQI values (or corresponding WFD EQR (Ecological Quality Ratio) values). The potential range of EQI values for any particular index is divided into classes corresponding to each status class. The user-supplied class limits are specified by providing the lower inclusive EQI (or EQR) value for each class. For example, lower class limits of 1.00 for 'high', 0.90 for 'good', 0.77 for 'moderate' and 0.65 for 'poor', means that all samples with EQI values for this index between 0.90 and up to, but not including, 1.00 would be classified as 'good'; while all samples with EQI values less than 0.65 would be classified as 'bad'.

Initially, it is suggested that the WFD status class limits for the adjusted O/E values (i.e. EQR) values for the WHPT indices are set to the same values as their BMWP counterparts. Specifically, WHPT NTAXA EQR class limits are set to the current RICT values for of BMWP NTAXA and the class limits for EQR WHPT ASPT are set to the current RICT values for BMWP ASPT.

Classifications based on multiple metrics (Multi-metrics)

The overall status class for sample/site is usually based on combining information from more than one index.
There are two main ways of combining information from two or indices:

(i) Combine the individual indices EQI/EQR values into a single multi-metric EQI/EQR. This can done as some form of averaging (perhaps weighted) of the individuals indices' EQI/EQR. (e.g. ICMi is a weighted average of the EQR values of six indices/metrics)

(ii) Determine the status class of the test site based on the EQI value of each index independently and then apply a pre-set rule for deriving the overall status class based on the classes for the individual metrics. (e.g. the current UK classification system rule (MINTA, which stands for "minimum of TAXA and ASPT") uses the worst of the two classes determined by EQI for NTAXA and EQI for ASPT for a test site – this is a simple example of what is known as the "worst case" rule.

The status class limits for the EQI/EQR values of individual indices and rules for multi-index classifications all need to be decided by the UK Environment Agencies. It is beyond the remit of this workstream of this project to set those limits and rules.

At present, we recommend that in the new RICT software, for site classifications based on the use of EQI/EQR values for abundance weighted WHPT NTAXA and WHPT ASPT, overall site status class should be based on the same MINTA type rule as used in the current RICT, namely the worst of the two classes based on the EQI/EQR values for these two abundance-based WHPT indices.

'Status classification method': For the purpose of defining the algorithms to assess uncertainty in such status class assignments (i.e confidence of class), we merely refer to the chosen methods of determining the status class of test sites, whether based on individual indices or a suite of indices combined in a simple or hierarchical manner, as the 'status classification method'.

Applying the 'status classification method' to the 'face value' EQI values derived from the observed sample gives the 'face value' class of the site based on each index and overall.

Applying the 'status classification method' to the average($AvEQI_{iB}$) of the simulated bias-corrected EQI values for the test site give the 'bias-corrected face value' class of the site based on each index and overall.

Confidence of class

The likelihood of the true status class (i.e. averaged across all possible samples) of a test site being each of the five possible WFD classes is estimated simply by applying the 'status classification method' to each simulation sample *r* in turn. Thus the class for simulation *r* is based on the EQI/EQR values for simulation *r*, namely the set of EQI_{ir}.

This can be done for EQI values uncorrected for bias and, if available, corrected for bias.

For each index and overall, the proportion of simulations assigned to a status class estimates the probability that the true (average) quality of the test site for that time period was of that ecological status class (based on its macroinvertebrates).

7. ACKNOWLEDGEMENTS

We would like to thank Dave Rendall (SEPA) as Project Manager, Dave Colvill (SEPA) and other members of the UK Agencies for their support during this project workstream.

We would like to thank the Environment Agency, the Scottish Environment Protection Agency the Northern Ireland Environment Agency, and Natural Resiurces Wales for permiting the used of several of their datasets in support of this research.

We would like to make a special thanks to James Pretty and Iwan Jones of Queen Mary University of London who provided access to historical RIVPACS macroinvertebrate sample external audit data and related advice.

8. REFERENCES

Armitage P.D. (2000) The potential of RIVPACS for predicting the effects of environmental change. In: Wright J.F., Sutcliffe D.W. & Furse M.T. (eds), *Assessing the Biological Quality of Freshwaters: RIVPACS and Similar Techniques*. Freshwater Biological Association, Ambleside. pg. 93-112.

Clarke R. T. (2000) Uncertainty in estimates of river quality based on RIVPACS. In: Wright J.F., Sutcliffe D.W. & Furse M.T. (eds), *Assessing the Biological Quality of Freshwaters: RIVPACS and Similar Techniques*. Freshwater Biological Association, Ambleside. pg. 39–54.

Clarke R.T., Cox R., Furse M.T., Wright J.F. & Moss D. (1997) RIVPACS III+ River Invertebrate Prediction and Classification System with Error Assessments. User Manual. Environment Agency R&D Technical Report E26.

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. & Hering D. (2006) Errors and uncertainty in bioassessment methods – major results and conclusions from the STAR project and their application using STARBUGS. *Hydrobiologia* 566, 433-440.

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.

Council of the European Communities (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, Luxembourg.

Davy-Bowker J., Clarke R., Corbin T., Vincent H., Pretty J., Hawczak A., Blackburn J., Murphy J. & Jones I. (2008). River Invertebrate Classification Tool. A report to the Scotland and Northern Ireland Forum for Environmental Research. [SNIFFER project WFD72C].

Extence C.A., Balbi D.M. & Chadd R.P. (1999) River flow indexing using British benthic macroinvertebrates: a framework for setting hydroecological objectives. *Regulated Rivers: Research & Management* 15: 543-574.

Extence C.A., Chadd R.P., England A.J., Dunbar M.J., Wood C.P.J. & Taylor E.D. (2011) The assessment of fine sediment accumulation in rivers using macro-invertebrate community response. *River Research and Applications.* DOI: 10.1002/rra.

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. R&D Note 412, Environment Agency, Bristol. 139 pp.

Haase P., Murray-Bligh J., Lohse S., Pauls S., Sundermann A., Gunn R. & Clarke R. (2006) Assessing the impact of errors in sorting and identifying macroinvertebrate samples. *Hydrobiologia* 566: 505-521.

Snedecor G. W. & Cochran W. G. (1980) *Statistical Methods*. Ames: Iowa State University Press.

9. APPENDICES

Appendix 1: LIFE index (Lotic Invertebrate Flow Evaluation)

Reference: Extence C.A., Balbi D.M. & Chadd R.P. (1999) River flow indexing using British benthic macroinvertebrates: a framework for setting hydroecological objectives. Regulated Rivers: Research & Management, 15: 543-574.

LIFE = mean of fs scores weighted by Flow Group and $\log_{10}$ abundance)
= (sum of abundance-based flow group scores (fs) of LIFE-scoring taxa present)
divided by (the number of LIFE-scoring taxa present)

Taxonomic level : TL2 – distinct families , TL1/2 – composite families

Family	Flow Group
Planariidae	***4***
Dugesiidae	***4***
Dendrocoelidae	4
Neritidae	2
Viviparidae	3
Valvatidae	4
Hydrobiidae	***4***
Bithyniidae	*4*
Physidae	4
Lymnaeidae	4
Planorbidae	4
Ancylidae	***2***
Acroloxidae	*4*
Margaritiferidae	2
Unionidae	4
Sphaeriidae	4
Dreissenidae	4
Piscicolidae	2
Glossiphoniidae	4
Hirudinidae	4
Erpobdellidae	4
Agelinidae	5
Chirocephalidae	6
Triopsidae	6
Astacidae	2
Mysidae	5
Asellidae	4
Corophiidae	3
Talitridae	6
Gammaridae	***2***
Crangonyctidae	*4*
Siphlonuridae	4
Baetidae	2
Heptageniidae	1

Family	Flow Group
Leptophlebiidae	2
Potamanthidae	3
Ephemeridae	2
Ephemerellidae	2
Caenidae	4
Taeniopterygidae	2
Nemouridae	4
Leuctridae	2
Capniidae	1
Perlodidae	1
Perlidae	1
Chloroperlidae	1
Platycnemididae	4
Coenagrionidae	4
Lestidae	4
Calopterygidae	3
Gomphidae	2
Cordulegastridae	2
Aeshnidae	4
Corduliidae	4
Libellulidae	4
Mesoveliidae	5
Hebridae	4
Hydrometridae	4
Veliidae	4
Gerridae	4
Nepidae	5
Naucoridae	4
Aphelocheiridae	2
Notonectidae	4
Pleidae	4
Corixidae	4
Haliplidae	4
Hygrobiidae	5

Family	Flow Group
Dytiscidae	***4***
Noteridae	*4*
Gyrinidae	4
Hydrophilidae	***4***
Hydraenidae	*4*
Scirtidae	4
Elmidae	2
Sialidae	4
Osmylidae	2
Sisyridae	4
Rhyacophilidae	***1***
Glossosomatidae	*2*
Hydroptilidae	4
Philopotamidae	1
Psychomyiidae	***2***
Ecnomidae	*3*
Polycentropodidae	4
Hydropsychidae	2
Phryganeidae	4
Brachycentridae	2
Lepidostomatidae	2
Limnephilidae	4
Goeridae	1
Beraeidae	2
Sericostomatidae	2
Odontoceridae	1
Molannidae	4
Leptoceridae	4
Tipulidae	4
Ptychopteridae	2
Chaoboridae	5
Culicidae	5
Simuliidae	2
Syrphidae	5

* BMWP composites italicised. Where BMWP composite families were used, the first family was used (emboldened) and the other member of the composite was ignored – as recommended by Extence *et al* (1999).

Log_{10} abundance category	1 (A)	2 (B)	3 (C)	4 (D)	5 (E)
Numerical abundance	1-9	10-99	100-999	1000-9999	10000+

LIFE scores (fs) for taxa in each Flow Group (1-6) in each $\log_{10}$ abundance category (A-E)

Flow Group	**Flow Group Description**	**Log_{10} Abundance Category**				
		A	**B**	**C**	**D**	**E**
1	Rapid	9	10	11	12	12
2	Moderate/fast	8	9	10	11	11
3	Slow/sluggish	7	7	7	7	7
4	Flowing/standing	6	5	4	3	3
5	Standing	5	4	3	2	2
6	Drought resistant	4	3	2	1	1

Appendix 2: WHPT index (Walley, Hawkes, Paisley, Trigg)

(Non-Abundance Weighted & Abundance Weighted)

Reference: No definitive reference exists. This index was supplied by email to John Davy-Bowker from John Murray-Bligh, 4th July 2007.

Indices:
Non-abundance weighted WHPT Score (sum of PO scores of families present)
Non-abundance weighted WHPT NTAXA (number of WHPT-scoring families present)
Non-abundance weighted WHPT ASPT (WHPT Score /NTAXA)

Abundance weighted WHPT Score (sum of abundance (AB_{1-4+}) scores of families present)
Abundance weighted WHPT NTAXA (number of WHPT-scoring families present)
Abundance weighted WHPT ASPT (WHPT Score /NTAXA)

Note: WHPT NTAXA is not dependent on the taxa abundances

PO = Presence only
Abundance categories: AB1 = 1-9, AB2 = 10-99, AB3 = 100-999, AB4+ = 1000+ individuals in sample

Individual family WHPT scores		**Log_{10} Abundance category**			
Family	**PO**	**AB1**	**AB2**	**AB3**	**AB4+**
Planariidae	4.90	4.70	5.40	5.40	5.4
Dugesiidae	2.90	2.80	3.10	3.10	3.1
Dendrocoelidae	3.00	3.00	2.60	2.60	2.6
Neritidae	6.40	6.40	6.50	6.90	6.9
Viviparidae	5.70	5.20	6.70	6.70	6.7
Valvatidae	3.20	3.30	3.10	2.70	2.7
Hydrobiidae	4.20	4.10	4.20	4.60	3.7
Bithyniidae	3.70	3.60	3.80	3.30	3.3
Physidae	2.40	2.70	2.00	0.40	0.4
Lymnaeidae	3.30	3.60	2.50	1.20	1.2
Planorbidae	3.10	3.20	3.00	2.40	2.4
Ancylidae	5.70	5.80	5.50	5.50	5.5
Acroloxidae	3.60	3.60	3.80	3.80	3.8
Unionidae	5.30	5.20	6.80	6.80	6.8
Sphaeriidae_Pea_mussels	3.90	4.40	3.50	3.40	2.3
Dreissenidae	3.70	3.70	3.70	3.70	3.7
Oligochaeta	2.70	3.60	2.30	1.40	-0.6
Piscicolidae	5.20	5.20	4.90	4.90	4.9
Glossiphoniidae	3.20	3.40	2.50	0.80	0.8
Hirudinidae	-0.80	-0.80	-0.80	-0.80	-0.8
Erpobdellidae	3.10	3.60	2.00	-0.80	-0.8
Astacidae	7.90	7.90	7.90	7.90	7.9
Asellidae	2.80	4.00	2.30	0.80	-1.6
Corophiidae	5.80	5.70	5.80	5.80	5.8
Crangonyctidae	3.90	3.80	4.00	3.60	3.6
Gammaridae	4.40	4.20	4.50	4.60	3.9
Niphargidae	6.30	6.30	6.30	6.30	6.3
Siphlonuridae	11.50	11.30	12.20	12.20	12.2

Individual family WHPT scores		**Log$_{10}$ Abundance category**			
Family	**PO**	**AB1**	**AB2**	**AB3**	**AB4+**
Baetidae	5.50	3.60	5.90	7.20	7.5
Heptageniidae	9.70	8.50	10.30	11.10	11.1
Leptophlebiidae	8.80	8.80	9.10	9.20	9.2
Potamanthidae	10.00	9.80	10.40	10.40	10.4
Ephemeridae	8.40	8.30	8.80	9.40	9.4
Ephemerellidae	8.20	7.90	8.50	9.00	9
Caenidae	6.50	6.50	6.50	6.50	6.5
Taeniopterygidae	11.30	11.00	11.90	12.10	12.1
Nemouridae	9.30	8.70	10.70	10.70	10.7
Leuctridae	10.00	9.30	10.60	10.60	10.6
Capniidae	9.60	9.70	9.40	9.40	9.4
Perlodidae	10.80	10.50	11.50	11.50	11.5
Perlidae	12.70	12.60	13.00	13.00	13.0
Chloroperlidae	11.60	11.40	12.20	12.20	12.2
Platycnemididae	6.00	6.00	6.00	6.00	6
Coenagriidae	3.50	3.40	3.80	3.80	3.8
Calopterygidae	6.00	5.90	6.20	6.20	6.2
Cordulegasteridae	9.80	9.80	9.80	9.80	9.8
Aeshnidae	4.70	4.70	4.70	4.70	4.7
Libellulidae	4.10	4.10	4.10	4.10	4.1
Mesoveliidae	4.70	4.70	4.70	4.70	4.7
Hydrometridae	4.30	4.30	4.30	4.30	4.3
Veliidae	4.50	4.50	3.90	3.90	3.9
Gerridae	5.20	5.20	5.50	5.50	5.5
Nepidae	2.90	2.90	2.90	2.90	2.9
Naucoridae	3.70	3.70	3.70	3.70	3.7
Aphelocheiridae	8.50	8.60	8.50	8.00	8
Notonectidae	3.40	3.40	3.90	3.90	3.9
Pleidae	3.30	3.30	3.30	3.30	3.3
Corixidae	3.80	3.70	3.90	3.70	3.7
Haliplidae	3.60	3.60	3.40	3.40	3.4
Hygrobiidae	3.80	3.80	3.80	3.80	3.8
Noteridae	3.20	3.20	3.20	3.20	3.2
Dytiscidae	4.50	4.50	4.80	4.80	4.8
Gyrinidae	8.20	8.10	9.00	9.00	9
Hydrophilidae	6.20	5.80	8.80	8.80	8.8
Hydraenidae	8.90	8.50	10.50	10.50	10.5
Scirtidae	6.90	6.90	6.80	6.80	6.8
Dryopidae	6.00	6.00	6.00	6.00	6
Elmidae	6.60	5.30	7.40	8.30	8.3
Sialidae	4.30	4.20	4.40	4.40	4.4
Sisyridae	5.70	5.70	5.70	5.70	5.7
Rhyacophilidae	8.40	8.10	9.20	8.30	8.3
Glossosomatidae	7.70	7.80	7.60	7.20	7.2
Hydroptilidae	6.20	6.10	6.50	6.80	6.8
Philopotamidae	11.20	11.20	11.10	11.10	11.1

Individual family WHPT scores		**Log$_{10}$ Abundance category**			
Family	**PO**	**AB1**	**AB2**	**AB3**	**AB4+**
Psychomyiidae	5.80	5.80	5.70	5.70	5.7
Polycentropodidae	8.10	8.20	8.10	8.10	8.1
Hydropsychidae	6.60	5.80	7.20	7.40	7.4
Phryganeidae	5.50	5.50	5.50	5.50	5.5
Brachycentridae	9.50	9.60	9.50	8.90	8.9
Lepidostomatidae	10.10	9.90	10.30	10.20	10.2
Limnephilidae	6.20	5.90	6.90	6.90	6.9
Goeridae	8.80	8.80	8.80	9.40	9.4
Beraeidae	8.70	8.80	7.30	7.30	7.3
Sericostomatidae	9.10	8.90	9.40	9.50	9.5
Odontoceridae	11.00	11.10	10.30	10.30	10.3
Molannidae	6.60	6.50	7.60	7.60	7.6
Leptoceridae	6.70	6.70	6.90	7.10	7.1
Tipulidae	5.90	5.40	6.90	6.90	7.1
Psychodidae	4.40	4.50	3.00	3.00	3
Ptychopteridae	6.40	6.40	6.40	6.40	6.4
Dixidae	7.00	7.00	7.00	7.00	7
Chaoboridae	3.00	3.00	3.00	3.00	3
Culicidae	2.00	2.00	1.90	1.90	1.9
Ceratopogonidae	5.50	5.40	5.50	5.50	5.5
Simuliidae	5.80	5.50	6.10	5.80	3.9
Chironomidae	1.10	1.20	1.30	-0.90	-0.9
Stratiomyidae	3.60	3.60	3.60	3.60	3.6
Rhagionidae	9.60	9.60	9.60	9.60	9.6
Tabanidae	7.10	7.10	7.30	7.30	7.3
Athericidae	9.30	9.30	9.50	9.50	9.5
Empididae	7.10	7.00	7.60	7.60	7.6
Dolichopodidae	4.90	4.90	4.90	4.90	4.9
Syrphidae	1.90	1.90	1.90	1.90	1.9
Sciomyzidae	3.40	3.40	3.40	3.40	3.4
Ephydridae	4.40	4.40	4.40	4.40	4.4
Muscidae	3.90	4.00	2.60	2.60	2.6
BMWP Composite taxa*					
Planariidae (incl. Dugesiidae)	5.0	4.8	5.4	5.3	5.3
Hydrobiidae (Incl. Bithyniidae)	4.2	4.1	4.2	4.5	3.7
Ancylidae (incl. Acroloxidae)	5.8	5.9	5.6	5.4	5.4
Gammaridae (incl. Crangonyctidae & Niphargidae)	4.5	4.3	4.7	4.7	3.9
Dytiscidae (incl. Noteridae)	4.7	4.7	5.0	5.0	5.0
Hydrophilidae (incl. Hydraenidae)	7.4	7.0	9.5	9.5	9.5
Rhyacophilidae (incl. Glossosomatidae)	8.2	7.9	8.8	7.5	7.5
Psychomyiidae (incl. Ecnomidae)	5.9	5.9	5.8	5.8	5.8

* BMWP composites italicised. Where BMWP composite families were used, the distinct families are ignored, as recommended by John Murray-Bligh, 4th July 2007, NB – Ecnomidae, as a distinct family do not score.

Appendix 3: PSI index (Proportion of Sediment-sensitive Invertebrates)

Reference:

The PSI (Proportion of Sediment-sensitive Invertebrates) index measures the abundance-weighted proportional frequency of taxa which are sensitive to fine sediment deposition (Extence *et al* 2011)

$$\text{PSI} = \frac{\text{Sum of Ss Scores for observed taxa in Sediment Sensitivity Groups A \& B}}{\text{Sum of Ss Scores for observed taxa in all Sediment Sensitivity Groups A-D}} \times 100$$

Sediment Sensitivty scores (Ss) for taxa in each Sensitivity Group (A-D) in each $\log_{10}$ abundance category (1 - 4+)

Sensitivty Group	**Sensitivity Group Description**	**Log_{10} Abundance Category (individuals)**			
		1 (1-9)	**2 (10-99)**	**3 (100-999)**	**4+ (1000+)**
A	Highly Sensitive	2	3	4	5
B	Moderately Sensitive	1	2	3	4
C	Moderately Insensitive	1	2	3	4
D	Highly Insensitive	2	3	4	5

Extence *et al* (2011) give a provisional interpretation of PSI scores as:

PSI range	**River bed Condition**
81-100	Minimally sedimented/unsedimented
61-80	Slightly sedimented
41-60	Moderately sedimented
21-40	Sedimented
0-20	Heavily sedimented

However, Extence *et al* (2011) acknowledge that the standardization of PSI scores is achievable by utilizing the UK reference condition model RIVPACS which can specify the unstressed invertebrate community expected at a site from the physical and chemical characteristics of that site. Importnatly, they note that, for RIVPACS predictive purposes, the overlying fine sediment is not (and should not be) used to characterize the composition of river bed substrata). The observed(O) PSI score of the sampled community can be directly compared to that expected (E), by deriving observed over expected (O/E) ratios as Environmental Quality Indices In the case of PSI, the lower the O/E ratio, the greater the sedimentation stress). Extence *et al* note that “this approach allows direct comparisons to be made spatially between sites on the same river and from different catchments/regions and also enables comparison between different types of fine sediment impacts (e.g., construction activities and bank erosion) or recovery (e.g., following natural spates or river restoration activities).”

PSI Sediment sensitivity group of families (taxonomic level TL3) : E denotes excluded taxa

Taxon	**PSI Group**
Spongillidae	B
Planariidae	D
Dugesiidae	D
Dendrocoelidae	C
Neritidae	C
Viviparidae	D
Valvatidae	C
Hydrobiidae	C
Bithyniidae	D
Physidae	D
Lymnaeidae	D
Acroloxidae	E
Planorbidae	D
Ancylidae	D
Margaritiferidae	A
Unionidae	D
Sphaeriidae	D
Dreissenidae	E
Lumbriculidae	D
Haplotaxidae	D
Naididae	D
Tubificidae	D
Enchytraeidae (incl. Propappidae)	D
Lumbricidae (incl. Glossoscolecidae)	D
Piscicolidae	B
Glossiphoniidae	C
Hirudinidae	D
Erpobdellidae	C
Astacidae	E
Asellidae	D
Corophiidae	D
Crangonyctidae	D
Gammaridae	B
Niphargidae	B
Siphlonuridae	C
Baetidae	A
Heptageniidae	A
Ameletidae	B
Leptophlebiidae	B
Potamanthidae	B
Ephemeridae	C
Ephemerellidae	A
Caenidae	D
Taeniopterygidae	A
Nemouridae	C
Leuctridae	A
Capniidae	A
Perlodidae	A
Perlidae	A
Chloroperlidae	A
Platycnemididae	D
Coenagriidae	E
Calopterygidae	C
Gomphidae	D
Cordulegasteridae	D
Aeshnidae	D
Libellulidae	D
Mesoveliidae	E
Hydrometridae	E
Veliidae	E
Gerridae	E
Nepidae	D
Naucoridae	E
Aphelocheiridae	A
Notonectidae	E
Corixidae	D
Haliplidae	D
Noteridae	D
Dytiscidae	D
Gyrinidae	E
Helophoridae	D
Hydrophilidae	D
Hydrochidae	D
Hydraenidae	B
Scirtidae	B
Dryopidae	D
Elmidae	B
Sialidae	D
Osmylidae	B
Sisyridae	B
Rhyacophilidae	A
Glossosomatidae	A
Hydroptilidae	E
Philopotamidae	A
Psychomyiidae	B
Ecnomidae	C
Polycentropodidae	B
Hydropsychidae	A
Phryganeidae	D
Brachycentridae	A
Lepidostomatidae	B
Limnephilidae	B
Goeridae	A
Beraeidae	A
Sericostomatidae	B
Odontoceridae	B
Molannidae	C
Leptoceridae	E
Apataniidae	A
Pyralidae	E
Tipulidae	B
Limoniidae	B
Pediciidae	B
Psychodidae	D
Ptychopteridae	D
Dixidae	B
Chaoboridae	E
Culicidae	E
Thaumaleidae	E
Ceratopogonidae	E
Simuliidae	A
Tanypodinae	E
Diamesinae	E
Prodiamesinae	E
Orthocladiinae	E
Chironomini	E
Tanytarsini	E
Stratiomyidae	C
Tabanidae	D
Athericidae	E
Syrphidae	D

www.fba.org.uk

The Freshwater Biological Association
The Ferry Landing
Far Sawrey, Ambleside
Cumbria, LA22 0LP, United Kingdom

The Freshwater Biological Association
River Laboratory
East Stoke, Wareham
Dorset, BH20 6BB, United Kingdom

Freshwater Biological Association