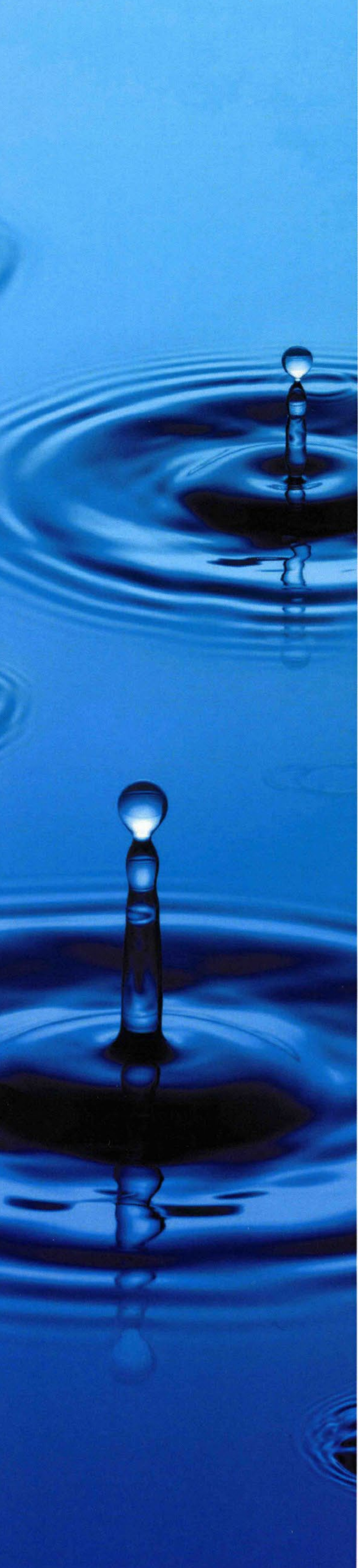


Queen Mary University of London

Freshwater Biological Association

River Invertebrate Classification Tool Science Development Project: Describing the impact of abstraction and fine sediment pressures on the biological communities in Scottish rivers

A Report to the
Scottish Government

The Scottish Government

J. Davy-Bowker
R.T. Clarke
J.I. Jones
J.F. Murphy

July 2014

Research Contractor

This document was produced by the Freshwater Biological Association:
John Davy-Bowker[†], Ralph T. Clarke[†], J. Iwan Jones* and John F. Murphy*

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

*School of Biological and Chemical Sciences, Queen Mary University of London, Mile End Road, London E1 4NS.

Project Funders

This project was funded by the Scottish Government.

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

Scottish Government Project Manager

Scottish Government's project manager for this contract was:
David Rendall (SEPA)

FBA Project Manager

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

FBA Project Code

S/0011/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 2
Impact of abstractions and fine sediment pressures on biological communities

Project funders: Scottish Government

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 of the first workstream of this project, Workstream 1, was 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 assessments. Workstream 1 was completed in March 2014 and was reported in Clarke and Davy-Bowker (2014).

The priority in this current project, Workstream 2, was to develop methods, algorithms and parameter estimates to enable the abundance-based LIFE and PSI indices to be incorporated into a further revision of RICT.

Objectives of research

- Develop algorithms and uncertainty parameter estimates for the incorporation of abundance-weighted classification indices LIFE and PSI (together with WHPT) into RICT.
- Develop algorithms and parameter estimates for the incorporation of sample biases of abundance-weighted indices LIFE and PSI, into RICT.
- Produce the basic statistical procedures needed to enable RICT classification for LIFE & PSI.

- Derive an initial set of statistically based WFD class boundaries for LIFE and PSI which take account of the range of pressures assessed by these metrics, namely water resource and fine sediment pressure on rivers.

Key findings and recommendations

Algorithms were derived to estimate river site ecological status based on macroinvertebrate indices involving the average of the single season sample observed (O) to expected (E) ratios of family-level LIFE ($LIFE_{fam}$) and/or family-level PSI (PSI_{fam}) indices. RIVPACS model predictions of both $LIFE_{fam}$ and PSI_{fam} were based on a new RIVPACS IV model which did not involve any of the time-varying environmental predictor variables measured at the time of macroinvertebrate sampling likely to be impacted by water resource and fine sediment pressures on rivers, namely stream width, depth and substratum composition.

Sampling uncertainty components in the abundance-weighted $LIFE_{fam}$ and PSI_{fam} 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 $LIFE_{fam}$ and PSI_{fam}, 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 provide algorithms to simulate the estimated sample processing biases in the abundance-weighted $LIFE_{fam}$ and PSI_{fam} indices from the observed index values and/or the audit-based estimates of the bias in BMWP NTAXA.

RIVPACS predictions of expected values and derived EQI (O/E ratios) were adjusted to be WFD-compliant Ecological Quality Ratios (EQRs) by first adjusting the raw Expected (E) values to adjusted Expected (E_{adj}). The adjusted EQI ($EQI_{adj} = O/E_{adj}$) were then calibrated to EQR by a calibration factor. Furthermore a new approach to combine the two steps by directly adjusting the raw E values to WFD reference condition expected values (E_{RC}) and then calculating WFD EQR values in one step as $EQR = O/E_{RC}$ was proposed.

The lower 5 percentile and lower 10 percentile of the distribution of EQI values for $LIFE_{fam}$ and PSI_{fam} amongst all individual single season samples and amongst the average of spring and autumn sample EQI_{adj} values for the reference sites from the 685 GB references sites was presented. These values could be used as trial values for the good-moderate WFD status class boundary for that index and can be compared to the proposed boundary values derived elsewhere in the project from datasets incorporating a range of sites across the stressor gradient of interest.

A detailed algorithms section is provided to enable the RICT software programmers to encode these new methods and uncertainty parameter estimates for the $LIFE_{fam}$ and PSI_{fam} indices into the next version of RICT for their use in river site status classifications.

Drawing upon existing Agency (SEPA, EA, NIEA and NRFA) datasets, and datasets held by our project team, we quantified the relationship between LIFE and PSI indices and variation in the level of the stress they were designed to diagnose.

Whilst a relationship between PSI and the gradient of pressure from fine sediment was apparent, both high and low EQI values were obtained for PSI at high levels of pressure from fine sediment. It would appear that both PSI_{fam} and mixed taxonomic level PSI (PSI_{sp}) EQIs are variable at high levels of pressure, with a wide range of EQI values being returned. This will ultimately make spatial interpretation of EQR values difficult.

Despite using large-scale and long-term data (434 sites from throughout GB, with multiple years' data at each site spanning the period 1994 to 2012) we have not been able to establish a clear relationship between EQI values for $LIFE_{fam}$ and low flow summary statistics for the antecedent period. Furthermore, we have not been able to consistently separate sites across a pressure gradient of impact by flow stress as defined by the UK Hydrometric Register. There was considerable overlap in EQI values for $LIFE_{fam}$ from natural and GB reference sites, with those where flow was significantly impacted by water abstraction, regulation (e.g. for hydroelectricity generation), and water oversupply. The relationship between EQI values for LIFE and the flow pressure gradient appears to be confounded by site-specific effects. Such site-specific relationships between EQI for LIFE and discharge will confound detection of sites suffering flow stress: we were unable to provide reliable class boundaries for LIFE based on the response to pressure on water resources.

These findings do not detract from the use of the $LIFE_{fam}$ and PSI_{fam} indices for detecting how individual sites or river systems respond to flow/sediment stress. Much existing work has demonstrated the utility of both LIFE and PSI in time series analyses (where the confounding problems of national-scale spatial classification do not apply). LIFE and PSI are used regularly in the setting of ecologically acceptable flow and sediment regimes and both have utility for incorporation into RICT in this capacity.

To guide selection of the classification boundaries for the LIFE and PSI indices we plotted the proportions of individuals classed as stressor-sensitive taxa and those classed as stressor-tolerant against the EQI for the index. Overlaying both relationships on the same graph it was possible to use the intersection point of their generalised additive modelled response as the proposed Good/Moderate boundary. Boundary values calculated using this approach were consistently more stringent (closer to 1) than those derived using the distribution of RIVPACS reference sample EQIs. For $LIFE_{fam}$ EQIs, a third approach to generating proposed classification boundaries was applied whereby the lower 5 and 10 percentile values of the distribution of EQI values for $LIFE_{fam}$ were calculated for the sites matched to gauging stations categorised by NRFA as having catchments experiencing natural flow. This third approach returned proposed $LIFE_{fam}$ EQIs boundaries intermediate between the two other approaches.

This project has produced an upgrade of the River Invertebrate Classification Tool (RICT) system to enable incorporation of the abundance-weighted indices $LIFE_{fam}$ and PSI_{fam} to provide Ecological Quality Ratio classifications. While $LIFE_{fam}$ and PSI_{fam} are both known to respond well to flow and sediment stress, the performance of the EQIs of the indices $LIFE_{fam}$ and PSI_{fam} against gradients of pressure suggest that at a national scale, in their current form, they are not able to accurately identify sites that are subject to water resource and sediment input pressures. Nor are they able to accurately assign reasons for waterbody downgrades. We conclude that while important work has been delivered to enable incorporation of index EQIs that assess flow and sediment stress, future refinement may be needed to achieve a more desirable national-scale EQI pressure gradient response.

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

Dedicated to the memory of the late Dr John F. Wright
who initiated and led the development of RIVPACS
for 25 years

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 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

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).

The aim of the original version of RICT software developed in 2008 was primarily to provide a modern version of the previous RIVPACS software and as such site assessments and uncertainty of WFD ecological status class were based on the original BMWP indices, BMWP NTAXA and BMWP ASPT.

In a new research project earlier this year, we (Clarke & Davy-Bowker 2014) developed the necessary information to include the alternative ability to base river site assessments on the newer abundance-weighted WHPT macroinvertebrate indices, WHPT NTAXA and WHPT ASPT.

The distribution of the observed (O) values and O/E values for the abundance-weighted WHPT NTAXA and WHPT ASPT indices for the GB references sites, grouped by end-group, are shown in Figure 3 and Figure 4. For illustration the observed values are for the autumn samples, but patterns of variation with RIVPACS end-group are similar for all seasons. The overall single season sample frequency distribution of the O/E (EQI) values for these indices across all GB reference sites and seasons is shown in Figure 5. These are adjusted EQI using the standard method adopted in RICT to adjust the raw O/E (EQI) values by adjusting the raw Expected values according the the Agencies biologists' assessment score (1-6) for the reference sites at the time of sampling for RIVPACS development (see section 5.2 of this report for more methodological and adjustment parameter details).

In addition the UK environment agenices have identified a need to update the RIVPACS IV models and the RICT software to address a major deficiency in these tools to take account of water resource pressures and fine sediments on rivers. This is needed both to more fully represent the range of pressures that should be taken into account for WFD classification, and also to help support decision making for several types of development activities on rivers (for example, hydroelectricity generation, water abstraction, and water oversupply). Recently developed indices to flow variation (LIFE; Extence *et al* 1999) and fine sediment stress (PSI; Extence *et al* 2013) are proposed as the diagnostic indices to be incorporated into RICT. The LIFE index (Lotic Index for Flow Evaluation) categorises macroinvertebrate taxa to one of six groups according to their association with different water velocities. Scores are assigned to each taxa present in a sample dependent on their log-abundance and water velocity association. The average of the assigned scores is the final LIFE index value. Higher water velocities should result in higher LIFE values. The PSI index (Proportion of Sediment-sensitive Invertebrates) is similar to LIFE in that it categorises macroinvertebrate taxa to one of four groups according to their reported tolerance of (or association with) fine sediment. The PSI value is calculated as the abundance-weighted proportion of sensitive taxa relative to all taxa (reported as a percentage). Higher PSI values indicate less fine sediment stress.

Furthermore, incorporation of the LIFE and PSI indices (alongside WHPT) into RICT will mean that WFD classification will be based on three family-level metrics. Between them, these three indices will report on general degradation (WHPT NTAXA), organic pollution (WHPT ASPT), flow stress (LIFE) and sedimentation (PSI). All three metrics will also incorporate abundance weighting, so that for the first time RICT will fully conform to the WFD requirement to assess not only the structure but also the abundance of freshwater communities.

The aim of this project is therefore to conduct the required underpinning science to allow RICT to better distinguish and report upon the extent of water resources and sedimentation pressures. Given the scarcity of public funds available for environmental improvements, this will assist the UK Environment Agencies in directing spend to when the maximum cost benefit can be achieved.

Being able to more accurately assess the extent of water resource stress, as well as the extent of sediment input problems, will also inform the future development of policy in this area. The work presented here will therefore contribute towards informing policy in areas such as hydroelectric power and the setting of ecologically acceptable flows, drinking and industrial water abstraction and its consequences for river flows, reservoir discharge and over supply, and farming and its contribution to diffuse sediment inputs. Decision making in all of these areas will be better informed by the addition of proven diagnostic indices to the RICT tool and the resultant improvement in understanding that this will provide.

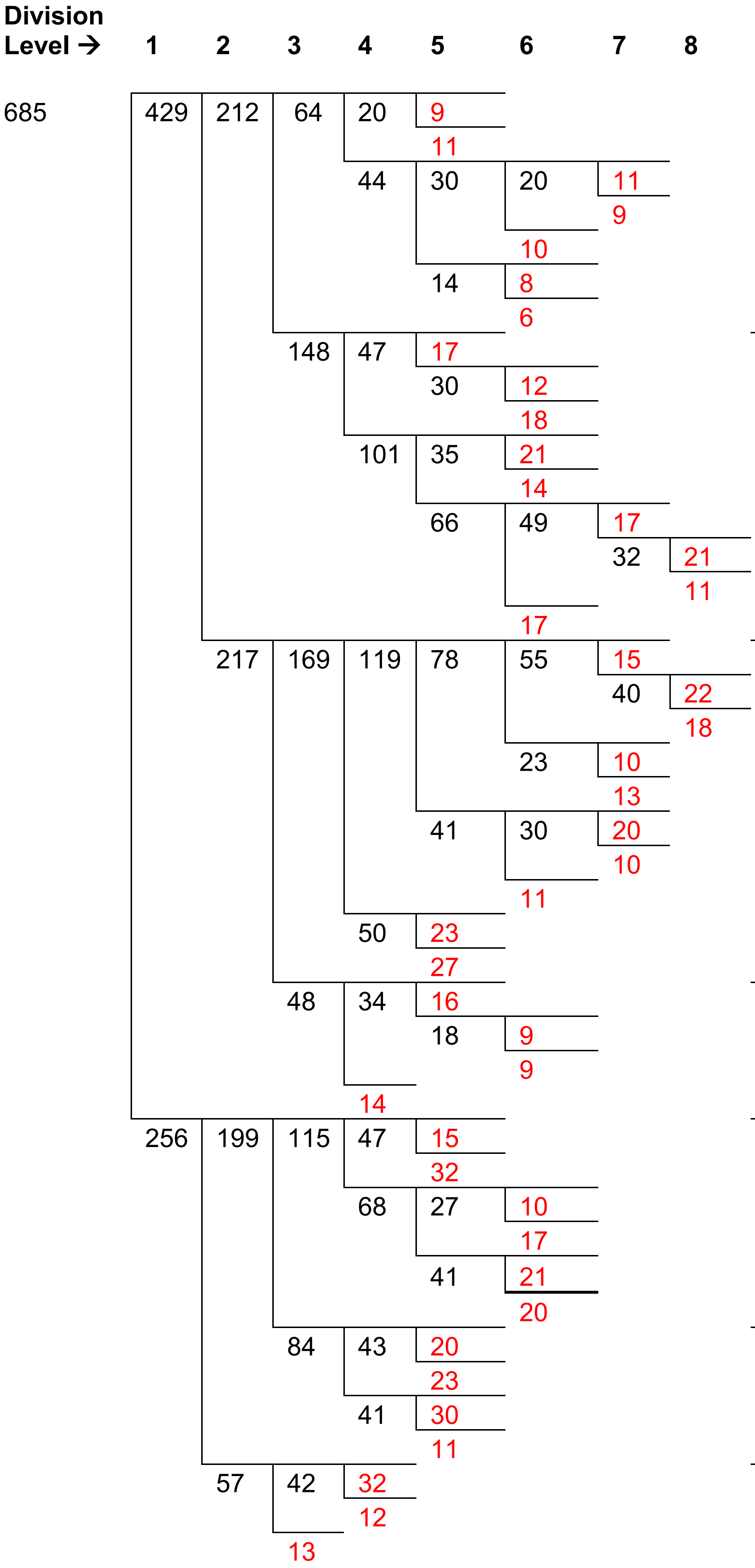
The present work will address four key objectives:

1. Devise algorithms to accurately predict spring and autumn taxa lists for the LIFE and PSI indices.
2. Develop an approach to express LIFE and PSI Ecological Quality Indices (EQIs) as Ecological Quality Ratios (EQR).
3. Derive an initial set of statistically based WFD class boundaries for LIFE and PSI. This should take account of the range of pressures assessed by these metrics, including elevated water levels.
4. Produce the basic statistical procedures needed to enable RICT classification for LIFE & PSI.

Chapters 2-6 of the report will detail the research undertaken to achieve objectives 1, 2 & 4. Chapter 7 will describe the data compilation and analysis done to complete objective 3.

Unless otherwise stated the terms ‘LIFE’ and ‘PSI’ are used to represent the family-level versions of these biotic indices. Where both family and species level indices are discussed (chapter 7), these are distinguished by the suffixes $LIFE_{fam}$; $LIFE_{sp}$; PSI_{fam} and PSI_{sp}.

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.

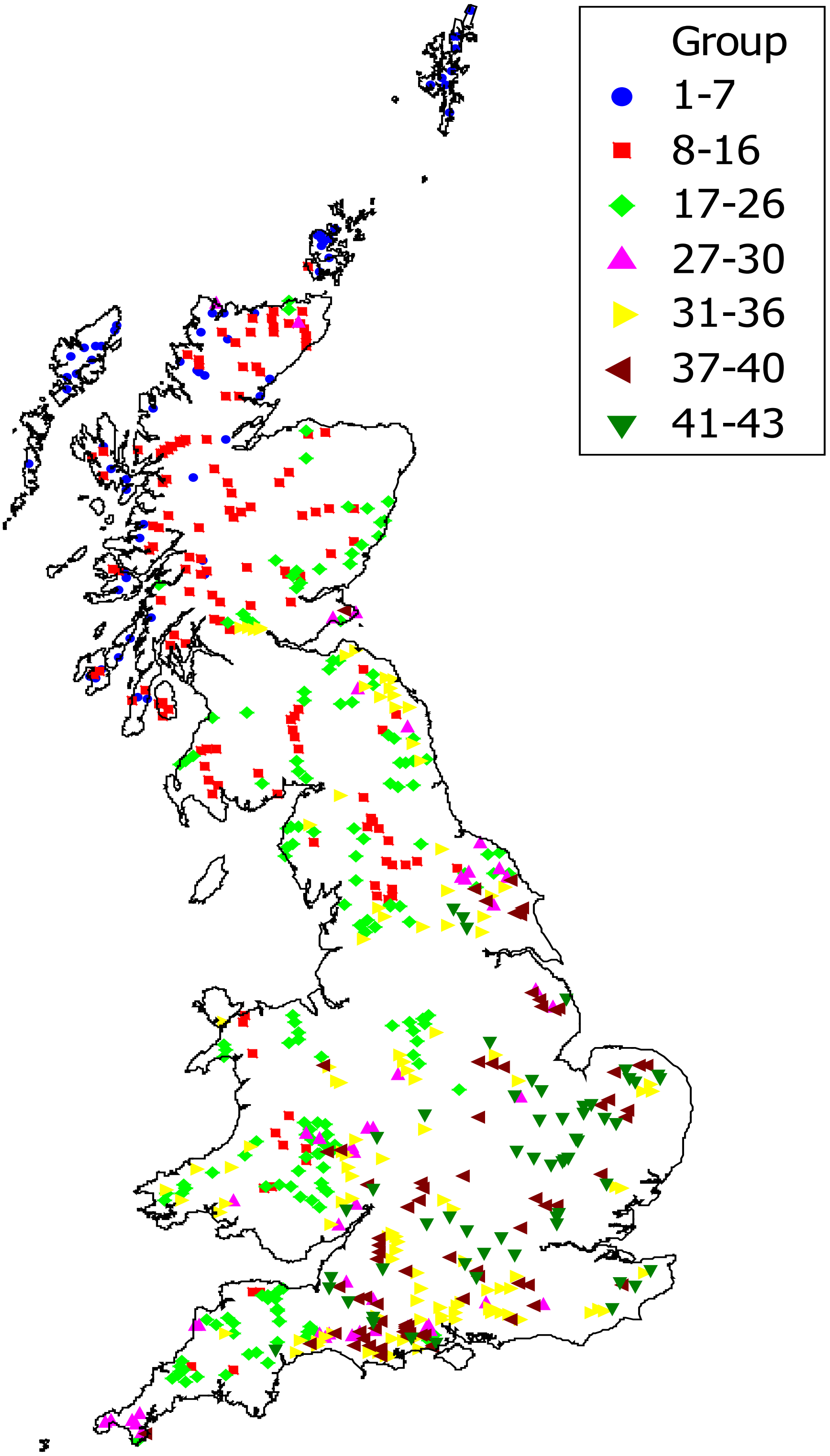


Figure 3. Observed (autumn sample) values of abundance-weighted (a) WHPT NTAXA and (b) WHPT ASPT for the 685 RICT reference sites, grouped by their end group (1-43). Vertical dashed lines separate the seven super-groups described in Table 1.

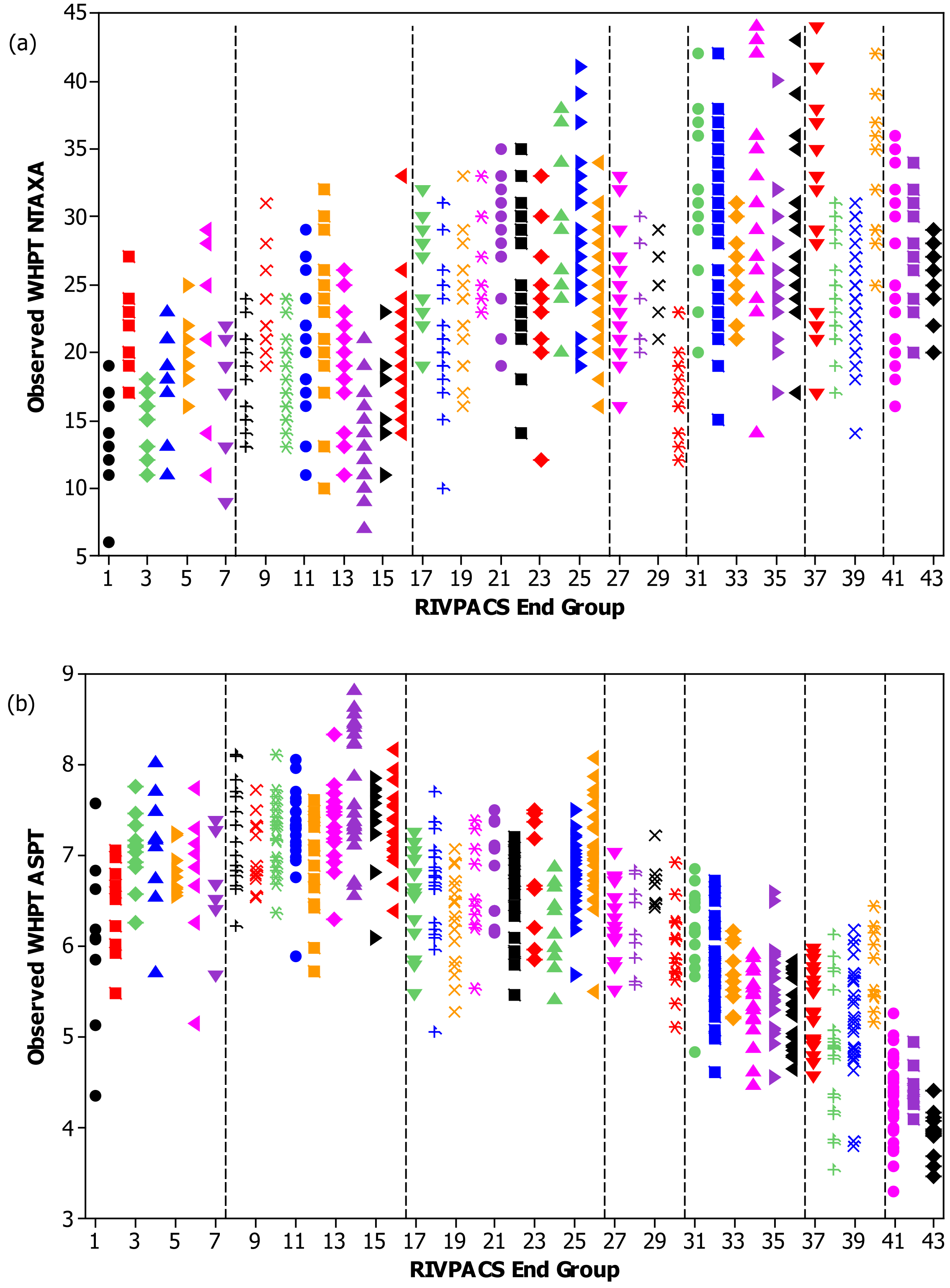


Figure 4. Distribution of O/E (EQI) values of the abundance-weighted (a) WHPT NTAXA and (b) WHPT ASPT indices for the 685 RICT reference sites, grouped by their end group (1-43). Dashed lines indicate O/E values of unity (1.0). Autumn sample values shown for illustration.

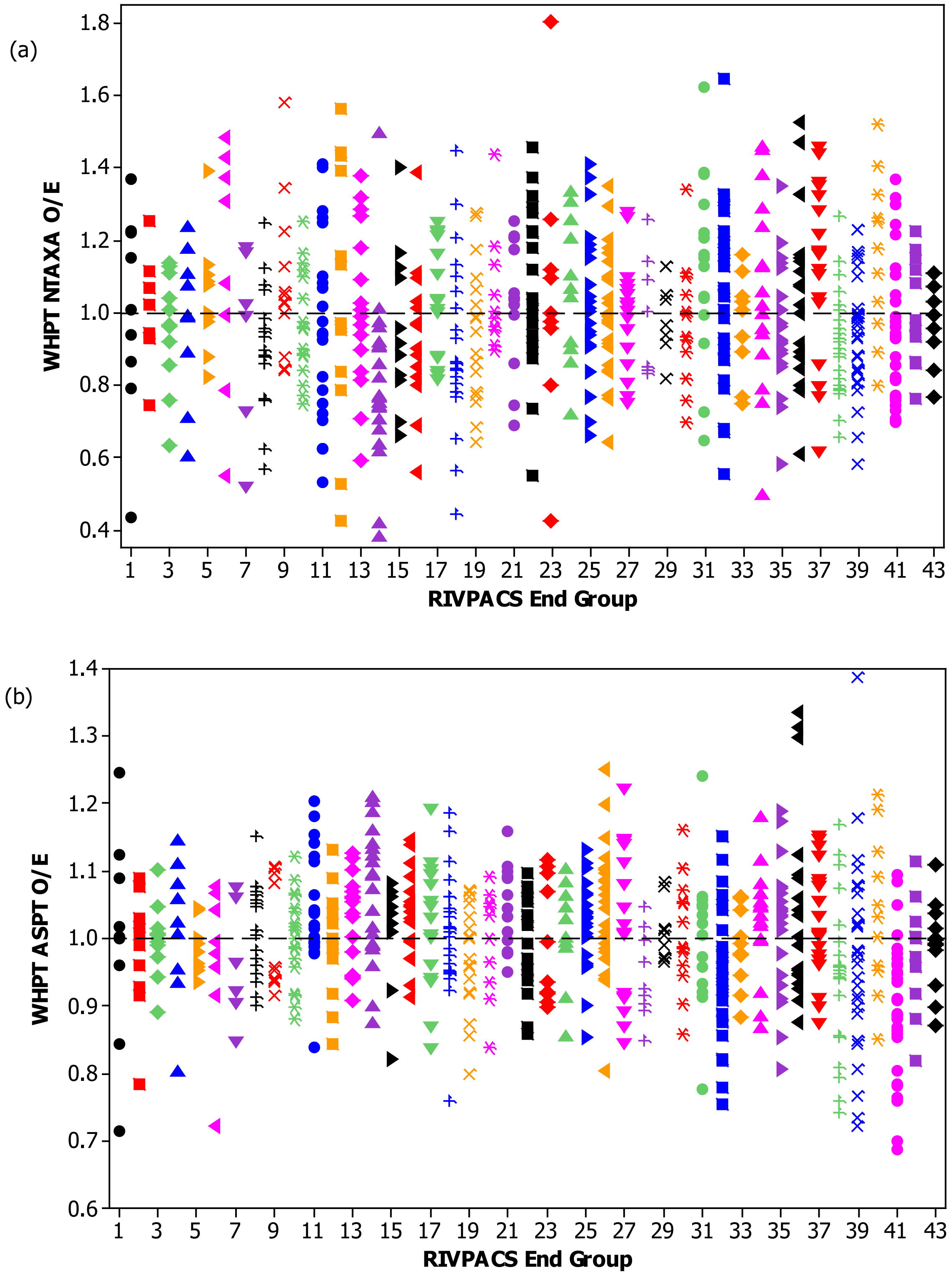
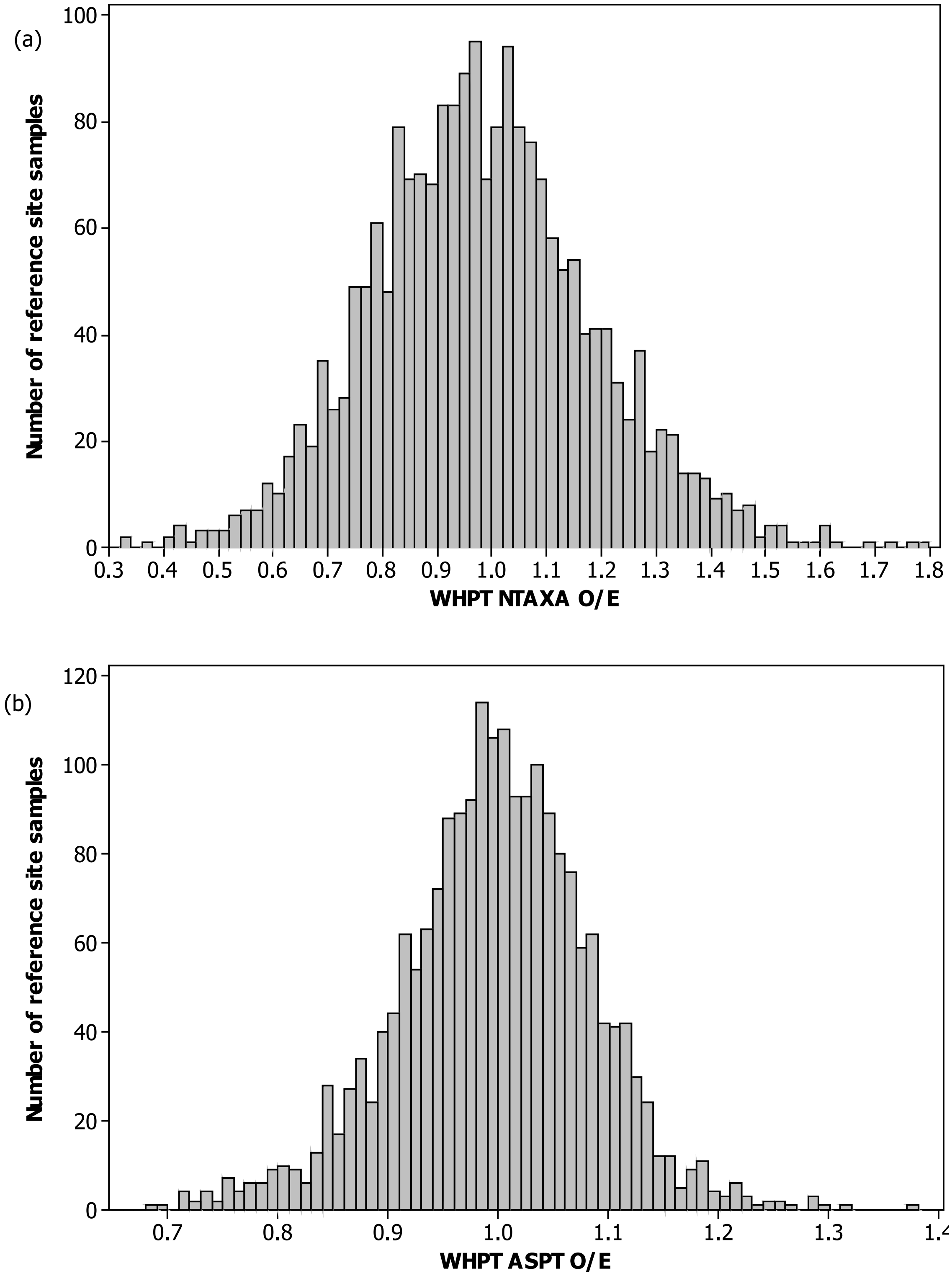


Figure 5. Frequency histogram showing the statistical distribution of the O/E (EQI) values of the abundance-weighted (a) WHPT NTAXA and (b) WHPT ASPT for the 685 reference sites for all single season samples. Expected values were based on the standard RIVPACS IV predictive model and then adjusted for perceived quality of the references sites involved.



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.

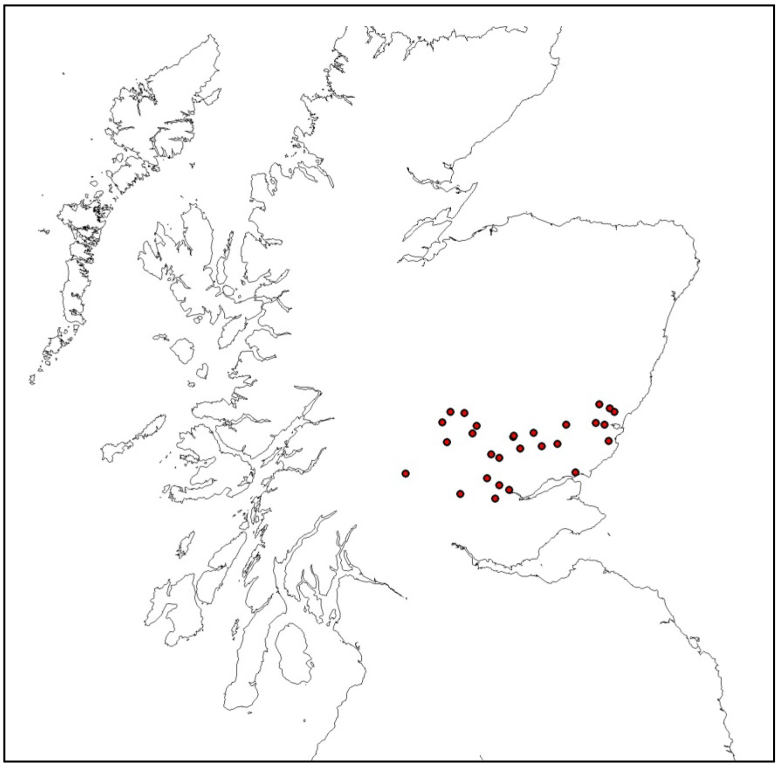


Figure 6. Map of the 28 sites in the Tay River Purification Board.

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, PSI 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).

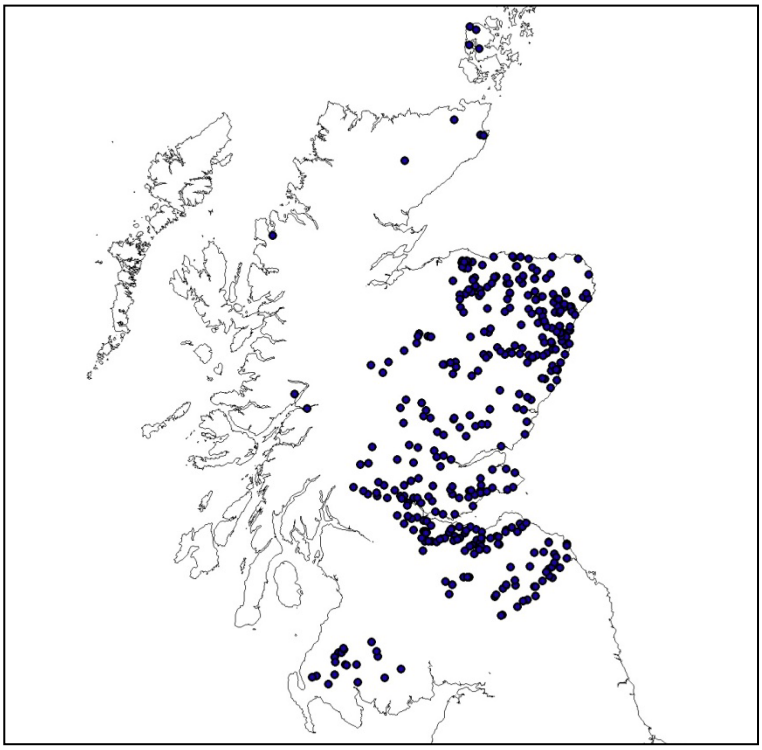


Figure 7. Map of the 416 sites in the East and North-East of Scotland.

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 $\log_{10}$ 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 $\log_{10}$ abundance RG allocated the mean numerical abundance value for that species at that $\log_{10}$ 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 $\log_{10}$ 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 $\log_{10}$ 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, PSI 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, PSI and both the non-abundance weighted WHPT and abundance weighted WHPT indices, BMWP composite families were used for scoring as opposed to distinct families.

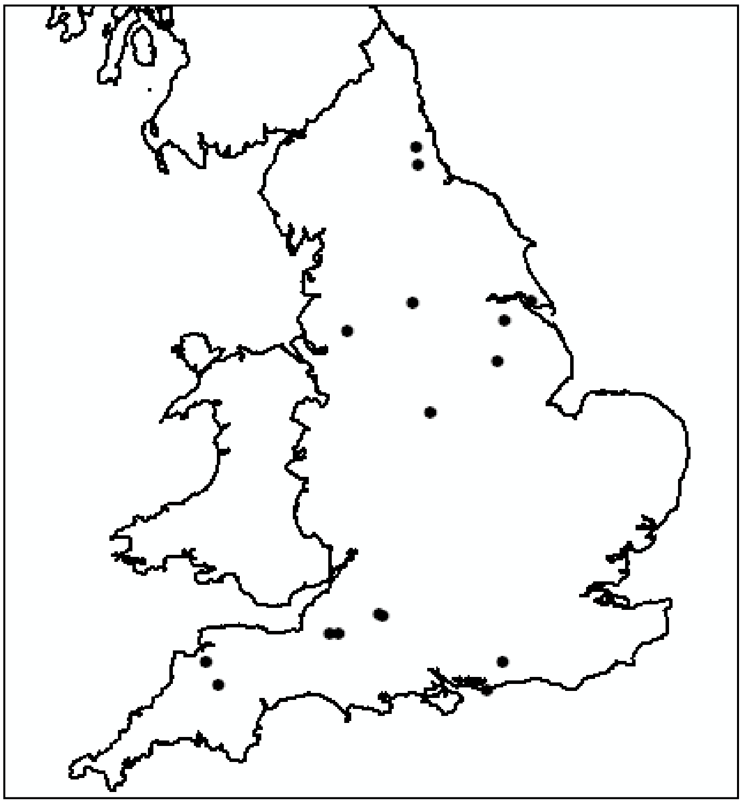


Figure 8. Map of the 16 BAMS sites in England and Wales.

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.

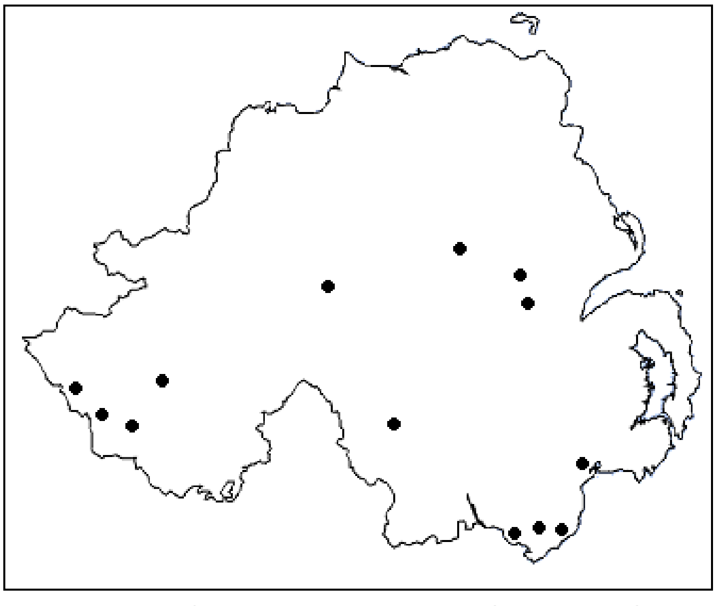


Figure 9. Map of the 12 Northern Ireland Community Change Study sites.

The Community Change Study dataset was of particular interest because unlike the other datasets it contained 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, PSI 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 (family level) index. The indices and their abundance-weighted taxonomic scoring systems are described in Appendices 1, 2 and 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 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.

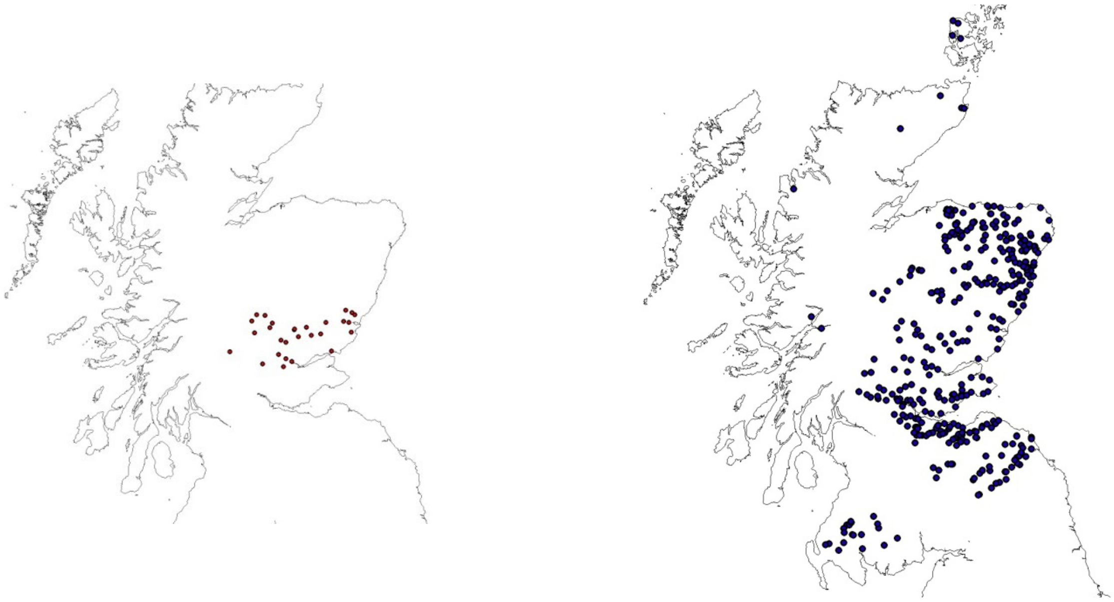
The focus of this current project (workstream 2) is to develop sampling uncertainty methods and estimates to help allow the WHPT, LIFE and PSI indices to be used and incorporated into assessments of site ecological status (see workstream 1, Clarke & Davy-Bowker (2014) for estimates of sampling variance for WHPT). 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 $\log_{10}$ categories of abundance (1 = 1-9, 2 =10-99, 3 = 100-999, 4 = 1000+ ; see in Appendices 1, 2 and 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 $\log_{10}$ 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.

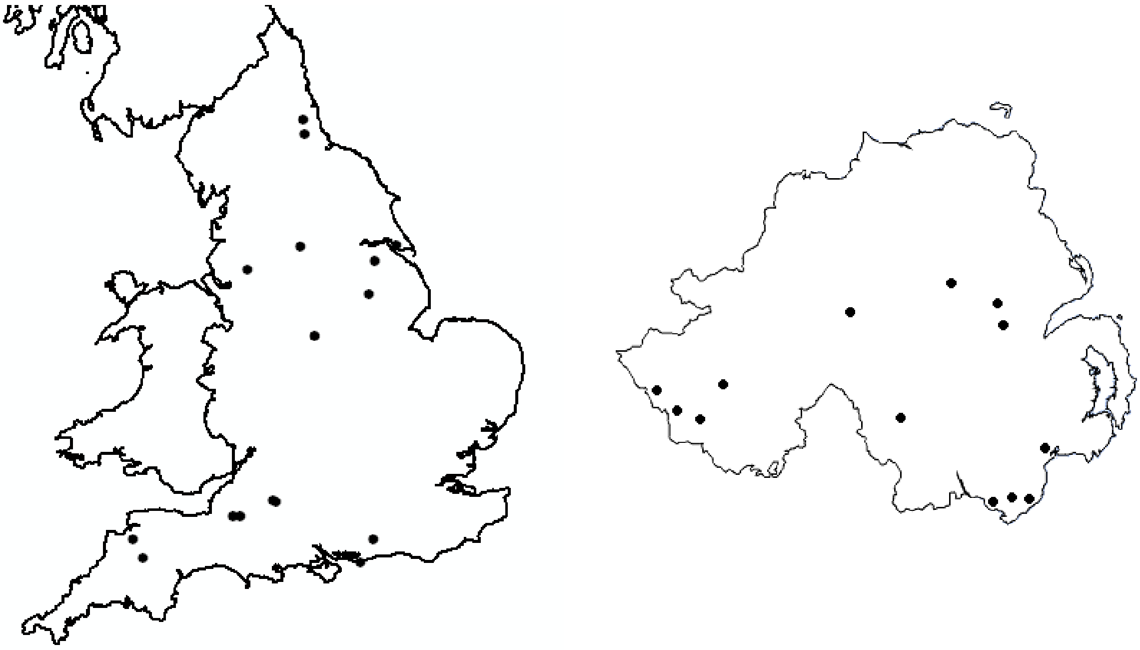
However, a separate reason for using single season samples, rather than combining them into two- or three-season lumped samples, is that there is additional information in calculating the O/E (and EQR) for each single sampled season and then defining the 'overall' WFD site quality for either a single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

Four datasets have been used in the estimation of the various error variance parameters for each of the biotic indices (Figure 10). The first two datasets were supplied by Robin Guthrie (SEPA), the third is the BAMS dataset (Furse *et al* 1995, Clarke *et al* 2002) and the fourth was supplied by Tommy McDermott, then of the Northern Ireland EHS. These are the same datasets (and statistical modelling approaches) as used by Davy-Bowker *et al* (2008) to estimate the variance components of the BMWP indices used in the development of the original RICT assessment software back in 2008.



(a) 28 Tay River Purification Board sites (b) 416 sites in the East and NE of Scotland



(c) 16 BAMS sites in England and Wales (d) 12 Northern Ireland sites

Figure 10. Geographic distribution of the sites used to estimate one or more biotic index variance parameters: (a) 28 Tay River Purification Board sites, (b) 416 SEPA sites in the East and NE of Scotland, (c) 16 BAMS sites in England and Wales, (d) 12 Northern Ireland sites.

The datasets vary in their extent of replicated sampling, spatial and temporal coverage as follows:

- 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

Further details of each dataset are provided in Chapter 2 and also in the Clarke and Davy-Bowker (2014) report to SEPA.

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 the three mainland UK datasets. 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.

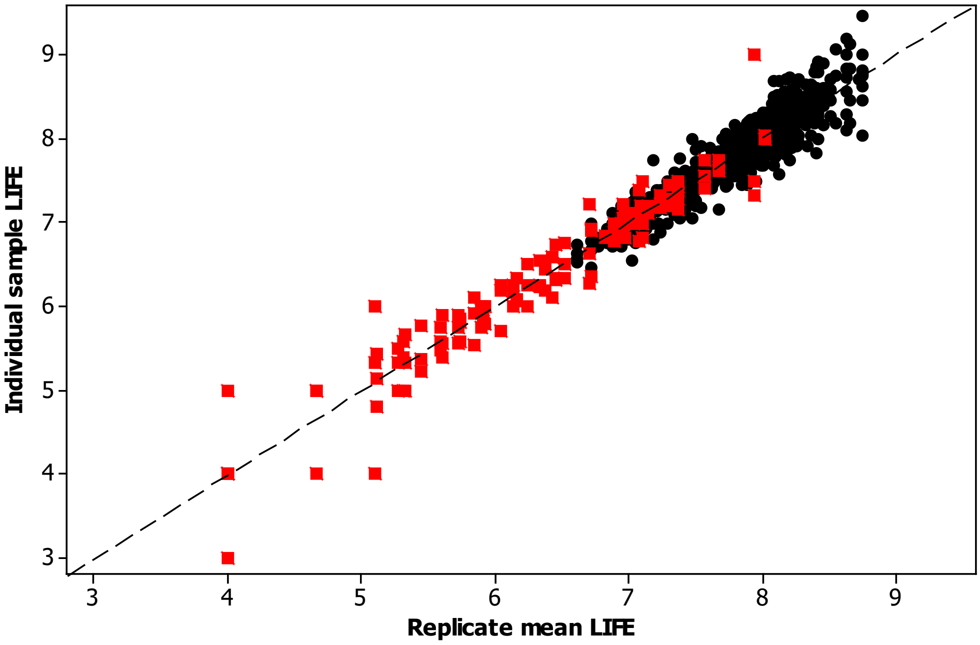
The first step in assessing the variability in biotic index values due to replicate sampling effects was therefore to plot the sampling standard deviation of replicate samples from the same site, year and season against the mean value of those replicate samples to help assess whether, and how, the variability 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 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 11).

Figure 11. 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.

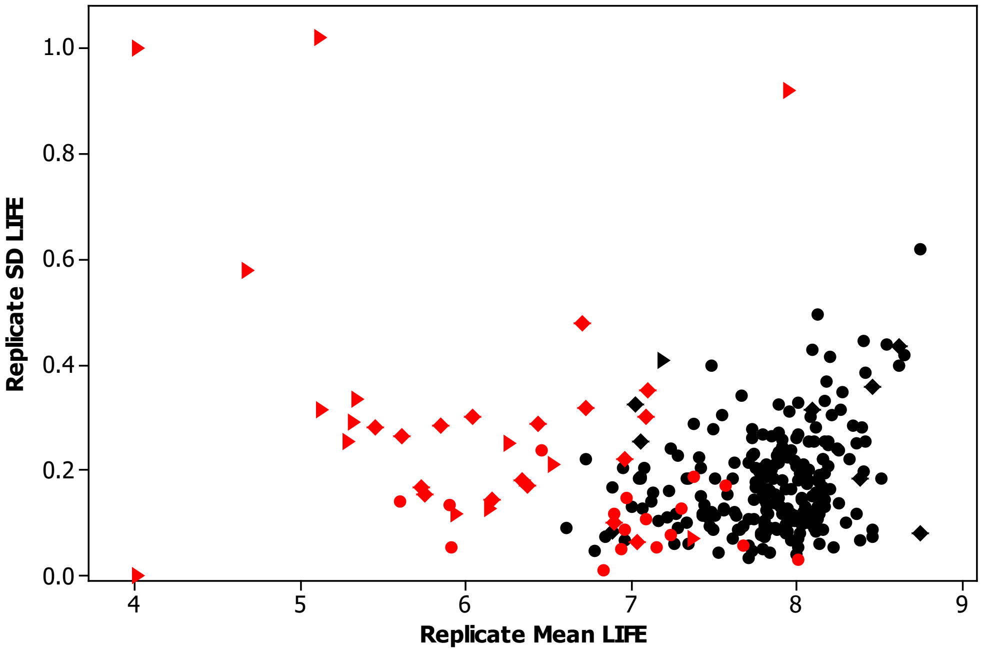
There was no apparent general relationship between the replicate SD of single season sample LIFE values and the mean of the replicate sample LIFE values; the log variance – log mean regression relationship did have negative slopes, suggesting some tendency for replicate sampling variance to decrease 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 12). 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.

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
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%

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 12. 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 currently 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 (NLIFE) 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 (NLIFE 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 12).

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 13). 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 13 (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 13 (b)) was excluded from the subsequent analyses (Figure 13 (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 13 (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 13 (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 (Eqn 3.1).

Equation (Eqn 3.1) 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 13. 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.

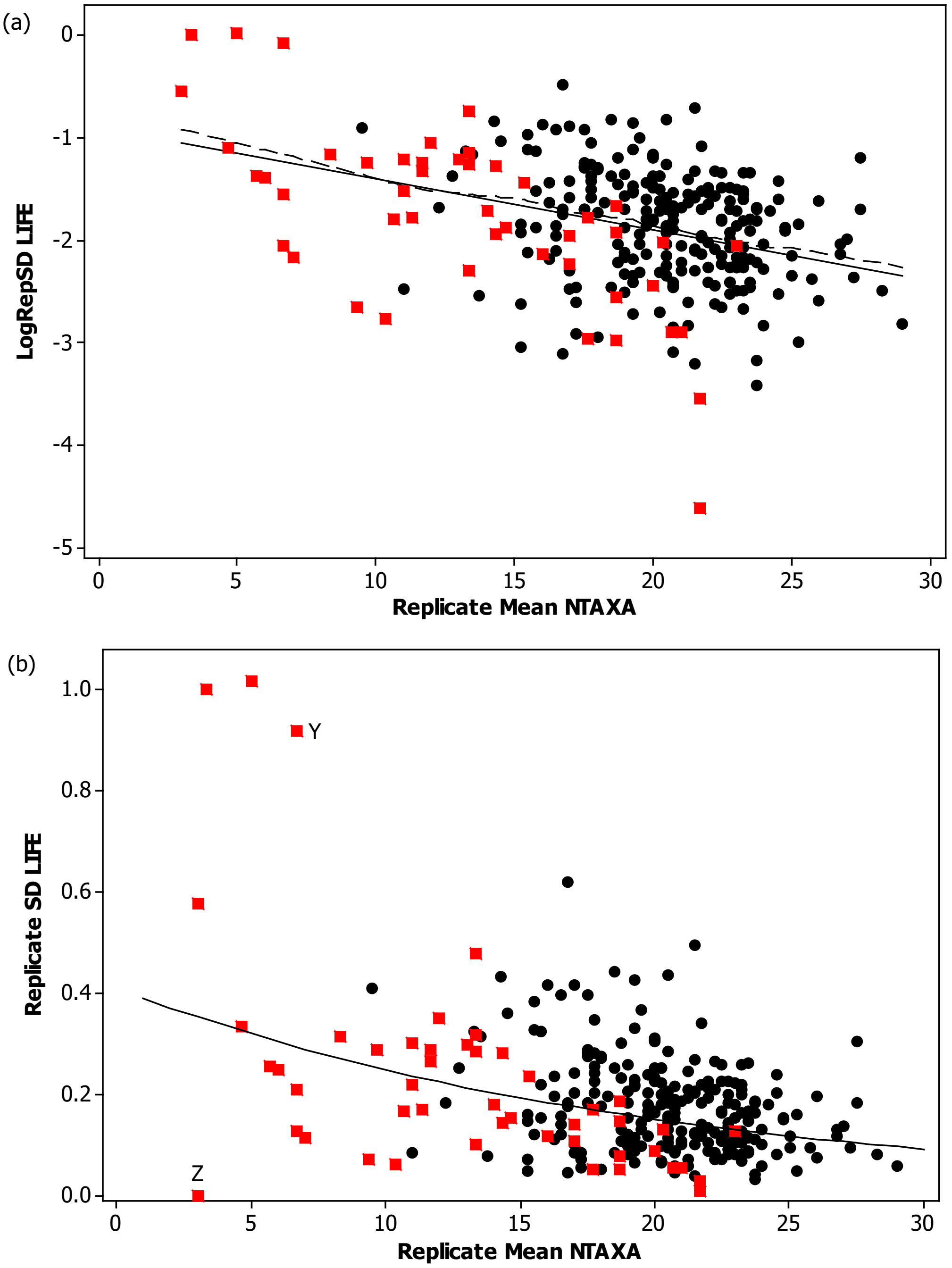


Table 4. Estimates of replicate sampling standard deviation (SD) of observed LIFE for sites based on the average number (*NTAXA*) of BMWP families present in a sample from that site (estimates based on equation (Eqn 3.1)).

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 13.

3.3.2 PSI (family) replicate sampling SD

The PSI (Proportion of Sediment-sensitive Invertebrates) index is a new index developed by Chris Extence and colleagues (Extence *et al* 2013) 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 2 contain 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:

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

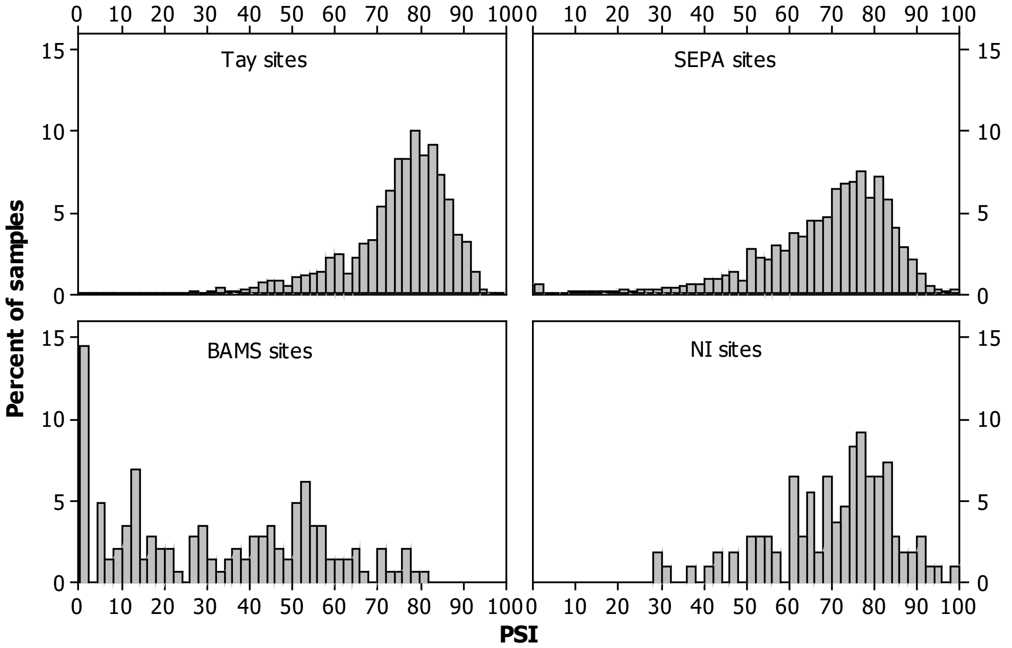
Table 5. Number of families in each PSI sediment sensitivity group.

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 datasets defined in section 2, as used to assess sampling variation in the WHPT and LIFE indices.

The distribution of the individual sample values of PSI of the four datasets are compared in Figure 14.

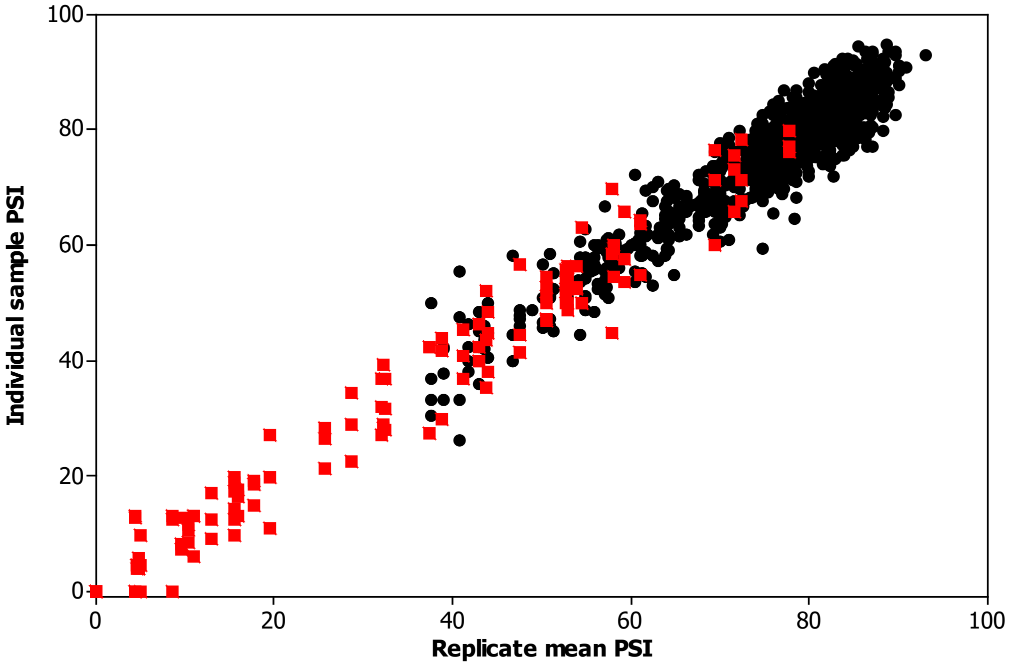
Figure 14. Histogram of the individual sample values of PSI for the four datasets.



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 14). 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 15. 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 16).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 mean 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 17). The fitted quadratic relationship (coefficient SE in brackets):

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

This 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 16).

Figure 16. 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 (●)).

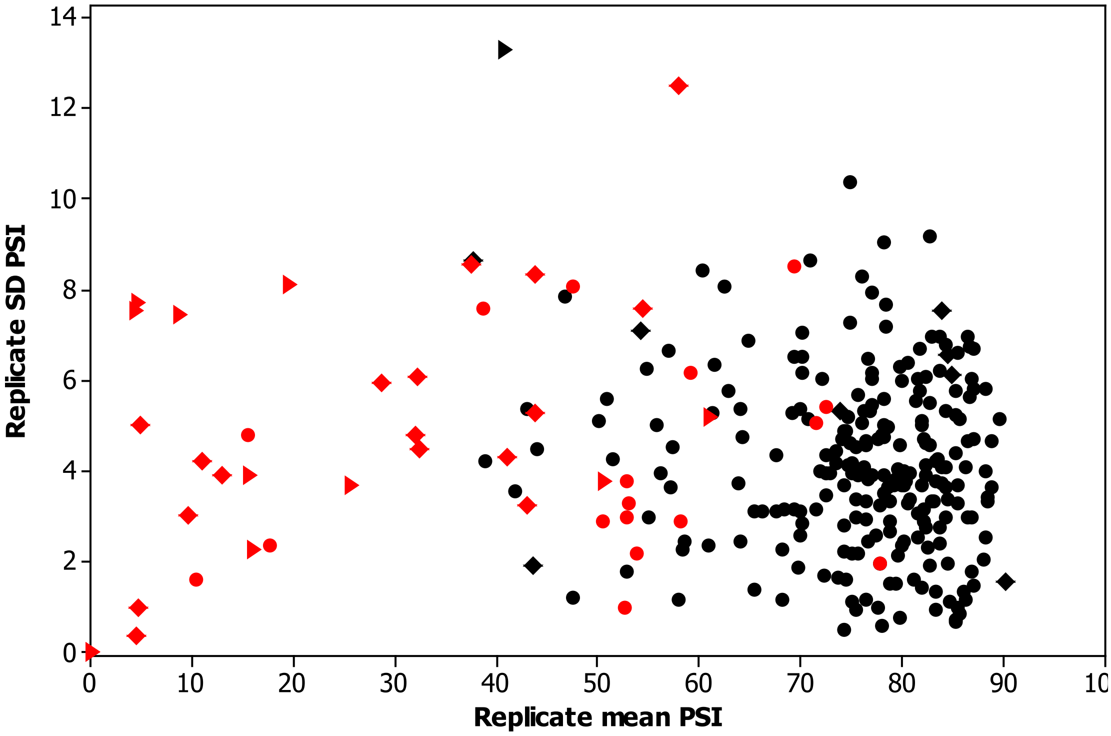
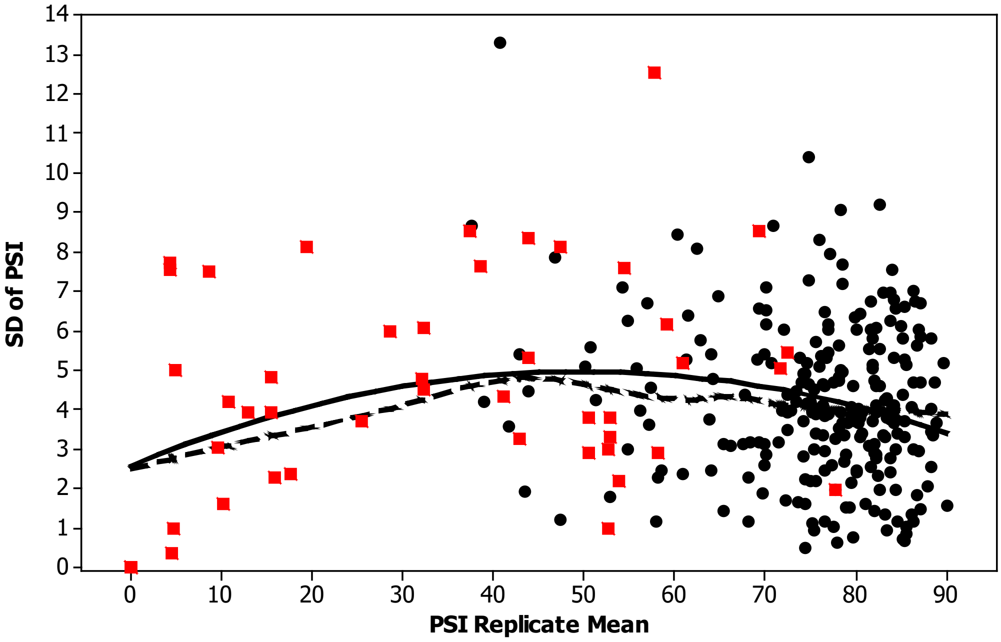


Figure 17. Plot of the relationship between replicate sampling SD of PSI and replicate mean PSI sample value for all available combinations of sites and seasons with replicate sampling for the 16 BAMS sites (■) and the 28 Tay sites (●). Lines denote quadratic (solid) and locally-weighted lowess (dashed) 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

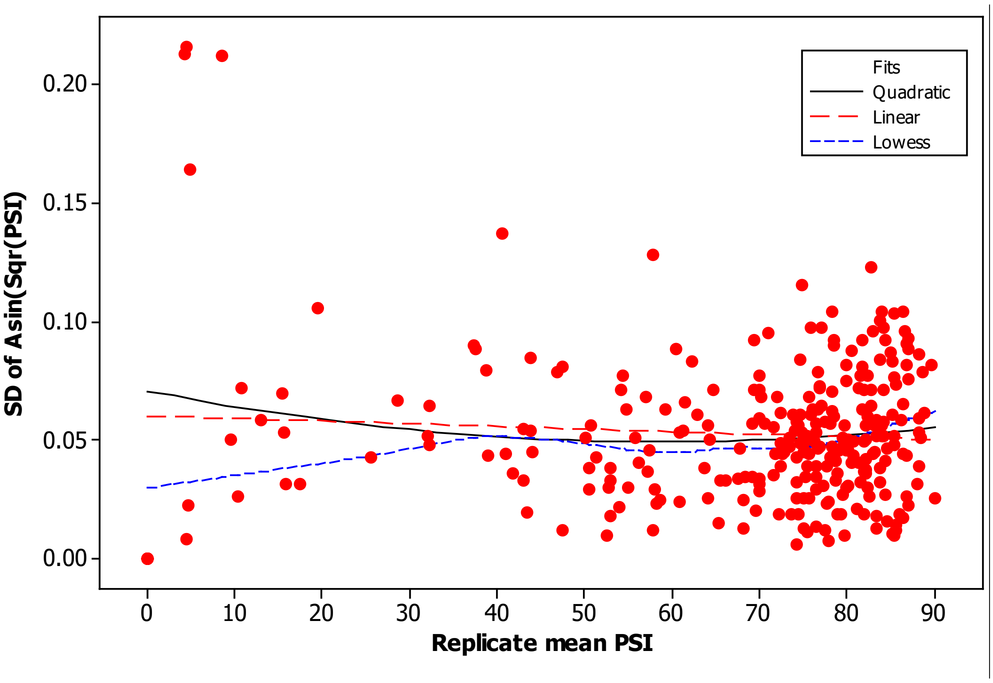
100 for percentages), can make sampling variability in values independent of sampling mean value. We applied this transformation to the individual sample PSI values as follows:

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 18). [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 18. 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 ($r = -0.32$, $p < 0.001$) 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:

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

This is superimposed as the solid line in Figure 19. 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 (as per Eqn 3.3) the observed sample PSI values, then create many (10,000) 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 (denoted Sin), as follows:

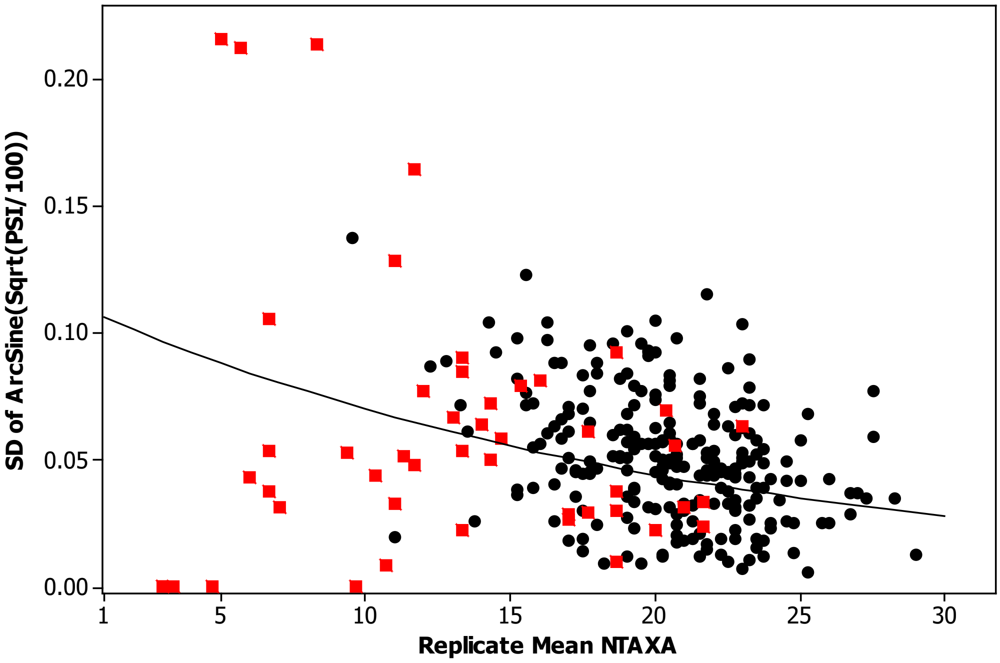
$$PSI_{Sim} = 100\ (\text{Sine}(PSI_{AsinSim}))^2$$

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 (10,000) 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 20 and Table 7.

Figure 20 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 19. 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 to 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 sites 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 20. 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 constant, or in relation to BMWP NTAXA (1,5,10,20,30; Eqn 3.4), or as quadratic between SD and mean on an untransformed scale (Eqn 3.2).

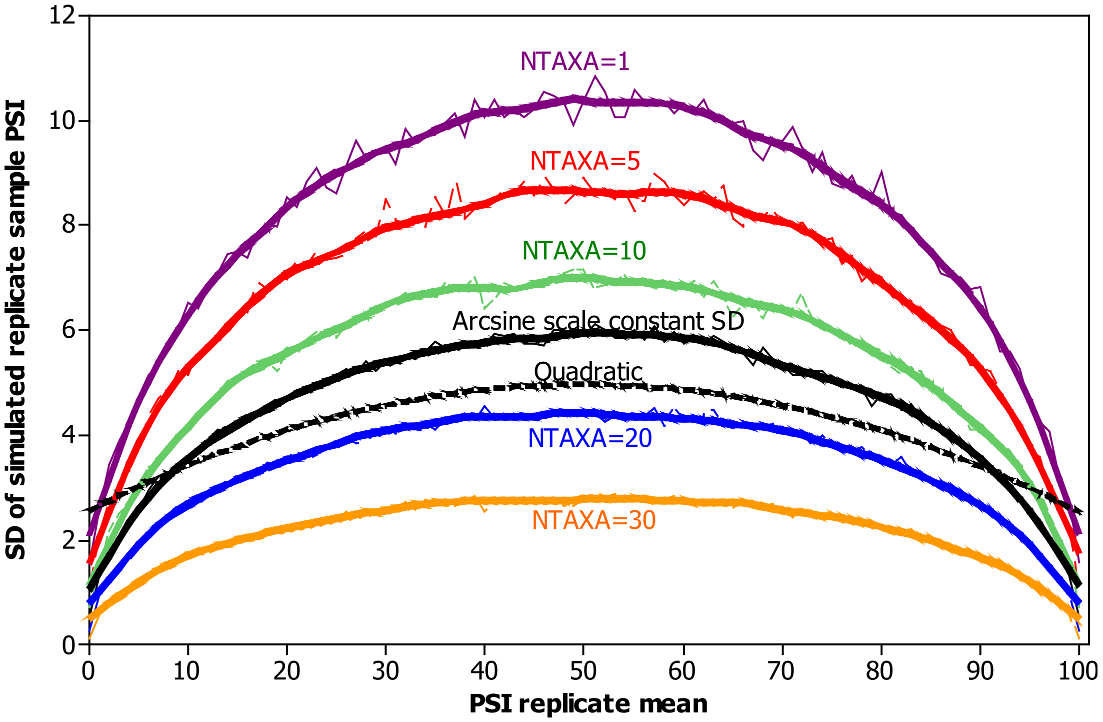


Table 7. Estimates of replicate sampling SD of PSI in relation to sample PSI values (0-100): (a) based of best estimate of constant SD (0.0596) on Arcsine scale, (b) based on relationship with NTAXA on Arcsine scale (Eqn 3.4), and (c) based on a quadaratic relationship with sample mean untransformed 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 20.

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 an 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 datasets 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
Abundance-weighted WHPT	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

Correlations between indices were assessed in two forms for the BAMS dataset which includes a wide range of types and qualities of sites and includes replicate variation for all sites and seasons. WHPT Score and WHPT NTAXA index values were analysed on the square root scale as this had already been found to make their variability more equal and less skewed (Clarke & Davy-Bowker 2014, section 3.3.1).

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 Clarke & Davy-Bowker 2014 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.

Firstly the correlations were assessed amongst the raw index values (transformed as appropriate). In this case the correlation measures the extent to which values of the two indices tend to vary together across all data whether co-varying between sites, covarying between seasons, or between replicates from the same site and season, or a mixture of all three.

The second approach was to remove all differences due to site and season by doing a one way ANOVA (Snedecor & Cochran 1980) on the raw index values allowing for a factor which was different for every site by season combination and then storing the residuals from this ANOVA. Correlations between these residuals then represent the average correlation between replicate sample values of the two indices from the same site and season. It is this correlation amongst replicates that is of main interest here in our assessment of replicate and other sampling variation.

Table 9. Pearson correlations between the abundance-weighted 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.

		WHPT Abundance-weighted			
(a) raw values		Score	Taxa	ASPT	LIFE
WHPT Abundance-weighted	NTAXA	**0.958**			
	ASPT	**0.929**	0.784		
	LIFE	0.703	0.587	0.807	
	PSI	0.754	0.593	0.877	0.895
(b) replicate residuals		Score	Taxa	ASPT	LIFE
WHPT Abundance-weighted	NTAXA	**0.920**			
	ASPT	0.708	0.414		
	LIFE	0.123	-0.051	0.481	
	PSI	0.102	-0.069	0.396	0.507

The correlations between WHPT NTAXA and WHPT ASPT were quite high (r = 0.784) in raw form, but moderately low (r = 0.414) among replicate residuals. Similar results for residual correlations amongst BMWP NTAXA and BMWP ASPT were first obtained by Furse *et al* (1995 - their Table 5.2), and 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 same logic applies to treating and simulating sampling errors in WHPT NTAXA and WHPT ASPT as independent for uncertainty simulation purposes in RICT.

In raw form, the abundance-weighted WHPT NTAXA index has moderate positive correlations with both LIFE and PSI with correlations of 0.587 and 0.593 respectively (Table 9 a). However, the correlation amongst replicate sample values of WHPT NTAXA and both LIFE and PSI are negligible (-0.051 and -0.069, Table 9 b) indicating that in terms of sampling variability PSI and LIFE are effectively independent of WHPT NTAXA.

In raw form, the abundance-weighted WHPT ASPT index has high positive correlations with both LIFE and PSI with correlations of 0.807 and 0.877 respectively (Table 9 a). The correlation amongst replicate sample values of WHPT NTAXA and replicate sample values of both LIFE and PSI are much smaller but still positive (0.481 and 0.396, Table 9 b)). Thus there is some mild tendency for the replicate sampling variability in both LIFE and PSI to be moderately correlated (but <0.5) with WHPT NTAXA.

In terms of sampling uncertainty simulation and confidence of class within RICT, the results of multi-metric site assessments based on WHPT NTAXA, WHPT ASPT and either LIFE or PSI are unlikely to be much affected by ignoring this moderate replicate correlation.

The correlation between LIFE and PSI was strongly positive (r = 0.895) amongst the raw index values for the BAMS dataset as a whole, indicating that they tend to show a similar pattern of variation between sites. Amongst replicate residuals the correlation was greatly reduced to 0.507. However, sampling correlations of this size and larger indicate that within a site and season, these two indices tend to vary together to some extent, when one is high in a sample, the other is high, relative to the true mean of all replicates for that site and season.

In the European Union Framework projects STAR (Clarke & Hering 2006) and WISER (Clarke 2013), Ralph Clarke developed the general STARBUGS and WISERBUGS software for assessing the uncertainty and confidence of WFD status class for multi-metric water body assessments. This was developed from Ralph Clarke’s previous experience in developing the uncertainty aspects of the RIVPACS III+ software system. The WISERBUGS software incorporates the option to include sampling correlations between biological indices in the uncertainty assessments for sites based on two or more biological indices. It may be possible to include such a facility within future versions of the RICT software.

Caroni *et al* (2013) also examined the general problem of the effect of correlated indices and stressors on multi-metric waterbody assessments in the context of the WFD. They recommended that the worst-case rule (one-out-all-out) should be used for determining the overall status class of a waterbody when basing an overall status class assessment on more than one type of index or biological quality element.

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 within 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
= $100\ Var_{TSeas}\ /\ (Var_{TSeas} + Var_{TYear})$

Where

Var_{TSeas} = Within season temporal variability
Var_{TYear} = Inter-year variability (within 3-year periods)

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 datasets, 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 (Var_{Rep} + Var_{TSeas} + 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
WHPT 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	
WHPT 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
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 (see section 3.2).

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:

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

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.

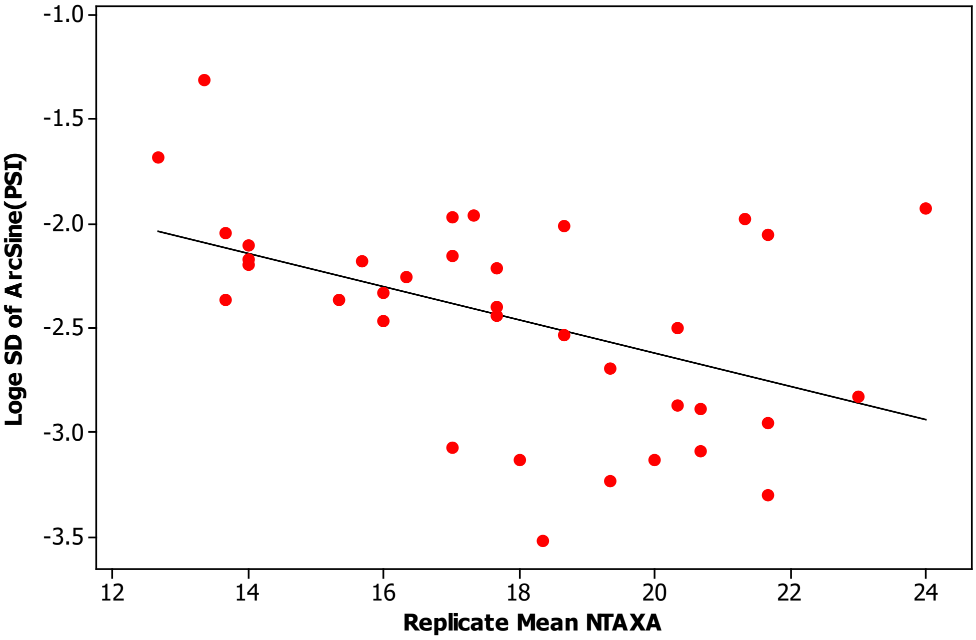
Table 11. Estimates of single season sample values for overall SD within a season (SD_{WSeas}), based on the combined effect of replicate and within-season temporal variability for (a) SEPA, Tay and BAMS datasets combined, and (b) NI monthly-sampled dataset.

		(a) SEPA+Tay+BAMS			(b) NI
Index	Index form	SD Rep	SD TSeas	SD WSeas	SD WSeas
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 $\log_e SD_{WSeas}$ of Arcsine transformed PSI values 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 21), adding further support to the use of this form of relationship to estimate sampling uncertainty of PSI values for UK river sites.

Figure 21. 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.



Summary

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 family-level 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))

This 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 decreased 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
Arcsine Sqr 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).

<u>Recommendation</u>
We recommend that for PSI the mid-value estimate of SD_{TYear} of 0.0288 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(b) for LIFE and Eqn 3.4(b) for PSI.

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 in RICT

The recommended set of sampling standard deviation (SD) parameter estimates for each form of abundance-weighted WHPT index, 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}
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 deviation (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, and K_{LIFE} therefore equal to 0.703, then for assessing the uncertainty in the average LIFE (or LIFE EQI) values for the site, we use:

SD_{Rep} = 0.703 x 0.211 = 0.148
SD_{TSeas} = 0.703 x 0.118 = 0.083
SD_{Tyear} = 0.703 x 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. Estimation of biases in abundance-weighted indices

4.1 Datasets used to estimate biases in abundance-weighted indices

The former RIVPACS III+ software and the current RICT software have the ability to adjust the BMWP NTAXA and BMWP ASPT observed index values and increase their uncertainty according to a User-supplied estimate of the typical average levels of under-estimation (i.e. bias) of the number of BMWP taxa present in samples from the same laboratory/region and year/period as the river site being assessed. Our initial approach to estimating bias values for LIFE, PSI and WHPT 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, the same simple user-supplied estimate of BMWP NTAXA 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 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 been 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 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

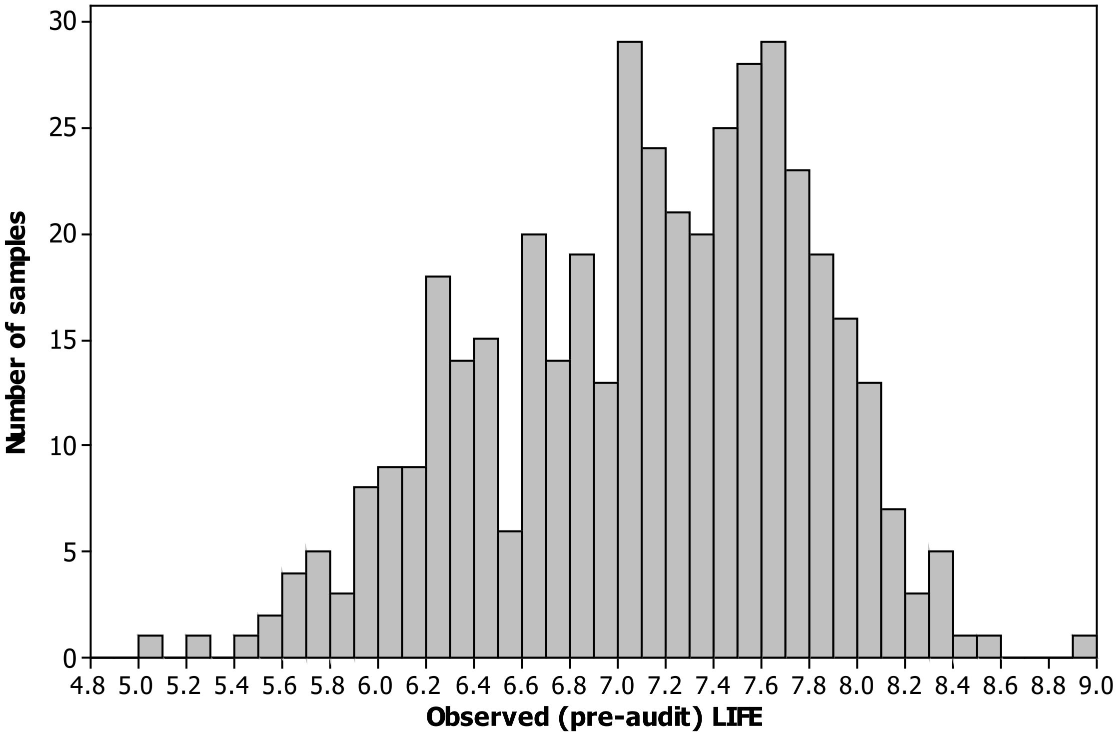
4.2 Family-level LIFE bias

These analyses are all based on the family-level abundance-weighted form of LIFE (hereafter referred to as LIFE), as this is what is intended for use within the new RICT software.

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. LIFE bias for a particular sample equals the audit-corrected (post-audit) sample LIFE value minus the pre-audit ‘observed’ LIFE value for the sample.

The (pre-audit) observed value of LIFE amongst the 427 audited sample dataset varied from 5.00 to 9.00 (which encompasses the major range of LIFE values observable in practice), with an inter-quartile range (i.e. middle 50%) of 6.67-7.64 and with a mean and median of around 7.2 (Figure 22).

Figure 22. Frequency histogram of the observed (pre-audited) values of LIFE amongst the 427 audited samples dataset.

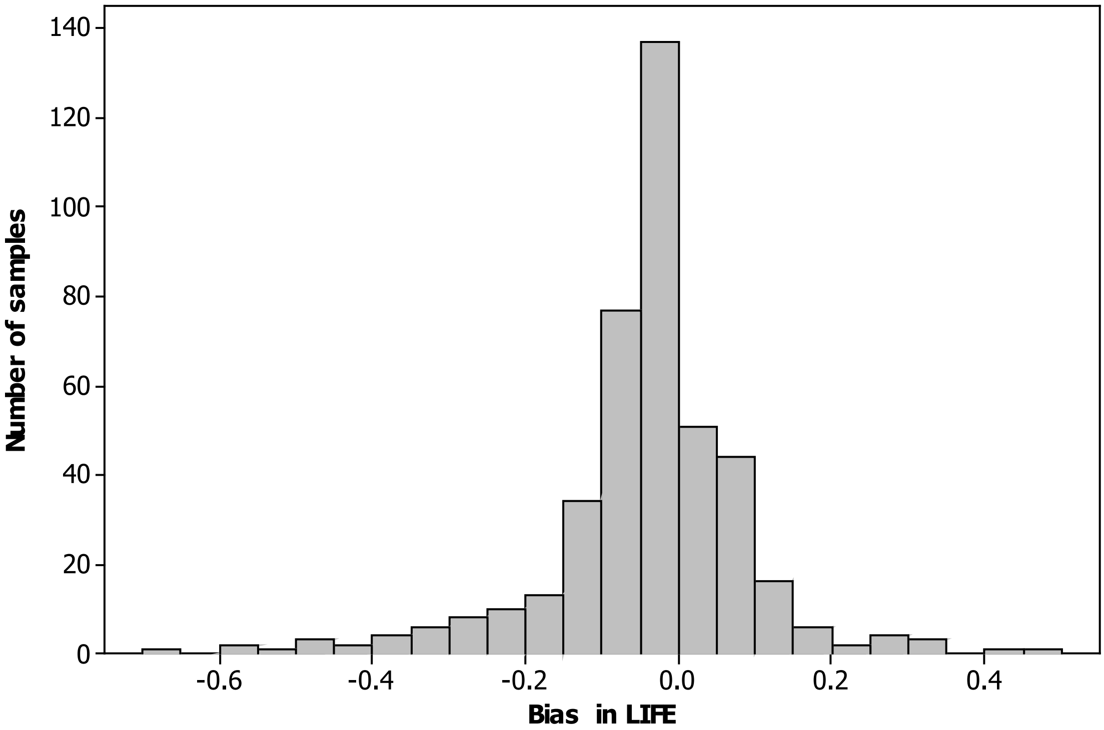


There was one audited exceptional sample with only two families present for which the bias in LIFE was -1.5 arising from a pre-audit LIFE value of 9.0 and an audit-corrected value of 7.5 once one missed family is included. As no other audited sample had less than five BMWP families observed and the next lowest bias in LIFE was only -0.67, this extreme outlier sample was excluded from the remaining analyses.

Amongst the remaining 426 audited samples, only 11 (2.6%) had less than 10 BMWP families recorded. The bias in LIFE varied from -0.67 up to +0.45, with an inter-quartile range (i.e. middle 50%) of -0.09 to +0.02 and with a mean and median of -0.04 and -0.01 respectively (Figure 23).

Overall, after audit correction, 51% of sample LIFE values went down, 19% stayed the same and 30% increased. Thus there is some overall tendency for the average LIFE score of the missed taxa to be slightly lower than that of the observed and recorded taxa, so that the bias adjustment (post-audit minus pre- audit LIFE) is more likely than not to be negative. However, the bias may vary with the taxonomic richness of the sample and site or with its observed LIFE value and this is investigated below.

Figure 23. Frequency histogram of the bias (post- minus pre- audited values) in LIFE amongst 426 audited samples



4.2.1 Factors affecting LIFE bias values

The sample bias in observed LIFE does not seem to vary systematically with sample taxonomic richness as represented by the observed (pre-audit) sample number of BMWP taxa and their Spearman rank correlation was -0.01 (Table 17). Thus samples with more taxa do not tend to have higher or large sized biases in observed LIFE values.

Table 17. Bias in LIFE in relation to observed (pre-audited) sample BMWP NTAXA.

Observed (pre-audit) BMWP NTAXA		Bias in LIFE					
	Samples	Mean	SD	Min	Median	Max	% positive
≤10	26	-0.03	0.19	-0.57	-0.02	0.29	35%
11-15	59	-0.01	0.13	-0.38	0.00	0.33	27%
16-20	113	-0.06	0.13	-0.57	-0.03	0.19	26%
21-25	154	-0.02	0.12	-0.67	0.00	0.45	35%
26-30	52	-0.07	0.13	-0.51	-0.04	0.09	21%
31-36	22	0.00	0.07	-0.12	0.00	0.24	41%
Overall	426	-0.04	0.13	-0.67	-0.01	0.45	30%

However, the sample bias in observed LIFE seems to decrease systematically with the observed (pre-audit) value of LIFE (Table 18). Amongst the 29 audited samples for which the observed LIFE was less than 6 the bias was positive in 66% of samples with a mean bias of +0.10. However, the average bias and the percentage of positive biases decreases with increasing observed LIFE, such that for the 22 samples with high (>8) observed LIFE, only 8% of LIFE biases were positive and the average bias was -0.18 (Table 18).

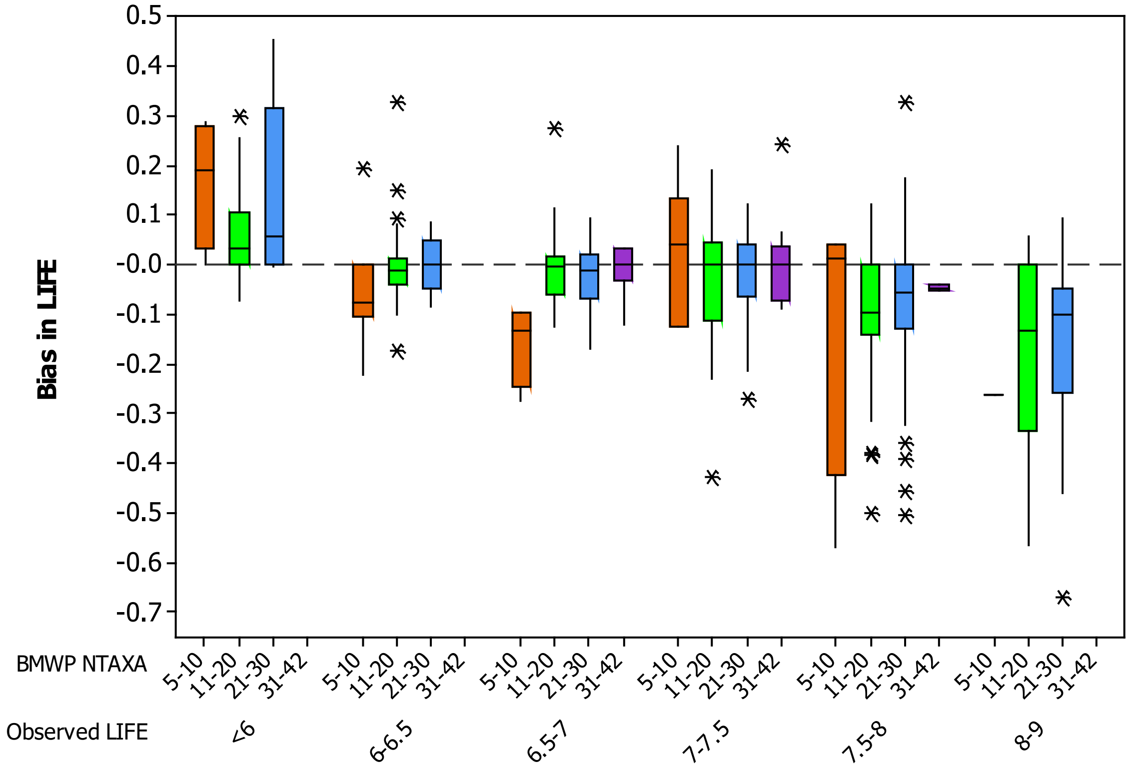
Table 18. Bias in LIFE in relation to observed (pre-audited) sample LIFE.

Observed (pre-audit) LIFE	Samples	Bias in LIFE					
		Mean	SD	Min	Median	Max	% positive
<6	29	0.10	0.14	-0.07	0.06	0.45	66%
6 .0- 6.5	63	-0.01	0.08	-0.22	-0.01	0.33	30%
6.5 - 7.0	86	-0.02	0.08	-0.27	-0.01	0.27	31%
7.0 - 7.5	112	-0.02	0.10	-0.43	0.00	0.24	37%
7.5 - 8.0	110	-0.09	0.15	-0.57	-0.06	0.33	18%
8.0 - 9.0	26	-0.18	0.19	-0.67	-0.13	0.09	8%
Overall	426	-0.04	0.13	-0.67	-0.01	0.45	30%

Thus for sites with low recorded LIFE, the average LIFE score of the missed taxa must be slightly higher than the average (abundance-weighted) LIFE score of the observed (i.e. recorded) taxa, whereas for sites with high recorded LIFE values, the average LIFE score of the missed taxa must usually be slightly lower than that of those recorded as present in the sample.

Thus overall, the bias in LIFE appears to decrease with the observed (pre-audited) sample LIFE values but not with observed sample taxonomic richness (Figure 24).

Figure 24. Boxplot of bias in LIFE in relation to the observed sample LIFE value and the observed BMWP NTAXA

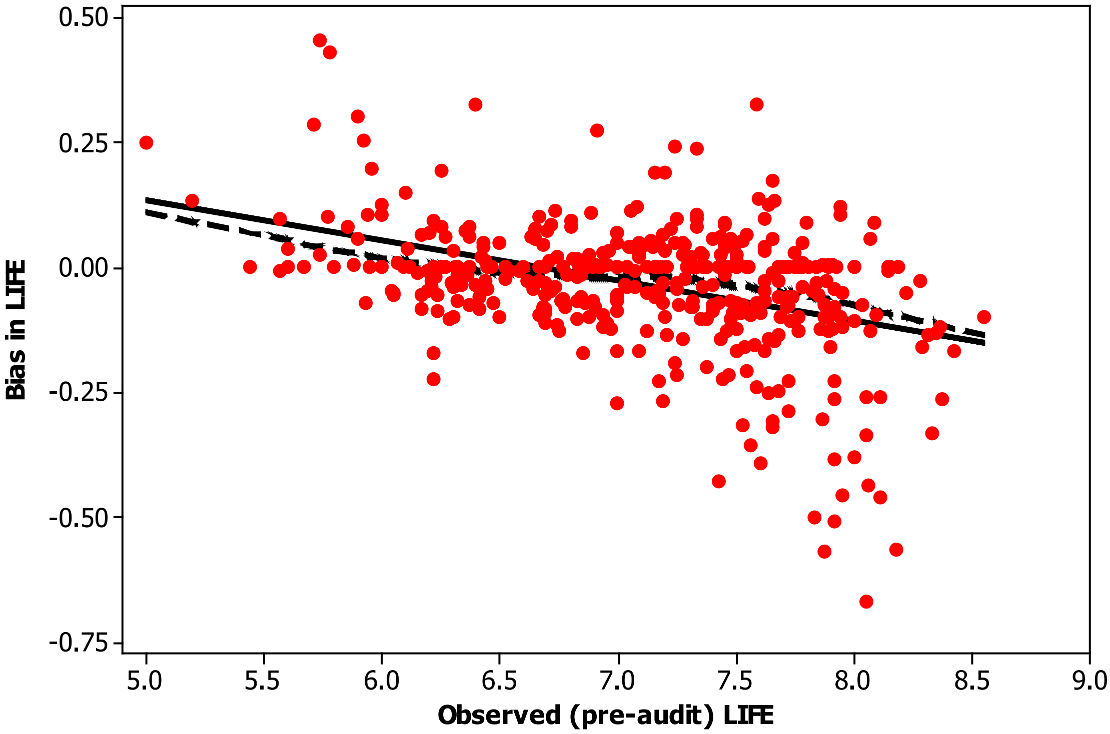


The best approach is judged to be to predict the bias in LIFE value for a sample from its observed LIFE value (Figure 25). The regression relationship is adequately linear and best fitted by the following linear regression relationship (coefficient standard errors in brackets):

$$\text{Bias in LIFE} = \underset{(0.062)}{0.531} - \underset{(0.0087)}{0.0797}\ \text{Observed LIFE} \qquad \text{(Eqn 5.1)}$$

Equation 5.1 explains 16% of the total variation in the values of LIFE biases amongst the audited sample and has a residual SD of 0.120 which can be used to estimate the uncertainty in the predictive estimates of LIFE bias for individual samples.

Figure 25. Linear regression relationship (Equation 5.1, solid line) between bias in LIFE and observed LIFE amongst the 426 audited samples; also shown is the non-linear optimal LOWESS regression line (dashed line).



Although only explaining a relatively small proportion of the total variation in sample bias in LIFE values, the average of the predicted values from Equation 5.1 follows a similar trend to the actual mean LIFE bias for samples in each band of observed LIFE values, indicating useful broad predictive power across a range of quality of sites (Table 19).

Table 19. Mean of actual and predicted (from Eqn 5.1) values of bias in LIFE in relation to observed (pre-audited) sample LIFE values.

Observed (pre-audit) LIFE	Samples	Mean bias in LIFE	
		Actual	Predicted from Equation 5.1
<6	29	0.102	0.072
6 .0- 6.5	63	-0.006	0.029
6.5 - 7.0	86	-0.020	-0.013
7.0 - 7.5	112	-0.021	-0.050
7.5 - 8.0	110	-0.085	-0.086
8.0 - 9.0	26	-0.178	-0.123

To assess whether the uncertainty in LIFE bias values is less when there are more missed taxa (i.e BMWP NTAXA sample bias is greater), we grouped the audited samples by both their observed LIFE value and the under-estimated number (M) of BMWP taxa and calculated the mean and SD of bias in LIFE values for the samples in each category (Table 20). The best single estimate of SD for each value of M was obtained from a one-way analysis of variance of LIFE bias on class of observed (pre-audit) LIFE to give similar estimates of SD of 0.105, 0.109 and 0.120 for M equal to 1, 2 and more than 2 respectively (Table 20).

Table 20. Mean and standard deviation (SD) of bias in LIFE values for samples grouped in relation to observed (pre-audited) sample LIFE and bias (M) in BMWP NTAXA (samples per category in brackets).

Observed (pre-audit) LIFE	Mean of LIFE Bias for M under-estimated BMWP taxa			SD of LIFE Bias for M under-estimated BMWP taxa		
	M=1	M=2	M>2	M=1	M=2	M>2
<6	0.114 (8)	0.071 (7)	0.294 (3)	0.107	0.047	0.121
6 .0- 6.5	0.034 (22)	-0.039 (11)	-0.026 (8)	0.089	0.108	0.063
6.5 - 7.0	-0.032 (29)	-0.012 (21)	-0.046 (13)	0.070	0.058	0.102
7.0 - 7.5	-0.033 (34)	-0.015 (19)	-0.057 (16)	0.085	0.115	0.104
7.5 - 8.0	-0.066 (37)	-0.041 (20)	-0.273 (15)	0.101	0.156	0.158
8.0 - 9.0	-0.199 (9)	-0.177 (7)	-0.181 (2)	0.247	0.090	0.216
	Overall average SD from ANOVA			0.105	0.109	0.120

Thus the variability (i.e. SD) of the LIFE bias does not not appear to depend on the number of missed taxa (at least in terms of the number (M) of missed BMWP taxa).

4.2.2 Relationship between bias in LIFE-scoring NTAXA and BMWP NTAXA bias

The 'bias in LIFE-scoring NTAXA' for a sample is defined to be the number of taxa with LIFE scores in the post-audit sample minus the number of taxa with LIFE scores in the observed (pre-audit) sample.

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 21). Two samples had two more taxa recorded than actually existed in the audited sample and 13 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 21).

Table 21. Two-way table summarising number of samples with each level of bias in BMWP NTAXA and bias in LIFE-scoring NTAXA (equal biases are highlighted).

		Bias in LIFE-scoring NTAXA										
		-2	-1	0	1	2	3	4	5	6	Total	%
Bias in BMWP NTAXA	-2	2	0	0	0	0	0	0	0	0	2	0.5
	-1	1	9	3	0	0	0	0	0	0	13	3.0
	0	0	6	113	10	1	0	0	0	0	130	30.4
	1	0	0	2	130	6	2	0	0	0	140	32.8
	2	0	0	2	5	65	11	2	0	0	85	19.9
	3	0	0	0	0	5	25	6	1	0	37	8.7
	4	0	0	0	0	0	1	10	3	0	14	3.3
	5	0	0	0	0	0	0	0	4	2	6	1.4
	Total	3	15	120	145	77	39	18	8	2	427	
	%	0.7	3.5	28.1	34.0	18.0	9.1	4.2	1.9	0.5		100

For the LIFE index, the maximum bias in estimating the number of LIFE-scoring taxa present was 6, but the median was one taxa and the average bias amongst the 427 samples analysed was 1.68 (Table 21). Three samples had two more LIFE taxa recorded than actually existed in the audited sample and 15 samples had one more LIFE taxa recorded than existed in the audited sample. Thus only 4% of samples actually had higher observed sample values of the number of LIFE-scoring taxa present than they should have. A further 28% of samples had the correct number of LIFE-scoring taxa. However, for the majority (68%) of samples, the recorded number of LIFE taxa present was less than the audit-corrected number (Table 21).

The correlation between the individual sample bias in number of LIFE-scoring taxa present and individual sample bias in BMWP NTAXA was high at 0.94 amongst all 427 samples. The best fit regression relationship was a straight line through the origin with a slope of 1.037, explaining 85% of the total variation and suggesting that on average, the under-estimation of the number of LIFE-scoring taxa is roughly the same (only 3.7% higher) as that of the number BMWP taxa. Moreover, for the vast majority (84%) of all samples the bias in number of LIFE-scoring taxa was the same as in BMWP NTAXA. In 11% of samples the (positive) under-estimation of number of LIFE-scoring taxa was greater than the under-estimation of BMWP NTAXA, whereas for the remaining 5% of samples it was less. Thus the bias in BMWP NTAXA is a very good correlate and representation of the pattern among samples of the under-estimation of the number of LIFE-scoring taxa present.

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.

4.2.3 LIFE of missed taxa

The LIFE index for a sample is defined to be the sum of the abundance-weighted LIFE scores of LIFE-scoring taxa present in the sample divided by the number of LIFE-scoring taxa present in the sample.

The ‘LIFE of missed taxa’ is defined to be:

$$\frac{\text{Post-audit sum of LIFE scores - Pre-audit sum of LIFE scores}}{\text{Post-audit number of LIFE-scoring taxa - Pre-audit number of LIFE-scoring taxa}}.$$

‘LIFE of missed taxa’ is only defined when the number of LIFE-scoring taxa is higher in the post-audit sample than the pre-audit sample.

This is analogous to the calculation of the BMWP ASPT of missed BMWP taxa in RIVPACS III+ bias estimation and of WHPT ASPT of missed WHPT taxa used for assessing uncertainty arising from sample processing errors in recent RICT assessments based on the WHPT indices (Clarke & Davy-Bowker 2014).

There is no clear relationship between the ‘LIFE of missed taxa’ and the observed (pre-audit) LIFE value of a sample (Table 22). The lower mean value of 5.47 amongst samples with observed LIFE score greater than 8.0 was largely due to one sample with one or more taxa misidentifed such that although its post audit number of LIFE-scoring taxa increased by one, the sum of LIFE scores of all taxa decreased by 6 giving a ‘LIFE of missed taxa’ value of -6. The median ‘LIFE of missed taxa’ for this group of high LIFE samples was actually the same as those for the group of samples with observed LIFE in the range 6.0 to 6.5 (Table 22). A one-way Kruskal-Wallis analysis of ranks did not detect any differences in ‘LIFE of missed taxa’ values in relation to class of observed LIFE values (test P = 0.358).

Therefore, we conclude that the ‘LIFE of missed taxa’ does not vary with the observed LIFE and can best be estimated as a constant equal to the observed mean value of the ‘LIFE of missed taxa’ amongst all 289 samples for which the audit increased the number of LIFE-scoring taxa, namely 6.55 (Table 22).

Table 22. LIFE of missed taxa in relation to observed (pre-audited) sample LIFE.

Observed (pre-audit) LIFE	Samples	LIFE of missed taxa				
		Mean	SD	Min	Median	Max
<6	20	6.80	1.12	5.50	6.58	10.50
6 .0- 6.5	40	6.53	1.01	4.50	6.00	10.00
6.5 - 7.0	64	6.58	0.86	4.00	6.42	9.00
7.0 - 7.5	74	6.78	1.36	3.00	6.78	9.00
7.5 - 8.0	72	6.51	1.95	-1.00	6.50	11.50
8.0 - 9.0	19	5.47	3.54	-6.00	6.00	9.00
Overall	289	6.55	1.63	-6.00	6.50	11.50

However, as was found in previous studies of the BMWP ASPT and WHPT ASPT of missed taxa, it is likely that the variability in sample values of 'LIFE of missed taxa' will decrease with the number (M) of missed taxa involved and this is what was found (Table 23).

Table 23. Mean and standard deviation (SD) of LIFE of missed taxa in relation to the audit-based under-estimated number (M) of LIFE-scoring taxa (samples per category in brackets).

	Under-estimated number (M) of LIFE-scoring taxa				
	M = 1	M = 2	M =3	M = 4	M > 4
Number of audited samples	145	77	39	18	10
Mean 'LIFE of missed taxa'	6.39	6.83	6.47	6.97	6.95
SD of 'LIFE of missed taxa'	2.07	1.17	0.77	0.81	0.64
Predicted SD = 1.8 / √ M	1.80	1.27	1.04	0.90	0.77

To enable prediction of the SD of 'LIFE of missed taxa', we suggest using a simple statistical formula to represent the main pattern of observed SD, namely:

$$\text{SD of 'LIFE of missed taxa'} = 1.8 / \sqrt{M} \qquad \text{(Eqn 4.1)}$$

where √ M indicates square root of M and M is the bias in BMWP NTAXA for this observed sample. The effect of using equation 4.1 is shown in Table 23.

Recommendation
In RICT, the LIFE value for the missed taxa of a sample with a simulated estimated bias of BMWP NTAXA of M should be estimated as random deviate from a statistical Normal distribution with a mean of 6.55 and a standard deviation equal to 1.8 divided by the square root of M.

4.2.4 Simulating bias in LIFE values

The bias-corrected LIFE value for a sample is by definition a weighted average of the observed (pre-audit) sample LIFE and the 'LIFE of missed taxa'. The proportion of all (post-audit) LIFE-scoring taxa which are missed is very closely correlated with the proportion of all post-audit BMWP taxa which are missed in the same sample (correlation = 0.93). Therefore, we conclude that it is reasonable to use the proportional under-estimation (P_{miss}) of the number of BMWP taxa in a sample to derive the weighted average bias-corrected LIFE values as:

$$\text{Bias-corrected LIFE} = (1-P_{miss})\ \text{Observed LIFE} + P_{miss}\ \text{'LIFE of missed taxa'} \qquad \text{(Eqn 4.2)}$$

Applying this predictive equation (Eqn 4.2) to all 429 audited samples by using a value of 6.55 for the 'LIFE of missed taxa' gives a reasonably accurate prediction of the actual bias in LIFE observed for the audited samples in relation to the observed (pre-audit) LIFE value (Table 24).

Table 24. Mean of actual and predicted (from Eqn 4.2 with LIFE of missed taxa set to 6.55) values of bias in LIFE in relation to observed (pre-audited) sample LIFE values (n = 426 samples)

Observed (pre-audit) LIFE		Mean bias in LIFE	
	Samples	Actual	Predicted from Equation 4.2
<6	29	0.102	0.070
6 .0- 6.5	63	-0.006	0.015
6.5 - 7.0	86	-0.020	-0.019
7.0 - 7.5	112	-0.021	-0.040
7.5 - 8.0	110	-0.085	-0.068
8.0 - 9.0	26	-0.178	-0.132

Recommendation

In RICT, the bias-corrected simulated value of LIFE should be estimated by:

Bias-corrected LIFE = { (1-P_{miss}) x Observed LIFE } + { P_{miss} x 'LIFE of missed taxa' }

where P_{miss} = M / (observed BMWP NTAXA + M)

M = simulated bias (M) in BMWP NTAXA

and 'LIFE of missed taxa' is simulated by a Normal distribution with a mean of 6.55
and standard deviation of 1.8 divided by the square root of M

4.3 Family-level PSI bias

The index PSI (Proportion of Sediment-sensitive Invertebrates) measures the abundance-weighted proportional frequency of taxa which are sensitive to fine sediment deposition (Extence *et al* 2013). PSI is defined as:

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

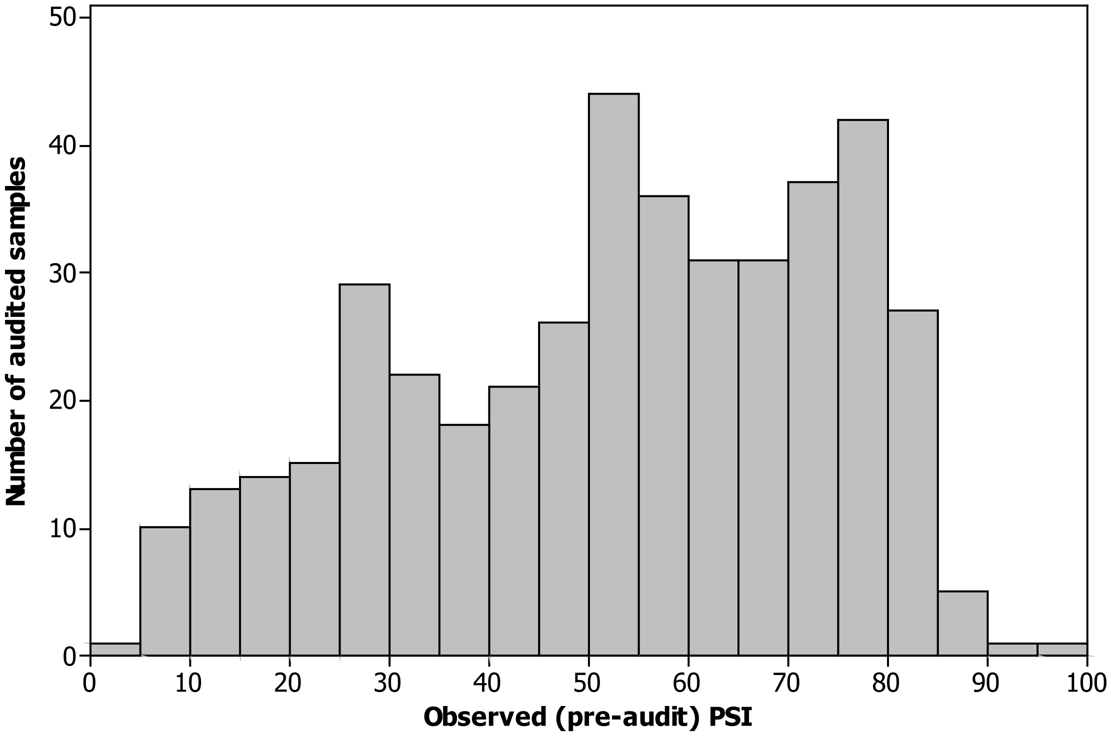
where the Sensitivity Group (A-D) of each family-level taxon and the Sediment Sensitivity scores (Ss) for taxa in each Sensitivity Group (A-D) in each $\log_{10}$ abundance category (1 - 4+) are given in Appendix 2.

These analyses are all based on the family-level abundance-weighted form of PSI (hereafter referred to as PSI), as this is what is intended for use within the new RICT software.

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. PSI bias for a particular sample equals the audit-corrected (post-audit) sample PSI value minus the pre-audit ‘observed’ PSI value for the sample.

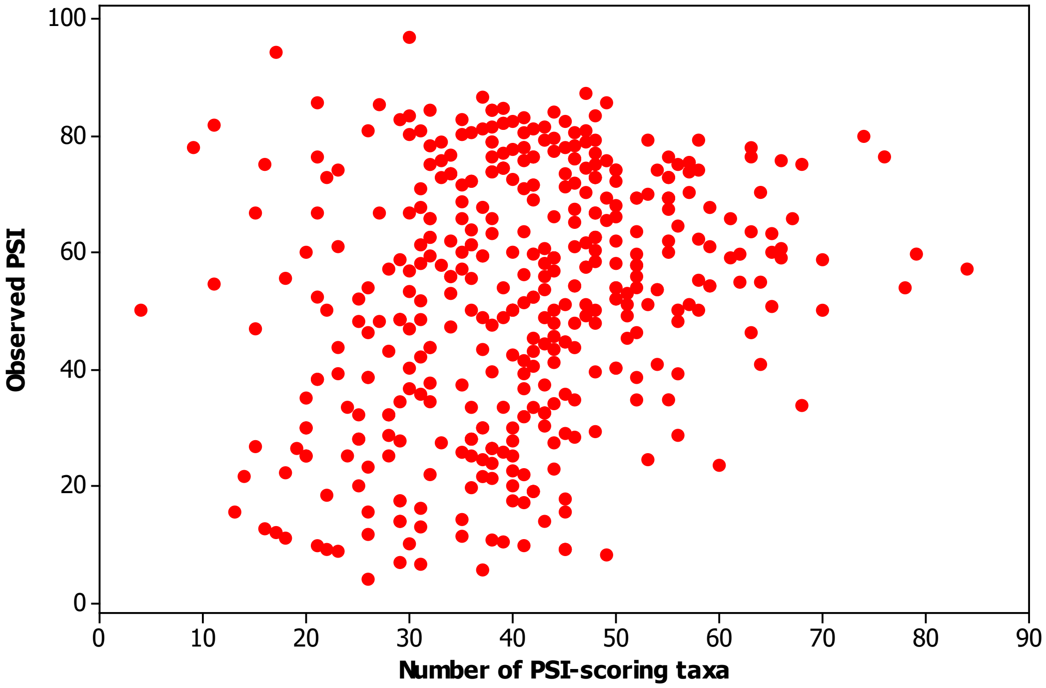
The (pre-audit) observed value of PSI amongst the 427 audited sample dataset varied from 3.8 to 96.7 (which encompasses almost the complete potential PSI range 0-100), with an inter-quartile range (i.e. middle 50%) of 35.5 to 71.4 and with a mean of 52.5 and median of 54.8 (Figure 26).

Figure 26. Frequency histogram of the observed (pre-audited) values of PSI amongst the 427 audited samples dataset.



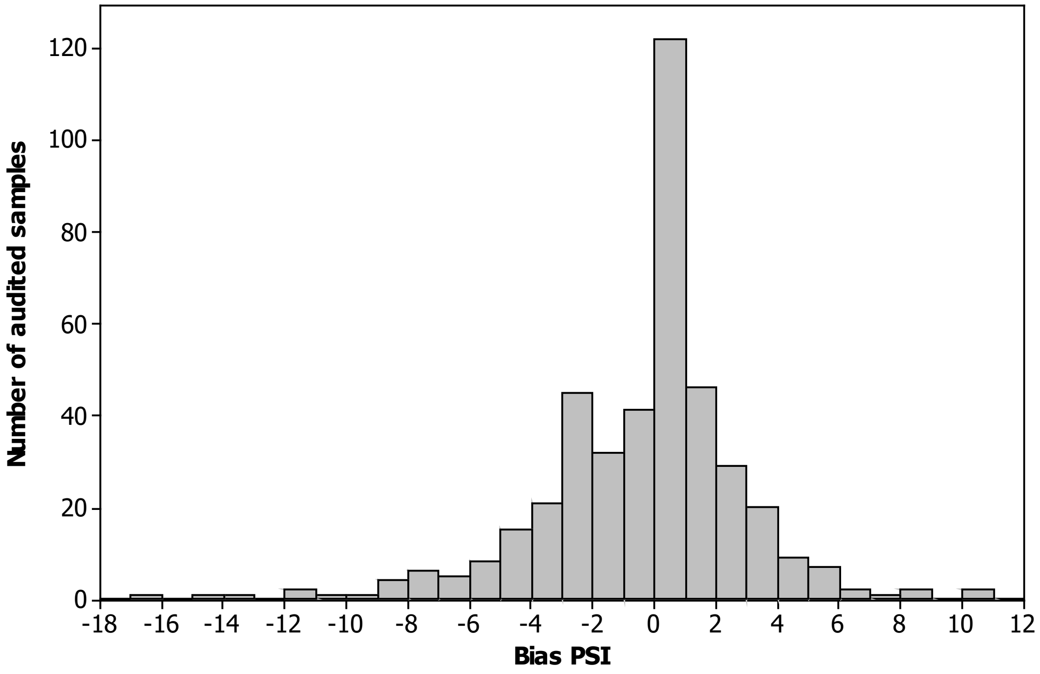
The observed PSI has only a weak positive correlation (0.27) with the total number of PSI-scoring taxa present (i.e. groups A+B+C+D in Appendix 2) (Figure 27).

Figure 27. Relationship between observed PSI and the number of PSI-scoring taxa recorded present in a sample.



The bias in PSI varied from -16.67 up to +10.91, with an inter-quartile range (i.e. middle 50%) of -2.08 to +1.26 and with a mean of -0.47 and median of exactly 0.00 (Figure 28). Overall, after audit correction, 20% sample PSI values stayed the same, 43% of went down and 37% increased. Thus overall, the bias adjustment (post-audit minus pre- audit PSI) is slightly more likely than not to be negative. However, the bias may vary with the taxonomic richness of the sample or with its observed PSI value and this is investigated below.

Figure 28. Frequency histogram of the bias (post- minus pre- audited values) in PSI amongst 427 audited samples.



4.3.1 Relationship between bias in PSI-scoring NTAXA and BMWP NTAXA bias

The ‘bias in PSI-scoring NTAXA’ for a sample is defined to be the number of taxa with PSI scores in the post-audit sample minus the number of taxa with PSI scores in the observed (pre-audit) sample.

For the PSI index, the maximum bias in estimating the number of PSI-scoring taxa present was 6, but the median was one taxa and the average bias amongst the 427 samples analysed was 1.22 (similar to the mean BMWP NTAXA bias of 1.15). One sample had three more PSI taxa recorded than actually existed in the audited sample, one sample had two more and 14 samples had one more PSI taxa recorded than existed in the audited sample (Table 25). Thus only 4% of samples actually had higher observed sample values of the number of PSI-scoring taxa present than they should have. A further 30% of samples had the correct number of PSI-scoring taxa recorded. However, for the majority (64%) of samples, the recorded number of PSI-scoring taxa present was less than the audit-corrected number (Table 25).

For 62% of all samples the bias in number of PSI-scoring taxa was the same as in BMWP NTAXA. In 19% of samples the under-estimation of number of PSI-scoring taxa was one greater than the under-estimation of BMWP NTAXA, whereas for 15% of samples it was one less (Table 25). The correlation between the individual sample bias in number of LIFE-scoring taxa present and individual sample bias in BMWP NTAXA was high at 0.85 amongst all 427 samples. The best fit straight line through the origin regression relationship had a slope of 0.991, explaining 71% of the total variation.

This suggests that, on average, the under-estimation of the number of PSI-scoring taxa is roughly the same as that of the number BMWP taxa.

Table 25. Two-way table summarising number of samples with each level of bias in BMWP NTAXA and bias in PSI-scoring NTAXA (equal biases are highlighted).

		Bias in PSI-scoring NTAXA											
		-3	-2	-1	0	1	2	3	4	5	6	Total	%
Bias in BMWP NTAXA	-2	1	0	1	0	0	0	0	0	0	0	2	0.5
	-1	0	1	6	6	0	0	0	0	0	0	13	3.0
	0	0	0	6	100	22	2	0	0	0	0	130	30.4
	1	0	0	1	20	92	22	5	0	0	0	140	32.8
	2	0	0	0	2	16	49	14	3	1	0	85	19.9
	3	0	0	0	0	0	15	10	11	1	0	37	8.7
	4	0	0	0	0	0	3	4	6	1	0	14	3.3
	5	0	0	0	0	0	0	0	0	2	4	6	1.4
	Total	1	1	14	128	130	91	33	20	5	4	427	
	%	0.2	0.2	3.3	30.0	30.4	21.3	7.7	4.7	1.2	0.9		100

4.3.2 Factors affecting PSI bias values

The sample bias in observed PSI does not seem to vary systematically with sample taxonomic richness as represented by the observed (pre-audit) sample number of BMWP taxa. Although there was some tendency for the percentage of positive biases to decrease with the number of BMWP taxa, the Spearman rank correlation between PSI Bias and observed BMWP NTAXA was -0.01 (Table 26). Thus samples with more taxa recorded do not tend to have higher or larger sized biases in observed PSI values.

Table 26. Bias in PSI in relation to observed (pre-audited) sample BMWP NTAXA.

Observed (pre-audit) BMWP NTAXA		Bias in PSI						
	Samples	Mean	SD	Min	Median	Max	% negative	% positive
≤10	26	-1.33	5.84	-16.67	0.00	8.55	38%	42%
11-15	57	-0.20	3.74	-10.81	0.00	10.91	47%	32%
16-20	113	-0.30	3.29	-14.91	0.00	7.22	35%	44%
21-25	154	-0.38	2.69	-8.44	0.00	6.50	46%	35%
26-30	52	-1.14	3.00	-11.35	-0.25	4.54	50%	33%
31-36	22	-0.12	1.40	-2.94	-0.07	3.84	50%	27%
Overall	424	-0.47	3.27	-16.67	0.00	10.91	43%	37%

However, the sample bias in observed PSI seems to decrease slightly with the observed (pre-audit) value of PSI (Table 27). Amongst the 40 audited samples for which the observed PSI was less than 20 the bias was positive in 47% of samples with a mean bias of +1.36. However, the mean bias and the percentage of positive biases decreases with increasing observed PSI, such that for the 32 samples with high (>80) observed PSI, only 28% of PSI biases were positive and the average bias was -1.74 (Table 27). However, the trend is not very strong or large as the median bias in PSI is zero for all class levels of pre-audit PSI and the changes in mean PSI bias are small relatively to the overall range of observed PSI values. On average around 20% of PSI bias values are zero.

Table 27. Bias in PSI in relation to observed (pre-audited) sample PSI.

Observed (pre-audit) PSI		Bias in PSI							
	Samples	Mean	SD	Min	Median	Max	% negative	% zero	% positive
0-20	40	1.36	2.48	-2.48	0.00	8.73	20%	33%	47%
20-40	85	0.22	2.81	-6.28	0.00	10.91	42%	20%	38%
40-60	129	-0.54	3.48	-16.67	0.00	10.25	47%	13%	40%
60-80	138	-1.08	3.15	-13.33	0.00	5.27	48%	19%	33%
80-100	32	-1.74	3.68	-14.91	0.00	2.05	44%	28%	28%
Overall	424	-0.47	3.27	-16.67	0.00	10.91	43%	20%	37%

Thus bias in PSI shows some tendency to decrease with the observed (pre-audited) sample PSI values but not with observed sample taxonomic richness (Figure 29).

Figure 29. Boxplot of bias in PSI in relation to the observed sample PSI value and the observed BMWP NTAXA (dashed line denotes zero bias).

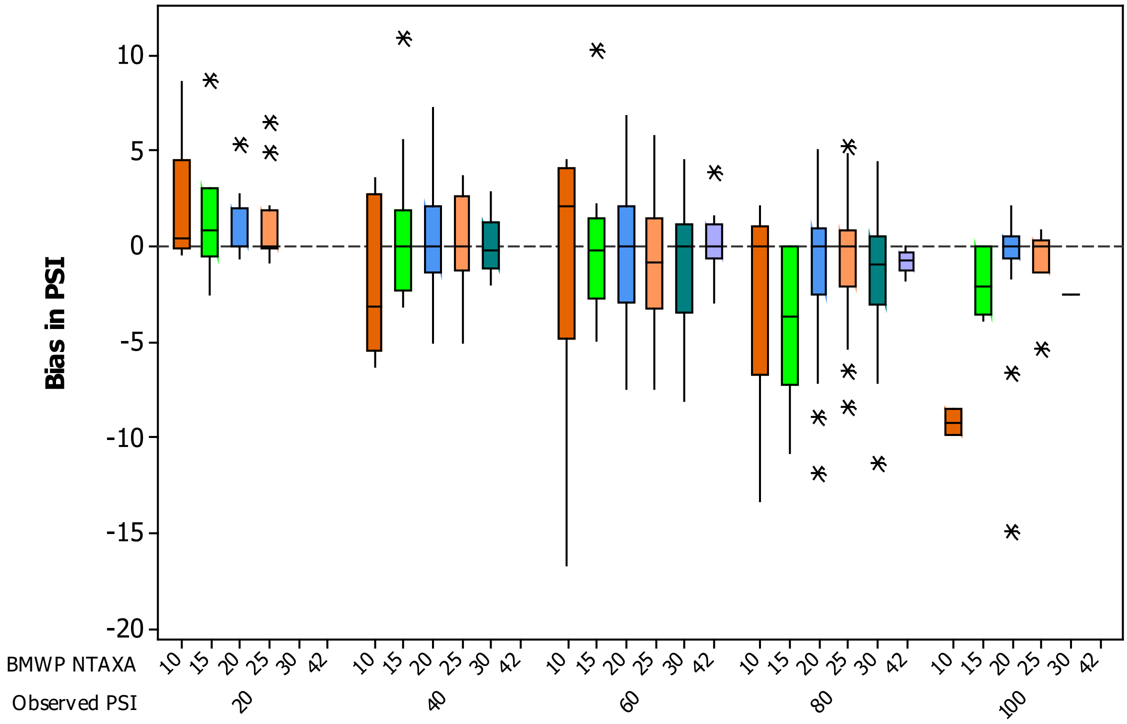
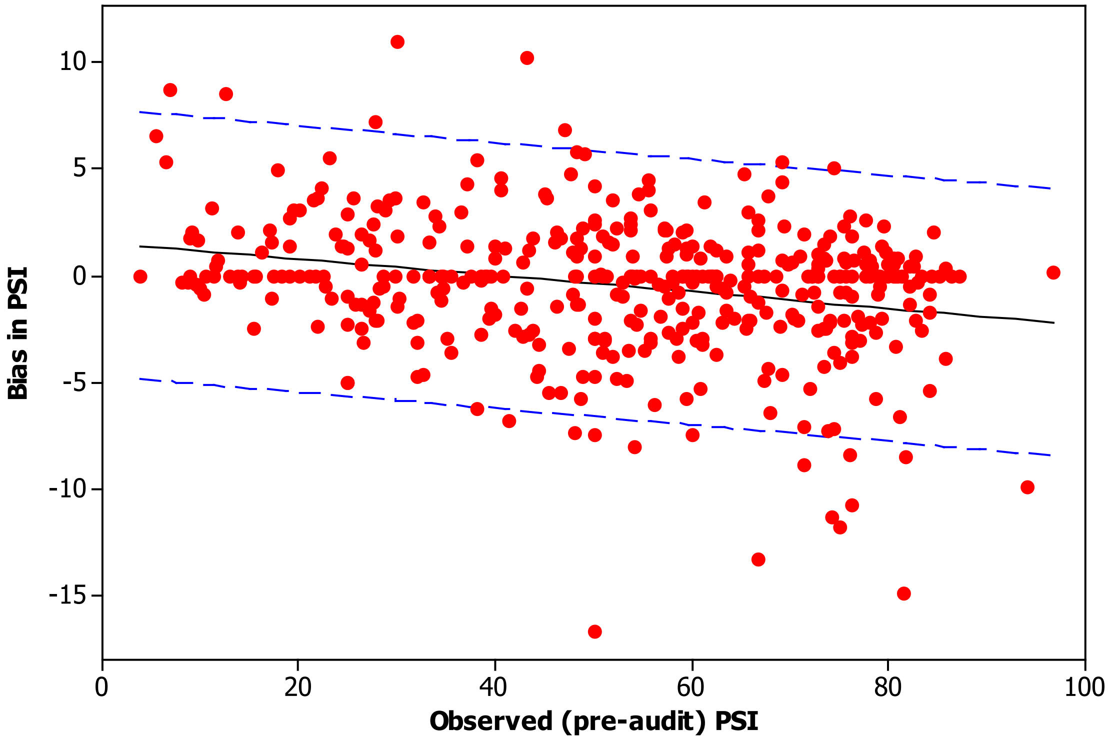


Figure 30. Linear regression relationship between bias in PSI and observed (pre-audit) PSI (fitted line (solid) and fitted 95% confidence interval of individual sample values (blue dashed).



4.3.3 Simulating bias in PSI in relation to observed (pre-audit) PSI value

The bias in PSI shows some tendency to be positive for low (<20) observed (pre-audit) values of PSI and to have a few large negative bias values when the observed PSI was high (Figure 30).

The statistically significant ($p<0.001$) fitted linear regression relationship (SE in brackets) is:

$$\text{Bias in PSI} = \underset{(0.408)}{1.559} - \underset{(0.0072)}{0.0387}\ \text{Observed PSI} \qquad \text{(Eqn 5.2)}$$

Although this only explains 6% of the total variation is individual sample PSI bias values, it does predict the observed change in average PSI bias from around +1.5 to +1.0 for small observed PSI down to -1.5 to -2.0 when observed PSI is large. The regression residual standard deviation is 3.16 and the regression 95% prediction confidence limits for individual samples (which are based on assuming a residual normal distribution), as shown in Figure 30 encompass the broad pattern of observed biases. Although there is some tendency for the prediction limits to miss too many of the relatively few large negative biases which occur when observed PSI is high, this regression relationship is an improvement on otherwise assuming a constant bias distribution regardless of observed PSI levels and provides the best currently available means of predicting and simulating sample biases in PSI values.

However, we have observed that across the range of observed PSI values around 20% of samples have PSI biases of zero. Substantial proportions of zero residuals are not allowed for in a normal regression relationship. Therefore the 87 audited samples with PSI biases of zero were excluded in a re-fitting of the linear regression relationship to give:

$$\text{Bias in PSI} = \underset{(0.515)}{2.067} - \underset{(0.0091)}{0.0504}\ \text{Observed PSI} \qquad \text{(Eqn 5.3)}$$

This statistically significant regression ($p<0.001$) explained 8% of the variation amongst the 340 non-zero PSI bias samples and gave, as expected a slightly larger residual SD of 3.50. This will give predicted average non-zero biases of around +1.5 for observed PSI of 10 and of -2.5 for observed PSI of 90.

Recommendation

PSI bias decreases with the observed (pre-audit) sample PSI value.

In RICT, simulate the bias in a sample PSI value as follows:

First set PSI bias to zero with probability 0.2

Non-zero biases are simulated using a Normal distribution with a mean of:

$$\text{Mean Bias in PSI} = 2.067 - 0.0504 \times (\text{Observed PSI value})$$

and a standard deviation (SD) of 3.5.

Any simulated bias-corrected values of PSI greater than 100 should be reset to 100 and any simulated bias-corrected PSI values less than zero should be reset to zero.

5. Deriving RIVPACS predicted and adjusted E values & calibration to EQR

5.1 RIVPACS predictions of Expected (E) values of family level LIFE & PSI

The LIFE and PSI index were designed to help assess the biological impacts of flow-related and fine sediment related stresses. The RIVPACS approach for river site assessment is to compare the observed sample biota and observed (O) values of biotic indices with the expected (E) values based on the site's physical characteristics as represented by a suite of environmental variables. If some of the physical characteristics of a site and thus the environmental variables' values for a site have already been altered by the time of their measurement at a site as a result of anthropogenic stresses operating at the site, then the predictions of expected index values for that site may be incorrect for that site in its 'natural' reference condition. For example increased levels of fine sediment deposition (or lack of occasional higher flows to wash out fine sediments) will lead to altered substrate particle size composition and altered RIVPACS model predictions of the expected values of biotic indices. This will be especially problematic for biotic indices such as LIFE and PSI whose O/E ratios are intended to help identify biotic impacts of such flow and sediment related stresses.

With the current environmental information available for all reference sites, the predictions of the site- and season- specific predictions for the two indices LIFE and PSI are best made using an alternative RIVPACS predictive model based on the same standard RIVPACS IV set of reference sites as used for predicting expected values of the BMWP and WHPT indices, but using a suite of RIVPACS environmental variables which exclude those variables measured in the field. Specifically, the reduced model excludes the three predictor variables of stream width, stream depth and substratum composition (i.e. mean substratum composition in phi units derived from percentage cover of the four substratum categories: boulders and cobbles; pebbles and gravel; sand; silt and clay).

This model was the same as the trial model denoted '15(2)' developed earlier by Clarke *et al* (2011) in a previous SNIFFER project to assess the potential to make RIVPACS models and predictions of expected index values without using any time-varying environmental predictor variables measured in the field. Although such models may give slightly poorer statistical fits to the biotic indices for the model's reference sites, they are (less) biased and more "fit for purpose" for predictions and WFD status assessments for sites which may have altered flow regimes and/or fine sediment related stresses.

All predictions of expected values and analyses of O/E (EQI) values of both PSI and LIFE reported here have been based on this reduced time-invariant model 15(2).

The predicted values for any site are based on the probability of that site belonging to each of the 43 RIVPACS model site End-groups and the means of the observed index values of reference sites in each of these 43 End-groups. Predictions of expected values arise from, and therefore must lie within, this range of End-group mean observed values.

The end-group means for the LIFE and PSI indices are shown for each of the three seasons in Table 28. For family-level LIFE, the End-group means vary from a minimum of 5.58 (for End-group 43 in summer) to 8.05 (for End-group 14 in autumn).

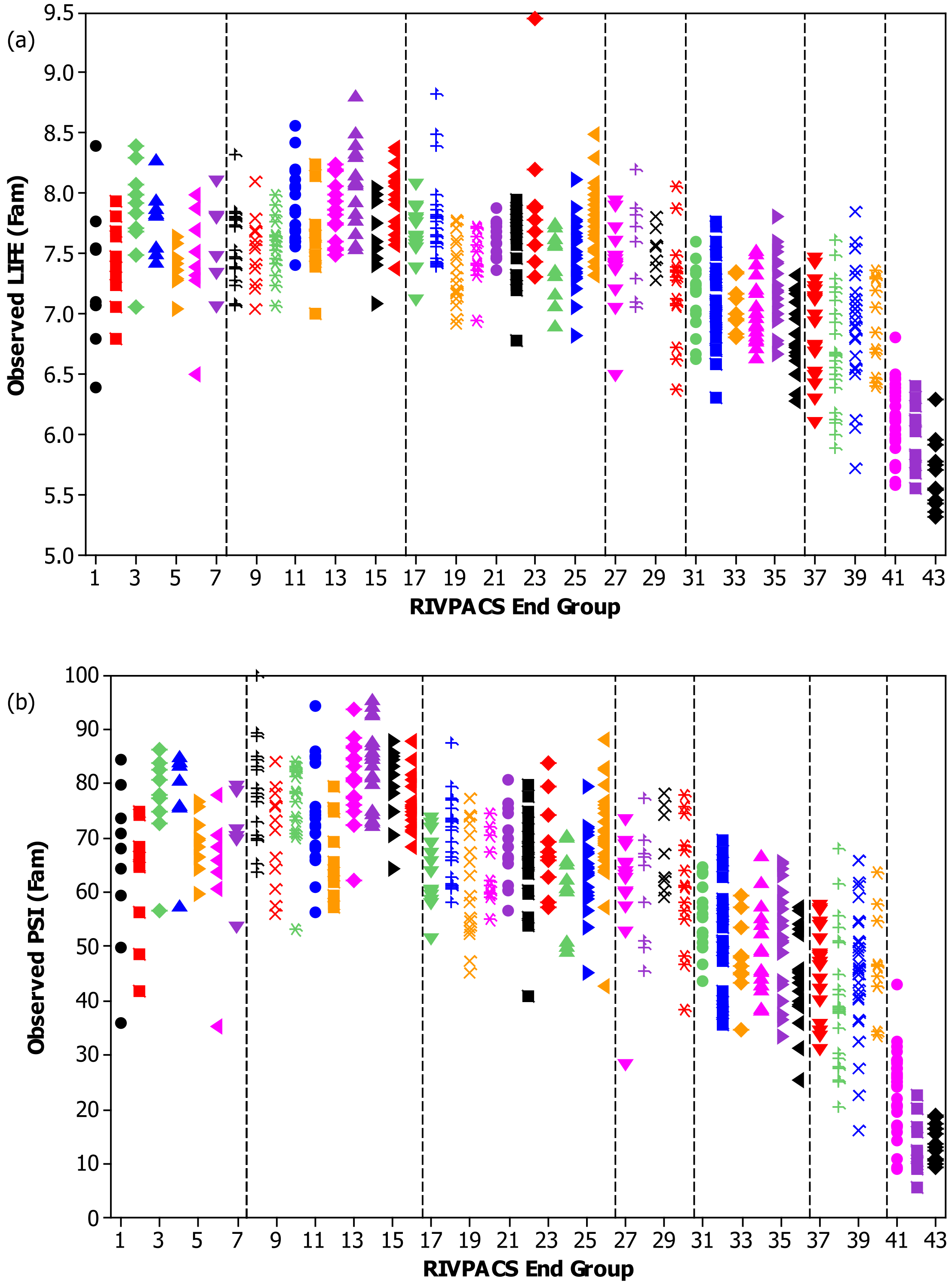
For family-level PSI, the End-group means vary from a minimum of 12.91 (for End-group 43 in spring and summer) to 83.29 (for End-group 14 in autumn) - a wide order of magnitude (Table 28). It is especially noticeable that the PSI End-group means for End-groups 41-43) are much lower than for any other groups of sites. This super-group (41-43) of sites from the RIVPACS TWINSPAN classification (see section 1 of this report) are mostly largish lowland fine-sediment sites from SE England, especially East Anglia (Figure 1 and Figure 2).

The distribution of individual sample observed LIFE and PSI values by reference site End-group are shown in Figure 31 for autumn samples, highlighting the pattern of variation with site End-group. Very similar patterns exist for the other two seasons.

Table 28. Mean of the observed family-level LIFE and PSI values for the 685 GB RIVPACS IV reference sites by End-group (1-43) and season.

End-group	Mean Observed LIFE (Fam)			Mean Observed PSI (Fam)		
	Spring	Summer	Autumn	Spring	Summer	Autumn
1	7.35	7.05	7.30	69.52	60.26	65.26
2	7.45	7.34	7.44	66.51	59.97	63.28
3	7.92	7.89	7.87	78.14	78.74	76.92
4	8.00	7.76	7.78	82.05	80.54	78.69
5	7.30	7.50	7.40	66.94	69.17	69.06
6	7.62	7.53	7.45	67.71	66.51	64.19
7	7.81	7.38	7.62	74.97	61.46	70.88
8	7.53	7.52	7.58	72.77	71.77	78.71
9	7.41	7.57	7.53	67.50	68.10	70.36
10	7.84	7.69	7.58	77.36	73.12	76.24
11	7.94	7.94	7.87	76.40	75.67	74.26
12	7.82	7.84	7.62	72.06	71.83	67.34
13	7.91	7.82	7.84	79.44	75.73	80.28
14	8.01	7.94	8.05	81.19	79.93	83.29
15	7.88	7.75	7.67	81.15	80.77	79.68
16	7.90	7.85	7.93	77.34	75.50	77.23
17	7.79	7.97	7.66	68.82	70.16	65.14
18	7.90	7.96	7.79	73.25	73.76	69.47
19	7.68	7.77	7.38	68.51	69.58	62.35
20	7.52	7.50	7.47	64.93	62.27	64.22
21	7.81	7.81	7.61	71.89	68.98	68.72
22	7.78	7.98	7.58	70.97	75.00	65.19
23	7.83	7.94	7.91	70.41	67.85	68.55
24	7.52	7.57	7.42	64.88	64.12	60.41
25	7.72	7.65	7.49	69.19	66.13	64.13
26	7.88	7.82	7.80	75.14	71.95	70.62
27	7.60	7.53	7.44	67.96	63.23	62.04
28	7.73	7.54	7.53	67.40	62.48	61.19
29	7.74	7.82	7.55	70.77	72.98	67.08
30	7.34	7.13	7.23	62.80	57.43	61.07
31	7.24	7.20	7.13	56.20	56.99	55.21
32	7.35	7.36	7.10	58.01	55.72	51.65
33	7.15	7.23	7.07	50.68	49.63	48.11
34	7.23	7.22	7.03	53.43	52.88	48.30
35	7.25	7.31	7.17	53.69	55.00	51.39
36	6.83	6.86	6.82	44.30	45.29	43.60
37	7.07	6.87	6.92	49.55	44.95	45.87
38	6.86	6.72	6.72	47.10	42.18	40.00
39	7.09	7.01	6.95	49.08	47.42	45.31
40	7.16	7.00	6.89	55.15	46.76	46.79
41	6.24	6.17	6.12	26.20	24.25	23.50
42	6.18	6.02	6.00	18.54	16.05	13.65
43	5.77	5.58	5.65	12.91	12.91	14.04

Figure 31. Observed (autumn sample) values of the family-level (a) LIFE and (b) PSI indices for the 685 GB RIVPACS IV reference sites, grouped by their end group (1-43). Vertical dashed lines separate the seven super-groups described in Table 1.

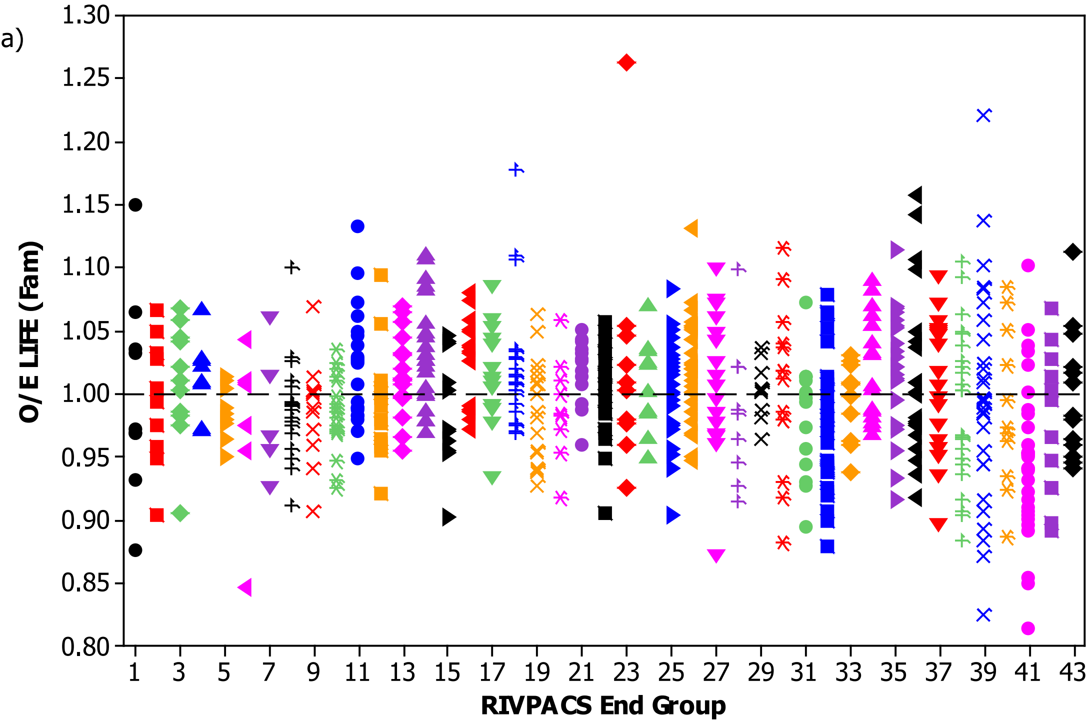


Once the observed (O) index values are standardised by their RIVPACS predictive model site- and season-specific expected (E) values as O/E ratios (EQI), the stream type-specific variation is largely removed and the distribution of EQI values is broadly centred around a mean of one for all end – groups, as intended and desired (Figure 32). However, considerable variation still exists due to natural sampling variation, but also to inevitable inefficiencies in the predictive model and variation in the quality of the reference sites at the time of sampling.

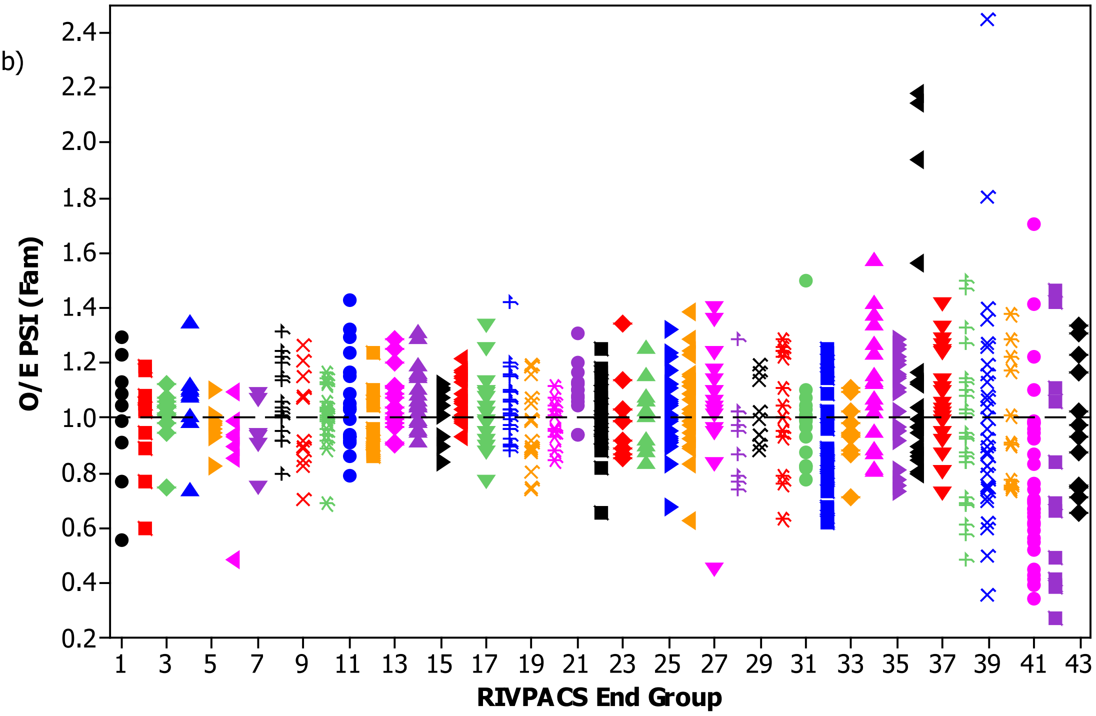
The variation in LIFE O/E values is largely constrained within a range of 0.9 to 1.1 across all type of site (i.e. End-groups). However, the O/E values for the new PSI index have a much greater inherent range, with 2-4 sites per season having O/E values between 2.0 and 2.55. The percentile distribution of O/E values for LIFE and PSI is assessed further in section 5.5.

Figure 32. Distribution of O/E (EQI) values of the family-level (a) LIFE and (b) PSI indices for the 685 GB RIVPACS IV reference sites, grouped by their end group (1-43). Dashed lines indicate O/E values of unity (1.0). Autumn sample values shown for illustration.

(a)



(b)



5.2 Adjustment of predicted E values based on RIVPACS Reference site quality

5.2.1 Background rationale and reference sites status assessment scores

During the development of RIVPACS, the river sites selected as reference sites were considered to be amongst the best available for each physical type of site. However, the environmental quality of RIVPACS reference sites, when they were sampled, has subsequently been judged to have varied from the top-end of WFD high status to occasionally beyond the lower end of good status. The distribution of quality was not even, but varied between stream types; good quality sites predominated in lowland England and a greater proportion of high quality sites were in upland Britain.

As RIVPACS has evolved from RIVPACS I to the current RIVPACS IV model, several original reference sites have subsequently been judged to be unsuitable and removed. However, RIVPACS model predictions of expected index values for a site will still depend on the status of the environmentally-similar reference sites actively involved in predictions for the site. In the current UK classification scheme, an EQI value of 1.0 is used to set the “high/good” boundary; a river site has to achieve biological index values equal to or in excess of those predicted by RIVPACS in order to be classified as the highest status. This means that, in effect, that roughly half of the RIVPACS references sites are assumed to be of “high” ecological status and roughly half of “good” status. The problem is that this assumption has been applied across the board for all types of river sites in the UK. Methods and modifications are required to allow for alterations in the predicted index values to reflect a standard level of status (i.e. high/good boundary) and these need to be integrated within the RICT software tool.

To help address this problem, the Centre for Ecology and Hydrology (CEH) collated and assessed the available pressure data on each site (SNIFFER project WFD46, Davy-Bowker *et al* 2007(a), 2007(b)). Subsequently UK agency regional biologists provided an assessment score (1-6) of the perceived WFD ecological status class of each reference site (1 = top of high, 2 = middle of high, 3 = high/good boundary, 4 = middle of good, 5 = good/moderate boundary, 6 = worse).

As part of SNIFFER project WFD72B (Clarke & Davy-Bowker 2006), Ralph Clarke developed a general statistical model (model M4), procedures and algorithms to make the adjustments and applied these procedures to estimate adjustment factors (A_1, A_2, A_3, A_4, A_5) for the existing BMWP indices. Through a subsequent project WFD72C (Davy-Bowker *et al* 2008), these adjustment algorithms and adjustment factors for the BMWP indices were then incorporated into the first version of the RICT software in 2008-09.

Subsequently, John Murray-Bligh (Environment Agency) was involved in an exercise to inter-calibrate national assessment systems to improve their standardisation in terms of the ecological quality represented by each WFD status class and especially the good/moderate boundary. Within this exercise, it became clear that Agency biologists’ assessments of the quality of RIVPACS reference sites were not in accord with the methods used for WFD inter-calibration. For the WFD, GB RIVPACS reference sites were instead screened against chemical and land-use criteria defined by a Central-Baltic and North Europe Geographical Intercalibration Group (GIG).

Sites that transgressed these GIG criteria were not considered to be in High Status, so any that had been assigned biologists’ assessment scores of 1 or 2 (which represent the top and mid-range of high status) were demoted to 3 (which represents the high-good boundary). Within the GB RIVPACS IV model of 685 reference sites, 10 of 120 sites with biologists scores of 1 and 112 of 242 sites with scores of 2 were downgraded to WFD definitive scores of 3 (high/good boundary). Within the NI RIVPACS IV model of 108 reference sites, 6 of 12 sites with biologists scores of 1 and 11 of 13 sites with scores of 2 were downgraded to WFD definitive scores of 3 (high/good boundary).

Appendix 4 gives the number of reference sites with each WFD assessment score in each RIVPACS IV End-group, separately for the GB and Northern Ireland models.

As part of the implementation of LIFE and PSI within RICT, it was crucial that EQI and EQRs involving these indices incorporate adjustment factors for the expected values within the RICT software. Adjusted expected values of LIFE and PSI were also needed within this project to obtain percentile distribution limits of the adjusted EQI values of the RIVPACS reference sites to guide status class limit setting. However, this work was not part of the current project specification. Having identified these requirements, Ralph Clarke subsequently agreed (without extra funding) to derive adjustment factors for LIFE and PSI and to develop his own interim code (prior to the functionality being available using RICT) to provide adjusted E values of the LIFE and PSI indices for the reference sites.

5.2.2 Statistical model, data and estimates of adjustment factors

It was concluded in project WFD72C (Davy-Bowker *et al* 2008) that the best statistical model to estimate the adjustment factors associated with each level (1-6) of the biologist assessment scores for the reference sites was model M4 (or more specifically M4*):

$$\log_{10} O_{isjk} = \log_{10} M_{is} + a_j + e_{isjk} \qquad \text{(M4*)}$$

where

O_{isjk} = Observed index value for the k^{th} site with assessment score j in site group i in season s

M_{is} = term for average index value for TWINSPAN site group i in season s

a_j = effect of assessment score j on $\log_{10}$ index values (re-scaled to give $a_3 = 0$).

e_{isjk} = residual value for the k^{th} site with assessment score j in group i in season s

Model (M4) is a multiplicative model, in that by analysing the observed index values on their logarithmic scale, the effect of a particular assessment score j on the observed index value for a site and season is assumed to be a constant multiple (i.e. proportion) of the value expected for that type of site (i.e. end group) and season if it was of high/good boundary quality (i.e. assessment score of 3). Model (M4) is also non-linear in that it does not assume that the (proportional) difference in index values between sites of similar types with scores of, say 1 and 3, is the same as the difference between sites with score of say 3 and 5. The non-linearity of effects of score was shown to be statistically significant in WFD72B (Clarke & Davy-Bowker 2006) for both BMWP TAXA and ASPT. This general non-linear model was therefore used in the current project to estimate adjustment factors for all of the indices.

In project WFD72B, it was concluded that there were insufficient sites with an assessment score of 6 (10 in total) for the estimates of parameter a_6 to be statistically significant or reliable. It was recommended that, in deriving adjustments to expected index values for test sites, those few reference sites given assessment scores of 6 by Agency biologist's should be treated as having assessment scores of 5. Note however, that the scores of 6 are retained in the actual fitting of model (M4*); the estimate of parameter a_6 is just not used in the subsequent algorithms to derived adjustments to expected index values for test sites.

In project WFD72C, it was also concluded that it was best to estimate a single set of adjustment parameters to be applied to test samples for any RIVPACS season option (based on analysing all possible season options samples together). For the BMWP indices, this was the combined observed sample data from all single season, two- and three- season combined samples.

However, the abundance-weighted WHPT, LIFE and PSI indices were designed and evaluated using single season samples; their values for combined season samples have not yet been assessed. Therefore the estimates of adjustment parameters in model (M4) for these indices were fitted using all the single season samples values for all 793 UK (685 GB and 108 NI) reference sites.

The fitted regression model (M4*) can be re-expressed as:

$$O_{isj} = M_{is} A_j \qquad \text{where} \quad A_j = 10^{a_j} \qquad \text{(M4**)}$$

The estimates the adjustment parameters (a_1 – a_5) obtained by fitting model (M4*) to the observed sample values of each biotic index are given in Table 29. The estimates of the multiplicative adjustment factors (A_1 – A_5) for RIVPACS expected values derived from UK-wide versions of model (M4*) (and thus M4**) are given for each index in Table 30.

The estimates of the multiplicative adjustment factors (A_1 – A_5) used to adjust expected index values are similar, but not identical, for the original BMWP and revised abundance-weighted WHPT forms of score, NTAXA and ASPT. Having an assessment score of 5, compared to reference sites of a similar type (i.e. end-group) with assessment scores of 3, has the effect of reducing the expected number of WHPT taxa by 8.5% (i.e. by 100(1-0.915)) (Table 30).

Table 29. Estimates of adjustment parameters (a_1 – a_5) for the effects of WFD assessment score (1-5) in model (M4*) for each biotic index based on using all single season samples for the 793 UK-wide reference sites (BMWP index parameters based on all possible single and combined season samples); *p* = model test probability value for effect of score.

	Original BMWP		
	BMWP Score	BMWP NTAXA	BMWP ASPT
p	<0.001	<0.001	<0.001
a_1	-0.01533 ± 0.00458	-0.01356 ± 0.00395	-0.00177 ± 0.00145
a_2	-0.00693 ± 0.00412	-0.00791 ± 0.00355	0.00098 ± 0.00131
a_3	0	0	0
a_4	-0.03079 ± 0.00403	-0.02025 ± 0.00347	-0.01053 ± 0.00128
a_5	-0.06424 ± 0.00653	-0.04158 ± 0.00562	-0.02267 ± 0.00207
	Abundance-weighted WHPT		
	WHPT Score	WHPT NTAXA	WHPT ASPT
p	<0.001	0.001	<0.001
a_1	-0.00962 ± 0.00823	-0.00912 ± 0.00724	-0.00050 ± 0.00274
a_2	-0.00720 ± 0.00739	-0.00645 ± 0.00651	-0.00078 ± 0.00246
a_3	0	0	0
a_4	-0.02867 ± 0.00723	-0.01803 ± 0.00636	-0.01065 ± 0.00240
a_5	-0.06657 ± 0.01171	-0.03872 ± 0.01031	-0.02787 ± 0.00389
	LIFE	PSI	
p	0.004	0.005	
a_1	-0.00285 ± 0.00166	-0.00306 ± 0.00675	
a_2	-0.00158 ± 0.00149	-0.00804 ± 0.00607	
a_3	0	0	
a_4	-0.00260 ± 0.00146	-0.01702 ± 0.00593	
a_5	-0.00560 ± 0.00236	-0.05080 ± 0.00961	

Table 30. Estimates of adjustment parameters (A_1 – A_5) for the effects of assessment score (1-5) in model (M4* - equivalent to M4**) for each biotic index based on using every possible combination of single and multiple season samples for the 793 UK-wide reference sites (LIFE and PSI estimates based on all single season samples only), (Note: A_j = 10 to the power a_j , where a_j is as in Table 29).

	Index	Estimate of adjustment parameter				
		A_1	A_2	A_3	A_4	A_5
Original BMWP	Score	0.965	0.984	1.000	0.932	0.863
	NTAXA	0.969	0.982	1.000	0.954	0.909
	ASPT	0.996	1.002	1.000	0.976	0.949
Abundance-weighted WHPT	Score	0.978	0.984	1.000	0.936	0.858
	NTAXA	0.979	0.985	1.000	0.959	0.915
	ASPT	0.999	0.998	1.000	0.976	0.938
	LIFE	0.993	0.996	1.000	0.994	0.987
	PSI	0.993	0.982	1.000	0.962	0.890

Few of the reference sites were identified as specifically having any signs of either flow or fine sediment related stress. Therefore, reference sites with high assessment scores (i.e. poorer perceived quality) may not tend to have systematically different values of either LIFE or PSI compared to other sites. However, the “innocent” statistical fitting of model (M4*) for these index values did show statistically significant effects of assessment score.

The estimates suggest that the PSI index values were, on average, 0.7% and 1.8% lower for reference sites with assessment scores of 1 and 2 respectively compared to reference sites of the same type (i.e. RIVPACS end-group) with assessment scores of 3 (i.e. on the high/good boundary of quality); however these reductions were not statistically significant from zero (i.e. parameters a_1 and a_2 estimates were not greater than twice their standard errors in Table 29). The estimates suggest that the PSI index values were, on average, 3.8% and 11.0% lower for reference sites with assessment scores of 4 and 5 respectively compared to reference sites of the same type (i.e. RIVPACS end-group) with assessment scores of 3 (i.e. on the high/good boundary of quality) (Table 30). Thus PSI is on average lower in the poorest quality RIVPACS reference sites.

The model (M4*) fit for the LIFE index suggests the weakest association with assessment score and reference sites with assessment scores of 5 were, on average, only 1.3% lower than those with assessment scores of 3 (although this difference was statistically significant).

5.2.3 Procedures and algorithms for adjusting RIVPACS Expected values of indices

In model (M4*), the effect of a unit change in assessment score is not assumed to be constant across the range of assessment scores; however, a given assessment score is assumed to have a constant multiplicative effect on the observed index values for the reference sites. Specifically an assessment score of *j* with a value of A_j less than one is assumed on average to decrease the observed index values of reference sites by a factor of ($1-A_j$). Therefore the expected index values of any sites based on reference sites with a score of *j* are on average under-estimated by a factor ($1-A_j$). In such cases, the correction should therefore be to divide the RIVPACS expected values by the same factor A_j, which will increase the expected values when A_j is less than one.

As the RIVPACS predictions for real test sites are always based on site groups and sites with reference sites with more than one assessment score, the multiplicative adjustment factor for the raw RIVPACS predicted expected (E) values are based on the reciprocal of a weighted average of the factors A_j as follows:

Adjusted expected value E for a test site = $E_{adj} = E / (\sum_{j=1}^{6} R_j A_j)$ (M4a)

where $R_j = \sum_{i=1}^{g} P_i Q_{ij}$ = weighted proportion of the reference sites involved in the prediction with an assessment score of *j*.

and

P_i = RIVPACS probability test site belongs to TWINSPAN site group *i*
Q_{ij} = Proportion of reference sites in group *i* with assessment score *j*
g = number of TWINSPAN site groups

These estimates of adjustment parameters $A_1 - A_5$ were used in equation (M4a) to adjust the RIVPACS predicted expected (E) values of each WHPT, LIFE and PSI index for each of the GB RIVPACS reference sites for subsequent analysis in section 5.4 of the distribution of adjusted EQI and eventual EQR values for the 685 GB reference sites. This was done using Ralph Clarke’s own code to derive RIVPACS expected (E) values, adjusted E values (E_{adj}) and thus adjusted EQI values ($EQI_{adj} = O / E_{adj}$).

The adjustment calculations represented by equation (M4a) are already coded into the original version of the RICT software.

Recommendation

The estimates of adjustment parameters $A_1 - A_5$ in Table 30 should be added to the RICT parameter files and software code to allow adjusted expected values (E_{adj}) based on existing code for equation (M4a) to be derived for the abundance-weighted WHPT, LIFE and PSI indices for any user-supplied test site.

5.3 Calibration of EQI to EQR for LIFE and PSI (family level)

RIVPACS predictions must represent the same environmental quality to be comparable. When used to determine the WFD ecological quality status, the predictions must represent the WFD reference state (reference condition). To achieve this, a further adjustment factor is needed for each index to convert the adjusted EQI values to Ecological Quality Ratios (EQRs) where the Expected value for a site is effectively that expected for the site if it was in WFD-defined reference condition.

To achieve this for the PSI and LIFE (and WHPT) indices, we followed the same procedure as developed and used by John Murray-Bligh (EA) in his previous calibration of EQI to EQR values for the BMWP indices.

All RIVPACS reference sites which met the WFD Geographic Inter-calibration Group (GIG) definition of Reference Condition (see section 5.2.1) and had a biologists' assessment score of 1 or 2, retained their biologists scoring and were classed as being in WFD Reference Condition

The calibration from adjusted EQI (EQI_{adj}) to EQR for an index is:

$$EQR = Calib * EQI_{adj}$$

where

$$Calib = \frac{\text{Median } EQI_{adj} \text{ for all RIVPACS Reference sites}}{\text{Median } EQI_{adj} \text{ for all RIVPACS reference sites in WFD Reference Condition}} = \frac{Med_{All}}{Med_{RC}}$$

The values of the EQI to EQR calibration parameter 'Calib' were derived from the distribution of EQI values of the 685 GB references sites using all single season sample (spring, summer and autumn).

Expected values, adjusted expected values and thus adjusted EQI values were based on the RIVPACS model using the standard RIVPACS suite of environmental variables for the BMWP and WHPT indices, but using the pressure-insensitive reduced RIVPACS model, model 15(2), (excluding width, depth and bed composition) for the LIFE and PSI indices (see section 5.2.1).

The values of the Calib parameter for each index are given in Table 31.

Table 31. Calibration coefficients (Calib = Med_{All} / Med_{RC}) to convert EQI to EQR for each biotic index; Med_{All} and Med_{RC} are the median adjusted EQI amongst all reference sites and amonst those in WFD reference condition respectively.

Index	Med_{All}	Med_{RC}	$Calib_i$
BMWP NTAXA	0.976	0.958	1.019
BMWP ASPT	0.999	1.005	0.993
WHPT NTAXA	0.975	0.962	1.014
WHPT ASPT	0.999	1.008	0.992
LIFE			0.999
PSI			0.992

5.4 Direct adjustment of RIVPACS Expected values to Reference Condition

In the past, RIVPACS predictions of expected values and especially the derived EQI (O/E ratios) were adjusted to be WFD-compliant EQRs by first adjusting the raw expected (E) values to adjusted expected (E_{adj}) (as detailed in section 5.2 of this report) and then the adjusted EQIs ($EQI_{adj} = O / E_{adj}$) were calibrated to EQRs by a calibration factor (as detailed in section 5.3 of this report).

During our research and data analyses within this project, Ralph Clarke had the idea that it was possible to combine the two steps by directly adjusting the raw RIVPACS predicted expected (E) values to WFD reference condition expected values (E_{RC}) and then calculating WFD EQR values in one step as $EQR = O / E_{RC}$.

The idea was to use the same logic as used formerly to adjust expected values to a reference site with a quality equal to a Biologist site assessment score of 3, which is the high-good boundary. The difference is that the adjustment is now to the quality of RIVPACS reference sites judged to be in WFD reference condition, namely all RIVPACS reference sites assigned a Biologists' assessment score of 1 or 2.

The adjustment coefficients (a_1, a_2, a_3, a_4 and a_5) are estimated by fitting a variation of model (M4*), which we will call model M8* defined by

$$log_{10}\, O_{isjk} = log_{10}\, M_{is} + a_j + e_{isjk} \qquad \text{(M8*)}$$

where

O_{isjk} = Observed index value for the k^{th} site with assessment score *j* in site group *i* in season *s*
M_{is} = term for average index value for TWINSPAN site group *i* in season *s*
a_j = effect of assessment score *j* on $\log_{10}$ index values
e_{isjk} = residual value for the k^{th} site with assessment score *j* in group *i* in season *s*

but where coefficients a_1 and a_2 are forced equal to zero.

This is done statistically by fitting a model which only has a term for sites in each of the assessments scores 3,4, 5 and occasionally 6. This means that the coefficients a_3 to a_5 (and a_6) represent, after allowance for site end-group and season, the average deviations of observed index values of reference sites with assessments scores of 3, 4 and 5 respectively from the average index values of reference sites with assessments scores of 1 or 2.

The adjustment coefficients obtained by fitting this model (M8*) to each index are given in Table 32 and the back-transformed estimates ($A_j = 10^{a_j}$) to be used in future RICT software are given in Table 33.

The adjustment procedures and algorithms of section 5.2.3 of this report can then be used directly to adjust the raw RIVPACS expected values to WFD reference condition expected values (E_{RC}).

For the PSI index, the adjustment for reference sites with an assessment score of 3 (high-good boundary) is actually slightly above one (1.014), whereas for poorer quality reference sites with scores of 4 or 5, the estimated reduction in PSI values is 2.5% and 9.8% (Table 33).

For the LIFE index, the adjustment factor is negligibly different from one for all site assessment scores 1-5, suggesting no detectable consistent change in LIFE amongst this set of reference sites with their perceived biological quality level at the time of sampling for RIVPACS.

Table 32. Estimates of adjustment parameters (*a1 – a5*) for the effects of WFD assessment score (1-5) in model (M8*) for each biotic index based on using all single season samples for the 793 UK-wide reference sites (BMWP index parameters based on all possible single and combined season samples); *p* = model test probability value for effect of score.

	Original BMWP	
	BMWP NTAXA	BMWP ASPT
p	<0.001	<0.001
a1	0	0
a2	0	0
a3	0.01012 ± 0.00324	0.00010 ± 0.00119
a4	-0.01005 ± 0.00433	-0.01039 ± 0.00160
a5	-0.03136 ± 0.00629	-0.02252 ± 0.00232
	Abundance-weighted WHPT	
	WHPT NTAXA	WHPT ASPT
p	0.001	<0.001
a1	0	0
a2	0	0
a3	0.00749 ± 0.00594	0.00067 ± 0.00224
a4	-0.01050 ± 0.00794	-0.00999 ± 0.00300
a5	-0.03117 ± 0.01154	-0.02721 ± 0.00436
	LIFE	PSI
p	0.009	<0.001
a1	0	0
a2	0	0
a3	0.00208 ± 0.00136	0.00609 ± 0.00554
a4	-0.00051 ± 0.00182	-0.01101 ± 0.00740
a5	-0.00350 ± 0.00264	-0.04479 ± 0.01076

Table 33. Estimates of adjustment to reference condition parameters (*A1 – A5*) for the effects of assessment score (1-5) in model (M8*) for each biotic index based on using every possible combination of single and multiple season samples for the 793 UK-wide reference sites (LIFE and PSI estimates based on all single season samples only), (Note: A_j = 10 to the power a_j , where a_j is as in Table 32).

Index	Estimate of adjustment parameter				
	A_1	A_2	A_3	A_4	A_5
BMWP NTAXA	1.000	1.000	1.024	0.977	0.930
BMWP ASPT	1.000	1.000	1.000	0.976	0.949
WHPT NTAXA	1.000	1.000	1.017	0.976	0.931
WHPT ASPT	1.000	1.000	1.002	0.977	0.939
LIFE	1.000	1.000	1.005	0.999	0.992
PSI	1.000	1.000	1.014	0.975	0.902

5.5 Distribution of EQI values for LIFE, PSI and WHPT for GB Reference sites

The overall frequency distribution of adjusted EQI values for family-level LIFE and PSI amongst all individual single season samples from the 685 GB references sites is shown in Figure 33.

The lower 5% and 10% percentile values (together with the mean, median, min and max) for each of the four abundance-weighted indices for the 685 GB reference sites are given in Table 32. The expected values for WHPT NTAXA and WHPT ASPT are based on the standard RIVPACS IV predictor model based on 13 environmental variables, whereas the expected values for the LIFE and PSI indices are based on the reduced 15(2) RIVPACS predictor model which does not involve stream substratum composition, stream width or depth.

Either the lower 5% or lower 10% percentile of the adjusted EQI values for an index for the reference sites could be used as trial values for the good-moderate WFD status class boundary for that index. Thus for PSI, setting the high-good boundary adjusted EQI to 0.604 would mean that based on their adjusted EQI for PSI for any single season sample, 5% of the reference sites would be classed as moderate status (or worse).

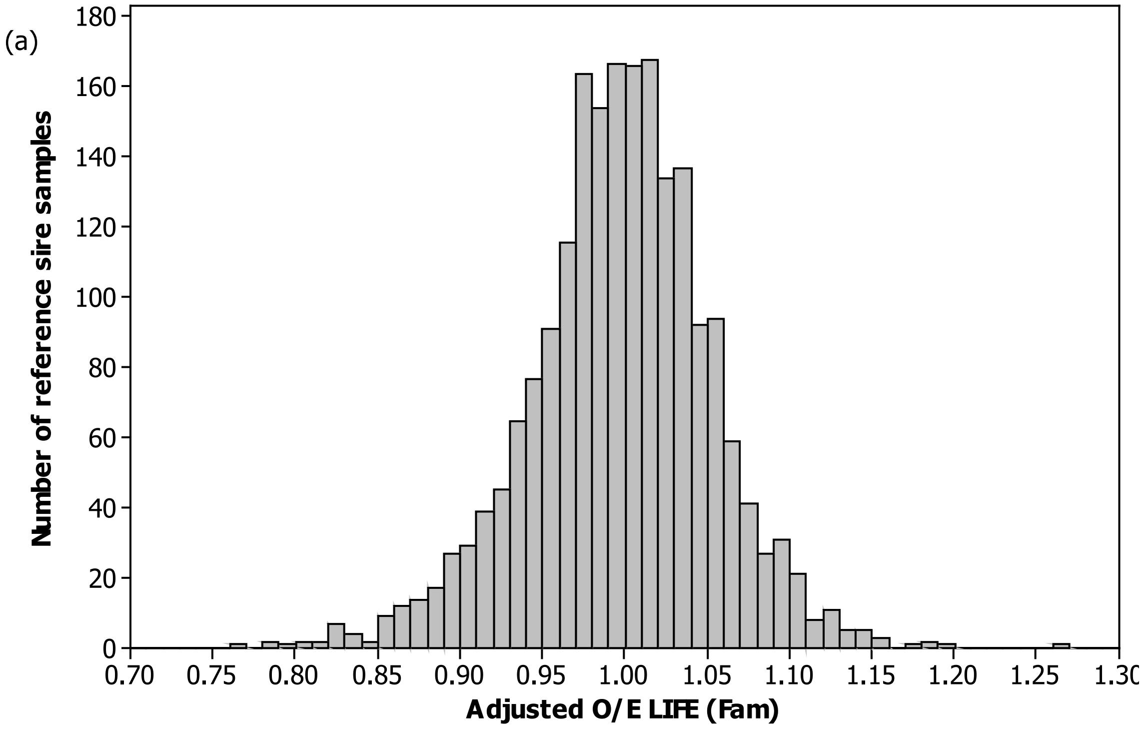
If the aim for the Agencies is to base annual estimates of site quality on the average of two single season samples adjusted EQI values, then it might be better to set the good-moderate class boundary on the percentile values of the distribution of the average of spring and autumn sample adjusted EQI values for the reference sites, as given in Table 34.

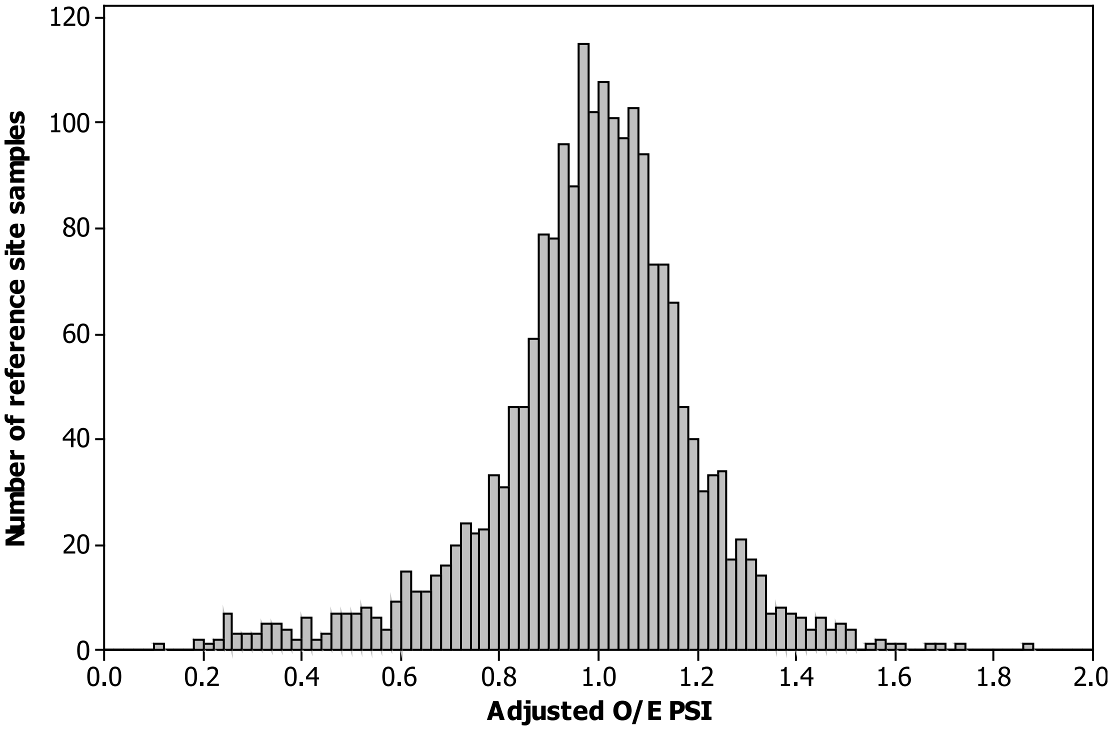
Notice that the lower 5% and 10% percentile values are higher and closer to one when based on the average of two single season sample EQI rather than just a single season sample EQI. This is because averaging across samples reduces the effect of the natural replicate sampling variability.

Table 34. Lower and upper 5% and 10% percentile frequency distribution values of the adjusted EQI values (EQI_{adj}) for the 685 GB reference sites (together with median, mean, min and max) for abundance-weighted WHPT NTAXA, WHPT ASPT, LIFE and PSI, both for single season EQI values (spring, summer and autumn) or for the average of spring and autumn adjusted EQI values; expected WHPT values based on standard 13 predictor variable RIVPACS model, expected LIFE and PSI values based on 10 variable stressor-independent 15(2) model (i.e. excluding stream width, depth and bed composition).

	Lower 5%	Lower 10%	Upper 10%	Upper 5%	Median	Mean	Min	Max
Single season EQI								
WHPT NTAXA	0.662	0.737	1.251	1.332	0.975	0.983	0.321	1.794
WHPT ASPT	0.847	0.886	1.102	1.130	0.999	0.997	0.686	1.374
LIFE	0.901	0.928	1.062	1.085	0.999	0.997	0.770	1.266
PSI	0.604	0.737	1.217	1.292	1.000	0.985	0.108	1.864
Average of 2 single seasons EQI								
WHPT NTAXA	0.706	0.755	1.217	1.297	0.978	0.983	0.504	1.783
WHPT ASPT	0.865	0.901	1.090	1.122	0.999	0.997	0.731	1.296
LIFE	0.913	0.938	1.052	1.069	0.999	0.997	0.809	1.154
PSI	0.636	0.786	1.185	1.271	1.002	0.985	0.179	1.514

Figure 33. Frequency histogram showing the statistical distribution of the adjusted O/E (EQI_{adj}) values of the family-level (a) LIFE and (b) PSI indices for the 685 reference sites for all single season samples. Expected values were based on the stressor-independent 15(2) RIVPACS predictive model (i.e. excluding stream width, depth and bed composition) and then adjusted for perceived quality of the reference sites involved.





6. Algorithms to simulate uncertainty for the family level indices 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 classifying 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 taking 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 equations 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 components (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 and those specific to the recommendations from this report on LIFE and PSI in yellow.

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. The statistical equations are given below in variance form, as is usual and more succinct. However, in the algorithms sections, the equations are given using the SD form, as this is the form of the input into the RICT software, and was also the form given in the algorithms sections of the orginal 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) in a 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 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) can be supplied for one, two or all three years in the period. It is generally assumed that the same season, or combination of seasons, have be sampled in each of the years included within a multi-year assessment. However, the overriding principle (as exemplified below) is that there must be the same number of seasons in each year (although these can be different seasons), and not all years need to be sampled. For example, the following combinations of samples are either appropriate (denoted by a ✓) or inappropriate (denoted by a ×) within any 3-year multi-year assessment):

Year 1	Year 2	Year 3	Appropriate	Description
Spr	Aut	Sum	✓	Same number of seasons in each year (these can be different seasons)
Spr	-	Aut	✓	Same number of seasons in each year (these can be different seasons) one year missing
Spr & Aut	Spr & Aut	Sum & Aut	✓	Same number of seasons in each year (these can be different seasons)
Spr & Aut	-	Spr & Aut	✓	Same number of seasons in each year (these can be different seasons) one year missing
Spr & Aut	Spr & Aut	Aut	×	Different number of seasons in each year (not currently supported)

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:

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

If different E values are available for each year then:

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

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 degrees 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 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 this estimated 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 within 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 composition 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 been standardised for differences 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 variation 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, $\mathrm{Var}(\{O/E\}_{Av})$, of the average O/E ($\{O/E\}_{Av}$) is estimated by:

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

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, $\mathrm{Var}(O_{Av})$, of the average observed values O_{Av}.

For single year assessments:

$$\mathrm{Var}(O_{Av}) = (\mathrm{Var}_{Rep} + \mathrm{Var}_{TSeas}) / N_{ObsSeas}$$

and for 3-year assessments:

$$\mathrm{Var}(O_{Av}) = ((\mathrm{Var}_{Rep} + \mathrm{Var}_{TSeas}) / N_{ObsSeas} + \mathrm{Var}_{Tyear} (1 – N_{ObsYear} / 3)) / N_{ObsYear}$$

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 measurements of stream width, depth and substratum 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 deviations (square root of variance estimates) for each component as 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 the 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

Where 'either 1 or 3 years' indicates either a single year assessment, or 3-years for a multi-year assessment.

$ObsIDX_i$ = 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 in the sub-sections that follow.

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

Estimates (see Table 14 of Clarke and Davy-Bowker, 2014) 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 (see Table 14 of Clarke and Davy-Bowker, 2014) 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 SDs 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 and Table 15), the sampling components SDs 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 SDs for a site were found to vary with the number of BMWP taxa present in the sample from that site and period.

In RICT, the various component SDs 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 r^th uncertainty simulation for LIFE (index 10):

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

(d) PSI index (index id 11):

For the PSI index, the estimates of the sampling component SDs 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 SDs 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:

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

On the Arcsine scale, the various component SDs (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 SDs for a site were found to vary with the number of BMWP taxa present in the sample from that site and period.

In RICT, the various component SDs 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:

$K_{PSI} = 0.955^{(NTAXAdiff)}$ = 0.955 to the power $NTAXA_{diff}$, 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 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):

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

As a check for the RICT software programmers in selecting the correct Arcsine 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 indices (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
= $u_{6a} + u_{6b} * ObsIDX_2 + Zbias_{6r} * (u_{6c} / \sqrt{Ubias_{2r}})$

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* :

$Ubias_{4r} = Ubias_{5r} * Ubias_{6r}$ = under-estimate of BMWP score for simulated sample r

$ObsIDX_{4rB} = ObsIDX_{4r} + Ubias_{4r}$ = bias-corrected observed unweighted WHPT Score (index 4)
$ObsIDX_{5rB} = ObsIDX_{5r} + Ubias_{5r}$ = bias-corrected observed unweighted WHPT NTAXA (index 5)
$ObsIDX_{6rB} = ObsIDX_{4rB} / ObsIDX_{5rB}$ = 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)

Definitions:

$ObsIDX_{2}$ = Observed sample value of BMWP NTAXA (index 2)

$ObsIDX_{10}$ = Observed sample value of LIFE (index 10)

$ObsIDX_{10r}$ = Simulation *r* Observed sample value of LIFE (index *10*) for current test site (uncorrected for bias)

$ObsIDX_{10rB}$ = Bias-corrected observed value of LIFE (index *10*) for simulation *r*

$Ubias_{2}r$ = RICT random estimate of bias for BMWP NTAXA (index 2) for simulated sample *r*,

P_{miss} = $Ubias_{2r}$ / ($ObsIDX_{2}$ + $Ubias_{2r}$) = Estimate of proportion of taxa missed for simulation *r*

$Zbias_{10r}$ = Random number deviate from a standard Normal distribution with a mean of 0.0 and SD of 1.0

$Ubias_{10r}$ = LIFE value of missed LIFE-scoring taxa for simulated sample r,

$= u_{10a} + Zbias_{10r} * (u_{10b} / \sqrt{Ubias_{2r}})$

where $u_{10a} = 6.55$, $u_{10b} = 1.8$

Then:

$ObsIDX_{10rB} = ObsIDX_{10} * (1 – P_{miss}) + Ubias_{10r} * P_{miss}$ = bias-corrected LIFE for simulated sample *r*

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

Definitions:

$ObsIDX_{2}$ = Observed sample value of BMWP NTAXA (index 2)

$ObsIDX_{11}$ = Observed sample value of PSI (index 11)

$ObsIDX_{11r}$ = Simulation *r* Observed sample value of PSI (index *11*) for current test site (uncorrected for bias)

$Ubias_{11r}$ = Bias in LIFE value for simulated sample r,

$ObsIDX_{11rB}$ = Bias-corrected observed value of PSI (index *11*) for simulation *r*

$Z_{null11r}$ = Random number deviate from a Uniform distribution over the range 0 -1

$Zbias_{11r}$ = Random number deviate from a standard Normal distribution with a mean of 0.0 and SD of 1.0

```
If $Z_{null11r} < u_{11d}$
  then
    $Ubias_{11r} = 0$
  else
    $Ubias_{11r} = u_{11a} - u_{11b} * ObsIDX_{11} + Zbias_{11r} * u_{11b}$
    where $u_{11a} = 2.067$ , $u_{11b} = 0.0504$ , $u_{11c} = 3.5$ , $u_{11d} = 0.2$
```

Then:

$ObsIDX_{11rB} = ObsIDX_{11} + Ubias_{11r}$ = Bias-corrected PSI value for simulated sample *r*

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 environmental 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)

$SDExp_{10}$ = 0.0 = Measurement error SD of Expected values of LIFE (index 10)

$SDExp_{11}$ = 0.0 = Measurement error SD of Expected values of PSI (index 11)

Note: RICT Expected values of LIFE and PSI are based on a RIVPACS predictive model that does not involve any of the field-measured environmental variables (stream width, depth and percentage substratum composition – the remaining measurement error SD for their Expected values is still not zero but will be small and is currently ignored).

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* (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 NTAXA 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. Confirmation of index response to stressor gradient and subsequent derivation of empirically based WFD class boundaries for family and species level LIFE and PSI.

Having first confirmed that the LIFE and PSI indices do indeed respond to variation in the level of the stress they purport to diagnose, we will then derive an initial set of empirically based WFD class boundaries for both indices at species and family-level. While the work described in section 5.5 used the distribution of adjusted EQI values from the RIVPACS reference dataset, WFD-compliant class boundaries for all classes should also be related to the level of stress (water flow or sediment) perceived to be operating at each site, especially at non-reference quality sites. This is particularly true for the Moderate/Poor and Poor/Bad boundaries, but also the critically important Good/Moderate boundary. Therefore, in addition to the simpler reference site-based percentile method, our approach is also to obtain observed (O), RICT expected (E) and thus EQI (O/E) values of LIFE and PSI for sites within suitable available datasets for which there is also recorded pressure data related to flow/water levels and sediment loadings. We drew upon both existing Agency (SEPA, EA, NIEA and NRFA) datasets, and datasets held by our project team.

7.1 Compilation of existing data

The first task was to request and compile existing macroinvertebrate community data from the UK environment Agencies and river discharge data from the National River Flow Archive. We also sought permission to use data from two previous research projects; (i) Defra WQ0128 - Extending the evidence base on the ecological impacts of fine sediment and developing a framework for targeting mitigation of agricultural sediment losses, (ii) Agri-environment Monitoring and Services Contract Lot 3 (183/2007/08) funded by the Welsh Government (Anthony *et al* 2012).

Macroinvertebrate community data and associated environmental data were collected and processed by Agency biologists using the standard RIVPACS field and laboratory protocols (Murray-Bligh, 1999).

7.1.1 Environment Agency and Natural Resources Wales

QMUL requested and received macroinvertebrate community data from all English and Welsh river sites, mostly at BMWP-family level, for samples taken between 1994 and 2012. These data, along with associated site and sample environmental data were compiled and formatted for analysis in a Microsoft Access 2010 relational database. Using a series of queries within this database, BMWP, WHPT, LIFE and PSI indices were calculated for all sampling occasions at all sites (Table 35). There was no formal labelling of samples that were identified to mixed taxonomic level and often samples were double-entered at both resolutions, therefore in the time-frame of the current project, it was not readily possible to calculate PSI_{sp} or $LIFE_{sp}$ values on this dataset.

Table 35. The number of samples and sites from England and Wales for which index values were calculated.

	BMWP (NTAXA & ASPT)	***$WHPT_{AB}$ (NTAXA & ASPT)***	***$LIFE_{fam}$***	***PSI_{fam}***
Samples	63560	63560	63494	63553
Sites	10714	10714	10708	10713

7.1.2 Scottish Environment Protection Agency

QMUL requested all routine macroinvertebrate monitoring data from SEPA for the period 2006-2012. The invertebrate community data were then compiled along with associated site and sample environmental data into an MS Access 2010 relational database.

The SEPA macroinvertebrate data came with the family-level and mixed taxonomic level (MTL) data on separate spreadsheets. This separation was maintained in the database. Using a series of queries within this database, BMWP, WHPT, LIFE and PSI indices were calculated for all sampling occasions at all sites (Table 36). It was possible to calculate the family-level indices on the family-level and the downgraded MTL data; in some cases this meant that two versions of same index were calculated for a sample. The species-level indices PSI_{sp} and $LIFE_{sp}$ were calculated using the MTL data. In addition, a recently developed macroinvertebrate species-level index to fine sediment stress, CoFSI (combined fine sediment index: Murphy *et al* 2015) was also calculated using the MTL data. This index was developed as part of the Defra WQ0128 project and in independent tests has been shown to perform better against the stress gradient than PSI (Murphy *et al* 2015).

Table 36. The number of samples from Scotland for which index values were calculated.

	BMWP (NTAXA & ASPT)	***$WHPT_{AB}$ (NTAXA & ASPT)***	***$LIFE_{fam}$***	***PSI_{fam}***	***$LIFE_{sp}$***	***PSI_{sp}***	***CoFSI***
MTL	3261	3261	3261	3261	3258	3261	3261
Family-level	9946	9946	9941	9945	-	-	-

7.1.3 Northern Ireland Environment Agency

QMUL requested and received the NIEA macroinvertebrate data which were compiled into an MS Access 2010 relational database. The database included family-level macroinvertebrate community data from 2007 to 2013 with associated site and sample environmental data. Using a series of queries within the database BMWP, WHPT, LIFE and PSI indices were calculated for all sampling occasions (n=874) at all sites (n=225).

7.1.4 Defra WQ0128: Extending the evidence base on the ecological impacts of fine sediment and developing a framework for targeting mitigation of agricultural sediment losses

This dataset assembled during Defra project WQ0128 (Collins *et al* 2012a) comprised measures of deposited fine sediment mass collected from 205 streams across England and Wales using the resuspension technique (Collins and Walling, 2007 a,b) and species-level macroinvertebrate community data collected from the same reach at the same time.

These 205 sites were selected as being unimpacted by sewage treatment discharges, urban run-off and upstream lakes or reservoirs. They were also chosen to represent as wide a range as possible of broad stream types across England and Wales e.g. western upland streams and south-eastern lowland streams. To focus the survey effort on those sites where fine sediment inputs were dominated by agricultural sources, a threshold was set at 75% for the proportion of the modelled total sediment input ($kg.ha^{-1}.yr^{-1}$) that was predicted to originate from agriculture (using recent updates to PSYCHIC (Collins *et al* 2007, 2009a,b, Davison *et al* 2008, Stromqvist *et al* 2008), a process-based model of suspended sediment mobilisation in surface run-off and drain flow and subsequent delivery to watercourses via both pathways now built into the APT (Agricultural Pollutant Transfer) framework) undertaken during Defra project WQ0128 (Collins et al., 2012a).

Each site was sampled once in either spring or autumn between 2010 and 2011. On each occasion a macroinvertebrate sample was taken using standard RIVPACS protocol (Murray-Bligh, 1999). Using a series of queries within the Defra WQ0128 relational database, BMWP, WHPT, LIFE, PSI and CoFSI indices were calculated for all sites. Data on reach-average total fine sediment mass (both surface drape and sub-surface combined) were used. These data were collected immediately upstream of the macroinvertebrate samples, using the disturbance technique described in Lambert and Walling (1988) and refined by Collins and Walling (2007a,b). Here, a steel cylinder (height 75 cm, diameter 48.5 cm) was inserted into an undisturbed section of the stream bed and the water column vigorously agitated for one minute, without touching the stream bed, to raise any fine sediment deposited on the surface of the stream bed. A pair of water samples was then collected quickly from within the cylinder. The stream bed was then disturbed to a depth of approximately 10 cm, and the water and bed vigorously agitated for one minute to raise any sub-surface fine sediment in addition to the re-suspended surface deposits. A second pair of water samples was then collected from within the cylinder. Four such sets of water samples (surface, and combined surface and subsurface) were collected from each site, two from erosional patches and two from depositional patches. The samples were then refrigerated and returned to the laboratory within 5 days, where they were processed for dry mass and organic content (i.e. volatile solids following combustion at 550°C). Reach-averaged values for total (combined surface and subsurface) deposited fine sediment were subsequently derived as the geometric mean for the reach. Recent research has confirmed that the resuspension technique performed as well as visual estimates of fine sediment cover in its ability to discriminate between rivers but, unlike visual estimates, was not affected by operator bias (Duerdoth *et al* 2015). These data provide a range of four orders of magnitude in reach average total deposited fine sediment mass (g m^{-2}), with which relationships with invertebrate indices were derived.

7.1.5 Welsh Government Agri-environment Monitoring and Services Contract Lot 3 (183/2007/08)

This dataset comprised 117 measures of deposited fine sediment mass collected from 62 Welsh streams using the resuspension technique (Collins and Walling, 2007) and species-level macroinvertebrate community data collected from the same reach at the same time. Most sites were sampled in spring and autumn (Anthony *et al* 2012).

The 62 sites were selected at random from a pool of headwater reaches with catchments characterised by less than 20% forestry cover, less than 10% urban cover, and greater than 40% in one of two categories of agri-environment scheme or >70% cover not in any scheme. Fifty five of the 62 sites were sampled in both spring and autumn; seven sites were only sampled in one season. Sampling and laboratory protocols were identical to those used in the Defra WQ0128 project. Using a series of queries within the WG-AES relational database, BMWP, WHPT, LIFE, PSI and CoFSI indices were calculated for all sites.

7.1.6 SEPARATE model outputs

To supplement the measured benthic deposited fine sediment data from the Defra WQ0128 and WG-AES projects we used the SEPARATE (SEctor Pollutant AppoRtionment for the AquaTic Environment) framework to generate predictions of the delivery (tonnes.yr^{-1}) of fine sediment to river reaches from their catchments.

SEPARATE is being developed by ADAS, Rothamsted Research-North Wyke and the Centre for Ecology and Hydrology as part of the Defra-funded project WQ0223 - Developing a field tool kit for ecological targeting of agricultural diffuse pollution mitigation measures project.

Project partners from WQ0223 kindly provided the current project with the modelled delivery of fine sediment to the river (apportioned to agricultural, natural (bank erosion), urban diffuse, and STW sources) from each of the 2023 English and Welsh Water Framework Directive waterbodies large enough for the model to make confident predictions (i.e. those with areas > 25 km^2). The predictions represent the delivery of fine sediment directly to the river from each WFD cycle 2 river waterbody. These predictions were based on the most up-to-date input data available and broadly represent conditions during the period 2010-present in the case of non-agricultural sources and the period 1991-2010 in the case of agriculture.

7.1.7 National River Flow Archive

To provide a measure of the extent of flow-stress on the macroinvertebrate community, QMUL requested summary flow data from the NRFA. Following discussions with NRFA staff at CEH Wallingford it was agreed that the NRFA would provide UK Hydrometric Register-style annual summaries for each gauging station with more than 5 years of data. Rather than using the conventional 'water year' (1st October to 30th September) or calendar years (1st January to 31st December), we requested that data was summarised for the 6 months preceding invertebrate sampling in spring and autumn. This time period was chosen to provide the best link between spring and autumn invertebrate samples and the preceding flow conditions and, therefore, provide the best opportunity for significant relationships.

We received summary flow statistics for 1436 UK gauging stations with data from 1993 up to 2012. These data included for each station, in each 6-month period, the mean, minimum and maximum recorded discharge (cumecs) as well as the Q_{95}, Q_{90}, Q_{70}, Q_{50}, Q_{30}, Q_{10} and Q_5 percentile data, where the Q-value denotes the discharge (cumecs) equalled or exceeded for x% of flow period, where x is the subscript. So, for example, Q_{95} provides a measure of low flow stress at the site in each 6-month period as it represents the discharge that was equalled or exceeded for 95% of that half-year's flow record. Hence, Q_{95} is a summary that incorporates both the extent and the duration of low flows.

To determine the relative impact of variation in discharge, we also calculated two versions of the summary flow statistics relative to the conditions typically experienced by the river. We used two approaches to derive relative statistics. First, we simply divided each half-year flow statistic for a given station by the average of that summary statistic derived from the total number of half-year values from the duration of that station's record. This gave a comparison of each half-year for a station relative to the long-term half-year average (full duration of the NRFA record for that station). For this term, the standardised summary statistic, a value of 1 indicates that the summary statistic for the half-year in question was the same as the long-term half-year average for that site. Second, we normalized each half-year flow statistic for a given station by subtracting from it the long-term half-year average (as defined above), and then dividing the difference by the standard deviation around the long-term average. This second approach gives what is known as the z-score, whereby half-year flow statistics for a station are normalized by their long-term average and inter-year variability. Here, a z-score of 0 indicates that the summary statistic for the half-year in question was the same as the long-term average for that site.

The NRFA also provide additional environmental information on the upstream catchment feeding into each gauging station, including an assessment of the factors that are acting on the catchment and are likely to affect the flow regime e.g. reservoirs, abstraction, discharges, etc.

7.2 Spatial and temporal matching of biological and stressor data

The next step in the data acquisition and compilation phase of the project was to spatially and temporally match the macroinvertebrate data to the measures of flow and fine sediment stress.

7.2.1 Fine Sediment Stress

For the WQ0128 and WG-AES datasets the matching was not necessary as the level of the stressor (deposited fine sediment) and the biological response were measured contemporaneously at the same location. These data (n=322 samples) provide a robust measure of fine sediment stress which we can relate to PSI and $CoFSI_{sp}$ index values.

An alternative approach (which sought to expand the range of river types included in the analysis) was to use the SEPARATE-estimates of delivery of fine sediment from the catchment in combination with some surrogate measure of retention at the site as a proxy for fine sediment stress to the benthic macroinvertebrate community. The SEPARATE outputs for England and Wales, however, needed to be matched in space and time with the most appropriate Environment Agency site and sampling occasions. We used ArcMap 10.2 to identify the most downstream EA macroinvertebrate site with more than three sampling occasions within each WFD cycle 2 waterbody. We then linked these 1843 WFD WBs with biotic index data to the SEPARATE data for the 2023 catchments with reliable modelling results (i.e. those with areas >25 km^2). We had 970 waterbody matches; the mismatches were due to the fact that the biological sites were assigned to an earlier version of the WFD waterbody codes, while the SEPARATE outputs used a more recent version designed for WFD cycle 2.

Whilst the SEPARATE predictions represent the delivery of fine sediment directly to the river from each WFD cycle 2 river waterbody, we were unable at this time to include the inputs of sediment from WFD cycle 2 river waterbodies upstream: the developers of the SEPARATE model have yet to receive details of waterbody connectivity from the Environment Agency, which would enable the model to compile delivery of fine sediment from all upstream sources.

For Scotland and Northern Ireland there is less scope for describing the fine sediment stress at each macroinvertebrate site as there are no available measurements of benthic fine sediment mass and SEPARATE or similar cross sector modelling has not been undertaken in either country.

7.2.2 Flow Stress

Using ArcMap 10.2 we linked each NRFA station to its closest or most appropriate EA and SEPA macroinvertebrate sampling location on the same watercourse. We only considered invert sites with more than three sampling occasions (n= 1012 in Scotland, n= 8651 in England & Wales).

We first identified the macroinvertebrate sampling sites within 1 km of each gauging station. Then those biology sites within 1km that were not on the same watercourse as the gauging station were excluded. Next we manually checked those gauging stations where there were multiple biology sites on the same watercourse to select the most appropriate match.

In England and Wales there were 4606 invert samples from 449 sites with derived biotic indices and spatially matched summary flow statistics. In Scotland there were 1206 invert samples from 123 sites with derived biotic indices and spatially matched summary flow statistics. In Northern Ireland we have 405 invert samples from 45 sites with derived biotic indices and spatially matched summary flow statistics. UK-wide we have a dataset for LIFE–flow stress analysis comprising 617 sites with multiple years' data at each site.

7.3 Generating stressor-independent EQI values

To achieve our objective we needed to relate the biotic index response to variations in the level of stress. The biotic index needed to be presented to the analysis as an Observed/Expected or EQI value, not as the raw observed value. EQI is a ratio, calculated as observed/expected; in other words EQI is the index score from a test site compared to a predicted index score based predominantly on data from environmentally-similar sites that are not subject to pressure from the stressor in question. Ordinarily we would use a RIVPACS IV-derived prediction of the expected index value in the absence of stress. However, the routine RIVPACS IV GB model uses measures of substrate, width and depth as predictor variables; variables that would be impacted by the stress being assessed. We therefore needed to use an alternative version of RIVPACS that omits these stress-related variables from the set of predictor variables. Such a model has been created by Ralph Clarke as part of the SNIFFER-funded WFD119 project; Model 15(2), (Clarke *et al* 2011). This Model is capable of making predictions of LIFE and PSI indices. Unfortunately there is at present no Model 15(2) for Northern Ireland. Therefore we were unable to proceed any further with these data as part of the index testing (n = 405 samples). Hence, the final dataset comprised 434 matched gauging stations and EQI values.

Currently there is no algorithm to provide EQI_{adj} or EQR values for test sites within RIVPACS IV Model 15(2). Hence, the values derived with this model are *unadjusted* EQIs. The distribution of unadjusted EQI values for the 685 RIVPACS IV GB reference sites is presented in Table 37 for comparison with their corresponding adjusted counterparts (Table 34). In general there is relatively little difference between the distributions of adjusted and unadjusted $LIFE_{fam}$ and PSI_{fam} EQI values. Descriptors of the distribution of EQIadj values were either equal or lower than those for EQI with an average difference of 0.006 (range 0.000 - 0.025).

Ralph Clarke ran RIVPACS IV Model 15(2) for all the GB spatially-matched macroinvertebrate samples for which there were the necessary environmental predictor data (national grid reference, altitude, distance from source, slope, discharge category and alkalinity) to generate expected values for NTAXA, ASPT, $WHPT_{ab}NTAXA$, $WHPT_{ab}ASPT$, $LIFE_{fam}$, $LIFE_{sp}$, PSI_{fam}, PSI_{sp} and CoFSI.

After this final step we had expected index values for 10,320 samples. These expected index values were imported into the respective databases to allow the calculation of EQI values and their linking to the measures of stress.

- WQ0128 dataset (n=188 samples/sites)
- WG-AES dataset (n=101 samples/57 sites)

- EA-SEPARATE dataset (samples taken 2010-present; linked to SEPARATE outputs, 702 samples, 204 sites)
- SEPA-flow dataset (n=1004 samples, 99 invert sites & gauging stations)
- EA-flow dataset (n=3172 samples, 335 invert sites & gauging stations)

These data were then analysed to quantify the relationships between EQIs for the various indices and their respective stressor gradients, with the objective of confirming that these indices responded appropriately to the pressure gradients for which they were designed. These data were also used to derive empirically based WFD class boundaries for both indices at species and family-level.

Table 37. Lower and upper 5% and 10% percentile frequency distribution values of the unadjusted EQI values (EQI) for the 685 GB Reference sites (together with median, mean, min and max) for abundance-weighted WHPT NTAXA, WHPT ASPT, $LIFE_{fam}$ and PSI_{fam}, both for single season samples (spring, summer and autumn) or for the average of spring and autumn unadjusted EQI values; expected WHPT values are based on standard 13 predictor variable RIVPACS model, expected $LIFE_{fam}$ and PSI_{fam} values are based on 10 variable stressor-independent model (i.e. excluding stream width, depth and bed composition).

	Lower 5%	Lower 10%	Upper 10%	Upper 5%	Median	Mean	Min	Max
Single season EQI								
$WHPT_{ab}NTAXA$	0.668	0.746	1.266	1.350	0.988	0.995	0.326	1.824
$WHPT_{ab}ASPT$	0.852	0.891	1.107	1.133	1.003	1.001	0.688	1.387
$LIFE_{fam}$	0.903	0.931	1.065	1.088	1.002	0.999	0.770	1.267
PSI_{fam}	0.614	0.748	1.232	1.309	1.009	0.996	0.110	1.864
Average of 2 single seasons EQI								
$WHPT_{ab}NTAXA$	0.712	0.755	1.240	1.311	0.991	0.995	0.511	1.813
$WHPT_{ab}ASPT$	0.872	0.905	1.094	1.127	1.003	1.001	0.737	1.308
$LIFE_{fam}$	0.916	0.940	1.054	1.072	1.002	0.999	0.811	1.156
PSI_{fam}	0.644	0.795	1.196	1.288	1.011	0.995	0.182	1.539

7.4 Quantifying the response of indices to stressor gradients

7.4.1 Fine sediment stress

The objective was to determine the relationship between EQI for PSI and a gradient of pressure due to fine sediment. This was necessary to verify that EQI for PSI is related to the level of stress perceived to be operating at a site. A range of invertebrate indices were tested against the gradient of pressure due to fine sediment in order to establish if indices developed to identify pressure from fine sediment added additional explanatory power compared with those already in use in RICT. The indices tested were:

- NTAXA (BMWP number of scoring taxa)
- ASPT (BWMP Average Score Per Taxon)
- $WHPT_{ab}NTAXA$ (Abundance weighted WHPT number of scoring taxa)
- $WHPT_{ab}ASPT$ (Abundance weighted WHPT Average Score Per Taxon)
- $LIFE_{fam}$ (lotic invertebrate index for flow evaluation based on family level data)
- $LIFE_{sp}$ (lotic invertebrate index for flow evaluation based on species level data)
- PSI_{fam} (Proportion of Sediment-sensitive Invertebrates based on family level data)
- PSI_{sp} (Proportion of Sediment-sensitive Invertebrates based on species level data)
- CoFSI (Combined Fine Sediment Index, developed as part of WQ0128: Murphy et al. 2015)

For each index EQI was established using RIVPACS IV Model 15(2) for all the GB, i.e. excluding those predictor variables that could be compromised by fine sediment pressure, namely bed composition, width and depth.

Two sets of data describing the pressure gradient were used to provide the best available representation across river types of the UK. As there is currently no consensus on the best approach to quantification of the pressure from fine sediment, two approaches to describing the pressure gradient were also used. Just as there are often different ways to measure the same chemical determinand, to date there is no consensus as to which aspect(s) of fine sediment pressure the biota respond to and, hence, which is the most appropriate measure of fine sediment to quantify this pressure (Collins & Anthony 2008, Collins *et al* 2011). Although the data on invertebrate community and deposited fine sediment are of high quality, WQ0128 was restricted to sites where agriculture is the dominant source of fine sediment pressure, and hence the data set is dominated by smaller rivers (upstream of urban areas) in agricultural areas. SEPARATE provides an estimate of fine sediment load which is a good measure of the fine sediment pressure (that portion of the fine sediment load that is additional to the natural background load) and covers a large spatial scale and a wide range of river types. Since the management of both rivers and fine sediment run-off must eventually take place at the catchment scale (Collins & Anthony 2008, Collins *et al* 2011), it is at this scale that investigations must take place. However, these data do not include a measure of retention: the most pronounced effects of fine sediment on invertebrates appear to be driven by deposited fine sediment, i.e. the portion of the load that is retained on the river bed (Jones *et al* 2012).

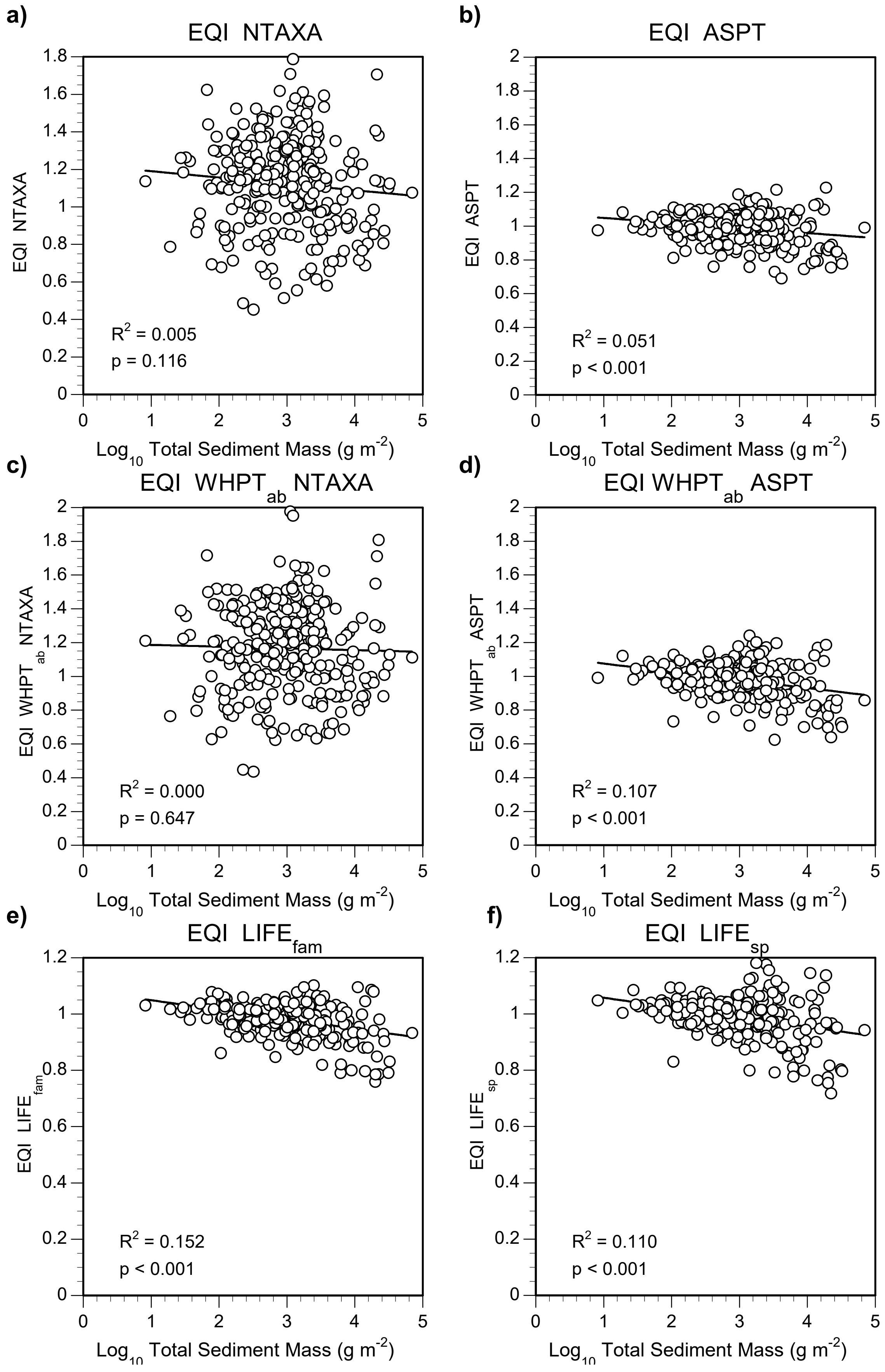
a) WQ0128 and WG-AES datasets

The WQ0128 and WG-AES datasets provided a range of 4 orders of magnitude in reach average total deposited fine sediment mass (g m^{-2}). These were used to derive relationships with invertebrate indices. EQI NTAXA and $WHPT_{ab}NTAXA$ showed no relationship with reach average total deposited fine sediment mass (Figure 34). EQI ASPT and $WHPT_{ab}ASPT$ showed significant negative relationships with reach scale average total deposited fine sediment mass, as did EQI of both LIFE indices and the sediment indices. However, the R^2 for EQI $LIFE_{fam}$ was greater than PSI_{fam}. There was a wide scatter in EQI PSI from sites with high reach average total deposited fine sediment mass, i.e. over 1,000 g m^{-2} fine sediment. It should also be noted that the relationship between PSI_{fam} and PSI_{sp} was not one-to-one: the level of taxonomic resolution used influences the PSI value returned, even where both PSI_{fam} and PSI_{sp} were derived from the same sample (see Appendix 5). The same was true for $LIFE_{sp}$ and $LIFE_{fam}$ although the influence of taxonomic resolution on these indices was less marked.

Whilst the WQ0128 and WG-AES data on reach average total deposited fine sediment mass provide a good snapshot estimate of the amount of fine sediment present in a reach, this does not necessarily represent a pressure from fine sediment on the biota. Rivers differ naturally in the amount of fine sediment present in the bed, even when unimpacted: steep mountain streams naturally have less deposited fine sediment than low gradient lowland rivers. The pressure from fine sediment should ideally be described in terms of the amount of deposited fine sediment in the bed of the river that is attributable to the impact of human activities in the catchment (which influences the load of sediment delivered to the river) or impacting the channel (which influences the retention of a portion of the fine sediment load).

River slope was strongly correlated with the reach average total deposited fine sediment mass (Figure 35), which may explain why the indices not specific to sediment were related to the reach average total deposited fine sediment mass, i.e. they were responding to river slope rather than fine sediment pressure *per se.* As the data collected in WQ0128 were from sites spread across a gradient of pressure, in terms of fine sediment load, there was considerable scatter around the relationship between river slope and reach average total deposited fine sediment mass (Figure 35 b): for the same river slope, samples were collected from sites with varying total deposited fine sediment mass. Hence, a better measure of the pressure from fine sediment at a site is the residual of the relationship between river slope and reach average total deposited fine sediment mass. This represents the deviation in total deposited fine sediment mass at the site from the average for rivers of the same slope (over the dataset): positive residuals represent more total deposited fine sediment mass than average, negative residuals represent less total deposited fine sediment mass than average.

The data from WQ0128 were used to establish a pressure gradient based on the residuals of the relationship between deposited fine sediment mass and river slope, and the relationships with the invertebrate based indices determined. Once more there was no relationship between this better measure of pressure from fine sediment and EQI for NTAXA and $WHPT_{ab}NTAXA$ (Figure 36 a & c). EQI for ASPT and $WHPT_{ab}ASPT$ showed significant negative relationships with the residuals of the relationship between deposited fine sediment mass and river slope, as did EQI for both LIFE indices and the sediment indices (Figure 36). However, the R^2 for EQI for $LIFE_{fam}$ was greater than either PSI_{fam} or PSI_{sp}. There was a wide scatter in EQI for PSI from sites with relatively high reach average total deposited fine sediment mass, i.e. those sites with positive residuals of the relationship between deposited fine sediment mass and river slope. Both high and low EQI values were obtained for PSI at high levels of pressure from fine sediment (Figure 36 g &h). A similar pattern was reported by Turley *et al* (2014) when they related observed PSI_{sp} to visual observations of the percentage fines (sand, silt and clay) in the stream bed using the RIVPACS reference dataset. Their analysis revealed such large variance around the relationship, particularly at the high end of the fine sediment stress gradient that it was not possible to discriminate between sites of varying percentage fines. Hence, they concluded further work was necessary to validate the PSI indices. In particular, they called for a more objective quantitative method of measuring deposited fine sediment. By quantifying fine sediment pressure based on the re-suspension technique and by accounting for stream slope in our analyses, the present work addresses these recommendations. In addition, by using EQI of PSI, we have factored out variation in PSI attributable to natural environmental gradients and, hence, enhanced the opportunity for detecting a robust, un-confounded and useful relationship between PSI and the fine sediment stress. Despite these efforts it would appear that both PSI_{fam} and PSI_{sp} are unstable at high levels of pressure, with a wide range of EQI values being returned which will make interpretation of EQI values difficult in terms of classification.


a)
EQI NTAXA
EQI NTAXA
$R^2 = 0.005$
$p = 0.116$
Log$_{10}$ Total Sediment Mass (g m^{-2})
b)
EQI ASPT
EQI ASPT
$R^2 = 0.051$
$p < 0.001$
Log$_{10}$ Total Sediment Mass (g m^{-2})
c)
EQI WHPT$_{ab}$ NTAXA
EQI WHPT$_{ab}$ NTAXA
$R^2 = 0.000$
$p = 0.647$
Log$_{10}$ Total Sediment Mass (g m^{-2})
d)
EQI WHPT$_{ab}$ ASPT
EQI WHPT$_{ab}$ ASPT
$R^2 = 0.107$
$p < 0.001$
Log$_{10}$ Total Sediment Mass (g m^{-2})
e)
EQI LIFE$_{fam}$
EQI LIFE$_{fam}$
$R^2 = 0.152$
$p < 0.001$
Log$_{10}$ Total Sediment Mass (g m^{-2})
f)
EQI LIFE$_{sp}$
EQI LIFE$_{sp}$
$R^2 = 0.110$
$p < 0.001$
Log$_{10}$ Total Sediment Mass (g m^{-2})
0
1
2
3
4
5
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2

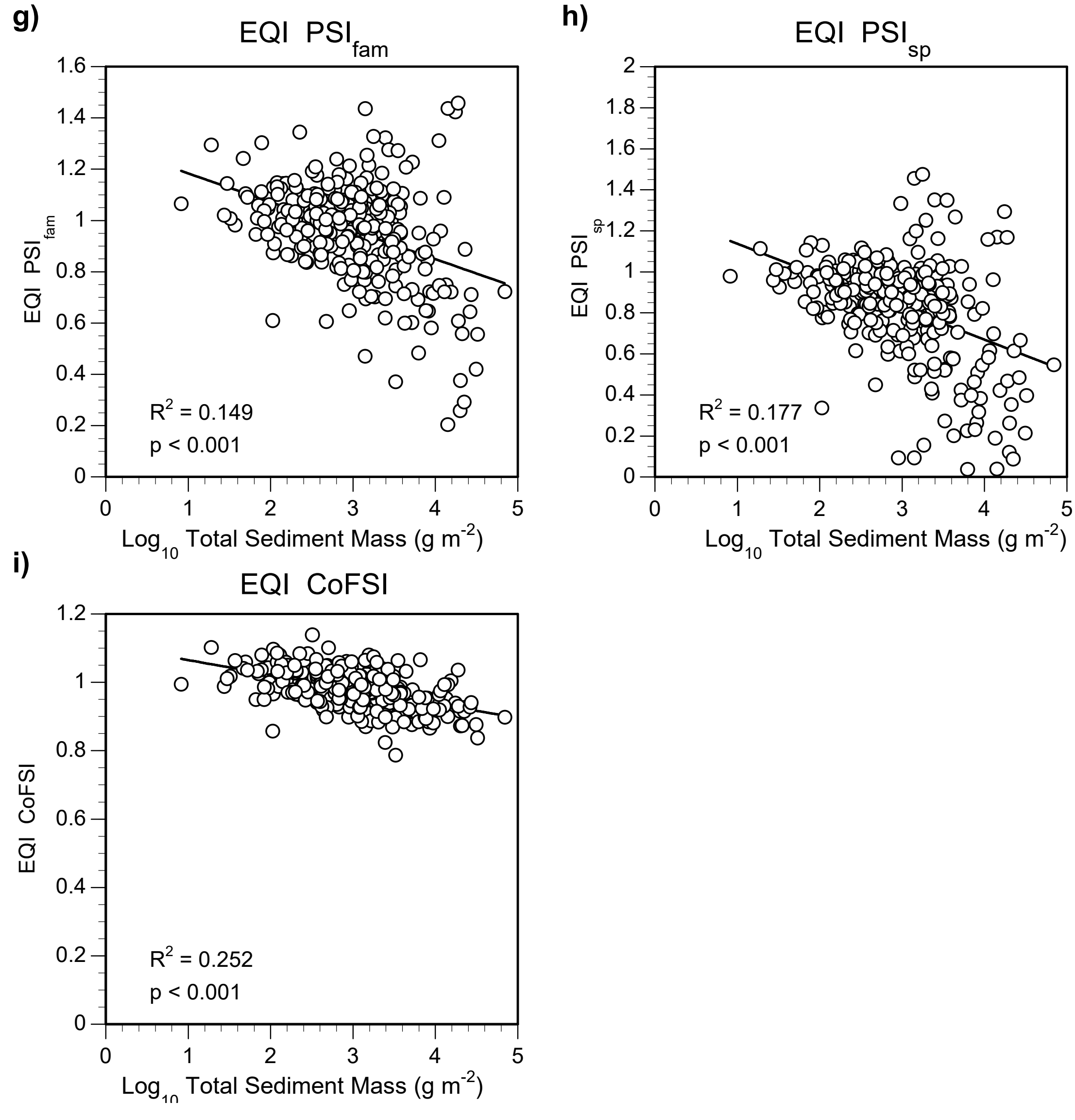
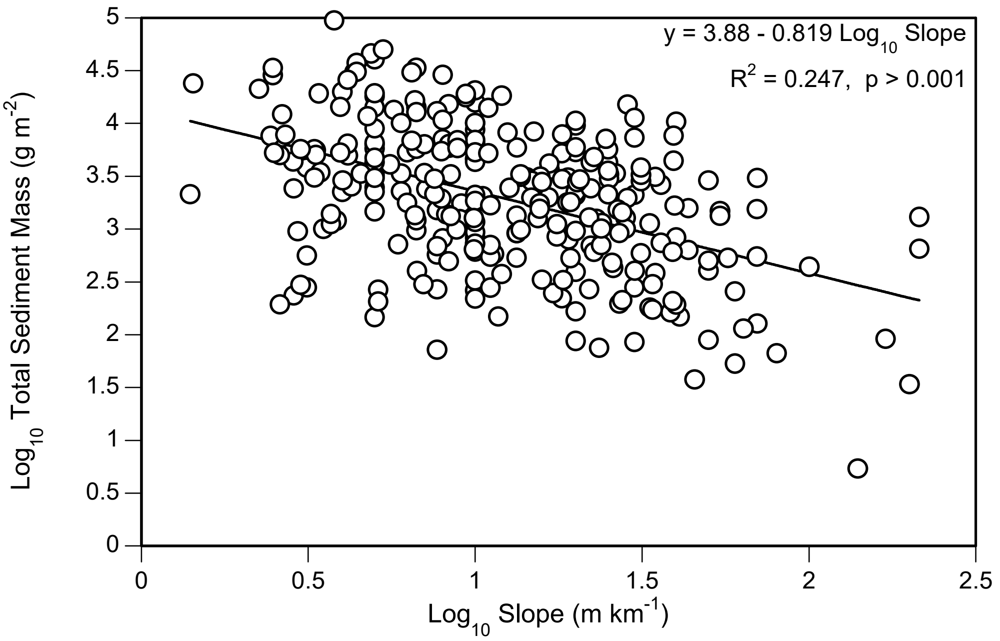


Figure 34. Relationships between reach scale average total deposited fine sediment mass (g m^{-2}) and EQI values for the invertebrate indices a) NTAXA, b) ASPT, c) WHPT$_{ab}$NTAXA, d) WHPT$_{ab}$ASPT, e) LIFE$_{fam}$, f) LIFE$_{sp}$, g) PSI$_{fam}$, h) PSI$_{sp}$, and i) CoFSI.

a)



b)

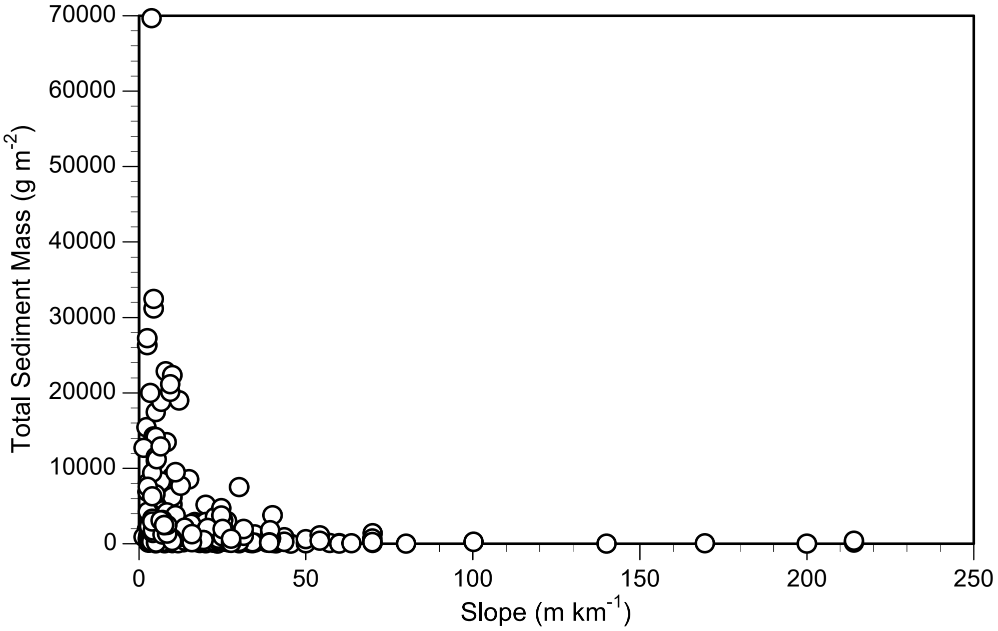
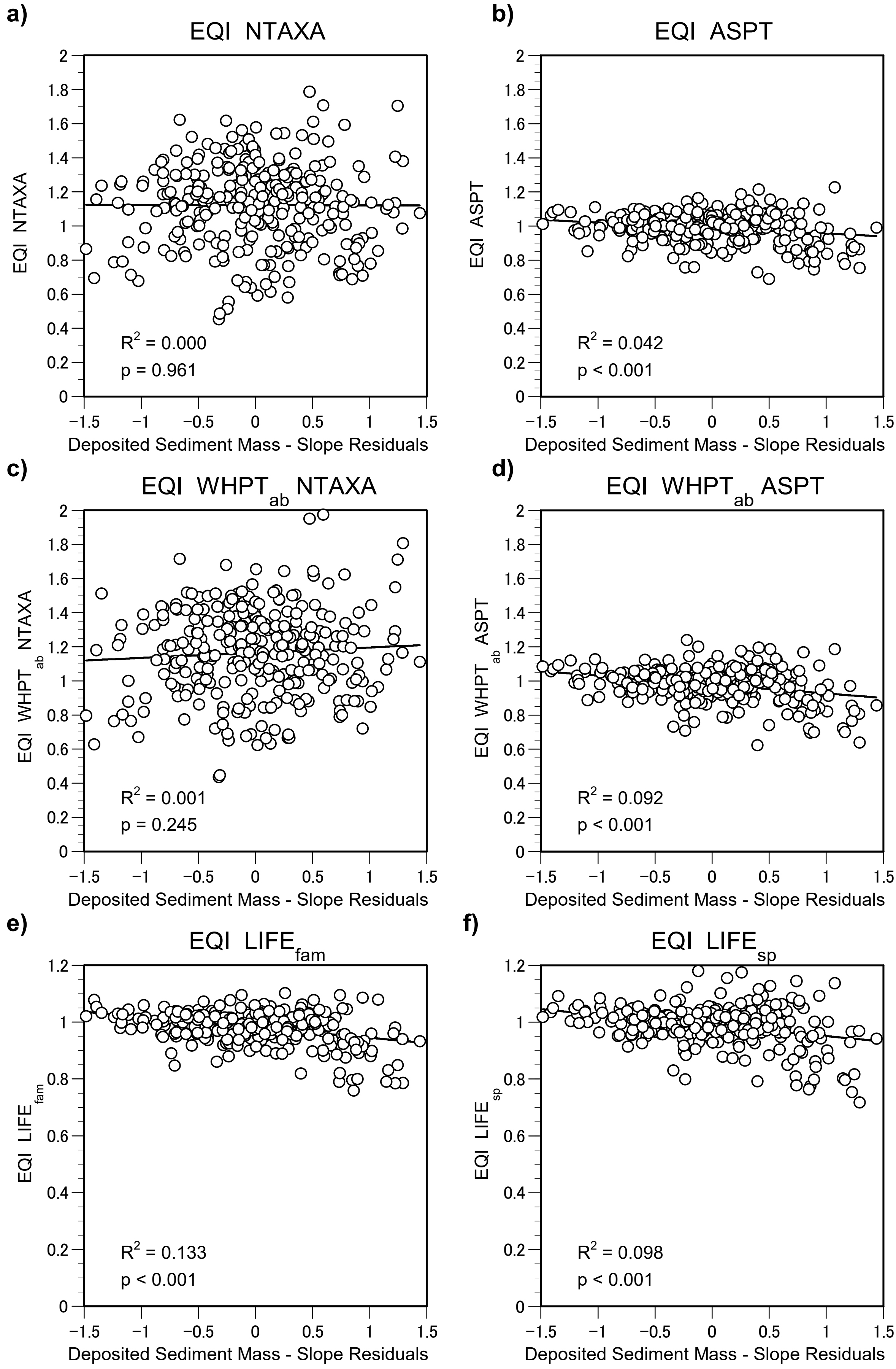


Figure 35. Relationship between river slope (m km^{-1}) and reach scale average total deposited fine sediment mass (g m^{-2}) as both a) log-log and b) untransformed values. As river slope influences the reach scale average total deposited fine sediment mass, the residual from the log-log relationship provides a better measure of sediment pressure than raw values of reach scale average total deposited fine sediment mass (i.e. deviation from average for that river type).


a)
EQI NTAXA
EQI NTAXA
$R^2 = 0.000$
$p = 0.961$
Deposited Sediment Mass - Slope Residuals
b)
EQI ASPT
EQI ASPT
$R^2 = 0.042$
$p < 0.001$
Deposited Sediment Mass - Slope Residuals
c)
EQI $WHPT_{ab}$ NTAXA
EQI $WHPT_{ab}$ NTAXA
$R^2 = 0.001$
$p = 0.245$
Deposited Sediment Mass - Slope Residuals
d)
EQI $WHPT_{ab}$ ASPT
EQI $WHPT_{ab}$ ASPT
$R^2 = 0.092$
$p < 0.001$
Deposited Sediment Mass - Slope Residuals
e)
EQI $LIFE_{fam}$
EQI $LIFE_{fam}$
$R^2 = 0.133$
$p < 0.001$
Deposited Sediment Mass - Slope Residuals
f)
EQI $LIFE_{sp}$
EQI $LIFE_{sp}$
$R^2 = 0.098$
$p < 0.001$
Deposited Sediment Mass - Slope Residuals
−1.5
−1
−0.5
0
0.5
1
1.5
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2

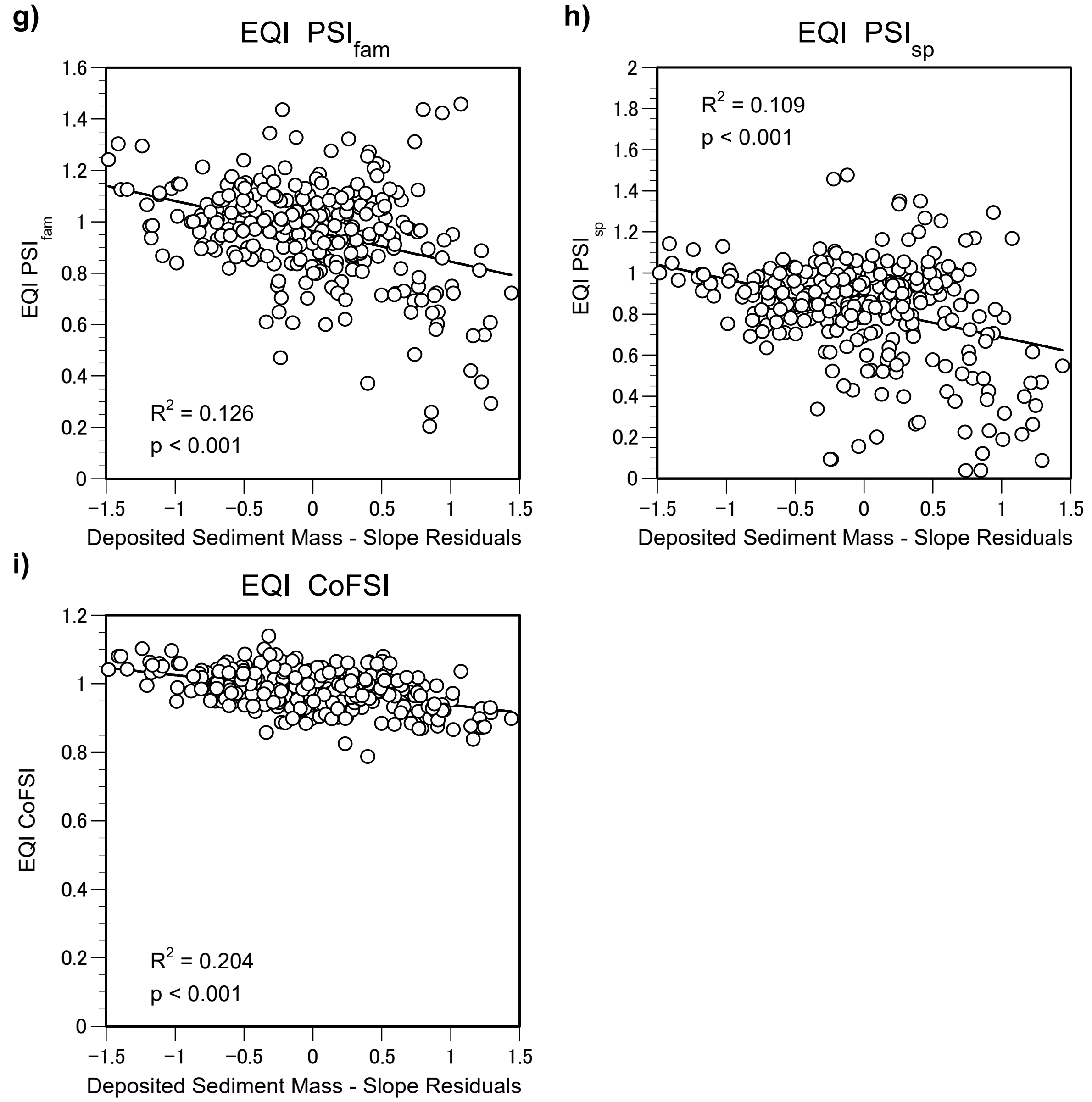


Figure 36. Relationships between the residuals of the relationship between deposited fine sediment mass and river slope, and EQI values for the invertebrate indices a) NTAXA, b) ASPT, c) $WHPT_{ab}NTAXA$, d) $WHPT_{ab}ASPT$, e) $LIFE_{fam}$, f) $LIFE_{sp}$, g) PSI_{fam}, h) PSI_{sp}, and i) CoFSI.

b) SEPARATE dataset

The SEPARATE data combine modelled and empirical information on sediment emissions based on data from a variety of sources including monitored sediment concentrations in STW outfalls. These data, matched with Environment Agency samples provided the second dataset for establishing the response of indices against a pressure gradient.

SEPARATE provides an estimate of fine sediment load (t yr^{-1}) to WFD cycle 2 waterbodies which is apportioned to agriculture, urban diffuse, sewage treatment works and bank erosion sources. SEPARATE can be used to provide an estimate of pressure from fine sediment by excluding that portion of the load that is natural (i.e. attributable to bank erosion) if the outputs are used in conjunction with the predictions of 'modern background sediment delivery to rivers' developed by Defra project WQ0128 (Foster *et al* 2011, Collins *et al* 2012a,b). These data provide a range of 2 orders of magnitude in fine sediment load (t yr^{-1}) to WFD cycle 2 waterbodies. However, the waterbodies vary in size, with larger waterbodies tending to receive larger loads of fine sediment (Figure 37). Hence, to standardize between waterbodies and to provide a better estimate of pressure from fine sediment, the load of fine sediment (excluding bank erosion) was divided by waterbody area to give an estimate of the yield of excess fine sediment from that catchment to the river.

Nevertheless, it should be noted that the measures of fine sediment pressure derived from SEPARATE refer to the fine sediment delivered to the river channel, and do not include a measure of retention: the most pronounced effects of fine sediment on invertebrates appear to be driven by deposited fine sediment, i.e. the portion of the load that is retained on the river bed (Jones *et al* 2012). When using the WQ0128 data, river slope was used as a surrogate for retentiveness, however, river slope is correlated with the yield of excess fine sediment (Figure 37 c). This is apparently a reflection of the fact that rivers with a high slope tend to occur in topographies with a high average gradient: topographies with a high gradient lead to better connectivity between the catchment and the river channel and, hence, high rates of estimated delivery of fine sediment. Whilst there is a strong negative relationship between percentage fines (sand, silt and clay) established by visual observation at the point of invertebrate sampling and river slope (Figure 37 d), the relationship with yield of excess fine sediment is also negative (Figure 37 e) reflecting both increased delivery and reduced retention in rivers with higher slopes.

EQI for $WHPT_{ab}$ASPT, $LIFE_{fam}$ and PSI_{fam} were not significantly related to variation in modelled excess fine sediment yield (Figure 38). EQI for NTAXA, ASPT and $WHPT_{ab}$NTAXA were negatively related to modelled excess fine sediment yield, although the amount of variation explained was low. EQI for CoFSI was positively correlated with modelled excess fine sediment yield (Figure 38), probably reflecting the lack of retentiveness of steeper rivers. Although the data describing fine sediment yield modelled using SEPARATE provide a good estimate of the pressure, without some measure of retention of fine sediment they do not describe the impact on the invertebrate community well; the invertebrates appear to respond to deposited fine material rather than suspended solids.

To explore further the influence of retention of fine sediment, the relationship between EQI of the invertebrate indices and percentage fines (sand, silt and clay) established by visual observation at the point of invertebrate sampling was established. Whilst percentage fines is not a particularly good measure of pressure from fine sediment, it does provide an estimate of the amount of deposited fine sediment at the site which can be used to compare with the EQI established using the RIVPACS IV stressor-independent Model 15(2).

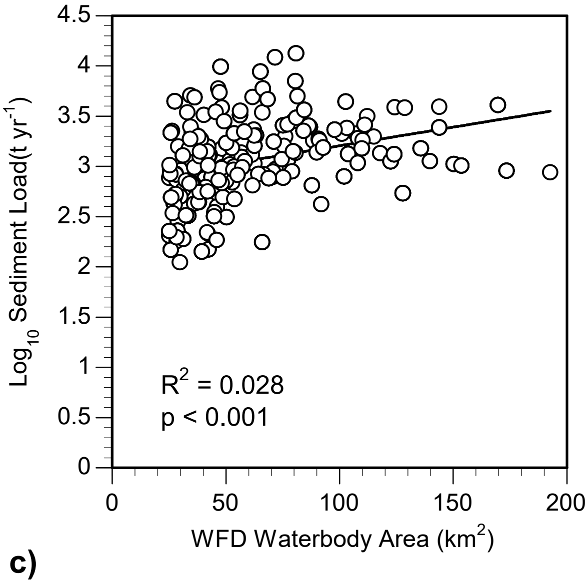
EQI for NTAXA, ASPT and $WHPT_{ab}$NTAXA were not significantly related to percentage fines (Figure 39). EQI for $WHPT_{ab}$ASPT was significantly related to percentage fines, but with low explanatory power. $LIFE_{fam}$, PSI_{fam} and CoFSI were significantly negatively related with percentage fines established by visual observation (Figure 39).

The results of this analysis should be treated with caution as the SEPARATE predictions only represent the delivery of fine sediment directly to the river from each WFD cycle 2 river waterbody. At this time we were unable to include estimates of the delivery of sediment from areas upstream of each WFD cycle 2 river waterbody: the developers of the SEPARATE model have yet to receive details of waterbody connectivity from the Environment Agency, which would enable the framework to compile delivery of fine sediment from all upstream sources. Hence, the pressure gradient used here only represents immediate pressure from within the WFD cycle 2 river waterbody. The pressure gradient will under-represent the total pressure from fine sediment, and, particularly in larger river catchments, this under-estimation will be significant.

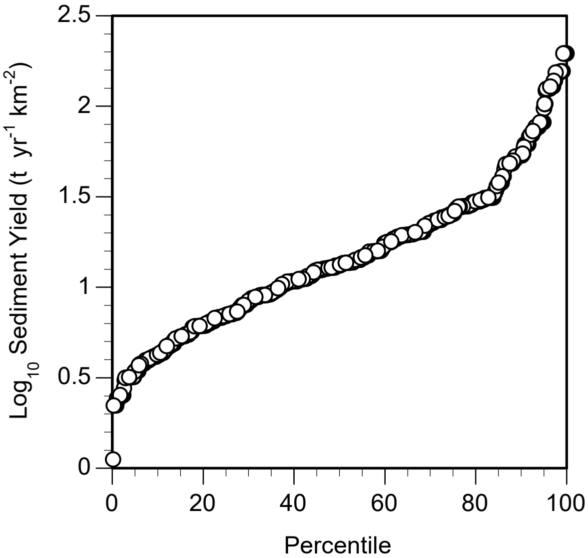
Summary:
Whilst a relationship between PSI and the gradient of pressure from fine sediment was apparent, both high and low EQI values were obtained for PSI at high levels of pressure from fine sediment. It would appear that both $\mathrm{PSI}_{\mathrm{fam}}$ and $\mathrm{PSI}_{\mathrm{sp}}$ are unstable at high levels of pressure, with a wide range of EQI values being returned. This instability will make interpretation of EQR values difficult in terms of classification if PSI were to be applied at the national scale.

Figure 37 (overleaf). Relationships in the data from the SEPARATE framework used to estimate pressure from fine sediment. a) relationship between WFD cycle 2 waterbody area and predicted total fine sediment load (t yr^{-1}), b) percentile distribution of the WFD cycle 2 waterbodies used according to predicted excess fine sediment yield (t yr^{-1} km^{-2}), c) relationship between predicted excess fine sediment yield (t yr^{-1} km^{-2}) and river slope at the point of invertebrate sampling, d) relationship between percentage fines (sand, silt, clay) observed visually and river slope at the point of invertebrate sampling, e) relationship between percentage fines (sand, silt, clay) observed visually at the point of invertebrate sampling and predicted excess fine sediment yield (t yr^{-1} km^{-2}).

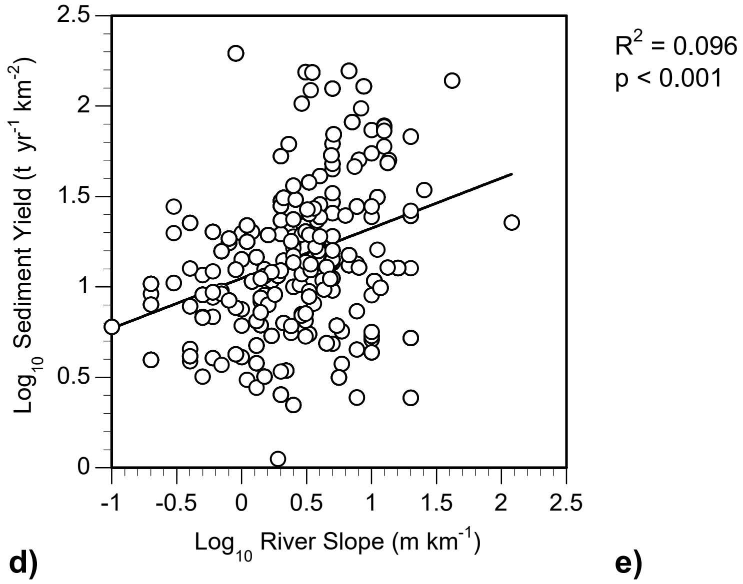
a)


Log10 Sediment Load(t yr-1)
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
R2 = 0.028
p < 0.001
0
50
100
150
200
WFD Waterbody Area (km2)

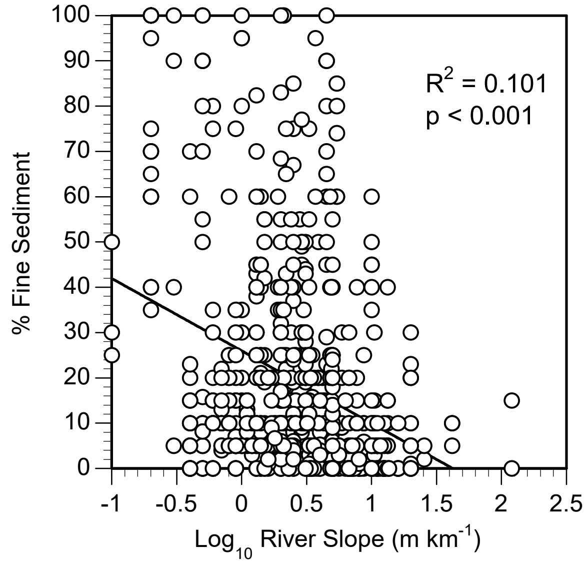
b)


Log10 Sediment Yield (t yr-1 km-2)
2.5
2
1.5
1
0.5
0
0
20
40
60
80
100
Percentile

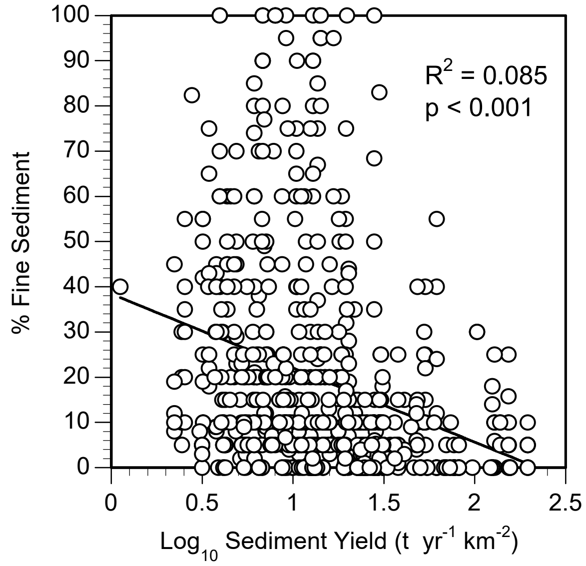
c)

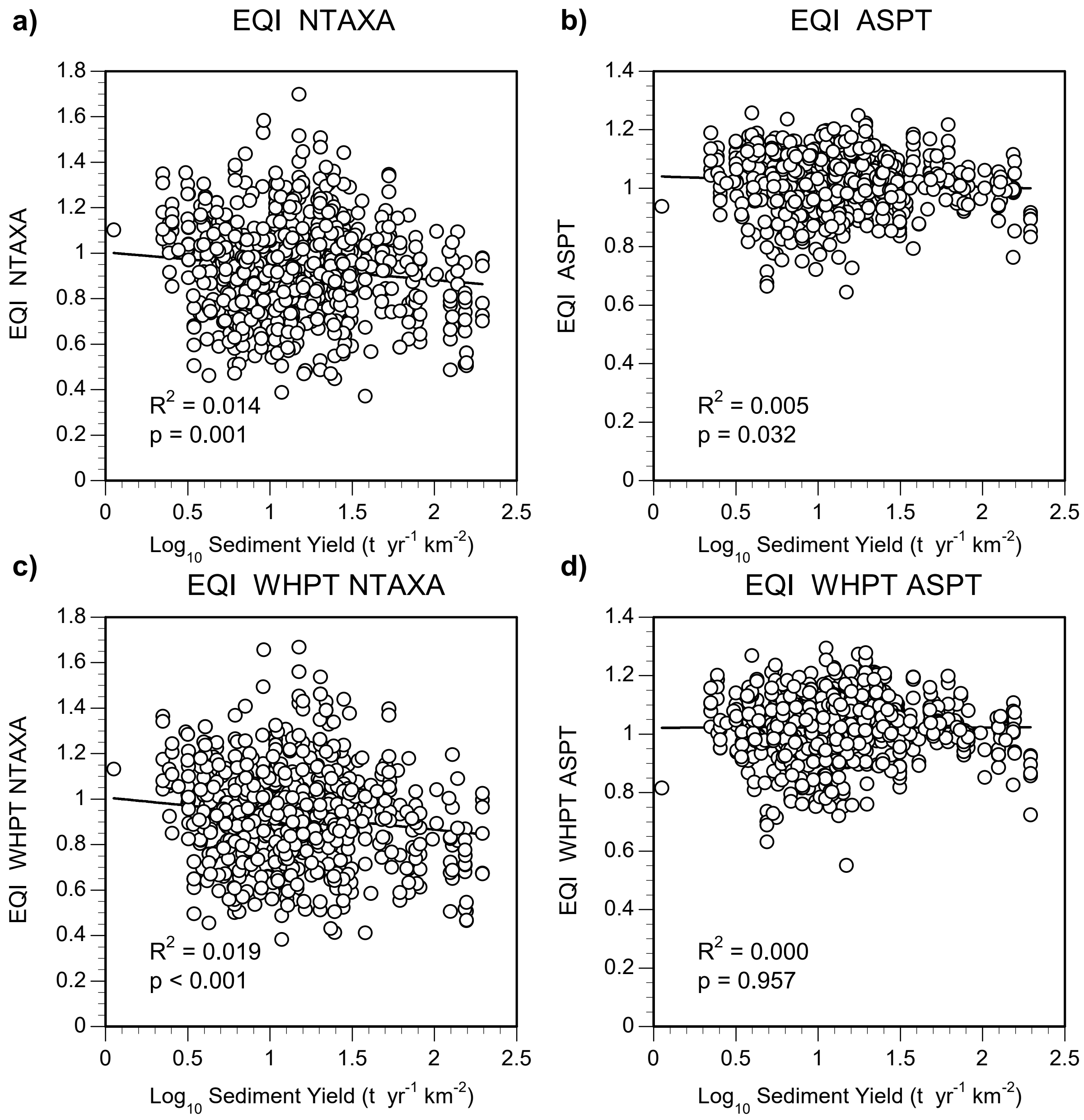

Log10 Sediment Yield (t yr-1 km-2)
2.5
2
1.5
1
0.5
0
-1
-0.5
0
0.5
1
1.5
2
2.5
Log10 River Slope (m km-1)
R2 = 0.096
p < 0.001

d)


% Fine Sediment
100
90
80
70
60
50
40
30
20
10
0
R2 = 0.101
p < 0.001
-1
-0.5
0
0.5
1
1.5
2
2.5
Log10 River Slope (m km-1)

e)


% Fine Sediment
100
90
80
70
60
50
40
30
20
10
0
R2 = 0.085
p < 0.001
0
0.5
1
1.5
2
2.5
Log10 Sediment Yield (t yr-1 km-2)


a)
EQI NTAXA
EQI NTAXA
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
R^2 = 0.014
p = 0.001
0
0.5
1
1.5
2
2.5
Log_{10} Sediment Yield (t yr^{-1} km^{-2})
b)
EQI ASPT
EQI ASPT
0
0.2
0.4
0.6
0.8
1
1.2
1.4
R^2 = 0.005
p = 0.032
0
0.5
1
1.5
2
2.5
Log_{10} Sediment Yield (t yr^{-1} km^{-2})
c)
EQI WHPT NTAXA
EQI WHPT NTAXA
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
R^2 = 0.019
p < 0.001
0
0.5
1
1.5
2
2.5
Log_{10} Sediment Yield (t yr^{-1} km^{-2})
d)
EQI WHPT ASPT
EQI WHPT ASPT
0
0.2
0.4
0.6
0.8
1
1.2
1.4
R^2 = 0.000
p = 0.957
0
0.5
1
1.5
2
2.5
Log_{10} Sediment Yield (t yr^{-1} km^{-2})

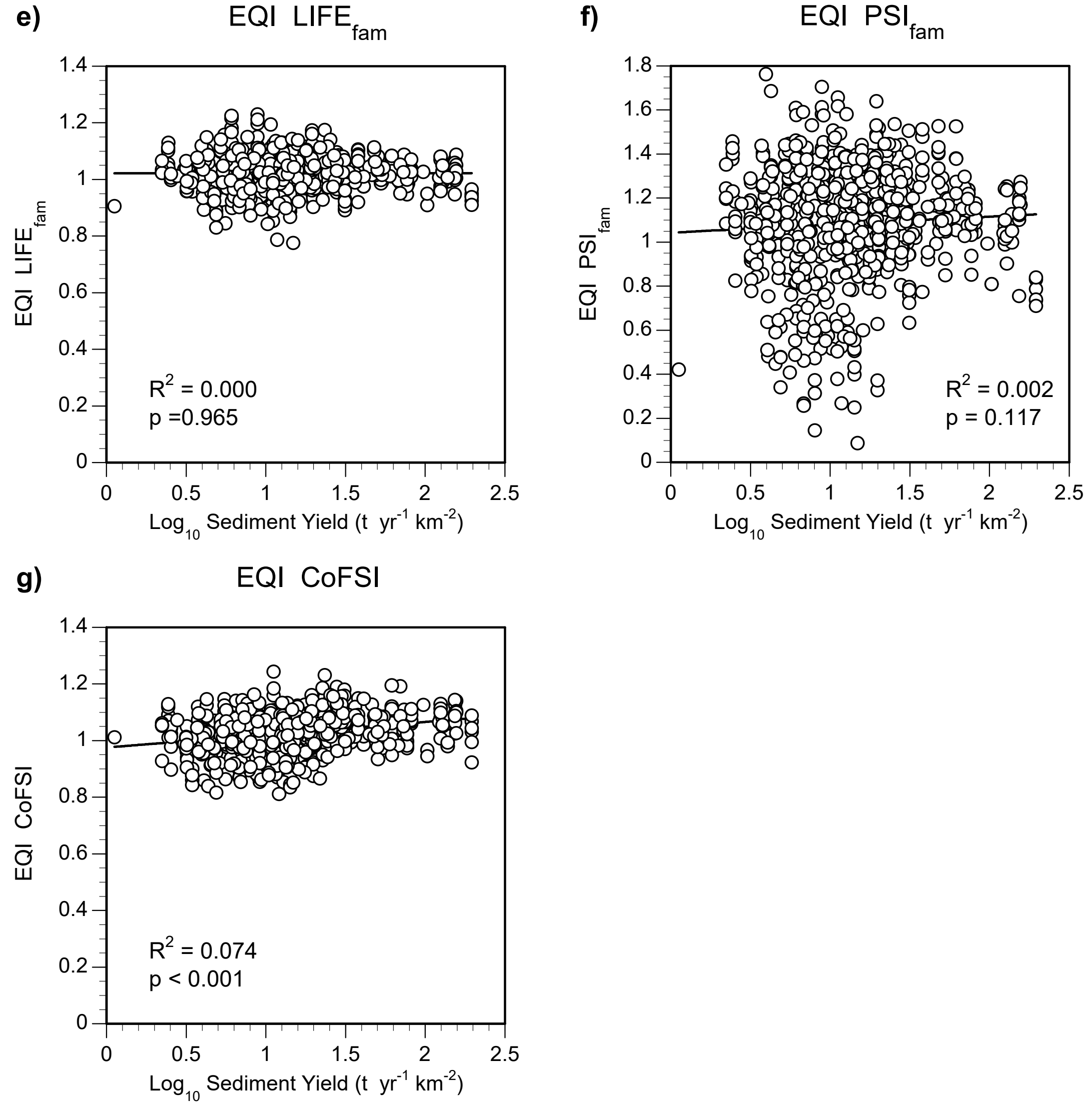
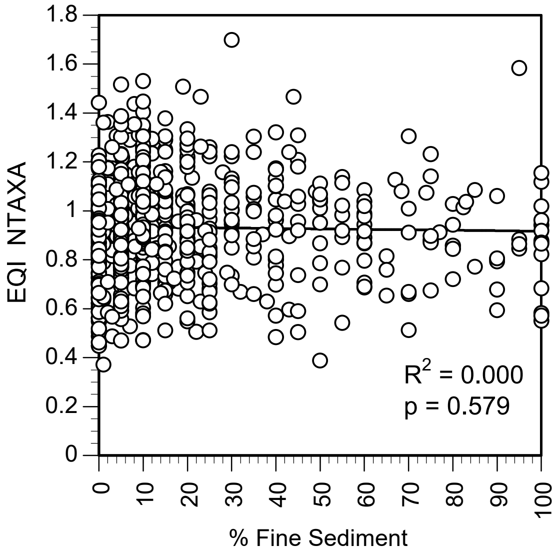
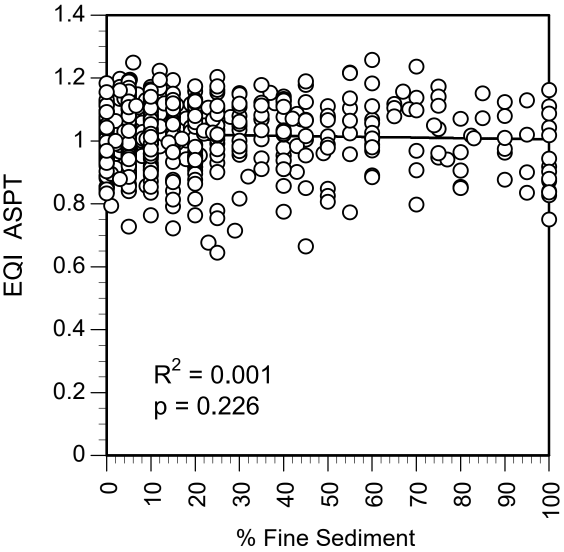


Figure 38. Relationships between excess fine sediment yield (t yr^{-1} km^{-2}) and EQI values for the invertebrate indices a) NTAXA, b) ASPT, c) WHPT$_{ab}$NTAXA, d) WHPT$_{ab}$ASPT, e) LIFE$_{fam}$, f) PSI$_{fam}$, and g) CoFSI.

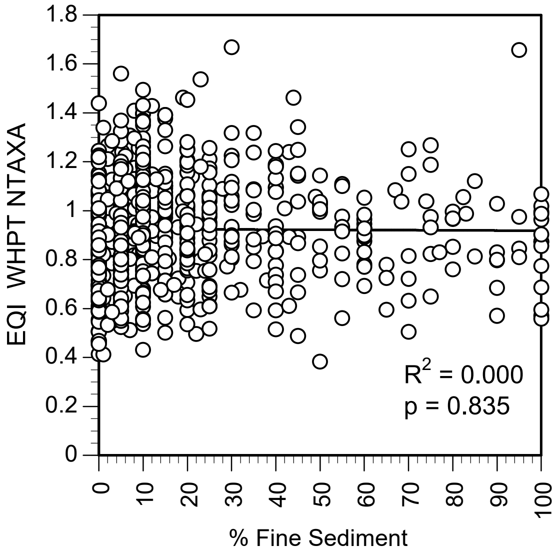
a) O/E NTAXA



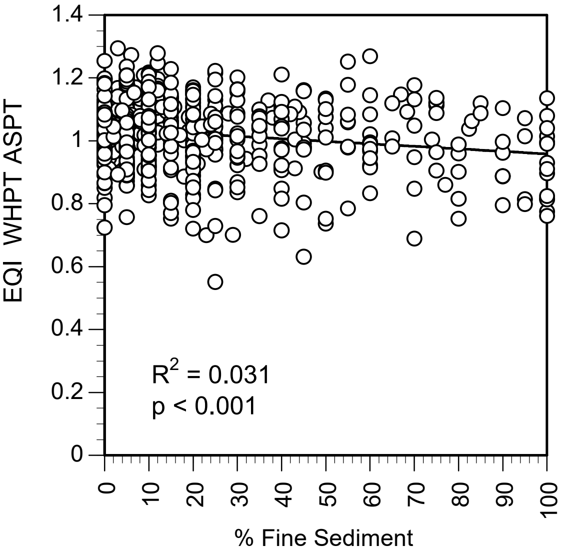
b) O/E ASPT



c) EQI WHPT NTAXA



d) EQI WHPT ASPT



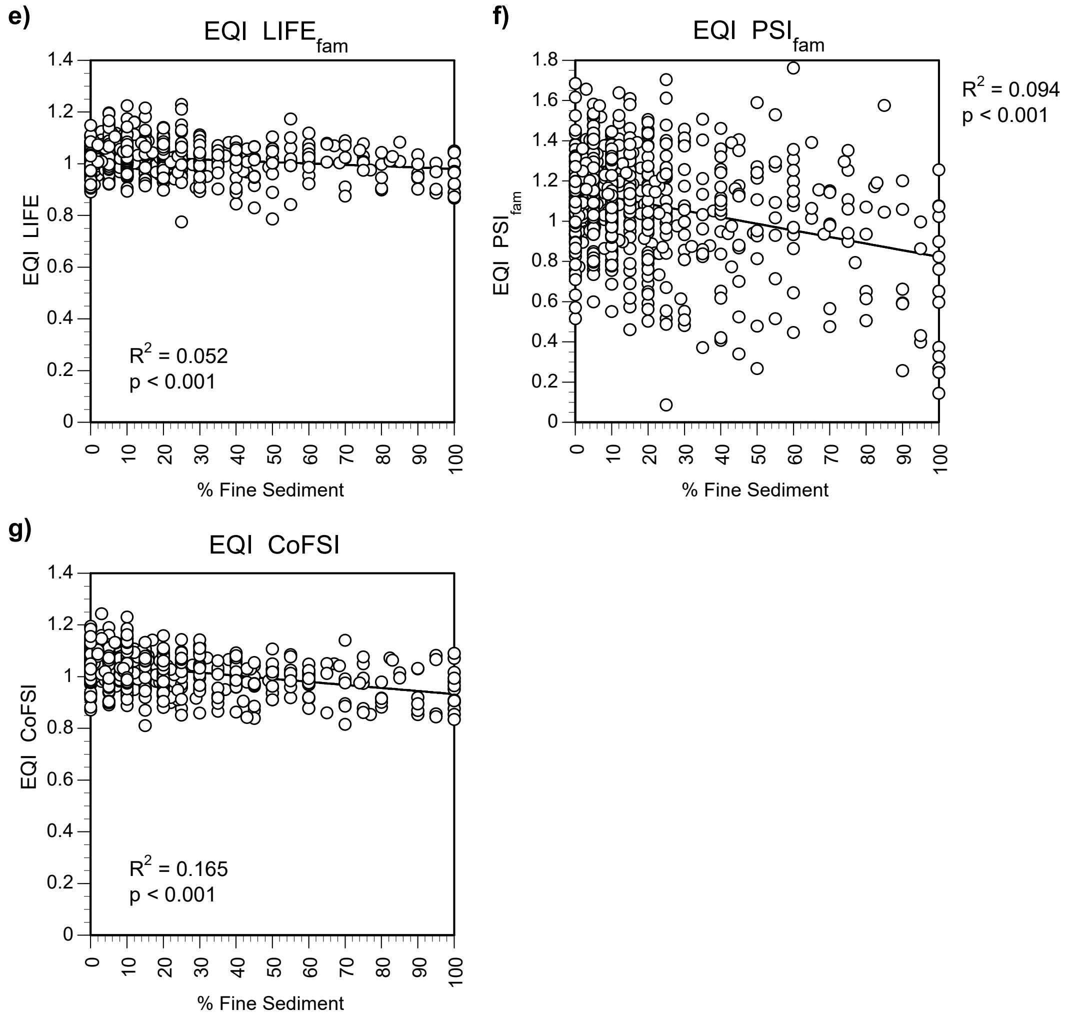


Figure 39. Relationships between percentage fines (sand, silt, clay) observed visually at the point of invertebrate sampling and EQI values for the invertebrate indices a) NTAXA, b) ASPT, c) WHPT$_{ab}$NTAXA, d) WHPT$_{ab}$ASPT, e) LIFE$_{fam}$, f) PSI$_{fam}$, and g) CoFSI.

7.4.2 Flow Stress

a) Relationship between LIFE and low flow summary statistics

The objective here was to determine the relationship between EQI for $LIFE_{fam}$ and a gradient of pressure from low flow. It is acknowledged here that variation in flow does not necessarily represent a human induced pressure. However, it is necessary to establish the response of $LIFE_{fam}$ to low flow conditions as a first step towards separating out those sites/occasions where flows are modified by human activities, such as abstraction or regulation. EQI for LIFE was used in these analyses as this measure represents the LIFE score relative to that which would be expected from an environmentally-similar site that is not subject to pressure from modification of flow (see Section 7.3).

A GB-wide dataset of EA and SEPA macroinvertebrate sampling sites spatially matched to gauging stations (n=434) was used for the analysis. A variety of summary flow statistics were derived from the NRFA data (see section 7.1.7). Of these summary flow statistics, two measures of low flow during the 6-month period preceding the spring and autumn invertebrate sampling windows were used, Q_{95} and minimum discharge. Q_{95} is the flow in cubic metres per second which was equalled or exceeded for 95% of the flow record. The Q_{95} flow is a measure of actual discharge and a significant low flow parameter, particularly relevant in the assessment of river water quality consent conditions. Minimum discharge is a less robust measure of low flow than Q_{95}: it is the lowest recorded discharge over the 6 month period, but does not incorporate any measure of the duration of low flow conditions which is incorporated into Q_{95}. To compare across rivers we also used both standardised and normalized (z-scores) versions of these measures of low flow (see section 7.1.7).

Relationships between EQI for $LIFE_{fam}$ and the low flow summary statistics were derived for all samples, irrespective of site and season (

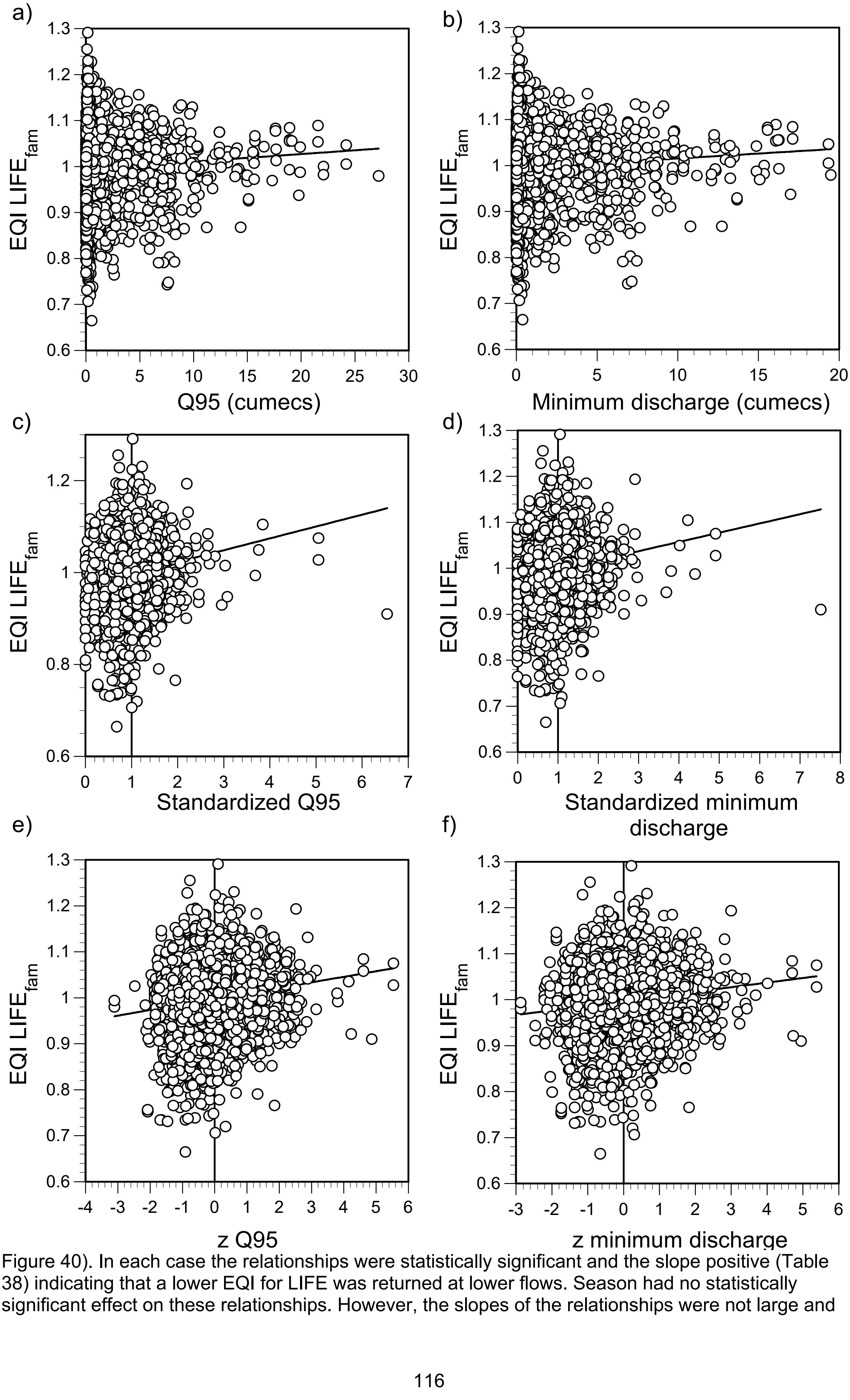


Figure 40). In each case the relationships were statistically significant and the slope positive (Table 38) indicating that a lower EQI for LIFE was returned at lower flows. Season had no statistically significant effect on these relationships. However, the slopes of the relationships were not large and

very little (less than 2%) of the variation in EQI for LIFE was explained by variation in either absolute or relative low flow summary statistics (Table 38). A wide scatter of EQI for LIFE was returned for sites and occasions where low flows were apparent, with values ranging from approximately 0.8 to 1.2 when flows were lowest. At this national scale of comparison encompassing a wide variety of river types over long time periods, EQI for LIFE was unable to detect occasions where flows were lowest.

To explore the mismatch between this apparent lack of response in EQI for LIFE to discharge and previous reports indicating a relationship between LIFE and discharge (Extence *et al* 1999; Dunbar *et al* 2010 a & b; Dunbar *et al* 2011), site was incorporated as a random factor into the statistical models relating low flow summary statistics to EQI for LIFE_{fam}.

The addition of site to the statistical models improved the power of the model to explain variation in EQI for LIFE_{fam} substantially ($R^2 > 0.80$: Table 38). However, these models indicated that the overall effect of the low flow summary statistics was not significant, whereas the interaction between the low flow summary statistics and site was highly significant in all cases. This suggests that the relationship between EQI for LIFE_{fam} and the low flow summary statistics was site specific (Figure 41). One possible explanation for such site-specific relationships between EQI for LIFE and low flow could be that the relationship between discharge and velocity is site specific: LIFE_{fam} may be detecting the response of the invertebrates to velocity rather than discharge *per se*. Whilst such site-specific relationships between EQI for LIFE_{fam} and discharge may confound detection of sites suffering flow stress at large scales, they do not detract from the use of LIFE_{fam} for detecting stress at individual sites.

Table 38. Summary of statistical results for relationships between EQI for LIFE_{fam} for individual samples and low flow summary statistics for all sites matched to NRFA stations. Results for models including just low flow summary statistics (Single factor models), and low flow summary statistics combined with sites (Models including site) are given. By adding 'Site' as a random factor to the models the importance of site-specific factors to the relationship between EQI for LIFE_{fam} and low flow summary statistics was quantified. The influence of season (spring or autumn invertebrate samples) on relationships was also assessed.

	p	**R^2**	**Estimate**	**SE**
Single factor models				
Minimum discharge	0.0004	0.0030	0.00142	0.00040
Q_{95}	0.0006	0.0028	0.00117	0.00034
Standardised minimum discharge	<.0001	0.0149	0.02018	0.00263
Standardised Q_{95}	<.0001	0.0150	0.02202	0.00285
Normalized minimum discharge	<.0001	0.0163	0.01006	0.00125
Normalized Q_{95}	<.0001	0.0163	0.01005	0.00125
Models including site				
Minimum discharge	0.7638		0.55924	1.04084
Season * minimum discharge	0.1357			
Site	<.0001			
Site * minimum discharge	0.0006			
Model	<.0001	0.8098		
Q_{95}	0.5795		0.21148	0.63654
Season * Q_{95}	0.2931			
Site	<.0001			
Site * Q_{95}	0.0210			
Model	<.0001	0.8118		
Standardised minimum discharge	0.7822		0.24712	0.36466
Season * Standardised minimum discharge	0.6792			
Site	<.0001			
Site * Standardised minimum discharge	0.0031			
Model	<.0001	0.8098		
Standardised Q_{95}	0.5783		0.80559	0.2427
Season * Standardised Q_{95}	0.4617			
Site	<.0001			
Site * Standardised Q_{95}	0.0068			
Model	<.0001	0.8118		
Normalized minimum discharge	0.5508		0.04209	0.0705
Season * Normalized minimum discharge	0.8632			
Site	<.0001			
Site * Normalized minimum discharge	0.0115			
Model	<.0001	0.8098		
Normalized Q_{95}	0.6831		0.03248	0.06708
Season * Normalized Q_{95}	0.5549			
Site	<.0001			
Site * Normalized Q_{95}	0.0235			
Model	<.0001	0.8118		

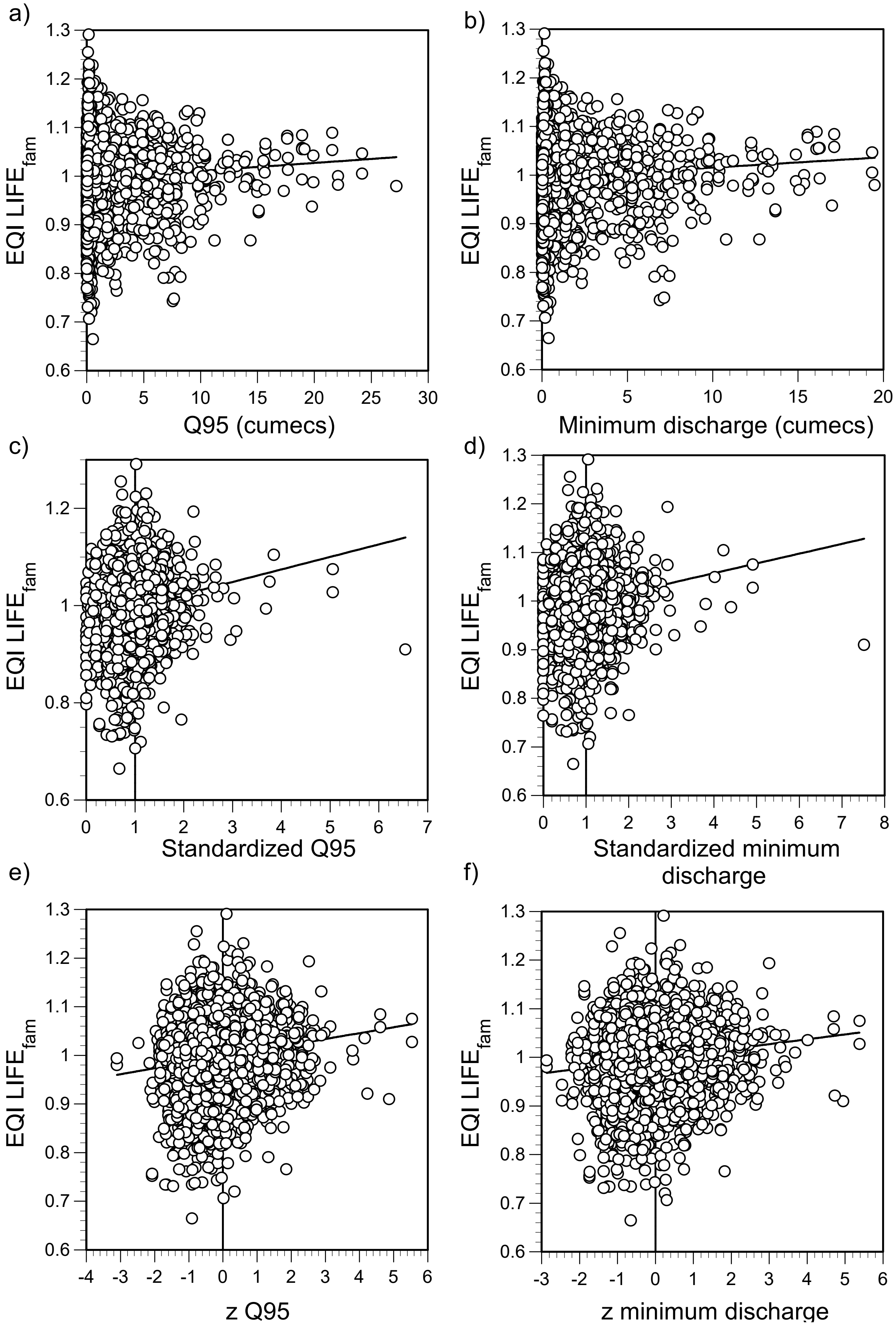


Figure 40. Relationships between EQI for LIFE$_{fam}$ for individual samples and low flow summary statistics (Q$_{95}$ and minimum discharge) for all sites matched to NRFA stations, a) & b) unstandardised, c) & d) standardised, e) & f) normalized (z-score). For details of relationships see 7.4.2 Flow Stress.

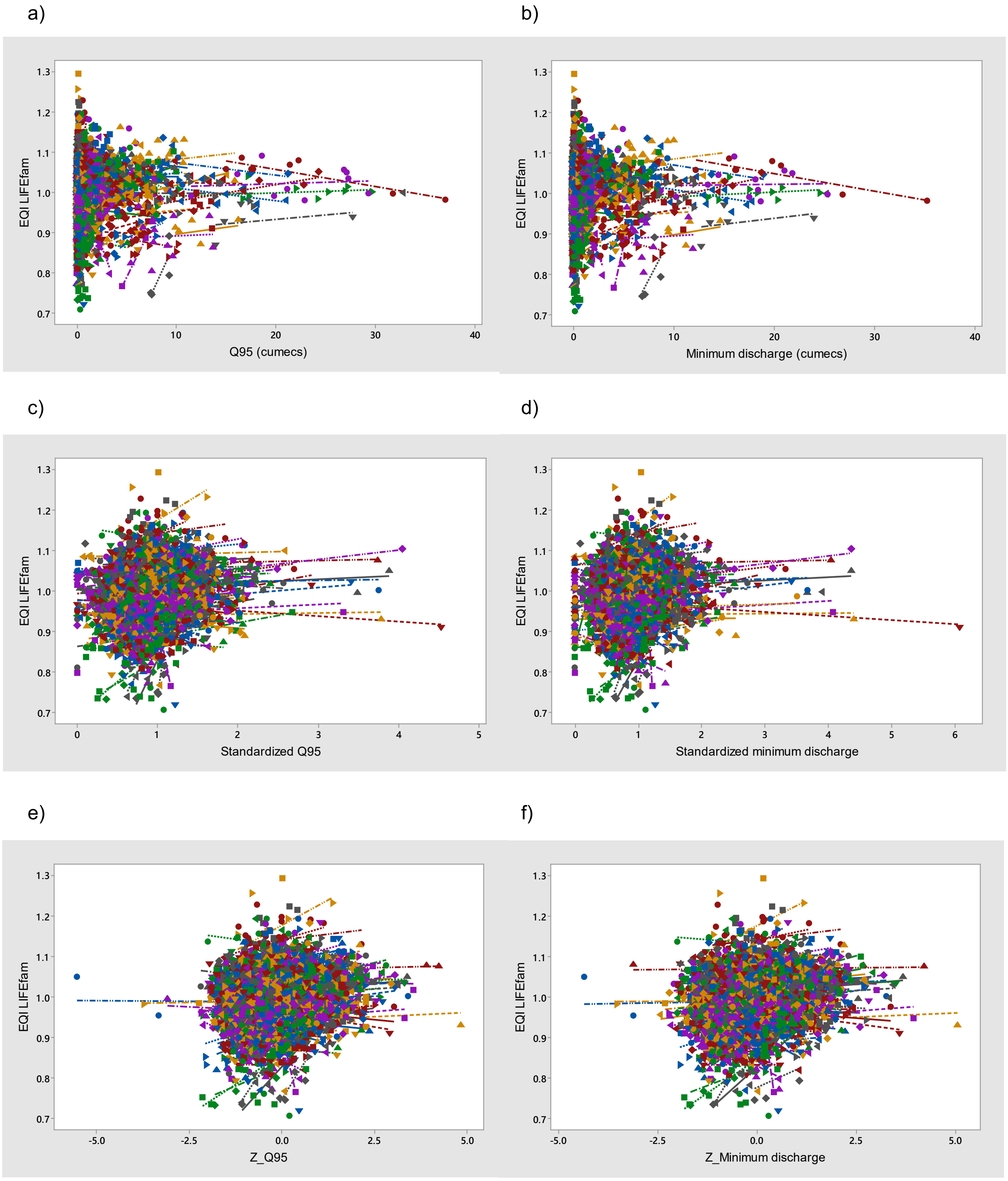


Figure 41. Relationships between EQI for LIFE$_{fam}$ for individual samples and low flow summary statistics for all sites matched to NRFA stations including site as a factor, a) Q_{95}, b) minimum discharge, c) standardised Q_{95}, d) standardised minimum discharge, e) normalized Q_{95}, and f) normalized minimum discharge. Individual sites represented by different symbols and lines. For details of relationships, see Table 38

b) Relationship between LIFE and the flow pressure gradient

The objective here was to determine the relationship between EQI for LIFE_{fam} and a gradient of pressure due to modification of flow by human activities. This is necessary to verify that EQI for LIFE_{fam} is related to the level of stress perceived to be operating at a site. The response of EQI for LIFE_{fam} to the pressure gradient could also be used to determine WFD-compliant class boundaries. Specifically, the distribution of EQI values for sites considered to have natural river flow could be used: either the lower/upper 5 percentile or lower/upper 10 percentile of the unadjusted EQI values for LIFE_{fam} for those sites with natural or near-natural flow could be used as trial values for the good-moderate WFD status class boundary. Furthermore, these boundaries could be assessed against EQI values for LIFE_{fam} for those sites where the flow is significantly affected by regulation, abstraction or augmentation (over-supply). Use of the upper 5 percentile or 10 percentile as a boundary would enable the issue of over-supply to be addressed if LIFE_{fam} returns unadjusted EQI values of greater than 1 where flows are augmented.

As the UK Hydrometric Register contains details of the factors affecting runoff within the catchment upstream of each gauging station it was possible to categorize the sites that had been matched to gauging stations by their Factors Affecting Runoff (F.A.R.) code.

The F.A.R. codes provide an indication of the various types of artificial influences operating within the catchment which alter the natural runoff. F.A.R. codes enable separation of those sites where flows are natural or near natural (i.e. there are no abstractions and discharges or the variation due to them is so limited that the gauged flow is considered to be within 10% of the natural flow at, or in excess of, the Q_{95} flow), from those that are impacted by regulation, abstraction, or augmentation to such an extent that flows are not within 10% of the natural flow at, or in excess of, the Q_{95} flow. For some areas the allocation of F.A.R. codes is incomplete and for all catchments the codes are subject to continuing review. As the absence of F.A.R. codes does not imply a natural flow regime, sites matched to a station that had no F.A.R. code were excluded. An explanation of the F.A.R. code letters is given below. With the exception of the induced loss in surface flow resulting from underlying groundwater abstraction, these codes and descriptions refer to quantifiable variations and do not include the progressive, and difficult to measure, modifications in flow regimes related to land use changes. Until recently, assignment of F.A.R. codes has been largely determined by expert local judgement of the magnitude of the impact of artificial influences at individual gauging stations. Access to the Low Flows 2000 (Young *et al* 2003) and other databases, is beginning to allow a more objective and quantitative approach to assignment of F.A.R. codes. Low Flows 2000 assessments of artificial influences have been used in the UK Hydrometric Register 2008, generally for stations commissioned during the last decade, to guide F.A.R. designations.

F.A.R. CODE EXPLANATION

S Storage or impounding reservoir. Natural river flows is affected by water stored in a reservoir situated in, and supplied from, the catchment above the gauging station.

R Regulated river. Under certain flow conditions the river is augmented from surface water and/or groundwater storage upstream of the gauging station.

P Public water supplies. Natural runoff is reduced by the quantity abstracted from a reservoir or by a river intake if the water is conveyed outside the gauging station’s catchment area.

G Groundwater abstraction. Natural river flow may be reduced or augmented by groundwater abstraction or recharge. This category includes the diminishing number of catchments where mine-water discharges influence the flow regime.

E Effluent return. Outflows from sewage treatment works that augment the river flow where the effluent originates from outside the catchment.

I Industrial and agricultural abstractions. Direct industrial and agricultural abstractions from surface water and from groundwater may reduce the natural river flow.

H Hydro-electric power. The river flow is regulated to suit the need for power generation; catchment to catchment diversions may also significantly affect average runoff.

Except for a very small set of gauging stations for which the net variation (i.e. the sum of abstractions and discharges) is assessed in order to derive the 'naturalised' flow from the gauged flow, the record of individual abstractions, discharges and changes in storage, as indicated in the F.A.R. codes is not available. Hence, sites were categorized into the following pressure categories:

Natural Flow

Stations with a natural flow as defined by the UK Hydrometric Register (2008).
F.A.R. code N
These sites are not under any pressure from modification of flow.

Near Natural Flow

Stations with a F.A.R. code including N, where other factors affecting runoff have been recorded, but the abstractions and discharges or the variation due to them is so limited that the gauged flow is considered to be within 10% of the natural flow at, or in excess of, the Q95 flow.
F.A.R. codes EN, GEN, GIN, GN, IN, PEIN.
These sites are not likely to be under substantial pressure from modification of flow.

Regulated Flow

Stations with a F.A.R. code indicating that the flows are regulated by augmentation from surface water and/or groundwater storage, storage in a reservoir situated in, and supplied from, the catchment above the gauging station, or to suit the need for power generation.
F.A.R. codes S, SR, SH, R, H.
These sites are under significant pressure from flow regulation.

Flow affected by Abstraction

Stations with a F.A.R. code indicating that the flows are significantly reduced from natural river flow by abstraction from a reservoir or by a river intake if the water is conveyed outside the gauging station's catchment area, reduced by groundwater abstraction, or by direct industrial and agricultural abstractions from surface water and from groundwater.
F.A.R. codes G, GI, I, P, PG, PGI, PI.
These sites are under significant pressure as a result of abstraction.

Augmented Flow

Stations with a F.A.R. code indicating that outflows from sewage treatment works augment the river flow as the effluent originates from outside the catchment.
F.A.R. code E.
These sites are under significant pressure from augmentation of flows (over-supply).

Further combinations of F.A.R. codes are given for stations listed in the UK Hydrometric Register 2008. However, without detailed assessment of the net variation (i.e. the sum of abstractions and discharges) from natural flow (not available at this time) it was not possible to determine the likely direction of impact of further combinations of Factors Affecting Runoff.

The lower and upper 5 and 10 percentile frequency distribution values, together with median, mean, min and max of $\mathrm{LIFE}_{\mathrm{fam}}$ EQI was calculated for the sites matched to gauging stations categorized according to F.A.R. codes, as natural flow, near natural flow, regulated flow, flow affected by abstraction and augmented flow (Table 39, Figure 42). Values for single samples and annual averages are given. The number of samples and number of sites used to derive the distributions are also shown. Expected $\mathrm{LIFE}_{\mathrm{fam}}$ values were based on the 10-variable stressor-independent model (i.e. excluding stream width, depth and bed composition).

Sites with a natural flow, as defined by the UK Hydrometric Register (2008), had a mean EQI for annual average $\mathrm{LIFE}_{\mathrm{fam}}$ of 1.01 and a median of 1.015, which suggests that the definition of natural flow is correct. The GB reference sites had a mean EQI for 2 season average $\mathrm{LIFE}_{\mathrm{fam}}$ of 0.999, and a median of 1.002, which is very close to those values returned from sites with a natural flow. These sites with a natural flow had a lower 10 percentile of 0.946 and 5 percentile of 0.922, which are only slightly higher than the EQI for 2 season average $\mathrm{LIFE}_{\mathrm{fam}}$ of the GB reference sites, 0.940 and 0.916

respectively (Table 37 and Table 39). However, the values for the GB reference sites are derived from a larger number of sites and samples, which may explain any slight difference.

The sites with near-natural flow had a mean annual average EQI for $LIFE_{fam}$ of 0.99, again suggesting that the definition was correct. These sites with near natural flow had a lower 10 and 5 percentile of 0.932 and 0.910, which was slightly lower than the sites with natural flow, but again similar to the GB reference sites. As the flow at these sites is only subject to minor modifications from a natural flow the lower 10 and 5 percentile could be used to provide a value for the Good-Moderate boundary. However, these values are derived from far fewer sample years within fewer sites, and it might be preferable to combine sites of natural and near natural flow to provide a value for the Good-Moderate boundary.

The mean EQI for annual average $LIFE_{fam}$ from sites that were under pressure as a consequence of regulation of flows was 0.992, with lower 5 and 10 percentiles of 0.902 and 0.937. These values were slightly lower than those for the sites with natural flow, but broadly equivalent to the corresponding values for the GB reference sites (taking into consideration the higher number of samples in the GB reference dataset).

The EQI for annual average $LIFE_{fam}$ from sites that were under pressure as a consequence of abstraction had a mean of 1.005, equivalent to sites with natural flow (1.010) and the GB reference sites (0.999). However, the 10 and 5 percentiles indicated a wider distribution of both low and high values: the lower 10 and 5 percentiles were 0.911 and 0.878, compared with 0.946 and 0.922 respectively for natural flow and 0.940 and 0.916 respectively for the GB reference sites. This suggests that the sites under significant pressure from abstraction may return some lower values than those with natural flow. However, the upper 10 and 5 percentiles, 1.088 and 1.126 respectively, were higher than the equivalent percentiles for sites with natural flow and the GB reference sites.

It had been expected that the EQI for $LIFE_{fam}$ of sites where the flow was augmented by water originating from outside the catchment would be in excess of 1, indicating over-supply. This was not the case: the maximum of the EQI for $LIFE_{fam}$, and the upper 10 and 5 percentiles were comparable to sites with natural flow. However, the mean for sites with augmented flow was 0.961, and the lower 10 and 5 percentiles were 0.879 and 0.851, the lowest of all the pressure categories. This suggests that the sites under significant pressure from augmentation (over-supply) may return some lower values than those with natural flow. This finding appears to be consistent with the limited data available from the Altnahinch Dam study (Murphy, 2012) where lower $LIFE_{fam}$ scores appeared to be associated with over-supply during low flow periods (see Appendix 6).

Despite the UK Hydrometric Register (2008) using F.A.R. codes to define those sites where regulation, abstraction and augmentation have a significant impact on flow, there was considerable overlap in the range of the EQI for $LIFE_{fam}$ for these three categories of flow pressure with EQI $LIFE_{fam}$ from sites with natural flow. Furthermore, there was considerable overlap between the pressure categories and the GB reference sites. Although sites with abstracted and augmented flows did return some lower EQI $LIFE_{fam}$ values, there was no clear indication that the pressures on flow (regulation, abstraction, augmentation) resulted in a consistently lower EQI for $LIFE_{fam}$ than where such pressures were absent.

Nevertheless, it is possible that the considerable overlap in the range of the EQI for $LIFE_{fam}$ may be a consequence of interannual variablity in runoff: it may be that the pressure from factors affecting runoff only resulted in substantial impact on flow in some years. For example, in years where precipitation is high, more water must be removed by abstraction to produce a reduction in flow relative to the long-term average for the site. Hence, statistical analysis was undertaken to determine if the flow pressure categories could be distinguished using EQI for $LIFE_{fam}$ whilst taking variation in discharge into account. Three measures of discharge were used; Q_{95}, standardised Q_{95} and normalized (z-score) Q_{95} (see section 7.1.7).

General Linear Models (GLM) were used to determine the significance of the flow pressure categories on variation in the EQI for $LIFE_{fam}$, whilst including either Q_{95} ,standardised Q_{95} or normalized Q_{95} as a covariable. Site was also included to take into account the fact that repeated measurements were made at each site.

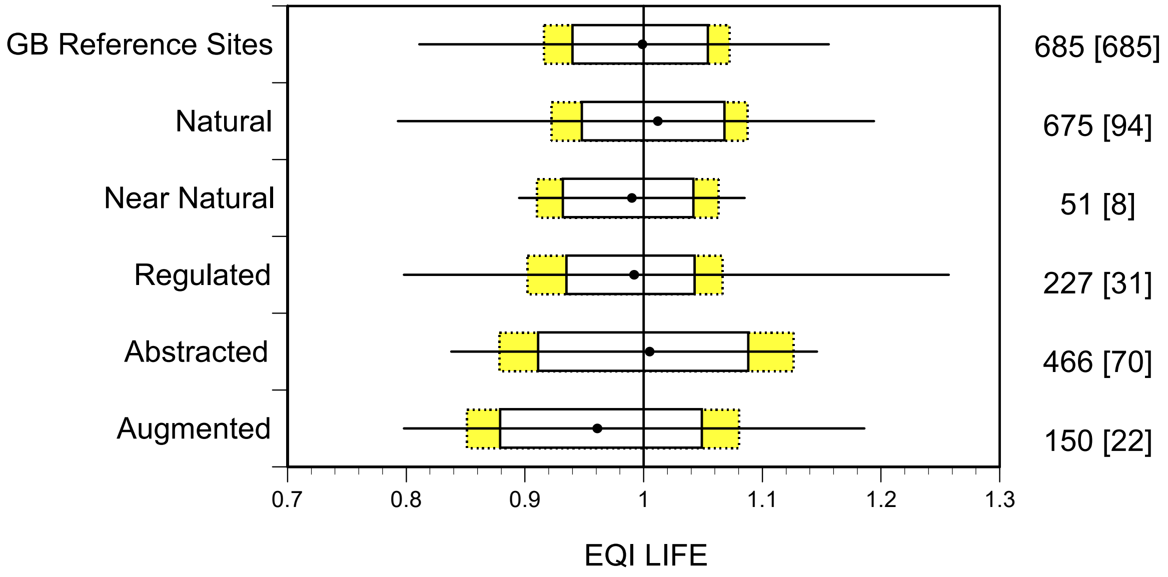


Figure 42. Lower and upper 5% (dashed box) and 10% percentile (solid box) frequency distribution values, together with mean (●) and range (line) of EQI LIFE$_{fam}$ (based on two season averages) for the sites matched to UK Hydrometric Register gauging stations by Flow Pressure Categories. The number of samples and number of sites (in parenthesis) used to derive the distributions are provided. Also shown are the frequency distribution values for the 685 GB reference sites.

Table 39. Lower and upper 5 and 10 percentile frequency distribution values, together with median, mean, min and max of the *unadjusted* EQI values for LIFE_{fam} for the sites matched to NRFA gauging stations categorized according to Factors Affecting Runoff (F.A.R.) codes, as natural flow (F.A.R. code N), near natural flow (F.A.R. codes EN, GEN, GIN, GN, IN, PEIN), regulated (F.A.R. codes S, SR, SH, R, H), abstracted (F.A.R. codes G, GI, I, P, PG, PGI, PI) and augmented flow (F.A.R. code E). The number of samples and sites used to derive the distributions are also given. For comparison the lower 5 and 10 percentile frequency distribution values, together with median, mean, min and max of the *unadjusted* EQI values for LIFE_{fam} for the GB reference sites are given (see section 5.5). Expected LIFE_{fam} values are based on 10-variable stressor-independent model (i.e. excluding stream width, depth and bed composition).

	Natural Flow	Near Natural Flow	Regulated	Abstracted	Augmented Flow	GB references sites
Annual average						
Lower 5%	0.922	0.910	0.902	0.878	0.851	0.916
Lower 10%	0.948	0.932	0.937	0.911	0.879	0.940
Upper 10%	1.068	1.042	1.043	1.088	1.049	1.054
Upper 5%	1.087	1.063	1.066	1.126	1.080	1.072
Median	1.016	0.995	0.993	1.011	0.961	1.002
Mean	1.012	0.990	0.992	1.005	0.961	0.999
Min	0.793	0.895	0.838	0.798	0.798	0.811
Max	1.194	1.085	1.146	1.257	1.186	1.156
Samples	675	51	227	466	150	685
Sites	94	8	31	70	22	685
Single season						
Lower 5%	0.919	0.908	0.901	0.880	0.849	0.903
Lower 10%	0.946	0.918	0.939	0.915	0.881	0.931
Upper 10%	1.073	1.057	1.047	1.093	1.050	1.065
Upper 5%	1.094	1.075	1.066	1.131	1.083	1.088
Median	1.016	1.002	0.994	1.011	0.954	1.002
Mean	1.012	0.994	0.992	1.008	0.958	0.999
Min	0.767	0.854	0.838	0.798	0.771	0.770
Max	1.293	1.143	1.146	1.257	1.186	1.267
Samples	991	82	315	650	199	2055
Sites	94	8	31	70	22	685

A significant (or near significant) difference was found in mean $LIFE_{fam}$ EQI values between the F.A.R. pressure categories when using Q_{95}, standardised Q_{95} or normalized Q_{95} as covariables (Table 40). The mean EQI for sites with augmented flow was significantly lower than for all other pressure categories (Figure 43). The mean EQI for sites with natural flow was not significantly different to the mean for sites where flow was affected by abstraction, but both were significantly different to the mean EQI for sites with regulated flows. The mean EQI for sites with near-natural flows was not significantly different to either of these pressure categories, although the number of replicates was low for this category. However, significant differences were also apparent when the EQIs of other indices were applied to these data. ASPT, $WHPT_{ab}ASPT$ and PSI_{fam} returned significant or near significant differences between the F.A.R. pressure categories, with particularly low values for sites with augmented flow (Figure 43). Although over-supply of high quality water appeared to be related to lower $LIFE_{fam}$ scores in the Altnahinch Dam study (Appendix 6), it is possible here that EQI for $LIFE_{fam}$ was responding to organic pollution rather than over-supply, as these sites were augmented with outflows from sewage treatment works.

However, a more substantial concern is the lack of a relationship between EQI for $LIFE_{fam}$ and any measure of low flow, Q_{95}, standardised Q_{95} or normalized Q_{95} (Table 40). The interaction between the low flow summary statistics and F.A.R. pressure category was not significant either, indicating that a lack of an overall response between EQI for $LIFE_{fam}$ and low flow summary statistics was not due to different responses for different pressure categories. However, the interaction between site and low flow summary statistics was significant (Table 40, Figure 44 and Figure 45), suggesting that the relationship between EQI for $LIFE_{fam}$ and these measures of low flow discharge was site-specific (Figure 44-Figure 45). Although previous works have shown relationships between LIFE indices and discharge (Extence *et al* 1999; Dunbar *et al* 2010 a & b, 2011), these works and the results here suggest that there are site-specific effects in these relationships. One possible explanation for such site-specific effects could be that the relationship between discharge and velocity is site-specific, and LIFE indices may be detecting the response of the invertebrates to velocity rather than discharge *per se*. This would not be surprising as the scores in LIFE are based on the response of taxa to velocity rather than discharge (Extence *et al* 1999). Whilst such site-specific relationships between LIFE and discharge may confound detection of sites suffering flow stress at large scales, they do not detract from the use of LIFE for detecting stress at individual sites.

Here we have compared variation in discharge both within sites and across the national scale. Furthermore, we have used the 10-variable stressor-independent model (i.e. excluding stream width, depth and bed composition) to establish expected LIFE values, which should improve the ability of RIVPACS to detect flow stress compared with previous models. Yet, at this scale we have not been able to consistently separate sites across a pressure gradient of impact by flow stress as defined by the UK Hydrometric Register (2008), where impacted sites were not within 10% of the natural flow at, or in excess of, the Q_{95} flow.

Table 40. Results of General Linear Model analysis to determine the statistical significance of F.A.R. Pressure Category on *unadjusted* EQI of $LIFE_{fam}$, whilst taking into account variation in discharge, as Q_{95}, standardised Q_{95} or normalized Q_{95} at each site.

	p
F.A.R. Pressure Category	0.0400
Site	<.0001
Q95	0.3744
F.A.R. Pressure Category * Q95	0.6550
Site * Q95	<.0001
F.A.R. Pressure Category	0.0545
Site	<.0001
Standardised Q95	0.8607
F.A.R. Pressure Category * Standardised Q95	0.2954
Site * Standardised Q95	0.0332
F.A.R. Pressure Category	0.0538
Site	<.0001
Normalized Q95	0.8037
F.A.R. Pressure Category * Normalized Q95	0.5676
Site * Normalized Q95	0.0276

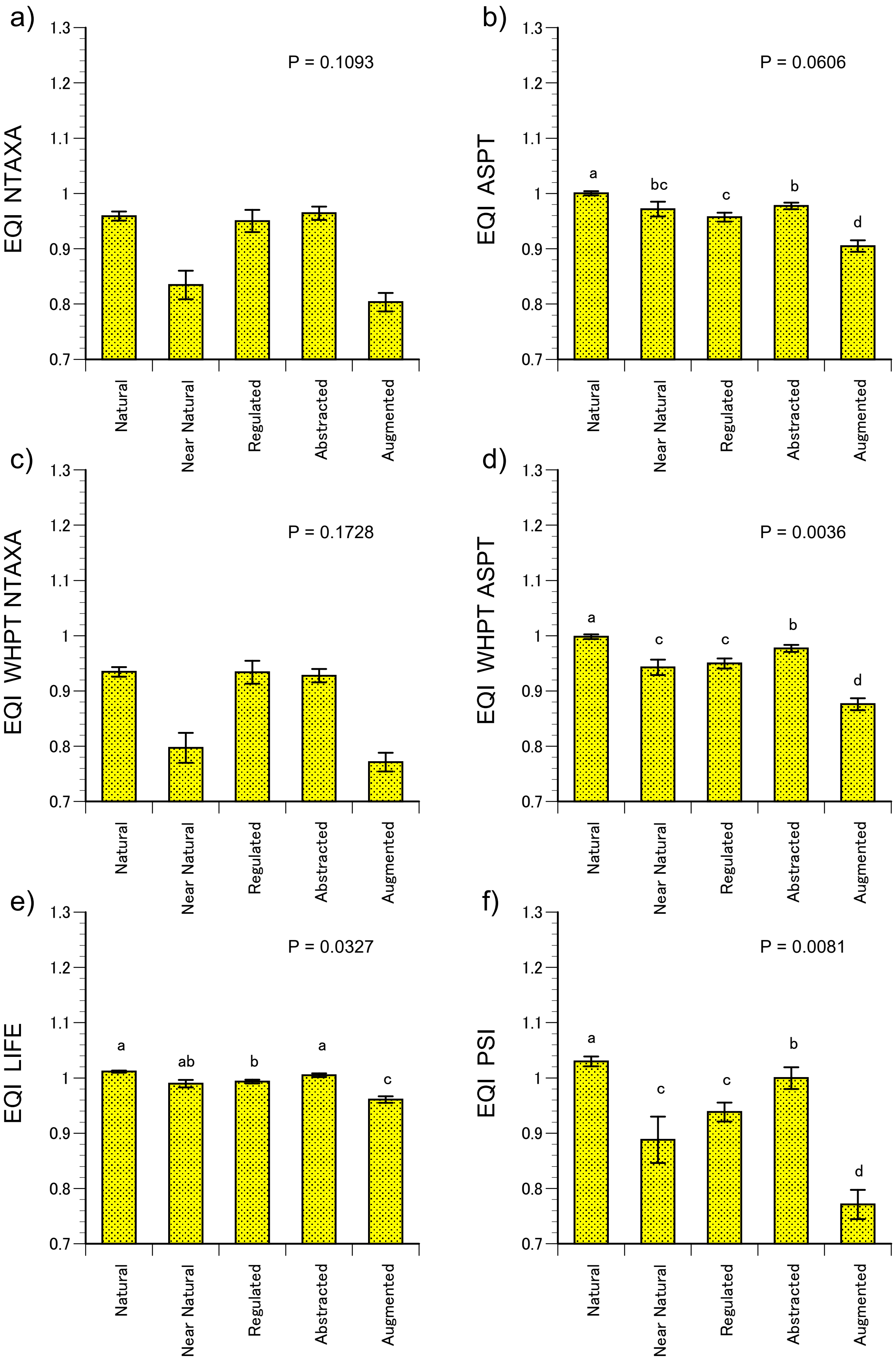


Figure 43. Mean EQI (±SE) for FAR Pressure Categories, from GLM including discharge as a covariable a) NTAXA, b) ASPT, c) $\mathrm{WHPT_{ab}NTAXA}$, d) $\mathrm{WHPT_{ab}ASPT}$, e) $\mathrm{LIFE_{fam}}$ and f) $\mathrm{PSI_{fam}}$. Differences among means identified by Tukey's test; means that share a letter are not significantly different.

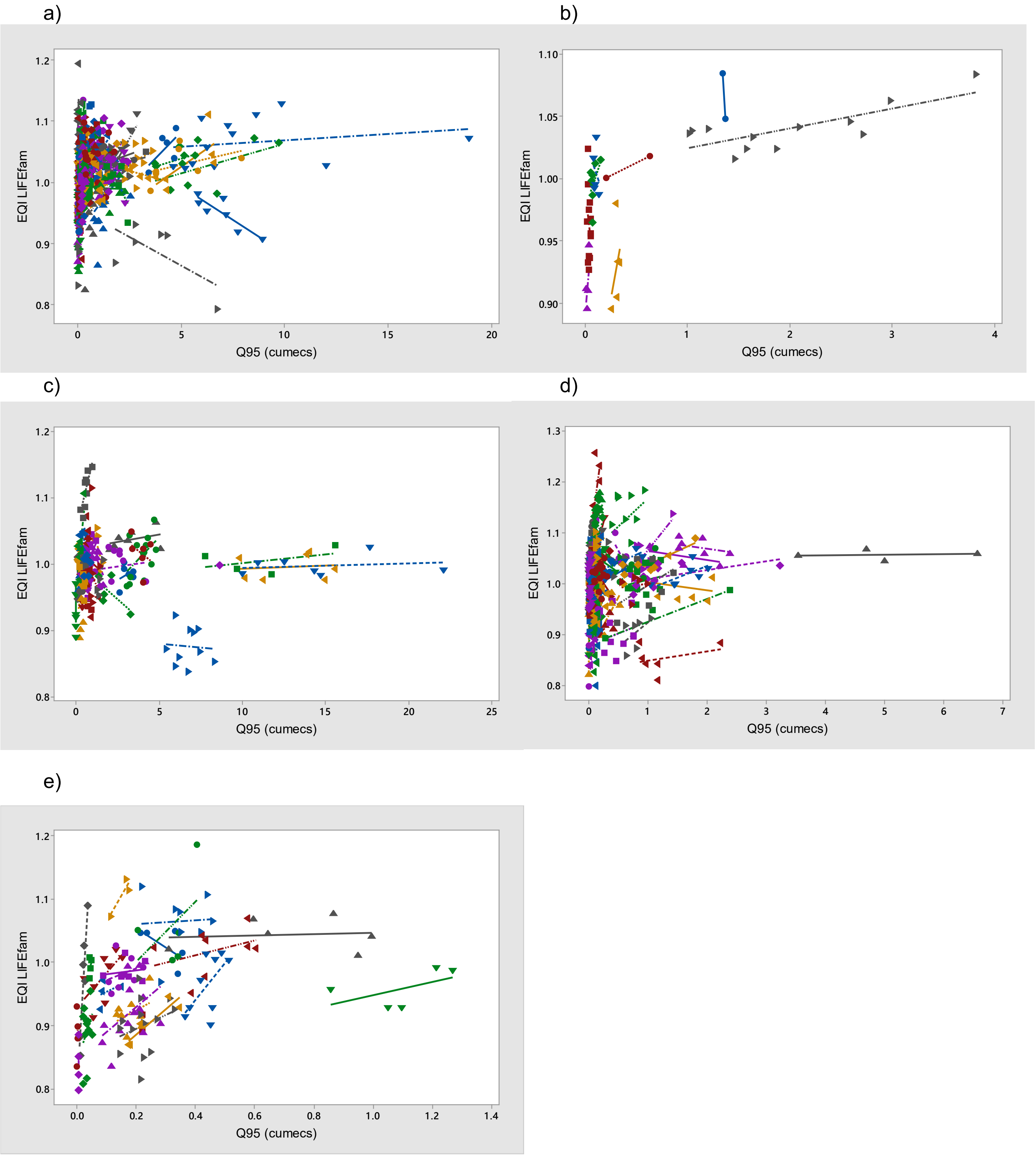


Figure 44. Relationships between EQI LIFE$_{fam}$ and Q95 for individual sites within each F.A.R. flow pressure category a) sites with natural flow, b) sites with near natural flow, c) sites with regulated flow, d) sites with flow affected by abstraction, and e) sites with augmented flow. Individual sites are represented by different symbols and lines.

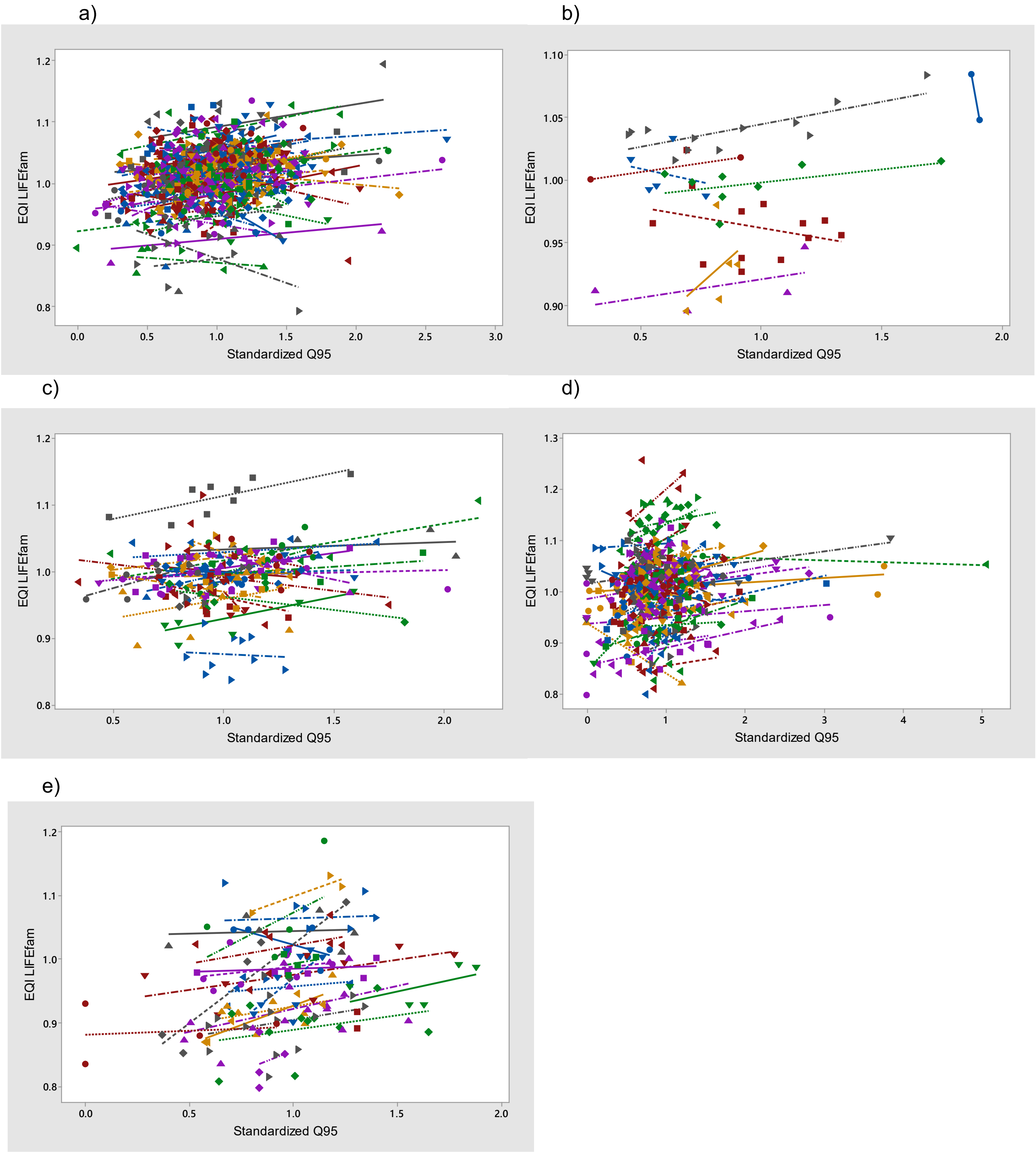
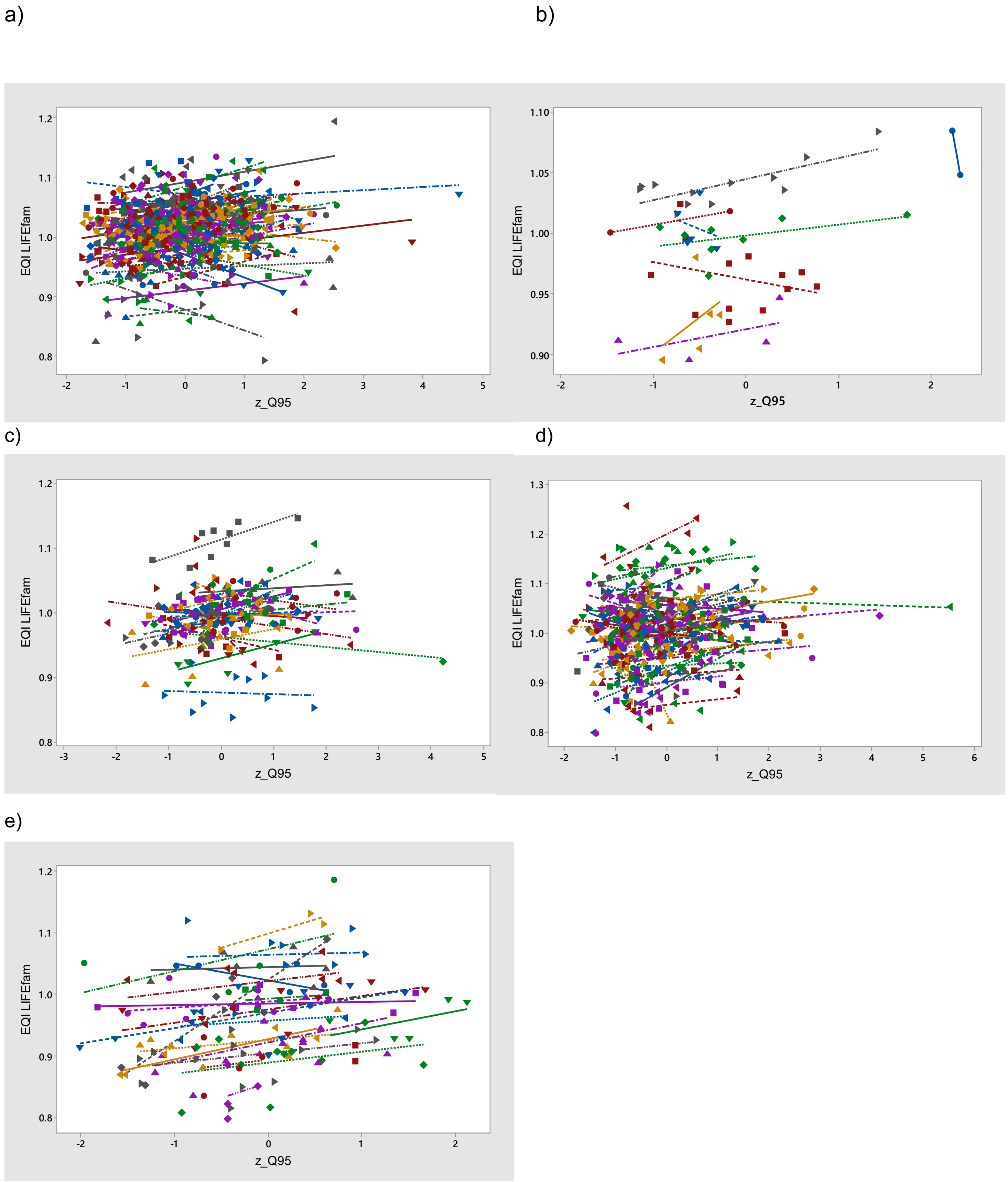


Figure 45. Relationships between EQR LIFE$_{fam}$ and standardised Q95 for individual sites within each F.A.R. flow pressure category a) sites with natural flow, b) sites with near natural flow, c) sites with regulated flow, d) sites with flow affected by abstraction, and e) sites with augmented flow. Individual sites are represented by different symbols and lines. Note, a standardised Q$_{95}$ value of 1 indicates that the Q$_{95}$ for the year in question was equal to the long term average for that site.


a)

b)

c)

d)

e)

Figure 46. Relationships between EQR LIFE$_{fam}$ and normalized (z-score) Q95 for individual sites within each F.A.R. flow pressure category a) sites with natural flow, b) sites with near natural flow, c) sites with regulated flow, d) sites with flow affected by abstraction, and e) sites with augmented flow. Individual sites are represented by different symbols and lines. Note, a normalized Q$_{95}$ value of 0 indicates that the Q$_{95}$ for the year in question was equal to the long term average for that site.

Summary of relationship between LIFE and flow pressure gradient

Here we have used a dataset for LIFE–flow stress analysis comprising 434 sites from throughout GB, with multiple years' data at each site spanning the period 1994 to 2012. EQI values have been established for LIFE_{fam} using a stressor-independent RIVACS model for each of the 4,176 sampling occasions, and average values derived for each Invert-Year to relate to low flow summary statistics provided by the NRFA. Furthermore, we have used an independent assessment of pressure due to modification of flow by human activities as assessed by the UK Hydrometric Register (2008). Yet, at this scale we have not been able to consistently separate sites across a pressure gradient of impact by flow stress as defined by the UK Hydrometric Register, where impacted sites were not within 10% of the natural flow at, or in excess of, the Q_{95} flow, nor have we been able to establish a clear relationship between EQI values for LIFE_{fam} and low flow summary statistics for the antecedent period. The relationship between EQI values for LIFE and the flow pressure gradient appears to be confounded by site specific effects. It should be noted that the expected value produced for each site by the stressor-independent RIVACS model is time-invariant; therefore, any variation in EQI within a site is due to variation in the observed LIFE values. One possible explanation for such site-specific relationships with low flow could be that the relationship between discharge and velocity is site specific: LIFE may be detecting the response of the invertebrates to velocity rather than discharge *per se*. Whilst such site-specific relationships between EQI for LIFE and discharge may confound detection of sites suffering flow stress at the national scale, they do not detract from the use of LIFE for detecting stress at individual sites.

Summary:
Despite using large scale and long-term data (434 sites from throughout GB, with multiple years' data at each site spanning the period 1994 to 2012) we have not been able to establish a clear relationship between EQI values for LIFE_{fam} and low flow summary statistics for the antecedent period. Furthermore, we have not been able to consistently separate sites across a pressure gradient of impact by flow stress as defined by the UK Hydrometric Register (2008). The relationship between EQI values for LIFE and the flow pressure gradient appears to be confounded by site-specific effects. Such site-specific relationships between EQI for LIFE and discharge will confound detection of sites suffering flow stress at the national scale. However, they do not detract from the use of LIFE for detecting how individual sites respond to flow stress.

7.5 Deriving empirically based WFD class boundaries

In order to guide selection of the classification boundaries for the LIFE and PSI indices we plotted the proportion of individuals classed as stressor-sensitive taxa and those classed as stressor-tolerant against the EQI for the index. Overlaying both relationships on the same graph it is possible to use the intersection point of their generalised additive modelled response as the Good/Moderate boundary. This approach has been used successfully in setting boundaries for the Trophic Diatom Index EQIs within the phytobenthos bioassessment tool (DARES) (Kelly *et al* 2008).

Generalised additive models (GAMs) do not impose a shape of relationship between the response and predictor variables. Instead, non-parametric smoothers are used to describe the relationship in the most parsimonious way possible, as judged by information criteria (Wood, 2006). All GAMs were fitted with Gaussian error terms and an "identity" link function; this assumes a normal distribution for the dependent variable and therefore does not apply any transformation. While the distribution of the % Sensitive and % Tolerant values was never truly Gaussian, by applying an Arcsine-square root transformation to the data it approximated a normal distribution. GAMS were fitted using the 'mgcv' package (Wood, 2006) within R 3.0.2 (R Development Core Team, 2013).

7.5.1 Family and Species level PSI classification boundaries

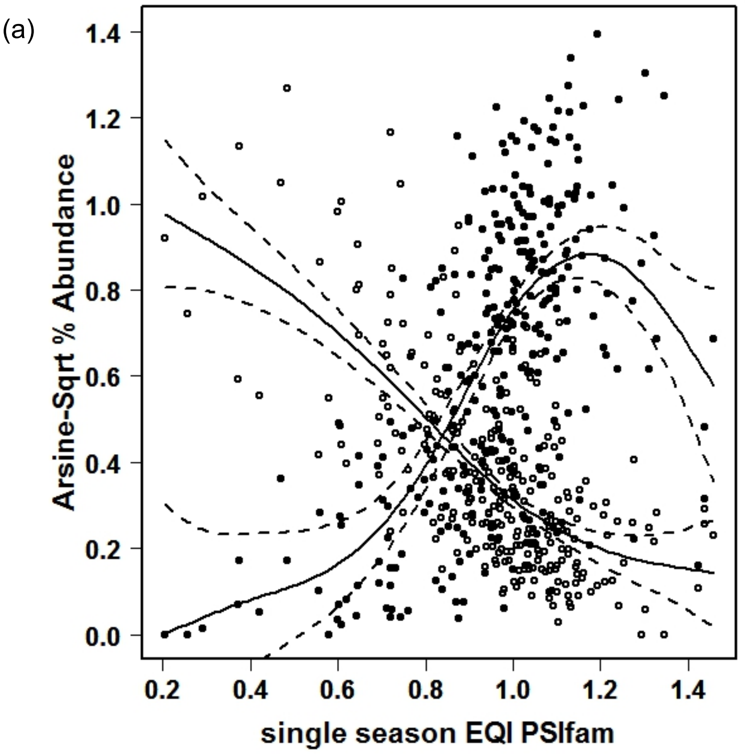
In order to derive a Good/Moderate boundary for the PSI indices we calculated the percentage of all individuals in each sample from the WQ0128 and WG-AES datasets that were categorised by the PSI scoring system as 'Highly Sensitive' and the percentage categorised as 'Highly Insensitive' (Extence *et al* 2013). Variation in both these percentage values (Arcsine square root transformed) was related to the RIVPACS IV Model 15(2)-generated EQIs and the cross-over point in the GAM line-of-best-fit through both datasets was taken as the EQI value above which highly sensitive taxa begin to dominate the macroinvertebrate assemblage and below which highly tolerant taxa begin to be more abundant. This point on the x-axis is the proposed Good/Moderate classification boundary.

This process was undertaken for both family and species-level PSI and using single season EQIs and 2-season-averaged EQIs.

The intersection for single season PSI_{fam} is at approximately 0.843 while when we ran these analyses on the two-season averaged PSI_{fam} EQI data the intersection was at approximately 0.835 (Figure 47).

The cross-over point of both single season % Sensitive and %Tolerant GAMs for PSI_{sp} was at an EQI of approximately 0.713 (Figure 48 a). For two season-averaged PSI_{sp} the two fitted GAM lines crossed at an EQI PSIsp value of approximately 0.755 (Figure 48 b).

The two season-averaged values for the Good/Moderate boundary are not that different from the single season values, but they are based on a relatively small sample size (n=44, WG-AES data only as WQ0128 dataset has only one sampling occasion at each site) with a consequent minor increase in uncertainty around the exact location of the intersection point. These four values are the proposed Good-Moderate EQI thresholds for single season and two season-averaged PSI_{fam} and PSI_{sp}.



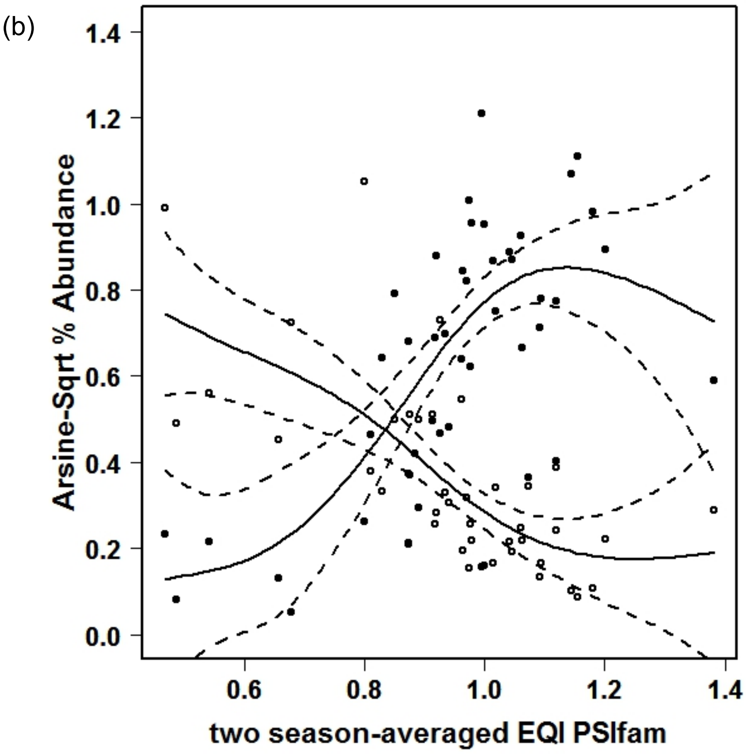
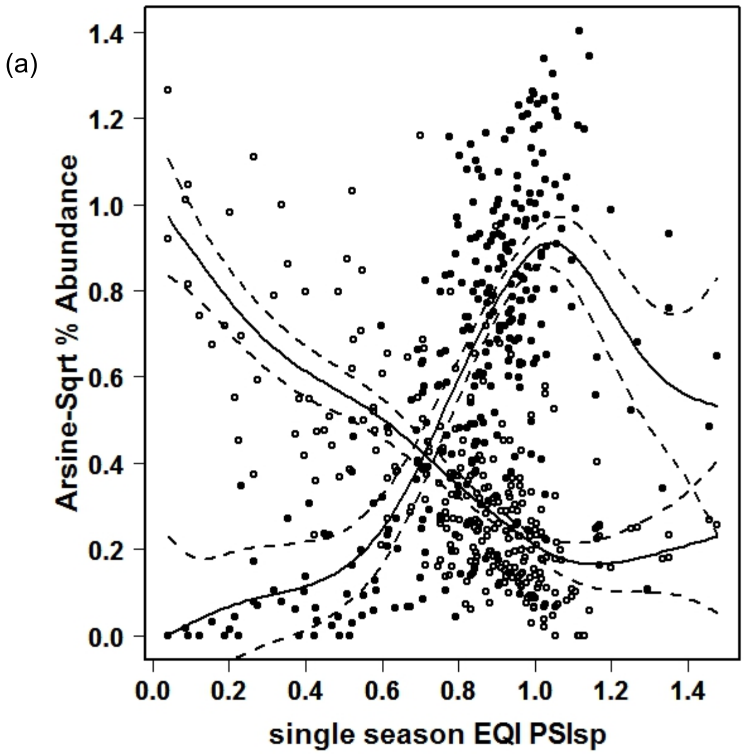


Figure 47. Variation in the relative abundance of fine sediment-sensitive (closed circles) and fine sediment-tolerant taxa (open circles) with (a) single season and (b) two-season averaged EQI PSI_{fam}. The solid lines are fitted generalised additive models with associated 95% confidence bands (dashed lines).



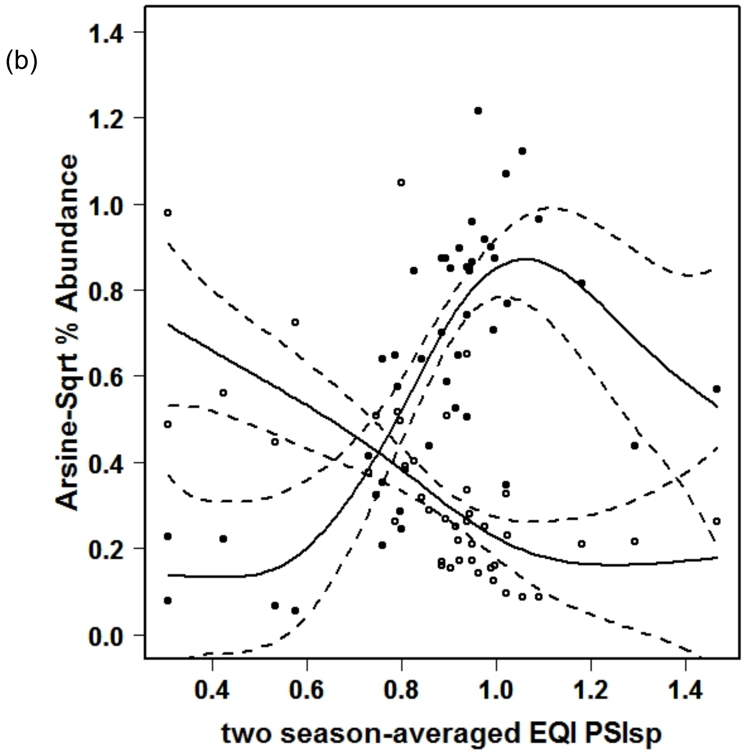


Figure 48. Variation in the relative abundance of fine sediment-sensitive (closed circles) and fine sediment-tolerant taxa (open circles) with (a) single season and (b) two-season averaged EQI PSI_{sp}. The solid lines are fitted generalised additive models with associated 95% confidence bands (dashed lines).

7.5.2 Family and Species level LIFE classification boundaries

We then repeated the analyses for $LIFE_{fam}$ using family-level data from SEPA & EA and for $LIFE_{sp}$ using MTL data from SEPA. For both indices, we calculated the percentage of individuals associated with moderate to rapid flows (LIFE Flow Groups I & II) and the percentage of individuals associated with slow-flowing, standing or drying waters (LIFE Flow Groups IV-VI) (Extence *et al* 1999).

Variation in the Arcsine square root transformed percentage abundance data was related to the RIVPACS IV Model 15(2)-generated EQIs and the cross-over point in the overlaid GAM fitted lines was taken as the EQI value above which taxa associated with faster flows begin to dominate the macroinvertebrate assemblage and below which taxa more associated with slow or no-flow conditions were more abundant. This point on the x-axis is the proposed Good/Moderate classification boundary.

This process was undertaken for both family and species-level LIFE and using single season EQIs and 2-season-averaged EQIs.

For single season $LIFE_{fam}$ the intersection of GAM-fitted lines was at an EQI $LIFE_{fam}$ of 0.959 and for the two season-averaged $LIFE_{fam}$ data the cross-over point was slightly higher at an EQI $LIFE_{fam}$ of 0.965 (Figure 49).

When this approach was applied to the $LIFE_{sp}$ index the intersection between the two modelled lines was too uncertain (Figure 50). This may be due to the poor model fit for moderate-fast flow taxa values (the GAM could explain only 18% of the variation). We therefore restricted these data to only those taxa associated with rapid flows (LIFE Flow Group I) to increase the focus on the taxa most sensitive to low-flow stress. The revised cross-over plot of the GAM fitted lines provided a better basis for determining the Good/Moderate boundary. The resultant zone of intersection was now at an EQI $LIFE_{sp}$ value of approximately 0.927. The two season-averaged GAM-fitted lines (with the flow-sensitive data based on LIFE Flow Group I taxa) intersected at an EQI $LIFE_{sp}$ value of approximately 0.934, only slightly higher than that for single season EQIs (Figure 51). However, there was a relatively high degree of uncertainty around the location of the intersection.

When compared to the proposed Good/Moderate boundaries derived using the distribution of RIVPACS Model 15(2) reference sample unadjusted EQIs (see Table 37), the single season PSI_{fam} value (0.843) lies closer to 1 than the lower RIVPACS 10%ile value (0.748). The latter distribution is quite wide (EQI PSI_{fam}: 0.110-1.864), due in part to a relatively high number of reference samples with values between 0.2 and 0.6, particularly in Super Group 7 reference sites (see Figure 32 b). The $LIFE_{fam}$ threshold for both single and two season-averaged EQIs was likewise closer to 1 than the 10%ile value derived from the distribution of reference sites values and the distribution of EQIs from the analysis of NRFA data (see section 7.4.2b) (Table 41). The GAM approach consistently returns more stringent thresholds than either of the other two approaches.

Table 41. Comparison of proposed Good/Moderate classification thresholds for LIFE and PSI indices, derived from three different approaches.

	Lower 10%ile of frequency distribution of *unadjusted* EQI values for 685 GB Reference sites (from Table 37)		Lower 10%ile of frequency distribution of *unadjusted* EQI values for Natural Flow catchments (see section 7.4.2b)		Intersection of % Sensitive and % Tolerant GAMs projected to *unadjusted* EQI axis	
	Single season	Two-season averaged	Single season	Two-season averaged	Single season	Two-season averaged
PSI_{fam}	0.748	0.795	-	-	0.843	0.835
PSI_{sp}	-	-	-	-	0.713	0.755
$LIFE_{fam}$	0.931	0.940	0.946	0.946	0.959	0.965
$LIFE_{sp}$	-	-	-	-	0.927	0.934

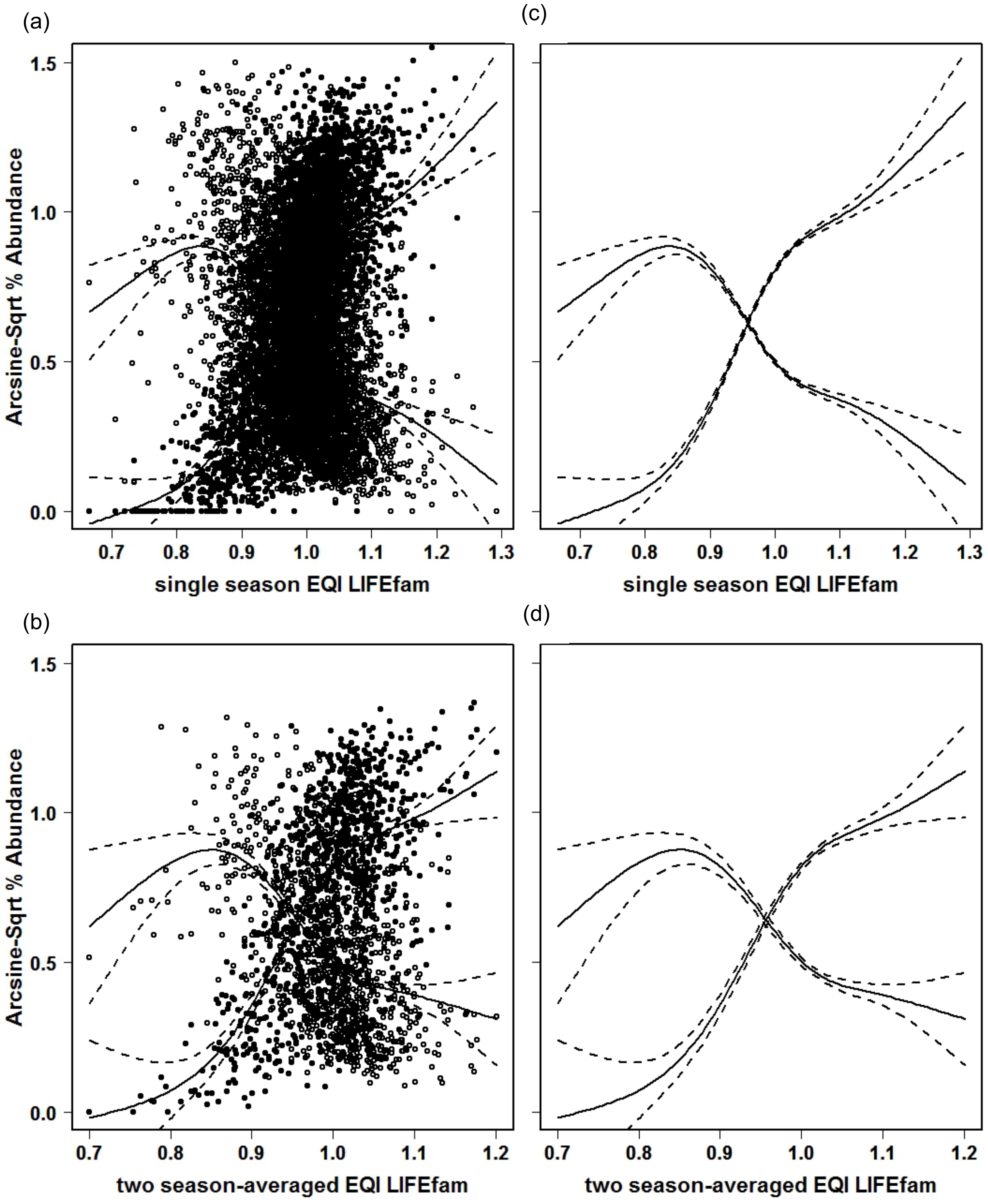


Figure 49. Variation in the relative abundance of taxa associated with moderate-rapid flowing water (closed circles) and taxa associated with slow & standing waters (open circles) with (a) single season and (b) two-season averaged EQI LIFE$_{fam}$. The solid lines are fitted generalised additive models with associated 95% confidence bands (dashed lines). Plots (c) and (d) are repeats of GAM-fitted lines in (a) and (b) respectively but with the individual sample points omitted for clarity.

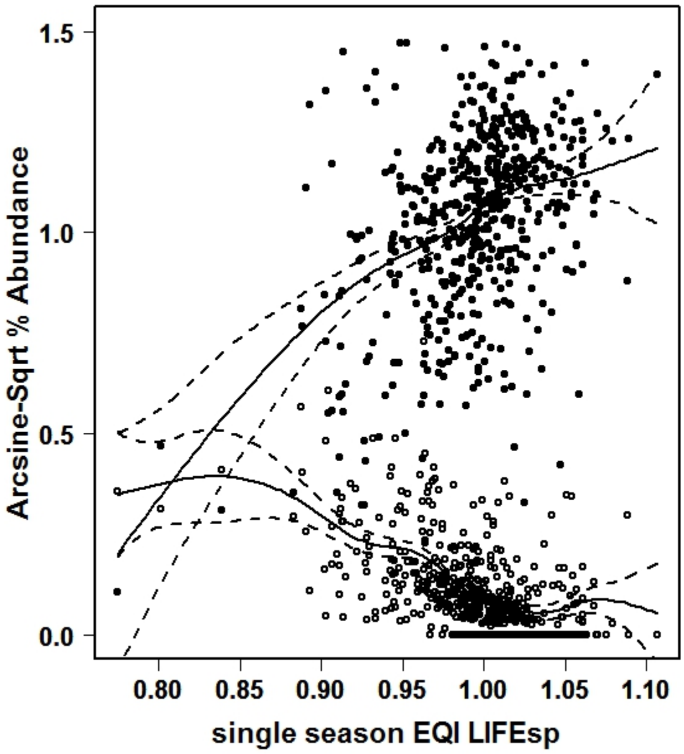
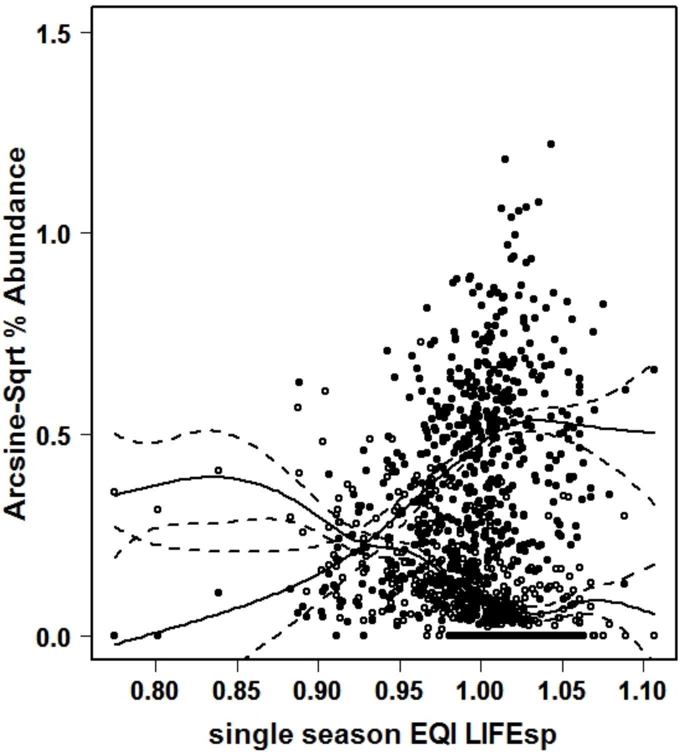


Figure 50. Variation in the relative abundance of taxa associated with moderate-rapid flowing water (closed circles) and taxa associated with slow & standing waters (open circles) with single season EQI LIFE$_{sp}$. The solid lines are fitted generalised additive models with associated 95% confidence bands (dashed lines).

(a)



(b)

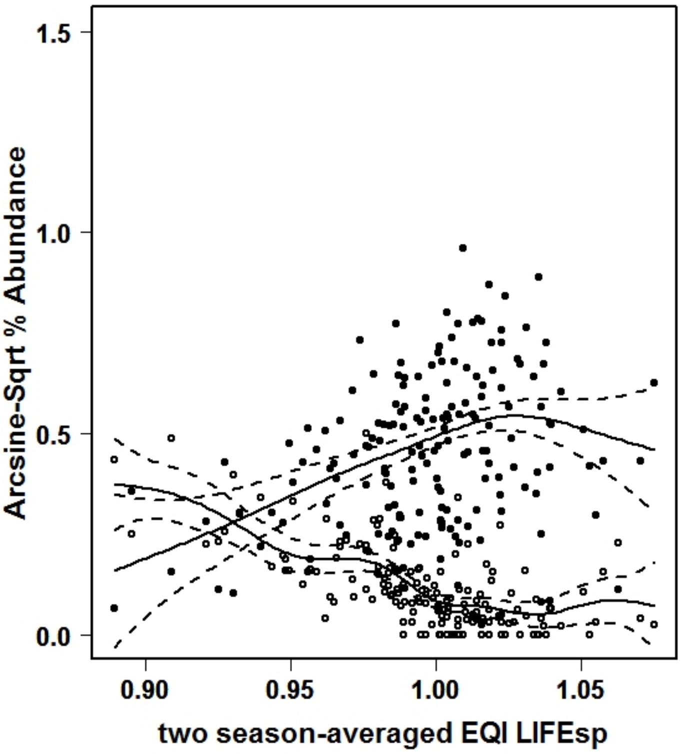


Figure 51. Variation in the relative abundance of taxa associated with rapid flowing water (closed circles) and taxa associated with slow & standing waters (open circles) with (a) single season EQI LIFE$_{sp}$ and (b) two season-averaged EQI LIFE$_{sp}$. The solid lines are fitted generalised additive models with associated 95% confidence bands (dashed lines).

8. ACKNOWLEDGEMENTS

We would like to thank Dave Rendall (SEPA) as Project Manager, Dave Colvill (SEPA), John Murray-Bligh (EA), Imelda O'Neil (NIEA), Mike Dunbar (EA) 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, Natural Resources Wales, Defra, The Welsh Government, ADAS, Rothamsted Research – North Wyke, Centre for Ecology and Hydrology and the National River Flow Archive for permitting the use of several of their datasets in support of this research. In particular we would like to thank Katie Muchan (CEH) for her assistance in extracting data from the NRFA, Adrian Collins (RR-NW) for access to the SEPARATE data and Graham Rutt (Natural Resources Wales) for helpful comments on an earlier version of the report. Adrian Collins contributed to the sections describing the SEPARATE framework used to predict fine sediment stress.

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

9. REFERENCES

Anthony S., Jones I., Naden P., Newell-Price P., Jones D., Taylor R., Gooday R., Hughes G., Zhang Y., Fawcett D., Simpson D., Turner D., Murphy J., Arnold A., Blackburn J., Duerdoth C., Hawczak A., Pretty J., Scarlett P., Liaze C., Douthwright T., Newell-Price P., Lathwood T., Jones M., Peers D., Kingston H., Chauhan M., Williams D., Rollett A., Roberts J. & Edwards-Jones G. (2012) *Contribution of the Welsh agri-environment schemes to the maintenance and improvement of soil and water quality, and to the mitigation of climate change*. Welsh Government, Agri-Environment Monitoring and Technical Services Contract Lot 3: Soil, Water and Climate Change (Ecosystems), No. 183/2007/08.

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.

Caroni R., van de Bund W., Clarke R.T. & Johnson R.K. (2013) Combination of multiple Biological Quality Elements into waterbody assessment of surface waters. *Hydrobiologia* 704: 437–451.

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. (2013) Estimating confidence of European WFD ecological status class and WISERBUGS (WISER Bioassessment Uncertainty Guidance Software). *Hydrobiologia* 704: 39–56.

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. & Davy-Bowker J. (2006) *Development of the scientific rationale and formulae for altering RIVPACS predicted indices for WFD Reference condition*. A report to the Scotland and Northern Ireland Forum for Environmental Research. [SNIFFER project WFD72B].

Clarke R. T., Davy-Bowker J., Dunbar M., Laize C., Scarlett P.M. & Murphy J.F. (2011) *Enhancement of the River Invertebrate Classification Tool (RICT).* SNIFFER Project WFD119 Report. 75pp Edinburgh. Scotland & Northern Ireland Forum for Environmental Research.

Clarke R.T. & Davy-Bowker J. (2014) *River Invertebrate Classification Tool Science Development Project: Modifications for WHPT and other abundance-weighted indices*. Report to the Scottish Environment Protection Agency, 85pp.

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.

Collins A.L. & Anthony S.G. (2008) Assessing the likelihood of catchments across England and Wales meeting 'good ecological status' due to sediment contributions from agricultural sources. *Environmental Science and Policy* 11: 163–170.

Collins, A.L., Stromqvist, J., Davison, P.S. & Lord, E.I. (2007) Appraisal of phosphorus and sediment transfer in three pilot areas identified for the Catchment Sensitive Farming initiative in England: application of the prototype PSYCHIC model. *Soil Use and Management* 23: 117-132.

Collins, A.L., Anthony, S.G., Hawley, J. & Turner, T. (2009a) Predicting potential change in agricultural sediment inputs to rivers across England and Wales by 2015. *Marine and Freshwater Research* 60: 626-637.

Collins, A.L., Anthony, S.G., Hawley, J. & Turner, T. (2009b) The potential impact of projected change in farming by 2015 on the importance of the agricultural sector as a sediment source in England and Wales. *Catena* 79: 243-250.

Collins A.L., Naden P.S., Sear D.A., Jones J.I., Foster I.D.L. & Morrow K. (2011) Sediment targets for informing river catchment management: international experience and prospects. *Hydrological Processes* 25: 2112–2129.

Collins, A.L., Jones, J.I., Sear, D.A., Naden, P.S., Skirvin, D., Zhang, Y.S., Gooday, R., Murphy, J., Lee, D., Pattison, I., Foster, I.D.L., Williams, L.J., Arnold, A., Blackburn, J.H., Duerdoth, C.P., Hawczak, A., Pretty, J.L., Hulin, A., Marius, M.S.T., Smallman, D., Stringfellow, A., Kemp, P.,Hornby, D., Hill, C.T., Naura, M. & Brassington, J. (2012a). Extending the evidence base on the ecological impacts of fine sediment and developing a framework for targeting mitigation of agricultural sediment losses. Final report to Defra, August 2012.

Collins, A.L., Foster, I., Zhang, Y., Gooday, R., Lee, D., Sear, D., Naden, P. & Jones, I. (2012b) Assessing 'modern background sediment delivery to rivers' across England and Wales and its use for catchment management. In: *Erosion and sediment yields in the changing environment* (pp 125-131). International Association of Hydrological Sciences (IAHS) Publication No. 356, Wallingford, UK, 451 pp.

Collins A.L. & Walling D.E. (2007a) Fine-grained bed sediment storage within the main channel systems of the Frome and Piddle catchments, Dorset, UK. *Hydrological Processes* 21: 1449–1459.

Collins, A.L. & Walling, D.E. (2007b) The storage and provenance of fine sediment on the channel bed of two contrasting lowland permeable catchments, UK. *River Research and Applications* 23: 429-450.

Davison P.S., Withers J.A., Lord E.I., Betson M.J. & Strömqvist J. (2008) PSYCHIC – A process-based model of phosphorus and sediment mobilisation and delivery within agricultural catchments. Part 1: Model description and parameterisation. *Journal of Hydrology* 350: 290–302.

Davy-Bowker J., Clarke R.T., Furse M.T., Davies C.E., Corbin T.A., Murphy J.F. & Kneebone N.T. (2007a) *RIVPACS Database Documentation*. A report to the Scotland and Northern Ireland Forum for Environmental Research. [SNIFFER project WFD46].

Davy-Bowker J., Clarke R.T., Furse M.T., Davies C.E., Corbin T.A., Murphy J.F. & Kneebone N.T. (2007b) *RIVPACS Pressure Data Analysis*. A report to the Scotland and Northern Ireland Forum for Environmental Research. [SNIFFER project WFD46].

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].

Duerdoth C.P., Arnold A., Murphy J.F., Naden P.S., Scarlett P., Collins A.L., Sear D.A. & Jones J.I. (2015) Assessment of a rapid method for quantitative reach-scale estimates of deposited fine sediment in rivers. *Geomorphology 230: 37*–50.

Dunbar M.J., Laize C., Baker R. & Scarlett P. (2011) Validation of Sediment Tool using DRIED-UP model. Draft Report to Environment Agency.

Dunbar M.J., Pedersen M.L., Cadman D., Extence C. Waddingham J., Chadd R. & Larsen S.E. (2010a) River discharge and local-scale physical habitat influence macroinvertebrate LIFE scores. *Freshwater Biology* 55: 226–242.

Dunbar M.J., Warren M., Extence C., Baker L., Cadman D., Mould D.J., Hall J. & Chadd R. (2010b) Interaction between macroinvertebrates, discharge and physical habitat in upland rivers. *Aquatic Conservation-Marine and Freshwater Ecosystems* 20 Supplement 1: S31–S44.

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. (2013) The assessment of fine sediment accumulation in rivers using macro-invertebrate community response. *River Research and Applications* 29: 17–55.

Foster, I.D.L., Collins, A.L., Naden, P.S., Sear, D.A., Jones, J.I. & Zhang, Y. (2011) The potential for palaeolimnology to determine historic sediment delivery to rivers. *Journal of Paleolimnology* 45, 287-306.

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.

Jones J.I., Murphy J.F., Collins A.L., Sear D.A., Naden P.S. & Armitage P.D. (2012) The impact of fine sediment on macro-invertebrates. *River Research and Applications* 28: 1055–1071.

Kelly M.G., Juggins S., Bennion H., Burgess A., Yallop M., Hirst H., King L., Jamieson B.J., Guthrie R. & Rippey B. (2008) *Use of diatoms for evaluating ecological status in UK freshwaters*. Science Report: SC030103/SR4, Environment Agency, Bristol, UK.

Lambert C.P. & Walling D.E. (1988) Measurement of channel storage of suspended sediment in a gravel-bed river. *Catena* 15: 65–80.

Murphy J.F., Jones J.I., Naden P.S., Pretty J.L., Duerdoth C.P., Hawczak A., Arnold A., Blackburn J.H., Old G., Sear D.A., Hornby D. & Collins A.L. (2015) Development and independent testing of a new biotic index of stream macroinvertebrate community response to deposited fine sediment stress. *Freshwater Biology* doi: 10.1111:fwb12627.

Murphy P. (2012) *Investigative monitoring into the impact of Altnahinch dam on Water Framework Directive quality elements (Flow, Dissolved Oxygen, Temperature, Chlorophyll, Hydromorphology, Invertebrates and River Continuity) in 2011*. Report to AMAP.

Murray-Bligh J. (1999) *Proceedure for collecting and analysing macroinvertebrate samples*. Environment Agency, Bristol, UK.

R Development Core Team (2013) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

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

Stromqvist, J., Collins, A.L., Davison, P.S., & Lord, E.I. (2008). PSYCHIC – a process-based model of phosphorus and sediment transfers within agricultural catchments. Part 2. A preliminary evaluation. *Journal of Hydrology* 350: 303-316.

Turley M.D., Bilotta G.S., Extence C.A. & Brazier R.E. (2014) Evaluation of a fine sediment biomonitoring tool across a wide range of temperate rivers and streams. *Freshwater Biology* 59: 2268–2277.

Wood S.N. (2006) *Generalized Additive Models: An Introduction with R*. Chapman and Hall/CRC, London, UK.

Young A. R., Grew R. & Holmes M. G. R. (2003) LF 2000: a national water resources assessment and decision support tool. *Water Science & Technology* 40(10): 119–126.

10. 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
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
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: 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* 2013).

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

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

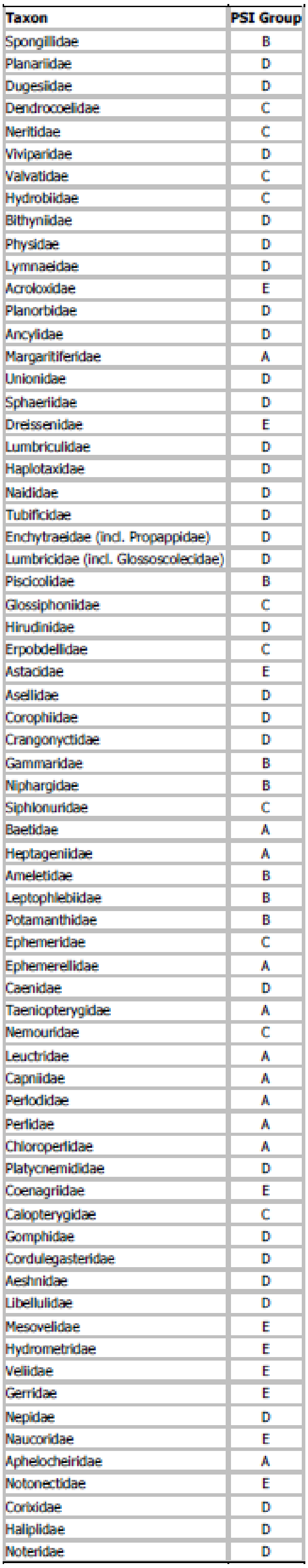
Sensitivty Group	**Sensitivity Group Description**	**Log_{10} Abundance Category (individuals)**			
		1 (1-9)	**2 (10-99)**	**3 (100-999)**	**4+ (1000+)**
A	Highly Sensitive	2	3	4	5
B	Moderately Sensitive	1	2	3	4
C	Moderately Insensitive	1	2	3	4
D	Highly Insensitive	2	3	4	5

Extence *et al* (2013) give a provisional interpretation of PSI scores as:

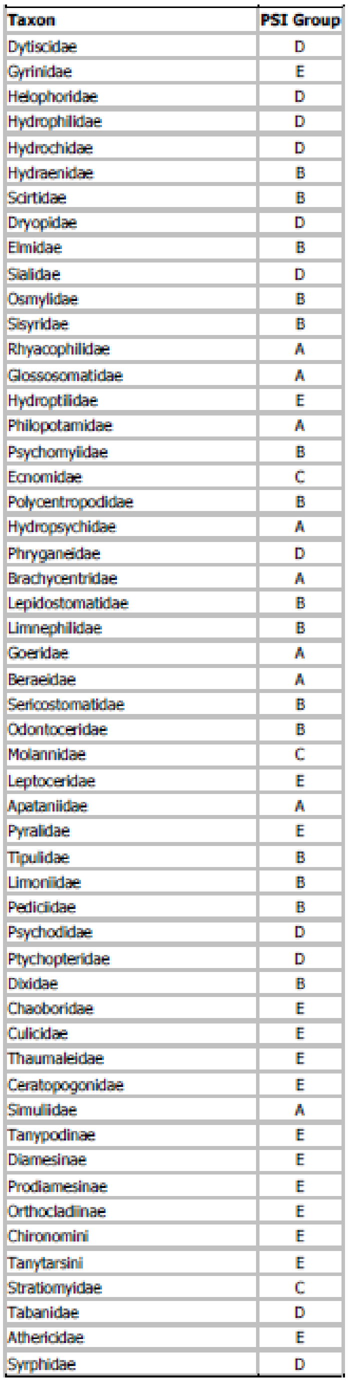
PSI range	**River bed Condition**
81-100	Minimally sedimented/unsedimented
61-80	Slightly sedimented
41-60	Moderately sedimented
21-40	Sedimented
0-20	Heavily sedimented

However, Extence *et al* (2013) 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. Importantly, 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* (2013) 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
Mesovelidae	E
Hydrometridae	E
Veliidae	E
Gerridae	E
Nepidae	D
Naucoridae	E
Aphelocheiridae	A
Notonectidae	E
Corixidae	D
Haliplidae	D
Noteridae	D



Taxon	PSI Group
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

Appendix 3: 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 4: Number of Reference sites with each WFD Assessment Score (1-6) in each End-Group for GB and Northern Ireland RIVPACS IV models

Module		WFD Assessment Score						
GB	End-group	1	2	3	4	5	6	Total sites
685 Reference sites	1	0	8	1	0	0	0	9
	2	4	5	1	1	0	0	11
	3	6	4	1	0	0	0	11
	4	5	4	0	0	0	0	9
	5	8	2	0	0	0	0	10
	6	3	5	0	0	0	0	8
	7	2	4	0	0	0	0	6
	8	10	2	4	1	0	0	17
	9	7	3	1	1	0	0	12
	10	8	8	2	0	0	0	18
	11	2	11	8	0	0	0	21
	12	3	5	6	0	0	0	14
	13	8	7	2	0	0	0	17
	14	12	7	2	0	0	0	21
	15	6	4	1	0	0	0	11
	16	6	5	6	0	0	0	17
	17	0	1	14	0	0	0	15
	18	1	1	16	4	0	0	22
	19	0	1	16	1	0	0	18
	20	1	3	6	0	0	0	10
	21	0	1	8	4	0	0	13
	22	1	6	12	1	0	0	20
	23	2	0	5	3	0	0	10
	24	0	2	9	0	0	0	11
	25	1	1	21	0	0	0	23
	26	6	5	15	1	0	0	27
	27	2	5	5	4	0	0	16
	28	0	3	3	3	0	0	9
	29	1	2	5	1	0	0	9
	30	5	7	0	1	1	0	14
	31	0	0	15	0	0	0	15
	32	0	1	29	2	0	0	32
	33	0	0	8	2	0	0	10
	34	0	0	16	1	0	0	17
	35	0	0	8	10	3	0	21
	36	0	0	15	4	1	0	20
	37	0	0	18	1	1	0	20
	38	0	0	18	3	2	0	23
	39	0	7	19	3	0	1	30
	40	0	0	10	0	0	1	11
	41	0	0	25	3	3	1	32
	42	0	0	5	6	1	0	12
	43	0	0	9	3	1	0	13
	Total							685

Northern Ireland 108 Reference sites		WFD Assessment Score						
	End-group	1	2	3	4	5	6	Total sites
	1	2	1	5	0	0	0	8
	2	1	0	3	1	2	0	7
	3	0	0	9	1	2	0	12
	4	1	0	2	4	0	0	7
	5	1	0	7	4	1	0	13
	6	0	0	3	6	3	0	12
	7	1	0	5	9	2	0	17
	8	0	0	3	2	5	0	10
	9	0	0	7	1	1	0	9
	10	0	1	5	1	0	0	7
	11	0	0	1	3	2	0	6
	Total							108

Definition of WFD Assessment scores:

1 = top of high
2 = middle of high
3 = high/good boundary
4 = middle of good
5 = good/moderate boundary
6 = worse

Appendix 5: Impact of taxonomic resolution on PSI and LIFE

Both LIFE and PSI scores can be calculated using family or mixed taxonomic resolution data. Due to the scoring systems used to calculate LIFE and PSI, it is possible that these indices may return different values dependent upon the level of taxonomic resolution used.

It is important to ensure that the taxonomic resolution of the data does not influence the returned EQR for a site when reporting at large spatial and temporal scales.

Here we have used mixed taxonomic resolution data to calculate the respective species ($LIFE_{sp}$ and PSI_{sp}) and family ($LIFE_{fam}$ and PSI_{fam}) versions of the indices from the same samples, to determine the effect of taxonomic resolution on returned EQI.

The two datasets used comprise:

a) mixed taxon data provided by SEPA
b) the WQ0128 and WG-AES dataset

Using the data derived from each sample the two versions of each index were calculated. To determine the effect of taxonomic resolution on index values returned, the relationships between the family level and species level versions of these indices were established. For each sample the EQI was established for family and mixed taxonomic resolution data using RIVPACS IV Model 15(2) (see section 7.3) using the appropriate level of taxonomic resolution to derive the expected value. To determine the effect of taxonomic resolution on index values returned, the relationships between the family level and species level versions of these indices were established for each dataset.

From the SEPA data it is apparent that when calculated from the same samples $LIFE_{sp}$ scores are consistently higher than the corresponding $LIFE_{fam}$ scores (Figure A5.1). Hence, it is clear that the appropriate level of taxonomic resolution must be used to establish expected scores.

When EQI values were derived for the SEPA dataset there was little difference between the values returned for $LIFE_{fam}$ and $LIFE_{sp}$. However, it should be noted that these data had a limited range of EQI values, focussed around 1.

In contrast to LIFE, PSI_{fam} returned consistently higher scores than PSI_{sp} in the SEPA data, with the two indices having a tendency to converge as they approached 100. It should be noted that it is not possible to achieve a score greater than 100 using the PSI system. When converted to an EQI, there was convergence between the two indices around 1. Again, it should be noted that these data had a limited range of EQI values, focussed around 1.

The WQ0128 and WG-AES dataset provided a wider range of values for LIFE and PSI. Whilst $LIFE_{sp}$ scores again tended to be consistently higher than the corresponding $LIFE_{fam}$ scores, the extent of divergence depended upon the sample score: the two indices converged at lower scores. This is likely to be caused by low scoring samples being dominated by families that are represented by individual species, whereas several species from the same family are likely to occur in high scoring samples.

When converted to EQI, the $LIFE_{sp}$ and $LIFE_{fam}$ indices diverged with increasing distance from EQI = 1: $LIFE_{sp}$ returned a higher EQI than $LIFE_{fam}$ for values greater than 1 and a lower EQI than $LIFE_{fam}$ for values less than 1.

In the WQ0128 and WG-AES dataset PSI_{fam} returned higher values than PSI_{sp} for low scoring samples and lower values than PSI_{sp} for higher scoring samples. This is likely to be caused by the distribution of scores within the PSI scoring system, and the fact that the score is a percentage.

When converted to EQI, the PSI_{sp} and PSI_{fam} indices also diverged with increasing distance from EQI = 1. At lower EQI values PSI_{sp} returned a consistently higher EQI than PSI_{fam}.

It is clear from these data that the taxonomic resolution used to derive LIFE and PSI influences both the raw score and EQI returned from a sample, with the effect becoming more pronounced with increasing distance from EQI=1.

Figure A5.1 Relationships between LIFE and PSI calculated using family level or mixed taxonomic level data for individual samples using SEPA data a) LIFE, b) EQI LIFE, c) PSI, and d) EQI PSI.

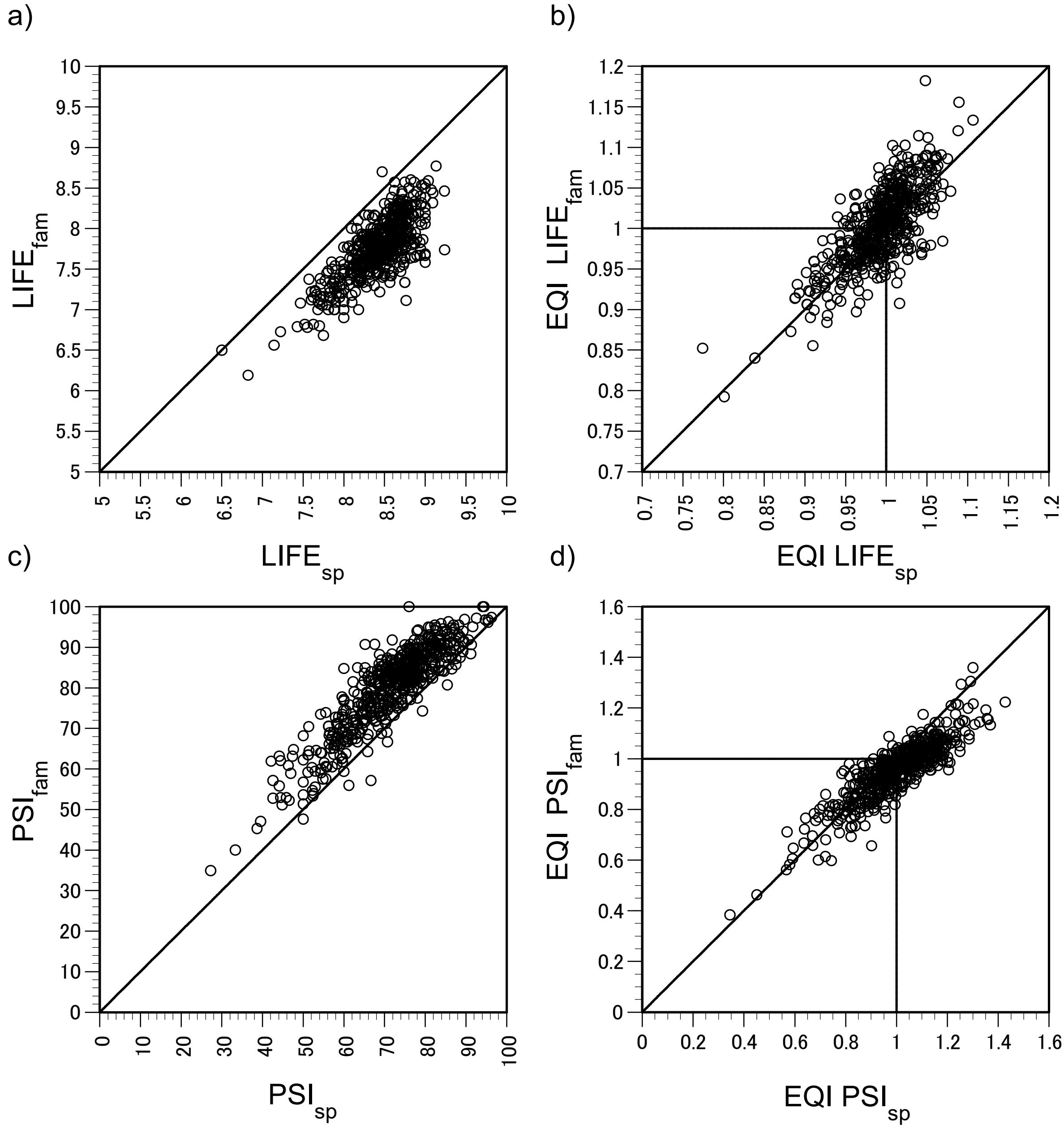
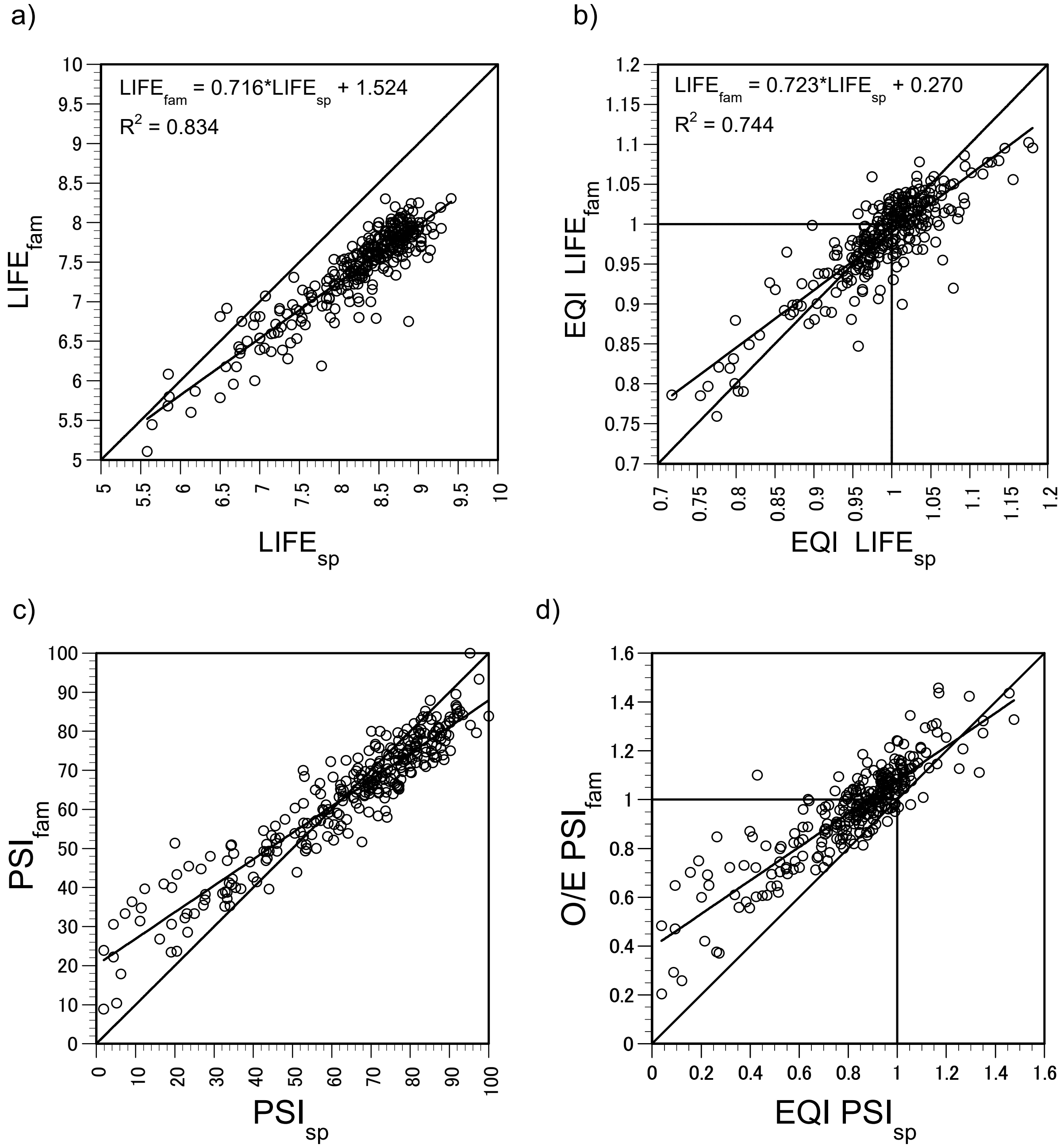


Figure A5.2 Relationships between LIFE and PSI calculated using family level or mixed taxonomic level data for individual samples using the WQ0128 and WG-AES data a) LIFE, b) EQI LIFE, c) PSI, and d) EQI PSI.



$PSI_{fam} = 1.312*PSI_{sp} -19.27$

$R^2 = 0.890$

$EQI\ PSI_{fam} = 1.136*EQI\ PSI_{sp} -0.265$

$R^2 = 0.779$

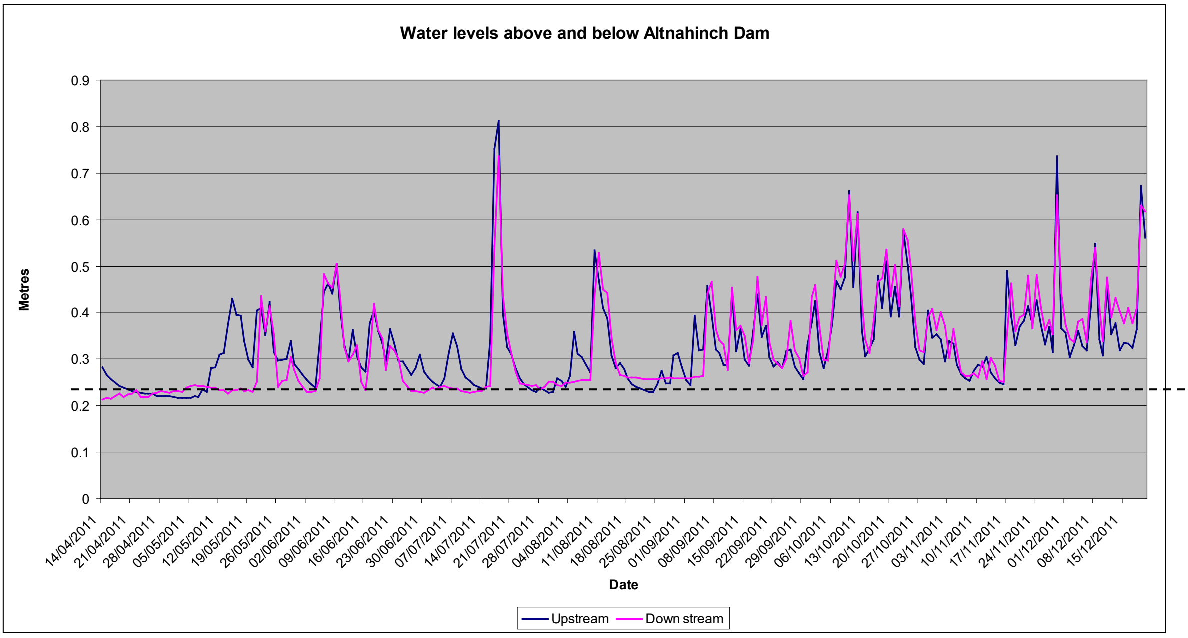
Appendix 6: Over-supply: Altnahinch Dam

To investigate the response of LIFE to over-supply of water, data from a study in Northern Ireland were kindly supplied by NIEA: Investigative monitoring into the impact of Altnahinch Dam on Water Framework Directive quality elements (Flow, Dissolved Oxygen, Temperature, Chlorophyll, Hydromorphology, Invertebrates and River Continuity) in 2011 (Murphy, 2012).

The study had been commissioned due to concerns that Altnahinch Dam did not release compensation flow as required by the Abstraction Impoundment Licence and that this resulted in the watercourse downstream being largely dry and interspersed with stagnant pools.

Water levels were monitored at 9 stations, both above and below the dam and, although a rating curve could not be constructed, it was apparent that not only was the dam releasing its required compensation flow volumes but that the dam was frequently full and regularly spilled water back in to the river during the study period (Figure A6.1). No problems were found in relation to DO and compensation flow was constant throughout the study period, although sediment composition appeared to be impacted by the structure and some peak flows appeared to be damped by the reservoir. Downstream flow is frequently augmented by spillages from the dam.

Figure A6.1 An overlay of the water level upstream and downstream of the Alnahinch Dam. Whilst the data is not directly comparable as it only records water depth changes at the site rather than flow. it does show when flood peaks occur at each location. The dotted black line is the approximate level of compensation flow released from the dam to the river downstream.



Invertebrate samples were collected in spring, summer and autumn from 5 sites in 2011-2012:

A. upstream of the reservoir upstream of the dam (Figure 6 Zone 1) which is a relatively natural river regarded as the “control” site,
B. 150 m downstream of the dam wall,
C approximately 1 km downstream of the dam wall, between sites B and C
D approximately 2.5 km downstream of the dam wall, by the confluence of a major tributary unimpacted by the dam, the Lewin Burn. Several other small tributaries enter the river between the Dam outfall and the Lewin Burn confluence.
E. at Ballyhoe Bridge (F10454), approximately 7 km downstream of the dam wall.

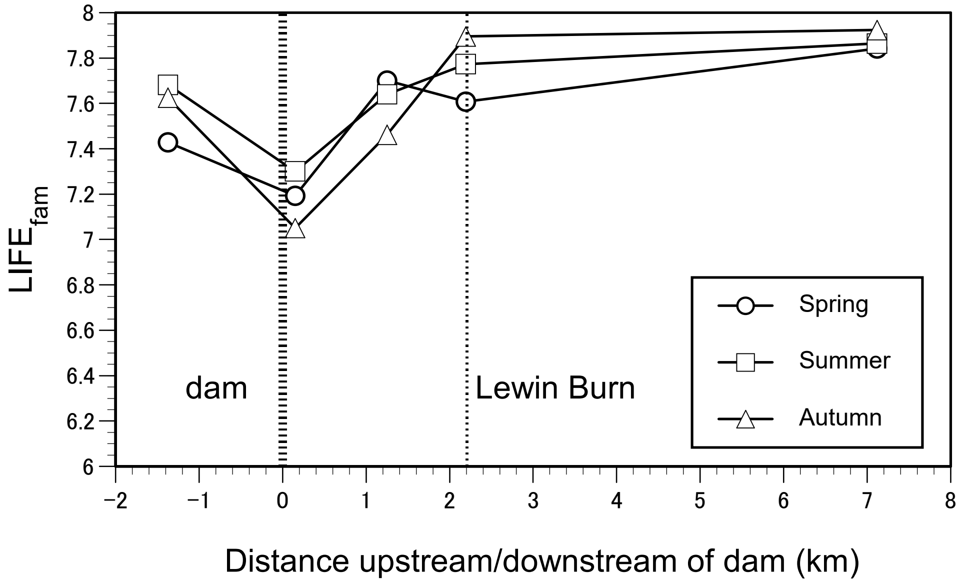
LIFE scores were calculated for each sampling occasion from the family level data provided.

Although the reservoir appeared to be releasing more water during low flow periods than upstream flows would suggest, $LIFE_{fam}$ scores were lowest just below the dam wall (Figure A6.2). Monitoring of oxygen concentrations indicate that the lower $LIFE_{fam}$ scores cannot be attributed to poor quality water being released from the reservoir. $LIFE_{fam}$ scores did not recover until some considerable distance downstream, and particularly after the confluence of undammed tributaries which would increase the likelihood of a more natural hydrological regime.

Despite the assumption that over-supply would result in an increase in LIFE scores, the data from Altnahinch Dam suggest that an increase in flow during typically low flow periods may be associated with reduced LIFE scores. However, the reservoir did appear to impact both low and high flows downstream of the dam, and without full ratings curves it is difficult to clearly attribute a cause to the decline in LIFE scores below the dam.

Whilst not conclusive, the limited data available from the Altnahinch Dam study suggest that over-supply of low flows does not result in increased LIFE scores.

Figure A6.2 $LIFE_{fam}$ scores for sites on the Bush River upstream (site A) and downstream (sites B, C, D & E) of Altnahinch Dam.



This page intentionally blank

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