Final Report **Project WFD72b**

Development of the scientific rationale and formulae for altering RIVPACS predicted indices for WFD Reference Condition

September 2006

Dissemination status
Unrestricted

Research contractor
This document was produced by:

Ralph Clarke and John Davy-Bowker
Centre for Ecology and Hydrology (CEH Dorset)
Winfrith Technology Centre
Winfrith Newburgh
Dorchester
Dorset DT2 8ZD

SNIFFER's project manager
SNIFFER's project manager for this contract is:

Ben McFarland, Environment Agency

SNIFFER's project steering group members are:

Gina Martin, SNIFFER
Robin Guthrie, Scottish Environment Protection Agency (SEPA)
Ben McFarland, Environment Agency
John Murray-Bligh, Environment Agency
Imelda O'Neill, Environment Heritage Service Northern Ireland
Peter Hale, Environment Heritage Service Northern Ireland

SNIFFER
First Floor, Greenside House
25 Greenside Place
EDINBURGH EH1 3AA
Scotland
UK

Company No: SC149513
Scottish Charity: SCO22375

www.sniffer.org.uk

EXECUTIVE SUMMARY

WFD72b: Development of the scientific rational and formulae for altering RIVPACS predicted indices for WFD Reference Condition (August, 2006)]

Project funders/partners: SNIFFER, Scottish Environment Protection Agency (SEPA), Environment Agency, Centre for Ecology and Hydrology (CEH)

Background to research

With the advent of the EU Water Framework Directive (WFD), the concept of the 'reference condition' (RC) against which assessments of biological degradation must be compared has become explicit within the legislative framework of the European Union (Council of the European Communities, 2000). It is therefore essential that member states can demonstrate that the biological datasets and models used to define RC meet the WFD criteria and set the same standards for all types of river site.

The selection of the RIVPACS (River InVertebrate Prediction And Classification System) reference sites and development of the RIVPACS methodology and software system for assessing the ecological status of UK rivers preceded the WFD. For several years it has been a concern that several, or even many, of these reference sites may not have been in WFD RC at the time of macroinvertebrate sampling, but merely represented the "best available" sites for each type of river site. In particular the predictions of expected fauna for some types of river site will be based on inadequate quality reference sites, leading to under-estimation of RC values for biotic indices and over-estimates of the RIVPACS observed to expected (O/E) ratios (termed Ecological Quality Indices (EQI)) and site quality for macroinvertebrates.

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). Subsequently UK agency aquatic ecologists 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).

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.

UK regulatory agency ecologists have been devising methods of adjusting the RIVPACS expected (E) values or EQI values of indices (ASPT and number of BMWP taxa (TAXA)) by determining the weighted average assessment score of the reference sites involved in the prediction for any particular test site. However, these approaches are, in effect, determined solely by the EQI values used to set the good/moderate boundary.

Objectives of research

To develop a robust defensible mechanism for adjusting the RIVPACS expected values of biotic indices for any specific test site according to the perceived ecological condition (at the time of sampling for RIVPACS) of the RIVPACS reference sites actively involved in the prediction for that test site. The resulting adjusted O/E values should then more evenly reflect the WFD concept of ecological status across all UK river types

To provide the necessary formulae for implementing this adjustment mechanism for the environmental regulatory agencies' classification sites.

Key findings and recommendations

The research approach used was based on statistical modeling to assess and quantify the actual relationship between assessment score (1-6) and the observed values of TAXA and ASPT amongst reference sites within RIVPACS site types (TWINSPAN groups).

Negligible differences in average index values were found between reference sites with assessment scores of 1,2 or 3. Observed index values are lower relatively for reference sites with assessment scores of 4 and especially 5. (There were insufficient reference sites with scores of 6 (worse than good/moderate boundary) to estimate its adjustment factor and it is recommended that any such sites are treated as having assessments scores of 5 in the adjusted of test site expected values.)

Models were fitted separately to data for each RIVPACS module (GB Northern Ireland (NI), Scottish Islands (SI) and Scottish Highlands (SH)) and for samples based on each the seven possible combinations of one, two or three RIVPACS seasons. There was insufficient range of scores in the high quality SI and SH sites to determine any significant relationships. Therefore a recommended single UK-wide adjustment model was fitted using all UK reference sites and samples. Although actual effects of changes in site quality on index values may vary with site type, to achieve adequate precision and for simplicity, the best single overall estimates for the adjustment factors were derived.

Models were also fitted based on grouping sites by WFD System A typology; some estimates of adjustment factors were greater, but overall model fits were poorer.

The recommended statistical model for estimating the adjustment factors was model M4 which estimates the average proportional increase or decrease in index values due to each level of site assessment score relative to sites with scores of 3 ("target" high/good boundary) within the same TWINSPAN group. Recommended adjustment factors are given in Table 30.

Procedures and formulae are derived to combine these score-specific adjustment factors with information on the proportion of reference sites in each site group with each assessment score to calculate the recommended adjustment factors for expected values for any test site based on its RIVPACS probability of belonging to each TWINSPAN group.

An EXCEL spreadsheet adjustment "calculator" with encoded formulae to automate this procedure for adjusting RIVPACS expected values of any UK RIVPACS test sites has been produced and is available as an project deliverable and output.

One advantage of the M4 type of model is that the approach could also be used in other European ecoregions where reference condition expected values of metrics for a test site are based on some average or percentile value of the reference sites in the same WFD System A or B stream typology (Council of the European Union, 2000) (i.e. stream type replaces TWINSPAN group in the models)).

Key words:
Water Framework Directive (WFD), Reference Condition, RIVPACS, Adjusted expected

TABLE OF CONTENTS

List of Tables

List of Figures

APPENDICES

1. INTRODUCTION

1.1 Background

With the advent of the WFD, the concept of the ‘reference condition’ has become explicit within the legislative framework of the European Union (Council of the European Communities, 2000). Reference Condition (RC) has to be established as a quality standard against which assessments of biological degradation must be compared. It is therefore essential that member states can demonstrate that the biological datasets used to define their reference conditions meet the criteria of the WFD, or that any derived model predictions of RC set these same standards for all types of river site. The WFD describes reference conditions as follows:

There are no, or only very minor, anthropogenic alterations to the values of the physicochemical and hydromorphological quality elements for the surface water body type from those normally associated with that type under undisturbed conditions.

The values of the biological quality elements for the surface water body reflect those normally associated with that type under undisturbed conditions, and show no, or only very minor, evidence of distortion.

The selection of the RIVPACS reference sites and development of the RIVPACS methodology and software system for assessing the ecological status of UK rivers preceded the European Commission’s Water framework Directive (WFD) and its definitions of ‘Reference Condition’ (RC) for water bodies. The RIVPACS approach is based on classifying these reference sites into biological types, developing predictive relationships between the site groups and their environmental characteristics, using the relationships to predict the expected fauna for test sites, and comparing the observed fauna with the expected fauna to derive standardised assessment of site ecological quality.

For several years it has been perceived and a is a matter of concern that several, or even many, of the RIVPACS reference sites may not have been in WFD RC at the time of macroinvertebrate sampling for RIVPACS. In particular, there is concern that the biological quality of the RIVPACS reference sites at the time of sampling for RIVPACS system development varied around the country according to river type, was dependent on the availability of high and good quality sites and that the RIVPACS sites generally represent the “best available” sites for each type of river site. This means that the RIVPACS predictive model for the expected macroinvertebrate fauna and expected values of biotic indices (e.g. number of taxa (TAXA) and Average Score Per Taxon (ASPT)) will be based on higher quality reference sites for some types of site than others. Where the expected fauna for a test site is based on relatively poorer condition reference sites, the expected (E) values of the biotic indices may not set a high enough target condition value for the biotic indices, and therefore the ratio (O/E) of observed (O) to expected (E) values of the indices may give an over-estimate of the ecological condition and status of the site and water body. The reverse may also be true when test site predictions are based on reference sites of relatively higher condition.

The RIVPACS model predicted macroinvertebrate fauna and predicted biotic index values for any particular site is, in effect, a weighted average of the observed fauna and index values of the “environmentally-similar” reference sites. This forces roughly half of the reference site samples to have Ecological Quality Indices (EQI), based on observed(O) to expected (E) ratios (O/E), of greater than 1.0 and roughly half to have values less than 1.0.

It is widely understood that true “reference condition” sites do not exist for all river types in the UK and that, in many instances, RIVPACS predictions will actually equate to a condition which is in fact lower than true “reference”. This is already acknowledged in the current UK classification scheme, whereby 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.

The “true” condition that the RIVPACS prediction for a site represents is dependent on the condition of those RIVPACS reference sites which have been actively utilised by RIVPACS in making the prediction (i.e. in reference site groups with non-zero probabilities of membership for the test site. In those instances where the RIVPACS prediction is based on sites in true reference condition, using the RIVPACS predicted value as the high/good boundary may be too stringent. Similarly, where the RIVPACS prediction is based on RIVPACS reference sites and samples which all, or mostly, represent a condition lower than the high/good boundary, then the predicted/expected index values may be unduly lax.

To overcome this, it is necessary to derive a prior evaluation of the status of each of the RIVPACS reference sites at the time of sampling for RIVPACS. In a recent SNIFFER project WFD46, led by John Davy-Bowker (CEH Dorset), pressure data from the UK RIVPACS reference sites were compiled and analysed to help indicate where in the condition spectrum these reference sites might lie in terms of the WFD concept of reference conditions.

Furthermore, Robin Guthrie (SEPA) devised an assessment score scale of 1-6, with accompanying descriptions, to represent the perceived quality of river site (Table 1, Appendix 1).

Table 1 – Definition of Aquatic ecologists’ assessment score (1-6)

Score	WFD Status Class position	Brief description
1	top of ‘high’	genuinely ‘pristine’
2	middle of ‘high’	basically ‘pristine’
3	‘high’/’good’ boundary	pressures might just be picked in the biology
4	middle of ‘good’	pressures have a noticeable effect on biology but most expected taxa still present
5	‘good’/’moderate’ boundary	starting to get noticeably impacted, probably missing quite a few taxa you’d expect
6	below ‘good’/’moderate’ boundary	

Appropriate UK Agency Regional Aquatic ecologists in the Environment Agency, SEPA and the Environment Heritage Service Northern Ireland (EHS) were asked by Robin Guthrie (SEPA) or Amanda Veal (EA) to give an assessment score (1-6) to each of the RIVPACS reference sites in their region, based on their best-available information and knowledge on the state of each site at the time of sampling for the development of RIVPACS.

1.2 Objectives

The objectives of the current SNIFFER project WFD72b are:

- To develop a robust defensible mechanism for adjusting the RIVPACS expected values of biotic indices for any specific test site according to the perceived ecological condition (at the time of sampling for RIVPACS) of the RIVPACS reference sites actively involved in the prediction for that test site. The resulting adjusted O/E values should then more evenly reflect the WFD concept of reference condition across all UK river types.

- To provide the necessary formulae for implementing this adjustment mechanism to the EQI values for the environmental regulatory agencies' classification sites.

1.3 Aquatic ecologists' Assessment scores and WFD Pressure data

The analysis of pressure data in WFD46 provided useful information on the condition of the RIVPACS reference sites at the time of sampling for RIVPACS. Some of this information may also have been used in the regional aquatic ecologists' determination of the assessment scores (1-6). As part of the current project, CEH had identified the following four sites where the pressure data analysis suggested that the aquatic ecologists' overall assessment could be modified, but the discrepancies were not major:

Module	Site ID	River	Site	Aquatic ecologist's Assessment score (1-6)	Suggested revised score	Unsuitability problem
NI	20201101	Owenrigh River	Carnanbane	2	3	Metals
		Total zinc mean of 14.14 μg/l is marginally highest in major site group + O/E TAXA of 0.79 is low				
GB	3309	Swale	Morton-on-Swale	2	3 or 4	Metals
		Dissolved zinc mean of 57.27 μg/l is twice site group 90%ile + O/E TAXA of 0.79 is low				
GB	5845	Unnamed	Dinmore Manor	2	3	Organic
		Suspended solids mean of 26.3 mg/l 105°C marginally high but not outlier + O/E ASPT of 0.90 is quite low				
GB	6845	Unnamed	Alton Common	2	3	Organic/Nutrients
		BOD ATU mean of 4.30 mg/l O_2 is highest in 9-group, Free & Saline Ammonia mean of 1.08 mg/l N is high, Nitrite mean 0.08 mg/l N is marginally high but not outlier , O/E ASPT of 0.91 is quite low				

At the project start-up meeting (3th May 2006), it was agreed that the development of the adjustment mechanism within the current project should be based solely on the aquatic ecologists' original assessment scores.

1.4 RIVPACS Modules

There are currently four UK RIVPACS modules:

Module	Module Code	Sites	Site groups
Great Britain	GB	614	35
Northern Ireland	NI	110	11
Scottish Islands	SI	55	5
Scottish Highlands	SH	108	10

Separate models were developed for reference sites in Northern Ireland and the Scottish Islands partly because it was thought that their geographic separation and isolation from mainland GB might have led to differences in macroinvertebrate community composition and taxonomic richness. A separate Scottish Highlands module was also developed recently in an attempt to improve RIVPACS predictions for sites in this region. Future developments may involve combining one or more of these modules and their reference sites.

Although the form of the derived adjustment mechanism was to be the same for each of these fours modules, the statistical model parameters involved in the adjustment formulae were initially estimated separately for each module and the extent of differences assessed.

2. GENERAL APPROACHES

2.1 Assessment score (S) for the site-specific Expected (E) value and O/E value

The expected fauna and expected values of biotic indices are based on the RIVPACS estimated probabilities of the test site belonging to each of the RIVPACS module's TWINSPAN site classification groups. Any adjustment of the expected (E) values and thus O/E values for a site are to be based on the (weighted) average perceived quality (S) of the references sites involved in the prediction of the expected fauna for the site. This is derived from the reference site assessment scores as follows:

P_i = RIVPACS probability that a test site belongs to TWINSPAN site group *i*

S_i = Average Assessment score for reference sites in TWINSPAN site group *i*

$S = \sum_i (P_i \, . \, S_i)$ = weighted average assessment score for reference sites in the prediction of the expected values for the test site

In all of the approaches developed, the assumption has been that an assessment score of 3 represents the high/good boundary in terms of WFD definitions of ecological status class. This was the intention in the environmental regulatory agencies' aquatic ecologists' assessment score for each of the UK reference sites (Table 1). Furthermore, the aim is that, after adjustment of the expected (E) values, an O/E value of 1.0 should represent the quality of sites on the high/good boundary – for any type of site in any module.

Therefore, if the weighted average quality (S) of the reference sites involved in the prediction for a test site is 3, there is no need to adjust the E and thus O/E values. If S < 3, then the references sites involved are on average, above the high/good boundary in quality and the E values needed to be adjusted downward (assuming highest values of biotic indices indicate highest quality). If S > 3, then the references sites involved are on average, below the high/good boundary in quality and the E values needed to be adjusted upward.

2.2 Previous proposals by Agency staff

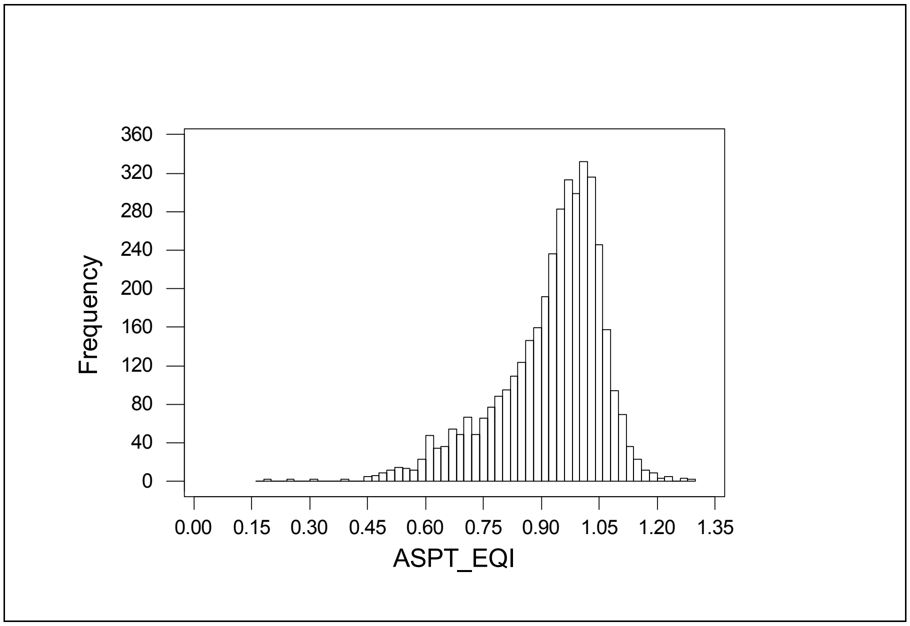
2.2.1 Robin Guthrie's method for adjusting O/E values

Robin Guthrie made the first attempt to devise a method of adjusting RIVPACS O/E values directly in autumn 2005. He worked solely on the O/E values for ASPT. He concluded from an examination of the distribution of all EQI_{ASPT} values from approximately 6000 sites in England and Wales that for those sites with an ASPT EQI value greater than 1.0 the distribution of values tailed off rapidly after an EQI value of 1.1 (Figure 1).

Thus it was considered on average that an ASPT EQI value of 1.05 reasonably represented the middle of high. This is consistent with the fact that the current General Quality Assessment (GQA) classification system upper and lower boundaries of the "good" class for EQI_{ASPT} are 1.0 and 0.9, representing a spread of 0.1 units and, therefore, the midpoint is 0.05 units below the upper boundary of 1.0, namely 0.95.

Each increase of one unit in Aquatic ecologists' assessment score (1-6) was intended to represent half a status class increase in quality. Therefore each unit increase in assessment score was assumed to lead to an increase in EQI_{ASPT} of 0.05.

Figure 1 – Frequency distribution of O/E$_{ASPT}$ values for GQA sites in England and Wales



Consequently, Robin suggested that all RIVPACS EQI$_{ASPT}$ values were "normalised" to the high/good boundary value of 1.0 by the following formula:

$$\text{Adjusted EQI}_{ASPT} = \text{unadjusted EQI}_{ASPT} + 0.05(3\text{-}S)$$

where S = weighted mean assessment score of RIVPACS sites used in prediction (as defined in Section 2.1)

Thus for example, a test site with an unadjusted EQI$_{ASPT}$ value of 0.93 and prediction sites weighted mean assessment score S of 4 would have an adjusted EQI$_{ASPT}$ value of:

$$0.93 + 0.05\ (3 – 4) = 0.93 + 0.05\ (-1) = 0.88.$$

In contrast, if S was say 1.6, the adjusted EQI$_{ASPT}$ value for the same site would be increased to:

$$0.93 + 0.05(\ 3 – 1.6) = 0.93 + 0.05\ (1.4) = 1.00.$$

2.2.2 Mathematical properties of adjustment methods

There is a mathematical flaw in this approach, which can be explained by the following illustrative example (Table 2). Suppose a test site has a true expected number of taxa of 25 in the sense of its prediction if based on reference sites with an average assessment score of 3. Furthermore, suppose that because its RIVPACS predictions is based on reference sites with an average assessment score of 5, it has a RIVPACS 'face' expected value of 20. Suppose this site is sampled in each of five years over which its quality has improved substantially, such that the observed number of taxa over the five years were 10, 20, 25, 30 and 35. Then the 'face' O/E values and the true O/E values, (assuming the E values are fixed over time) are as given in Table 2.

Table 2 – Illustrative example explaining effect of bias in expected values and adjustment procedures

Observed (O)	10	20	25	30	35
True E	25	25	25	25	25
True O/E	0.4	0.8	1.0	1.2	1.4
Face E	20	20	20	20	20
Face O/E	0.5	1.0	1.25	1.5	1.75
Bias in O/E Face O/E – True O/E	0.1	0.2	0.25	0.3	0.35
True E / Face E	1.25	1.25	1.25	1.25	1.25
(True O/E) / (Face O/E)	0.8	0.8	0.8	0.8	0.8

Using Robin's idea in this scenario with an average assessment score of 5, sites with a true EQI of 1.0, would have an EQI of 0.8 and therefore all such sites should have their EQI increased by 0.2. However, this logic only works for test sites whose true quality was the high/good boundary (i.e. EQI =1.0). The illustration above shows that for sites whose true quality is less than or greater than 1.0, the appropriate adjustment for bias in the face EQI value is less than or greater than 0.2 respectively.

The correct adjustment for all of the face EQI values is a multiplicative factor, in this example, it is 0.8. In general the appropriate correction factor to multiply face EQI would be the ratio of the true EQI value to the RIVPACS predicted face EQI value. (This is explained in further detail in Section 2.2.3 below)

2.2.3 John Murray-Bligh's method for adjusting O/E values

Beginning in early 2006, John Murray-Bligh has independently evolved a procedure to adapt Robin's original idea to make the appropriate multiplicative adjustment to the EQI values, equivalent to the ideas explained in Section 2.2.2 above. John's approach makes use of the same assumption as Robin that the good/moderate boundary for EQI_{ASPT} of 0.90 should be the (average) EQI of reference sites with an assessment score of 5. As a score of 3 is intended to represent a site whose quality lies on the high/good boundary, a unit deviation in assessment score from 3 is still taken to represent a change in face EQI value of 0.05. However, John's method acknowledges that the effect of under-estimating the expected value for a test site on the degree of over-estimation of its true EQI value depends on the observed index value and thus true quality of the site. The appropriate correction is not a constant change in EQI for a given deviation of average assessment score from 3, but rather a multiplicative factor, akin to that explained in Section 2.2.2.

John's method determines the value of EQI, denoted D_S, assumed for reference sites with an assessment score of S and a true EQI of 1.0. The multiplicative adjustment factor applied to the face expected value (E_{face}) of a test site with a weighted average assessment score of S is then intended to be $1/D_S$, as detailed below for the case of EQI ASPT with the good/moderate boundary set at an EQI of 0.9:

Score (S)	1	2	3	4	5
D_S	1.10	1.05	1.0	0.95	0.90
multiplicative adjustment factor ($1/D_S$) for the Expected value	0.909	0.952	1.000	1.053	1.111

John Murray-Bligh fitted a quadratic regression between S and $1/D_S$ to describe the expected value adjustment factor for non-integer S. However, this can be done mathematically as follows:

Suppose more generally that the good/moderate boundary of EQI is set to EQI_{GM}, (it is assumed implicitly that the high/good boundary EQI is 1.0) and that a test site has a weighted average assessment score for its prediction of S (not necessarily an integer). Then the adjusted expected value (E_{adj}) is estimated from the 'face' expected value (E_{face}) by:

$$E_{adj} = E_{face} / (1+ 0.5(1-EQI_{GM}) (3-S)) \qquad \text{(equation J1)}$$

The adjusted EQI value, EQI_{adj} is then estimated by:

$$EQI_{adj} = O / E_{adj} \qquad \text{(equation J2)}$$

These formula works for any values of S between 1 and 5 and for any setting of the good/moderate boundary of EQI below unity.

Note: this is not how John Murray-Bligh expressed his method in his explanation in Excel, but is I understand what he was aiming for and intended.

2.2.4 Critique of previous methods

Both Robin Guthrie's and John Murray-Bligh's proposed methods are dependent on the values of the EQIs chosen for the ecological status class boundaries. Furthermore they assume that the indices increase uniformly with site quality in that a unit decrease in assessment score is implicitly assumed to lead to a constant increase in the EQI value of sites. This assumed "linear" response needs testing.

A considerable part of the recorded variation in O/E values amongst reference sites will be due to sampling variation. This is greater for single season samples than for two or three season combined samples (Clarke *et al* ., 2002) and will affect the amount of "noise" in the observed distribution and range of EQI values.

John Murray-Bligh's method is an improvement of Robin Guthrie's original idea, as it correctly involves a multiplicative adjustment to the EQI value based on the perceived degree of under- or over- prediction of the expected index values.

The ecological quality represented by a particular EQI value (e.g. 0.7) is likely to vary between biotic indices, and appropriate ecological status class boundaries of EQI need to be determined separately for each index. However, for any particular index, both Robin Guthrie's and John Murray-Bligh's proposed methods are dependent on the values chosen for the ecological status class boundaries of the EQI. If the class boundaries and class widths were changed, then the degree of adjustment to all test site values of that EQI would be changed. This does not seem logical when the degree of under- or over-estimation of the RIVPACS expected values and thus of the EQI values for the test sites has not changed. Also the distribution of EQI values for the reference sites has not changed, only the interpretation of what EQI value represents the middle or 'top end' of 'high' class and bottom of 'good' class.

It seems more robust for the adjustment factors to be based on the distribution of the observed index values for the reference sites in relation to their assessment score, independent of any subsequent division of the EQI scale into WFD status classes.

3. STATISTICAL DATA-BASED APPROACHES TO ADJUSTING EXPECTED VALUES

The main purpose of the research project WFD72b is the "development of the scientific rationale and formulae for altering RIVPACS predicted indices". The approach is to develop an adjustment mechanism based on deriving a statistical relationship between the observed (O) values of biotic indices for the individual RIVPACS reference sites and their assessment score (1-6), allowing for stream type.

Reference sites differ naturally in their fauna and typical values of derived biotic indices according to their physical and environmental characteristics. RIVPACS allows for this in assessing site quality by classifying the references sites into groups based on the TWINSPAN multivariate classification method. These groups are then used to determine the expected fauna and expected values of indices. These natural differences in observed values of biotic indices need to be allowed for in our statistical analyses. Within a TWINSPAN group, one might expect the reference sites with assessment scores 1 or 2 to have higher values of the biotic indices than reference sites with scores of 4-6. Our adjustment approach is based on this idea and estimates the average extent to which the observed values of the reference sites change with assessment score within site types (i.e. TWINSPAN groups). There is not enough data to estimate this separately for each site group, and moreover, expected values of indices are based on a weighted average of several groups. Therefore the aim is to derive one set of estimates of adjustment factors for each RIIVPACS module.

3.1 Distribution of assessment scores for reference sites and TWINSPAN groups

Tables 3-6 and Figures 2-5 show the distribution and range of assessment scores (1-6) that occur within TWINSPAN group of each of the four RIVPACS modules GB, NI, SI and SH. Major differences between the modules are apparent. The map of assessment score for GB reference sites in Figure 6 highlights that the perceived ecological condition of the RIVPACS reference sites varies regionally with, few example, fewer "high" quality reference sites in south-eastern England. As the RIVPACS prediction of expected values for a site is based on environmentally similar sites, the average assessment score (and perceived ecological condition) of the reference sites on which predictions are made will vary between test sites.

In the GB and Northern Ireland modules, only 35% and 23% respectively of reference sites were assessed as having been better than the high/good boundary status (i.e. score of 1 or 2) at the time of sampling for RIVPACS development. In contrast, 91% of Scottish Islands and 95% of Scottish Highlands reference sites were assessed as being better than the high/good boundary status.

In the Scottish Islands and Highlands modules, none and one respectively of the reference sites were assessed as being worse than high/good boundary status. As nearly all Scottish reference sites were assessed as being above the high/good boundary, predictions for such modules will need to be downgraded. However, the lack of range in perceived qualities of the reference sites is likely to make it more difficult to derive a quantitative relationship between the values of the biotic indices and assessment score within either of these two modules. It may be necessary to use adjustment relationships derived by involving sites from other RIVPACS modules; this is assessed below.

Both the GB and Northern Ireland modules had reference sites with a wide range of perceived qualities in each of several TWINSPAN groups.

Table 3 – Number of GB module reference sites in each assessment score (1-6) and mean score (Q_i) for sites in each TWINSPAN site group *i*, and overall.

GB Module TWINSPAN group	Assessment Score 1	2	3	4	5	6	Total Sites	Mean Score Q_i
1	19	10	4	1	0	0	34	1.62
2	2	1	3	0	0	0	6	2.17
3	4	7	6	3	0	0	20	2.40
4	6	3	2	0	0	0	11	1.64
5	5	6	0	0	1	0	12	1.83
6	1	5	3	5	0	0	14	2.86
7	1	8	4	3	0	0	16	2.56
8	0	13	5	1	1	2	22	2.82
9	0	2	6	2	0	0	10	3.00
10	6	4	2	1	0	0	13	1.85
11	3	3	4	0	0	0	10	2.10
12	4	4	0	0	0	0	8	1.50
13	11	7	2	0	0	0	20	1.55
14	1	18	12	1	0	0	32	2.41
15	2	6	4	0	0	0	12	2.17
16	5	9	17	0	0	0	31	2.39
17	1	13	11	3	0	0	28	2.57
18	0	11	2	0	0	0	13	2.15
19	0	8	5	2	1	0	16	2.75
20	1	9	9	1	0	0	20	2.50
21	0	3	10	2	1	0	16	3.06
22	1	11	16	10	1	0	39	2.97
23	0	6	9	0	0	0	15	2.60
24	0	2	15	0	0	0	17	2.88
25	1	9	10	1	0	0	21	2.52
26	0	3	7	2	0	0	12	2.92
27	0	3	16	5	1	0	25	3.16
28	0	2	3	4	1	0	10	3.40
29	0	0	5	2	2	0	9	3.67
30	1	3	12	4	3	1	24	3.33
31	0	3	4	2	0	1	10	3.20
32	0	5	4	0	0	1	10	2.80
33	0	1	20	2	3	5	31	3.71
34	1	1	7	3	1	0	13	3.15
35	0	2	4	6	2	0	14	3.57
All	76	201	243	66	18	10	614	2.64
% of sites	12.4%	32.7%	39.6%	10.7%	2.9%	1.6%	100.0%	

Table 4 – Number of Northern Ireland (NI) module reference sites in each assessment score (1-6) and mean score (Q_i) for sites in each TWINSPAN group *i*

NI Module TWINSPAN group	Assessment Score 1	2	3	4	5	6	Total Sites	Mean Score Q_i
1	2	2	4	0	0	0	8	2.25
2	1	0	3	1	2	0	7	3.43
3	2	3	4	1	2	0	12	2.83
4	2	0	1	4	0	0	7	3.00
5	2	1	5	4	1	0	13	3.08
6	0	0	3	6	4	0	13	4.08
7	2	2	2	9	2	0	17	3.41
8	0	1	2	2	5	0	10	4.10
9	0	2	5	1	1	0	9	3.11
10	1	2	3	1	0	0	7	2.57
11	0	0	2	3	2	0	7	4.00
All	12	13	34	32	19	0	110	3.30
% of sites	10.9%	11.8%	30.9%	29.1%	17.3%	0.0%	100.0%	

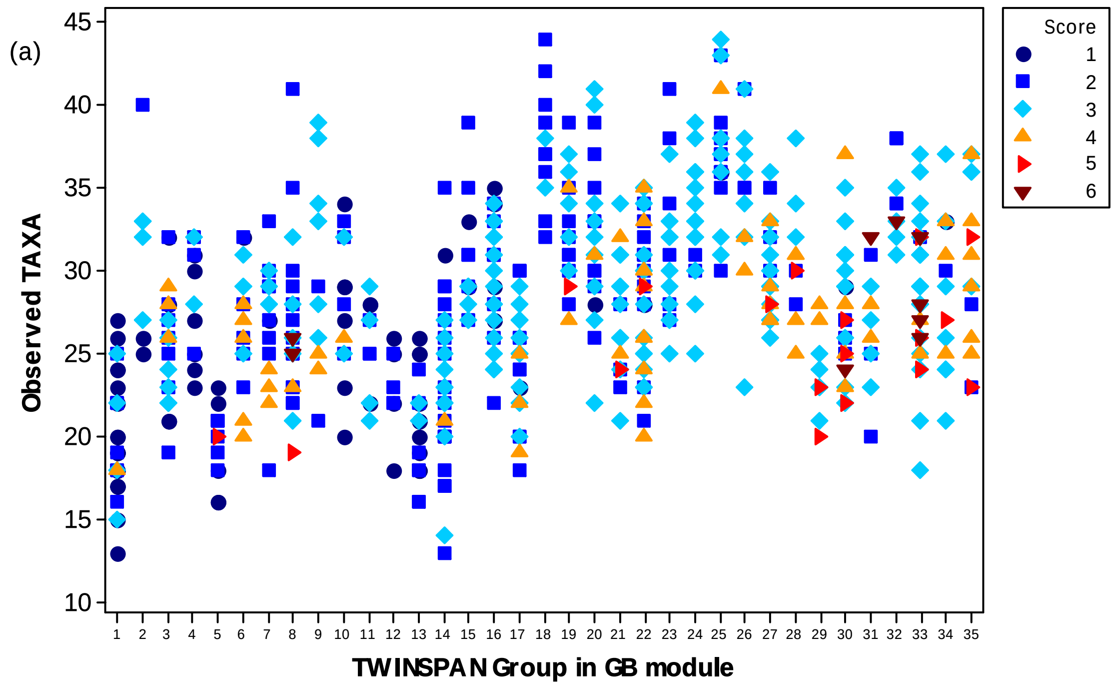
Table 5 – Number of Scottish Islands (SI) module reference sites in each assessment score (1-6) and mean score (Q_i) for sites in each TWINSPAN group *i*

SI Module TWINSPAN group	Assessment Score 1	2	3	4	5	6	Total Sites	Mean Score Q_i
1	3	5	1	0	0	0	9	1.78
2	1	9	1	0	0	0	11	2.00
3	5	7	1	0	0	0	13	1.69
4	8	4	1	0	0	0	13	1.46
5	0	8	1	0	0	0	9	2.11
All	17	33	5	0	0	0	55	1.78
% of sites	30.9%	60.0%	9.1%	0.0%	0.0%	0.0%	100.0%	

Table 6 – Number of Scottish Highlands (SH) module reference sites in each assessment score (1-6) and mean score (Q_i) for sites in each TWINSPAN group *i*

SH Module TWINSPAN group	Assessment Score 1	2	3	4	5	6	Total Sites	Mean Score Q_i
1	1	6	3	0	0	0	10	2.20
2	5	1	0	0	0	0	6	1.17
3	9	6	0	0	0	0	15	1.40
4	6	2	0	0	0	0	8	1.25
5	7	4	0	0	0	0	11	1.36
6	2	7	1	0	0	0	10	1.90
7	10	4	0	0	0	0	14	1.29
8	10	4	0	0	0	0	14	1.29
9	5	2	0	0	0	0	7	1.29
10	9	3	0	1	0	0	13	1.46
All	64	39	4	1	0	0	108	1.46
% of sites	59.3%	36.1%	3.7%	0.9%	0.0%	0.0%	100.0%	

Figure 2 - Plot of (a) Observed TAXA and (b) Observed ASPT for the GB reference sites in each RIVPACS TWINSPAN group, coded by their site assessment score (1-6)



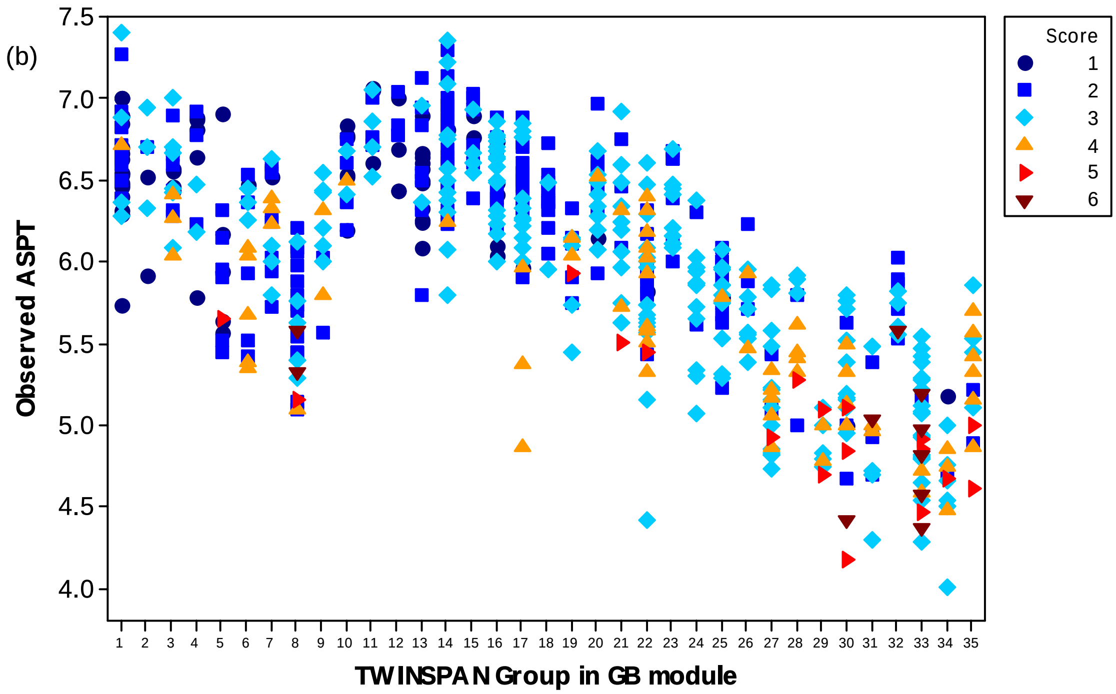
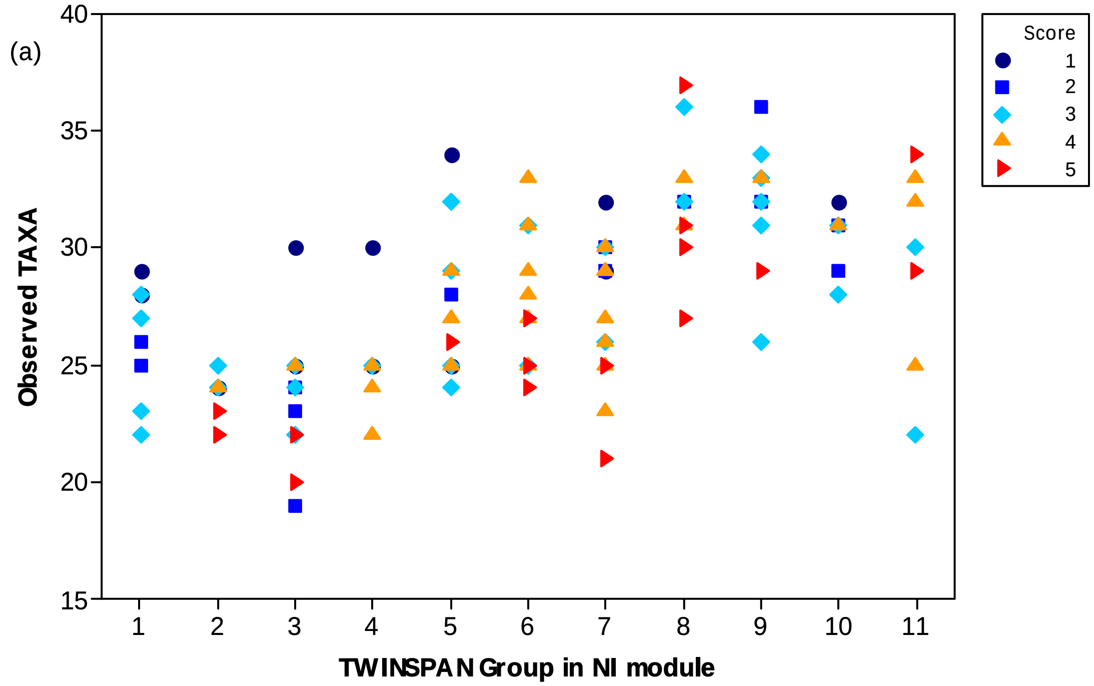


Figure 3 - Plot of (a) Observed TAXA and (b) Observed ASPT for the Northern Ireland reference sites in each TWINSPAN group, coded by their site assessment score (1-6)



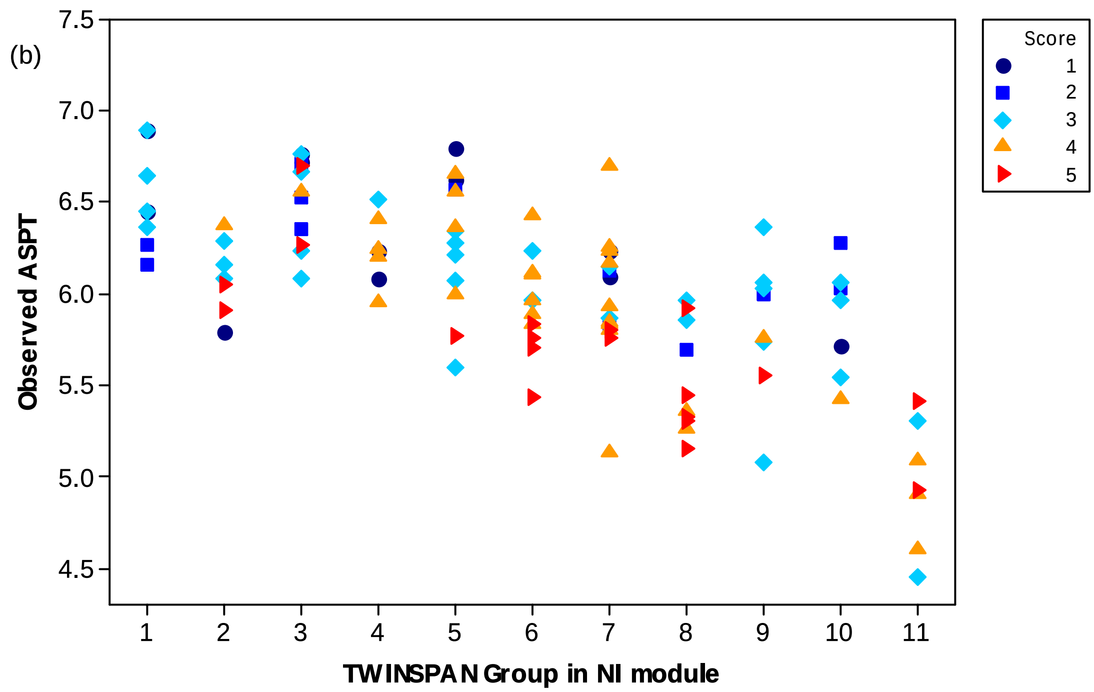
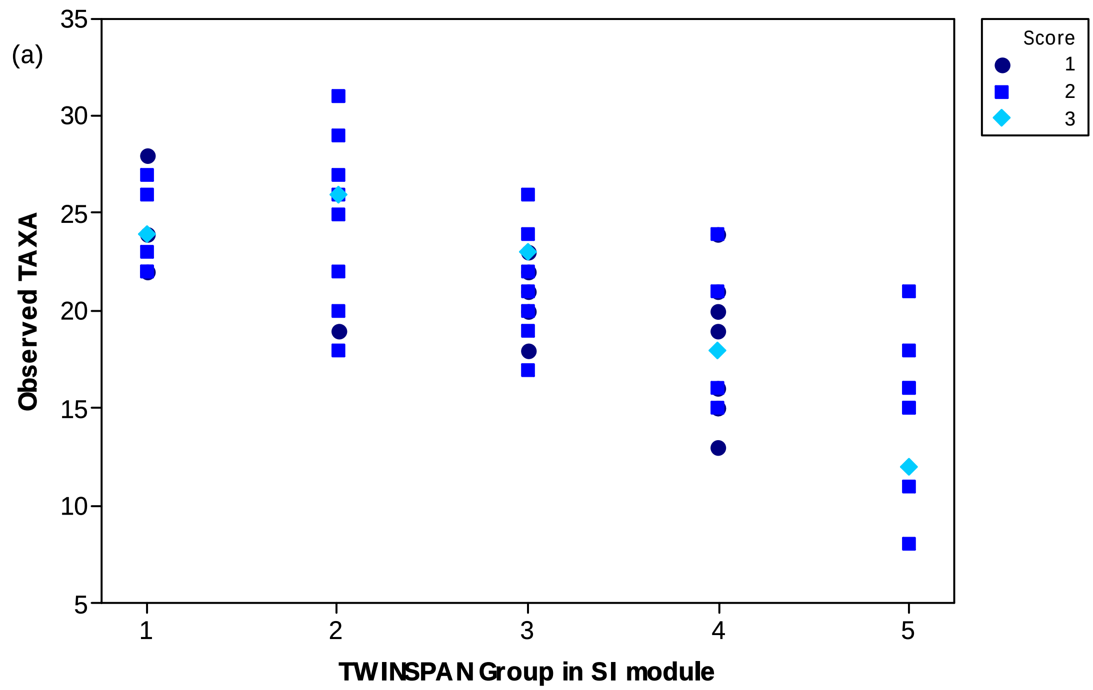


Figure 4 - Plot of (a) Observed TAXA and (b) Observed ASPT for the Scottish Islands reference sites in each TWINSPAN group, coded by their site assessment score (1-6)



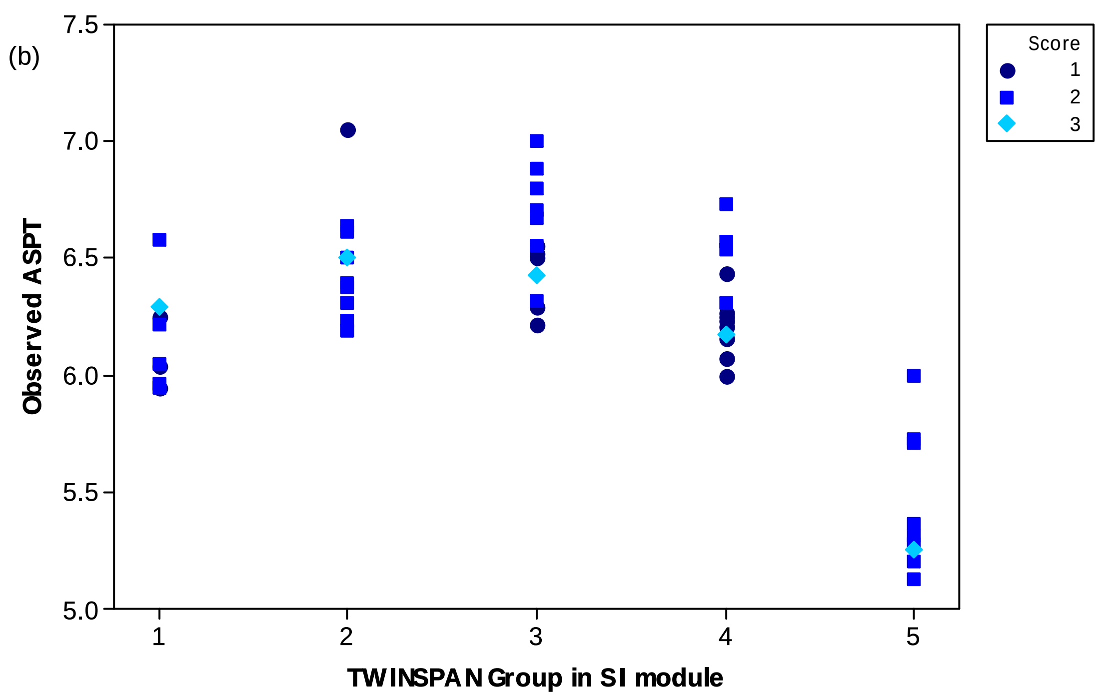


Figure 5 - Plot of (a) Observed TAXA and (b) Observed ASPT for the Scottish Highlands reference sites in each TWINSPAN group, coded by their site assessment score (1-6)

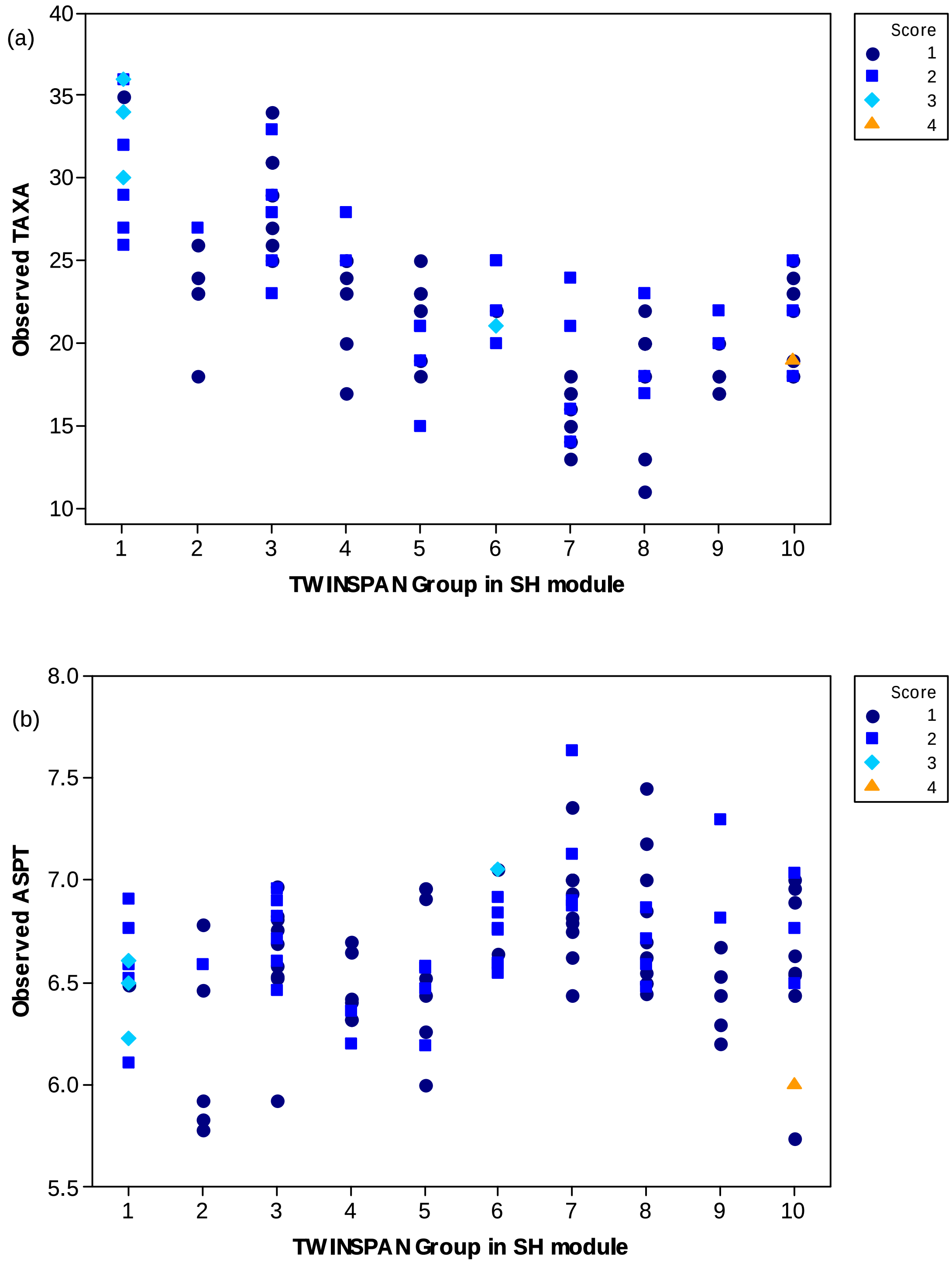
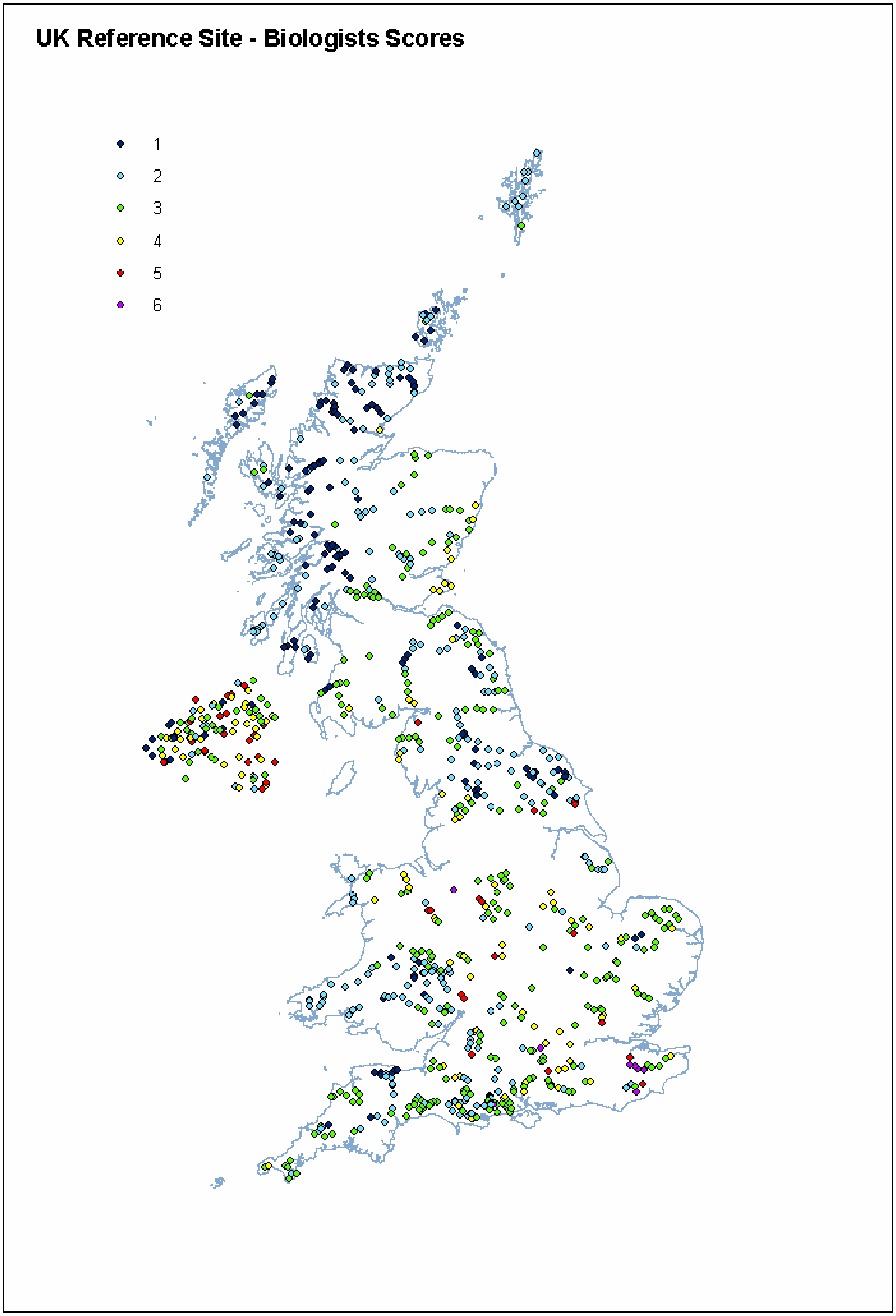


Figure 6 – Map of GB reference sites coded by their assessment score (1 = dark-blue, 2 = light-blue, 3 = green, 4 = yellow, 5 = red, 6 = purple)



The weighted average assessment score (S) (see Section 2.1) is based on a weighted average of the mean assessment scores (Q_i) of the reference sites in each TWINSPAN group *i*. Therefore, the value of S for any test site can only, at most, range between the minimum and maximum values of Q_i for the relevant RIVPACS module. The mean assessment score for reference sites in each TWINSPAN group of each module are given in the right-hand columns of Table 3-6 and summarised in Table 7. degree of adjustment of expected values of biotic indices,

Within the GB module, the average assessment score per TWINSPAN groups varies from 1.50 (group 12) to 3.71 (group 33), spanning the high/good boundary equivalent score of 3.00, and with an overall score amongst all reference sites of 2.64 (Tables 3 and 7).

The reference sites in Northern Ireland were assessed to have been in the poorest average ecological condition amongst the four UK RIVPACS modules, with mean assessment score per site group varying from 2.25 (group 1) to 4.10 (group 8) with an overall mean score of 3.30, slightly worse than the high/good boundary (Tables 4 and 7).

Table 7 – Overall percentage of reference sites with each assessment score (1-6), together with the overall mean and range of average within-group scores, separately for each module

	% of sites with Assessment Score:						Average score within a TWINSPAN group		Overall Mean
Module	1	2	3	4	5	6	Min	Max	
GB	12.4%	32.7%	39.6%	10.7%	2.9%	1.6%	1.50	3.71	2.64
NI	10.9%	11.8%	30.9%	29.1%	17.3%	0.0%	2.25	4.10	3.30
SI	30.9%	60.0%	9.1%	0.0%	0.0%	0.0%	1.46	2.11	1.78
SH	59.3%	36.1%	3.7%	0.9%	0.0%	0.0%	1.17	2.20	1.46

The average score amongst reference sites within each TWINSPAN group in both the Scottish Islands and Scottish Highlands modules were all well below 3, ranging from 1.46 to 2.11 for the Islands modules and from 1.17 to 2.20 for the Highlands module (Tables 5-7). The current RIVPACS predicted values for all test sites using both of these Scottish modules are therefore based on reference sites whose (weighted) average quality is above the high/good boundary. Assuming the aim for UK site assessments is for an EQI of 1.0 to be equivalent to the high/good boundary, then current expected index values for these two modules may need to be adjusted “downwards” for a lower quality.

3.2 Within group association of assessment score with observed index values

Before developing statistical models of the relationship between observed biotic index values (TAXA and ASPT) and assessment score within groups, the strength and consistency of any relationship across TWINSPAN groups was assessed. In particular the calculated the mean value of a biotic index of sites with each assessment score within each group and calculated the percentage of groups for which the mean index value of sites with one assessment score was greater than the mean index value for sites with a second score (Table 8-12).

For the GB module, mean observed TAXA was higher for sites with score 1 compared to sites with score 2 in only 7 (35%) of the 20 TWINSPAN groups which had one or more sites with scores of both 1 and 2 (Table 8). If there was no real change in average index value with score one would expect the percentage greater by chance (due to sampling and other stochastic variation) to be about 50%. However, the mean value for sites with an assessment score of 1 (top of high) was greater than the mean scores for sites with a score of 3 (high/good boundary) in only 39% of the 18 comparable groups. This suggests that, at least in some types of site, taxonomic richness may actually increase slightly with initial departures from pristine conditions. To allow for this would require separate adjustment factors for different stream types and different modules, but our policy, as with the previously proposed adjustment methods of Section 2, was to derive a single set of adjustment factors for the UK, or at least for any one RIVPACS module.

In contrast, the mean TAXA for sites with score 2 was greater than that for sites with score 3 in 59% of the 32 TWINSPAN groups for which comparisons were possible. However, the mean TAXA for sites with a score of 1, or of 2, or of 3 were greater than the mean TAXA for sites with a score of 4 in 73-74% of TWINSPAN groups. Mean TAXA for sites with a score of 5 (good/moderate boundary) was less than mean TAXA for sites with a score of 3 (high/good boundary) in all 11 groups for which comparisons could be made (Table 8(a)).

This suggests that the number of BWMP TAXA does not change dramatically between the top and bottom of the high ecological status class, may actually peak at mid-high, but declines below the high/good boundary, as desired.

This tendency for the mean index value of sites with score 1 to be greater than the mean for sites with score 2 in fewer than 50% of the TWINSPAN groups was repeated for ASPT in the GB module (40% of groups), TAXA and ASPT in the Scottish Islands module (25%) and TAXA (30%) and ASPT (40%) in the Scottish Highlands module (Tables 9, 11 and 12). However, in the Northern Ireland module, the mean value for sites with score of 1 was greater than the mean value for sites with a score of 2 in 80% and 100% respectively of the 5 TWINSPAN groups for which sites with both scores occurred.

The mean index value for sites with a score of 1, 2 or 3 (at or above the high/good boundary) was greater than the mean value for sites with a score of 4 or 5 (i.e. below the high/good boundary) in the majority of TWINSPAN groups for both TAXA and ASPT in all modules and cases where comparisons were possible (Tables 8-12). This suggests that there is a fairly consistent decline in TAXA and ASPT amongst reference sites once quality declines below the high/good boundary and that it should be possible to develop statistical models to derive one or more correction factors to adjust for this.

Table 8 – Mean value of observed TAXA for GB module reference sites with each assessment score (1-6) in each TWINSPAN group and consistency across groups of differences in mean TAXA in relation to score.

	Assessment score						Mean	Sites in
Group	1	2	group	4	5	6	TAXA	group
1	19.8	20.2	20.0	18.0			19.9	34
2	25.5	40.0	30.7				30.5	6
3	25.5	25.7	24.0	27.7			25.5	20
4	26.7	29.3	30.0				28.0	11
5	19.0	20.0			20.0		19.6	12
6	32.0	26.8	28.3	24.4			26.6	14
7	27.0	27.1	28.8	23.0			26.8	16
8		27.4	26.4	23.0	19.0	25.5	26.4	22
9		25.0	33.0	24.5			29.7	10
10	27.0	29.5	28.5	26.0			27.9	13
11	26.0	25.7	24.8				25.4	10
12	22.8	23.0					22.9	8
13	21.2	19.0	21.5				20.5	20
14	31.0	22.8	22.6	21.0			22.9	32
15	31.0	32.0	28.3				30.6	12
16	30.4	29.7	29.1				29.5	31
17	23.0	25.7	25.3	22.0			25.0	28
18		38.0	36.5				37.8	13
19		32.4	33.8	31.0	29.0		32.4	16
20	28.0	32.4	32.1	31.0	*		32.0	20
21		25.0	27.9	28.5	24.0		27.2	16
22	28.0	29.4	29.3	27.5	29.0		28.8	39
23		33.2	30.3				31.5	15
24		30.5	32.7				32.5	17
25	36.0	36.8	37.3	41.0			37.2	21
26		37.0	34.4	31.0			34.5	12
27		32.3	29.3	28.8	28.0		29.5	25
28		29.0	34.7	28.5	30.0		30.6	10
29		*	23.2	27.5	21.5		23.8	9
30	29.0	26.0	29.1	28.3	24.7	24.0	27.8	24
31		25.3	26.0	27.0		32.0	26.6	10
32		36.4	32.8			33.0	34.6	10
33		32.0	29.0	26.0	27.3	29.0	28.7	31
34	33.0	30.0	27.7	29.7	27.0		28.7	13
35		25.5	34.5	30.2	27.5		30.4	14
All	24.1	28.3	29.2	27.4	25.6	28.5	28.0	614

Difference in mean TAXA	% of *n* groups (amongst those with both scores)
Mean Score 1 > Mean Score 2	35% (7 of 20)
Mean Score 1 > Mean Score 3	39% (7 of 18)
Mean Score 2 > Mean Score 3	59% (19 of 32)
Mean Score 1 > Mean Score 4	75% (9 of 12)
Mean Score 2 > Mean Score 4	73% (16 of 22)
Mean Score 3 > Mean Score 4	74% (17 of 23)
Mean Score 3 > Mean Score 5	100% (11 of 11)

Table 9 – Mean value of observed ASPT for GB module reference sites with each assessment score (1-6) in each TWINSPAN group and consistency across groups of differences in mean ASPT in relation to score.

	Assessment score						Mean	Sites in
Group	1	2	group	4	5	6	ASPT	group
1	6.60	6.72	6.73	6.72			6.65	34
2	6.22	6.70	6.66				6.52	6
3	6.51	6.55	6.56	6.24			6.50	20
4	6.63	6.64	6.33				6.58	11
5	6.04	5.88			5.65		5.93	12
6	6.47	5.95	6.36	5.71			5.99	14
7	6.52	6.24	6.13	6.32			6.24	16
8		5.70	5.64	5.09	5.16	5.45	5.61	22
9		5.80	6.29	6.06			6.14	10
10	6.60	6.48	6.55	6.50			6.55	13
11	6.91	6.83	6.78				6.83	10
12	6.64	6.85					6.75	8
13	6.48	6.58	6.66				6.53	20
14	6.81	6.72	6.60	6.24			6.67	32
15	6.83	6.78	6.69				6.76	12
16	6.36	6.57	6.53				6.51	31
17	5.96	6.49	6.37	5.40			6.31	28
18		6.39	6.22				6.37	13
19		6.11	5.91	6.09	5.93		6.03	16
20	6.14	6.48	6.38	6.52			6.42	20
21		6.43	6.22	6.02	5.50		6.19	16
22	5.82	5.88	5.79	5.89	5.45		5.83	39
23		6.41	6.30				6.34	15
24		5.96	5.79				5.81	17
25	5.78	5.80	5.73	5.78			5.77	21
26		5.94	5.64	5.70			5.73	12
27		5.32	5.18	5.13	4.93		5.17	25
28		5.40	5.87	5.45	5.27		5.55	10
29			4.89	4.89	4.90		4.89	9
30	5.00	5.10	5.36	5.24	4.71	4.42	5.17	24
31		5.00	4.80	4.98		5.03	4.92	10
32		5.81	5.68			5.58	5.73	10
33		5.16	5.03	4.66	4.74	4.78	4.94	31
34	5.18	4.70	4.57	4.69	4.67		4.66	13
35		5.06	5.48	5.34	4.81		5.26	14
All	6.46	6.24	5.88	5.60	5.02	4.98	6.00	614

Difference in mean ASPT	% of *n* groups (amongst those with both scores)
Mean Score 1 > Mean Score 2	40% (8 of 20)
Mean Score 1 > Mean Score 3	56% (10 of 18)
Mean Score 2 > Mean Score 3	72% (23 of 32)
Mean Score 1 > Mean Score 4	58% (7 of 12)
Mean Score 2 > Mean Score 4	59% (13 of 22)
Mean Score 3 > Mean Score 4	61% (14 of 23)
Mean Score 3 > Mean Score 5	73% (8 of 11)

Table 10 – Mean value of (a) observed TAXA and (b) observed ASPT for NI module reference sites with each assessment score (1-6) in each TWINSPAN group and consistency across groups of differences in mean TAXA in relation to score.

(a) TAXA	Assessment score						Mean	Sites in
Group	1	2	3	4	5	6	TAXA	group
1	28.5	25.5	25.0				26.0	8
2	24.0		24.3	24.0	22.5		23.7	7
3	27.5	22.0	23.3	25.0	21.0		23.4	12
4	27.5		25.0	23.3			24.7	7
5	29.5	28.0	28.4	27.0	26.0		27.9	13
6			27.0	28.8	25.8		27.5	13
7	30.5	29.5	28.0	27.6	23.0		27.7	17
8		32.0	34.0	32.0	31.0		31.9	10
9		34.0	31.2	33.0	29.0		31.8	9
10	32.0	30.0	29.0	31.0			30.0	7
11			26.0	30.0	31.5		29.3	7
All	28.6	28.0	27.4	27.8	26.8		27.6	110

Difference in mean TAXA	% of *n* groups (amongst those with both scores)
Mean Score 1 > Mean Score 2	100% (5 of 5)
Mean Score 1 > Mean Score 3	86% (6 of 7)
Mean Score 2 > Mean Score 3	57% (4 of 7)
Mean Score 1 > Mean Score 4	83% (5 of 6)
Mean Score 2 > Mean Score 4	50% (3 of 6)
Mean Score 3 > Mean Score 4	50% (5 of 10)
Mean Score 3 > Mean Score 5	87% (7 of 8)

(b) ASPT	Assessment score						Mean	Sites in
Group	1	2	3	4	5	6	ASPT	group
1	6.67	6.22	6.59	*	*		6.52	8
2	5.79	*	6.18	6.38	5.98		6.09	7
3	6.75	6.53	6.44	6.56	6.49		6.53	12
4	6.16	*	6.52	6.20	*		6.23	7
5	6.71	6.57	6.10	6.40	5.77		6.30	13
6	*	*	6.05	6.06	5.68		5.94	13
7	6.17	5.98	6.01	5.99	5.78		5.99	17
8	*	5.69	5.92	5.31	5.43		5.53	10
9	*	6.00	5.85	5.76	5.55		5.84	9
10	5.72	6.16	5.85	5.42	*		5.86	7
11	*	*	4.88	4.87	5.17		4.96	7
All	6.37	6.20	6.07	5.94	5.69		6.01	110

Difference in mean ASPT	% of *n* groups (amongst those with both scores)
Mean Score 1 > Mean Score 2	80% (4 of 5)
Mean Score 1 > Mean Score 3	57% (4 of 7)
Mean Score 2 > Mean Score 3	57% (4 of 7)
Mean Score 1 > Mean Score 4	67% (4 of 6)
Mean Score 2 > Mean Score 4	67% (4 of 6)
Mean Score 3 > Mean Score 4	60% (6 of 10)
Mean Score 3 > Mean Score 5	75% (6 of 8)

Table 11 – Mean value of (a) observed TAXA and (b) observed ASPT for SI module reference sites with each assessment score (1-6) in each TWINSPAN group and consistency across groups of differences in mean TAXA in relation to score.

(a) TAXA	Assessment score						Mean	Sites in
Group	1	2	3	4	5	6	TAXA	group
1	24.7	24.0	24.0				24.2	9
2	19.0	25.4	26.0				24.9	11
3	20.8	21.3	23.0				21.2	13
4	17.9	19.0	18.0				18.2	13
5		15.0	12.0				14.7	9
All	20.0	21.0	20.6				20.7	55

Difference in mean TAXA	% of *n* groups (amongst those with both scores)
Mean Score 1 > Mean Score 2	25% (1 of 4)
Mean Score 1 > Mean Score 3	25% (1 of 4)
Mean Score 2 > Mean Score 3	40% (2 of 5)

(b) ASPT	Assessment score						Mean	Sites in
Group	1	2	3	4	5	6	ASPT	group
1	6.08	6.15	6.29				6.143	9
2	7.05	6.38	6.50				6.454	11
3	6.42	6.70	6.43				6.572	13
4	6.20	6.54	6.17				6.304	13
5	*	5.47	5.25				5.441	9
All	6.29	6.21	6.13				6.230	55

Difference in mean ASPT	% of *n* groups (amongst those with both scores)
Mean Score 1 > Mean Score 2	25% (1 of 4)
Mean Score 1 > Mean Score 3	50% (2 of 4)
Mean Score 2 > Mean Score 3	60% (3 of 5)

Table 12 – Mean value of (a) observed TAXA and (b) observed ASPT for SH module reference sites with each assessment score (1-6) in each TWINSPAN group and consistency across groups of differences in mean TAXA in relation to score.

(a) TAXA	Assessment score						Mean	Sites in
Group	1	2	3	4	5	6	TAXA	group
1	35.0	30.3	33.3	*			31.7	10
2	22.8	27.0	*	*			23.5	6
3	29.2	27.7	*	*			28.6	15
4	22.3	26.5	*	*			23.4	8
5	21.7	19.0	*	*			20.7	11
6	22.0	23.0	21.0	*			22.6	10
7	15.5	18.8	*	*			16.4	14
8	17.5	20.3	*	*			18.3	14
9	18.0	21.0	*	*			18.9	7
10	21.2	21.7	*	19.0			21.2	13
All	21.1	23.8	30.3	19.0			22.4	108

Difference in mean TAXA	% of *n* groups (amongst those with both scores)
Mean Score 1 > Mean Score 2	30% (3 of 10)
Mean Score 1 > Mean Score 3	100% (2 of 2)
Mean Score 2 > Mean Score 3	50% (1 of 2)
Mean Score 1 > Mean Score 4	100% (1 of 1)
Mean Score 2 > Mean Score 4	100% (1 of 1)

(b) ASPT	Assessment score						Mean	Sites in
Group	1	2	3	4	5	6	ASPT	group
1	6.49	6.58	6.45	*			6.53	10
2	6.15	6.59	*	*			6.23	6
3	6.62	6.75	*	*			6.67	15
4	6.47	6.28	*	*			6.42	8
5	6.58	6.45	*	*			6.53	11
6	6.85	6.72	7.05	*			6.78	10
7	6.87	7.14	*	*			6.95	14
8	6.78	6.67	*	*			6.75	14
9	6.43	7.06	*	*			6.61	7
10	6.58	6.77	*	6.00			6.58	13
All	6.61	6.71	6.60	6.00			6.64	108

Difference in mean ASPT	% of *n* groups (amongst those with both scores)
Mean Score 1 > Mean Score 2	40% (4 of 10)
Mean Score 1 > Mean Score 3	50% (1 of 2)
Mean Score 2 > Mean Score 3	50% (1 of 2)
Mean Score 1 > Mean Score 4	100% (1 of 1)
Mean Score 2 > Mean Score 4	100% (1 of 1)

3.3 Statistical Adjustment models based on observed values for reference sites

In the first models developed, the observed value of a biotic index for a reference site is assumed to depend on its biological type (as represented by its TWINSPAN site group *i*) and its assessment score *j* (1-6). By allowing for the effect of site type, the models could be used to estimate the parameters of the (partial) relationship of assessment score with the index values. The site-specific expected values of biotic indices are based on some weighted average of the observed values for the reference sites. Therefore the regression-type estimates of these parameters can be used to adjust the expected values for test sites.

In the following models:

O_{ijk} = Observed index value for the k^{th} site with assessment score *j* in site group *i*
e_{ijk} = residual value for the k^{th} site with assessment score *j* in group *i*
M_i = term for average index value for TWINSPAN site group *i*

P_i = RIVPACS probability test site belongs to TWINSPAN site group *i*
g = number of TWINSPAN site groups

Q_{ij} = Proportion of reference sites in group *i* with assessment score *j*
S = weighted average assessment score for reference sites in the prediction of the expected values for the test site

3.3.1 Model M1: Additive linear

$$O_{ijk} = M_i + b_1 \times j + e_{ijk} \quad \text{(M1)}$$

where b_1 = average effect of a unit increase in assessment score on index values

In this model (M1), the a unit increase in assessment score is assumed to have a constant average additive effect (b_1) on the index values. For example, the difference in index values between sites with scores of 1 versus those with scores of 2, is assumed to be the same as between sites with scores of 3 versus those with scores of 4. The adjustment factor estimate b_1 is expected to be negative for both the TAXA and ASPT indices.

The RIVPACS predicted expected (E) value for a test sites would be adjusted as follows:

$$\text{Adjusted E} = \text{E} + b_1 \times (3 - S) \quad \text{(M1a)}$$

Note that this model M1 is not the same as Robin Guthrie's proposed adjustment method (Section 2.2.1), as Robin's method use the value of S to determine the level of linear adjustment to EQI values rather than to the RIVPACS expected E values.

Results of model M1 fits for each module separately and for all four modules combined are given in Table 13. Model M1 gave a statistically significant (all $p \leq 0.012$) linear effect of score on index values for the GB and NI modules only for both TAXA and ASPT (Table 13). Lack of statistical relationship for the SI and SH modules was not surprising as most of their reference sites were assessed as score 1-3 and our analyses in Tables 11-12 had found no systematic differences in mean index values between sites with this range of scores. The Scottish Islands module also contains relatively few reference sites.

Although statistically significant linear effects of score could only be detected for the GB and NI modules, there were no statistically significant differences in the linear effects between the four modules (i.e. no interaction between score and module) (p = 0.148 and 0.227 for the TAXA and ASPT models M1 respectively). Therefore a single estimate of effect b_1 in model M1 is valid for all modules.

When model M1 was fitted to the data from the reference sites of all four modules combined, but allowing for all TWINSPAN group differences for all modules, the estimates of the average linear effect (b_1) of a unit change in assessment score was -0.48 for TAXA and -0.062 for ASPT (Table 13). Therefore the adjustment to the expected (E) values for test sites would be:

TAXA: $E_{adj} = E - 0.48 \times (3 - S)$, ASPT: $E_{adj} = E - 0062 \times (3 - S)$

3.3.2 Model M2: Additive non-linear

$$O_{ijk} = M_i + A_j + e_{ijk} \qquad \text{(M2)}$$

where A_j = effect of assessment score j on index values (re-scaled to give $A_3 = 0$)

In Model (M2) the effect of a unit change in assessment score is not assumed to be constant across the range of assessment score. This model effectively estimates the average effect of each particular assessment score j on the index values, the parameters are re-scaled so that the parameter A_j represents the average difference in index values for reference sites with score j relative to the values for sites with a score of 3 (the standard for the high/good boundary). This allows the difference in index values between the perceived highest quality sites with scores 1 versus 2, to be less (or more) than between sites with scores 3 versus 4. The adjustment to the expected value for a test site then tries to correct for the proportions of reference sites involved in determining its predicted value which have each assessment score above and below 3.

The adjustment to the RIVPACS predicted expected (E) value is therefore calculated as follows:

Adjusted expected value E for a test site = $E - \sum_{j=1}^{6} R_j A_j$ (M2a)

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

Results of model (M2) fits for each module separately and for all four modules combined are given in Table 13. Model M2, which allows for non-constant effects of a unit change in assessment score on index values, was also only statistically significant for the GB and NI modules, for both TAXA and ASPT (Table 13). This not surprising, as model M1 is a special case of model M2 with constant effects across the full range (1-6) of scores.

When Model M2 was fitted to the data for all four modules combined, there were no statistically significant interactions between the effect of score and module (p = 0.481 and 0.700 for TAXA and ASPT respectively). Therefore a single UK-wide version of model M2 is appropriate for all four modules. The estimates of the model M2 terms A_j for the single UK model M2 are given in the right-hand column of Table 13. The additive factors ($-A_j$) for adjusting expected values for each assessment score are plotted in Figure 7. Assessment score of >5 are treated as 5 for the purposes of adjusting expected values of test sites

As an example, if a test site's prediction was 10% based on reference sites with an assessment score of 2, 20% with a score of 3, 40% with a score of 4 and 30% with score of 5 (i.e. $R_1 = 0$, $R_2 = 0.1$, $R_1 = 0.2$, $R_4 = 0.4$, $R_5 = 0.3$), then the adjusted E value for TAXA is:

$E_{adj} = E - (0.1 \times 0.35 + 0.2 \times 0 + 0.4 \times (-1.15) + 0.3 \times (-2.64)) = E - 1.217$

The estimates of the adjustment factors derived from Models M1 and M2 are plotted together for comparison in Figure 7. Additive linear model M1 estimates that test sites with average assessment scores of 1 should have their expected values reduced by 0.96 (taxa) for TAXA and 0.124 for ASPT, whereas, using model M2, the expected values of such sites would be unchanged (ASPT) or changed very little (TAXA) (Figure 7 & Table 13). As Model M1 is a special case of model M2, it is possible to test for improvement in using model M2 over M1. Model M2 was found to give a statistical significant improvement in fit over Model M1 for both TAXA (F=2.73 with 4 and 821 d.f. ; p = 0.028) and ASPT (F=5.80 with 4 and 821 d.f. ; p <0.001). Thus Model M2 is a better description of the data-based relationship between assessment score and index values for the reference sites, and should be used in preference to the other additive model M1 for adjusting RIVPACS estimates of expected values for both TAXA and ASPT.

Table 13 – Statistical estimates (± SE) of average within-site-group linear adjustment factor b_1 in model (M1) and) and effect (A_i) of each assessment score in model (M2) and for (a) observed TAXA and (b) observed ASPT for the spring and autumn combined samples for the reference sites in each module; *p* = significance probability in test for effect of score or test for interaction with RIVPACS module

(a) TAXA Model (M1)	GB	NI	SI	SH	All ref sites
$p(b_1)$	0.012	<0.001	0.514	0.380	0.001
b_1	-0.47 ± 0.18	-0.92 ± 0.24	0.57 ± 0.86	0.48 ± 0.54	-0.48 ± 0.15
Model (M2)					
p	0.012	<0.001	0.647	0.480	0.001
A_1	-0.06 ± 0.63	3.12 ± 0.92	-0.71 ± 1.84	-1.80 ± 1.84	0.11 ± 0.46
A_2	0.19 ± 0.40	0.64 ± 0.89	0.41 ± 1.69	-0.96 ± 1.75	0.35 ± 0.34
A_3	0.00	0.00	0.00 ±	0.00	0.00
A_4	-1.44 ± 0.57	0.36 ± 0.72	---	-3.92 ± 3.63	-1.15 ± 0.46
A_5	-2.83 ± 0.98	-1.65 ± 0.82	---	---	-2.64 ± 0.69
A_6	-0.35 ± 1.32	---	---	---	-0.26 ± 1.24
(b) ASPT Model (M1)	GB	NI	SI	SH	All ref sites
$p(b_1)$	<0.001	0.010	0.360	0.854	<0.001
b_1	-0.069 ± 0.015	-0.068 ± 0.026	0.055 ±0.059	0.010 ± 0.054	-0.062 ± 0.012
model (M2)					
p	<0.001	<0.001	0.177	0.112	<0.001
A_1	-0.005 ± 0.051	0.116 ± 0.102	-0.040 ± 0.124	-0.083 ± 0.179	0.000 ± 0.039
A_2	0.058 ± 0.033	0.054 ± 0.098	0.107 ± 0.114	0.010 ± 0.170	0.067 ± 0.029
A_3	0.000	0.000	0.000	0.000	0.000
A_4	-0.118 ± 0.046	-0.027 ± 0.079	---	-0.685 ± 0.353	-0.108 ± 0.039
A_5	-0.354 ± 0.080	-0.194 ± 0.090	---	---	-0.298 ± 0.058
A_6	-0.243 ± 0.107	---	---	---	-0.235 ± 0.103

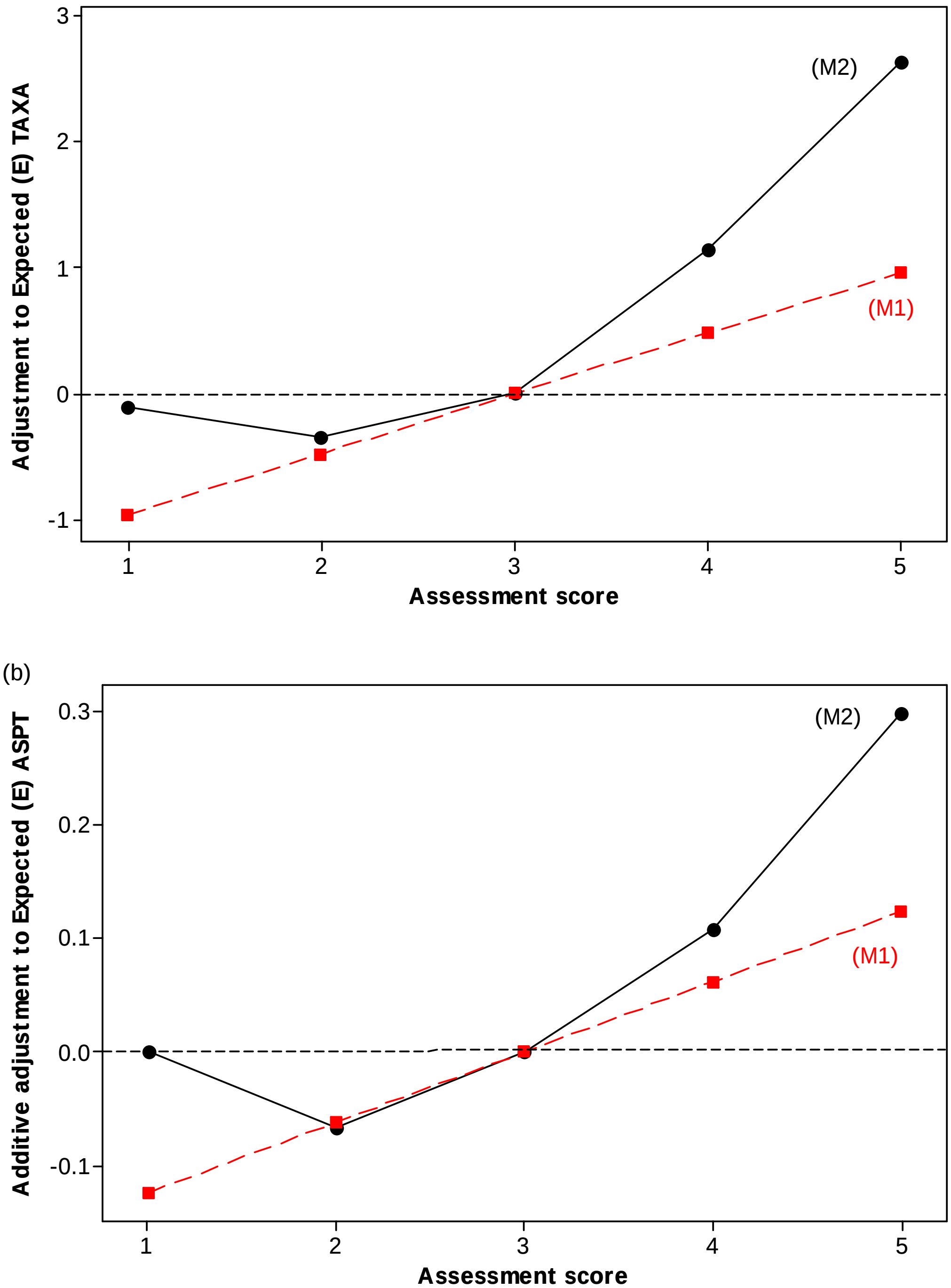
p value in test for interaction between assessment score effects and RIVPACS module

	Model (M1)	Model (M2)
(a) TAXA	0.148	0.481
(b) ASPT	0.227	0.700

Figure 7 – Models M1 and M2 estimates of additive adjustments to Expected (E) values of (a) TAXA and (b) ASPT for all UK RIVPACS modules combined for the spring and autumn combined samples; estimates are $-b_1$ for M1 and $-A_j$ for M2.

(a)

(b)



3.3.3 Model M3: Multiplicative linear

$log_{10}\, O_{ijk} = \log_{10} M_i + b_3 \times j + e_{ijk}$ (M3)

where b_3 = average effect of a unit increase in assessment score on $\log_{10}$ index values (b_3 is expected to be negative)

The fitted regression model can be re-expressed as:

$O_{ij} = M_i (B_3)^j$ where $B_3 = 10^{b_3}$

In this model, a unit increase in assessment score *j* is assumed to have a constant average multiplicative effect (B_3) on the index value. The adjustment factor estimate b_3 is expected to be negative and hence B_3 is expected to be <1.

Adjusted expected value E for a test site = E $(B_3)^{3-S}$ (M3a)

For example, if b_3 = -0.02 then B_3 = 0.955; a test site with a prediction based on value of *S* of 4.5 (i.e. reference sites on average towards the bottom of "good") would have its expected index value multiplied by $(0.955)^{3-4.5} = (0.955)^{-1.5} = 1.071$; namely increased by 7.1%

Results of model (M3) fits for each module separately and for all four modules combined are given in Table 14. The estimates of b_3 were negative, as expected, and statistically significant for the GB module (b_3 = -0.007, p =0.025) and NI module (b_3 = -0.015, p <0.001), but were non-significant and positive for the SI and SH modules. However, there was no statistically significant difference in the values of b_3 between the four modules (Interaction test p = 0.110 for TAXA and 0.221 for ASPT; Table 14).The version of model M3 based on all UK references sites combined gave an estimate of b_3 for TAXA of 0.007 with a SE of 0.003 (p = 0.006). This gives an estimate of $B_3 = 10^{-0.007} = 0.984$. Using this single overall model M3, the adjustment to the RIVPACS predicted expected (E) value is therefore calculated as:

Adjusted expected E value for TAXA for a test site = E $(0.984)^{3-S}$ (M3b)

This means that, for example, if S=2 for a test site the RIVPACS expected number of taxa is multiplied by 0.984 (i.e. decreased by 1.6%) and if S=4, the E for TAXA is multiplied by $0.984^{-1} = 1.0163$ (i.e. increased by 1.63%).

A similar pattern of significant results was found for ASPT. For ASPT, the UK-wide estimate of b_3 = -0.005 (p<0.001), equivalent to $B_3 = 10^{-0.005} = 0.989$, such that the expected values of ASPT for test sites with values of S=2 and S=4 would be decreased by 1.1% and increased by 1.2% respectively.

3.3.4 Model M4: Multiplicative non-linear

$log_{10}\, O_{ijk} = log_{10}\, M_i + a_j + e_{ijk}$ (M4)

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

The fitted regression model can be re-expressed as:

$O_{ij} = M_i A_j$ where $A_j = 10^{a_j}$

In this model, the effect of a unit change in assessment score is not assumed to be constant across the range of assessment score; however, a given assessment score is assumed to a constant multiplicative effect on the observed index values for the reference sites. Specifically an assessment score of *j* is assumed on average to increase the observed index values of reference sites by a factor of A_j (i.e. increase index values by 100(A_j -1) percent). Therefore the expected index values of any sites based on reference sites with a score of *j* are on average over-estimated by a factor A_j. In such cases, the correction should therefore be to divide the RIVPACS expected values by the same factor A_j. 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 expected 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 / $(\sum_{j=1}^{6} R_j A_j)$ (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.*

Results of model (M4) fits for each module separately and for all four modules combined are given in Table 14. Model M4, which allows for non-constant effects of a unit change in assessment score on index values, was statistically significant for the GB and NI modules for TAXA and for the GB module only for ASPT (Table 14). However, when model M4 was fitted to the data for all four modules combined, there were no statistically significant interactions between the effect of score and module (p = 0.422 and 0.709 for TAXA and ASPT respectively). Therefore a single UK-wide version of model M4 is appropriate for all four modules; the model is statistically significant for the effect of score (p = 0.003 for TAXA and <0.001 for ASPT) and the parameter estimates for both indices are given in the right-hand column of Table 14.

The estimates of the multiplicative adjustment factors for RIVPACS expected values derived from UK-wide versions of models M3 and M4 are given in Table 15 and plotted together for comparison in Figure 8. As Model M3 is a special case of model M4, it is possible to test for improvement in using model M4 over M3. Model M4 was found to give a statistical significant improvement in fit over Model M3 for ASPT (F=2.70 with 4 and 821 d.f.; p = 0.030) but not for TAXA (F=0.60 with 4 and 821 d.f. ; p = 0.663). Overall, it is recommended that the adjustment approach used is better based on model M4 than on model M3.

Under UK-wide model M4, no adjustment would be needed if the expected value for a test site was based solely on reference sites with a score of 1 (Table 15 and Figure 8). This is supported by the preliminary analyses in tables 8-12 which showed that within site groups, reference sites with scores of 1 or 3 tended to have the same average index values. If RIVPACS expected values for a test were based are reference sites all with a score of *j* ,then to adjust expected values to a score of 3 (high/good boundary), expected values of TAXA would be left unchanged (j = 1), reduced by 1.1% (j = 2), increased by 4.2% (j = 4) and increased by 10.2% (j = 5), while expected values of ASPT would be left unchanged (j = 1), reduced by 1.1% (j = 2), increased by 1.9% (j = 4) and increased by 5.4% (j = 5) (Table 16). The actual prediction for test sites based on reference sites with a mixture of assessment scores would be based on equation (M4a) above (see Section 4 for a detailed worked example).

Table 14 – Statistical estimates (± standard error) of average within-site-group linear adjustment factor b_3 in multiplicative linear model (M3) and) and effect each assessment score (1-6) in multiplicative non-linear model (M4) and for (a) observed TAXA and (b) observed ASPT for the spring and autumn combined samples for the reference sites in each module; *p* = significance probability in test for effect of score or test for interaction with RIVPACS module

(a) TAXA Model (M3)	GB	NI	SI	SH	All ref sites
$p(b_3)$	0.025	<0.001	0.567	0.311	0.006
b_3	-0.007 ± 0.003	-0.015 ± 0.004	0.016 ± 0.020	0.012 ± 0.011	-0.007 ± 0.003
Model (M4)					
p	0.022	0.001	0.704	0.405	0.003
A_1	-0.001 ± 0.010	0.050 ± 0.015	-0.014 ± 0.043	-0.031 ± 0.039	0.000 ± 0.008
A_2	0.002 ± 0.007	0.009 ± 0.014	0.009 ± 0.039	-0.009 ± 0.037	0.005 ± 0.006
A_3	0.00	0.00	0.00	0.00	0.00
A_4	-0.022 ± 0.009	0.006 ± 0.011	---	-0.072 ± 0.076	-0.018 ± 0.008
A_5	-0.044 ± 0.016	-0.027 ± 0.013	---	---	-0.042 ± 0.012
A_6	-0.002 ± 0.021	---	---	---	-0.001 ± 0.021

(b) ASPT Model (M3)	GB	NI	SI	SH	All ref sites
$p(b_3)$	<0.001	0.012	0.377	0.879	<0.001
b_3	-0.005 ± 0.001	-0.005 ± 0.002	0.004 ±0.004	0.001 ± 0.004	-0.005 ± 0.001
model (M4)					
p	<0.001	0.121	0.188	0.099	<0.001
A_1	-0.000 ± 0.004	0.008 ± 0.008	-0.003 ± 0.009	-0.005 ± 0.012	0.000 ± 0.003
A_2	0.004 ± 0.002	0.005 ± 0.007	0.007 ± 0.008	0.001 ± 0.011	0.005 ± 0.002
A_3	0.000	0.000	0.000	0.000	0.000
A_4	-0.009 ± 0.003	-0.002 ± 0.006	---	-0.046 ± 0.023	-0.008 ± 0.003
A_5	-0.029 ± 0.006	-0.014 ± 0.007	---	---	-0.023 ± 0.004
A_6	-0.021 ± 0.008	---	---	---	-0.020 ± 0.008

p value in test for interaction between assessment score effects and RIVPACS module

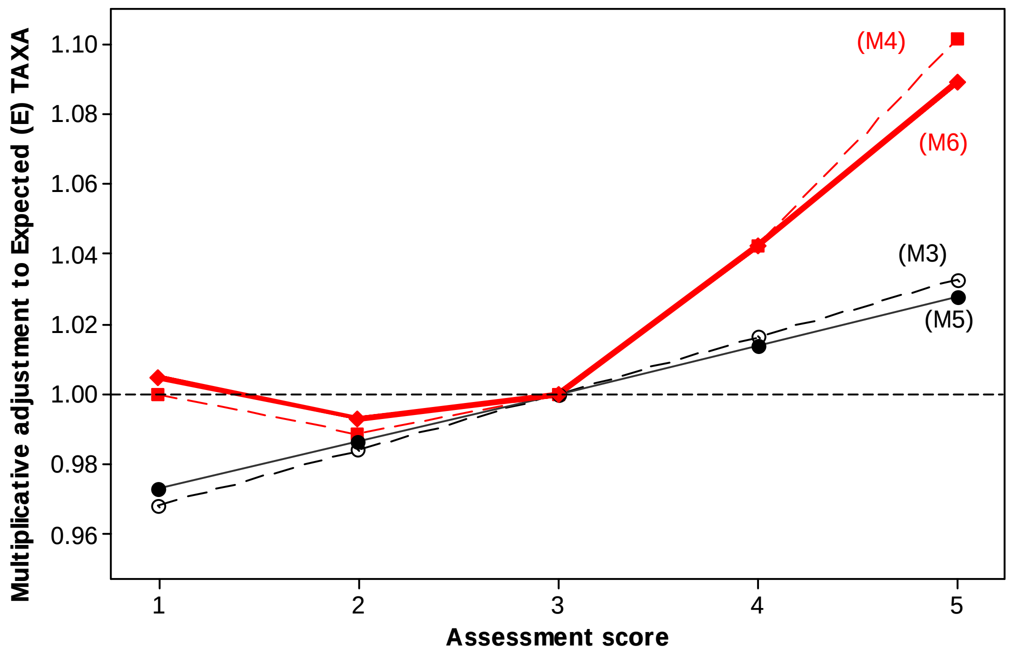
	Model (M3)	Model (M4)
(a) TAXA	0.100	0.422
(b) ASPT	0.221	0.709

Table 15 – Estimates of multiplicative adjustment factors for RIVPACS expected TAXA and expected ASPT based on UK-wide versions of models M3 and M4 using spring and autumn combined samples, and for model M4 based on all possible season combinations ('M4 All').

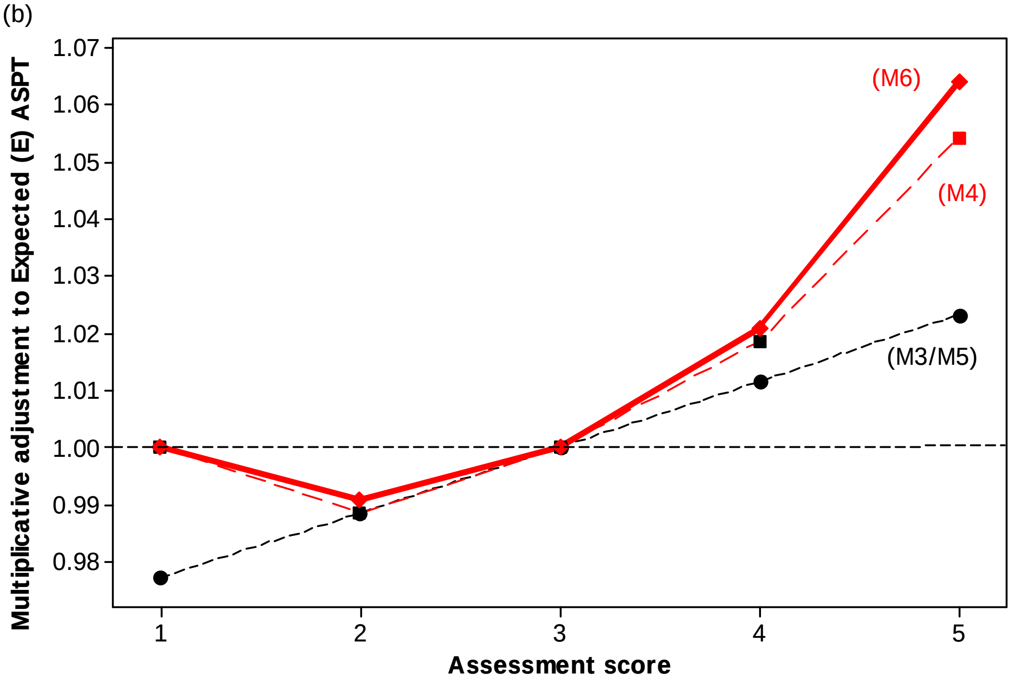
Assessment score	TAXA			ASPT		
	M3	M4	M4 All	M3	M4	M4 All
1	0.968	1.000	1.000	0.977	1.000	0.991
2	0.984	0.989	0.993	0.989	0.989	0.986
3	1.000	1.000	1.000	1.000	1.000	1.000
4	1.016	1.042	1.038	1.012	1.019	1.009
5	1.033	1.102	1.086	1.023	1.054	1.054

Figure 8 – Models M3, M4, M5 and M6 estimates of multiplicative adjustments to Expected (E) values of (a) TAXA and (b) ASPT for all UK RIVPACS modules combined, based on spring and autumn combined samples.

(a)



(b)



3.3.5 Consistency of parameters of models M3 and M4 across RIVPACS season(s)

The calibration of statistical models has been concentrated on fitting the models and estimating the parameters and adjustment factors using the reference sites data for the spring and autumn combined samples. This is because this is the recommended and most commonly used combination of seasons for General Quality Assessment (GQA) and other river quality surveys. However, RIVPACS predictions of the site-specific expected fauna, the expected values of biotic indices and O/E ratios can be made for any single RIVPACS season, Spring (March-May), Summer (June-August) or Autumn (September-November), or for the combined sample obtained from any two or all three of the individual seasons' samples.

Each reference site has been given a single assessment score. Unless the reference sites dramatically changed quality within the year of sampling for RIVPACS, there is no logical reason to expect the parameters of our adjustment models (M1-M4 and others) to vary systematically with the choice of season(s). The estimates of preferred model M4 based on samples for each of the seven possible season combinations are given in Table 16 and the derived multiplicative adjustment factors ($1/A_j$) are plotted in Figure 9.

The adjustment parameter estimates of model M4 do vary to some extent between the RIVPACS season combinations. However, they generally show the same patterns, (Figure 9). In general, for TAXA, the estimated adjustment factors for assessment scores of 1 or 2 are minor, and considerably less than for assessment scores of 4 and especially 5. The general pattern of estimates of adjustment factors for ASPT, regardless of season(s), is to reduce RIVPACS Expected ASPT by 0-2% for assessment scores of 1 or 2, increase Expected ASPT values by 0-2% for scores of 4 and increase by expected values by 2-6% for scores of 5. For both TAXA and ASPT, the greatest discrepancy in estimates occurred when based purely on autumn samples (Figure 9).

It should be remembered that all of the samples from the reference sites on which our models must be calibrated are subject to the same degree of sampling variation as any other RIVPACS samples, as quantified by the Biological Assessment Methods Study (Furse *et al*. 1995) and published in Clarke *et al*. (2002). Clarke *et al*. (2002) showed that sampling variation is considerably greater for single season samples than for two and three season combined samples, and therefore the parameters of our adjustment models are likely to be more variable and less precise when based on single season samples.

We believe it is best to derive a single set of adjustment factors for any one biotic index (e.g. TAXA or ASPT) to be used to adjust RIVPACS expected values regardless of the season(s) on which the assessments of the test sites are to be based. One solution would be to use the estimates derived using the spring and autumn combined season sample data, as this is the seasonal combination most commonly used in national assessments. However, an alternative is to derive some form of average using the data from all of the potential season combinations. This was achieved by fitting an extension of model M4 using the UK-wide sample data for all seven possible season combinations (i.e. 7 x 887 reference sites) which allowed for TWINSPAN group differences in average index values (M_i terms in model M4) separately for each season combination. These parameter estimates for model M4 based on all available data from all season combinations are given in the right-hand column of Table 16 and the derived adjustment factors plotted as the solid black line labelled 'All' in Figure 9.

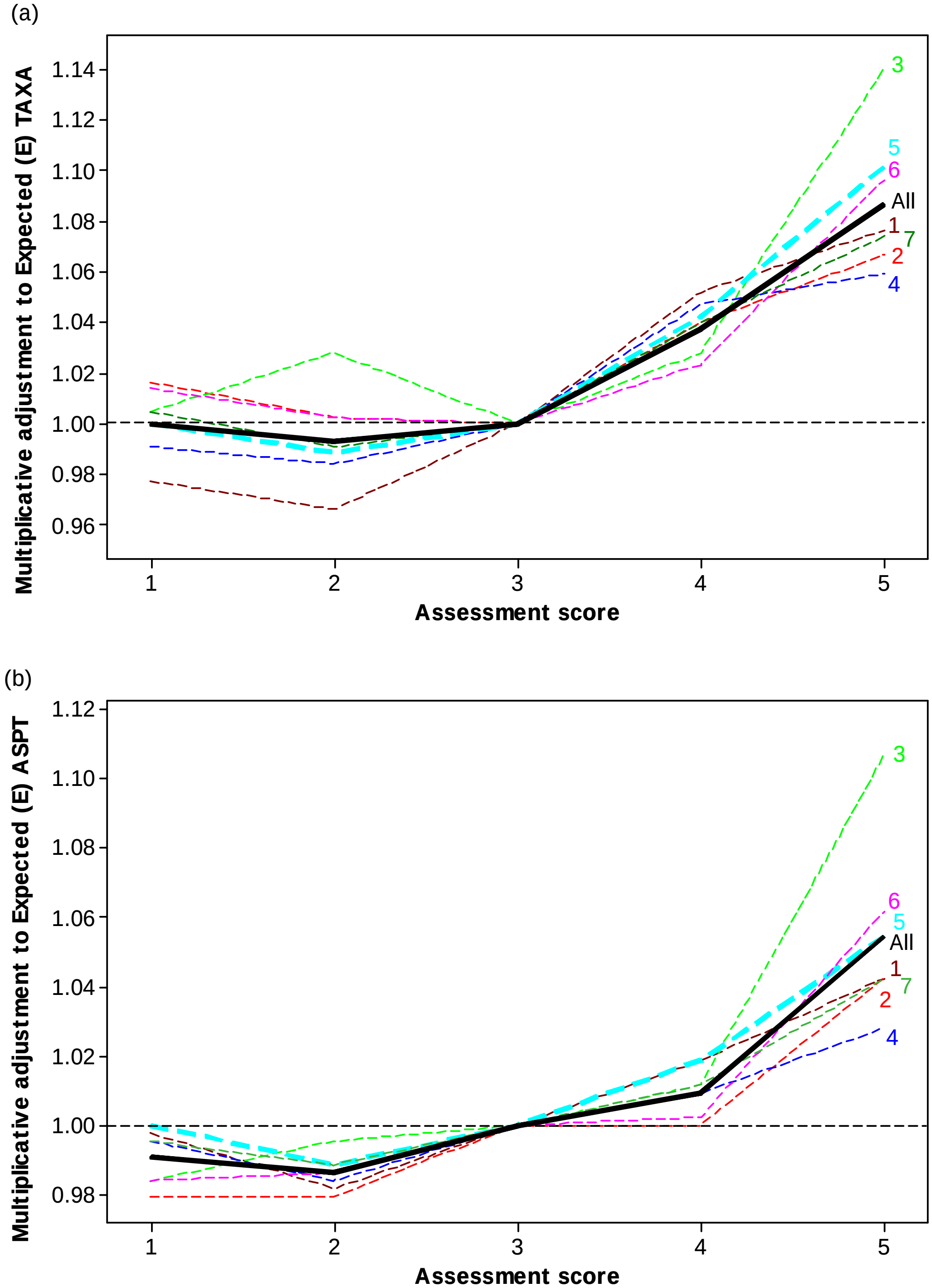
When compared to the previous estimates for model M4 based on the spring and autumn combined sample data, the 'All' combinations data gave similar estimates of multiplicative adjustment factors for expected ASPT and slightly smaller adjustments for TAXA (Table 15). Either set of estimates could be used to adjust RIVPACS expected values for test sites.

Table 16 – Consistency across RIVPACS single and multiple season(s) combinations of the estimates of parameters for the effects of assessment score (1-6) in models (M3) and (M4) for (a) observed TAXA and (b) observed ASPT for UK-wide reference sites; 'All' denotes estimates using data from all possible season(s) options together; *p* = model test probability value,

	Season(s) combination							
(a) TAXA	Spr	Sum	Aut	Spr/Sum	Spr/Aut	Sum/Aut	Spr/Sum /Aut	All seasons options
	1	2	3	4	5	6	7	All
Model (M3)								
$p(b_3)$	0.011	0.122	0.185	0.004	0.006	0.052	0.009	<0.001
b_3	-0.009	-0.005	-0.006	-0.007	-0.007	-0.005	-0.006	-0.006
Model (M4)								
p	0.017	0.212	0.095	0.023	0.003	0.025	0.008	<0.001
A_1	0.010	-0.007	-0.002	0.004	0.000	-0.006	-0.002	0.000
A_2	0.015	-0.001	-0.012	0.007	0.005	-0.001	0.004	0.003
A_3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A_4	-0.022	-0.017	-0.012	-0.020	-0.018	-0.010	-0.017	-0.016
A_5	-0.032	-0.028	-0.057	-0.025	-0.042	-0.040	-0.031	-0.036
A_6	0.025	-0.047	-0.021	-0.013	-0.001	-0.036	-0.012	-0.015

(b) ASPT	Spr	Sum	Aut	Spr/Sum	Spr/Aut	Sum/Aut	Spr/Sum /Aut	All seasons options
Model (M3)	1	2	3	4	5	6	7	All
$p(b_3)$	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
b_3	-0.005	-0.007	-0.007	-0.004	-0.005	-0.006	-0.004	-0.0055
Model (M4)								
p	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
A_1	0.001	0.009	0.007	0.002	0.000	0.007	0.002	0.004
A_2	0.008	0.009	0.002	0.007	0.005	0.006	0.005	0.006
A_3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A_4	-0.008	0.000	-0.005	-0.004	-0.008	-0.001	-0.005	-0.004
A_5	-0.018	-0.018	-0.044	-0.012	-0.023	-0.026	-0.018	-0.023
A_6	-0.018	-0.047	-0.025	-0.025	-0.020	-0.034	-0.021	-0.027

Figure 9 – Seasonal variation in model M4 estimates of multiplicative adjustments (1/A_j) to Expected (E) values of (a) TAXA and (b) ASPT for all UK RIVPACS modules combined; 1=spring, 2=summer, 3=autumn, 4=spr/sum, 5=spr/aut, 6=sum/aut, 7=spr/sum/aut, All = data from all 7 season combinations.



3.3.6 Model M5: Multiplicative linear on Expected values

Models M3 and M4 were based on analysing variation in the observed index values of reference sites in relation to assessment score (1-6) after allowing for differences between TWINSPAN groups in average values of the indices for the reference sites. RIVPACS expected values for sites (including the reference sites) are based on weighted averages of these means of the observed values for the reference sites in each TWINSPAN group. The adjustments derived in this project are to be applied to RIVPACS predictions of expected values.

Therefore, models M3 and M4 were modified slightly to analyse variation in assessment score about the RIVPACS expected values, rather than about their TWINSPAN group averages. This gave rise to corresponding models M5 and M6 (see Section 3.3.7):

$$log_{10} O_{ijk} = \log_{10} E_{ijk} + b_3 \times j + e_{ijk} \qquad \text{(M5)}$$

where b_5 = average effect of a unit increase in assessment score on $\log_{10}$ index values (b_5 is expected to be negative)

This model can be re-expressed and most easily and appropriately fitted as:

$$log_{10} O_{ijk} - \log_{10} E_{ijk} = \log_{10} (O_{ijk} / E_{ijk}) = b_3 \times j + e_{ijk} \qquad \text{(M5a)}$$

The fitted regression model can be re-expressed as:

$$O_{ij} k = E_{ijk} (B_5)^j \qquad \text{where } B_5 = 10^{b_5}$$

In this model, a unit increase in assessment score *j* is assumed to have a constant average multiplicative effect (B_5) on the index value. The adjustment factor estimate b_5 is expected to be negative and hence B_5 is expected to be <1.

$$\text{Adjusted expected value E for a test site = E } (B_5)^{3-S} \qquad \text{(M5b)}$$

Results of fitting model (M5) to the reference site data for all four UK modules combined are given in Table 17. When based on the data from all seven season options, the version of model estimate of b_5 for TAXA was -0.0041 with a SE of 0.00093 ($p < 0.001$). This gives an estimate of $B_5 = 10^{-0.0041} = 0.991$, and therefore the adjustment to the RIVPACS predicted expected (E) value is calculated as:

$$\text{Adjusted expected E value of TAXA for a test site = E } (0.991)^{3-S} \qquad \text{(M5c)}$$

The expected values of TAXA for test sites with values of S of 1 or 2 would be decreased by 1.8% and 0.9% respectively and test sites with values of S of 4 or 5 would be increased by 0.9% and 1.8% respectively.

A similar pattern of results was obtained for ASPT (Table 17). For ASPT, the equivalent all-data estimate of b_5 in model M5 was -0.0049 (p<0.001), equivalent to $B_5 = 10^{-0.0049} = 0.989$, the same as found for B_3 in model M3 for ASPT, such that, as with model M3, the expected values of ASPT for test sites with values of S=2 and S=4 would be decreased by 1.1% and increased by 1.2% respectively.

3.3.7 Model M6: Multiplicative non-linear on Expected values

Model M6 is a variant of model M4 which allows for non-linear effects of assessment score about the RIVPACS expected values, rather than about the TWINSPAN group average index values:

$$log_{10} O_{ijk} = log_{10} E_{ijk} + a_j + e_{ijk} \quad \text{(M6)}$$

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

This model can be re-expressed and most easily and appropriately fitted as:

$$log_{10} O_{ijk} - \log_{10} E_{ijk} = \log_{10} (O_{ijk} / E_{ijk}) = a_j + e_{ijk} \quad \text{(M6a)}$$

The fitted regression model can be re-expressed as:

$$O_{ij} = E_{ijk} A_j \quad \text{where} \quad A_j = 10^{a_j}$$

As with model M4, when based on model M6, the multiplicative adjustment factor for the expected values are based on a weighted average of the factors ($1/A_j$), as follows:

Adjusted expected value E for a test site = E $(\sum_{j=1}^{6} R_j / A_j)$ (M6b)

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.*

Results of fitting model (M5) to the reference site data for all four UK modules combined are given in Table 17. The estimates of the multiplicative adjustment factors for RIVPACS expected values derived from UK-wide versions of model M6 fitted to the data for each of the seven RIVPACS season combinations are given in Table 17 and plotted together for comparison in Figure 10. In practical terms, the estimates of adjustment factors are similar for all season, the only exception being for autumn samples (RIVPACS season combination 3), which gives rise to higher adjustment factors, $1/A_2$ and $1/A_5$, for sites with assessment scores of 2 and 5 (Figure 10).

As Model M5 is a special case of model M6, it is possible to test for improvement in using model M6 over model M5. Based on the all season combinations combined data, model M6 was found to give a statistical significant improvement in fit over model M5 for both TAXA (F=9.53 with 4 and 6199 d.f. ; p <0.001) and ASPT (F=3.61 with 4 and 6199 d.f.; p =0.006).

Overall, we recommend that the adjustment approach used is better based on the non-linear model M6 than on model M5.

Under UK-wide model M6 based on spring and autumn combined samples (season option 5), almost no adjustment would be needed if the expected value for a test site was based solely on reference sites with a score of 1 or 2 ($1/A_1$ = 1.007 and $1/A_2$ = 0.991 in Table 18 and Figure 10). Overall, using these model estimates, if RIVPACS expected values for a test were based are reference sites all with a score of *j*, then to adjust expected values to a score of 3 (high/good boundary), expected values of TAXA would be increased by 0.7% (*j* = 1), reduced by 0.9% (*j* = 2), increased by 4.2% (*j* = 4) and increased by 6.7% (*j* = 5) (Table 18).

Similarly, expected values of ASPT, based on the same spring and autumn combined sample fit to model M6, would be increased by 0.5% (*j* = 1), reduced by 0.9% (*j* = 2), increased by 1.4% (*j* = 4) and increased by 5.2% (*j* = 5) (Table 18 and Figure 10).

Figure 10 – Seasonal variation in model M6 estimates of multiplicative adjustments (1/A_j) in relation to assessment score *j* for Expected (E) values of (a) TAXA and (b) ASPT for all UK RIVPACS modules combined; 1=spring, 2=summer, 3=autumn, 4=spr/sum, 5=spr/aut, 6=sum/aut, 7=spr/sum/aut, All = data from all 7 season combinations.

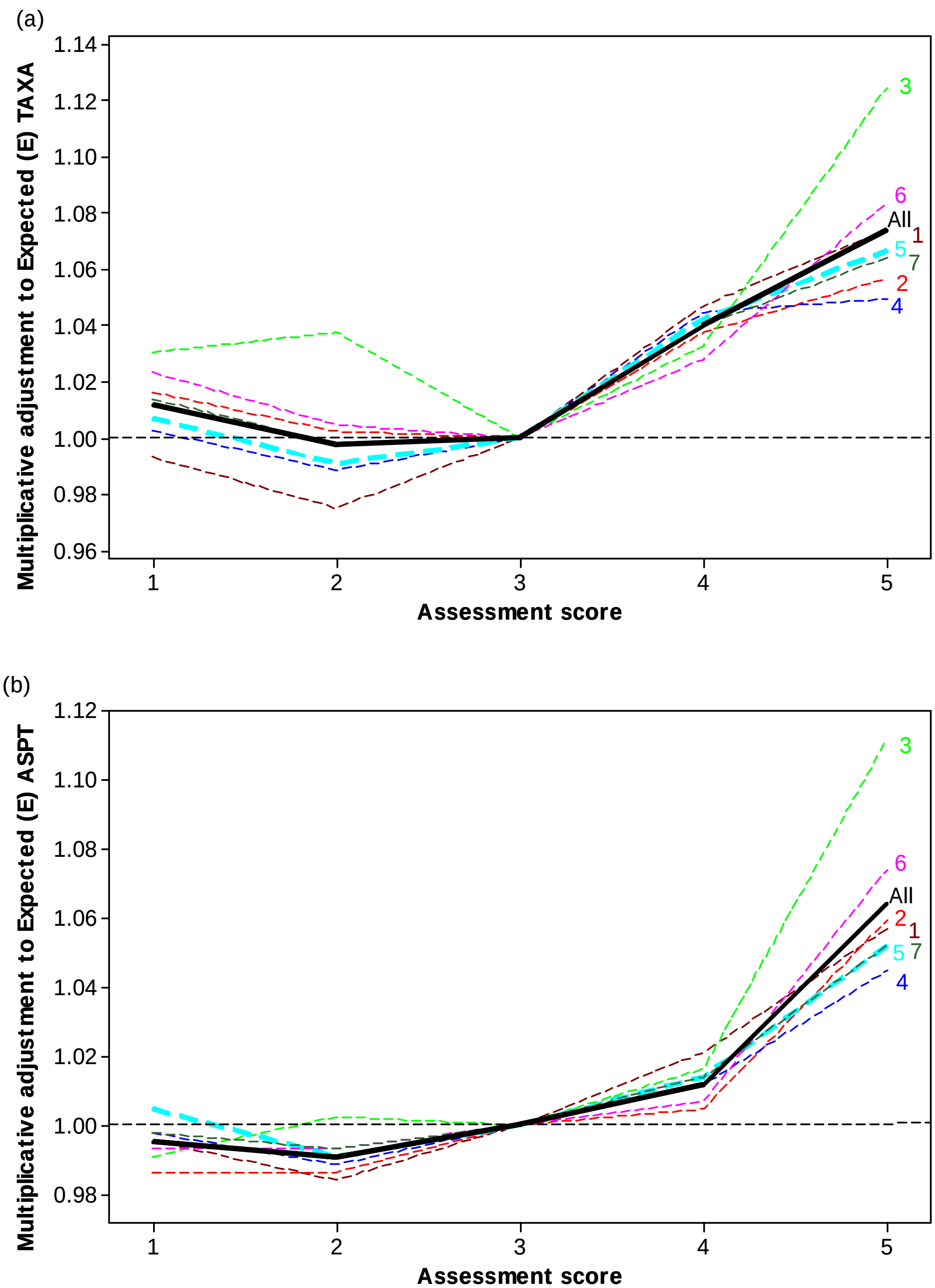


Table 17 – Consistency across RIVPACS single and multiple season(s) combinations of the estimates of parameters for the effects of assessment score (1-6) in models (M5) and (M6) for (a) observed TAXA and (b) observed ASPT for UK-wide reference sites; 'All' denotes estimates using data from all possible season(s) options together; *p* = model test probability value,

	Season(s) combination							
(a) TAXA	Spr	Sum	Aut	Spr/Sum	Spr/Aut	Sum/Aut	Spr/Sum /Aut	All seasons options
	1	2	3	4	5	6	7	All
Model (M5)								
$p(b_5)$	0.012	0.126	0.703	0.013	0.022	0.820	0.422	<0.001
b_5	-0.0070	-0.0041	-0.0013	-0.0051	-0.0048	-0.0031	-0.0036	-0.0041
Model (M6)								
p	0.020	0.195	0.137	0.035	0.023	0.035	0.012	<0.001
A_1	0.003	-0.007	-0.013	-0.001	-0.003	-0.010	-0.006	-0.005
A_2	0.011	-0.001	-0.016	0.005	0.004	-0.002	0.001	0.001
A_3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A_4	-0.020	-0.016	-0.014	-0.019	-0.018	-0.012	-0.017	-0.017
A_5	-0.031	-0.024	-0.051	-0.021	-0.028	-0.035	-0.027	-0.031
A_6	-0.016	-0.044	=0.016	-0.015	-0.011	-0.034	-0.013	-0.017
(b) ASPT	Spr	Sum	Aut	Spr/Sum	Spr/Aut	Sum/Aut	Spr/Sum /Aut	All seasons options
	1	2	3	4	5	6	7	All
Model (M5)								
$p(b_5)$	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
b_5	-0.0051	-0.0060	-0.0067	-0.0040	-0.0036	-0.0052	-0.0040	-0.0049
model (M6)								
p	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
A_1	0.002	0.006	0.004	0.001	-0.002	0.003	0.001	0.002
A_2	0.007	0.006	-0.001	0.005	0.004	0.003	0.003	0.004
A_3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A_4	-0.009	-0.002	-0.007	-0.005	-0.006	-0.003	-0.006	-0.005
A_5	-0.024	-0.025	-0.046	-0.019	-0.022	-0.031	-0.022	-0.027
A_6	-0.022	-0.047	-0.028	-0.027	-0.030	-0.035	-0.024	-0.030

Table 18 – Estimates of multiplicative adjustment factors ($1/A_j$) for RIVPACS expected TAXA and ASPT based on UK-wide versions of non-linear models M4 and M6 estimated from spring and autumn combined sample data (Spr/Aut) and from the samples obtained from all seven possible combinations of one or more seasons.

Assessment score	TAXA				ASPT			
	M4 Spr/Aut	M4 All	M6 Spr/Aut	M6 All	M4 Spr/Aut	M4 All	M6 Spr/Aut	M6 All
1	1.000	1.000	1.007	1.012	1.000	0.991	1.005	0.995
2	0.989	0.993	0.991	0.998	0.989	0.986	0.991	0.991
3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
4	1.042	1.038	1.042	1.040	1.019	1.009	1.014	1.012
5	1.102	1.086	1.067	1.074	1.054	1.054	1.052	1.064

3.4 Summary of statistical models and recommendations

The multiplicative models M3-M6 are preferred over the additive models M1 and M2, as the latter, adjust the RIVPACS expected values by multiplying by a factor which depends on the assessments score of the reference sites involved in determining the predictions. Adjustments derived from models M3-M6 are proportional to the original expected values, as seems logical, as, for a given set of assessment scores, the under-prediction of, say, expected taxa, is likely to be numerically greatest for test sites which have the highest expected taxa richness.

Models M3 and M5 assume that the effect of a unit change in assessment score is always to reduce the RIVPACS expected value by the same proportion, regardless of the assessment score values. Non-linear multiplicative models M4 and M6 allow the proportional reduction in RIVPACS expected value to vary across the range 1-6 of assessment scores. Models M4 and M6 have been shown to give statistically significant improvements in fit, over models M3 and M5 respectively, to the observed data for the reference sites.

In particular, the data modelling suggests that there are only negligible systematic differences in the observed values of TAXA and ASPT between reference sites of the same type with assessment scores of 1, 2 or 3. Observed index values are lower relatively for reference sites with assessment scores of 4 and especially 5. (There were insufficient reference sites with scores of 6 (worse than good/moderate boundary) to estimate its adjustment factor and it is recommended that any such sites are treated as having assessments scores of 5 in the adjusted of test site expected values.)

Model M4 estimates the effects of assessment score after allowing for differences in mean levels of index values of reference sites between TWINSPAN sites groups. Model M6 estimates the impact of assessment score on systematic deviations of observed values for reference sites from the RIVPACS expected values, assessed on the logarithmic scale. Although both models M4 and M6 have their merits, model M4 based on allowing for TWINSPAN group differences is recommended. (One advantage of the M4 type of model is that the approach could also be used in other European ecoregions where reference condition expected values of metrics for a test site are based on some average or percentile value of the reference sites in the same site type according to a WFD System A or B stream typology (Council of the European Union, 2000) (i.e. group sites by WFD site type rather than TWINSPAN group in the models).

Any derived adjustment factors may need to be used to adjust RIVPACS predictions based on macroinvertebrate samples from any single (spring, summer or autumn) or combination of RIVPACS seasons. Adjustment estimates have been derived using each of the seven possible season combinations, and they broadly give similar results. The recommended procedure for GQA type national surveys is to use spring and autumn combined samples.

We therefore recommend using adjustment factors derived from model estimates based on the spring and autumn combined sample data for the reference sites. However, the adjustment factors based on fitting the models to the data from 'All' seven season combinations may seem more generally applicable and give broadly similar levels of adjustment to expected values.

The actual prediction for test sites based on reference sites with a mixture of assessment scores would be based on equation (M4a) above. We recommend that no adjustment is needed for the proportion of reference sites which have assessment scores of 1, 2 or 3, and that an increase in expected values is made using equation M4a according to the weighted proportions, R_4 and R_5, of reference sites with assessment scores of 4 and 5 respectively (see Section 5.2 for a detailed worked example).

4. ADJUSTMENT ESTIMATES BASED ON WFD SYSTEM A SITE GROUPS

4.1 Rationale

Models M1-M4 were based on analysing variation in the observed index values of reference sites in relation to assessment score (1-6) after allowing for differences between TWINSPAN groups in average values of the indices. Models M5-M6 were based on analysing variation in the observed index values of reference sites in relation to assessment score (1-6) after allowing for differences in RIVPACS expected values for sites; expected values being based on weighted averages of the TWINSPAN group mean values.

In the development of the RIVPACS approach of the years, these TWINSPAN groups have been used to represent the major "natural" types of UK river sites. This assumption was fine when the reference sites were taken to represent, in a pre-WFD sense, high quality unstressed sites. However, if the RIVPACS reference sites have been "re-assessed" to be of varying quality in terms of WFD definitions of ecological status class criteria (Appendix 1), then it is possible that some slightly or moderately stressed reference sites may have had their macroinvertebrate fauna sufficiently altered by the anthropogenic stresses operating at the time of sampling for RIVPACS that the biological composition led to be assigned to a different (although relatively similar) TWINSPAN site group. If some types of the reference sites vary sufficiently in biological quality, then the TWINSPAN groups may no longer provide a completely valid "natural" grouping of river sites. This may affect our estimates of the adjustment factors for expected values based on models M1-M6.

An alternative might be to use a purely environmental classification of river sites which is completely independent of the biological composition and quality of the reference sites at the time of sampling. The obvious readily-available choice is the WFD System A classification based on broad categories of altitude, catchment area and geology.

However, in a recent paper in the journal Hydrobiologia, Davy-Bowker *et al.* (2006) showed that the System A classification of GB reference sites led to much poorer prediction of the expected fauna and expected values of biological indices than predictions based on the RIVPACS TWINSPAN site classification and subsequent multiple discriminant analysis (MDA). In particular, compared to a null model whereby all site predictions are the same overall mean observed value of all 614 reference sites, the standard deviation (SD) of O/E for TAXA is reduced by only 5-6% when based on System A groups but by 17-27% based on RIVPACS, while SD of O/E for ASPT is reduced by 11-15% using System A and 34-45% using RIVPACS. Thus index values are far more variable in System A site groups than in TWINSPAN groups and thus predictions based on the mean observed index values of System A reference sites groups are far less precise than those based on RIVPACS expected values. A wide range of stream types and biological communities can be encompassed within one System A type. However, this alternative adjustment approach was also assessed.

The category of each UK RIVPACS reference site for (a form of) each of these three variables and thus the WFD System A site group of each site was available from the RIVPACS reference site database developed within SNIFFER project WFD46 (Davy-Bowker *et al*. 2006) (This database will soon be freely available off the web). This purely-environmental classification of the reference sites was used as an alternative grouping of the reference sites.

4.2 Statistical analysis and model fitting

The distribution of assessment scores (1-6) within each WFD System A site group is given for each of the four RIVPACS modules in Tables 19-22. Although the 614 GB sites form 16 System A groups, two groups are very large with 117 sites (<200m, 10-100 km^2, calcareous) and 116 sites (<200m, 100-1000 km^2, calcareous). Twenty of the 28 GB reference sites with assessment scores of 5 or 6 occur in just two groups and the remainder in three other groups and therefore any derived adjustment factor for the effect of scores of 5 or 6 are predominantly based on the extent of differences in average index value with assessment scores in these two site groups. In the other three modules, the reference sites come from13 (NI), 6 (SI) and 11 (SH) WFD System A site groups.

Table 19 – Number of GB module reference sites in each assessment score (1-6) and mean score (Q_i) for sites in each WFD System A site group *i*, and overall.

WFD Type A group			Assessment Score						Total Sites	Mean Score
Alt (m)	Area (km^2)	Geol	1	2	3	4	5	6		Q_i
<200	10-100	Sil	7	18	29	5	2	3	64	2.78
		Cal	8	34	46	19	9	1	117	2.92
		Org	3	0	0	0	0	0	3	1.00
	100-1000	Sil	1	6	12	3	1	0	23	2.87
		Cal	1	30	59	16	4	6	116	3.09
		Org	1	7	1	0	0	0	9	2.00
	>1000	Sil	0	0	3	0	0	0	3	3.00
		Cal	1	3	11	10	2	0	27	3.33
200-800	10-100	Sil	23	19	14	2	0	0	58	1.91
		Cal	14	19	8	1	0	0	42	1.91
		Org	6	2	0	0	0	0	8	1.25
	100-1000	Sil	8	33	26	5	0	0	72	2.39
		Cal	3	21	21	2	0	0	47	2.47
		Org	0	3	0	0	0	0	3	2.00
	>1000	Sil	0	2	8	2	0	0	12	3.00
		Cal	0	4	5	1	0	0	10	2.70
	All		76	201	243	66	18	10	614	2.64
	% of sites		12.4%	32.7%	39.6%	10.8%	2.9%	1.6%	100.0%	

Table 20 – Number of Northern Ireland (NI) module reference sites in each assessment score (1-6) and mean score (Q_i) for sites in each WFD System site group *i*, and overall.

WFD Type A group			Assessment Score						Total Sites	Mean Score
Alt (m)	Area (km^2)	Geol	1	2	3	4	5	6		Q_i
<200	10-100	Sil	1	1	0	4	2	0	8	3.63
		Cal	5	2	7	9	9	0	32	3.47
		Org	1	0	2	1	0	0	4	2.75
	100-1000	Sil	1	0	2	3	2	0	8	3.63
		Cal	0	3	11	6	2	0	22	3.32
	>1000	Cal	0	0	1	0	0	0	1	3.00
200-800	<10	Org	0	0	0	0	1	0	1	5.00
	10-100	Sil	0	2	3	2	1	0	8	3.25
		Cal	1	2	5	4	1	0	13	3.15
		Org	2	3	1	1	1	0	8	2.50
	100-1000	Sil	0	0	1	0	0	0	1	3.00
		Cal	1	0	0	2	0	0	3	3.00
		Org	0	0	1	0	0	0	1	3.00
	All		12	13	34	32	19	0	110	3.30
	% of sites		10.9%	11.8%	30.9%	29.1%	17.3%	0.0%	100.0%	

Table 21 – Number of Scottish Islands (SI) module reference sites in each assessment score (1-6) and mean score (Q_i) for sites in each WFD System A site group *i*, and overall.

WFD Type A group			Assessment Score						Total Sites	Mean Score
Alt (m)	Area (km^2)	Geol	1	2	3	4	5	6		Q_i
<200	<10	Org	1	2	0	0	0	0	3	1.67
	10-100	Sil	2	9	0	0	0	0	11	1.82
		Cal	2	7	1	0	0	0	10	1.90
		Org	7	8	4	0	0	0	19	1.84
200-800	10-100	Sil	2	3	0	0	0	0	5	1.60
		Cal	3	4	0	0	0	0	7	1.57
			17	33	5	0	0	0	55	1.78
			30.9%	60.0%	9.1%	0.0%	0.0%	0.0%	100.0%	

Table 22 – Number of Scottish Highlands (SH) module reference sites in each assessment score (1-6) and mean score (Q_i) for sites in each WFD System A site group *i*, and overall.

WFD Type A group			Assessment Score						Total Sites	Mean Score
Alt (m)	Area (km^2)	Geol	1	2	3	4	5	6		Q_i
<200	<10	Sil	0	0	0	1	0	0	1	4.00
	10-100	Sil	7	1	0	0	0	0	8	1.13
		Cal	3	1	0	0	0	0	4	1.25
		Org	3	0	0	0	0	0	3	1.00
	100-1000	Org	1	6	0	0	0	0	7	1.86
200-800	<10	Sil	2	0	0	0	0	0	2	1.00
	10-100	Sil	31	11	0	0	0	0	42	1.26
		Org	2	1	0	0	0	0	3	1.33
	100-1000	Sil	15	14	2	0	0	0	31	1.58
		Org	0	1	0	0	0	0	1	2.00
	>1000	Sil	0	4	2	0	0	0	6	2.33
			64	39	4	1	0	0	108	1.46
			59.3%	36.1%	3.7%	0.9%	0.0%	0.0%	100.0%	

Tables 23-27 summarise the differences in mean observed TAXA and ASPT between references in each WFD System A group, separately for each RIVPACS module. Average number of TAXA does not tend to be consistently higher for reference sites with assessment scores of 1 than for sites in the same System A group with scores of 2 or 3; the only exception being for Northern Ireland where mean TAXA for sites with scores of 1 or 2 are higher than those for sites with scores of 3 in 71% and 75% of System A groups (Table 25).

In the GB module there is some tendency for mean TAXA in a group to be higher for sites with scores of 1, 2 or 3 than for sites with scores of 4, 5 or 6 (Table 23). In contrast, for the Northern Ireland reference sites, there is no consistent tendency for mean TAXA to be higher for sites with scores of 1, 2 or 3 than for sites with scores of 4 or 5 (Table 25). No sites have assessment scores of greater than 3 in the Scottish Islands module and only one site has a score of 4 in the Scottish Highlands module (Tables 26-27).

Table 23 – Mean value of observed TAXA for GB module reference sites with each assessment score (1-6) in each WFD System A group and consistency across groups of differences in mean TAXA in relation to score.

WFD Type A group			Assessment Score						Mean TAXA	Sites in group
Alt (m)	Area (km^2)	Geol	1	2	3	4	5	6		
<200	10-100	Sil	26.1	32.4	29.6	25.4	25.5	32.3	29.7	64
		Cal	23.0	27.3	28.1	26.2	25.0	25.0	26.9	117
		Org	22.7						22.7	3
	100-1000	Sil	28.0	35.5	31.7	24.3	28.0		31.4	23
		Cal	36.0	32.3	30.4	30.5	26.8	27.2	30.7	116
		Org	26.0	26.9	25.0				26.6	9
	>1000	Sil			30.0				30.0	3
		Cal	33.0	27.7	33.0	30.4	24.5		30.8	27
200-800	10-100	Sil	23.0	25.8	27.1	24.5			25.0	58
		Cal	26.3	26.8	29.9	30.0			27.3	42
		Org	20.5	18.5					20.0	8
	100-1000	Sil	23.3	27.3	27.0	22.4			26.4	72
		Cal	22.3	26.5	29.0	25.5			27.3	47
		Org		20.0					20.0	3
	>1000	Sil		24.5	26.3	25.5			25.8	12
		Cal		29.3	32.2	24.0			30.2	10
	All									614

Difference in mean TAXA	% of *n* groups (amongst those with both scores)
Mean Score 1 > Mean Score 2	27% (3 of 11)
Mean Score 1 > Mean Score 3	30% (3 of 10)
Mean Score 2 > Mean Score 3	42% (5 of 12)
Mean Score 1 > Mean Score 4	56% (5 of 9)
Mean Score 2 > Mean Score 4	73% (8 of 11)
Mean Score 3 > Mean Score 4	82% (9 of 11)
Mean Score 3 > Mean Score 5	100% (5 of 5)
Mean Score 3 > Mean Score 6	67% (2 of 3)

In both the GB and NI modules, mean ASPT for sites with scores of 1 was greater than mean ASPT for sites with scores of 2 or 3 in the majority of System A groups (Tables 24-25), but this was not the case for either the Scottish Islands or Scottish Highlands modules (Tables 26-27). Mean ASPT for sites with scores of 2 was greater than mean ASPT for sites with scores of 3 in the majority of groups only for the GB module (71% of the 12 groups). However, in both the GB and NI modules where scores of 4 or more were available, mean ASPT was higher for sites with scores of either 1, 2 or 3 than for sites with scores of 4, 5 or 6 in the majority of System A groups in all cases (Table 25) .

In summary, this exploratory analysis suggests that there is some inconsistent evidence for mean TAXA or mean ASPT to be higher for sites with scores of 1 (top of high) or 2 (middle of high) than for sites with scores 3 (high/good boundary). More importantly, there does appear to be a decrease in mean ASPT and, to a less consistent extent, TAXA with assessment scores of 4 (middle of good), 5 (good/moderate boundary) or 6 (worse), as found in the earlier statistical analyses based on TWINSPAN groups (Section 3).

Table 24 – Mean value of observed ASPT for GB module reference sites with each assessment score (1-6) in each WFD System A group and consistency across groups of differences in mean ASPT in relation to score.

WFD Type A group			Assessment Score						Mean ASPT	Sites in group
Alt (m)	Area (km^2)		1	2	3	4	5	6		
<200	10-100	Sil	6.40	6.14	6.21	6.16	5.06	5.27	6.12	64
		Cal	5.81	5.99	5.52	5.58	5.14	5.32	5.66	117
		Org	6.60						6.60	3
	100-1000	Sil	5.82	6.44	6.35	5.39	4.93		6.16	23
		Cal	5.78	5.72	5.32	5.36	4.88	4.79	5.39	116
		Org	6.69	6.43	5.80				6.39	9
	>1000	Sil			5.42				5.42	3
		Cal	5.18	5.18	5.55	5.37	4.73		5.37	27
200-800	10-100	Sil	6.68	6.60	6.60	6.62			6.63	58
		Cal	6.62	6.37	6.44	5.93			6.46	42
		Org	6.61	6.15					6.50	8
	100-1000	Sil	6.43	6.53	6.43	5.91			6.44	72
		Cal	6.39	6.50	6.19	5.45			6.31	47
		Org		6.50					6.50	3
	>1000	Sil		6.65	6.46	5.92			6.40	12
		Cal		6.18	5.88	5.50			5.96	10
	All									614

Difference in mean ASPT	% of *n* groups (amongst those with both scores)
Mean Score 1 > Mean Score 2	55% (6 of 11)
Mean Score 1 > Mean Score 3	75% (7.5 of 10)
Mean Score 2 > Mean Score 3	71% (8.5 of 12)
Mean Score 1 > Mean Score 4	89% (8 of 9)
Mean Score 2 > Mean Score 4	73% (8 of 11)
Mean Score 3 > Mean Score 4	73% (8 of 11)
Mean Score 3 > Mean Score 5	100% (5 of 5)
Mean Score 3 > Mean Score 6	100% (3 of 3)

To derive quantitative estimates of adjustment factors based on WFD System A groups, models (M3) and (M4) were modified accordingly to give models (M7) and (M8) respectively as follows:

$$log_{10} O_{ijk} = \log_{10} M_i + b_3 \times j + e_{ijk} \quad \text{(M7)}$$

where M_i = term for average index value for WFD System A site group *i*
b_3 = average effect of a unit increase in assessment score on $\log_{10}$ index values

The fitted regression model can be re-expressed as: $O_{ij} = M_i (B_7)^j$ where $B_7 = 10^{b_7}$

Adjusted expected value E for a test site = E $(B_7)^{3-S}$ (M7a)

$$log_{10} O_{ijk} = log_{10} M_i + a_j + e_{ijk} \quad \text{(M8)}$$

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

The fitted regression model M8 can be re-expressed as: $O_{ij} = M_i A_j$ where $A_j = 10^{a_j}$.

Results of fitting models (M7) and (M8) for each module separately and for all four modules combined are given in Table 28.

Table 25 – Mean value of (a) observed TAXA and (b) observed ASPT for Northern Ireland (NI) module reference sites with each assessment score (1-6) in each WFD System A group and consistency across groups of differences in mean TAXA in relation to score.

(a) TAXA

WFD Type A group			Assessment Score						Mean TAXA	Sites in group
Alt (m)	Area (km^2)		1	2	3	4	5	6		
<200	10-100	Sil	25.0	30.0		28.8	27.0		28.0	8
		Cal	30.0	30.5	27.3	26.6	27.4		27.8	32
		Org	25.0		29.5	33.0			29.3	4
	100-1000	Sil	32.0		28.5	30.0	26.5		29.0	8
		Cal		32.0	29.9	28.0	30.0		29.7	22
	>1000	Cal			28.0				28.0	1
200-800	<10	Org					20.0		20.0	1
	10-100	Sil		22.5	22.0	23.0	30.0		23.4	8
		Cal	30.0	30.0	26.0	28.3	22.0		27.3	13
		Org	26.5	24.0	24.0	27.0	23.0		24.9	8
	100-1000	Sil			25.0				25.0	1
		Cal	28.0			29.0			28.7	3
		Org			23.0				23.0	1
	All									110

(b) ASPT

WFD Type A group			Assessment Score						Mean ASPT	Sites in group
Alt (m)	Area (km^2)		1	2	3	4	5	6		
<200	10-100	Sil	6.80	6.13	*	5.76	5.37	*	5.84	8
		Cal	6.18	5.92	5.97	6.08	5.60	*	5.93	32
		Org	6.72	*	6.36	6.12	*	*	6.39	4
	100-1000	Sil	6.09	*	6.10	5.27	5.75	*	5.70	8
		Cal	*	6.10	5.73	5.87	5.59	*	5.80	22
	>1000	Cal	*	*	5.54	*	*	*	5.54	1
200-800	<10	Org	*	*	*	*	6.70	*	6.70	1
	10-100	Sil	*	6.40	6.48	6.40	5.33	*	6.30	8
		Cal	6.77	6.13	6.44	6.23	6.27	*	6.34	13
		Org	6.12	6.41	6.29	6.26	5.91	*	6.24	8
	100-1000	Sil	*	*	6.52	*	*	*	6.52	1
		Cal	6.89	*	*	5.58	*	*	6.01	3
		Org	*	*	6.65	*	*	*	6.65	1
	All									110

Difference in mean values	% of *n* groups (amongst those with both scores)	
	TAXA	ASPT
Mean Score 1 > Mean Score 2	37% (1.5 of 4)	75% (3 of 4)
Mean Score 1 > Mean Score 3	80% (4 of 5)	60% (3 of 5)
Mean Score 2 > Mean Score 3	90% (4.5 of 5)	40% (2 of 5)
Mean Score 1 > Mean Score 4	43% (3 of 7)	86% (6 of 7)
Mean Score 2 > Mean Score 4	67% (4 of 6)	58% (3.5 of 6)
Mean Score 3 > Mean Score 4	29% (2 of 7)	71% (5 of 7)
Mean Score 3 > Mean Score 5	50% (3 of 6)	100% (6 of 6)

Table 26 – Mean value of (a) observed TAXA and (b) observed ASPT for Scottish Islands (SI) module reference sites with each assessment score (1-6) in each WFD System A group and consistency across groups of differences in means in relation to score.

(a) TAXA

WFD Type A group			Assessment Score						Mean TAXA	Sites in group
Alt (m)	Area (km^2)		1	2	3	4	5	6		
<200	<10	Org	28.0	15.0					19.3	3
	10-100	Sil	19.5	23.9					23.1	11
		Cal	20.0	24.0	24.0				23.2	10
		Org	20.6	17.9	19.8				19.3	19
200-800	10-100	Sil	14.0	19.7					17.4	5
		Cal	20.3	19.8					20.0	7
	All									55

(b) ASPT

WFD Type A group			Assessment Score						Mean ASPT	Sites in group
Alt (m)	Area (km^2)		1	2	3	4	5	6		
<200	<10	Org	6.04	5.98					6.00	3
	10-100	Sil	6.78	6.48					6.53	11
		Cal	6.09	6.36	6.29				6.30	10
		Org	6.18	5.49	6.09				5.87	19
200-800	10-100	Sil	6.25	6.73					6.54	5
		Cal	6.50	6.54					6.52	7
	All									55

Difference in mean values	% of *n* groups (amongst those with both scores)	
	TAXA	ASPT
Mean Score 1 > Mean Score 2	50% (3 of 6)	50% (3 of 6)
Mean Score 1 > Mean Score 3	50% (1 of 2)	50% (1 of 2)
Mean Score 2 > Mean Score 3	25% (0.5 of 2)	50% (1 of 2)

For TAXA, model M7 was not statistically significant for any of the four RIVPACS modules, and model M8 was only significant for the fits based on GB sites and thus for all UK reference sites combined ($p<0.001$) (Table 28) For TAXA, models M7 and M8 were both statistically significant for GB and NI module sites and thus when based on all UK reference sites ($p<0.001$). Non-linear effects model M8 was always a better fit to the data than log-linear model M7 (Table 28).

The estimates of the adjustment factors (A_j) obtained from fitting model (M8) to the UK-wide reference sites based on the spring and autumn combined samples and based on the data from all possible single and combined season sample options are given in Table 29, with the equivalent estimates based on model (M4) for comparison.

Table 27 – Mean value of observed TAXA for Scottish Highlands (SH) module reference sites with each assessment score (1-6) in each WFD System A site group and consistency across groups of differences in means in relation to score.

(a) TAXA

WFD Type A group Alt (m)	Area (km²)		Assessment Score 1	2	3	4	5	6	Mean TAXA	Sites in group
<200	<10	Sil				19.0			19.0	1
	10-100	Sil	21.6	15.0					20.8	8
		Cal	24.3	25.0					24.5	4
		Org	22.7						22.7	3
	100-1000	Org	26.0	26.3					26.3	7
200-800	<10	Sil	17.5						17.5	2
	10-100	Sil	20.4	20.8					20.5	42
		Org	21.0	18.0					20.0	3
	100-1000	Sil	21.7	24.1	28.5				23.2	31
		Org		25.0					25.0	1
	>1000	Sil		30.3	32.0				30.8	6
	All									108

(b) ASPT

WFD Type A group Alt (m)	Area (km²)		Assessment Score 1	2	3	4	5	6	Mean ASPT	Sites in group
<200	<10	Sil				6.00			6.00	1
	10-100	Sil	6.48	6.47					6.48	8
		Cal	6.07	6.92					6.28	4
		Org	6.60						6.60	3
	100-1000	Org	6.69	6.72					6.71	7
200-800	<10	Sil	6.63						6.63	2
	10-100	Sil	6.68	6.77					6.70	42
		Org	6.85	6.50					6.73	3
	100-1000	Sil	6.61	6.77	6.83				6.70	31
		Org		6.20					6.20	1
	>1000	Sil		6.50	6.37				6.45	6
	All									108

Difference in mean values	% of *n* groups (amongst those with both scores) TAXA	ASPT
Mean Score 1 > Mean Score 2	33% (2 of 6)	33% (2 of 6)
Mean Score 1 > Mean Score 3	0% (0 of 1)	0% (0 of 1)
Mean Score 2 > Mean Score 3	0% (0 of 2)	50% (1 of 2)

Table 28 – Statistical estimates (± standard error) of average within WFD A site group linear adjustment factor b_7 in multiplicative linear model (M7) and effect of each assessment score (1-6) in multiplicative non-linear model (M8) for (a) observed TAXA and (b) observed ASPT for the spring and autumn combined samples for the reference sites in each module; *p* = significance probability in test for effect of score

(a) TAXA Model (M7)	GB	NI	SI	SH	All ref sites
$p(b_7)$	0.966	0.122	0.958	0.170	0.897
b_7	0.000 ± 0.004	-0.007 ± 0.004	-0.001 ± 0.026	0.028 ± 0.020	-0.000 ± 0.003
Model (M8)					
p	<0.001	0.453	0.947	0.185	<0.001
A_1	-0.058 ± 0.013	0.025 ± 0.020	-0.007 ± 0.059	-0.081 ± 0.058	-0.039 ± 0.010
A_2	-0.006 ± 0.008	0.026 ± 0.019	-0.015 ± 0.057	-0.060 ± 0.056	-0.006 ± 0.008
A_3	0.00	0.00	0.00	0.00	0.00
A_4	-0.033 ± 0.011	0.005 ± 0.014	---		-0.028 ± 0.010
A_5	-0.063 ± 0.021	-0.003 ± 0.017	---	---	-0.047 ± 0.016
A_6	-0.029 ± 0.028	---	---	---	-0.027 ± 0.028

(b) ASPT Model (M7)	GB	NI	SI	SH	All ref sites
$p(b_7)$	<0.001	<0.001	0.555	0.112	<0.001
b_7	-0.014 ± 0.002	-0.011 ± 0.003	-0.004 ± 0.006	0.007 ± 0.004	-0.013 ± 0.001
model (M8)					
p	<0.001	0.002	0.076	0.650	<0.001
A_1	0.017 ± 0.005	0.021 ± 0.016	-0.007 ± 0.013	-0.008 ± 0.012	0.015 ± 0.004
A_2	0.016 ± 0.004	0.007 ± 0.011	-0.022 ± 0.013	0.001 ± 0.012	0.014 ± 0.003
A_3	0.000	0.000	0.000	0.000	0.000
A_4	-0.011 ± 0.005	-0.007 ± 0.008	---		-0.011 ±0.004
A_5	-0.051 ± 0.009	-0.028 ± 0.010	---	---	-0.041 ± 0.006
A_6	-0.053 ± 0.012	---	---	---	-0.053 ± 0.011

Table 29 – Estimates of multiplicative adjustment factors (1/A_j) for RIVPACS expected TAXA and ASPT based on UK-wide versions of non-linear models M4 (TWINSPAN site groups) and M8 (WFD System A site groups) estimated from spring and autumn combined sample data (Spr/Aut) and from the samples obtained from all seven possible combinations of one or more seasons.

Assessment score	TAXA				ASPT			
	M4 Spr/Aut	M4 All	M8 Spr/Aut	M8 All	M4 Spr/Aut	M4 All	M8 Spr/Aut	M8 All
1	1.000	1.000	1.093	1.107	1.000	0.991	0.966	0.962
2	0.989	0.993	1.014	1.021	0.989	0.986	0.967	0.966
3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
4	1.042	1.038	1.068	1.067	1.019	1.009	1.025	1.017
5	1.102	1.086	1.115	1.105	1.054	1.054	1.100	1.104

Model M8 leads to estimates of adjustment factors for RIVPACS expected values of TAXA (spring and autumn combined samples) which equate to 6.7% and 11.5% increases in expected TAXA for predictions based on reference sites with assessment scores of 4 and 5 respectively. The equivalent adjustment factors for model M4 for TAXA were slightly lower at 4.2% and 10.2% respectively (Table 29).

However, under model M8 references sites with assessment scores of 1 appear, overall, to have lower observed TAXA than reference sites in the same WFD System A group with assessments scores of 3, and thus the adjustment factor for predictions based on reference sites with assessments scores of 1 is to increase expected TAXA by 9.3% (the corresponding adjustment for assessment scores of 2 is only 1.4%). This equivalent to implying that taxonomic richness within a site type is greatest for sites on high/good boundary status (i.e. assessments score of 3). This is quite different from the results of model M4 from which no (or very little) adjustment to expected TAXA (or ASPT) is recommended for the part of predictions based on reference sites with assessment scores of 1 or 2 (Table 29).

The differences between models M8 and M4 in adjustment factors for expected ASPT were greater. For assessment scores of 4 or 5, model M8 led to adjustment factors increases of expected ASPT of 2.5% and 10.0% respectively, compared to corresponding increases of only 0.9% and 5.4% under model M4. Moreover, model M8 led to adjustment factors decreases of expected ASPT for assessment scores of 1 and 2 of 3.8% and 3.4% respectively, compared to corresponding minor decreases of only 0.9% and 1.4% under model M4.

Similar results were obtained when model M8 was fitted to the data for all seven possible single and multiple season sample values combined (Table 29).

4.3 Summary of models and recommendations

In summary, adjustment factors derived from model M8 lead to greater percentage reductions than model M4 in RIVPACS values of expected TAXA and expected ASPT for that part of the prediction based on reference sites with assessment scores of 4 and 5. For expected TAXA, model M4 leads to almost no adjustment for scores of 1 or 2, whereas adjustment factors from model M8 give rise to nearly 10% increases in expected TAXA for the part based on assessment scores of 1. This is completely at odds with John Murray-Bligh’s method which would decrease expected TAXA for a test site by 18% if based on assessment scores of 1 (assuming a high/good boundary of EQI_{TAXA} of 0.78).

For ASPT, adjustment factors derived from model M8 are still non-linear, with greater percentage decreases for assessment scores of 5 than increases for assessment scores of 1. However, the adjustment factors for model M8 are generally greater and more similar to those proposed by John Murray-Bligh (Section 2.2.3) than those of model M4; model M4 led to recommendations of almost no adjustment for assessment scores of 1 or 2.

After presentation and discussion of all the above modelling approaches (M1-M8) and derived estimates of adjustment factors at the WFD72b final project meeting (21 September 2006), it was collectively agreed that the final recommended approach and adjustment estimates should be based on statistical model M4 which estimates average non-linear effects of assessment scores with TWINSPAN groups. Estimates provided with the Excel adjustment calculator are based on spring and autumn combined samples, but can be used to adjust RIVPACS expected values of TAXA and ASPT based on any single or combined season sample.

5. PROCEDURES FOR ADJUSTING EXPECTED VALUES

5.1 Statistical logic

The recommended procedures for adjusting RIVPACS expected values are based on fitting regression model M4:

$$log_{10} O_{ijk} = log_{10} M_i + a_j + e_{ijk} \qquad \text{(M4)}$$

where

O_{ijk} = Observed index value for the k^{th} site with assessment score j in group i
M_i = term for average index value for TWINSPAN group i
a_j = effect of assessment score j on $\log_{10}$ index values (re-scaled to give $a_3 = 0$).
e_{ijk} = residual value for the k^{th} site with assessment score j in group i

The fitted regression model can be re-expressed as:

$$O_{ij} = M_i A_j$$

where $A_j = 10^{a_j}$ = proportional effect of assessment score j on index values (relative to index values for reference sites with a score or 3)

Note that the parameters a_j can be positive or negative depending on whether an assessment score of j, on average, increases or decreases the index values. If $a_j < 0$ then A_j <1 and if $a_j > 0$ then A_j >1. By definition $A_3 = 1.0$. Any assessment scores of 6 are treated as scores of 5 for the purposes of adjustment.

In this model, the effect of a unit change in assessment score is not assumed to be constant across the range of assessment scores (1-5). However, a given assessment score is assumed to a constant multiplicative effect on the observed index values for the reference sites. Specifically an assessment score of j is assumed, on average, to increase the observed index values of reference sites by a factor of A_j (i.e. increase index values by $100(A_j-1)$ percent). Therefore the expected index values of any sites based on reference sites with a score of j are on average over-estimated by a factor A_j. In such cases, the correction should therefore be to divide the RIVPACS expected values by the same factor A_j.

RIVPACS predictions for real test sites are usually based on more than one site group each with reference sites with more than one assessment score.

The general adjustment factor is derived by the following logic:

If E^i_{face} = Mean index value of the reference sites in TWINSPAN site group i

E^i_{true} = True Mean index value of reference sites of high/good boundary quality (score=3) in TWINSPAN site group i

Q_{ij} = Proportion of reference sites in group i with assessment score j

P_{ij} = RIVPACS probability test site belongs to TWINSPAN site group i

then $$E^i_{face} = E^i_{true}\left(\sum_{j=1}^{5} Q_{ij} A_j\right)$$

The unadjusted or 'face' RIVPACS expected value (E_{face}) of a biotic index for a test site is, on average, given by:

$$E_{face} = \sum_{i=1}^{g} P_i E_{face}^{i} = \sum_{i=1}^{g} P_i E_{true}^{i} (\sum_{j=1}^{5} Q_{ij} A_j)$$

As the prediction of the expected value for a test site is intended to be based on environmental similar reference site groups, the true expected values (E_{true}^{i}) of the actively involved groups can be approximated by the true (unknown) expected index value E_{true} for the test site, to give:

$$E_{face} = \sum_{i=1}^{g} P_i E_{true} (\sum_{j=1}^{5} Q_{ij} A_j) = E_{true} \sum_{j=1}^{5} (A_j \sum_{i=1}^{g} (P_i Q_{ij}))$$

This can be written as :

$$E_{face} = E_{true} \sum_{j=1}^{5} (A_j R_j) \qquad \text{(equation A1)}$$

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.*

From equation A1, the adjusted expected value (E_{adj}) for the test sites is estimated by:

$$E_{adj} = E_{face} F_{adj} \qquad \text{(equation A2)}$$

where $F_{adj} = 1 / \sum_{j=1}^{5} (A_j R_j)$ and $R_j = \sum_{i=1}^{g} (P_i Q_{ij})$

The multiplicative adjustment factor can be less than or greater than one depending the weighted assessment quality of the reference sites involved in the prediction and the effect (A_j) of each assessment score (1-5).

5.2 Detailed worked example

The example test site has an observed number of taxa of 15, a RIVPACS prediction 'face' expected number of taxa (E_{face}) of 20, and thus a face EQI value of 15/20 = 0.75.

Suppose there are just three TWINSPAN groups in the RIVPACS module appropriate for this test site. The test site has RIVPACS probabilities of 0.55, 0.34 and 0.11 of belonging to groups 1, 2, and 3. The number and proportion (Q_{ij}) of reference sites in each group *i* with assessment score *j* is as follows:

TWINSPAN site group	Ref sites in group	Proportion (Q_{ij}) of ref sites with assessment score *j* (Number of sites in brackets)					Mean score S
		1	2	3	4	5	
1	20	0.20 (4)	0.10 (2)	0.15 (3)	0.05 (1)	0.50 (10)	3.55
2	12	0.08 (1)	0.00 (0)	0.17 (2)	0.33 (4)	0.42 (5)	4.01
3	25	0.12 (3)	0.24 (6)	0.28 (7)	0.24 (6)	0.12 (3)	3.00

The mean assessment score (S) of reference sites in each group is given in the right-hand column. For example, for group 1,

S = 0.20 x 1 + 0.10 x 2 + 0.15 x 3 + 0.05 x 4 + 0.50 x 5 = 3.55

The estimates of the a_j parameters in model M4 fitted to the data based on all season combination options are given in the right-hand column of Table 16 and repeated in Table 30 below, together with the derived values of $A_j = 10^{a_j}$ for each assessment score (1-5):

Table 30 – Recommended estimates of multiplicative factors (A_j) to be used in equation A2 to adjust RIVPACS predictions of expected values of TAXA and ASPT for all UK RIVPACS sampling sites.

		Assessment score				
		1	2	3	4	5
a_j	TAXA	0.000	0.003	0.000	-0.016	-0.036
	ASPT	0.004	0.006	0.000	-0.004	-0.023
A_j	TAXA	1.000	1.007	1.000	0.964	0.920
	ASPT	1.009	1.014	1.000	0.991	0.948

From equation A2,

$R_1 = (P_1 \times Q_{11}) + (P_2 \times Q_{21}) + (P_3 \times Q_{31})$

$= (0.55 \times 0.20) + (0.34 \times 0.08) + (0.11 \times 0.12) = 0.1504$

Similarly $R_2 = (0.55 \times 0.10) + (0.34 \times 0.00) + (0.11 \times 0.24) = 0.0814$

$R_3 = (0.55 \times 0.15) + (0.34 \times 0.17) + (0.11 \times 0.28) = 0.1711$

$R_4 = (0.55 \times 0.05) + (0.34 \times 0.33) + (0.11 \times 0.24) = 0.1661$

$R_5 = (0.55 \times 0.50) + (0.34 \times 0.42) + (0.11 \times 0.12) = 0.4310$

The adjustment factor (F_{adj}) for this test site is therefore:

$F_{adj} = 1 / (A_1 \times R_1 + A_2 \times R_2 + A_3 \times R_3 + A_4 \times R_4 + A_5 \times R_5)$

$= 1 / (1.000 \times 0.1504 + 1.007 \times 0.0814 + 1.000 \times 0.1711 + 0.964 \times 0.1661 + 0.920 \times 0.4310)$

$= 1 / 0.96027 = 1.041$

The adjusted expected number of taxa is therefore

$E_{adj} = E_{face}$ x 1.041 = 20 x 1.041 = 20.82

For this test site, a greater weight of reference sites involved in its prediction were assessed as being of worse (rather than better) than high/good boundary quality, and therefore the RIVPACS expected number of taxa is likely to be an under-estimate of the that target (i.e. high good boundary) for such sites.

The correction factor is 1.041, indicating that the adjustment is to increase the expected number of taxa by 4.1%, which seems reasonable.

The statistical analysis of the reference site data has shown that the effect of site quality (i.e. assessment score) on the TAXA and ASPT index values is non-linear. This is why it is not valid in the recommended approach to simply base the adjustment on the average assessment score (S) of reference sites in each TWINSPAN group as used in the previous methods of Robin Guthrie and John Murray-Bligh (see Section 2).

For example, consider two cases where the value of S is 3, which would both lead to no adjustment under a model assuming linearity of assessment score effects. In the first case, if all reference sites involved in a test site prediction have a score of 3, then no adjustment to test site expected values is needed. However, in the second case, if half of the reference sites involved have a score of 1 and half a score of 5, also giving an average assessments score of 3, those sites with score of 1 will have little or no impact on average index values while those with a score of 5 will significantly reduce index values, such that the RIVPACS prediction, being based on the (weighted) average of the reference sites' observed index values, will be an under-estimate. In this specific case, the recommended adjustment for expected TAXA would be to multiply the RIVPACS expected value by a factor of:

F_{adj} = 1 / (1.000 x 0.50 + 0.920 x 0.50) = 1 / 0.960 = 1.042.

5.3 Distribution of adjustment factor for UK sites

The range and distribution of values for the adjustment factors (F_{adj}) based on recommended model M4 and equation A2 for the combined UK reference sites and 1995 GQA sites are given in Table 31 and Figure 11.

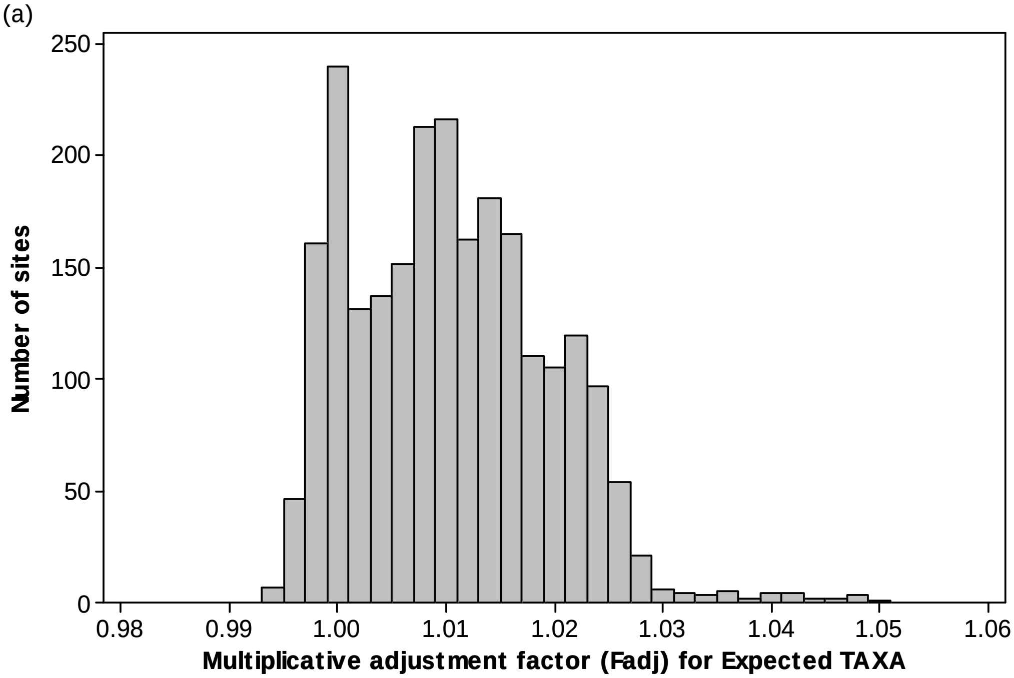
Table 31 – Distribution of the multiplicative adjustment factors (F_{adj} in equation A2) for (a) Expected TAXA and (b) Expected ASPT amongst the combined UK reference sites and 1995 GQA sites.

	Min	Lower 5%	Lower 25%	Median	Upper 25%	Upper 5%	Max
TAXA	0.994	0.998	1.002	1.009	1.016	1.024	1.049
ASPT	0.987	0.990	0.995	1.001	1.005	1.012	1.028

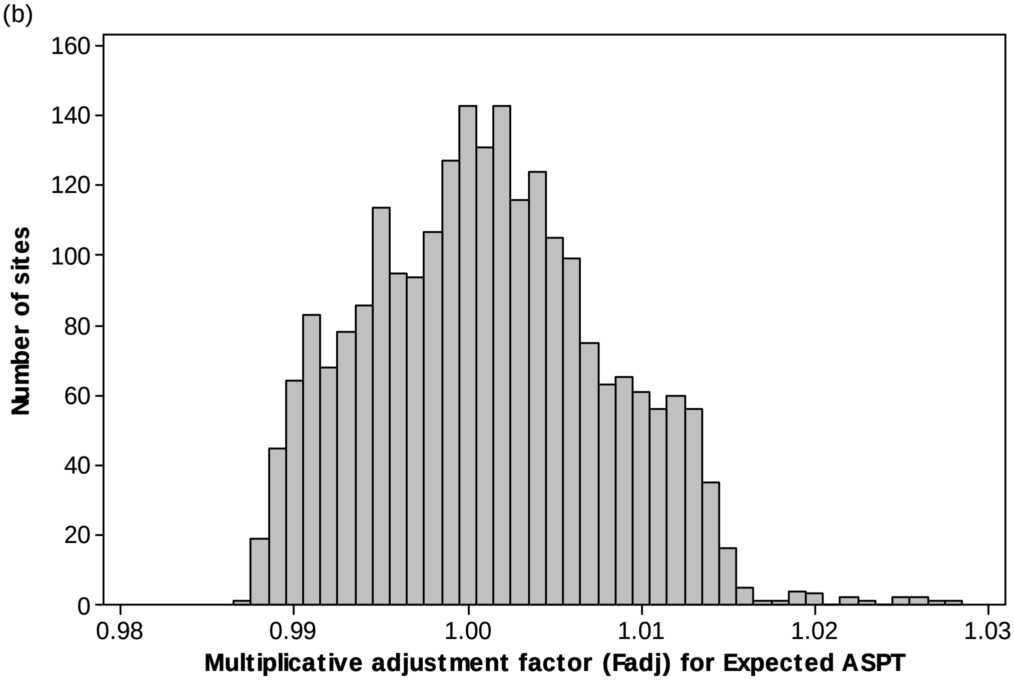
There are both increases and decreases to expected values, but most adjustments are within ± 1% for ASPT and ± 2% for TAXA. The distribution of upward and downward adjustments does vary with RIVPACS module; although not shown here, the adjustments for all four UK modules individual reference sites are supplied as example data within the EXCEL spreadsheet adjustment calculator (see Section 5).

Figure 11 – Frequency distribution of the multiplicative adjustment factors (F_{adj} in equation A2) for (a) Expected TAXA and (b) Expected ASPT for spring and autumn combined season samples amongst the combined UK reference sites and 1995 GQA sites.

(a)



(b)



6. EXCEL SPREADSHEET ADJUSTMENT CALCULATOR

As part of this SNIFFER project WFD72b,.an EXCEL spreadsheet adjustment “calculator” with encoded formulae to automate this procedure for adjusting RIVPACS expected values of any UK RIVPACS test sites has been produced and is available as an project deliverable and output. Detailed instructions for use are given in Section 6.2.

6.1 Inter-compatibility with John Murray-Bligh’s procedure

It can be shown that using Murray-Bligh’s method of adjusting expected values, as described in Section 2.2.3, is mathematically equivalent to using the recommended procedure proposed here based on equation A2, but using the values of A_1, A_2, A_3, A_4 and A_5 given in Table 32.

Table 32 – Values of the multiplicative factors (A_j) to be used in equation A2 which would give the same adjustments as using John Murray-Bligh’s method (Section 2.2.3) of adjusting RIVPACS predictions of expected values of TAXA and ASPT (assuming good/moderate boundary of EQI_{TAXA} of 0.78 (John Murray-Bligh *pers.comm.*).

		Assessment score				
		1	2	3	4	5
A_j	TAXA	1.220	1.110	1.000	0.890	0.780
	ASPT	1.100	1.050	1.000	0.950	0.900

Therefore John’s Murray-Bligh’s method could also be implemented using the same EXCEL spreadsheet calculator.

<u>*Mathematical proof*</u>

In Section 2.2.3, John Murray-Bligh’s method of adjusting the ‘face’ E values (E_{face}) to give an adjusted E value (E_{adj}) was shown to be equal to:

$E_{adj} = E_{face} / (1+ 0.5(1-EQI_{GM}) (3-S))$ (equation J1)

where

S_i = Average Assessment score for reference sites in TWINSPAN site group *i*

$S = \sum_i (P_i . S_i)$ = weighted average assessment score for reference sites in the prediction of the expected values for the test site

This can be re-written as:

$E_{adj} = E_{face} / (1+ k(3-S))$ where $k = 0.5(1-EQI_{GM})$ (equation J1a)

With John’s method, the A_i in equation (A2) are given by:

$A_i = 1 + k(3-i)$ where $k = 0.5(1-EQI_{GM})$ (equation J1b)

Substituting equation (J1b) into the proposed adjustment method equation (A2) gives:

$$E_{adj} = E_{face}F_{adj} \quad \text{where} \quad F_{adj} = 1/\sum_{j=1}^{5}((1+k(3-i))R_j) \quad \text{and} \quad R_j = \sum_{i=1}^{g}(P_iQ_{ij})$$

Re-arranging gives:

$$F_{adj} = 1/\sum_{i=1}^{g} P_i (\sum_{j=1}^{5} (Q_{ij}(1 + k(3 - i)))) = 1/\sum_{i=1}^{g} P_i (1 + k(3 - \sum_{j=1}^{5} (Q_{ij} i)))$$

$$= 1/\sum_{i=1}^{g} P_i (1 + k(3 - S_i) = 1/(1 + k(3 - \sum_{i=1}^{g} P_i S_i)) = 1/(1 + k(3 - S)$$

exactly as in equation (J1a) above for John's method

6.2 Instructions for using the EXCEL spreadsheet adjustment calculator

The EXCEL workbook to carry out the adjustments is called:

"WFD72B RIVPACS prediction adjustment calculator v1.xls"

The EXCEL calculator has been written (with agreement of the project board) so that there is just one workbook for all four UK RIVPACS modules.

At present the workbook is set up to adjust the expected values for the two current GQA indices of number of BMWP taxa (TAXA) and ASPT. However, by suitable copying of existing columns the cell formulae can easily be extended to other indices, such as LIFE and AWIC, or other ICMi metrics.

The workbook contains two spreadsheets named "User data" and "Internal data"

6.2.1 "Internal data" spreadsheet

This "Internal data" spreadsheet should be left unchanged if you only want to run the calculator on a set of samples/sites using the existing adjustment factors.

Cell block A1:BJ8 (highlighted in green) contains the number of reference sites with each assessment score (1-5) in each TWINSPAN group of each module in turn (GB, NI, SI, and SH).

Cell block A10:BJ14 (also highlighted in green) contains formulae to convert these numbers into proportion of sites with each score for each group, which should not be deleted or altered as these are used in the calculation of the adjustments.

Cell block A17:G19 (highlighted in orange) holds the values of the adjustment factors, A_1, A_2, A_3, A_4 and A_5, for each assessment score 1-5, which are to be used to adjust the RIVPACS expected values.

As supplied, the adjustment factors (A_j) are for TAXA and ASPT, and as estimated from UK-wide RIVPACS reference sites and given in Table 30. The adjustment results that would be produced by John Murray-Bligh's method can be obtain by using the values of the adjustment factors A_1 - A_5, given in Table 32.

6.2.2 "User data" spreadsheet

The "User data" spreadsheet is where the User supplies the required input data for each sample/site (one per row) for which they requires estimates of the adjustments to the RIVPACS expected (E) values of TAXA and ASPT (or other indices if edited).

The spreadsheet begins with four header rows, which are supplied informatively named, but are not essential to the calculator and could be changed to other useful text. if required.

The sample/site data begins in row 5. All of the input and output information for any one sample/site occurs on a single row of the spreadsheet.

The cell formulae to derive the adjustment factors are contained with cells on the row.

As supplied, the spreadsheet has been setup with the following sample/site data:

(i) All 614 GB reference sites spring and autumn combined samples
(ii) All 110 Northern Ireland (NI) reference sites spring and autumn combined samples
(iii) All 55 Scottish Islands (SI) reference sites spring and autumn combined samples
(iv) All 108 Scottish Highlands (SH) reference sites spring and autumn combined samples
(v) 6016 England and Wales 1995 GQA sites spring and autumn combined samples

INPUT DATA

The columns requiring user input data are highlighted in yellow.

Columns A-C (highlighted yellow) contain user-supplied input values of :

A = Sample/Site code B = Site name C = RIVPACS module

Columns D-I (highlighted yellow) contain user-supplied RIVPACS unadjusted values of:

Observed TAXA, Observed ASPT, Expected TAXA, Expected ASPT, O/E_{TAXA}, O/E_{ASPT}

Columns S-CA (highlighted yellow) contain user-supplied RIVPACS probabilities of the test site belonging to each of the RIVPACS TWINSPAN site groups:

These probabilities can be obtained from a tabular fixed format text file called 'Groups.out' which is automatically output to your computer's 'Windows temporary directory' by RIVPACS software program RPBATCH (Clarke *et al*. 2005).

The probabilities of group membership must be pasted into the correct columns for the relevant RIVPACS module (see column headings)

The probabilities for all groups of all other RIVPACS modules must be set to zero (0.000).

Columns CC-CG are intermediate working columns with cell formulae which must be left unaltered.

OUTPUT DATA

Columns K-P gives the adjustment results (highlighted in blue)

K = multiplicative adjustment factors (F_{adj}) for expected TAXA
L = multiplicative adjustment factors (F_{adj}) for expected ASPT
M = adjusted value of expected TAXA
N = adjusted value of expected ASPT
O = adjusted value of O/E for TAXA
P = adjusted value of O/E for ASPT

References

Council of the European Communities, 2000. Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy, Official Journal of the European Communities L327 (43): 1-72.

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

Clarke, R.T., Furse, M.T., Davy-Bowker, J. & Gunn, I.D.M. 2005. RPBATCH: RIVPACS III+ River InVertebrate Prediction And Classification System with error assessments. Release 3.3 (January 2005). User Manual. Produced for the Environment Agency, Scottish Environment Protection Agency (SEPA) and Environment and Heritage Service Northern Ireland. 64pp & 14 Appendices.

Davy-Bowker J., Clarke R.T., Furse M.T., Davies C.E., Corbin T.A., Murphy J.F. & Kneebone N.T. 2005 RIVPACS pressure data analysis. Scotland & Northern Ireland Forum for Environmental Research report. Davy-Bowker,J., Clarke, R.T., Johnson, R.K., Kokes, J., Murphy, J.F. & Zahrádková, S. 2006.. A comparison of the European Water Framework Directive physical typology and RIVPACS-type models as alternative methods of establishing reference conditions for benthic macroinvertebrates. Hydrobiologia, 566, 91-108.

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, National Rivers Authority, Bristol.

APPENDICES

Appendix I Detailed description of each Reference site assessment score (1-6)

Appendix I Detailed description of each Reference site assessment score (1-6)
(development led by Robin Guthrie)

1. Pristine. Virtually no development in the catchment upstream of the sampling point. No, or completely insignificant, amounts of intensive land use such as arable or intensive forestry in the catchment. Typical land use would have included “natural” or “semi-natural” such as montaine, open moorland type non-intensive grazing, unimproved grassland, semi-natural woodlands, heathland, scrub, bog, fen etc. No known abstractions or other hydrological disturbances (other than very minor such as private domestic water supply). Morphologically the river reach would be almost completely unimpacted, without any alterations of any significance. Footbridges, minor fords, minor road bridges are acceptable. A diverse riparian bank vegetation structure compatible with the nature of the river Chemically, there were no artificially elevated parameters such as BOD or nutrients and there was almost certainly no pesticide or herbicide use in the catchment upstream of this point (other than a bit of domestic use potentially). Extremely unlikely to show any effects of acidification as there was sufficient natural buffering capacity. Biologically, the invertebrate fauna would be completely “natural”. The river macrophytes would also be natural. This is about as good as it gets in the UK! Although some of the land uses identified above are not truly natural in the context of the ancient wild wood etc. we’re not trying to go back that far in time. What we are really looking at is a “pre-intensification” type land use. Top of GQA class A.

2. Some of the above pressure parameters would be more elevated compared to the pristine situation. Human development may have been more obvious and more intensive land uses more abundant, but certainly not dominant. For example, semi-improved and improved pasture may have been fairly common in the catchment, and perhaps some limited intensive forestry. Morphological or hydrological alterations such as very occasional bank reinforcement or set back embankments may have been present, but fundamentally the channel diversity and flow variability would be the same as for a natural river, with predominantly a diverse riparian bank vegetation structure compatible with the nature of the river. However, the effects of this increased development and morphological/hydrological alterations would not be apparent in the river biology and the fauna/flora would still be indiscernible from the pristine situation. Mid GQA Class A.

3. Morphological and hydrological alterations were likely, but again these would not have been dominant features. Some level of urbanisation likely. This is approximately the point at which the effects of increasing development pressures are starting to become apparent on the biology, although the biology would still be regarded as being well towards the upper end of the quality spectrum compared to many rivers. For example, tolerant taxa may have been slightly more abundant than expected at pristine. However, none of the usually abundant sensitive taxa would have been absent. In other words, if you would expect lots of stoneflies and sensitive mayflies such as Heptageniidae these would still be common although numbers of some taxa may be slightly depressed. The sorts of things which might have been missing are one or two of the typically lower abundance sensitive taxa – the taxa which tend to boost ASPT values by being represented by just one or two individuals. Equally, one or two taxa typically indicative of enrichment might be present. Boundary of GQA Classes A and B.

4. This represents the point where there were almost certainly some expected sensitive families missing and increased numbers of tolerant taxa. However, the overall appearance was still healthy. For example, if this is the sort of river dominated by stoneflies and mayflies, they would still have been well represented, albeit with a few families missing. More intensive land use was probably common and a fair degree of urbanisation might have been present. There's a good chance the river had some noticeable morphological and hydrological alterations, but it's certainly wasn't a completely dredged or re-sectioned channel and still retained the predominant characteristics of the type of river. Mid GQA Class B.

5. This is the point where you might find, for example, that there are no or very few stoneflies even though they would be expected for a river of this type. The same may apply to high scoring mayflies or cased caddis. In other words, "sensitive" orders, such as Ephemeroptera, Plecoptera or Trichoptera which you would expect for this sort of river were possibly weakly represented, with perhaps only their most tolerant families well represented (e.g. Baetidae, Nemouridae, Hydropsychidae) and perhaps occasional individuals of some of the higher scoring families. For faster flowing rivers it is possible that taxa such as Asellus, leeches and various molluscs were turning up in reasonable numbers in riffle samples. Typically, the number of "sensitive" taxa (roughly those scoring more than 6) may only be about half of the number expected at pristine. If organic pollution pressures were a significant influence, there's a good chance that thin films of sewage fungus organisms would have been present at least on the undersides of stones, heavily shaded areas etc. Typically BOD and ammonia levels would have been quite elevated, quite possibly enough to take the site down into GQA class "C". (B in Scotland). Boundary of GQA Classes B & C.

6. Anything below Class 5.

Appendix 2 MINITAB code to fit statistical model M4 and illustrative MINITAB output

Minitab macro code

```
gmacro
FitM4

note
note Fits Model M4 in SNIFFER WFD72b Final Report:
note

name c1 'SiteCode' c2 'RiverName' c3 'SiteName' c4 'Group' c5 'Score' c6 'ObsIndex' c7 'LogObs'

note c1 holds the Reference site code
note c2 holds the Reference site River name
note c3 holds the Reference site name
note
note c4 hold the Group id for the reference site
note In the supplied example for all UK reference sites
note this column is unique for each TWINSPAN group of each of the four RIVPACS modules
note where : 1 = GB, 2 = Northern Ireland, 3 = Scottish Islands, 4 = Scottish Highlands
note c4 = 100 x (Module id) + TWINSPAN group within module
note
note c5 holds the Ecologist's assessment score (1-6) of the WFD condition of each reference site
note     1 = top of high status
note     2 = middle of high
note     3 = high/good boundary
note     4 = middle of good status
note     5 = good/moderate boundary
note     6 = worse than good/moderate boundary
note
note c6 holds the observed value of the biotic indxe of interest for the reference site
note c7 holds the Logarithm (to base 10) of the observed index value

note
note Columns c11-c16 holds 0/1 indicator variables of the scores 1,2,3,4,5,6 respectively
indicator c5 c11-c16
name c11 'Score1' c12 'Score2' c13 'Score3' c14 'Score4' c15 'Score5' c16 'Score6'

let c7=logten(c6)

table 'Group' 'Score';
means 'ObsIndex';
n 'ObsIndex'.

GLM 'LogObs' = 'Group' 'Score';
  Brief 1.

GLM 'LogObs' = 'Group' 'Score1' 'Score2' 'Score4' 'Score5' 'Score6';
        Covariates 'Score1' 'Score2' 'Score4' 'Score5' 'Score6';
  Brief 2.

endmacro
```

Example dataset

An example dataset is given in MINITAB project file 'FitM4.MPJ'
supplied to SNIFFER with the WFD72b project Final report.

This gives all the required information specified in the macro above to fir model M4 to the observed number of TAXA for all of the UK RIVPACS reference sites (all four modules combined).

Running the macro

To run the macro, first open MINITAB and load the example project file 'FitM4.MPJ'
(or your own equivalently formatted dataset)

At the MINITAB command prompt 'MTB>', type: %FitM4
to run the macro

Macro Output

The macro will generate the following output:

```
MTB > %FitM4
Executing from file: FitM4.MA
Fits Model M4 in SNIFFER WFD72b Final Report:

c1 holds the Reference site code
c2 holds the Reference site River name
c3 holds the Reference site name

c4 hold the Group id for the reference site
In the supplied example for all UK reference sites
this column is unique for each TWINSPAN group of each of the four RIVPACS
modules
where : 1 = GB, 2 = Northern Ireland, 3 = Scottish Islands, 4 = Scottish
Highlands
c4 = 100 x (Module id) + TWINSPAN group within module

c5 holds the Ecologist's assessment score (1-6) of the WFD condition of
each reference site
     1 = top of high status
     2 = middle of high
     3 = high/good boundary
     4 = middle of good status
     5 = good/moderate boundary
     6 = worse than good/moderate boundary

c6 holds the observed value of the biotic indxe of interest for the
reference site
c7 holds the Logarithm (to base 10) of the observed index value

Columns c11-c16 holds 0/1 indicator variables of the scores 1,2,3,4,5,6
respectively
```

The following tabular output is used to form Tables 3-6 and parts of Tables 8-12

Tabulated statistics: Group, Score

Rows: Group Columns: Score

	1	2	3	4	5	6	All
101	19.84	20.20	20.00	18.00	*	*	19.91
	19	10	4	1	0	0	34
102	25.50	40.00	30.67	*	*	*	30.50
	2	1	3	0	0	0	6
103	25.50	25.71	24.00	27.67	*	*	25.45
	4	7	6	3	0	0	20
104	26.67	29.33	30.00	*	*	*	28.00
	6	3	2	0	0	0	11
105	19.00	20.00	*	*	20.00	*	19.58
	5	6	0	0	1	0	12
106	32.00	26.80	28.33	24.40	*	*	26.64
	1	5	3	5	0	0	14
107	27.00	27.13	28.75	23.00	*	*	26.75
	1	8	4	3	0	0	16
108	*	27.38	26.40	23.00	19.00	25.50	26.41
	0	13	5	1	1	2	22
109	*	25.00	33.00	24.50	*	*	29.70
	0	2	6	2	0	0	10
110	27.00	29.50	28.50	26.00	*	*	27.92
	6	4	2	1	0	0	13
111	26.00	25.67	24.75	*	*	*	25.40
	3	3	4	0	0	0	10
112	22.75	23.00	*	*	*	*	22.88
	4	4	0	0	0	0	8
113	21.18	19.00	21.50	*	*	*	20.45
	11	7	2	0	0	0	20
114	31.00	22.78	22.58	21.00	*	*	22.91
	1	18	12	1	0	0	32
115	31.00	32.00	28.25	*	*	*	30.58
	2	6	4	0	0	0	12
116	30.40	29.67	29.06	*	*	*	29.45
	5	9	17	0	0	0	31
117	23.00	25.69	25.27	22.00	*	*	25.04
	1	13	11	3	0	0	28
118	*	38.00	36.50	*	*	*	37.77
	0	11	2	0	0	0	13

119	*	32.38	33.80	31.00	29.00	*	32.44
	0	8	5	2	1	0	16
120	28.00	32.44	32.11	31.00	*	*	32.00
	1	9	9	1	0	0	20
121	*	25.00	27.90	28.50	24.00	*	27.19
	0	3	10	2	1	0	16
122	28.00	29.36	29.25	27.50	29.00	*	28.79
	1	11	16	10	1	0	39
123	*	33.17	30.33	*	*	*	31.47
	0	6	9	0	0	0	15
124	*	30.50	32.73	*	*	*	32.47
	0	2	15	0	0	0	17
125	36.00	36.78	37.30	41.00	*	*	37.19
	1	9	10	1	0	0	21
126	*	37.00	34.43	31.00	*	*	34.50
	0	3	7	2	0	0	12
127	*	32.33	29.25	28.80	28.00	*	29.48
	0	3	16	5	1	0	25
128	*	29.00	34.67	28.50	30.00	*	30.60
	0	2	3	4	1	0	10
129	*	*	23.20	27.50	21.50	*	23.78
	0	0	5	2	2	0	9
130	29.00	26.00	29.08	28.25	24.67	24.00	27.79
	1	3	12	4	3	1	24
131	*	25.33	26.00	27.00	*	32.00	26.60
	0	3	4	2	0	1	10
132	*	36.40	32.75	*	*	33.00	34.60
	0	5	4	0	0	1	10
133	*	32.00	28.95	26.00	27.33	29.00	28.71
	0	1	20	2	3	5	31
134	33.00	30.00	27.71	29.67	27.00	*	28.69
	1	1	7	3	1	0	13
135	*	25.50	34.50	30.17	27.50	*	30.36
	0	2	4	6	2	0	14
201	28.50	25.50	25.00	*	*	*	26.00
	2	2	4	0	0	0	8
202	24.00	*	24.33	24.00	22.50	*	23.71
	1	0	3	1	2	0	7
203	27.50	22.00	23.25	25.00	21.00	*	23.42
	2	3	4	1	2	0	12

204	27.50	*	25.00	23.25	*	*	24.71
	2	0	1	4	0	0	7
205	29.50	28.00	28.40	27.00	26.00	*	27.92
	2	1	5	4	1	0	13
206	*	*	26.67	28.83	25.75	*	27.38
	0	0	3	6	4	0	13
207	30.50	29.50	28.00	27.56	23.00	*	27.65
	2	2	2	9	2	0	17
208	*	32.00	34.00	32.00	31.00	*	31.90
	0	1	2	2	5	0	10
209	*	34.00	31.20	33.00	29.00	*	31.78
	0	2	5	1	1	0	9
210	32.00	30.00	29.00	31.00	*	*	30.00
	1	2	3	1	0	0	7
211	*	*	26.00	30.00	31.50	*	29.29
	0	0	2	3	2	0	7
301	24.67	24.00	24.00	*	*	*	24.22
	3	5	1	0	0	0	9
302	19.00	25.44	26.00	*	*	*	24.91
	1	9	1	0	0	0	11
303	20.80	21.29	23.00	*	*	*	21.23
	5	7	1	0	0	0	13
304	17.88	19.00	18.00	*	*	*	18.23
	8	4	1	0	0	0	13
305	*	15.00	12.00	*	*	*	14.67
	0	8	1	0	0	0	9
401	35.00	30.33	33.33	*	*	*	31.70
	1	6	3	0	0	0	10
402	22.80	27.00	*	*	*	*	23.50
	5	1	0	0	0	0	6
403	29.22	27.67	*	*	*	*	28.60
	9	6	0	0	0	0	15
404	22.33	26.50	*	*	*	*	23.38
	6	2	0	0	0	0	8
405	21.71	19.00	*	*	*	*	20.73
	7	4	0	0	0	0	11
406	22.00	23.00	21.00	*	*	*	22.60
	2	7	1	0	0	0	10
407	15.50	18.75	*	*	*	*	16.43
	10	4	0	0	0	0	14

```
408     17.50  20.25       *       *       *       *  18.29
           10      4       0       0       0       0     14

409     18.00  21.00       *       *       *       *  18.86
            5      2       0       0       0       0      7

410     21.22  21.67       *   19.00       *       *  21.15
            9      3       0       1       0       0     13

All     22.88  26.81   28.87   27.42   26.19   28.50  26.79
          169    286     286      99      37      10    887

Cell Contents:  ObsIndex  :  Mean
                ObsIndex  :  Nonmissing


General Linear Model: LogObs versus Group, Score

Factor  Type   Levels  Values
Group   fixed      61  101, 102, 103, 104, 105, 106, 107, 108, 109, 110,
111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121,
                       122, 123, 124, 125, 126, 127, 128, 129, 130, 131,
132, 133, 134, 135, 201, 202, 203, 204, 205, 206, 207,
                       208, 209, 210, 211, 301, 302, 303, 304, 305, 401,
402, 403, 404, 405, 406, 407, 408, 409, 410
Score   fixed       6  1, 2, 3, 4, 5, 6


Analysis of Variance for LogObs, using Adjusted SS for Tests

Source   DF    Seq SS    Adj SS    Adj MS      F      P
Group    60  6.154723  4.978871  0.082981  21.60  0.000
Score     5  0.070409  0.070409  0.014082   3.67  0.003 <--(Test p value)
Error   821  3.153921  3.153921  0.003842                   in Table 14
Total   886  9.379053

S = 0.0619803   R-Sq = 66.37%   R-Sq(adj) = 63.71%

General Linear Model: LogObs versus Group

Factor  Type   Levels  Values
Group   fixed      61  101, 102, 103, 104, 105, 106, 107, 108, 109, 110,
111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121,
                       122, 123, 124, 125, 126, 127, 128, 129, 130, 131,
132, 133, 134, 135, 201, 202, 203, 204, 205, 206, 207,
                       208, 209, 210, 211, 301, 302, 303, 304, 305, 401,
402, 403, 404, 405, 406, 407, 408, 409, 410


Analysis of Variance for LogObs, using Adjusted SS for Tests

Source   DF    Seq SS    Adj SS    Adj MS      F      P
Group    60  6.154723  4.978871  0.082981  21.60  0.000
Score1    1  0.000235  0.000012  0.000012   0.00  0.955
Score2    1  0.014375  0.003266  0.003266   0.85  0.357
Score4    1  0.007836  0.018848  0.018848   4.91  0.027
Score5    1  0.047960  0.047818  0.047818  12.45  0.000
Score6    1  0.000004  0.000004  0.000004   0.00  0.974
Error   821  3.153921  3.153921  0.003842
Total   886  9.379053
```

```
S = 0.0619803   R-Sq = 66.37%   R-Sq(adj) = 63.71%


Term            Coef    SE Coef       T      P
Constant     1.41589    0.00432  327.51  0.000
Score1     -0.000440   0.007852   -0.06  0.955   The coefficients in red
Score2      0.005392   0.005847    0.92  0.357   are rounded (up) to 3
Score4     -0.017458   0.007882   -2.22  0.027   decimal places, as
Score5      -0.04174    0.01183   -3.53  0.000   reported in Table 14
Score6      -0.00069    0.02105   -0.03  0.974   of WFD72b Final report
```

These coefficients are the terms a_1, a_2, a_4, a_5, a_6 in equation M4

```
Unusual Observations for LogObs

Obs   LogObs      Fit   SE Fit  Residual  St Resid
  6  1.60206  1.48297  0.02565   0.11909      2.11 R
 62  1.56820  1.42978  0.01410   0.13842      2.29 R
 ...
879  1.11394  1.25087  0.01669  -0.13693     -2.29 R
884  1.04139  1.25087  0.01669  -0.20948     -3.51 R

R denotes an observation with a large standardized residual.

MTB >
```