Development of the Acid Water Indicator Community (AWIC) macroinvertebrate family and species level scoring systems
Monitoring Acid Waters - Phase I

R&D Technical Report P2-090/TR1

J Davy-Bowker, M T Furse, J F Murphy, R T Clarke, R Wiggers, H M Vincent

Research Contractor:
CEH Dorset

Publishing Organisation
Environment Agency, Rio House, Waterside Drive, Aztec West, Almondsbury
Bristol BS32 4UD
Tel: 01454 624400 Fax: 01454 624409

ISBN: 1844321320

Dissemination Status
Internal: Released to Regions
External: Public Domain

Statement of Use
This is the final report of the Environment Agency commissioned Monitoring Acid Waters Phase 1 research and development project, undertaken by the Centre for Ecology and Hydrology (formerly the Institute of Freshwater Ecology), a component part of the Natural Environment Research Council. Two freshwater macroinvertebrate scoring systems for use in the biomonitoring of surface water acidification in England and Wales are presented. The scoring systems, AWIC*(fam)* and AWIC*(sp)*, are intended to complement the established BMWP (organic pollution) scoring system. It is anticipated that the AWIC scoring systems will contribute to the achievement of forthcoming statutory obligations to report the acidification status of streams and rivers for the European Union Water Framework Directive. It is also hoped that the AWIC scoring systems will be useful in permissive monitoring programmes and special investigations of acidification.

Keywords
acidification; macroinvertebrates; water quality; AWIC, AWIC*(fam)*, AWIC*(sp)*

Research Contractor
This document was produced under an Environment Agency R&D Project by:

CEH Dorset
Winfrith Technology Centre
Winfrith Newburgh
DORCHESTER
Dorset DT2 8ZD

Tel : 01305 213500 **Fax :** 01305 213600

Environment Agency Project Manager
The Environment Agency's Project Manager for this R&D Project was:
John Steel, Thames Region

Further copies of this report are available from:
Environment Agency R&D Dissemination Centre
WRc, Frankland Road, Swindon, Wilts. SN5 8YF

Tel: 01793 865000 Fax: 01793 514562 E-mail: publications@wrcplc.co.uk

ACKNOWLEDGEMENTS

The authors wish to acknowledge R. Atkins (Environment Agency Contract Development Officer) for his support in the initial stages of the project, J. Steel (Environment Agency) for managing the project, Graham Rutt (Environment Agency, Wales) for his scientific appraisal of the work, Alistair Gordon and Darrell Sturley (Environment Agency) for chemical data, Jeanette Collette, Shelley Howard and Anne Lewis for their comments on the University of Wales System, and Dr. J.A.D. Murray-Bligh (Environment Agency) for his valuable comments. We would also like to acknowledge the assistance of a large number of additional Environment Agency personnel who provided datasets and comments on proposed sampling sites.

We wish to acknowledge the following members of CEH, Dorset for their assistance with field work, sample processing and data entry and collation: J.H. Blackburn, V. Boyden, G. Cooper, R.J.M. Gunn, D. Hardie, N. Kneebone, K.L. Symes, R. Wiggers, Dr. J.M. Winder, and P. Wittenoom.

We would also like to thank Dr. J.F. Wright (CEH, Dorset) for comments on the manuscript and Professor S.J. Ormerod (University of Wales,) for his scientific appraisal of an early version of the AWIC*(fam)* scoring system.

This work was funded by an Environment Agency research and development contract awarded to the Natural Environment Research Council.

EXECUTIVE SUMMARY

The forthcoming EU Water Framework Directive will place new statutory requirements on the Environment Agency to assess and remediate pressures on surface water bodies including the acidification of streams and rivers. Biomonitoring will play a key role because target quality will be set using ecological criteria. It is therefore of great importance that a suitable index of acidification stress is developed in readiness for the implementation of the Directive. This report reviews one of the most commonly used macroinvertebrate acidification metrics (the University of Wales system) and presents two new scoring systems developed by CEH during this R&D contract 'AWIC*(fam)*' and 'AWIC*(sp)*' (the **A**cid **W**aters **I**ndicator **C**ommunity, **fam**ily and **Sp**ecies scoring systems).

In developing the new family level AWIC*(fam)* scoring system, it was important that the scoring system would be applicable across the whole of England and Wales and that the level of invertebrate identification was the same as the BMWP scoring system (developed principally to assess the impact of organic pollution).

Regression analysis of the AWIC*(fam)*-ASPT versus mean pH gave rise to a predictive equation enabling mean pH (within defined confidence limits) to be calculated from the AWIC*(fam)*-ASPT at a test site. These predictions compared favourably with the observed mean pH in a substantial partially independent data set.

A species level scoring system, AWIC*(sp)*, has also been developed. Because of the difficulty in obtaining a large independent species level data set (with pH) for testing, an equation predicting the mean pH from AWIC*(sp)*-ASPT scores has not yet been developed. However, the AWIC*(sp)* looks very promising and a testing exercise will be carried out and included in a paper on both indices to be submitted for publication in a scientific journal later in 2003.

CEH recommend that the AWIC*(fam)*-ASPT scoring system is included in statutory monitoring programmes carried out by the Environment Agency. CEH also recommend that permissive monitoring programmes carried out in areas of England and Wales that are susceptible to acidification, should make use of AWIC*(fam)*-ASPT to assess acidity. In future, both the AWIC*(fam)*-ASPT and AWIC*(sp)*-ASPT should be included in special investigations of acidification so that the effectiveness of both indices can be evaluated as a precursor to possible use in the Water Framework Directive reporting of acidic stress in England and Wales.

In order to make the AWIC scoring systems available to Environment Agency biologists, it is recommended that both scoring systems should be incorporated into the Environment Agency Biology for Windows database. CEH also intend to include the AWIC scoring systems in the next version of the RIVPACS software.

CONTENTS

1. INTRODUCTION

1.1 Objectives

In October 1996 the Centre for Ecology and Hydrology, CEH (then The Institute of Freshwater Ecology, IFE) were commissioned by the Environment Agency to carry out the research and development project:

Monitoring Acid Waters, Phase 1, Indicator Populations.

The objectives were defined in the first progress report of the project (Furse & Symes, 1998) and are reproduced below:

The overall objective was:

- To produce a standard methodology that enables the Environment Agency to assess the extent of ecological damage caused by acidification in controlled surface waters in order that they can make considered comment on short and longer term effects and on the likely effects of changes in land use.

The specific objectives were:

- To produce an algorithm to differentiate biological communities into groups which reflect the effects of acidification on their environment.
- To test the algorithm using field data
- To propose monitoring guidelines for applying the algorithm nationally
- To produce an R&D Technical Report and Project Record in accordance with the Environment Agency's Guidelines for Reporting.
- To use the project output to produce a paper for publication in a relevant scientific journal

CEH were also asked to review comments from the Environment Agency on the effectiveness of the University of Wales system (also known as the Rutt Key), a family level macroinvertebrate indicator system developed by Rutt *et al.* (1990).

1.2 Review of acid water assessment methods

In acidic conditions, freshwater macroinvertebrates exhibit reduced density and species richness (Feldman & Connor, 1992; Townsend *et al*., 1983; Rosemond *et al*., 1992). Of the Ephemeroptera, Plecoptera and Trichoptera ('EPT' taxa), the Ephemeroptera are widely noted for their susceptibility to acidic conditions, with low densities and species richness observed in several countries (Mackay & Kersey, 1985; Raddum & Fjellheim, 1984; Rosemond *et al.*, 1992; Smith *et al*., 1990), while the Plecoptera are noted for their resilience to acidic conditions. Consistent differences in acid tolerance are also evident at family, genus and species levels in most groups of freshwater macroinvertebrates.

The inadequacies of intermittent chemical monitoring programmes (which are liable to miss ecologically important low pH episodes), have led many researchers to attempt to develop operational tools for the assessment of acidity levels in streams and rivers using macroinvertebrates. This has mirrored the well-established development of indices and metrics for assessing the impact of organic pollution, such as the BMWP (**B**iological **M**onitoring **W**orking **P**arty) scoring system (Hawkes, 1997). Drawing principally from the work of European researchers, where the acidification of surface waters has been most intensively studied, the most significant and operationally useful indices of acidification using macroinvertebrates are reviewed below.

1.2.1 The Wade Key

Wade *et al.* (1989) classified samples from 104 upland sites in Wales using TWINSPAN (Hill, 1979) and related their chemistry (and other environmental factors) to the observed species assemblages using correlation and Multiple Discriminant Analysis. Samples from the spring and summer were used to form two separate, 4-group TWINSPAN classifications. These then formed the basis for 2 dichotomous keys (Figure 1) with indicator species appropriate for spring and summer, the spring key having the greater differences between end-groups. These keys were then proposed for the rapid detection and assessment of acidity throughout Wales.

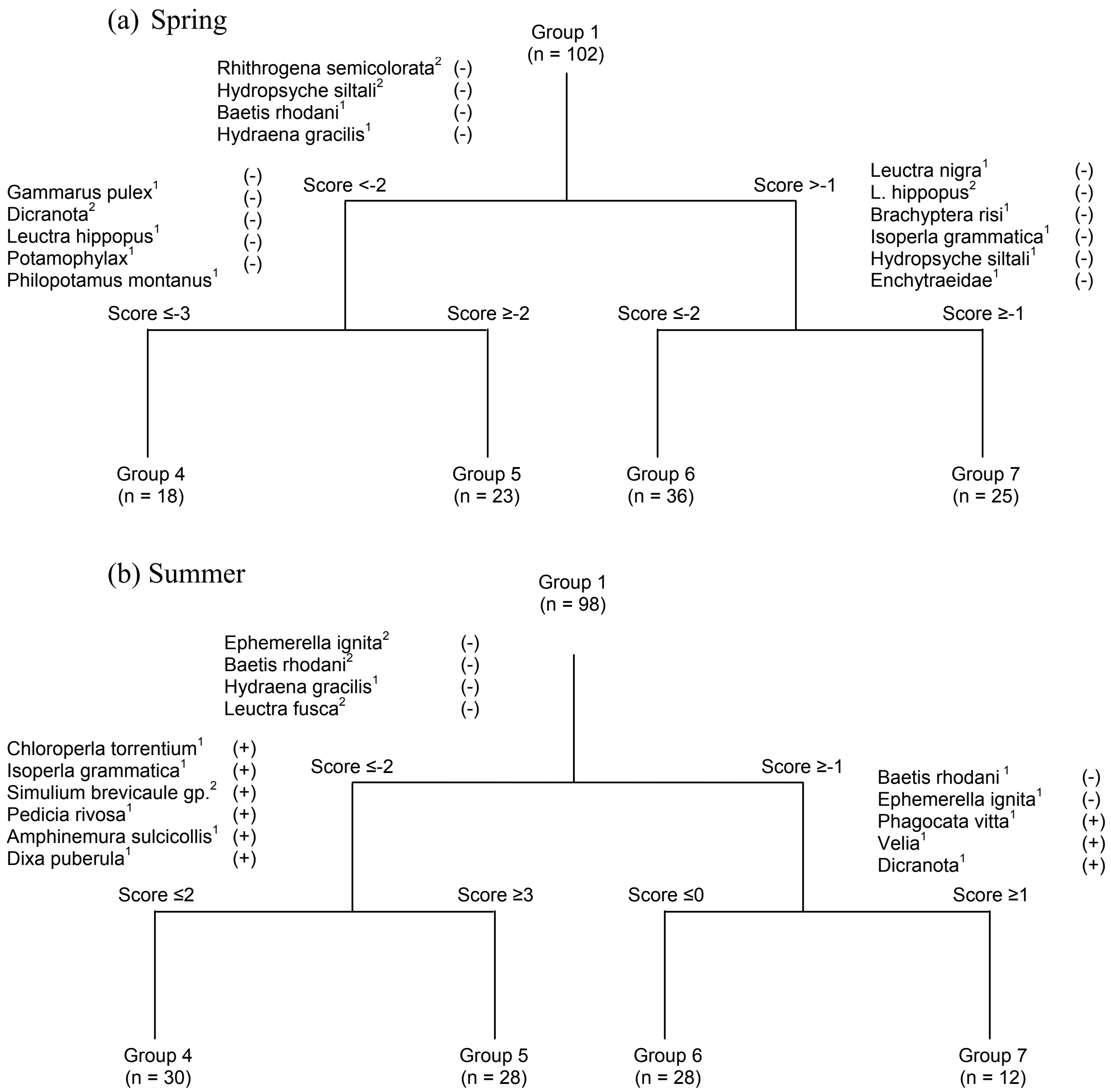


Figure 1. Macroinvertebrate indicator keys proposed by Wade *et al.* (1989) for (a) spring and (b) summer. Abundance categories; 1 = 1-9, 2 = 10-99. Indicator scores; (+) = +1, (-) = -1.

1.2.2 The University of Wales System

Using a similar approach to Wade *et al.* (1989), Rutt *et al.* (1990) used TWINSPAN to classify macroinvertebrate data from 368 upland streams in Wales, Scotland and N.W. England into a 2-group and a 4-group classification of sites that differed markedly in their invertebrate fauna and acidity. The 4-group classification was then used as the basis for a macroinvertebrate family level indicator key (Figure 2) where the key indicator families from the TWINSPAN classification were used to allow new sites to be placed into one of four groups each with distinct pH, aluminium concentrations and buffering capacity.

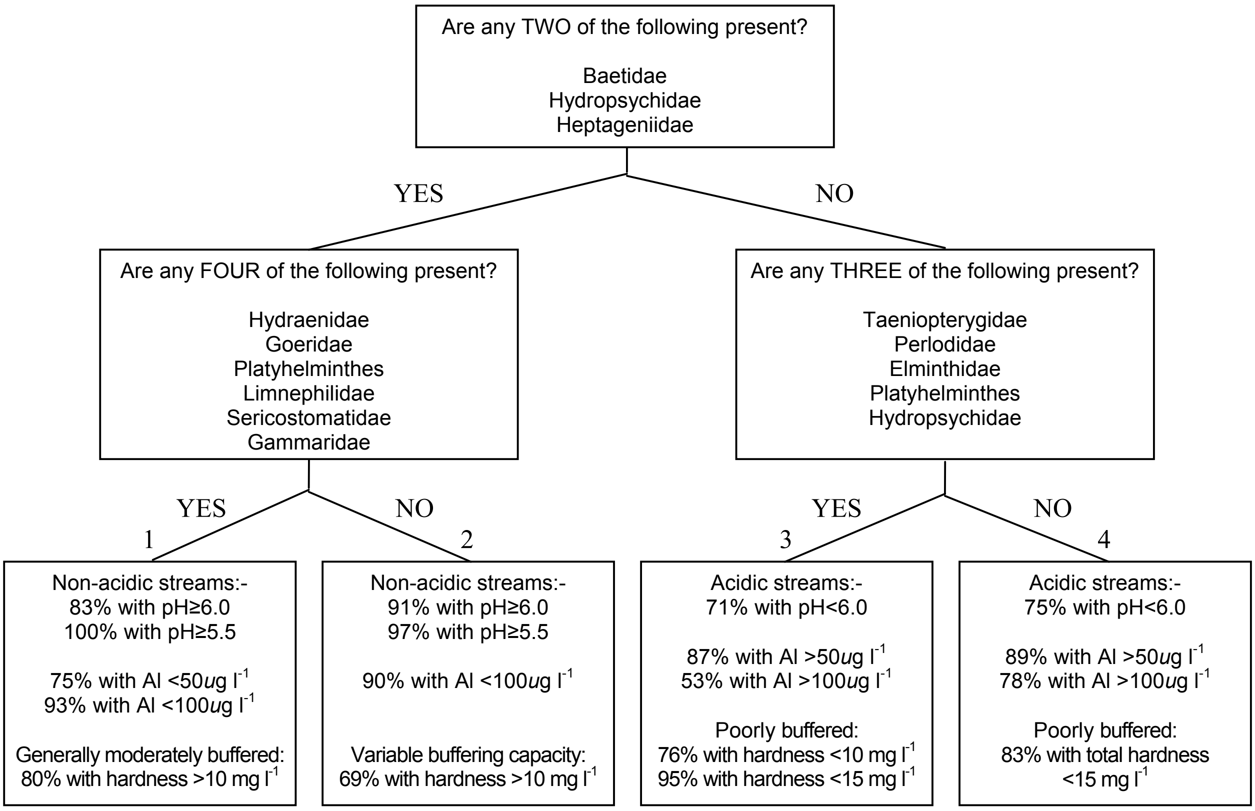


Figure 2. The University of Wales System devised by Rutt *et al.* (1990). Boxes indicate the likely stream chemistry in each of the groups.

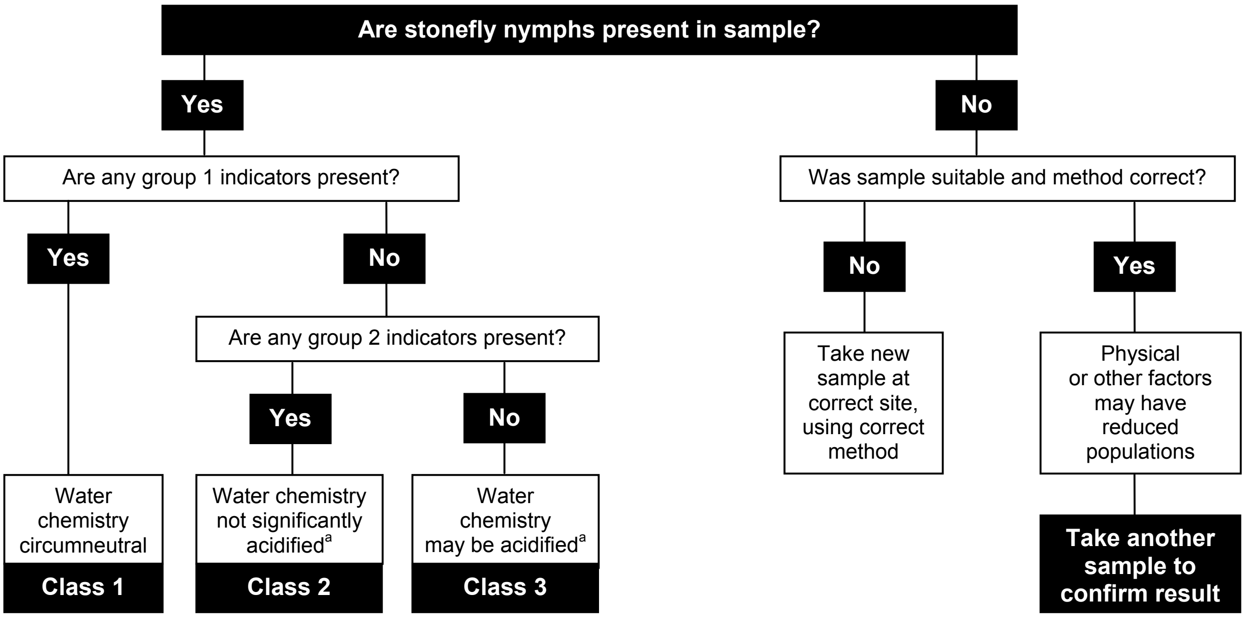
Rutt *et al.* (1990) also used Multiple Discriminant Analysis to show that pH, aluminium and calcium concentrations alone were effective in discriminating between these TWINSPAN groups. Models were then devised to predict the TWINSPAN group from physiochemical data alone. Testing of these models was conducted on a reserved subset of data revealing 75-80% agreement when attempting to predict group membership at the 2-group level, and 53-55% agreement with 4 groups. The distributions of pH and aluminium concentrations at the test sites classified using the indicator key also agreed reasonably well with the distributions of these parameters in the four original TWINSPAN groups. It was concluded that a macroinvertebrate indicator system, derived from TWINSPAN, could be used in surveys of stream acidity and susceptibility to further acidification.

1.2.4 The Patterson and Morrison key

In Scotland, Patterson & Morrison (1993) devised a 3-class system using macroinvertebrates, enabling forest managers to assess the acidity and biological diversity of streams for the purposes of forest management. Lists of acid-tolerant and acid-intolerant macroinvertebrate taxa (Table 1) are used in a decision chart (Figure 3) to allocate sites to one of the 3 classes. The definition and ecological significance of each class is shown in Table 2.

Table 1. Acid-tolerant and intolerant macroinvertebrates (Patterson and Morrison, 1993).

Indicator Group	Animal group/species	Indicator Group	Animal group/species
Acid-intolerant	Gammarus pulex (freshwater shrimp)	Acid-tolerant	Plecoptera (stoneflies)
	Glossosoma and Agapetus species		Trichoptera (caddis flies)
Group 1	Ancylus fluviatilis (freshwater limpet)		Leptophlebiidae (mayflies)
	Lymnaea peregra (wandering snail)		Siphlonuridae (mayflies)
	Asellus aquaticus (freshwater slater)		Dytiscidae (swimming beetles)
Acid-intolerant Group 2	Hydropsyche species (caddis flies)		
	Baetis species (mayflies)		
	Family Heptageniidae (flat bodied mayflies)		



[a]Further samples at the same site and others on the same stream may be considered to increase the confidence of the diagnosis, where the number of Group 2 indicator species present in the first sample is 0 or 1.

Figure 3. Decision chart for assessing the acidity of sites (Patterson and Morrison, 1993).

Table 2. Definition and ecological significance of water chemistry status classes (Patterson & Morrison, 1993).

Class 1	Streams with a water chemistry suitable for the great majority of plants and animals. They are sufficiently alkaline to be buffered against most acid spate waters and pH is unlikely to drop below 5.6 with the mean probably exceeding 6.0. Salmonid fish would not suffer any significant stress from the water chemistry.
Class 2	Streams which are not sufficiently acidified. They could possibly be circumneutral like Class 1, but merely lack group 1 indicators through the chance location of the sample. Their mean pH is unlikely to be much less than 5.6 although where aluminium and heavy metals are low and/or organic content is high the mean could be down to about 5.3. These streams are likely to be suitable for most wildlife species except perhaps the most sensitive taxa. The water chemistry is likely to be suitable for fish populations and aquatic birds including the dipper. These streams could be vulnerable to acidification in future however.
Class 3	Streams which may be acidic to the point where wildlife is significantly affected across a wide range of groups. These effects include the reduction of populations of salmonid fish, especially salmon, and reduced invertebrate animal diversity. The diagnosis of Class 3 cannot be certain because lack of acid-sensitive indicators in samples could be due to sampling error or causes other than water chemistry. Taking further samples would help to eliminate these factors and increase certainty. Chemical analysis of streams identified as Class 3 should be considered to further improve the diagnosis.

1.2.5 The Acidification Number system

In Norway, Raddum & Fjellheim (1984) and Raddum *et al*. (1988) describe the 'Acidification Number' system for assessing stream acidification. The acidification number was determined through a hierarchical system (Table 3) where indicator organisms have been assigned scores based on acidification sensitivity (Table 4).

Table 3. Norwegian Acidification Number system – the four levels of acidification tolerance (Raddum *et al*., 1988).

Category a)	Species tolerating pH >5.5 are given the number	1
Category b)	Species tolerating pH >5.0 are given the number	0.5
Category c)	Species tolerating pH >4.7 are given the number	0.25
Category d)	Species tolerating pH <4.7 are given the number	0

Table 4. Norwegian Acidification Number system - invertebrate indicator taxa with acidification sensitivity scores (Raddum *et al*., 1988).

Category	Species/Group	Score
a	Gastropoda Crustacea: *Gammarus lacustris* *Lepidurus arcticus* Ephemeroptera: *Baetis* spp.	1
b	Cladocera: *Daphnia* spp. Ephemeroptera: *Siphlonurus* spp. *Ameletus inopinatus* Plecoptera: *Isoperla* spp. *Diura* spp. *Capnia* spp. *Leuctra fusca* *Arcynopteryx compacta* *Dinocras cephalotes* Trichoptera: *Apatania* spp. *Hydropsyche* spp. *Philopotamus montanus* *Lepidostoma hirtum* *Ithytrichia lamellaris* *Glossosoma* spp.	0.5
c	Small Mussels (Sphaeriidae)	0.25
d	No registration of above mentioned species or groups	0

A new sample is evaluated by firstly looking for species in category 'a'. If any of these are present, the sample is assigned to category 'a'. If no species in category 'a' are present, the sample is examined secondly against category 'b' and then category 'c'. If none of the species are present, the sample is allocated to category 'd'. This classification scheme has been adopted by several European countries and is included in the "International Co-operative Programme on the Assessment and Monitoring of Acidification of Lakes and Rivers" (NIWR, 1991).

1.2.6 The Expanded Acidification Number system

Fjellheim & Raddum (1990) also presented an expanded version of the Norwegian Acidification Number system of Raddum *et al*., (1988), which we have called the Raddum Index. This used the same four levels of acidification tolerance of Raddum *et al*. (1988) (Table 3), together with an expanded taxon list (Table 5).

Table 5. The Raddum Index - invertebrate indicator taxa with acidification index values (Fjellheim & Raddum, 1990).

Invertebrate Taxa	Index	Invertebrate Taxa	Index
Turbellaria		*Taeniopteryx nebulosa* (L.)	0
Crenobia alpina (Dana)	0.5	*Brachyptera risi* (Mort.)	0
Otomesostoma auditivum (Pless.)	0.5	*Amphinemura standfussi* (Ris)	0
		Amphinemura borealis (Mort.)	0
Bivalvia		*Amphinemura sulcicollis* (Stph.)	0
Anodonta spp.	1	*Nemoura cinerea* (Retz.)	0
Margaritana margaritifera L.	1	*Nemoura avicularis* Mort.	0
Sphaerium spp.	0.5	*Nemurella picteti* Klap.	0
Pisidium spp.	0.25	*Protonemura meyeri* (Pict.)	0
		Capnia atra Mort.	0.5
Gastropoda		*Capnia pygmaea* (Zett.)	0.5
Lymnaea peregra (Muller)	1	*Leuctra fusca* (L.)	0
Planorbis spp.	1	*Leuctra hippopus* Kempny	0
		Leuctra nigra (Oliv.)	0
Hirudinea			
Helobdella stagnalis (L.)	0.5	Trichoptera	
Theromyzon tessulatum (O.F. Muller)	1	*Rhyacophila nubila* (Zett.)	0
Glossiphonia complanata (L.)	1	Glossosoma intermedium Klap.	1
Haemopis sanguisuga (L.)	1	*Ithytrichia lamellaris* Eaton	0.5
		Oxyethira spp.	0
Crustacea		*Philopotamus montanus* (Donovan)	0.5
Lepidurus arcticus Kroyer	1	*Tinodes waeneri* (L.)	0.5
Gammarus lacustris Sars	1	*Cyrnus flavidus* McL.	0
Asellus aquaticus (L.)	0.5	*Cyrnus trimaculatus* (Curtis)	0
Daphnia magna Strauss	0.5	*Holocentropus dubius* (Rambur)	0
Daphnia longispina O.F. Muller	0.5	*Neureclipsis bimaculata* (L.)	0
		Plectrocnemia conspersa (Curtis)	0
Ephemeroptera		*Polycentropus flavomaculatus* (Pict.)	0
Ameletus inopinatus Eaton	0.5	*Polycentropus irroratus* (Curtis)	0
Siphlonurus aestivalis (Eaton)	0.5	*Hydropsyche angustipennis* (Curtis)	0.5
Siphlonurus linnaeanus (Eaton)	0.5	*Hydropsyche pellucidula* (Curtis)	0.5
Baetis rhodani (Pictet)	0.5	*Hydropsyche siltalai* Dohler	0.5
Baetis fuscatus (L.)	1	*Agrypnia obsoleta* Hagen	0
Baetis lapponicus (Bengts.)	1	*Phryganea grandis* L.	0
Baetis macani (Kimmins	1	*Lepidostoma hirtum* (Fabr.)	0.5
Baetis muticus (L.)	1	*Apatania zonella* (Zett.)	0.5
Baetis niger (L.)	1	*Apatania stigmatella* (Zett.)	0.5
Baetis scambus (Eaton)	1	*Chaetopteryx villosa* (Fabr.)	0
Baetis subalpinus Bengts.	1	*Limnophilus centralis* (Fabr.)	0
Baetis vernus Curtis	1	*Limnephilus centralis* Curtis	0
Heptagenia sulphurea (Muller)	0.5	*Limnephilus extricatus* Curtis	0
Heptagenia fuscogrisea (Retz.)	0	*Limnephilus flavicornis* (Fabr.)	0
Leptophlebia vespertina (L.)	0	*Limnephilus lunatus* Curtis	0
Leptophlebia marginata (L.)	0	*Limnephilus rhombicus* (L.)	0
Ephemerella aurivilli (Bengts.)	1	*Limnephilus stigma* Curtis	0
Ephemerella mucronata (Bengts.)	0	*Limnephilus vittatus* (Fabr.)	0
Ephemerella ignita Bengts.	0	*Halesus radiatus* (Curtis)	0
Caenis horaria (L.)	1	*Micropterna lateralis* (Steph.)	0
		Potamophylax cingulatus (Steph.)	0
Plecoptera		*Potamophylax latipennis* (Curtis)	0
Arcynopteryx compacta (McL.)	0.5	*Stenophylax permistus* McL.	0
Diura nanseni (Kempny)	0.5	*Notidobia ciliaris* (L.)	0
Diura bicaudata (L.)	0.5	*Sericostoma personatum* (K & Sp.)	0.5
Isoperla grammatica (Poda)	0.5	*Molanna angustata* Curtis	0
Isoperla obscura (Zett.)	0.5	*Molannodes tinctus* (Zett.)	0
Dinocras cephalotes (Curt.)	0.5	*Adicella reducta* McL.	0
Siphonoperla burmeisteri (Pict.)	0	*Athripsodes aterrimus* (Steph.)	0
		Athripsodes cinereus (Curtis)	0
		Mystacides azurea (L.)	0

The Raddum Index for a sample is taken as the mean of the index values for the taxa in Table 5 which were found in the sample.

1.2.7 Weighted Averaging

In Finland, Hämäläinen & Huttunen (1990) compared two stream acidity assessment methods based on macroinvertebrates:

1) Tolerance limit (TL) method. This used the presence/absence of indicator species where the tolerance limit for each indicator species was defined as the lowest measured pH value among those streams where the species was present. The pH tolerance limits were arranged in four categories: (1) pH <4.5, (2) pH 4.5-4.9, (3) pH 5.0-5.4, (4) pH >5.4. The minimum pH tolerance limits given to taxa in Hämäläinen & Huttunen (1990) are shown in Table 6.

2) Weighted averaging based on species optima and tolerances. These were estimated by both the maximum likelihood (ML) and weighted averaging (WA) methods (see Hämäläinen & Huttunen, 1990 for further details).

Table 6. Minimum pH tolerance limits (TL) assigned by Hämäläinen & Huttunen (1990).

Category	Invertebrate Taxa	TL
1	*Nemurella picteti* Klp.	4.3
	Leuctra digitata (Ol.)	4.3
	Leuctra nigra (Ol.)	4.3
	Taeniopteryx nebulosa (L.)	4.3
	Nemoura cinerea (Retz.)	4.3
	Micropterna spp.	4.3
	Leuctra hippopus Kmp.	4.3
	Nemoura avicularis Morton	4.3
	Plectrocnemia conspersa (Curtis)	4.3
	Leptophlebia marginata (L.)	4.3
	Chaetopteryx spp.	4.3
	Agabus spp.	4.3
	Sialis fuliginosus Pict.	4.3
	Micrasema gelidum McLachlan	4.3
2	*Baetis vernus* coll.	4.5
	Potamophylax spp.	4.5
	Leptophlebia vespertina (L.)	4.5
	Rhyacophila fasciata Hagen	4.5
	Asellus aquaticus L.	4.5
	Rhyacophila nubila (Zett.)	4.7
	Halesus spp.	4.7
	Polycentropus flavomaculatus (Pict.)	4.7
	Diura bicaudata (L.)	4.7
	Elmis aenea Ph.Müll.	4.7
	Pisidium spp.	4.7
	Neureclipsis bimaculata (L.)	4.7
	Sialis lutaria (L.)	4.8
	Oxyethira spp.	4.8
	Agrypnia spp.	4.8
	Heptagenia fuscogrisea (Retz.)	4.8
	Oulimnius tuberculatus (Pict.)	4.9
	Baetis rhodani (Pict.)	4.9
	Paraleptophlebia spp.	4.9
	Limnephilus spp.	4.9
	Molannodes tinctus Zett.	4.9
3	*Leuctra fusca* (L.)	5.0
	Baetis niger (L.)	5.0
	Heptagenia sulphurea (Mull.)	5.0
	Lepidostoma hirtum (Fabr.)	5.0
	Nemoura flexuosa Aubert	5.0
	Hydropsyche pellucidula (Curtis)	5.0
	Ceraclea annulicornis (Steph.)	5.0
	Amphinemura borealis (Morton)	5.0
	Hydropsyche siltalai Döhler	5.0
	Isoperla spp.	5.0
	Protonemura meyeri (Pictet)	5.0
	Hydraena palustris L.	5.0
	Erpobdella octoculata L.	5.0
	Ephemerella ignita (Poda)	5.0
	Molanna angustata Curtis	5.1
	Glossiphonia complanata (L.)	5.1
	Somatochlora metallica (Linden)	5.1
	Helobdella stagnalis (L.)	5.1
	Helodes sp.	5.1
	Lymnea sp.	5.1
	Calopteryx virgo (L.)	5.1
	Hydraena gracilis Germ.	5.1
	Pericoma sp.	5.3
	Hydropsyche angustipennis (Curtis)	5.3
	Gyraulus sp.	5.3
	Bathyomphalus contortus L.	5.3
	Hydraena riparia Kugel.	5.4
	Sericostoma personatum (K. & Sp.)	5.4
4	*Polycentropus irroratus* (Curtis)	5.6
	Cordulegaster boltoni (Donovan)	5.6
	Lype spp.	5.6
	Limnius volckmari Panzer	5.6
	Centroptilum luteolum L.	5.8
	Sphaerium corneum L.	5.8
	Caenis horaria (L.)	6.0
	Silo pallipes (Fabr.)	6.4

In a further comparison of weighted averaging and tolerance limit methods, Hämäläinen & Huttunen (1996) found that weighted averaging performed better than the tolerance limit method in inferring minimum pH. The weighted averaging methods used in Hämäläinen & Huttunen (1996) gave rise to estimates of minimum pH optima and tolerance values that can be used be infer the minimum pH of a stream (if the approximate relative abundances of the taxa in a sample are known). Hämäläinen & Huttunen (1998) also compared inferred stream water acidity determined by weighted averaging and tolerance limit methods for streams in north eastern Finland. The minimum pH of the study sites was derived by weighted averaging calibration using the species optima and tolerances derived for southern and eastern Finland (Hämäläinen & Huttunen, 1996). Again weighted averaging performed better than the tolerance limit methods. Weighted averaging has also been used in Norway (Larsen *et al.*, 1996) to examine the usefulness of macroinvertebrates as predictors of pH.

2. METHODS

2.1 Review of the University of Wales system (the Rutt Key)

In December 1997 a letter was circulated to all regional and most area biologists in the Environment Agency requesting their views, both practical and theoretical, on whether the recently developed University of Wales System (Rutt *et al.*, 1990), worked effectively in their respective regions (Furse & Symes 1998). This system, a user-friendly key for the rapid detection and assessment of acidification, had recently been trialled throughout the Environment Agency. Comments on the system were collated by CEH as part of this research and development project and are reviewed in this report.

2.2 Data collection and database construction

The development of a standard methodology to assess the extent of acidification required the collection of a considerable quantity of site-matched biological and chemical data both at species and family levels. The data requests made to the Environment Agency, the acquisition and review of these data and the decisions leading to the final selection of sites for inclusion in the analyses are detailed in the progress reports Furse & Symes (1998), Furse (1998) and Furse *et al*. (2000). The biological data sets selected for analysis, for use in the construction of a scoring system and in testing the system are given in the sections below.

2.2.1 The IFE Test data set

The 15 samples comprising the IFE Test dataset (see section 2.5) were collected in April & May 1998. The sites were selected using the following criteria (Furse, 1998):

- they encompassed a broad geographic spread
- they included sites across a broad range of acidification
- appropriate chemical data were available for the site for each of the previous five years, including monthly measurements of pH and conductivity, frequent measurements of total hardness and, preferably frequent measurements of aluminium, manganese and iron concentrations
- they were substantially free from other forms of environmental stress
- they were easy to access, preferably including available Agency “site sheets”

Environment Agency staff were consulted to assist in the selection process.

2.2.2 The Welsh Acid Waters Survey

Data collected for the substantial Welsh Acid Waters Surveys of 1984 and 1995 (Stevens *et al*., 1997) was supplied to CEH in 1998. The samples were taken from a wide variety of river systems in North, mid and South Wales and included many headwater sites that are rare in data sets supplied by the Environment Agency.

2.2.3 The 1990 River Quality Survey

IFE selected 100 of the reconstituted 1990 RQS samples (which are still held in store at CEH Dorset and other locations in Dorset) for reanalysis at species level. These samples were selected to cover both a geographical and an acid-alkali range in order to enhance the coverage of the combined data set.

2.2.4 The North West Water Authority 1982 –1986 data set

Eighty-three samples from a North West Water research project entitled 'Acidification of Surface Waters in Cumbria and South Pennines' (Crawshaw *et al*. 1989) were obtained from Graham Rutt (*pers. comm.*). These samples were collected in 1982 and 1986 from sites on catchments with a variety of susceptibilities to acidification.

2.2.5 RIVPACS reference sites

An additional data set of 64 RIVPACS reference sites was chosen to further improve coverage of areas in England and Wales thought to be at potential risk from acidification.

2.2.6 The 1995 General Quality Assessment

Family level data from the 1995 General Quality Assessment (GQA) survey undertaken by National Rivers Authority (approximately 6000 sites) had already been collated by CEH for a previous research and development project (Davy-Bowker *et al*., 2000). This was used in the testing of the final scoring system by correlation with the mean pH data available for each site (which had also been collated and matched to the biological sites in the same R&D project).

2.2.7 Chemical data

In 1998, after the list of sites with biological data had been decided, requests were made to the Environment Agency for chemical data from the same sites (Furse *et al*., 2000). These data were converted to a standardised format in spreadsheets and then transferred to the same Microsoft® Access database as the species data (the CEH, National Invertebrate Database). The combined dataset could then be queried and formatted to create files suitable for analysis.

Within the full data set, while a high proportion of the biological samples had pH data supplied, less than half of them had conductivity and aluminium information in addition to pH. The number of sites with additional chemical variables (e.g. manganese, iron, alkalinity, calcium and copper) was lower still. In order to maximise the geographical coverage of the biological dataset, it was decided that the only chemical variable that would be used in the analysis would be pH.

2.2.8 Data standardisation

Because data had been obtained from several sources, it was necessary to standardise file structures. It was also necessary to standardise the level of taxonomic resolution of the 5 species level data sets. In the case of some taxonomic groups, some of the most precise levels of determination had to be downgraded to achieve a uniform level of identification. In addition to the species level data set, a standard 487 sample biological data set at BMWP 'family' level was produced by downgrading the 'species' data. This included the removal of non-BMWP taxa and combining species records into BMWP family records using the maximum of the $\log_{10}$ abundances of the species in each family to represent each family's abundance.

2.2.9 Acquisition of GIS variables

In order to increase the number of environmental variables available for each site, a semi-automated process in the ARCView geographical information system software was carried out to obtain geographical variables such as altitude, slope, and distance from source for all of the sites in a consistent manner. Together, the five biological data sets (sections 2.2.1 – 2.2.5) with matching chemistry and environmental data formed a substantial data set of 353 samples, which encompassed most areas in England and Wales where acidification was considered to be a potential environmental stress (Figure 4).

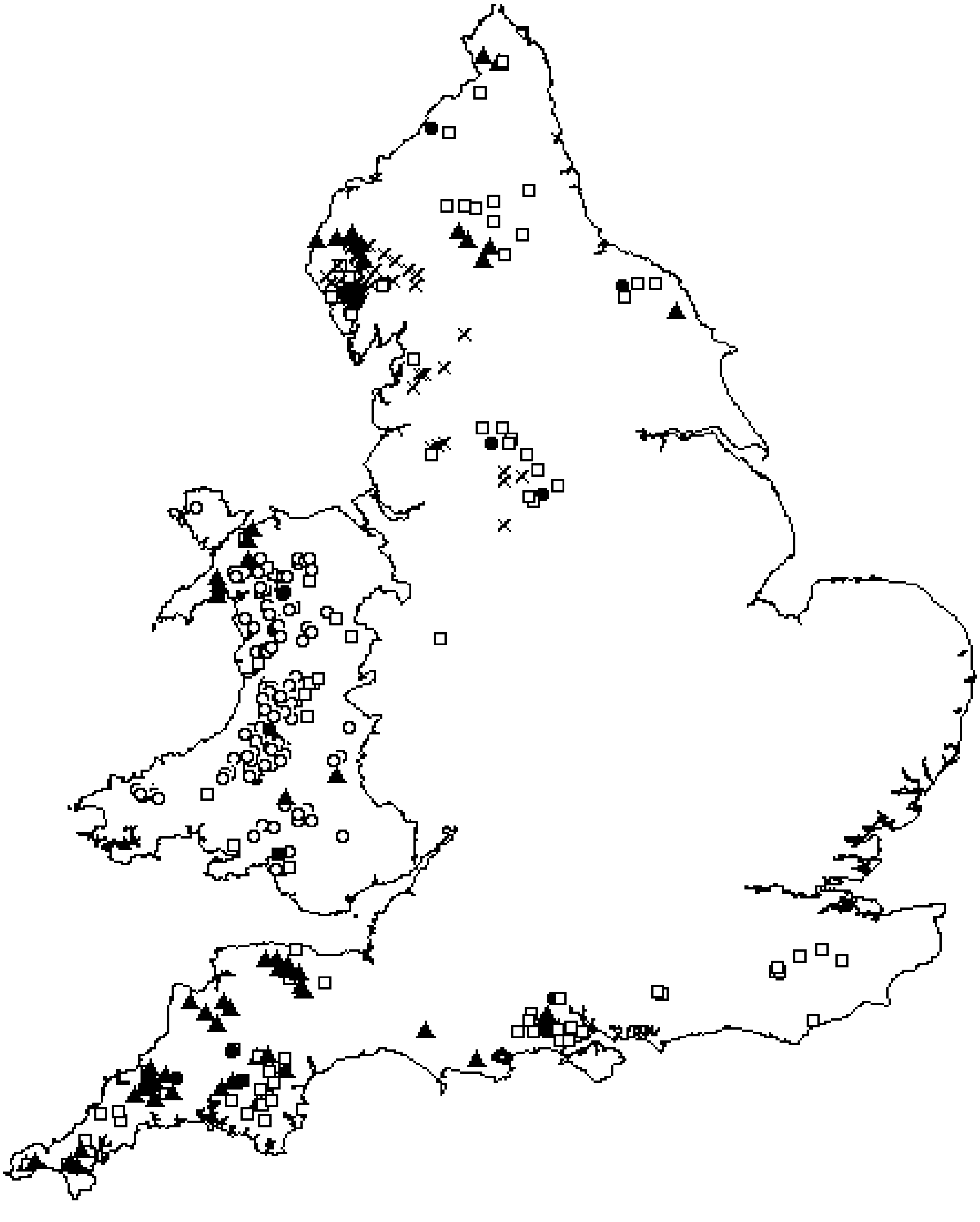


Figure 4. Site map of 353 samples used to develop the AWIC scoring system. Samples are shown by their project: ● IFE Test; O Welsh Acid Waters Survey; □ River Quality Survey; X North West Water Authority; ▲ RIVPACS reference sites.

2.3 Family level analysis and AWIC*(fam)* scoring system construction

The primary aim of the research project was to produce a standard methodology enabling the Environment Agency to assess the extent of ecological damage caused by acidification. In order to do this, it was necessary to determine the extent to which the macroinvertebrate communities in the combined dataset were affected by the primary chemical variable pH, expressed as mean pH, minimum pH, maximum pH, pH range etc.

In considering the development of a family level scoring system, ('AWIC*(fam)*' - the **A**cid **W**aters **I**ndicator **C**ommunity, **fam**ily scoring system), some thought was given to the choice of using taxonomically defined families or BMWP families which contain a number of artificial families (e.g. Dytiscidae including Noteridae). BMWP family level was chosen in preference to taxonomically defined families as it was felt that this would give greater compatibility with the standard BMWP-oriented level of data collection in widespread and well-established use throughout the Environment Agency. It was felt that this would lead to a greater likelihood of acceptance of an acidification scoring system and allow the metric to be calculated on historical national datasets extending back to at least the early 1990s by which time the currently accepted definition of the BMWP families had been widely agreed. All the family level analyses were performed using $\log_{10}$ abundance data.

In an initial investigation of possible analytical approaches to devising an index or scoring system for acidified streams, TWINSPAN was used to form a classification of acid waters in the hope that a distinct acidified community would emerge. In the resultant classification it proved difficult to identify a distinct acid waters community, probably because acidification tends to be characterised by removal of species that would otherwise be expected in non-acidic streams of similar environmental characteristics but with circum-neutral pH. The same effect has also been observed when attempting to use TWINSPAN to isolate an 'organically polluted' community type from a data set of organically polluted and unpolluted sites as part of another research and development project (Davy-Bowker *et al*., 2002). While TWINSPAN is highly suited to classifying unstressed sites (such as in RIVPACS) or for classifying sites that are physically very similar because they are in a restricted geographical area (e.g. Wade *et al*. 1989), it is perhaps less suited to the identification of 'stressed' communities where these are characterised by species loss, especially where the sites are physically diverse.

As an alternative to classification, ordination was used as a means of determining the response of the macroinvertebrate communities to pH. It was important to isolate the community level response of the macroinvertebrates to pH from the wide range of additional variables that have been demonstrated as correlates of macroinvertebrate assemblages (Wright *et al*., 1984). This type of problem is well suited to analysis with Canonical Correspondence Analysis (ter Braak & Smilauer, 1998). CCA is a widely used method for direct gradient ordination that assumes a unimodal model for the relationship between the responses of each taxon to environmental gradients (ter Braak & Prentice 1988). Partial CCA (pCCA) was used in order to factor out the variability in assemblage composition due to environmental variables other than acidity. The residual variability was then assessed in relation to the pH variables. This procedure allowed the response of the macroinvertebrate assemblages to the acidity gradient to be isolated from the over-riding upstream-downstream longitudinal gradient within catchments. CCA and pCCA calculate taxa scores along each ordination axis. These are interpreted as the taxa 'optima' along that axis, assuming that the abundance of a taxon is a symmetrical unimodal function of its position along environmental gradients (ter Braak 1987). CCA has been shown to be quite robust to deviations from this assumption (Palmer 1993). CCA also calculates the tolerance or standard deviation around each taxon optimum along each axis, which can be interpreted as a measure of niche breadth for the taxon along that environmental gradient (ter Braak & Verdonschot 1995).

After an initial examination in CCA to exclude any out-lying taxa that would otherwise skew the analyses, the first stage of the BMWP family level analysis was to run a CCA to determine the amount of variability in the family data that could be accounted for with the available environmental variables.

The second analysis then used CCA with forward selection of environmental variables to determine the relative power of each variable in explaining the variation in the taxon data. Also within this analysis, CCA inflation factors were used to identify (and sequentially remove) variables that were found to be highly co-related to each other (because these add little or nothing to the explanatory power of the analysis yet add to the complexity of subsequent interpretation). This process therefore resulted in a set of largely independent variables that each accounted for a significant amount of variation in the taxon data.

The third stage of analysis, leading to the construction of a scoring system, was to run a 'partial CCA' (pCCA) with the taxon data and environmental variables remaining from the forward selection procedure. pCCA allows each of the physical variables to be allocated into one of two classes, 'constrained' and 'unconstrained'. The variability in the taxon data purely attributable to the unconstrained variable(s) can be isolated from that due the constrained variables. This process was used to separate the effects of pH from the other environmental variables in the combined dataset.

The fourth and final stage of the family level analysis (leading to the construction of a scoring system) was to plot the taxon scores for each family along the primary axis in the pCCA above. The position of the families along this primary axis (which is strongly related to pH since the effects of the other available variables have been partialled out) therefore represents the response of the families to the pH variable(s) alone. This ranking of families in relation to pH response was then subdivided in proportion to their positions along axis 1 so that each family was allocated a score leading to the development of the AWIC*(fam)* scoring system.

In practice this four-step procedure of CCA, forward selection of environmental variables, pCCA and construction of a scoring system was repeated several times over because different combinations of data and available environmental variables needed to be assessed and the resultant scoring systems evaluated.

2.4 Species level analysis and AWIC*(sp)* scoring system construction

The species level analysis, leading to the development of the species scoring system AWIC(*sp*) (the **A**cid **W**aters **I**ndicator **C**ommunity, **sp**ecies scoring system) was carried out in the same way as the family analysis with a four-step procedure of CCA, forward selection of environmental variables, pCCA and construction of a scoring system.

2.5 Preliminary testing of the family and species scoring systems

In the original project design it was envisaged that a new data set would be collected by IFE (now CEH) specifically for testing the new acidity algorithm. The Environment Agency were consulted to assist in the selection process and a small set of 'test' sites were sampled in April and May 1998 (section 2.2.1). The final data set collated for the development of the scoring system was far more extensive than this test data set, and hence it was decided to include these 15 sites in the data set for model development and seek a further independent (and more substantial) data set for model testing. The National Rivers Authority, 1995 GQA data set was therefore used to test the correlation of the family scoring system values against the mean pH for these sites in 1995 (section 2.2.6). This data set was also used to examine the geographical spread of family level scores to assess the extent to which high scores were biased to certain geographical regions of England and Wales and to examine the number of false positives arising from regions where acidification was thought to be highly unlikely.

2.6 Mapping and banding

To assist in the visual interpretation of the spatial behaviour of the AWIC(*fam*) scoring system, a preliminary banding system is proposed. This is merely an aid to interpretation and it is not intended that these bands should in any way define target values for the index value. In the same way that BMWP scores vary in relation to geographical location (e.g. high BMWP-ASPT in lotic streams that are characterised by many stonefly and mayfly families, and low BMWP-ASPT in lentic streams that support elements of a slow or standing water community), it was anticipated that an index developed to assess acidification would also exhibit variation in index values due to physical differences between stream types.

3. RESULTS

3.1 Review of the University of Wales System (the Rutt Key)

The consultation exercise on the effectiveness of the University of Wales System designed by Rutt *et al.*, (1990) was presented in Furse *et al.*, (2000) and is reproduced here. Comments from the Environment Agency were gathered by CEH's three separate meetings with biologists from Welsh, Midlands and South West Regions, in the form of written correspondence from biologists in Anglian and North East Regions and from additional verbal communication with biologists from Midlands and South West Regions.

The University of Wales System was a family level indicator key developed from a data set from the Welsh and North West Regions of the National Rivers Authority and from several of the Scottish River Purification Boards. The key was the first family level index proposed for widespread use in England and Wales by the Environment Agency. It rapidly gained acceptance in Welsh Region for use in special investigations, especially those related to forestry issues, and in routine monitoring programmes, where it proved useful in reporting the extent of acidification in LEAPs (Local Environment Agency Plans) and their predecessors CMPs (Catchment Management Plans). Graham Rutt also applied the system to sites with low total hardness (<30 mgl^{-1} $CaCo_3$) in the National Rivers Authority, 1990 River Quality Survey and sent the results to each Region for evaluation.

The comments collated by CEH fall into two principle types:

- The key failed to distinguish sites affected by metal mine drainage (and some other stress types such as domestic and industrial discharges, salinity and low flow) from those affected by acidification. It was felt that the key would be enhanced by the inclusion of some positive factors to distinguish sites affected by acidification rather than a solely negative set of indicators.

- The key worked poorly in Regions that were beyond the scope of the data set from which it was developed, e.g. in Anglian Region many of the taxa required to make an assessment are absent as a result of their natural distribution patterns.

It was also widely acknowledged throughout the Environment Agency that the University of Wales system was scientifically sound and had made a positive contribution to the development of an index for the assessment of a stress type that had received far less attention than organic pollution.

These comments on the University of Wales System are discussed in more detail in section 4.

3.2 pH tolerances of the BMWP families

To examine the pH tolerances of the BMWP scoring families, box and whisker plots were constructed (Figure 5). These were based on the pH of chemical samples taken within a four-year period (up to 3 years previously and one year after) the biological sample in which each family was found. The samples were drawn from the entire data set and therefore represent the pH tolerances of the BMWP taxa (those found at 6 or more sites in the data set) across the full geographical extent of the sample sites. What is particularly apparent is the wide pH range of most of the families in terms of maximum and minimum pH range. It is also clear from the boxes indicating upper and lower quartiles that while the upper and lower quartile pH range is much narrower than the maximum and minimum pH range, the upper and lower quartiles do not span a mutually exclusive pH range, i.e. the lower quartile for the family with

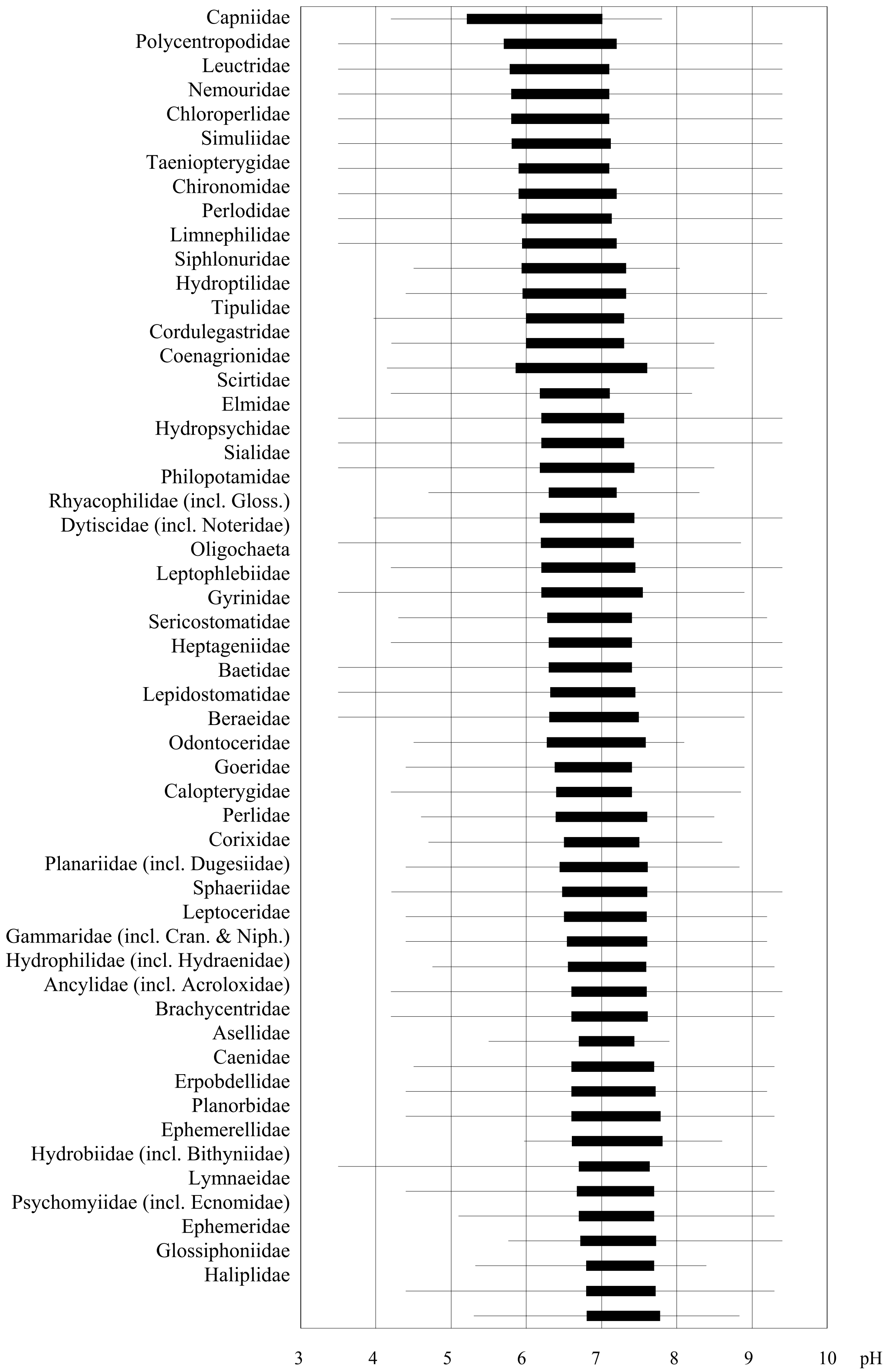


Figure 5. pH tolerances of the BMWP taxa found at 6 or more sites in the data set. Boxes indicate upper and lower quartiles, Whiskers indicate minimum and maximum values.

the greatest sensitivity to low pH (Haliplidae) overlaps the upper quartile for the family with the least sensitivity to low pH (Capniidae).

Although these box and whisker plots are useful in visualising a simple ranking of the BMWP families in terms of pH sensitivity, it is important to realise that the sites within this data set also encompass a large range of variation in terms of many other physical variables and that a scoring system based on this ranking of the taxa would make no allowance for possible relationships between pH and, for example, altitude or distance from source.

3.3 AWIC*(fam)* scoring system

As described in section 2.3, our data analysis with CCA was repeated several times because different combinations of biological data and available environmental variables needed to be analysed and the resultant scoring systems evaluated. The initial analyses are presented in some detail to illustrate relative importance of the environmental variables.

3.3.1 Acid Waters Data Set pH analysis

The biological data set used in the initial analysis at family level comprised the 487 samples (drawn from 410 sites) as shown below:

• IFE Test data set	15 samples	(section 2.2.1)
• Welsh Acid Waters Survey	225 samples	(section 2.2.2)
• River Quality Survey	100 samples	(section 2.2.3)
• North West Water Authority	83 samples	(section 2.2.4)
• RIVPACS reference sites	64 samples	(section 2.2.5)

We obtained pH data for 353 of these biological sampling sites. The minimum requirement for pH was the availability of at least five pH measurements taken within the period of 3 years prior to and 1 year after the biological sample (although in most cases there were many more than 5 readings), and that the pH was measured on the same watercourse and in close proximity to the biological sampling site. The pH data was used as the following variables in the analyses:

- Mean pH
- Minimum pH
- Maximum pH
- Minimum - maximum pH range
- Standard deviation of pH

The following physical variables from the ARCView GIS software were also obtained (section 2.2.9) for all 353 biological samples:

- Altitude (m)
- Slope (m km^{-1})
- Distance from source (km)
- Strahler stream order

After the preliminary examination of the data set with CCA to exclude out-lying taxa, the first stage of the BMWP family level analysis revealed the amount of variability in the family data that could be accounted for with all of the available environmental variables. The total inertia of this analysis was 1.417 and the inertia explained by all of the available environmental variables (including all of the above summary statistics of pH) was 0.273, indicating that 19.3% of the variability in the biological communities had been accounted for by the variables above. This might seem low, but it is reasonable given the unavailability of variables such as width, depth, substratum composition and flow that are known to be important explanatory variables in RIVPACS.

In the second analysis CCA was used with forward selection of environmental variables to determine the relative power of each variable in explaining the variation in the taxon data and to remove highly correlated variables. The following sequence of variables is arranged in order of power in explaining the variation in the taxon data. They reveal mean pH to be the most powerful single variable (out of those available) in accounting for the species variation in the data set:

> Mean pH, altitude (m), slope (m km^{-1}), distance from source (km), maximum pH, standard deviation of pH, minimum pH, Strahler stream order
>
> NB. Minimum - maximum pH range was rejected by the CCA in forward selection of environmental variables because of its very high correlation with the separate variables minimum and maximum pH.

Examination of the inflation factors in this CCA analysis was then used to identify highly correlated variables. There is currently little guidance available to assist in choosing a cut off value for inflation factors and several published analyses report the use of a cut off values as high as 20. A more conservative cut off point of 3.0 (e.g. Magalhães *et al.,* 2002) for inflation factors was used in this analysis. The following variables, which could therefore be regarded as sufficiently independent and at the same time contributed significant explanatory power remained:

> Mean pH, altitude (m), slope (m km^{-1}), distance from source (km), standard deviation of pH, Strahler stream order

Interestingly minimum pH, (which is thought to be particularly important for invertebrate communities), was removed because it was highly correlated with mean pH but was much weaker as an explanatory variable, while standard deviation of pH (a measure of pH variability) remained. Removal of the two variables; minimum pH and maximum pH resulted in only a slight reduction in the power of the analysis as a whole in accounting for variability in the biological communities (18.1% compared to 19.3%).

The third stage of analysis was to run a ‘partial CCA’ (pCCA) using only the environmental variables remaining from the forward selection procedure, with the two variables mean pH and standard deviation pH as explanatory variables and the other variables (altitude, slope, distance from source and Strahler stream order) as co-variables. The fourth stage of the family level analysis was to plot the taxon scores for each family in this analysis along the primary axis in the pCCA, (Figure 6), which was accounted for primarily by mean pH and to a lesser extent by standard deviation of pH).

3.3.2 1995 GQA data set pH analysis

The analysis presented in section 3.3.1 was performed on a data set that was collated specifically for the examination of biological communities in relation to acid stress. This data set was biased towards a moderate to low pH range with less representation of neutral to high

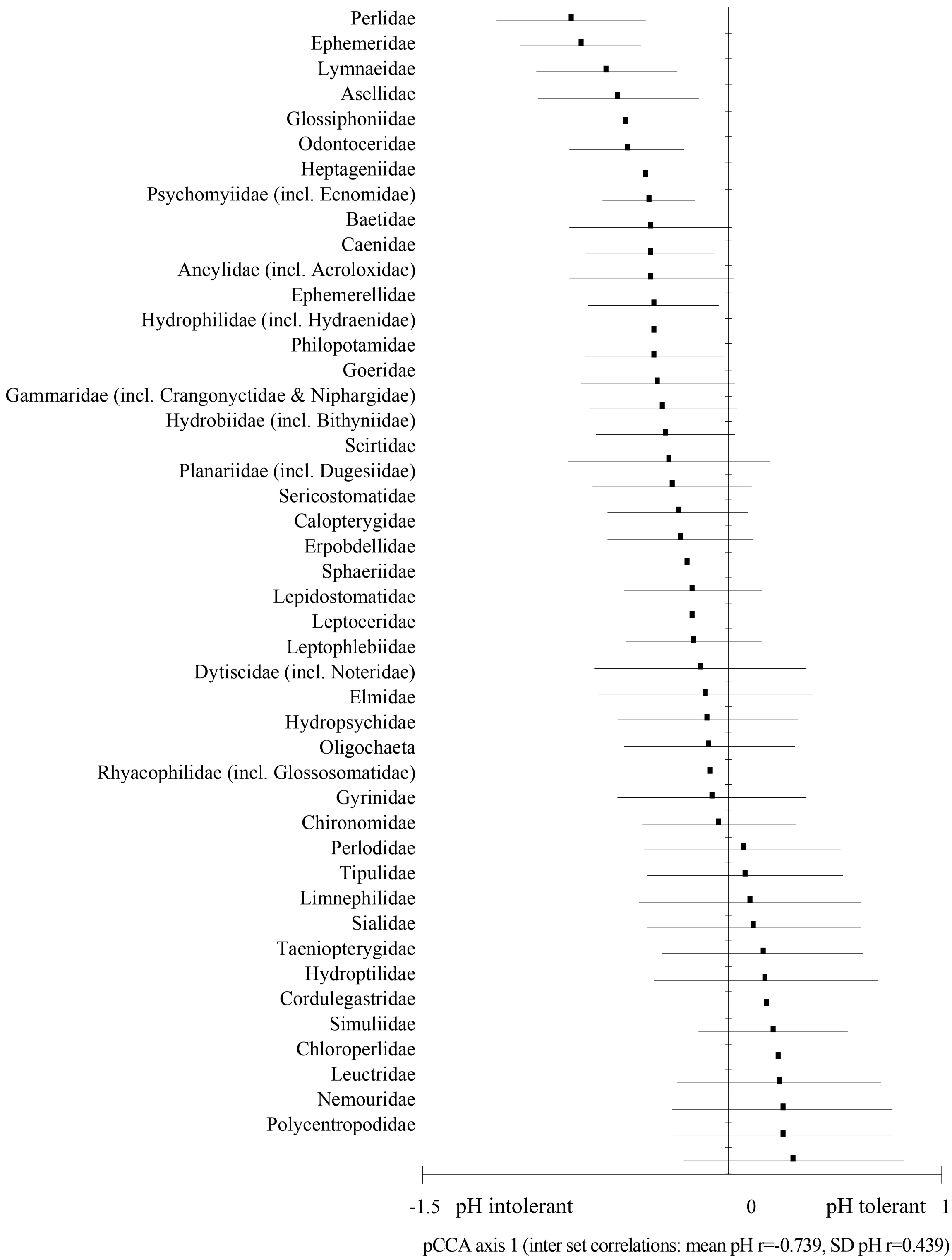


Figure 6. Ranking of BMWP taxa along axis 1 of the pCCA analysis on the 353 sample Acid Waters Data Set (see section 3.3.1 for further explanation).

pH sites. To construct a ranking of acidity tolerance including sites with high pH we used the National Rivers Authority GQA 1995 data set (section 2.2.6). Of the 5837 sites where biological data was available, a sub set of 956 sites had either 'no perceived stress' or only 'acidic stress' (based on a questionnaire described in Davy-Bowker *et al*., 2000). These 956 sites could therefore be regarded as free from other significant types of stress. Of these sites, 689 also had pH data. Four pH summary statistics were used (N.B. minimum – maximum pH range was dropped because it is too highly correlated with minimum and maximum pH):

- Mean pH
- Minimum pH
- Maximum pH
- Standard deviation of pH

The GQA data had more physical variables than the combined acid waters data set used in the previous analysis:

- Altitude (m)
- Slope (m km^{-1})
- Distance from source (km)
- Discharge category
- Average depth (cm)
- Width (m)
- Substratum composition (Phi scale)

Because there were more physical variables in this data set than in the first analysis, a forward selection of environmental variables was used to find and remove highly correlated variables that were weak in terms of explanatory power. The following sequence of variables is arranged in order of power in explaining the variation in the taxon data. Substratum composition, followed by mean pH were the most powerful variables (out of those available) in accounting for the species variation in the data set:

> Substratum composition (Phi), mean pH, distance from source (km), average depth (cm), altitude (m), discharge category, slope (m km^{-1}), width (m), standard deviation pH, maximum pH, minimum pH

A CCA analysis was carried out in order to identify and remove highly correlated variables. Four variables were clearly correlated, with inflation factors ranging between 14 and 37. These variables were the four pH summary statistics mean pH, standard deviation pH, maximum pH and minimum pH. Because of their weak explanatory power, maximum pH and minimum pH were removed. Of the seven non-pH variables, none of these exceeded an inflation factor of 4.1, which was considered acceptable.

The third stage of analysis was to run a 'partial CCA' (pCCA) using only the environmental variables remaining from the forward selection procedure, with the two variables mean pH and standard deviation pH as explanatory variables and the other variables (substratum composition, distance from source, average depth, altitude, discharge category, slope and width) as co-variables. The fourth stage of the analysis was to plot the taxon scores for each family in this analysis along the primary axis in the pCCA, (Figure 7), which was accounted for primarily by mean pH. Compared with the previous analysis, this set of environmental variables accounted for over 26% of the total variation in the taxon data.

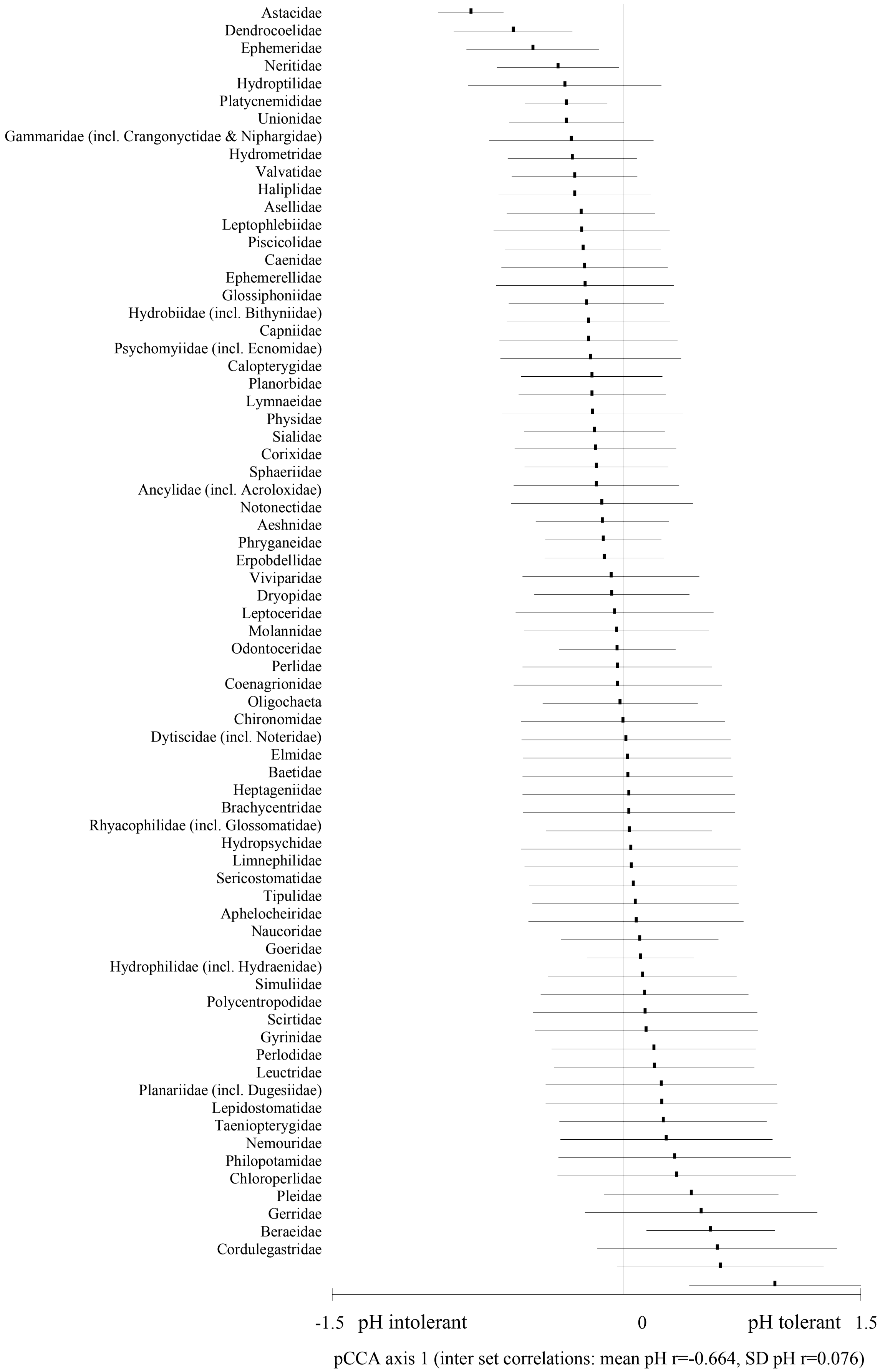


Figure 7. Ranking of BMWP taxa along axis 1 of the pCCA analysis on the 689 sample GQA 1995 data (see section 3.3.2 for further explanation).

3.3.3 Combined Acid Waters and 1995 GQA data set pH analysis

Examination of the previous two analyses reveals close agreement between the two taxon ranks (Spearman's Rank Correlation Coefficient = 0.508, P-value = <0.01), with a few notable exceptions (e.g. Perlidae, Sialidae and Polycentropodidae). Both data sets were then merged into a combined data set of 1042 samples comprising:

- Acid Waters Data Set 353 samples (section 3.3.1)
- 1995 data set 689 samples (section 3.3.2)

The following pH summary statistics and physical variables were available for the whole data set:

- Mean pH
- Minimum pH
- Maximum pH
- Standard deviation of pH
- Altitude (m)
- Slope (m km^{-1})
- Distance from source (km)

After an initial analysis to identify and remove rare (occurring in less than 5% of sites) and outlying taxa, a CCA with forward selection of environmental variables was used to identify and remove highly correlated variables, again resulting in the removal of maximum pH and minimum pH.

We then performed pCCA analysis with mean pH and standard deviation of pH as variables and altitude, slope and distance from source as co-variables. In this final analysis of the family level data, the mean and standard deviation pH accounted for 4.5% of the taxon variability, the variation accounted for by the co-variables (altitude, slope and distance from source) was 11.4% and the variation accounted for by all of the environmental variables combined was 15.9%. The pCCA diagram for this analysis is presented in Figure 8. The positions of the families along this primary axis (which is strongly related to mean pH and, to a lesser extent, standard deviation of pH since the effects of the other available variables have been partialled out) therefore, primarily represents the response of the families to pH. The taxa at either extreme of this ranking were then allocated values of zero and 100, and the taxa in between were given percentage positions along this axis. These percentage positions were then divided into bands at each 10% interval to form a scoring system where each family was allocated an integer score of 1 (acid tolerant) to 10 (acid intolerant).

The initial 10-band scoring system was then investigated by plotting the AWIC*(fam)* and AWIC*(fam)*-ASPT (where ASPT is the average score per taxon) against mean pH for the 3393 grade a-c sites in the 1995 GQA data set (section 2.2.6) where mean pH was available. The AWIC(*fam*)-ASPT versus mean pH plot had a wide scatter of points, especially for neutral to alkaline sites (Figure 9) such that prediction of mean pH from a given AWIC(*fam*)-ASPT score was imprecise. To address this problem, the variation in scores given to high scoring taxa in the positive side of Axis 1 (Figure 8) was reduced resulting in the scoring system given in Table 7. Compared to the 1-10 band system, the 1-6 band system is less likely to differentiate change in mean pH at a given site but this has to be balanced against the greater predictive power of a given AWIC*(fam)* –ASPT score in terms of mean pH.

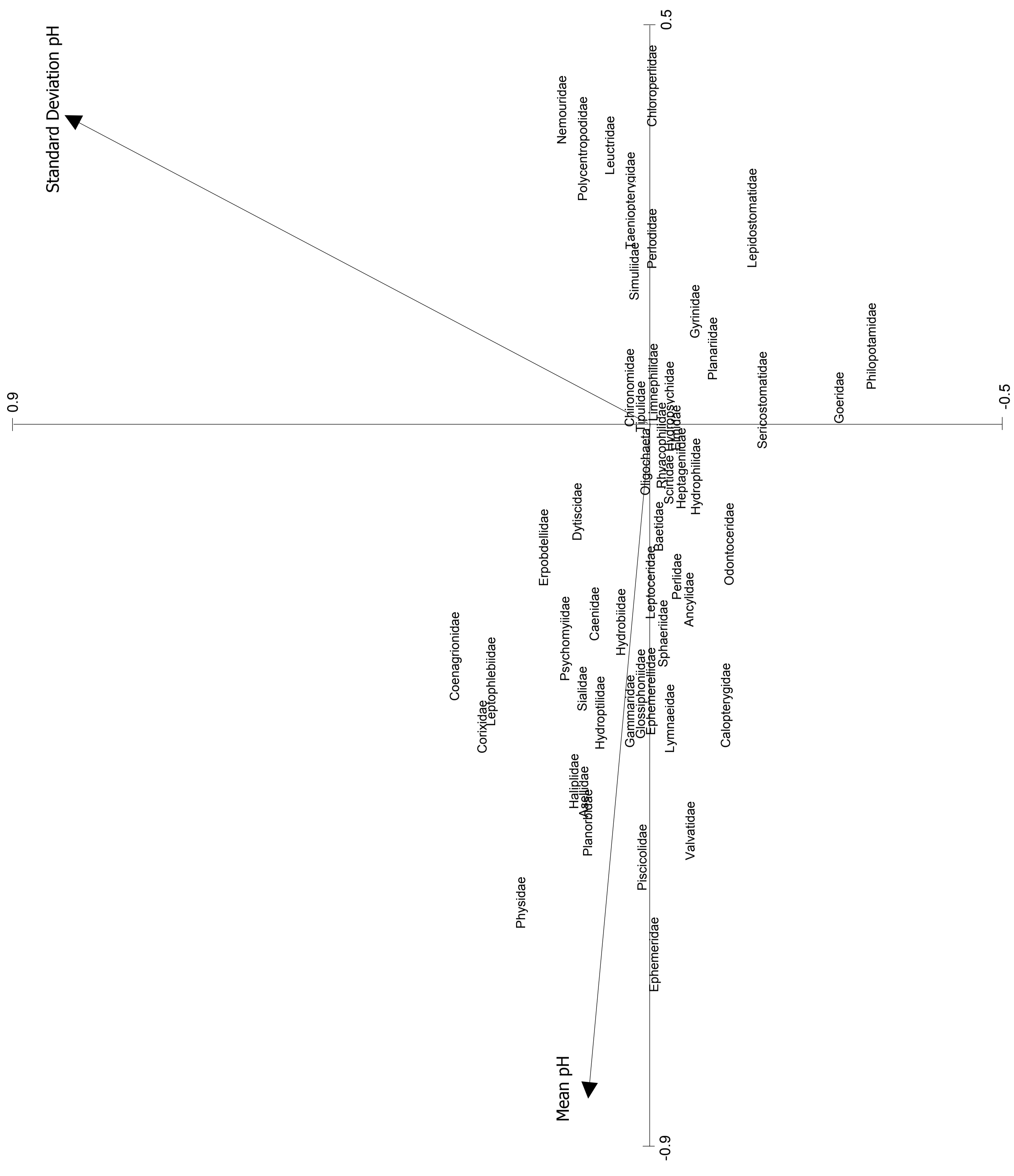


Figure 8. Partial CCA analysis using the two variables mean and standard deviation of pH as explanatory variables and altitude, slope, distance from source and Strahler stream order as co-variables. Artificial BMWP families are shown by their primary name. Inter set correlations: mean pH r=-0.670, SD pH r=0.287).

Table 7. The AWIC*(fam)* scoring system. AWIC*(fam)*-ASPT is the average score per taxon.

Furse code	'Family' name	AWIC*(fam)* scores
40320000	Ephemeridae	6
16210000	Physidae	
22110000	Piscicolidae	
16130000	Valvatidae	
16230000	Planorbidae	
36110000	Asellidae	
45110000	Haliplidae	
43610000	Corixidae	
16220000	Lymnaeidae	
48130000	Hydroptilidae	
371Z0000	Gammaridae (incl. Crangon. & Nipharg.)	
42140000	Calopterygidae	
22120000	Glossiphoniidae	
40410000	Ephemerellidae	
40210000	Leptophlebiidae	
46110000	Sialidae	
42120000	Coenagrionidae	
482Z0000	Psychomyiidae (incl. Ecnomidae)	
17130000	Sphaeriidae	
161Z0000	Hydrobiidae (incl. Bithyniidae)	
40510000	Caenidae	
162Z0000	Ancylidae (incl. Acroloxidae)	
48410000	Leptoceridae	
41220000	Perlidae	
22310000	Erpobdellidae	
48380000	Odontoceridae	
40120000	Baetidae	
451Z0000	Dytiscidae (incl. Noteridae)	
45510000	Scirtidae	
453Z0000	Hydrophilidae (incl. Hydraenidae)	
40130000	Heptageniidae	
20000000	Oligochaeta	
481Z0000	Rhyacophilidae (incl. Glossosomatidae)	
45630000	Elmidae	
48250000	Hydropsychidae	4
50100000	Tipulidae	
48370000	Sericostomatidae	
48350000	Goeridae	
50400000	Chironomidae	
48340000	Limnephilidae	
051Z0000	Planariidae (incl. Dugesiidae)	
48210000	Philopotamidae	3
45150000	Gyrinidae	
50360000	Simuliidae	
41210000	Perlodidae	2
48330000	Lepidostomatidae	
41110000	Taeniopterygidae	
48240000	Polycentropodidae	1
41130000	Leuctridae	
41120000	Nemouridae	
41230000	Chloroperlidae	

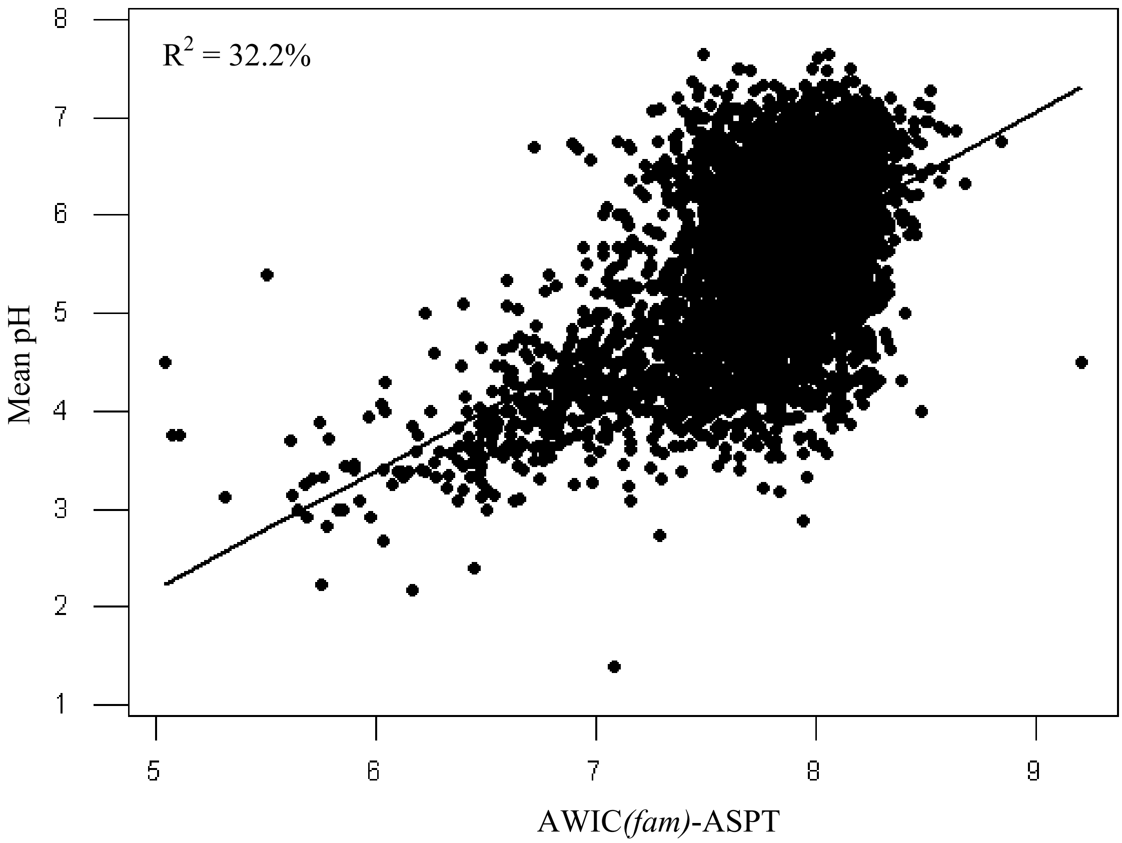


Figure 9. Correlation of AWIC*(fam)*-ASPT against mean pH (for 3393 1995 GQA grade a-c sites) for the initial 10-band version of the AWIC*(fam)*-ASPT scoring system. Individual family scores were integers at 10% intervals along axis 1 of the pCCA analysis in section 3.3.3.

3.4 Deriving the relationship between AWIC*(fam)*-ASPT and mean pH

The relationship between the AWIC*(fam)*-ASPT and mean pH was examined in the data used to derive the AWIC scoring system. AWIC*(fam)*-ASPT was chosen for this analysis rather than the AWIC*(fam)* score itself because AWIC*(fam)* score was shown to have a poorer correlation with mean pH than AWIC*(fam)*-ASPT in an initial application of both indices to the 4174 1995 GQA sites where pH data was available. The relationship between mean pH was chosen, rather than standard deviation of pH, because mean pH was the more powerful explanatory variable of the two and mean pH is a commonly calculated statistic in reporting chemical water quality whereas standard deviation of pH is not.

Linear regression analysis to derive the relationship between AWIC*(fam)*-ASPT versus mean pH gave the regression equation below:

Mean pH = 3.93416 + 0.76094 AWIC*(fam)*-ASPT $\quad\quad$ R^2 = 67.8%, $P<0.001$)

Mean pH and AWIC*(fam)*-ASPT are strongly correlated, as expected, given that this is based on the data used to derive the scoring system. The regression plot is given in Figure 10 together with 95 percentiles.

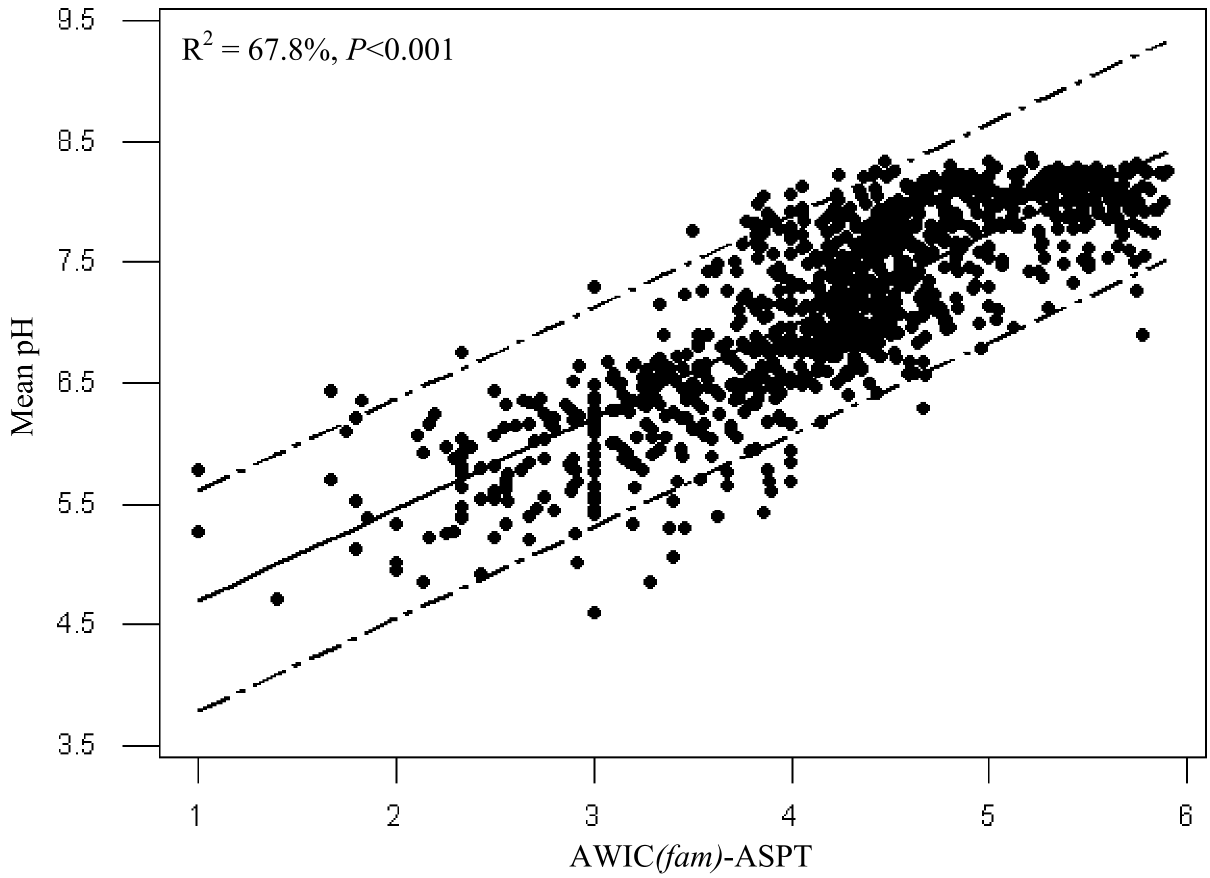


Figure 10. Regression plot of AWIC*(fam)*-ASPT versus mean pH with 95 percentile limits for the 1042 samples used to derive the AWIC*(fam)* scoring system.

Predicted mean pH values for given AWIC*(fam)*-ASPT values are given in Table 8 together with 95 percentile limits (AWIC*(fam)*-ASPT values below 2.0 are not given because there were too few samples in this region).

Table 8. Predicted mean pH values for given AWIC*(fam)*-ASPT with 95 percentiles.

AWIC	Mean pH	Lower 95 Percentile	Upper 95 Percentile
2.0	5.46	4.55	6.37
2.5	5.84	4.93	6.75
3.0	6.22	5.31	7.12
3.5	6.60	5.69	7.50
4.0	6.98	6.07	7.88
4.5	7.36	6.45	8.27
5.0	7.74	6.83	8.65
5.5	8.12	7.21	9.03
6.0	8.50	7.59	9.41

The 95 percentiles of predicted mean pH for a given AWIC*(fam)*-ASPT in Table 8 are quite wide, so an alternative model is presented below with 51 percentiles (Figure 11).

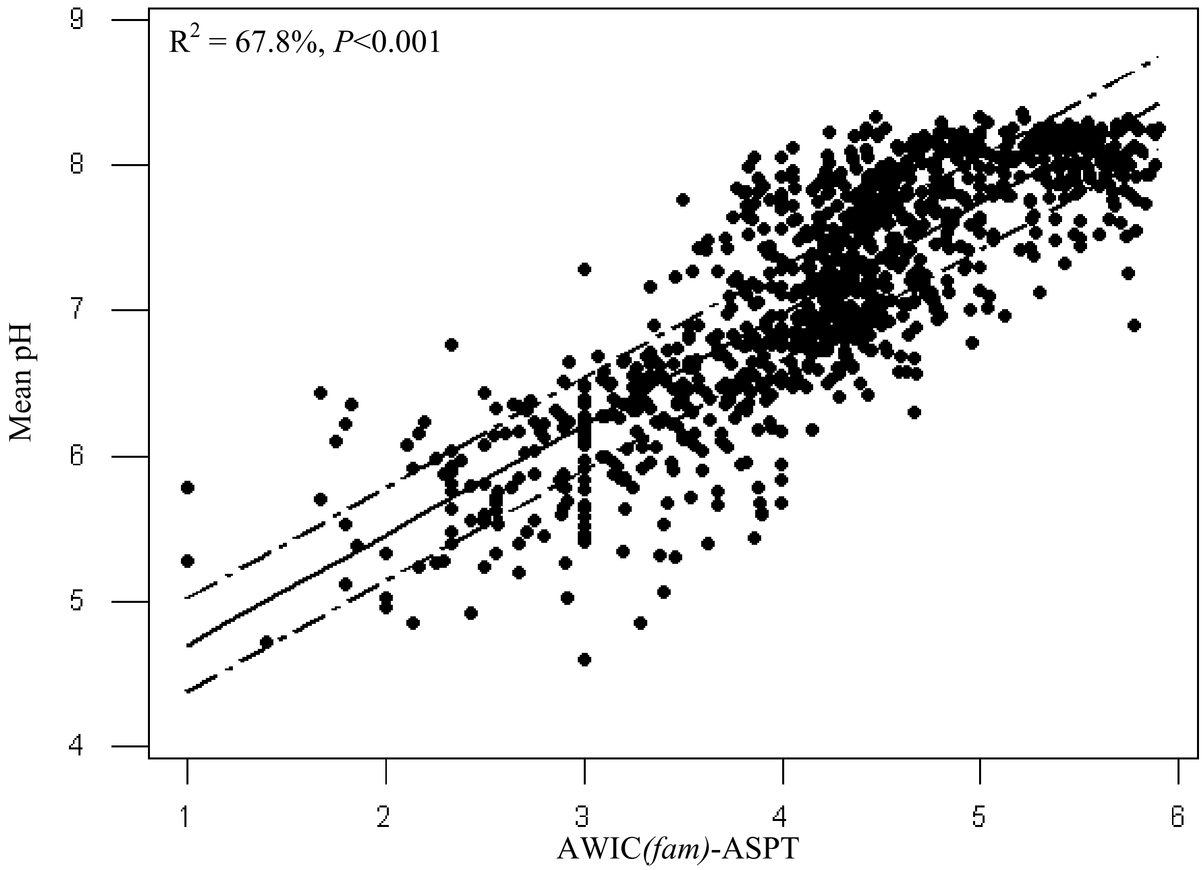


Figure 11. Regression plot of AWIC*(fam)*-ASPT versus mean pH with 51 percentile limits for the 1042 samples used to derive the AWIC scoring system.

The 51 percentile is an indication that it is more likely than not that a mean pH predicted from a given AWIC*(fam)*-ASPT lies between the 51 percentile limits. The predicted mean pH values for given AWIC*(fam)*-ASPT values, together with the 51 percentile limits, are given in Table 9 (again AWIC*(fam)*-ASPT values below 2.0 are not given because there were too few samples in this region). To further illustrate the relationship between AWIC*(fam)*-ASPT and mean pH, the probabilities that each AWIC*(fam)*-ASPT class will have a certain mean pH are given in Table 10.

Table 9. Predicted mean pH values for given AWIC*(fam)*-ASPT with 51 percentiles.

AWIC*(fam)*-ASPT	Mean pH	Lower 51Percentile	Upper 51Percentile
2.0	5.46	5.14	5.78
2.5	5.84	5.52	6.16
3.0	6.22	5.90	6.54
3.5	6.60	6.28	6.92
4.0	6.98	6.66	7.30
4.5	7.36	7.04	7.68
5.0	7.74	7.42	8.06
5.5	8.12	7.80	8.44
6.0	8.50	8.18	8.82

Table 10. The upper portion shows the number of samples contributing to the prediction and the lower portion shows the probability of that each AWIC*(fam)*-ASPT class will have a certain mean pH. Intervals shown are the upper class limits and values are rounded. AWIC*(fam)*-ASPT class 1.5 is excluded due to insufficient data.

Counts		AWIC*(fam)*-ASPT Class									Total
		2	2.5	3	3.5	4	4.5	5	5.5	6	
pH Class	5	1	2	1	1	-	-	-	-	-	6
	5.5	4	7	11	4	2	-	-	-	-	28
	6	2	16	24	16	13	-	-	-	-	71
	6.5	4	6	30	37	44	8	1	-	-	130
	7	-	1	2	23	56	94	15	1	1	193
	7.5	-	-	1	2	27	85	48	11	1	175
	8	-	-	-	2	25	86	75	39	28	255
	8.5	-	-	-	-	2	15	51	63	51	182
	Total	11	32	69	85	169	288	190	114	81	1042

Probabilities		AWIC*(fam)*-ASPT Class								
		2	2.5	3	3.5	4	4.5	5	5.5	6
pH Class	5	9.1	6.3	1.5	1.2	-	-	-	-	-
	5.5	36.4	21.9	15.9	4.7	1.2	-	-	-	-
	6	18.2	50.0	34.8	18.8	7.7	-	-	-	-
	6.5	36.4	18.8	43.5	43.5	26.0	2.8	0.5	-	-
	7	-	3.1	2.9	27.1	33.1	32.6	7.9	0.9	1.2
	7.5	-	-	1.5	2.4	16.0	29.5	25.3	9.7	1.2
	8	-	-	-	2.4	14.8	29.9	39.5	34.2	34.6
	8.5	-	-	-	-	1.2	5.2	26.8	55.3	63.0

It is important to note that the predictions presented in section 3.4 are based on the relationship between AWIC*(fam)*-ASPT and mean pH in the data used to derive the scoring system itself and that these equations need to be tested on an independent data set to assess the validity of the model (section 3.5).

3.5 Testing the relationship between AWIC family and mean pH

To test the predictions of the *AWIC*(fam)-*ASPT* regression equation with mean pH we used a partially independent data set of 3393 GQA 1995 grade a-c sites (where grade was the banded EQI of BMWP-ASPT). We calculated the *AWIC*(fam)-*ASPT* for each site and used the regression equation (section 3.4) to predict the 95 and 51 percentile ranges of pH at each site. We then took the actual mean pH of each site to see if this lay between the predicted percentile ranges. The predictions of the regression equation were confirmed at the 51 percentile confidence level, where the mean pH of the partially independent test data lay within the predicted limits on 61.4% of occasions. They were also confirmed at the 95 percentile confidence level where the test data lay within the limits on 97.5% of occasions.

3.6 AWIC*(fam)*-ASPT correlations with other indices

To compare the behaviour of the AWIC*(fam)*-ASPT scoring system and the BMWP-ASPT, regressions were performed using the 3393 three-season combined grade a-c samples from the Environment Agency 1995 General Quality Assessment survey where pH data was available (section 2.2.6). These are presented in Figures 12 and 13.

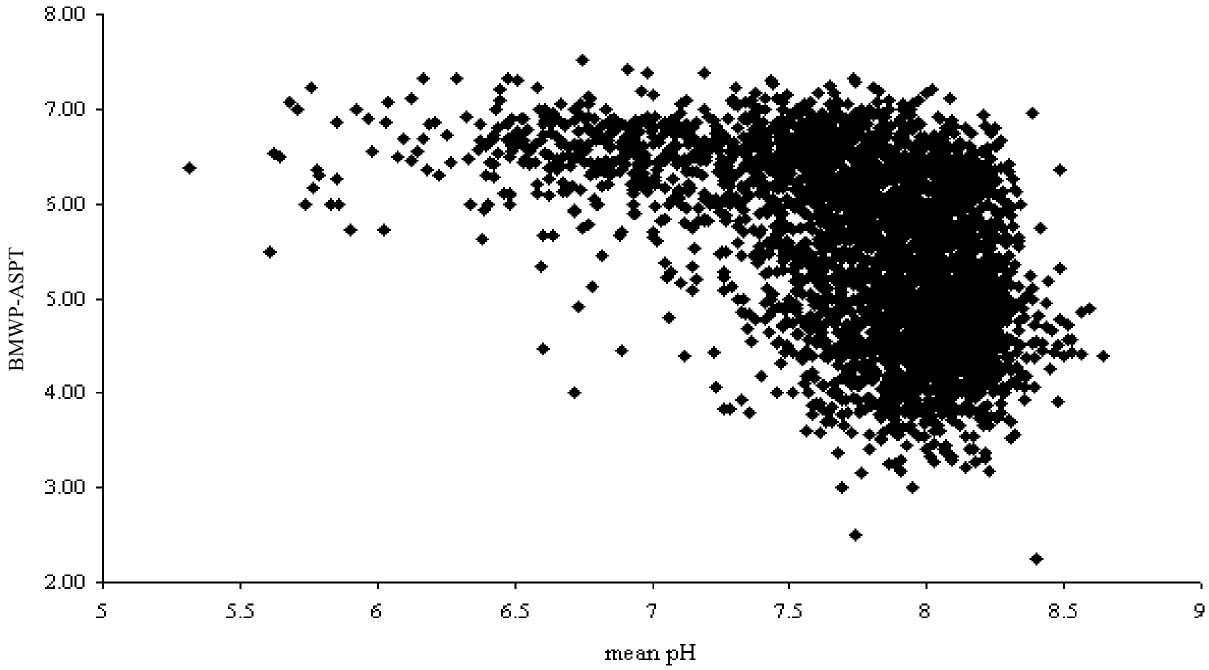


Figure 12. Correlation of BMWP-ASPT against mean pH (grade a-c 1995 GQA data, n=3393).

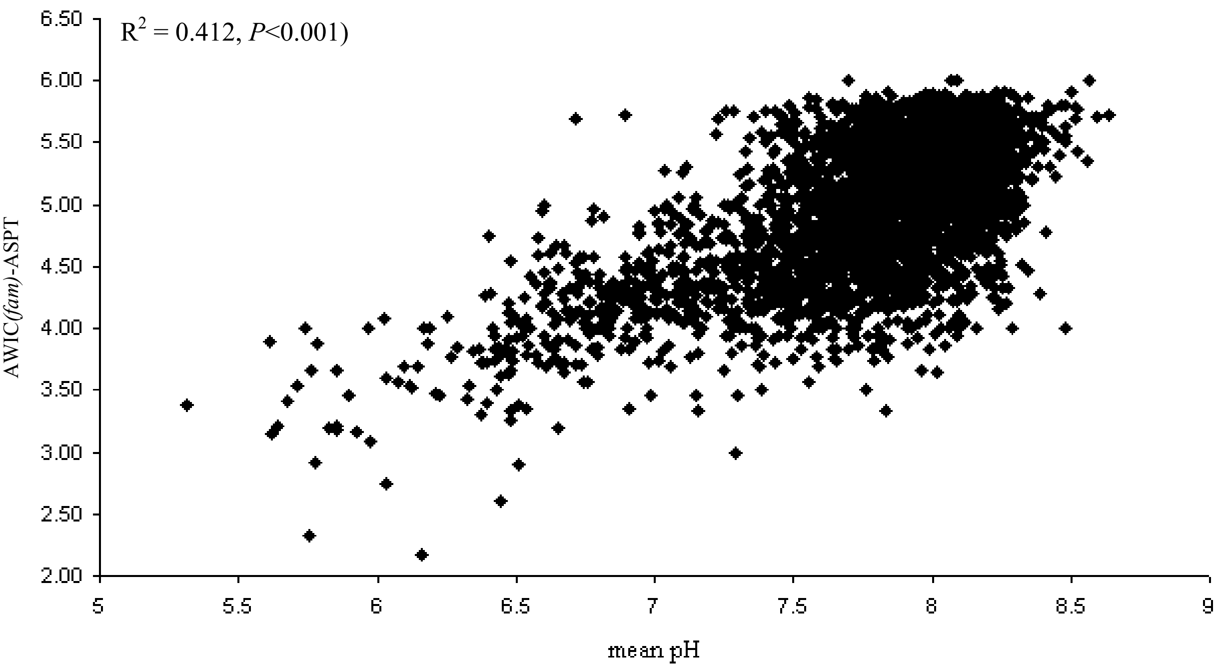


Figure 13. Correlation of AWIC*(fam)*-ASPT against mean pH (grade a-c 1995 GQA data, n=3393).

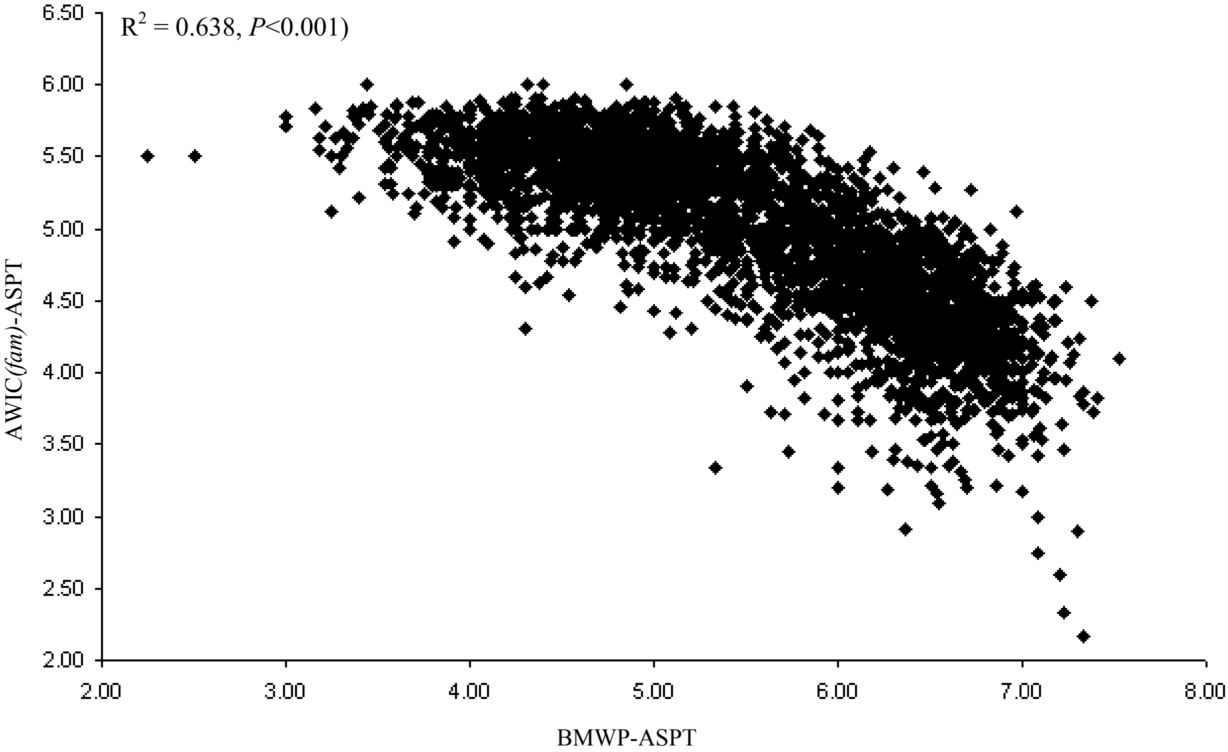


Figure 14. Correlation of AWIC*(fam)*-ASPT against BMWP-ASPT (grade a-c 1995 GQA data, n=3393).

The high BMWP-ASPT values commonly encountered at low pH are shown in Figure 12. This feature of BMWP-ASPT has been recognised by Environment Agency biologists as a possible indication of acidity or metal pollution. At pH values above ca. 7, BMWP-ASPT varies considerably. In contrast AWIC*(fam)*-ASPT has a more linear relationship with mean pH (Figure 13) and the R^2 value of 0.412 indicates that 41.2% of the variation in AWIC*(fam)*-ASPT is accounted for by mean pH. There is also a strong negative relationship between BMWP-ASPT and AWIC*(fam)*-ASPT (Figure 14) so that sites with low AWIC*(fam)*-ASPT (indicating more acidic condition) are also those sites where BMWP-ASPT is higher (indicating less organic pollution).

3.7 Geographical distribution of AWIC*(fam)*-ASPT scores

The geographical distribution of AWIC*(fam)*-ASPT scores is presented in Figure 15. AWIC*(fam)*-ASPT values are displayed according to a provisional banding system at one-point intervals through the range one to six. Sites with low AWIC*(fam)*-ASPT values were present in Wales, the North West of England, the Pennines, the South West and Ashdown Forest, but generally absent from central, southern and eastern England. Sites in the lowest category of AWIC*(fam)*-ASPT (1.00-1.99) were rare in the GQA dataset (although they were present in the data used to derive the index (see Figure 9).

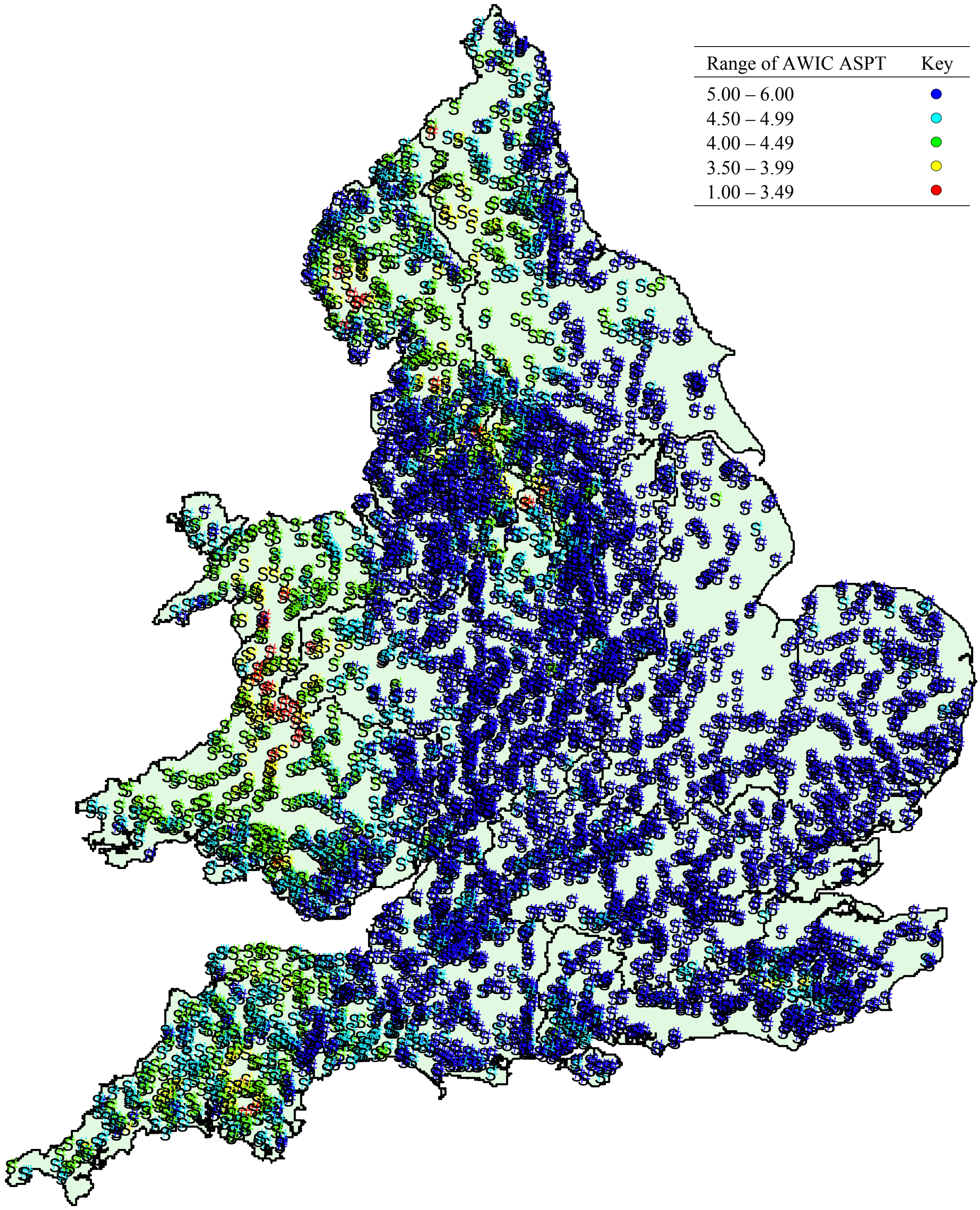


Figure 15. Geographical distribution of AWIC*(fam)*-ASPT scores in England and Wales based on the National Rivers Authority 1995 General Quality Assessment survey (n=6022).

3.8 AWIC*(sp)* scoring system

In addition to the family level work already described, CEH undertook a further analysis with the aim of producing a species level scoring system for the assessment of stream acidification. This analysis was conducted in the same way as the family level work, although a notable difference in the data set was that it was in the form of presence/absence data only, in contrast to the family level $\log_{10}$ abundance data. The analysis was performed using the biological data set described in detail in section 3.3.1, which comprised 353 biological samples from five sources. The five summary statistics of pH and 4 environmental variables obtained from the ARCView GIS software were as follows:

- Mean pH
- Minimum pH
- Maximum pH
- Minimum - maximum pH range
- Standard deviation of pH
- Altitude (m)
- Slope (m km^{-1})
- Distance from source (km)
- Strahler stream order

After taxonomic standardisation and removal of all rare taxa occurring in less than 10 of the 353 samples (leaving 91 species in the analysis), the first stage of the species level analysis was to run a CCA with forward selection of environmental variables to identify and remove variables that were highly co-related and weak in explanatory power. As in the family level analysis, minimum - maximum pH range, minimum pH and maximum pH were removed by the forward selection procedure leaving the following variables which were both independent (inflation factors below 3.0) and contributed significant explanatory power:

- Mean pH
- Standard deviation of pH
- Altitude (m)
- Slope (m km^{-1})
- Distance from source (km)
- Strahler stream order

The next stage was to use a partial CCA to separate the variation in the species data due to pH from the other environmental variables. Mean and standard deviation pH were therefore chosen as variables whilst altitude, slope, distance from source and Strahler stream order were the co-variables. The ranking of species along the primary axis of this pCCA is shown in Figure 16. The horizontal bars of this plot (tolerances) represent niche width and in this analysis, where the level of taxonomic resolution is more precise than the family level work, the tolerances have been corrected for bias due to varying numbers of occurrence across all taxa following the method recommended in ter Braak & Smilauer, (1998).

The final stage, resulting in the species level scoring system, was to allocate scores of 1 and 10 to the two extremes of the ranking and then to divide the species list into bands at 10% intervals (Table 11).

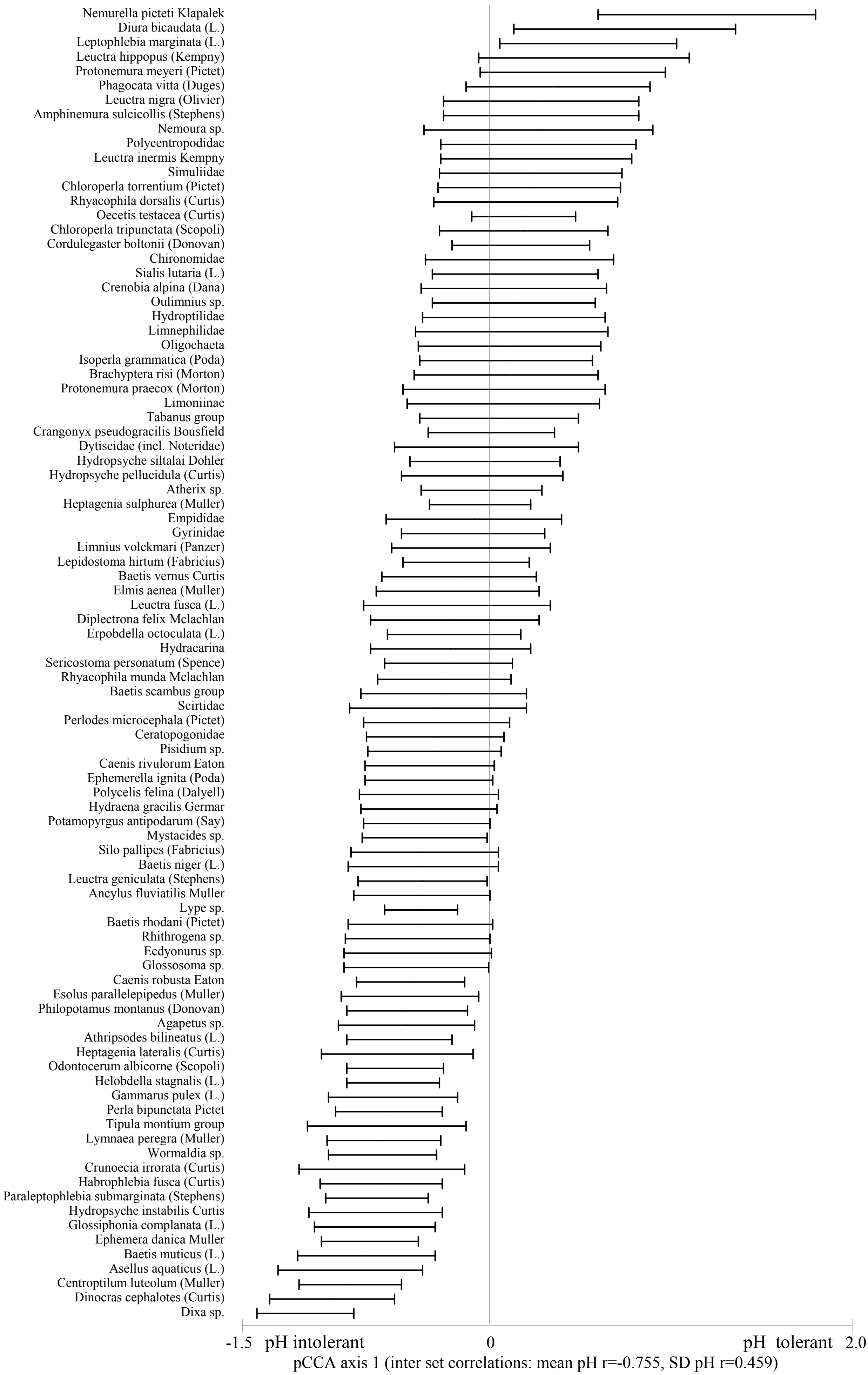


Figure 16. pCCA axis 1 ranking of the species level analysis.

Table 11. The AWIC*(sp)* scoring system. AWIC*(sp)*-ASPT is the average score per taxon.

<table>
<tr><td>Dixa sp.</td><td rowspan="2">10</td></tr>
<tr><td>Dinocras cephalotes (Curtis)</td></tr>
<tr><td>Centroptilum luteolum (Muller)</td><td rowspan="11">9</td></tr>
<tr><td>Asellus aquaticus (L.)</td></tr>
<tr><td>Baetis muticus (L.)</td></tr>
<tr><td>Ephemera danica Muller</td></tr>
<tr><td>Glossiphonia complanata (L.)</td></tr>
<tr><td>Hydropsyche instabilis Curtis</td></tr>
<tr><td>Paraleptophlebia submarginata (Stephens)</td></tr>
<tr><td>Habrophlebia fusca (Curtis)</td></tr>
<tr><td>Crunoecia irrorata (Curtis)</td></tr>
<tr><td>Wormaldia sp.</td></tr>
<tr><td>Lymnaea peregra (Muller)</td></tr>
<tr><td>Tipula montium group</td><td rowspan="21">8</td></tr>
<tr><td>Perla bipunctata Pictet</td></tr>
<tr><td>Gammarus pulex (L.)</td></tr>
<tr><td>Helobdella stagnalis (L.)</td></tr>
<tr><td>Odontocerum albicorne (Scopoli)</td></tr>
<tr><td>Heptagenia lateralis (Curtis)</td></tr>
<tr><td>Athripsodes bilineatus (L.)</td></tr>
<tr><td>Agapetus sp.</td></tr>
<tr><td>Philopotamus montanus (Donovan)</td></tr>
<tr><td>Esolus parallelepipedus (Muller)</td></tr>
<tr><td>Caenis robusta Eaton</td></tr>
<tr><td>Glossosoma sp.</td></tr>
<tr><td>Ecdyonurus sp.</td></tr>
<tr><td>Rhithrogena sp.</td></tr>
<tr><td>Baetis rhodani (Pictet)</td></tr>
<tr><td>Lype sp.</td></tr>
<tr><td>Ancylus fluviatilis Muller</td></tr>
<tr><td>Leuctra geniculata (Stephens)</td></tr>
<tr><td>Baetis niger (L.)</td></tr>
<tr><td>Silo pallipes (Fabricius)</td></tr>
<tr><td>Mystacides sp.</td></tr>
<tr><td>Potamopyrgus antipodarum (Say)</td><td rowspan="18">7</td></tr>
<tr><td>Hydraena gracilis Germar</td></tr>
<tr><td>Polycelis felina (Dalyell)</td></tr>
<tr><td>Ephemerella ignita (Poda)</td></tr>
<tr><td>Caenis rivulorum Eaton</td></tr>
<tr><td>Pisidium sp.</td></tr>
<tr><td>Ceratopogonidae</td></tr>
<tr><td>Perlodes microcephala (Pictet)</td></tr>
<tr><td>Scirtidae</td></tr>
<tr><td>Baetis scambus group</td></tr>
<tr><td>Rhyacophila munda Mclachlan</td></tr>
<tr><td>Sericostoma personatum (Spence)</td></tr>
<tr><td>Hydracarina</td></tr>
<tr><td>Erpobdella octoculata (L.)</td></tr>
<tr><td>Diplectrona felix Mclachlan</td></tr>
<tr><td>Leuctra fusca (L.)</td></tr>
<tr><td>Elmis aenea (Muller)</td></tr>
<tr><td>Baetis vernus Curtis</td></tr>
</table>

<table>
<tr><td>Lepidostoma hirtum (Fabricius)</td><td rowspan="14">6</td></tr>
<tr><td>Limnius volckmari (Panzer)</td></tr>
<tr><td>Gyrinidae</td></tr>
<tr><td>Empididae</td></tr>
<tr><td>Heptagenia sulphurea (Muller)</td></tr>
<tr><td>Atherix sp.</td></tr>
<tr><td>Hydropsyche pellucidula (Curtis)</td></tr>
<tr><td>Hydropsyche siltalai Dohler</td></tr>
<tr><td>Dytiscidae (incl. Noteridae)</td></tr>
<tr><td>Crangonyx pseudogracilis Bousfield</td></tr>
<tr><td>Tabanus group</td></tr>
<tr><td>Limoniinae</td></tr>
<tr><td>Protonemura praecox (Morton)</td></tr>
<tr><td>Brachyptera risi (Morton)</td></tr>
<tr><td>Isoperla grammatica (Poda)</td><td rowspan="19">5</td></tr>
<tr><td>Oligochaeta</td></tr>
<tr><td>Limnephilidae</td></tr>
<tr><td>Hydroptilidae</td></tr>
<tr><td>Oulimnius sp.</td></tr>
<tr><td>Crenobia alpina (Dana)</td></tr>
<tr><td>Sialis lutaria (L.)</td></tr>
<tr><td>Chironomidae</td></tr>
<tr><td>Cordulegaster boltonii (Donovan)</td></tr>
<tr><td>Chloroperla tripunctata (Scopoli)</td></tr>
<tr><td>Oecetis testacea (Curtis)</td></tr>
<tr><td>Rhyacophila dorsalis (Curtis)</td></tr>
<tr><td>Chloroperla torrentium (Pictet)</td></tr>
<tr><td>Simuliidae</td></tr>
<tr><td>Leuctra inermis Kempny</td></tr>
<tr><td>Polycentropodidae</td></tr>
<tr><td>Nemoura sp.</td></tr>
<tr><td>Amphinemura sulcicollis (Stephens)</td></tr>
<tr><td>Leuctra nigra (Olivier)</td></tr>
<tr><td>Phagocata vitta (Duges)</td><td rowspan="3">4</td></tr>
<tr><td>Protonemura meyeri (Pictet)</td></tr>
<tr><td>Leuctra hippopus (Kempny)</td></tr>
<tr><td>Leptophlebia marginata (L.)</td><td rowspan="2">3</td></tr>
<tr><td>Diura bicaudata (L.)</td></tr>
<tr><td>Nemurella picteti Klapalek</td><td>1</td></tr>
</table>

Taken together, the two explanatory variables mean pH and standard deviation of pH accounted for 4.6% of the total variation in the species data whilst all of the available environmental variables together accounted for 13.8%

3.9 Deriving the relationship between AWIC*(sp)*-ASPT and mean pH

The relationship between the AWIC*(sp)*-ASPT and mean pH was examined in the data used to derive the AWIC scoring system. As in the family level work, the relationship between mean pH was chosen, rather than standard deviation of pH (because mean pH was the more powerful explanatory variable and mean pH is a commonly calculated statistic in reporting chemical water quality. Linear regression analysis to derive the relationship between AWIC*(sp)*-ASPT versus mean pH gave the regression equation below:

Mean pH = 1.52 + 0.852 AWIC(*sp*)-ASPT **R^2 = 69.4%, *P*<0.001)**

Mean pH and AWIC *(sp)*-ASPT were strongly correlated, as expected, given that this is based on the data used to derive the scoring system. The regression plot with 95 percentiles is given in Figure 17.

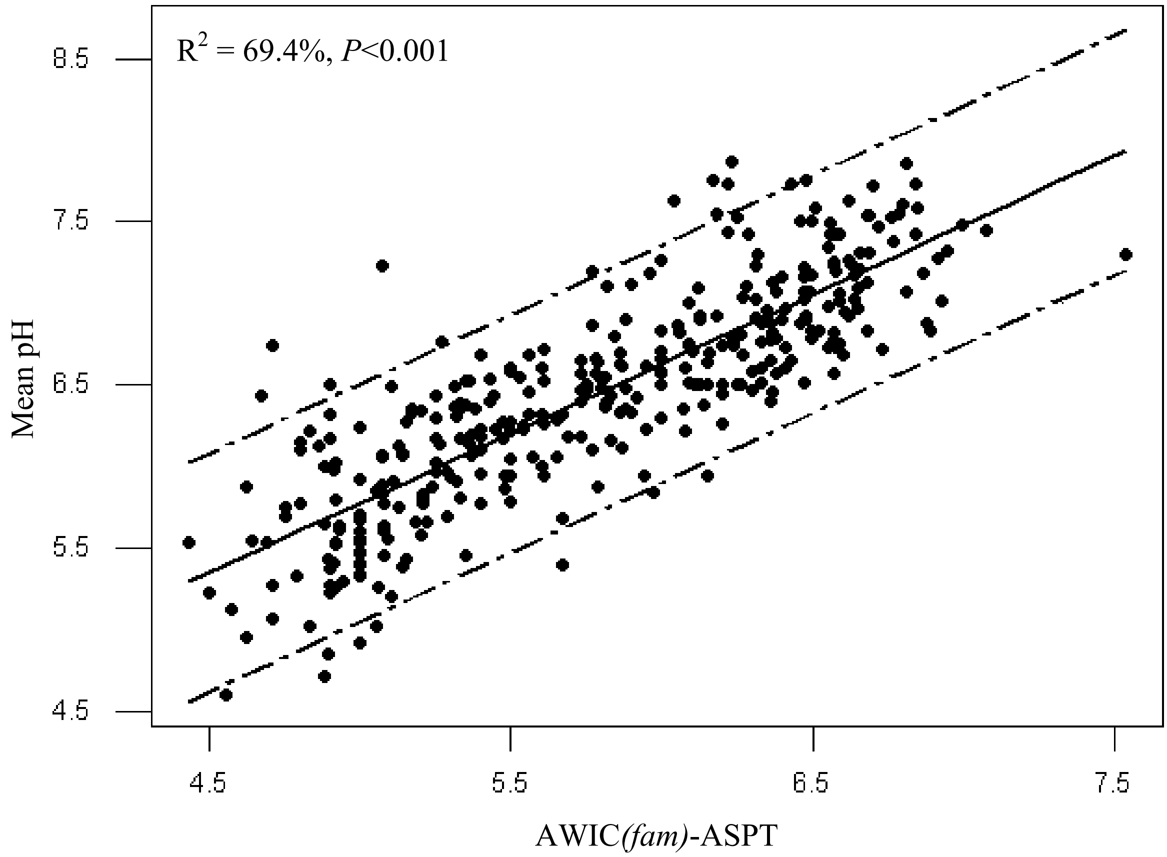


Figure 17. Regression plot of AWIC*(sp)*-ASPT versus mean pH with 95 percentile limits for the 353 samples used to derive the AWIC*(sp)* scoring system.

Predicted mean pH values for given AWIC*(sp)*-ASPT values are given in Table 12 together with 95 percentile limits (AWIC*(sp)*-ASPT values below 5.0 are not given because there were too few samples in this region).The 95 percentiles of predicted mean pH for a given AWIC*(sp)*-ASPT in Table 12 are quite wide, so an alternative model is presented below with 51 percentiles (Figure 18). The 51 percentile is an indication that it is more likely than not that a mean pH predicted from a given AWIC*(sp)*-ASPT lies between the 51 percentile limits.

Table 12. Predicted mean pH values for given AWIC*(sp)*-ASPT with 95 percentiles.

AWIC	Mean pH	Lower 95 Percentile	Upper 95 Percentile
5.0	5.78	5.05	6.51
5.5	6.20	5.47	6.93
6.0	6.63	5.90	7.36
6.5	7.05	6.33	7.79
7.0	7.48	6.75	8.27
7.5	7.91	7.17	8.64

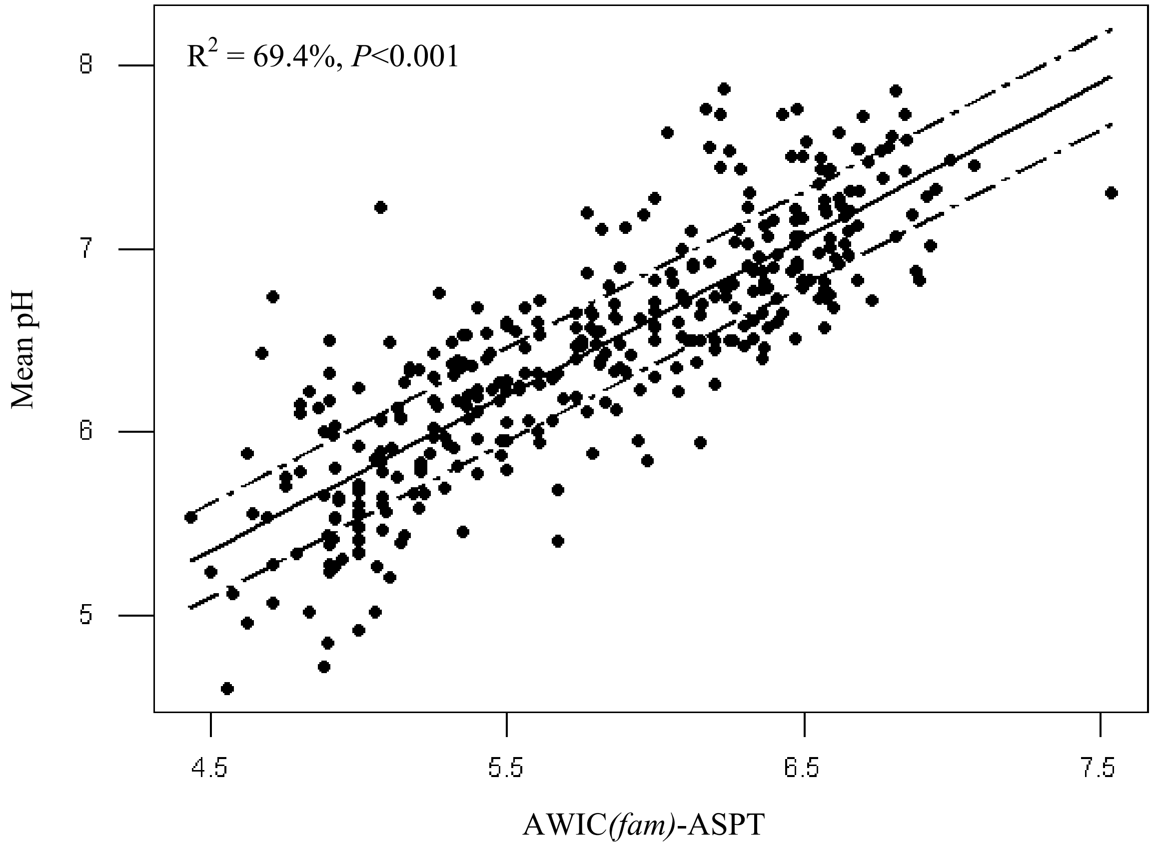


Figure 18. Regression plot of AWIC*(sp)*-ASPT versus mean pH with 51 percentile limits for the 353 samples used to derive the AWIC*(sp)* scoring system.

The predicted mean pH values for given AWIC*(sp)*-ASPT values, together with the 51 percentile limits, are given in Table 13 (AWIC*(sp)*-ASPT values below 4.5 are not given because there were too few samples in this region). To further illustrate the relationship between AWIC*(sp)*-ASPT and mean pH, the probabilities that each AWIC*(sp)*-ASPT class will have a certain mean pH are given in Table 14.

It is important to note that the predictions presented in section 3.4 are based on the relationship between AWIC*(sp)*-ASPT and mean pH in the data used to derive the scoring system itself and that these equations need to be tested on an independent data set to assess the validity of the model (section 3.10).

Table 13. Predicted mean pH values for given AWIC*(sp)*-ASPT with 51 percentiles.

AWIC	Mean pH	Lower 95 Percentile	Upper 95 Percentile
5.0	5.78	5.52	6.03
5.5	6.20	5.95	6.46
6.0	6.63	6.37	6.89
6.5	7.05	6.80	7.31
7.0	7.48	7.22	7.74
7.5	7.91	7.65	8.16

Table 14. The upper portion shows the number of samples contributing to the prediction and the lower portion shows the probability of that each AWIC*(sp)*-ASPT class will have a certain mean pH. Intervals shown are the upper class limits and values are rounded. AWIC*(sp)*-ASPT classes below 4.5 are excluded due to insufficient data.

Counts		AWIC*(sp)*-ASPT Class								Total
		4.5	5	5.5	6	6.5	7.0	7.5	8.0	
pH Class	5	-	6	-	-	-	-	-	-	6
	5.5	1	19	7	1	-	-	-	-	28
	6	1	22	28	6	1	-	-	-	58
	6.5	-	11	39	32	14	-	-	-	96
	7	-	1	7	24	43	16	-	-	91
	7.5	-	-	1	5	19	28	1	1	55
	8	-	-	-	-	8	11	-	-	19
	Total	2	59	82	68	85	55	1	1	353

Probabilities		AWIC*(sp)*-ASPT Class							
		4.5	5	5.5	6	6.5	7.0	7.5	8.0
pH Class	5	-	10.2	-	-	-	-	-	-
	5.5	50.0	32.2	8.5	1.5	-	-	-	-
	6	50.0	37.3	34.2	8.8	1.2	-	-	-
	6.5	-	18.7	47.6	47.1	16.5	-	-	-
	7	-	1.7	8.5	35.3	50.6	29.1	-	-
	7.5	-	-	1.2	7.4	22.4	50.9	100	100
	8	-	-	-	-	9.4	20.0	-	-

3.10 Testing the relationship between AWIC species and mean pH

To test the predictions of the AWIC(*sp*)-ASPT regression equation with mean pH we used an independent data set of 374 samples that had been collected for potential inclusion as RIVPACS reference sites. We calculated the AWIC(*sp*)-ASPT for each site and used the regression equation (section 3.9) to predict the 95 and 51 percentile ranges of pH at each site. We then took the actual mean pH of each site to see if this lay between the predicted percentile ranges. Observed mean pH was within the predicted 51%ile range on only 43.6% of occasions and within the predicted 95%ile range on only 90.4% of occasions. AWIC(*sp*)-ASPT was therefore less successful than AWIC(*fam*)-ASPT in predicting mean pH.

4. DISCUSSION

4.1 Review of the University of Wales system (the Rutt Key)

Following trials of the University of Wales System (Rutt *et al.*, 1990), comments from Environment Agency biologists were sorted into two broad categories (section 3.1). These are discussed below.

- The key failed to distinguish sites affected by metal mine drainage (and some other stress types such as domestic and industrial discharges, salinity and low flow) from those affected by acidification. It was felt that the key would be enhanced by the inclusion of some positive factors in order to distinguish sites affected by acidification rather than the sole reliance on a negative set of indicators.

The University of Wales System was regarded as being affected by stresses other than acidification. In this respect it is not unlike the BMWP scoring system because this is affected by some stresses which reduce BMWP scores and ASPT, and by acidification which also reduces BMWP score but elevates BMWP-ASPT. It would probably be extremely difficult to devise a scoring system or key that is specific to one stress type because most taxonomic groups are intolerant to a greater or lesser extent to many types of stress. The BMWP scoring system and University of Wales System respond well to organic and acidic stress respectively but expert interpretation of both index values is needed when they are applied in mixed stress situations or in streams with stresses other than organic pollution or acidification.

The heavy reliance on the BMWP scoring system (BMWP-ASPT) in assessing the biological quality of streams and rivers in Great Britain has resulted in the use of what is primarily an index designed to respond to organic pollution becoming used as general index of biological quality. However, the BMWP scoring system (and BMWP-ASPT), are not suitable for the assessment of heavily acidified sites because while ASPT is elevated in these cases, it is also high at organically unpolluted sites (Figure 12). There would be merit in moving away from sole reliance on BMWP indices and adoption of additional indices for site assessment. In the case of BMWP and the University of Wales System, a test site would need to attain satisfactory scores with both indices to be regarded as unstressed.

It was also noted that the University of Wales system contained few positive indicators of acidification. This weakness is difficult to address because few if any positive indicator taxa exist, especially at family level. At species level, there may be some taxa that respond positively to acidification stress, and the inclusion of abundances in the data used to develop the University of Wales System might have led to a more powerful indicator key. Indices of acidification based on macroinvertebrates have by necessity to be largely based on negative indicators because acidification tends to act mainly by removal of intolerant taxa.

- The key worked poorly in Regions that were beyond the scope of the data set from which it was developed. For example, in Anglian Region many of the taxa required to make an assessment are absent as a result of their natural distribution patterns.

The data set used to construct the University of Wales System was from Wales, N.W. England and Scotland. It is therefore unsurprising that the key was judged as less successful when applied to watercourses outside these areas. Additionally Rutt *et al.*, (1990) recommended that the key should only be used in streams consisting principally of pebbles and cobbles, and samples should be taken in riffles.

These limitations could only be solved by further analysis with a more geographically widespread data set, although a negative effect would be the loss of sensitivity in detection of acidification in Wales, N.W. England and Scotland because other physical gradients would become more significant in accounting for the variation in the taxon data over an increased geographical range. Looking again at the data used to construct the system, Rutt *et al.*, (1990) noted that sites in the Scottish data set were more physically diverse than those from Wales and N.W. England and that a lower proportion of the Scottish sites were acidified. This resulted in a less clear-cut relationship between the fauna and variables related to acidity in Scotland compared to Wales and N.W. England. The University of Wales system is therefore best suited to assessing the degree of acidification in streams in Wales and N.W. England, and is reasonably effective for use in Scotland. However, it is less suitable for use in rivers in East, Central and Southern England. It is likely that the University of Wales System would be effective in South West England because most of the taxa used in the key are found in this region.

It is interesting to contrast the University of Wales System, which has just 12 taxonomic groups, with the BMWP scoring system, which incorporates 82 'families'. BMWP achieves its geographically widespread applicability by incorporating a large number of families. Davy-Bowker *et al.*, (2000) categorised the BMWP families into 3 groups depending on geographical distribution; western upland families (12), eastern and southern lowland families (32) and rare or ubiquitous families (38). In the BMWP scoring system there are no regions of England or Wales where there are too few taxa to make an assessment of unpolluted sites.

In conclusion, it is important to appreciate that the University of Wales System was the first family level national indicator system proposed for use in Great Britain, unlike BMWP which represents the culmination of decades of refinement of organic pollution indices (e.g. Chesters 1980, Hawkes 1997). The University of Wales System is taxonomically compatible with the BMWP scoring system making it quicker and easier to use in both the laboratory and field than species level acidity indices (e.g Wade *et al.*, 1989). It was also widely acknowledged that the University of Wales system was scientifically sound.

4.2 The AWIC*(fam)* scoring system

Traditional analysis of the pH tolerance of freshwater macroinvertebrates, especially minimum pH tolerance, has been used in the development of several existing assessment methodologies (section 1.2.5, 1.2.6 and 1.2.7). Our analysis of pH tolerances (Figure 5) revealed very wide and overlapping pH ranges. We also considered the pH minimum values to be highly dependent on the frequency and timing of sampling. While the upper to lower quartile pH ranges were narrower and less susceptible to extreme samples, they overlapped with each other indicating that families at either extreme of pH tolerance (Capniidae and Haliplidae) could still, in theory be found in the same sample. Another potential pitfall in using pH tolerances alone to generate an index of acidity is that apparent pH tolerances of each family may be artefacts of other environmental factors. For example, while 75% of the Haliplidae records were from watercourses with pH above 6.8, this may be because Haliplidae have other physiochemical habitat preferences that limit their distribution from encompassing sites with low pH. In addition, Sandin & Johnson (2000) noted that pH metrics based on lower pH tolerance limits may have high power to detect improvement, because the colonisation of a site by just one pH-sensitive taxon will improve the score, but they may have a low power to detect deterioration because all of the sensitive taxa will have to be lost before the index will respond.

Using partial CCA to obtain a ranking of taxa in relation to pH has enabled us to produce a scoring system (Table 7) that should reflect the response of the families to pH (primarily mean pH) without the difficulties of the tolerance limit-based systems described above. Our regression plots of AWIC*(fam)*-ASPT versus mean pH (Figures 10 and 11) are fitted with linear regression equations that enables us to make testable predictions at both 95 and 51 percentile levels of confidence. Our rational for focusing on mean rather than standard deviation pH is justified both by the high correlation of axis 1 and mean pH in all of our pCCA analyses and by the difficulties in the integration of measures of pH variability in predictive models reported by other researchers. Whilst Ormerod *et al*. (1987) found that acid pulses had a direct toxic effect on invertebrates, Weatherley & Ormerod (1991) found that mean pH and aluminium were the most effective statistics in macroinvertebrate models and inclusion of variables such as pH minimum and aluminium maximum only gave moderate increases in precision at the expense of model complexity.

Close examination of the mean pH regression plot (Figure 9) indicates that there may be a slight tendency for AWIC*(fam)*-ASPT to underestimate the mean pH of very acidic sites. The mean pH predicted by the linear regression equation (section 3.4) for a given AWIC*(fam)*-ASPT, also had quite large 51 and 95 percentiles so that it was difficult to be precise about what a given AWIC(*fam*)-ASPT meant in terms of mean pH. Possible approaches to address these problems include reduction of the number of samples used to derive the relationship that have pH>7 (thereby reducing the influence of high pH sites on the linear regression equation), further adjustment to the individual taxon scores to reduce scatter in the relationship between AWIC(*fam*)-ASPT and mean pH (as in section 3.3.3) and examination of sigmoid (or other non-linear) relationships between AWIC*(fam)*-ASPT and mean pH. This might enable reductions in the confidence limits of predicted mean pH for a given AWIC*(fam)*-ASPT to be achieved. Despite these weaknesses, the mean pH of a partially independent data set (section 3.5) lay within the 51 and 95 percentiles of predicted mean pH calculated using AWIC(*fam*)-ASPT scores and the linear regression equation.

The correlations of AWIC*(fam)*-ASPT and BMWP-ASPT with mean pH (Figures 12 and 13) showed that AWIC*(fam)*-ASPT is more useful that BMWP-ASPT in indicating the acidity of a given site. AWIC*(fam)*-ASPT is highly correlated with mean pH. In contrast, whilst almost all sites with mean pH below 7 had BMWP-ASPT scores above 6, many unpolluted non-acidic sites also had BMWP-ASPT in the same range.

The geographical distribution of AWIC*(fam)*-ASPT scores (Figure 15) shows that sites in upland Wales, the north-west and south-west of England, together with the Ashdown Forest in the south east of England are identified as acidic. Note how the AWIC*(fam)*-ASPT is also successful in classifying the majority of the sites in the east, south and south east of England as non-acidified. The inclusion of families in the AWIC scoring system that are commonly found in the south and east of England has enabled AWIC to be applied in these areas with confidence. An approach of only including samples from potentially acidified areas in the data used to develop the index would have led to difficulties in the application of the scoring system outside of this geographical range.

Further improvement to the AWIC*(fam)* scoring system may be possible by the addition of more variables in the pCCA analysis to better isolate the effect of pH, and also by the investigation of weighted averaging techniques (section 1.2.7) to further refine the ranking of families.

4.3 The AWIC*(sp)* scoring system

Johnson *et al*. (1993) described the ideal attribute of an indicator species as its narrow and specific environmental tolerance. One of the weaknesses of the AWIC*(fam)* scoring system was the loss of pH discrimination where acid sensitive and acid tolerant species occur within the same family. It was hoped that the species level analysis would reveal more specific pH tolerances and hence a more precise ranking of taxa in the first axis of pCCA. Additionally, comparison of the ranking of taxa in the AWIC species scoring system with the same taxa in the pH tolerance limit system devised for Finland by Hämäläinen and Huttunen (1990) (section 1.2.7), revealed a very high correlation between the two approaches (correlation coefficient = 0.978).

Testing the AWIC(*sp*)-ASPT regression equation showed that AWIC(*sp*)-ASPT was less successful than the AWIC(*fam*)-ASPT in the predicting mean pH of an independent data set. AWIC(*sp*)-ASPT tended to predict mean pH values that were higher than expected for sites with an observed mean pH below 6.0. This may have been due to the low number of sites with mean pH less than 6.0 in the data set used to derive the AWIC species scoring system. The family level scoring system was also based on $\log_{10}$ abundance data while the species index was only based on presence/absence data. There are two potential avenues the improve the correlation of AWIC(*sp*)-ASPT with mean pH. Firstly, it may be possible to reduce scatter in the relationship between AWIC(*sp*)-ASPT and mean pH by adjustment of individual taxon scores (as was achieved with the AWIC(*fam*) scoring system - section 3.3.3). Secondly, and in addition to the first point, examination of non-linear relationships between AWIC*(fam)*-ASPT and mean pH might enable reductions in the confidence limits of predicted mean pH for a given AWIC*(sp)*-ASPT. These possibilities will be explored in delivering the last specific objective identified in section 1.1 (producing a paper for publication in a relevant scientific journal).

4.4 Proposals for Monitoring Guidelines

Proposals for acidification monitoring guidelines need to address statutory monitoring obligations, permissive monitoring programmes and special investigations.

There is currently no statutory requirement for the Environment Agency to monitor acidification, although this will change in the next few years with the advent of the EU Water Framework Directive (European Commission, 2000). The Water Framework Directive requires that management plans are formulated for each catchment to ensure that watercourses achieve ‘good’ ecological status in terms of their biology, hydromorphology and physicochemistry (which includes acidification status). While chemical monitoring of pH can be used to assess acidity directly, appropriate biological indices will be needed to determine the extent to which acidity contributes to failure in achieving ‘good’ ecological status. It is proposed that the AWIC*(fam)* scoring system may be suitable for this role and that AWIC*(fam)* scores should be reported (and subsequently evaluated) in statutory monitoring programmes as a precursor to possible use in Water Framework Directive reporting of acidic stress in England and Wales.

It is proposed that permissive monitoring programmes carried out in areas of England and Wales that are susceptible to acidification, should also make use of the AWIC*(fam)* scoring system to assess acidity.

In addition, it is proposed that the AWIC*(fam)* and AWIC*(sp)* indices should be applied in special investigations of potential acidification such as the assessment of afforestation and clear felling.

To make the AWIC*(fam)* and AWIC*(sp)* scoring systems available to Environment Agency biologists it is proposed that both scoring systems should be routinely calculated by the Environment Agency's new data archive and reporting database, 'Biology for Windows'.

CEH intend to incorporate the AWIC*(fam)* and AWIC*(sp)* indices into the next release of RIVPACS, which is anticipated to be a Windows version (Wright *et al*. 2002), so that it will be possible to examine the extent of deviation in observed AWIC scores from RIVPACS predicted values. This is an important precursor to the development of a suitable grading system for AWIC(fam) and AWIC(sp) EQI values.

5. RECOMMENDATIONS

The following recommendations are made to the Environment Agency:

- **In readiness for forthcoming statutory obligations under the Water Framework Directive, the usefulness of AWIC*(fam)* and AWIC*(sp)* scoring systems should be evaluated in Environment Agency statutory monitoring programmes.**

- **Permissive monitoring programmes in acid susceptible areas of England and Wales should trial the AWIC*(fam)* scoring system.**

- **Special investigations of potential acidification issues (e.g. forestation and clear felling) should incorporate the AWIC*(fam)* and AWIC*(sp)* scoring systems.**

- **The facility to calculate AWIC*(fam)* and AWIC*(sp)* scoring systems should be made available in the Environment Agency Biology for Windows database.**

CEH shall undertake the following:

- **AWIC*(fam)* and AWIC*(sp)* scoring systems will be incorporated into the next (Windows)* version of RIVPACS.**

*CEH will incorporate the AWIC scoring systems into the next version of the RIVPACS software as soon as funding becomes available.

LIST OF FIGURES

LIST OF TABLES

REFERENCES

Chesters R.K. (1980). Biological Monitoring Working Party. The 1978 national testing exercise. Department of the Environment, Water Data Unit Technical Memorandum 19: 1-37.

Crawshaw D.H., Diamond M., Prigg R.F., Cragg-Hine D. and Robinson J.F. (1989). Acidification of surface waters in Cumbria and the South Pennines. North West Water, final report to the Department of the Environment.

Davy-Bowker J., Furse M.T., Clarke R.T. & Gravelle M.J. (2000). Analysis of 1995 survey data. Phase 2. Post – survey appraisal. Unit I: Taxon Distribution Studies. Environment Agency R&D Technical Report E103.

Davy-Bowker J., Furse M.T., Wright J.F. & Clarke R.T. (2002). Development of a dynamic model to predict the biological consequences of changes in water quality. Testing and further development of RIVPACS, Stage 5 report. Environment Agency R&D Technical Report E1-007/TR2.

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

Feldman R.S.& Connor E.F. (1992). The relationship between pH and community structure of invertebrates in streams of the Shenandoah National Park, Virginia, U.S.A. *Freshwater Biology*, 27: 261-276.

Fjellheim A. & Raddum G.G. (1990). Acid precipitation: biological monitoring of streams and lakes. *Science of the Total Environment*, 96: 57-66.

Furse M.T. & Symes K.L. (1998). Monitoring of Acid Waters. Phase 1. Indicator Populations. R&D Progress Report P2/090/1 for the period 1st October 1997 to 31st December 1997. Institute of Freshwater Ecology, Wareham, UK.

Furse M.T. (1998). Monitoring of Acid Waters. Phase 1. Indicator Populations. R&D Progress Report P2/090/2 for the period 1st January 1998 to 31st March 1998. Institute of Freshwater Ecology, Wareham, UK.

Furse, M.T., Davy-Bowker J., Gunn R.J.M., Blackburn J.H., Winder J.M., Wiggers R., Wittenoom P., Cooper G., Hardie D., Boyden V. & Kneebone N.T. (2000). Monitoring of Acid Waters. Phase 1. Indicator Populations. R&D Progress Report P2/090/. Centre for Ecology & Hydrology, Dorchester, UK.

Hämäläinen H. & Huttunen P. (1990). Estimation of Acidity in streams by means of benthic invertebrates: evaluation of two methods. In: *Acidification in Finland* (Eds. P. Kauppi, P. Anttila & K. Kenttämies), pp. 1051-1070. Springer, Berlin.

Hämäläinen H. & Huttunen P. (1996). Inferring the minimum pH of streams from macroinvertebrates using weighted averaging regression and calibration. *Freshwater Biology*, 36: 697-709.

Hämäläinen H. & Huttunen P. (1998). Macroinvertebrate-inferred stream-water acidity in north eastern Finland along a sulphur deposition gradient. *Water Air and Soil Pollution*. 104(3-4): 223-236.

Hawkes H.A. (1997). Origin and development of the Biological Monitoring Working Party score system. *Water Research* 32: 964-968.

Hill M.O. (1979). *TWINSPAN – a FORTRAN programme for Detrended Correspondence Analysis and Reciprocal Averaging*. Ecology and Systematics, Ithaca, New York.

Johnson R.K., Wiederholm T. & Rosenberg D.M. (1993). Freshwater biomonitoring using individual organisms, populations and species assemblages of benthic macroinvertebrates. In: Rosenberg D.M & Resh V.H. (eds.) *Freshwater Biomonitoring and Benthic Macroinvertebrates*. Chapman Hall, New York, 488 pp.

Larsen J., Birks H.J.B., Raddum G.G. & Fjellheim A. (1996). Quantitative relationships of invertebrates to pH in Norwegian river systems. *Hydrobiologia*, 328: 57-74.

Mackay R.J. & Kersey K.E. (1985). A preliminary study of aquatic insect communities and leaf decomposition in acid streams near Dorset, Ontario. *Hydrobiologia*, 122: 3-11.

Magalhães M.F., Batalha D.C. & Collares-Pereira M.J. (2002). Gradients in stream fish assemblages across a Mediterranean landscape: contributions of environmental factors and spatial structure. *Freshwater Biology*, 47: 1015-1031.

NIWA (1991). International Co-operative Programme on Assessment and Monitoring of Acidification of Rivers and Lakes: Program Manual. Norwegian Institute for Water Research (NIWA), Oslo, Norway.

Ormerod S.J., Boole P., McCahon C.P., Weatherley N.S., Pascoe D. & Edwards R.W. (1987). Short-term experimental acidification of a Welsh stream: comparing the biological effects of hydrogen ions and aluminium. *Freshwater Biology*, 17: 341-356.

Palmer M.W. (1993). Putting things in even better order: the advantages of canonical correspondence analysis. *Ecology* 74:2215-30.

Patterson G. & Morrison B. (1993*).* Invertebrate Animals as Indicators of Acidity in Upland Streams*.* Forestry Commission Field Book 13.

Raddum G.G. & Fjellheim A. (1984). Acidification and early warning organisms in freshwater in western Norway. *Verh. Internat. Verein. Limnol*. 22: 1973-1980.

Raddum G.G., Fjellheim A. & Hesthagen T. (1988). Monitoring of acidification by the use of aquatic organisms. *Verh. Internat. Verein. Limnol*. 23: 2291-2297.

Rosemond A.D., Reice, S.R., Elwood J.W. & Mulholland P.J. (1992). The effects of stream acidity on benthic invertebrate communities in the south-eastern United States. *Freshwater Biology*, 27: 193-209.

Rutt G.P., Weatherley N.S. & Ormerod S.J. (1990). Relationships between physiochemistry and macroinvertebrates in British upland streams: the development of modelling and indicator systems for predicting fauna and detecting acidity. *Freshwater Biology*, 24: 463-480.

Sandin L. & Johnson R.K. (2000). The statistical power of selected indicator metrics using macroinvertebrates for assessing acidification and eutrophication of running waters. *Hydrobiologia*, 422/423: 233-243.

Smith M.E., Wyskowski B.J., Brooks C.M., Driscoll C.T & Cosentini C.C. (1990). Relationships between acidity and benthic invertebrates of low order woodland streams in the Adirondack Mountains, New York. *Canadian Journal of Fisheries and Aquatic Sciences*, 47: 1318-1329.

Stevens P.A., Ormerod S.J. & Reynolds B. (1997). Final report on the acid waters survey for Wales, Volume 1 Main Text. Institute of Terrestrial Ecology, Bangor, UK.

ter Braak C.J.F. (1987). Ordination. P. 91-173 In: Jongman, R.H., ter Braak C.J.F & van Tongeren O.F.R. (eds.) *Data Analysis in Community Ecology*. Pudoc, Wageningen, The Netherlands.

ter Braak C.J.F. & Prentice I.C. (1988). A theory of gradient analysis. *Adv. Ecol. Res*. 18:271-313.

ter Braak C.J.F. & Šmilauer P. (1998). *Canoco Reference Manual and User's Guide to Canoco for Windows: Software for Canonical Community Ordination Version 4.0.* Microcomputer Power, Ithaca, New York.

ter Braak C.J.F. & Verdonschot P.F.M (1995). Canonical correspondence analysis and related multivariate methods in aquatic ecology. *Aquatic Sciences* 57:255-289.

Townsend C.R., Hildrew A.G. & Francis J. (1983). Community structure in some southern English streams: the influence of physiochemical factors. *Freshwater Biology*, 13: 521-544.

Wade K.R., Ormerod S.J. & Gee A.S. (1989). Classification and ordination of macroinvertebrate assemblages to predict stream acidity in upland Wales. *Hydrobiologia*, 171: 59-78.

Weatherley N.S. & Ormerod S.J. (1991). The importance of acid episodes in determining faunal distribution in Welsh streams. *Freshwater Biology*, 25: 71-84.

Wright J.F., Moss D., Armitage P.D. & Furse M.T. (1984). A preliminary classification of running water sites in Great Britain based on macroinvertebrate species and the prediction of community type using environmental data. *Freshwater Biology*, 14: 221-256.

Wright J.F., Winder J.M., Clarke R.T. & Davy-Bowker J (2002). Testing and Further Development of RIVPACS. Stage 3 Report. R&D Technical Report E1-007/TR to the Environment Agency.