Acid Waters macroInvertebrate Status Tool

Draft Final Report

6 May 2009

Contributors:

Gavin L. Simpson
Simon Turner
Stephen Brooks
Malcolm Greenwood
Hong Yang
Don Monteith
Simon Patrick

Executive Summary

WFD60: Macroinvertebrate diagnostic tool development (May 2009)

Project funders/partners: SEPA and SNIFFER

Background to research

This project (WFD60) forms part of the UK Strategy for the implementation of the EC Water Framework Directive (WFD: European Union, 2000). Within its broad remit the WFD requires the development of ecological classification tools for the purpose of determining ecological status with reference to specific environmental pressures. The WFD requires that these tools should assign lakes to one of five categories (High, Good, Moderate, Poor, Bad) to indicate conditions relative to what is considered "high status". This report focuses on the further development of a tool with which to determine the extent of the pressure of acidification on lake macroinvertebrate communities.

Objectives of research

The primary objective is the development of a method and tool with which to assess the pressure of acidification (a major threat to the ecology of acid-sensitive freshwaters, particularly in the UK uplands) on the benthic macroinvertebrate assemblage of lakes. This phase of the WFD60 project focuses on the development of a new, improved version of the original WFD60 classification tool made possible by a programme of new data collection.

Key findings and recommendations

Development of the new WFD60 tool was based closely on the observations from and the results of the first phase of the WFD60 project (Monteith and Simpson 2007). Briefly, knowledge on lake acidification status was combined into a simple measure of acidification damage based upon acid neutralising capacity (ANC) and lake water Calcium concentration (Ca^{2+}), and a simple classification scheme known as the "damage matrix" was produced. This damage matrix represents our current understanding of ecosystem damage resulting from acidification, and is used to provide an assessment of the true acidification status of lakes in the WFD60 data set.

We then employed Random Forests, a sophisticated statistical data mining methodology to build an ensemble (collection) of classification trees that contained decision rules that attempt to predict the a prior assigned damage matrix classification using only the macro species data. We supplement the raw macroinvertebrate data with "meta taxa"; aggregations of subsets of the species data that reflect aspects of the community such as total species richness or richness of important acid tolerant or acid intolerant indicator groups.

We show that the random forest approach can produce a classification tool that is able to correctly predict the damage matrix class for 99% of the samples in the WFD60 data set, incorrectly assigning the wrong class to a single observation. This represents the apparent performance of the tool and will over-estimate the expected performance when the tool is applied to new samples that have not taken part in the model building. Cross-validated performance estimates suggest that the correct class is assigned 50% of the time. Whilst this figure appears low, it is twice as good as randomly guessing the class and is comparable to other WFD tools developed for acidification status (LAMM and CPET). However, due to uneven

sampling of the four status classes predicted by the new tool we believe this performance estimate to be an under-estimate due to problems performing the bootstrap sampling in cases where some classes contain relatively few samples. We believe the true performance of the tool lies somewhere between the apparent and cross-validated error and as such the new tool performs favourably compared to LAMM and much better than CPET for the subset of samples to which the three tools have been applied.

A major addition to the new tool is the calculation of sample EQRs from the output form the random forest. The four classes predicted by the new tool are assigned a base score. The final EQR for an individual sample or site is computed as a weighted average of these base scores, with weights for each base score given by the probability that the sample belong to that status class. We show that this EQR performs favourably compared to WFD normative definitions and measures of acidification damage. Furthermore, we show that the EQR contains an element of uncertainty in the assigned classification; a site that could just as likely be assigned Good status or Moderate status should have a lower EQR than a site that is clearly in Good status, even if both sites are ultimately assigned to Good status on the basis of majority votes from the random forest.

A further development of the new tool has allowed sample-specific estimates of prediction uncertainty or confidence of class to be determined. The confidence of class is taken as the proportion of votes for the assigned class out of the total votes for all classes for individual samples. This measure is computed for all four classes. As such, to simplify the confidence of class information, we use Shannon's entropy measure to combine the four confidence of class measures. A low value (close to 0) of Shannon's entropy indicates high confidence of class and a high value (close to 1) indicates low confidence of class.

Several assessments of the new tool have been performed as part of this phase of WFD60. The first assessment looked at performance of the tool against temporal data from the lake sites of the UK Acid Waters Monitoring Network (UK AWMN). The second assessment compared the results from AWIST with data from palaeoecological studies at two acidified sites and one minimally impacted site. Despite complications arising from the nature of the individual data (large temporal variability in the macroinvertebrate time series data, and incomplete preservation of the macroinvertebrate community in the palaeoecological data), the results from applying AWIST to the test sites show that the new tool can track recovery in the macroinvertebrate community through time in several UK AWMN sites and that the tool predicts acidification status that reflects the acidification history of the two acidified test lakes as demonstrated by diatom and macrofossil remains.

In conclusion, the new tool developed as part of the second phase of WFD60 is capable of providing relevant information for WFD purposes and that AWIST is able to capture and reflect known acidification status and track temporal recovery in macroinvertebrate assemblages where this is occuring in UK AWMN sites.

Additional work should be performed to investigate whether and how LAMM, CPET and AWIST predictions can be combined to provide an ensemble classification tool that draws upon the best features of each of the tools. This will require further collaborative work between the tool developers and the relevant agencies to conduct relevant comparisons and testing. This additional work may allow time to tweak the underlying AWIST random forest to improve the cross-validation performance statistics by performing stratified bootstrap resampling within with random forest algorithm.

Table of Contents

1 Introduction

The WFD60 project forms part of the UK Strategy for the implementation of the EC Water Framework Directive (WFD; European Union, 2000). Within this broad remit the WFD requires the development of ecological classification tools for the purposes of determining ecological status, with reference to specific environmental pressures. Our primary objective is the development of a method and tool with which to assess the pressure of acidification (a major threat to the ecology of acid-sensitive fresh waters – particularly in the UK uplands) on the benthic macroinvertebrate assemblage of lakes.

The WFD requires that tools should assign sites, here lakes, to one of five categories (High, Good, Moderate, Poor and Bad) relating to WFD normative definitions, to indicate conditions relative those considered to be "high status". Inherent errors are to be defined and quality assurance provided, such as the confidence of assigned classification. The tool is provided as a simple application that allows the funding responsible agencies to apply the tool to further datasets.

The first phase of the WFD60 project produced a tool based on classification tree methods, which was able to predict three WFD classes (High-Good, Moderate, and Poor-Bad) and used only a series of meta taxa, aggregated measures of the original species data into indicative species groups. There were insufficient samples in the original data set (105) to identify rules that could predict the five status classes individually. However the tool was thought not ready for use and a further round of data collection and tool development was recommended. This report describes this additional developmental work.

Ca^{2+} (µeq/l)	ANC group (µeq/l)								
	100-80	80-60	60-40	40-20	20-10	10-0	0-(-10)	-10-(-20)	-20-(-40)
0-20	H	H	H	H	G	M	P	B	B
20-40	H	H	H	G	M	P	B	B	B
40-60	H	H	H	G	M	P	B	B	B
60-80	H	H	H	M					
80-100	H	H	G	M					

Figure 1: Damage matrix, based on understanding of relationships between ANC and calcium concentrations and evidence from palaeoecological and hydrochemical models of acidification, and contemporary relationships with Al_{lab} and macroinvertebrate assemblage characteristics. Letters represent expert judgement on likely ecological status with respect to damage from acidification. H = High, G = Good, M = Moderate, P = Poor and B = Bad.

The background to the current study is presented in detail in Monteith & Simpson (2007). In this report we describe the third phase of tool development and how the underlying statistical machine learning methodology is used to develop a WFD-compliant tool. We do not repeat material from the report on the first phase of tool development (Monteith & Simpson, 2007), except where such information is necessary to inform the current discussion. Readers are refered to the earlier report.

1.1   Damage Matrix

The first phase of tool development under WFD60 described the generation of a "damage matrix". The damage matrix encapsulates our understandings of the relationships between current ANC and calcium concentrations and relates this to mobilisation of inorganic monomeric aluminium (labile aluminium, Al_{lab}) concentrations, predictions of the extent of acidification (from critical loads and palaeoecological studies) and our understanding of critical biological thresholds. The damage matrix is split into the five WFD status classes (High, Good, Moderate, Poor and Bad) on the basis of the above criteria. The damage matrix, therefore, represents the current expert knowledge on state change over reference conditions, but expressed through two simple hydrochemical properties (ANC and calcium concentration), which allows each site in the WFD60 training set to be assigned an a priori WFD status class. It is these classes that we aim to predict through the application of decision trees using the macro-invertebrate species data only. The damage matrix is shown in Figure 1.

In the first phase of the WFD60 project a classification tree, a statistical machine learning tool was employed to identify simple decision rules that allowed the prediction of class status for each site in the training set, assigned from the damage matrix. Here we retain classification trees to produce the decision rules, but extend the single tree of the first phase to a forest of trees, fit to bootstrap samples of the training, to produce a tool that is more consistent and robust with greater predictive ability. In addition, we make use of an expanded data set of macroinvertebrate counts and high-quality chemistry.

1.2   Structure of the report

First, we describe the approach taken to develop a new tool for WFD60 based upon classification trees and a new technique known as Random Forests (Breiman, 2001). We discuss how random forests work and how the output from the computations is made WFD compliant through the calculation of ecological quality ratios (EQRs) and confidence of class. We present results of applying the tool to the WFD60 macroinvertebrate dataset and subsequently compare the results of the tool with other tools developed from different biological groups.

We then describe the results of a palaeoecological study of two acidified and one acid-sensitive, yet un-acidified, lakes into changes in macroinvertebrate assemblages and relate these changes to damage status predicted for these three sites from the present-day macroinvertebrate species assemblage.

Finally, we discuss the results of the current phase of the work programme and present conclusions and recommendations for future work.

2 Tool Development

The first phase of the WFD60 project utilised a classification tree to identify decision rules that best predicted *a priori* defined WFD classes (High, Good, Moderate, Poor, Bad). These classes were assigned to each site in the WFD60 training set through the use of the "damage matrix". The damage matrix

combined acid neutralising capacity and calcium concentrations to apportion sites into one of the five WFD classes. The construction of the damage matrix was informed from critical load methodologies and palaeoecological information on state change over reference conditions in relation to contemporary water chemistry. The project report for the first phase of WFD60 (Monteith and Simpson, 2007) contains detailed descriptions of the construction of the damage matrix and the use of classification trees.

Following additional sample collection in Phase II of WFD60, a new tool has been constructed, fitted to a combined training set of existing macro-invertebrate data and newly collected samples. The enlarged training set allows a newly-developed statistical machine learning technique, random forests, to be used to improve the predictions arising from the development of a single tree through the creation of a forest a decision trees.

2.1 Design Principles

We first repeat the design principles behind development of the WFD60 tool presented in the phase I report. These design decisions remain relevant to the new tool described in this report.

2.1.1 Other WFD Schemes under development

Several WFD classification tools are currently being developed in the UK under the supervision of UKTAG. The design methodology for most of these is similar and involves four main steps.

1. Reference sites (i.e. a sub-set of those assumed to be of High status) are identified for each typology, using "expert judgement".

2. Biological assemblages in a "training set", including those for reference sites, are related to a physico-chemical pressure gradient, such as phosphorus concentration, using multivariate ordination methods such as canonical correspondence analysis (CCA).

3. Sample scores derived from the ordination procedure are divided by sample scores for reference lakes within the same typology to provide an ecological quality ratio or EQR for each lake.

4. EQRs are then related to biological normative definitions, such as the relationship between stress tolerant and intolerant species, and this is used to divide up the gradient into the five WFD classes introduced in Section 1.

Class membership is then subjected to uncertainty analysis to ascertain the likelihood that a biological sample will be allocated the appropriate damage class, given the known susceptibility of the sample data to variability in sampling effort, time, space, etc..

Thus EQR derivation follows WFD guidance outlined in Annex 5 in a very literary manner, i.e.:

"*...the results …...shall be expressed as ecological quality ratios for the purposes of classification of ecological status. These ratios shall represent the relationship between the values of the biological parameters observed for a given body of surface water and the values for these parameters in the reference conditions applicable to that body.*"

However, there is no explicit requirement in the Directive that an EQR must be calculated mathematically by dividing sample scores in the way outlined above. We argue that this clause may alternatively be

interpreted as a qualitative requirement that the EQR must be based on comparisons of biological condition of a site with what is considered to be reference condition. If certain biological characteristics of high status can be considered universal, then any deviation from these characteristics may be used to derive an EQR score.

We argue that the commonly adopted procedure is prone to *a priori* uncertainties which are not subjected to rigorous analysis. First, no two lake ecosystems are identical now or in the past, however WFD "reference lakes" tend to be few. Uncertainties arise immediately, therefore, with regard to representativity of these lakes and hence the relative level of damage for lakes within the same typology for which the same reference condition is used. While a continuous EQR score is generated there are no grounds to believe a lake which has been allocated an EQR of, for example, 0.5 is less damaged than one with an EQR of 0.4, even within the same typology. The apparent continuity of the score, therefore, has limited ecological merit with respect to damage assessment.

Second, the procedure by which the EQR gradient is divided into damage classes is often based on subjective criteria such as the cross-over point between two biological indicator classes (themselves defined by relating to "pressure gradients" used in previous steps), rather than a mechanistic understanding of how the pressure is likely to influence the biological community.

Finally, this cross-over point, whilst being the optimal discriminator between the two groups of biological indicators, may not be an optimal decision threshold for discriminating between site damage classes. Again, the use of the cross-over point to fix the Good-Moderate boundary is not required by Annex V of the Directive and we argue that the widespread use of this criterion for setting this important boundary results from an overly prescriptive adoption of statements in the WFD. The cross-over point is used solely to describe what the Commission meant by "moderate" change over reference, not that this will necessarily be adopted as part of member states classification schemes. Overall, therefore, we feel this process is self-referential and of restricted ecological validity.

2.1.2   Classification under WFD60

We have proposed an alternative approach to lake classification to that discussed above. This is based on the following observations outlined in previous sections of this report:

1. In contrast to other pressures of concern to the WFD, lakes of "reference condition" may be particularly difficult to identify within the same biogeographic region; the most appropriate "reference conditions" for acidified lakes in north Wales, the Pennines and the English Lake District, may only be found in un-acidified parts of the far north-west of Scotland, but these may not be sufficiently analogous to lakes much further south for climatic, geological and biogeographic reasons;

2. unlike other pressures of concern to the WFD the pressure of acidification can be predicted from current physico-chemistry;

3. high status, according to physico-chemical normative definitions, can be identified with some confidence on the basis of ANC;

4. a physico-chemical "good-moderate" boundary may be defined according to our understanding of the importance of Ca^{2+} in determining the likely ANC threshold for biological damage through acidification (based on physico-chemical and palaeoecological models); this threshold can be

considered to be the point at which biological communities begin to differ "moderately" from reference and where major taxonomic groups are first likely to disappear, according to WFD normative definitions. The relationship between ANC and Ca^{2+} determines the likelihood of acidification, not biological damage;

5. "poor" to "bad" status may also be defined on the basis of biological toxicity thresholds for aluminium, and relationships between Al_{lab} and ANC. Our data demonstrate that many species and taxonomic groups typical of acid sensitive lakes are excluded from lakes with an ANC below zero and the dominant reason for this is likely to be due to the coincidence of this threshold with substantially elevated levels of Al_{lab} in addition to low pH.

We are confident, therefore, that we can classify lakes in our training set using physico-chemistry in a manner that accords with biological normative definitions and evidence of the degree of departure of biological communities from reference state. In this report we go on to show how such classes can be predicted by the macroinvertebrate community, and how this can then be used as the basis for the WFD60 tool.

2.1.3   WFD60 proposed classification approach

Rather than generate EQRs for individual sites using assumptions on appropriate reference sites, and then attempt to divide the EQR gradient according to normative definitions into the five WFD classes, we set out to test whether it was possible to use our understanding of the relationship between physico-chemical indicators of damage and aquatic biology to derive *a priori* EQR compatible classes for sites, and then use a classification approach to predict membership of a class based on the macroinvertebrate characteristics of each site.

For example, according to our preliminary analysis of the data and information available in the scientific literature, it is reasonable to assume that the macroinvertebrate community of sites with an annual mean ANC of >60 µeq l^{-1} are unlikely to differ significantly from reference condition (with respect to acidification). Providing there are no other major environmental constraints (and out training set was designed to avoid these wherever possible): taxonomic composition should correspond totally or nearly totally to un-acidified conditions; there should be no sign of alteration in the ratio of sensitive to insensitive taxa; and, there should be no sign of any reduction in diversity from that found in acidified sites. The EQR of such a site should approach a value of 1, and could, for the sake of convenience, be allocated a score of 0.9.

Conversely, the macroinvertebrate community of any site with an ANC of < -50 µeq l^{-1}, is highly likely to exhibit very low pH and highly toxic concentrations of Al_{lab}. Such a site is highly likely to support a very limited number of highly acid-tolerant taxa only and will deviate profoundly from reference condition with respect to taxonomic composition, abundance, ratios of sensitive to insensitive taxa and diversity. Its EQR must therefore approach zero and could again, for convenience, be allocated a score of 0.1. If we can derive classification rules which can determine the likelihood of a biological assemblage falling into these classes then we have the basis for a robust classification system which is compatible with WFD requirements.

The damage matrix arises naturally from these design decisions and expert knowledge gleaned from numerous sources within the broad field of acidification effects on lake ecosystems. We aim, therefore, to use statistical machine learning tools designed to be good classifiers to predict the *a priori* assigned classes for each site in the WFD60 training set.

2.2   Model Fitting

To improve upon the original WFD60 tool Random Forests, a sophisticated machine learning algorithm was employed to classify sites into WFD status classes using features of the macro-invertebrate data. Additional sites were sampled for macro-invertebrates and hydrochemistry and these were added to the data set available for analysis. Here we describe the data processing and analysis steps involved in producing the AWIST tool.

2.2.1   WFD60 Dataset

In addition to the sites used in the first phase of the WFD60 project, newly sampled sites were added to the data set used to produce the AWIST tool. Details of these sites are shown in Table 1, below.

WBID	Site name	Alt (m)	Area (ha)	Grid ref
Galloway				
27827	Loch Dungeon	308	35.4	NX524844
27777	Loch Harrow	248	13.7	NX527866
27699	Loch Macaterick	284	74.4	NX440914
27801	Loch Minnoch	271	6.6	NX530857
27693	Loch Fannie	296	1.2	NX446924
27823	Loch Dow (by Moan)	226	1.3	NX353848
28370	Plantain Loch	27	3.7	NX841601
27923	Loch Dow (by Round)	478	0.5	NX457807
28076	Loch Gower	231	0.6	NX549735
27888	Loch Dow (by Narroch)	498	0.5	NX461825
Lake District				
28905	Scales Tarn	598	1.2	NY328281
290811	Bleaberry Tarn	497	1.1	NY165154
29179	Angle Tarn	568	3.5	NY244077
29181	Greendale Tarn	407	1.8	NY147074
29181	Eel Tarn	144	1.6	NY189019
29290	Seathwaite Tarn	374	24.1	SD253987
29285	Levers Water	413	13.6	SD279993
29045	Floutern Tarn	377	1.3	NY124171
29246	Stoney Tarn	298	1.4	NY200025
29008	High Nook Tarn	217	0.6	NY124199
Wales (in current EA survey				
33974	Llyn Cwmffynnon	385	10.7	SH649562
35180	Llyn Cwm Bychan	158	13.6	SH640312

35426	Llyn Hywel	539	5.1	SH663266
38240	Llyn Fyrddon Fawr	519	12.0	SN800707
38282	Llyn Cerrigllwydion Isaf	498	6.0	SN843699
37834	Llynnoedd Ieuan	529	5.6	SN795814
Wales Snowdonia				
34668	Llyn Mair	79	6.1	SH652412
34635	Llyn y Garnedd	219	2.1	SH656419
34665	Llyn Hafod-y-llyn	128	1.9	SH645413
34375	Llyn Cwm-corsiog	533	3.2	SH663470
34382	Llyn Cwm-y-foel	440	2.7	SH655468
34387	Llyn Bowydd	478	7.3	SH725467
34372	Llyn Newydd	478	5.7	SH722471
34632	Llyn y Dywarchen	503	2.8	SH762420
34039	Llyn Teyrn	375	1.9	SH641547
34249	Llyn Cwm Dulyn	238	13.4	SH492495
34061	Llyn Nadroedd	539	1.0	SH594543

Table 1: Additional sites analysed as part of the new data collection in this phase of the WFD60 project

2.2.2   Data processing

Following preliminary data analysis, despite the additional data collection exercise that formed Phase II of the WFD60 project, a paucity of sites in the Poor and Bad classes remained and as such there was insufficient numbers of sites within the two classes individually to allow the random forest technique to be applied consistently. As such, we merged the Poor and Bad classes to form a single "Poor-Bad" class. This merging has no implications for the use of the tool in WFD settings as the boundary between these two classes does not require formal definition for WFD purposes. The resulting number of observations within the remaining four classes are given in Table 2.

High	Good	Moderate	Poor-Bad
55	36	30	14

Table 2: Number of samples within each WFD status class

Unlike Phase I, the additional data collection, coupled with the more sophisticated nature of the random forest method, we were able to retain separately the High and Good classes. Consequently, the classification tool presented below allowed prediction of four WFD status classes.

Macroinvertebrate data were processed for subsequent analysis in as follows:

Total abundances were first calculated for each Mixed Taxon Level taxa for each sample date for each site. The bulked data thus represented differing sample sizes depending on the origin of the samples. In

order to account for this, species data were converted to proportions of the full species count.

In addition, the following summary data were collated for each bulked sample:

1) The minimum possible number of species present (described as minimum species richness or MSR; i.e. all species plus genus and family level groups where there are no higher order members present) within the following groups);

- The whole assemblage
- Trichoptera
- Ephemeroptera
- Ephemeroptera not of the family Leptophlebiidae (known to be an acid tolerant family)
- Plecoptera
- Odonata
- Hemiptera
- Coleoptera

2) Total number of individuals within the following groups (identified to any level) expressed as a proportion of the total number of individuals in the sample;

- Trichoptera
- Ephemeroptera
- Plecoptera
- Trichoptera + Ephemeroptera + Plecoptera (EPT)
- Odonata
- Hemiptera
- Coleoptera
- Tricladida
- Chironomidae

2.2.3   Classification Trees

Decision trees are popular in many fields as a way of encapsulating and structuring the knowledge of experts for use by the less experienced. They are commonly used in botany and medical decision making for example. Automatic tree construction was first developed in the social sciences, but the work of Breiman and colleagues in the late 1970's and early 1980's (encapsulated in their monograph; Breiman et al., 1984) placed tree-based models firmly within a statistical framework. Since then, the properties of tree-based classifiers have been well studied and are widely regarded as being a powerful tool for supervised classification purposes. Recent advances such as bagging (Breiman 1996), boosting (Freund and Schapire, 1997) and random forests (Breiman, 2001) have been developed that extend the tree concept to so-called ensemble methods to improve predictions from trees, but do so at the expense of simplicity and, to some extent, interpretability.

Tree-based methods partition the "feature space" into a set of regions and then fit a simple model, such as a constant one, in each one. Tree-based models are computationally intensive methods that are ideally suited to situations where there are many explanatory variables to choose from and it is not known a priori which of them are most important. The main virtues of tree-based models are that they are that they are excellent for initial data inspection, they give a clear picture of the structure of the data and they

provide a highly intuitive insight into the kinds of interactions between variables.

Models are fitted using binary recursive partitioning, where the data are successively split along features of the environmental data so that at any node the split which maximally distinguishes the response variable in the left and right branches is selected. Splitting continues until the nodes are "pure", i.e. comprising one class only, or the data are too sparse.

Where the response variable is a factor (i.e. grouped in classes), the tree is known as a classification tree. Where the response is continuous, the tree is known as a regression tree. Because the recursive partitioning continues until pure nodes are reached or the samples in each node are too sparse, there is a danger of over fitting the response. Tree-based models are generally pruned back to a minimal, adequate model. This is done via a cross-validation procedure to obtain "honest" estimates of the true prediction error. Plotting this prediction error against tree-size allows the selection of the tree with the minimum error. An alternative is to select, as the best tree, the smallest tree whose estimated error rate is within one standard deviation of the minimum error rate. A simple introduction to the use of classification and regression trees in ecological data analysis is given by De'ath and Fabricius (2000).

Splits are determined via minimising the Gini index measure of node impurity

$$D_i = 1 - \sum_{k=1}^{m} p_{ik}^2$$

where p_{ik} is the observed proportion of class k within node i, and D_i is the node impurity for node i. The total measure of node impurity is then

$$D = \sum_{i=1}^{n} D_i$$

Trees are generally fitted to their maximal extent and then a 10-fold cross-validation (CV) procedure is used to identify the smallest size of tree within 1 standard error of the tree with the lowest cross-validated misclassification error. In 10-fold CV, the training data are randomly assigned to one of ten groups. In turn, each group is excluded from the CV training set whilst the remaining nine groups are used to grow an "unpruned" tree. This tree is then used to predict the class membership of the samples in the group left out. This is repeated for each group in turn. At each of the 10 stages, the misclassification error is computed for each tree of size s, $s = 1, ..., m$, m = number of samples. The average error across the 10 CV stages is used as a measure of tree performance, with an associated standard error.

Predictions from the tree are governed by the terminal nodes or leaves of the tree. Predictions are based on a "majority rule" concept, whereby the predicted class for a target sample is determined by the most abundant class of the node the target sample ends up in. The proportion of samples classified into a particular node can also be used as an indicator of class probability; a target sample ending up in a node containing samples mainly of class k will have a high probability of belonging to class k. Conversely, a target sample ending up in a node with a mixture of classes will have a lower probability of belonging to the majority class.

2.2.4   Random Forests

Single classification tree models are known to be unstable predictors; they are sensitive to small changes

in the input or training data, with minor changes potentially leading to very different resulting split selections and subsequently different trees.

To overcome these issues recent developments in statistical data mining have focused on building ensembles of trees, each one fitted to a perturbed version of the training data, and aggregating the predictions from each tree into a final classification. Typically the perturbations involve bootstrap resampling of the training data to produce many sets of data of the same size as the training set. On average, each bootstrap sample contains 63% of the samples from the training set. The remaining 36% form a bootstrap test set against which the model fitted to that bootstrap sample is tested. This process is know as bagging.

The predictions from the many trees amount to a series of votes for the class of each sample in the training set. The votes are collected from all trees in the ensemble and the class that receives the largest number of votes for each sample is the predicted class for that sample.

A recent development in statistical machine learning is the method of Random Forests (Breiman, 2001). In addition to the stochastic perturbation introduced by repeated bootstrap sampling of the training data to build a large number of tree models, additional randomisation is added through the random selection of a small number of predictor variables with which to build each tree.

In Random Forests, a number of trees are grown. Each tree is grown from a bootstrap sample of the training set. Also, each tree is grown using a randomly chosen subset, usually small, of the available predictors, in this case the macro-invertebrate species compositional data and several "meta-taxa" (minimum species richness and total abundance aggregations for the entire sample or specific species groups within samples; see Section2.2.2 and Monteith and Simpson (2007) for more details). A large tree is grown without pruning and hence over fits the training data. The samples not included in the bootstrap sample, know as "out-of-bag" (OOB) samples, are introduced to the newly grown tree and the predicted class for the OOB samples recorded; this yields an estimate of the uncertainty inherent in model prediction likely to exist when the model is applied to new data not included in the training set. This procedure is repeated a large number of times. In the case of the Phase III tool, 81 trees were grown. The votes for the class of each training set sample, determined from the times when they were OOB, are aggregated and the class with the majority vote is taken as the predicted class for each sample.

Despite the potential for each individual tree in the random forest to over fit the data, it has been shown that the final random forest model does not. Furthermore, random forests have been shown to be comparable with the current leading machine learning classifiers in terms of predictive ability.

For new data, not used in training, the predicted class is determined by passing each sample in the new data down each of the trees in the random forest to arrive at a set of votes as to the predicted class for each new sample. Again, the final predicted class for each sample in the new data is the class receiving the majority vote over all trees in the forest. If required, the decision rules from the random forest can be coded externally in a GIS or database for use within other tools through interrogation of the individual tress within the R statistical software. Alternatively, it may be easier to use the command-line version of the tool to automate the process of saving a machine-readable version set of results from applying the tool and load this into external tools.

The sampling bias of the various classes can cause significant problems for building any classification tool as the tool could naively optimise fitting the predominant class and still achieve good prediction performance but do so at the expense of the less well-sampled classes. To help reduce this effect we

altered the prior probabilities of class membership to be equal across all classes when fitting the random forest model.

We now describe the transformation of the output from the random forest routine to meet the requirements of the WFD, and subsequently report the results of applying the tool to the training set samples.

2.3   Tool EQRs

To calculate EQRs, we adopt a novel approach based on a weighted average of scores assigned to the damage matrix. The damage matrix is derived from our understanding of how ANC in the context of Ca2+ can be used to infer acidification and biological damage (on the basis of critical load and palaeolimnological data pertaining to state change over reference conditions). As such, deviations from reference condition are explicitly contained with the damage matrix. Therefore, we argue that predicted class from the AWIST tool contains the underlying concepts that motivate EQRs, without having to compute them prior to assigning class.

However, EQRs are needed for the presentation of results and inter-calibration of tools across other member states. Therefore we have assigned initial scores to each of the damage matrix classes fitted by the random forest, the reflect a an ecological quality ratio on a scale of 0,...,1, where 1 represents no deviation from reference and 0 complete deviation from reference. As we have merged the Poor and Bad classes, we have four scores to assign. These scores are shown in Table 3.

High	Good	Moderate	Poor-Bad
0.9	0.7	0.5	0.3

Table 3: Initial scores assigned to damage matrix classes

We do not assign a score of 1 to the High class as it is unlikely that a site will today be exactly equivalent to its reference condition, as indicated by subtle changes in diatom assemblages over reference samples from pristine lakes in clean areas of the UK, away from strongly elevated atmospheric deposition (Simpson, unpublished data). We then compute an EQR for each site using these scores and the probability that a sample belongs to each of the classes. This probability is derived from the number of trees assigning the sample to each class as a proportion of all trees. See Section2.3.2 on the calculation of confidence of class for more details on these probabilities.

The EQR is calculated as a weighted average of the scores shown in Table 3, with weights taken as the probability that a sample belongs to a given class. As a worked example, a given sample is classified as Good status by the majority vote rule from the random forest, with probability (High = 0.2, Good = 0.7, Moderate = 0.15, and Poor-Bad = 0.05). The EQR for this sample, calculated as a weighted average of the base scores = 0.691.

It could be argued that the score assigned to the Poor-Bad class should be 0.2, the mid point between 0.3 and 0.1, which are the scores that would have been assigned to the Poor and Bad classes individually if they had not been merged. The current version of AWIST uses 0.3 but this can be changed without affecting the underlying forest the underpins the model predictions.

2.3.1   Tool Boundary EQRs

For the purposes of inter-calibration and reporting, EQRs may be required for key boundaries, such as the High-Good and Good-Moderate boundaries. The WFD60 tool is not based explicitly on an underlying EQR or metric scale. As such, there are no explicit tool boundaries along this underlying scale. Instead, we must consider what a boundary between two classes represents within the framework of the random forest approach.

It is natural to think of a boundary between two classes as the situation where a sample has equal probability of belonging to either of the classes on each side of the boundary. For example, the High-Good boundary can be thought of as occurring where a sample has a 50% probability of belonging to the High class and 50% probability of belonging to the Good class. In reality, given the odd number of trees grown in the WFD60 random forest, the observed probabilities will never attain this equal 50%; one class would receive at least one more vote than the other class, but for purposes of exposition, we retain this equal probability interpretation.

As we have seen, the actual EQR assigned is determined by the spread of class probabilities, which are used as weights in the weighted average calculation. Tool boundary EQRs can be assigned in the same manner to the definition of a boundary for the WFD60 tool described above. For example, the High-Good boundary, representing a 50:50 probability of a sample belonging to the two classes either side of the boundary would be assigned an EQR score of 0.8. This represents the mid-point between the two scores given to the respective classes (0.9 and 0.7) in Table 3. Similarly, the Good-Moderate boundary would be assigned an EQR score of 0.6, the mid-point between the two scores assigned to the Good and Moderate classes (0.7 and 0.5).

Whilst it is useful to define these boundaries for reporting or inter-comparison between tools it should be noted that the theoretical definition of the boundary within the WFD60 is somewhat artificial as the decision boundaries between classes will depend on the distribution of votes (probabilities) across all four classes predicted by the tool, not just the two classes representing the theoretical boundary.

2.3.2   Confidence of Classification

Confidence of class is computed from the posterior probabilities assigned to each of the four classes for each individual sample. The posterior probabilities are computed as the number of trees assigning a sample to class k divided by the total number of trees, n. A sample that is regularly predicted to be in say the High status class by trees in the forest is a sample in which we have a high degree of confidence in the assigned class. Where the assigned class is determined from a relatively small proportion of the alloted votes for that class it represents a situation where a number of the votes were cast for other classes and correspondingly will acquire a posterior probability similar to that of the assigned class. This represents a borderline case or a sample that is not well predicted by the random forest, and in which we would have a much reduced confidence in the assigned class.

For example, a given sample is assigned to the High status class on the basis of the majority vote rule. Out of 100 votes, 80 are for High, 10 for Good and 5 each for Moderate and Poor-Bad. The probability or confidence of class for this sample is then 80/100 = 0.8 or 80%. A second sample is also assigned to the High class, this time with only 55 votes for High, 30 for Good, 10 for Moderate and 5 for Poor. The confidence of the classification into High status for this sample is only 55/100 = 0.55 or 55%.

The WFD60 tool produces these confidence of class measures for each of the four classes modelled by

the random forest directly from the posterior probabilities assigned and are reported in the tool output.

As a summary of the confidence of class, we provide Shannon's entropy (H) measurement

$$H(X) = -\sum_{i=1}^{n} p(x_i)\log_2(p(x_i))$$

where $p(x_i)$ is the probability that x belongs to the nth class. H takes values between 0 and 1, with 0 indicating that the probability of belonging to a single status class is 100% and 0% for all other classes. Likewise, H = 1 indicates a sample for which there is equal probability of belonging to all classes. Here we apply Shannon's entropy measure to the probability votes for the four WFD status classes computed from the random forest.

2.4   Results of model fitting

The WFD60 training set data were applied to the AWIST random forest tool to determine predicted WFD status class as well as EQRs and confidence of class. Table 4 Shows a cross classification table of the original damage matrix classes and the classes predicted for each training set sample, aggregated to the class level. The individual sample results are shown in Table 5 below. For each sample (identified by the WBID for the lake from which the sample was collected) the posterior probability that that sample is a member of the stated class is given, along with the AWIST predicted class and the computed EQR generated from the weighted average of the posterior probabilities.

AWIST	Damage Matrix				
	High	Good	Moderate	Poor	Bad
High	57	0	0	0	0
Good	0	35	0	0	0
Moderate	0	1	30	0	0
Poor-Bad	0	0	0	9	5

Table 4: Cross classification table comparing the Damage Matrix class against that predicted by AWIST

Table 5: Results of the application of AWIST to the WFD60 training set. WBID is the Water-Body ID for each sample (from UKLakes.net); High, Good, Moderate, Poor-Bad are the probabilities that the sample belongs to the stated class; Class is the assigned AWIST classification; EQR is the computed EQR for the sample.

WBID	High	Good	Moderate	Poor-Bad	Class	EQR
4204	0.92	0.04	0.04	0	HIGH	0.88
5073	0.75	0.08	0.16	0.01	HIGH	0.82
5350	0.88	0.05	0.07	0	HIGH	0.86
6140	0.92	0.06	0.03	0	HIGH	0.88
6405	0.88	0.06	0.05	0.01	HIGH	0.86
7824	0.73	0.09	0.17	0.01	HIGH	0.81
8266	0.07	0.69	0.15	0.09	GOOD	0.65

WBID	High	Good	Moderate	Poor-Bad	Class	EQR
9070	0.86	0.09	0.05	0	HIGH	0.86
10307	0.78	0.08	0.12	0.02	HIGH	0.83
11189	0.85	0.08	0.04	0.03	HIGH	0.85
11238	0.77	0.08	0.13	0.03	HIGH	0.82
11315	0.23	0.68	0.09	0	GOOD	0.73
11338	0.93	0.06	0.02	0	HIGH	0.88
11424	0.04	0.64	0.23	0.09	GOOD	0.63
11611	0.98	0.02	0.01	0	HIGH	0.89
11862	0.74	0.08	0.16	0.03	HIGH	0.81
12469	0.1	0.12	0.73	0.06	MOD	0.55
12578	0.93	0.05	0.03	0	HIGH	0.88
12606	0.85	0.13	0.03	0	HIGH	0.87
13791	0.13	0.18	0.65	0.03	MOD	0.58
14057	0.9	0.08	0.02	0	HIGH	0.88
14202	0.77	0.13	0.07	0.03	HIGH	0.83
14293	0.83	0.11	0.02	0.05	HIGH	0.84
15316	0.88	0.04	0.05	0.03	HIGH	0.85
15551	0.89	0.08	0.03	0	HIGH	0.87
16443	0.84	0.12	0.04	0	HIGH	0.86
16530	0.88	0.1	0.01	0.01	HIGH	0.87
17147	0.19	0.75	0.05	0.01	GOOD	0.73
17334	0.09	0.83	0.04	0.03	GOOD	0.7
17379	0.82	0.1	0.04	0.04	HIGH	0.84
17619	0.7	0.18	0.13	0	HIGH	0.82
18209	0.15	0.74	0.08	0.03	GOOD	0.7
18305	0.28	0.26	0.26	0.2	HIGH	0.63
18767	0.82	0.14	0.03	0.02	HIGH	0.85
18825	0.93	0.06	0.02	0	HIGH	0.88
19381	0.94	0.05	0.01	0	HIGH	0.89
19540	0.91	0.08	0.02	0	HIGH	0.88
20633	0.78	0.15	0.04	0.03	HIGH	0.84
20647	0.86	0.12	0.02	0.01	HIGH	0.87
20657	0.91	0.08	0.02	0	HIGH	0.88
20712	0.75	0.16	0.08	0.02	HIGH	0.83
20725	0.69	0.13	0.15	0.03	HIGH	0.8
20735	0.11	0.39	0.43	0.08	MOD	0.61
20739	0.69	0.1	0.17	0.04	HIGH	0.79
20922	0.7	0.06	0.24	0	HIGH	0.79

WBID	High	Good	Moderate	Poor-Bad	Class	EQR
21490	0.96	0.02	0.02	0.01	HIGH	0.89
21723	0.03	0.03	0.13	0.81	POOR-BAD	0.36
21790	0.7	0.05	0.11	0.14	HIGH	0.76
21848	0.38	0.62	0.01	0	GOOD	0.77
22125	0.08	0.8	0.06	0.07	GOOD	0.68
22223	0.16	0.68	0.14	0.02	GOOD	0.7
22259	0.92	0.05	0.03	0.01	HIGH	0.88
22308	0.24	0.07	0.68	0.01	MOD	0.61
22395	0.82	0.16	0.01	0.02	HIGH	0.86
22577	0.81	0.15	0.04	0	HIGH	0.85
22782	0.95	0.05	0	0	HIGH	0.89
23361	0.17	0.68	0.09	0.06	GOOD	0.69
24020	0.89	0.06	0.04	0.01	HIGH	0.87
24744	0.88	0.09	0.03	0	HIGH	0.87
24745	0.73	0.06	0.17	0.04	HIGH	0.8
24754	0.17	0.79	0.02	0.03	GOOD	0.72
24758	0.93	0.04	0.02	0.01	HIGH	0.88
24892	0.84	0.09	0.04	0.03	HIGH	0.85
27693	0.07	0.1	0.81	0.03	MOD	0.54
27699	0.11	0.22	0.68	0	MOD	0.59
27777	0.16	0.78	0.04	0.02	GOOD	0.72
27801	0.14	0.72	0.13	0.02	GOOD	0.7
27808	0.05	0.14	0.78	0.03	MOD	0.54
27823	0.68	0.04	0.11	0.17	HIGH	0.75
27827	0.1	0.74	0.09	0.07	GOOD	0.68
27849	0.03	0.73	0.2	0.05	GOOD	0.65
27872	0.1	0.75	0.11	0.04	GOOD	0.68
27888	0.07	0.09	0.82	0.03	MOD	0.54
27900	0.06	0.13	0.76	0.05	MOD	0.54
27912	0.05	0.12	0.77	0.07	MOD	0.53
27922	0.12	0.83	0.05	0.01	GOOD	0.71
27923	0.08	0.71	0.1	0.12	GOOD	0.65
27927	0.12	0.09	0.66	0.13	MOD	0.54
28076	0.01	0.09	0.16	0.74	POOR-BAD	0.37
28130	0.1	0.17	0.14	0.59	POOR-BAD	0.46
28370	0.05	0.08	0.78	0.09	MOD	0.52
28905	0.04	0.08	0.13	0.74	POOR-BAD	0.39
28965	0.86	0.12	0.03	0	HIGH	0.87

WBID	High	Good	Moderate	Poor-Bad	Class	EQR
28986	0.88	0.06	0.05	0.01	HIGH	0.86
29000	0.2	0.79	0.01	0	GOOD	0.74
29008	0.07	0.11	0.1	0.73	POOR-BAD	0.4
29021	0.72	0.17	0.08	0.04	HIGH	0.81
29045	0.05	0.72	0.1	0.13	GOOD	0.64
29052	0.2	0.73	0.04	0.03	GOOD	0.72
29062	0.24	0.75	0.01	0	GOOD	0.75
29081	0.09	0.77	0.1	0.04	GOOD	0.68
29153	0.02	0.13	0.15	0.71	POOR-BAD	0.39
29179	0.58	0.13	0.16	0.13	HIGH	0.73
29181	0.02	0.71	0.21	0.07	GOOD	0.64
29183	0.26	0.71	0.02	0.02	GOOD	0.74
29184	0.94	0.02	0.03	0.02	HIGH	0.88
29215	0.78	0.16	0.04	0.03	HIGH	0.84
29246	0.13	0.8	0.07	0	GOOD	0.71
29252	0.08	0.65	0.19	0.08	GOOD	0.64
29285	0.04	0.1	0.14	0.72	POOR-BAD	0.39
29290	0.05	0.08	0.69	0.18	MOD	0.5
31104	0.07	0.05	0.12	0.77	POOR-BAD	0.38
33730	0.87	0.08	0.05	0.01	HIGH	0.86
33803	0.19	0.76	0.05	0	GOOD	0.73
33836	0.22	0.78	0	0	GOOD	0.74
33843	0.03	0.12	0.81	0.04	MOD	0.53
33998	0.15	0.13	0.71	0.02	MOD	0.58
34002	0.19	0.76	0.04	0.01	GOOD	0.73
34039	0.02	0.13	0.81	0.05	MOD	0.52
34061	0.04	0.03	0.7	0.23	MOD	0.48
34243	0.02	0.11	0.85	0.03	MOD	0.52
34249	0.05	0.11	0.75	0.09	MOD	0.52
34319	0.12	0.07	0.79	0.03	MOD	0.56
34363	0.03	0.12	0.8	0.06	MOD	0.52
34375	0.04	0.63	0.12	0.22	GOOD	0.6
34382	0.03	0.64	0.29	0.04	GOOD	0.63
34390	0.04	0.13	0.72	0.11	MOD	0.52
34400	0.03	0.09	0.2	0.68	POOR-BAD	0.39
34632	0.06	0.68	0.22	0.05	GOOD	0.65
34635	0.03	0.07	0.85	0.05	MOD	0.52
34668	0.05	0.11	0.77	0.08	MOD	0.53

WBID	High	Good	Moderate	Poor-Bad	Class	EQR
34987	0.87	0.05	0.08	0	HIGH	0.86
35233	0.07	0.04	0.88	0.01	MOD	0.53
35262	0.07	0.09	0.18	0.67	POOR-BAD	0.41
35561	0.06	0.12	0.17	0.66	POOR-BAD	0.42
35578	0.07	0.1	0.81	0.03	MOD	0.54
35650	0.08	0.09	0.8	0.03	MOD	0.54
36267	0.08	0.16	0.7	0.07	MOD	0.55
36405	0.25	0.72	0.03	0	GOOD	0.74
38390	0.08	0.03	0.04	0.85	POOR-BAD	0.37
38394	0.12	0.09	0.79	0	MOD	0.57
38409	0.08	0.09	0.08	0.75	POOR-BAD	0.4
38422	0.8	0.18	0.02	0.01	HIGH	0.85
38525	0.13	0.09	0.76	0.02	MOD	0.57
38907	0.18	0.78	0.03	0.01	GOOD	0.73
46232	0.12	0.08	0.73	0.08	MOD	0.55
46279	0.08	0.73	0.11	0.08	GOOD	0.67
99999	0.05	0.12	0.13	0.71	POOR-BAD	0.4

Table 4 indicates that only a single site in the training set is misclassified; damage matrix class is Good, AWIST predicted is Mod. 136 out of 137 sites were correctly predicted on the basis of the AWIST random forest yielding an apparent error rate of less than 1%. This is extremely high and illustrates the predictive ability of the tool. However, we should be cautious in concluding that the tool will act with the same degree of accuracy when predicting WFD status class for new samples not used in the training process.

A better guide to the real predictive ability of a tool is achieved using cross-validation, such as *k*-fold or bootstrapping (bagging). Recall that each tree in the random forest was grown with a separate bootstrap sample of the original training data, and that, on average 33% of the training samples are not included in a bootstrap sample and are termed out-of-bag samples. For each tree that an individual sample was out-of-bag we can use the predictions of that tree to assess the predictive ability of the model under cross-validation. A cross-classification table showing the AWIST out-of-bag predicted class against the original damage matrix class is shown in Table 6.

AWIST	Damage Matrix				
	High	**Good**	**Moderate**	**Poor**	**Bad**
High	46	14	6	1	0
Good	3	4	5	3	1
Moderate	5	14	13	3	2
Poor-Bad	1	3	6	2	2

Table 6: Cross classification table comparing the Damage Matrix class against that predicted by AWIST when considering only the occasions that samples were out-of-bag

The misclassification (error) rate for the OOB samples is 0.5 (50%). This is indicates the model is twice as good as a naive predictor selecting the same class for all samples. This values seems quite low, especially given the very low apparent error rate of <1%. A clue to this apparent discrepancy is given by the trace plot showing the change in the OOB prediction error rate as new trees are added to the forest. This plot is shown in Figure 2.
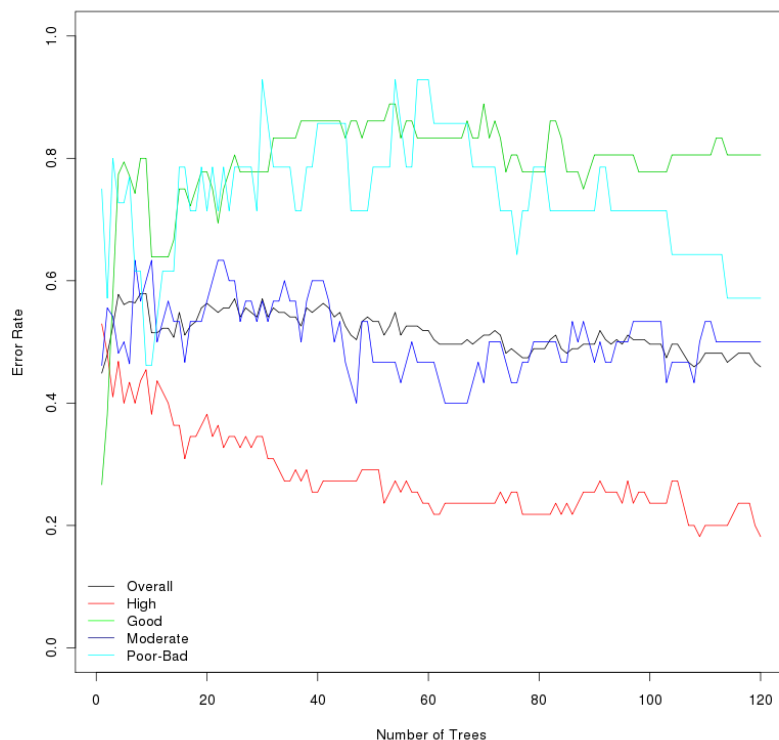
Figure 2: Trace of the OOB error rate as trees are added to the random forest

This plot shows large fluctuations in the OOB error rate for the all classes other than High periodically as trees are added to the forest. This, we believe, is due to the sampling bias in the training data. Recall that the High class was best represented, with low numbers of Moderate and Poor-Bad samples. As bootstrap samples are drawn from this training set, it is likely that many of these bootstrap samples contain few, if any, of the less well sampled classes are selected, thus making it difficult for an individual tree fitted to such data to find sufficient rules to predict samples of the classes poorly represented in that bootstrap

sample. We believe that whilst the apparent error rate is unrealistically low, the OOB error rate is unrealistically high given the problem described above, and that the actual tool will have a real error rate somewhere between these two extremes. With an independent test set it would be possible to quantify the real error rate of the random forest.

One disadvantage of ensemble methods such as random forests over a single decision tree is that the ensemble models loose the simple interpretability that is a feature of single trees. It is possible to interrogate individual trees in the forest, but with 120 trees in the AWIST forest this is not practically feasible. An alternative approach is to to compute the average change in two accuracy measures for each variable for those trees that a variable was not used to grow the tree. The two measures computed are i) mean decrease in accuracy and ii) mean decrease in the Gini index (node impurity). A variable that is important in determining the predictions from, and the predictive ability of, the random forest will have high values for these two statistics.
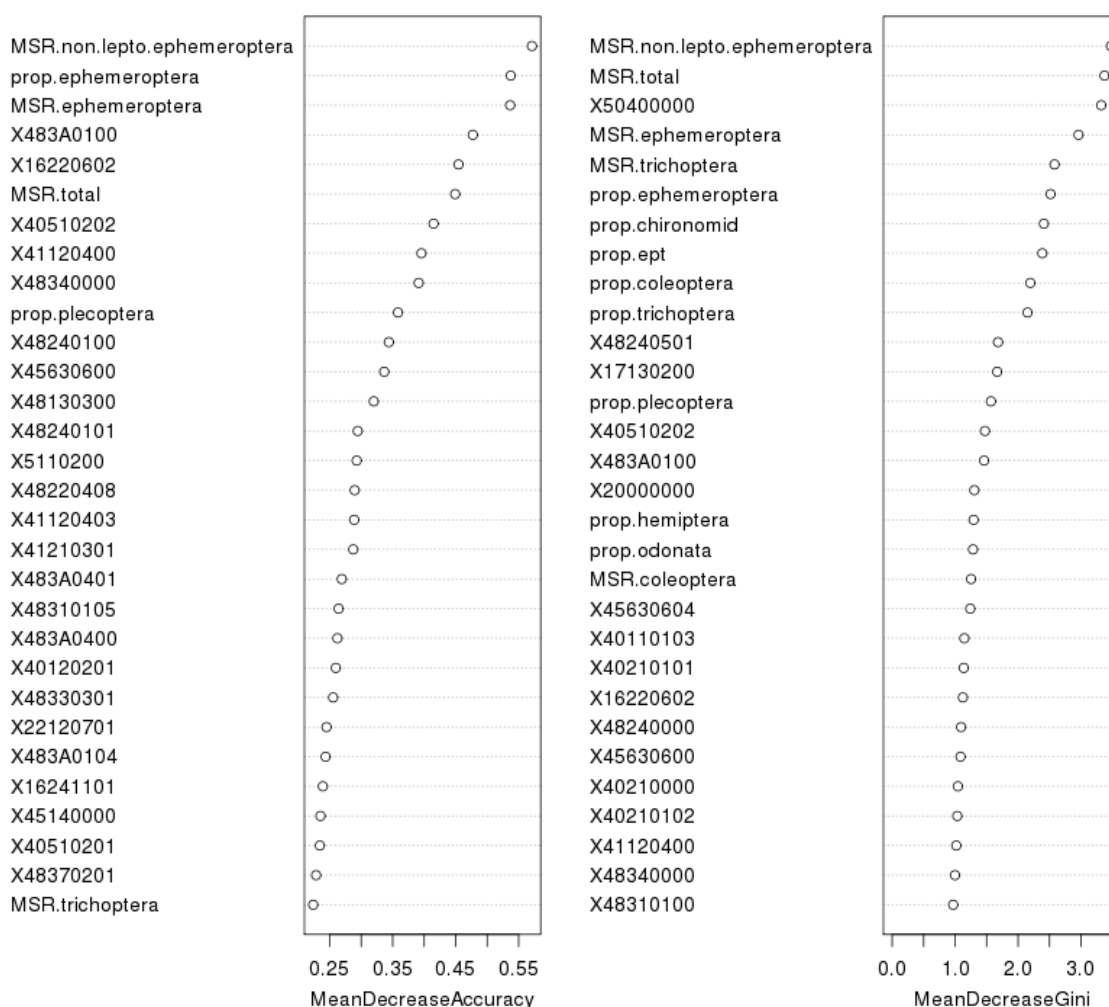
Figure 3: Variable importance statistics for the AWIST random forest. See text for details.

Figure 3 shows the variable importance measures for the top 30 variables. The meta taxa dominate these two measures, suggesting that these summaries of the macroinvertebrate assemblage are particularly useful in predicting damage matrix classes. The minimum species richness (MSR) of non-leptophlebid mayflies is the most important variable on both measures of importance. Other variables with high
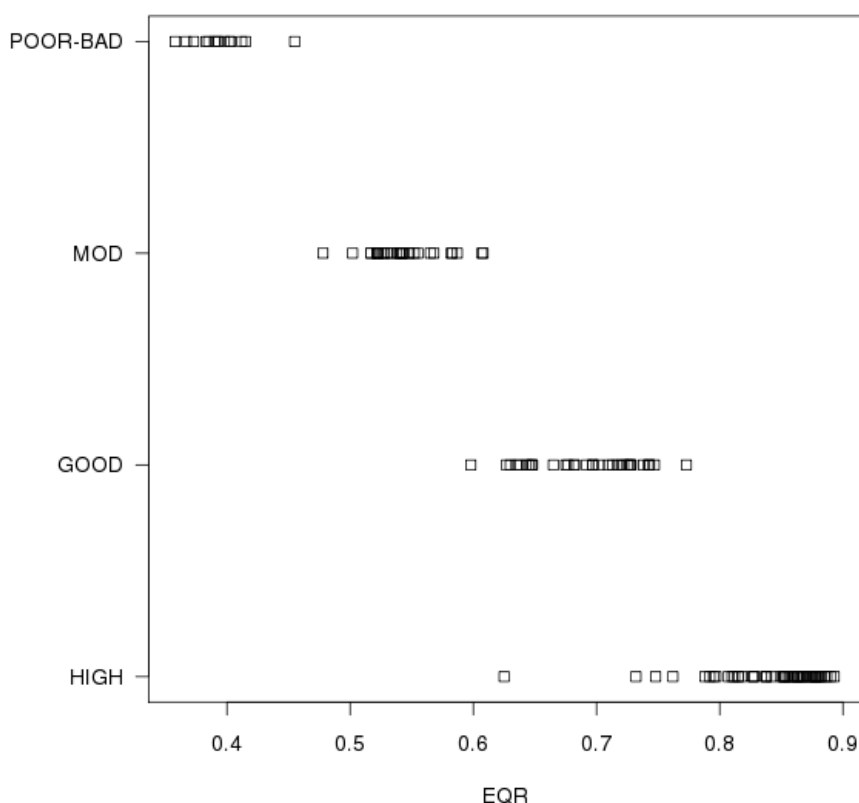
Figure 4: Stripplot showing the distribution of EQRs and predicted WFD status class for AWIST. Note the point to the extreme left of the High status samples. This is WBID 18305 (Carslub), and is an example of a site for which a high degree of uncertainty exists. The low EQR for this site derives from the similar spread of votes across all four status classes.

importance include the MSR of the total assemblage the proportion of the total assemblage comprising Ephemoptera and MSR of Ephemoptera. A number of taxa are also important in determining the predicted classes.

The WFD normative definitions place importance upon the loss of major taxonomic groups in defining WFD status class. As such, the results of the variable importance measures for the AWIST random forest indicate that the tool is consistent with the spirit and definition of the WFD.

Tool EQR's should discriminate between WFD status class, allowing comparison of EQR boundaries between the tools of other member states'. Figure X shows a stripplot of the EQR assigned to each site in the WFD60 data set on the basis of AWIST, against the predicted class. In general there is good discrimination between predicted class on the basis of EQR scores, however, in some cases sites with differing predicted class have similar EQRs. It is easy to see why this may occur; poorly predicted sites, those for which the tool is uncertain and trees in the forest produce votes for several classes for a single site, are given an EQR based on a weighted average of the class votes. For a poorly predicted site, it is not infeasible that a similar number of votes will be cast for several or all status classes. As the predicted class is determined by the majority vote, a prediction must be made by AWIST, but the resulting EQR may be inconsistent with the predicted class. Overlap between the EQRs of sites in different classes will be most apparent at the boundaries between classes on the EQR scale.
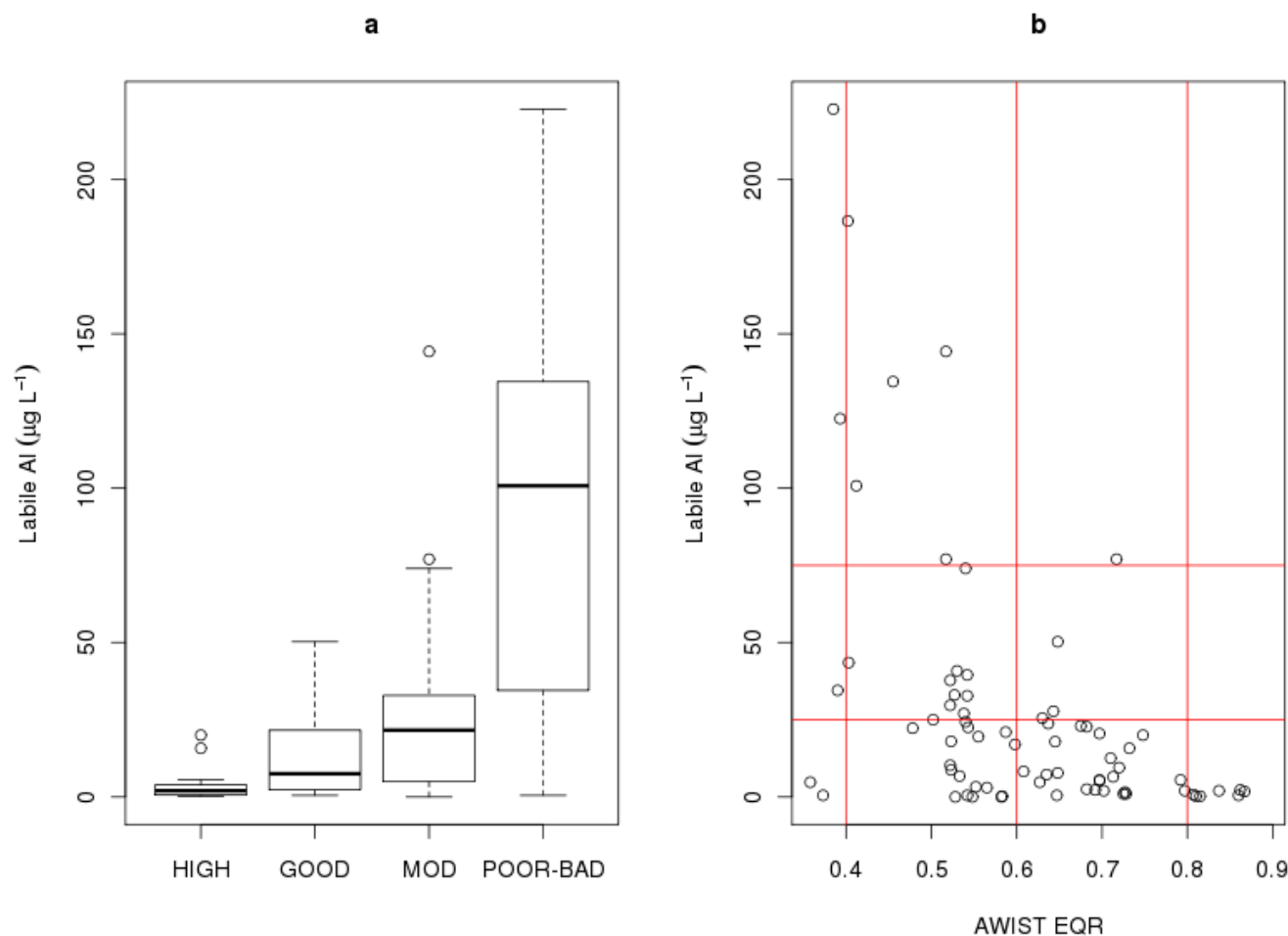
Figure 5: Distribution of labile aluminium concentration against AWIST status class (a) and AWIST EQR (b) for the WFD60 data set. The vertical red lines are the theoretical EQR boundaries for AWIST,whilst the horizontal lines are thresholds of ecological damage for fish populations.

A clear example of this is shown in Figure 4, where WBID 18305 (Caslub) has an EQR of 0.625 yet is predicted to be in HIGH status. The Damage Matrix class for this site is High. The proportion of votes cast for each class for Carslub was High (0.283), Good (0.258), Moderate (0.258), and Poor-Bad (0.2). Clearly there is much uncertainty about which class this site belongs to on the basis of the observed macroinvertebrate assemblage and AWIST. In this sense, the AWIST EQR encapsulates uncertainty around the predictions from the random forest component of the tool that generates the predictions; the tool highlights cases where, despite a class status being assigned, there is high uncertainty in that prediction. We consider this feature of our tool to be beneficial to managers of lake systems.

Figure 5 shows plots of Al_{lab} against (a) AWIST class and (b) AWIST EQR. The boxplots of A_{llab} against AWIST class show that labile aluminium concentrations in waters increase considerably as status drops from High to Poor-Bad, indicating that the predictions from AWIST, that are based upon rules derived from the macroinvertebrate community data, are consistent with our understanding of potential for biological damage. As labile aluminium concentration was not used directly to produce the damage matrix, this represents a semi-independent test of AWIST and demonstrates results consistent with WFD normative definitions. To investigate this relationship formally, ANOVA was used to determine if there
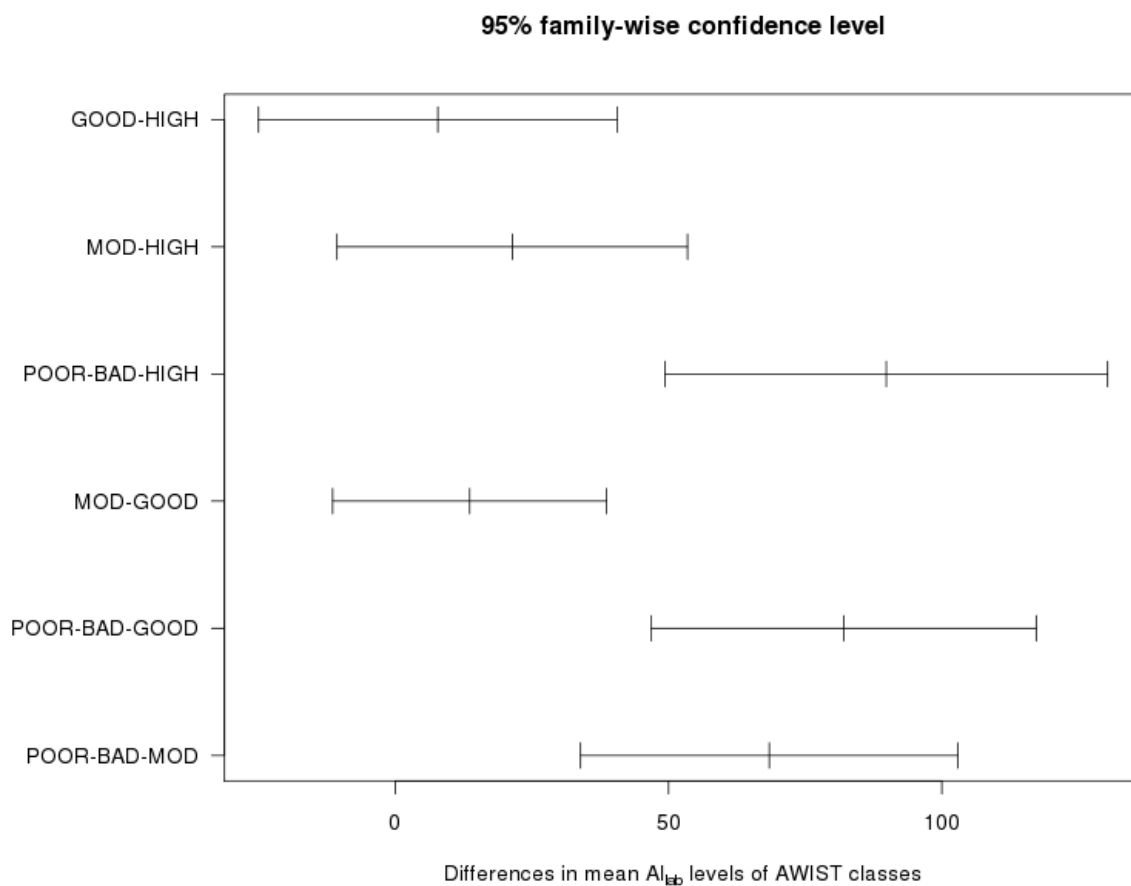
Figure 6: Tukey's honest significant differences between pair-wise comparisons of mean Al_{lab} concentration for AWIST classes

was a statistically significant difference between the mean Al_{lab} concentration of the four AWIST classes.

ANOVA showed a statistically significant difference between the mean Al_{lab} concentration of the four AWIST status classes ($F = 14.695$, $p = $ <<0.00001). This results indicates that when taken as a whole, there are statistically significant differences between the mean Al_{lab} concentration of some of the AWIST status classes. To investigate which classes have different mean concentrations at the 95% level we must individually compare each pair-wise grouping of status classes, resulting in six comparisons. This presents a multiple comparisons problem in that if we naively computed *t*-tests for each pair-wise grouping of classes and noted if each pairing were significantly different at the 95% level, the overall significance of this test would not by at the 95% level. To allow for multiple comparisons we use Tukey's honest significance differences between each pair-wise grouping, a method that controls for the overall significance level.

The results of this analysis are shown in Figure 6. The difference in mean concentration is shown by the middle tick mark on each bar, whilst the extreme ticks are 95% confidence intervals on the difference in means. Where the confidence interval includes 0 there is insufficient evidence to say that the means of the two classes are different. From this figure, we see that the High-Good, High-Mod, and Good-Mod pairings have similar mean concentrations, whilst the Poor-Bad class has a statistically significantly different mean Al_{lab} concentration to the other classes. Lack of a difference between several classes likely stems from two observations; firstly, in Figure 5, we note that the Al_{lab} concentrations cover a similar
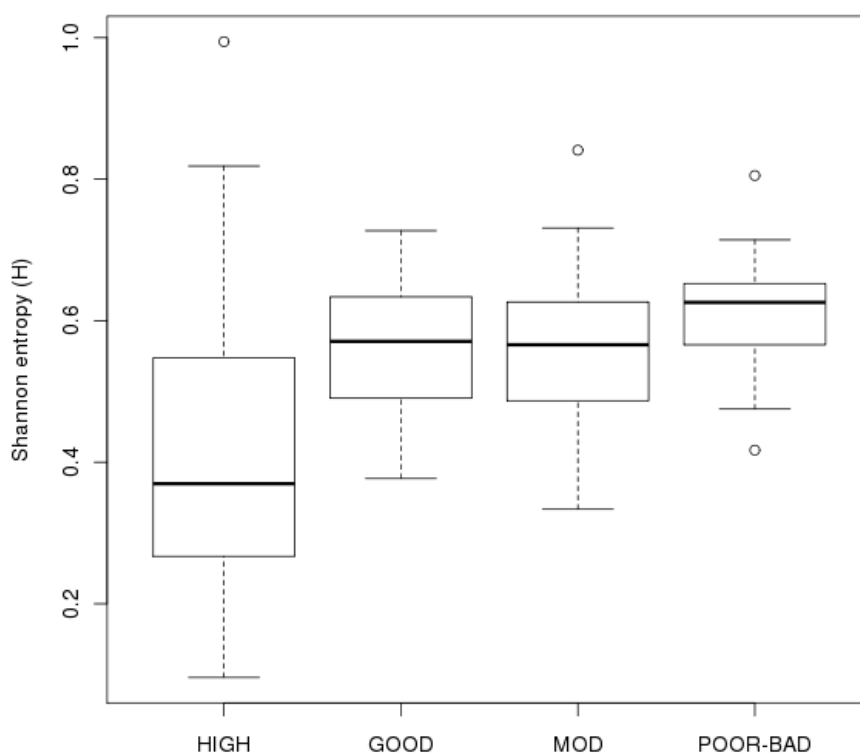
*Figure 7: Boxplot showing the distribution of Shannon entropy (H) within
each AWIST status class.*

range down to 0 (or undetectable) concentration, and secondly, there are marked differences in the
variability and numbers of samples in each comparison. So whilst quantitatively there is no significant
difference between High-Moderate and Good-Moderate on this measure of ecosystem status (Al_{lab}),
qualitatively the tool is producing logical results consistent with our understanding of acidification
damage. Furthermore, given the noisy data and relatively limited number of samples in the overall data
set and in some classes in particular, it is encouraging that at least qualitatively we are able to demonstrate
such differentiation in a key measure of damage. A similar impression is given when we compare
AWIST EQR scores with Al_{lab} concentration for the WFD60 data set (Figure 5b).

Figure 7 shows boxplots of Shannon entropy for each AWIST status class. The High class has the lowest
observed H values and the lowest median H value of all the four classes. The median H values for each of
the three remaining classes are quite similar about c. 0.6, indicating relatively larger uncertainty (of lower
confidence of class) for each of these classes in general than High status. The High status class has the
greatest range of H.

We would expect uncertainty to be highest around the boundaries between classes. To illustrate this, we
plot Shannon entropy against AWIST EQR scores in X. This figure shows aspects of the results already
seen in Figure 7, namely that for High status predictions, the uncertainty, expressed as Shannon entropy,
is generally lower than that of the three remaining classes. There is a high degree of scatter in the Figure
X, so a LOESS smoother has been added to highlight pattern in the relationship between EQR and H. As
expected, confidence of class is, in general, higher at the boundaries between classes on the EQR scale.
For the Poor-Bad class, this relationship breaks down somewhat, a reflection of the paucity of samples in
this class and hence the greater uncertainty attached to this class.
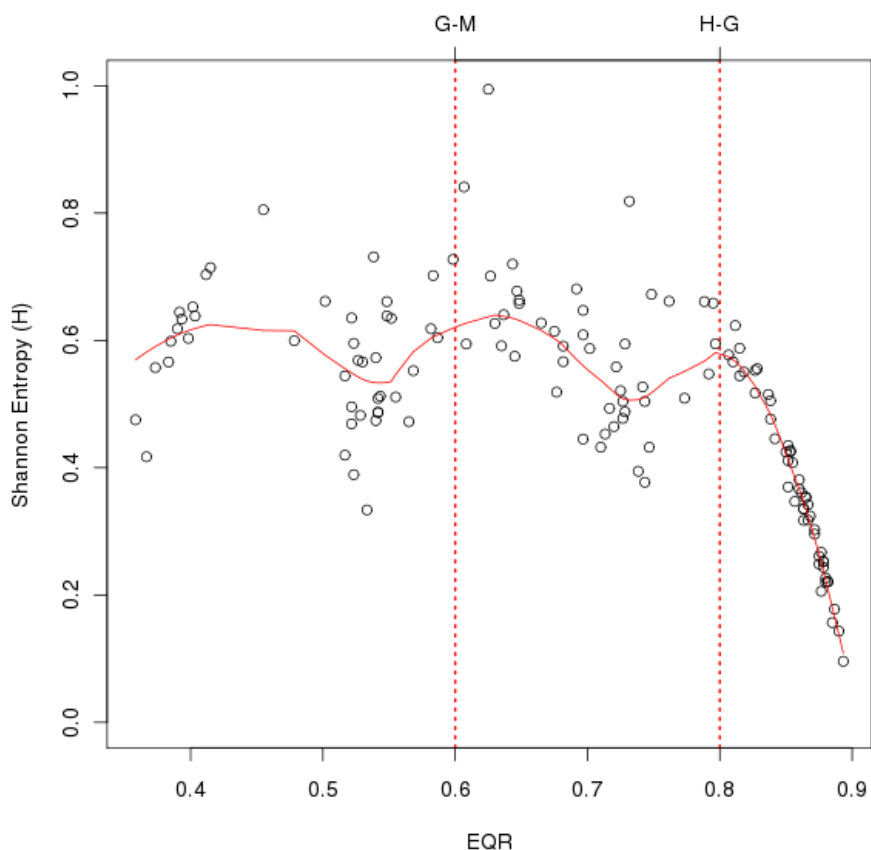
Figure 8: AWIST EQR and Shannon entropy (H) for the WFD60 data set. The solid line is a Loess smoother which highlights the pattern in the relationship. The dashed vertical lines are the theoretical boundaries between High and Good (H-G) and Good and Moderate (G-M) classes for AWIST.

In each of these examples where AWIST EQRs are compared with predicted class, uncertainty or other measures of ecosystem status (e.g. Al_{lab}) there is a degree of variability in EQRs within status classes or groupings of these other measures. This variability arises, in part due to the way in which the EQRs are computed from the distribution of votes for each of the four classes. It is important to note that the AWIST EQR includes, as a result of the way in which it is computed, the inherent uncertainty in the predicted class. This should be considered an advantage of the EQR computed by AWIST over other tools; clearly sites that are likely to belong to other groups possess features of the macroinvertebrate community that are common to other classes and the EQR should be adjusted accordingly.

2.5   Comparison with other WFD tools for acid waters

Several other WFD tools have also been developed for acid waters based on macroinvertebrate assemblages, including CPET, which is based on chironomid pupal exuviae, and LAMM, a tool developed by the Environment Agency that is based on species metrics. It is useful therefore to compare the results of these tools and assess whether combining the results of two or more tools into a final prediction of class status.

As there is no true class it is difficult to compare the tools in a consistent manner. In some cases, there is
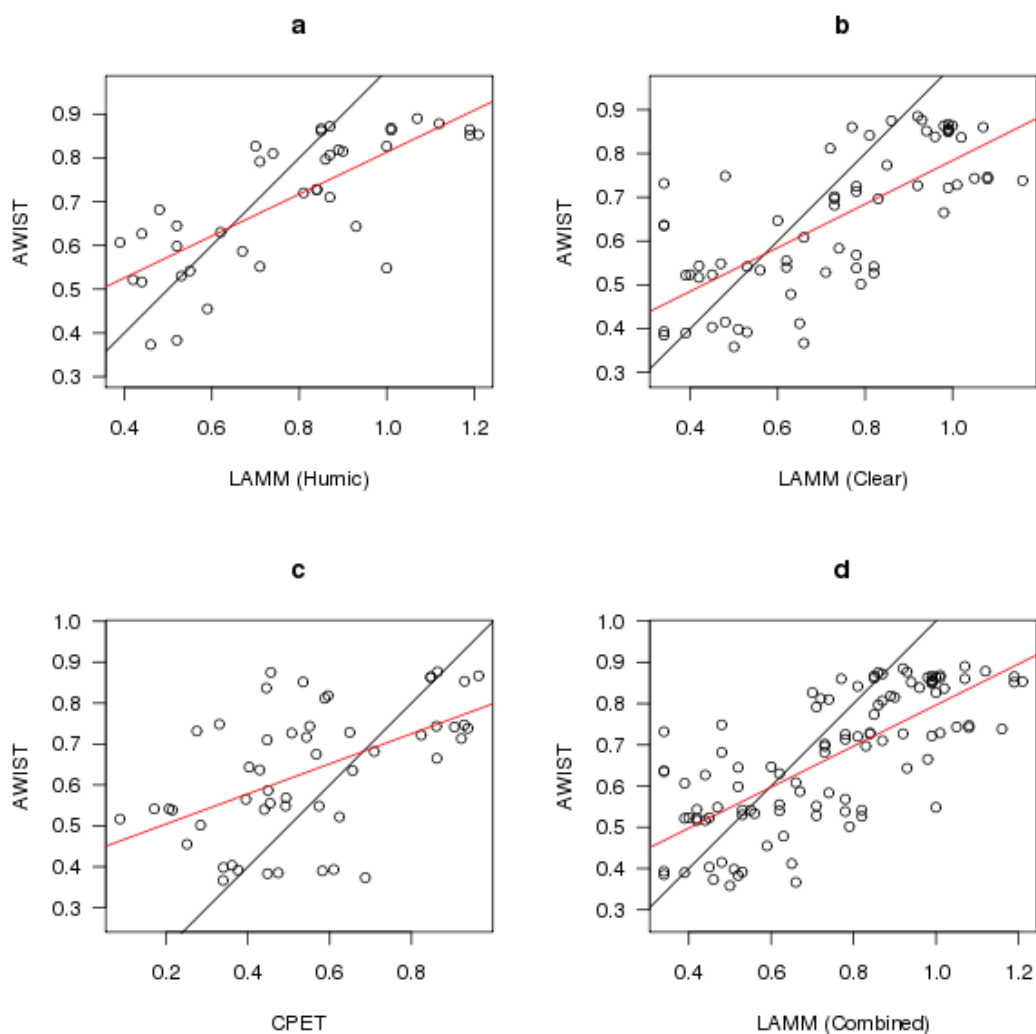
Figure 9: Comparison of EQRs generated by AWIST and LAMM humic (a), LAMM clear (b), CPET (c), and LAMM combined (d). In each plot the solid black line is a 1:1 line and the solid red line is the linear, least-squares regression line between each pair of EQRs.

currently no prediction for a sample for one or more of the tools. Instead, here we choose to compare each of the tools to the damage matrix as this classifier represents the best currently available measure of WFD status as it incorporates information from a wide range of sources as well as data on state change over reference from critical loads and palaeoecological evidence. Table 7 contains the predicted classes for sites in the WFD60 training set for each of the CPET, LAMM and AWIST tools, as well as those from the damage matrix. It is clear that for CPET in particular, currently there are no predictions for many of the WFD60 samples.

We proceed to generate cross-classification tables for CPET and LAMM compared to the damage matrix. Reference should be made to Table 4 and Table 6 whilst interpreting the tables below as they present the same data for AWIST as well as the OOB predictions for AWIST. Table 8, Table 9and Table 10 show cross-classification tables comparing LAMM (clear), LAMM (humic) and CPET, respectively with the damage matrix for a subset of the samples for which damage matrix class is known. A summary of the comparisons of the four tools (including the OOB predictions from AWIST) is shown in Table 11.

Table 7: Comparison between CPET, LAMM and AWIST WFD tools class predictions for named sites and those of the damage matrix. The code "-" is used to indicate where a prediction for that tool-site combination is not currently available. The AWIST predicted classes are based on applying the tool to the training set data and not from the OOB predictions.

WBID	Damage	AWIST	LAMM	CPET
4204	HIGH	HIGH	-	MOD
5073	HIGH	HIGH	HIGH	-
5350	HIGH	HIGH	HIGH	-
6140	HIGH	HIGH	-	-
6405	HIGH	HIGH	HIGH	HIGH
7824	HIGH	HIGH	GOOD	-
8266	GOOD	GOOD	MOD	-
9070	HIGH	HIGH	HIGH	-
10307	HIGH	HIGH	GOOD	-
11189	HIGH	HIGH	HIGH	-
11238	HIGH	HIGH	HIGH	GOOD
11315	GOOD	GOOD	HIGH	-
11338	HIGH	HIGH	-	-
11424	GOOD	GOOD	MOD-POOR-BAD	-
11611	HIGH	HIGH	-	-
11862	HIGH	HIGH	HIGH	-
12469	MOD	MOD	GOOD	-
12578	HIGH	HIGH	-	-
12606	HIGH	HIGH	HIGH	-
13791	MOD	MOD	GOOD	-
14057	HIGH	HIGH	HIGH	HIGH
14202	HIGH	HIGH	HIGH	-
14293	HIGH	HIGH	GOOD	-
15316	HIGH	HIGH	HIGH	GOOD
15551	HIGH	HIGH	-	-
16443	HIGH	HIGH	GOOD	-
16530	HIGH	HIGH	HIGH	-
17147	GOOD	GOOD	GOOD	-
17334	GOOD	GOOD	GOOD	-
17379	HIGH	HIGH	HIGH	-
17619	HIGH	HIGH	-	-
18209	GOOD	GOOD	GOOD	-
18305	HIGH	HIGH	-	-
18767	HIGH	HIGH	HIGH	-

WBID	Damage	AWIST	LAMM	CPET
18825	HIGH	HIGH	-	-
19381	HIGH	HIGH	-	-
19540	HIGH	HIGH	-	-
20633	HIGH	HIGH	-	-
20647	HIGH	HIGH	-	-
20657	HIGH	HIGH	HIGH	-
20712	HIGH	HIGH	-	-
20725	HIGH	HIGH	-	-
20735	GOOD	MOD	MOD-POOR-BAD	-
20739	HIGH	HIGH	-	-
20922	HIGH	HIGH	GOOD	-
21490	HIGH	HIGH	HIGH	-
21723	POOR	POOR-BAD	POOR-BAD	-
21790	HIGH	HIGH	-	-
21848	GOOD	GOOD	GOOD	-
22125	GOOD	GOOD	-	-
22223	GOOD	GOOD	GOOD	-
22259	HIGH	HIGH	HIGH	-
22308	MOD	MOD	MOD	-
22395	HIGH	HIGH	-	-
22577	HIGH	HIGH	HIGH	HIGH
22782	HIGH	HIGH	HIGH	-
23361	GOOD	GOOD	-	-
24020	HIGH	HIGH	HIGH	-
24744	HIGH	HIGH	HIGH	-
24745	HIGH	HIGH	HIGH	-
24754	GOOD	GOOD	HIGH	-
24758	HIGH	HIGH	-	-
24892	HIGH	HIGH	HIGH	-
27693	MOD	MOD	MOD-POOR-BAD	BAD
27699	MOD	MOD	GOOD	MOD
27777	GOOD	GOOD	-	GOOD
27801	GOOD	GOOD	-	-
27808	MOD	MOD	POOR-BAD	-
27823	HIGH	HIGH	POOR-BAD	MOD
27827	GOOD	GOOD	-	GOOD
27849	GOOD	GOOD	MOD-POOR-BAD	-
27872	GOOD	GOOD	MOD-POOR-BAD	-

WBID	Damage	AWIST	LAMM	CPET
27888	MOD	MOD	-	-
27900	MOD	MOD	MOD	MOD
27912	MOD	MOD	MOD-POOR-BAD	-
27922	GOOD	GOOD	HIGH	MOD
27923	GOOD	GOOD	-	-
27927	MOD	MOD	GOOD	POOR
28076	BAD	POOR-BAD	MOD-POOR-BAD	GOOD
28130	POOR	POOR-BAD	MOD-POOR-BAD	POOR
28370	MOD	MOD	MOD-POOR-BAD	BAD
28905	POOR	POOR-BAD	POOR-BAD	MOD
28965	HIGH	HIGH	HIGH	HIGH
28986	HIGH	HIGH	HIGH	HIGH
29000	GOOD	GOOD	HIGH	HIGH
29008	POOR	POOR-BAD	POOR-BAD	MOD
29021	HIGH	HIGH	GOOD	GOOD
29045	GOOD	GOOD	POOR-BAD	MOD
29052	GOOD	GOOD	HIGH	HIGH
29062	GOOD	GOOD	HIGH	HIGH
29081	GOOD	GOOD	GOOD	GOOD
29153	POOR	POOR-BAD	POOR-BAD	GOOD
29179	HIGH	HIGH	POOR-BAD	POOR
29181	GOOD	GOOD	POOR-BAD	GOOD
29183	GOOD	GOOD	HIGH	HIGH
29184	HIGH	HIGH	-	-
29215	HIGH	HIGH	HIGH	MOD
29246	GOOD	GOOD	GOOD	HIGH
29252	GOOD	GOOD	HIGH	MOD
29285	POOR	POOR-BAD	POOR-BAD	GOOD
29290	MOD	MOD	GOOD	POOR
31104	BAD	POOR-BAD	MOD-POOR-BAD	MOD
33730	HIGH	HIGH	HIGH	-
33803	GOOD	GOOD	HIGH	GOOD
33836	GOOD	GOOD	-	GOOD
33843	MOD	MOD	GOOD	-
33998	MOD	MOD	-	-
34002	GOOD	GOOD	HIGH	-
34039	MOD	MOD	-	GOOD
34061	MOD	MOD	MOD	-

WBID	Damage	AWIST	LAMM	CPET
34243	MOD	MOD	POOR-BAD	-
34249	MOD	MOD	POOR-BAD	-
34319	MOD	MOD	MOD	MOD
34363	MOD	MOD	POOR-BAD	-
34375	GOOD	GOOD	MOD-POOR-BAD	-
34382	GOOD	GOOD	GOOD	-
34390	MOD	MOD	MOD-POOR-BAD	-
34400	BAD	POOR-BAD	POOR-BAD	MOD
34632	GOOD	GOOD	-	-
34635	MOD	MOD	POOR-BAD	-
34668	MOD	MOD	GOOD	-
34987	HIGH	HIGH	HIGH	-
35233	MOD	MOD	MOD	-
35262	POOR	POOR-BAD	MOD	-
35561	BAD	POOR-BAD	POOR-BAD	-
35578	MOD	MOD	GOOD	POOR
35650	MOD	MOD	POOR-BAD	-
36267	MOD	MOD	POOR-BAD	MOD
36405	GOOD	GOOD	HIGH	HIGH
38390	POOR	POOR-BAD	MOD	MOD
38394	MOD	MOD	-	MOD
38409	BAD	POOR-BAD	POOR-BAD	MOD
38422	HIGH	HIGH	HIGH	-
38525	MOD	MOD	GOOD	MOD
38907	GOOD	GOOD	HIGH	MOD
46232	MOD	MOD	HIGH	GOOD
46279	GOOD	GOOD	HIGH	HIGH
99999	POOR	POOR-BAD	-	-

LAMM Humic	Damage Matrix				
	High	Good	Moderate	Poor	Bad
High	15	5	1	0	0
Good	3	1	2	0	0
Moderate -Poor- Bad	10	5	4	1	2

Table 8: Cross classification table comparing predictions from LAMM for humic waters with those from the damage matrix

LAMM Clear	Damage Matrix				
	High	**Good**	**Moderate**	**Poor**	**Bad**
High	14	8	0	0	0
Good	3	7	7	0	0
Moderate	0	1	5	2	0
Poor-Bad	2	2	7	5	3

Table 9: Cross classification table comparing predictions from LAMM for clear waters with those from the damage matrix

CPET	Damage Matrix				
	High	**Good**	**Moderate**	**Poor**	**Bad**
High	5	7	0	0	0
Good	3	5	2	2	1
Moderate	3	4	6	3	3
Poor	1	0	3	1	0
Bad	0	0	2	0	0

Table 10: Cross classification table comparing predictions from CPET with those from the damage matrix

	CPET	**LAMM**	**AWIST**	**AWIST OOB**
No. of comparisons with damage matrix	52	105	138	134
% of comparisons match with damage matrix	34.62%	54.29%	99.27%	50.00%

Table 11: Summary of comparisons between CPET, LAMM (clear and humic waters combined), AWIST and AWIST OOB and the damage matrix

The number of comparisons is low for CPET, currently, and as a result it is difficult to place too much weight on the low agreement between this method and the damage matrix. We note that CPET predicts five classes and as such will tend to have lower correctly classified % than the other tools. 111 samples in the WFD60 training set also had predictions of status class from LAMM, with approximately 50% of these assigned to the same class by LAMM and the damage matrix. The OOB predictions from AWIST perform similarly to LAMM, whilst, as we saw earlier, the actual AWIST predictions correctly identify the damage matrix class in 99% of cases. The difference in the number of comparisons between AWIST and AWIST OOB is due to the deletion of 3 outliers from the training data used to fit the random forest in AWIST, and consequently for which OOB predictions are available.

Given the biased sampling in the training set and the problems this causes for resampling and cross-validation methods to assess predictive ability we feel that the AWIST OOB is a worst case measure of the predictive ability of the tool, whilst the actual predictions represent, logically a best case. The real predictive ability of the tool likely lies somewhere beteen these two extremes. Despite this, the AWIST OOB error rate is comparable with LAMM and is significantly better than CPET in predicting the damage

matrix class. It should be noted that LAMM was produced using data from Phase I of WFD60 and as such, the performance statistics in Table 11 for this tool also represent a partial best case arising from the use of the training data to test the tool. As a result, the apparent performance of AWIST (99% correctly classified) is remarkably high.

We should caution however in over interpretation of these results. AWIST was designed to predict the damage matrix classes and employs a state of the art statistical machine learning tool (random forests) to identify an appropriate set of decision rules to achieve this. The other two tools, CPET and LAMM, were not explicitly designed to predict damage matrix classes. Indeed, these tools were not designed to predict any classes (categories) directly. Instead, an underlying metric scale is split into regions corresponding to the normative definitions of the WFD. As such, we should expect AWIST to do somewhat better than the other tools in predicting the damage matrix classes. How better we expect AWIST to perform, though is unclear.

A further, and related, complication is that is this comparison we have assumed that the damage matrix is the gold-standard in terms of being indicative of WFD status class. Whilst Monteith and Simpson (2007) presented a substantial body of evidence to support the decisions that went into the construction of the damage matrix and the boundaries it contains, we have not formally tested the damage matrix against sites of know status. This is because we currently do not have an independent, direct measurement of state change over reference from palaeoecological data for sites that do not take part in the creation of the damage matrix, nor a well-defined metric to equate species turn-over in diatoms assemblages between the reference sample and the present day core-top assemblage. As such, we do not know the true status of sites in terms of state change over reference conditions with which to evaluate the various tools and we must rely upon inter-comparisons between the various raters (tools).

It is also relevant to compare the EQRs produced by the various tools. Figure 9 Shows comparisons between the EQRs generated by AWIST and each of LAMM (humic and clear), CPET and LAMM combined. In general, the EQRs generated by AWIST are higher than those of the other tools at the low end of the range, and lower and high end of the range, suggesting some truncation relative to the other tools. This is to be expected to some degree as AWIST EQRs are computed as a weighted average of the initial scores assigned to each of the four classes. As such there will be some regression towards the average of these initial scores. A further difficulty in comparing the sets of EQRs in Figure 9a-c is that there are relative few data in each. In all cases there is general agreement between the tools, though much scatter is produced, resulting in part due to the unique nature of the AWIST EQR and the much lower performance of the other two tools in predicting the Damage matrix classes.

2.6   Temporal assessment of AWIST status

To investigate how AWIST performs when supplied with macroinvertebrate samples collected through time at individual sites, we ran the tool on the lakes sites that form part of the UK Acid Waters Monitoring Network (UK AWMN). In total 12 lake sites we used in this analysis, however Loch Coire Fionnaraich is only been sampled as part of the UK AWMN since 2002, so represents less of a test of AWIST in this regard. Details of the analysed sites are shown in Table 12.

Data from each site were processed as per the WFD60 training set, by aggregating counts from individual kick samples taken on a single site visit per year. AWIST was then run on the samples and the results presented in EQR form visually to represent a time series. This plot is shown in Figure 10.

Site	Code	UK Grid Ref	Altitude (m)	Lake Area (ha)
Loch Coire nan Arr	ARR	NG808422	125	14.4
Lochnagar	NAGA	NO252859	788	9.9
Loch Chon	CHN	NN421051	92	105.7
Loch Tinker	TINK	NN445068	418	11.1
Round Loch of Glenhead	RLGH	NX450804	298	12.7
Loch Grannoch	LGR	NX542700	214	111.4
Scoat Tarn	SCOATT	NY159104	598	4.3
Burnmoor Tarn	BURNMT	NY184044	253	23.9
Llyn Llagi	LAG	SH649483	375	5.1
Llyn Cwm Mynach	MYN	SH678238	278	5.7
Blue Lough	BLU	J327252	335	2.1
Loch Coire Fionnaraich	VNG9402	NG945498	236	9.3

Table 12: UK AWMN lake sites used in the temporal assessment of AWIST.

The UK AWMN sites represent a range of acid status and acid sensitivities as well as site types across a wide gradient of acid deposition. As such the AWIST predicted EQRs range from low (NAGA, BLU, and SCOATT) in the most severely acidified sites to high (ARR, BURNMT, and VNG9402) in the least sensitive or minimally impacted sites.

Several sites exhibit trends in the AWIST predicted EQR, particularly TINK, BURNMT, CHON and LAG, towards higher status throughout the period of monitoring, presumably in response to hydrochemical recovery from acidification following marked sulphur emissions reductions. These results are in general agreement with observed changes in macroinvertebrate assemblages and trends towards recovery observed in the raw UK AWMN data (Monteith *et al.* 2005).

The results for individual sites shown considerable inter-annual variability, of the order of c. 0.1 EQR in some sites. This is likely the result of inter-annual variations in macroinvertebrate assemblages arising from inter-annual variations in climate and the effects of episodic acidification via increased precipitation. For the two control sites on the UK AWMN, ARR and VNG9402, there is good consistency in the overall EQR score despite this inter-annual variability, as one would expect to find in minimally impacted sites where the only forcing factors will be climatically related, and therefore, to some degree random.

These results highlight an important point of note; assessment of site status should be based on repeated samples over several years of monitoring in order to quantify the range of natural variation in the macroinvertebrate community and the effect this has on the predicted EQR and WFD class status.
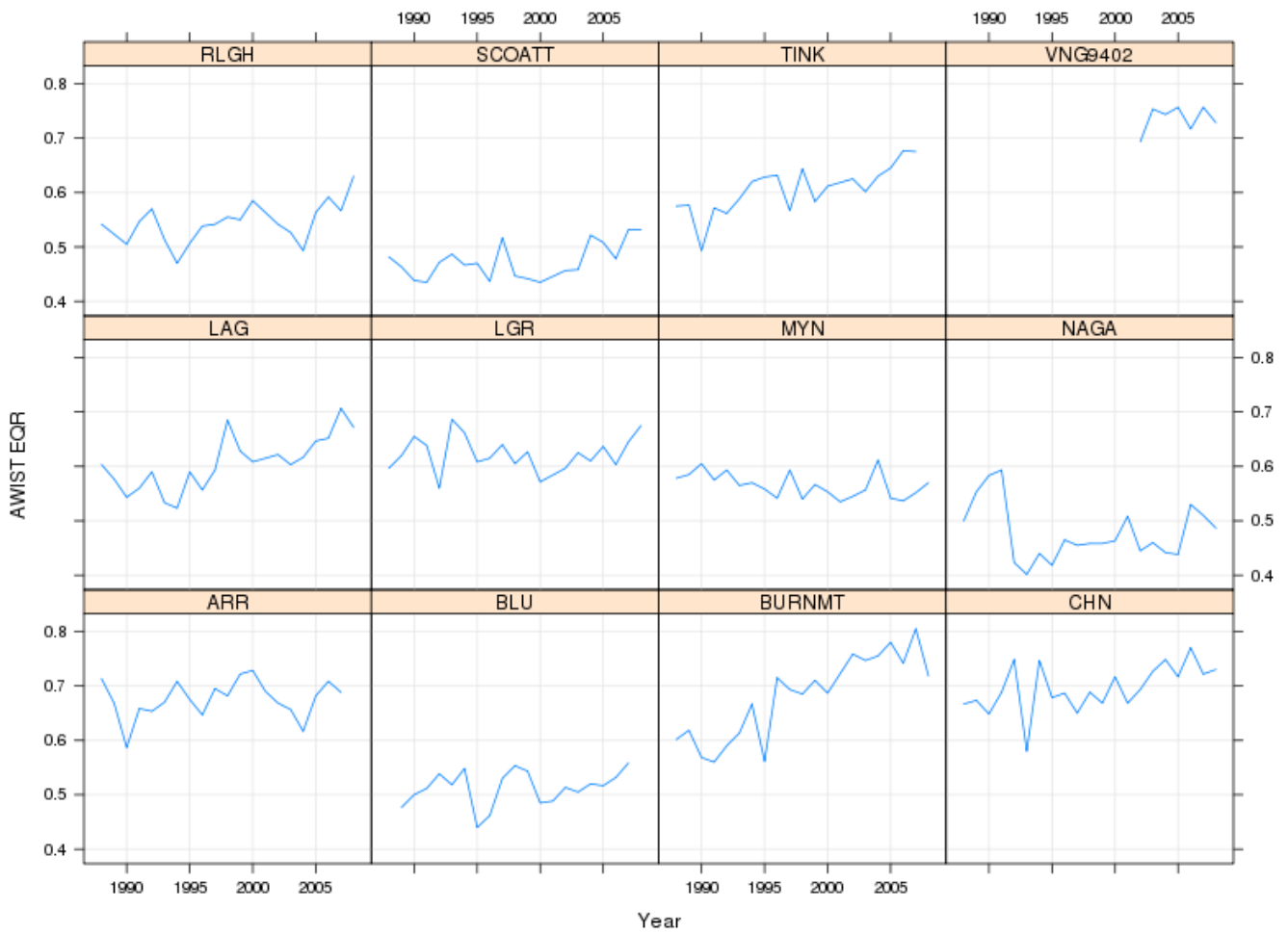
Figure 10: Trellis time-series plots of AWIST EQR for macroinvertebrate time-series from UK AWMN lakes sites for the period 1988 to 2008.
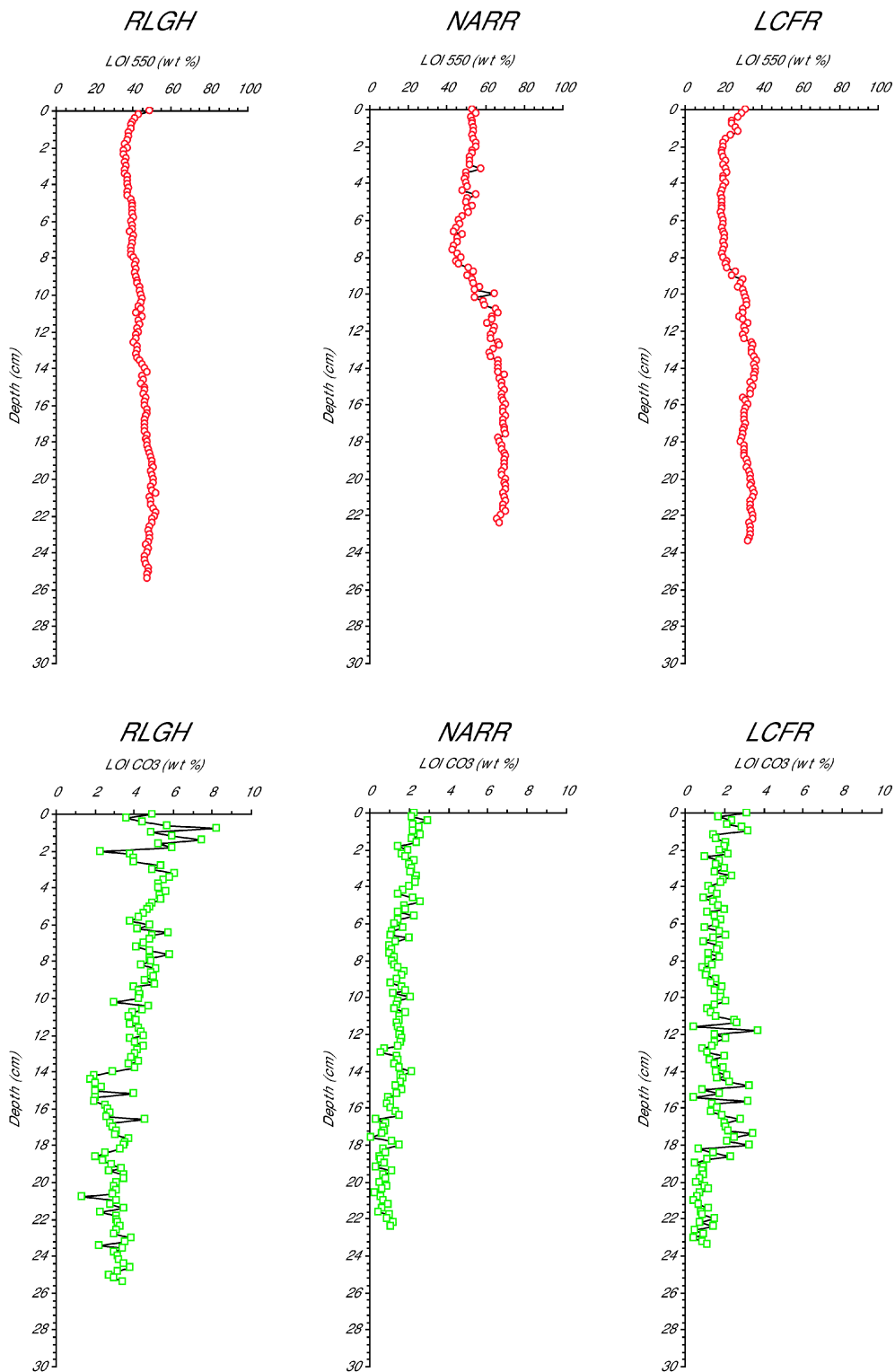
Figure 11: Loss on ignition profiles from Round Loch of Glenhead (RLGH-A), Loch Narroch (NARR-A) and Loch Coire Fionnaraich (LCFR-A)

3 Palaeoecological Study

3.1   Introduction

The only true method of determining what status a water body is in is to utilise the palaeoecological approach and assess state change over reference conditions from sub-fossil remains or aquatic organisms or other proxies extracted from sediment core records. In this section we report upon a multiproxy study of invertebrate remains and diatom sub-fossils at three acid-sensitive lakes in Scotland and compare these results to the predicted status from AWIST.

3.2   Study Sites

Paleoecological analyses were undertaken at three sites, one minimally impacted yet acid sensitive site, Loch Coire Fionnaraich, from north west Scotland and two acidified sites from the acid sensitive Galloway region of southern Scotland (the Round Loch of Glenhead and Loch Narroch).

3.2.1   Loch Coire Fionnaraich (NG945498, 236 m a.s.l.)

Loch Coire Fionnaraich (LCFR) in north-west Scotland, covers an area of 9.3 hectares and is fed by five minor streams, and one large one, the Allt Bealach Bàn. Dammed by moraine, the loch drains five and a half kilometres, almost due south, into the River Carron via the Fionn-abhainn outflow stream. The bathymetry of the loch shows two basins in the centre of the loch, the larger and deeper of which reaches a depth of 14m (AWMN Ref). Palaeolimnological analysis of cores taken in 2001 show that the lake has not been subject to significant acidification (Pla et al. 2009).

3.2.2   Round Loch of Glenhead (NX449803, 298 m a.s.l.)

The Round Loch of Glenhead (RLGH) in the Galloway region of south-west Scotland is 12.7 ha in area and receives drainage from minor streams and catchment blanket peats. The outflow drains to the south-west into the Glenhead Burn and Loch Trool. The loch bathymetry indicates a single deep basin (maximum depth 13.5 m) offset to the south with slopes rising gently away from the southern shore. An island is located just off the western shore some 250 m from the outflow (Monteith et al 2005).

Palaeolimnological evidence from previous research (Allott, Harriman, and Battarbee 1992; Flower, Battarbee and Appleby 1987) indicates that the loch has been acidic throughout the post-glacial period, although a stable pH of circa 5.5 declined rapidly from approximately 1870 to around 4.7 by the late 1960s and that a very slight rise has occurred since the late 1970s.

The loch is chronically acid (mean pH 4.93), poorly buffered and exhibits negative alkalinity and low calcium concentrations (mean 0.63 mg 1-1). Observed trends within the chemistry over 15 years include decreases in base cations, aluminium species and SO^4 as well as an increase in DOC.

3.2.3   Loch Narroch (NX452815, 328 m a.s.l.)

Loch Narroch (NARR) (3.3 ha), 1km north of Round Loch of Glenhead in the Galloway region of south-west Scotland. Similarly, the loch and its catchment are situated on the Loch Doon granite intrusion.

3.3   Coring and sampling strategy

To collect sufficient material for the palaeoecological analyses, multiple cores were taken from the deepest point of each of the lochs using a Renberg gravity corer. A master 'A' core used for 210Pb dating and four other cores were taken from the same position on the loch. Cores were sliced and bagged in the field; the 'A' core at 0.2 cm intervals, cores B-E at 0.5 cm and refrigerated until sub-sampling for subsequent analyses.

3.4   Loss-on-ignition (LOI)

In the ECRC laboratory all core sediments were sub-sampled and processed for loss-on-ignition. Wet samples were weighed, oven dried at 105 ºC for 12 hours and combusted in a muffle furnace at 550 ºC for 2 hours and at 950 ºC for 4 hours to calculate organic matter and carbonate content (Dean, 1974, Heiri et al, 2001). The results of the LOI analyses are shown in Figure 11.

3.4.1   Round Loch of Glenhead (RLGH)

In RLGH the higher surface value (48%) decreases down to 35% by 2cm depth. From 2cm depth LOI 550 C values gradually but steadily increase with depth to ~22-24cm. The low carbonate (<10%) values are not unexpected considering the known historic and monitored low pH. Nonetheless there is a visible trend of decreasing carbonate with depth.

3.4.2   Loch Narroch (NARR)

The surface LOI 550 value (53%) of NARR gradually decreases down to 45% at 8 cm. Below 8cm LOI 550 values increase slightly to 65% at 11cm and remain similar to the bottom of the core. Carbonate values are lower (<4.5%) than RLGH but also show a gradual decrease with depth

3.4.3   Loch Coire Fionnaraich (LCFR)

The surface LOI 550 value (31%) of LCFR decreases to 20% by 2cm and remain similar down to 8 cm. Below this is a slow increase to a maximum of 37% at 14cm. Carbonate values are similarly low in the cores (<4%).

3.4.4   Cross-Correlation between cores at each site

To obtain enough material for the separate analyses, multiple cores were taken at each site and LOI measurements were used to cross-correlate approximate sediment depths of equal age to tie-in to the dated 'A' core. Sub-samples from the identified intervals in the cores were extracted, homogenised and then sub-sampled again for diatoms, chironomids, and coleoptera/trichoptera. The separate analysts received 10 bagged samples from each site in descending stratigraphic order; 1 being the surface.

	LCFR			RLGH			NARR	
	Depth (cm)	**Approx. Year**		**Depth (cm)**	**Approx. Year**		**Depth (cm)**	**Approx. Year**
Bag 1	0-0.2	2008	Bag 1	0-0.2	2008	Bag 1	0-0.2	1963
Bag 2	0.8-1	2002	Bag 2	1.4-1.6	~1995	Bag 2	0.8-1	1955
Bag 3	1.8-2	1992	Bag 3	2.4-2.6	~1985	Bag 3	1.4-1.6	~1950
Bag 4	2.6-2.8	1983	Bag 4	3.4-3.6	~1970	Bag 4	1.8-2	~1946
Bag 5	3-3.2	1980	Bag 5	4.4-4.6	~1960	Bag 5	2-2.2	~1944
Bag 6	5-5.2	1954	Bag 6	6.2-6.4	~1944	Bag 6	3.8-4	~1925
Bag 7	7.4-7.6	~1920	Bag 7	8-8.2	~1926	Bag 7	4.8-5	~1913
Bag 8	7.8-8	~1914	Bag 8	8.6-8.8	~1920	Bag 8	9.8-10	~1840
Bag 9	8.8-9	~1902	Bag 9	9.4-9.6	~1912	Bag 9	12.8-13	< 1830
Bag 10	10.4-10.6	~1890	Bag 10	10.8-11	~1895	Bag 10	16.2-16.4	< 1800

Table 13: Core depths and ages of samples used for the suite of palaeoecological analyses.

3.5   Radiometric dating of cores from LCFR-A, NARR-A and RLGH-A

[210]Pb (half-life 22.3 years) is a naturally-occurring radionuclide, derived from atmospheric fallout (termed unsupported ^{210}Pb). [137]Cs (half-life 30 years) and [241]Am are artificially produced radionuclides, introduced to the environment by atmospheric fallout from nuclear weapons testing and nuclear reactor accidents. They have been extensively used in the dating of recent sediments.

Dried sediment samples from cores LCFR-A, NARR-A and RLGH-A were analysed for ^{210}Pb, [226]Ra, ^{137}Cs and [241]Am by direct gamma assay in the Bloomsbury Environmental Isotope Facility (BEIF) at University College London, using ORTEC HPGe GWL series well-type coaxial low background intrinsic germanium detector. [210]Pb was determined via its gamma emissions at 46.5keV, and [226]Ra by the 295keV and 352keV gamma rays emitted by its daughter isotope [214]Pb following 3 weeks storage in sealed containers to allow radioactive equilibration. [137]Cs and [241]Am were measured by their emissions at 662kev and 59.5kev (Appleby et al, 1986). The absolute efficiencies of the detector were determined using calibrated sources and sediment samples of known activity. Corrections were made for the effect of self absorption of low energy gamma rays within the sample (Appleby et al, 1992).

3.5.1   Loch Coire Fionnaraich: LCFR-A

Equilibrium of total ^{210}Pb activity with the supporting ^{210}Pb measured as [226]Ra activity appears to occur at c. 15 cm of the core (Figure 12a). Unsupported ^{210}Pb activities, calculated by subtracting [226]Ra activity from total ^{210}Pb activity, decline more or less exponentially with depth (Figure 12b), suggesting that sediment accumulation rates have been relatively uniform in this section of the core.

The ^{137}Cs activity versus depth profile (Figure 12c) has a well-resolved peak at 4.9 cm, and this peak

should be derived from the fallout maximum from the atmospheric testing of nuclear weapons. The ^{137}Cs peak showing the 1963 layer was confirmed by detectable [241]Am at this depth.

[210]Pb chronologies were calculated using the CRS dating model (Appleby and Oldfield, 1978). The CRS dating model places the 1963 layer at 4.5 cm. which is in relatively good agreement with the ^{137}Cs and [241]Am records, suggesting the CRS dates are reasonable. Sedimentation rates calculated by unsupported ^{210}Pb data show relatively stable sediment accumulations in the core with an average sedimentation rate at 0.019 g cm^{-2} yr^{-1}.

Depth	Dry Mass	[210]Pb						Cum. Unsupported [210]Pb	
		Total		Supported		Unsupported			
cm	g cm^{-2}	Bq Kg^{-1}	±	Bq Kg^{-1}	±	Bq Kg^{-1}	±	Bq m^{-2}	±
0.8-1	0.0911	633.89	27.29	32.52	4.35	601.37	27.63	599.5	34.9
2.8-3	0.4388	377.87	19.36	28.01	3.34	349.86	19.65	2213.7	112.9
4.8-5	0.8503	186.53	15.68	25.14	3.09	161.39	15.98	3216.2	144.8
5.8-6	1.1094	106.94	15.48	26.04	3.17	80.9	15.8	3518.1	153.1
8.8-9	1.8799	53.18	10.01	35.85	2.28	17.33	10.27	3836	177.6
10.4-10-6	2.1774	51.12	11.17	33.66	2.41	17.46	11.43	3887.8	182.5
11.8-12	2.4475	44.66	10.19	37.51	2.21	7.15	10.43	3919	185.1
13.4-13.6	2.7179	50.16	10.16	34.21	2.34	15.95	10.43	3948.6	187.2
14.8-15	2.9753	42.25	8.25	36.45	1.93	5.8	8.47	3974.4	189
17.8-18	3.524	27.86	7.36	30.81	1.76	-2.95	7.57		
20.8-21	4.0647	36.08	8.71	35.9	2.05	0.18	8.95		

Table 14: [210]Pb concentrations in core LCFR-A

Depth	[137]Cs		[241]Am	
cm	Bq Kg^{-1}	±	Bq Kg^{-1}	±
0.8-1	230.84	6	0	0
2.8-3	283.61	5.76	0	0
4.8-5	330.36	5.73	4.65	1.39
5.8-6	115.76	3.9	0	0
8.8-9	17.71	1.42	0	0
10.4-10-6	23.91	1.7	0	0
11.8-12	13.39	1.33	0	0
13.4-13.6	11.69	1.37	0	0
14.8-15	7.53	1.02	0	0
17.8-18	1.27	0.89	0	0
20.8-21	1.44	0.96	0	0

Table 15: Artificial fallout radionuclide concentrations in core LCFR-A

Depth	Dry mass	Chronology			Sedimentation Rate		
cm	g cm^{-2}	Date AD	Age yr	±	g cm^{-2} yr^{-1}	cm yr^{-1}	± %
0	0	2008	0				
1	0.1077	2002	6	2	0.0174	0.118	7.8
2	0.2732	1992	16	2	0.0166	0.101	9.6
3	0.4388	1982	26	3	0.0158	0.083	11.4
4	0.6554	1968	40	4	0.0153	0.074	17.3
5	0.8739	1954	54	6	0.0151	0.067	23.9
6	1.1094	1939	69	10	0.0179	0.071	36.2
7	1.3662	1927	81	15	0.0208	0.085	57.4
8	1.6231	1914	94	20	0.0237	0.098	78.6
9	1.8799	1902	106	26	0.0265	0.112	99.9
10	2.0782	1893	115	29	0.0202	0.097	112.5
11	2.2674	1884	124	33	0.018	0.097	125.5
12	2.4475	1875	133	36	0.0199	0.111	139

Table 16: ^{210}Pb chronology of core LCFR-A. Depth/Ages calculated by CRS dating model
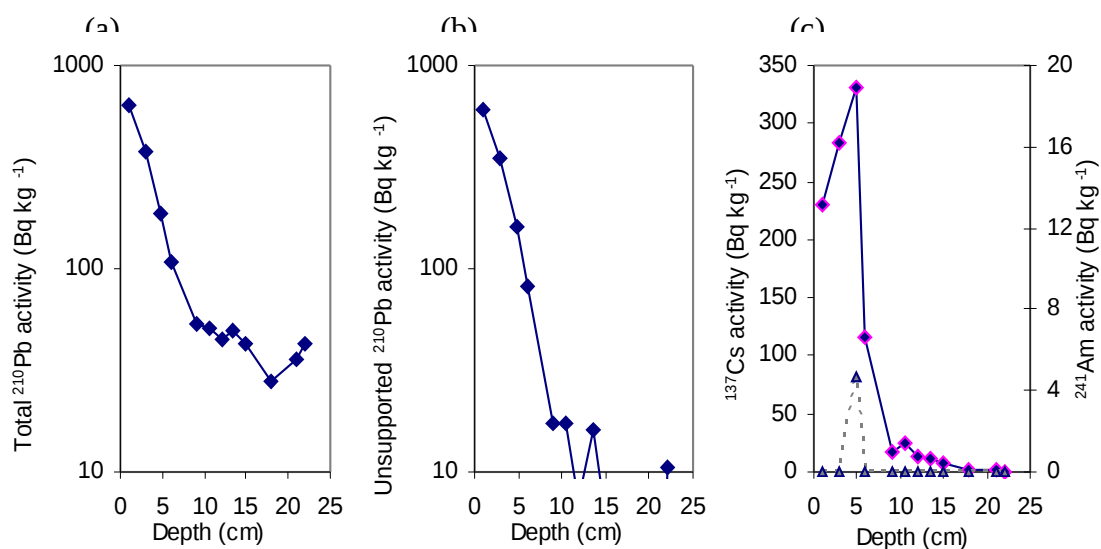
Figure 12: Fallout radionuclide concentrations in core LCFR-A showing (a) total ^{210}Pb, (b) unsupported ^{210}Pb, and (c) ^{137}Cs and ^{241}Am concentrations versus depth

Figure 13: Radiometric chronology of core LCFR-A showing the CRS model ^{210}Pb dates and sedimentation rates (dashed line, right hand axis)

3.5.2   Loch Narroch: NARR-A

Total ^{210}Pb activity reaches equilibrium depth with the supporting ^{210}Pb at c. 12 cm (Figure 14a). Unsupported ^{210}Pb activities also decline more or less exponentially with depth (Figure 14b) in this core, suggesting that sediment accumulation rates are relatively uniform in this section of the core.

The ^{137}Cs activity versus depth profile (Figure 14c) shows that ^{137}Cs activities decline from the surface downwards, the ^{241}Am activities also decline from the surface of the core, suggesting the high activity of ^{137}Cs in the surface sediments is derived from the fallout maximum from the atmospheric atomic bomb testing in 1963. This means that the sediments in the core have been deposited before 1963 and the contemporary surface of the core is missing. It was also noted during the extraction of the NARR cores that the sediment-water interface was visibly disturbed.

^{210}Pb chronologies were calculated using the CRS and CIC dating models (Appleby, 2001), and the results of the two models agreed well with each other. As the ^{137}Cs and ^{241}Am records suggested that the surface sediments were missing in the core, the final chronologies were calculated using the CRS model and taking the surface of the core as 1963. Calculation base on unsupported ^{210}Pb data by using the CRS model shows that sediment accumulations were fairly stable at c. 0.0092 g cm^{-2} yr^{-1}.

Depth	Dry Mass	^{210}Pb Total		Supported		Unsupported		Cum. Unsupported ^{210}Pb	
cm	g cm^{-2}	Bq Kg^{-1}	±	Bq Kg^{-1}	±	Bq Kg^{-1}	±	Bq m^{-2}	±
0-0.2	0.0074	692.53	44.11	77.28	9.45	615.25	45.11	46.1	3.3
0.8-1	0.0707	586.44	30.58	58.24	6.17	528.2	31.2	407.2	26.2
1.8-2	0.1632	396.19	29.5	76.42	6.71	319.77	30.25	791.4	42
2.8-3	0.2451	350.58	24.05	68.44	5.68	282.14	24.71	1037.5	50.1
3.8-4	0.3559	251.46	12.12	85.85	3.17	165.61	12.53	1280	56.4
4.8-5	0.454	222	18.46	72.77	5.01	149.23	19.13	1434.3	58.8
5.8-6	0.5846	163.07	16.02	81.93	4.18	81.14	16.56	1580.2	63.4
6.8-7	0.7053	132.3	19.14	84.23	5.59	48.07	19.94	1656.5	67
7.8-8	0.8661	112.83	14.44	78.96	4	33.87	14.98	1721.6	72.7
9.8-10	1.1406	84.6	12.46	70.13	3.54	14.47	12.95	1784.3	80.8
11.8-12	1.388	88.77	10.31	83.82	3.12	4.95	10.77	1806.2	86.8
12.8-13	1.5022	96.82	10.65	74.27	3.08	22.55	11.09	1819.5	88.2
13.8-14	1.6358	88.17	12.68	89.39	3.89	-1.22	13.26	1833.7	89.5

Table 17: ^{210}Pb concentrations in core NARR-A

Depth	^{137}Cs		^{241}Am	
cm	Bq Kg^{-1}	±	Bq Kg^{-1}	±
0-0.2	398.15	12.46	12.64	3.77
0.8-1	363.3	9.07	4.5	2.63
1.8-2	288.78	8.88	4.83	2.91
2.8-3	218.09	6.8	0	0
3.8-4	121.78	2.89	1.74	0.66
4.8-5	107.64	4.49	0	0
5.8-6	63.48	3.28	0	0
6.8-7	67.76	4.27	0	0
7.8-8	46.93	2.91	0	0
9.8-10	30.07	2.23	0	0
11.8-12	27.85	1.82	0	0
12.8-13	28.79	1.81	0	0
13.8-14	23.61	2.16	0	0

Table 18: Artificial fallout radionuclide concentrations in core NARR-A

Depth	Dry mass	Chronology			Sedimentation Rate		
cm	g cm^{-2}	Date AD	Age yr	±	g cm^{-2} yr^{-1}	cm yr^{-1}	± %
0	0	1963	45				
0-0.2	0.0074	1962	46	2	0.009	0.114	9.2
0.8-1	0.0707	1955	53	2	0.0083	0.102	8.9
1.8-2	0.1632	1945	63	2	0.01	0.115	12.7
2.8-3	0.2451	1936	72	3	0.0087	0.09	13.8
3.8-4	0.3559	1924	84	4	0.0102	0.098	16.9
4.8-5	0.454	1913	95	6	0.0081	0.071	24.1
5.8-6	0.5846	1898	110	9	0.0093	0.074	37.3
6.8-7	0.7053	1886	122	13	0.0108	0.076	59.3
7.8-8	0.8661	1870	138	20	0.0093	0.066	77.2
9.8-10	1.1406	1839	169	30	0.0082	0.063	113.8

Table 19: ^{210}Pb chronology of core NARR-A. Depth/Ages calculated by CRS dating model
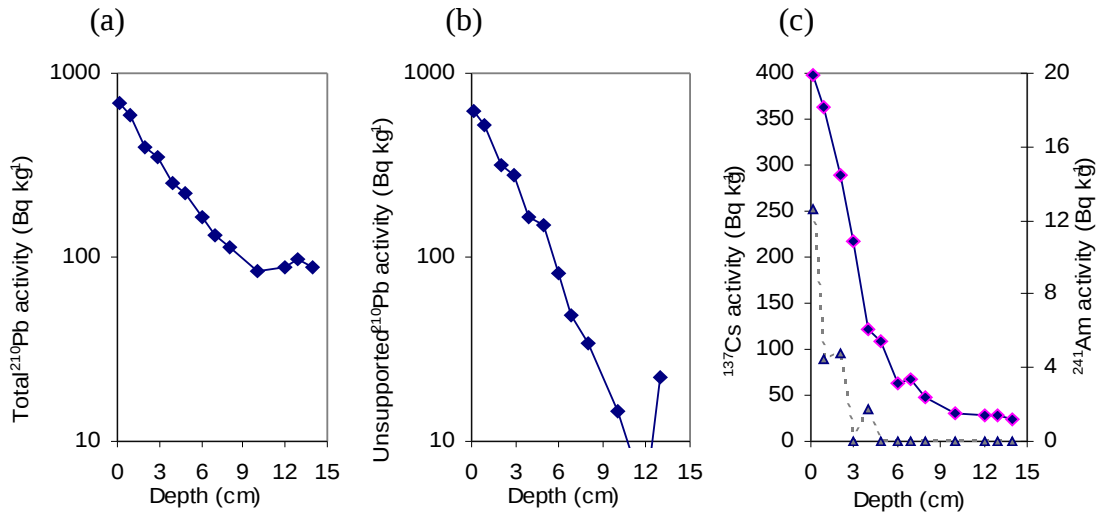
Figure 14: Fallout radionuclide concentrations in core NARR-A showing (a) total ^{210}Pb, (b) unsupported ^{210}Pb, and (c) ^{137}Cs and ^{241}Am concentrations versus depth
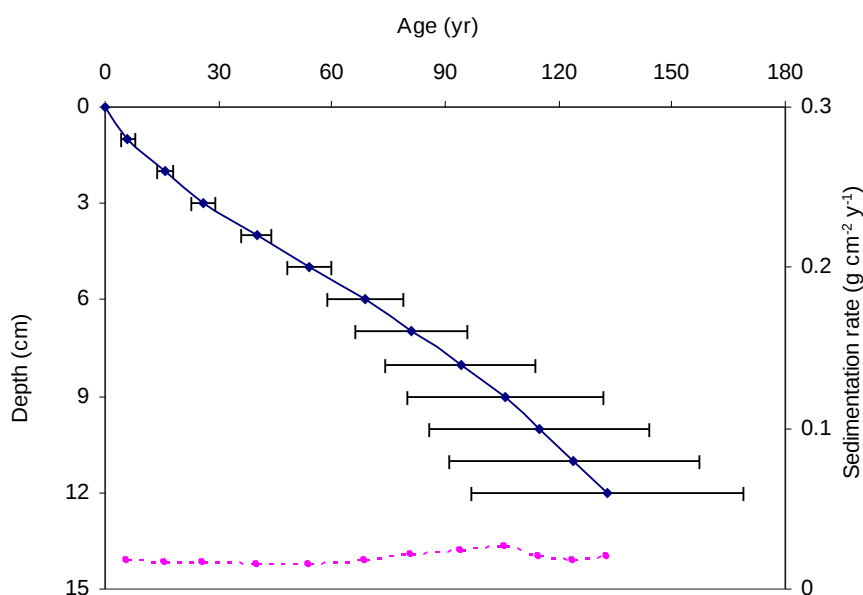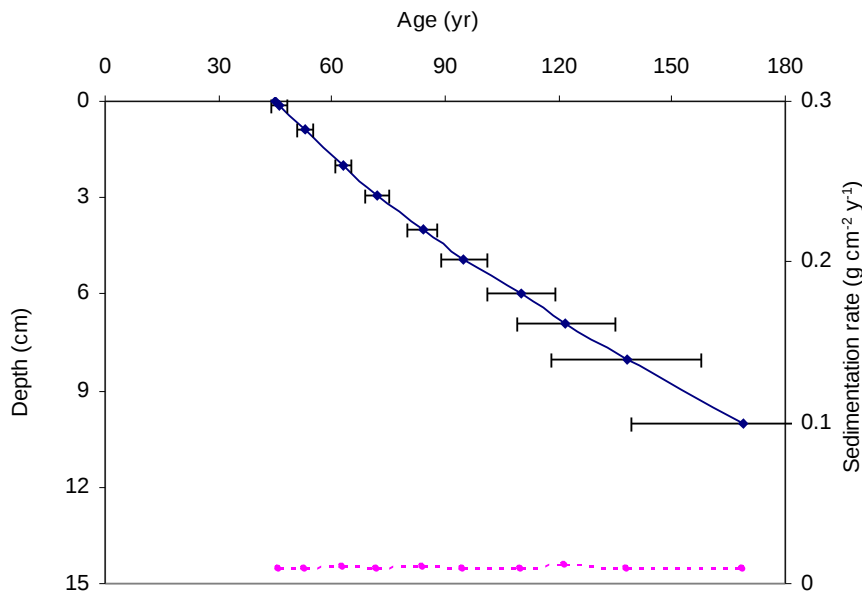

Figure 15: Radiometric chronology of core NARR-A showing the CRS model ^{210}Pb dates and sedimentation rates (dashed line, right hand axis)

3.5.3   Round Loch of Glenhead: RLGH-A

Total ^{210}Pb activity reaches equilibrium depth with the supporting ^{210}Pb at c. 16 cm (Figure 16a). Unsupported ^{210}Pb activities decline again more or less exponentially with depth (Figure R b) in this core, suggesting that sediment accumulation rates are relatively uniform in this part of the core.

The ^{137}Cs activity versus depth profile (Figure 16b) shows that ^{137}Cs activities decline from the surface downwards. The ^{241}Am activity versus depth profile (Figure 16c) has a relatively well-resolved peak at 3.9 cm, which was recording the 1963 fallout maximum from the atmospheric testing of nuclear weapons.

This suggests that the high ^{137}Cs activity in the surface sediments was derived from the 1986 Chernobyl accident, and the 1963 peak of ^{137}Cs fallout was obscured by the high fallout of the Chernobyl accident.

The raw CRS model dated 1963 layer at 5.1 cm, slightly deeper than the depth suggested by the ^{241}Am record in the core. ^{210}Pb chronologies were calculated using the CRS dating model and referring to the ^{241}Am record and shown in Table 9. Calculation of sediment accumulations also shows that they were fairly stable at c. 0.0077 g cm^{-2} yr^{-1}.

| Depth | Dry Mass | ^{210}Pb | | | | | | Cum. Unsupported ^{210}Pb | |
| | | Total | | Supported | | Unsupported | | | |
cm	g cm^{-2}	Bq Kg^{-1}	±	Bq Kg^{-1}	±	Bq Kg^{-1}	±	Bq m^{-2}	±
0.8-1	0.0469	1294.24	72.53	39.12	16	1255.12	74.27	632.7	40.1
2.8-3	0.2149	952.02	28.42	60.47	5.02	891.55	28.86	2419	128.1
3.8-4	0.2922	709.02	34.19	53.99	6.18	655.03	34.74	3012.3	134.7
4.8-5	0.3788	500.04	31.76	57.22	6.12	442.82	32.34	3481.3	139.7
5.8-6	0.4724	378.42	35.91	66.81	7.85	311.61	36.76	3831	144.2
7.4-7.6	0.5997	310.25	29.18	64.97	6.1	245.28	29.81	4183.7	151
8.8-9	0.7349	217.97	30.49	98.89	7.4	119.08	31.38	4419.7	156.8
10.4-10.6	0.8686	186.71	16.24	75.57	3.72	111.14	16.66	4573.6	161.6
11.8-12	1.0089	139.78	17.95	88.39	4.56	51.39	18.52	4682.3	163.4
13.4-13.6	1.1445	130.23	17.94	80.89	4.28	49.34	18.44	4750.6	165.4
14.8-15	1.2792	118.3	14.57	98.36	3.61	19.94	15.01	4794.3	167.1
17.8-18	1.5696	106.28	11.59	110.71	3.12	-4.43	12	4816.8	170.7
19.4-19.6	1.7102	126.51	14.2	108.3	3.58	18.21	14.64	4826.5	172.1
20.8-21	1.8659	128.31	25.26	122.82	6.52	5.49	26.09	4843	174.3

Table 20: ^{210}Pb concentrations in core RLGH-A

| Depth | ^{137}Cs | | ^{241}Am | |
cm	Bq Kg^{-1}	±	Bq Kg^{-1}	±
0.8-1	1040.53	25.32	0	0
2.8-3	838.61	9.81	10.92	2.41
3.8-4	457.04	9.64	12.41	2.71
4.8-5	253.2	7.36	10.7	2.53
5.8-6	187.27	7.61	6.84	3.13
7.4-7.6	150.06	6.2	0	0
8.8-9	86.04	5.35	0	0
10.4-10.6	68.45	2.75	0	0
11.8-12	46.94	2.95	0	0
13.4-13.6	40.68	2.78	0	0
14.8-15	25.48	2.14	0	0
17.8-18	13.34	1.42	0	0
19.4-19.6	9.71	1.61	0	0
20.8-21	11.48	3.1	0	0

Table 21: Artificial fallout radionuclide concentrations in core RLGH-A

Depth	Dry mass	Chronology			Sedimentation Rate		
cm	g cm^{-2}	Date AD	Age yr	±	g cm^{-2} yr^{-1}	cm yr^{-1}	± %
0	0	2008	0				
1	0.0549	2002	6	2	0.0085	0.118	10.9
2	0.1349	1990	18	3	0.0072	0.095	14.2
3	0.2149	1979	29	4	0.0059	0.072	17.4
4	0.3009	1964	44	7	0.0071	0.0826	28.8
4.9	0.3788	1956	52	3	0.0095	0.1111	10.6
6	0.4724	1946	62	3	0.0092	0.1083	15
7.5	0.5997	1932	76	4	0.0090	0.1034	17.8
9	0.7349	1917	91	6	0.0087	0.0967	31.5
10.5	0.8686	1901	107	9	0.0076	0.0833	30.6
12	1.0089	1881	127	15	0.0064	0.0697	57.1
13.5	1.1445	1858	150	27	0.0059	0.0652	91.7

Table 22: ^{210}Pb chronology of core RLGH-A. Depth/Ages calculated by CRS dating model
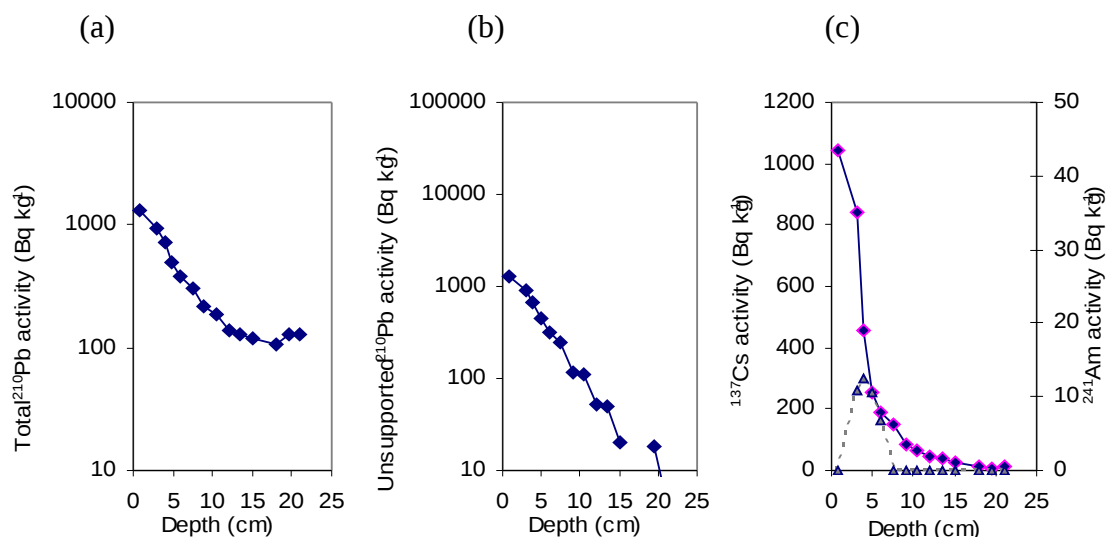
Figure 16: Fallout radionuclide concentrations in core RLGH-A showing (a) total ^{210}Pb, (b) unsupported ^{210}Pb, and (c) ^{137}Cs and ^{241}Am concentrations versus depth
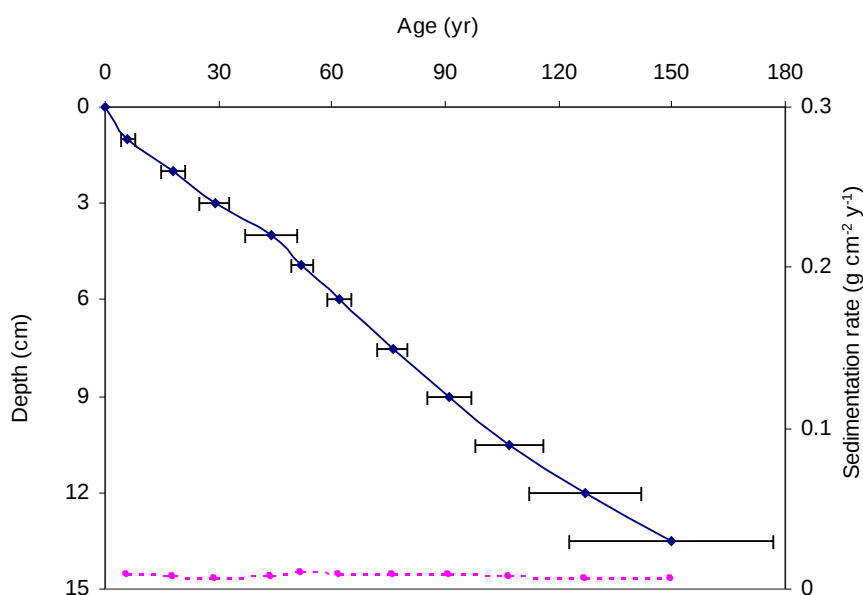
Figure 17: Radiometric chronology of core RLGH-A showing the CRS model ^{210}Pb dates and sedimentation rates (dashed line, right hand axis)

3.6 Chironomid Analysis

3.6.1 Introduction

Chironomid (Insecta: Diptera) non-biting midges are well-established as proxy indicators of palaeoenvironmental change (Walker, 2001). The group is particularly effective as a quantitative indicator of past changes in summer temperature (Brooks, 2006), but chironomids have also been used to detect trophic change (Brooks et al; 2001), changes in lake depth (Korhola et al., 2000) and the impact of acidification (Brodin and Gransberg, 1993). Chironomids have several attributes that make them useful in palaeolimnological studies (Brooks et al., 2007). Their larval head capsules are ubiquitous, diverse and well-preserved in most lake sediment samples. Most specimens are identifiable at least to generic level and often to species morphotype. Many taxa are stenotopic and relatively well-known ecologically so conclusions can be drawn from the characteristic larval assemblages about prevailing environmental conditions when the remains were deposited. By virtue of the winged adult stages, which are blown long distances across the landscape, chironomids respond rapidly to environmental change and the high abundance of their larval head capsules in most lake sediments means that just a few grams of sediment will yield sufficient head capsules for studies at high temporal resolution.

3.6.2 Methods

Preparation of the chironomid samples followed standard laboratory procedures (Brooks et al., 2007). A few grams of sediment of known weight were de-flocculated for 5-10 minutes in KOH heated to 70°C. The sediment was strained through nested sieves of mesh size 212µm and 90µm. The material trapped on the sieves was washed into a petri dish and pipetted in small amounts into a grooved sorting tray. Under a 25-40x dissecting microscope chironomid larval head capsules were picked using fine forceps and stored

Figure 18: Chironomid taxa from Round Loch of Glenhead

in 80% ethanol. Picking chironomid head capsules from sediment samples is time-consuming but previous work (Heiri and Lotter, 2001) has shown that a minimum of 50 are required from each sample to produce statistically meaningful results. For this reason sufficient material was processed to exceed this number. The chironomid larval head capsules were slide mounted in Euparal and identified under a compound microscope at 400x magnification using Brooks et al. (2007) and Weiderholm (1983) and other standard taxonomic references cited therein. Most specimens were identified to generic level and where possible to species morphotype level, following the protocols described in Brooks et al. (2007).

The chironomid assemblage diagrams were made using the program TILIA v. 2.2 (Grimm, 2004). Significant changes in the chironomid assemblages were identified by optimal partitioning using the program ZONE (Juggins, 1991) and by comparison with the broken stick model using the program BSTICK (Bennett, 1996; JM Line and HJB Birks, unpublished).

3.6.3 Results and Interpretation

Round Loch of Glenhead

A total of 41 taxa was identified from the whole sequence. No significant changes in the chironomid assemblage were identified by numerical analysis but a change above sample 7 is visible and has been indicated in Figure 18. Taxon diversity was greatest below this depth and 13 taxa were absent from the upper part of the sequence. Two taxa, *Acamptocladius* and *Eukiefferiella*, were restricted to the upper part of the sequence. Five taxa in the assemblage are characteristic of hygropteric, semi-terrestrial or lotic environments and so cannot necessarily be considered as indicative of limnological conditions. These taxa were relatively rare and largely restricted to the lower part of the sequence. The chironomid fauna was dominated throughout most of the sequence by acidophilic taxa, especially *Psectrocladius sordidellus*-type. However, there was a marked increase in the abundance of acidophilic taxa above sample 7 in the profile, especially *P. sordidellus*-type, *Heterotanytarsus apicalis*-type,
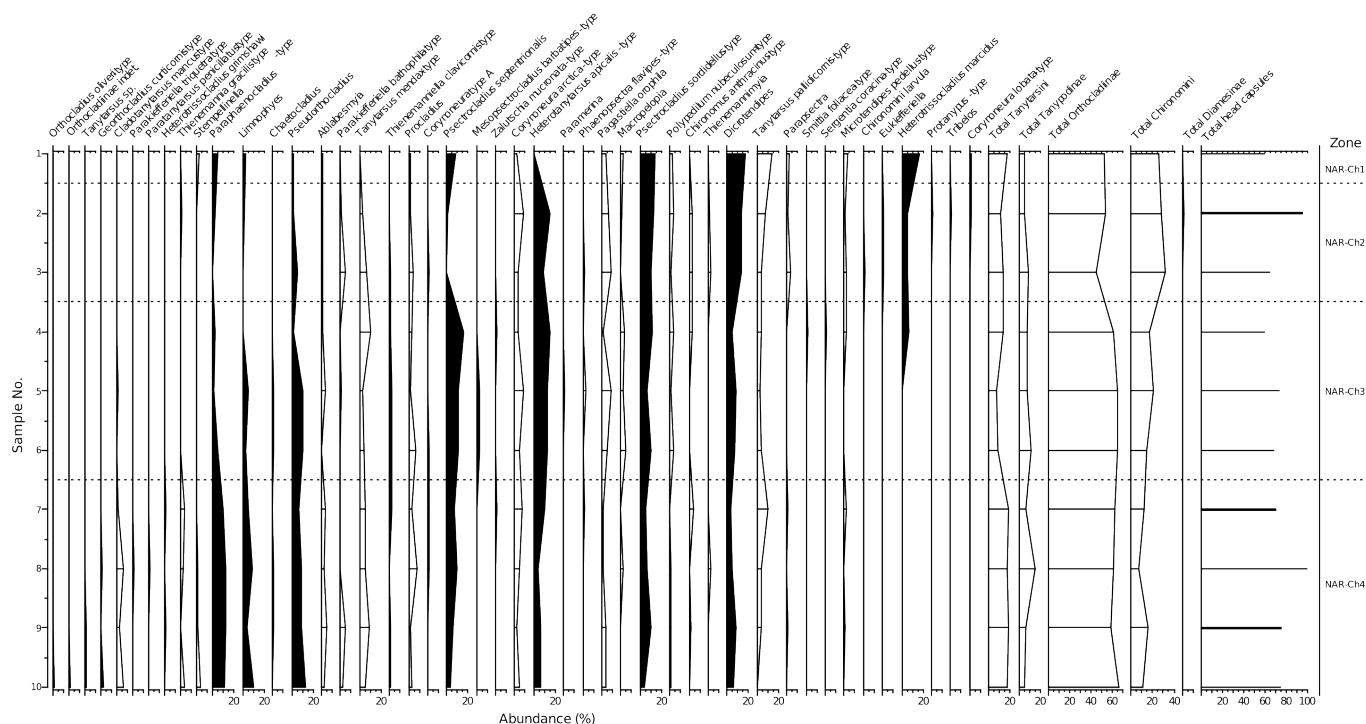
Figure 19: Chironomid taxa from Loch Narroch

Heterotrissocladius marcidus-type, *Dicrotendipes* and *Sergentia coracina*-type. At the same point in the sequence there was a decline in Tanytarsini, considered acidophobic by (Henrikson et al., 1982), in particular *Tanytarsus mendax*-type and *T. pallidicornis*-type, and *Stictochironomus rosenschoeldi*-type, which is also known to be acidophobic (Ilyashuk et al., 2000).

The changes in the chironomid assemblage suggest a response to increasing acidity, especially above sediment sample 7. This has resulted in an overall decrease in chironomid diversity, an increase in the abundance of acidophilic taxa and a decline in the abundance of acidophobic taxa.

Loch Narroch

A total of 42 taxa were identified from the whole sequence (Figure 19). Ten of these taxa are derived either from inflowing streams, hygropetric or semi-terrestrial environments around the lake and so are less likely to have been influenced by lake water quality than the rest of the assemblage. The three commonest of these taxa, *Paraphaenocladius*, *Limnophyes* and *Pseudorthocladius* are most abundant in the lower part of the core, so their subsequent decline may be in response to changes in lake level or the volume of water entering the lake from feeder streams. The fauna includes several acidophilic taxa, such as *Psectrocladius septentrionalis*-type, *P. sordidellus*-type, *Heterotanytarsus apicalis*-type, *Dicrotendipes* and *Heterotrissocladius marcidus*-type, which are relatively abundant in the sequence.

Loch Coire Fionnaraich

A total of 56 taxa was identified from the whole sequence. No significant changes in the chironomid assemblage were identified by numerical analysis but a change above sample 7 is visible and has been indicated in Figure 20. Seventeen lotic, hygropetric or semi-terrestrial taxa were present in the

LOCH COIRE FIONNARAICH: Chironomids
Solid = acidophiles
Pattern = hygropetric or lotic

Figure 20: Chironomid taxa from Loch Coire Fionnaraich

assemblage indicating a considerable influence of stream inflow. There is an increase in *Thienemanniella clavicornis*-type, the most abundant of these taxa, above sample 6. There are only two acidophilic taxa that occur at abundances above 5% but one of these, *Heterotanytarsus apicalis*-type, dominates the chironomid fauna below sample 6, having an abundance of 15-20%. However, above a depth of sample 6, H. *apicalis*-type declines markedly in abundance until it finally disappears from the fauna at sample 3. *H. apicalis*-type is acidophilic and typically occurs in humic lakes, so the decline in the abundance of this taxon suggests a response to falling acidity in the lake above sample 7, which is particularly marked above sample 3.

Three statistically significant changes in the fauna were identified but they appear to be largely driven by changes in the abundance of lotic, semi-terrestrial and hygropetric taxa so have limited value in interpreting responses in the lake fauna to environmental change. One of the most notable changes is the first appearance of *H. marcidus*-type above sample 5 and its subsequent increase in abundance until it reaches 20% at the top of the sequence. *Dicrotendipes*, another acidophile (Brodin, 1986), also increases notably in abundance above sample 4. However, there are also sharp reductions in *P. septentrionalis*-type between sample 2 and 3 and *H. apicalis*-type at the very top of the sequence and since these are also acidophilic taxa it is difficult to interpret these responses. Other taxa also increase in the upper part of the sequence, such as *Pagastiella orophila*-type, which is considered acidophobic (Pinder & Morley, 1995), but there is a general decline towards the top of the sequence in Tanytarsini which are also generally acidophobic (Henrikson et al., 1982). The general indication is one of the chironomid fauna responding to slightly increasing acidity towards the top of the sequence. However, the response is ambiguous and is obscured somewhat by the overall decline in abundance of lotic and hygropteric taxa towards the top of the sequence which may drive the increase in other taxa due to the 'percentage effect'.

53

3.7   Invertebrate analysis (with special emphasis given to Caddisflies (Trichoptera) and Beetles (Coleoptera)

Trichoptera and Coleoptera are used in biological monitoring programs at the organism (as physiological indicators), at the population (as part of biotic indices), at the community level, (in diversity indices and multivariate approaches,) and in palaeoecology (Resh1992, Coope, 1986).

Caddisflies are associated with almost all types of water body, from ephemeral pools to large lowland rivers (Wallace, 1991) and are widely used by contemporary freshwater ecologists as descriptors of aquatic habitats (e.g. Roux and Castella, 1987), longitudinal spatial patterns in stream networks (e.g. Edington and Hildrew, 1995), flow conditions (e.g. Extence et al., 1999), water quality (e.g. Wright et al. 1996). More recently, fossil caddisfly assemblages have been used as descriptors of past riverine habitats and in the reconstruction of flow environments from floodplain palaeochannels (Greenwood et al. 2003, 2006; Howard et al, 2008).

In palaeoecological studies it is the aquatic larval stage (growth stage) of the life cycle that is important. Larval development, usually of 1 to 2 year duration, goes through a number of instars before pupation and final metamorphosis. The head capsule and thoracic plates are sclerotised and robust and these accumulate in the sediments as the final (death) assemblage. These fragments have a distinctive shape, colour pattern and microsculpture that can be matched against a verified reference collection, allowing specific identification to be made.

At the organism level, experimental work to describe limiting factors for single species of caddisfly and for combinations of different species, is scarce (see Resh, 1992) but a limited number of species-factor relationships, derived from fieldwork have been undertaken, so as to infer optimum tolerance curves (e.g.Verdonschot and Higler, 1992).

At the population level, the tolerances of various species of macroinvertebrates to various forms of pollution have long been used to develop biotic indices. In such indices, Trichoptera, often only used at the family level, are considered pollution sensitive, along with Ephemeroptera (mayflies) and Plecoptera (stoneflies). Aquatic Coleoptera are considered more pollution tolerant. In the UK for example, the BMWP scoring system (a system based upon expert knowledge) is used where 56% of Trichoptera families are given the highest clean water score (10); those families of waterbeetles, being given a score no greater than 5 (Armitage et al, 1983).

At the community level, caddisflies and beetles form important structural and functional components of the benthic community. Diversity indices, in many forms can be used to consider the relationship between the abundance of individuals and the species richness of the community (Magurran, 1988). The many factors, both abiotic and biotic, of which acidity is part, affect the community structure in both transient and complex ways and derived characteristics of diversity are seen as indicators of the well being of ecological systems.

The evidence derived from palaeoecological studies raises an added issue in that the species richness of a specific lake is a function of a number of processes operating at different spatial scales. The interaction of processes at the different scales can be viewed as the result of a series of different filters that operate at the regional to the local scale (Rahel, 2002). Local diversity is then a function of both processes within the lake and the filters operating at higher levels; these stages determining the species found n the final death assemblage.

This report is based upon data derived from deposited death assemblages found in sediments from Loch Coire Fionnaraich, Round Loch of Glenhead and, Loch Narroch. To date no transfer function relating caddisflies or beetle species to acidity has been developed that is comparable with those for chironomids and diatoms. Information regarding specific tolerances to pH for both insect orders is embedded in broader ecological studies of communities: as such a community level approach has been taken in this study.

3.7.1   Methods

Wet lake sediment samples (1-10) from the three sites (Loch Coire Fionnaraich, Round Loch of Glenhead and Loch Narroch) were weighed. The whole sample was then wet sieved at 250 microns and all recognisable fragments removed and examined under a stereomicroscope at 16x magnification. For Trichoptera, fragments of different sizes representing different instars of the same taxon were often observed. For consistency, the numbers of frontoclypeal fragments, the central part of the head capsule, were simply added together, with some additional reference being made to pro and mesonotal thoracic plates to confirm species identifications. The preserved fragments were mounted on microscope slides in Hoyer's medium. Identification was based primarily on the shape, size and colour pattern of the frontoclypeal apotome, as matched against reference material and figures from standard texts. Nomenclature follows that of Barnard (1985).

Fragments of Coleoptera, principally the elytra (wing cases), head and thorax, were collected and stored in 70% IMS. Identification of coleopteran fragments was also undertaken using standard texts and use of the Gorham and Girling reference collections, at University of Birmingham. Nomenclature follows that of Duff (2008).

Other recognisable invertebrate taxa were also recorded. Real counts were recorded down to and including *Sialis lutaria* (see data tables): other taxa below this species were recorded simply as presence/absence. Calculations describing community structure only, used real counts.

Community structure for each site using weighted averaging, was diagrammatised using the programme C2 (Juggins, 2007) and described using Species richness, Indices of diversity (Shannon-Wiener (H'), Evenness (H'/H'max) and a measure of dominance (Berger-Parker index). Species richness, whilst giving valuable insight into diversity, can mask shifts in dominance and evenness (May, 1975). An intuitively simple dominance measure, the Berger-Parker index 'd', is used to express the proportional importance of the most abundant species and in order to ensure that the index increases with increasing diversity, the reciprocal form of the measure has been adopted here. Significant changes in the assemblages were identified by optimal partitioning using the program ZONE (Juggins, 1991) and by comparison with the broken stick model (Ref). A description of habitat preference is also indicated for each taxon.

3.7.2   Results and interpretation

Round Loch of Glenhead

A total of 39 taxa was identified from the whole sequence (10 Trichoptera, 28 Coleoptera (18 of terrestrial origin), 1 Megaloptera (alderfly). In addition, other taxonomic groups were recorded (presence/ absence) which included Ephemeroptera, Corixidae, Hymenoptera (ant mandibles), Simuliidae, ephippia from planktonic crustacea, mites and small bivalves of the genus Pisidium (Table 23).

Applying ZONE software to the full dataset, no significant zones were identified, however there is

ROUND LOCH OF GLENHEAD (RLGH)

			1	2	3	4	5	6	7	8	9	10	Habitat
TRICHOPTERA													
HYDROPTILIDAE					2	1	1			1			
POLYCENTROPODIDAE	*Cyrnus flavidus*	Cyrnflav	42	31	28	17	71	43	52	69	79	24	A/L
	Holocentropus picicornis	Holopici	1										A/L
	Plectrocnema conspersa	Pleccons	2	2	5	1	5		2	1			A/S
	Polycentropus flavomaculatus	Polyflav	3	3	3	2	13	9	7	9	17	3	A/S/L
	Polycentropus irroratus	Polyirro						2	1		2	1	A/S
PHRYGANEIDAE	*Phryganea cf bipunctata*	Phrybipu	2	5	1		1	1	4	1	1		A/L
LIMNEPHILIDAE	*Potamophylax cf latipennis*	Potalati	1			1	1						A/S
SERICOSTOMATIDAE	*Sericostoma personatum*	Seripers						1	1	3	1	2	A/S/L
LEPTOCERIDAE	*Athripsodes aterrimus*	Athrater							1	1			A/L
	Mystacides azurae	Mystazur		2	1			1					A/S
COLEOPTERA													
DYTISCIDAE	*Hydroporus sp*	Hydropor							1		1		A/M
	Nebrioporus depreesus/elegans	Nebreleg	1			1							A/S/L
	Stictotarsus duodecimpustulatus	Sticduod	2	1				1					A/S/L
	Laccophilus sp	Laccophi								1			A
CARABIDAE	*Leistus montanus*	Leismont									1		T
	Nebria sp	Nebriasp									1		T
	Trechus obtusus/quadristriatus	Trecquad							1	1			T
	Pterostichus strenuus	Pterstre				1							T
HELOPHORIDAE	*Helophorus sp.*	Helophor						1					A
HYDROPHILIDAE	*Cymbiodyta marginellus*	Cymbmarg					1						A/L
	Coelostoma orbiculare	Coelorbi						1					M
STAPHYLINIDAE	*Lesteva sp*	Lestevas					1			1		1	T
	Olophrum sp	Olophrum								1	1		T
	Omalium sp	Omaliums								1			T
	Tachinus sp	Tachinus							1				T
	Aleocharinae Gen & sp indet.	Aleoinde		1	1		1	2	1	1			T
	Anotylus rugosus	Anotrugo			1		1			1			T
	Platystethus arenarius	Plataren								1			T
	Lathrobium sp	Lathrobi			1				1				T
SCARABAEIDAE	*Aphodius sp*	Aphodius	1										T
	Phyllopertha horticola	Phylhort							1				T
SCIRTIDAE	*Cyphon sp*	Cyphonsp	1								1		M
DASCILLIDAE	*Dascillus cervinus*	Dasccerv							1	1			T
	Oulimnius cf tuberculatus	Oulitube			1	1	1		2	1			A/S/L
ELATERIDAE		Elaterid	1										T
NITIDULIDAE	*Meligethes sp*	Meligeth						1				1	T
CHRYSOMELIDAE	*Donacia/Plateumaris*	Donaplat				1							M
CURCULIONIDAE	*Rhamphus sp*	Rhamphus									1		T
Megaloptera SIALIDAE	*Sialis lutaria*	Silaluta	3	1	1	1	2		2	2	1		A
Ephemeroptera		Ephemero	1	1	1	1	1	1	1	1		1	
Hemiptera cf CORIXIDAE		Corixasp				4		4					
Hymenoptera: Ant spp		Hymenant						1	1				
Diptera SIMULIIDAE		Diptsimu	1		1			1		1	1	2	
Crustacea DAPHNIIDAE	*Ephippia: 2 egg*	Ephiptwo				1		1	1	1		1	
Crustacea FAMILY indet.	*Ephippia: 1 egg*	Ephipone	1			1	1	1		1		1	
Arachnida		Arachnid				1		1					
Acari	*Orobatid mites*	Orobmite		1				1	1		1	1	
	Water mites	Watemite		1			1						
Mollusca SPHAERIIDAE	*Pisidium sp*	Pisidium		1									
Wings: various		Wingvari		1		1	1	1				1	
Caddisfly larval cases: spp in det.		Caddcase						1					

Table 23: Round Loch of Glenhead: Raw data (Real counts, presence/absence). Also the preferred habitat characteristics are indicated; A (aquatic), S (stream), L (lake), M (marginal), T (terrestrial).

indication of change taking place between 4 and 5, also to be seen in Figure 21.

Descriptors of community structure (Table 24) show the range in number of taxa (6-17), abundance (26-107), diversity (0.98-1.66), dominance (0.58-0.75) and reciprocal dominance (1.33-1.72). The total number of individuals was highest (655 individuals) for the three Lochs.

Overall, species diversity values (H') are low. From the core profile, descriptors of community structure show a degree of similarity between samples but there does appear to be improvement in both diversity
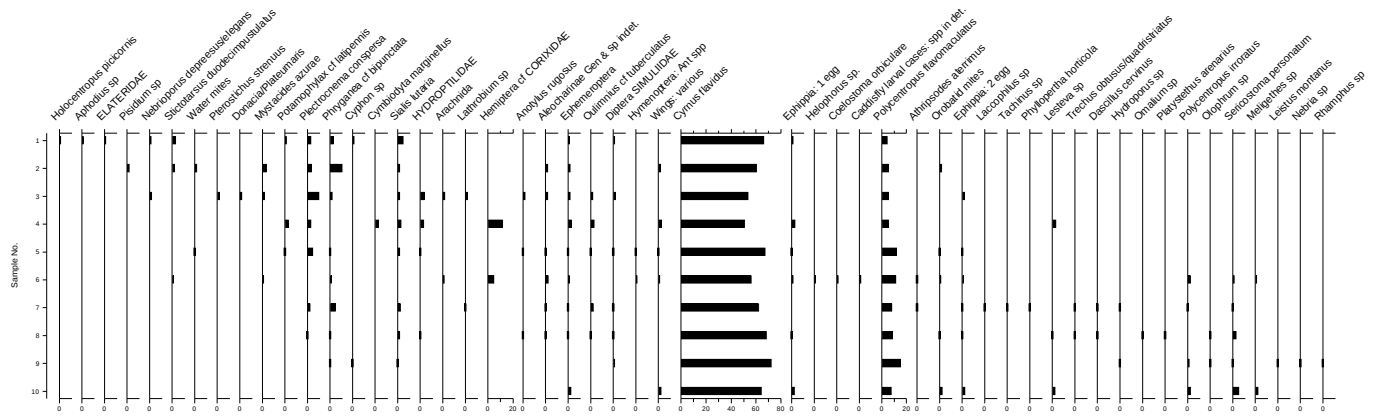
Figure 21: Round Loch of Glenhead: C2 plot of the community structure based upon weighted averaging.

Sample No.	N of Taxa	Abundance	Shannon Wiener H	H/Hmax	Berger-Parker	1 over d	Sample wt(g)
1	12	60	1.3	0.52	0.7	1.43	91.7
2	8	46	1.21	0.58	0.67	1.48	82.8
3	14	48	1.66	0.63	0.58	1.72	86
4	9	26	1.35	0.62	0.65	1.53	94.8
5	10	97	1.01	0.44	0.73	1.37	94.3
6	12	64	1.28	0.51	0.67	1.49	83.7
7	17	80	1.52	0.54	0.65	1.54	101.9
8	16	95	1.22	0.44	0.73	1.38	92.9
9	12	107	0.98	0.4	0.74	1.36	104.6
10	6	32	0.94	0.52	0.75	1.33	101.1

Table 24: Round Loch of Glenhead: Community descriptors, Numbers of Taxa, Abundance, Species Diversity (Shannon-Wiener H'), Evenness (H'/Hmax), Measure of Dominance (Berger-Parker), Reciprocal dominance (1/'d') and Sample wet weight (g).
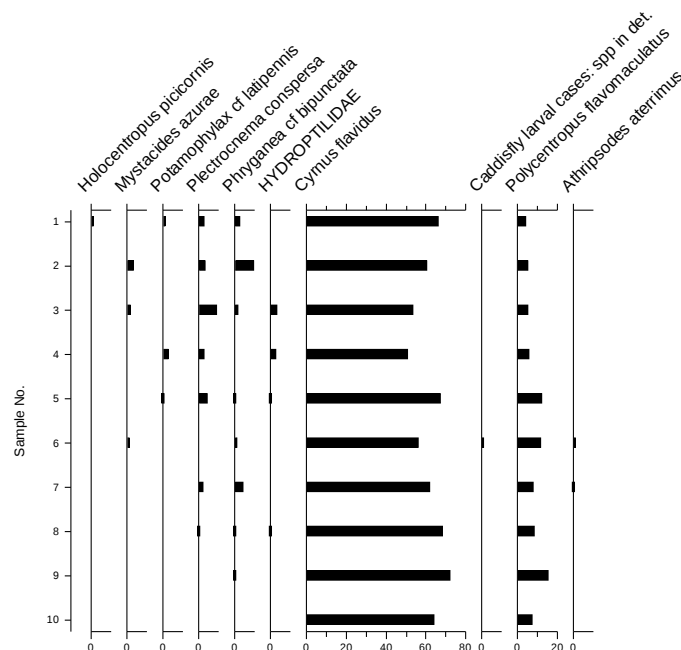

Figure 22: Round Loch of Glenhead: C2 plot of the caddisfly community based upon weighted averaging

57

Loch Coire Fionnaraich

			Samples 1	2	3	4	5	6	7	8	9	10	Habitat
TRICHOPTERA													
PHILOPOTAMIDAE	Philopothamus montanus	Philmont		1	3			3	2	1	1	5	A/S
POLYCENTROPODIDAE	Cyrnus flavidus	Cyrnflav	7	11	12	2		3	10	4	8	11	A/L
	Plectrocnema conspersa	Pleccons	4	3	2	1	5	3	1	3	2	4	A/S/L
	Polycentropus flavomaculatus	Polyflav	3	7	1				3	1	5	2	A/S/L
HYDROPSYCHIDAE		Hydropsy		1	1				1				A
	Hydropsyche cf instabilis	Hydrinst					1						A/S
LIMNEPHILIDAE	Ecclisopteryx sp.	Ecclisop									1		A/S
	Halesus cf digitatus	Haledigi					1	1	1				A/S
	Potamophylax cf latipennis	Potalati		1			1						A/S/L
	Limnephilus cf lunatus	Limnluna		1							1		A/S/L
	Limnephilidae indet.	Limninde		2									A
GOERIDAE	Silo pallipes	Silopall						1			1		A/S
SERICOSTOMATIDAE	Sericostoma personatum	Seripers	1					1		1	1		A/S/L
LEPTOCERIDAE	Mystacides azurae	Mystazur						1	1		1		A/S/L
COLEOPTERA													
DYTISCIDAE	Hydroporus sp	Hydropor			1	1							A/M
CARABIDAE	Trechus obtusus/quadristriatus	Trecquad				1	1				1		T
	Trechus sp.	Trechuss							1				T
	Pterostichus sp.	Pterosti	1										T
	Harpalus sp.	Harpalus				1							T
	Bradycellus ruficollis	Bradrufi		1									T
	Stenolophus sp.	Stenolop		1									T
HYDROPHILIDAE	Cercyon sp.	Cercyons									1		M
SCYDMAENIDAE		Scydmaen								1			T
STAPHYLINIDAE	Lesteva elongoelytrata	Lestelon		1			1						T
	Lesteva sp	Lestevas				1		1					T
	Olophrum assimile	Olopassi						1					T
	Olophrum fuscum	Olopfusc						1					T
	Olophrum sp	Olophrum		1		1							T
	Eusphalerum sp.	Eusphale						1			1		T
	Lordithon exoletus	Lordexol		1									T
	Tachyporus sp.	Tachypor							1		1		T
	Aleocharinae Gen & sp indet.	Aleoinde		1	1		1			1	1	1	T
	Stenus sp.	Stenussp							1				T
	Lathrobium sp	Lathrobi							1				T
	Philonthus sp.	Philonth						1					T
GEOTRUPIDAE	Geotrupes sp.	Geotrupe					1						T
SCARABAEIDAE	Aphodius sp	Aphodius						1					T
SCIRTIDAE	Cyphon sp	Cyphonsp						1		1			M
ELMIDAE	Elmis aenea	Elmiaene	1	1		1	1				1		A/S
	Oulimnius cf tuberculatus	Oulitube		1		1					1		A/S/L
CANTHARIDAE	Malthodes sp.	Malthode							1	1			T
CHRYSOMELIDAE	Donacia/Plateumaris	Donaplat				1							M
	Micrelus ericae	Micreric							1				T
Megaloptera SIALIDAE	Sialis lutaria	Silaluta				1	1		1				
Ephemeroptera		Ephemero	1				1	1	1	1	1		
Hemiptera cf CORIXIDAE		Corixasp	1										
Hymenoptera: Ant spp		Hymenant				1					1		
Diptera SIMULIIDAE		Diptsimu	1	6	1			1	3	4	1	1	
Diptera larvae		Diptlarv				1							
Crustacea DAPHNIIDAE	Ephippia: 2 egg	Ephiptwo	1	1	1	1	1	1		1	1	1	
Crustacea FAMILY indet.	Ephippia: 1 egg	Ephipone			1			1					
Acari	Orobatid mites	Orobmite	1		1			1		1		1	
Mollusca SPHAERIIDAE	Pisidium sp	Pisidium	3		1								
Wings: various		Wingvari	1	1	1		1		1	1		1	
Caddisfly larval cases: spp in det.		Caddcase						3		7			

Table 25: Fionnaraich: Raw data (Real counts, presence/absence). Also the preferred habitat characteristics are indicated; A (aquatic), S (stream), L (lake), M (marginal), T (terrestrial).

and evenness in samples 1-4, a feature also reflected in the reciprocal dominance measure (1/d). Eighteen of the beetle taxa are of terrestrial origin giving a strong indication of stream inflow.

The caddisfly community (Figure 22) is dominated by *Cyrnus flavidus*, a the net spinning species restricted to still water habitats in the UK, along with other members of the family Polycentropodidae (*Holocentropus picicornis, Plectrocnema conspersa, Polycentropus flavomaculatus* and *P. irroratus*). All members of this family are predatory and operate their nets as snares, which capture live prey. Other families present are, Hydroptilidae, Leptoceridae (*Mystacides azurae, Athripsodes aterrimus*), Limnephilidae (*Potamophylax latipennis*) and Phryganeidae (*Phryganea bipunctata*).

Figure 23: Fionnaraich: C2 plot of the community structure based upon weighted averaging.

Sample No.	N of taxa	Abundance	Shannon Wiener H	H/Hmax	Berger-Parker	1/d	Sample wt(g)
1	6	17	1.51	0.84	0.41	2.43	61.3
2	16	35	2.28	0.82	0.31	3.19	74.6
3	9	23	1.64	0.74	0.52	1.92	83.8
4	11	12	2.37	0.99	0.17	5.99	94.1
5	12	16	2.27	0.91	0.31	3.2	98.6
6	11	17	2.25	0.94	0.18	5.65	96.6
7	13	25	2.11	0.82	0.4	2.5	90.7
8	9	14	2.01	0.91	0.29	3.5	96.7
9	13	25	2.18	0.85	0.32	3.13	92.1
10	8	26	1.67	0.8	0.42	2.36	90.7

Table 26: Fionnaraich: Community descriptors, Numbers of Taxa, Abundance, Species Diversity (Shannon-Wiener H'), Evenness (H'/Hmax), Measure of Dominance (Berger-Parker), Reciprocal dominance (1/'d') and Sample wet weight (g). Plots are also given for Species diversity, Evenness, Dominance and Reciprocal dominance

Figure 24: Fionnaraich: C2 plot of the caddisfly community based upon weighted averaging.

59

Of the Coleoptera only 7 taxa can be considered aquatic.e.g. *Cymbiota marginalis, Stictotarsus duodecimpustulatus* but all are in low numbers and little can be derived as to their affinity to changing levels of pH.

Loch Coire Fionnaraich

A total of 44 taxa was identified from the whole sequence (14 Trichoptera, 29 Coleoptera (23 of terrestrial origin), 1 Megaloptera). In addition, other taxonomic groups were recorded (presence/ absence) which included Ephemeroptera, Corixidae, Hymenoptera (ant mandibles), Simuliidae, other Dipterous larvae, ephippia from planktonic crustacea, mites and small bivalves of the genus Pisidium (Table 25). Applying ZONE software to the full dataset, no significant zones were identified however there is indication of change taking place between 3 and 4, also to been seen in Figure 23.

Descriptors of community structure (Table 26) show the range in the number of taxa (6-16), abundance (12-35), diversity (1.51-2.37), dominance (0.167-0.52) and reciprocal dominance (1.92-5.99).

No clear pattern of change in the community structure is apparent within the core but there is a possible increase in species diversity and evenness below 3 cm. This is also reflected in a high dominance score at this level. Twenty three of the beetle taxa are of terrestrial origin indicating a considerable influence of stream inflow.

The caddisfly community (Figure 24) comprises those taxa inhabiting the loch and its littoral edge, together with taxa most commonly found in streams. The most common taxon and one only indicative of the loch itself, is *Cynus flavidus*, a net spinning caseless caddisfly. Within the profile *Cyrnus flavidus* is absent in sample 5. Other species of the same Family Polycentropodidae are *Polycentropus flavomaculatus* and *Plectronemia conspersa*, these latter species also being found in adjacent streams. Below the level of sample 3 other taxa appear in the samples but these are predominantly in small number and represent taxa only found in stream habitats e.g. *Philopotamus montanus, Halesus digitatus*.

Of the coleopteran taxa, only 6 can be considered aquatic or semiaquatic. All are found in small number but *Elmis aenea* and *Oulimnius cf tuberculatus* are found in more than one sample, the former being an inhabitant of streams, the latter of both running water and lakes. No conclusions as to affinity to changes in acidity can be made from this group.

Loch Narroch

Sample weights are low for this site (23.8 -50.8 g) but a total of 39 taxa was identified from the whole sequence (9 Trichoptera, 29 Coleoptera (20 of terrestrial origin), 1 Megaloptera), along with the presence of other invertebrate groups (Table 27). Applying ZONE software to the full dataset, no significant zones were identified however there is indication of change taking place at 6.5, also to been seen in Figure 25.

Descriptors of community structure (Table 28) show the range in the number of taxa (4-18), abundance (10-27), diversity (1.09-2.58), dominance (0.18-0.6) and reciprocal dominance (1.67-5.49).

Low numbers of taxa occur in the uppermost 4 samples. Below this level species diversity appears to increase towards the base of the profile from a lowest value of 1.089 at sample 4. High dominance values are also found in this upper zone.

Although there is evidence of the substantial addition of a terrestrial fauna, there are invertebrate descriptors of the aquatic habitat that differ from e.g. Fionnaraich. The alderfly *Sialis lutaria* is present in

Loch Narroch

			Samples 1	2	3	4	5	6	7	8	9	10	Habitat
TRICHOPTERA													
POLYCENTROPODIDAE	Cyrnus flavidus	Cyrnflav	6	7	4	6	9	9	3	2	9	8	A/L
	Plectrocnema conspersa	Pleccons					1	1				2	A/S
	Polycentropus flavomaculatus	Polyflav								1	1	1	A/S/L
HYDROPSYCHIDAE	Hydropsyche cf instabilis	Hydrinst											A/S
PHRYGANEIDAE	Phryganea cf bipunctata	Phrybipu		1	1	1	2	1	1				A/L
LIMNEPHILIDAE	Halesus cf digitatus	Haledigi	1										A/S
	Limnephilus rhombicus	Limnrhom	1										A/L
SERICOSTOMATIDAE	Sericostoma personatum	Seripers		1							1	1	A/S
LEPTOCERIDAE	Athripsodes aterrimus	Athrater	2								1	1	A/L
COLEOPTERA													
DYTISCIDAE	Agabus sp.	Agabussp			1								A
	Rhantus sp.	Rhantuss					1						A
	Hydroporus sp	Hydropor				1					1	1	A/M
	Hydroporus melanarius	Hydrmela						2	1	1			M
	Nebrioporus depreesus/elegans	Nebreleg			1								A/S/L
CARABIDAE	Dyschirius globosus	Dyscglob							1			1	T
	Trechus obtusus/quadristriatus	Trecquad		1				1					T
	Pterostichus minor	Ptermino							1				T
	Pterostichus diligens	Pterdili						1				1	T
	Bradycellus ruficollis	Bradrufi			1					1			T
HYDROPHILIDAE	Coelostoma orbiculare	Coelorbi										1	M
STAPHYLINIDAE	Acidota crenata	Acidcren	1										T
	Lesteva sicula	Lestsicu									1		T
	Lesteva sp	Lestevas					3		1		1	1	T
	Olophrum fuscum	Olopfusc										1	T
	Olophrum sp	Olophrum					2	1	1	1	1		T
	Eusphalerum sp.	Eusphale						1					T
	Omalium sp	Omaliums									1		T
	Lordithon exoletus	Lordexol					1						T
	Tachyporus sp.	Tachypor		1				1	1	1		1	T
	Aleocharinae Gen & sp indet.	Aleoinde		1			1				3	1	T
	Stenus sp.	Stenussp		1					1				T
	Lathrobium sp	Lathrobi								1			T
	Philonthus sp.	Philonth										1	T
	Quedius sp.	Quediuss	1		1					1			T
SCIRTIDAE	Cyphon sp	Cyphonsp										1	M
ELMIDAE	Oulimnius cf tuberculatus	Oulitube			1								A/S/L
CHRYSOMELIDAE	Donacia/Plateumaris	Donaplat					1				1	1	M
CURCULIONIDAE	Micrelus ericae	Micreric						1				1	T
Megaloptera SIALIDAE	Sialis lutaria	Silaluta	3	1	2	2	3	3	2	2	2	2	A
Ephemeroptera		Ephemero	1	1	1	1	1	1	1		1	1	
Hemiptera cf CORIXIDAE		Corixasp					1						
Diptera SIMULIIDAE		Diptsimu			1	1		1			1	1	
Crustacea DAPHNIIDAE	Ephippia: 2 egg	Ephiptwo		1			1	1	1	1	1	1	
Crustacea FAMILY indet.	Ephippia: 1 egg	Ephipone		1	1	1	1	1	1				
Acari	Orobatid mites	Orobmite	1	1	1		1	1	1	1	1	1	
	Water mites	Watemite							1				
Wings: various		Wingvari									1	1	
Caddisfly larval cases: spp in det.		Caddcase	1										

Table 27: Loch Narroch: Raw data (Real counts, presence/absence). Also the preferred habitat characteristics are indicated; A (aquatic), S (stream), L (lake), M (marginal), T (terrestrial).

all samples from this site. The larvae live in ponds, lakes and sluggish parts of streams and rivers where there is an abundance of silt. They are tolerant to changes in pH are often numerous in the benthos of acid lakes (Elliott, 1996).

The caddisfly community (Figure 26), whilst still dominated by the Polycentropodid *Cyrnus flavidus* also has numbers of *Phryganea bipunctata* (Phryganeidae). The larvae of this family live in tubular cases made of pieces of plant stem or leaves and are found in a wide variety of still and slow-flowing waters (Wallace et al, 1990). This species is also considered to be acid tolerant (Fjellheim pers com.). *Athripsodes aterrimus* (Leptoceridae), a case building species is also commonly found in lakes,
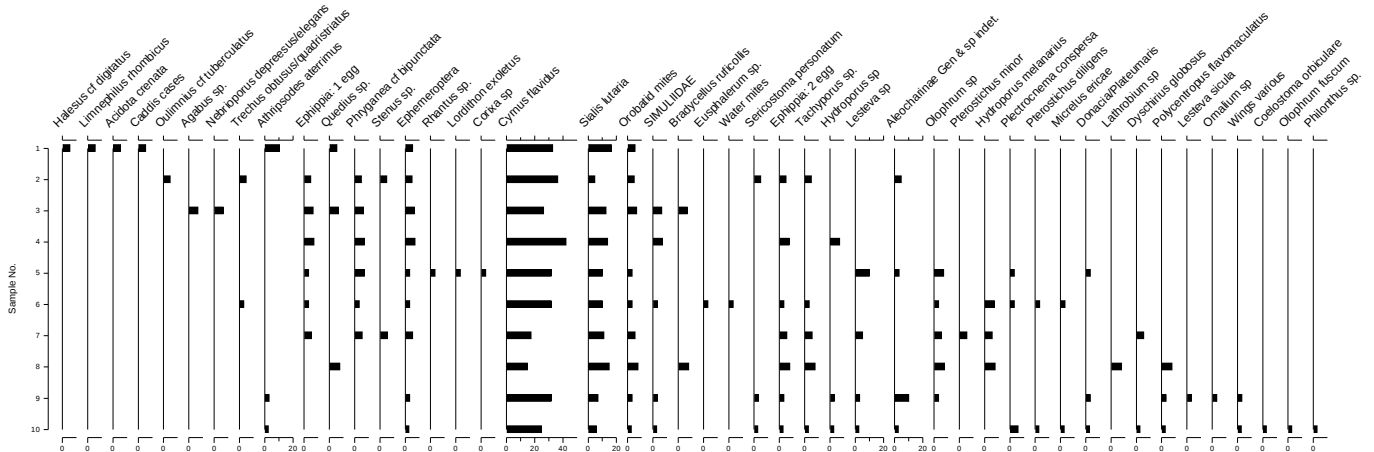
Figure 25: C2 plot of the community structure based upon weighted averaging.

Sample No.	N of taxa	Abundance	Shannon Wiener H	H/Hmax	Berger-Parker	1/d	Sample wt(g)
1	7	15	1.6792	0.8629	0.4000	2.50	50.8
2	9	15	1.8000	0.8192	0.4667	2.14	35.7
3	7	11	1.7678	0.9084	0.3636	2.75	23.8
4	4	10	1.0889	0.7855	0.6000	1.67	35.3
5	10	24	1.9639	0.8529	0.3750	2.66	39.6
6	11	22	1.9794	0.8255	0.4091	2.44	45.0
7	10	13	2.2048	0.9575	0.2308	4.31	44.6
8	9	11	2.1454	0.9766	0.1818	5.49	44.9
9	12	23	2.2072	0.8338	0.3913	2.55	46.8
10	18	27	2.5770	0.8916	0.2963	3.38	40.0

Table 28: Loch Narroch community descriptors, Numbers of Taxa, Abundance, Species Diversity (Shannon-Wiener H'), Evenness (H'/Hmax), Measure of Dominance (Berger-Parker), Reciprocal dominance (1/'d') and Sample wet weight (g). Plots are also given for Species diversity, Evenness, Dominance and Reciprocal dominance
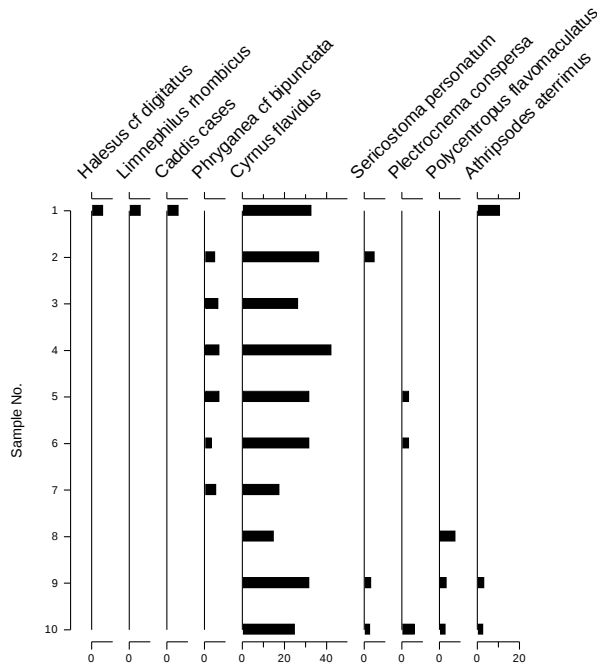
Figure 26: C2 plot of the caddisfly based upon weighted averaging

particularly in the flowing water of the outlet.

Of the coleopteran fauna only 9 can be considered aquatic or semiaquatic. The only true water beetles (Family Dytiscidae) are *Agabus sp.*, *Rhantus sp.* and *Nebrioporus depressus elegans* . All are found in small numbers and no conclusions as to relationships to pH can be made.

No transfer functions exist for caddisflies or beetles as to their association with changes in acidity of freshwaters, comparable with those determined for chironomids or diatoms. As such the changes in community structure throughout the profiles described above are going to reflect the interaction of many variables, both abiotic and biotic.

From ecological studies and the monitoring of water quality of lakes, especially in Scandinavia, some information regarding sensitivity to pH is available.

Most of the species in this study are considered acid tolerant: *Cyrnus flavidus* (pH 4.1), *Polycentropus flavomaculatus* (4.2), *Plectrocnema conspersa* (3.8), *Holocentropus picicornis* (4.2), *Mystacides azurae* (5.1), *Sericostoma personatum* (4.5), *Hydropsyche instabilis* (4.5), *Silo pallipes* (4.5) (J. Brittain per comm.). There are no specific tolerance values for *Phryganea bipunctata*, *Limnephilus rhombicus* and *Halesus digitatus* but these species are also considered to be acid tolerant (A. Fjellheim pers comm.). It is worth noting that for these invertebrates, pH tolerances will vary with water quality e.g. in lakes of varying Calcium concentrations.

Benthic invertebrates are a diverse and generally abundant group with a wide range of environmental tolerances which can act as indicators of environmental quality (Rosenberg and Resh, 1993). The presence/absence of certain species has therefore been used to assess the effects of acid stress on stream ecosystems but with less emphasis being placed on acid stress effects in lakes. An assessment of the effects of using existing metrics, developed for assessment of river acidification, in the context of monitoring lake acidification, is given in Schartau et al., 2008.

3.8   Diatom Analysis

3.8.1   Methods

Sediment sub-samples from the three lochs were treated in the laboratory with H_2O_2 and mounted in Naphrax following the method of Battarbee *et al.* (2001). Between 600-700 diatom valves were counted on each slide. The diatom data presented here is relative abundance.

pH was reconstructed using the Surface Water Acidification Project (SWAP) diatom calibration data-set (Stevenson et al., 1991) using Weighted Averaging Partial Least Squares (WA-PLS) method in C2. The overall root mean square error of prediction (RMSEP) of the transfer-function was estimated using leave-one-out cross-validation.

The best model was chosen as a combination of a high coefficient of determination (R^2) between observed and predicted values, a low mean and a maximal bias, and a low root mean squared error of prediction (Birks, 1998) and the best model was used to reconstruct the historical pH from fossil diatoms preserved in lake sediment cores.
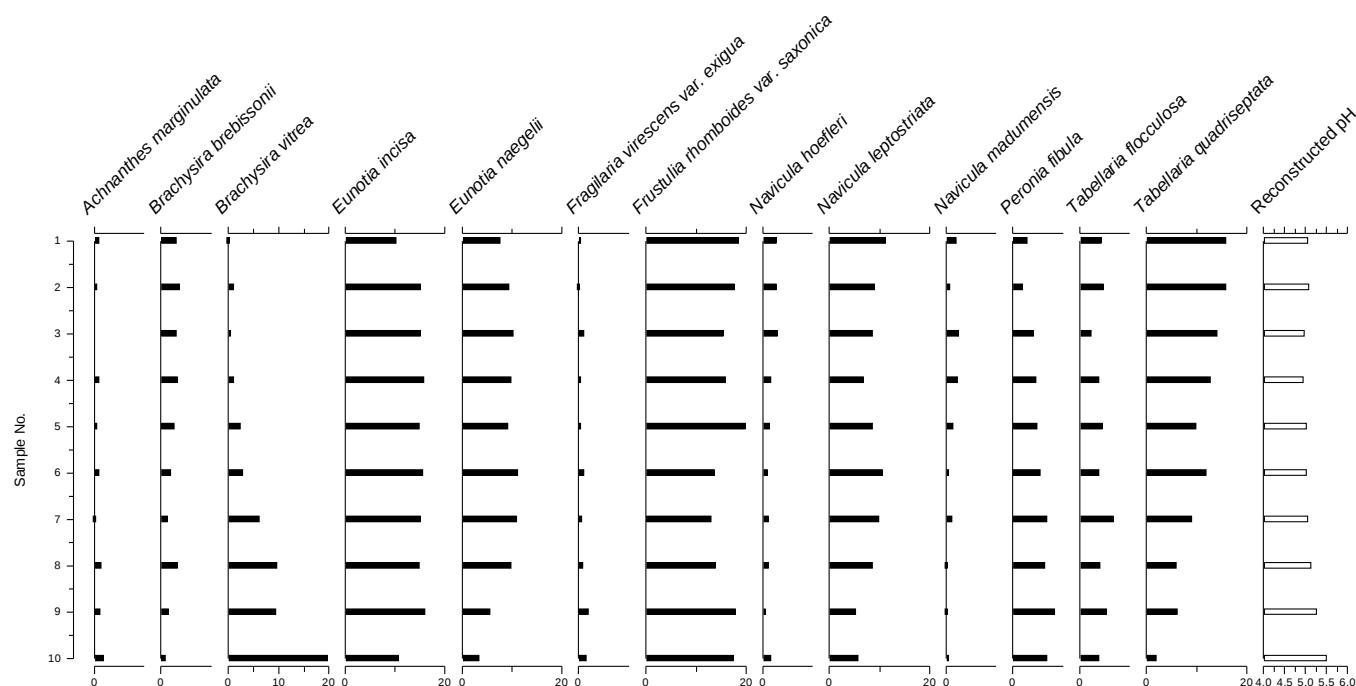
Figure 27: Stratigraphic plot of diatom counts from RLGH.

3.8.2   Results and interpretation

Round Loch of Glenhead

Figure 27 shows the dominant diatoms found in the sediment samples from RLGH. Acidophilous taxa (*Eunotia incisa*, E. naegelii, *Frustulia rhomboides* var. *saxonica*, *Tabellaria quadriseptata* and *Navicula leptostriata*) reflect the low pH of the lake.

The sequence of samples record a visible shift between *Brachysira vitrea/Fragilaria rhombiodes* var. *saxonica* to more acidophilc *Tabellaria quadriseptata/Navicula madumensis/N.hoefleri*. Previous diatom data from cores show that this occurred from the mid- to end 19th Century.

The SWAP diatom pH reconstruction for RLGH shows this trend, decreasing from 5.5 in sample 10 to 4.8 in sample 4. There has been some amelioration in pH at RLGH since the 1990's and this is seen in the last few samples that have a reconstructed pH of 5.

Loch Narroch (NARR)

Figure 28shows the dominant diatoms found in the sediment samples from Loch Narroch. Similar to RLGH, there is clear evidence of increased acidity, shown by the trends in *Navicula hoefleri, Tabellaria binalis* and *Eunotia bactriana.* Previous core work at Loch Narroch has shown that *Eunotia bactriana* is first present in sediments dating from the 1930's and *Tabellaria binalis* occurs around 1960 with increased acidification of the lake. The SWAP diatom pH reconstruction for the NARR samples shows only a slight acidity increase. Considering that these diatom samples are from subsamples taken from a number of cores and that both the radiometric dating and observations during coring indicate significant post-depositional disturbance, the diatom data here is surprisingly comparable with previous core work.
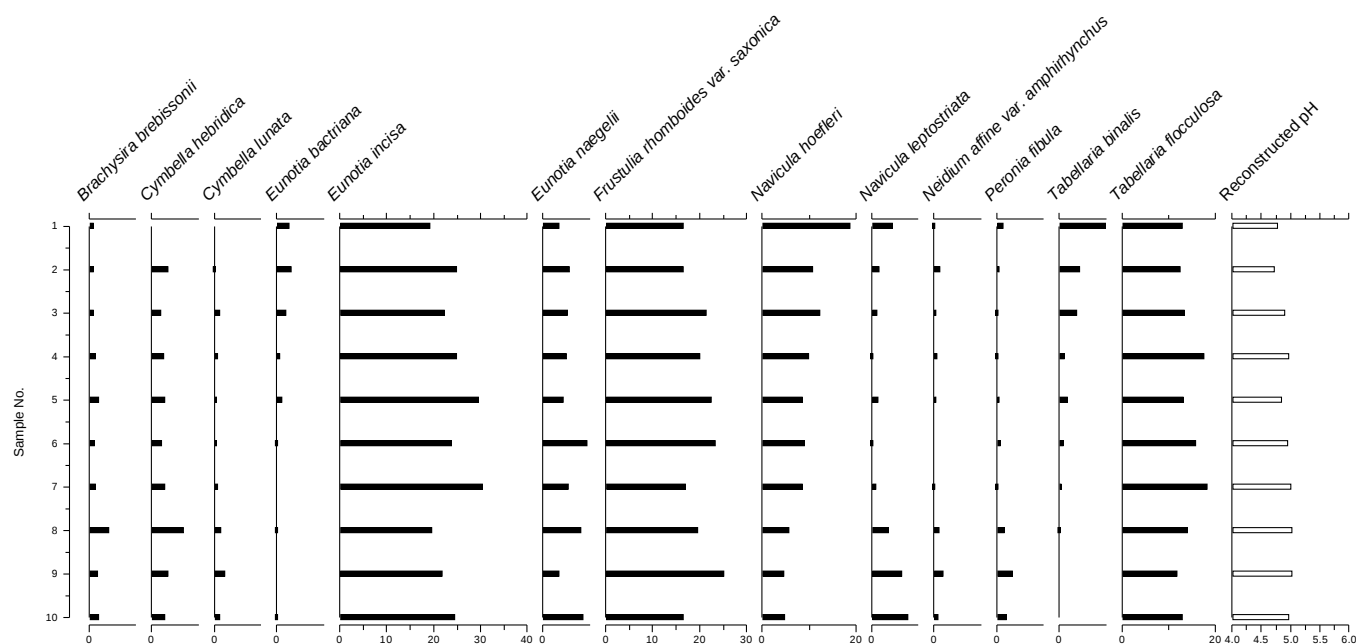
Figure 28: Stratigraphic plot of diatom counts from Loch Narroch
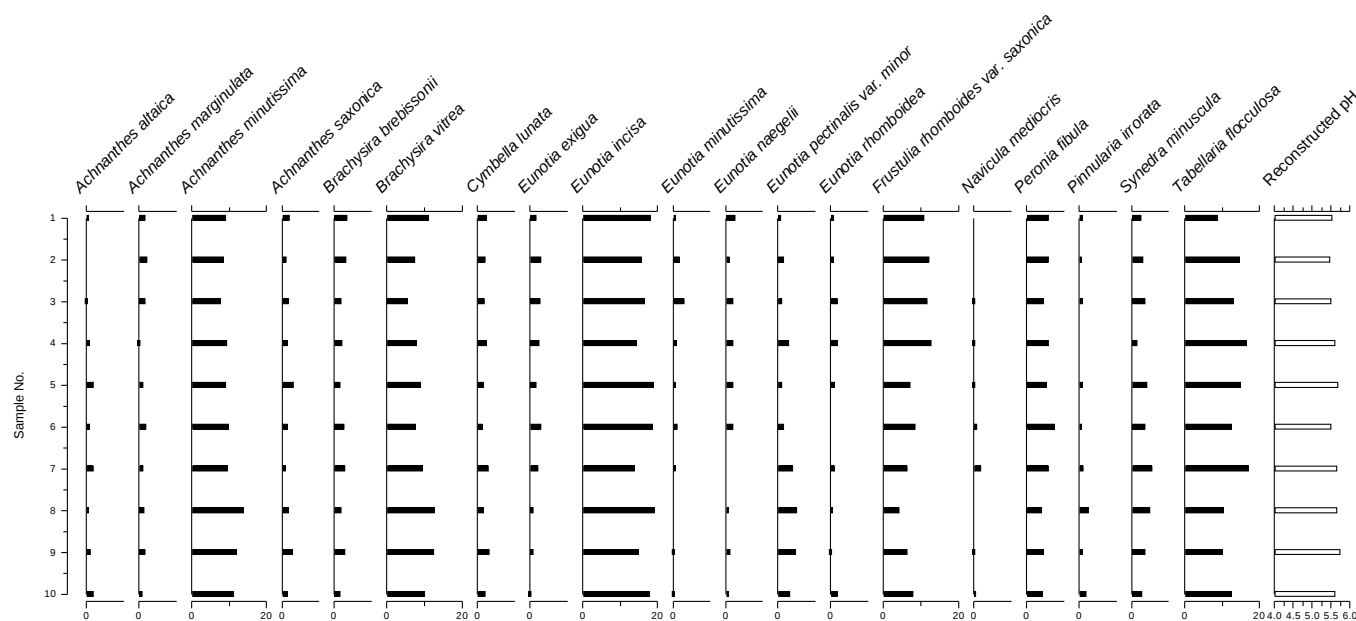
Figure 29: Stratigraphic plot of diatom counts from Loch Coire Fionnaraich.

Loch Coire Fionnaraich (LCFR)

Figure 29 shows the dominant diatoms found in the sediment samples from Loch Coire Fionnaraich. The diatom assemblage is dominated by *Brachysira vitrea* (pH optima 5.9), *Eunotia incisa* (pH optima 5.1), *Peronia fibula* (pH optima 5.3) and *Tabellaria flocculosa* (pH optima 5.4) and the less acid species *Achnanthes minutissima* (pH optima 6.3).

There is no indication from contemporary water chemistry monitoring and analysis of cores taken in 2001 (Pla *et al.* 2009) to show that the loch has been subject to significant acidification.

3.9 Comparison with AWIST Status

The diatom data from Loch Narroch and Round Loch of Glenhead suggest that significant acidification of these surface waters has taken place over the last 200 years. In comparison, Loch Coire Fionnaraich is minimally impacted with now sign of acidification, though Pla et al. (2009) demonstrate changes in siliceous algal communities that, whilst not acidity related, do appear to be due to factors related to atmospheric deposition, such as nitrogen enrichment. Here, we address the acidification impacts in relation to AWIST predictions.

The AWIST predicted class and EQR for the three sites at which palaeoecological data were compiled are shown in Table 29.

WBID	Name	AWIST Class	AWIST EQR
27192	Loch Narroch	Moderate	0.530
29927	Round Loch of Glenhead	Moderate	0.538
17334	Loch Coire Fionnaraich	Good	0.696

Table 29: AWIST predicted class and EQR for the three sites in the palaeoecological study

Whilst it is difficult to relate specific changes in the palaeolimnological record of chironomid, Coleoptera and caddisfly remains, the results described in the previous sections above suggest that the macroinvertebrates community of Loch Narroch and the Round Loch of Glenhead has responded to acidification, demonstrated by increases in acidophilous taxa or loss of species diversity. In contrast, Loch Coire Fionnaraich has not acidified and the macroinvertebrate shows little sign of change.

The observations from the palaeoecological record for the three sites are largely borne out in the AWIST predictions. The acidification at Loch Narroch and the Round Loch of Glenhead is reflected in the low EQR and assignment of Moderate status from AWIST (Table 29). The present-day (core-top) assemblage represents a period following chemical recovery at these two sites as sulphur emissions have declined. As such the AWIST prediction relates to present-day conditions not the height of acidification experienced by these two lochs.

The assigned class for Loch Coire Fionnaraich is Good status. This is perhaps an under-estimate of the invertebrate assemblage status at this site considering the palaeoecological data. However, it must be remembered that the sediment record is an incomplete record of the macroinvertebrate assemblage whereas AWIST uses the entire assemblage. Furthermore, AWIST is based upon spatial relationships between measured hydrochemistry and species assemblages, and logically will include differences due to variation in site type and size, and habitat availability as well as other disturbances such as enrichment effects of N deposition. Whilst the WFD60 training set was chosen to minimise the effects of these other factors they could not be eliminated completely, especially for factors related to the size of site and habitat availability as this would have necessitated loosing too many sites from the training set.

In conclusion, given the uncertainties inherent in both the AWIST methodology and the palaeoecological approach the results of the palaeoecological analysis compare well with the contemporary status of the three studied sites as determined using AWIST.

4 Recommendations for future work

This report reflects substantial amounts of new work and tool development to produce a robust, working macroinvertebrate lake classification tool for acidification status. New, high quality macroinvertebrate and hydrochemical data were collected from a range of acid sensitive sites across the UK. The additional data collection exercise provided a data set of a size at the lower end of the size range expected for applying the random forest methodology.

We recommend that the agencies collect additional samples that could be used in the future to update AWIST to make the tool more robust and to reduce the problems the small sample size and biased sampling of High and Good status sites cause for model fitting and evaluation. Where planned or new agency sampling is to be undertaken at acid sensitive sites, we recommend that the agencies collect data that could be usefully employed in a future revision of AWIST. To this end, we recommend that WFD60 protocols (described in Monteith and Simpson 2007) be adopted for macroinvertebrate sampling and that high quality water chemistry data be collected that will allow calculation of Cantrell ANC and also provide measurement of labile aluminium concentrations.

Without further, substantial additions of new samples there is little further that can be done to improve the current tool, with the exception of some minor improvements to the fitting of the random forest.

There is considerable potential from pooling output from the three acidification classification tools for lakes into an ensemble predictor, which draws upon the good features of the individual tools. We recommend that future efforts be directed towards investigating combining the three tools and evaluating the performance of such a produced ensemble classification tool. Such work may also provide a vehicle for implementing the minor improvements in the random forest algorithm that might provide a more realistic cross-validated performance measure by implementing stratified bootstrap resampling.

5 Conclusions

In this report we have presented continued development of a novel classification scheme for macroinvertebrate-based WFD classification of damage due to lake acidification. Hydrochemical parameters (ANC and Ca^{2+} concentration) are used as a simple summary of lake status and damage from acidification – this classification is known as the "Damage Matrix". Random forests, a modern, powerful statistical data mining tool, is then used to find decision rules from the macroinvertebrate species data that best predict the Damage Matrix class for each site in the WFD60 training set. Random forests are an ensemble method and incorporate many classification trees fitted to bootstrap samples of the training data, and predictions from the forest are based on majority votes from the ensembles of trees. Due to a paucity of samples in the Poor and Bad classes, it proved impossible to separate these two groups and therefore these classes were merged into a Poor-Bad class.

The new tool correctly predicts the Damage Matrix class for training set samples 99% of the time. AWIST performs considerably better than CPET (34.62% of damage matrix class predicted correctly) and LAMM (54.29%). Whilst the true performance for new samples will be lower than this value, a potential lower limit is given by the out-of-bag predictions for the training set samples which was correctly identified the status class 50% of the time. The out-of-bag predictions are probably an over-estimate of the tool error owing to problems of drawing bootstrap samples from a data set with biased sampling of the status classes, such as is the case with the WFD60 training set. We expect the true tool performance to lie somewhere between the apparent (99%) and the out-of-bag (50%). If the true error lies closer to the out-of-bag error, AWIST still compares favourably to CPET and performs similarly to LAMM, and we expect true performance to be considerably better than the out-of-bag performance.

The original WFD60 macroinvertebrate classification tool did not produce EQRs and provided only a coarse grained measure of prediction uncertainty or confidence of class. Drawing on the ensemble nature of the random forest, the new tool computes an EQR for each site based on a weighted average of base scores for each of the four WFD status classes predicted by the tool, with weights taken as the proportion of votes for each class returned for a sample by the ensemble of trees. We demonstrate that this EQR reflects ecosystem damage due to acidification by relating the EQR to labile aluminium concentrations in a subset of the WFD60 training set where this hydrochemical parameter was reported. Thus, the EQR reported by AWIST conforms to WFD normative definitions of damage. The EQR produced also incorporates uncertainty in predicted class as it is based on the proportions of votes returned for each class.

Confidence of class measures are now produced by the new tool, based upon the proportion of votes returned for each class for individual samples. The votes can be used as a probability that an individual sample belong to each of the WFD status classes predicted by the new tool. As four classes can be predicted by AWIST, we summarise these four probabilities using Shannon's entropy measure to give a single measure of confidence of class.

To assess temporal variability of outputs from the new tool, we applied AWSIT to time series of macroinvertebrate data from the UK AWMN lake sites for the period 1988 to 2008. This analysis showed considerable inter-annual variability in site EQR, resulting from substantial inter-annual variability in the macroinvertebrate species assemblage. As such, we recommend that several samples are collected over time and used to provide a consensus EQR based on the individual EQRs returned by AWIST for the individual annual samples. Despite the inter-annual variability in EQRs, the results from the time series application of AWIST are in agreement with other, direct assessments of recovery of the

macroinvertebrate assemblage observed at several sites in the UK AWMN data. The relative spread of EQRs across the 12 studies sites also reflect the acidity status of these well-studies lakes. These observations suggest that the new tool performs appropriately when tested using independent data and is able to capture and reflect important aspects of acidity status and differentiate between sites that differ in acidity status.

As an additional test of the AWIST tool a palaeoecological study was performed at three sites in Scotland; two acidified sites and one minimally impacted site. The AWIST predictions for the present-day macroinvertebrate assemblage suggest that the two acidified sites are currently in Moderate status, whilst the minimally impacted site (with respect to acidification) is in Good status. Whilst it is difficult to identify direct acidification-related changes for the wider macroinvertebrate subfossil assemblage, the results from the palaeoecological study suggest moderate damage to the macroinvertebrate communities of the two acidified lakes via increases in acidophilous species and / or changes in species diversity measures.

We believe that AWIST is now ready to be used for WFD purposes and to be inter-calibrated with other member tools.

6 References

Allott, T.E.H., Harriman, R. & Battarbee, R.W. (1992) Reversibility of acidification at the Round Loch of Glenhead in Galloway, Scotland. *Environmental Pollution* **77**(2): 219-225.

Appleby, P G, 2001. Chronostratigraphic techniques in recent sediments. In W.M. Last and J.P. Smol (eds.) Tracking Environmental Change Using Lake Sediments. Vol. 1: Basin Analysis, Coring, and Chronological Techniques. Kluwer Academic Publishers, Dordrecht. p171-203.

Appleby, P G, Richardson, N, Nolan, P J, 1992. Self-absorption corrections for well-type germanium detectors. *Nucl. Inst. & Methods B,* 71: 228-233.

Appleby, P G, Nolan, P J, Gifford, D W, Godfrey, M J, Oldfield, F, Anderson, N J & Battarbee, R W, 1986. 210Pb dating by low background gamma counting. Hydrobiologia, 141: 21-27.

Appleby, P.G. & Oldfield, F., 1978. The calculation of ^{210}Pb dates assuming a constant rate of supply of unsupported ^{210}Pb to the sediment. Catena, 5:1-8.

Armitage PD, Moss, D., Wright, JF and Furse, MT (1983). The performance of a new biological water quality score system based on macroinvertebrates over a wide range of unpolluted running-water sites. Water Research 17: 333-347.

Barnard PC (1985). An annotated check-list of the Trichoptera of Britain and Ireland. Entomologist's Gazette 36: 31-45.

Battarbee, R.W., Jones V.J., Flower R.J., Cameron N.G., Bennion H., Carvalho L. and Juggins S. (2001) Diatoms. In: Tracking Environmental Change Using Lake Sediments. Volume 3: Terrestrial, Algal, and Siliceous Indicators (Eds J.P. Smol, H.J.B. Birks and W.M. Last), Kluwer Academic Publishers, Dordrecht. p. 155- 202.

Birks, H.J.B., 1998. D.G. Frey & E.S. Deevey review #1 - Numerical tools in palaeolimnology - Progress, potentialities, and problems. Journal of Paleolimnology 20, 307-332.

Brodin, Y.W. 1986. The postglacial history of Lake Flarken, southern Sweden, interpreted from subfossil insect remains. Internationale Revue der Gesamten Hydrobiologie 71: 371-432.

Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J. 1984. Classification and Regression Trees. Chapman & Hall/CRC.

Breiman, L. 1996. Bagging predictors. Machine Learning 26(2): 123-140.

Breiman, L. 2001. Random Forests. Machine Learning 45, 5-32.

Brodin, Y.W. & Gransberg, M. 1993. Responses of insects, especially Chironomidae (Diptera), and mites to 130 years of acidification in a Scottish lake. Hydrobiologia 250: 201-212. Brooks, S.J. 2006. Fossil midges (Diptera: Chironomidae) as palaeoclimatic indicators of the Eurasian region. Quaternary Science Reviews 25: 1894-1910.

Brooks, S.J., Bennion, H. & Birks, H.J.B. 2001. Tracing lake trophic history with a chironomid-total

phosphorus inference model Freshwater Biology 46: 511-532.

Brooks, S.J., Langdon, P.G. and Heiri, O. 2007. The identification and use of Palaearctic Chironomidae larvae in palaeoecology. Quaternary Research Association Technical Guide no. 10. 276 pp.

Coope GR (1986). Coleoptera analysis. In handbook of Holocene paleoecology and Paleohydrology, Berglund BE (ed.). Wiley: Chichester.

De'Ath G., and Fabricius, K.E., 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. Ecology 81(11), 3178-3192.

Duff AG (2008). Checklist of Beetles of the British Isles. 2008 edition. Wells. Somerset.

Edington JM and Hildrew AG, (1995). Caseless Caddis Larvae of the British Isles. Freshwater Biological Association Scientific Publication No. 53.

Elliott JM (1996). British Freshwater Megaloptera and Neuoptera: A Key with Ecological Notes. Freshwater Biological Association Scientific Publication No. 54.

Extence CA, Balbi DM, Chadd RP (1999). River flow indexing using British macroinvertebrates: a framework for setting hydroecological objectives. Regulated Rivers 15: 543-574.

Flower, R.J., Battarbee, R.W. & Appleby, P.G. (1987) The recent palaeolimnology of acid lakes in Galloway, Southwest Scotland: diatom analysis, pH trends, and the role of afforestation. *Journal of Ecology* **75**: 797-824.

Freund, Y. and Schapire, R.E. 1997. A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences 55(1): 119-139.

Greenwood MT, Agnew MD, Wood PJ (2003).. The use of caddisfly fauna (Insecta: Trichoptera) to characterise the Late-glacial River Trent, England. Journal of Quaternary Science 18: 645-661.

Greenwood MT, Wood PJ, Monk WA (2006). The use of fossil caddisfly assemblages in the reconstruction of flow environments from floodplain paleochannels of the River Trent, England. Journal of Paleolimnology 35: 747-761

Grimm, E.C. 2004. TGView version 2.2. Illinois State Museum.

Heiri O. and Lotter A.F. 2001. Effects of low count sums on quantitative environmental reconstructions: an example using subfossil chironomids. Journal of Paleolimnology 26: 343-350.

Henrikson, L., Olofsson, J.B. & Oscarson, H.G. 1982. The impact of acidification on Chironomidae (Diptera) as indicated by subfossil stratification. Hydrobiologia 86: 223-229.

Howard LC, Wood PJ, Greenwood MT. (2008). Reconstructing riverine paleoflow regimes using subfossil insects (Coleoptera and Trichoptera): the application of the LIFE methodology to paleochannel sediments. Journal of Paleolimnology (in press).

Il'yashuk, B.P. and Il'yashuk, E.A. 2000. Paleoecological analysis of chironomid assemblages of a mountain lake as a source of information for biomonitoring. Russian Journal of Ecology 31: 353-358.

Juggins, S. (1991). ZONE, version 1.2. University of Newcastle, UK.

Juggins S (2007). C2 version 1.5. Software for ecological and palaeoecological data analysis and visualisation. Newcastle University. Newcastle upon Tyne. UK

Korhola, A., Olander, H. & Blom, T. (2000). Cladoceran and chironomid assemblages as quantitative indicators of water depth in subarctic Fennoscandian lakes. Journal of Paleolimnology 24: 43-54.

Magurran, AE (1988). Ecological Diversity and its measurement. Princetown University press, Princetown, New Jersey. 179pp.

May RM (1975). Patterns of species abundance and diversity. In Ecology and Evolution of Communities, Cody ML, Diamond JM (eds). Harvard University Press: Cambridge, MA: 81-120.

Monteith, D.T., Hildrew, A.G., Flower, R.J., Raven, P.J., Beaumont, W.R.B., Collen, P., Kreiser, A.M., Shilland, E.M., and Winterbottom, J.H. 2005. Biological responses to the chemical recovery of acidified fresh waters in the UK. Environmental Pollution 137: 83-101.

Monteith, D.T. And Simpson, G.L. 2007. Macroinvertebrate classification diagnostic tool development. Final report of the WFD60 project to the Scotland and Northern Ireland Forum For Environmental Research (SNIFFER).

Pinder, L.C.V. and Morley, D.J. 1995. Chironomidae as indicators of water quality – with a comparison of the chironomid faunas of a series of contrasting Cumbrian Tarns. pp. 272-293. In: Harrington, R. and Stork, N.E. (eds) Insects in a changing environment. Academic Press, London.

Pla,S., Monteith, D., Flower, R. and Rose, N. 2009. The recent palaeolimnology of a remote Scottish loch with special reference to the relative impacts of regional warming and atmospheric contamination. Freshwater Biology 54: 505-523.

Rahel FJ (2002). Homogenisation of freshwater faunas. Annual Review of Ecology and Systematics 33: 291-315.

Resh, V (1992). Recent Trends in the use of Trichoptera in water quality monitoring. Proc.7th Int. Symp. Trichoptera., Umea Sweden. Backhuys Publishers. Leiden, 1993.

Rosenberg DM and Resh VH. (1993). Freshwater biomonitoring and benthic macroinvertebrates. Chapman and Hall, New York.

Roux C and Castella E, (1987). Les Peuplements Larvaires de trichopteres des Anciens Lits Fluviaux dans trios Secteurs de la Plaine Alluviale du Haut-Rhone Francais. In proceedings of the 5th International Symposium On Trichoptera, Bournaud M, tachet H (eds), Lyon. 305-312.

Schartau AK, Moe SJ, Sandin L, McFarland B, Raddum GG (2008). Macroinvertebrate indicators of lake acidification: analysis of monitoring data from UK, Norway and Sweden. Aquatic Ecology 42: 293-305.

Stevenson, A.C., Juggins, S., Birks, H.J.B., Anderson, D.S., Anderson, N.J., Battarbee, R.W., Berge, F., Davis, R.B., Flower, R.J., Haworth, E.Y., Jones, V.J., Kingston, J.C., Kreiser, A.M., Line, J.M., Munro, M.A.R., Renberg, I., 1991. The surface waters acidification project palaeolimnology programme: modern diatom/ lake-water chemistry data-set. ENSIS Ltd, London.

Verdonschot PFM and Higler LWG (1992). Optima and tolerances of Trichoptera larvae for key factors in Dutch inland waters. Proc.7th Int. Symp. Trichoptera., Umea Sweden. Backhuys Publishers Leiden, 1993.

Wallace ID, Wallace, B and Philipson GN (1990), A Key to the case-bearing Caddis Larvae of Britain and Ireland. Freshwater Biological Association Scientific Publication No.51.

Wallace ID (1991). A Review of Trichoptera of Great Britain. Research and Survey in Nature Conservation, No. 32, Nature Conservancy Council, Peterborough.

Walker, I.R., 2001. Midges: Chironomidae and related Diptera. pp. 43-66. In: Smol, J.P. Birks, H.J.B. and Last, W.M. (eds), Tracking Environmental Change Using Lake Sediments. Volume 4. Zoological Indicators. Kluwer Academic Publishers, Dordrecht.

Wiederholm, T., editor, 1983: Chironomidae of the Holarctic region. Keys and diagnoses. Part 1. Larvae. Entomologica Scandinavica Supplement 19: 1-457.

Wright JF, Blackburn, JH,Gunn RJM, Furse MT, Armitage PD, Winder JM, Symes KL (1996). Macroinvertebrate frequency data for the RIVPACS III sites in Great Britain and their use in conservation evaluation.. Aquatic Conservation: Marine and Freshwater Ecosystems 6: 141-167.

7 Appendices

AWIST was built using the open source statistical software R (www.r-project.org), and the **randomForest**, **vegan** and **analogue** packages. The user interface (GUI) was built using the cross platform TCL-TK tool kit, that interfaces to a console based version of the tool. Users can choose to operate the tool via the GUI (function **awistGUI()**) or via the console based version (function **awist()**) directly within R.

The software is provided in the form of an R package for which the source code is available under the GNU General Public Licence version 2. Binaries (compiled versions of the package) are currently available for the Windows operating system, and could be built for the MacOS X operating system, although compiling from source is easier on MacOS X. The Windows binary is in the form of a zip archive.

To install the tool, a user will need to:

1. Install a recent version of the R software. Version 2.8.1 is current and can be downloaded from the UK CRAN mirror (http://www.stats.bris.ac.uk/R/bin/windows/base/). The file downloaded is a executable file that launches a standard installation wizard. To have the TCL-TK tool kit interface available in R, make sure that you select the User Installation option or specifically check the Tcl-Tk option in the list of available install options.

2. Once R is installed and running, three additional packages are required, all of which are available via CRAN. The easiest option, if internet access is available, is to install the packages from within R via the Packages menu > Install packages. The **randomForest** and **vegan** packages should be installed this way.

3. Currently AWIST requires a development version of the **analogue** package. The development version of this package will be sent to CRAN in the next few days, but until then, it can be installed from the development repository on R-Forge (http://analogue.r-forge.r-project.org) by running the following command at the R console prompt:

    ```
    install.packages("analogue", repos="http://R-Forge.R-project.org")
    ```

    In future, it will be sufficient to install **analogue** alongside **randomForest** and **vegan** as per step 2.

4. The final step to install AWIST is to install the AWIST zip archive. Again, the easiest way to do this is to go to the Packages menu again, but this time select the **Install packages from local zip file...** option. Select the **AWIST_0.1-0.zip** file and allow R to install the package. The AWIST zip file should have been made available to you with this report.

5. To check the installation has worked, run the following command to load the AWIST package into the R session:

    ```
    require("AWIST")
    ```

6. If installation has proceeded correctly, a series of message statements loading various packages

should be printed at the console and no errors reported.

At this point, AWIST is ready to be used. In future, to use the package a user will need only to start R and run the `require("AWIST")` command.

To run the GUI, the user needs to issue this command at the R console prompt:

    awistGUI()

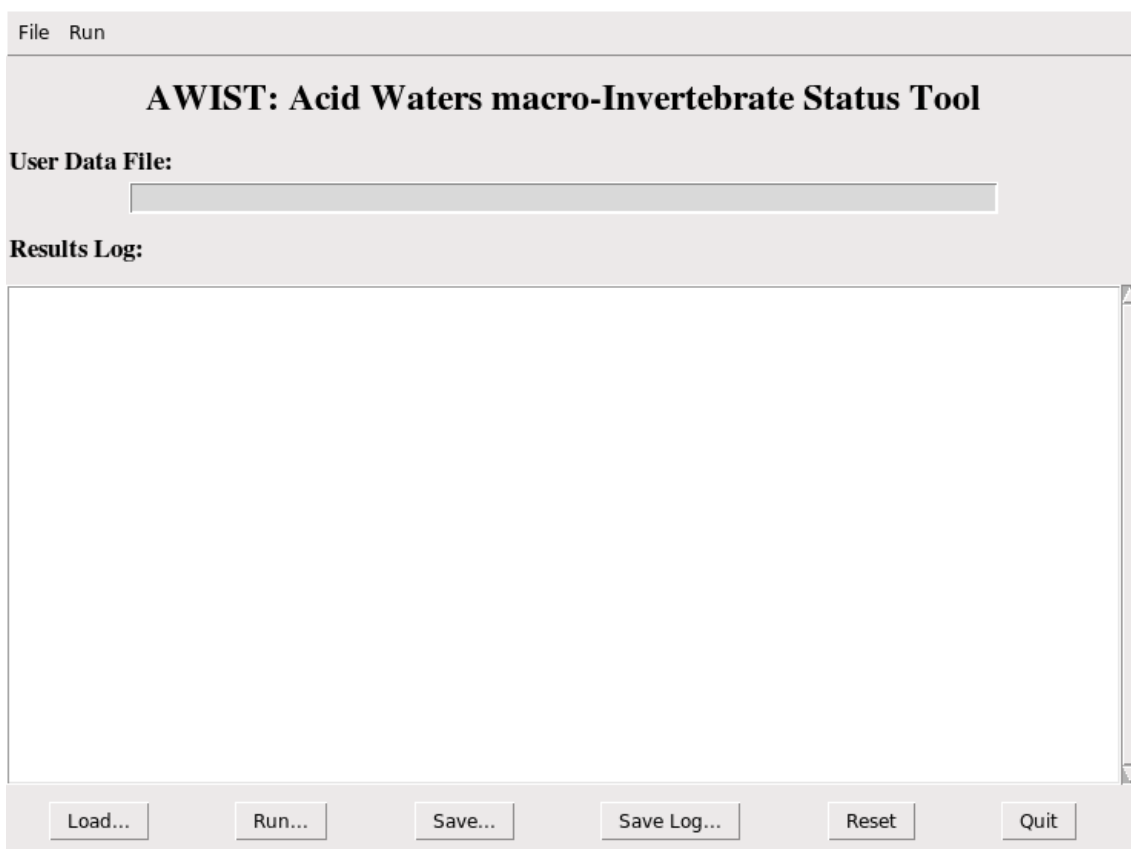This will launch a window similar to the one shown in Figure 3.

Figure 30: The AWIST user interface running within R

The tool is very simple, consisting of a field that displays the name of a loaded macro-invertebrate data file. Beneath this is a results log window, where the tool output is displayed. Underneath the log window is a series of buttons that operate the tool.

The tool is very easy to use, and consists of:

1. Loading a data file of macro-invertebrate counts on one or more sites into the tool.

2. Running the AWIST tool to generate predictions from the random forest and EQR scores and

confidence of class measures.

3. Saving the numerical output from the tool in CSV format.

4. Saving the textual log file of results if required.

5. Reset the tool to start another analysis, or Quit the tool.

These steps are initiated via the buttons, moving from left to right. The menus can also be used to control these steps.

The data files loaded by the tool are in standard 3 column data base format, stored as comma separated value (CSV)files. These files consist of three columns:

1. The site ID

2. The Taxon ID in standard format following Furse codes. Details of which are in Monteith and Simpson (2007).

3. The count for the indicated taxon in the indicated sample

An example CSV file can be found in the AWIST zip archive, in the data directory within the Rdata.zip file. It is called testdata.csv. A snippet of this file is included below:

```
"WBID","SPECIES","ABUN"
10307,10000000,1
10307,16220602,2
10307,20000000,5
10307,20110000,2
10307,24000000,1
10307,40210000,83
10307,40210102,45
10307,45140000,1
10307,45630000,1
10307,45630600,41
10307,48000000,2
10307,48240000,2
10307,48240402,7
10307,48240501,7
10307,48310105,1
10307,48340000,12
10307,48340600,1
10307,48370201,1
10307,"483A0401",3
10307,50140500,1
10307,50400000,28
```

The snippet include the counts for a sample collect at waterbody ID 10307. The first row should contain textual headers for the columns but the names of these columns are not important. The easiest way to

generate appropriate files is to set up the three columns in Microsoft Excel, enter the data and save the spreadsheet as a CSV file. These files should open natively in Excel on Windows so none of the benefits of storing data in Excel spreadsheets are lost, but the resulting CSV files are compact and can be read on any system.

It should be noted that the current version of the tool does very little checking of the user's data file. If there are problems with the file, the tool will simply print error messages to the R console.

A data file is loaded into the tool by clicking the Load... button. This opens a standard file selection dialogue box. The user navigates to the file they wish to run through the tool and clicks OK. The tool will then display the full path and name of the selected file in the User Data File box. Note this box cannot be edited.

AWIST is run on the loaded data file by clicking the Run... button. Depending on the number of samples and species in the loaded data file it may take a few seconds to process the data file, to produce the meta taxa computed as summaries of the abundance counts, before the results are displayed.

The results from the tool are displayed in the Results Log window. The sample ID is displayed, along with the predicted class. The next four columns give the probability that the sample belongs to each of the four classes. The EQR column contains the computed EQR score, whilst the final column states the number of actual taxa within the sample. Some meta data about the analysis is displayed above the results.

A CSV version of the results can be saved by clicking on the Save... button. The log file itself can be exported in TXT format by clicking on the Save Log... button. In both cases a file selection dialogue is displayed and the user is required to navigate to where they wish to save the results/log file and provide a name to save the results/log file under.

At any time, the tool can be reset, clearing the loaded data file and any results by clicking on the Reset button. To exit the tool, click the Quit button.