**Final Report** **Project WFD38**

# Phytoplankton Classification Tool for UK Lakes

**October 2006**

ENVIRONMENT AGENCY

SNIFFER
SCOTLAND & NORTHERN IRELAND FORUM FOR ENVIRONMENTAL RESEARCH

## Dissemination status

Unrestricted

## Authors and Research contractor

This document was produced by:
L. Carvalho, I. Dodkins, F. Carse, B. Dudley & S. Maberly

The document was reviewed by SNIFFER's project steering group.
The work formed part of a contract to the Centre for Ecology & Hydrology (CEH), Bush Estate, Penicuik, Midlothian, EH26 0QB. Email: laca@ceh.ac.uk

## SNIFFER's project manager

SNIFFER's project manager for this contract was:

Sian Davies, Environment Agency

**SNIFFER's project steering group members were:**

Sian Davies, Environment Agency
Geoff Phillips, Environment Agency
Bill Brierly, Environment Agency
Jan Krokowski, SEPA
Ian Fozzard, SEPA
Wayne Trodd, EPA Ireland
Kirsty Irvine, SNIFFER

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

Company No: SC149513
Scottish Charity: SCO22375

**www.sniffer.org.uk**

## EXECUTIVE SUMMARY

**WFD38:** Phytoplankton Classification Tool for UK Lakes: Phytoplankton Composition (October, 2006)

Project funders/partners: SNIFFER & Environment Agency

**Background to research**
The Environment Agency and SNIFFER have commissioned this R & D project to develop a method to classify the ecological status of lakes on the basis of phytoplankton. As part of this assessment, metrics need to be developed for phytoplankton community composition.

**Objectives of research**
Specific objectives for the project were to develop a robust classification, incorporating:

1. Prediction of reference scores for UK lakes based on phytoplankton composition
2. Developing criteria for defining the good/moderate boundary
3. Classifying the ecological status of a water body in to one of five status classes (High/Good/Moderate/Poor/Bad), based on the calculation of an Ecological Quality Ratio (EQR). An EQR being calculated from the relationship between current observed and reference phytoplankton community composition for a site
4. Determining uncertainty associated with the classification result, based on statistical confidence or probability of class

**Key findings and recommendations**
Following collation of a dataset of matching phytoplankton and environmental data from 189 lake samples, a multivariate approach to metric development was adopted. CCA was used to develop a species-environment model for phytoplankton, with the main typology variables (alkalinity, altitude, mean depth, lake area) included as significant explanatory variables in the model alongside two variables indicative of eutrophication pressure (chlorophyll and TP concentrations).

The model indicated strong correlations between a number of the explanatory variables, with the eutrophication pressure gradients (Chlorophyll and TP) closely correlated with alkalinity. This highlighted the potential problem of developing simple univariate optima of phytoplankton taxa against pressure gradients.

The current model explained only 6.8% of the variance in the phytoplankton composition data. This is low, but fairly typical of ecological datasets with large numbers of taxa with greatly varying biomass. The use of abundance data improved the model performance a little, but may need to be re-considered after estimates of analysis error of biovolume measurements have been quantified. A simpler model using presence/absence data may be acceptable. Further enlargement of the dataset, addition of further explanatory variables, such as colour or flushing rate, and taxonomic harmonisation should all help improve the model.

Optima were derived for 66 of the most common phytoplankton genera along both eutrophication gradients (chlorophyll and TP) using reciprocal averaging. Although this was still a univariate approach, the correlative effect with alkalinity (and to a lesser extent other typology variables) was removed through the calculation of an EQR by taking account of a site's typology in the reference score. This was done through the development of a regression model relating reference site scores to typology variables.

Internal validation of the derived metric showed a fairly strong correlation with chlorophyll concentrations in the water column ($r^2$ = 0.53). External validation on an independent dataset is, however, required to more accurately reflect the strength of the relationship.

Currently to obtain an EQR between 0 and 1 involves two transformation steps. Further guidance from LTT or ECOSTAT as to what forms of EQR scaling and transformation are acceptable would be beneficial.

Currently the H/G boundary is determined from the 75% of reference site scores, giving an EQR of 0.65. The remaining boundaries were derived from an equal division of the EQR scale between 0 and 0.65. This does, however, assume that the maximum impact score observed represents bad status with an EQR of 0.

**Data gaps and Further Work**

There is great scope for improving the phytoplankton model through further data collection. Despite 380 phytoplankton samples being counted, only 189 samples had matching chemistry and typology data, with Scottish samples (and many reference lakes) having a particularly poor match. Further data collection from all reference lake types are required (particularly shallow and deep medium and high alkalinity lakes). Across the whole pressure gradient, further data are required from very shallow low and medium alkalinity lakes and deep medium and high alkalinity lakes. Currently the nutrient pressure gradient is not spanned evenly with particular lake types either having few reference sites (e.g. high alkalinity lakes) or few highly impacted sites (e.g. low alkalinity lakes). Currently no data from Northern Ireland and the Republic of Ireland are included in the model. Phytoplankton samples from these regions need to be counted following the project standard guidelines and taxonomy and need to incorporate biovolume measurements.

Additional environmental data also needs recording and collating. Mean depth and alkalinity data are needed from all sampled sites and additional data on colour (Hazen units) and (modelled) flushing rate would be beneficial to examine if they added additional, independent, explanatory power to the model.

As well as model and metric development, the possibility of identifying class boundaries based on ecological thresholds (e.g. ratio of positive to negative indicators) needs to be examined. Further work is also required on estimates of error and consequent uncertainties in classification. Errors associated with sampling variability (by location and season) and analytical (counter) error requires further data collection. Development of a standard harmonised list of commonly recorded taxa would be of benefit alongside regular (annual) taxonomic workshops to minimise counter error.

Key words: phytoplankton, WFD, classification, lake, ecological status

## List of Contributors

**Laurence Carvalho**
Centre for Ecology & Hydrology
Bush Estate
Penicuik
Midlothian
EH26 0QB
Email: laca@ceh.ac.uk

**Ian Dodkins**
University of Ulster
Coleraine

**Fiona Carse & Stephen Maberly**
Centre for Ecology & Hydrology
Lancaster

**Bernard Dudley**
Centre for Ecology & Hydrology
Edinburgh

**Sian Davies**
Environment Agency
Haddiscoe

## Acknowledgements

The following phytoplankton counters provided the data essential for tool development and also offered advice on developing counting guidance:
Jo Girvan, Jane Fisher, Nadia Solvieva, Genevieve Madgwick, Sarah Pritchard, Bill Brierly & Tom Barker

The authors would also like to thank Nigel Willby & Robert Ptacnik for discussions on methodological approaches

## TABLE OF CONTENTS

**List of Tables**



**List of Figures**



**APPENDICES**

## 1. INTRODUCTION

The EC Water Framework Directive (WFD) is the most significant piece of European water legislation for over twenty years. A key component of the Directive is the development of ecological classification tools for determining the ecological status of waterbodies. Such tools need to be sensitive to specific environmental pressures. These include immediate pressures such as point and diffuse chemical inputs, and longer-term pressures such as climate change.

The Environment Agency and SNIFFER are funding the development of a package of new classification methods in order to satisfy the requirements of the WFD. WFD requires the ecological status of water bodies to be assessed on the condition of their biological quality elements (Article 8, annex V). For lakes this includes phytoplankton.

For this purpose, SNIFFER have commissioned this R & D project (WFD 38) to develop a method to classify the ecological status of lakes on the basis of phytoplankton communities. To fit with the requirements of the WFD, the phytoplankton classification scheme needs to ensure that it:

- distinguishes the 5 status classes High/Good/Moderate/Poor/Bad using reference-based Ecological Quality Ratios
- includes actual or relative abundance as well as composition
- is applicable to all UK lakes
- considers error in classification and potential of misclassification

Phytoplankton is widely considered to be the first biological community to respond to eutrophication pressures and is the most direct indicator of all the WFD Biological Quality Elements (BQEs) of nutrient concentrations in the water column (Carvalho et al., 2002). The phytoplankton community is, however, notoriously diverse and dynamic. Developing an ecological classification specifically in relation to nutrient pressures requires minimising the effects of seasonal variability associated with the changing physical and biological structure of the water column and magnifying the signal related to nutrient pressures.

Individual species or taxa can be positive or negative indicators in relation to nutrient pressures. Positive indicators include species of chrysophytes (e.g. *Dinobryon*), desmids (e.g. *Cosmarium*) and diatoms (e.g. *Cyclotella comensis*). Negative indicators include species of green algae (e.g. *Scenedesmus*), diatoms (e.g. *Stephanodiscus*) and many groups of cyanobacteria, such as the large colonial and filamentous genera *Microcystis*, *Aphanizomenon* and *Anabaena*. The latter are favoured by relatively stable stratification and high alkalinity and can, therefore, also form a significant natural component of the phytoplankton community in deep alkaline lakes, i.e. they do not necessarily always indicate impacted conditions. As taxonomic status at the phylum/class level does not consistently represent positive or negative indicators, higher taxonomic resolution to genus or species level may be necessary for classification tool development.

Alternatively, phytoplankton composition may be considered in terms of functional groups. 32 functional groupings have been identified by Reynolds et al. (2002) using a combination of experimental evidence, empirical data and expert opinion to group species with consistent functional properties or attributes. In principle, functional groups are more predictable than individual species or genera in terms of their response to nutrient conditions under a broad set of physical conditions and, therefore, can potentially be developed to indicate impacts of nutrient pressures more consistently. Assigning taxa to functional group does, however, still generally require taxonomic resolution to the genus level, and species-level identification is required for some taxa.

Phytoplankton composition data can be summarised at Species, Genus or higher taxonomic level groupings (Class, Order), or in terms of functional groups (Reynolds et al., 2002). 'Species' will be used in this report as a generic term of composition data, unless specific terms such as Genus, Class or functional group are directly applicable. Following comparisons of phytoplankton count data from different analysts, it was decided that, in general, species-level tools would be difficult to implement across UK (and even more so across Europe) due to lack of an agreed or harmonised taxonomy; Genus or higher taxonomic level groupings (Family, Class, Order) do, however, appear possible. Functional groups are also preferred to higher taxonomic class groupings (Class, Order) as they are slightly more diverse and should be more predictable in terms of nutrient pressures.

## 1.1 1.1 Approaches available

Ecological classification for the WFD requires the comparison of the biological composition (species, genera, classes or functional groups) of an individual lake to an expected reference condition (site or type-specific). The approaches available for classification tool development are largely dependent on the type of data available and whether the pressure being assessed is correlated with another natural environmental gradient (Table 1). Previous attempts to classify phytoplankton in terms of nutrient pressures have been based on a combination of experimental evidence and expert knowledge/judgement (Reynolds in Carvalho et al., 2002), largely due to the lack of consistent quantitative empirical data.

As part of this project, a relatively harmonised dataset of phytoplankton composition data has been assembled alongside matched environmental and pressure data, enabling more sophisticated quantitative approaches to be considered. It is widely acknowledged, however, that nutrient pressures are correlated with the natural alkalinity of freshwaters; both being greatest in lowland areas, with their more alkaline geology, more intensive agriculture and higher population densities compared with upland areas. This correlation means that either a type-specific classification of phytoplankton must be considered, distinguishing responses in lakes of different alkalinity types, or a multivariate approach should be adopted (Table 1).

**Table 1 Approaches available for biological classification**

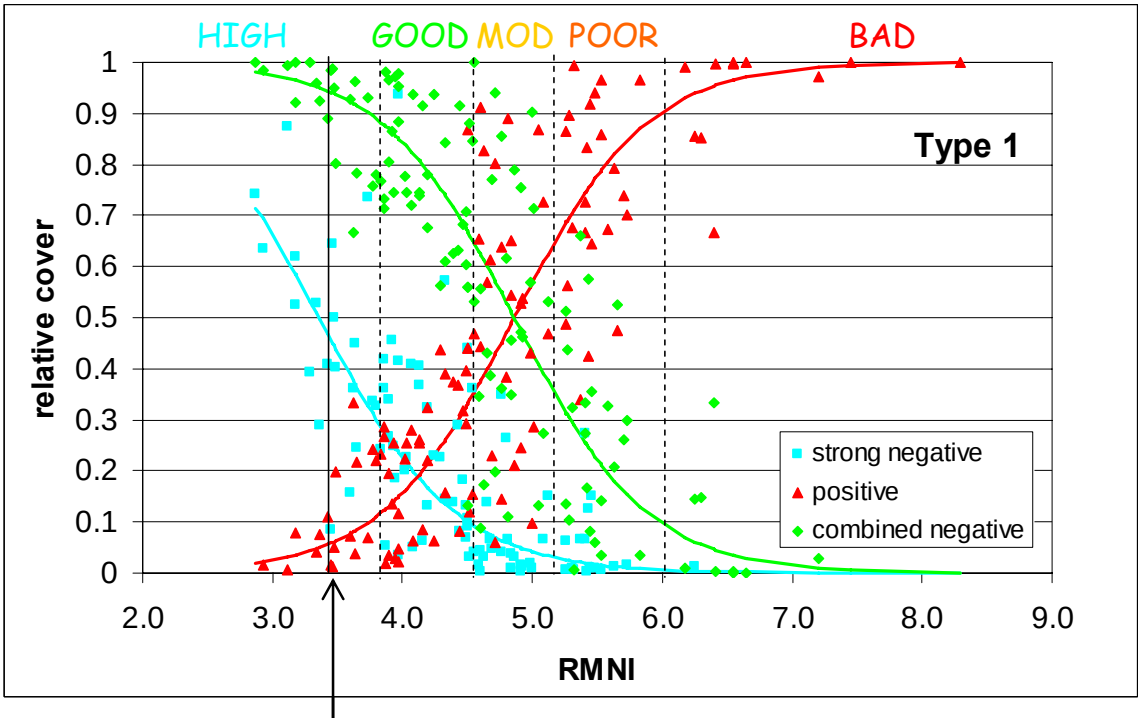
| Data Available | Amount of Data | Correlation with Pressure | Approach |
|---|---|---|---|
| Expert Knowledge Only | None | Unlikely | Expert Score (binary) |
| | | Probable | Type-specific Expert Score (binary) |
| Expt'l Evidence or Some Quantified Knowledge | Limited | Unlikely | Expert Score (scaled) |
| | | Probable | Type-specific Expert Score (scaled) |
| Matched Biology - Pressure - Environmental Data | Limited | Unlikely | Re-scaled expert score |
| | | Probable | Type-specific re-scaled expert score |
| | Extensive | Not present | Univariate Metric |
| | | Present | Multivariate / combined metric |

Some approaches being developed for the WFD take a type-specific approach (e.g. LEAFPACS, Nordic phytoplankton metric) using the relative abundance, or balance, of positive and negative indicators to construct a measure of ecological status. A predominance of positive indicators (taxonomic or functional group) of low pressure in a given waterbody type should be representative of reference conditions – although more widely tolerant taxa may be common and negative indicators may also be present in low abundance. A predominance of negative indicators suggests an impacted state, i.e. less

than good status. This approach requires taxa to be classified at a waterbody type level as positive or negative indicators based on expert scores or their optima and tolerance to particular pressures. The LEAFPACS Project re-scaled expert scores for aquatic macrophytes (Ellenberg Scores) using empirical data on species distributions. Although this approach has led to clear consistent rules on defining status class boundaries (Fig. 1), it was felt to include an additional unnecessary step in classifying species as positive or negative indicators on a type-specific basis. This requires a lot of data within a lake type and, is potentially problematic for sites that are borderline between two types.

**Figure 1 Boundary-setting River Macrophyte Nutrient Index**

**Potential class boundaries inferred from the changes in relative cover of macrophyte species identified as responding negatively and positively to a pressure gradient (RMNI – River Macrophyte Nutrient Index). The arrowed line indicates the median of the High class and is considered to equate to an EQR of 1.0. (Figure taken from Willby et al., 2006)**




Considering all this, a number of general points can be made on the philosophy adopted by this project for classification tool development:

- Type-specific classifications are restrictive and, in particular, problematic for sites close to type boundaries. The choice (and optima) of indicators will be restricted by where type boundaries are set. For this reason, a 'global' lake classification tool is preferred. Currently a global classification is necessary for phytoplankton due to data limitations within most lake types
- All taxa contain information, not just 'reference' or 'impact' taxa, or groups of recognised indicator value (e.g. chrysophytes or desmids)
- 'Species' optima contain more information if based on a continual scale along the impact gradient, rather than a binary system of reference/impact (or positive/negative)
- Using a community response is more robust as it does not necessarily rely on identifying all taxa in a sample and does not rely on the indicator values of just a few taxa that may not always be present

With community or assemblage data, there are three general approaches that can be adopted for this purpose (US EPA, 1999):

1. Multimetric assessment using an index that is the sum of several metrics. This is the basis of the Index of Biotic Integrity (IBI) (Karr et al. 1986).
2. Multimetric assessment using an index that is developed from a multivariate model to discriminate reference from impaired sites. This is the basis of the estuarine invertebrate indices developed by the EMAP-Estuaries program (USEPA 1993).
3. Multivariate assessment using ordination of species abundances. This methodology has been used widely in the assessment of UK rivers and streams, through the development of RIVPACS (e.g., Wright et al. 1984).

These three general approaches are outlined and compared in a report by the US EPA (1999). They are not the only possible approaches and there are numerous possible variations of the three general approaches above. According to the US EPA (1999), approaches 1 & 2 are "easy to apply in a continuing operational monitoring program because data from an individual site are entered into a formula, and the site's deviation from reference conditions can be known immediately. The ordination approach (3) requires reanalysis of the reference data set for each new batch of monitoring sites." The metric approach is also the easiest to explain to managers and the public as it only requires simple mathematics to use and is highly pressure-specific.

In their favour, multivariate approaches (e.g. DCA, CCA, etc.) provide useful exploratory tools for investigating and visualising patterns in compositional data. They allow testing of which environmental variables help explain significant variance in the composition data. Unlike basic multimetric approaches, they also allow for correlations between typology and impact variables to be taken into account (i.e. the widely recognised correlation between alkalinity and nutrient pressures).

After much discussion and consideration of the various approaches by a number of people across UK and Europe involved in WFD classification tool development, it was decided a combination multivariate-derived metric approach was most suitable, particularly given the limited phytoplankton data currently available in the UK. The combination approach adopted is the "CBAS methodology" (Dodkins et al., 2005) and can be briefly summarised as follows:

Multivariate methods are used to:

- Identify potential significant environmental drivers of phytoplankton composition. These environmental drivers include lake typology factors (alkalinity, altitude, depth, etc.) and pressure variables, or proxies of pressure, such as TP or chlorophyll
- Determine the correlations between significant typology and pressure variables.

Univariate methods (Reciprocal-Averaging) are then used to derive simple species 'optima' and 'tolerances' to the pressure variables identified in the multivariate analysis as being significant. These species optima and tolerances can then be used as a metric to measure the impact of those pressures at monitoring sites. Reciprocal Averaging maximises the spread of species optima along the impact gradient. This approach also produces metrics that are to the same scale ('standard deviations of species turnover') and thus can be added together (taking the correlations into account). The comparison of a site's metric score with a site's reference metric score is used to remove the correlation(s) between pressure and other significant environmental gradients.

All aspects of the approach require a relatively large matched phytoplankton and chemistry dataset spanning all lake types with a good coverage along the whole potential pressure gradient. A global model applicable to all lake types is being developed for phytoplankton as the datasets would have to be much larger to develop lake type-specific models.

Further details of the methodology being adopted is given in the methods section below and has also been described in Dodkins et al. (2005) and Dodkins & Rippey (2006).

## 2. METHODS

### 1.2 2.1 CBAS Methodology

CBAS is an acronym for CCA (Canonical Correspondence Analysis) Based Assessment System, developed initially for river macrophytes for the WFD (Dodkins et al., 2005). The CBAS methodology is both a multivariate and a (multi)metric approach. The methodology has undergone a number of refinements since the original published version, as documented in Dodkins & Rippey (2006). The development of a CBAS methodology for lake phytoplankton (ppCBASlak) can be summarised as requiring a number of steps:

1. Develop a multivariate model (CCA) using 'species' and environmental data, including lake typology and impact parameters that explain significant species variance.
2. Determine the univariate optima and tolerances of species along the impact gradients (through reciprocal averaging), in this case TP and chlorophyll$_a$ concentration gradients.
3. Calculate a 'metric score' at each reference site using the same approach as the Trophic Diatom Index (Kelly 1998), based on the optima, weighted by the abundance and indicator value (from tolerance) of the species present (for each impact gradient), using equation 1 below.

$$SiteScore = \frac{\sum a_i s_i v_i}{\sum a_i v_i} \qquad \text{(Equation 1)}$$

Where:
$a_i$ = abundance of $i$th taxon at the site
$s_i$ = optimum of $i$th taxon at the site
$v_i$ = indicator value of $i$th taxon at the site
[latter is inverse of tolerance]

4. Use Stepwise Multiple Linear Regression to derive a predictive equation relating these reference site metric scores to statistically significant lake typology variables. This predictive equation can then be used to derive site-specific expected reference metric values to be determined at new monitoring sites.

Following model development, determining ecological status at a site is simply done by:

1. Calculating a 'metric score' based on the optima, abundance and indicator value of the species present (for each impact gradient) (as in equation 1 above)
2. Using lake typology (alkalinity, altitude and mean depth) data from the monitoring site, determine the site-specific reference metric value, and then subtracting this from the site's observed metric value to produce an 'impact metric'* value.

3. As different impact metrics (e.g. TP and chlorophyll metrics) are correlated the correlation between them, as identified in the original CCA model, is removed. The uncorrelated eutrophication impacts can then be added together to produce a 'total ecological change' (TEC)** value for phytoplankton at a site in response to eutrophication pressures.
4. The TEC value is reversed and scaled between 0 and 1 to produce the EQR value
5. The gradient between the EQR values derived from the maximum and minimum TEC values for all known sites within the type is then divided into a number of equally spaced categories of ecological status; the number dependent on adequate coverage of sites along the impact gradient.

(*) The 'impact metric' value is a diagnostic measure, suggesting the impact that is occurring at that monitoring site. All metrics are scaled to the same units (standard deviations of species turnover).
(**) Total ecological change is also measured in SD of species turnover. Sites with a change of 4SDs are unlikely to have any species in common with the reference conditions.

### 1.3 2.2 Data availability

To ensure a reasonable level of taxonomic consistency, only phytoplankton composition data counted following a counting/taxonomy workshop was used to develop the tool. Counts were available from 380 samples taken from 167 lake basins. Of these samples, only 236 (111 sites) had matching chemistry information.

The majority of data available were for the months July to September (Fig. 2). For this reason, only a 'summer' model has been developed based on these three months. A model based on all samples, including the winter and spring samples, was rejected in order to reduce the effects of seasonality on composition. This left 189 samples available for model development, 27 of which (14%) were from reference sites (38 samples from Scotland, 151 from England & Wales).

Samples from high alkalinity shallow or very shallow lakes predominated, with very few matched samples from deep lakes (Table 2). A similar representation of lake types was not, however, present in the set of reference lakes used in the model, with the majority of reference lake samples taken from low alkalinity shallow or deep lakes (Table 3). Both the full dataset and the reference lake dataset highlight lake types that need further sampling to provide a balanced and adequate coverage across typology and pressure gradients.

Phytoplankton data were summarised for each sample as total biovolume by genus, although higher taxonomic units were used where genus level identification was not possible (e.g. unicellular centric diatoms, unidentified cyanobacteria, etc.). Genera occurring in less than 3 samples were excluded from the analysis.

Typology variables used in model development included: lake surface area, altitude, mean depth and mean alkalinity. Total phosphorus and chlorophyll concentrations were also included in model development as measures of eutrophication impact. Month was also included in the analysis to represent seasonal effects.

Typology, chemistry and phytoplankton biovolume data were all log transformed to normalise the data. Multivariate analysis (DCA and CCA) was carried out using CANOCO version 4.5.

**Figure 2 Number of phytoplankton samples counted by month**

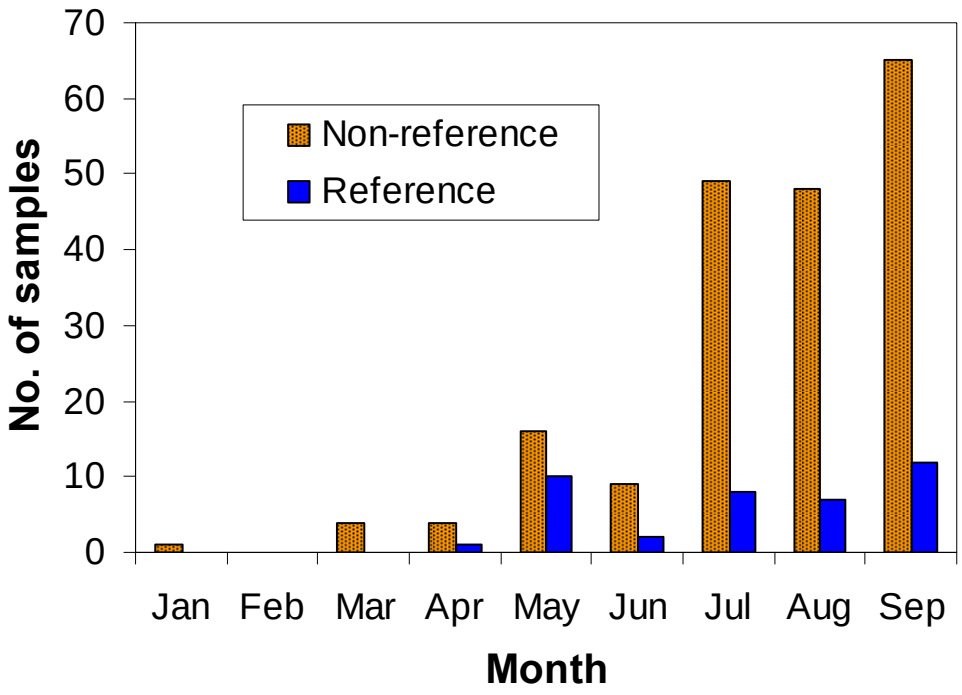



**Table 2 Samples in model classified according to GB depth and alkalinity classes**

| | Alkalinity Type | | | |
|---|---|---|---|---|
| | Low | Medium | High | Total |
| Very Shallow | 14 | 19 | 51 | 84 |
| Shallow | 24 | 29 | 31 | 84 |
| Deep | 14 | 7 | 0 | 21 |
| Total | 52 | 55 | 82 | 189 |

**Table 3 Samples from reference lakes in model classified according to GB depth and alkalinity classes**

| | Alkalinity Type | | | |
|---|---|---|---|---|
| | Low | Medium | High | Total |
| Very Shallow | 2 | 4 | 4 | 10 |
| Shallow | 9 | 1 | 1 | 11 |
| Deep | 6 | | | 6 |
| Total | 17 | 5 | 5 | 27 |

# 3. RESULTS

## 1.4 3.1 Model Development

Initial DCA of the Genus data alone revealed gradient lengths intermediate between those for which linear (<2) and unimodal (>3) responses would be expected (axis 1 gradient length 2.90 and axis 2 2.53). CCA was selected since it is a unimodal model which is also robust with linear gradients (Ter Braak and Šmilauer, 2002).

The output of the initial CCA analysis indicates log alkalinity explains the most variance in the phytoplankton composition data (2.5%) with the pressure variables, TP and chlorophyll, the next most important (2.1% and 2% respectively); total inertia was 5.338 (Table 4). The analysis also indicated the absence of a strong 'seasonal' effect in the July-September dataset, sample month being the variable that explained the least variance in the species data (0.7%) (Table 4)

**Table 4 Environmental variables in model and associated eigenvalue (λ) and significance**

| Variable | Eigenvalue (λ) | Significance (P-value) |
|---|---|---|
| log Alkalinity | 0.135 | 0.0001 |
| log TP | 0.113 | 0.0001 |
| log Chlorophyll | 0.109 | 0.0001 |
| log Depth | 0.070 | 0.0001 |
| log Area | 0.061 | 0.0001 |
| log Altitude | 0.054 | 0.0001 |
| Month | 0.036 | 0.0460 |

Stepwise manual forward selection was used in CANOCO to produce a species-environment model. After first including alkalinity in the model, altitude, mean depth and area were the next most important in explaining significant additional variance in the Genus data and were, therefore, all included in the CCA model (Table 5). There was then little difference between the choice of pressure variable to select, TP and chlorophyll were highly correlated and explained similar levels of variance in the Genus data after the typology variables had been selected. After selecting one, the other did not explain significant additional variance in the composition data taking into account Bonferonni-correction (Table 5). Chlorophyll was taken out first in the forward selection due to its slightly higher eigenvalue. Despite TP not explaining significant additional variance it was still retained in the model since it was felt that the difference between a TP and Chlorophyll metric response at certain sites would be important. Bonferroni correction is overly severe in downweighting significance and it was considered that a larger dataset would result in a TP gradient explaining significant additional variance.

**Table 5 Results of forward selection of environmental variables for model, highlighting conditional effects in CCA model**

| Variable | Eigenvalue (λ) | Significance (ρ) | Bonferonni-corrected significance required | Significant? |
|---|---|---|---|---|
| log Alkalinity | 0.135 | 0.0001 | 0.05 | Yes |
| log Altitude | 0.055 | 0.0001 | 0.025 | Yes |
| log Depth | 0.051 | 0.0001 | 0.0125 | Yes |
| log Area | 0.044 | 0.0009 | 0.00625 | Yes |
| log Chlorophyll | 0.044 | 0.0006 | 0.003125 | Yes |
| log TP | 0.037 | 0.0186 | 0.0015625 | No |

The environmental variables included in the model explained only 6.8% of the variance in the phytoplankton genus data – this is low, but typical of 'noisy' datasets with large numbers of taxa and rapidly varying biomass. The model was based on 189 samples with 116 active taxa and 6 selected environmental variables This is relatively few samples and environmental variables, reducing the magnitude of variance explained.

**Table 6 Summary statistics for the first four axes of CCA**

| CCA Axes | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Eigenvalues | 0.156 | 0.057 | 0.049 | 0.046 |
| Species-environment correlations | 0.846 | 0.648 | 0.683 | 0.695 |
| Cumulative percentage variance | | | | |
| of species data | 2.9 | 4.0 | 4.9 | 5.8 |
| of species-environment relationship | 43 | 58 | 72 | 84 |

The 'species'-environment biplot of the first two axes of the final CCA model is shown in Figure 3a. This highlights the close correlation between alkalinity, chlorophyll a and TP along the first axis, with altitude, mean depth and surface area more correlated with the 2$^{nd}$ axis. The correlation between alkalinity and the pressure variables highlighted in the plot illustrate clearly why a simple univariate metric alone cannot be used to assess impact, as the phytoplankton may simply be representing changes along an alkalinity gradient. It is for this reason that the metric score must be compared with a reference metric score that takes into account the typology, particularly alkalinity, and that ideally this reference score should be site-specific. Ordination of the 27 reference lakes data alone (Fig 3b), highlights that chlorophyll is less correlated with alkalinity than TP in undisturbed lakes, potentially making it a more suitable choice as a metric.

**Figure 3 Species-environment bi-plot of the first two axes of the CCA (a) full model and (b) reference lake model**

**Triangles represent location of 'genera'**

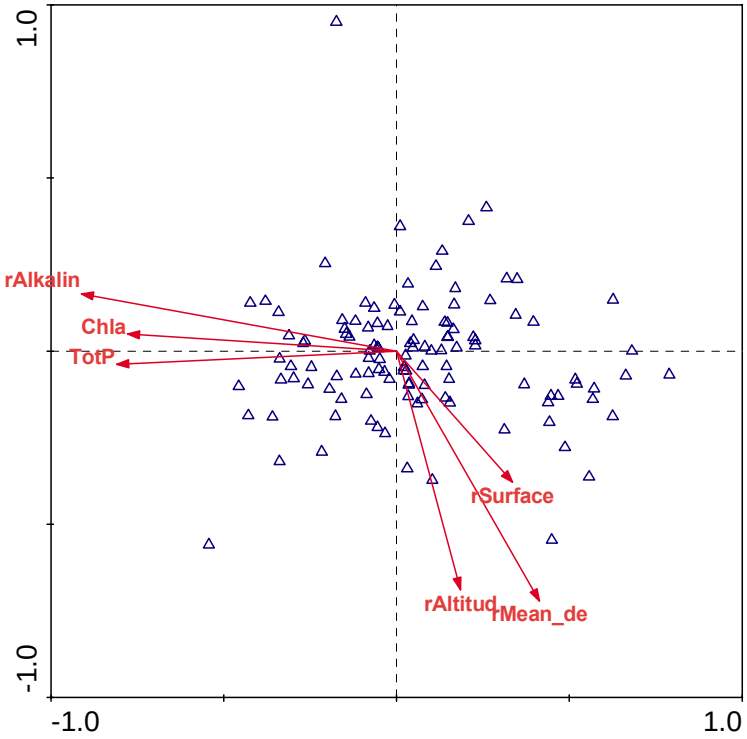
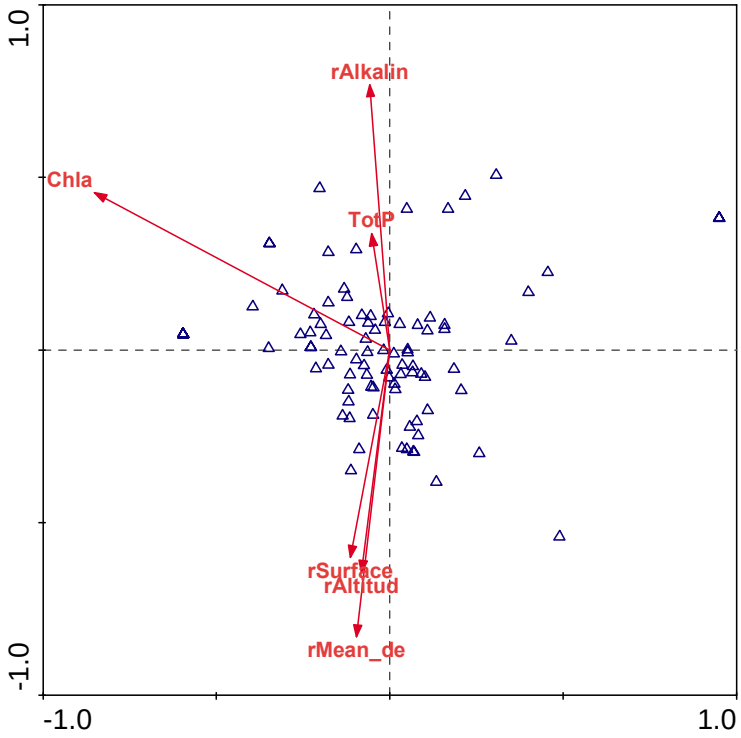


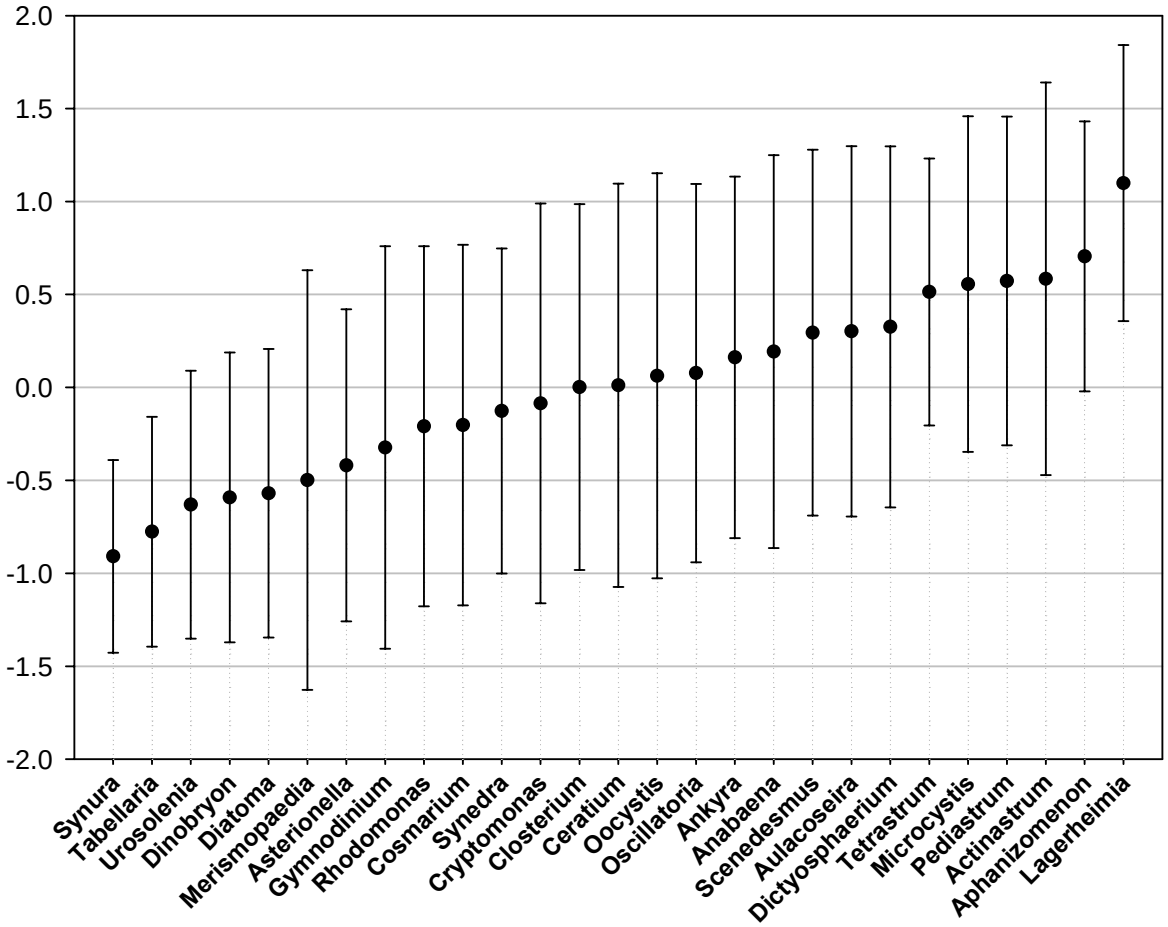

**Figure 3b**



### 1.5 3.2 Development of Metrics

'Genus' optima were obtained along both impact gradients (TP and Chlorophyll) by CCA analysis, with each considered in the model individually (i.e. univariate analysis). These optima are equivalent to abundance weighted-averages (WA), iteratively adjusted to ensure sites are weighted averages of species as well as species being the weighted averages of sites (reciprocal averaging (RA)). Hill's scaling was used to produce optima measured in standard deviations of species turnover. These allow comparable measures of ecological change within the dataset, allowing impacts of both metrics to be compared or combined. The 'chlorophyll' optima and tolerances of a few selected commonly occurring phytoplankton genera are illustrated in Figure 4. These are abundance-weighted and scaled to species turnover. The RA procedure in CCA is iterative and adjusts optima's based on species occurrences at sites to maximise their spread – potentially maximising the 'chlorophyll' signal they can differentiate.

**Figure 4 Weighted-average optima and tolerances of selected phytoplankton genera scaled to standard deviations of species turnover**




Indicator values were obtained from species 'tolerance' scores, by subtracting the tolerance score from the maximum tolerance score. This ensures that a small niche breadth (tolerance) produces a high indicator value.

**Table 7 Phytoplankton genera metrics: optima (chl-opt), tolerance (chl-tol) and indicator value (IndVal) listed from lowest (least impact) to highest (most impact)**

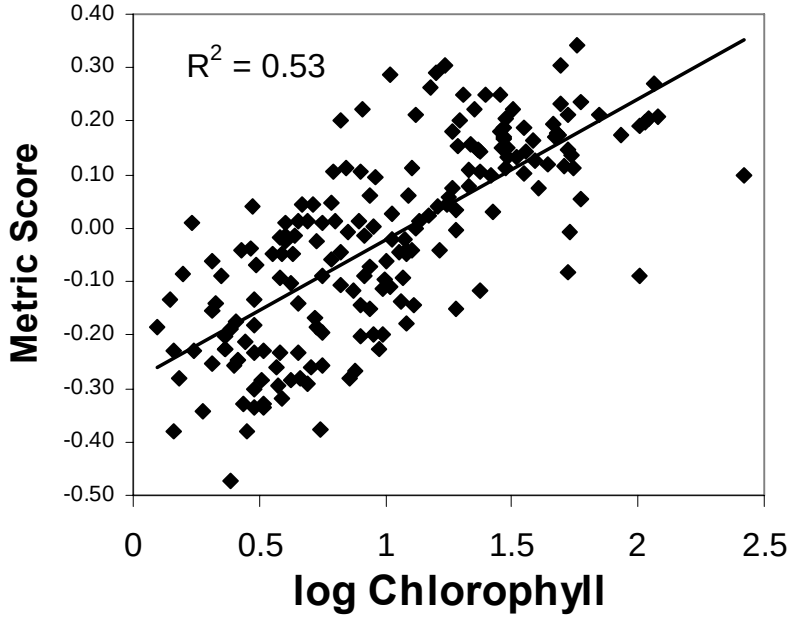
**The number of samples (N) each genus was recorded in is also indicated**

| GenusCode | Genus | N | Chl-opt | Chl-tol | IndVal |
|---|---|---|---|---|---|
| 1778 | Quadrigula | 9 | -1.3256 | 0.4398 | 0.8602 |
| 0801 | Chrysochromulina | 8 | -1.0261 | 0.6514 | 0.6486 |
| 0953 | Synura | 6 | -0.9091 | 0.5187 | 0.7813 |
| 2739 | Staurodesmus | 14 | -0.8619 | 0.6025 | 0.6975 |
| 1382 | Tabellaria | 31 | -0.7765 | 0.618 | 0.682 |
| 2721 | MOUGEOTIA | 17 | -0.6538 | 1.0799 | 0.2201 |
| 1220 | Urosolenia | 34 | -0.6309 | 0.7211 | 0.5789 |
| 1677 | Volvox | 6 | -0.6198 | 0.7275 | 0.5725 |
| 0923 | DINOBRYON | 62 | -0.592 | 0.7794 | 0.5206 |
| 1326 | Diatoma | 16 | -0.5698 | 0.7763 | 0.5237 |
| 0146 | MERISMOPEDIA | 26 | -0.4984 | 1.1286 | 0.1714 |
| 1308 | Asterionella | 50 | -0.4194 | 0.8394 | 0.4606 |
| 1626 | Eudorina | 16 | -0.3382 | 0.9404 | 0.3596 |
| 1725 | Crucigenia | 34 | -0.3266 | 1.1424 | 0.1576 |
| 0607 | GYMNODINIUM | 52 | -0.3235 | 1.0826 | 0.2174 |
| 1714 | CHLOROCOCCUM | 19 | -0.3004 | 0.9938 | 0.3062 |
| 0931 | Mallomonas | 50 | -0.2923 | 1.0557 | 0.2443 |
| 0402 | Euglena | 10 | -0.2443 | 1.1701 | 0.1299 |
| 0510 | RHODOMONAS | 132 | -0.2088 | 0.9685 | 0.3315 |
| 2737 | STAURASTRUM | 24 | -0.2047 | 0.8726 | 0.4274 |
| 2705 | COSMARIUM | 41 | -0.2025 | 0.9701 | 0.3299 |
| 2501 | Elakatothrix | 42 | -0.1649 | 1.0355 | 0.2645 |
| 1381 | Synedra | 62 | -0.1268 | 0.8743 | 0.4257 |
| 0935 | OCHROMONAS | 14 | -0.1163 | 1.1302 | 0.1698 |
| 1618 | CHLAMYDOMONAS | 115 | -0.1066 | 1.0878 | 0.2122 |
| 0504 | CRYPTOMONAS | 140 | -0.0859 | 1.0751 | 0.2249 |
| 0115 | COELOSPHAERIUM | 26 | -0.0633 | 0.9436 | 0.3564 |
| 1791 | Sphaerocystis | 33 | -0.0526 | 0.9553 | 0.3447 |
| 1743 | Golenkinia | 29 | -0.0365 | 1.1775 | 0.1225 |
| 0410 | Trachelomonas | 88 | -0.0257 | 1.113 | 0.187 |
| 0502 | Chroomonas | 50 | -0.0235 | 1.0127 | 0.2873 |
| 2704 | Closterium | 23 | 0.0016 | 0.9842 | 0.3158 |
| 0602 | Ceratium | 40 | 0.0112 | 1.0853 | 0.2147 |
| 0611 | Peridinium | 33 | 0.0171 | 1.2324 | 0.0676 |
| 0106 | Aphanothece | 46 | 0.0199 | 1.0818 | 0.2182 |
| 1337 | FRAGILARIA | 41 | 0.0468 | 0.8938 | 0.4062 |
| 1764 | Oocystis | 84 | 0.0625 | 1.0894 | 0.2106 |
| 0153 | Oscillatoria | 89 | 0.0773 | 1.0179 | 0.2821 |
| 1721 | COENOCHLORIS | 21 | 0.098 | 0.8304 | 0.4696 |
| 1717 | Closteriopsis | 58 | 0.1032 | 1.1682 | 0.1318 |
| 1758 | MONORAPHIDIUM | 127 | 0.1037 | 1.035 | 0.265 |
| 0407 | Phacus | 29 | 0.1042 | 1.1709 | 0.1291 |
| 1751 | KIRCHNERIELLA | 12 | 0.1272 | 1.0941 | 0.2059 |
| 1706 | Ankyra | 30 | 0.1618 | 0.9726 | 0.3274 |
| 0105 | APHANOCAPSA | 35 | 0.1692 | 1.1983 | 0.1017 |
| 0113 | Chroococcus | 68 | 0.1918 | 1.0841 | 0.2159 |
| 0102 | ANABAENA | 78 | 0.1929 | 1.0568 | 0.2432 |
| 1705 | ANKISTRODESMUS | 18 | 0.213 | 0.7308 | 0.5692 |
| 1783 | Schroederia | 29 | 0.2701 | 1.0755 | 0.2245 |
| 1781 | Scenedesmus | 105 | 0.2947 | 0.9842 | 0.3158 |
| 1203 | Aulacoseira | 57 | 0.3019 | 0.9962 | 0.3038 |
| 2738 | Staurastrum | 25 | 0.3145 | 1.1224 | 0.1776 |
| 1757 | MICRACTINIUM | 12 | 0.3176 | 1.082 | 0.218 |
| 1733 | Dictyosphaerium | 55 | 0.326 | 0.9713 | 0.3287 |
| 1796 | Tetraedron | 44 | 0.3412 | 0.9889 | 0.3111 |
| 0132 | Gomphosphaeria | 33 | 0.4327 | 0.9486 | 0.3514 |
| 1797 | Tetrastrum | 20 | 0.5136 | 0.718 | 0.582 |
| 0149 | MICROCYSTIS | 24 | 0.5559 | 0.9026 | 0.3974 |
| 1768 | Pediastrum | 52 | 0.5725 | 0.8843 | 0.4157 |
| 1702 | ACTINASTRUM | 19 | 0.5842 | 1.0561 | 0.2439 |
| 1720 | Coelastrum | 41 | 0.6248 | 1.0003 | 0.2997 |
| 0154 | Pannus | 6 | 0.6337 | 0.8414 | 0.4586 |
| 1769 | Planktosphaeria | 8 | 0.6968 | 0.9246 | 0.3754 |
| 0104 | Aphanizomenon | 30 | 0.7051 | 0.7259 | 0.5741 |
| 1801 | TREUBARIA | 14 | 0.9218 | 0.9417 | 0.3583 |
| 1754 | Lagerheimia | 16 | 1.0993 | 0.7427 | 0.5573 |

### 1.6 3.3 Validation of Metrics

Site scores were calculated using the log transformed 'Genus' abundance data for a site and the relevant Genus Chlorophyll and TP Optima and Indicator Values, using equation 1. Site scores were also calculated with no weighting given for abundance (i.e. data treated as presence/absence) and/or no weighting given to Indicator Value to examine whether these factors affect metric performance. The metric values were then compared with the measured values for the site to assess model performance, as illustrated in Figure 5.

**Figure 5 Scatterplot showing relationship between log chlorophyll and phytoplankton metric weighted by both abundance and indicator value**




The strength of the relationship between the metrics and observed chlorophyll and TP measured values is indicated by the coefficient of determination, $r^2$, the higher the value the better. This can be calculated from the original dataset (internal validation $r^2$) (Table 6) or from an independent test set (external validation $r^2$). An independent test set was not available, so for external validation the dataset was randomly split into two datasets with one used to generate new optima and tolerances and the 2nd dataset to validate it on. Because of the reduced dataset size, the true external $r^2$ is likely to be somewhere in between the reduced dataset $r^2$ and the internal $r^2$.

**Table 8 Coefficient of determination between metric site scores and observed values obtained from internal validation**

| Metric | Weighting of Optima | $r^2$ |
|---|---|---|
| Chlorophyll | Abundance and Indicator Value | 0.535 |
| | Abundance | 0.523 |
| | Indicator Value | 0.518 |
| | Optima only | 0.505 |
| TP | Abundance and Indicator Value | 0.499 |
| | Abundance | 0.498 |
| | Indicator Value | 0.492 |
| | Optima only | 0.492 |

The error of the metric is described by the root mean square error (RMSE). As for the $r^2$ values, these can be calculated based on the original dataset (the apparent RMSE) but more realistically on an independent test set (the RMSE of prediction or RMSEP). The lower the error, the better the metric performs.

### 1.7 3.4 Site-specific Reference Conditions

Site scores for 27 reference lake samples were calculated with the strongest metric model (incorporating abundance and Indicator Value). Stepwise regression was then carried out using these reference lake site scores to develop a model for predicting site-specific reference chlorophyll metric scores. Depth, alkalinity, altitude and area were all considered as potential predictor variables. The best predictive model, assessed by the model having the highest $r^2$ (53.1%), lowest PRESS statistic (0.303) and highest $r^2$(pred.) (44.3%) included depth, alkalinity and altitude as predictors (Table 7). This predictive equation can then be used to derive site-specific expected reference metric values to be determined at new monitoring sites.

**Table 9 Coefficients for reference condition optimal model**

| Predictor | Coef | SE Coef. | t | p |
|---|---|---|---|---|
| Constant | -0.00927 | 0.03913 | -0.24 | 0.814 |
| logDepth | -0.04436 | 0.03281 | -1.35 | 0.185 |
| logAlk | 0.10436 | 0.03484 | 3 | 0.005 |
| logAlt | -0.04735 | 0.02313 | -2.05 | 0.048 |

### 1.8 3.5 Site Impact Scores & EQR

To calculate a site impact score for a new monitoring site, the site metric score was calculated using the phytoplankton composition data and the site's reference metric score using typology data. These scores were transformed to an EQR value between 0 and 1 by the following:

1) Transform all reference and observed values to positive values by adding maximum negative value

2) EQR = E/O [Reference metric score] / [Observed metric score]

3) Transform E/O score obtained to an EQR between 0 and 1….

$EQR_{0\text{-}1}$ = (E/O) - (min'm E/O) / (max'm E/O) - (min'm E/O)

Or alternatively, transform E/O score in relation to median reference value

$EQR_{0\text{-}1+}$ = (E/O) - (min'm E/O) / (median ref E/O) - (min'm E/O)

## 1.9 3.6 Determining Boundary Values and Status Class

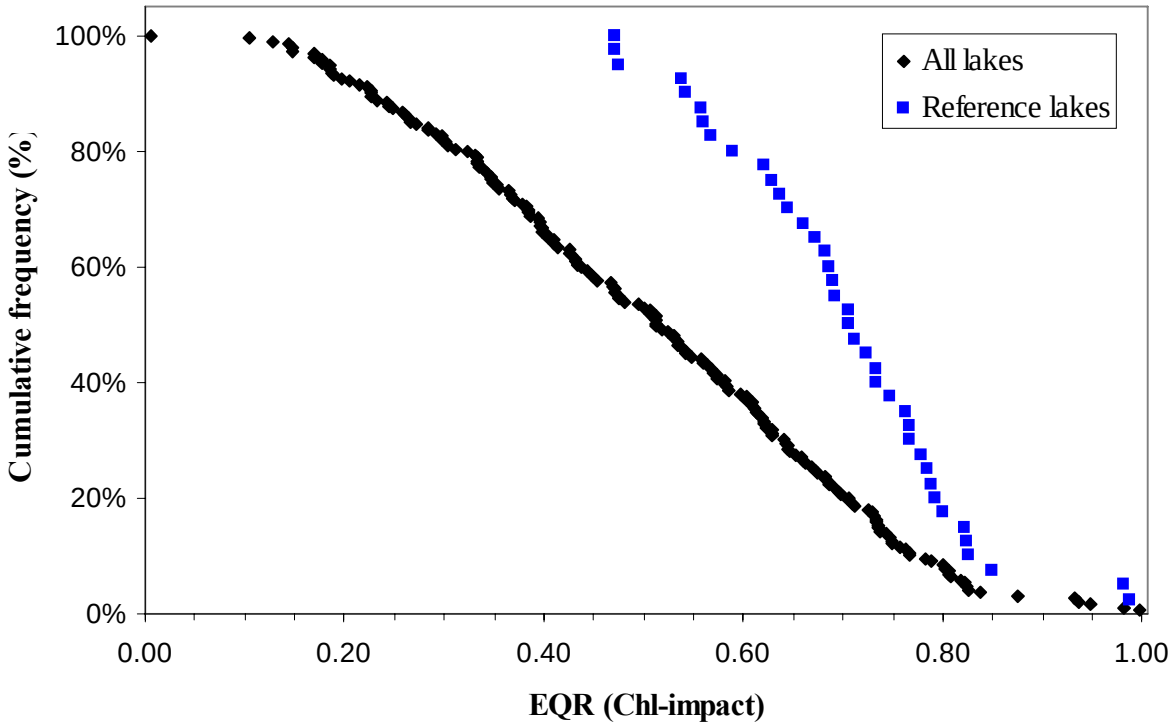
The H/G boundary was determined from the 75% of reference site scores, giving an EQR of 0.65 (Table 8). The remaining boundaries were derived from a geometric division of the EQR scale between 0 and 0.65. This does, however, assume that the maximum impact score observed represents bad status with an EQR of 0.

**Table 10 Proposed Interim EQR boundary values**

| Class Boundary | EQR |
|---|---|
| H/G | 0.65 |
| G/M | 0.49 |
| M/P | 0.33 |
| P/B | 0.16 |

The distribution of reference samples and all samples along the EQR scale highlights the generally higher EQR values obtained for reference lakes (Fig. 6). It also suggests that a reasonable part of the impact gradient was spanned. One non-reference lake (Ennerdale Water) had an EQR value higher than any reference lake

**Figure 6 Cumulative distribiution of EQR values for reference lakes and all lakes**




These interim boundary classes can be applied to the phytoplankton samples to derive a status class based on phytoplankton composition (Table 9). The results show good consistency of samples from different months from the same lake.

**Table 11 EQR and Status Class of samples based on Phytoplankton Composition**

| WBID | NAME | Year | Month | EQR | Status class |
|---|---|---|---|---|---|
| 2088 | Loch of Mey | 2004 | 9 | 0.80 | **High** |
| 2358 | Loch Calder | 2003 | 9 | 0.57 | **Good** |
| 2490 | Loch Hope | 2004 | 7 | 0.82 | **High** |
| 2499 | Loch Scarmclate | 2004 | 8 | 0.77 | **High** |
| 4204 | Loch Meadie | 2004 | 9 | 0.75 | **High** |
| 5222 | Loch Meadie | 2004 | 7 | 0.70 | **High** |
| 5350 | Loch Stack | 2004 | 7 | 0.82 | **High** |
| 6234 | Loch Culaidh | 2004 | 9 | 0.73 | **High** |
| 6405 | Loch Naver | 2004 | 7 | 0.68 | **High** |
| 11189 | Loch Osgaig | 2003 | 9 | 0.65 | **Good** |
| 11611 | Loch Brora | 2004 | 8 | 0.47 | **Moderate** |
| 12578 | Loch an Lagain | 2004 | 9 | 0.73 | **High** |
| 14057 | Loch Maree | 2004 | 9 | 0.63 | **Good** |
| 16456 | Loch Ussie | 2004 | 9 | 0.47 | **Moderate** |
| 18682 | Loch Druidibeag | 2004 | 8 | 0.77 | **High** |
| 20860 | Loch Insh | 2003 | 8 | 0.63 | **Good** |
| 22839 | Loch Laidon | 2004 | 7 | 0.75 | **High** |
| 23559 | Loch of Lowes | 2004 | 9 | 0.60 | **Good** |
| 24132 | Loch Earn | 2004 | 7 | 0.74 | **High** |
| 24459 | Loch Lubnaig | 2003 | 8 | 0.82 | **High** |
| 24919 | Lake of Menteith | 2004 | 7 | 0.63 | **Good** |
| 28165 | Greenlee Lough | 2004 | 7 | 0.45 | **Moderate** |
| 28165 | Greenlee Lough | 2004 | 8 | 0.45 | **Moderate** |
| 28165 | Greenlee Lough | 2004 | 9 | 0.18 | **Poor** |
| 28200 | Woodhall loch | 2004 | 9 | 0.47 | **Moderate** |
| 28386 | Talkin Tarn | 2004 | 8 | 0.21 | **Poor** |
| 28386 | Talkin Tarn | 2004 | 9 | 0.19 | **Poor** |
| 28847 | Bassenthwaite lake | 2004 | 7 | 0.36 | **Moderate** |
| 28847 | Bassenthwaite lake | 2004 | 9 | 0.76 | **High** |
| 28955 | Ullswater | 2004 | 7 | 0.53 | **Good** |
| 28965 | Derwent Water | 2004 | 8 | 0.93 | **High** |
| 28986 | Loweswater | 2004 | 8 | 0.67 | **High** |
| 28986 | Loweswater | 2004 | 9 | 0.67 | **High** |
| 29000 | Crummock water | 2004 | 9 | 0.54 | **Good** |
| 29021 | Thirlemere | 2004 | 9 | 0.62 | **Good** |
| 29052 | Buttermere | 2004 | 9 | 0.69 | **High** |
| 29062 | Ennerdale Water | 2004 | 8 | 0.74 | **High** |
| 29062 | Ennerdale Water | 2004 | 9 | 1.00 | **High** |
| 29183 | Wastwater | 2004 | 7 | 0.56 | **Good** |
| 29183 | Wastwater | 2004 | 9 | 0.66 | **High** |
| 29184 | Grasmere | 2004 | 7 | 0.61 | **Good** |
| 29184 | Grasmere | 2004 | 8 | 0.66 | **High** |
| 29184 | Grasmere | 2004 | 9 | 0.84 | **High** |
| 29222 | Elterwater | 2004 | 7 | 0.58 | **Good** |
| 29222 | Elterwater | 2004 | 8 | 0.64 | **Good** |
| 29222 | Elterwater | 2004 | 9 | 0.53 | **Good** |

**Table11 continued. EQR and Status Class of samples based on Phytoplankton Composition**

| WBID | NAME | Year | Month | EQR | Status class |
|---|---|---|---|---|---|
| 29233 | Windermere | 2004 | 7 | 0.43 | **Moderate** |
| 29233 | Windermere | 2004 | 9 | 0.43 | **Moderate** |
| 29270 | Blelham Tarn | 2004 | 7 | 0.49 | **Good** |
| 29270 | Blelham Tarn | 2004 | 9 | 0.70 | **High** |
| 29321 | Coniston Water | 2004 | 7 | 0.45 | **Moderate** |
| 29321 | Coniston Water | 2004 | 8 | 0.55 | **Good** |
| 29321 | Coniston Water | 2004 | 9 | 0.75 | **High** |
| 29328 | Esthwaite Water | 2004 | 7 | 0.18 | **Poor** |
| 29328 | Esthwaite Water | 2004 | 8 | 0.33 | **Moderate** |
| 29647 | Hawes Water | 2004 | 7 | 0.73 | **High** |
| 29647 | Hawes Water | 2004 | 9 | 0.98 | **High** |
| 29844 | Malham tarn | 2004 | 7 | 0.60 | **Good** |
| 30030 | Stocks Reservior | 2004 | 8 | 0.62 | **Good** |
| 30030 | Stocks Reservior | 2004 | 9 | 0.70 | **High** |
| 32359 | Derwent Reservior, Midlanc | 2004 | 7 | 0.67 | **High** |
| 32435 | Llyn Llygerian | 2004 | 8 | 0.52 | **Good** |
| 32435 | Llyn Llygerian | 2004 | 9 | 0.94 | **High** |
| 32459 | Ladybower Reservior | 2004 | 7 | 0.81 | **High** |
| 32459 | Ladybower Reservior | 2004 | 8 | 0.31 | **Poor** |
| 32459 | Ladybower Reservior | 2004 | 9 | 0.61 | **Good** |
| 32538 | Llyn Alaw | 2004 | 7 | 0.74 | **High** |
| 32538 | Llyn Alaw | 2004 | 8 | 0.43 | **Moderate** |
| 32538 | Llyn Alaw | 2004 | 9 | 0.35 | **Moderate** |
| 32650 | Rostherne mere | 2004 | 7 | 0.30 | **Poor** |
| 32650 | Rostherne mere | 2004 | 8 | 0.18 | **Poor** |
| 32650 | Rostherne mere | 2004 | 9 | 0.19 | **Poor** |
| 32744 | Mere mere | 2004 | 7 | 0.35 | **Moderate** |
| 32744 | Mere mere | 2004 | 8 | 0.26 | **Poor** |
| 32744 | Mere mere | 2004 | 9 | 0.41 | **Moderate** |
| 32761 | Llyn yr Wyth Eidion | 2004 | 7 | 0.73 | **High** |
| 32804 | Tatton Mere | 2004 | 7 | 0.39 | **Moderate** |
| 32804 | Tatton Mere | 2004 | 8 | 0.23 | **Poor** |
| 32804 | Tatton Mere | 2004 | 9 | 0.30 | **Poor** |
| 32948 | Llyn Dinam | 2004 | 8 | 0.40 | **Moderate** |
| 32960 | Tabley Mere | 2004 | 8 | 0.29 | **Poor** |
| 32961 | llyn Helyg | 2004 | 7 | 0.57 | **Good** |
| 32961 | llyn Helyg | 2004 | 8 | 0.47 | **Moderate** |
| 32961 | llyn Helyg | 2004 | 9 | 0.65 | **Good** |
| 32968 | Llyn Penrhyn | 2004 | 7 | 0.62 | **Good** |
| 32968 | Llyn Penrhyn | 2004 | 8 | 0.38 | **Moderate** |
| 32968 | Llyn Penrhyn | 2004 | 9 | 0.53 | **Good** |
| 33337 | llyn Coron | 2004 | 7 | 0.48 | **Moderate** |
| 33337 | llyn Coron | 2004 | 8 | 0.21 | **Poor** |
| 33337 | llyn Coron | 2004 | 9 | 0.47 | **Moderate** |
| 33474 | Oak Mere | 2004 | 8 | 0.14 | **Bad** |
| 33474 | Oak Mere | 2004 | 9 | 0.23 | **Poor** |
| 33627 | Llyn Rhos-ddu | 2004 | 7 | 0.69 | **High** |
| 33627 | Llyn Rhos-ddu | 2004 | 8 | 0.40 | **Moderate** |
| 33627 | Llyn Rhos-ddu | 2004 | 9 | 0.71 | **High** |

**Table 11 continued. EQR and Status Class of samples based on Phytoplankton Composition**

| WBID | NAME | Year | Month | EQR | Status class |
|---|---|---|---|---|---|
| 33730 | Llyn Padarn | 2004 | 7 | 0.43 | **Moderate** |
| 33730 | Llyn Padarn | 2004 | 8 | 0.28 | **Poor** |
| 33784 | Rudyard reservoir | 2004 | 8 | 0.27 | **Poor** |
| 33784 | Rudyard reservoir | 2004 | 9 | 0.35 | **Moderate** |
| 33803 | llyn Ogwen | 2004 | 8 | 0.34 | **Moderate** |
| 33803 | llyn Ogwen | 2004 | 9 | 0.44 | **Moderate** |
| 33836 | Llyn Idwal | 2004 | 8 | 0.62 | **Good** |
| 34002 | Llyn Cwellyn | 2004 | 7 | 0.33 | **Moderate** |
| 34002 | Llyn Cwellyn | 2004 | 8 | 0.37 | **Moderate** |
| 34002 | Llyn Cwellyn | 2004 | 9 | 0.33 | **Moderate** |
| 34400 | Llyn Conwy | 2004 | 7 | 0.71 | **High** |
| 34400 | Llyn Conwy | 2004 | 9 | 0.53 | **Good** |
| 34622 | Llyn Glasfryn | 2004 | 7 | 0.30 | **Poor** |
| 34622 | Llyn Glasfryn | 2004 | 8 | 0.61 | **Good** |
| 34622 | Llyn Glasfryn | 2004 | 9 | 0.57 | **Good** |
| 34780 | Hanmer Mere | 2004 | 8 | 0.40 | **Moderate** |
| 34780 | Hanmer Mere | 2004 | 9 | 0.59 | **Good** |
| 34987 | Llyn Tegid | 2004 | 8 | 0.51 | **Good** |
| 34987 | Llyn Tegid | 2004 | 9 | 0.37 | **Moderate** |
| 34990 | The Mere | 2004 | 7 | 0.52 | **Good** |
| 34990 | The Mere | 2004 | 9 | 0.24 | **Poor** |
| 35091 | White Mere | 2004 | 7 | 0.38 | **Moderate** |
| 35091 | White Mere | 2004 | 9 | 0.26 | **Poor** |
| 35211 | Crose Mere | 2004 | 7 | 0.60 | **Good** |
| 35211 | Crose Mere | 2004 | 9 | 0.51 | **Good** |
| 35640 | Hickling broad | 2004 | 7 | 0.48 | **Moderate** |
| 35640 | Hickling broad | 2004 | 8 | 0.51 | **Good** |
| 35724 | Aqualate Mere | 2004 | 7 | 0.34 | **Moderate** |
| 35724 | Aqualate Mere | 2004 | 8 | 0.32 | **Poor** |
| 35953 | Wroxham Broad | 2004 | 9 | 0.87 | **High** |
| 35981 | Rollesby Broad | 2004 | 7 | 0.68 | **High** |
| 35981 | Rollesby Broad | 2004 | 8 | 0.51 | **Good** |
| 36202 | Upton Broad | 2004 | 7 | 0.79 | **High** |
| 36202 | Upton Broad | 2004 | 8 | 0.57 | **Good** |
| 36331 | Cropston Reservoir | 2004 | 7 | 0.26 | **Poor** |
| 36331 | Cropston Reservoir | 2004 | 9 | 0.23 | **Poor** |
| 36405 | Tal y llyn lake | 2004 | 9 | 0.61 | **Good** |
| 36523 | Chasewater | 2004 | 8 | 0.39 | **Moderate** |
| 36523 | Chasewater | 2004 | 9 | 0.51 | **Good** |
| 36544 | Bomere | 2004 | 7 | 0.10 | **Bad** |
| 36566 | Betton Pool | 2004 | 8 | 0.28 | **Poor** |
| 38214 | Craig Goch Reservoir | 2004 | 7 | 0.58 | **Good** |
| 38214 | Craig Goch Reservoir | 2004 | 8 | 0.73 | **High** |
| 38214 | Craig Goch Reservoir | 2004 | 9 | 0.81 | **High** |
| 38310 | Grafham water | 2004 | 7 | 0.13 | **Bad** |
| 38310 | Grafham water | 2004 | 8 | 0.25 | **Poor** |
| 38310 | Grafham water | 2004 | 9 | 0.15 | **Bad** |

**Table 11 continued. EQR and Status Class of samples based on Phytoplankton Composition**

| WBID | NAME | Year | Month | EQR | Status class |
|---|---|---|---|---|---|
| 38409 | Llyn Egnant | 2004 | 7 | 0.64 | Good |
| 38409 | Llyn Egnant | 2004 | 9 | 0.15 | Bad |
| 38422 | Llyn Eidwen | 2004 | 7 | 0.30 | Poor |
| 38422 | Llyn Eidwen | 2004 | 8 | 0.23 | Poor |
| 38422 | Llyn Eidwen | 2004 | 9 | 0.50 | Good |
| 38525 | llyn Gynon | 2004 | 7 | 0.80 | High |
| 38525 | llyn Gynon | 2004 | 8 | 0.73 | High |
| 38525 | llyn Gynon | 2004 | 9 | 0.62 | Good |
| 38907 | Llyn Berwyn | 2004 | 9 | 0.70 | High |
| 39450 | Stewartby Lake | 2004 | 7 | 0.27 | Poor |
| 39450 | Stewartby Lake | 2004 | 8 | 0.51 | Good |
| 39450 | Stewartby Lake | 2004 | 9 | 0.40 | Moderate |
| 39967 | Usk reservoir | 2004 | 7 | 0.38 | Moderate |
| 39967 | Usk reservoir | 2004 | 8 | 0.54 | Good |
| 39967 | Usk reservoir | 2004 | 9 | 0.65 | Good |
| 40067 | Llangorse lake | 2004 | 7 | 0.37 | Moderate |
| 40067 | Llangorse lake | 2004 | 8 | 0.35 | Moderate |
| 40067 | Llangorse lake | 2004 | 9 | 0.20 | Poor |
| 40755 | Stanborough Lake | 2004 | 9 | 0.34 | Moderate |
| 41427 | Hanningfield reservoir | 2004 | 9 | 0.19 | Poor |
| 41559 | Cotswold Water Park | 2004 | 9 | 0.56 | Good |
| 41602 | Bosherston Central | 2004 | 7 | 0.66 | High |
| 41602 | Bosherston Central | 2004 | 8 | 0.65 | High |
| 41602 | Bosherston Central | 2004 | 9 | 0.95 | High |
| 43096 | Chew Valley Lake | 2004 | 9 | 0.38 | Moderate |
| 43348 | Cheddar Reservoir | 2004 | 8 | 0.81 | High |
| 43348 | Cheddar Reservoir | 2004 | 9 | 0.82 | High |
| 43602 | Bough Beech reservoir | 2004 | 9 | 0.22 | Poor |
| 43909 | Shear Water | 2004 | 7 | 0.17 | Poor |
| 43909 | Shear Water | 2004 | 8 | 0.01 | Bad |
| 43943 | Frensham Little Pond | 2004 | 7 | 0.41 | Moderate |
| 43943 | Frensham Little Pond | 2004 | 8 | 0.43 | Moderate |
| 43943 | Frensham Little Pond | 2004 | 9 | 0.40 | Moderate |
| 44031 | Frensham Great Pond | 2004 | 7 | 0.33 | Moderate |
| 44031 | Frensham Great Pond | 2004 | 8 | 0.44 | Moderate |
| 44031 | Frensham Great Pond | 2004 | 9 | 0.24 | Poor |
| 44518 | Fonthill Lake | 2004 | 7 | 0.41 | Moderate |
| 44518 | Fonthill Lake | 2004 | 9 | 0.35 | Moderate |
| 45108 | Burton Mill pond | 2004 | 9 | 0.69 | High |
| 45652 | Hatchett Pond | 2004 | 9 | 0.78 | High |
| 46102 | Little Sea Lake | 2004 | 7 | 0.56 | Good |
| 46102 | Little Sea Lake | 2004 | 8 | 0.54 | Good |
| 46102 | Little Sea Lake | 2004 | 9 | 0.76 | High |
| 46232 | Dozmary Pool | 2004 | 9 | 0.43 | Moderate |
| 46279 | Burrator Reservoir | 2004 | 9 | 0.58 | Good |
| 46501 | Stithians Reservoir | 2004 | 9 | 0.40 | Moderate |
| 46556 | The Loe | 2004 | 9 | 0.17 | Poor |

# 4. DISCUSSION

## 1.10 4.1 Species-environment model

Given the collation of a dataset of matching phytoplankton and environmental data, a multivariate approach to metric development was considered most appropriate. CCA was used to develop a species-environment model for phytoplankton, with the main typology variables (alkalinity, altitude, mean depth, lake area) included as significant explanatory variables in the model alongside two variables indicative of eutrophication pressure (chlorophyll and TP concentrations).

The model indicated strong correlations between a number of the explanatory variables, with the eutrophication pressure gradients (Chlorophyll and TP) closely correlated with alkalinity. The fact that phytoplankton composition changes are likely to be similar in response to increasing alkalinity as they are with increasing nutrient gradients is well established (Lund 1961; Shapiro 1990) and highlight the potential problem of developing simple univariate metric scores of phytoplankton taxa against pressure gradients.

Optima were derived for 66 of the most common phytoplankton genera along both eutrophication gradients (chlorophyll and TP) using reciprocal averaging, which is similar to weighted averaging, but maximises the 'pressure signal' that taxa can differentiate. Although this was still a univariate approach, the correlative effect with alkalinity (and to a lesser extent other typology variables) is removed through the calculation of an EQR by taking account of a site's typology in the reference score. This was done through the development of a regression model relating reference site scores to typology variables.

The current model explained only 6.8% of the variance in the phytoplankton composition data. This is low, but fairly typical of ecological datasets with large numbers of taxa and rapidly varying biomass in a relatively small dataset (compared to number of taxa). The use of abundance data improved the model performance a little, but may need to be re-considered after estimates of analysis error have been quantified. A simpler model using presence/absence data only may be acceptable.

The phytoplankton community is clearly very dynamic with many key factors that shape the community (stability of water column, grazer densities, humic content) are not incorporated in the current model. Further enlargement of the dataset, addition of further explanatory variables, such as colour or flushing rate, and taxonomic standardisation should all help improve the model.

Despite the pressure variables only explaining about 2% of the variability in the composition data, the internal validation of the derived metric showed a fairly strong correlation with chlorophyll concentrations in the water column ($r^2$ = 0.53). External validation on an independent dataset is, however, required to more accurately reflect the strength of the relationship.

## 1.11 4.2 EQR calculation and boundary-setting

Currently species optima are untransformed CCA axis scores ranging from (-1.33 to 1.10). This results in many site scores from relatively unimpacted lakes often having negative values. Currently, therefore, a transformation of scores is needed to derive an EQR. It may be more user-friendly to transform species optima to standardised scores (such as a range between 0-10) or scores reflecting an equivalent chlorophyll-optima the species is most likely to be found at.

Furthermore to obtain an EQR between 0 and 1 involves a second transformation. Various transformation options are possible that affect the EQR value and hence the boundary EQR values. Further guidance is required from LTT or ECOSTAT as to what level of EQR scale and transformations are acceptable.

Currently the H/G boundary is determined from the 75% of reference site scores, giving an EQR of 0.65. The remaining boundaries were derived from an equal division of the EQR scale between 0 and 0.65. This does, however, assume that the maximum impact score observed represents bad status with an EQR of 0. The possibility of identifying boundaries based on ecological thresholds (e.g. ratio of negative to positive indicators, rise in % cyanobacteria) needs to be examined further.

## 1.12 4.3 Data gaps

There is great scope for improving the phytoplankton model through further data collection. Despite 380 phytoplankton samples being counted, only 189 samples had matching chemistry and typology data, with Scottish samples (and many reference lakes) having a particularly poor match

Further data collection from all reference lake types are required (particularly shallow and deep medium and high alkalinity lakes) (see Table 3). Across the whole pressure gradient, further data are required from very shallow low and medium alkalinity lakes and deep medium and high alkalinity lakes (see Table 2). Currently the nutrient pressure gradient is not spanned evenly with particular lake types either having few reference sites (e.g. high alkalinity lakes) or few highly impacted sites (e.g. low alkalinity lakes).

Additional environmental data also needs recording and collating. Mean depth and alkalinity data are needed from all sampled sites and additional data on colour (Hazen units) and (modelled) flushing rate would be beneficial from all sites.

Currently no data from Northern Ireland and the Republic of Ireland are included in the model. Phytoplankton samples from these regions need to be counted following the project standard guidelines and taxonomy and need to incorporate biovolume measurements.

## 1.13 4.4 Further work

As well as further work on model development, validation and boundary-setting described above, further work is required on error estimates: particularly quantification of sampling and analytical (counter) error using replicate samples and ring-test data.

Data quality checks need to be carried out on existing and new data to assess individual counter's taxonomy is not a major issue. Development of a standard harmonised list of commonly recorded taxa would be of benefit alongside regular (annual) taxonomic workshops.

Further analysis of the datasets could also be carried out examining the taxonomic resolution required for optimal model performance: identifying which genera show similar responses (and potentially could be lumped) and which species within a genus should ideally be discriminated (e.g. *Anabaena* spp.). The value of identifying planktonic diatoms to higher resolution could also be considered, but would require samples to be split and analysed separately for diatom identities.

## 5. REFERENCES


Carvalho L, Bennion H, Dawson H, Furse M, Gunn I, Hughes R, Johnston A, Luckes S, Maitland P, May L, Monteith D, Taylor R, Trimmer M and Winder J, 2002. *Nutrient conditions for different levels of ecological status and biological quality in surface waters*. R&D Technical Report P2-260/4/TR to the Environment Agency, June 2002, 298 pp.

Carvalho L, Maberly S, May L, Reynolds C, Hughes M, Brazier R, Heathwaite L, Liu S, Hilton J, Hornby D, Bennion H, Elliott A, Willby N, Dils R, Pope L, Phillips G & Fozzard I. 2004. *Risk Assessment Methodology for Determining Nutrient Impacts in Surface Freshwater Bodies*. Environment Agency of England and Wales and Scottish Environment Protection Agency.

Dillon PJ & Rigler FH. 1974. A test of a simple nutrient budget model predicting the phosphorus concentration in lake water. *J. Fish. Res. Bd. Can.*, **31**: 1771-1778.

Dodkins I, Rippey B & Hale P, 2005. An application of canonical correspondence analysis for developing ecological quality assessment metrics for river macrophytes. *Freshwater Biology*, **50**, 891-904.

Dodkins I & Rippey B, 2006. *mCBASriv: A Method for Assessing the Ecological Status of Rivers Using Macrophytes. Scientific Summary v3.0*, Internal Report to NSShare, March 2006.

Kelly M.G., 1998. Use of the trophic diatom index to monitor eutrophication in rivers. *Water Research*, **32**, 236-242

Kirchner WR & Dillon PJ. 1975. An emperical method of estimating the retention of phosphorus in lakes. *Water Resources Research*, **11**: 182-183.

Lund J W G, 1961 The algae of the Malham Tarn district. *Field Studies*, **1**, 85-115

McGowan S, Britton G, Haworth E & Moss B, 1999. Ancient blue-green blooms. *Limnology & Oceanography*, **44**, 436-439.

Moss B, Stephen D, Alvarez C, Becares E, van de Bund W, Collings SE, van Donk E, de Eyto E, Feldmann T & Fernandez-Alaez C. 2003. The Determination of Ecological Status in Shallow Lakes-a Tested System (ECOFRAME) for Implementation of the European Water Framework Directive /. *Aquatic Conservation*, **13**: 507-550.

OECD. 1982. Eutrophication of Water: Monitoring, Assessment and Control. Organisation for Economic Cooperation and Development, Paris.

Oliver RL & Ganf GG, 2000. Freshwater Blooms, In Whitton & Potts (Eds), *The Ecology of Cyanobacteria: Their Diversity in Time and Space*, Kluwer Academic, Dordrecht.

Phillips, G. 2005. *Proposed boundaries for Phytoplankton Biomass and supporting Phosphorus standards*. Version 2. Lakes Task Team Paper 084, March 2005.

Reynolds CS. 1984. *The ecology of freshwater phytoplankton*. Cambridge University Press.

Reynolds CS. 1999. Modelling phytoplankton dynamics and its application to lake management. *Hydrobiologia*, **396**: 123-131.

Reynolds CS, Huszar V, Kruk C, Naselli-Flores L & Melo S. 2002. Towards a functional classification of the freshwater phytoplankton. *Journal of Plankton Research*, **24**: 417-428.

Sas H. 1989. *Lake restoration by reduction of nutrient loading*. Academia Verlag Richartz GmbH. St. Augustin.

Scottish Executive Health Department, 2002. *Blue-Green Algae (Cyanobacteria) in Inland Waters: Assessment and Control of Risks to Public Health. Revised Guidance,* Compiled by the Scottish Executive Health Department Blue-green Algae Working Group.

Shapiro J, 1990 Current beliefs regarding dominance by blue-greens: the case for the importance of CO2 and pH. *Verhandlungen Internationale Vereinigung Limnologie*, **24**, 38-54

Solheim AL, Andersen T, Brettum P, Bækken T, Bongard T, Moy F, Kroglund T, Olsgard F, Rygg B & Oug E. 2004. BIOKLASS - Klassifisering av økologiske status i norske

vannforekomster: Forslag til aktuelle kriterier og foreløpige grenseverdier mellom god og moderat økologisk status for utvalgte elementer og påvirkninger. Norsk institutt for vannforskning (NIVA), Oslo.

Sondergaard M, Jeppesen E, Jensen JP, Bradshaw E, Skovgaard H & Grunfeld S. 2003. Vandrammedirektivet og danske soer Undertitel: Del 1: Sotyper, referencetilstand og okologiske kvalitetsklasser. Danmarks Miljøministeriet, Copenhagen.

Sondergaard M, Jeppesen E, Jensen JP & Amsinck SL, 2005. Water Frameork Directive: ecological classification of Danish lakes. *Journal of Applied Ecology*,

Ter Braak, C.J.F. and Šmilauer P. 2002: *CANOCO Reference Manual and CanoDraw for Windows User's Guide: Software for Canononical Community Ordination* (version 4.5). Ithaca N.Y.: Microcomputer Power, 500 pp.

Tilzer MM. 1988. Secchi Disk-Chlorophyll Relationships in a Lake with Highly Variable Phytoplankton Biomass. In.

US EPA, 1999. Statistical Analysis Methods for Biological Assessment. Appendix E in US EPA "*Lake and Reservoir Bioassessment and Biocriteria Technical Guidance Document*" EPA 841-B-98-007, US EPA, Washington

Van der Berg M, Baretta-Bekker H, Bijkerk R, van Dam H, Ietswaart T, van der Molen J & Wolfstein K. 2004. Achtergrondrapportage referenties en maatlatten fytoplankton. Rapportage van de expertgroep fytoplankton.

Vighi M. & Chiaudani G., 1985. A simple method to estimate lake phosphorus concentrations resulting from natural background loadings. *Water Research*, **19**, 987-991.

Willby N., Pitt J. and Phillips G., 2006. *Summary of approach used in LEAFPACS for defining ecological quality of rivers and lakes using macrophyte composition* Draft Report to Environment Agency, January 2006

Wunsam S & Schmidt R. 1995. A diatom-phosphorus transfer function for Alpine and pre-alpine lakes. *Memorie Ist. ital. Idrobiol.*, **53**: 85-99.

# APPENDICES

## Appendix I Recommended sampling, storage and analysis methods

| | |
|---|---|
| Sampling | Open water, integrated sample or bottle on rope thrown from edge – by outflow, pier<br>Volume of water – dependent on phytoplankton abundance – 1 litre generally sufficient<br>Regular sampling – recommended monthly during growth season (April to September) |
| Storage | Preservative – acidified lugol's for short-term storage only (less than 1 year). Recommend additional sample taken and preserved with formaldehyde for longer-term storage |

## Appendix II Guidance on Counting and Analysis of Freshwater Phytoplankton Samples

***Introduction***

The following guidance has been developed with reference to the CEN water quality guidance standard for the routine analysis of phytoplankton abundance and composition (CEN, 2004), standard operating procedures developed for Irish lakes (Donnelly, 2004) and lakes in Northern Ireland (Girvan, 2003), phytoplankton methods summarised in Wetzel and Likens (2000) and comments from Bill Brierly (Environment Agency).

Two accepted methods are described for counting phytoplankton:
Using sedimentation chambers on an inverted microscope (Utermöhl technique)
Using a Lund Chamber with a conventional compound microscope

CEN guidance (CEN, 2004) is focused on the use of sedimentation chambers with inverted microscopy, although much of the guidance is applicable to other counting methods. The use of Lund Chambers with conventional compound microscopes is detailed in Annex E of the CEN guidance document.

The CEN guidance does not explicitly state that one method is more suitable than another, although implicit in its focus on sedimentation chambers, is that these are more widely adopted and accepted. CEN guidance does describe a number of advantages and disadvantages for both methods. The main advantage of the sedimentation chamber is that their use may circumvent the need for an initial sample concentration step (unless algal densities are low), reducing the errors associated with concentrating samples. Sedimentation chambers are also closed during analysis and so evaporation does not occur during counting, whereas evaporation can occur from the open ends of the Lund Chamber during counting; the effect of this needs to be minimised by sufficiently quick counting or counting a smaller number of fields in several re-fills of the Lund Chamber. The other advantage of sedimentation chambers is that their should be no size bias with counts of both small and large algae, whereas very large colonial algae can be restricted entering Lund Chambers through the bore of pipette and/or the depth of the cell.

The advantage of using a Lund Chamber is that they can be viewed using objectives of up to x40 magnification on conventional compound microscopes, although these still require microscope objectives with long working distances. The other advantage of using a Lund

Chamber is that it is quick to set up with a settling time of 2-5 minutes, although a prior concentration step is usually required (CEN, 2004: Annex E.4). This compares favourably with 4-24 hours using sedimentation chambers varying between 1 to 6 cm in depth (CEN, 2004: Section 6.5). CEN (2004) also highlights that because of the shallow nature of a Lund Chamber, random distribution of phytoplankton cells is more likely than with the use of deeper sedimentation chambers. Clumping around the edge is more common in sedimentation chambers, making random counting methods less appropriate.
It is unlikely that many counters will be able to switch from one counting chamber to another as each requires a different microscope, so either chamber is considered acceptable.

***Sample preparation and Filling***
Sample preparation and filling of the chambers should follow CEN (2004). To promote random distribution of cells in counting chambers, it is very important that samples and chambers are first allowed to acclimatise to room temperature over a period of 12 hours or so. Just before filling chambers, samples should be well mixed through gentle shaking (rolling and turning upside down) for at least 1 minute.

***Sedimentation Chambers (Utermöhl technique)***
The description below is modified from Donnelly (2004), incorporating aspects of CEN (2004).

Using a wide-bore pipette, or pouring into a measuring cylinder if dilution is required, dispense the temperature-acclimatized, well-mixed sample into the sedimentation chamber.
Fill the chamber completely in one instance, then slide a thick cover glass over the top of the chamber to close it, making sure you avoid any air bubbles.
Allow contents to settle on a flat surface for at least 4 hours per cm height of chamber before examination. Usually 16 hours is sufficient (CEN, 2004).
After settling, place carefully on microscope stage.

***Lund Chambers***
The description below is modified from CEN (2004: Annex E.4).

Place a coverslip diagonally across the rim of the chamber
Using a Pasteur/large-bore pipette, without a pipette bulb, dispense very carefully by capillarity, the temperature-acclimatized, well-mixed sample into the chamber from one of the open ends. Slide the coverslip into position and fill completely ensuring no air bubbles are trapped. When full the coverslip should fit tightly, not slipping if pushed.
Take care not to overfill - as removing excess liquid with a pipette or tissue will interfere with cell distribution. Overfilled Lund cells are better emptied and re-filled.
Place chamber on the microscope stage and allow contents to settle for 5 minutes (with the illumination off to reduce evaporation).

***Counting Procedure***
Under low magnification (x4 or x10 objectives) check that the phytoplankton appear randomly distributed (e.g. large forms are not concentrated near edges of counting chamber), scan the chamber, and make a list of the dominant taxa – in particular noting large algae that may be rarer and require counting under low magnification (x10 objective)
Under high power (x40 or x50 objectives), check that the abundance of the phytoplankton cells in the field of view is neither too over-crowded or too sparse (if so, adjust sub-sample accordingly). Aim to have around 3-5 algal units per field of view

***Counts of all taxa at high magnification***

It is recommended that random fields of view are counted, counting a minimum of 100 fields of view. For sedimentation chambers, transect counting across the widest part of the chamber can be carried out if there is a suspicion of clumping around the edges, If clumping is very obvious, chambers should be re-filled.

Count all live cells, filaments or colonies of identified plankton per field until 100 fields of view have been observed, ideally >400 units (cells, filaments, or colonies) should be counted in total. If cells have lost >50% of their cell contents they should be considered dead and not counted. Note that for some diatom species, such as *Rhizosolenia*, cell contents can only take up a small proportion of the frustule and should not be considered dead (cf. Fig 10h, Cox, 1996).

For cells/colonies/filaments that cross the edges of the counting field apply a consistent rule as to whether it is included in the count (e.g. count cells crossing left and bottom boundaries only - see Fig. 2 in CEN (2004) guidance). Note guidance below on estimating biovolume of filaments or colonies that are not entirely in the counting field.

If counting using transects in sedimentation chambers, continue until a full transect has been completed and >400 units (cells, filaments, or colonies) have been counted. Several transects may be required and chambers can be turned between transects so new areas can be counted.

If 100 individual units (cells, filaments or colonies) of an individual taxon have been reached, then counting can be stopped for that specific taxon – but it is very important to note down the number of fields of view counted for this specific taxon (including fields of view where absent). If transect counting, full transects should always be completed before counting of an individual taxon is stopped – total transect length should be noted.

***Counts of larger taxa at low magnification***

Large species that are identifiable at lower magnification are often uncommon under high magnification, but can contribute proportionally more to total biovolume. For this reason, counts of large taxa should also be carried out at low magnification (e.g. x10 objective lens) to ensure sufficient numbers are observed. Whole chamber counts should be carried out.

***Biovolume measurements***

To estimate biovolumes, it is important to measure linear dimensions of at least three individual of all taxa observed in the sample. For taxa of more variable size, at least 10 individuals should be measured to estimate mean dimensions. For some species with external skeletons much larger than cell contents, e.g. *Dinobryon*, *Rhizosolenia*, etc. the dimensions of the organic cell contents should be measured, not the external skeleton dimensions.

For filamentous taxa, filament lengths should be measured for <u>all</u> filaments observed. Filament width/diameter is normally relatively fixed and generally only needs to be measured once. Only measure the filament length that is contained within the counting field. **Do not measure the whole filament length if it extends outside the counting field**

For colonial taxa count cell numbers and multiply by mean cell dimensions (often single measure of dimensions needed). If the colony is very large or cells are very small, mean cell numbers may have to be estimated. This is best done by estimating cell numbers in a more restricted area of the colony and estimating how many similar areas are contained within the counting field. Remember to take into account in estimates colony depth and hidden cells. **Do not measure the whole colony volume if it extends outside the counting field.**

Use representative formulae to estimate biovolume as illustrated in Wetzel and Likens (2000: Figure 10.9). Check biovolume estimate with published biovolumes in

spreadsheets provided. If biovolume estimates are very different with published literature for many species, check the calibration of your microscope.

***Recording data***

Make a log of all results on counting sheets (copies to be sent with results to Sian) and input into the standard spreadsheet provided [**Action: Laurence to circulate update**] to calculate cells/ml and biovolume/ml for individual taxa and main phyla

When counts are based on low magnification observations of the whole chamber enter the equivalent fields of view in the spreadsheet (= total area of chamber / area of field of view)

If some identifications were uncertain these should be flagged on the spreadsheet

Produce a summary count spreadsheet for inputing into the phytoplankton database by using an autofilter (Data/Filter) on the "count" column to remove rows where the count was zero (filter for "NonBlanks") and copy to the separate "Species list" worksheet

Send all completed spreadsheets to Sian and Laurence.

***Identification and Coding***

The standard flora for identification is the Freshwater Algal Flora of the British Isles (Whitton et al., 2003) and use of the blue-green and green algae CD-ROMs produced for the Environment Agency. Please record species codes noted in the Whitton et al. (2003) flora on the recording sheet – please also note old codes (in brackets) where appropriate as currently the phytoplankton database is using the old Whitton et al. (1998) codes.

Identification should be carried out to the highest possible taxonomic level, although for monitoring purposes the genus level is often sufficient for placing taxa within a functional group. The exceptions to this are *Peridinium*, *Staurastrum* and certain diatom (*Aulacoseira*, *Cyclotella*, *Fragilaria*, *Stephanodiscus*, *Synedra* and *Tabellaria*) and cyanobacteria (*Anabaena*, *Microcystis*, *Oscillatoria* and *Planktothrix*) genera and which, if possible, should be identified to species level. For centric diatoms this may only be possible if dead cells are visible, for which specialist floras are currently required (series by Krammer & Lange-Bertalot), although *Aulacoseira* species can sometimes be identified from 'live' material if characteristic spines are present (Cox, 1996)

Photosynthetic picoplankton are not distinguishable from non-photosynthetic prokaryotes in Lugol's-preserved samples (Brian Whitton, pers. comm.) and are not consistently counted across Europe (In Finland counted "when abundant" Liisa Lepisto, pers. comm.). It has yet to be decided what, if anything, can be concluded from picoplankton counts, but for the time being it is recommended that picoplankton <u>are</u> routinely counted and measured to include in the total biovolume estimates. Small unidentified cells of <2 µm size without a flagella or obvious cell structures should be recorded as prokaryotic picoplankton (code 90000002) and those with flagella are recorded as small unidentified flagellates or eukaryotic picoplankton (code 90000001)

If possible, for all unidentified taxa that are relatively abundant, digital/photographic images should be taken for circulation amongst counting teams and storage in an image database associated with the phytoplankton database.

***References***

CEN, 2004. *Water quality – Guidance standard for the routine analysis of phytoplankton abundance and composition using inverted microscopy (Utermöhl technique)* CEN TC 230/WG 2/TG 3/N83, May 2004.

Cox EJ, 1996. *Identification of Freshwater Diatoms from Live Material*. Chapman & Hall, London, 158 pp.

Donnelly K, 2004. *Standard operating procedure for the collection, preservation and analysis of lake phytoplankton*. Report to Irish EPA, May 2004

Girvan J, 2003, PhD Thesis, University of Belfast

John DM, Whitton BA and Brook AJ (Eds), 2003. *The Freshwater Algal Flora of the British Isles: an identification guide to freshwater and terrestrial algae*, Cambridge University Press

Wetzel RG and Likens GE, 2000. *Limnological Analysis*, 3rd edition

Whitton et al., 1998. *A Coded List of Freshwater Algae of the British Isles* Second Edition. Available from: http://www.ceh.ac.uk/data/dict/algae/

## Appendix III Phytoplankton database structure and quality checks

### Main Database Tables

Phytoplankton count data was collated together in a single Oracle database using a similar structure to the Environmental Change Network (ECN) database and using Whitton codes (Whitton et al., 1998). A Microsoft Access front-end has also been constructed with basic queries to deliver data summaries. The phytoplankton database includes accompanying lake typology data but no chemistry data. The main phytoplankton tables are:

D1FPP: Freshwater Phytoplankton: Species data
D2FPP: Freshwater Phytoplankton: Sample information
DQFPP: Freshwater Phytoplankton: Quality codes

These three tables relate to other data tables (Figure A3.1). These other tables include temporary tables set up by the phytoplankton project that are derived from other databases (e.g. site typology) and other tables used by ECN for quality checking purposes. Details of the data field required for the three main phytoplankton tables are provided in the ECN freshwater data transfer protocol document (Lane, 2000), a slightly modified version of which is reproduced here in Appendix 4.

### Data quality check procedures

The data are uploaded from standard Excel data entry spreadsheets provided by counters into temporary Oracle tables. The data are then screened for the following potential problems:

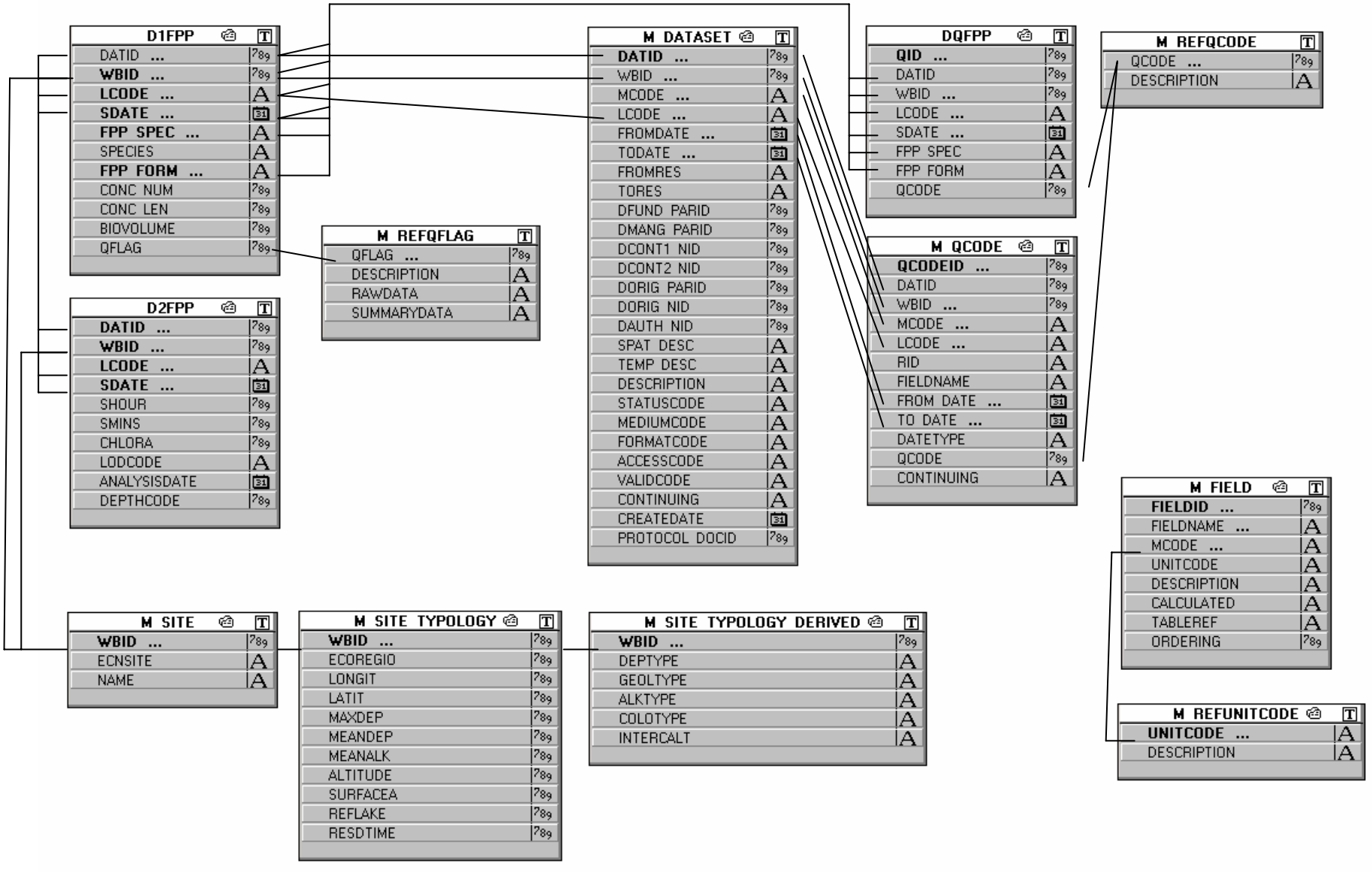
- mis-matches between sampling info and species info
- duplicate records
- errors in recording of sampling dates
- records without counts/concentrations
- records without forms
- records with forms where the form code doesn't exist in the reference list
- records without phytoplankton species codes
- records with phytoplankton species codes that don't exist in the definitive Whitton list
- records with phytoplankton codes that do exist in the list but the taxon description text doesn't match that in the Whitton list

A list of minimum and maximum sampling dates is also produced, by sampling location, this gives a quick output to check.

All errors are then looked into one by one and queried with the data originator if necessary. In the case of acceptable synonyms, these records are flagged and kept as they are. Most errors tend to be spelling mistakes/typos in the originator's taxon text. When fully cleaned, the data are loaded into permanent tables. A final check is run on the permanent tables to look for duplicates with data that have been loaded previously

Whitton et al., 1998. A Coded List of Freshwater Algae of the British Isles. Second Edition. Available from: http://www.ceh.ac.uk/data/dict/algae/

Figure A3.1 Diagram of data tables, field names and relationships in phytoplankton database



## Appendix IV ECN FRESHWATER DATA TRANSFER: Phytoplankton (FPP)

This document explains the standard format for transfer of data from freshwater phytoplankton sampling (ECN Protocol: FPP, Sykes et al, 1999) to the ECN Central Co-ordination Unit (CCU) at CEH Lancaster (see contact details at end).

### ECN Data Transfer Procedures

Please see first the document: "ECN Data Transfer: Guidelines for All Datasets" (Lane, 1999), for an overview of the types of information, formats and procedures required for the ECN Programme. Data should be sent via a single contact person responsible for handling data from all the organisation's or region's ECN sites, and it should be sent in ASCII text form - i.e. not in any internal software format.

### 2. FPP Datasets and Formats

Two datasets are associated with this protocol:

1. Sample information and chlorophyll-*a* results (rivers and lakes)
2. Phytoplankton species concentration (lakes only)

In addition, sites should send the following details with their first dataset, in the information message/file (see section 3.):

The Location Code and coordinates (GB National Grid or Irish Grid, as appropriate), of each ECN FPP sampling location to 10m resolution for rivers and as accurately as possible (normally 30-40m) for lake sites. The Location Code identifies different FPP sampling locations within a single ECN site; set it to 01 where a single sampling location only is used.
Details of any difference in methodology from that given by the FPP Protocol.

### SAMPLE INFORMATION

This dataset provides sampling dates, chlorophyll-*a* results and sample quality information. It is assumed that the same field sample is then sub-sampled for chlorophyll-*a* analysis and phytoplankton counts. Data for more than one site may be included in the same dataset.

### 2.1.1 Dataset Header - Samples

Each dataset should begin with 4 lines of header information, as follows:

ECN FPP - SAMPLES
Site Name(s) (separated with a '/')
Site Code(s), Location Code(s) (separate sites with a '/')
Names of people responsible for sample collection (in same order as site names/codes)

For example, for ECN Sites at Windermere and Esthwaite:
ECN FPP - SAMPLES
Windermere/Esthwaite
L04,01,02/L05,01
D.Smith, P.Brown

For the phytoplankton project the GBLakes.net WBID should be used instead of the ECN site code

**Data Records – Samples**

Data records should follow on immediately from the header information, on the next line. Each record should contain information for a single sample taken at a single location on a given collection date - this information should uniquely identify each sample. The record should be in free-format, comma-separated and consist of the following variables, in the order given below:

| Variable | Units | Format/Reporting Precision |
|---|---|---|
| 1 Site Code | 3-character code | e.g. L04 |
| 2 Location Code[1] | 2-digit code | e.g. 01 |
| 3 Sampling Date | Day-Month-Year | DD-MON-YYYY e.g. 02-JUN-2000 |
| 4 Sampling Time | GMT (24hr) | hh:mm e.g. 09:30 |
| 5 Chlorophyll *a* | μg/litre | precision: 3 sig. figs. |
| 6 Analysis Date | Day-Month-Year | DD-MON-YYYY |
| 7 Quality Code separator[2] | 1-character code | Q |

Quality fields follow on (see section 3. and Appendix I):
8 onwards:
Quality Codes 3-digit codes e.g. 227

Notes:
This identifies different FPP sampling locations within a single site. Set to 01 where a single sampling location only is used.

Please include the Q separator even if no Quality Codes have been recorded, as it provides a check for the database when dealing with a variable number of fields.

Example dataset:

ECN FPP - SAMPLES
Windermere/Esthwaite
L04,01,02/L05,01
D.Smith, P.Brown
L04,01,1-MAR-2001,09:30,3.01,2-MAR-2001,Q,000
L04,02,1-MAR-2000,12:20,3.55,2-MAR-2001,Q,000
L05,01,1-MAR-2001,10:00,2.99,2-MAR-2001,Q,000

Missing Data

Any missing data values should be indicated simply by including only the separating comma in the data record where the data value would be, i.e. setting the field to 'null'. For example, if the time of sampling had not been recorded for some reason, then the first data record in the example above should appear as:

L04,01,1-MAR-2001,,3.01,2-MAR-2001,Q,000

It is **most important** that these commas appear in the data record in the place where values would have been, otherwise it will not be clear which data field refers to which variable.

Please see the document: ECN Data Transfer: Guidelines for All Datasets (Lane, 1999) for more complete information on handling missing data.

## PHYTOPLANKTON SPECIES CONCENTRATION

This dataset provides information on concentration of phytoplankton by species. Data for more than one site may be included in the same dataset.

### Dataset Header - Species

Each dataset should begin with 4 lines of header information, as follows:

ECN FPP - SPECIES
Site Name(s) (separated with a '/')
Site Code(s), Location Code(s) (separate sites with a '/')
Names of people responsible for species identification (in same order as site names/codes)

For example, for the ECN Site at Windermere

ECN FPP - SPECIES
Windermere
L04,01,02
S.Owens, H.Kelly

**Data Records - Species**

Data records should follow on immediately from the header information, on the next line.

Each record should contain information for a single species sampled at a single location at a given collection date. The record should be in free-format, comma-separated, and consist of the following variables, in the order given below:

| Variable | | Units | Format/Reporting Precision |
|---|---|---|---|
| 1 | ECN Site Code | 3-character code | e.g. L04 |
| 2 | Location Code[1] | 2-digit code | e.g. 01 |
| 3 | Sampling Date | Day-Month-Year | DD-MON-YYYY |
| 4 | Species Code[2] | 8-digit code | e.g. 13080010 |
| 5 | Species name[3] | Genus species | e.g. Asterionella formosa |
| 6 | Species Type Code[4] | 2-character code | e.g. CE for Cells |
| 7 | Concentration[4] | individuals/ml | precision: 1 |
| | | or mm/ml if filamentous | precision: 1 |
| 8 | Quality Code separator[5] | 1-character code | Q |

Quality fields follow on (see section 3. and Appendix I):

| | | | |
|---|---|---|---|
| 9 | onwards: | | |
| | Quality Codes | 3-fig codes | e.g. 204 |

Notes:

The Location Code identifies different FPP sampling locations within a single site. Set to 01 where a single sampling location only is used.

ECN uses the Whitton, B A et al (1998) coded list of freshwater algae. A machine-readable version of this list is available from the NERC Land-Ocean Interaction Study Web site: http://www.nwl.ac.uk/~loissys/algal_coded_list.htm, or via the ECN CCU.

Species names: Ideally, data should be entered using software which allows codes to be entered and automatically checked against names. Until this becomes available, a taxonomic name as well as a code is necessary as a cross-check.

Phytoplankton species are categorised as:

CE – Cellular
CO – Collonial
FI – Filamentous

The Species Type Code determines which units of measurement are used for reporting concentration. Please include the Q separator even if no Quality Codes have been recorded, as it provides a check for the database when dealing with a variable number of fields.

Include records only for taxa present in the sample.

Example dataset:

```
ECN FPP - SPECIES
Windermere
L04,01,02
S.Owens, H.Kelly
L04,01,1-MAR-2001,13080100,Asterionella formosa,CO,255,Q,000
L04,01,1-MAR-2001,05100010,Rhodomonas lacustris,CE,2040,Q,000
L04,02,1-MAR-2001,05100010,Rhodomonas lacustris,CE,966,Q,000
```

**Missing Data**
Any missing data values should be indicated simply by including only the separating comma in the data record where the data value would be, i.e. setting the field to 'null'. For example, if for some reason the concentration value is missing, the first record above should appear

L04,01,1-MAR-2001,13080100,Asterionella formosa,CO,,Q,506

(where, for example, quality code 506 indicates a problem with the equipment)

It is **most important** that these commas appear in the data record in the place where values would have been, otherwise it will not be clear which data field refers to which variable.

Please see the document: ECN Data Transfer: Guidelines for All Datasets (Lane, 1999) for more complete information on handling missing data.

**Data Quality Codes and Comments**
Quality information may be provided as pre-defined codes attached to appropriate data records or, where necessary, as free-text. A list of Quality Codes most appropriate to this Protocol is given in Appendix I of this document. If there is no suitable Quality Code to select from this list, then free-text comments should be recorded and reported in the first e-mail message (or accompanying information file on disk). Please do **not** add your own code to the data, as it will not be recognised by the database on input. If the reported condition is repeated regularly, then a new Quality Code will be added to the database and a new list issued.

Note that Quality Code 000 should be added where no other codes apply, to differentiate from no Quality Codes recorded (e.g. forgotten to record, or lost information).

**References**
Lane A (1999) *ECN Data Transfer: Guidelines for All Datasets*. ECN internal document. (Also reproduced in Chapter 3, section 2 of the ECN Freshwater Protocols book - see below)

Sykes J M, Lane A M J and George D G (1999) *The United Kingdom Environmental Change Network: Protocols for Standard Measurements at Freshwater Sites*. Abbots Ripton: Institute of Terrestrial Ecology

Whitton B A, John D M, Johnson L R, Boulton P N G, Kelly M G and Haworth E Y (1998) *A Coded List of Freshwater Algae of the British Isles* LOIS Publication Number 222. Wallingford: Institute of Hydrology.

**Contact details:**
If you have any queries or difficulties with the data transfer format, please contact the ECN Data Manager:

Mandy Lane
ECN Data Manager
CEH Lancaster, Lancaster Environment Centre, Library Avenue, Bailrigg, Lancaster, LA1 4AP
Tel: 01524 595800, E-Mail: mlane@ceh.ac.uk

## APPENDIX I - Quality Codes most applicable to Phytoplankton Sampling

(see ECN Data Transfer Guidelines for complete list)

| | |
|---|---|
| 100 | No information available – data lost |
| 101 | No sampling made because of equipment failure or inability to reach site |
| 102 | Sample lost or inadvertently discarded |
| 103 | Partial loss of sample |
| 118 | Crop spraying in vicinity |
| 119 | Construction work in vicinity |
| 120 | Liming in vicinity |
| 121 | Change of land use in vicinity |
| 126 | River/lake frozen - no sample |
| 127 | River/lake dry – no sample |
| 134 | Significant disturbance at sampling site |
| 139 | Muck/slurry/slag application in vicinity |
| 140 | Application of chemicals in vicinity |
| 203 | No flow observed in river – standing water only |
| 204 | Material inadequately preserved |
| 205 | Supplementary samples taken |
| 206 | Unidentified material archived |
| 222 | Non-standard sampling date |
| 227 | Sample taken from lake outflow |
| 228 | Sample taken from jetty/dam |
| 229 | High river flow |
| 233 | Lake level high |
| 234 | Lake level low |
| 502 | Laboratory: Sample lost or inadvertently discarded |
| 503 | Laboratory: Partial loss of sample |
| 504 | Laboratory: Sample discarded because of contamination |
| 505 | Laboratory: Insufficient sample for measurement |
| 506 | Laboratory: Measurement not made because of equipment failure |
| 999 | Free-text information associated with this data record |
| 000 | No problems with sampling/analysis - no quality codes or text apply |