



Article (refereed) - postprint

Thackeray, Stephen J.; Noges, Peeter; Dunbar, Michael J.; Dudley, Bernard J.; Skjelbred, Birger; Morabito, Giuseppe; Carvalho, Laurence; Phillips, Geoff; Mischke, Ute; Catalan, Jordi; de Hoyos, Caridad; Laplace, Christophe; Austoni, Martini; Padedda, Bachisio M.; Maileht, Kairi; Pasztaleniec, Agnieszka; Järvinen, Marko; Lyche Solheim, Anne; Clarke, Ralph T. 2013.
**Quantifying uncertainties in biologically-based water quality assessment: a pan-European analysis of lake phytoplankton community metrics.**

**Highlights for Thackeray et al submission**

- Phytoplankton ecological quality metrics were calculated for 32 European lakes.
- We modelled sources of variability (within and among lakes) in these metrics.
- Metrics varied more among lakes, than within lakes or due to sampling variation.
- Metrics varied significantly with eutrophication and lake depth.
- Three metrics are considered robust for Water Framework Directive Intercalibration.

# Quantifying uncertainties in biologically-based water quality assessment: a pan-European analysis of lake phytoplankton community metrics

Stephen J. Thackeray[a], Peeter Nõges[b], Michael J. Dunbar[c], Bernard J. Dudley[d], Birger Skjelbred[e], Giuseppe Morabito[f], Laurence Carvalho[d], Geoff Phillips[g], Ute Mischke[h], Jordi Catalan[i], Caridad de Hoyos[j], Christophe Laplace[k], Martina Austoni[f], Bachisio M. Padedda[l], Kairi Maileht[b], Agnieszka Pasztaleniec[m], Marko Järvinen[n], Anne Lyche Solheim[e] & Ralph T. Clarke[o].

[a]*Centre for Ecology & Hydrology, Lancaster Environment Centre, Library Avenue, Bailrigg, Lancaster, LA1 4AP, UK. Tel: 00 44 (0)1524 595852, Fax: 00 44 (0)1524 61536, Email: sjtr@ceh.ac.uk*

[b]*Centre for Limnology, Institute of Agricultural and Environmental Sciences, Estonian University of Life Sciences, 61117 Rannu, Tartumaa, Estonia.*

[c]*Centre for Ecology & Hydrology, Maclean Building, Benson Lane, Crowmarsh Gifford, Wallingford, Oxfordshire, OX10 8BB, UK.*

[d]*Centre for Ecology & Hydrology, Bush Estate, Penicuik, Midlothian, EH26 0QB, UK.*

[e]*Norsk Institutt for Vannforskning, Gaustadalléen 21, NO-0349 Oslo, Norway.*

[f]*CNR Institute for Ecosystems Study, Largo V. Tonolli 50, 28922 Verbania Pallanza, Italy.*

[g]*Environment Agency, Kings Meadow House, Kings Meadow Road, Reading, RG1 8DQ, UK.*

[h]*Leibniz Institute of Freshwater Ecology and Inland Fisheries, Justus-von-Liebig-Straße 7, 12489, Berlin, Germany.*

[i]*Center for Advanced Studies of Blanes (CEAB-CSIC), Accés Cala St. Francesc 14, Blanes 17300, Spain.*

[j]*Centro de Estudios Hidrográficos del CEDEX, PO Bajo de la Virgen del Puerto, 3 28005, Madrid, Spain.*

[k]*Research Institute for Agricultural and Environmental Engineering, CEMAGREF, av de Verdun 50, 33612 Cestas-Gazinet, France.*

[l]*University of Sassari, Department of Sciences for Nature and Territory, Località Piandanna, 07100 Sassari, Italy.*

[m]*Institute of Environmental Protection-National Research Institute, 01-692 Warszawa, Kolektorska 4, Poland.*

[n]*Finnish Environment Institute (SYKE), The Jyväskylä Office, Survontie 9, FI-40500 Jyväskylä, Finland.*

[o]*Conservation Ecology and Environmental Sciences (CEES), School of Applied Sciences, Bournemouth University, Talbot Campus, Fern Barrow, Poole, Dorset, BH12 5BB, UK.*

## Abstract

Lake phytoplankton are adopted world-wide as a sensitive indicator of water quality. European environmental legislation, the EU Water Framework Directive (WFD), formalises this, requiring the use of phytoplankton to assess the ecological status of lakes and coastal waters. Here we provide a rigorous assessment of a number of proposed phytoplankton metrics for assessing the ecological quality of European lakes, specifically in response to nutrient enrichment, or eutrophication, the most widespread pressure affecting lakes. To be useful indicators, metrics must have a small measurement error relative to the eutrophication signal we want them to represent among lakes of different nutrient status. An understanding of variability in metric scores among different locations around a lake, or due to sampling and analytical variability can also identify how best this measurement error is minimised.

To quantify metric variability, we analyse data from a multi-scale field campaign of 32 European lakes, resolving the extent to which seven phytoplankton metrics (including chlorophyll *a*, the most widely used metric of lake quality) vary among lakes, among sampling locations within a lake and through sample replication and processing. We also relate these metrics to environmental variables, including total phosphorus concentration as an indicator of eutrophication.

For all seven metrics, 65 - 96% of the variance in metric scores was among lakes, much higher than variability occurring due to sampling/sample processing. Using multi-model inference, there was strong support for relationships between among-lake variation in three metrics and differences in total phosphorus concentrations. Three of the metrics were also related to mean lake depth. Variability among locations within a lake was minimal (<4%), with sub-samples and analysts accounting for much of the within-lake metric variance. This indicates that a single sampling location is representative and suggests that

sub-sample replication and standardisation of analyst procedures should result in increased precision of ecological assessments based upon these metrics.

For three phytoplankton metrics being used in the WFD: chlorophyll *a* concentration, the Phytoplankton Trophic Index (PTI) and cyanobacterial biovolume, > 85% of the variance in metric scores was among-lakes and total phosphorus concentration was well supported as a predictor of this variation. Based upon this study, we can recommend that these three proposed metrics can be considered sufficiently robust for the ecological status assessment of European lakes in WFD monitoring schemes.

**Keywords:** cyanobacteria, ecological quality assessment, eutrophication, linear mixed effects models, multi-model inference, Water Framework Directive

## 1. Introduction

The Water Framework Directive [WFD; (EC, 2000)] has revolutionised the assessment of anthropogenic impacts upon fresh- and coastal-transitional waters of the member states of the European Union. The central tenet of the Directive is that the assessment of human impacts on the surface water environment, rather than being based solely upon chemical parameters, should be based upon the attributes of key communities (Biological Quality Elements, BQEs). In turn, these BQEs should be sensitive to environmental pressures such as eutrophication and physical habitat modification.

For lakes, the phytoplankton has been identified as a key BQE to be used in ecological status assessment (Carvalho et al., 2012) and is already widely used as an important early-warning indicator of water quality changes. This is because of rapid replication rates (ensuring rapid responses to environmental stressors), direct sensitivity to physical and chemical environmental factors, and high diversity with species and/or functional types showing markedly variable responses to changes in the surrounding environment (Murphy et al., 2002; Reynolds, 2006). Furthermore, sampling of these communities is simple and inexpensive, with minimal impacts on co-existing biota. As a result of these features, phytoplankton was included in the WFD monitoring scheme as a relevant quality element for all surface water categories. As parameters to be studied, the WFD prescribes phytoplankton abundance, composition, and the frequency and intensity of blooms. While phytoplankton community composition and diversity are regulated by a complex interplay of intrinsic and extrinsic drivers such as climate, resource availability, patterns of competition and predation, and dispersal (Reynolds, 2006) they may also act as sensitive indicators of environmental pressures such as eutrophication as a result of increased nutrient loading (Kümmerlin, 1998; Padisák and Reynolds, 1998). Phytoplankton abundance, composition and the frequency/intensity of blooms are all considered to undergo changes along this pressure

97 gradient (Carvalho et al., 2006; 2012). The WFD explicitly requires robust quantitative high-
98 level indicators, or metrics, of the phytoplankton community which can be used to monitor
99 the status of freshwater communities in the face of anthropogenic pressures, and identify
100 improvements to ecological status as a result of management interventions. As part of the EU
101 project WISER (http://www.wiser.eu/) a number of existing, or newly developed, metrics
102 have been considered for this purpose (Mischke et al., 2010; Phillips et al., 2010).

103 However, there is a WFD requirement to assess the uncertainty in ecological status
104 assessments when using such metrics (Hering et al., 2010). Phytoplankton communities show
105 marked spatial heterogeneity within lakes, over a range of spatial scales, as a result of
106 patterns in lake circulation and mixing, and spatial gradients in flushing, grazing and nutrient
107 availability (Pinel-Alloul and Ghadouani, 2007). In addition, variation in phytoplankton
108 metrics may occur due to differences in the analysts processing samples and sub-sampling
109 procedures (Vuorio et al., 2007). Therefore, it is highly likely that the choice of sampling
110 location within a lake and sample processing will affect the values of metrics based upon
111 phytoplankton community data. Where metric values fall close to ecological status class
112 boundaries, then these variations may fundamentally influence the overall assessment of a
113 waterbody (Clarke et al., 2006b; Clarke, 2012). This has led to suggestions that results of
114 ecological status classification should be given in terms of probabilities (Hering et al., 2010).
115 Analyses of riverine macroinvertebrate community metrics have shown that the level of
116 metric variability due to sampling may itself change with the ecological quality of a site
117 (Clarke et al., 2002; Clarke et al., 2006a). If the candidate phytoplankton metrics are to be
118 used to distinguish between lakes of differing ecological quality, then among-lake variations
119 in metric scores must be maximised and variation due to sampling/sample-processing
120 minimised. This would give the best chance for the former to be related to differences in the
121 intensity of key ecological pressures acting upon those lakes. It is also important to know

whether these metrics become inherently more or less variable (uncertain) along this pressure gradient.

Until now, there has not been a formal assessment of the multiple sources of uncertainty that are inherent in phytoplankton metrics, even for widely adopted metrics, such as chlorophyll *a*. The statistical tools to make this assessment exist (Carvalho et al., 2006; Clarke and Hering, 2006; Clarke, 2012) but there has been a need for new data, collected according to a sampling design that allows distinction of different and independent sources of variability in metric scores. Knowledge of the relative importance of different sources of metric variability will guide the design of sampling campaigns aimed at ecological quality assessment. For example if a large component to the total variance in a metric is associated with sub-sampling of field samples, then the precision of assessments based upon this metric could be improved by analysing a larger number of sub-samples to derive a more representative average metric score for the lake. Herein, we present the results of a novel analysis of seven established phytoplankton community metrics based on a pan-European field sampling campaign of 32 lakes. Rigorous standardisation of sampling and sample processing procedures, along with a hierarchical sampling design targeted at uncertainty estimation, allow an entirely consistent analysis of sources of variation in phytoplankton metrics within and between European lakes. Specific objectives address the following questions; do candidate phytoplankton community metrics:

Q1: show greater variability among lakes than within lakes or as a result of differences in sample processing?

Q2: differ significantly along a gradient in lake nutrient status, after accounting for within-lake and sample-processing variation?

Q3: show systematic changes in their level of variability along gradients in physical, chemical and geographic attributes of lakes?

## 2. Materials and methods

### *2.1 Field survey*

The analysis is based upon water samples collected from 32 lakes in eleven European countries during the spring and summer of 2009 (Table 1). These collectively represent lake types found within Member States and Norway comprising the Alpine, Northern, Central/Baltic and Mediterranean Geographical Intercalibration Groups [GIGs (WISE 2008)]. All lakes were less than 10 $km^2$ in surface area, but varied widely in mean depth (3.5 - 34 m) and altitude (15 – 970 m a.s.l.). The lakes also differed markedly in productivity/trophic status, with wide variation in alkalinity (0.06 – 4.40 meq $L^{-1}$) and total phosphorus concentration (4 - 151 mg $m^{-3}$) at the time of sampling.

Each lake was sampled according to the same standardised protocol. The sampling design allowed the total variability in phytoplankton community structure, as indicated by a range of metrics, to be decomposed into a series of independent variance components, each indicating a potential source of uncertainty. The sampling design was as follows (Fig. 1):

(i) Within each lake, water samples were collected at three stations. These were above the deepest point of the open water zone, and at points representing the mean depth of the lake and a depth intermediate to the mean and maximum depths. This allowed quantification of within-lake spatial heterogeneity in phytoplankton community composition and metric scores, at the basin scale.

(ii) Two water samples were collected at each of the three stations. This allowed quantification of errors associated with repeated sampling at a specific location, as a result of smaller-scale heterogeneity in the phytoplankton community.

(iii) Each sample was sub-sampled in order to quantify variations in phytoplankton metric scores due to sub-sampling errors and differences in the analyst identifying and enumerating phytoplankton in the sub-samples. For analyses of phytoplankton composition, three sub-samples were collected from the first sample. Two of these were processed by the same analyst (revealing sub-sampling error), while the third was processed by a different analyst (to evaluate variability in metric scores due to differences in the approach used by different analysts). This is similar to the sampling design used by Clarke et al. (2002) to separate field replicate sampling variation from operator effects for river macroinvertebrate community metrics. From the second sample, only one sub-sample was collected, to allow comparison with metric scores derived from the first sample. Prior to microscopic examination an aliquot (sub- sub-sample) of each sub-sample was collected and put into a sedimentation chamber. Any variation associated with this sub-sub sampling is of course confounded with sub-sample variation in what follows, as no replication is available at this level of the hierarchy. For chlorophyll *a* (Chl-*a*) analysis, which followed a rigorously standardised spectrophotometric protocol, the effect of the analyst was not addressed and only two sub-samples were taken from the first sample to evaluate the sub-sampling error.

For reasons of cost the hierarchical sampling design was unbalanced at the within-station level: it was not feasible for both analysts to assess every replicate sub-sample of every sample at every station. However, by using appropriate statistical modelling approaches (see section 2.5) it was possible to use this design to identify elements of field sampling

campaigns that, through greater replication or standardisation, could be modified in order to improve the precision of ecological status assessments. For example, would the precision of such assessments be improved if we collected more samples, samples from more stations throughout the lake, processed more sub-samples or standardised taxonomic skills among analysts?

At each station, water samples were collected using an integrated tube sampler. If a lake was thermally stratified samples were taken from the euphotic layer (estimated as 2.5 x Secchi depth). When the water column was mixed samples were collected from throughout the whole water column, down to 0.5m above the sediment surface. Sub-samples were collected from each sample after thorough mixing. If immediate extraction of Chl-*a* samples was not possible, they were stored in a refrigerator or ice box for as short a time as possible. Samples for microscopic analysis were preserved using a solution of Lugol's iodine (final concentration approximately 0.5% by volume) and stored in the dark.

A further separate water sample was collected at the deepest point of each lake and analysed for alkalinity and concentrations of total phosphorus (TP). TP was measured following sulphuric acid-potassium persulphate digestion of unfiltered samples, according to Murphy and Reilly (1962). For some lakes multiple determinations of each variable were made and these were averaged prior to statistical analyses. Whilst data on total phosphorus concentrations were available for all lakes, alkalinity values were missing for some lakes and so representative values were necessarily derived from data collected under a parallel hierarchical macrophyte survey (Dudley et al., 2010). Secchi depth was also recorded at the deepest point of each lake.

In the following analyses TP concentrations were used to indicate where the sampled lakes fell on a gradient of nutrient enrichment. Latitude, longitude and altitude of each lake were also included, as proxies for broad climatic gradients that might impact upon phytoplankton communities via effects on lake physical processes. Alkalinity and mean lake depth were included in the study as they are the primary determinants of the fundamental lake "types" described in the WFD. Different combinations of high-low alkalinity and mean depth have been used to categorise these lake "types". This captures the fact that lakes show natural variability in their phytoplankton communities, due to their catchment setting and morphometry, irrespective of differences in nutrient enrichment (Pinel-Alloul et al., 1990).

### *2.2 Sample processing for Chl-a analysis*

A fixed volume of water, dependent on the amount and type of seston present in each lake, was filtered through 47-mm GF/F filters and the filter was placed into 10 ml of 96% ethanol for pigment extraction at 4 °C for 24 hours. Analysis then followed the International Standard method ISO10260 (1992).

### *2.3 Sample processing for microscopic examination of phytoplankton*

Microscopic examination of phytoplankton followed the same standardised protocol across Member States, and was based upon procedures outlined in CEN 15204 (2006), National Rivers Authority (1995) and Brierley et al. (2007). Briefly, samples were examined in sedimentation chambers with an inverted microscope, according to the Utermöhl technique (Utermöhl, 1958). For each sample, a low magnification (40x or 100x) whole chamber count, two intermediate magnification (200x or 250x) transect counts and 50-100 field of view counts at high magnification (400x or greater) were completed. Phytoplankton taxa were identified to the highest possible level. Counts of each taxon were converted to biovolumes

by measuring cell/colony dimensions and approximating each taxon to a simple geometric shape (Brierley et al., 2007). Phytoplankton cells were measured using eye-piece graticules, after calibration with a stage micrometer. All subsequent phytoplankton metric calculations were based upon the biovolume data.

### *2.4 Phytoplankton metrics*

Seven candidate phytoplankton metrics are considered herein, a brief description of which is given below. Full details on each metric are provided in Phillips et al. (2010) and Mischke et al. (2010). These metrics have been categorised according to whether they relate to phytoplankton abundance or composition, or to features of blooms.

1. Chl-*a* concentration (Abundance metric, in mg $m^{-3}$) is a measure of phytoplankton abundance, commonly used to represent the ecological status of a lake with respect to eutrophication pressures.
2. Phytoplankton Trophic Index (PTI, Composition metric). This has been developed, using an independent data set, from the "trophic scores" of phytoplankton taxa along a eutrophication gradient (Phillips et al., 2010). After a Canonical Correspondence Analysis (CCA) constrained by total phosphorus, taxa optima on the first ordination axis were derived indicating the TP concentration for the mean occurrence of each taxon. For each sub-sample, PTI was calculated as the weighted average of these taxa optima, where the weighing factor is the proportional biovolume of each taxon. The PTI increases with increasing lake trophic state.
3. Size Phytoplankton Index (SPI, Composition metric). The phytoplankton taxa within a sub-sample are grouped into a series of size categories, each one encompassing a doubling of cell biovolume e.g. $\leq 0.5 \mu m^3$, 0.5-1.0 $\mu m^3$, 1.0-2.0 $\mu m^3$, 2.0-4.0 $\mu m^3$ etc

(Kamenir and Morabito, 2009). The SPI is then calculated as a function of the size categories and "trophic scores"/"indicator values" for those categories (Phillips et al., 2010). Trophic scores indicate the position of a size class along the trophic spectrum and indicator values estimate the "power" of each size class as a biotic indicator. The SPI tends to increase with increasing lake trophic state, due to a shift towards increased dominance of larger, rather than smaller, phytoplankton (Phillips et al., 2010).

4. Morpho-Functional Group Index (MFGI, Composition metric). The phytoplankton taxa within a sub-sample are grouped into a series of categories ("Morpho-Functional Groups") based upon their morphological attributes e.g. presence/absence of flagella, colonial or unicellular, large or small size (Salmaso and Padisak, 2007). The MFGI is then calculated as a function of the Morpho-Functional Groups and the "trophic scores"/"indicator values" for those groups (Phillips et al., 2010). The MFGI tends to increase with increasing lake trophic state, due to an increase in the dominance of colonial cyanobacteria, large diatoms/chlorophytes/conjugatophytes, and unicellular/colonial chlorococcales (Phillips et al., 2010).
5. Functional Traits Index (FTI, Composition metric). This is the arithmetic mean of the SPI and MFGI, and thus combines information on both the size spectrum and morpho-functional traits of the phytoplankton community. Phillips et al. (2010) recommend the use of the FTI for water quality assessment.
6. Evenness metric (Bloom metric). This is Pielou's evenness index, which expresses the ratio between the Shannon diversity of a sub-sample and the maximum possible value of the Shannon diversity index (Pielou, 1969, 1975). Evenness has been shown to decline under bloom conditions in more productive lakes, due to an increase in the

dominance of a small number of tolerant species with high growth rates (Mischke et al., 2010).

7. Cyanobacterial abundance (Bloom metric). This is the total cyanobacterial biovolume ($mm^{-3}$ $L^{-1}$) within a sub-sample, and is expected to increase with increasing lake trophic status (Mischke et al., 2010).

### *2.5 Statistical modelling*

*Q1: Do metrics show greater variability among lakes than within lakes or as a result of differences in sample processing?*

These analyses aimed to resolve whether metrics had the potential to be sensitive to variations in the intensity of environmental pressures acting at the lake level. This potential was to be estimated by the relative size of the among-lake variance in metric values and the within-lake variance components). Furthermore, we aimed to identify aspects of sampling campaigns that might be modified to improve the precision of ecological status assessments (by comparison of components of within-lake metric variance). A nested random effects statistical model structure was used to emulate the hierarchical nature of the sampling campaign. In this structure, lake was nested within country, sampling station within lake, sample within station, and sub-sample within sample was modelled implicitly as the lowest level "residual" variability. Each analyst could not process sub-samples from all samples or all stations or all lakes, even though some analysts processed samples from more than one country. Therefore the model factor 'Analyst' was included (except for analyses of Chl-*a* concentration) as a random effect which was, in mixed model technical terms, partially crossed with the other factors and variables. However, it was still possible for the mixed model functions in R to estimate the separate variance components. These variance

components are (as usual in most mixed models) estimates of the average size of that source of variance averaged over the other factors; it was not feasible to investigate interactions in factor variance components. Our variance estimates provide the best available information on the relative typical (i.e. average) sizes of the different sources of metric total and within-lake variance. More formally, the model structure can be denoted:

$$m_{austlc} = \beta_0 + v_{ustlc} + v_{stlc} + v_{tlc} + v_{lc} + v_c + v_a + e_{austlc} \quad (1)$$

where $m_{austlc}$ is the value of the metric $m$ for analyst a, for sub-sample u, in sample s, in station t, in lake l, in country c. Thus, $m_{austlc}$ is the sum of a series of components that each contribute to the total metric variation about an overall mean $\beta_0$. The components of metric variation are modelled as independent, normally distributed, variance components for analyst ($\sigma^2_a$=Var($v_a$)), sub-sample ($\sigma^2_u$=Var($v_{ustlc}$)), sample ($\sigma^2_s$=Var($v_{stlc}$)), station ($\sigma^2_t$=Var($v_{tlc}$)), lake ($\sigma^2_l$=Var($v_{lc}$)) and country ($\sigma^2_c$=Var($v_c$)).

Sub-sampling variance, being the lowest level in the hierarchical sampling, is estimated implicitly by the fitted model residual variance. Having fitted random effects model equation 1 to our data, the relative sizes of the estimated variance components were used to determine the levels of the sampling hierarchy at which each metric's values showed the greatest variability. In particular, the total variance among all lakes is $\sigma^2_A = \sigma^2_c + \sigma^2_l$, the average total variance within lakes is $\sigma^2_W = \sigma^2_t + \sigma^2_s + \sigma^2_u + \sigma^2_a$ and therefore the total variance in all metric values is $\sigma^2_T = \sigma^2_A + \sigma^2_W$. The percentage of the total metric variance ($\sigma^2_T$) occurring at each level in the sampling hierarchy was calculated from these variance parameter estimates (e.g. percentage among lakes = 100 $\sigma^2_A$ /$\sigma^2_T$). The hierarchical and crossed random effect models of equation 1 were all fitted to the unbalanced datasets using the standard Restricted Maximum Likelihood (REML) method of model fitting in order to give unbiased estimates of the random effects. Whenever subsequent truly mixed effects models with different fixed effects structures (i.e. different combinations of predictors) were compared, models were re-

fit using the Maximum Likelihood (ML) method of model fitting (Crawley, 2007). Unlike many traditional ANOVA techniques, REML fitting of models with fixed and random (i.e. variance component) hierarchical and/or crossed factors can cope with unbalanced datasets with unequal replication at some levels, providing the sampling design gives some subsets of information within the data which enable the REML algorithm to distinguish and estimate each variance component (Crawley, 2007; Clarke, 2012). This is the case for our lake sampling design.

*Q2: Do metrics differ significantly along a gradient in lake nutrient status, once accounting for within-lake and sample-processing variation?*

We investigated whether relationships between phytoplankton metrics and measured morphometric, chemical and geographical features of lakes could be detected against the "background" of methodological variation resolved in stage 1 of the analysis. It is convenient here to refer to the pure random effects models as the "null model" in terms of having no environmental predictor variables. These pure random effect null models were augmented to include the measured environmental variables (TP, alkalinity, mean lake depth, latitude, longitude and altitude) as fixed effects and fitted as linear mixed effects models. Secchi depth was omitted since the direction of causality between this variable and the phytoplankton community is equivocal. In order to explicitly take into account uncertainty and parameter bias due to model selection, arising since both model formulation and parameters are estimated from the sample data, we used multi-model inference (Burnham and Anderson, 2002). For each metric, a "global" linear mixed effects model was constructed containing the same within-lake random effects structure and all the predictor variables (alkalinity, latitude, longitude, altitude, mean depth and TP). These environmental predictor variables have single

values for each lake and therefore can only explain aspects of the null model total among lake variance. Models were then run including all possible subsets of these variables, and ranked by the Akaike Information Criterion (AIC). A subset of top models, receiving progressively lower levels of statistical likelihood support from the data, was determined by finding the model with the most optimal combination of environmental predictor variables (i.e. lowest AIC value) and other candidate models with AIC values differing from this "top" model by ≤ 4 (Burnham and Anderson, 2002; Zuur et al., 2009). Model-averaged parameters (with 95% confidence intervals) were calculated using the parameter estimates in models within this top model subset. Maximum likelihood (ML) estimation was used when fitting models with different combinations of predictor variables.

To estimate the proportion ($Prop_e$) of the total among-lake variation in metric scores that could be "explained" by the selected environmental variables we compared the residual among-lake metric variance ($\sigma^2_{l,fitted}$) estimated by the model with the most optimal combination of environmental predictors (i.e. lowest AIC value), with the total among-lake variance ($\sigma^2_{l,null}$) estimated in the corresponding null model (i.e. with no environmental predictors) thus:

$$Prop_e = 1-(\sigma^2_{l,\,fitted}/\,\sigma^2_{l,\,null}) \qquad (2)$$

$\sigma^2_{l,\,fitted}$ therefore represents the among lake variation in a metric that cannot be explained by the predictor variables in the top fitted model, while $\sigma^2_{l,null}$ represents the total among-lake variation in that metric. This approach is conceptually similar to that employed by Clarke et al. (2006b) to compare variance components of invertebrate metric scores gathered from hierarchical sampling designs. Since $\sigma^2_{l,\,fitted}$ and $\sigma^2_{l,null}$ are themselves estimated parameters, and therefore each have a level of uncertainty associated with them, $Prop_e$ must also be

394 considered an estimate with a level of uncertainty. Herein, we do not calculate the uncertainty
395 associated with the estimate of $Prop_e$ and merely use the values as broadly indicative of the
396 explanatory power of the selected predictor variables.
397
398 During the model fitting exercise, it was necessary to simplify the random effects structure to
399 retain only crossed effects of "Lake" and "Analyst". Preliminary analyses revealed that the
400 inclusion of the full random effects hierarchy when comparing models with different fixed
401 effect structures resulted in convergence errors, due to high levels of model complexity.
402 Furthermore, fitting of null models (see results) demonstrated that the omitted random effects
403 consistently accounted for little of the total metric variance.
404
405 *Q3: Do metrics show systematic changes in their level of variability along gradients in*
406 *physical, chemical and geographic attributes of lakes?*
407
408 As a final step in the analysis, we examined whether metric scores became more or less
409 variable as a function of between-lake changes in predictor variables, such as TP
410 concentration or mean depth. If metric variability is not constant across lakes with different
411 environmental attributes, then this could mean that sampling campaign design (in terms of
412 sample replication, level of standardisation) might also need to vary between lakes. This was
413 done by adding additional variance structures to previously fitted models that allowed for
414 changes in residual metric variability as a function of the measured environmental predictors.
415 For each metric, we worked with the model with the most optimal combination of
416 environmental predictor variables (lowest AIC) and added these extra variance structures
417 based upon each of the predictors within this top model. These structures took the form (Zuur
418 et al., 2009):

$$\mathrm{var}(\varepsilon) = \sigma^2 e^{2\delta x} \quad (3)$$

so that the residual variance [var(ε)] was allowed to vary as an exponential function of explanatory variable $x$ and the estimated parameter δ. For each metric, we compared the top fitted model with none of these additional variance structures, with models including structures that allowed for residual "spreading" with respect to each of the explanatory variables present in the top model. So, for example, if the top model for a particular metric included predictors $x_1$ and $x_2$, we compared models i) without structures to capture spreading of residual metric variation, ii) with residual spreading as a function of $x_1$, iii) with residual spreading as a function of $x_2$ and, iv) with residual spreading as a function of $x_1$ and $x_2$. The most optimal solution was found by comparing the AIC values of each of these models, after fitting using REML estimation.

All analyses were conducted using the *base*, *gplots*, *lme4*, *MuMIn* and *nlme* packages of R version 2.13.1 (Pinheiro et al., 2010; Warnes, 2010; Barton, 2011; Bates et al., 2011; R Development Core Team, 2011) and the Variance Estimation and Precision (VEPAC) package of STATISTICA 8.0 (StatSoft. Inc. 1984-2007).

## 3. Results

### *3.1 Sources of metric variability*

Exploratory analyses of the metrics data revealed that Chl-*a* and total cyanobacterial biovolume were positively skewed and so, prior to statistical modelling, we $\log_{10}(x+0.1)$ transformed these metrics in order to reduce the potential influence of the minority of

444 relatively high values in the dataset. Results from null models of all seven metrics (Table 2)
445 suggest that the majority of metric variance occurred between lakes. The Country ($\sigma^2_c$) and
446 Lake ($\sigma^2_l$) random effects together accounted for between 65% and 96% of the total metric
447 variance, with the majority of this variability found among lakes rather than among
448 Countries. This suggested that metric scores varied more among lakes (which were
449 distributed along a pressure gradient) than within lakes. It is noteworthy that the Analyst ($\sigma^2_a$)
450 and Error (sub-sample level, $\sigma^2_u$) variance components were the major contributors to the
451 within-lake component. Therefore, metric variation due to analyst differences and sub-
452 sampling exceeded variation due to within-lake spatial heterogeneity in the phytoplankton.
453

454 ***3.2 Relationships between metrics and lake characteristics***

455 The seven metrics varied widely in their relationship to total phosphorus concentration;
456 highlighting different strengths of the metrics for indicating the primary among-lake pressure
457 gradient of nutrient enrichment (Fig. 2). Visual inspection of the data suggested that metric-
458 phosphorus relationships were strongest for the abundance metric Chl-*a*, PTI composition
459 metric and total cyanobacterial biovolume bloom metric. This was confirmed by the structure
460 of the most optimal models for these metrics, which included fixed effects of total
461 phosphorus concentration and mean lake depth (Table 3). Delta AIC values for these models,
462 all ≥13.5, indicated a significant improvement in model fit compared to (null) models with no
463 predictors. Therefore a detectable increase in all three of these metrics was observed in lakes
464 with higher phosphorus concentrations, and in shallower lakes. This was observed despite
465 methodological uncertainty arising due to sampling and sample processing. Top models for
466 the three remaining composition metrics (MFGI, SPI and FTI) suggested that all three metrics
467 were higher in shallow lakes and in lakes at higher altitudes. While ΔAIC values ≥9 indicated

that top models were considerably better supported than null models for MFGI and FTI, this was not the case for SPI (ΔAIC =2). Similarly the top model for the evenness metric, suggestive of a reduction in this bloom metric with increasing phosphorus concentration and at low alkalinity, represented only a modest improvement on a model with no fitted predictor variables (ΔAIC = 2.3). The majority of the among-lake variance in Chl-*a* concentration was accounted for by the fitted predictors in the top model, as indicated by $Prop_e$ (Table 3, Fig. 3). For total cyanobacteria and the PTI metric, the amount of among-lake variance "explained" by the fitted predictors in the top model was less, at 43-47%, while for the remaining metrics <40% of the among lake metric variance was accounted for in the fitted models.

However, relatively low Akaike weights for the top models for all metrics (0.06-0.19, Table 3) suggested that the top models did not receive overwhelming support within each model set and that, for each metric, other candidate models collectively received likelihood support from the data. We used a multi-model inference approach to calculate model averaged parameters for the relationships between each metric and the selected environment predictors. This confirmed strong support for an increase in Chl-*a* concentration, PTI and total cyanobacterial biovolume at high phosphorus concentrations, despite methodological metric variation (positive slope parameters, Figs. 4-6). Across many of the metrics there was a support for an effect of mean lake depth on metric scores. With the exception of evenness, all metrics decreased with an increase in mean lake depth i.e. a negative slope parameter for their relationship (Figs. 4-6). For MFGI, FTI and total cyanobacterial biovolume there was strong support for this effect, while for the remaining metrics support for this effect was relatively weaker. With the exception of Chl-*a* concentration there was also consistent, though weak, support for an effect of altitude on metric scores. Tables summarising the

492 model sets used to derive these averaged parameters for each metric can be found in the
493 Supplementary Information.

494

495 ***3.3 Changes in metric variability as a function of among-lake variations in physical,***
496 ***chemical and geographical attributes***

497 For all but one of the metrics (FTI) the fit of the most optimal statistical model (from Table 3)
498 was improved by allowing residual metric values to vary as a function of certain explanatory
499 variables (phosphorus concentration, lake depth, Table 4). In general, this supported the idea
500 that metric scores were more variable in some limnological contexts than in others. In the
501 case of SPI and MFGI the difference in AIC between models including and excluding these
502 structures (5.7 and 2.7 respectively) was much lower than for Chl-*a* concentration, PTI,
503 evenness and total cyanobacteria biovolume (20.9 - 44.8). While residual Chl-*a*
504 concentrations and evenness appeared to become more variable at lower phosphorus
505 concentration (negative δ estimates), cyanobacterial biovolume showed the reverse pattern;
506 with residuals being more variable at higher phosphorus concentrations (positive δ estimate).
507 Residual Chl-*a* concentrations also became more variable at greater mean lake depths
508 (positive δ estimate), while residual PTI and MFGI became less variable in these deeper lakes
509 (negative δ estimates). Both residual SPI and PTI became more variable in higher altitude
510 lakes (positive δ estimates). The model selection process, using multi-model inference to find
511 the most well supported predictors of between-lake variations in each of these metrics, was
512 repeated after including these additional variance structures, although the final parameter
513 estimates for the fixed effects were affected minimally (results not shown).

514

515 **4. Discussion**

Comparison of sources of variation in metric scores showed that among-lake variation was by far the dominant component of variability for all seven metrics. This suggested that, all other things being equal, the capability of the metrics to respond to pressures acting at the lake level should not be limited by sampling variation arising from within-lake spatial variation. Differences in locations around a lake, or sampling and analytical variability, only accounted for a relatively small proportion of the variance in metric scores. These results are especially true for the three candidate phytoplankton metrics adopted by many European Member States: chlorophyll, PTI, and cyanobacterial abundance. For these metrics, 88% or more of the variance in metric scores occurred at the among-lake level of the sampling hierarchy. Between-analyst and between sub-sample variation accounted for most of the within-lake variation. Little variation was attributable to within-lake spatial heterogeneity i.e. differences among lake stations and repeated sampling from each station. This was despite the fact that lake stations were treated as "random" in the modelling approach even though they were selected: which should lead to an over-estimate of the station-to-station variability. Lake stations were selected to represent water columns of mean depth or greater in the present study, and it is plausible that a greater station level effect might have been observed if stations had been selected from a wider range of water depths and/or including from outflow or edge samples. Processes in inshore regions of lakes, such as flushing by influent waters (Mackay et al., 2011), enhanced zooplankton grazing facilitated by structurally complex macrophyte refugia (Schriver et al., 1995) or chemical interactions with macrophytes (Wium-Andersen et al., 1982; Jasser, 1995) may generate differences in phytoplankton communities between these areas and the deeper, open-water, zone. If sampling stations are distributed among the multiple interconnected basins of some lakes, it is conceivable that more station-level metric variation would be observed, but any resulting uncertainty can be minimised by using the facility within the WFD to treat such basins as separate waterbodies.

Though within-lake metric variance was relatively low compared to among-lake variance, the relative magnitude of the components of the former indicates potential areas for the refinement of field sampling campaigns, which could improve the precision of ecological assessments of lakes. Increasing the number of open water sampling stations visited, or the number of samples collected at each station, would do little to improve the precision of ecological assessments based upon these phytoplankton metrics. The representativeness of ecological assessments based upon the metrics, with respect to the impact of lake level pressures, could instead be improved by processing greater numbers of replicate sub-samples from each sample and standardising either *i)* analyst identity for samples from different lakes, or *ii)* taxonomic skills and laboratory procedures among different analysts (e.g.Vuorio et al., 2007). In fact, the majority of analysts had attended workshops that aimed to standardise sample processing techniques and algal identification/enumeration. Furthermore, counters followed standard procedures based upon ISO 10260 (1992), CEN 15204 (2006), National Rivers Authority (1995) and Brierley et al. (2007). It may therefore be that analyst variability was lower than normal. Nevertheless, the results of this study indicate that rigorous standardisation of sample mixing and sedimentation protocols, as well as of taxonomic procedures, can help minimise sampling and analytical variability. In turn, this would permit more meaningful comparisons of ecological status between different lakes.

We should also note that, in the current sampling design, the effects of analyst and sub-sampling variation were crossed. Therefore, it was not possible to compare results derived from different analysts counting exactly the same fields of view from the same sub-sample, or the same analyst counting different fields of view from the same sub-sample. Furthermore,

565 the sub-samples were actually sub-sub-sampled prior to microscopic examination; another
566 source of potential metric variability that was unquantifiable in this study. It is, therefore,
567 difficult to truly isolate the effect of analyst variation upon metric scores in this study. Future
568 studies targeting sources of variation arising from sampling processing and analyst variation
569 alone would allow more accurate assessment of the extent to which metrics are influenced by
570 these factors.

571

572 Taking a multi-model inference approach, there was strong support for a response of metric
573 scores to phosphorus concentrations for three of the seven metrics: Chl-*a* concentration, PTI
574 and total cyanobacterial biovolume. This would suggest that these proposed metrics are
575 indeed responsive to the eutrophication pressure gradient apparent across the lakes sampled.
576 Furthermore, this would suggest that such relationships are detectable, despite metric
577 variation arising due to sampling/sample processing decisions. These relationships suggested
578 a general increase in Chl-*a* concentration and cyanobacterial abundance with increased
579 phosphorus availability. The finding that Chl-*a* concentration increases with lake phosphorus
580 concentration is consistent with the idea that the availability of this nutrient determines the
581 supportive capacity of a lake system for phytoplankton biomass (Reynolds, 2006); a
582 relationship embodied in the results of previous empirical (Dillon and Rigler, 1974;
583 Schindler, 1978; Phillips et al., 2008; Sondergaard et al. 2011), and process-based modelling
584 studies (Elliott et al., 2006). Indeed, between lake variations in total phosphorus
585 concentration have been found to be more powerful predictors of phytoplankton biomass than
586 similar variations in total nitrogen concentrations (Brown et al., 2000; Phillips et al., 2008;
587 Sondergaard et al., 2011), though this difference may be dependent on the relative availability
588 of these two nutrients (McCauley et al., 1989; Brown et al., 2000; Phillips et al., 2008). The
589 observation of increased cyanobacterial biomass at higher phosphorus concentrations is

similarly consistent with the findings of previous studies (Smith, 1985; Watson et al., 1997; Elliott et al., 2006). PTI scores were also higher in lakes with higher phosphorus concentrations, as shown by Phillips et al. (2010), due to increases in the biomass of cyanobacteria, and some members of the Chlorophyceae and Bacillariophyceae.

Comparison of results across metrics also revealed consistent support for an effect of mean lake depth, particularly for FTI, MFGI and total cyanobacterial biovolume (though there was also weaker support for this effect for PTI, SPI and Chl-*a* concentration). Mean lake depth acts as a surrogate for a variety of physical and chemical attributes, such as maximum depth, the likelihood of thermal stratification, flushing rate, underwater light availability and the likelihood of internal nutrient loading (Kalff, 2002). Furthermore, inverse relationships between among-lake variations in lake depth and Chl-*a* concentrations/cyanobacterial abundance have been noted in a number of previous studies (Pridmore et al., 1985; Smith, 1985; Smith et al., 1987; Phillips et al., 2008). The fact that lake depth covaries with so many other physical and chemical determinants of phytoplankton production, renders hypothesising the mechanism behind the observed relationships difficult. That depth and total phosphorus concentration co-occur as independent predictors in the top models for Chl-*a* concentration and total cyanobacterial biovolume would suggest that depth offers "unique" explanatory power for these phytoplankton metrics compared to phosphorus availability. The higher observed Chl-*a* concentrations and cyanobacterial biovolumes in shallower lakes could be related to the increased average nutrient supply in these systems. This would occur due to frequent mixing-induced internal nutrient loading. In addition, in shallow lakes sedimented phytoplankton may be resuspended back into the water column. However, it is also true that in deep lakes, simply mixing at times during the summer and subsequent light limitation of

614 primary production may result in a lower phytoplankton/cyanobacterial biomass (Sakamoto,
615 1966; Berger et al., 2006; Phillips et al., 2008).
616
617 Effects of mean depth were also strongly supported in analyses of functional composition
618 metrics (MFGI, FTI), suggesting systematic changes in community structure and trait
619 representation with changes in lake depth. High values of MFGI (such as in shallow lakes)
620 indicate an increasing biomass of large, colonial and buoyant Chroococcales or Nostocales
621 cyanobacteria. Low MFGI values (deep lakes) indicate an increasing biomass of non-motile
622 xanthophytes, small pennate diatoms, small centric diatoms or Oscillatoriales. The inverse
623 relationship between MFGI and depth seems to be driven by the trophic preferences of these
624 functional groups, with the most eutrophic colonial Chroococcales and Nostocales being
625 more abundant in shallow lakes. The results for these trait metrics may therefore suggest that
626 the effect of mean depth is via correlated changes in the frequency of episodic nutrient
627 release, as hypothesized above for Chl-*a* and cyanobacterial biovolume.
628
629 However, for each metric, considerable among-lake variation remained unexplained by the
630 available environmental data. This was particularly the case for the composition (PTI, MFGI,
631 SPI, FTI) and bloom (total cyanobacterial biovolume, evenness) metrics. While some of this
632 variation might arise due to measurement errors in some of the environmental variables, this
633 would also suggest the existence of important unmeasured drivers of phytoplankton
634 community structure. Geographic variables were included in the analysis as a proxy for the
635 effects of broad climatic gradients upon community structure, via lake physical processes, but
636 the effects of grazing, flushing, water colour (DOC), silica or even other parameters
637 associated with eutrophication pressure, such as dissolved nitrogen and turbidity, are all

likely to be influential. However, these variables were not recorded consistently enough to include their effects in the current analysis.

Unexplained among-lake variability is also likely to arise due to the temporal dimension inherent in phytoplankton-environment interactions. Current phytoplankton community structure is a biological response to previous environmental conditions (Madgwick et al., 2006), with the time lag of the relationship determined by the time-scale over which phytoplankton gather resources and replicate. It is therefore to be expected that phytoplankton communities (and thus metrics) will show within-year temporal variation, and that the results of waterbody assessment will vary accordingly. However, waterbody assessment must ultimately depend upon sampling programmes that produce "snapshots" of this temporal variation. It is therefore important to know the uncertainties associated with such samples if we are to understand how well sample metric scores represent *current* conditions. Once sampling uncertainty is resolved for samples collected at a single point in time (the aim of this study), the next step would be to examine the temporal uncertainties associated with waterbody assessment. To this end, the relationship between metrics and environmental drivers could be resolved by integrating these variables over the growing season. In lakes with suitable time-series data it would, in principle, be possible to model temporal variability in metric scores as a further source of uncertainty, and also include the temporal relationship between metrics and drivers. Explicit consideration of these temporal aspects could not be achieved here due to the sampling design, but this is highly recommended for future research.

For six of the seven metrics there was evidence that not only mean values, but also variability, changed systematically with among-lake variations in physical, chemical and

663 geographical attributes. Residual variability in metrics was not constant with respect to total
664 phosphorus concentration (Chl-*a*, evenness, total cyanobacterial biovolume), mean depth
665 (Chl-*a*, PTI, MFGI) or altitude (PTI, SPI). Furthermore, the association of this variability
666 with specific drivers differed among metrics e.g. increases in total phosphorus concentration
667 led to increased variability in total cyanobacterial biovolume, but decreases in variability in
668 evenness and Chl-*a*. These findings are similar to the observations of Clarke et al. (2006a),
669 who found that the sampling variability of macroinvertebrate community metrics can vary as
670 a function of the overall ecological quality of a site (i.e. the average metric score). Plots of
671 residual metric variability against predictor variables for some of the metrics in the present
672 analysis suggested that a greater spread of metric variation for only a small proportion of the
673 32 study lakes compared to the rest was sufficient for the inclusion of these variance
674 structures to result in an improvement in overall model fit, as judged by AIC. If a future study
675 were to compile data from a larger number of lakes it would be possible to assess how robust
676 these among-lake gradients in metric variability are. For now, the present results suggest that
677 phytoplankton metric variability, and therefore uncertainty, may differ with attributes of the
678 environment from which the phytoplankton samples were drawn and that this may be an
679 important consideration when planning monitoring programmes.
680

681 **5. Conclusion**

682 By analysing the results of a unique pan-European hierarchical sampling programme we have
683 shown that seven candidate phytoplankton community metrics, being considered for
684 intercalibration under the Water Framework Directive, show the potential to indicate among
685 lake variations in the effects of environmental pressures. This is particularly true for Chl-*a*
686 concentration, PTI and total cyanobacterial biovolume, which appear to respond to variations
687 in total phosphorus concentration as a proxy of eutrophication. These metrics are clearly also

688 responsive to among-lake variations in other attributes such as mean depth, and other
689 unidentified factors. In order to further assess the performance of such metrics, it is essential
690 to examine the temporal dimension of their variability (Sondergaard et al., 2011) and also the
691 extent to which uncertainty in water body assessment may vary systematically among lakes
692 differing in their physico-chemical and ecological attributes. These should be considered
693 priorities for future research into freshwater ecological quality assessment.
694

695 **Acknowledgements**

696 This paper is a result of the project WISER (Water bodies in Europe: Integrative Systems to
697 assess Ecological status and Recovery) funded by the European Union under the 7th
698 Framework Programme, Theme 6 (Environment including Climate Change) (contract No.
699 226273), www.wiser.eu. We are indebted to all who processed the samples from the field
700 campaign: Ana Negro (University of Salamanca), Tatiana Caraballo (CEAB), Roser Farres
701 (CEAB), Pierisa Panzani (CNR-ISE), Maria Antonietta Mariani (University of Sassari),
702 Tomasa Virdis (ENAS), Aimar Rakko (EMU), Małgorzata Poniewozik (The John Paul II
703 Catholic University of Lublin), Annette Tworeck (LBH Freiburg), Camilla H. C. Hagman
704 (NIVA), Mitzi De Ville (NERC-CEH), Rene Groben (NERC-CEH), Sarah Pritchard (Beacon
705 Biological), Bill Brierley (EA), Reija Jokipii (SYKE) & Maija Niemelä (SYKE).

706 **References**


707 Barton, K., 2011. MuMIn: Multi-model inference. R package version 1.0.0. http://CRAN.R-
708 project.org/package=MuMIn.

709 Bates, D., Maechler, M., Bolker, B., 2011. lme4: Linear mixed-effects models using S4
710 classes. R package version 0.999375-41. http://CRAN.R-project.org/package=lme4.

711 Berger, S. A., Diehl, S., Kunz, T. J., Albrecht, D., Oucible, A. M., Ritzer, S., 2006. Light
712 supply, plankton biomass, and seston stoichiometry in a gradient of lake mixing
713 depths. Limnol. Oceanogr. 51, 1898-1905.

714 Brierley, B., Carvalho, L., Davies, S., Krokowski, J., 2007. Guidance on the quantitative
715 analysis of phytoplankton in freshwater samples. Report to SNIFFER (Project
716 WFD80), Edinburgh.

717 Brown, C. D., Hoyer, M. V., Bachmann, R. W., Canfield Jr., D. E., 2000. Nutrient-
718 chlorophyll relationships: an evaluation of empirical nutrient-chlorophyll models
719 using Florida and north-temperate lake data. Can. J. Fish. Aquat. Sci. 57,1574-1583.

720 Burnham, K. P., Anderson, D. R., 2002. Model selection and multimodel inference: a
721 practical information-theoretic approach. Springer, New York, 488pp.

722 Carvalho, L., Lepisto, L., Rissanen, J., Pietiläinen, O.-P., Rekolainen, S., Torok, L., Lyche
723 Solheim, A., Saloranta, T., Ptacnik, R., Tartari, G., Cardoso, A. C., Premazzi, G.,
724 Gunn, I., Penning, E., Hanganu, J., Hellsten, S., Orhan, I., Navodaru, I., 2006.
725 Nutrients and eutrophication in lakes, in: Solimini, A., Cardoso, A.C., Heiskanen, A.-
726 S., (Eds.), Indicators and methods for the ecological status assessment under the
727 Water Framework Directive: Linkages between chemical and biological quality of
728 surface waters. European Commission, Luxembourg, pp3-32.

729 Carvalho, L., Poikane S., Lyche Solheim A., Phillips G., Borics G., Catalan J., De Hoyos C.,
730 Drakare S., Dudley B., Jarvinen M., Laplace-Treyture C., Maileht K., McDonald C.,

731 Mischke U., Moe J., Morabito G., Nõges P., Nõges T., Ott I., Pasztaleniec, A.,
732 Skjelbred B. & Thackeray S. 2012. Strength and uncertainty of lake phytoplankton
733 metrics for assessing eutrophication impacts in lakes. Hydrobiologia, online early,
734 DOI 10.1007/s10750-012-1344-1
735 CEN EN 15204. 2006. Water quality – Guidance standard for the routine analysis of
736 phytoplankton abundance and composition using inverted microscopy (Utermöhl
737 technique).
738 Clarke, R.T., Furse, M.T., Gunn, R.J.M., Winder, J.M., Wright, J.F., 2002. Sampling
739 variation in macroinvertebrate data & implications for river quality indices.
740 Freshwater Biology 47, 1735-1751.
741 Clarke, R. T., Davy-Bowker, J., Sandin, L., Friberg, N., Johnson, R. K., Bis, B., 2006a.
742 Estimates and comparisons of the effects of sampling variation using "national"
743 macroinvertebrate sampling protocols on the precision of metrics used to assess
744 ecological status. Hydrobiologia 566, 477-503.
745 Clarke, R. T., Hering, D., 2006. Errors and uncertainty in bioassessment methods – major
746 results and conclusions from the STAR project and their application using
747 STARBUGS. Hydrobiologia 566, 433-439.
748 Clarke, R. T., Lorenz, A., Sandin, L., Schmidt-Kloiber, A., Strackbein, J., Kneebone, N. T.,
749 Haase, P., 2006b. Effects of sampling and sub-sampling variation using the STAR-
750 AQEM sampling protocol on the precision of macroinvertebrate metrics.
751 Hydrobiologia 566, 441-459.
752 Clarke, R. T., 2012. Estimating confidence of European WFD ecological status class and
753 WISER Bioassessment Uncertainty Guidance Software (WISERBUGS).
754 Hydrobiologia. doi:10.1007/s10750-012-1245-3.
755 Crawley, M.J., 2007. The R Book. Wiley, Chichester, 942pp.

Dillon, P. J., Rigler, F. H., 1974. The phosphorus-chlorophyll relationship in lakes. Limnol. Oceanogr. 19,767-773.

Dudley, B. J., Dunbar, M. J., Penning, E., Kolada, A., Hellsten, S., Kanninen, A., 2010. WISER Deliverable D3.2-2: Report on uncertainty in macrophyte metrics, 29pp.

EC, 2000. Directive 2000/60/EC of the European Parliament and of the council of 23 October 2000 establishing a framework for community action in the field of water policy. Official Journal of the European Communities L327, 1-72.

Elliott, J. A., Jones, I. D., Thackeray, S. J., 2006. Testing the sensitivity of phytoplankton communities to changes in water temperature and nutrient load, in a temperate lake. Hydrobiologia 559,401-411.

Hering, D., Borja, A., Carstensen, J., Carvalho, L., Elliott, M., Feld, C. K., Heiskanen, A.-S., Johnson, R. K., Moe, J., Pont, D., Solheim, A. L., van de Bund, W., 2010. The European Water Framework Directive at the age of 10: a critical review of the achievements with recommendations for the future. Sci. Total Environ. 408, 4007-4019.

ISO 10260, 1992. Water quality – Measurement of biochemical parameters – Spectrometric determination of the chlorophyll-*a* concentration. Page 6 Int. Org. Standard., Geneva, 1st edn. 1992–07–15.

Jasser, I., 1995. The influence of macrophytes on a phytoplankton community in experimental conditions. Hydrobiologia 306, 21-32.

Kalff, J., 2002. Limnology. Prentice-Hall Inc., Upper Saddle River, NJ, 592pp.

Kamenir, Y., Morabito, G., 2009. Lago Maggiore oligotrophication as seen from the long-term evolution of its phytoplankton taxonomic size structure. J. Limnol. 68, 146-161.

779 Kümmerlin, R. E., 1998. Taxonomical response of the phytoplankton community of Upper
780 Lake Constance (Bodensee-Obersee) to eutrophication and re-oligotrophication. Adv.
781 Limnol. 53, 109-117.
782 Mackay, E. B., Jones, I. D., Folkard, A. M., Thackeray, S. J., 2011. Transition zones in small
783 lakes: the importance of dilution and biological uptake on lake-wide heterogeneity.
784 Hydrobiologia 678, 85-97.
785 Madgwick, G., Jones, I. D., Thackeray, S. J., Elliott, J. A., Miller, H. J., 2006. Phytoplankton
786 communities and antecedent conditions: high resolution sampling in Esthwaite Water.
787 Freshwater Biol. 51, 1798-1810.
788 McCauley, E., Downing, J. A., Watson, S., 1989. Sigmoid relationships between nutrients
789 and chlorophyll among lakes. Can. J. Fish. Aquat. Sci. 46, 1171-1175.
790 Mischke, U., Carvalho, L., McDonald, C., Skjelbred, B., Solheim, A. L., Phillips, G., de
791 Hoyos, C., Borics, G., Moe, J., 2010. WISER Deliverable D3.1-2: Report on
792 phytoplankton bloom metrics, 45pp, http://www.wiser.eu/
793 Murphy, J., Reilly, J. P., 1962. A modified single solution method for the determination of
794 phosphate in natural waters. Anal. Chim, Acta 27, 31-36.
795 Murphy, K. J., Kennedy, M. P., McCarthy, V., Ó'Hare, M. T., Irvine, K., Adams, C., 2002. A
796 review of ecology based classification systems for standing freshwaters. SNIFFER
797 Project Number: W(99)65, Environment Agency R&D Technical Report: E1-091/TR.
798 National Rivers Authority, 1995. Test Methods and Procedures: Freshwater Phytoplankton.
799 National Rivers Authority, Peterborough.
800 Padisák, J., Reynolds, C. S., 1998. Selection of phytoplankton associations in Lake Balaton,
801 Hungary, in response to eutrophication and restoration measures, with special
802 reference to the cyanoprokaryotes. Hydrobiologia 384, 41-53.

803 Phillips, G., Pietiläinen, O.-P., Carvalho, L., Solimini, A., Lyche Solheim, A., Cardoso, A.
804 C., 2008. Chlorophyll–nutrient relationships of different lake types using a large
805 European dataset. Aquat. Ecol. 42, 213-226.
806 Phillips, G., Skjelbred, B., Morabito, G., Carvalho, L., Solheim, A. L., Andersen, T.,
807 Mischke, U., de Hoyos, C., Borics, G., 2010. WISER Deliverable D3.1-1: Report on
808 phytoplankton composition metrics, including a common metric approach for use in
809 intercalibration by all GIGs, 56pp, http://www.wiser.eu/
810 Pielou, E. C., 1969. An introduction to mathematical ecology. Wiley-Interscience, New York,
811 294pp.
812 Pielou, E. C., 1975. Ecological diversity. Wiley Interscience, New York, 165pp.
813 Pinel-Alloul, B., Ghadouani, A., 2007. Spatial heterogeneity of planktonic microorganisms in
814 aquatic systems: mutiscale patterns and processes, in: Franklin, R., Mills, A. (Eds.),
815 The Spatial Distribution of Microbes in the Environment. Springer.
816 Pinel-Alloul, B., Méthot, G., Verrault, G.,Vigneault, Y., 1990. Phytoplankton in Quebec
817 lakes: Variation with lake morphometry, and with natural and anthropogenic
818 acidification. Can. J. Fish. Aquat. Sci. 47, 1047-1057.
819 Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., R Development Core Team, 2010. nlme:
820 Linear and nonlinear mixed effects models. R package version 3.1-97.
821 Pridmore, R. D., Vant, W. N., Rutherford, J. C., 1985. Chlorophyll-nutrient relationships in
822 North Island lakes (New Zealand). Hydrobiologia 121, 181-189.
823 R Development Core Team, 2011. R: A language and environment for statistical computing.
824 R Foundation for Statistical Computing, ISBN 3-900051-07-0, url = http://www.R-
825 project.org/, Vienna, Austria.
826 Reynolds, C. S., 2006. Ecology of phytoplankton. Cambridge University Press, Cambridge.

827 Sakamoto, M., 1966. Primary production by phytoplankton community in some Japanese
828 lakes and its dependence on lake depth. Arch.Hydrobiol. 62, 1-28.
829 Salmaso, N., Padisak, J., 2007. Morpho-Functional Groups and phytoplankton development
830 in two deep lakes (Lake Garda, Italy and Lake Stechlin, Germany). Hydrobiologia
831 578, 97-112.
832 Schindler, D. W., 1978. Factors regulating phytoplankton production and standing crop in the
833 world's freshwaters. Limnol. Oceanogr. 23, 478-486.
834 Schriver, P., Bogestrand, J., Jeppesen, E., Sondergaard, M., 1995. Impact of submerged
835 macrophytes on fish-zooplankton-phytoplankton interactions: large-scale enclosure
836 experiments in a shallow eutrophic lake. Freshwater Biol. 33, 255-270.
837 Smith, V. H., 1985. Predictive models for the biomass of blue-green algae in lakes. Water
838 Resour. Bull. 21, 433-439.
839 Smith, V. H., Willen, E., Karlsson, B., 1987. Predicting the summaer peak biomass of four
840 species of blue-green algae (Cyanophyta/Cyanobacteria) in Swedish lakes. Water
841 Resour. Bull. 23, 397-402.
842 Sondergaard, M., Larsen, S. E., Jorgensen, T. B., Jeppesen, E., 2011. Using chlorophyll *a* and
843 cyanobacteria in the ecological classification of lakes. Ecol. Indic. 11, 1403-1412.
844 Utermöhl, H., 1958. Zur Vervollkommnung der quantitativen Phytoplankton-Methodik. Mitt.
845 Int. Ver. Theor. Angew. Limnol. 9, 1-38.
846 Vuorio, K., Lepistö, L., Holopainen, A.-L., 2007. Intercalibrations of freshwater
847 phytoplankton analyses. Boreal Environ. Res. 12, 561-569.
848 Warnes, G. R., 2010. gplots: Various R programming tools for plotting data. R package
849 version 2.8.0, url = {http://CRAN.R-project.org/package=gplots.

Watson, S. B., McCauley, E., Downing, J. A., 1997. Patterns in phytoplankton taxonomic composition across temperate lakes of different nutrient status. Limnol. Oceanogr. 42, 487-495.

WISE. 2008. Water Note 7. Intercalibration: A common scale for Europe's waters. European Commission, 2pp.

Wium-Andersen, S., Anthoni, U., Christophersen, C., Houen, G., 1982. Allelopathic effects on phytoplankton bysubstances isolated from aquatic macrophytes (Charales). Oikos 39, 187-190.

Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A., Smith, G. M., 2009. Mixed effects models and extensions in ecology with R. Springer, New York, 574pp.

Table 1. Lakes sampled in the field campaign. GIG indicates the Geographical Intercalibration Group within which each lake falls: AL = Alpine, CB = Central/Baltic, M = Mediterranean, N = Northern. Only Chl-*a* data were available for lakes marked with an asterisk.

| **Lake** | **Country** | **GIG** | **Latitude (°N)** | **Longitude (°W)** | **Mean depth (m)** | **Maximum depth (m)** | **Altitude (m a.s.l.)** | **Total phosphorus (mg $m^{-3}$)** | **Alkalinity (meq $L^{-1}$)** |
|---|---|---|---|---|---|---|---|---|---|
| Nordborgsø | Denmark | CB | 55.06 | 9.76 | 5.0 | 8.5 | 20 | 62.67 | 2.30 |
| Fussingsø | Denmark | CB | 56.47 | 9.88 | 12.6 | 31.0 | 15 | 45.67 | 1.50 |
| Saadjärv | Estonia | CB | 58.54 | 26.65 | 8.0 | 21.7 | 85 | 14.00 | 2.53 |
| Viljandi | Estonia | CB | 58.35 | 25.60 | 5.5 | 9.5 | 75 | 21.50 | 4.40 |
| Sääksjärvi | Finland | N | 62.17 | 25.73 | 9.3 | 15.2 | 121 | 12.00 | 0.23 |
| Vuojärvi | Finland | N | 62.41 | 25.94 | 4.4 | 10.2 | 91 | 35.5 | 0.54 |
| Iso-Jurvo | Finland | N | 62.60 | 25.93 | 8.6 | 29.6 | 139 | 8.00 | 0.06 |
| Salagou | France | M | 43.66 | 3.40 | 15.6 | 49.3 | 139 | 21.76 | 2.77 |
| Caramany | France | M | 42.74 | 2.59 | 14.5 | 36.0 | 170 | 26.80 | 2.96 |
| Glindower See | Germany | CB | 52.36 | 12.92 | 4.9 | 14.3 | 24 | 151.00 | 2.40 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Grienericksee | Germany | CB | 53.10 | 12.89 | 4.7 | 11.5 | 55 | 19.00 | 2.20 |
| Roofensee | Germany | CB | 53.11 | 13.02 | 9.0 | 19.1 | 59 | 18.00 | 2.00 |
| Alserio | Italy | AL | 45.78 | 9.21 | 5.0 | 8.0 | 243 | 24.00 | 2.34 |
| Bidighinzu | Italy | M | 40.56 | 8.66 | 7.5 | 21.8 | 330 | 65.00 | 2.24 |
| Candia | Italy | AL | 45.33 | 7.92 | 5.0 | 7.5 | 226 | 16.50 | 1.00 |
| Monate | Italy | AL | 45.80 | 8.66 | 18.0 | 34.0 | 266 | 8.50 | 0.88 |
| Segrino | Italy | AL | 45.83 | 9.27 | 3.5 | 8.0 | 374 | 12.50 | 2.23 |
| Nøklevann | Norway | N | 59.88 | 10.88 | 19.0 | 31.0 | 163 | 4.00 | 0.17 |
| Longumvatnet | Norway | N | 58.49 | 8.76 | 14.0 | 35.5 | 34 | 7.50 | 0.28 |
| Temse | Norway | N | 58.38 | 8.64 | 6.0 | 10.2 | 15 | 17.00 | 0.32 |
| Rumian | Poland | CB | 53.38 | 20.00 | 6.0 | 14.0 | 152 | 88.00 | 2.60 |
| Lidzbarskie | Poland | CB | 53.26 | 19.80 | 10.0 | 24.0 | 128 | 56.50 | 2.45 |
| Kiełpińskie | Poland | CB | 53.35 | 19.79 | 5.8 | 10.0 | 120 | 63.50 | 2.90 |
| Vencías, Las | Spain | M | 41.43 | -3.96 | 8.0 | 14.8 | 869 | 20.46 | 2.43 |
| Vega de Jabalón | Spain | M | 38.76 | -3.79 | 6.6 | 10.8 | 635 | 54.65 | 2.26 |
| Arquillo de San Blas | Spain | M | 40.36 | -1.21 | 34.0 | 38.0 | 970 | 6.90 | 2.80 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Fiolen* | Sweden | N | 57.08 | 14.53 | 3.8 | 10.0 | 226 | 10.00 | 0.10 |
| Skirösjön* | Sweden | N | 57.36 | 15.38 | 5.2 | 8.0 | 146 | 45.33 | 0.63 |
| Västra Solsjön* | Sweden | N | 59.08 | 12.29 | 12.3 | 40.0 | 147 | 10.00 | 0.16 |
| Loweswater | UK | N | 54.58 | -3.36 | 8.0 | 14.8 | 125 | 9.97 | 0.22 |
| Grasmere | UK | N | 54.45 | -3.02 | 8.4 | 19.4 | 61 | 9.15 | 0.21 |
| Rostherne Mere | UK | CB | 53.35 | -2.39 | 11.5 | 29.7 | 27 | 121.00 | 2.44 |

Table 2. Proportions of metric variance at different levels in the sampling hierarchy, for null models of the seven different metrics. Total among = Country + Lake, Total within = Station + Sample + Analyst + Error (sub-sample). Models fitted using REML estimation.

| **Metric** | **Country** | **Lake** | **Station** | **Sample** | **Analyst** | **Error (sub-sample)** | **Total within** | **Total among** |
|---|---|---|---|---|---|---|---|---|
| $Log_{10}$ Chl-*a* | 0.00 | 0.96 | 0.01 | 0.01 | - | 0.02 | 0.04 | 0.96 |
| PTI | 0.00 | 0.88 | <0.01 | 0.00 | 0.04 | 0.07 | 0.12 | 0.88 |
| SPI | 0.00 | 0.65 | 0.03 | 0.00 | 0.19 | 0.13 | 0.35 | 0.65 |
| MFGI | 0.00 | 0.86 | 0.02 | <0.01 | 0.05 | 0.08 | 0.14 | 0.86 |
| FTI | 0.00 | 0.81 | 0.02 | 0.00 | 0.09 | 0.08 | 0.19 | 0.81 |
| Evenness | 0.00 | 0.69 | 0.04 | 0.00 | 0.17 | 0.10 | 0.31 | 0.69 |
| $Log_{10}$ total cyanobacteria | 0.09 | 0.86 | 0.01 | 0.00 | 0.02 | 0.03 | 0.06 | 0.94 |

Table 3. Relationships between metrics and environmental drivers, in the most optimal linear mixed-effects models for each metric. Shown are the number of estimated model parameters (*k*), the predictors present in the model, the difference in AIC between the most optimal model and the corresponding null model ($\Delta AIC_{null}$) and the Akaike weight; a measure of the relative level of support for the most optimal model, compared to other candidate models, given the data. For the Akaike weight, values close to 1 indicate overwhelming support for the corresponding model, while lower values indicate the presence of other models with similar levels of support. See Figures 4-6 for model averaged estimates of the parameters for each metric-lake attribute relationship, based upon all models with similar levels of support for each metric. Note that *k* includes the global intercept and parameters for both the fitted predictors and the random effects variances. For each predictor, the sign of the corresponding relationship is given as positive (+) or negative (-). Models fitted using ML estimation.

| **Metric** | ***k*** | **Predictors** | **$\Delta AIC_{null}$** | **Akaike weight** |
|---|---|---|---|---|
| $Log_{10}$ Chl-*a* | 6 | $Log_{10}$ Mean lake depth (-)<br>$Log_{10}$ total phosphorus (+)<br>Latitude (+) | 35.5 | 0.12 |
| PTI | 7 | $Log_{10}$ Mean lake depth (-)<br>$Log_{10}$ total phosphorus (+)<br>$Log_{10}$ Altitude (+) | 13.5 | 0.11 |
| SPI | 6 | $Log_{10}$ Mean lake depth (-)<br>$Log_{10}$ Altitude (+) | 2.0 | 0.12 |
| MFGI | 6 | $Log_{10}$ Mean lake depth (-)<br>$Log_{10}$ Altitude (+) | 10.0 | 0.12 |
| FTI | 6 | $Log_{10}$ Mean lake depth (-)<br>$Log_{10}$ Altitude (+) | 9.0 | 0.19 |
| Evenness | 6 | $Log_{10}$ total phosphorus (-) | 2.3 | 0.06 |

| | | | | |
|---|---|---|---|---|
| | | Alkalinity (+) | | |
| $Log_{10}$ total cyanobacteria | 6 | $Log_{10}$ Mean lake depth (-) | 16.2 | 0.13 |
| | | $Log_{10}$ total phosphorus (+) | | |

Table 4. Models examining metric variability as a function of environmental drivers. AIC comparison of the most optimal linear mixed-effects models for each of the seven phytoplankton metrics (see Table 3), when including/excluding variance structures to account for changes in metric variability (residual metric variance) as a function of the fitted predictors. Shown are the predictors that residual variability is modeled as a function of (Predictor), the estimated delta parameter for the exponential function describing the relationship between residual variance and the named predictor (δ) and the AIC for each model. For each metric, the most optimal model is indicated in bold. Models fitted using REML estimation.

| **Metric** | **Model No.** | **Predictor** | **δ** | **AIC** |
|---|---|---|---|---|
| $Log_{10}$ Chl-*a* | 1 | None | - | -195.1 |
| | 2 | $Log_{10}$ Mean lake depth | 0.88 | -205.7 |
| | 3 | $Log_{10}$ total phosphorus | -0.70 | -230.7 |
| | 4 | Latitude | 0.02 | -198.3 |
| | **5** | **$Log_{10}$ Mean lake depth** | **0.57** | **-233.8** |
| | | **$Log_{10}$ total phosphorus** | **-0.65** | |
| | 6 | $Log_{10}$ total phosphorus | -0.70 | -228.8 |
| | | Latitude | <0.01 | |
| | 7 | $Log_{10}$ Mean lake depth | 0.75 | -205.3 |
| | | Latitude | 0.01 | |
| | 8 | $Log_{10}$ Mean lake depth | 0.57 | -231.8 |
| | | $Log_{10}$ total phosphorus | -0.66 | |
| | | Latitude | <-0.01 | |
| PTI | 1 | None | - | -138.7 |
| | 2 | $Log_{10}$ Mean lake depth | -0.44 | -144.9 |
| | 3 | $Log_{10}$ total phosphorus | -0.40 | -147.9 |

| | | | | |
|---|---|---|---|---|
| | 4 | $Log_{10}$ Altitude | 0.66 | -180.4 |
| | 5 | $Log_{10}$ Mean lake depth | -0.53 | -156.3 |
| | | $Log_{10}$ total phosphorus | -0.43 | |
| | 6 | $Log_{10}$ total phosphorus | -0.11 | -179.0 |
| | | $Log_{10}$ Altitude | 0.62 | |
| | **7** | **$Log_{10}$ Mean lake depth** | **-0.39** | **-183.5** |
| | | **$Log_{10}$ Altitude** | **0.65** | |
| | 8 | $Log_{10}$ Mean lake depth | -0.43 | -183.1 |
| | | $Log_{10}$ total phosphorus | -0.17 | |
| | | $Log_{10}$ Altitude | 0.59 | |
| SPI | 1 | None | - | -1682.9 |
| | 2 | $Log_{10}$ Mean lake depth | 0.19 | -1682.8 |
| | **3** | **$Log_{10}$ Altitude** | **0.23** | **-1688.6** |
| | 4 | $Log_{10}$ Mean lake depth | -0.06 | -1686.7 |
| | | $Log_{10}$ Altitude | 0.25 | |
| MFGI | 1 | None | - | -1760.6 |
| | **2** | **$Log_{10}$ Mean lake depth** | **-0.43** | **-1763.3** |
| | 3 | $Log_{10}$ Altitude | -0.12 | -1760.7 |
| | 4 | $Log_{10}$ Mean lake depth | -0.42 | -1763.3 |
| | | $Log_{10}$ Altitude | -0.12 | |
| FTI | **1** | **None** | **-** | **-1854.2** |
| | 2 | $Log_{10}$ Mean lake depth | -0.15 | -1853.1 |
| | 3 | $Log_{10}$ Altitude | 0.01 | -1852.2 |
| | 4 | $Log_{10}$ Mean lake depth | -0.19 | -1851.3 |
| | | $Log_{10}$ Altitude | 0.04 | |
| Evenness | 1 | None | - | -621.7 |
| | **2** | **$Log_{10}$ total phosphorus** | **-0.51** | **-642.6** |

| | | | | |
|---|---|---|---|---|
| | 3 | Alkalinity | -0.13 | -633.8 |
| | 4 | $Log_{10}$ total phosphorus | -0.42 | -641.6 |
| | | Alkalinity | -0.04 | |
| $Log_{10}$ total cyanobacteria | 1 | None | - | -171.6 |
| | 2 | $Log_{10}$ Mean lake depth | -0.52 | -177.1 |
| | **3** | **$Log_{10}$ total phosphorus** | **0.71** | **-214.4** |
| | 4 | $Log_{10}$ Mean lake depth | -0.23 | -214.0 |
| | | $Log_{10}$ total phosphorus | 0.67 | |

**Figure legends**

Fig. 1. The sampling design employed in each lake. Samples were collected from three stations, above the deepest point ($z_{max}$), the mean depth ($z_{mean}$) and a depth intermediate between the maximum and mean depths ($z_{int}$). Two samples (S1, S2) were collected at each station. At each station, three sub-samples (Sub1, Sub2, Sub3) were collected from sample 1 and one sub-sample from sample 2. In each case, two sub-samples from the first sample and the only sub-sample from the second sample were processed by one analyst (An1 or An2), while the third sub-sample from sample one was processed by a different analyst (An1 or An2).

Fig. 2. Scatterplots of lake-averaged values of the seven phytoplankton metrics against $\log_{10}$ total phosphorus concentration.

Fig. 3. The proportion ($Prop_e$, equation 2) of the total among-lake variance in metric scores "explained" in top models, with the most optimal combination of environmental predictor variables. REML estimation used in model fitting.

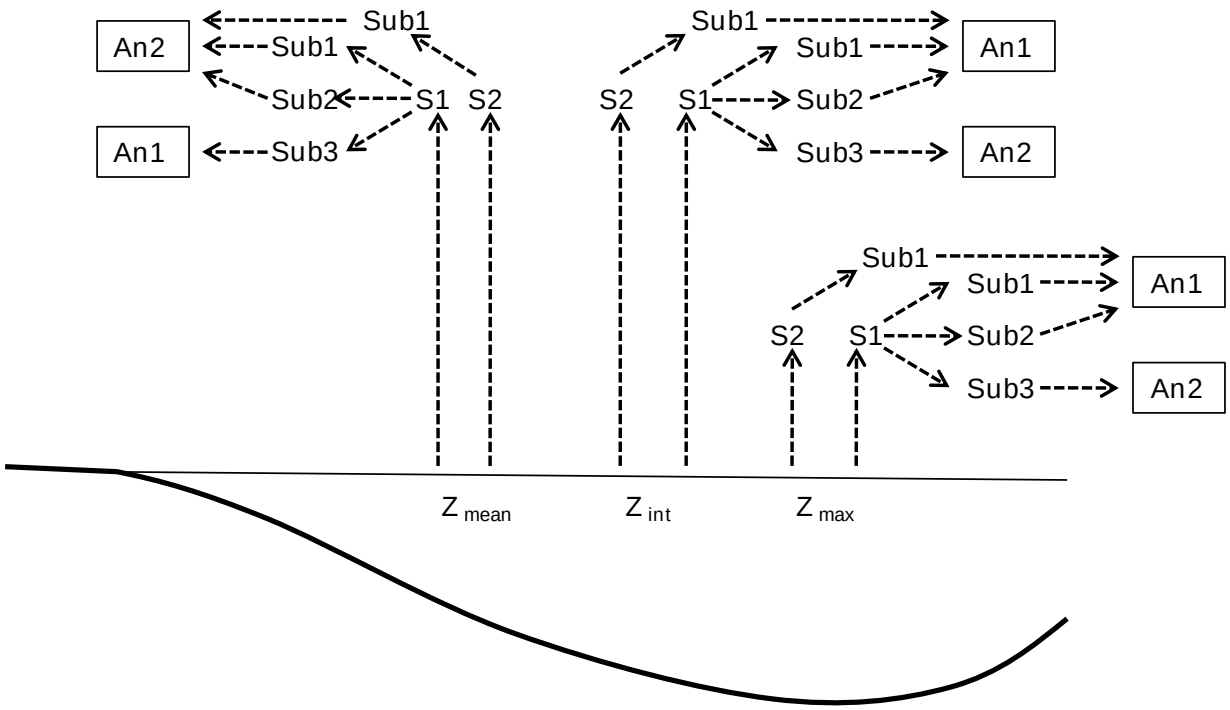
Fig. 4. Model-averaged slope parameters for the relationships between the modelled environmental predictors and the phytoplankton abundance metric ($\log_{10}$ Chl-*a* concentration). Filled circles indicate the model-averaged slope parameter estimate for each metric-predictor relationship, and whiskers indicate the 95% confidence interval for the estimate. Dashed horizontal line indicates zero. ML estimation used in model fitting.

Fig. 5. Model-averaged slope parameters for the relationships between the modelled environmental predictors and the four phytoplankton composition metrics. Filled circles indicate the model-averaged slope parameter estimate for each metric-predictor relationship, and whiskers indicate the 95% confidence interval for the estimate. Dashed horizontal line indicates zero. ML estimation used in model fitting.

Fig. 6. Model-averaged slope parameters for the relationships between the modelled environmental predictors and the two phytoplankton bloom metrics. Filled circles indicate the model-averaged slope parameter estimate for each metric-predictor relationship, and whiskers indicate the 95% confidence interval for the estimate. Dashed horizontal line indicates zero. ML estimation used in model fitting.

**Figure(s)**

Fig. 1



**Figure(s)**

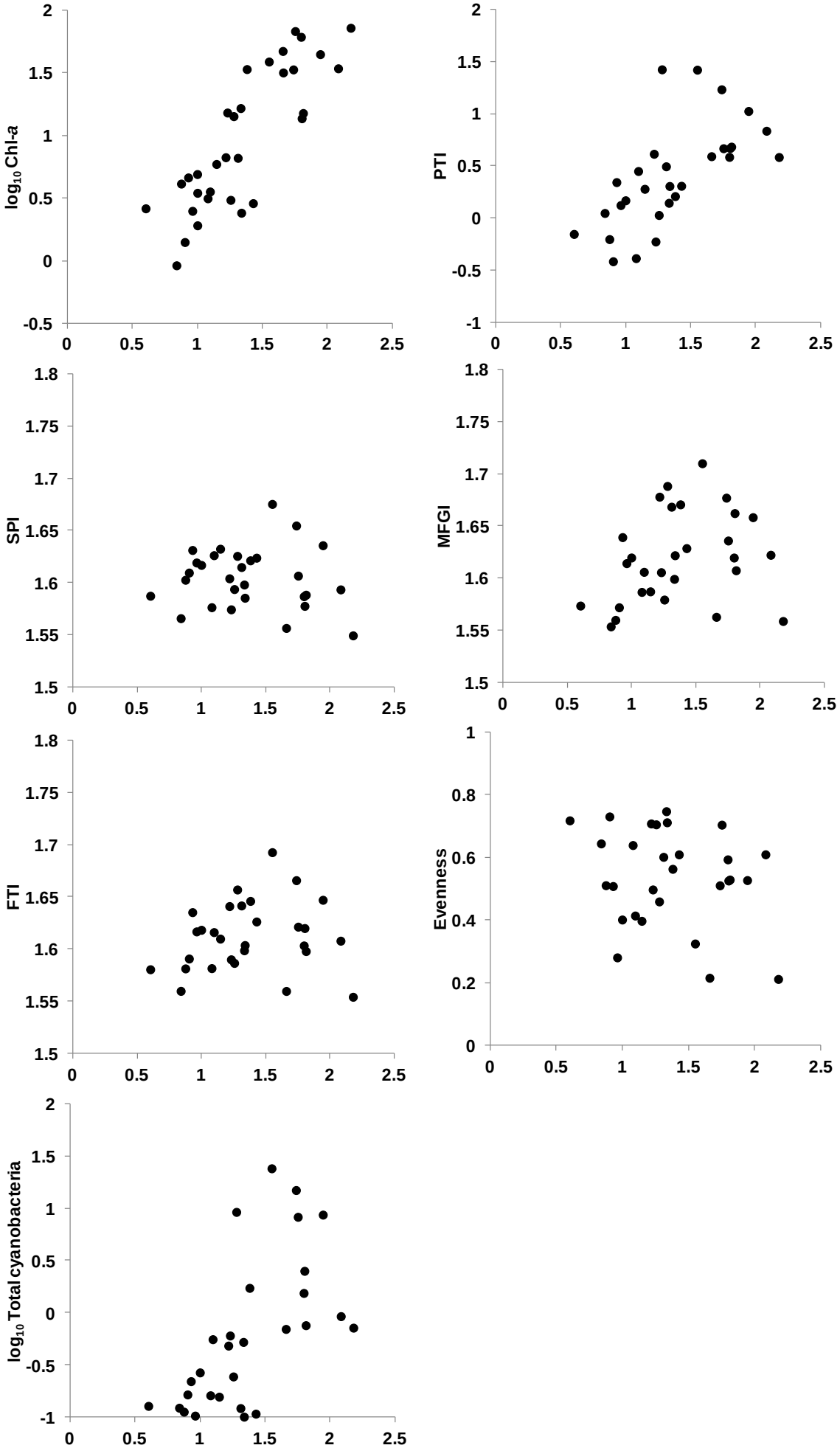
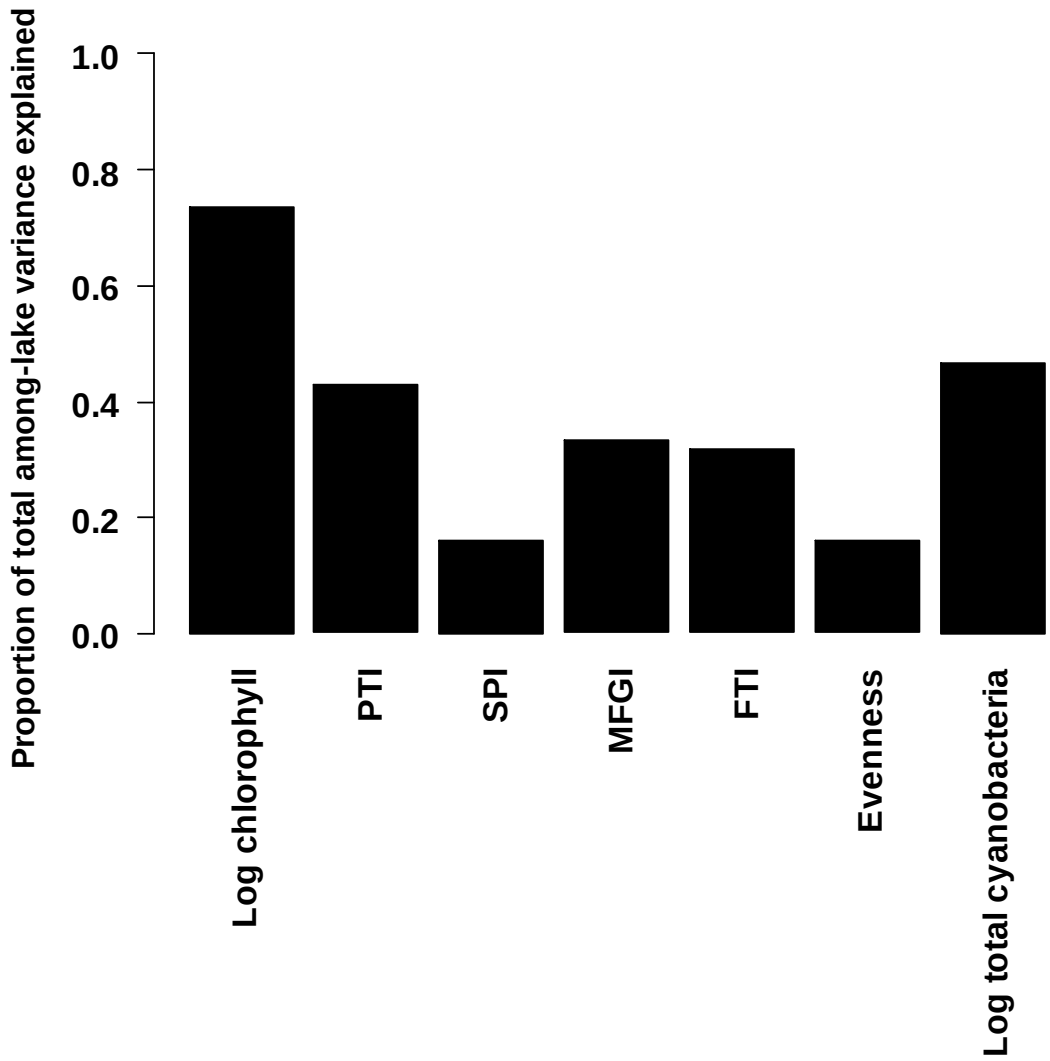



Fig. 2.

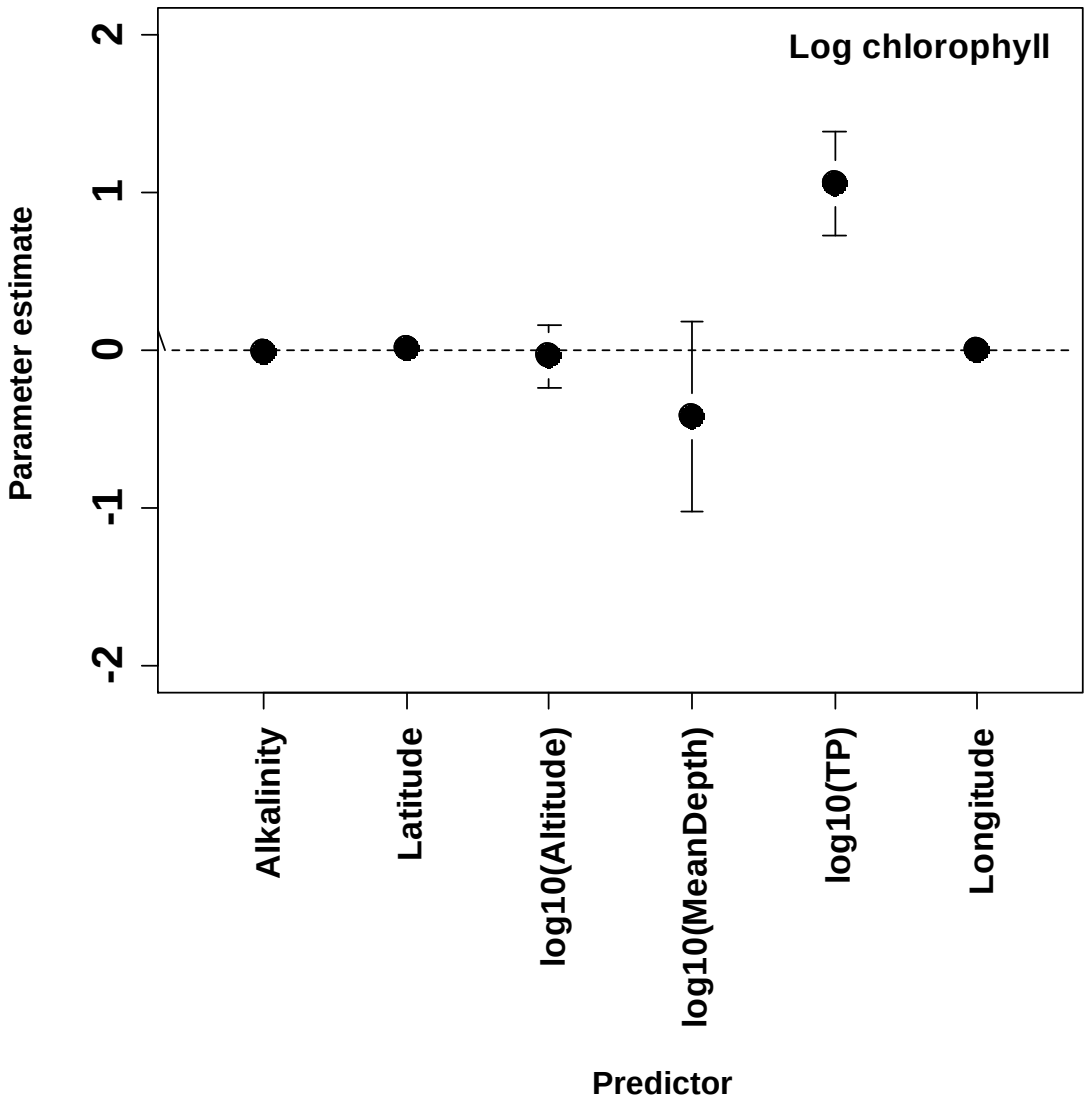
**Figure 3**

Fig. 3



**Figure(s)**

Fig 4



**Figure(s)**

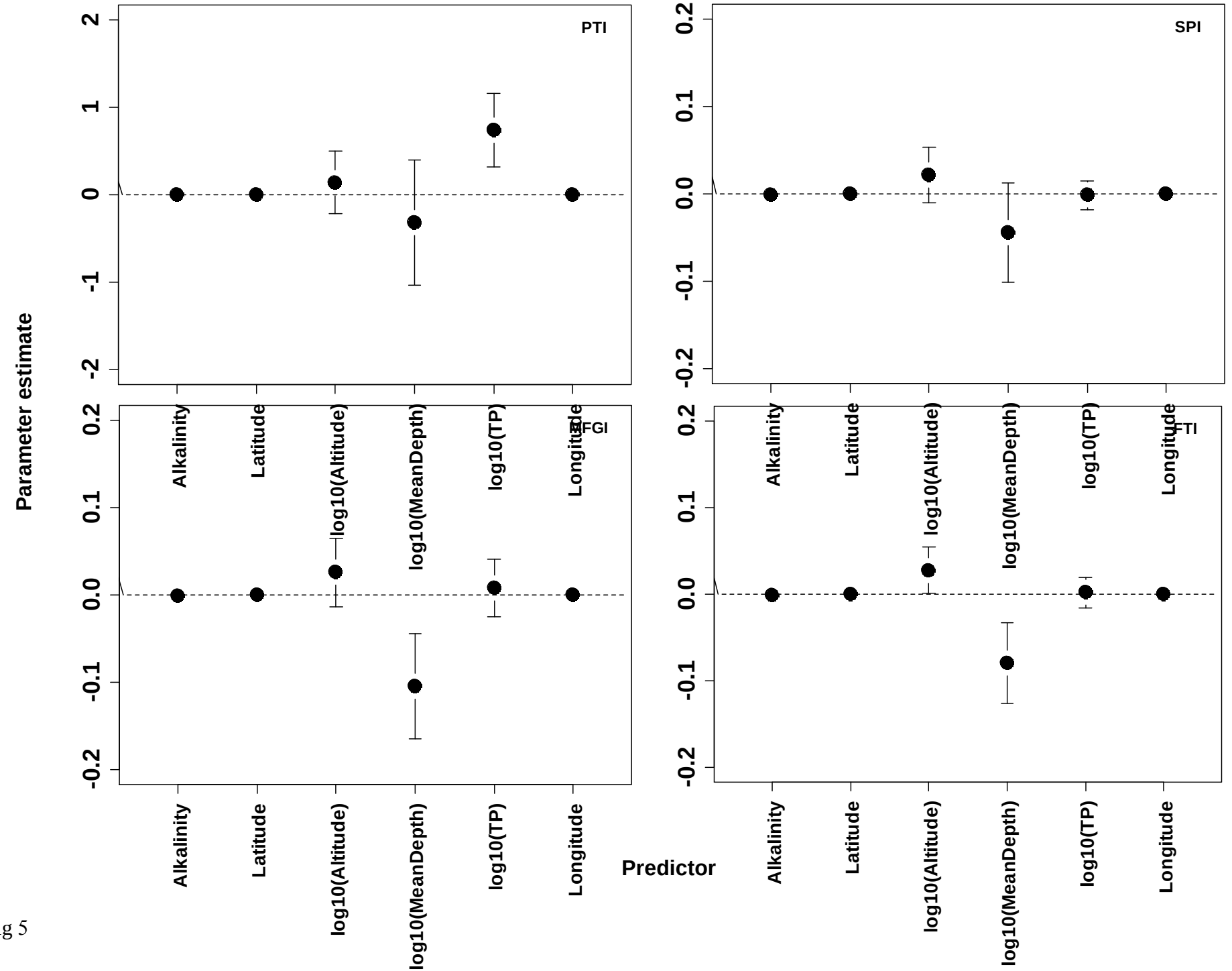
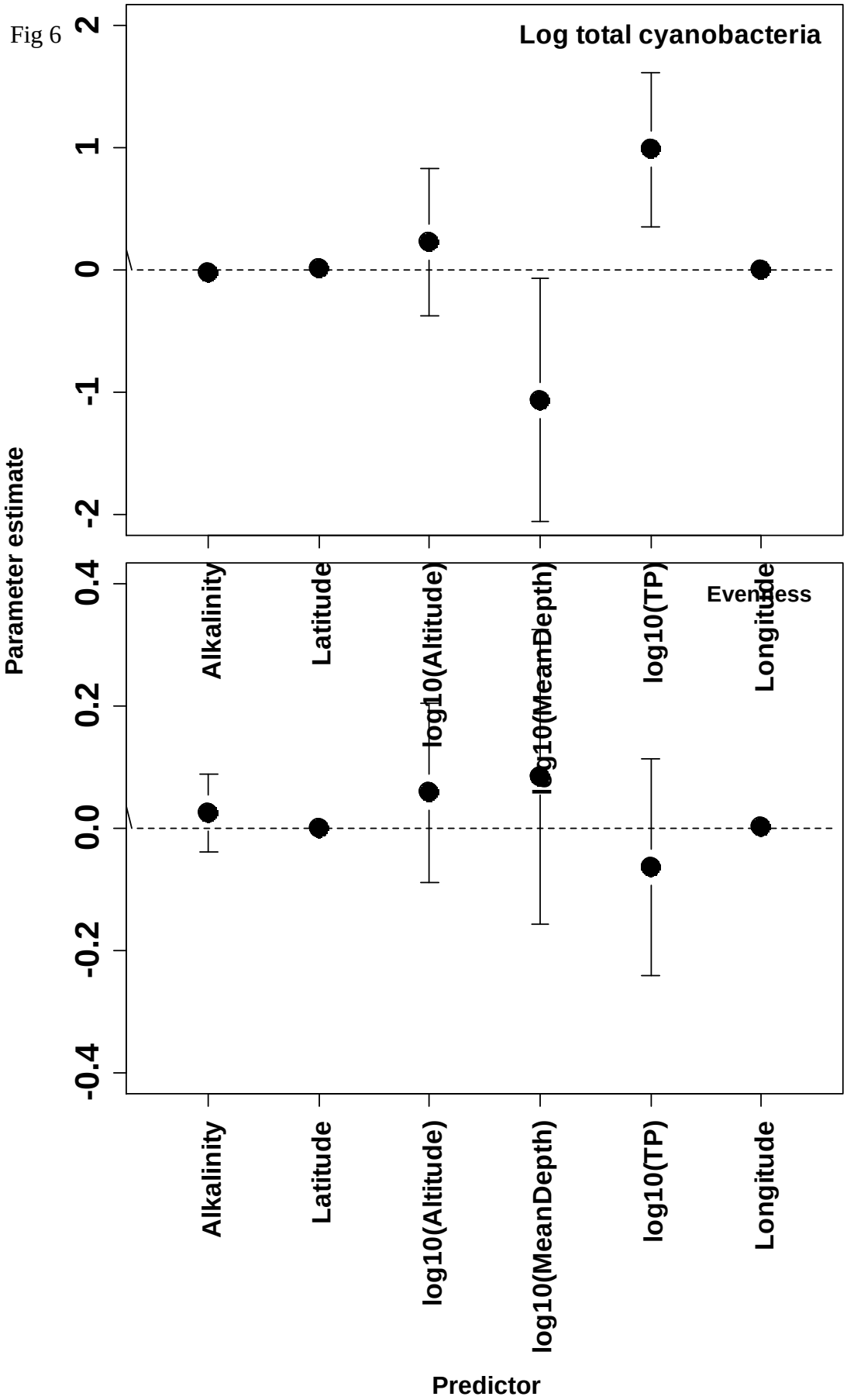



Fig 5

**Figure(s)**

Fig 6



## Supplementary information

Table S1. Model selection table for the total cyanobacterial biovolume metric. For all of the models in the top model set (ΔAIC≤4) the table includes estimates of the model intercept and slope parameters for relationships between the metric and alkalinity (Alk), latitude (Lat), $\log_{10}$ transformed altitude [$\log_{10}$(Alt)], $\log_{10}$ transformed mean lake depth [$\log_{10}$(MnD)], $\log_{10}$ transformed total phosphorus concentration [$\log_{10}$(TP)] and longitude (Long). Also shown are the number of parameters estimated in each model (k), the model deviance (Dev.), AIC, ΔAIC and Akaike weight (weight). Note that k includes the global intercept and parameters for both the fitted predictors and the random effects

| Intercept | Alk | Lat | $\log_{10}$(Alt) | $\log_{10}$(MnD) | $\log_{10}$(TP) | Long | k | Dev. | AIC | ΔAIC | weight |
|---|---|---|---|---|---|---|---|---|---|---|---|
| -0.217500 | | | | -1.205 | 0.8181 | | 6 | -239.6 | -227.6 | 0.0000 | 0.126 |
| -3.611000 | | 0.032610 | 0.556500 | -1.049 | 1.1270 | | 8 | -243.4 | -227.4 | 0.2413 | 0.111 |
| -0.675600 | | | 0.200400 | -1.240 | 0.8769 | | 7 | -240.6 | -226.6 | 1.0110 | 0.076 |
| -0.469300 | | | | -1.070 | 0.8397 | 0.0093050 | 7 | -240.4 | -226.4 | 1.1640 | 0.070 |
| -1.108000 | | | 0.249000 | -1.076 | 0.9188 | 0.0118800 | 8 | -242.0 | -226.0 | 1.6510 | 0.055 |
| -0.262500 | -0.05064 | | | -1.175 | 0.8972 | | 7 | -239.9 | -225.9 | 1.7340 | 0.053 |
| -3.529000 | -0.06749 | 0.029670 | 0.577900 | -1.036 | 1.2260 | | 9 | -243.8 | -225.8 | 1.8090 | 0.051 |
| -0.534600 | | 0.004875 | | -1.167 | 0.8399 | | 7 | -239.7 | -225.7 | 1.8740 | 0.049 |
| -0.953500 | -0.10280 | | 0.282000 | -1.194 | 1.0610 | | 8 | -241.6 | -225.6 | 2.0310 | 0.046 |
| -3.685000 | | 0.033770 | 0.565700 | -1.053 | 1.1330 | -0.0008116 | 9 | -243.4 | -225.4 | 2.2380 | 0.041 |
| -1.489000 | -0.12040 | | 0.350800 | -1.001 | 1.1400 | 0.0134100 | 9 | -243.3 | -225.3 | 2.2630 | 0.041 |
| -0.523000 | -0.05405 | | | -1.035 | 0.9245 | 0.0095170 | 8 | -240.8 | -224.8 | 2.8520 | 0.030 |
| -0.202100 | | -0.005112 | | -1.074 | 0.8224 | 0.0117200 | 8 | -240.5 | -224.5 | 3.0790 | 0.027 |
| -5.789000 | | 0.044360 | 0.650300 | | 1.4310 | | 7 | -238.2 | -224.2 | 3.4170 | 0.023 |
| -0.404500 | -0.04371 | 0.002278 | | -1.162 | 0.8966 | | 8 | -239.9 | -223.9 | 3.7110 | 0.020 |
| -3.123000 | -0.08095 | 0.023050 | 0.533700 | -1.010 | 1.2140 | 0.0042450 | 10 | -243.9 | -223.9 | 3.7380 | 0.019 |
| -1.853000 | | | | | 1.0780 | 0.0165800 | 6 | -235.7 | -223.7 | 3.9370 | 0.018 |

Table S2. Model selection table for the PTI metric. For all of the models in the top model set (ΔAIC≤4) the table includes estimates of the model intercept and slope parameters for relationships between the metric and alkalinity (Alk), latitude (Lat), $log_{10}$ transformed altitude [$log_{10}$(Alt)], $log_{10}$ transformed mean lake depth [$log_{10}$(MnD)], $log_{10}$ transformed total phosphorus concentration [$log_{10}$(TP)] and longitude (Long). Also shown are the number of parameters estimated in each model (k), the model deviance (Dev.), AIC, ΔAIC and Akaike weight (weight). Note that k includes the global intercept and parameters for both the fitted predictors and the random effects

| Intercept | Alk | Lat | $log_{10}$(Alt) | $log_{10}$(MnD) | $log_{10}$(TP) | Long | k | Dev. | AIC | ΔAIC | weight |
|---|---|---|---|---|---|---|---|---|---|---|---|
| -0.66800 | | | 0.26310 | -0.5137 | 0.7356 | | 7 | -226.0 | -212.0 | 0.0000 | 0.110 |
| -1.25700 | | | 0.24550 | | 0.8477 | | 6 | -223.4 | -211.4 | 0.5402 | 0.084 |
| 0.88010 | | -0.014520 | | -0.5802 | 0.5917 | | 7 | -224.9 | -210.9 | 1.1200 | 0.063 |
| -0.64120 | | | | | 0.7644 | | 5 | -220.8 | -210.8 | 1.1290 | 0.062 |
| -0.06112 | | | | -0.4702 | 0.6561 | | 6 | -222.8 | -210.8 | 1.1870 | 0.061 |
| -0.49590 | | | 0.24390 | -0.5779 | 0.7180 | -0.0047540 | 8 | -226.4 | -210.4 | 1.5860 | 0.050 |
| -0.75110 | -0.030910 | | 0.28770 | -0.4996 | 0.7904 | | 8 | -226.1 | -210.1 | 1.8300 | 0.044 |
| -0.40760 | | -0.002898 | 0.23160 | -0.5304 | 0.7133 | | 8 | -226.0 | -210.0 | 1.9600 | 0.041 |
| -0.05729 | | -0.010530 | | | 0.7361 | | 6 | -221.9 | -209.9 | 2.0710 | 0.039 |
| -1.35200 | -0.043450 | | 0.28080 | | 0.9205 | | 7 | -223.7 | -209.7 | 2.2280 | 0.036 |
| 0.13480 | | | | -0.5737 | 0.6381 | -0.0072750 | 7 | -223.7 | -209.7 | 2.2650 | 0.035 |
| -1.51300 | | 0.003077 | 0.27950 | | 0.8675 | | 7 | -223.5 | -209.5 | 2.4960 | 0.032 |
| -1.24000 | | | 0.24160 | | 0.8471 | -0.0008480 | 7 | -223.4 | -209.4 | 2.5270 | 0.031 |
| -0.60960 | | | | | 0.7672 | -0.0033800 | 6 | -221.0 | -209.0 | 2.9250 | 0.025 |
| 0.95630 | -0.025450 | -0.016030 | | -0.5767 | 0.6241 | | 8 | -225.0 | -209.0 | 3.0140 | 0.024 |
| -0.04169 | 0.022730 | | | -0.4835 | 0.6212 | | 7 | -222.9 | -208.9 | 3.0900 | 0.023 |
| -0.63990 | 0.009379 | | | | 0.7512 | | 6 | -220.9 | -208.9 | 3.1140 | 0.023 |
| 0.85900 | | -0.013880 | | -0.5861 | 0.5927 | -0.0007501 | 8 | -224.9 | -208.9 | 3.1140 | 0.023 |
| -1.12800 | | 0.008274 | 0.32130 | -0.5718 | 0.7704 | -0.0078400 | 9 | -226.5 | -208.5 | 3.4300 | 0.020 |
| -0.57480 | -0.024900 | | 0.26510 | -0.5620 | 0.7633 | -0.0044240 | 9 | -226.5 | -208.5 | 3.4760 | 0.019 |
| -0.36710 | -0.035990 | -0.004426 | 0.24370 | -0.5228 | 0.7653 | | 9 | -226.2 | -208.2 | 3.7410 | 0.017 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.03818 | -0.029760 | -0.012320 | | 0.7730 | | 7 | -222.0 | -208.0 | 3.9380 | 0.015 |
| 0.06028 | | -0.013130 | | 0.7267 | 0.0028800 | 7 | -222.0 | -208.0 | 3.9840 | 0.015 |

Table S3. Model selection table for the SPI metric. For all of the models in the top model set (ΔAIC≤4) the table includes estimates of the model intercept and slope parameters for relationships between the metric and alkalinity (Alk), latitude (Lat), $log_{10}$ transformed altitude [$log_{10}$(Alt)], $log_{10}$ transformed mean lake depth [$log_{10}$(MnD)], $log_{10}$ transformed total phosphorus concentration [$log_{10}$(TP)] and longitude (Long). Also shown are the number of parameters estimated in each model (k), the model deviance (Dev.), AIC, ΔAIC and Akaike weight (weight). Note that k includes the global intercept and parameters for both the fitted predictors and the random effects

| Intercept | Alk | Lat | $log_{10}$(Alt) | $log_{10}$(MnD) | $log_{10}$(TP) | Long | k | Dev. | AIC | ΔAIC | weight |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1.601 | | | 0.02449 | -0.05126 | | | 6 | -1819 | -1807 | 0.0000 | 0.116 |
| 1.608 | -0.0054800 | | 0.02691 | -0.05396 | | | 7 | -1821 | -1807 | 0.8761 | 0.075 |
| 1.545 | | 0.0007997 | 0.03161 | -0.05068 | | | 7 | -1820 | -1806 | 1.4270 | 0.057 |
| 1.620 | | | 0.02279 | -0.05591 | -0.0078440 | | 7 | -1820 | -1806 | 1.7430 | 0.049 |
| 1.600 | | | 0.02456 | -0.05105 | | 2.196e-05 | 7 | -1819 | -1805 | 1.9990 | 0.043 |
| 1.643 | | | | -0.04271 | | | 5 | -1815 | -1805 | 2.0930 | 0.041 |
| 1.563 | | | 0.01998 | | | | 5 | -1815 | -1805 | 2.4330 | 0.035 |
| 1.580 | -0.0046510 | 0.0003839 | 0.02996 | -0.05328 | | | 8 | -1821 | -1805 | 2.7670 | 0.029 |
| 1.604 | -0.0058700 | | 0.02753 | -0.05294 | 0.0020430 | | 8 | -1821 | -1805 | 2.8640 | 0.028 |
| 1.607 | -0.0054870 | | 0.02702 | -0.05362 | | 3.663e-05 | 8 | -1821 | -1805 | 2.8730 | 0.028 |
| 1.604 | | | | | | | 4 | -1813 | -1805 | 2.9410 | 0.027 |
| 1.510 | | 0.0014890 | 0.03605 | -0.05569 | | -5.849e-04 | 8 | -1821 | -1805 | 2.9590 | 0.027 |
| 1.675 | | | | -0.05329 | -0.0157700 | | 6 | -1816 | -1804 | 3.0890 | 0.025 |
| 1.556 | | 0.0007181 | 0.03040 | -0.05206 | -0.0022230 | | 8 | -1820 | -1804 | 3.4120 | 0.021 |
| 1.676 | | -0.0005986 | | -0.04501 | | | 6 | -1816 | -1804 | 3.6250 | 0.019 |
| 1.650 | -0.0032260 | | | -0.04385 | | | 6 | -1816 | -1804 | 3.7400 | 0.018 |
| 1.621 | | | 0.02266 | -0.05632 | -0.0079950 | -3.464e-05 | 8 | -1820 | -1804 | 3.7400 | 0.018 |
| 1.502 | | 0.0008742 | 0.02789 | | | | 6 | -1816 | -1804 | 3.8330 | 0.017 |
| 1.567 | -0.0041300 | | 0.02163 | | | | 6 | -1816 | -1804 | 3.8730 | 0.017 |

Table S4. Model selection table for the MFGI metric. For all of the models in the top model set (ΔAIC≤4) the table includes estimates of the model intercept and slope parameters for relationships between the metric and alkalinity (Alk), latitude (Lat), $log_{10}$ transformed altitude [$log_{10}$(Alt)], $log_{10}$ transformed mean lake depth [$log_{10}$(MnD)], $log_{10}$ transformed total phosphorus concentration [$log_{10}$(TP)] and longitude (Long). Also shown are the number of parameters estimated in each model (k), the model deviance (Dev.), AIC, ΔAIC and Akaike weight (weight). Note that k includes the global intercept and parameters for both the fitted predictors and the random effects

| Intercept | Alk | Lat | $log_{10}$(Alt) | $log_{10}$(MnD) | $log_{10}$(TP) | Long | k | Dev. | AIC | ΔAIC | weight |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1.650 | | | 0.03321 | -0.10740 | | | 6 | -1846 | -1834 | 0.0000 | 0.118 |
| 1.600 | | | 0.03714 | -0.09385 | 0.021560 | | 7 | -1847 | -1833 | 0.5446 | 0.090 |
| 1.671 | | | 0.03058 | -0.11610 | | -7.710e-04 | 7 | -1847 | -1833 | 0.8187 | 0.078 |
| 1.822 | | -0.0020610 | | -0.10420 | | | 6 | -1845 | -1833 | 0.9816 | 0.072 |
| 1.725 | | -0.0010390 | 0.02337 | -0.10840 | | | 7 | -1846 | -1832 | 1.2540 | 0.063 |
| 1.624 | | | 0.03445 | -0.10280 | 0.019110 | -6.606e-04 | 8 | -1848 | -1832 | 1.6610 | 0.051 |
| 1.583 | -6.414e-03 | | 0.04223 | -0.09091 | 0.032940 | | 8 | -1848 | -1832 | 1.7350 | 0.049 |
| 1.650 | -6.668e-05 | | 0.03324 | -0.10750 | | | 7 | -1846 | -1832 | 2.0000 | 0.043 |
| 1.636 | | -0.0004012 | 0.03278 | -0.09617 | 0.018470 | | 8 | -1847 | -1831 | 2.4600 | 0.034 |
| 1.849 | -4.271e-03 | -0.0023810 | | -0.10770 | | | 7 | -1845 | -1831 | 2.5730 | 0.033 |
| 1.687 | | -0.0002624 | 0.02846 | -0.11510 | | -6.634e-04 | 8 | -1847 | -1831 | 2.7930 | 0.029 |
| 1.671 | 6.872e-05 | | 0.03054 | -0.11600 | | -7.712e-04 | 8 | -1847 | -1831 | 2.8190 | 0.029 |
| 1.820 | | -0.0019910 | | -0.10490 | | -8.399e-05 | 7 | -1845 | -1831 | 2.9730 | 0.027 |
| 1.818 | | -0.0020460 | | -0.10320 | 0.001236 | | 7 | -1845 | -1831 | 2.9770 | 0.027 |
| 1.606 | -5.615e-03 | | 0.03921 | -0.09921 | 0.029330 | -5.866e-04 | 9 | -1849 | -1831 | 3.0360 | 0.026 |
| 1.749 | -3.032e-03 | -0.0013200 | 0.02214 | -0.11070 | | | 8 | -1847 | -1831 | 3.0430 | 0.026 |
| 1.536 | | 0.0011490 | 0.04518 | -0.10190 | 0.026380 | -1.089e-03 | 9 | -1848 | -1830 | 3.3260 | 0.022 |
| 1.644 | -7.229e-03 | -0.0007101 | 0.03517 | -0.09464 | 0.028900 | | 9 | -1848 | -1830 | 3.4770 | 0.021 |
| 1.706 | | | | -0.09485 | | | 5 | -1840 | -1830 | 3.4790 | 0.021 |
| 1.729 | | | | -0.10800 | | -1.045e-03 | 6 | -1842 | -1830 | 3.5600 | 0.020 |

Table S5. Model selection table for the FTI metric. For all of the models in the top model set (ΔAIC≤4) the table includes estimates of the model intercept and slope parameters for relationships between the metric and alkalinity (Alk), latitude (Lat), $log_{10}$ transformed altitude [$log_{10}$(Alt)], $log_{10}$ transformed mean lake depth [$log_{10}$(MnD)], $log_{10}$ transformed total phosphorus concentration [$log_{10}$(TP)] and longitude (Long). Also shown are the number of parameters estimated in each model (k), the model deviance (Dev.), AIC, ΔAIC and Akaike weight (weight). Note that k includes the global intercept and parameters for both the fitted predictors and the random effects

| Intercept | Alk | Lat | $log_{10}$(Alt) | $log_{10}$(MnD) | $log_{10}$(TP) | Long | k | Dev. | AIC | ΔAIC | weight |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1.627 | | | 0.02841 | -0.07941 | | | 6 | -1990 | -1978 | 0.000 | 0.188 |
| 1.637 | | | 0.02715 | -0.08351 | | -3.894e-04 | 7 | -1990 | -1976 | 1.546 | 0.087 |
| 1.630 | -0.0027980 | | 0.02969 | -0.08108 | | | 7 | -1990 | -1976 | 1.663 | 0.082 |
| 1.612 | | | 0.02962 | -0.07558 | 0.006206 | | 7 | -1990 | -1976 | 1.817 | 0.076 |
| 1.635 | | -1.238e-04 | 0.02727 | -0.07951 | | | 7 | -1990 | -1976 | 1.984 | 0.070 |
| 1.596 | -0.0059950 | | 0.03440 | -0.07274 | 0.016640 | | 8 | -1991 | -1975 | 2.770 | 0.047 |
| 1.641 | -0.0027290 | | 0.02842 | -0.08506 | | -3.824e-04 | 8 | -1990 | -1974 | 3.220 | 0.038 |
| 1.598 | | 6.404e-04 | 0.03223 | -0.08572 | | -6.513e-04 | 8 | -1990 | -1974 | 3.320 | 0.036 |
| 1.625 | | | 0.02819 | -0.08024 | 0.004789 | -3.594e-04 | 8 | -1990 | -1974 | 3.438 | 0.034 |
| 1.749 | | -1.322e-03 | | -0.07458 | | | 6 | -1986 | -1974 | 3.468 | 0.033 |
| 1.666 | -0.0038550 | -4.771e-04 | 0.02576 | -0.08211 | | | 8 | -1990 | -1974 | 3.472 | 0.033 |
| 1.600 | | 1.320e-04 | 0.03104 | -0.07484 | 0.007230 | | 8 | -1990 | -1974 | 3.804 | 0.028 |

Table S6. Model selection table for the evenness metric. For all of the models in the top model set (ΔAIC≤4) the table includes estimates of the model intercept and slope parameters for relationships between the metric and alkalinity (Alk), latitude (Lat), $log_{10}$ transformed altitude [$log_{10}$(Alt)], $log_{10}$ transformed mean lake depth [$log_{10}$(MnD)], $log_{10}$ transformed total phosphorus concentration [$log_{10}$(TP)] and longitude (Long). Also shown are the number of parameters estimated in each model (k), the model deviance (Dev.), AIC, ΔAIC and Akaike weight (weight). Note that k includes the global intercept and parameters for both the fitted predictors and the random effects

| Intercept | Alk | Lat | $log_{10}$(Alt) | $log_{10}$(MnD) | $log_{10}$(TP) | Long | k | Dev. | AIC | ΔAIC | weight |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.669200 | 0.06012 | | | | -0.179600 | | 6 | -782.4 | -770.4 | 0.000 | 0.063 |
| 0.305700 | | | 0.10990 | | | | 5 | -779.4 | -769.4 | 0.974 | 0.039 |
| 0.529300 | 0.05659 | | | 0.1128 | -0.148900 | | 7 | -783.2 | -769.2 | 1.144 | 0.036 |
| 0.186300 | | | 0.09657 | 0.1587 | | | 6 | -781.2 | -769.2 | 1.169 | 0.035 |
| 0.081130 | | | 0.10890 | 0.1989 | | 0.0042190 | 7 | -783.1 | -769.1 | 1.237 | 0.034 |
| 0.540200 | 0.05079 | | 0.05056 | | -0.148900 | | 7 | -783.0 | -769.0 | 1.368 | 0.032 |
| 0.652400 | 0.05984 | | | | -0.179100 | 0.0016430 | 7 | -782.7 | -768.7 | 1.674 | 0.027 |
| 0.353900 | | | | 0.1918 | | | 5 | -778.5 | -768.5 | 1.843 | 0.025 |
| -0.483900 | 0.04348 | 0.0086550 | 0.15470 | 0.1870 | | | 8 | -784.5 | -768.5 | 1.874 | 0.025 |
| 0.648300 | 0.06123 | 0.0003777 | | | -0.180000 | | 7 | -782.4 | -768.4 | 1.993 | 0.023 |
| 0.251800 | | | 0.12130 | | | 0.0029950 | 6 | -780.4 | -768.4 | 2.007 | 0.023 |
| 0.435500 | | | 0.09208 | | -0.068200 | | 6 | -780.3 | -768.3 | 2.099 | 0.022 |
| 0.152600 | 0.02477 | | 0.08573 | 0.1714 | | | 7 | -782.2 | -768.2 | 2.223 | 0.021 |
| -0.012160 | | 0.0045240 | 0.15090 | | | | 6 | -780.1 | -768.1 | 2.256 | 0.020 |
| 0.531100 | | | | | | | 4 | -776.1 | -768.1 | 2.291 | 0.020 |
| 0.050390 | 0.02400 | | 0.09814 | 0.2104 | | 0.0041480 | 8 | -784.1 | -768.1 | 2.309 | 0.020 |
| 0.285900 | 0.03196 | | | 0.2037 | | | 6 | -780.1 | -768.1 | 2.322 | 0.020 |
| 0.670300 | | | | | -0.101800 | | 5 | -778.1 | -768.1 | 2.326 | 0.020 |
| -0.150100 | | 0.0047600 | 0.13910 | 0.1622 | | | 7 | -782.1 | -768.1 | 2.327 | 0.020 |
| 0.459600 | 0.05508 | | | 0.1469 | -0.138600 | 0.0026590 | 8 | -784.0 | -768.0 | 2.342 | 0.020 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.285600 | 0.02052 | | 0.10180 | | | | 6 | -780.0 | -768.0 | 2.356 | 0.019 |
| 0.400800 | 0.04730 | | 0.05041 | 0.1126 | -0.118300 | | 8 | -783.9 | -767.9 | 2.498 | 0.018 |
| -0.281000 | 0.03720 | 0.0078380 | 0.16600 | | | | 7 | -781.8 | -767.8 | 2.554 | 0.018 |
| 0.129100 | 0.05660 | 0.0051090 | 0.10070 | | -0.124100 | | 8 | -783.7 | -767.7 | 2.631 | 0.017 |
| 0.286200 | | | | 0.2281 | | 0.0033640 | 6 | -779.7 | -767.7 | 2.692 | 0.016 |
| 0.487200 | 0.04827 | | 0.06215 | | -0.141100 | 0.0022780 | 8 | -783.6 | -767.6 | 2.762 | 0.016 |
| 0.276500 | | | 0.08840 | 0.1360 | -0.038340 | | 7 | -781.5 | -767.5 | 2.916 | 0.015 |
| 0.605300 | | -0.0069620 | | 0.2361 | | 0.0066400 | 7 | -781.4 | -767.4 | 2.963 | 0.014 |
| 0.489400 | | | | 0.1463 | -0.068350 | | 6 | -779.3 | -767.3 | 3.043 | 0.014 |
| 0.460500 | 0.05971 | 0.0011180 | | 0.1183 | -0.148600 | | 8 | -783.3 | -767.3 | 3.080 | 0.014 |
| 0.217400 | 0.03219 | | | 0.2401 | | 0.0033880 | 7 | -781.3 | -767.3 | 3.103 | 0.013 |
| 0.133400 | | | 0.10410 | 0.1853 | -0.020580 | 0.0040710 | 8 | -783.2 | -767.2 | 3.164 | 0.013 |
| 0.480400 | 0.02832 | | | | | | 5 | -777.2 | -767.2 | 3.185 | 0.013 |
| 0.267900 | 0.04219 | | 0.06759 | 0.1561 | -0.094730 | 0.0034100 | 9 | -785.2 | -767.2 | 3.226 | 0.013 |
| 0.376500 | | | 0.10390 | | -0.063650 | 0.0028090 | 7 | -781.1 | -767.1 | 3.231 | 0.013 |
| 0.105300 | | -0.0003993 | 0.10580 | 0.2002 | | 0.0043820 | 8 | -783.1 | -767.1 | 3.234 | 0.013 |
| -0.184900 | 0.05396 | 0.0067760 | 0.11680 | 0.1455 | -0.076390 | | 9 | -785.1 | -767.1 | 3.235 | 0.013 |
| 0.235000 | 0.01929 | | 0.11320 | | | 0.0028780 | 7 | -780.9 | -766.9 | 3.451 | 0.011 |
| 0.802000 | 0.05088 | -0.0029590 | | | -0.175300 | 0.0030260 | 8 | -782.9 | -766.9 | 3.456 | 0.011 |
| 1.076000 | | -0.0083670 | | | -0.124000 | 0.0056530 | 7 | -780.7 | -766.7 | 3.634 | 0.010 |
| 0.852500 | | -0.0032900 | | | -0.110800 | | 6 | -778.7 | -766.7 | 3.724 | 0.010 |
| 0.429800 | | -0.0013720 | | 0.1865 | | | 6 | -778.6 | -766.6 | 3.737 | 0.010 |
| 0.829300 | | -0.0080920 | | 0.1774 | -0.082060 | 0.0067340 | 8 | -782.6 | -766.6 | 3.775 | 0.010 |
| -0.363100 | 0.03863 | 0.0065310 | 0.14210 | 0.1967 | | 0.0014360 | 9 | -784.6 | -766.6 | 3.782 | 0.010 |
| 0.208100 | | 0.0027440 | 0.12160 | | -0.049980 | | 7 | -780.5 | -766.5 | 3.893 | 0.009 |
| 0.138800 | | 0.0017850 | 0.13480 | | | 0.0023020 | 7 | -780.4 | -766.4 | 3.946 | 0.009 |
| 0.512700 | | | | | | 0.0017980 | 5 | -776.4 | -766.4 | 3.965 | 0.009 |
| 0.652000 | | | | | -0.101800 | 0.0017890 | 6 | -778.4 | -766.4 | 3.982 | 0.009 |

Table S7. Model selection table for the Chl-*a* metric. For all of the models in the top model set (ΔAIC≤4) the table includes estimates of the model intercept and slope parameters for relationships between the metric and alkalinity (Alk), latitude (Lat), $\log_{10}$ transformed altitude [$\log_{10}$(Alt)], $\log_{10}$ transformed mean lake depth [$\log_{10}$(MnD)], $\log_{10}$ transformed total phosphorus concentration [$\log_{10}$(TP)] and longitude (Long). Also shown are the number of parameters estimated in each model (k), the model deviance (Dev.), AIC, ΔAIC and Akaike weight (weight). Note that k includes the global intercept and parameters for both the fitted predictors and the random effects

| Intercept | Alk | Lat | $\log_{10}$(Alt) | $\log_{10}$(MnD) | $\log_{10}$(TP) | Long | k | Dev. | AIC | ΔAIC | weight |
|---|---|---|---|---|---|---|---|---|---|---|---|
| -0.648300 | | 0.0124400 | | -0.4843 | 1.0460 | | 6 | -222.5 | -210.5 | 0.0000 | 0.124 |
| 0.532100 | | | -0.162300 | -0.5542 | 0.9437 | | 6 | -221.7 | -209.7 | 0.7108 | 0.087 |
| 0.167200 | | | | -0.5853 | 0.9887 | | 5 | -219.6 | -209.6 | 0.8830 | 0.080 |
| 0.106200 | -0.06437 | | | -0.5471 | 1.0920 | | 6 | -221.0 | -209.0 | 1.4330 | 0.061 |
| -1.428000 | | 0.0157700 | | | 1.1650 | | 5 | -218.8 | -208.8 | 1.6200 | 0.055 |
| -0.763700 | | 0.0159000 | | -0.5175 | 1.0530 | -0.0040550 | 7 | -222.8 | -208.8 | 1.6570 | 0.054 |
| -0.554600 | -0.03175 | 0.0105500 | | -0.4808 | 1.0880 | | 7 | -222.8 | -208.8 | 1.6890 | 0.053 |
| -0.344800 | | 0.0097090 | -0.055380 | -0.4958 | 1.0180 | | 7 | -222.6 | -208.6 | 1.8770 | 0.049 |
| 0.422000 | -0.04038 | | -0.130400 | -0.5363 | 1.0170 | | 7 | -222.3 | -208.3 | 2.1940 | 0.041 |
| 0.070150 | | | | -0.5320 | 0.9965 | 0.0035280 | 6 | -220.0 | -208.0 | 2.4850 | 0.036 |
| 0.461200 | | | -0.154300 | -0.5266 | 0.9502 | 0.0019240 | 7 | -221.9 | -207.9 | 2.5900 | 0.034 |
| 0.002016 | -0.06553 | | | -0.4898 | 1.1020 | 0.0037470 | 7 | -221.5 | -207.5 | 2.9610 | 0.028 |
| -1.320000 | -0.03467 | 0.0136800 | | | 1.2100 | | 6 | -219.2 | -207.2 | 3.2930 | 0.024 |
| -0.095200 | | | -0.183700 | | 1.0630 | | 5 | -217.0 | -207.0 | 3.4750 | 0.022 |
| -0.668900 | -0.02120 | 0.0136700 | | -0.5058 | 1.0790 | -0.0029210 | 8 | -222.9 | -206.9 | 3.5440 | 0.021 |
| -0.305200 | -0.02995 | 0.0083660 | -0.046480 | -0.4906 | 1.0620 | | 8 | -222.9 | -206.9 | 3.6020 | 0.020 |
| -1.465000 | | 0.0165700 | | | 1.1690 | -0.0008813 | 6 | -218.8 | -206.8 | 3.6050 | 0.020 |
| -1.356000 | | 0.0151000 | -0.014040 | | 1.1590 | | 6 | -218.8 | -206.8 | 3.6130 | 0.020 |
| -0.691000 | | 0.0150900 | -0.011900 | -0.5178 | 1.0460 | -0.0037920 | 8 | -222.8 | -206.8 | 3.6520 | 0.020 |
| -0.551600 | | | | | 1.1220 | | 4 | -214.6 | -206.6 | 3.8430 | 0.018 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| -0.568800 | -0.07928 | | | 1.2380 | | 5 | -216.5 | -206.5 | 3.9470 | 0.017 |
| 0.323300 | -0.04346 | -0.117700 | -0.4999 | 1.0310 | 0.0024500 | 8 | -222.5 | -206.5 | 3.9970 | 0.017 |