Accepted refereed manuscript of: Kelly MG, Phillips G, Juggins S & Willby NJ (2020) Re-evaluating expectations for river phytobenthos assessment and understanding the relationship with macrophytes. *Ecological Indicators*, 117, Art. No.: 106582. https://doi.org/10.1016/j.ecolind.2020.106582

Re-evaluating expectations for river phytobenthos assessment and understanding the relationship with macrophytes

M.G. Kelly[1,2 *], G. Phillips[3], S. Juggins[4] & N.J. Willby[3]

[1] Bowburn Consultancy, 11 Monteigne Drive, Bowburn, Durham DH6 5QB, UK

[2] School of Geography, University of Nottingham, Nottingham NG7 2RD, UK

[3] Department of Biological and Environmental Sciences, University of Stirling, Stirling, FK9 4LA, UK

[4] School of Geography, Politics and Sociology, Newcastle University, Newcastle NE1 7RU

* Corresponding author: MGKelly@bowburn-consultancy.co.uk

Abstract

The reference model underlying the UK phytobenthos (diatom) tool for Water Framework Directive assessments is revisited and a new approach is proposed which uses quantile regression to predict the lowest values of the Trophic Diatom Index (equating to the best available condition) at any level of alkalinity to be predicted. Whilst a reference model based on least disturbed or minimally impacted conditions would be preferable in theory, in practice the absence lowland high alkalinity streams in a minimally impacted condition in the UK precludes the use of these approaches. Having proposed a revised reference model for phytobenthos, we then go on to examine the relationship between phytobenthos and macrophytes. These two groups respond to nutrients and other stressors in different ways with phytobenthos being more sensitive to nutrients whilst macrophytes better reflect the extent to which secondary effects are likely. We argue that averaging the two sub-elements of the "macrophytes and phytobenthos" biological quality element is a more realistic option than the current approach of taking the lower of the two assessments. It is, however, possible, to predict the value of the combined quality element from either sub-element, though we recognize that this also risks misclassifications.

Keywords: diatoms, phytobenthos, reference concept, Water Framework Directive, macrophytes, rivers

1. Introduction

The central objective of managing Europe's aquatic environment is to ensure sustainable water resources (Water Framework Directive (WFD) Article 1; European Union, 2000). "Sustainable" is a difficult notion to encapsulate in objective terms so a practical understanding has evolved that assumes that water bodies in a natural or near natural state possess sufficient resilience to ensure availability of the resource to future generations. The WFD therefore sets a target ("good ecological status") for water bodies to show no more than slight differences from their physical, chemical and biological condition at the natural or near natural state.

The concept of "good ecological status" raises two philosophical challenges, along with a host of methodological issues related both to its measurement (Borja & Rodríguez, 2010; Birk et al., 2012; Kelly, 2013) and to the harmonization of concepts between 28 (now 27) countries of varying biogeographical characteristics operating within a federal union (Birk et al., 2013; Poikane et al., 2015). The philosophical challenges are, first, achieving a practical understanding of the natural or near natural state and, second, developing a meaningful understanding of a "slight change" from this state that ensures that key structural and functional aspects are protected. The first of these, a workable notion of the natural or near-natural state, so-called 'reference conditions', is the subject of this paper.

An understanding of 'reference conditions' firstly requires an appreciation that there are natural differences in ecological communities that relate to landscape and climatic factors. It also has to recognize that the European environment is constantly changing in geological time (Flannery, 2018) and that, for the past seven millennia, also shows evidence of alteration due to human activities (Behre, 1988). This alteration is profound, but also varies geographically in its severity. It is against this reality that we need to establish a reference point against which future change is measured. In some water body types (e.g. lakes) it is possible, in theory at least, to use sediment records to establish historical baselines that could serve as reference points (Bennion et al., 2004); however, such opportunities are rare and a more profitable option is to seek out contemporary locations which, by way of low population density and an absence of significant human activity in the catchment, are close to their natural state (Stoddard et al., 2006; Pardo et al., 2012).

Stoddard et al. (2006) proposed four different views of the reference state: "minimally disturbed condition" refers to the absence of significant human disturbance and "historical condition" is a point in the past when this state was achieved (relevant in lakes, for example, where palaeolimnology can be used to define the reference state). "Least disturbed condition" refers to contemporary sites that do not conform to “minimally disturbed condition” but where human disturbance is deemed to fall below thresholds likely to impact ecological condition. Finally, “best attainable (or “best available”) condition” recognizes situations where none of the other criteria are met but “where the impact on biota of inevitable land use is minimized” (Stoddard et al., 2006). The WFD itself does not specify in detail which of these is most appropriate for defining the “expected” condition of the biota, and differences in assumptions about the reference state complicate comparisons between national approaches to WFD implementation and harmonizing ambitions (Birk et al., 2013). For this reason, a common approach to defining reference conditions for rivers was developed (Pardo et al., 2012), roughly aligned to the “least disturbed condition” of Stoddard et al. (2006). This was a valuable step towards ensuring a harmonized implementation of the WFD in those regions where there are still catchments with relatively low population densities but, for many other regions, this European reference concept only served to highlight the impaired nature of the landscape.

A further complication in the WFD is that ‘macrophytes’ and ‘phytobenthos’ are treated as a single ‘Biological Quality Element’ (BQE), meaning that information about the condition of two ecosystem components which respond to change on very different spatial and temporal scales has to be combined when reporting outcomes. In practice, most Member States (UK included) developed separate methods for macrophytes and phytobenthos (the latter often using diatoms as a proxy), combining outputs only at the final stage before reporting. However, there is no reason, in theory, why two such different ecosystem components should respond similarly along a stressor gradient. Differences in growth rates, in their use of sediment nutrient pools and susceptibilities to other stressors will all contribute to differences in assessment outcomes even before differences in the tools themselves are considered. The UK macrophyte tool, for example, uses two different measures of diversity as well as a direct measure of the impact of nutrients on the macrophyte community (Willby et al., 2009) whereas the UK phytobenthos tool, like most European methods, depends solely on a measure of nutrient impact (Kelly et al., 2008; Kelly, 2013). Indeed, the use of diatom diversity as part of a status/condition assessment is questionable (Denicola & Kelly, 2014) and rarely insightful (e.g. Blanco et al., 2012). There are, in other words, both theoretical and methodological issues besetting the combination of macrophytes and phytobenthos into a single BQE. Whether or not this is relevant will depend upon how the two sub-elements are combined (either the average or the more stringent – the latter accords to the ‘one out, all out’ principle that is used when comparing BQEs). In the UK, because the first iteration of the phytobenthos tool was consistently more stringent than the macrophyte tool, particularly in high alkalinity rivers, a decision to apply the ‘one out all out’ rule within the BQE effectively made the macrophyte tool redundant.

This, in turn, exposed methodological differences between the two approaches, particularly in the way that expectations (i.e. "reference conditions") were calculated. In brief, many high alkalinity (> 125 mg L^{-1} $CaCO_3$) lowland sites (especially chalk streams) that failed to achieve good status for phytobenthos supported rich macrophyte floras as well as (in many cases) good quality invertebrate and fish faunas. The initial response to this divergence between macrophytes and phytobenthos involved recalibration of the phytobenthos reference model to bring it in line with the approach used to define reference conditions for macrophytes, along with rules about how the tools should be used (Kelly et al., 2014). However, subsequent experience suggests that these administrative and methodological 'fixes' really need to be underpinned by (i) a better theoretical understanding of how both sub-elements respond to the abiotic variables from which "expected" metric values (reference values) are calculated, and (ii) how the respective tools reflect target stressors.

This paper presents an alternative approach to determining reference metric values , but also addresses the fundamental differences between phytobenthos and macrophytes that need to be considered when using the two groups as part of a combined 'macrophytes and phytobenthos' assessment for the WFD. Though based around UK experiences, the lessons we describe are appropriate for any country within the EU that is currently revising methods, as well as for those countries wishing to join the EU. In particular, we discuss the limitations of using the concept of "least disturbed condition" to select sites from which the denominator for ecological status estimates can be calculated, especially in regions of high population density and intensive agriculture. We revisit the possibility of using 'best available' sites as an alternative. Whilst the limitations of this approach identified by Stoddard et al. (2008) still apply, the availability of larger datasets and better knowledge of the limitations of the other possible approaches means that we now have the information necessary to use the 'best available' concept to produce valid predictions of the expected condition that will, in turn, inform better management of the UK's rivers. The more realistic phytobenthos assessments that result then form the basis for a robust comparison with macrophyte assessments. Finally, we argue the case for taking the average of the two assessments rather than the most stringent of these.

More specifically, the objectives are:

1. To test the performance of the phytobenthos reference model currently used in the UK, with a particular focus on how it responds to variations in non-stressor variables such as alkalinity;

2. To develop a new reference mode using the conceptual approach described above;

3. As the WFD requires macrophyte and phytobenthos assessments to be reported together, we also examine the consequences of the new phytobenthos reference model on combined macrophyte and phytobenthos assessments; and,
4. To explore the potential for using a single sub-element (e.g. macrophytes or phytobnethos) to predict the EQR of the combined biological quality element.

2. Methods

2.1 Dataset

The datasets used in this report consist of 1505 benthic diatom samples from 843 locations throughout the UK, all of which are linked to hydrochemistry data and 443 of which are also linked to macrophyte survey data. Environmental variables included are PO_4-P, NO_3-N, alkalinity, conductivity and pH. Hydrochemistry data are expressed as annual means using either the arithmetic mean (alkalinity and pH) or geometric mean (all other variables) of all available data for the period 2012-2017. Determinations less than the detection limit were taken as half the detection limit. This may overestimate actual values at low concentrations but water chemistry data was used primarily to validate diatom metrics and only used as a guide to modify the indicator values of a few, rare taxa (see below).

Diatom samples were collected and analysed following Kelly et al. (2008) using methods that conform to EN13986 and EN14407 (CEN, 2014a,b); minor modifications to the Trophic Diatom Index of Kelly et al. (2008: TDI3) are described in UK TAG (2014a: TDI4) and in Kelly et al. (2018: TDI5). Macrophytes were surveyed and evaluated following UK TAG (2014b) which corresponds to EN14184 (CEN, 2014c). In both cases, to comply with WFD criteria, results are expressed as observed metric values divided by expected metric values, so-called Ecological Quality Ratios (EQRs). For diatoms, EQR is calculated as (100 – Observed TDI) / 100 – Expected TDI), as the TDI increases as ambient nutrient availability increases). EQR = 1 indicates observed = expected condition, suggesting little or no anthropogenic impact on the biota.

2.2 Statistics

In order to aid comparisons between diatoms and macrophytes, EQR scales were normalized so that status class boundaries occurred at regular intervals (0.2, 0.4, 0.6 and 0.8) along the EQR scale.

Analyses were performed with the R software package (R Development Core Team 2012) with the following packages: mgcv (Wood, 2017) for generalised additive modelling (GAM) and quantreg (Koenker, 2017) for quantile regression. The GAM models were fitted using thin plate splines, with the number of knots set at 7. This value was chosen as the main purpose of the model was to provide a general indication of a non-linear trend. We checked the models using the gam.check function to examine the distribution of residuals and to ensure that the basis dimension was adequate.

2.3 Development of reference models

The phytobenthos reference model currently used in the UK was developed by building a linear regression model of the response of the biological metric to abiotic properties of a site, using a sub-set of sites deemed to have no or minimal levels of anthropogenic pressure (Table 1). For the first iteration of the UK phytobenthos tool, only alkalinity and season made significant contributions to the relationships (Kelly et al., 2008) in contrast to the situation for macrophytes where distance from source, source altitude and slope also play a significant role (Willby et al., 2009):

$$TDI_{exp} = (56.83 * \log_{10}(\text{alkalinity}) - (12.96 * (\log_{10}(\text{alkalinity})^2 + 3.21(\text{season}) - 25.36 \quad [1]$$

Where: "TDI_{exp}" = expected value of TDI and "season" is 0 for samples collected in January to June and 1 for samples collected from July to December.

A shortcoming of the above approach was that for the phytobenthos there were very few sites from high alkalinity rivers that passed the screening criteria for reference sites and thus the expected TDI of such sites was determined by extrapolation. This reference model produced relatively low TDI values in rivers of high alkalinity resulting in very few of these sites achieving better than moderate status. Due to the divergence between macrophyte and phytobenthos assessments, a second version of the equation for phytobenthos was derived using a larger subset of sites that met the screening criteria used for macrophytes, which included additional sites from high alkalinity rivers that still contained a range of sensitive macrophyte taxa (Table 1). This yielded the following equation:

$$TDI_{exp} = 9.933 \times \exp(\log_{10}(\text{alkalinity}) \times 0.81) \quad [2]$$

This equation, however, also performed poorly in high alkalinity sites, effectively preventing phytobenthos from being used in status assessments in such situations. These two approaches demonstrate the difficulties of using the small data sets available at the time along with an incomplete understanding of reference conditions.

In this paper, rather than use a subset of data from “least disturbed sites” we have applied Stoddard et al. (2008)’s concept of “best attainable (available) sites”. Such sites do not necessarily equate to the WFD’s definition of reference sites as having “no or minimal anthropogenic [alteration]” so care is needed when using values derived from these sites as the denominator in EQR calculations. However, where true reference conditions no longer exist the approach provides a more robust method of providing the benchmark that the WFD requires. This issue will be dealt with in greater length in the Discussion.

Table 1. Screening criteria applied in selection of reference sites for original (equation 1: Kelly et al., 2008) and revised (equation 2) diatom reference models. The screening criteria for the revised reference model were also used to define expected values of metrics within the UK macrophyte tool (Willby et al. 2009).

Original reference model (Kelly et al., 2008)	Revised reference model (Willby et al., 2009)
	Filamentous algal cover <5%
	Number of macrophyte taxa >4
Alkalinity < 50 mg L^{-1} $CaCO_3$: 20 µg L^{-1};Soluble reactive phosphorus (SRP): 2 mg L^{-1} Nitrate-N	Predictions of number of invertebrate taxa or average score per taxon > middle of good status
alkalinity ≥ 50 mg L^{-1} $CaCO_3$: 30 µg L^{-1} SRP: 4 mg L^{-1}	Total oxidised nitrogen: type specific: Low alkalinity (upland and lowland): ≤ 1mg L^{-1} High alkalinity: ≤ 2 mg L^{-1}
Samples with TDI >50 removed	SRP type specific: Low alkalinity lowland: ≤ 20 µg L^{-1} Low alkalinity upland: ≤ 15 µg L^{-1} High alkalinity lowland: ≤ 40 µg L^{-1} High alkalinity upland: ≤ 30 µg L^{-1} Very high alkalinity: ≤ 50 µg L^{-1}

Notes:

1 Low alkalinity: < 50 mg L^{-1} $CaCO_3$ (based on long-term average at site); high alkalinity: ≥50 mg L^{-1} $CaCO_3$; very high alkalinity: > 150 mg L^{-1} $CaCO_3$.

2 Upland: >80 m above sea level; lowland: ≤ 80 m

3 Predictions of invertebrate status were based on practice current at time of site selection, using approaches described in Wright et al., 1989)

3. Results

3.1 Overview of data

A total of 525 non-planktonic diatom taxa were identified in the 1505 samples. These samples covered a wide range of water quality (Table 2) within which both PO_4-P and NO_3-N were correlated with alkalinity and with each other (Fig. 1a., b., c.). There was also a significant correlation between pH and alkalinity (Fig. 1d).

Table 2: Summary statistics of selected environmental variables for the dataset.

Variable	N	Mean	Median	Min	Max
PO_4-P (µg L^{-1})	1505	81.18	28.23	1.0	3600
NO_3-N (mg L^{-1})	1505	2.47	1.45	0.05	27.33
Conductivity (µS cm^{-1})	1357	320	238	26	2162
Alkalinity (mg L^{-1} $CaCO_3$)	1505	79.9	56.2	1.7	382
pH	1373	7.70	7.78	5.77	8.44

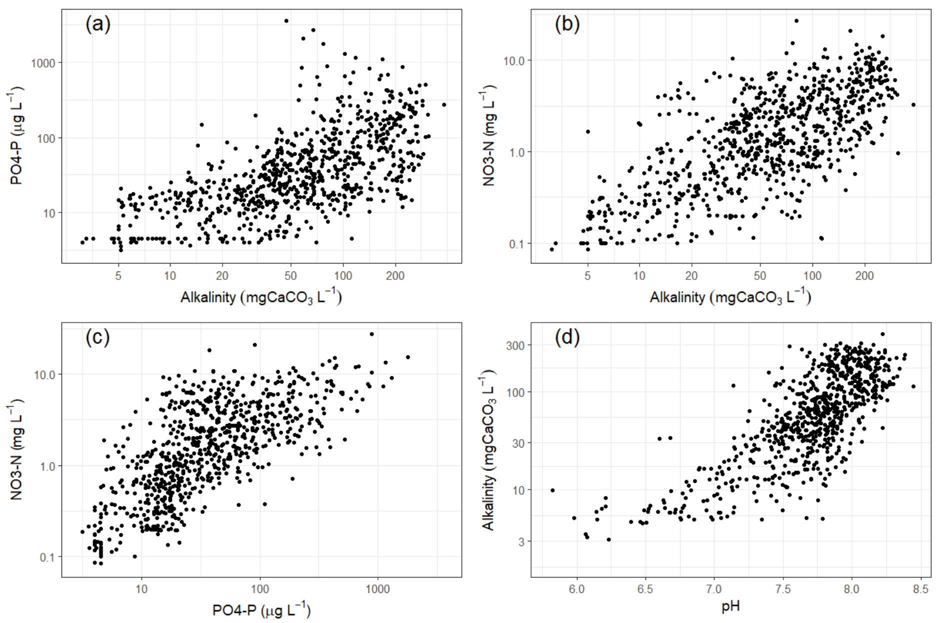

Figure 1. Relationships between hydrochemical variables within the dataset used for this study.

3.2 Evaluation of existing reference model

TDI values generated by both the original (equation 1) and revised (equation 2) phytobenthos reference models hug the lower edge of the data cloud (composed of all samples, whether or not they fulfil reference screening criteria) at low and moderate alkalinity (Fig. 2). However, they diverge at high alkalinity, with equation 1 levelling off when TDI = ~40 whilst equation 2 continues to increase, so that "expected" TDI values of 60 or more are possible in very high alkalinity water. Both predicted lines lie within the data cloud; however, at higher alkalinity values equation 1 returns values greater than many of the observed TDI values, meaning that a large number of sites would have a diatom EQR value exceeding 1 and thus be classified at high status.

The limitations of the reference model can clearly be seen when the TDI values are converted to EQRs and modelled using a multivariate GAM with soluble nutrients (PO4-P and NO3-N) and alkalinity as independent variables. TDI EQR has a clear response to phosphorus (Figure 3a) and a weaker response to nitrate-nitrogen (Figure 3b), as expected, but there is also a marked increase of EQR at higher alkalinity values (Figure 3c). By comparison, the same relationships for macrophytes (Figure 3f) show a much smaller EQR response to alkalinity. The consequence of this on EQR values for these two BQEs is shown in Figure 4, with diatoms generally providing more stringent classifications (negative differences) at lower alkalinity (< 50 mg $CaCO_3$ L^{-1}), but less stringent classifications (positive differences) at very high alkalinity (> 125 mg $CaCO_3$ L^{-1}). A further complication is that high "expected" values mean that five status classes have to be compressed into less than half of the total

TDI scale and, as a result, phytobenthos could not be used for ecological status classification in high alkalinity rivers in the UK. The rest of the paper, therefore, is focused on developing a more effective reference tool for diatoms.

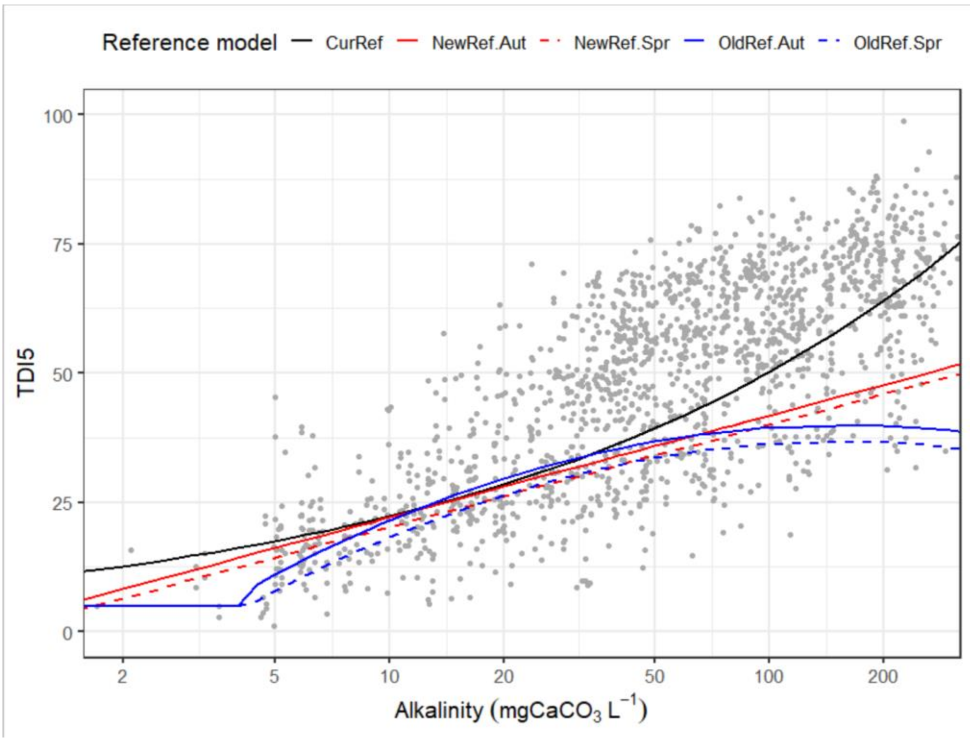

Figure 2 Modelled reference TDI values overlain on scatter plot of observed TDI. Blue line = original reference equation (equation 1: spring dashed, autumn solid); black line = current reference equation (equation 2); red line = new reference equation (equation 4: spring dashed, autumn solid)

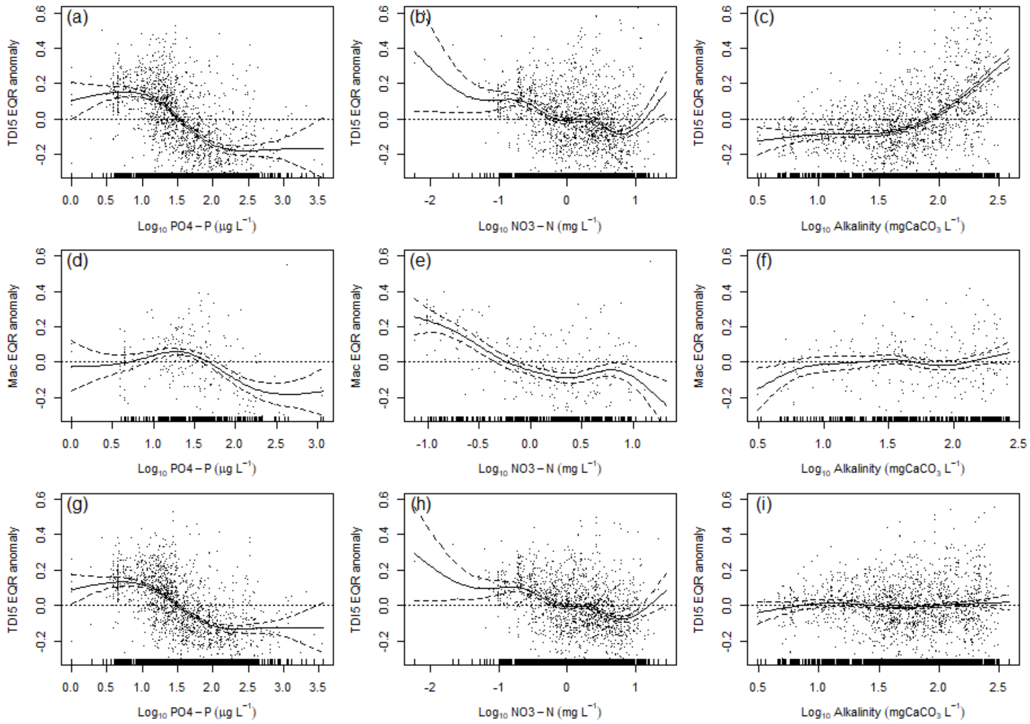

Figure 3: Relationship between TDI5 (light microscopy) EQR (a-c), macrophyte final EQR ("Mac EQR", d-f), and TDI using the new reference TDI equation (with season) (g-i) and soluble reactive phosphorus, nitrate-nitrogen and alkalinity, showing GAM smooths. EQR values are expressed as an anomaly, i.e. the values are relative to the overall mean EQR (the zero dotted line) , points show distribution of residuals.

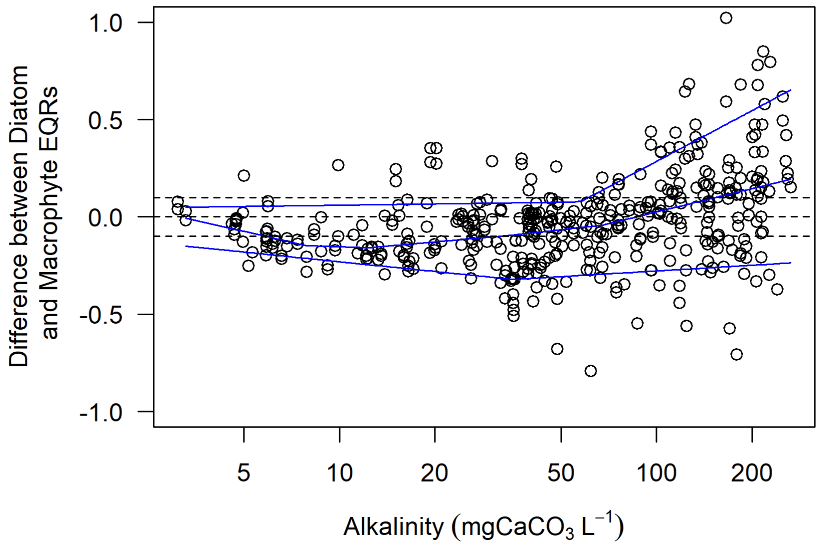

Figure 4: Difference between diatom (TDI5) and macrophyte EQR values plotted against alkalinity. Blue lines show regressions fitted to 90th, 50th and 10th quantiles, dotted lines mark an EQR of ±0.1 (a WFD class) and the zero value (no class difference).

3.3 Development of new diatom reference model

It is clear that selecting reference sites in rivers is a difficult process and, inevitably, very few sites from lowland higher alkalinity rivers survive the screening process. An alternative strategy is to fit regression models to a sub-set of sites that have the lowest observed TDI values for a given alkalinity, the so-called “best available” approach. This can be done by fitting a regression to a lower quantile of the relationship between observed TDI and alkalinity. As alkalinity is correlated with the soluble phosphorus concentration (Figure 1a), fitting a regression between alkalinity and a lower quantile of TDI allows for the effect of increasing background (natural) phosphorus along this gradient. However, alkalinity is also correlated with nitrate-nitrogen (Figure 1b). Although background phosphorus is likely to be correlated with alkalinity as sources of both are related to catchment geology, this is unlikely to be true for background nitrogen, which should be low across the range of alkalinity. To allow for this, nitrate-nitrogen concentration has also been included as a predictor variable in a quantile regression equation.

A regression model was fitted to the 25^{th} quantile using the $\log_{10}$ of alkalinity and nitrate nitrogen as predictor variables (Equation 3). In addition, sample season was included as a categorical variable (spring = 0, autumn = 1) with season split before/after July (Equation 4), as this was found to be a significant variable in the original diatom reference equation.

Both models show highly significant effects of alkalinity and nitrate-nitrogen, and model 2 showed a just significant effect of season ($p = 0.03$). The resulting equations are shown in Table 3 and Fig. 2. These parameters were then used to predict reference TDI values, taking a nitrate-nitrogen concentration of 0.5 mg NO_3-N L^{-1}, a conservative value less than the threshold of 1 mg L^{-1} used by Willby et al. (2009) and thus assumed to be consistent with reference conditions across the range of alkalinity.

Table 3: Quantile regression equations for revised reference model.

term	estimate	std. error	statistic	p.value
Equation 3				
(Intercept)	7.216	2.025	3.563	0.0004
Log_{10} Alkalinity (mg L^{-1} $CaCO_3$)	18.9	1.231	15.35	<0.0001
Log_{10} NO_3-N (mg L^{-1})	15.15	1.013	14.95	<0.0001
Equation 4				
(Intercept)	5.061	2.105	2.404	0.0163
Log_{10} Alkalinity (mg L $CaCO_3$)	19.69	1.239	15.89	<0.0001
Log_{10} NO_3-N (mg L^{-1})	14.95	1.008	14.83	<0.0001
Season	1.856	0.8629	2.151	0.0317

The EQR for TDI5 was calculated using the new reference equation, including season as a predictor; Fitting GAM models including nutrients and alkalinity demonstrates that the new reference equation results in EQRs that still respond to nutrients (Figure 3g.,h.) but where the effect of alkalinity on EQR has been removed (Figure 3i, Table 4) and should thus give a more reliable reflection of the ecological response to the nutrient stressor gradient.

Table 4: GAM model for TDI5 EQR using new reference TDI model against nutrients and alkalinity.

Parametric coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.616	0.003	179.115	< 0.001

Approximate significance of smooth terms (df = degrees of freedom):

	Estimated df	Reference df	F	p-value
s(LogP)	4.761	5.494	56.824	< 0.001
s(LogN)	5.820	5.983	18.000	< 0.001
s(LogAlk)	3.839	4.632	1.722	0.122

3.4 Implications for combining macrophytes and phytobenthos

The current rule for combining macrophytes and diatoms in the UK is that the lowest of the two EQR values determines the overall macrophyte and phytobenthos BQE classification (in effect, applying the “one-out, all-out” rule within the BQE). Using the current UK TDI reference model (equation 2), diatoms tended to be more stringent at low alkalinity and macrophytes at high alkalinity (Figure 5a). However, the new diatom reference equation shifts this balance, leading to consistently more stringent classifications being obtained using diatoms across the entire alkalinity range (Figure 5b). This means that, in effect, macrophytes will rarely determine final classifications so will have less direct relevance to the river basin management process. There is a case, therefore, for re-examining the manner in which results from macrophytes and diatoms are combined and, in particular, to consider whether averaging the metrics might provide a better approach.

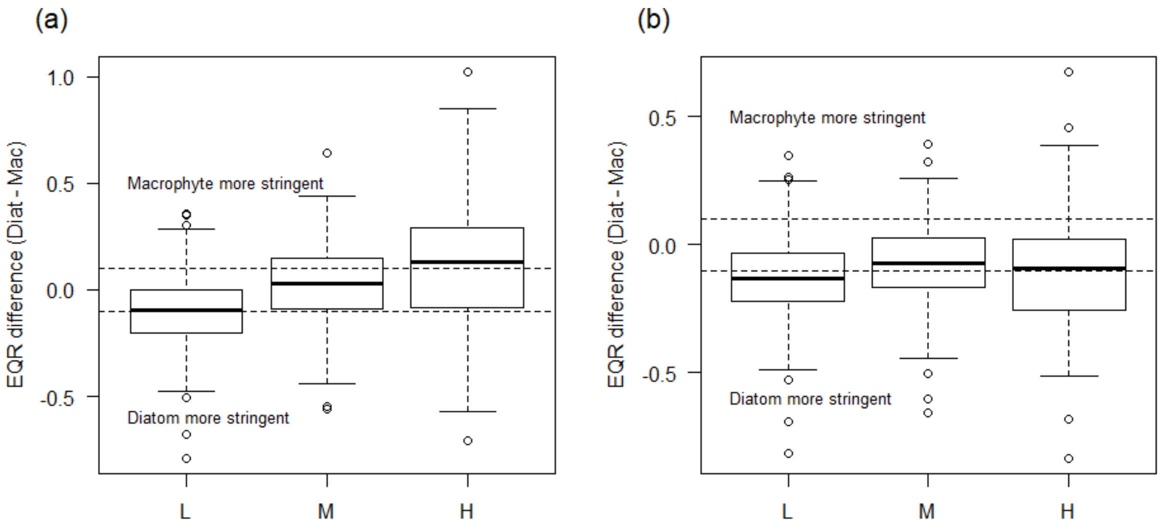

Figure 5: Difference between diatom (Diat) and macrophyte (Mac) EQR values using (a) current TDI reference and (b) new TDI reference, split by alkalinity type (L = < 75 mg $CaCO_3$ L^{-1}, M = 75 - 125 mg $CaCO_3$ L^{-1}, >125 mg $CaCO_3$ L^{-1}). Horizontal lines mark ±0.1 EQR units i.e. 1 WFD class.

Comparing the relationships between each of the metric EQRs, the average and the lowest of either metric EQR with PO_4-P concentration (Figure 6) clearly demonstrates that the new diatom reference equation has a better relationship with phosphorus gradient than the current equation (linear regression: r^2 0.372 compared to r^2=0.123) and somewhat outperforms macrophytes (r^2=0.313), whilst the average of the new diatom EQR and macrophyte EQR gives the strongest relationship of all (r^2=0.453), although only slightly stronger than that using the lowest of diatom and macrophyte EQRs (r^2=0.430). The uplift in sensitivity from considering both macrophytes and diatoms together rather than in isolation is, however, quite clear.

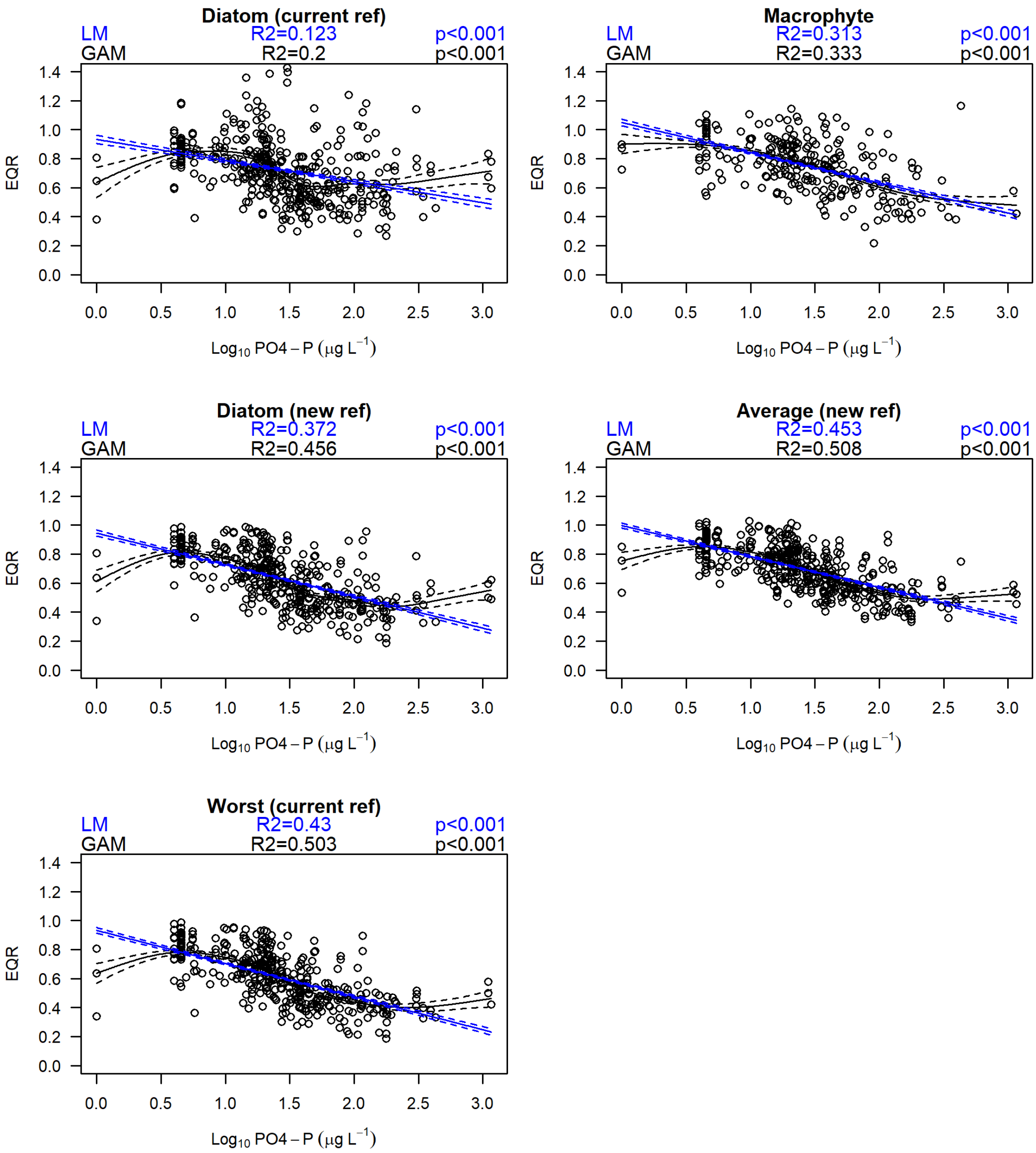

Figure 6: Relationship between EQR and SRP concentration, with linear model (LM) and GAM model fits.

3.5 Prediction of average EQR from single metrics

The average EQR of the new diatom and macrophyte EQRs will be lower than one of the two individual metrics (typically macrophytes, as diatoms are now, on average, more stringent). The extent to which either sub-element departs from the average of the two can be estimated by modelling the average EQR from the individual metric EQRs and alkalinity (Tables 5 & 6, Figure 7). On average the Diatom EQR values are increased by 0.06 EQR units and the macrophyte EQR values are decreased by 0.06 EQR units (slightly more than a quarter of a class), though the extent of the change depends upon the position along the gradient.

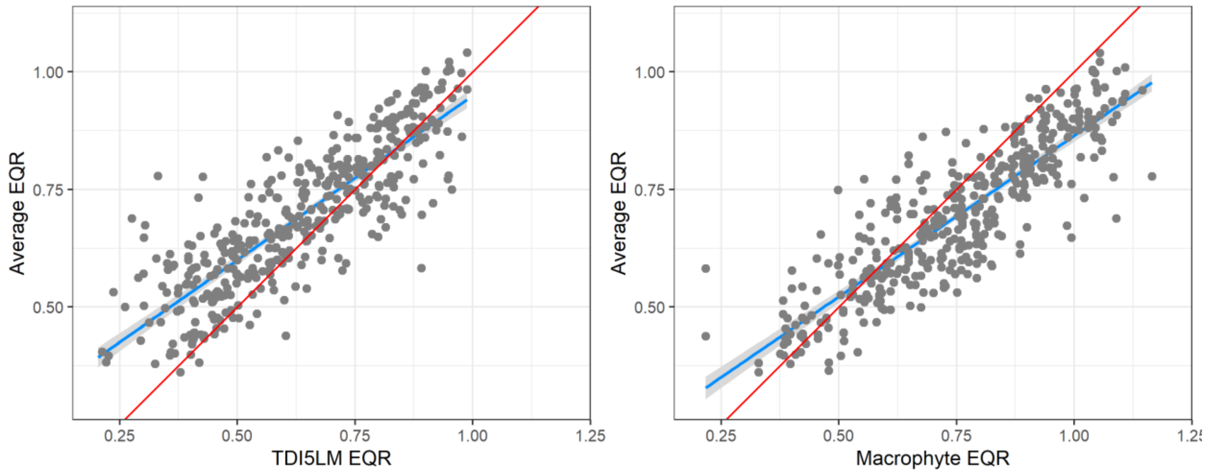

Figure 7: Conditional regression plots, showing relationship between Average EQR and TDI5-EQR (left) and Macrophyte-EQR (right)) at median alkalinity for models listed in Tables 4 and 5. Red line = 1:1 fit.

Table 5: Linear model predicting Average EQR from Diatom TDI5 EQR and log_{10} Alkalinity.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.3426	0.0258	13.25	< 0.001
TDI5_EQR	0.6988	0.0233	30.05	< 0.001
Log10_Alkalinity	-0.0545	0.0090	-6.024	< 0.001

Table 6: Linear model predicting Average EQR from Macrophyte final EQR and log_{10} Alkalinity.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.2694	0.0266	10.13	< 0.001
Macrophyte_EQR	0.6849	0.02162	31.67	< 0.001
Log10_Alkalinity	-0.0533	0.00871	-6.122	< 0.001

4. Discussion

4.1 Combining macrophyte and phytobenthos assessments

The WFD presents some genuine scientific challenges, as well as frustrations such as the decision to include both macrophytes and phytobenthos as a single BQE, which makes little logical sense. "Macrophytes and Phytobenthos" includes photosynthetic organisms with a wide range of growth strategies. Trying to reconcile differences in classifications produced by macrophytes and phytobenthos needs some recognition of how these respond at different spatial and temporal scales.

"Macrophytes" encompass a range of growth forms, including filamentous algae, mosses, free floating and rooted vascular plants, the latter including species that are wholly-submerged and emergent. There is also a range of sizes, from a few millimetres to several metres in length. Macrophytes exploit a range of habitats within a stream, some growing directly on rocks, whilst others are rooted in fine or coarse sediments. Life-cycles range from a few weeks (in the case of some of the algae) to a year or longer, in the case of vascular plants. This means that the macrophyte assemblage as a whole is exposed to a variety of sediment and water column nutrient pools, and responds to change at different temporal scales.

"Phytobenthos", on the other hand, is sampled from a single habitat (biofilms on rocks and/or plant surfaces). The assemblage is dependent primarily on water column nutrients, and individual organisms are smaller and shorter-lived. Studies have shown that diatom assemblages are shaped by in-stream nutrient and flow conditions over the preceding two to three weeks (Lavoie et al., 2008; Snell et al., 2014).

It is important to acknowledge these differences in order to develop a robust approach to dealing with the combined "Macrophyte and Phytobenthos" BQE. They also help to explain the problems encountered with the present approach, in which high alkalinity reference sites for phytobenthos were selected using expert judgement based on an understanding of the macrophyte communities. Well-developed and species-rich macrophyte communities will be better-buffered against consequences of occasional nutrient pulses than phytobenthos and we believe that using phytobenthos from such sites may have led to inflated predictions of expected TDI in high alkalinity rivers.

This leaves the problem of how reference conditions for phytobenthos should be set in high alkalinity rivers. Having exhausted other options, we have adopted a new approach based on the "best available" results obtained from ongoing monitoring. The lower edge of the data cloud produced when TDI is plotted against alkalinity, regardless of stressor state, should indicate the best possible conditions that are encountered. That the current reference model follows a line closer to the median, particularly at high alkalinity, suggests a problem with this model. We have, therefore, fitted a new relationship to this data cloud using quantile regression. The result is a model that is more stringent than the current one, particularly at high alkalinity, but is a better reflection of the distribution of the data. Our experience of UK conditions is that, at low and moderate alkalinity, “best available” equates to “least disturbed conditions” (sensu Stoddard et al., 2006) but diverges from this at high alkalinity where geological conditions have often resulted in situations well-suited to settlement and agriculture.

However, this means that we now have different reference concepts for the two sub-elements within a single BQE. Can this be justified? Given the differences between macrophytes and phytobenthos, different responses to stressors are to be expected and this will extend to the appropriate variables used to screen reference sites. In particular, the sensitivity of phytobenthos to nutrients at a temporal scale finer than that used for routine monitoring raises issues about the use of a chemical screening threshold that cannot be supported by land use screening criteria.

All of our work to date suggests that a ‘significant change’ in community composition occurs at lower nutrient concentrations for phytobenthos than it does for macrophytes. Therefore, the way in which the two sub-elements are combined into the final BQE is critical. Using the “one out all out” rule **within** the macrophytes and phytobenthos BQE with a stringent diatom model will lead to some high alkalinity sites (such as chalk streams) failing to achieve GES despite other, more conspicuous (and valued) elements of the biota (invertebrates, fish, macrophytes) being at high or good status. In particular, this does not acknowledge the key habitat structuring role of macrophytes, especially at higher alkalinities, or recognize the basic biological differences between the sub-elements. Averaging the sub-elements means that information from both sub-elements contributes to the final decision, and “one out, all out” still applies **between** BQEs.

A possible problem with averaging is that this will reduce the likelihood of detecting ecological impacts in situations where macrophyte status is lower than that of diatoms due to non-nutrient stressors. A final possibility (not considered here) would be to introduce a more complex rule (e.g. to use the average of the two sub-elements in cases where macrophyte EQR > diatom EQR, but, otherwise, to use the worst case). This type of rule is already in use for lake phytoplankton analyses in the UK where cyanobacteria abundance is combined with the other constituent metrics by averaging when they are worse than the other metrics but are ignored when they are better (UK TAG, 2014b).

Applying a similar rule to macrophytes and phytobenthos could be considered but would require additional testing.

Another possibility is that the observed differences are the result of different means of assessing the status of the two sub-elements. The UK macrophyte tool uses a multimetric approach to compare sites against expectations that are based on several predictor variables (Willby et al., 2009) whilst phytobenthos relies on a single metric which is compared to predictions based on a single predictor variable, alkalinity (Kelly et al., 2008). These other predictors (slope, elevation, elevation of source and distance from source) are necessary due to the variation in size of macrophytes and their sensitivity to flow and substratum. We have repeatedly tested a wider range of abiotic predictor variables on datasets of phytobenthos reference sites and always find alkalinity to be the only one with a significant response. Developing multimetric models based on diatoms is also not straightforward: diversity, for example, is a problematic measure when applied just to diatoms (Denicola & Kelly, 2014) whilst metrics developed to evaluate other stressor gradients (e.g. Diatom Acidification Metric; Juggins et al., 2016) will not necessarily buffer the response to the primary nutrient gradient. The approach to phytobenthos assessment used in the UK is broadly in line with that elsewhere in Europe (Poikane et al., 2016), so we do not believe that the issues addressed here are unique to the UK. However, this does mean that phytobenthos is not well-equipped to detect "undesirable disturbances" (i.e. secondary effects arising from the impact of nutrients on the photosynthetic biota) so, again, a more nuanced means of interpretation than either averaging or taking the lowest might be appropriate, particularly at lower EQRs.

4.2 Revisiting "reference conditions"

The concept of ecological status, which is integral to the WFD, requires comparisons of the observed biota and supporting elements with those expected in conditions of no or minimal anthropogenic alteration (see Annex V, section 1.2). Reference conditions, in other words, define an ideal situation, whilst "good ecological status" recognizes a condition where key ecosystem services are delivered, fulfilling the WFD's ambition to "promote sustainable water use based on a long-term protection of aquatic resources" (Article 1).

The issue of reference conditions has been controversial throughout the lifetime of the WFD, with purists arguing for a very high standard of reference screening (Moss, 2008; Demars et al., 2012) whilst pragmatists point out that very strict criteria reduces the pool of qualifying sites to the point where there are insufficient to provide a reliable baseline for assessments that form the basis of statutory regulation. In practice, the definition of reference conditions is critical only if good ecological status is defined in terms of the amount of species turnover (or change in an index value) from this baseline (about half of all ecological assessment methods developed for the WFD define status classes by dividing the scale below reference into five equally-spaced sections: Birk et al., 2012). Both phytobenthos and macrophyte assessment tools in the UK use a more nuanced definition of good status (Kelly et al., 2008; Willby et al., 2009), which, whilst still expressed as a distance from reference, does not depend directly on the value of the index at reference conditions. We argue that ensuring a definition of "good status" that is consistent with sustainable use of aquatic resources is, actually, more important than conceptualizing the "perfect" stream.

A second important role of reference conditions within WFD implementation has been to act as a benchmark for international comparisons. The WFD required Member States to "intercalibrate" their high and good status boundaries in order to ensure a common level of ambition (Birk et al., 2013; Poikane et al., 2015) and a common understanding of "reference" was regarded as a necessary first step towards this goal (Pardo et al., 2012). However, whilst some exercises profited from an agreed understanding of the reference state (Kelly et al., 2009; Bennett et al., 2011), many other exercises foundered because reference conditions simply did not exist for many types of water body within Europe, prompting the development of a range of alternative approaches to ensure that status class boundaries were consistent between countries (Birk et al., 2013; Kelly et al., 2014). Time has, therefore, shown that the pursuit of a hypothetical ideal needs to be tempered by pragmatism if the long-term objectives of the WFD are to be secured.

This is the background against which the present study unfolded. "Best attainable condition" is, intrinsically, the weakest of the reference concepts discussed by Stoddard et al. (2008); however, by using a very large dataset and modelling lower quantiles of the relationship between biology and physico-chemical variables, we believe that it is possible to identify the lowest values of the biological metric that are achievable in practice. Whether these should be considered as "reference conditions" is a moot point. When a similar approach was applied to lowland Romanian lakes, there were concerns that the extent of degradation was such that the quantile approach should only be used to identify the high/good status boundary rather than reference conditions (Kelly et al., 2019). However, our judgment (based on knowledge of high alkalinity streams in the UK) is that this quantile can be treated as being close to "reference conditions" and can therefore serve as a denominator in EQR calculations. Use of the 25th, rather than a more extreme, percentile, ensures that the quantile regression equation is "anchored" by a substantial quantity of data and is offset by the use of a stringent concentration of nitrate-N when estimating the expected value of metrics. The decision about whether "best available" is good enough to ensure the sustainable use of water resources is, inevitably, a judgement call. What is clear from our study is that setting "expected" conditions for phytobenthos in this manner creates a substantial challenge for regulators, faced with a large number of sites which clearly still not meet the criterion of "good ecological status".

4.3 Conclusion: the "art of the possible"

This work was performed at a time when UK environment agencies were working with reduced budgets due to a central government policy of fiscal austerity. This has resulted in a "more with less" ethos in which all aspects of the organization were examined to ensure that the best possible use was made of limited resources. This scrutiny has extended to monitoring and assessment practice, and one question that naturally arises is whether assessments of both macrophytes and phytobenthos are necessary in order to make an assessment of the overall condition of "macrophytes and phytobenthos". In reality, poor performance of the current phytobenthos reference model had precluded its use in high alkalinity streams, even though there are situations (very small streams, ditches, canalized rivers etc.) where the macrophyte method cannot be applied or requires a bespoke approach (Willby et al., 2009). However, budgetary circumstances do not necessarily mean that the UK statutory agencies will now be enthusiastic about adopting the new model (however convincing its performance) alongside macrophyte assessments as this will require additional monitoring effort.

We have shown that it is possible to predict the value of the combined "macrophytes and phytobenthos" BQE from either of the sub-elements (Fig. 7; Table 5). Whether it is always wise to do this is another matter. Redundancy of phytobenthos assessments has been shown for lakes (Kelly et al., 2016; Schneider et al., 2019) though care is needed when extrapolating this conclusion to rivers as phytobenthos is often one of three elements of the photosynthetic biota (along with macrophytes and phytoplankton) evaluated in lakes, whilst only two are routine in all but the larger rivers. In practice, phytobenthos and macrophytes together capture different aspects of ecosystem complexity and we would recommend that both should be used wherever possible, not least because macrophytes, especially, respond to stressors such as hydromorphological alteration to which phytobenthos is less sensitive (Schneider et al., 2012). If decision-making is devolved to biologists who know the local context then we suspect that there will be situations where either macrophytes or phytobenthos alone will provide robust assessments and insights into key stressors. Where decisions are taken at higher levels, or evidence gathering is driven by budget constraints, then misclassifications should be expected.

The challenges presented by the WFD were large and, in the two decades since it was passed, much has been learned about the science underlying effective implementation. It is, therefore, appropriate, that methods that were developed in the first years of the WFD should be revisited and their performance reevaluated against the larger datasets now available. This will, inevitably, lead to shifts in classifications which may not be popular with the bureaucrats responsible for water management. However, the greater risk is that poor decisions will result from the use of tools that, for perfectly understandable reasons, may not be as effective as they should be. Compared to capital investment in wastewater treatment, for example, the cost of periodic “servicing” of WFD assessment tools is miniscule.

Acknowledgements

Many thanks to Jan Krokowski for support during this project.

Funding: This work was supported by the Scottish Environment Protection Agency (SEPA) on behalf of the UK Water Framework Directive Technical Advisory Group (UK TAG) (Project 4057194).

References

Behre, K.-E., 1988. The rôle of man in European vegetation history, in Huntley, B., Webb, T, (Eds), Vegetation history (. Kluwer, Dordrecht, pp. 633-672.

Bennett, C., Owen, R., Birk, S., Buffagni, A., Erba, S., Mengin, N., Murray-Bligh, J., Ofenböck, G., Pardo, I., van der Bund, W., Wagner, F., Wasson, J. G., 2011. Bringing European river quality into line: An exercise to intercalibrate macro-invertebrate classification methods. Hydrobiologia 667, 31-48. https://doi.org/10.1007/s10750-011-0635-2

Bennion, H., Fluin, J., Simpson, G. L., 2004. Assessing eutrophication and reference conditions for Scottish freshwater lochs using subfossil diatoms. Journal of Applied Ecology. 41, 124-138. https://doi.org/10.1111/j.1365-2664.2004.00874.x

Birk, S., Bonne, W., Borja, Á., Brucet, S., Courrat, A., Poikane, S., Solimini, A., Van De Bund, W., Zampoukas, N., Hering, D., 2012. Three hundred ways to assess Europe’s surface waters: An almost complete overview of biological methods to implement the Water Framework Directive. Ecological Indicators 18, 31–41. doi:10.1016/j.ecolind.2011.10.009

Birk, S., van Kouwen, L., & Willby, N., 2012. Harmonising the bioassessment of large rivers in the absence of near-natural reference conditions - a case study of the Danube River. Freshwater Biology 57, 1716-1732.. https://doi.org/10.1111/j.1365-2427.2012.02831.x

Birk, S., Willby, N. J., Kelly, M. G., Bonne, W., Borja, A., Poikane, S., & van de Bund, W., 2013. Intercalibrating classifications of ecological status: Europe's quest for common management objectives for aquatic ecosystems. Science of the Total Environment, 454–455, 490-499. https://doi.org/10.1016/j.scitotenv.2013.03.037

Blanco, S., Cejudo-Figueiras, C., Tudesque, L., Bécares, E., Hoffmann, L., & Ector, L., 2012. Are diatom diversity indices reliable monitoring metrics? Hydrobiologia 695, 199-206.. https://doi.org/10.1007/s10750-012-1113-1

Borja, Á., Rodríguez, J. G.2010. Problems associated with the "one-out, all-out" principle, when using multiple ecosystem components in assessing the ecological status of marine waters. Marine Pollution Bulletin 60, 1143-1146. https://doi.org/10.1016/j.marpolbul.2010.06.026

CEN,2014a. EN13946: Water quality - Guidance standard for the routine sampling and pretreatment of benthic diatoms from rivers. Comité European de Normalisation, Geneva.

CEN, 2014b. EN14407: Water quality - Guidance standard for the identification, enumeration and interpretation of benthic diatom samples from running waters. Comité European de Normalisation, Geneva.

CEN, 2014c. EN 14184: Water quality - guidance standard for the surveying of aquatic macrophytes in runnIng waters. Comité European de Normalisation, Geneva.

Demars, B., Potts, J.M., Trémolières, M., Thiébaut, G., Gougelin, N., Nordmann, V., 2012. River macrophyte indices: not the Holy Grail! Freshwater Biology 57: 1745-1759.

Denicola, D. M., Kelly, M., 2014. Role of periphyton in ecological assessment of lakes. Freshwater Science, 33: 33: 619-638. **https://doi.org/10.1086/676117**

European Union, 2000. Directive 2000/60/EC of the European Parliament and of the Council of 23rd October 2000 Establishing a Framework for Community Action in the Field of Water Policy Official Journal of the European Communities, European Commission, Brussels (2000) (22 December, L 327/1)

Flannery, T., 2018. Europe: the First 100 Million Years. Penguin Books, Hamondsworth.

Hustedt, F., 1930. Susswasserflora von Mitteleuropa 10: Bacillariophyceae. Gustav Fischer, Jena.

Juggins, S., Kelly, M., Allott, T., Kelly-Quinn, M., Monteith, D., 2016. A Water Framework Directive-compatible metric for assessing acidification in UK and Irish rivers using diatoms. Science of the Total Environment 568: 671-678.

Kelly, M., 2013. Data rich, information poor? Phytobenthos assessment and the Water Framework Directive. European Journal of Phycology, 48, 437-450. https://doi.org/10.1080/09670262.2013.852694

Kelly, M., Juggins, S., Guthrie, R., Pritchard, S., Jamieson, J., Rippey, B., Hirst, H., Yallop, M., 2008. Assessment of ecological status in U.K. rivers using diatoms. Freshw. Biol. 53, 403-422.

Kelly, M.G., Bennett, C., Coste, M., Delgado, C., Delmas, F., Denys, L., Ector. L., Fauville, C., Ferreol, M., Golub, M., Jarlman, A., Kahlert, M., Lucey, J., Ní Chatháin, B., Pardo, I., Pfister, P., Picinska-Faltynowicz, J., Schranz, C., Schaumburg, J., Tison, J., van Dam, H., Vilbaste, S., 2008. A comparison of national approaches to setting ecological status boundaries in phytobenthos assessment for the European Water Framework Directive: results of an intercalibration exercise. *Hydrobiologia* 695, 109-124.

Kelly, M., Urbanic, G., Acs, E., Bennion, H., Bertrin, V., Burgess, A., Denys, L., Gottschalk, S., Kahlert, M., Karjalainen, S.-M., Kennedy, B., Kosi, G., Marchetto, A., Morin, S., Picinska-Fałtynowicz, J., Poikane, S., Rosebery, J. Schoenfelder, I., Schoenfelder, J., Varbiro, G., 2014. Comparing aspirations: Intercalibration of ecological status concepts across European lakes for littoral diatoms. Hydrobiologia, 734, 125-141. https://doi.org/10.1007/s10750-014-1874-9

Kelly, M.G., Birk, S., Willby, N.J., Denys, L., Drakare, S., Kahlert, M., Karjalainen, S.-M., Marchetto, A., Pitt, J.-A., Urbanič, G., Poikane, S., 2016. Redundancy in the ecological assessment of lakes: Are phytoplankton, macrophytes and phytobenthos all necessary? *Science of the Total Environment* 568: 594-602.

Kelly, M. G., Chiriac, G., Soare-Minea, A., Hamchevici, C., Juggins, S., 2019. Use of phytobenthos to evaluate ecological status in lowland Romanian lakes. Limnologica 77, 125682. https://doi.org/10.1016/j.limno.2019.125682

Kelly, M.G., Juggins, S., Phillips, G., Willby, N., 2018. Evaluation of benthic diatom classification in UK rivers using LM and NGS methods. Report number E18-56. Scottish Environment Protection Agency, Stirling.

Kelly, M., Yallop, M., 2012. A streamlined taxonomy for the Trophic Diatom Index. Report SC070034/TR1. Bristol: Environment Agency.

Kelly, M., Willby, N., Phillips, G., Benstead, R., 2014. The integration of macrophyte and phytobenthos surveys as a single biological quality element for the Water Framework Directive. Bristol, Environment Agency, .

Koenker, R., 2017. quantreg: Quantile Regression. R package version 5.36. http://CRAN.R-project.org/package=quantreg.

Lavoie, I., Campeau, S., Darchambeau, F.,Cabana, G., Dillon, P.J., 2008. Are diatoms good integrators of temporal variability in stream water quality? Freshwater Biology 53, 827-841.

Lloyd, C.E.M., Johnes, P.J., Freer, J.E., Carswell, A.M., Jones, J.I., Stirling, M.W., Hodgkinson, R.A., Richmond, C., Collins, A.L., 2018. Determining the sources of nutrient flux to water in headwater catchments: Examining the speciation balance to inform the targeting of mitigation measures. Science of the Total Environment 648, 1179-1200.

Moss, B., 2008. The Water Framework Directive: total environment or political compromise? Science of the Total Environment 400, 43-41.

Pardo, I., Gómez-Rodríguez, C., Wasson, J.-G., Owen, R., van de Bund, W., Kelly, M., Bennett, C., Birk, S., Buffagni, A., Erba, S., Mengin, N., Murray-Bligh, J., Ofenböeck, G., 2012. The European reference condition concept: A scientific and technical approach to identify minimally-impacted river ecosystems. Sci. Total Environ. 420, 33-42.

Poikane, S., Birk, S., Böhmer, J., Carvalho, L., De Hoyos, C., Gassner, H., Hellsten, S., Kelly, M., Solheim, A.L., Olin, M., Pall, K., Phillips, G., Portielje, R., Ritterbusch, D., Sandin, L., Schartau, A.-K., Solimini, A.G., van den Berg, M., Wolfram, G., Van De Bund, W., 2015. A hitchhiker's guide to European lake ecological assessment and intercalibration. Ecological Indicators, 52, 522-544. https://doi.org/10.1016/j.ecolind.2015.01.005

Poikane, S., Kelly, M.G., Cantonati, M., 2016. Benthic algal assessment of ecological status in European lakes and rivers: challenges and opportunities.   *Science of the Total Environment* 568: 603-613.

R Development Core Team, 2006. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Schneider, S.C., Kahlert, M., Kelly, M.G., 2013. Interactions between pH and nutrients on benthic algae in streams and consequences for ecological status assessment and species richness patterns. *Science of the Total Environment* 444: 73-84.

Schneider, S. C., Hjermann, D. O., & Edvardsen, H. (2019). Do benthic algae provide important information over and above that provided by macrophytes and phytoplankton in lake status assessment? – Results from a case study in Norway. Limnologica. 76., 28-40. https://doi.org/10.1016/j.limno.2019.02.001

Snell, M. A., P. A. Barker, B. W. Surridge, A. R. Large, J. Jonczyk, C. M. Benskin, S. Reaney, M. T. Perks, G. J. Owen, W. Cleasby, C. Deasy, S. Burke, P. M. Haygarth, 2014. High frequency variability of environmental drivers determining benthic community dynamics in headwater streams. Environ. Sci. Processes Impacts 16, 1629-1636.

Stoddard, J.L., Larsen, D.P., Hawkins, C.P., Johnson, R.K., Norris, R.H., 2006. Setting expectations for the ecological condition of streams: the concept of reference condition. Ecology Applications 16, 1267–1276.

Willby, N., Pitt, J.-A., Phillips, G., 2009. The Ecological Classification of UK Rivers using Aquatic Macrophytes. Science Report SC010080/SR1. Bristol: Environment Agency.

UK TAG, 2014a. UK TAG river assessment method: macrophytes and phytobenthos: phytobenthos – Diatoms for Assessing River and Lake Ecological Quality (River DARLEQ2). http://www.wfduk.org/resources/rivers-phytobenthos

UK TAG, 2014b. UK TAG river assessment method: macrophytes and phytobenthos – macrophytes (river LEAFPACS2). United Kingdom Technical Advisory Group on the Water Framework Directive. http://www.wfduk.org/resources/rivers-macrophytes

UK TAG, 2014c. UK TAG lake assessment method: phytoplankton – Phytoplankton Lake Assessment Tool with Uncertainty Assessment (PLUTO). United Kingdom Technical Advisory Group on the Water Framework Directive. http://www.wfduk.org/resources/lakes-phytoplankton

Wood, S.N., 2017. Generalized Additive Models: An Introduction with R (2nd edition). Chapman and Hall/CRC.

Wright, J.F., Armitage, P.D., Furse, M.T., Moss, D., 1989. Prediction of invertebrate communities using stream measurements. Regulated Rivers: Research & Management 4: 147-155.