Establishing ecologically- relevant nutrient thresholds: A tool-kit with guidance on its use

Martyn G. Kelly[1,2], Geoff Phillips[3], Heliana Teixeira[4], Gábor Várbíró[5], Fuensanta Salas Herrero[6], Nigel J. Willby[3] and Sandra Poikane[6]

1. Bowburn Consultancy, 11 Monteigne Drive, Bowburn, Durham DH6 5QB, UK.
2. School of Geography, Nottingham University, Nottingham NG7 2RD, UK.
3. Biological and Environmental Sciences, University of Stirling, Stirling FK9 4LA, UK.
4. Department of Biology & CESAM, University of Aveiro, Campus de Santiago, 3810-193 Aveiro, Portugal.
5. Department of Tisza Research, Institute of Aquatic Ecology, Centre for Ecological Research, Bem t´er 18/c, H-4026 Debrecen, Hungary.
6. European Commission Joint Research Centre (JRC), I-21027, Ispra, Italy.

Highlights

- A tool kit has been developed to derive ecologically-relevant nutrient thresholds Type II regression recommended where relationships are strong
- Binomial and classification mismatch approaches recommended for weaker relationships
- Most methods have limited applicability when other stressors are present
- Final choice of method also depends upon regulatory and enforcement regime.

Accepted refereed manuscript of: Kelly MG, Phillips G, Teixeira H, Várbíró G, Herrero FS, Willby NJ & Poikane S (2021) Establishing ecologically-relevant nutrient thresholds: A tool-kit with guidance on its use. *Science of the Total Environment*. https://doi.org/10.1016/j.scitotenv.2021.150977
© 2021, Elsevier. Licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International http://creativecommons.org/licenses/by-nc-nd/4.0/

Abstract

One key component of any eutrophication management strategy is establishment of realistic thresholds above which negative impacts become significant and provision of ecosystem services is threatened. This paper introduces a toolkit of statistical approaches with which such thresholds can be set, explaining their rationale and situations under which each is effective. All methods assume a causal relationship between nutrients and biota, but we also recognise that nutrients rarely act in isolation. Many of the simpler methods have limited applicability when other stressors are present. Where relationships between nutrients and biota are strong, regression is recommended. Regression relationships can be extended to include additional stressors or variables responsible for variation between water bodies. However, when the relationship between nutrients and biota is weaker, categorical approaches are recommended. Of these, binomial regression and an approach based on classification mismatch are most effective although both will underestimate threshold concentrations if a second stressor is present. Whilst approaches such as changepoint analysis are not particularly useful for meeting the specific needs of EU legislation, other multivariate approaches (e.g. decision trees) may have a role to play. When other stressors are present quantile regression allows thresholds to be established which set limits above which nutrients are likely to influence the biota, irrespective of other pressures. The statistical methods in the toolkit may be useful as part of a management strategy, but more sophisticated approaches, often generating thresholds appropriate to individual water bodies rather than to broadly defined “types”, are likely to be necessary too. The importance of understanding underlying ecological processes as well as correct selection and application of methods is emphasised, along with the need to consider local regulatory and decision-making systems, and the ease with which outcomes can be communicated to non-technical audiences.

Keywords: nutrients, Water Framework Directive, standards, aquatic ecosystems, nitrogen, phosphorus

Introduction

If visions of long-term sustainable water resources are to be achieved it is necessary to understand the links between degraded ecosystems and the stressors responsible. This enables appropriate management actions to be taken to restore those ecosystems to a point where they have sufficient resilience to be sustainable. Many of the decisions involved will be specific to individual water bodies; however, there is a case for national and international frameworks that can convert the broad ambition of legislation into quantifiable objectives. This, in turn, helps professionals identify those water bodies within a region in need of restoration, prioritise those with the greatest need, and gauge progress towards these objectives.

If water bodies in need of restoration are to be identified and prioritised, then we need to know both the condition of the ecosystem in relation to legislative targets (in Europe this is "good ecological status", as defined by the Water Framework Directive, WFD: European Union, 2000, or "good environmental status" for the Marine Strategy Framework Directive, MSFD, European Union, 2008) as well as the stressors likely to be responsible for their degradation. A key principle behind the WFD is that ecological status, though primarily focussed on biological structure, is also dependent on physico-chemical and hydromorphological conditions, which are in turn influenced by pressures in the catchment. In theory, if the sensitivities of different groups of organisms to these physico-chemical conditions can be quantified, then it should be possible to infer a threshold above which good status is unlikely to be achieved.

Much attention in recent years has focussed on interactions between stressors, recognising that part of the uncertainty observed in relationships with a single stressor is due to interactions (additive, synergistic or antagonistic) with other stressors (Nõges et al., 2016; Torres et al., 2017). Subsequently, models have begun to incorporate this complexity within catchment-level decision making processes (Spears et al., 2021). Such approaches, however, sit within broader screening exercises that, in effect, evaluate a wide range of potential stressors against estimates of "no observable effect concentrations" (borrowing a phrase from ecotoxicology) in order to focus attention of regulators on stressor combinations likely to be significant within a particular region. These threshold concentrations may have regulatory significance and are often referred to as "standards" or "criteria". In practice, however, uncertainty in relationships between biology and individual stressors means that predictions of the benefits of remediation currently lack precision (Moe et al., 2015; Prato et al., 2014). This is now recognised as a major weakness of WFD implementation (Hering et al., 2010; 2015; Carvalho et al., 2019).

Eutrophication (the negative biological consequences of elevated nutrient concentrations) is one of the key pressures affecting waters - both freshwater and marine (e.g. European Environment Agency, 2018). The ability to set realistic targets to guide catchment managers would therefore be an important step towards achieving environmental quality objectives. However, recent reviews of nutrient targets adopted by Member States revealed that a wide range of concentrations are currently used (Poikane et al., 2019a). Some of this variation reflects the substantial differences in background concentrations and the sensitivities of water bodies to nutrient enrichment that exist within and between Member States. However, it is also possible that some nutrient standards are not fit for the purpose of protecting good ecological status, both in the water body itself and in water bodies further downstream. Recent predictions, for example, suggest that MSFD objectives are unlikely to be achieved even after proposed nutrient reduction measures are in place, and more ambitious steps may thus be required (Piroddi et al., 2021; Friedland et al., 2021; Grizzetti et al., 2021). Any such steps will have implications for various industrial and agricultural sectors and therefore need to be based on a firm understanding of what concentrations are necessary to achieve WFD and MSFD nutrient targets.

Nutrients are also good candidates for a broader consideration of how thresholds for physico-chemical stressors should be derived. There are situations (e.g. phytoplankton in deep lakes) where phosphorus, in particular, is frequently the sole or most important stressor whilst in other circumstances (e.g. rivers), nutrients are almost always just one ingredient of a “cocktail” of stressors (Birk et al., 2020). In both cases, however, decisions by regulators have substantial real-world consequences, requiring public or private investment, in the context of legislation for which public consultation and transparency are prerequisites. The science behind such decisions, therefore, needs to be clear and uncertainty well explained.

In this paper, we present a toolkit for establishing ecologically-relevant nutrient thresholds. The toolkit is available either as a series of R scripts (https://publications.jrc.ec.europa.eu/repository/handle/JRC112667) or as a Shiny app (http://phytoplanktonfg.okologia.mta.hu:3838/Tkit_nutrient/). These approaches have been tested for lakes (Free et al., 2016; Poikane et al., 2019b; Kagalou et al., 2021), rivers (Canning et al., 2021; Poikane et al., 2021), coastal and transitional waters (Salas Herrero et al., 2019) as well as with simulated data (Phillips et al., 2019). Alongside statistical approaches, we also provide a brief guide on how to choose the most suitable approach and how to interpret the results.

General principles

There are many potential approaches to defining boundaries for nutrients and other physico-chemical variables. Conclusions from experimental studies could be used but are potentially highly context-specific, so the most common approach is to derive standards from monitoring data (Dodds et al, 2010; Free et al., 2016; Hausmann et al., 2016; HELCOM, 2013, Poikane et al., 2019b; Phillips et al., 2019). This, however, presumes that a stressor gradient is present which, though usually the case, is not universally true. It will be difficult to apply many of the methods in this toolkit in situations where there is no appreciable stressor gradient or, conversely, where all sites are so degraded that there are no high or good quality sites against which thresholds can be calibrated. The appropriate method for any situation will depend upon particular regulatory needs as well as the statistical properties of the data. In the case of the WFD, boundaries for "supporting elements" need to be linked to boundaries between ecological status classes for one or more Biological Quality Elements (BQEs). As the WFD adopts the "one out, all out" principle (Borja and Rodriguez, 2010; Ojaveer and Eero, 2011) for defining overall status, the BQE that is most sensitive to a given stressor is the best candidate for establishing a protective threshold. High statistical significance should be combined with theoretical justification or experimental evidence to demonstrate a causal relationship between ecological condition and nutrients, including determination of whether phosphorus, nitrogen, or phosphorus and nitrogen are limiting nutrients (Dolman et al., 2016; Guildford and Hecky, 2000; Phillips et al., 2008; Søndergaard et al., 2017). However, the overwhelming conclusion from many studies is that phosphorus reduction alone, without concomitant reduction in nitrogen, will not provide efficient eutrophication control. In the best case, this might displace the effects of eutrophication in space or time whilst, in the worst case, it may increase the potential for algal blooms and associated toxicity (Conley et al., 2009; Glibert, 2017; Paerl, 2009; Paerl et al., 2016).

Approaches in this toolkit should also protect particular levels on the "biological condition gradient", as used in the USA (Davies & Jackson, 2006; Charles et al., 2021). It is also possible to derive nutrient boundaries from ecological data without the need to summarise the latter as a metric (e.g. Roubeix et al., 2016, 2017; Tibby et al., 2019). This is less appropriate in the context of the WFD or MSFD as there is no link with measured ecological condition, although it may be appropriate in situations where the link with ecology is defined differently and is also a valuable means of validating boundaries obtained by other means (Taylor et al., 2018; Kelly et al., 2019b).

The prerequisite for all the methods described here is a dataset comprising biological samples summarised as a metric with each matched to water chemistry (preferably several samples aggregated as a mean or median). Samples in the dataset should be drawn from water bodies of a similar type so that the response of the biota throughout the dataset is not influenced significantly by major geological or geographical factors. Typically, these samples are drawn from separate water bodies conforming to these properties within a territory, spanning a long gradient that encompasses the biological boundaries of interest. In practice, multiple samples from the same water body but separated temporally, can also be used, though there are risks of pseudoreplication (Hurlbert, 1984) and spatial autocorrelation (Diniz-Filho et al., 2003; Legendre, 1993) if the ratio of water bodies to samples is low. An essential feature of the data is that it should span a sufficiently wide pressure gradient to allow robust characterisation of the ecological response. To achieve this there may be situations where different types of water body within a country can be merged to produce larger datasets, or where collaboration with neighbouring countries may be the most productive option.

The general situation can conveniently be envisaged as a scatter plot between biology (expressed as an Ecological Quality Ratio, EQR) and nutrient concentrations for similar water bodies, to which a regression line is fitted (Fig. 1). The threshold concentration for nutrients to support good status may be set at the point where the biological threshold intersects the chemistry (Fig. 1a) or at a position above or below this point (the upper or lower 95% confidence limit, for example). The use of the upper limit gives a low probability of restoring water bodies back to good status, but minimises the risk of a water body being wrongly downgraded (i.e. chemical threshold is exceeded despite biology at good status; Fig 1b). The lower limit is more precautionary, giving a high probability of restoring water bodies back to good status, but will result in more water bodies being wrongly downgraded (Fig 1c). There are, in other words, trade-offs between the “false positives” and “false negatives” that a particular threshold will produce. The scale of this problem will decrease as the predictive power of the regression equation increases, and when pressures other than nutrients have less influence on biological status (Phillips et al., 2019).

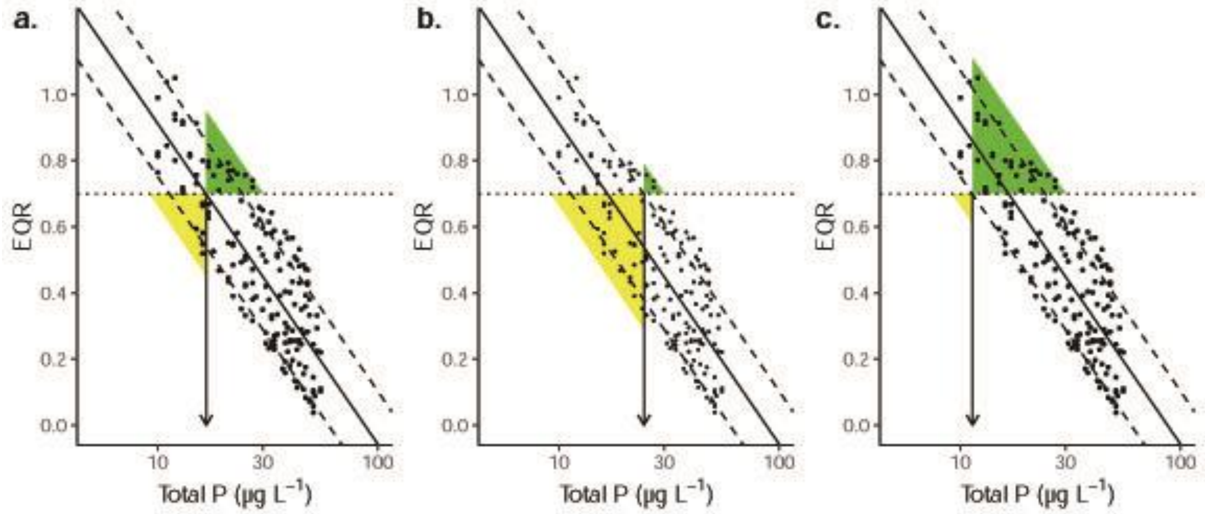

Figure 1: Hypothetical relationship between total phosphorus and biological EQR, showing regression line with confidence intervals (dotted lines). Horizontal line shows the biological good/moderate threshold (0.7 in this example), vertical lines show intersection with regression line ± confidence intervals marking potential good/moderate threshold values for total phosphorus using, a) intersection with best fit line, b) upper confidence line, c) lower confidence line. Triangles mark areas where classification mismatches occur, green (biology Good but phosphorus Moderate) and yellow (biology Moderate or worse but phosphorus Good) using three different approaches to interpretation.

The situation shown in Fig. 1 is typical for the relationship between phytoplankton and total phosphorus in lakes, where nutrients are typically the principal pressure. By contrast, there is often much greater scatter in the pressure response relationships in rivers, estuaries and coastal waters (Salas Herrero et al., 2019). There are many potential reasons (Page et al., 2012; Harris and Heathwaite, 2012; O'Hare et al., 2018) including interactions with other stressors (Van den Brink et al., 2019) or by interactions amongst species (Pérez-Ruzafa et al., 2002). In such cases, relationships between nutrient concentration and biological status have a high level of uncertainty. Appropriate target values therefore become difficult to establish and carry greater risks of false positive or negative classifications.

Scatter plots often reveal patterns that clearly do not conform to a simple linear relationship. In the extreme they can show a 'wedge'-type relationship to which an upper-quantile line can be fitted, providing an estimate of the highest level of nutrient that is theoretically consistent with good status (Figure 2a). Such a pattern would be caused where other stressors (e.g. hydromorphological alteration) are present, depressing ecological status independently of nutrients. An inverted wedge

(Figure 2b) can also occur where other factors mitigate the effect of nutrient enrichment. In lakes and coastal waters this might be grazing by zooplankton or zebra mussels (Caraco et al., 2006; ; Higgins et al., 2011; Pérez-Ruzafa et al., 2002); in rivers and estuaries it might be shade or flow reducing primary production, or the toxic effects of herbicides (e.g. Polazzo & Rico, 2021) or metals. In this case a lower quantile line could be fitted and used to generate a target concentration derived from the lowest concentration of nutrient associated with good status.

There is an ongoing debate on how to set nutrient targets when other stressors are present and definitive guidance cannot yet be offered. In the meantime, Feld et al. (2016) provide a toolkit for investigating the role of multiple stressors whilst Phillips et al. (2019) use synthetic datasets to examine the extent to which interactions amongst stressors might affect relationships. The complexity of multiple stressor interactions has also raised interest in the use of more sophisticated approaches such as null models that consider underlying mechanistic assumptions for better predicting multi-stressor effects at different organisational levels from individual to communities (e.g. Schäfer and Piggott, 2018). More recently, a general framework to aid identification and assessment of the interactive effects of multiple stressors on aquatic ecosystems (Van der Brink et al. 2019) was tested in anthropogenic influenced environments such as ditches (Bracewell et al., 2019), floodplains (Monk et al., 2019) and estuaries (O'Brien et al., 2019).

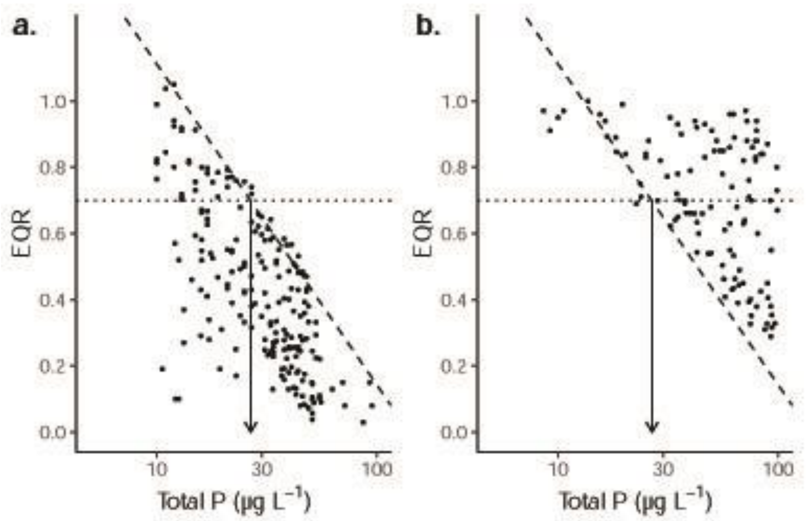

Figure 2: Hypothetical relationship between total phosphorus and biological EQR where multiple pressures occur. a) Regression of an upper quantile (e.g. 95th percentile); b) regression of a lower quantile (e.g. 5th percentile). Horizontal lines show the biological good/moderate threshold, vertical lines show intersection with line marking potential good/moderate threshold values for total phosphorus.

Preliminary visualisation and overview of method selection

The first step of any process of developing nutrient thresholds is visualisation of the data. Preliminary data visualisation does not need any complicated software – basic functions in Excel may suffice – but it provides the insights into the distribution of data along the gradient of interest that will guide subsequent method selection (Zuur et al. 2010). This visualisation will also reveal whether or not transformation of axes is necessary to ensure linearity, and the extent to which heteroscedasticity is an issue that will complicate analyses (see above). Some curvature may remain even after axes have been transformed, in which case visualisation will help to identify the linear range (but see below for statistical approaches for identifying “breakpoints”). All methods described in this paper have advantages and disadvantages, depending on circumstances and the most appropriate method for any situation is summarised in Figure 3. Application of causal analysis principles (Grace & Irvine 2019) may also be helpful. We recommend, however, that as many approaches as possible are applied to the data and results evaluated with an awareness of the statistical properties of the dataset prior to selecting a regulatory threshold. For example, a dataset for which type II regression is a suitable approach could also be analysed using categorical methods. Each will generate a different threshold but together, and when combined with knowledge of the water bodies under examination, as well as local regulatory needs, will give a more nuanced insight into the most appropriate threshold.

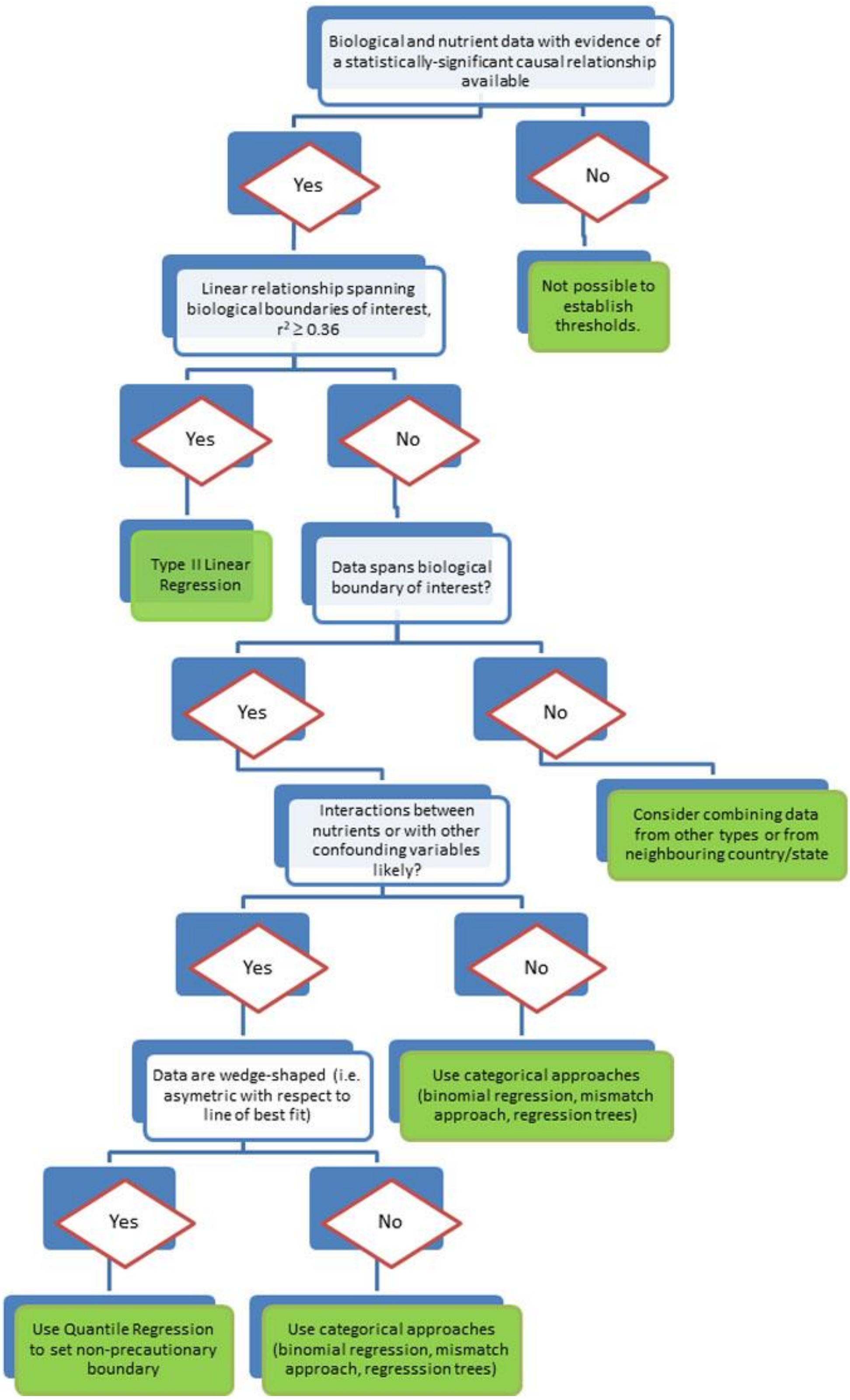

Figure 3. A flow-chart to select the most appropriate method in the toolkit for situations where nutrient thresholds need to be established.

Statistical approaches to establishing thresholds

Linear regression

Where there is a strong relationship between biology and nutrients, fitting regression models to data that span the pressure gradient is recommended. These models assume a linear response between variables which can often be achieved by log transformation of nutrient concentration data. Even after this, however, visual inspection may reveal nonlinearity, often with sigmoid responses (i.e. with regions at the extremes of the distribution, where there is little response of the biology to changed nutrient concentrations). Preliminary visualisation of the data using generalized additive modelling, followed by segmented regression (Muggeo, 2021) to identify breakpoints is recommended. Thresholds of interest need to be within the linear portion of the graph if linear regression is to be effective.

It is also important that there is not a high proportion of 'less than' values in the stressor data set (due to limits of detection) as these constitute 'censored' data which incorrectly 'anchor' regression relationships and exert undue influence on the modelled gradient (Helsel, 2010). Where this is the case specialist advice should be obtained. As the WFD requires status to be expressed as an EQR on a 0-1 scale, it is also common practice for values that are >1.0 to be rounded down ("capped") to 1.0. This, too, is a form of censoring that can distort natural gradients, introducing curvature and increasing uncertainty. Wherever possible, we recommend the use of uncapped data and, where this is not possible, alternative approaches such as generalized linear models with logit link functions, or binomial regression should be considered

Ordinary least squares (OLS) regression models establish a relationship between nutrients and biological status by minimising the variation in the dependent variable whilst assuming no error in the predictor variable. When using such models to establish nutrient thresholds changing nutrient concentrations are assumed to influence ecological condition, suggesting that the former is the independent variable whilst the latter is dependent. However, for this particular purpose we are inferring the chemical concentration at a particular point on the biological scale, in effect inverting this assumption. Furthermore, nutrient concentrations are also influenced by the biology through uptake, especially when dissolved inorganic nutrients are used in the regression. This means that neither is, strictly, independent of the other. In practice, however, as neither biological nor chemical condition is measurable without error, OLS regression will underestimate the true slope of the relationship (Legendre, 2013) and thus influence the estimation of a nutrient concentration at the biological threshold.

The alternative is to use a type II regression (Sokal and Rohlf, 1995), which minimises the variation of both dependent and independent variables. The disadvantages of a type II regression are that it is less appropriate where the purpose of the model is to make predictions (Legendre and Legendre, 2012), and, secondly, it is more difficult to interpret uncertainty (Smith, 2009). It is also important to only apply type II regression to relationships with a strong correlation ($r \geq 0.6$; $r^2 = 0.36$) as suggested by Jolicoeur (1990) as the method will generate a line with a slope significantly different from zero with random data. It should be noted however, that if the threshold EQR being predicted is close to the mean EQR of the data, the choice of regression method will have little effect as both type I (i.e. OLS regression) and type II fitted lines pass through the mean of x and y. Where r^2 values are high (>0.6) there is little practical difference in the nutrient boundaries resulting from type I or type II, but for less certain relationships differences are more substantial.

When type II reduced major axis regression was applied to a dataset of macrophyte communities from streams in NW Europe, predictions of total phosphorus concentrations to support high and good ecological status using the line of best fit (i.e. Fig. 1a) were 14 and 37 μg L^{-1} respectively (Fig. 4). When predictions were based on the upper quartile of residuals, the corresponding figures were 25 μg L^{-1} for high status and 66 μg L^{-1} for good status (Poikane et al., 2021).

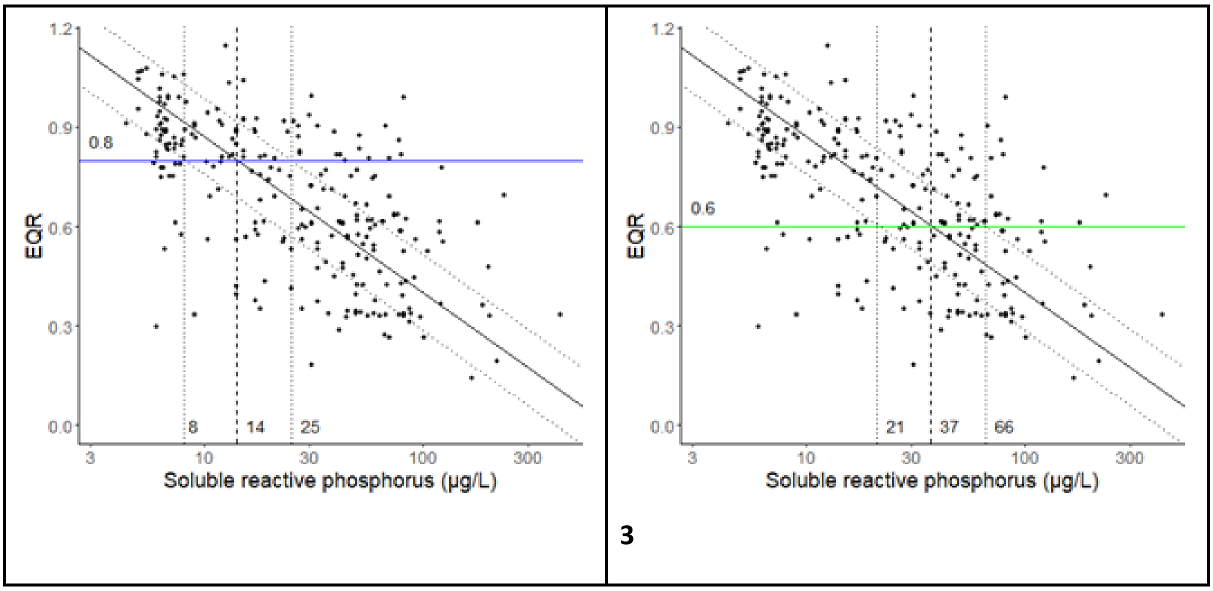

Figure 4. Relationship between EQRs for macrophytes and soluble reactive phosphorus for low alkalinity lowland rivers in NW Europe. Estimates of threshold concentrations for high/good and good/moderate status assume EQRs of 0.8 and 0.6 respectively. Solid line shows type II RMA regression and dashed lines show upper and lower quartiles of residuals. Modified from Poikane et al. 2021.

Multivariate regressions

A development from the use of bivariate regressions is the inclusion of extra predictor variables into the models from which thresholds are obtained. These could include variables that account for natural variability of the dependent ecological variable, such as alkalinity and altitude, in order to increase precision. This approach can also bypass the need for artificial divisions of water bodies into "types".

This does not necessarily require multivariate modelling if such variables can be combined within a single index value. In the United Kingdom, for example, river phosphorus standards are based on models which use the alkalinity and altitude of the site, along with the biological EQR (macrophytes and phytobenthos combined, in this case) to set standards (UK TAG, 2014).

The first step in deriving these phosphorus standards was to predict the concentration of phosphorus expected if a site were at 'reference condition' — an estimate of the natural condition of the site. The prediction used values of alkalinity and altitude to represent key geological and geographic factors that determine a site's natural phosphorus concentration. The next step was to calculate the ratio between the estimated 'natural' phosphorus concentration and the concentration actually measured at the site (this is, in effect, a phosphorus 'EQR'). A regression equation was then developed to describe the link between the biological data (also expressed as an EQR) and these phosphorus ratios. Provided a site's alkalinity and altitude are known, this model, following rearrangement of the equation, can estimate the likely ranges of phosphorus concentrations for each status class at any site (Figure 5).

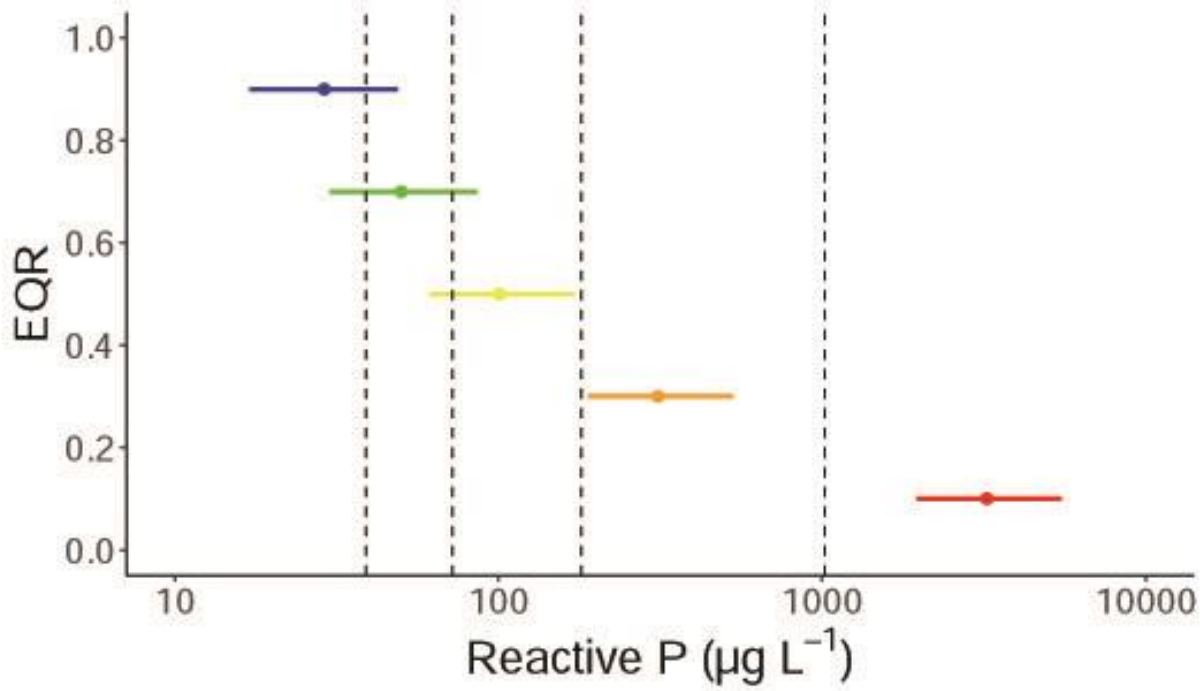

Figure 5. The relationship between reactive P concentration and EQR (minimum of macrophytes and phytobenthos) for a typical lowland high alkalinity river in England. Phosphorus standards are shown as vertical dotted lines and are set at the midway point of the overlapping error bars for the five classes (blue = high; green = good; yellow = moderate; orange = poor; red = bad). This position represents a concentration at which there is equal statistical confidence (P = 0.5) of the biology being in adjacent classes.

For any site, the phosphorus concentrations at the midpoint of the biological class are calculated using the following equation:

P concentration =

$$10^{\wedge}((1.0497 \times \log10\ (\mathrm{EQR}) + 1.066) \times (\log10\ (\text{reference condition RP}) - \log10(3{,}500)) + \log10(3{,}500)).$$

where:

EQR = class midpoint ecological quality ratio (minimum of macrophytes and phytobenthos), i.e. 0.9, 0.7, 0.5, 0.3, 0.1 for High, Good, Moderate, Poor and Bad respectively.

Reference condition RP = phosphorus concentration expected at reference condition, calculated as:

$$\text{Reference condition RP} = 10^{\wedge}(0.454\ (\log10\mathrm{alk}) - 0.0018\ (\mathrm{altitude}) + 0.476)$$

where:

Alk = alkalinity (as mg L^{-1} CaCO3)

Altitude = height above sea level (metres)

For a hypothetical lowland (28 m above sea level) high alkalinity (2.1 meq L^{-1}) river in the UK, the midpoint of high status, estimated by this method, is 29 µg L^{-1}, with a likely range of 17 – 48 µg L^{-1} whilst the midpoint and range of good and moderate status are 50 (30 – 85) µg L^{-1} and 69 (54 - 85) µg L^{-1} respectively moderate status. These error bars represent the range in the estimates of the phosphorus concentrations predicted by the regression model. As the ranges of adjacent status classes often overlap it is not possible to use these to set thresholds. Instead, the recommended phosphorus standards are set at the midway point of the overlapping error bars since this position represents a concentration at which there is equal statistical confidence ($P = 0.5$) of the biology being in adjacent classes.

A benefit of the approach described here is that it does not rely on dividing rivers into “types”. By using the alkalinity and altitude of the site concerned, the method derives phosphorus standards that are, in principle, specific to each point in a river. In contrast, most of the other approaches specify a single threshold applicable to the continuum of waters within a type, which could vary widely depending on how types are defined. By working with EQRs for both biology and nutrients this approach also has the advantage of extending the available gradient lengths for both stressor and response beyond what is likely to be available within individual river types. On the other hand, care is needed when applying such models in regions where calcium carbonate or related materials (‘lime’) are applied to agricultural land (or to mitigate acidification in low alkalinity rivers), as this may raise the alkalinity of the receiving water and indirectly influence the phosphorus target (Tappin et al., 2018). In theory, the natural alkalinity of a river could be modelled from underlying geology but this has not yet been incorporated into this assessment scheme, and would, in itself, be prone to uncertainty.

Multivariate modelling can also include additional pressure variables. For example, Poikane et al. (2019b) used models that included both TP and TN to derive nutrient threshold values for lakes based on their relationships with macrophytes. These models had higher precision and thus greater confidence in the resulting threshold values. Multivariate models have the potential disadvantage that they generate an unlimited range of potential pairs of threshold values which can complicate their use for management. However, Poikane et al. (2019b) provided a solution by determining the threshold for the most likely TP:TN ratio using a bivariate plot overlain by the good/moderate threshold EQR value expressed as a contour (Figure 6).

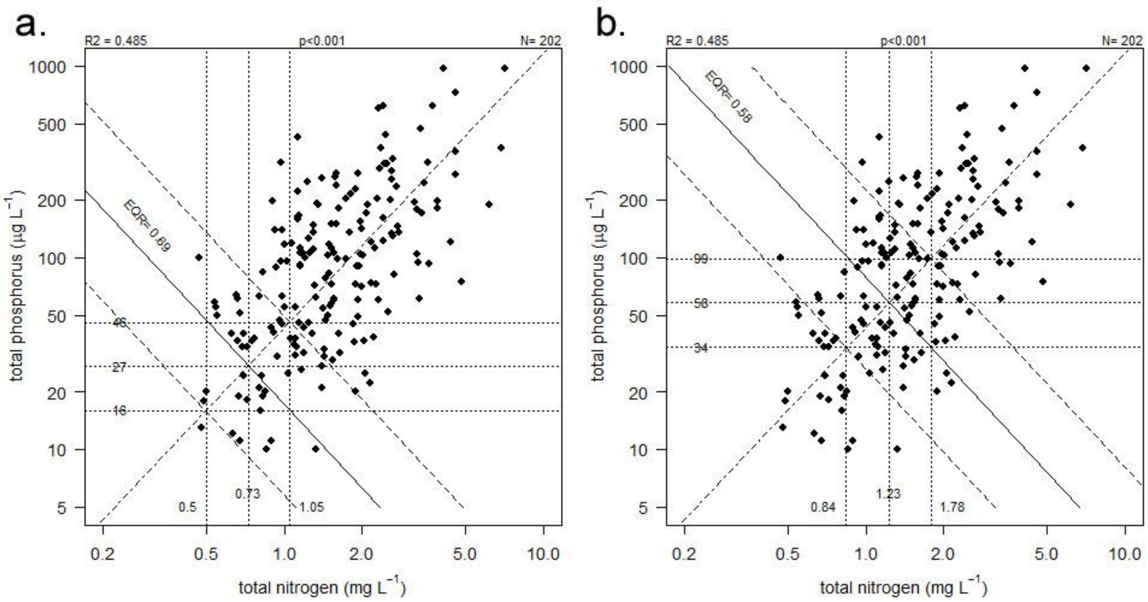

Figure 6. Relationship between mean TP and TN in high alkalinity very shallow lakes (L-CB2). Dotted lines show contours of predicted TN and TP concentration when macrophyte EQR is at a) high/good and b) good/moderate threshold (±25th & 75th residuals of prediction). Horizontal and vertical lines show intersection with RMA regression of observed TP and TN showing threshold concentrations for good status.

Binomial regression

In practice, ecological status assessment collapses the EQR, a continuous variable, into five ecological status classes and it is also possible to derive nutrient thresholds directly from these. Binomial logistic regression offers a method for fitting a logistic model to categorical data using a binary response, either side of the threshold of interest (e.g. “moderate or worse” = 1 and “good or better” = 0). This approach has the advantage of being applicable in situations where the relationship between nutrients and biology is weak and is less sensitive to the position of the data cloud relative to the threshold of interest. It also overcomes the limitations of EQR values capped at an upper value of 1.0. The quality of the statistical model can be tested using a variety of methods and binomial regression can be combined with other approaches.   For example, it could be applied after linear regression, to determine the probability that predicted nutrient concentrations will protect ecological status. Furthermore, logistic regression could also be applied for risk assessment of management practices, while allowing the effect of nutrient reduction targets proposed by authorities in relation to Ecological Status classification to be tested.

Results obtained using simulated data (Phillips et al., 2019) suggest it is likely to be the best alternative to linear regression models, provided that other stressors are not also influencing biological status. The resulting model can however also be used to determine threshold values at different levels of probability of being ‘moderate or worse’, providing an adequate alternative when the size of classes (i.e. “biology good or better” vs. “biology moderate or worse”) is not balanced and when there are multiple pressures or unaccounted environmental factors (see Wallace et al., 2011).

Bivariate regression was not appropriate for deriving thresholds for dissolved inorganic nitrogen (DIN) using phytoplankton in estuaries (“transitional waters”) from five EU Member States (ES, IE, NL, PT, UK) bordering the North East Atlantic due to the weak relationship between biology and chemistry ($r^2 = 0.22$). Instead, binomial regression gave estimated threshold concentrations with a 50% probability of being in either category were 44 µM for high/good status and 80 µM for good/moderate status (Fig. 7). These estimates are equivalent to the “line of best fit” in a bivariate regression (i.e. Fig. 1a) and, by adjusting the probability it is also possible to estimate precautionary boundaries (20 and 32 µM respectively) and non-precautionary boundaries (102 and 196 µM respectively). Once again, interactions from other stressors is a key consideration when deciding whether this method is appropriate (Phillips et al., 2019).

a. b.

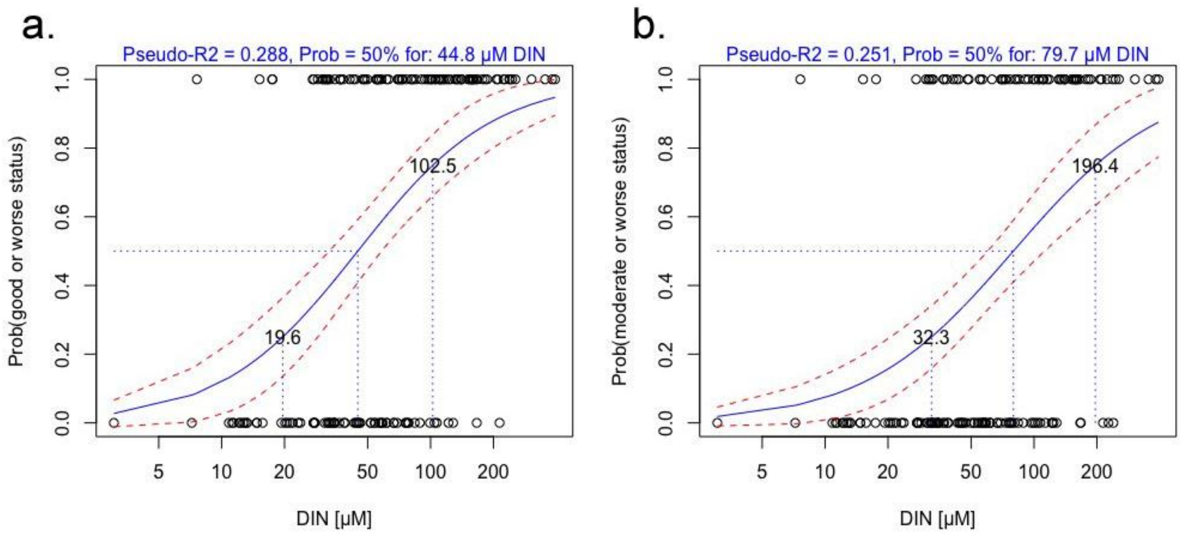

Figure 7. Binomial logistic regression showing the probability of ecological status being a. “good or lower status” and b) “moderate or lower status” for phytoplankton in estuaries (“transitional waters”) in five countries bordering the NE Atlantic. Lines show potential threshold values for DIN at different probabilities of being in good or worse status and moderate or worse. Modified from Salas-Herrero et al. (2019).

Other categorical methods

Another approach is simply to set a nutrient threshold that minimises the mismatch between ecological status and the supporting element (Figure 8a). Use of bootstrap sampling and a LOESS curve fit make the approach more robust and testing using synthetic data has shown that it is more sensitive to data uncertainty than logistic regression (Phillips et al., 2019) and requires a relatively large data set. This approach is conceptually similar to the conditional probability approach which uses non-parametric deviance reduction in order to determine the change point (Paul and McDonald, 2005).

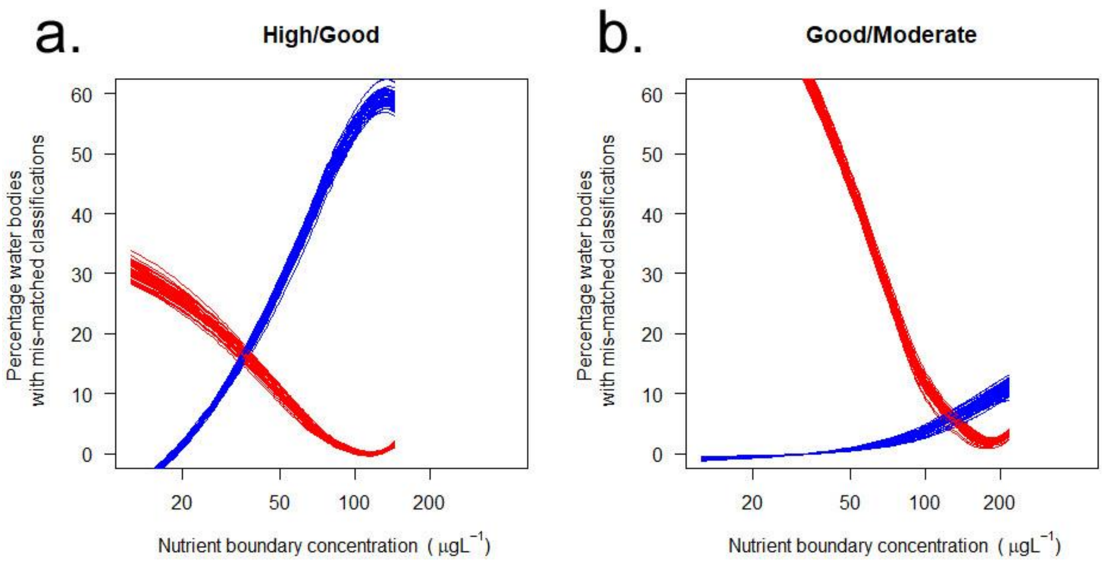

Figure 8: Minimisation of mis-match between nutrient and biology for H/G and G/M boundaries respectively as a means of setting nutrient boundaries, based on the European very large river dataset (Kelly et al., 2019a). The y axis shows the percentage of misclassified records when biological and nutrient classifications are compared, vertical lines mark the range of crossover points where the misclassification is minimised, together with the mean nutrient concentration, after bootstrap iterations (each line indicates a sub-sample of the data set selected at random).

Categorical methods, in other words, are a valid option in situations where there are well defined states that need to be protected but there are few heavily impacted sites with which to ‘anchor’ a regression model. However, the precision of estimates will not be any greater when the relationship is very noisy than would be the case if a regression was used. The categorical approach is similar to a type 1 regression of nutrients on biology, because it assumes that all the uncertainty is in nutrients and that biology is the (error-free) 'predictor'. Problems will also arise if there are few water bodies in each category or if there are missing categories.

Decision trees

Decision tree methods such as classification and regression trees can also be used as alternatives to logistic regression. These work by iteratively splitting the data into distinct subsets, with the splits chosen in such a way that entropy in the resulting subsets is minimised. Decision tree outputs typically have high accuracy and stability and should be straightforward to understand even for people with non-statistical backgrounds. Use of decision trees is also possible in the presence of multiple stressors and they can be used to model complex datasets (Mori et al., 2019). In contrast to other modelling approaches such as neural networks, techniques such as classification and regression trees are able to handle different types of predictor variables and accommodate missing data and outliers. They can fit complex nonlinear relationships and handle interactions between predictors (Lemm et al., 2021).

In the simplest case, decision trees can generate likely thresholds for a single variable. Kagalou et al. (2021) used this approach to derive thresholds for TP in deep natural lakes in Greece) were 13 $\mu g\ L^{-1}$ and 49 $\mu g\ L^{-1}$ TP respectively for high and good status (Kagalou et al., 2021). However, these methods can also be used for simultaneous generation of thresholds for several variable. When used to derive TP and TN in Hungarian lakes, for example (Figure 9; G. Varbiro, unpublished data), threshold values for high status were TP < 128 $\mu g\ L^{-1}$ and TN < 960 $\mu g\ L^{-1}$ whilst for good status these were TP ≥ 128 $\mu g\ L^{-1}$ and TN < 1709 $\mu g\ L^{-1}$. The importance of cross-validation to indicate the size of the tree that is appropriate for the decision to be made increases as the number of variables increases but is always recommended in order to avoid overfitting (Flach, 2019).

In case of classification and regression trees, the accuracy of the model may be increased by bootstrapping methods such as fitting multiple trees to minimise the risk of overfitting. Multiple tree models such as boosted regression trees (Ridgeway, 2006) or random forest methods (Breiman, 2001) increase diversity among the classification trees by resampling the data with replacement, bootstrapping and random changes in the predictive variable sets over the different tree induction processes. The validity of the models can be evaluated through the use of misclassification or confusion matrices which summarizes the performance of the final classifications using metrics such as accuracy, misclassification rate, null error rate or Cohen's Kappa (Liu et al., 2011).

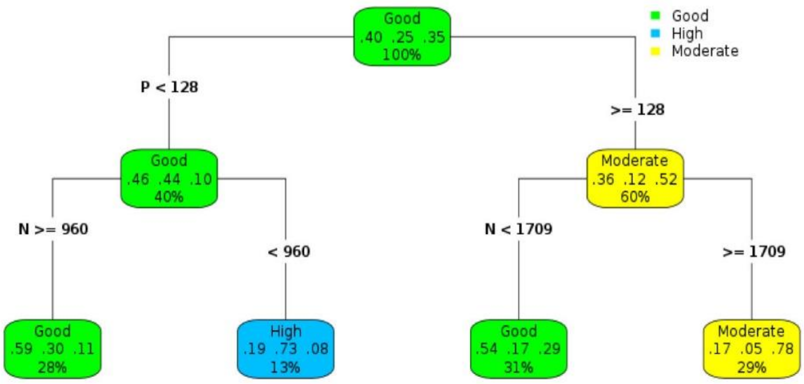

Figure 9: a) Classification decision tree of total phosphorus (TP) on biological classes (High, Good, Moderate) for phytoplankton in deep natural lakes in Greece (Kagalou et al., 2021). Each node shows (from left) the predicted class, the predicted probability of each class and the percentage of observations in the node (High, Good, Moderate). b) Classification decision tree of total phosphorus(P) and total nitrogen (N) on biological classes (High, Good, Moderate) for phytoplankton of Hungarian lakes.

Quantile regression

Most of the methods described above are unlikely to yield meaningful precautionary boundaries when other stressors confound nutrient-biology relationships (Fig. 2; Phillips et al., 2019). In such cases the variance around the mean of the response variable is itself a function of the explanatory variable, leading to a wedge-shaped distribution. Under these circumstances, quantile regression may be more appropriate. This is a variant of conventional least squares regression analysis. Whereas least squares regression aims to predict the mean of the response variable for a given value of the predictor variable, quantile regression aims to predict different aspects of the statistical dispersion of points.

Quantile regression can be implemented through packages such as ‘quantreg’ (Koenker, 2016) within R and the toolkit includes some scripts that could be adapted for other uses. The values produced by an upper quantile of a relationship between EQR and nutrients will be inherently less precautionary than those produced by a conventional “line of best fit”. In effect, an upper quantile defines the maximum value of a response variable likely at any given value of the explanatory variable and is useful where one or more additional pressures drive the response variable, overriding the influence of nutrients to reduce status.

As a result, the use of quantile regression for setting thresholds needs to be considered with care. A wedge-shaped distribution might, for example, indicate that nutrients are not the primary factor influencing the biota for sites included in the data set. This, in turn, might provoke investigations into the role of other stressors and better regulation of these might need to take priority over nutrient control (Spears et al., 2021). The upper quantile will, nonetheless, provide a value that can serve as an interim target, by identifying thresholds above which nutrients are almost certainly driving ecological status. In a few cases (e.g. sites of high conservation interest), the use of a lower quantile, which will produce a precautionary threshold value, may be appropriate.

The confidence with which the slope and intercept of a quantile function can be estimated will decrease towards the extreme of the distribution, due to a likely variation of the 'conditional density of the response' (Koenker, 2011). The selection of an appropriate quantile for threshold setting is essentially a value judgement, partially conditioned by dataset size, data distribution, but it should be based on knowledge of the importance of nutrients versus other pressures and of how their interactions affect the sensitivity of the BQEs to nutrients. We suggest that values of the 25th and 75th percentiles are most likely to be appropriate for data with inverted wedge- or wedge-shaped scatter plots, respectively. Where an upper-quantile approach is used, leading to less precautionary thresholds, it is particularly important that the threshold is validated by independent evidence (Phillips et al. 2018).

Data from phytoplankton in estuaries draining into the NE Atlantic has a clear wedge-shape distribution (Salas-Herrero et al., 2019; Figure 10). Boundaries obtained using quantile regression were of a similar order, albeit slightly more lenient, as the upper (less precautionary) ranges obtained using logistic regression (Fig. 7). Bear in mind, however, that data from five countries, each with slightly different approaches to collecting both chemical and biological data, had to be merged and harmonised in order to obtain a dataset covering a sufficiently large range to permit estimates to be made, particularly for countries not covering the full gradient of disturbance (e.g. PT, which only had High status samples).

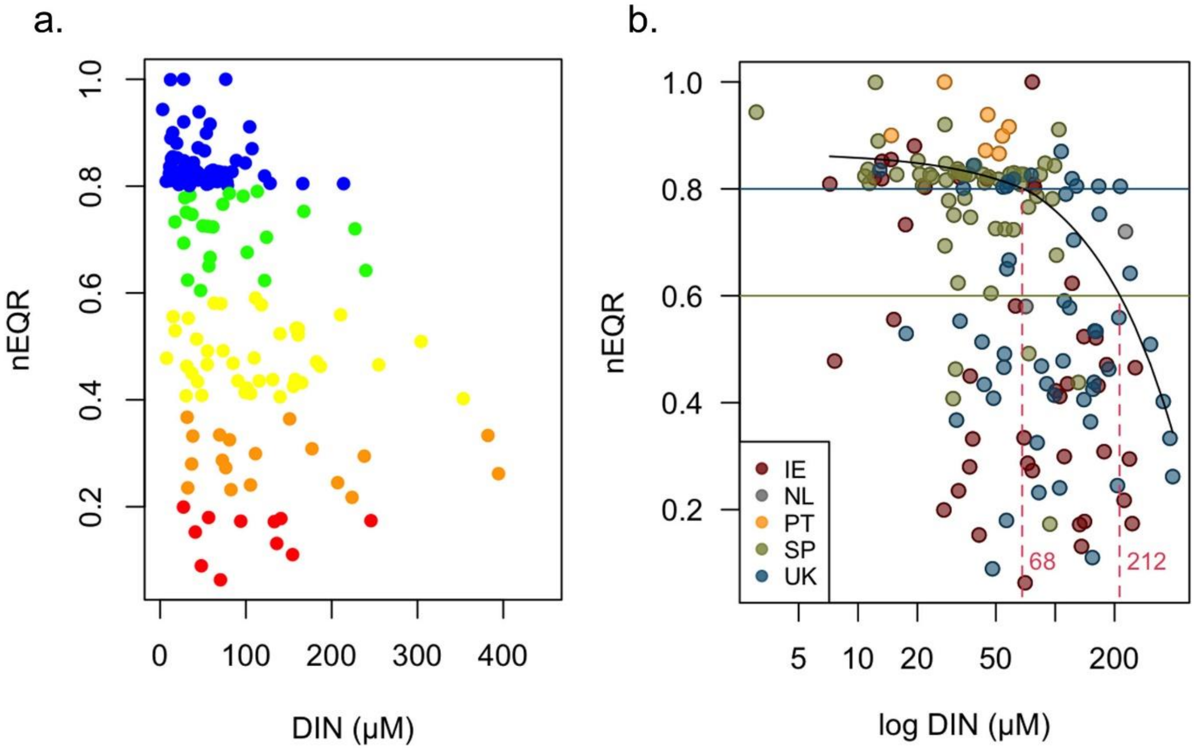

Figure 10. Relationship between dissolved inorganic nitrogen (DIN) concentrations (µM) and normalised phytoplankton EQRs (nEQR) in NE Atlantic estuaries. Observations coloured by WFD ecological status (High to Bad, n=160) (a.) and quantile regression (Additive Quantile Regression Smoothing rqss using quantreg; Koenker, 2016) fit of nutrient with nEQR (b.) based on 160 observations from Ireland (IE), Netherlands (NL), Portugal (PT), Spain (SP) and the United Kingdom (UK). Horizontal lines indicate nEQR boundaries at H/G and G/M, and vertical lines the nutrient boundaries, respectively for H/G and G/M, at the 70th quantile. Modified from Salas-Herrero et al. 2019.

Discussion: selecting appropriate threshold values

Setting targets for nutrients (and, indeed, other physico-chemical variables that influence ecological condition) for aquatic systems is rarely straightforward. Applying a range of approaches to the same dataset can result in a wide range of potential threshold values with very different implications for regulators and, by extension, for the type of developments permitted within river basins, or the programmes of measures intended to reduce such pressures. It is important, therefore, that any exercise to develop nutrient thresholds includes rigorous validation steps to ensure that regulatory boundaries are robust. These steps may include checking threshold estimates against values published in the literature (including those based on experimental studies) and with boundaries used by other countries with similar water bodies, as well as examining the condition of other components of the biota (Piroddi et al., 2021). The data from which nutrient targets are obtained often contains considerable uncertainty and heteroscedasticity which confounds attempts to use simple statistical methods. Yet, at the same time, the use of nutrient targets is linked to the regulatory regime within which they operate and, as there are likely to be significant financial implications, they need to be established using approaches that are not just statistically robust but which can be readily understood at all levels within organisations (not just by technical specialists) and by the wider public. Our discussion is, therefore, framed around four themes: ecology, statistics, regulation and communication, all of which overlap with each other, and all of which need to be considered when setting nutrient targets.

Ecological aspects of setting nutrient targets

In many respects, this is the most straightforward aspect of the process: setting nutrient targets assumes that there is a causal relationship between nutrients and biology, even though demonstrating this causality in the field may, in practice, be challenging (Poikane et al., 2021). Whilst this has been demonstrated many times in lakes (e.g. Anonymous, 1982), the story is more nuanced in other ecosystems where the nutrient signal is likely to be confounded by other pressures (Matthaei et al., 2010; Piggott et al., 2012, Gameiro and Brotas, 2010; Salas-Herrero et al., 2019; Polazzo and Rico, 2021) and where retention times are lower. Our experience is that interactions from these other stressors frequently complicate the process of setting targets, due to limitations of predicting the combined effect of stressors (Orr et al., 2020). Of the techniques included in the toolkit, quantile regression allows boundaries (albeit non-precautionary) to be set in the face of additional stressors whilst decision trees and multivariate models also both show potential. However, it is also likely that solutions to achieve desirable ecological states will have to be worked out for each water body separately, with outcomes depending upon the “cocktail” of stressors present and the tractability of each of these to remediation.

Measurements of both the environmental chemistry and the biological communities from which these targets are derived are, necessarily, greatly simplified expressions of the complex interactions which occur in reality and, consequently, both prone to uncertainties. A discussion of chemical sampling frequencies and design (Kreyling et al., 2018) and appropriate determinands (e.g. Ptacnik et al., 2010; Poikane et al., 2021) is beyond the scope of this paper although we recognise both as potentially significant contributors to the overall uncertainty in relationships. Similarly, biological communities are collapsed into summary metrics calibrated against principal pressure gradients

(Borja et al., 2011). Whilst this is far from ideal from the point of view of understanding ecosystem dynamics, one positive consequence of the WFD is that these high-level expressions of ecological health have been subject to intercalibration, to ensure that Member States share a similar level of ambition towards WFD targets (Birk et al., 2013; Kelly et al., 2014; Lopez y Royo et al., 2011; Simboura et al., 2008). It is generally assumed that photosynthetic components of the biota are used to set nutrient targets although there is no reason why heterotrophic organisms should not also be used and, indeed, secondary effects of eutrophication such as hypolimnetic deoxygenation (Winfield et al., 2008) or habitat alteration (Law et al., 2019) can be sensitive indicators of condition.

In addition, there should be reasonable grounds for expecting a causal relationship between nutrients and biology without significant interference from other stressors. This means that it can be assumed that a water body with a biota consistent with elevated nutrient concentrations is, in theory, capable of being restored back to pre-impact conditions (presumed at or close to the "natural" state). All the approaches considered in this paper are, in other words, underpinned by a "space-for-time" substitution (Pickett, 1988). The limitations of this with respect to setting nutrient targets are considered in Taylor et al. (2018); however, we argue that the use of large spatial datasets does, at least, mean that between-water body variation can be acknowledged in ways that are not possible using experimental approaches.

A further question that should be asked is whether metrics that are developed as broad indicators of ecological integrity are appropriate for deriving nutrient standards. Another recurring theme in this paper is the importance of acknowledging the role played by other stressors and appreciating the scale of inherent uncertainty. Thus, whilst the relationship between nutrients and ecological status cannot be ignored (as it is the basis by which the overall success of national and regional management programs will be judged under existing frameworks), there is also a case for developing alternative metrics focussed on particular stressors. Leboucher et al. (2020), for example, recognise the role played by mass effect and dispersal processes on phytobenthos assemblages in rivers and this raises the possibility that variants of metrics that are capable of filtering out "noise" from such processes may permit purer insights into biology-nutrient relationships.

Statistical aspects of setting nutrient targets

Much of this paper has addressed the issues around uncertainty in the datasets from which nutrient targets are derived. Whilst this uncertainty can be reduced by using adequate data sets (see General Principles, above) and categorising water bodies into similar types, the complexity of the ecological interactions involved, coupled with stochastic effects, will always result in a variation in biological status (or EQR) at any nutrient concentrations for any water body. This uncertainty can be broken down into three components: adequate data, statistical approach and model uncertainty.

Adequate data

Methods described in this paper, and in many others that suggest means of setting nutrient targets (e.g. Dodds et al., 2010; Hausmann et al. 2016; Poikane et al., 2019b) depend upon datasets derived from sampling that captures the spatial and temporal variability of water bodies of similar types within a region. It is possible to use long-term datasets (e.g. HELCOM, 2013); however, our experience is that there are few locations where appropriate data have been collected in a consistent manner for long enough for this to represent a viable alternative to approaches based on spatial datasets. Similarly, experimental approaches (e.g. Taylor et al., 2018) are also possible but require considerable investment in resources at a few locations, results of which then have to be extrapolated to cover all water bodies in a region. By contrast, spatial datasets allow standards to be set that take account of the range of variation within a region so long as:

- there is a means of grouping water bodies into ecologically meaningful types such that their response to nutrients will be similar (Lyche Solheim et al., 2019);
- data capture the full range of spatial and temporal variation, including the part of the gradient where biology is most sensitive to nutrients; and,
- there are analytical procedures for both explanatory and response variables, with means for accounting for differences between laboratories (as large datasets invariably involve several analysts). In the context of target-setting for the WFD, the use of biological metrics with harmonised status class boundaries (Birk et al., 2013; Poikane et al., 2015) should mean that targets represent similar levels of ambition between Member States.

Whilst data that fulfil these criteria should be available from national monitoring programmes, there will be situations where individual Member States do not have enough data, and collaboration between countries is necessary (Salas-Herrero et al., 2019).

Choice of statistical approach

Each of the methods described in this paper will differ in suitability depending upon the particular circumstances associated with each exercise. For example, type II regression is the preferred regression model, as it minimises deviations along both EQR and nutrient axis. Similarly, estimates derived from categorical methods depend upon factors such as the relative number of water bodies in each biological class and the width of that class. Thus, these categorical estimates are also uncertain, and users need to be sure that their data sets are representative of the regions to which they will be applied. Uncertainty can be estimated by fitting a binary logistic model, or by the use of bootstrapping when estimating misclassification rates but results are dependent on the reliability of the underlying biological status classification.

In view of these factors, we recommend that the flow chart (Figure 3) is followed but, wherever possible, as many methods as possible are applied to the data and that the predictions (which represent a range of possible threshold values) are compared. The range in thresholds reflects differences in concepts and assumptions underpinning the statistical methods used. Data where r^2 values are low will have higher uncertainty and some relationships may be so uncertain it is impossible to make a reliable or useful prediction. In such cases, the answer may be to return to the field and gather new – likely different – data and address the problem from a different perspective.

Model uncertainty

Regression models provide the best estimate of the 'average' response of water bodies in a data set. Individual water bodies will fall above or below that line, partly due to data and statistical uncertainty, but also because of uncertainty inherent in the model itself. This can be expressed using the interquartile range of the residuals of the regression models, from which a further range of threshold values, the 'possible range', can be predicted. The magnitude of the possible range depends on the quality of our conceptual model. For example, in mesotrophic deep lakes phytoplankton biomass is highly dependent on phosphorus and thus the relationship between phytoplankton EQR and TP is normally very good ($r^2 > 0.65$: Phillips et al., 2008). Conversely, in rivers phytobenthos and macrophytes will respond to many other pressures and be subject to other influences such as grazing, shade or variation in substratum and simple pressure-response models will result in boundaries with very large uncertainty bands. Until it is possible to improve our conceptual models to include a mechanistic understanding of multi-stressor effects and develop statistical models that incorporate a wider range of variables (Schäfer and Piggot, 2018), we need to recognise and manage this variation when we set threshold values for management.

Regulatory aspects of setting nutrient targets

The uncertainty described above is more than just an interesting ecological and statistical paradox for academic scientists to unravel at their leisure: it has to work within regulatory structures governed by national and international legislation. Those involved in regulation stress clarity and stability as two key factors that need to be considered: the former gives managers an indication of the benefits that can be expected when a particular target is applied whilst the latter enables the likely investment (e.g. in improved wastewater treatment) to be calculated and costed. Bearing this in mind, we recognise three types of ecological target that can be achieved using the approaches described here, and suggest possible applications of each within the EU:

- **Most likely threshold value** derived from regression best fit lines (Figure 1a), and the mismatch approach. The likelihood of achieving good status with the mean nutrient concentration as the threshold would be 50 % and there would be a moderate risk of downgrading a water body despite biology being at good status, when the 'one out, all out' rule is applied.
- **Most certain that biology dictates status** derived from either an upper quantile of linear regression residuals (Fig. 1b) or higher probability value of logistic regression. Only 25 % of water bodies would be classified as not being at good status based on nutrients when their biological status was good. However, the benefits of reducing unnecessary downgrades due to the "one out all out" rule are offset by the low level of precaution in the target. This would be a good option if many water bodies in a region were not achieving good status, and the primary roles of the target are to prioritise water bodies for remediation, and to establish the importance of nutrients relative to other pressures. It would, however, not be a good option if the purpose was to prevent deterioration of water bodies that were already at good status. Where multiple stressors are suspected this approach would indicate nutrient concentration which would be relatively certain of causing a downgrade of biological status. Whilst achieving this target should ensure a reduction in secondary effects, further interventions may be required before good status is achieved.
- **Most protective threshold value** derived from the lower quantiles of the linear regression residuals (Figure 1c), a lower quantile of a quantile regression, or a lower probability value from binary logistic regression should be used. This ensures that a majority of water bodies within a type will achieve good status but will result in unnecessary downgrades of status using the 'one out all out' rule, with implications for expenditure on programmes of measures unless additional safeguards in the decision-making process can be applied.

Communication of nutrient targets

A recurring theme of this paper has been the complex interactions between biology and nutrients that occur in many natural systems, and the advanced statistical approaches required to deal with this. However, these targets then have to be implemented within regulatory regimes, with cost implications that may run into millions of Euros. The final element to be considered, therefore, is the communication of results from those who develop the standards to those who are affected by their implementation. Ecological targets may well push the capability of "best available technology" as well as testing consumer's enthusiasm for changes in land-use practices, so those engaged in setting them should be prepared for their results to face close scrutiny from those responsible for their implementation.

Our experience is that box and whisker plots and mismatch plots (Figure 8) are the easiest visual means for explaining nutrient targets. Scatter plots (Figure 5) are also useful, so long as the relationship between nutrients and biology is strong enough for the position of the line of best fit within the data cloud to be obvious to a non-specialist. Advanced statistical methods such as TITAN undoubtedly have a role to play in setting nutrient targets (e.g. Roubeix et al., 2016; 2017; Hausmann et al., 2016) but the output from these methods can be difficult for those without prior knowledge of the method to interpret. Whilst we have encouraged the use of binomial logistic regression for setting standards, interpretation of results produced using unbalanced datasets has been difficult for those without a strong statistical background.

Once we start to consider the role of multiple stressors the situation becomes considerably more complex, particularly if multivariate models are used. Such models typically generate multiple potential target values contingent on other predictor variables included within the model (Poikane et al., 2019b) but do not remove the difficulty of communicating targets derived by this method. Extreme climatic events, such as droughts, floods and strong winds, are expected to exacerbate nutrient pollution effects by influencing the nutrient load and concentration in aquatic ecosystems (Wetz and Yoskowitz, 2013; Malta et al. 2017). Nutrient targets set in current conditions must not be communicated as static thresholds as they might need adjustments in the future in order to reflect these additional stressors and protect from such, likely to intensify, future scenarios.

Conclusions

Whilst we have dwelt at length on the problems associated with setting nutrient targets, our final message is one of hope rather than despair. An appreciation of the uncertainties associated with spatial datasets, coupled with a willingness to collaborate with neighbours where necessary and an awareness of how targets will be used should allow plausible estimates to be established for many, if not most, types of water body. These can be corroborated by comparison with targets set for similar water bodies elsewhere (see Tables 4.1 – 1.14 in Phillips et al., 2018) and, in turn, provide a basis for strategic planning for nutrient management within Member States. Recognition of the limitations of these methods, at the same time, sets an agenda for research, firstly to better understand the interactions between nutrients and other stressors, but also to broaden the toolkit (perhaps looking beyond the established suite of ecological metrics) in order to gain better insights into the needs of individual water bodies.

Author contributions

Martyn G. Kelly: Conceptualization, Writing- Original draft preparation, Writing – reviewing and editing; **Geoff Phillips:** Conceptualization, Formal analysis, Software, Writing- Original draft preparation; **Heliana Teixeira:** Conceptualization, Formal analysis, Software, Writing- Original draft preparation; **Gabor Varbiro:** Conceptualization, Formal analysis, Software, Writing- Original draft preparation; **Fuensanta Salas Herrero:** Conceptualization, Writing- Original draft preparation, Project administration, Funding acquisition; **Nigel J. Willby:** Formal analysis, Writing – Review and Editing; **Sandra Poikane:** Conceptualization, Writing – Original draft preparation, Project administration, Funding acquisition.

Acknowledgements

We are very grateful to three referees for useful comments on the manuscript. HT thanks further support within CESAM (UIDB/50017/2020+UIDP/50017/2020).

References

Anonymous, 1982. Eutrophication of waters: monitoring, assessment and control. Organisation for Economic Cooperation and Development Environmental Directorate, Paris.

Birk, S., Chapman, D., Carvalho, L., Spears, B.M., Andersen, H.E., Argillier, C., Auer, S., Baattrup-Pedersen, A., Banin, L., Beklioğlu, M., Bondar-Kunze, E., et al. 2020. Impacts of multiple stressors on freshwater biota across spatial scales and ecosystems. *Nat. Ecol. Evol. 4*, 1060-1068.

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. Sci. Total Environ. 454-455, 490-499.

Borja, A., 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. Mar. Pollut. Bull. 60, 1143-1146.

Borja, A., Barbone, E., Basset, A., Borgersen, G., Brkljacic, M., Elliott, M., Garmendia, J.M., Marques, J.C., Mazik, K., Muxika, I., Neto, J.M., 2011. Response of single benthic metrics and multi-metric methods to anthropogenic pressure gradients, in five distinct European coastal and transitional ecosystems. Mar. Pollut. Bull. 62, 499-513.

Bracewell, S., Verdonschot, R.C., Schäfer, R.B., Bush, A., Lapen, D.R., Van den Brink, P.J., 2019. Qualifying the effects of single and multiple stressors on the food web structure of Dutch drainage ditches using a literature review and conceptual models. Sci. Total Environ. *684*, 727-740.

Breiman, L., 2001. Random forests. Machine Learning 45, 5–32.

Cade, B.S., Noon, B.R., 2003. A gentle introduction to quantile regression for ecologists. Front. Ecol. Environ. 1, 412-420.

Canning, A.D., Joy, M.K., Death, R.G., 2021. Nutrient criteria to achieve New Zealand's riverine macroinvertebrate targets. PeerJ 9:e11556.

Caraco, N.F., Cole, J.J., Strayer, D.L., 2006. Top down control from the bottom: regulation of eutrophication in a large river by benthic grazing. Limnol. Oceanogr. 51,, 664-670

Carvalho, L., Mackay, E.B., Cardoso, A.C., Baattrup-Pedersen, A., Birk, S., Blackstock, K.L., Borics, G., Borja, A., Feld, C.K., Ferreira, M.T., Globevnik, L. et al., 2019. Protecting and restoring Europe's waters: An analysis of the future development needs of the Water Framework Directive. Sci. Total Environ. *658*, 1228-1238.

Charles, D.F., Kelly, M.G., Stevenson, R.J., Poikane, S., Theroux, S., Zgrundo, A., Cantonati, M., 2021. Benthic algae assessments in the EU and the US: Striving for consistency in the face of great ecological diversity. *Ecol. Indic.* 121, e107082.

Conley, D.J., Paerl, H.W., Howarth, R.W., Boesch, D.F., Seitzinger, S.P., Havens, K.E., Lancelot, C., Likens, G.E., 2009. Controlling eutrophication: nitrogen and phosphorus. *Science 323*, 1014-1015.

Davies, S.P., Jackson, S.K., 2006. The biological condition gradient: a descriptive model for interpreting change in aquatic ecosystems. Ecol. Appl. 16, 1251-1266.

Debeljak M., Džeroski S., 2011. Decision Trees in Ecological Modelling. In: Jopp F., Reuter H., Breckling B. (Eds) Modelling Complex Ecological Dynamics. Springer, Berlin.

Diniz-Filho J.A.F., Bini, L.M., Hawkins, B.A., 2003. Spatial autocorrelation and red herrings in geographical ecology. Glob. Ecol. Biogeogr. 12, 53-64.

Dodds, W.K., Clements, W.H., Gido, K., Hilderbrand, R.H., King, R.S., 2010. Thresholds, breakpoints, and nonlinearity in freshwaters as related to management. J. N. Amer. Benthol. Soc. 29, 988-997.

Dolman, A.M., Mischke, U., Wiedner, C., 2016. Lake-type-specific seasonal patterns of nutrient limitation in German lakes, with target nitrogen and phosphorus concentrations for good ecological status. *Freshwat. Biol. 61*, 444-456.

European Environment Agency, 2018. European waters assessment. EEA Report No 7/2018. ISBN: 978-92-9213-947-6

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. O. J. L327, 1-88.

European Union, 2008. Directive 2008/56/EC of the European Parliament and of the Council of 17 June 2008 establishing a framework for community action in the field of marine environmental policy (Marine Strategy Framework Directive). O. J. L164, 19-40.

Feld, C.K., Segurado, P., Gutiérrez-Cánovas, C., 2016. Analysing the impact of multiple stressors in aquatic biomonitoring data: a cookbook with applications in R. Sci. Total Environ. 573, 1320-1339.

Flach, P., 2019. Performance evaluation in machine learning: the good, the bad, the ugly, and the way forward. In: Proceedings of the AAAI Conference on Artificial Intelligence 33, 9908-9914.

Free, G., Tierney, D., Little, R., Kelly, F., Kennedy, B., Plant, C., Trodd, W., Wynne, C., Caroni, R., Byrne, C., 2016. Lake ecological assessment metrics in Ireland: relationships with phosphorus and typology parameters and the implications for setting nutrient standards. Biol. Environ. 116B, 191-204.

Friedland. R., Macias, D., Cossarini, G., Daewel, U., Estournel, C., Garcia-Gorriz, E., Grizzetti, B., Grégoire, M., Gustafson, B., Kalaroni, S., Kerimoglu, O., et al., 2021. Effects of Nutrient Management Scenarios on Marine Eutrophication Indicators: A Pan-European, Multi-Model Assessment in Support of the Marine Strategy Framework Directive. Front. Mar. Sci. 8: 596126.

Gameiro, C., Brotas, A., 2010. Patterns of Phytoplankton Variability in the Tagus Estuary (Portugal). Estuar. Coast. 33, 311-323.

Glibert, P.M., 2017. Eutrophication, harmful algae and biodiversity—Challenging paradigms in a world of complex nutrient changes. Mar. Pollut. Bull. *124*, 591-606.

Grace, J.B., Irvine, K.M., 2019. Scientist's guide to developing explanatory statistical models using causal analysis principles. Ecology 101, e02962.

Grizzetti, B., Vigiak, O., Udias, A., Aloe, A., Zanni, M., Bouraoui, F., Pistocchi, A., Dorati, C., Friedland, R., De Roo, A., Benitez Sanz, C., Leip, A., Bielza, M., 2021. How EU policies could reduce nutrient pollution in European inland and coastal waters. EarthArXiv. doi:10.31223/X5CC91

Guildford, S.J., Hecky, R.E., 2000. Total nitrogen, total phosphorus, and nutrient limitation in lakes and oceans: Is there a common relationship? Limnol. Oceanogr. *45*, 1213-1223.

Harris, G.P., Heathwaite, A.L., 2012. Why is achieving good ecological outcomes in rivers so difficult? Freshwat. Biol. 57, 91-107.

Hausmann, S., Charles, D.F., Gerritsen, J., Belton, T.J., 2016. A diatom-based biological condition gradient (BCG) approach for assessing impairment and developing nutrient criteria for streams. Sci. Total Environ. 562, 914-927.

HELCOM, 2013. Approaches and methods for eutrophication target setting in the Baltic Sea region. Baltic Sea Environment Proceedings No. 133.

Helsel, D., 2010. Much ado about next to nothing: incorporating nondetects in science. Ann. Occup. Hyg. 54, 257–262.

Hering, D., Borja, A., Carstensen, J., Carvalho, L., Elliott, M., Feld, C.K., Heiskanen, A.-S., Johnson, R.K., Moe, J., Pont, D., Lyche-Solheim, A., v. 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. doi:10.1016/j.scitotenv.2010.05.031

Hering, D., Carvalho, L., Argillier, C., Beklioglu, M., Borja, A., Cardoso, A.C., Duel, H., Ferreira, T., Globevnik, L., Hanganu, J., Hellsten, S., Jeppesen, E., Kodeš, V., Solheim, A.L., Nõges, T., Ormerod, S., Panagopoulos, Y., Schmutz, S., Venohr, M., Birk, S.,2015. Managing aquatic ecosystems and water resources under multiple stress — An introduction to the MARS project. Sci. Total Environ. 503-504: 10-21.

Higgins, S.N., Vander Zanden, M.J., Joppa, L.N., Vadeboncoeur, Y., 2011. The effect of dreissenid invasions on chlorophyll and the chlorophyll: total phosphorus ratio in north-temperate lakes. Can. J. Fish. Aquat. Sci., 68, 319-329.

Hurlbert, S.H., 1984. Pseudoreplication and the design of ecological field experiments. Ecol. Monogr. 54, 187-211.

Jolicoeur, P., 1990. Bivariate allometry: interval estimation of the slopes of the ordinary and standardized normal major axes and structural relationship. J. Theor. Biol. 144, 275-285.

Kagalou, I., Ntislidou, C., Latinopoulos, D., Kemitzoglou, D., Tsiaoussi, V., & Bobori, D. C., 2021. Setting the Phosphorus Boundaries for Greek Natural Shallow and Deep Lakes for Water Framework Directive Compliance. *Water 13*, e739.

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.

Kelly, M.G., Chiriac, G., Soare-Minea, A., Hamchevici, C., Birk, S., 2019a. Defining ecological status of phytobenthos in very large rivers: A case study in practical implementation of the Water Framework Directive in Romania. *Hydrobiologia*, *828*, 353-367.

Kelly, M.G., Chiriac, G., Soare-Minea, A., Hamchevici, C., Juggins, S., 2019b. Use of phytobenthos to evaluate ecological status in lowland Romanian lakes. Limnologica 77, 125682

Koenker, R., 2011. Additive models for quantile regression: model selection and confidence bandaids. Braz. J. Probab. Stat. 25, 239-262.

Koenker, R., 2016. Package ‘quantreg’: Quantile Regression. R package version 4.76.

Kreyling, J., Schweiger, A.H., Bahn, M., Ineson, P., Migliavacca, M., Morel-Journel, T., Christiansen, J.R., Schtickzelle, N., Larsen, K.S., 2018. To replicate, or not to replicate – that is the question: how to tackle nonlinear responses in ecological experiments. Ecol. Lett. 21, 1629-1638. doi:10.1111/ele.13134

Law, A, Baker, A, Sayer, C, Foster, G., Gunn, I.D.M., Taylor, P., Pattison, Z., Blaikie, J, Willby, N.J., 2019. The effectiveness of aquatic plants as surrogates for wider biodiversity in standing fresh waters. *Freshw. Biol*. 64, 1664– 1675. https://doi.org/10.1111/fwb.13369

Leboucher, T., Tison-Rosebery, J., Budnick, W.R., Jamoneau, A., Vyverman, W., Soininen, J., Boutry, S., Passy, S.I., 2020. A metacommunity approach for detecting species influenced by mass effect. *J. Appl. Ecol. 57*, 2031-2040.

Legendre, P., 1993. Spatial autocorrelation: trouble or new paradigm?. *Ecology 74*, 1659-1673.

Legendre, P., 2013. Model II regression user's guide, R edition. (A tutorial within R session, R package lmode2) http://cran.r-project.org/web/packages/lmodel2/vignettes/mod2user.pdf.

Legendre, P., Legendre, L.F., 2012. Numerical ecology. Elsevier, Amsterdam.

Lemm, J. U., Venohr, M., Globevnik, L., Stefanidis, K., Panagopoulos, Y., van Gils, J., Posthuma L., Kristensen P., Feld K. P., Mahnkopf J., Hering D., Birk, S., 2021. Multiple stressors determine river ecological status at the European scale: Towards an integrated understanding of river status deterioration. *Glob. Change Biol. 27*, 1962-1975.

Liu, C., White, M., Newall, G., 2011. Measuring and comparing the accuracy of species distribution models with presence/absence data. Ecography 34, 232-243.

Lopez y Royo, C., Pergent, G., Alcoverro, T., M.C.Buia, M.C., Casazza, G., Martínez-Crego, B..Silvestre, F., Perez, M., Romero., J. 2011. The seagrass *Posidonia oceanica* as indicator of coastal water quality: Experimental intercalibration of classification systems. Ecol. Indic. 11, 557-563.

Lyche Solheim, A., Globevnik, L., Austnes, K., Kristensen, P., Moe, S. J., Persson, J. et al., 2019. A new broad typology for rivers and lakes in Europe: Development and application for large-scale environmental assessments. *Sci. Total Environ. 697*, e134043.

Malta, E.J., Stigter, T.Y., Pacheco, A., Dill, A.C., Tavares, D., Santos, R., 2017. Effects of external nutrient sources and extreme weather events on the nutrient budget of a southern European coastal lagoon. Estuar. Coast. 40, 419-436.

Matthaei, C.D., Piggott, J.J., Townsend, C.R., 2010. Multiple stressors in agricultural streams: interactions among sediment addition, nutrient enrichment and water abstraction. J. Appl. Ecol. 47, 639-649.

Moe, S.J., Lyche-Solheim, A., Soszka, H., Gołub, M., Hutorowicz, A., Kolada, A., Picinska-Fałtynowicz, J., Białokoz, W., 2015. Integrated assessment of ecological status and misclassification of lakes: The role of uncertainty and index combination rules. Ecol. Indic. 48, 605-615.

Monk, W.A., Compson, Z.G., Choung, C.B., Korbel, K.L., Rideout, N.K., Baird, D.J., 2019. Urbanisation of floodplain ecosystems: Weight-of-evidence and network meta-analysis elucidate multiple stressor pathways. *Sci. Total Environ. 684*, 741-752.

Mori, N., Debeljak, B., Škerjanec, M., Simčič, T., Kanduč, T., Brancelj, A., 2019. Modelling the effects of multiple stressors on respiration and microbial biomass in the hyporheic zone using decision trees. Wat. Res. 149, 9-20.

Muggeo, V.M.R., 2021. Package ‘segmented’. 1.3-4. https://cran.r-project.org/web/packages/segmented/segmented.pdf [date accessed: 08.06.2021]

Nõges, P., Argillier, C., Borja, Á., Garmendia, J.M., Hanganu, J., Kodeš, V., Pletterbauer, F., Sagouis, A., Birk, S., 2016. Quantified biotic and abiotic responses to multiple stress in freshwater, marine and ground waters. Sci. Total Environ. 540, 43-52.

O'Brien, A.L., Dafforn, K.A., Chariton, A.A., Johnston, E.L., Mayer-Pinto, M., 2019. After decades of stressor research in urban estuarine ecosystems the focus is still on single stressors: a systematic literature review and meta-analysis. Sci. Total Environ. *684*, 753-764.

O’Hare, M.T., Baattrup-Pedersen, A., Baumgarte, I., Freeman, A., Gunn, I.D.M., Lázár, A.N., Wade, A.J., Bowes, M.J., 2018. Responses of aquatic plants to eutrophication in rivers: a revised conceptual model. Front. Plant Sci. 9:451.

Ojaveer, H., Eero, M., 2011. Methodological Challenges in Assessing the Environmental Status of a Marine Ecosystem: Case Study of the Baltic Sea. PLoS ONE 6, e19231.

Orr, J.A., Vinebrooke, R.D., Jackson, M.C., Kroeker, K.J., Kordas, R.L., Mantyka-Pringle, C., Van den Brink, P.J., De Laender, F., Stoks, R., Holmstrup, M., Matthaei, C.D., Monk, W.A., Penk, M.R., Leuzinger, S., Schäfer, R.B., Piggott J.J., 2020. Towards a unified study of multiple stressors: divisions and common goals across research disciplines. Proc. R. Soc. Lond. B, 287, e20200421.

Page, T., Heathwaite, A.L., Moss, B., Reynolds, C., Beven, K.J., Pope, L., Willows, R., 2012. Managing the impacts of nutrient enrichment on river systems: dealing with complex uncertainties in risk analysis. Freshwat. Biol. 57, 108-123.

Paerl, H.W., 2009. Controlling eutrophication along the freshwater–marine continuum: dual nutrient (N and P) reductions are essential. Estuar. Coast. *32*, 593-601.

Paerl, H.W., Scott, J.T., McCarthy, M.J., Newell, S.E., Gardner, W.S., Havens, K.E., Hoffman, D.K., Wilhelm, S.W., Wurtsbaugh, W.A., 2016. It takes two to tango: when and where dual nutrient (N & P) reductions are needed to protect lakes and downstream ecosystems. Env. Sci. Tech. *50*, 10805-10813.

Paul, J.F., McDonald, M.E., 2005. Development of empirical, geographically specific water quality criteria: a conditional probability analysis approach. *J. Am. Water Resour. Assoc. 41*, 1211-1223.

Pérez-Ruzafa, A., Gilabert, J., Gutiérrez, J.M., Fernández, A.I., Marcos, C., Sabah, S., 2002. Evidence of a planktonic food web response to changes in nutrient input dynamics in the Mar Menor coastal lagoon, Spain. Hydrobiologia 475, 359–369.

Phillips, G., Kelly, M.G., Teixeira, Salas, F., Free, G., Leujak, W., Pitt, J.-A., Solheim, A.L., Várbíró, G. & Poikane, 2018. *Best practice for establishing nutrient concentrations to support good ecological status.* European Commission Joint Research Centre Science for Policy Report, Publications Office of the European Union, Luxembourg. doi:10.2760/84425

Phillips, G., Teixeira, H., Poikane, S., Salas, F., Kelly, M.G**.,** 2019. Establishing nutrient thresholds in the face of uncertainty and multiple stressors: a comparison of approaches using simulated data sets. *Sci. Total Environ.* 684, 425-433.

Phillips, G., Pietiläinen, O.P., Carvalho, L., Solimini, A., Solheim, A.L. and Cardoso, A.C., 2008. Chlorophyll–nutrient relationships of different lake types using a large European dataset. *Aquatic Ecol. 42*, 213-226.

Pickett, S.T.A., 1988. Space-for-time substitution as an alternative to long-term studies. In: In: Likens, G.E. (Ed.). Long-term Studies in Ecology, Approaches and Alternatives. Springer-Verlag, New York. pp. 110-135.

Piggott, J.J., Lange, K., Townsend, C.R. & Matthaei, C.D., 2012. Multiple stressors in agricultural streams: a mesocosm study of interactions among raised water temperature, sediment addition and nutrient enrichment. PLoS ONE 7, e49873.

Piroddi, C., Akoglu, E., Andonegi, E., Bentley, J.W., Celić, I., Coll, M., Dimarchopoulou, D., Friedland, R., de Mutsert, K., Girardin, R., Garcia-Gorriz, E., Grizzetti, B., Hernvann, P.-Y., Heymans, J.J., Müller-Karulis, B., Libralato, S., Lynam, C.P., Macias, D., Miladinova, S., Moullec, F., Palialexis, A., Parn, O., Serpetti, N., Solidoro, C., Steenbeek, J., Stips, A., Tomczak, M.T., Travers-Trolet, M., Tsikliras, A.C.,

2021. Effects of nutrient management scenarios on marine food webs: a pan-European assessment in support of the Marine Strategy Framework Directive. Front. Mar. Sci. 8, e596797.

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., 201). A hitchhiker's guide to European lake ecological assessment and intercalibration. Ecol. Indic. 52, 533-544.

Poikane, S., Kelly, M.G., Salas Herrero, F., Pitt, J.A., Jarvie, H.P., Claussen, U., Leujak, W., Lyche Solheim, A., Teixeira, H., Phillips, G., 2019a. Nutrient criteria for surface waters under the European Water Framework Directive: Current state-of-the-art, challenges and future outlook. *Sci. Total Environ. 695*, e133888.

Poikane, S., Phillips, G., Birk, S., Free, G., Kelly, M., Willby, N., 2019b. Deriving nutrient criteria to support "good" ecological status in European lakes: an empirically based approach to linking ecology and management. *Sci. Total Environ.* 650, 2074-2084.

Poikane, S., Várbíró, G., Kelly, M.G., Birk, S., Phillips, G., 2021. Estimating river nutrient concentrations consistent with good ecological condition: More stringent nutrient thresholds needed. *Ecol. Indic*. *121*, e107017.

Polazzo, F., Rico, A., 2021. Effects of multiple stressors on the dimensionality of ecological stability. Ecol. Lett. 24, 1594-1606.

Prato, S., La Valle, P., De Luca, E., Lattanzi, L., Migliore, G., Morgana, J.G., Munari, C., Nicoletti, L., Izzo, G., Mistri, M., 2014. The ''one-out, all-out'' principle entails the risk of imposing unnecessary restoration costs: A study case in two Mediterranean coastal lakes. Mar. Pollut. Bull. 80, 30-40.

Ptacnik, R., Anderson, T., Tamminen, T., 2010. Performance of the Redfield ratio and a family of nutrient limitation indicators as thresholds for phytoplankton N vs. P limitation. Ecosystems 13, 1201-1214.

Ridgeway, G. 2006 Generalized boosted regression models. Documentation on the R package 'gbm', version 1.5-7. http://www.ipensieri.com/gregr/gbm.shtml

Roubeix, V., Danis, P.-A., Feret, T., Baudoin, J.-M., 2016. Identification of ecological thresholds from variations in phytoplankton communities among lakes: contribution to the definition of environmental standards. Environ. Monit. Assess. 188, e246.

Roubeix, V., Daufresne, M., Argillier, C., Dublon, J., Maire, A., Nicolas, D., Raymond, J.-C., Danis, P.-A., 2017. Physico-chemical thresholds in the distribution of fish species among French lakes. Knowl. Manag. Aquat. Ecosyst. 418, e41.

Salas-Herrero, F ., Teixeira, H., Poikane, S., 2019. A Novel Approach for Deriving Nutrient Criteria to Support Good Ecological Status: Application to Coastal and Transitional Waters and Indications for Use. Front. Mar. Sci. 6, e255.

Schäfer, R.B., Piggott, J.J., 2018. Advancing understanding and prediction in multiple stressor research through a mechanistic basis for null models. Glob. Change Biol. 24, 1817-1826.

Simboura, M., Reizopolu, S., 2008. An intercalibration of classification metrics of benthic macroinvertebrates in coastal and transitional ecosystems of the Eastern Mediterranean ecoregion (Greece). Mar. Pollut. Bull. 56, 116-126

Smith, R. J., 2009. Use and misuse of the reduced major axis for line-fitting. Am. J. Phys. Anthropol. 140, 476-86.

Sokal, R. R., Rohlf, F. J., 1995. Biometry: the principles and practice of statistics in biological research. WH Freeman & Co, San Francisco.

Spears, B.M., Chapman, D., Carvalho, L., Rankinen, K., Stefanidis, K., Ives, S., Vuorio, K., Birk, S., 2021. Assessing multiple stressor effects to inform climate change management responses in three European catchments. *Inland Waters* DOI: 10.1080/20442041.2020.1827891.

Søndergaard, M., Lauridsen, T.L., Johansson, L.S., Jeppesen, E., 2017. Nitrogen or phosphorus limitation in lakes and its impact on phytoplankton biomass and submerged macrophyte cover. *Hydrobiologia 795*,35-48.

Tappin, A.D., Navarro-Rodriguez, A., Comber, S.D., Worsfold, P.J., 2018. The role of alkalinity in setting water quality metrics: phosphorus standards in United Kingdom rivers. *Environ. Sci.: Process. Impacts 20*, 1361-1372.

Taylor, J.M., Back, J.A., Brooks, B.W., King, R.S., 2018. Spatial, temporal, and experimental: three study design cornerstones for establishing defensible numeric criteria for freshwater ecosystems. J. Appl. Ecol. 55, 2114-2123.

Tibby, J., Richards, J., Tyler, J.J., Barr, C., Fluin, J., Goonan, P., 2019. Diatom-water quality thresholds in South Australian streams indicate a need for more stringent water quality guidelines. Mar. Freshwater Res. 71, 942-952.

Tomal, J. H., & Ciborowski, J. J., 2020. Ecological models for estimating breakpoints and prediction intervals. Ecol. Evol. 10, 13500-13517.

Torres, M.A., Casini, M., Huss, M., Otto, S. A., Kadin, M., Gårdmark, A., 2017. Food-web indicators accounting for species interactions respond to multiple pressures. Ecol. Indic. 77, 67-79.

UK TAG, 2014. 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.

Van den Brink, P.J., Bracewell, S.A., Bush, A., Chariton, A., Choung, C.B., Compson, Z.G., Dafforn, K.A., Korbel, K., Lapen, D.R., Mayer-Pinto, M., Monk, W.A., 2019. Towards a general framework for the assessment of interactive effects of multiple stressors on aquatic ecosystems: Results from the Making Aquatic Ecosystems Great Again (MAEGA) workshop. *Sci. Total Environ. 684*, 722-726.

Wallace, B. C., Small, K., Brodley, C. E., Trikalinos, T. A., 2011. Class imbalance, redux. In 2011 IEEE 11th international conference on data mining (pp. 754-763). IEEE. doi: 10.1109/ICDM.2011.33

Wetz, M.S., Yoskowitz, D.W., 2013. An ‘extreme’ future for estuaries? Effects of extreme climatic events on estuarine water quality and ecology. Mar. Pollut. Bull. 69, 7-18.

Winfield, I.J., Fletcher, J.M., James, J.B., 2008. The Arctic Charr (*Salvelinus alpinus*) populations of Windermere, UK: population trends associated with eutrophication, climate change and increased abundance of roach (*Rutilus rutilus*). Environ. Biol. Fishes 83, 25-35.

Zuur, A.F., Ieno, E.N., Elphick, C.S., 2010. A protocol for data exploration to avoid common statistical problems. Methods Ecol. Evol. 1, 3-14.