ISSN XXXX-XXXX

European Commission

JRC TECHNICAL REPORT

Establishing supporting element standards. A revised approach and applications.

Report to ECOSTAT by the Task Group on Supporting Physico-Chemical Elements, 2024

Phillips, G., Teixeira, H., Kelly, M. G., Lyche Solheim, A., Free, G., Salas Herrero, M.F., Kolada, A., Varbiro, G. and Poikane, S.

2024

Joint Research Centre

EUR XXXXX XX

Contact information
Name: Sandra Poikane
Address: European Commission, Joint Research Centre (JRC), Ispra, Italy.
Email: sandra.poikane@ec.europa.eu (or ENV-Water@ec.europa.eu)

EU Science Hub
https://joint-research-centre.ec.europa.eu

JRC137201

EUR XXXXX XX

Print	ISBN XXX-XX-XX-XXXXX-X	ISSN XXXX-XXXX	doi:XX.XXXX/XXXXXX	XX-XX-XX-XXX-XX-C
PDF	ISBN XXX-XX-XX-XXXXX-X	ISSN XXXX-XXXX	doi:XX.XXXX/XXXXXX	XX-XX-XX-XXX-XX-N

Add hyperlink to PDF doi: https://doi.org/XXXXXXX

Luxembourg: Publications Office of the European Union, 2024

How to cite this report: Phillips, G., Teixeira, H., Kelly, M. G., Lyche Solheim, A., Free, G., Salas Herrero, M.F., Kolada, A., Varbiro, G. and Poikane, S. (2024). *Establishing supporting element standards. A revised approach and applications*. Publications Office of the European Union. Luxembourg, 2024, doi:XX.XX/XXXXX, JRC137201.

Contents

Abstract

The Water Framework Directive requires member states to determine thresholds for a range of supporting chemical quality elements that will support good ecological status. This is a fundamental step in maintaining natural biodiversity as well as ensuring ecosystem service provision. All countries have defined thresholds, which have been used to report status in river basin planning cycles. The original intention of the work was to determine from empirical analysis a likely range of boundary values using pan-European data, which could be used to help inform any revision of current national boundary values. However, the analysis revealed relatively high levels of uncertainty, emphasising the challenge in establishing these boundary values. This highlighted the need for a more robust method. Binary logistic regression presented advantages over other methods, especially for complex data. However, a probability value must be seleted objectively to derive the nutrient threshold. This can be achieved by comparing the proportions of matching and mismatching status classifications of nutrients and a BQE using a confusion matrix. The method is applied to selected supporting elements for rivers and lakes.

.

Acknowledgements

We are very grateful to individuals and organisations who sumitted data for use in this report. We thank all member states for comments that improved this document.

.

Authors

Geoff Phillips: University of Stirling, UK

Heliana Teixeira: University of Aveiro, Portugal

Martyn Kelly: Bowburn Consultancy, Durham, UK

Anne Lyche Solheim: NIVA, Norway

Gary Free: EC JRC, Ispra, Italy

Fuensanta Maria Salas Herrero: EC JRC, Ispra, Italy

Agnieszka Kolada: Institute of Environmental Protection, Warsaw, Poland

Gabor Varbiro: Centre for Ecological Research, Debrecen, Hungary

Sandra Poikane: EC JRC, Ispra, Italy

1 Introduction

The Water Framework Directive requires member states to determine thresholds for a range of supporting chemical quality elements (SEs) that will support good ecological status (GES). All countries have defined thresholds, which have been used to report status for the last two river basin planning cycles. Unlike the methods used to assess biological status these thresholds are not required to be formally compared through an intercalibration and a relatively wide range of values are currently in use (Kelly et al. 2022). These values compiled by Kelly et al. (2022) were used in this report and don't incorporate any more recent changes at MS level. The original intention of the work was to determine from empirical analysis a likely range of boundary values for river total phosphorus, using pan-European data, that could be used to help inform any revision of current national boundary values. However, the analysis revealed relatively high levels of uncertainty, emphasising the challenge in establishing these boundary values and highlighting the need for a more objective consideration of the resulting misclassification rates when biotic and supporting element classes are compared. Thus in this report we present first a revised approach to setting boundary values and second, in separate documents, the results of its application to selected supporting elements for rivers, lakes and TRAC.

2 Approach to setting thresholds

2.1 Key Points

- We propose that a comparison of biota and SE classifications should be included in the process of establishing SE boundary values as this allows an assessment to be made of the relative rates of misclassification. In particular is the proportion of sites predicted to be in good or better biological status by the proposed SE boundary value, but that actually had a biota class that was moderate or worse, appropriate given the need for boundary values to be sufficiently precautionary to achieve WFD goals?

- While this step should be carried out following any modelling approach, we also suggest that using a binary logistic model is ideally suited to this approach.

2.2 Background

Previous guidance (Kelly et al. 2021) has outlined several statistical methods that can be used to determine the concentration of a supporting element that should ensure good ecological status. For example, a regression model between EQR and nutrient concentration enables a nutrient boundary to be determined by inverse prediction using the boundary of the EQR value. Once established the supporting element boundary is used to classify water bodies on the basis of observed concentrations of that element. Where water bodies are found to have nutrient concentrations lower than the boundary threshold they are assumed to be capable of supporting good status, where they are higher then remedial action is needed to reduce nutrients. This classification approach tends to forget the uncertainties associated with the boundary values used and inevitably there will be situations where the nutrient and biological classifications disagree. Provided we fit reliable models the comparative mis-classification rates should be low and the balance between false positive and negative rates will be similar. For example, “false positives” (i.e., chemistry indicates no enrichment when biology is challenged) might be catastrophic if the objective is conservation of rare taxa that are sensitive to enrichment. On the other hand, Phillips et al. (2019) demonstrated that interactions amongst stressors may complicate setting precautionary thresholds for nutrients and an approach that minimises “false negatives”, where biology is not impacted despite nutrients being higher than the threshold, at least ensures that only genuinely impacted sites are subject to regulation. Thus, the balance between false positives and negatives becomes an important consideration as boundary selection inevitably imposes either societal costs in achieving nutrient reduction or benefits of greater environmental protection. The need to consider the purpose to which thresholds will be put, as well as the properties of the relationship between biology and nutrients was stressed by (Kelly et al. 2021) However, this also raises questions about how the purpose can be expressed in objective manner that can inform threshold selection. Model uncertainty is clearly key to this, but given that the management objective is to determine boundaries for use in classification, we propose that a formal consideration of the proportions of true and false categorical predictions should be given a greater priority.

2.3 Comparing classifications

To fully understand the approach, we need to remind ourselves that in setting boundary values we are assuming that the range of concentration of the supporting element is different where the biology status is good in comparison to when it is not good. Taking rivers and total phosphorus (TP) concentration as an example Figure 2.1 shows an example where there is a clear difference in the TP distributions. Despite this there is a significant overlap (shaded area) and clearly wherever we place the boundary some sites that are biologically good will have TP concentrations higher than the boundary and others that are not in good status biologically will have TP concentrations below the boundary. We refer to these sites as being mis-classified and ideally, we should aim to minimise these, more on this below. For now, we can consider two simplistic approaches. First a precautionary approach would be to select the lowest TP concentration recorded from sites that were not in good status. In our example taking the minimum is clearly not appropriate but using the lower 10th percentile might be an appropriate value (Figure 2.1 red line). A less precautionary approach would be to take the highest or more realistically the 90th percentile of TP found in sites classified as good or better (Figure 2.1 blue line). In this example these values are relatively similar, but where the overlap of TP between the classes is greater (Figure 2.2) the differences are greater.

This raises question, should we select the lower or the higher thresholds? Alternatively, we could take the average of the values, the approach suggested in the current guidance by taking the average of the 25th and 75th quantiles (shown by the boxes in Figure 2.1 & Figure 2.2). Both are reasonable options, but it would be better to consider the implications more carefully. Which option provides the best overall comparative classification (least mis-classifications); what is the balance between the two types of mis- classification; are there more mis-classifications where the biology is moderate or worse but TP predicts they should be good or better (consequence of a less precautionary boundary) than those where the biology is in good or better status but TP is predicting it should not be (consequence of a precautionary boundary)?

It is challenging to provide objective advice, but in section 2.4 we introduce a method that quantifies mis-classification rates and thus helps establish the consequences of setting a particular threshold concentration. The approach is used extensively in other areas (e.g. drug testing) and is focused on the comparison of the biota and supporting element classifications.

Additionally, we emphasise the use of binary logistic models (section 2.5). These were included in the current guidance, but we now consider that they are the most useful modelling approach. Unlike linear modelling methods the model predicts for each concentration of the supporting element a probability of achieving good biological status. Thus, a decision is required to select this probability threshold which can be directly related to the relative proportions of positive and negative mis-classifications. Additionally, the model uses the biological class rather than the EQR, avoiding issues associated with different EQR class boundaries when combining data from several countries.

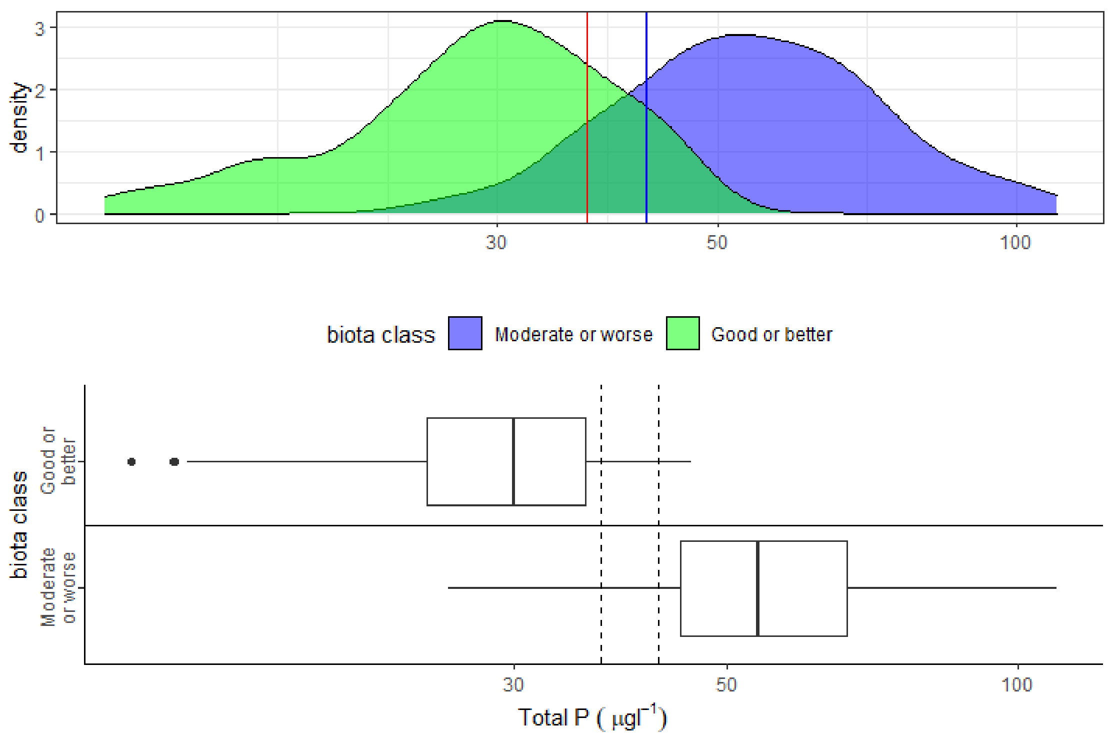

Figure 2.1. Density distribution and box plots showing the range of total phosphorus concentration in sites classified biologically into good or better and moderate or poor status. Data are synthesised to illustrate a good separation of total phosphorus concentration. vertical lines mark a range of potential good/moderate boundary values; 10th quantile of total phosphorus in sites classified as moderate (red line) and 90th quantile for sites classified as good (blue line).

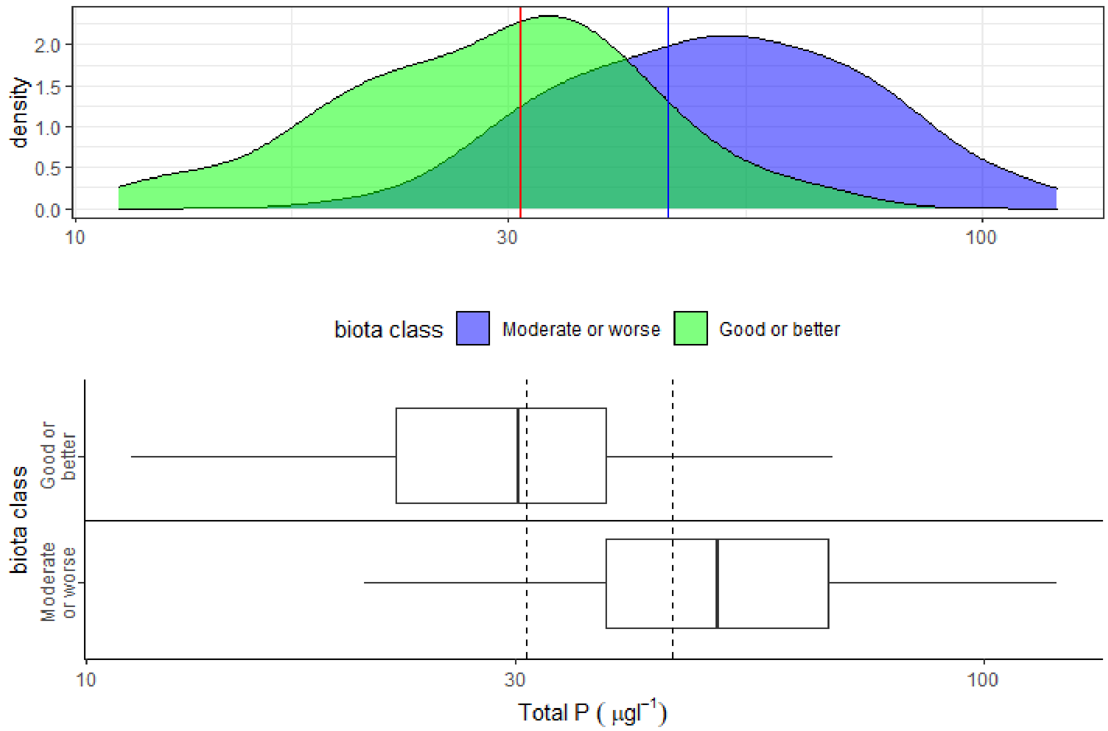

Figure 2.2. Density distribution and box plots showing the range of total phosphorus concentration in sites classified biologically into good or better and moderate or poor status. Data are synthesised to illustrate a poor separation of total phosphorus concentration. vertical lines mark a range of potential good/moderate boundary values; 10th quantile of total phosphorus in sites classified as moderate (red line) and 90th quantile for sites classified as good (blue line).

2.4 The confusion matrix

To understand our approach it is important to clarify the following 2 points.

1) **We are predicting Biology from the Supporting Elements**, it means that we create expectations if a supporting element is within, or is not within, a good range of values. Based on those supporting element values we predict biological status and compare them with observed biological status.

2) Secondly, when predictions and observations match we have TRUE predictions and where there is a mismatch we have FALSE predictions. In our case we split the TRUE predictions into True Positive and True Negative when the predicted and observed match respectively for good and not good status. Thus linking it to a condition and not with the presence/absence of a supporting element as usually applied. Therefore we define:

    - **TRUE Positive** status (**both biota and SE are good**, a positive acceptable condition

    - **TRUE Negative** status (**both biota and SE are not good**, a negative non-acceptable condition)

    So, if we think of the error in terms of the predictions, then:

- **FALSE Positive predictions** – would be the **SE falsely predicting biota as good** (positive) status

- **FALSE Negative predictions** – would be the **SE falsely predicting biota as not good** (negative) status.

Thus, for any data-set we can visualise the biological and supporting element classifications in a simple 2 x 2 matrix (Table 2.1). The elements of the matrix show the number of the same or correct classifications (*true positive* and *true negative* results) and different or mis-classifications (*False positive* and *False negative* results). There are different ways of arranging such a matrix, but in the context of SE boundaries we are used to a visual representation of biota status (e.g. EQR) as a vertical y axis, with the SE as a horizontal x axis. Thus, for our confusion matrix we show the observed (true) biota status on the vertical side and the SE status on the horizontal. This arrangement also simplifies comparing pressure response plots to the confusion matrix (Figure 2.4).

Table 2.1. *Confusion Matrix (or Classification Error Matrix) comparing biological and nutrient status classes. +ve = positive, -ve = negative.*

		Predicted	Predicted
		Good(1) nutrient	Not Good(0) nutrient
Observed	Good(1) biota	True +ve (1,1)	False -ve (1,0)
Observed	Not Good(0) biota	False +ve (0,1)	True -ve (0,0)

A number of different measures that can be calculated from the confusion matrix are commonly used (Fielding and Bell, 1997) to assess the success of models predicting classifications. Those most useful for our purposes are shown in Table 2.2 These measures can be split into three groups, the first two measures *miss-classification rate* (*Misclass*) and *kappa* (*kp*) are measures of overall classification success, the following two *FPR* and *FNR* are overall proportions of false positive (Fp) and negative (Fn) classification for nutrients. These were the measures used in the toolkit for the misclassification method, which determined a boundary where these measures were equal. The other two measures are conditional proportions of false positive (Fp) and negative (Fn) classifications *Commission* and *Omission*, both expressed as the proportions of the “true” (observed biological) classes. In the medical literature *Sensitivity* the proportion of correctly predicted positive values (biota good, nutrients good) and the inverse *Specificity* the proportion of correctly predicted negative values (biota not good, nutrients not good). However, in the context of boundary setting we are more interested in the incorrect predictions, results where the biology would be incorrectly classified, we propose using *Commission* the proportion of false positives (biota not good, nutrients good) and *Omission* the proportion of false negatives (biota good, nutrients not good). The final two measures, the Over and Under Prediction Rates (*OPR* and *UPR*) are similar, but the proportions are expressed relative to the SE classes, rather than the true biota class and are thus less useful.

Table 2.2. Measures to assess confusion matrix. N= total number of classifications.

Measure	Abrev	Calculation	Explanation	Criteria
Correct classification rate	*CCR*	*(Tp+Tn)/N*	proportion of correct classifications	maximum value
Kappa	*kp*	*[(Tp+Tn)-(((Tp+Fn)(Tp+Fp)+(Fp+Tn)(Fn+Tn))/N)] / [N-(((Tp+Fn)(Tp+Fp)+(Fp+Tn)(Fn+Tn))/N)]*	index of classification power expressed as a ratio of observed agreement (Po) and expected agreement (Pe) Kappa = (Po-Pe)/(1-Pe)	maximum value
False positive rate (non conditional)	*FPR*	*Fp/N*	predicted proportion of false positives	Optimum balance *FNR=FPR* (toolkit mis-match method)
False negative rate (non conditional)	*FNR*	*Fn/N*	predicted proportion of false negatives	Optimum balance *FNR=FPR*
False positive rate (conditional on observed biota)	*Commission* (*1-Specificity*) true -ve	*Fp/(Fp+Tn)*	A measure of "badness of fit" predicted false positives as proportion of all true (biota) negatives	Optimum balance, *Commission=Omission* or consider minimise either to value of 0.1 for skewed data
False negative rate (conditional on observed biota)	*Omission* (*1-Sensitivity*) true +ve	*Fn/(Tp+Fn)*	A measure of "badness of fit" predicted false negatives as proportion of all true (biota) positives	Optimum balance, *Commission=Omission* or consider minimise either to value of 0.1 for skewed data
Over prediction rate (conditional on predicted nutrient boundary)	*OPR*	*Fp/(Tp+Fp)*	predicted false positives as proportion of all predicted positives.	Optimum balance, *OPR=UPR*
Under prediction rate (conditional on predicted nutrient boundary)	*UPR*	*Fn/(Tn+Fn)*	predicted false negatives as proportion of all predicted negatives.	Optimum balance, *OPR=UPR*

2.4.1 An example of a confusion matrix

Using the example data from Figure 2.2 and selecting an arbitrary boundary (in this case where the 75th and 25th quantiles are equal) to categorise the SE total phosphorus, the resulting confusion matrix is shown in Figure 2.3. In this example the data are evenly balanced with similar number of sites classified by biota as good (79+27=106) and not good (23+71= 94), a prevalence of 0.53 (the proportion of good biota in the data set). As a result, the proportions of false +ve (*Commission*) and false -ve (*Omission*) are similar with a relatively high proportion of sites correctly classified (*CCR* = 0.75).

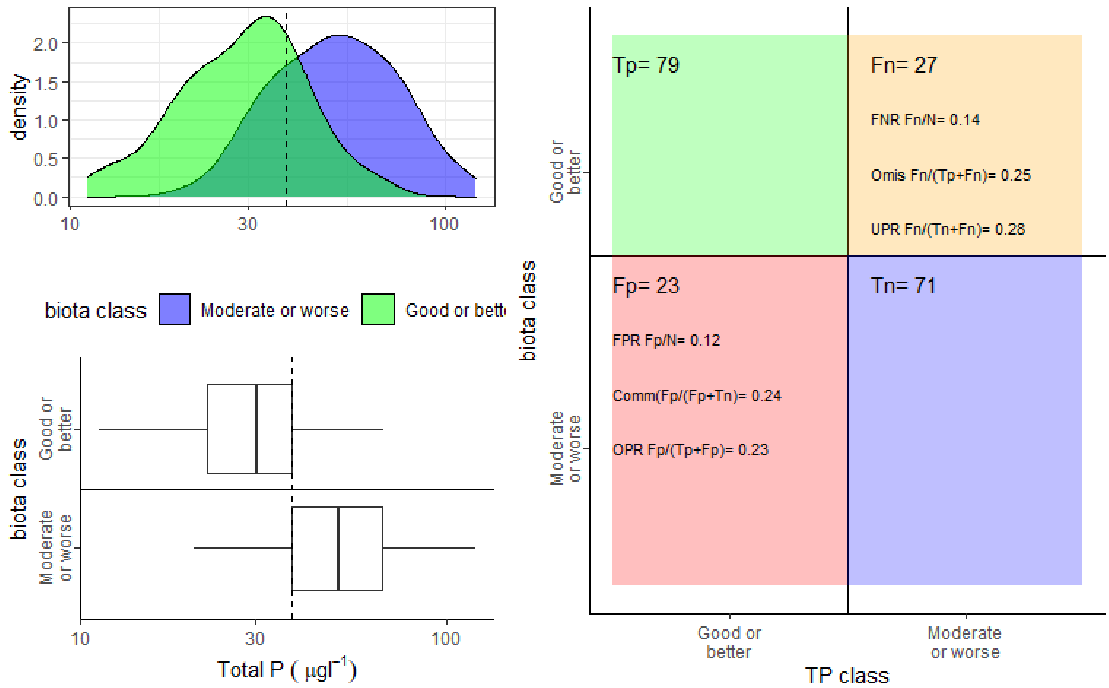

Figure 2.3. Density distribution and box plots showing the range of total phosphorus concentration in sites classified biologically into good or better and moderate or poor status using an arbitrary TP boundary (vertical line). Confusion matrix showing a comparing the classifications. Data are synthesised as in Figure 2.2.

2.4.2 Example of confusion matrix with less ideal data

An ideal data set will have relatively few mis-classifications and these will be evenly distributed between false negative and false positive results, although for many data sets this may not be the case. In the best practice guide we point out that a scatter plot showing the relationship between a continuous measure of biological status (EQR) and nutrient concentration may resemble a *wedge*, where for example other additional pressures reduce EQR values. Alternatively, an *inverted wedge* may occur where additional factors, such as grazing or light limitation, counteract the effect of nutrients. In these situations fitting linear models is not always appropriate as the variance of model residuals is unlikely to be constant across the pressure gradient (the data exhibit *heteroscedasticity*) and the mis-classifications in the classification matrix that will be generated when the predicted nutrient boundary values is used for classification can be different, with a greater proportion of false negatives (inverted wedge) or false positives (wedge).

An example of such a data set is shown in Figure 2.4. The data set used was synthetic, with a known relationship between EQR and TP plus a random error term, following the method in (Phillips et al. 2019). The data set included an additional factor simulating a counteracting influence of an additional unknown variable thus generating a typical inverted wedge shaped data cloud. The range of TP values in each of the biota classes overlap substantially, but based on the known relationship the true good moderate boundary of the data set was 40 ugL^{-1}. A conventional linear model predicts a boundary value of 70 ugL-1 (Figure 2.4a). With this boundary, the confusion matrix (Figure 2.4b) shows a very high Commission (false +ve) rate of 0.7, contrasting with a very low Omission rate of 0.03 (false -ve). Thus with this boundary there is a substantial risk that sites predicted to be at good biological status based on the TP class will actually be not in good status (false +ve). Thus, the boundary is likely to be insufficiently precautionary and a lower value is probably appropriate.

2.5 The binary logistic model

The confusion matrix is not dependent on the method of establishing the boundary but one modelling approach, the binary logistic model (BLM), is well suited to the technique. While a GLM model generates a single boundary value using the intersection of the fitted model to the good/moderate boundary values (EQR = 0.6), the BLM predicts the probability of being in good status over the range of nutrient concentration values in the data set. The boundary concentration is determined from a particular *threshold* probability. In the best practice guide we suggested using the probability of 0.5, as this is the concentration where there is an equal chance of being in good or not good status, however other probabilities can be selected. For example if we wish to be more certain that good status can be achieved we could select a higher probability value. This requires subjective decisions, but we can use the classification (confusion) matrix and the measures used to quantify mis-classifications as a guide. This is greatly simplified using the R package *modEva* which calculates a wide range of measures for all potential probability thresholds for a particular binary logistic model (Table 2.2).

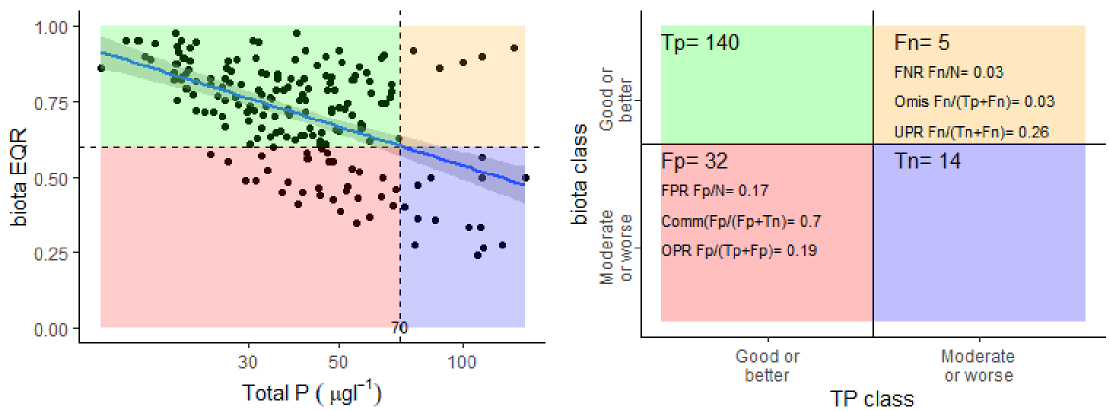

Figure 2.4. a) Scatter plot showing relationship between EQR and total phosphorus; b) confusion matrix comparing biotic and TP classification resulting from default p threshold. Synthetic data with an inverted wedge distribution simulating the effect of an unknown factor mitigating the impact of phosphorus.

An example of fitting a BLM to the same data set as that used in Figure 2.4 (the inverted wedge-shaped distribution) is shown in Figure 2.5a. Using the default probability threshold of 0.5, as suggested in the last guidance, produces a similar boundary value (71ugL^{-1} boundary value as the linear model and thus the same confusion matrix. The box plots (Figure 2.5c, d) illustrate this very clearly, with the boundary value almost as high as the 75th quantile of TP in the mode rate or worse sites. These results clearly show how fitting a linear model or always using a threshold probability of 0.5 with a BLM is not necessarily appropriate.

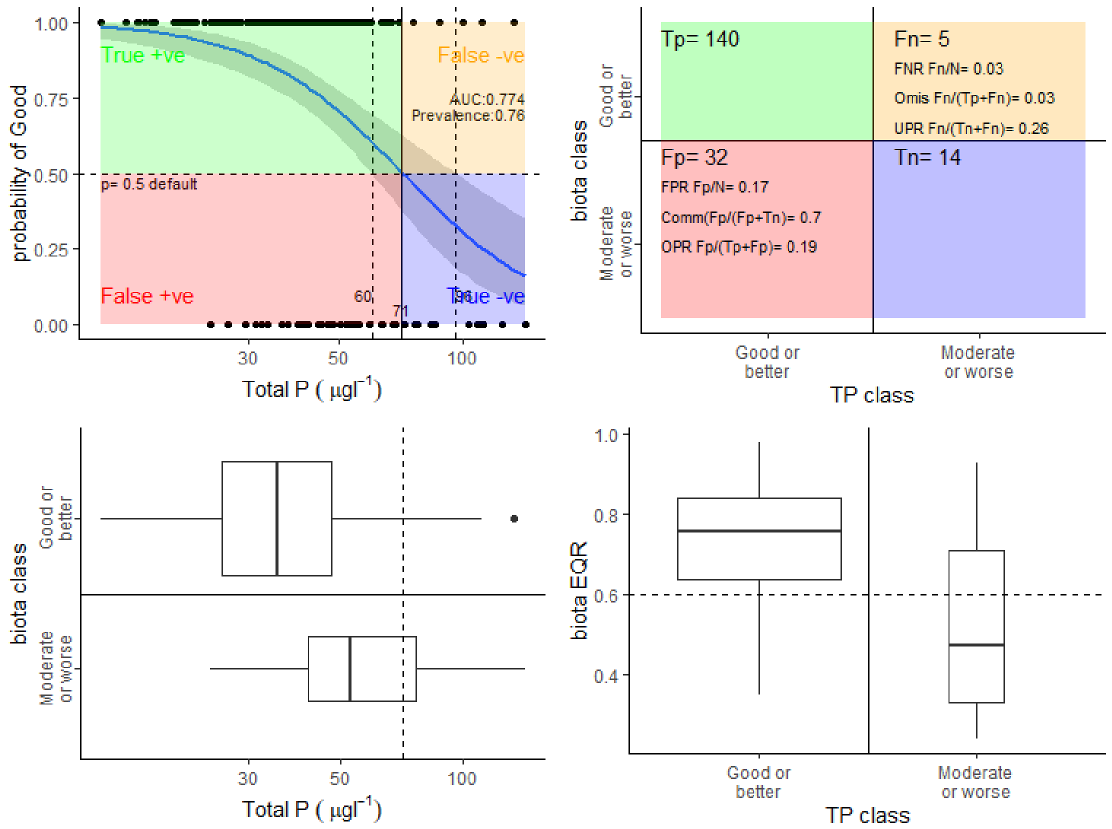

Figure 2.5. a) logistic model fitted to the data shown in Figure 2.4, with default (p=0.5) probability threshold marked; b) confusion matrix comparing biotic and TP classification resulting from default p threshold; c) box plots comparing range of TP in the biota classes; d) box plots comparing the range of EQR in the TP classes.

Using the *modEva* package we can inspect the effect of using different threshold probability values to determine the TP boundary for these data (Figure 2.6). The correct classification rate (CCR) is the least useful metric as at low and high probabilities its value will reflect the proportion of sites in the most or least common class (Prevalence). The kappa statistic is a more useful indicator of accuracy with a maximum value of 0.35 at a probability of 0.54 (point b Figure 2.6, Table 2.3). At this point the commission (false +ve) rate is still 0.652. Given the purpose of boundary setting (supporting good status for the BQE), we suggest that reducing commission is more important than reducing omission. The omission rate (false -ve) increases steadily from p=0.6, equaling commission at a probability of 0.76 (point e Figure 2.6). At this point commission is substantially reduced to 0.283. A further increase of p would reduce this further, but that will also increase omission.

Selecting the most appropriate threshold probability is thus a compromise between maximising overall classification accuracy, reducing commission without excessive omission. It seems reasonable to suggest that the most appropriate probability threshold lies between that given by the maximum kappa and a maximum of 0.9 or the probability where the omission rate is not greater than twice the commission. Additionally the value of kappa should ideally be ≥ 0.21, the threshold for *fair* agreement (Landis and Koch 1977). By listing the different measures by ascending probability (Table 2.3) it is possible to identify this range p 0.54-0.8, boundary 67-40 ugL^{-1}. Given that these data had clear evidence of an inverted wedge, then a precautionary boundary would be obtained using p=0.8, a boundary of 40 ugL^{-1}, which was the true boundary for this synthetic data set (Figure 2.7). It should also be noted that where data distributions are wedge shaped the distribution of EQR values by the SE class when the SE class boundary is set as described above (Figure 2.7d) will not reflect the normalised EQR boundary value of 0.6. This highlights that the mis-classification rates conditional on the predicted (i.e. SE) class (OPR and UPR) are not useful, note the very high predicted boundary (107 ugL^{-1}) for the probability threshold where OPR = UPR (Figure 2.8).

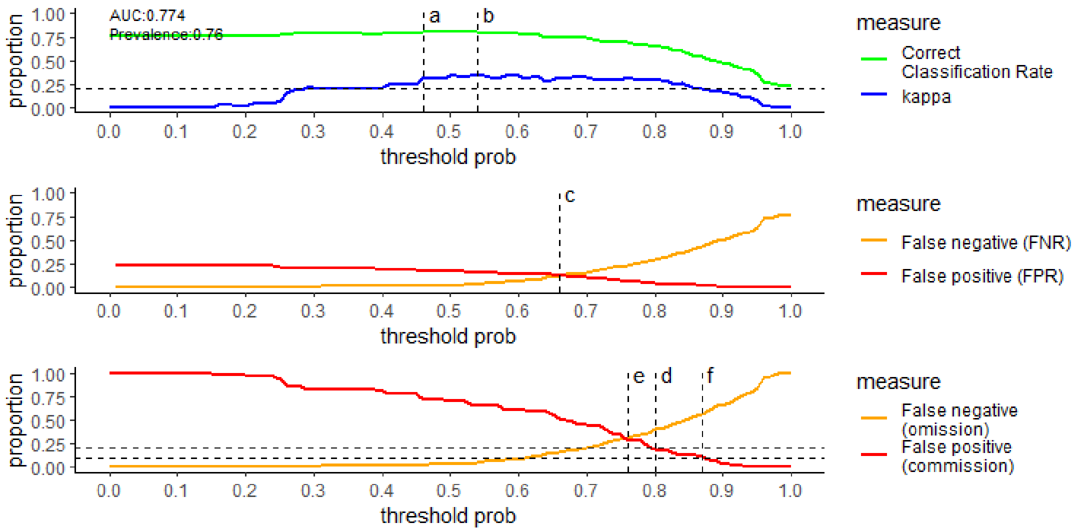

Figure 2.6. Change in measures used for assessing confusion matrix, with five possible cut-points marked; a) max CCR, b) max kappa, c) the intersection of FNR/FPR, d) commission = 0.2, e) cross-over commission/omission, f) commission = 0.1. Measures created using binary logistic regression fitted to data used for Figure 2.4.

Table 2.3. Predicted boundary values, together with key measures from confusion matrix, from binary logistic models fitted to synthetic data used in Figure 2.4 and 2.5.

measure	criteria	p threshold	GM boundary	LCL	UCL	commission	omission	kappa
CCR	maximum	0.46	76	63	106	0.717	0.028	0.325
default	NA	0.50	71	60	96	0.696	0.034	0.337
kappa	maximum	0.54	67	57	87	0.652	0.055	0.350
TKitMM	Cross-over FPR/NPR	0.66	54	47	66	0.522	0.159	0.322
OmisComm	Cross-over Omis/Commis	0.76	44	37	52	0.283	0.317	0.321
Com0.2	equal 0.2	0.80	40	33	47	0.196	0.400	0.297
Com0.1	equal 0.1	0.87	33	24	39	0.109	0.566	0.207

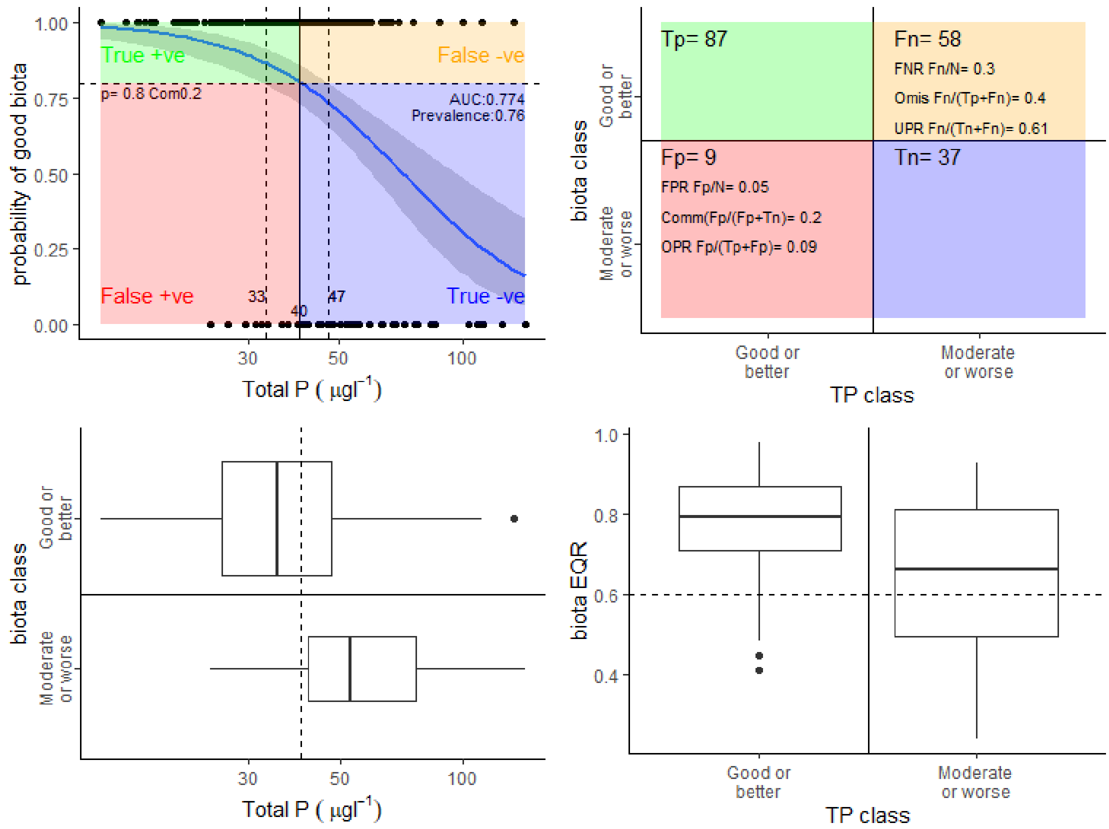

Figure 2.7. Results of fitting binary logistic model to data used for figure 2.4; a) scatter plot with model fit and predicted boundary concentrations for p threshold determined by commission = 0.2; b) confusion matrix showing number of true and false records and measures, c) boxplots showing range of TP for waterbodies classified by biota. d) boxplots showing range of EQR for waterbodies classified using the predicted TP boundary. Dotted lines show boundary values.

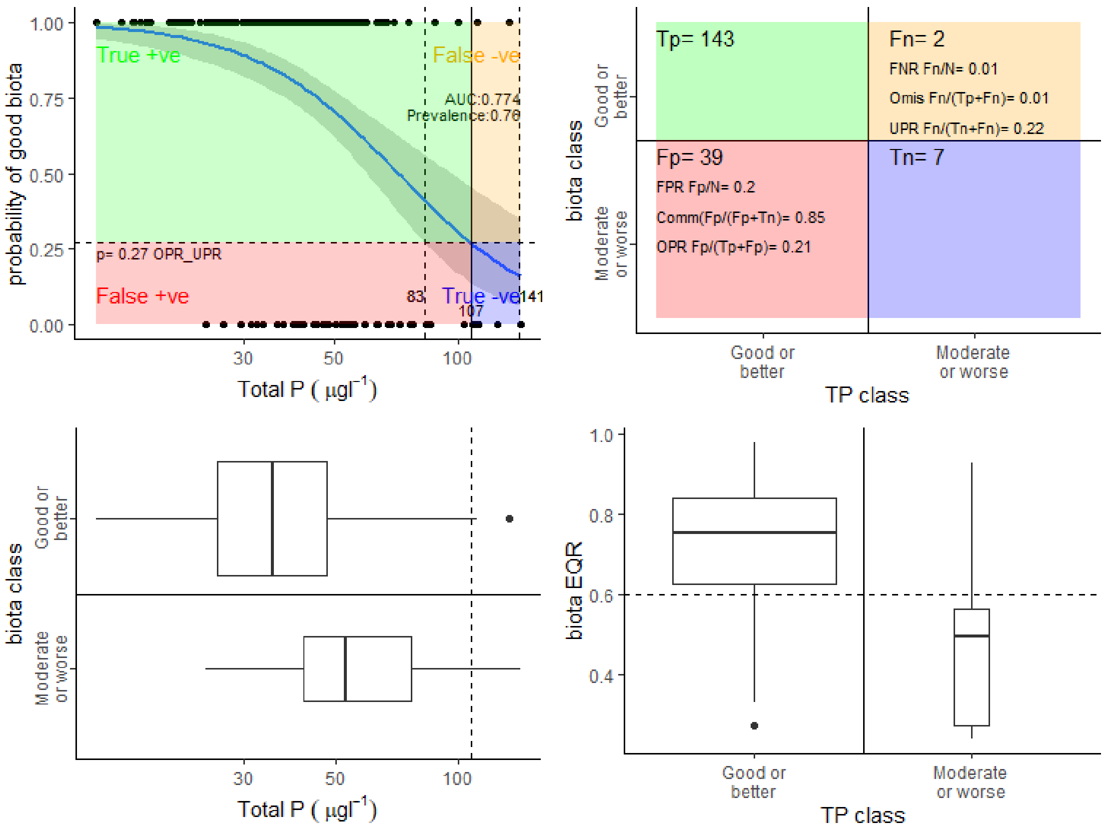

Figure 2.8. Results of fitting binary logistic model to data used for figure 2.4; a) scatter plot with model fit and predicted boundary concentrations for p threshold determined by OPR = UPR; b) confusion matrix showing number of true and false records and measures, c) boxplots showing range of TP for waterbodies classified by biota. d) boxplots showing range of EQR for waterbodies classified using the predicted TP boundary. Dotted lines show boundary values.

2.6 Assessing model fit

Another important consideration is the overall fit of the model. We have already illustrated the difference between data sets with high and low overlap of SE distributions for the two biota classes (Figure 2.1 & Figure 2.2). Where a continuous modelling approach is used current guidance proposes using r^2 to assess model performance. This metric is not available for a BLM, although a pseudo r^2 can be calculated. However, given the focus on comparative classifications we can also use a *receiver operating characteristic* (ROC) plot (Figure 2.9a). This relates the sensitivity (true +ve) to commission (false +ve) rates for all p thresholds, a straight line represents no relationship and the area under the curve (AUC) provides an overall index of fit varying from 0.5 - 1.0 (complete to no overlap of distributions). The test data set had an AUC of 0.774 indicating that for 77% of the time a random selection from the good biota group would be more likely to be predicted as good than a random selection from the moderate biota group. Generally, an AUC or 0.7-0.8 is adequate, 0.8-0.9 good, while >0.9 is excellent. The pseudo r^2 and AUC are related (Figure 2.9b), although clearly when prevalence is high (>0.8) the AUC may be higher than expected. Kappa is a threshold dependent measure of classification success and the maximum value of kappa is also related to AUC (Figure 2.9c) with a kappa of 0.32 and an AUC of 0.7. As a general guide an adequate model would be expected to have an $AUC \geq 0.70$, a pseudo r^2 of ≥ 0.15 and a maximum kappa of ≥ 0.32.

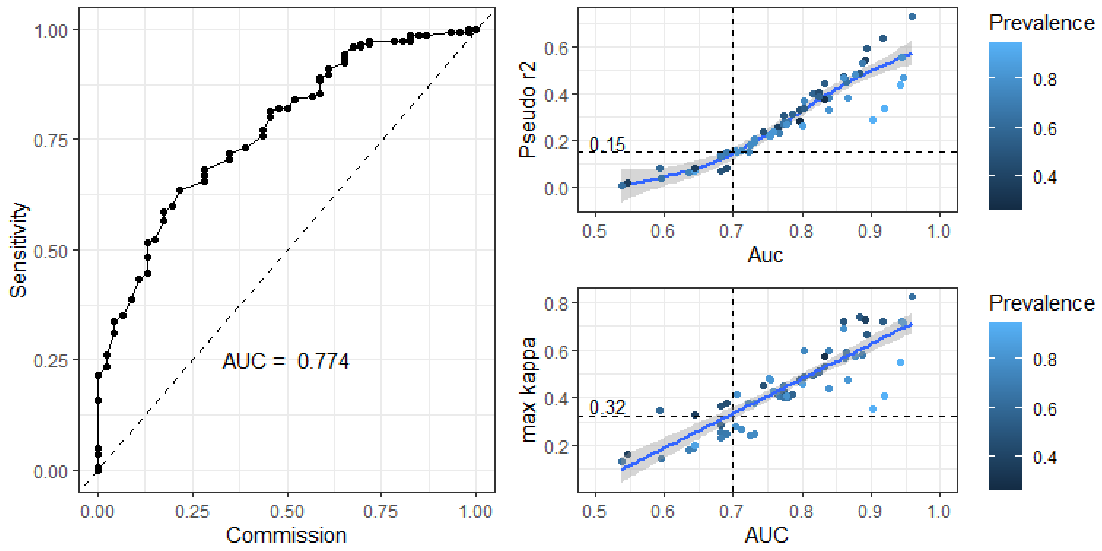

Figure 2.9. a) ROC curve for data used for figure 2.4; b) relationship between AUC and pseudo r2; c) relationship between AUC and maximum kappa. Data for b and c were from different subsets of the river data described in the associated document on river boundaries.

Although not unexpected it should be noted that mis-classification rates will be dependent on AUC, with the value of commission for maximum kappa and Omission = Commission declining as AUC increases (Figure 2.10c). Data prevalence can also be an important factor, with higher threshold probabilities as prevalence increases (Figure 2.10b). Increasing prevalence also increases the value of commission, while having no influence on the intersection of omission and commission (Figure 2.10d), thus typically omission will be lower when the critical threshold is based on omission = commission than at maximum kappa when prevalence is high, similar when it is close to 0.5 and lower when prevalence is low.

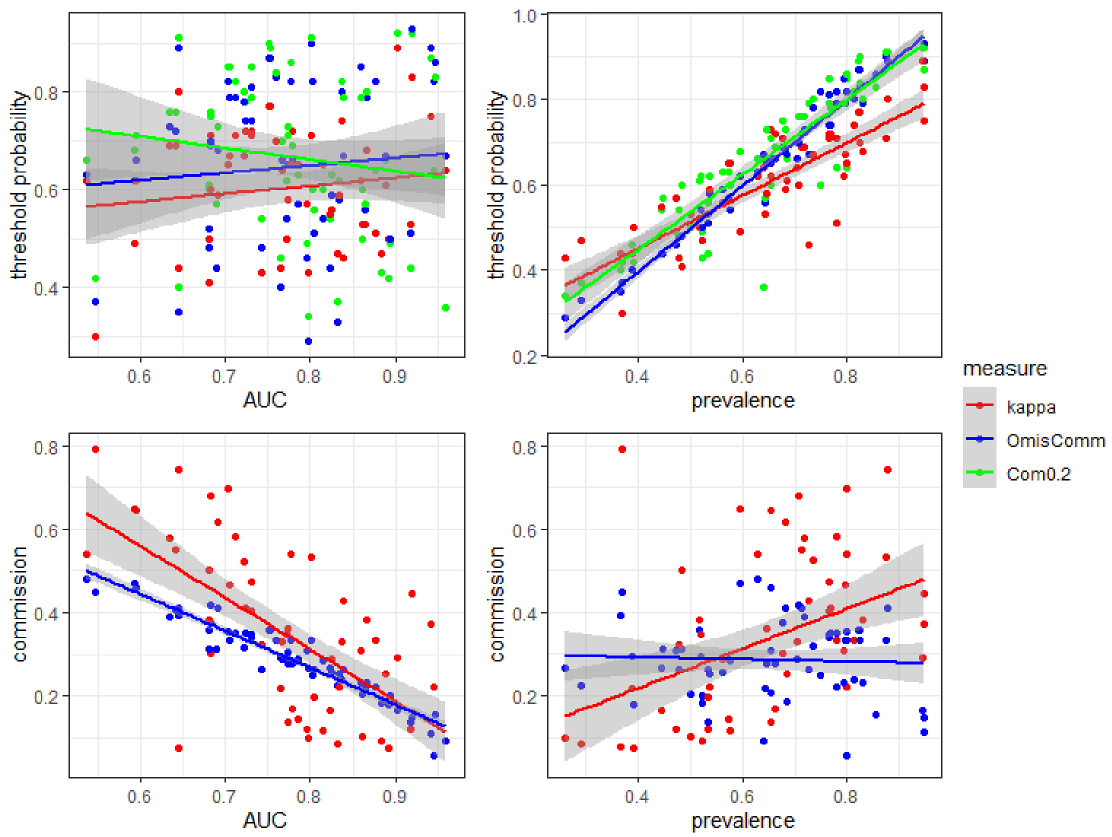

Figure 2.10. Relationship between a) threshold probability and AUC; b) threshold probability and data prevalence; c) value of commission and AUC; d) value of commission and prevalence. Each split by different measures, maximum value of kappa, omission = commission and commission = -0.2.

2.7 Summary

To determine supporting element boundaries we recommend

1. Regardless of the method used to determine a boundary value construct a confusion matrix comparing the resulting binary classifications and calculate at least the following metrics (*kappa*, *omission* and *commission* rates) to enable an assessment of mis-classification rates.
2. Use a binary logistic model to determine a range of potential boundary values using different probability cut levels.
3. Check model performance using r^2 for continuous models, pseudo r^2 (≥ 0.15 ,based on river TP models) and AUC (≥ 0.70).
4. Select the most appropriate probability cut level, after considering how important it is to minimise commission (false +ve) or omission (false -ve) rates.
5. The range of appropriate threshold probability values is likely to be between that provided by the maximum value of kappa and the point where omission = commission.
6. Except where there is clear evidence of secondary pressures reducing the status of the biota, aim to minimise commission rates.
7. In general select the lowest commission value where omission has not increased to more than double the commission value, the threshold probability ≤ 0.9 and kappa is ≥ 0.21.

3 Lake boundaries

This section is a revised version of a report written in 2021. It has been updated to reflect changes to the approach used to select the most appropriate probability thresholds used for determining the predicted boundary values for binary logistic models (BLMs) as described in Section 2 of this report.

3.1 Approach to modelling

Two types of model are fitted to the data.

1. Where comparable EQR values (e.g Normalised EQRs) are available a continuous regression modelling approach is used. The best practice guide highlights the need to select linear parts of the pressure response gradient when fitting models. This typically requires filtering the data using an upper nutrient threshold above which EQR values are less responsive. However where EQR values are truncated, as occurs during normalisation (to make national EQRs comparable), curvature also occurs as EQR values approach 1.0. To overcome this I have fitted a Generalised linear model (GLM) with a logit link function to the data which better allows for this curvature. This model is not currently available in the tool-kit, but can be added in the next revision.

2. Where normalised EQR values are not available, a binary logistic model is fitted to the data using an almost identical approach to that in the current toolkit. The only differences is that for this work we define the binary categories differently, as we are predicting good or better status, rather than moderate or worse. Thus good or better status = 1, while moderate or worse = 0.

For both modelling approaches we use the predicted nutrient boundary value to classify the data and create a classification matrix. We then compare the mis-classification rates using a range of measures widely used for this purpose (Fielding and Bell 1997).

The GLM models generate a single boundary value, together with uncertainty estimates (confidence limit), based on the intersection of the fitted model to the good/moderate boundary values (EQR = 0.6). However, the binary logistic model predicts the probability of being in good status for the range of nutrient concentration values in the data set. The boundary concentration is determined from a particular *threshold* probability. In the best practice guide we suggest using the probability of 0.5, as this is the concentration where there is an equal chance of being in good or not good status. However, other probabilities can be selected and where for example we wish to be more certain that good status can be achieved we could select a higher probability value. This requires a subjective decisions, but we can use the classification (confusion) matrix and the measures used to quantify mis-classifications as a guide. We use the r package *modEva* to achieve this and full details of the approach are provided in section 2 of this report.

3.2 Data

For lakes, models were fitted to three data sets.

3. **SoE data-set**. Data collated by the EEA for state of the environment reporting. Data at water body/Year level for phytoplankton status (NEQR) were matched to mean annual nutrient concentrations. These data were screened to remove records where the calculated NEQR values did not match status originally reported and excluded EQR values where the method was specific to acidification pressure. For this data-set both GLM and binary logistic models were fitted (using binary categories calculated from the NEQR).

4. **WFD data-set**. Data obtained from the WFD classification database and linked by water body to the EEA SoE nutrient concentrations. As this data-set only contained categorical (WFD class) data only binary logistic regression models were applied.

5. **IC data-set**. Data collated for the phytoplankton inter calibration process and subsequently used to determine boundary values using the IC typology (Poikane et al. 2015). Although these data included both national and common metric EQR values these are not normalised and thus cannot easily be used for lakes grouped using the broad typology (Lyche-Solheim *et al.*, 2019). Thus as for the WFD data set only binary logistic regression models are used.

The IC data set was the largest and while it contained fewer lake types it was originally compiled regionally and had sufficient TP and TN records (>10 in each binary class & >50 total records) for most lake types to be modeled (Figures 3.1 & 3.2) and for some types to model different regions. The WFD data set was the smallest but together with the SoE data set provided useful additional data which might reflect a wider range of lake conditions within each of the broad types. Where possible lakes falling into one of the main broad types were used, but in some cases the aggregated lake types were needed to provide sufficient records (LA-03 lowland & mid-altitude calcareous and LA-04 lowland & mid-altitude siliceous).

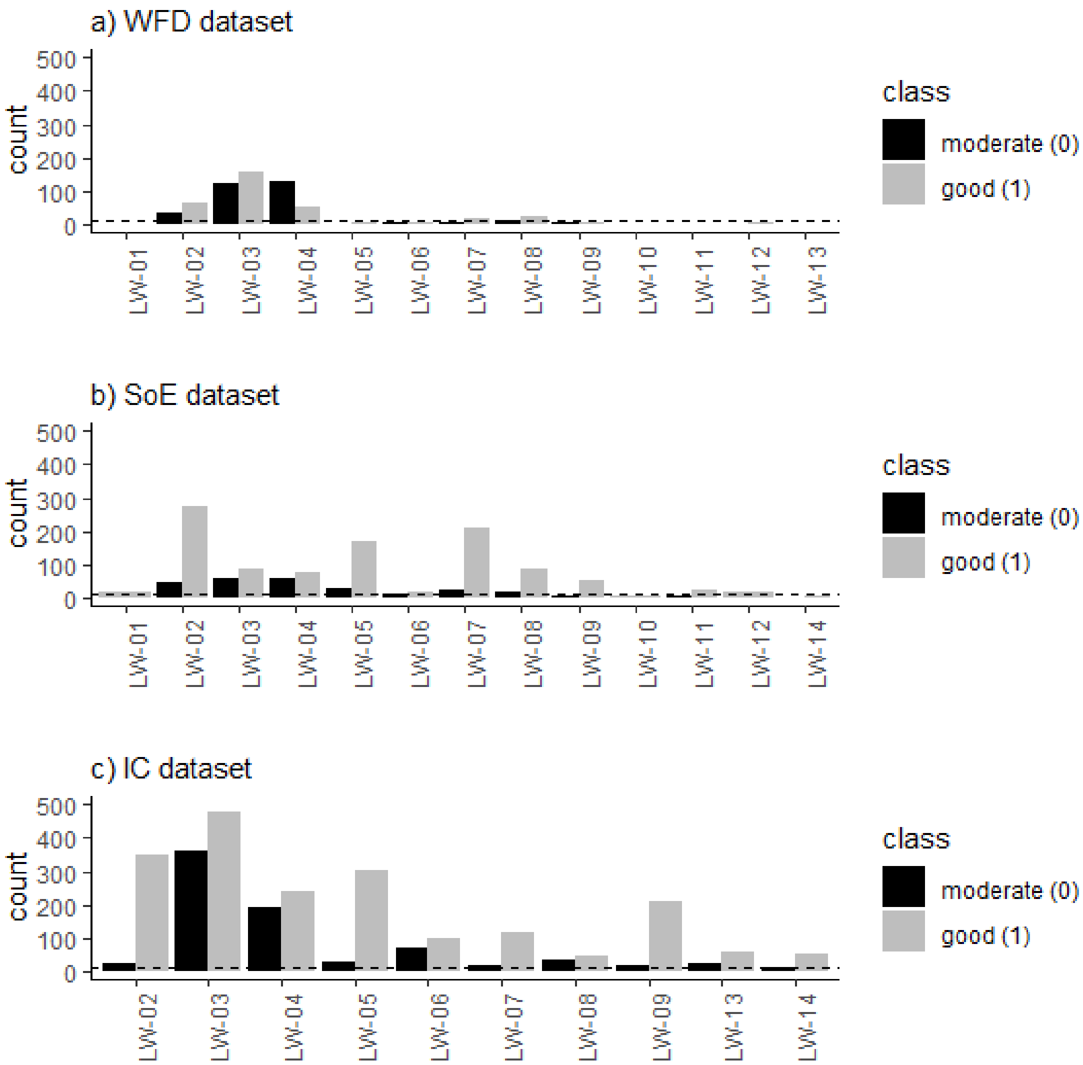

Figure 3.1. Comparison of number of total phosphorus records available in the three data sets by broad type

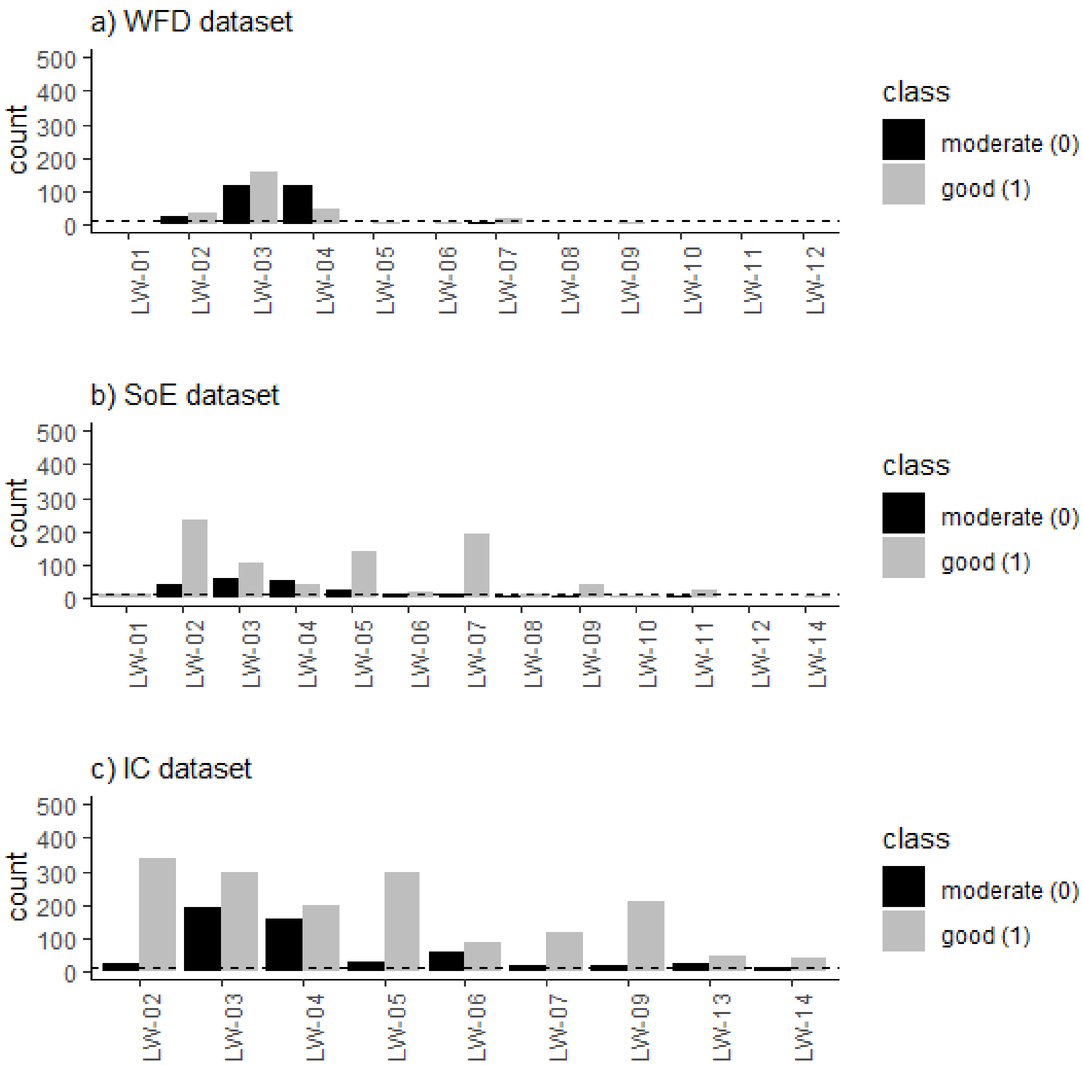

Figure 3.2. Comparison of number of total nitrogen records available in the three data sets by broad type.

3.2.1 Data prevalence

The balance of the data, in terms of the proportion of records where the observed phytoplankton class is in good or not good status (data prevalence) is important as it can influence model outcomes. For a well balanced data set (prevalence =0.5) most modelling approaches will generate similar predictions. However, when the data are less well balanced it is important to be aware that for BLMs the predicted boundary will tend to be drawn towards the most numerous class. The majority of the datasets are well balanced with prevalence values between 0.4 and 0.6, but some lake types have data with relatively high values >0.8 (LA-04, LW-02, LW-05, LW-07, LW-08, LW-14), the siliceous , mid-altitude and highland lakes. This imbalance in the data does not prevent modelling, but the resulting predicted boundaries might be lower than they would had a more balanced data set been available.

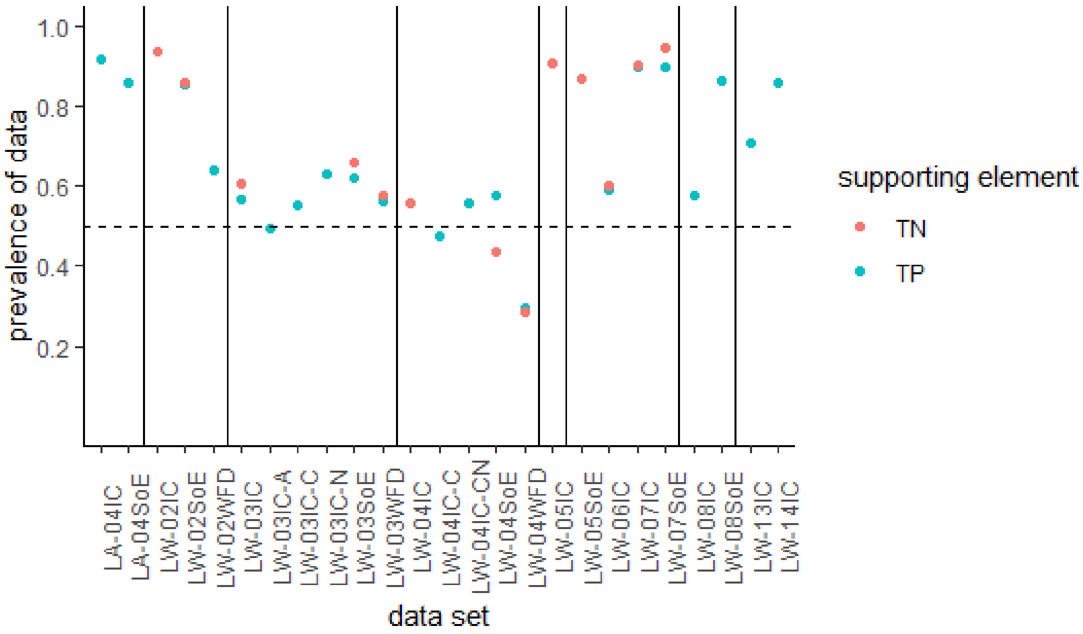

Figure 3.3. Prevalence (biota good:not good) of TP and TN data for each dataset by lake type

4 Lake total phosphorus boundaries, model results

Models were fitted to each data set by lake type. Full details of the models, the predicted boundary values and the resulting confusion matrices and the measures used to select appropriate cut levels are available in an appendix, as this report only presents an overview of the results.

4.1 Model performance

For linear modelling the strength of the relationship between biology and supporting element is assessed using the r^2 value, with models generating low values for r^2 being rejected. With logistic models it is only possible to calculate a pseudo r^2 and while this is indicative of the strength of the relationship it is less comparable between models which use different data.

However, by using this discrete modelling approach, which requires a probability threshold to be selected, a greater emphasis is placed on the relative classifications of biology and TP, via a consideration of the confusion matrix. This provides an opportunity to assess the overall strength of the relationship as an alternative assessment can be generated from the confusion matrix, where the relationship between the proportion of true positive and negative values are plotted against each other to form an ROC (Receiver Operating Characteristic) curve. The area under this curve (AUC) is a measure of the overall success of detecting, in this case, a difference between the two biological categories using 2 categories of TP. A high value (AUC > 0.8) suggests that TP is a good predictor, a lower value (AUC > 0.7) a fair predictor, while a value of AUC = 0.5 indicates that a random flip of a coin would be as good as using TP, thus no relationship between biological status and TP. The majority of the TP data sets provided AUC values >0.7, the median AUC for the SoE data was > 0.8, while 75% of the IC data sets had values >0.8 (Figure 4.4c).

The AUC statistic cannot be generated for a GLM, but to compare the relative performance of the GLM and BLM approaches we compared how successful the resulting TP classifications were in comparison to the *true* classification derived from phytoplankton using kappa, a measure of overall agreement. The GLM models clearly generated classifications with an overall higher level of agreement (i.e. higher value of kappa) than the BLM models (Figures 4.1a). However, the GLM models also produced a much higher commission (false +ve) classification rate, i.e. many sites that where phytoplankton were actually worse than good were incorrectly predicted to be in good status (Figures 4.1b). Thus with these data the GLM models typically generate boundary values that are unlikely to be sufficiently precautionary, while by selecting an appropriate probability threshold the BLM model predicted boundary concentrations that give a more balanced ratio of mis-classifications, albeit with a lower overall accuracy.

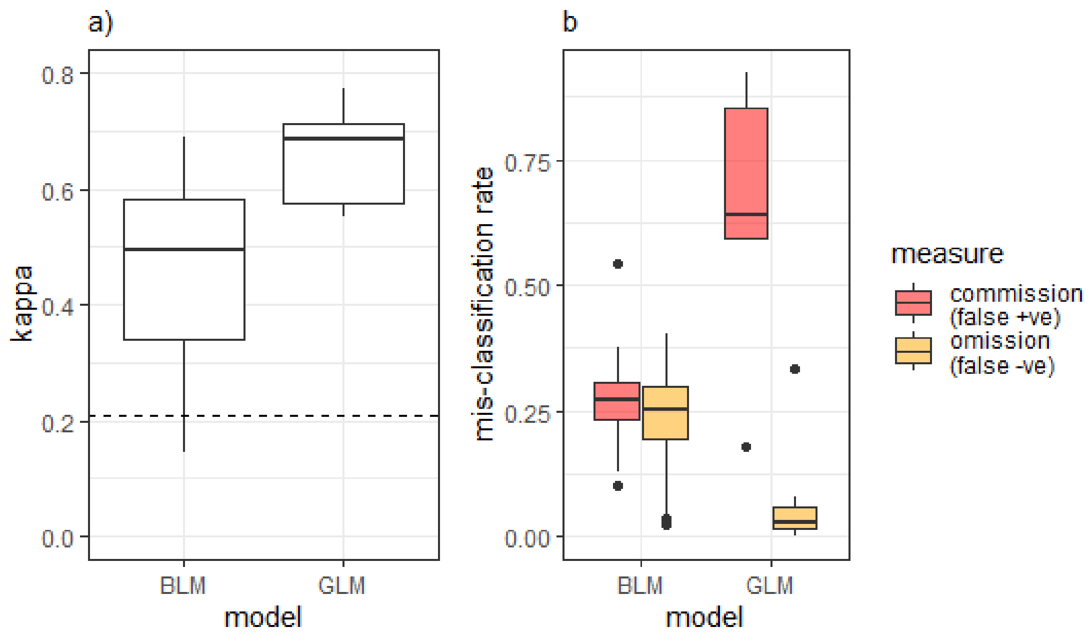

Figure 4.1 Comparison of mis-classification outcomes using different modelling (GLM and BLM) approaches a) value of kappa, b) commission (false+ve) and omission (false -ve) rates. (horizontal dotted line marks kappa = 0.21, a suggested level for acceptable accuracy, see 2.7).

4.2 Selecting appropriate threshold probability thresholds for BLM models

Table 4.2 details the lake TP values derived from the BLM models applied to data from different lake types, together with the mis-classification rates and model fit parameters. For each type four potential measures are shown (toolkit mismatch method where false +ve & -ve rates are equal; maximum kappa; omission & commission rates are equal, commission rate = 0.1.). The range of probability thresholds for each of these criteria, the resulting TP boundary and relative mis-classification rates are shown in Figure 4.2. In most cases selecting either the tool-kit mis-match method or the probability threshold that maximised correct classifications (maximum kappa) generated the lowest probability threshold and thus higher boundary value (Figure 4.2ab). However, these measures generated higher commission (false +ve) rates. Given the need to establish a relatively precautionary boundary a lower threshold was selected that minimised the commission (false +ve) rate, subject to that probability being ≤ 0.9, kappa ≥ 0.21 and omission < double commission rate. These values are shown in table 4.1 and in figure 4.3. They can be compared with boundary concentrations predicted using the other potential probability thresholds in table 4.2.

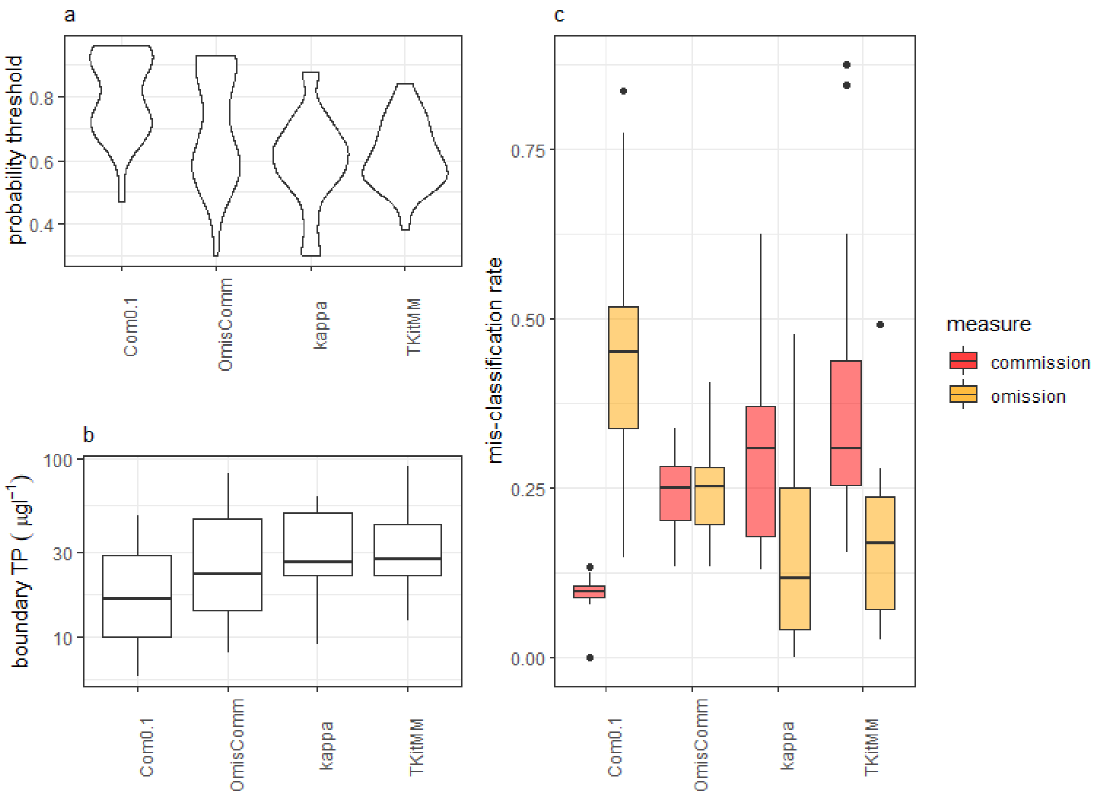

Figure 4.2. Effect of different thershold probability measures on a) the probability threshold, b) predicted TP boundary values, c) mis-classification rates (commission false +ve, omission false -ve).

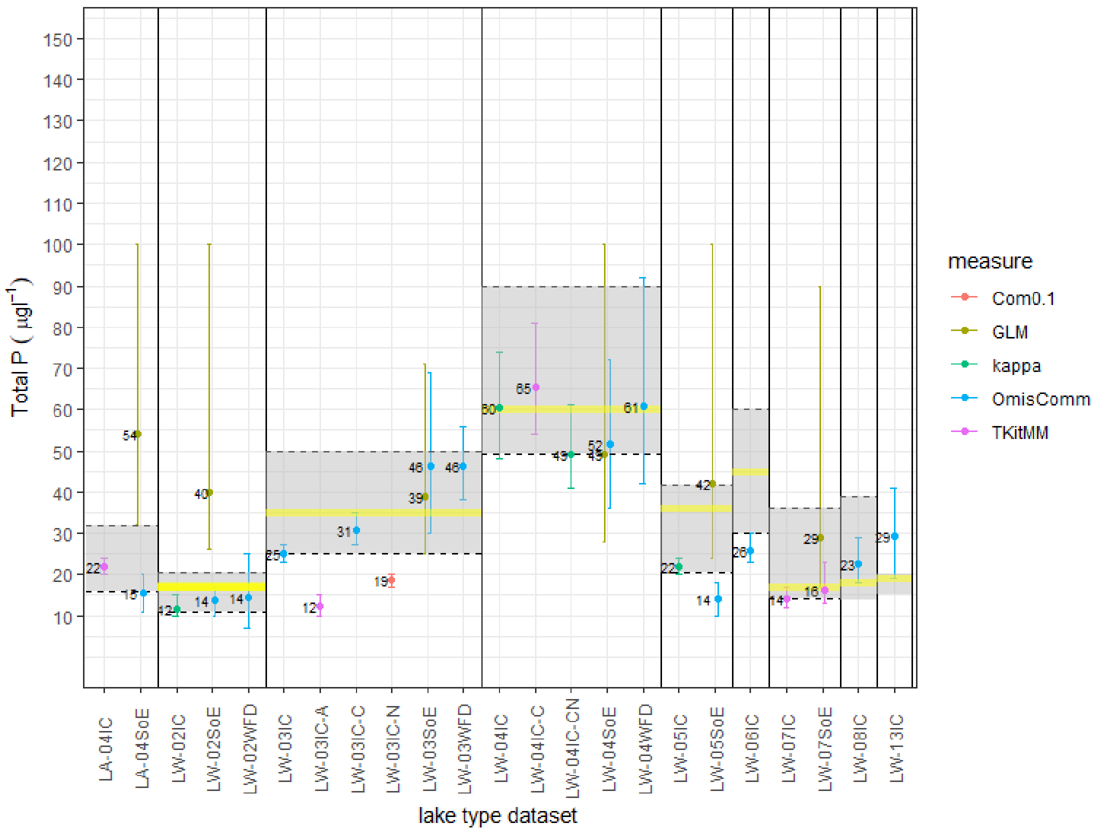

Figure 4.3. Predicted total phosphorus boundary values for lakes by broad type using models fitted to data from all available countries, compared to upper and lower quartiles of MS boundary values (grey shading) and estimated pan-European boundary values reported by (Nikolaidis et al. 2021) (yellow shading). (Vertical lines show 95% confidence limits).

4.3 Lake TP boundary values

The predicted good moderate boundary values varied from 11 - 65ugL^{-1}. There were clear differences between lake types, with the highest values occurring in the lowland calcareous very shallow lakes (LW-04), followed by the lowland calcareous shallow stratified lakes (LW-03), with the lowest found in the lowland siliceous (LW-02) and mid-altitude siliceous (LW07) lakes. The range of predicted boundary values was generally similar to the inter quartile range of the values reported by Member States (grey shaded area in Figure 4.3), suggesting that the modelling was broadly appropriate and reflected the collective Member States view of the TP concentrations that support good status.

However, as expected the different modelling approaches and data sets generated different type specific boundary values. Thus, to determine a typical boundary value we need to consider the factors that influence the specific model predictions, the modelling approach, and the representativeness and distributions of the data sets used.

4.3.1 Effect of model used

Predictions using the continuous GLM model were in most, but not all, cases higher than those made using the binary logistic models (Figures 4.3 with values often exceeding the interquartile range of MS boundary values. The scatter plots of EQR v TP (see appendix) suggest the reason was heteroscedasticity, with evidence of an inverted wedge pressure response curve, where unknown environmental factors were reducing the impact of elevated TP on phytoplankton and thus reduce the slope of the response. This illustrates the consequence of fitting continuous models without considering the distribution of residuals and consequently the imbalance of false positive/negative classifications.

The differences between the boundaries predicted using the measures used to determine probability cut thresholds for the discrete binary models was smaller (Figure 4.2b), but higher boundaries were predicted when the maximum of *kappa* was used in comparison to the equivalence of *omission* and *commission*. This is as expected, as using the former measure (*kappa*) seeks to minimise mis-classification, while the latter measure balances the false positive and negative classifications which, with an inverted wedge distribution of the pressure response relationship, reduces the proportion of false positives (TP predicts good but biology was actually not good).

4.3.2 Effect of prevalence

For the lake data in most, but not all cases, using the equivalence of *omission* and *commission* generated a higher probability threshold and a more precautionary boundary value. This was because the prevalence (ratio of Good:Not Good sites) was greater than 0.5. Where prevalence is <0.5 (i.e. the data is weighted towards not good sites) then using the measure of maximum kappa generates a higher threshold probability and lower boundary value, illustrating how with binary models minimising commission (false +ve) rates tends to pull the predicted boundary towards the dominant class in the dataset. Thus, where the prevalence is very high (>0.9) there is a risk of under predicting the boundary. For the lake TP data the following lake data sets exceeded this threshold (LA-04IC, LW-02IC, LW-05IC, LW-07IC). However, the selection rule that specified that the optimum probability threshold should be ≤ 0.9 ensured that for these data sets a lower probability threshold (e.g. maximum kappa) was selected (Figure 4.3 & Table 4.2) and thus excessively low boundary concentrations were avoided.

4.3.3 Importance of data set used

4.3.3.1 Representativeness

Model predictions are clearly dependent on the data sets used. The WFD data set was smaller and only had sufficient data to allow the three most common lowland lake types to be modeled (siliceous LW-02, calcareous stratified LW-03 and calareous very shallow L-04). However, this data set contained records from more countries which covered a wider range of conditions (prevalence values closer to 0.5, a value with an equal balance between good and not good categories) (Figure 4.4a). The IC data was a large data set and contained sufficient records for the majority of lake types to be modeled. It also generated the highest AUC values (Figure 4.4c), although it had records from fewer countries and may thus be less representative. Additionally both this and the SoE data sets were less balanced with higher prevalence values indicating dominance of records where phytoplankton was good or better (Figure 4.4b). Despite these differences there is no evidence that there was any clear bias in the different data sets (Figure 4.5).

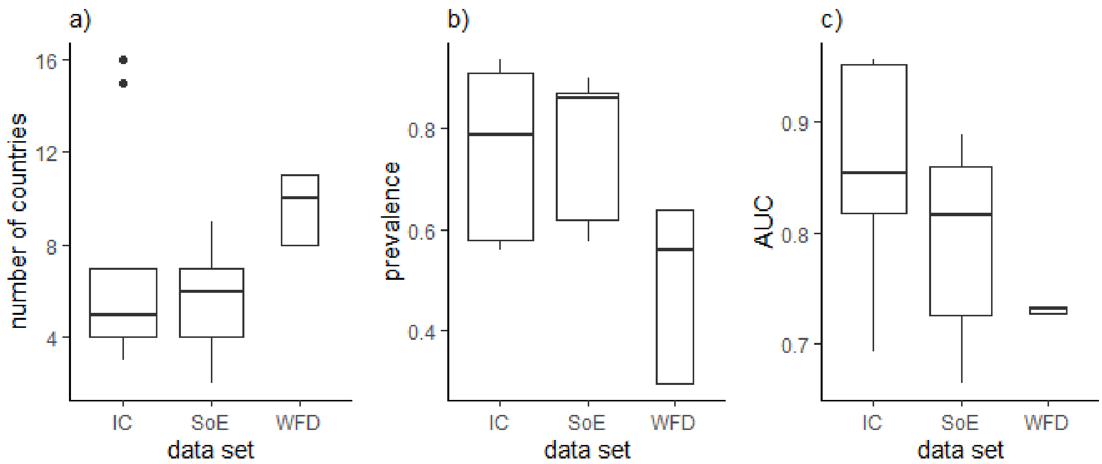

Figure 4.4. Range of a) number of countries in dataset; b) prevalence of the data; c) AUC of data.

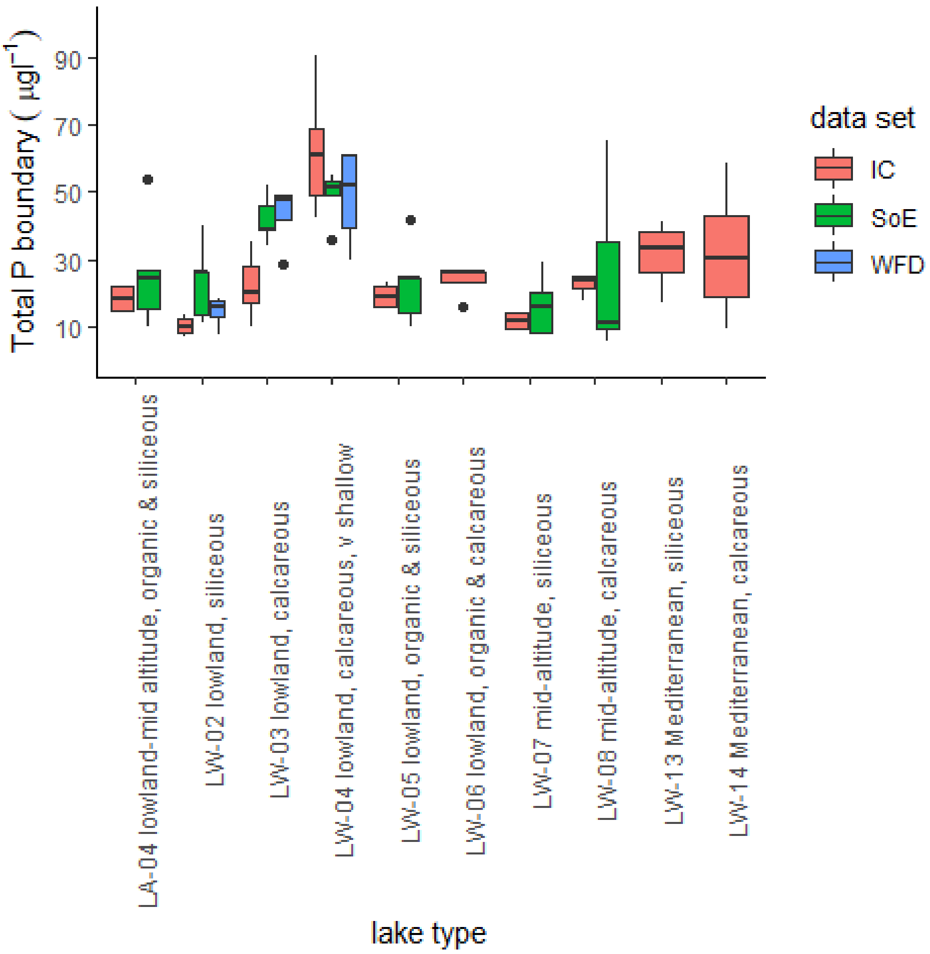

Figure 4.5. Range of predicted TP good moderate boundary values by lake type and data set.

4.3.3.2 *Regional variation*

There is considerable interest in whether there is regional variability in biological responses which needs to be considered in determining boundary values for the supporting elements. There were only sufficient data to model regions separately for the three common lowland European lake types (siliceous LW-02, calcareous stratified LW-03 & calcareous very shallow LW-04). There was a substantial overlap between the distributions of TP within each biological class in the siliceous lakes (LW-02) from the northern (N) and central (C) regions and consequently there was no significant difference between the predicted boundary values (Figure 4.6a).

For the calcareous stratified lakes (LW-03) there were clear differences in the TP distributions and predicted boundary values (Figure 4.6b). For the calcareous very shallow lakes (LW-04) the Eastern region had clearly elevated TP which overlapped in both biological classes, resulting in a non significant model. There were too few impacted lakes to model the Northern region separately, so it was not possible to test the effect of region on the northern and central region data and boundaries were thus modelled for these two regions combined (Figure 4.6c).

The only regions where a significant regional effect could be shown was for the shallow calcareous lakes. This is a common lake type and the definition of “calcareous” encompasses a wide range of lake alkalinities. For example, it contains the Northern IC type L-N1 lake type (moderate alkalinity) the Central L-CB1 calcareous (high alkalinity) and the lowland or mid-altitude moderate-high alkalinity Alpine lakes. Thus while there were regional differences between boundaries for lake type LW-04 it is suggested that it is the wide definition of the broad type, rather than a regional climatic influence that is responsible for these differences.

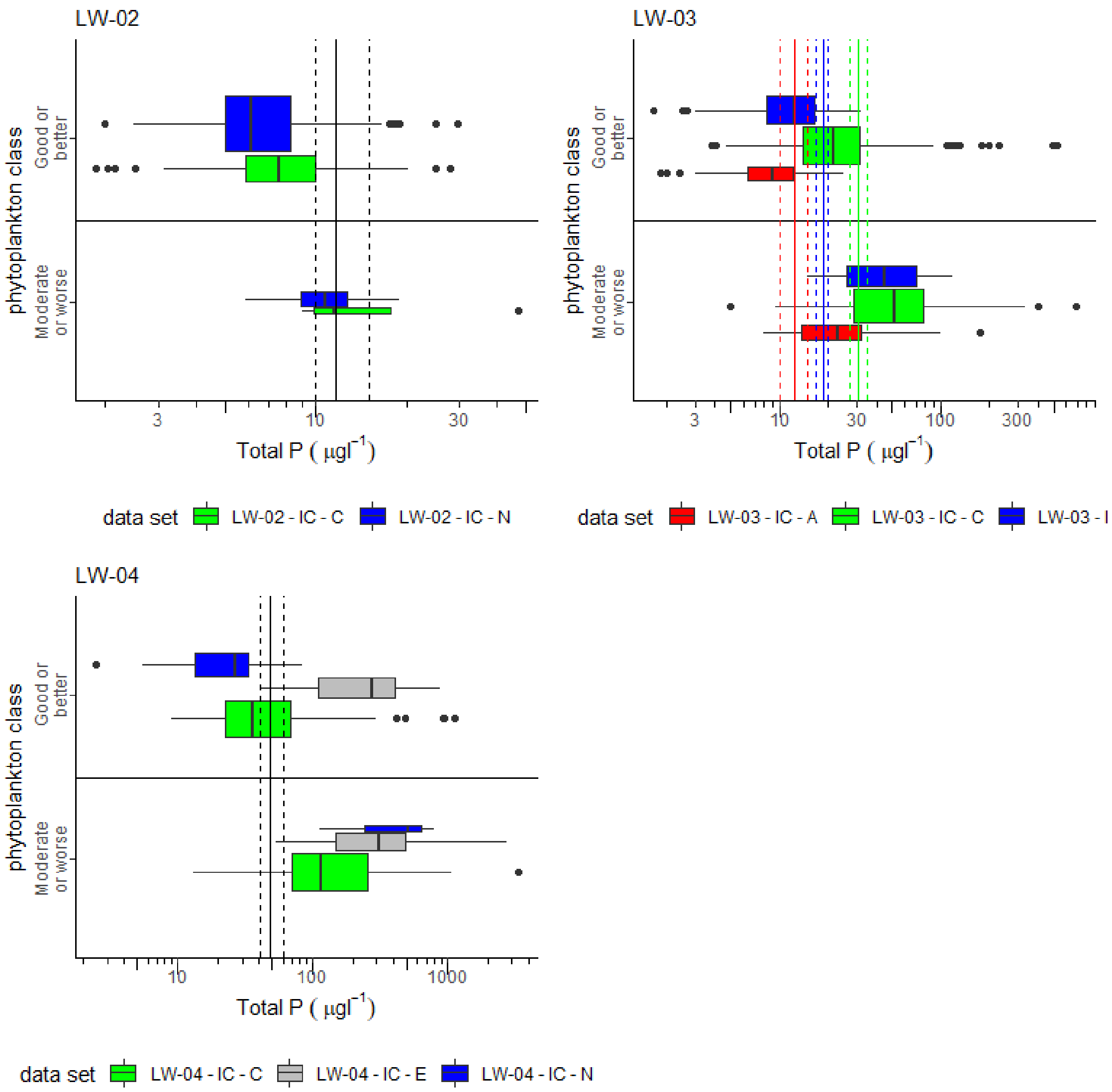

Figure 4.6. Regional distribution of TP by phytoplankton class overlain by vertical lines showing predicted boundary values (± 95% confidence limits) from binomial logistic models. a) LW-02 lowland siliceous, b)LW-03 lowland calcareous shallow, c) lowland calcareous very shallow.

4.4 Most likely TP boundary ranges for lakes

The resulting boundary values from each of the data sets are provided in Table 4.1. To simplify these and provide a most likely range of boundary values for European lakes is clearly challenging. There was insufficient data to develop region specific models and it is clear that the available data sets provided a range of variability and representativeness. The IC data produced the better models, but are biased towards higher status sites, while the WFD data are more representative but generate lower accuracy models, probably as the sites used are subject to a wider range of pressures and the data will contain some estimated classifications. The SoE data fall between these extremes and might provide the better data set to use for the purpose of generating *most likely* boundary values.

However, rather than making this assumption it is suggested that a better alternative is to use all of the predicted boundary values from the binary logistic models, presenting these as box plots together with the highest and lowest of the boundary 95% confidence intervals (Figure 4.7). The most likely boundary ranges are then indicated by either the interquartile ranges (boxes in Figure 4.7, or where these are very small, due to restricted data, the range provided by the 95% confidence limits (blue bars in Figure 4.7).

To place these values into context the range provided by the maximum and minimum values of the 95th percentiles are compared with the distribution of TP of the combined data sets and the most recently reported MS boundary values (Figure 4.8).

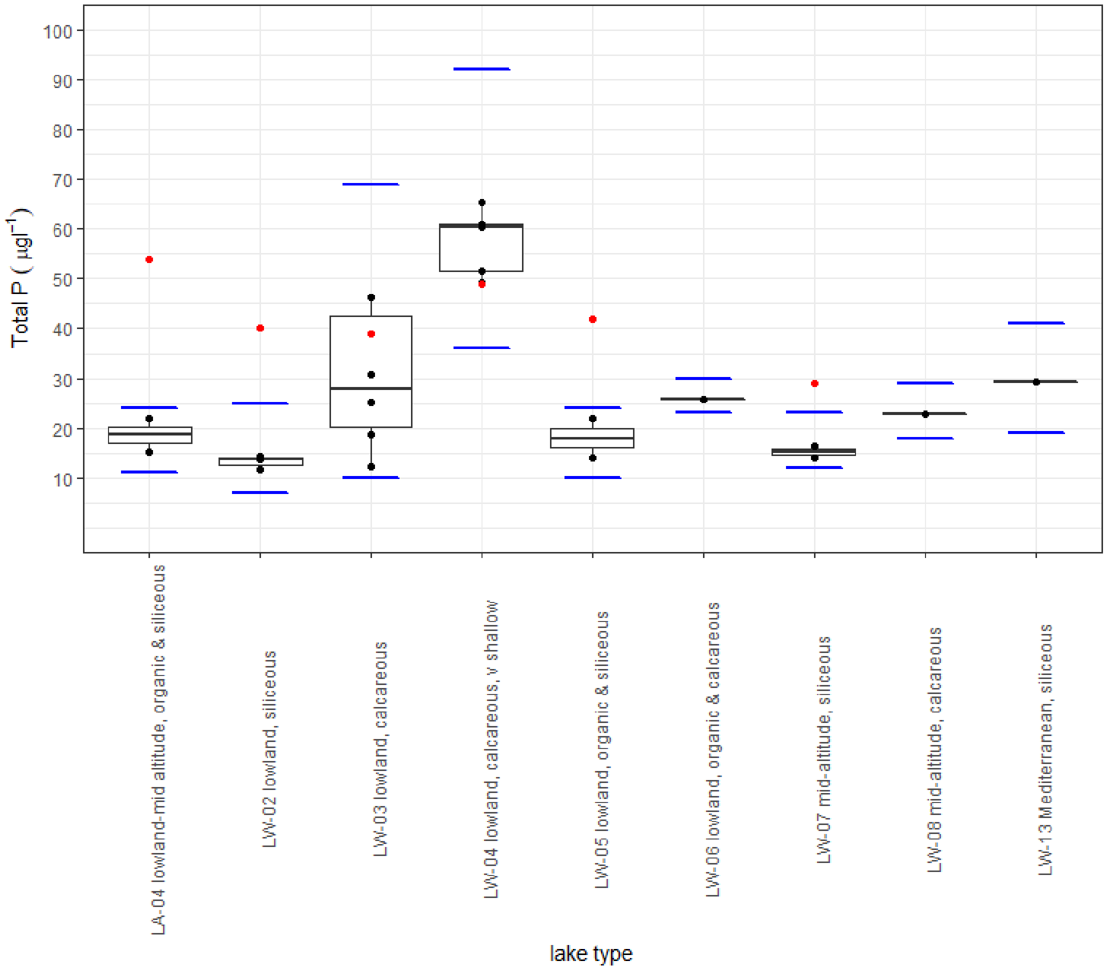

Figure 4.7. Range of predicted boundary values, overlaid on points showing values coloured by the probability threshold measure used, for broad lake types using 3 data sets (IC, SoE, WFD). Red points show the boundary values predicted using a GLM for comparison.

Table 4.1. Predicted type specific lake TP boundary values derived using binary logistic models, together with key measures from confusion matrix

lake type	dataset	boundary	lcl	ucl	p thresh	commission	omission	kappa	AUC	pseudo r2	countries
LA-04	IC	22	20	24	0.66	0.31	0.03	0.67	0.95	0.60	5
LA-04	SoE	15	11	20	0.88	0.28	0.26	0.30	0.84	0.34	5
LW-02	IC	12	10	15	0.88	0.54	0.11	0.23	0.82	0.17	7
LW-02	SoE	14	10	17	0.89	0.26	0.24	0.35	0.86	0.39	7
LW-02	WFD	14	7	25	0.67	0.32	0.30	0.36	0.73	0.21	8
LW-03	IC	25	23	27	0.59	0.25	0.25	0.49	0.83	0.40	16
LW-03	IC-A	12	10	15	0.55	0.24	0.25	0.51	0.87	0.56	3
LW-03	IC-C	31	27	35	0.57	0.27	0.26	0.46	0.80	0.31	10
LW-03	IC-N	19	17	20	0.77	0.10	0.18	0.69	0.96	0.80	3
LW-03	SoE	46	30	69	0.61	0.31	0.33	0.34	0.73	0.15	9
LW-03	WFD	46	38	56	0.56	0.34	0.33	0.33	0.73	0.21	11
LW-04	IC	60	48	74	0.64	0.16	0.32	0.50	0.79	0.29	15
LW-04	IC-C	65	54	81	0.51	0.23	0.26	0.51	0.82	0.37	10
LW-04	IC-CN	49	41	61	0.66	0.13	0.24	0.62	0.86	0.46	12
LW-04	SoE	52	36	72	0.57	0.24	0.23	0.53	0.82	0.34	7
LW-04	WFD	61	42	92	0.30	0.30	0.30	0.36	0.73	0.18	10
LW-05	IC	22	20	24	0.71	0.23	0.04	0.69	0.95	0.62	5
LW-05	SoE	14	10	18	0.88	0.31	0.28	0.25	0.81	0.26	2
LW-06	IC	26	23	30	0.56	0.18	0.20	0.62	0.88	0.53	5
LW-07	IC	14	12	17	0.65	0.31	0.03	0.66	0.96	0.62	3
LW-07	SoE	16	13	23	0.70	0.38	0.04	0.58	0.89	0.54	6
LW-08	IC	23	18	29	0.58	0.23	0.21	0.56	0.89	0.57	3
LW-13	IC	29	19	41	0.72	0.27	0.28	0.40	0.82	0.36	4

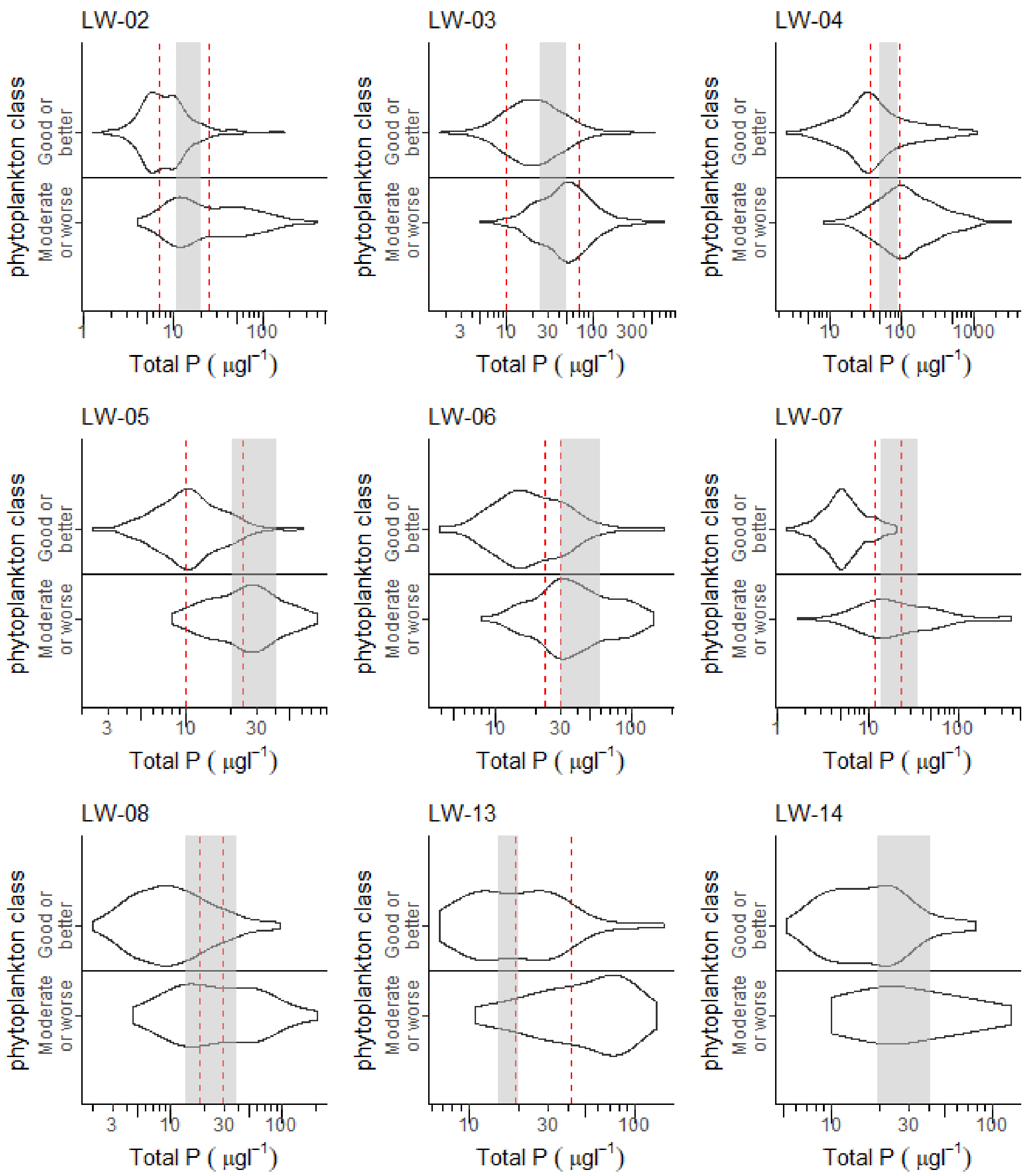

Figure 4.8. Violin plots showing the range of TP concentrations in lakes split by lake broad types overlain by proposed most likely boundary range (vertical red lines) derived from the highest and lowest confidence intervals of all predicted values. Grey shading marks interquartile range of Member State reported boundary values.

Table 4.2. Predicted type specific lake TP boundary values and key confusion matrix measures derived from binary models using different probability thresholds.

Select	dset	typology	Bound	lcl	ucl	measure	threshold	commission	omission	Prev	kappa	Auc	pseudo.r2	Ncountries
0	LA-04IC	LA-04	22	20	24	kappa	0.66	0.31	0.03	0.92	0.67	0.95	0.60	5
1	LA-04IC	LA-04	22	20	24	TKitMM	0.66	0.31	0.03	0.92	0.67	0.95	0.60	5
0	LA-04IC	LA-04	15	13	17	OmisComm	0.93	0.13	0.13	0.92	0.45	0.95	0.60	5
0	LA-04IC	LA-04	15	13	16	Com0.1	0.94	0.09	0.15	0.92	0.43	0.95	0.60	5
0	LA-04SoE	LA-04	27	21	36	TKitMM	0.73	0.44	0.07	0.86	0.49	0.84	0.34	5
0	LA-04SoE	LA-04	25	19	33	kappa	0.76	0.38	0.08	0.86	0.52	0.84	0.34	5
1	LA-04SoE	LA-04	15	11	20	OmisComm	0.88	0.28	0.26	0.86	0.30	0.84	0.34	5
0	LA-04SoE	LA-04	10	6	13	Com0.1	0.94	0.09	0.50	0.86	0.18	0.84	0.34	5
0	LW-02IC	LW-02	14	11	19	TKitMM	0.84	0.88	0.05	0.94	0.08	0.82	0.17	7
1	LW-02IC	LW-02	12	10	15	kappa	0.88	0.54	0.11	0.94	0.23	0.82	0.17	7
0	LW-02IC	LW-02	9	7	11	OmisComm	0.93	0.21	0.26	0.94	0.20	0.82	0.17	7
0	LW-02IC	LW-02	7	5	8	Com0.1	0.96	0.08	0.47	0.94	0.11	0.82	0.17	7
0	LW-02SoE	LW-02	26	21	33	kappa	0.73	0.36	0.06	0.85	0.58	0.86	0.39	7
0	LW-02SoE	LW-02	26	21	33	TKitMM	0.73	0.36	0.06	0.85	0.58	0.86	0.39	7
1	LW-02SoE	LW-02	14	10	17	OmisComm	0.89	0.26	0.24	0.85	0.35	0.86	0.39	7
0	LW-02SoE	LW-02	11	8	14	Com0.1	0.92	0.09	0.35	0.85	0.31	0.86	0.39	7
0	LW-02WFD	LW-02	18	11	35	TKitMM	0.62	0.44	0.23	0.64	0.33	0.73	0.21	8
0	LW-02WFD	LW-02	18	10	33	kappa	0.63	0.38	0.25	0.64	0.36	0.73	0.21	8
1	LW-02WFD	LW-02	14	7	25	OmisComm	0.67	0.32	0.30	0.64	0.36	0.73	0.21	8
0	LW-02WFD	LW-02	8	3	13	Com0.1	0.78	0.09	0.70	0.64	0.17	0.73	0.21	8
0	LW-03IC	LW-03	35	31	38	kappa	0.45	0.37	0.11	0.57	0.53	0.83	0.40	16
0	LW-03IC	LW-03	27	24	30	TKitMM	0.56	0.27	0.21	0.57	0.52	0.83	0.40	16
1	LW-03IC	LW-03	25	23	27	OmisComm	0.59	0.25	0.25	0.57	0.49	0.83	0.40	16
0	LW-03IC	LW-03	17	16	19	Com0.1	0.73	0.10	0.45	0.57	0.42	0.83	0.40	16
0	LW-03IC-A	LW-03	17	14	24	kappa	0.31	0.32	0.03	0.49	0.66	0.87	0.56	3
0	LW-03IC-A	LW-03	12	10	15	OmisComm	0.55	0.24	0.25	0.49	0.51	0.87	0.56	3
1	LW-03IC-A	LW-03	12	10	15	TKitMM	0.55	0.24	0.25	0.49	0.51	0.87	0.56	3
0	LW-03IC-A	LW-03	10	7	12	Com0.1	0.71	0.10	0.40	0.49	0.50	0.87	0.56	3

Select	dset	typology	Bound	lcl	ucl	measure	threshold	commission	omission	Prev	kappa	Auc	pseudo.r2	Ncountries
0	LW-03IC-C	LW-03	35	31	41	kappa	0.52	0.30	0.19	0.55	0.51	0.80	0.31	10
0	LW-03IC-C	LW-03	33	28	38	TKitMM	0.55	0.27	0.24	0.55	0.49	0.80	0.31	10
1	LW-03IC-C	LW-03	31	27	35	OmisComm	0.57	0.27	0.26	0.55	0.46	0.80	0.31	10
0	LW-03IC-C	LW-03	18	15	22	Com0.1	0.74	0.09	0.57	0.55	0.32	0.80	0.31	10
0	LW-03IC-N	LW-03	24	22	27	kappa	0.39	0.18	0.02	0.63	0.83	0.96	0.80	3
0	LW-03IC-N	LW-03	21	19	23	TKitMM	0.62	0.16	0.09	0.63	0.75	0.96	0.80	3
0	LW-03IC-N	LW-03	19	18	21	OmisComm	0.72	0.14	0.14	0.63	0.70	0.96	0.80	3
1	LW-03IC-N	LW-03	19	17	20	Com0.1	0.77	0.10	0.18	0.63	0.69	0.96	0.80	3
0	LW-03SoE	LW-03	52	36	82	TKitMM	0.58	0.46	0.26	0.62	0.28	0.73	0.15	9
1	LW-03SoE	LW-03	46	30	69	OmisComm	0.61	0.31	0.33	0.62	0.34	0.73	0.15	9
0	LW-03SoE	LW-03	39	23	56	kappa	0.65	0.13	0.43	0.62	0.39	0.73	0.15	9
0	LW-03SoE	LW-03	34	19	49	Com0.1	0.68	0.09	0.50	0.62	0.36	0.73	0.15	9
0	LW-03WFD	LW-03	49	41	61	kappa	0.54	0.37	0.28	0.56	0.35	0.73	0.21	11
0	LW-03WFD	LW-03	49	41	61	TKitMM	0.54	0.37	0.28	0.56	0.35	0.73	0.21	11
1	LW-03WFD	LW-03	46	38	56	OmisComm	0.56	0.34	0.33	0.56	0.33	0.73	0.21	11
0	LW-03WFD	LW-03	29	21	36	Com0.1	0.71	0.10	0.64	0.56	0.24	0.73	0.21	11
0	LW-04IC	LW-04	90	74	115	TKitMM	0.55	0.31	0.25	0.56	0.44	0.79	0.29	15
0	LW-04IC	LW-04	83	68	108	OmisComm	0.57	0.26	0.26	0.56	0.48	0.79	0.29	15
1	LW-04IC	LW-04	60	48	74	kappa	0.64	0.16	0.32	0.56	0.50	0.79	0.29	15
0	LW-04IC	LW-04	47	34	61	Com0.1	0.69	0.10	0.40	0.56	0.47	0.79	0.29	15
0	LW-04IC-C	LW-04	69	54	88	OmisComm	0.49	0.25	0.24	0.48	0.52	0.82	0.37	10
1	LW-04IC-C	LW-04	65	54	81	TKitMM	0.51	0.23	0.26	0.48	0.51	0.82	0.37	10
0	LW-04IC-C	LW-04	49	41	61	kappa	0.60	0.13	0.33	0.48	0.55	0.82	0.37	10
0	LW-04IC-C	LW-04	43	34	54	Com0.1	0.64	0.10	0.41	0.48	0.50	0.82	0.37	10
0	LW-04IC-CN	LW-04	68	54	81	TKitMM	0.54	0.24	0.18	0.56	0.58	0.86	0.46	12
0	LW-04IC-CN	LW-04	61	54	74	OmisComm	0.58	0.20	0.19	0.56	0.60	0.86	0.46	12
1	LW-04IC-CN	LW-04	49	41	61	kappa	0.66	0.13	0.24	0.56	0.62	0.86	0.46	12
0	LW-04IC-CN	LW-04	42	34	48	Com0.1	0.71	0.10	0.34	0.56	0.55	0.86	0.46	12
0	LW-04SoE	LW-04	55	39	77	TKitMM	0.55	0.27	0.21	0.58	0.51	0.82	0.34	7
0	LW-04SoE	LW-04	53	38	75	kappa	0.56	0.24	0.21	0.58	0.55	0.82	0.34	7
1	LW-04SoE	LW-04	52	36	72	OmisComm	0.57	0.24	0.23	0.58	0.53	0.82	0.34	7

Select	dset	typology	Bound	lcl	ucl	measure	threshold	commission	omission	Prev	kappa	Auc	pseudo.r2	Ncountries
0	LW-04SoE	LW-04	36	22	49	Com0.1	0.68	0.09	0.35	0.58	0.54	0.82	0.34	7
0	LW-04WFD	LW-04	61	42	92	kappa	0.30	0.30	0.30	0.29	0.36	0.73	0.18	10
1	LW-04WFD	LW-04	61	42	92	OmisComm	0.30	0.30	0.30	0.29	0.36	0.73	0.18	10
0	LW-04WFD	LW-04	43	27	60	TKitMM	0.38	0.20	0.49	0.29	0.30	0.73	0.18	10
0	LW-04WFD	LW-04	30	16	44	Com0.1	0.47	0.10	0.77	0.29	0.15	0.73	0.18	10
0	LW-05IC	LW-05	23	21	26	TKitMM	0.64	0.30	0.03	0.91	0.68	0.95	0.62	5
1	LW-05IC	LW-05	22	20	24	kappa	0.71	0.23	0.04	0.91	0.69	0.95	0.62	5
0	LW-05IC	LW-05	16	14	18	OmisComm	0.93	0.13	0.15	0.91	0.45	0.95	0.62	5
0	LW-05IC	LW-05	16	14	18	Com0.1	0.93	0.13	0.15	0.91	0.45	0.95	0.62	5
0	LW-05SoE	LW-05	25	19	35	kappa	0.72	0.50	0.07	0.87	0.44	0.81	0.26	2
0	LW-05SoE	LW-05	25	19	35	TKitMM	0.72	0.50	0.07	0.87	0.44	0.81	0.26	2
1	LW-05SoE	LW-05	14	10	18	OmisComm	0.88	0.31	0.28	0.87	0.25	0.81	0.26	2
0	LW-05SoE	LW-05	10	6	13	Com0.1	0.93	0.12	0.46	0.87	0.19	0.81	0.26	2
0	LW-06IC	LW-06	27	23	32	TKitMM	0.53	0.25	0.18	0.59	0.57	0.88	0.53	5
0	LW-06IC	LW-06	26	23	30	kappa	0.56	0.18	0.20	0.59	0.62	0.88	0.53	5
1	LW-06IC	LW-06	26	23	30	OmisComm	0.56	0.18	0.20	0.59	0.62	0.88	0.53	5
0	LW-06IC	LW-06	16	12	19	Com0.1	0.84	0.10	0.46	0.59	0.40	0.88	0.53	5
0	LW-07IC	LW-07	14	12	18	kappa	0.61	0.31	0.02	0.90	0.72	0.96	0.62	3
1	LW-07IC	LW-07	14	12	17	TKitMM	0.65	0.31	0.03	0.90	0.66	0.96	0.62	3
0	LW-07IC	LW-07	10	7	12	OmisComm	0.92	0.15	0.15	0.90	0.45	0.96	0.62	3
0	LW-07IC	LW-07	9	6	11	Com0.1	0.93	0.08	0.17	0.90	0.46	0.96	0.62	3
0	LW-07SoE	LW-07	20	16	30	kappa	0.57	0.42	0.00	0.90	0.72	0.89	0.54	6
1	LW-07SoE	LW-07	16	13	23	TKitMM	0.70	0.38	0.04	0.90	0.58	0.89	0.54	6
0	LW-07SoE	LW-07	8	6	10	OmisComm	0.93	0.17	0.16	0.90	0.44	0.89	0.54	6
0	LW-07SoE	LW-07	8	6	10	Com0.1	0.94	0.12	0.19	0.90	0.41	0.89	0.54	6
0	LW-08IC	LW-08	25	20	33	kappa	0.51	0.23	0.17	0.58	0.60	0.89	0.57	3
0	LW-08IC	LW-08	25	20	33	TKitMM	0.51	0.23	0.17	0.58	0.60	0.89	0.57	3
1	LW-08IC	LW-08	23	18	29	OmisComm	0.58	0.23	0.21	0.58	0.56	0.89	0.57	3
0	LW-08IC	LW-08	18	12	22	Com0.1	0.73	0.11	0.33	0.58	0.53	0.89	0.57	3
0	LW-08SoE	LW-08	36	10	47	TKitMM	0.79	0.85	0.11	0.87	0.05	0.66	0.04	4
1	LW-08SoE	LW-08	11	4	47	OmisComm	0.87	0.31	0.40	0.87	0.15	0.66	0.04	4

Select	dset	typology	Bound	lcl	ucl	measure	threshold	commission	omission	Prev	kappa	Auc	pseudo.r2	Ncountries
0	LW-08SoE	LW-08	9	4	47	kappa	0.88	0.15	0.48	0.87	0.17	0.66	0.04	4
0	LW-08SoE	LW-08	6	4	25	Com0.1	0.90	0.08	0.61	0.87	0.12	0.66	0.04	4
0	LW-13IC	LW-13	41	30	65	kappa	0.58	0.32	0.07	0.71	0.63	0.82	0.36	4
0	LW-13IC	LW-13	37	27	56	TKitMM	0.62	0.32	0.15	0.71	0.53	0.82	0.36	4
1	LW-13IC	LW-13	29	19	41	OmisComm	0.72	0.27	0.28	0.71	0.40	0.82	0.36	4
0	LW-13IC	LW-13	17	8	25	Com0.1	0.87	0.09	0.52	0.71	0.29	0.82	0.36	4
0	LW-14IC	LW-14	58	31	130	kappa	0.65	0.62	0.04	0.86	0.40	0.69	0.15	4
0	LW-14IC	LW-14	38	19	130	TKitMM	0.76	0.62	0.10	0.86	0.27	0.69	0.15	4
1	LW-14IC	LW-14	23	1	61	OmisComm	0.86	0.25	0.39	0.86	0.19	0.69	0.15	4
0	LW-14IC	LW-14	9	1	19	Com0.1	0.95	0.00	0.84	0.86	0.05	0.69	0.15	4

5 Lake total nitrogen boundaries, model results

5.1 Model performance

As for the TP models, to assess performance we compared how successful the resulting TN classifications were in comparison to the true classification derived from phytoplankton using kappa, a measure of overall agreement. The GLM models again generated classifications with an overall higher level of agreement (i.e. higher value of kappa) than the BLM models, but with a higher commission (false +ve) rate (Figures 5.1). The overall performance of the TN models was slightly worse than for TP with lower kappa values, although most of the BLM models were above the 0.21 threshold suggested as an acceptable value (see 2.7).

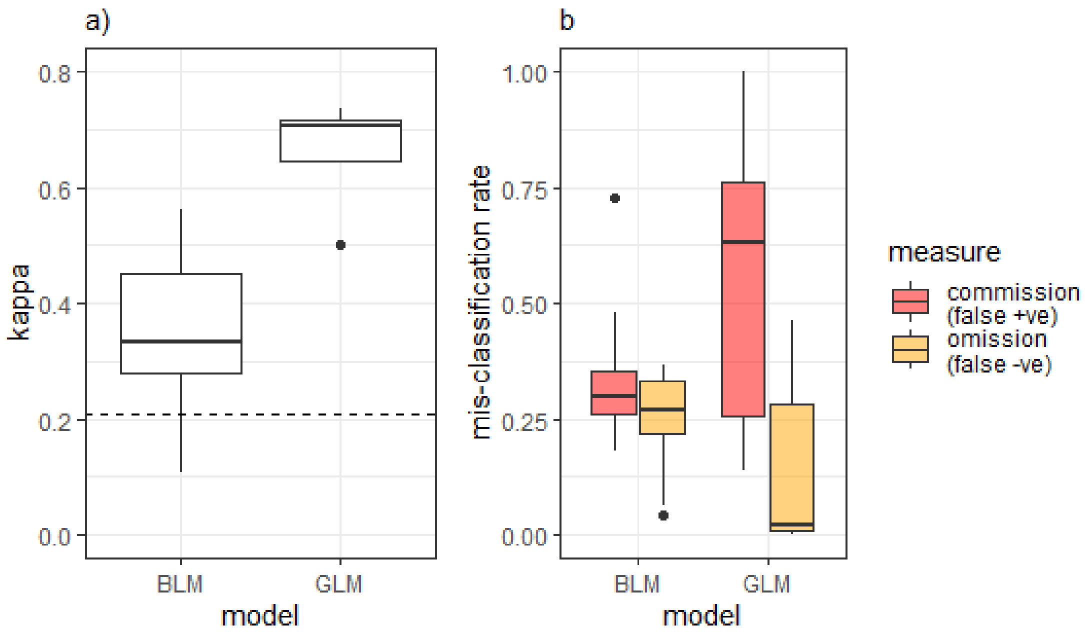

Figure 5.1. Comparison of mis-classification outcomes using different modelling (GLM and BLM) approaches a) value of kappa, b) commission (false+ve) and omission (false -ve) rates. (horizontal dotted line marks kappa = 0.21, a suggested level for acceptable accuracy, see 2.7).

5.2 Selecting appropriate threshold probability thresholds for BLM models

Table 5.2 details the lake TN values derived from the BLM models applied to data from different lake types, together with the mis-classification rates and model fit parameters. For each type four potential measures are shown (toolkit mismatch method where false +ve & -ve rates are equal; maximum kappa; omission & commission rates are equal, commission rate = 0.1.). The range of probability thresholds for each of these criteria, the resulting TN boundary and relative mis-classification rates are shown in Figure 5.2. In most cases the selecting either the tool-kit mis-match method or the probability threshold that maximised correct classifications (maximum kappa) generated the lowest probability threshold and thus higher boundary value (Figure 5.2ab). However, these measures generated higher commission (false +ve) rates. Given the need to establish a relatively precautionary boundary a lower threshold was selected that minimised the commission (false +ve) rate, subject to that probability being ≤ 0.9, kappa ≥ 0.9 and omission < double commission rate. These values are shown in table 5.1and in figure 5.3. They can be compared with boundary concentrations predicted using the other potential probability thresholds in table 5.2.

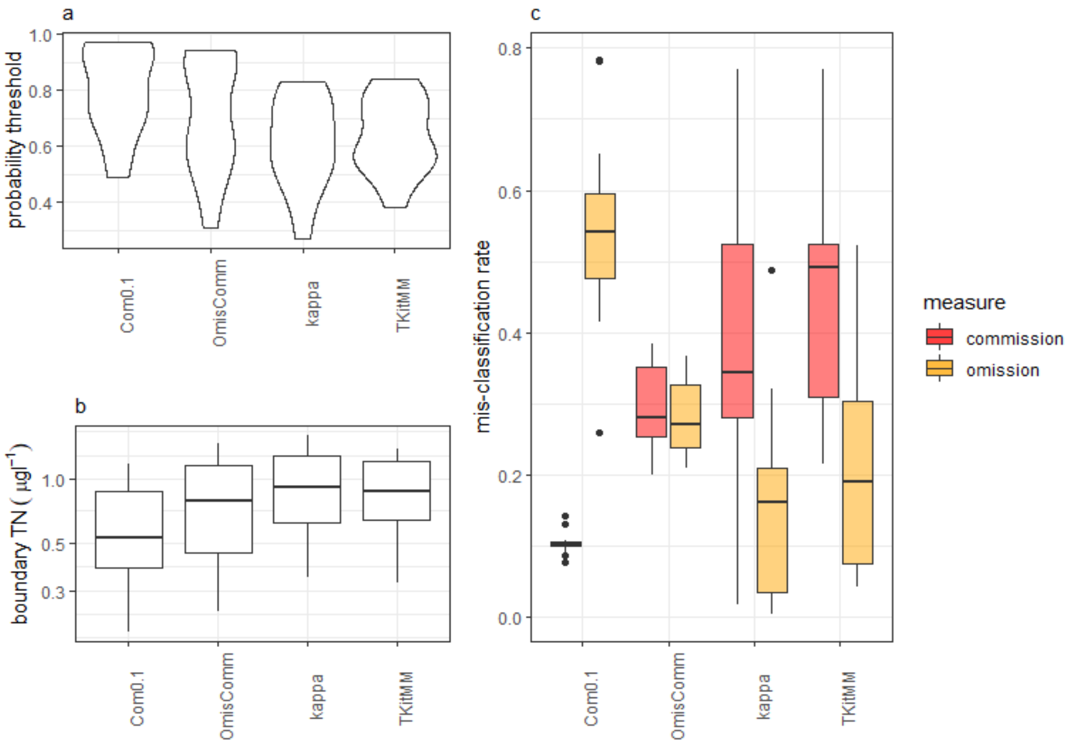

Figure 5.2. Effect of different thershold probability measures on a) the probability threshold, b) predicted TN boundary values, c) mis-classification rates (commission false +ve, omission false -ve).

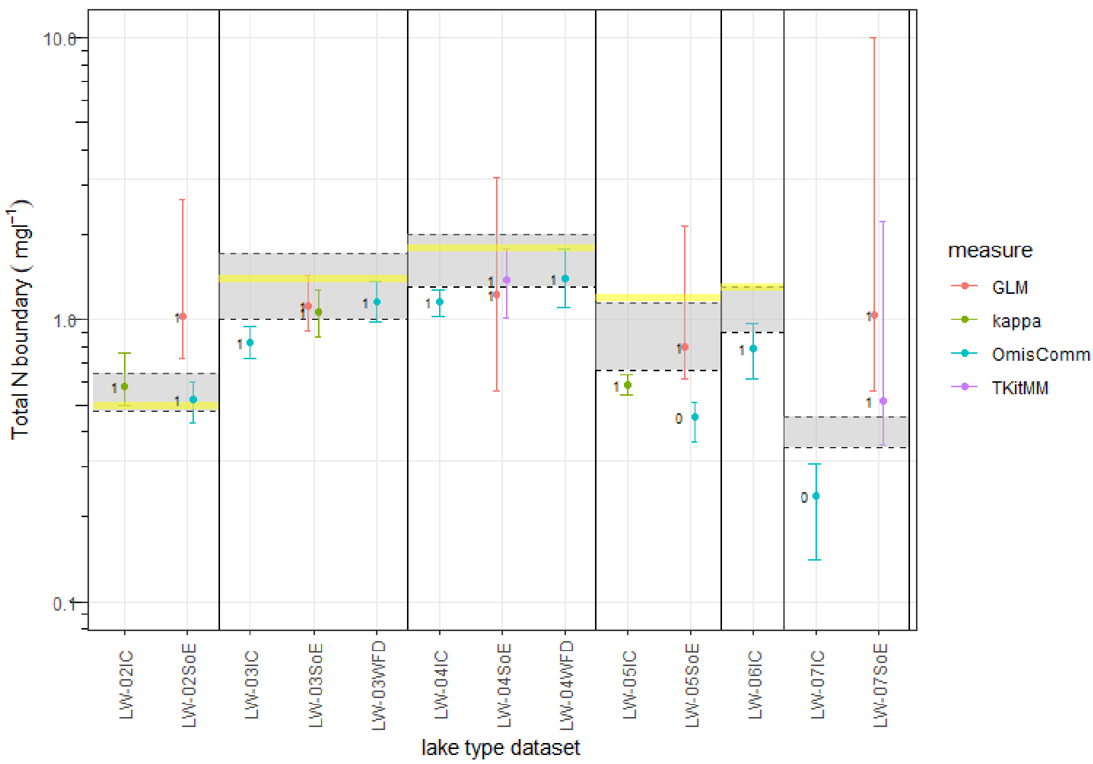

Figure 5.3. Predicted total nitrogen boundary values for rivers by broad type using models fitted to data from all available countries, compared to upper and lower quartiles of MS boundary values (grey shading) and estimated pan-European boundary values reported by (Nikolaidis et al. 2021) (yellow shading). (Vertical lines show 95% confidence limits).

5.3 Lake TN boundary values

The predicted good moderate boundary values varied from 0.2 - 1.4 mgL^{-1}. The differences between types were smaller than for the TP boundaries, but the pattern was similar. The highest values occurred in the lowland calcareous very shallow lakes (LW-04), followed by the lowland calcareous shallow stratified lakes (LW-03), with the lowest found in the lowland siliceous (LW-02) and mid-altitude siliceous (LW07) lakes. As for the TP boundaries the range of the TN predicted boundary values was generally similar to the inter quartile range of the values reported by Member States (grey shaded area in Figure 5.3).

5.3.1 Effect of model used

Predictions using the continuous GLM model were in most, but not all, cases higher than those made using the binary logistic models (Figures 5.3) with values often exceeding the interquartile range of MS boundary values.

The differences between the boundaries predicted using the measures used to determine probability cut thresholds for the discrete binary models was smaller (Figure 5.2b), but as for TP higher boundaries were predicted when the maximum of *kappa* was used in comparison to the equivalence of *omission* and *commission*.

5.3.2 Importance of data set used

5.3.2.1 Representativeness

As was the case for TP the boundary values there was no evidence any bias resulting from the different data sets (Figure 5.5). The WFD data set was the smallest and only had sufficient data to allow modelling of the lowland calcareous lakes (LW-03, LW-04). For TN, although fewer countries reported TN, the IC data set had more countries represented in the data sets (Figure 5.4a) than was the case for TP. Both this and the SoE data sets were however less balanced with higher prevalence values indicating dominance of records where phytoplankton was good or better (Figure 5.4b).

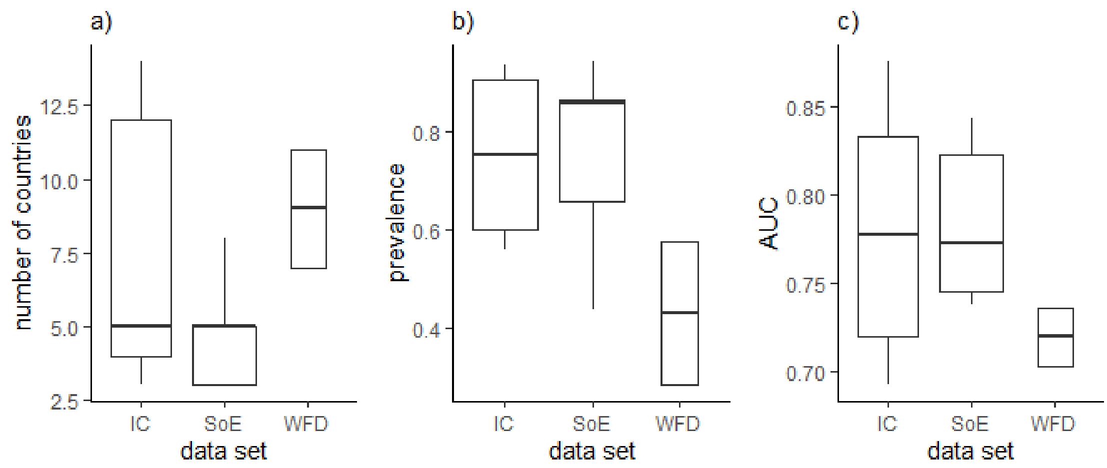

Figure 5.4. Range of a) number of countries in dataset; b) prevalence of the data; c) AUC of data.

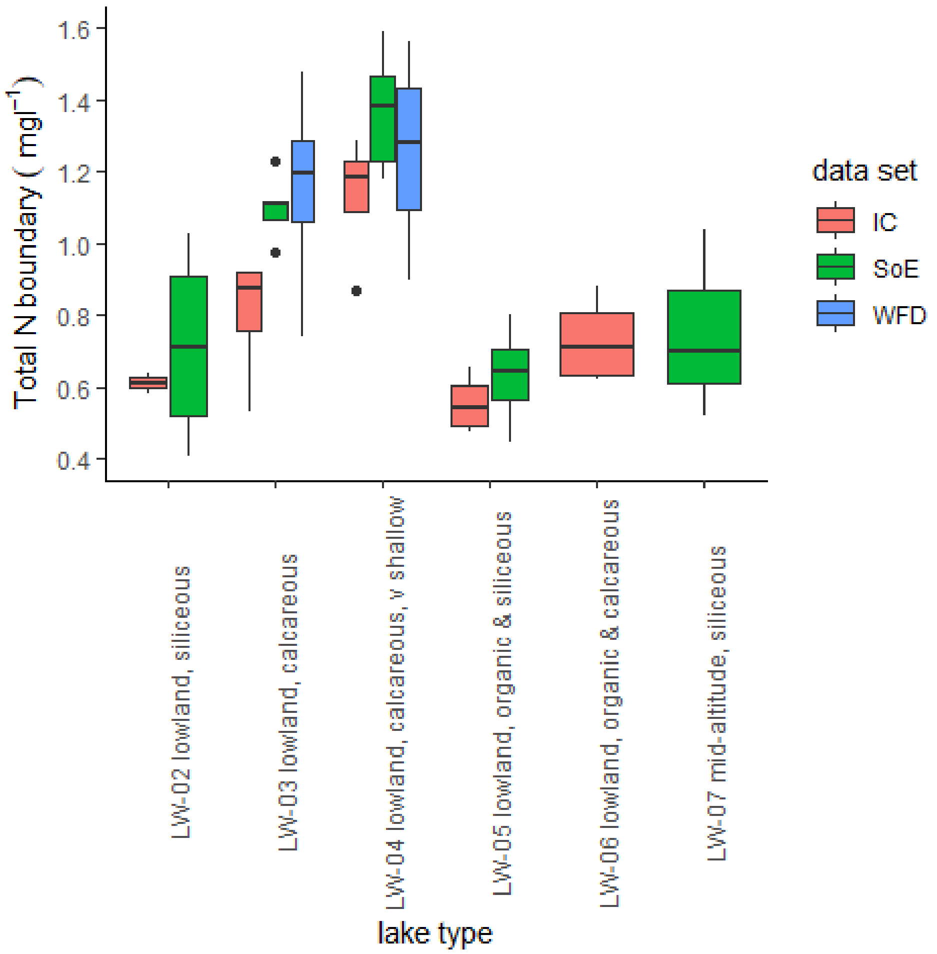

Figure 5.5. Range of predicted TN good moderate boundary values by lake type and data set.

5.3.2.2 Regional variation

The distribution of TN in the two phytoplankton status classes for the three common lowland lake types (siliceous LW-02, calcareous stratified LW-03 & calcareous very shallow LW-04) is shown in Figure 5.6. Regional differences for TN were less clear than they were for TP. There were only sufficient data to model regions separately for of these types (siliceous LW-02 and calcareous stratified LW-03) and for both of these there was no significant effect of including region.

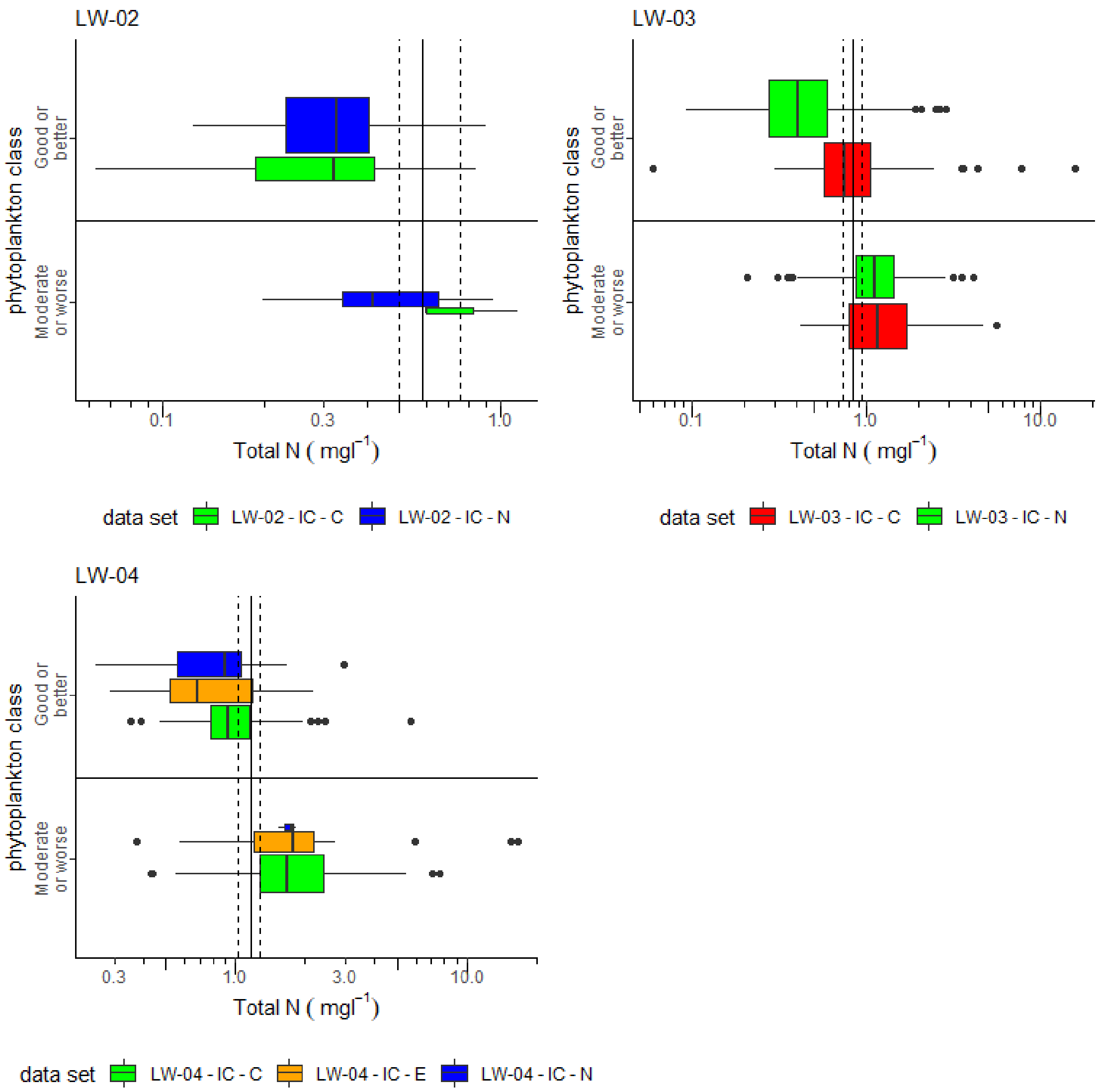

Figure 5.6. Regional distribution of TN by phytoplankton class overlain by vertical lines showing predicted boundary values (± 95% confidence limits) from binomial logistic models.

5.4 Most likely TN boundary ranges

The resulting boundary values from each of the data sets are provided in Table 5.1. Using the same approach as that used for TP, all of the predicted boundary values from the binary logistic models are presented as box plots together with the highest and lowest of the boundary 95% confidence intervals (Figure 5.7). The most likely boundary ranges are then indicated by either the interquartile ranges (boxes in Figure 5.7, or where these are very small, due to restricted data, the range provided by the 95% confidence limits (blue bars in Figure 5.7).

To place these values into context the range provided by the maximum and minimum values of the 95th percentiles are compared with the distribution of TN of the combined data sets and the most recently reported MS boundary values (Figure 5.8).

Table 5.1. Predicted type specific TN boundary values (mg l^{-1}) derived using binary logistic models, together with key measures from confusion matrix.

lake type	dataset	boundary	lcl	ucl	p thresh	commission	omission	kappa	AUC	pseudo r2	countries
LW-02	IC	0.6	0.5	0.8	0.83	0.48	0.06	0.38	0.76	0.19	6
LW-02	SoE	0.5	0.4	0.6	0.87	0.26	0.26	0.30	0.84	0.39	5
LW-03	IC	0.8	0.7	0.9	0.60	0.26	0.27	0.45	0.79	0.29	12
LW-03	SoE	1.1	0.9	1.3	0.66	0.18	0.32	0.45	0.77	0.21	8
LW-03	WFD	1.2	1.0	1.4	0.58	0.35	0.37	0.27	0.70	0.17	11
LW-04	IC	1.2	1.0	1.3	0.59	0.22	0.22	0.56	0.83	0.40	14
LW-04	SoE	1.4	1.0	1.8	0.48	0.26	0.33	0.41	0.74	0.24	5
LW-04	WFD	1.4	1.1	1.8	0.31	0.32	0.33	0.31	0.74	0.19	7
LW-05	IC	0.6	0.5	0.6	0.81	0.30	0.08	0.51	0.88	0.34	4
LW-05	SoE	0.4	0.4	0.5	0.89	0.24	0.24	0.33	0.82	0.32	3
LW-06	IC	0.8	0.6	1.0	0.60	0.35	0.36	0.28	0.72	0.19	4
LW-07	SoE	0.5	0.4	2.2	0.84	0.73	0.04	0.23	0.75	0.10	3

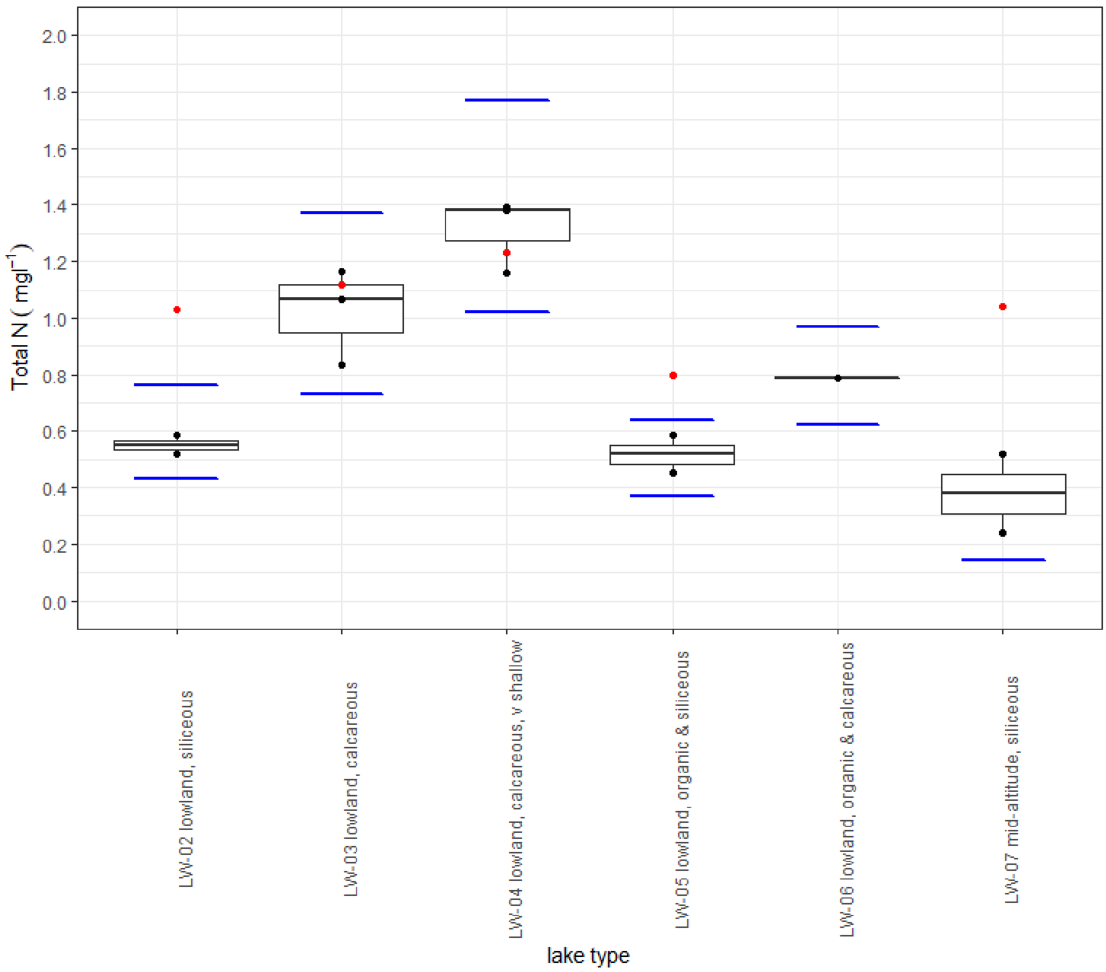

Figure 5.7. Range of predicted boundary values, overlaid on points showing values coloured by the probability threshold measure used, for broad lake types using 3 data sets (IC, SoE, WFD).

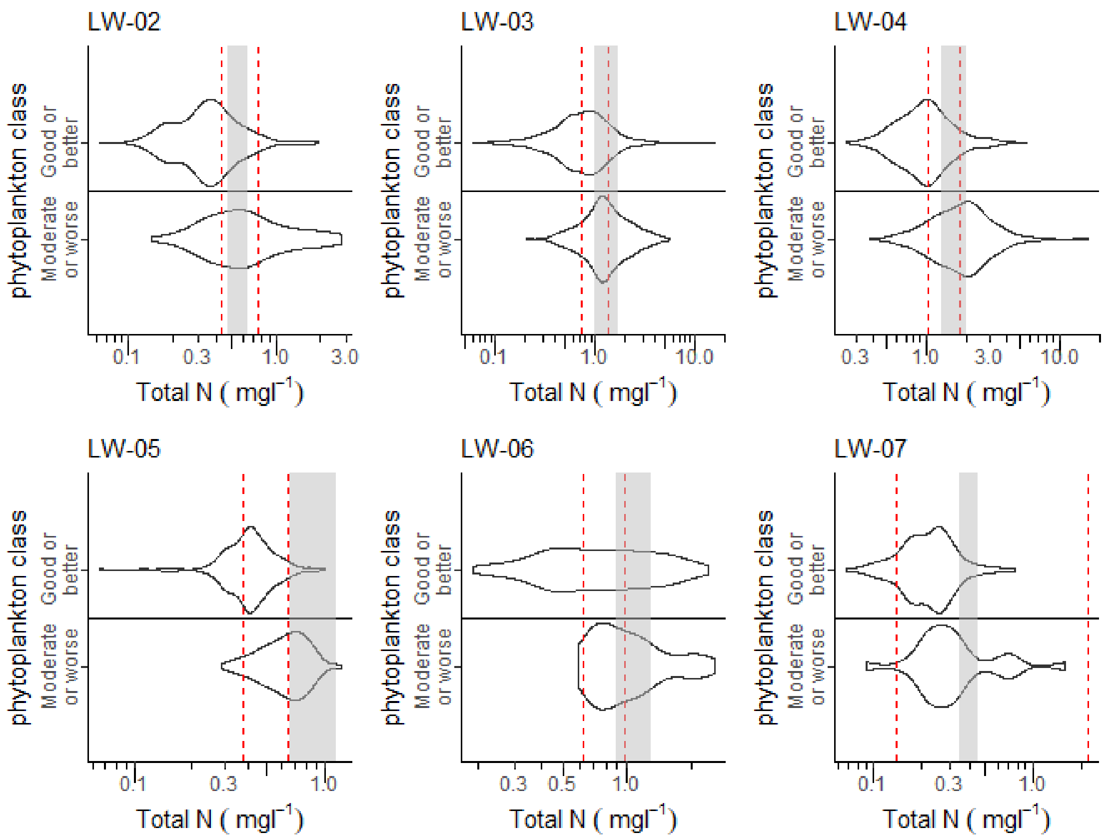

Figure 5.8. Violin plots showing the range of TN concentrations in lakes split by lake broad types overlain by proposed most likely boundary range (vertical red lines) derived from the highest and lowest confidence intervals of all predicted values. Grey shading marks interquartile range of Member State reported boundary values.

Table 5.2. Predicted type specific lake TN boundary values (mg l^{-1}) and key confusion matrix measures derived from binary models using different probability thresholds.

Select	typology	Bound	lcl	ucl	measure	threshold	commission	omission	Prev	kappa	Auc	pseudo.r2	Ncountries
0	LW-02	0.9	0.8	1.2	kappa	0.57	0.58	0.01	0.86	0.52	0.84	0.39	5
0	LW-02	0.7	0.6	0.8	TKitMM	0.73	0.50	0.08	0.86	0.42	0.84	0.39	5
0	LW-02	0.6	0.5	0.9	TKitMM	0.79	0.70	0.04	0.94	0.27	0.76	0.19	6
1	LW-02	0.6	0.5	0.8	kappa	0.83	0.48	0.06	0.94	0.38	0.76	0.19	6
1	LW-02	0.5	0.4	0.6	OmisComm	0.87	0.26	0.26	0.86	0.30	0.84	0.39	5
0	LW-02	0.4	0.3	0.5	Com0.1	0.93	0.11	0.43	0.86	0.23	0.84	0.39	5
0	LW-02	0.4	0.3	0.4	OmisComm	0.94	0.35	0.33	0.94	0.11	0.76	0.19	6
0	LW-02	0.3	0.2	0.4	Com0.1	0.97	0.13	0.60	0.94	0.05	0.76	0.19	6
0	LW-03	1.5	1.3	1.8	kappa	0.48	0.52	0.16	0.58	0.33	0.70	0.17	11
0	LW-03	0.9	0.8	1.0	kappa	0.56	0.31	0.21	0.61	0.48	0.79	0.29	12
0	LW-03	0.9	0.8	1.0	TKitMM	0.56	0.31	0.21	0.61	0.48	0.79	0.29	12
0	LW-03	1.2	1.0	1.4	TKitMM	0.56	0.41	0.30	0.58	0.29	0.70	0.17	11
1	LW-03	1.2	1.0	1.4	OmisComm	0.58	0.35	0.37	0.58	0.27	0.70	0.17	11
0	LW-03	1.2	1.0	1.5	TKitMM	0.59	0.42	0.20	0.66	0.39	0.77	0.21	8
1	LW-03	0.8	0.7	0.9	OmisComm	0.60	0.26	0.27	0.61	0.45	0.79	0.29	12
0	LW-03	1.1	0.9	1.3	OmisComm	0.64	0.25	0.28	0.66	0.43	0.77	0.21	8
1	LW-03	1.1	0.9	1.3	kappa	0.66	0.18	0.32	0.66	0.45	0.77	0.21	8
0	LW-03	1.0	0.8	1.2	Com0.1	0.70	0.11	0.42	0.66	0.41	0.77	0.21	8
0	LW-03	0.7	0.5	0.9	Com0.1	0.75	0.11	0.78	0.58	0.10	0.70	0.17	11
0	LW-03	0.5	0.4	0.6	Com0.1	0.76	0.10	0.55	0.61	0.31	0.79	0.29	12
0	LW-04	1.6	1.3	2.1	kappa	0.27	0.34	0.20	0.28	0.38	0.74	0.19	7
1	LW-04	1.4	1.1	1.8	OmisComm	0.31	0.32	0.33	0.28	0.31	0.74	0.19	7
0	LW-04	1.2	0.9	1.4	TKitMM	0.38	0.22	0.52	0.28	0.26	0.74	0.19	7
0	LW-04	1.6	1.2	2.1	kappa	0.41	0.28	0.23	0.44	0.48	0.74	0.24	5
0	LW-04	1.5	1.1	1.9	OmisComm	0.45	0.28	0.26	0.44	0.46	0.74	0.24	5
1	LW-04	1.4	1.0	1.8	TKitMM	0.48	0.26	0.33	0.44	0.41	0.74	0.24	5
0	LW-04	0.9	0.6	1.1	Com0.1	0.49	0.10	0.78	0.28	0.14	0.74	0.19	7

Select	typology	Bound	lcl	ucl	measure	threshold	commission	omission	Prev	kappa	Auc	pseudo.r2	Ncountries
0	LW-04	1.3	1.2	1.4	kappa	0.52	0.25	0.17	0.56	0.57	0.83	0.40	14
0	LW-04	1.2	1.1	1.3	TKitMM	0.56	0.25	0.19	0.56	0.56	0.83	0.40	14
0	LW-04	1.2	0.8	1.5	Com0.1	0.56	0.10	0.51	0.44	0.40	0.74	0.24	5
1	LW-04	1.2	1.0	1.3	OmisComm	0.59	0.22	0.22	0.56	0.56	0.83	0.40	14
0	LW-04	0.9	0.7	1.0	Com0.1	0.76	0.10	0.57	0.56	0.31	0.83	0.40	14
0	LW-05	0.7	0.6	0.8	kappa	0.59	0.52	0.04	0.87	0.50	0.82	0.32	3
0	LW-05	0.7	0.6	0.7	TKitMM	0.70	0.50	0.05	0.91	0.45	0.88	0.34	4
0	LW-05	0.6	0.5	0.7	TKitMM	0.70	0.52	0.08	0.87	0.40	0.82	0.32	3
1	LW-05	0.6	0.5	0.6	kappa	0.81	0.30	0.08	0.91	0.51	0.88	0.34	4
1	LW-05	0.4	0.4	0.5	OmisComm	0.89	0.24	0.24	0.87	0.33	0.82	0.32	3
0	LW-05	0.5	0.4	0.5	OmisComm	0.91	0.20	0.21	0.91	0.32	0.88	0.34	4
0	LW-05	0.5	0.4	0.5	Com0.1	0.93	0.10	0.26	0.91	0.30	0.88	0.34	4
0	LW-05	0.4	0.3	0.4	Com0.1	0.94	0.14	0.48	0.87	0.17	0.82	0.32	3
0	LW-06	0.9	0.7	1.1	TKitMM	0.55	0.49	0.33	0.60	0.18	0.72	0.19	4
1	LW-06	0.8	0.6	1.0	OmisComm	0.60	0.35	0.36	0.60	0.28	0.72	0.19	4
0	LW-06	0.6	0.5	0.8	Com0.1	0.69	0.09	0.48	0.60	0.39	0.72	0.19	4
0	LW-06	0.6	0.4	0.8	kappa	0.70	0.02	0.49	0.60	0.44	0.72	0.19	4
0	LW-07	0.7	0.5	5.0	kappa	0.75	0.73	0.01	0.94	0.38	0.75	0.10	3
0	LW-07	0.3	0.3	1.0	kappa	0.77	0.77	0.03	0.90	0.25	0.69	0.12	3
0	LW-07	0.3	0.3	0.7	TKitMM	0.80	0.77	0.08	0.90	0.16	0.69	0.12	3
1	LW-07	0.5	0.4	2.2	TKitMM	0.84	0.73	0.04	0.94	0.23	0.75	0.10	3
1	LW-07	0.2	0.1	0.3	OmisComm	0.90	0.38	0.36	0.90	0.11	0.69	0.12	3
0	LW-07	0.2	0.1	0.3	Com0.1	0.93	0.08	0.54	0.90	0.12	0.69	0.12	3
0	LW-07	0.3	0.1	0.4	OmisComm	0.94	0.36	0.24	0.94	0.14	0.75	0.10	3
0	LW-07	0.2	0.0	0.3	Com0.1	0.97	0.09	0.65	0.94	0.04	0.75	0.10	3

6 Comparison of lake TP & TN boundary values

The range of proposed boundary values for TP and TN are shown in Figure 6.1, by lines joining the lower and upper proposed TP and TN boundaries. Clearly the TN and TP boundaries are related to each other and given that they lie above the line where the N:P ratio suggests P limitation it suggests that the TN boundary is unlikely to control status, indicating the greater importance of the TP boundary values for these lake types.

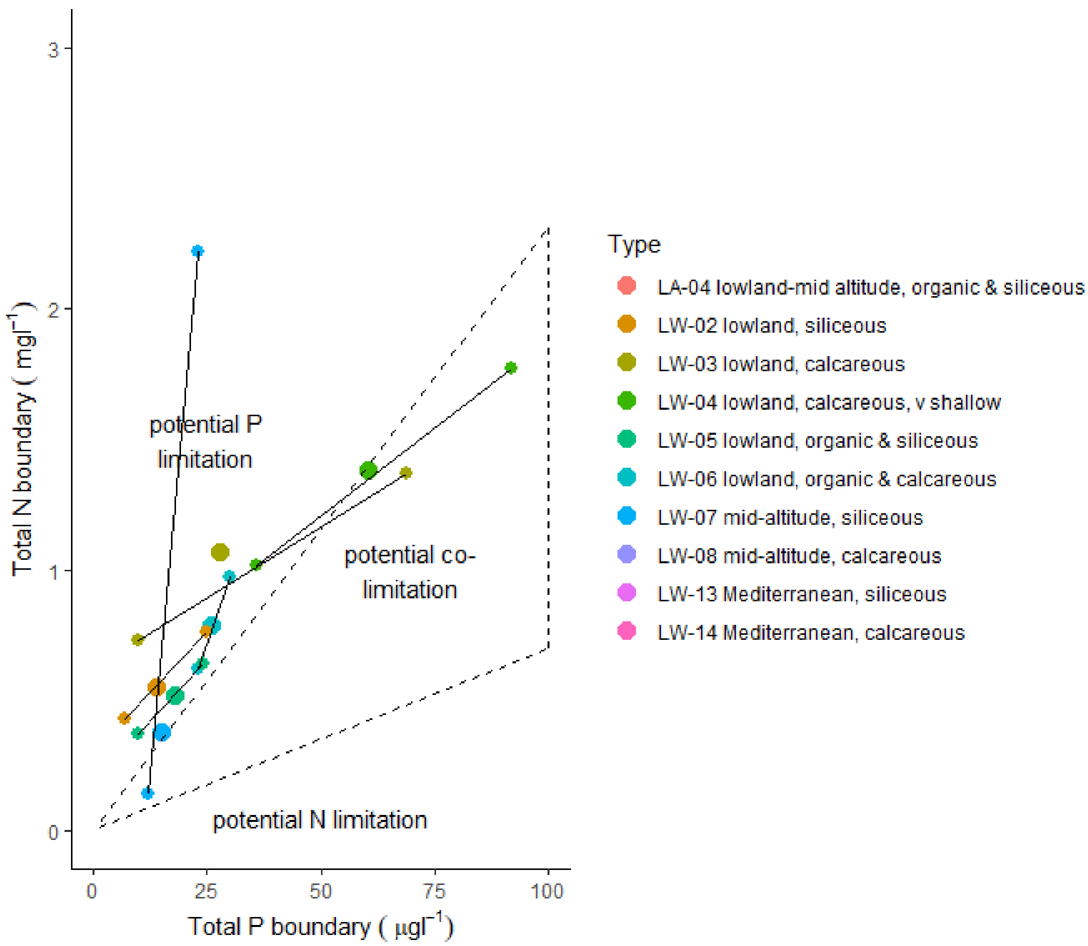

Figure 6.1. Comparison of modelled boundary values for different lake types. Lines mark the upper and lower values of the most likely boundaries. Dotted lines show N:P rations where N, P or co N&P limitation may occur: <16 molar ratio N limitation, >53 molar ratio P limitation (Ptacnik et al. 2010).

Table 6.1 summarises the results of the modelling providing a type specific median boundary value derived from all models together with a typical range of values based on the highest and lowest 95% confidence limits of the predicted values.

Table 6.1. Summary of type specific lake TP and TN boundary values, derived from the range of confidence limits of predicted boundary values using binary logistic models.

	TP (ugL^{-1})		TN (mgL^{-1})	
typology	median	range	median	range
LA-04 lowland-mid altitude, organic & siliceous	19	11-24		
LW-02 lowland, siliceous	14	7-25	0.6	0.43-0.76
LW-03 lowland, calcareous	28	10-69	1.1	0.73-1.37
LW-04 lowland, calcareous, v shallow	60	36-92	1.4	1.02-1.77
LW-05 lowland, organic & siliceous	18	10-24	0.5	0.37-0.64
LW-06 lowland, organic & calcareous	26	23-30	0.8	0.62-0.97
LW-07 mid-altitude, siliceous	15	12-23	0.4	0.14-2.22
LW-08 mid-altitude, calcareous	23	18-29		
LW-13 Mediterranean, siliceous	29	19-41		
LW-14 Mediterranean, calcareous	23	1-61		

7 River total phosphorus boundaries

7.1 Data

To investigate the potential range of river TP boundary values for European rivers data from the EEA were collated. Normalised biology EQR values were linked to summary TP data (annual means matched to biological data collected in the same year) from the nearest water quality site using spatial coordinates and by reference to their WFD water bodies. To determine TP boundary values binary logistic models were fitted using the status of phytobenthos, determined by categorising EQR values into two binary categories (NEQR ≥ 0.6 good or better = 1, NEQR < 0.6 moderate or worse = 0). Only EQR marked as sensitive to eutrophication or general degradation were used. Phytobenthos was used as it was the most frequently reported biological metric and has been shown to have the best relationship to TP. (A document / section is in preparation on BRT (Boosted Regression Trees) models and SEM (structural equation modeling) analysis regarding this). An additional national data set was made available by NO and this was incorporated into the analysis. All sites were related to the EU broad typology (Lyche Solheim, 2019) using their spatial locations and a GIS map layer to facilitate a type specific analysis.

The data set contained records for 18 countries although not all had TP data available. The distribution of TP data in the two biological status categories is shown in Figure 7.1, note the values are scaled to proportion, thus the size of the two peaks does not reflect the relative numbers of records in each class (prevalence). They provide an overview of the data, the range and the degree of overlap of TP concentrations and show clear differences between countries. This reflects differences in both typology and pressure and highlights the need to determine type specific boundaries. However, it also reflects conclusions from previous analysis that demonstrated that country was a more significant explanatory variable than river type. Additionally, it is clear that there are only marginal differences in the TP distributions for some countries, suggesting that in these countries the phytobenthos metric is not sufficiently responsive to TP. An example is PL, but similar poor separation is evident for EE, IT and LT.

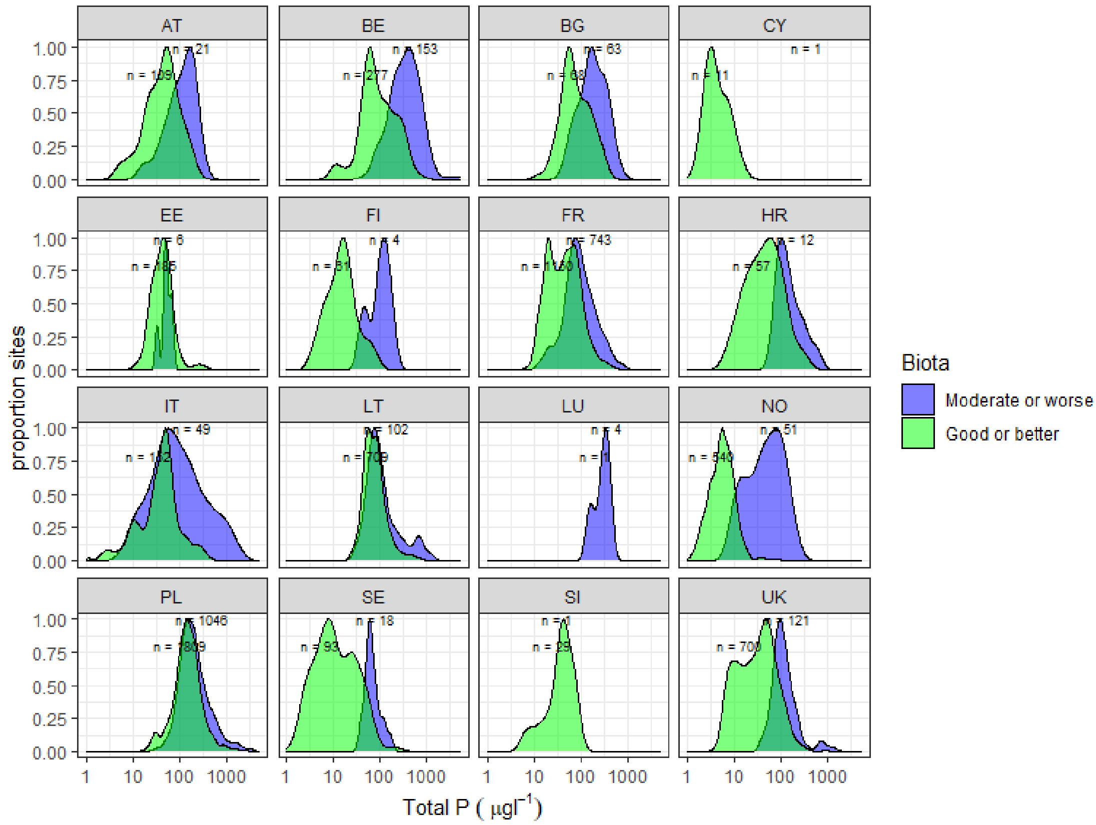

Figure 7.1. Density distribution and box plots showing the range of total phosphorus concentration in sites classified biologically into good or better and moderate or poor status for all rivers by country (data set SoE).

Figure 7.2 provides a similar plot but split by the broad typology for types with sufficient data in both biotic class. It illustrates that overlap is high in many types and thus potential difficulties fitting significant models.

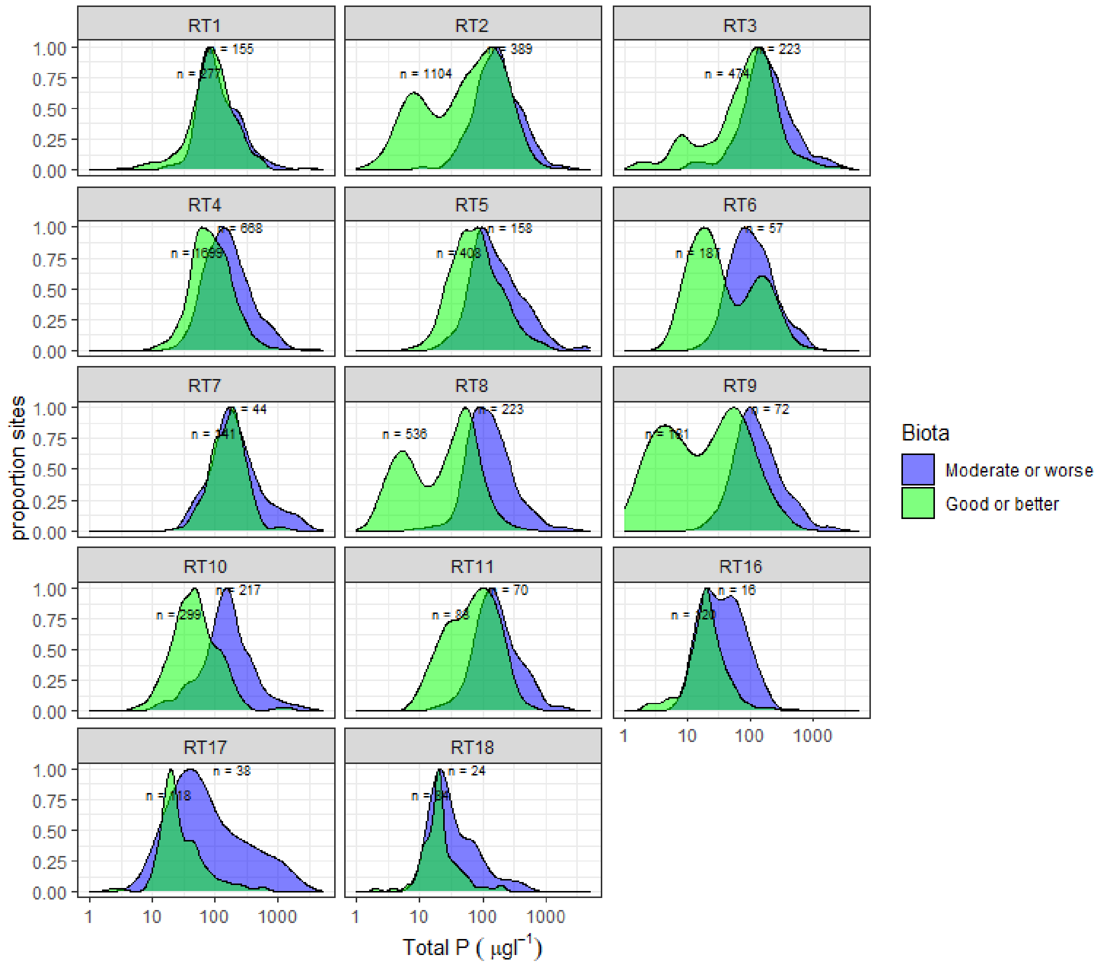

Figure 7.2. Density distribution and box plots showing the range of total phosphorus concentration in sites classified biologically into good or better and moderate or poor status for all rivers by broad type. (data set SoE).

7.2 Approach to modelling

The initial examination of the data showed that for some countries there were only marginal differences between the TP distributions in the good and not good biotic classes. Thus, BLM models were fitted to the country specific data sets for all countries with sufficient data ($N \geq 10$ for each binary biological category) to identify countries where model fit was inadequate (AUC < 0.7, pseudo r^2 <0.15). These countries were excluded from subsequent model runs. The data were then split into groups based on the EU broad typology by altitude (lowland, mid-altitude, highland) and geology (calcareous, siliceous). Within each of these groups the data were tabulated by type and the distributions of TP examined using box plots to assess relative country and type specific differences and to determine whether there were sufficient data for national or regional models. Where data were sufficient, BLM models that included categorical variables for type and region were then fitted to identify whether types could be combined (e.g., small and large rivers) and whether regional differences could be identified. A final set of models were then run on those types/regions with sufficient data, including for some types a grouping referred to as *not Northern* (notN) comprising all countries not in the *Northern* region. From the output of each model a confusion matrix was created, the key comparative measures were calculated and the optimum probability threshold identified and used to determine the predicted boundary and confidence limit values.

7.3 Model results

Full details for each river type are available in a separate document (html file), as only summary details are provided here.

7.3.1 Model performance

The AUC values for all of the models fitted to data split by country and/or river type are shown in Figure 7.3 There were sufficient data to run national models for 11 countries (AT, BE, BG, FR, HR, IT, LT, NO, PL, SE, UK). Of these, LT, PL had poor model performance. For the national type specific data sets models fitted to data from 6 countries (AT, BE, HR, NO, SE, UK) all had adequate performance. Those from 4 countries (BG, FR, IT, PL) were more variable, while no models from LT were adequate. Thus, data from LT was excluded from all models and those from PL for all models except for the mid-altitude rivers where the PL data performed well.

Using data from all countries models could be fitted to the majority of river types, but only a minority of types could be run by region. For example, in the northern region there were insufficient impacted records, and for Alpine and Eastern regions too few records or only records from a single country and thus not representative. Where all available data (excluding countries considered to have inadequate separation of TP) were used all but one model (very large rivers) met the performance criteria of an AUC > 0.7. Similarly, the majority of the regional data sets performed adequately.

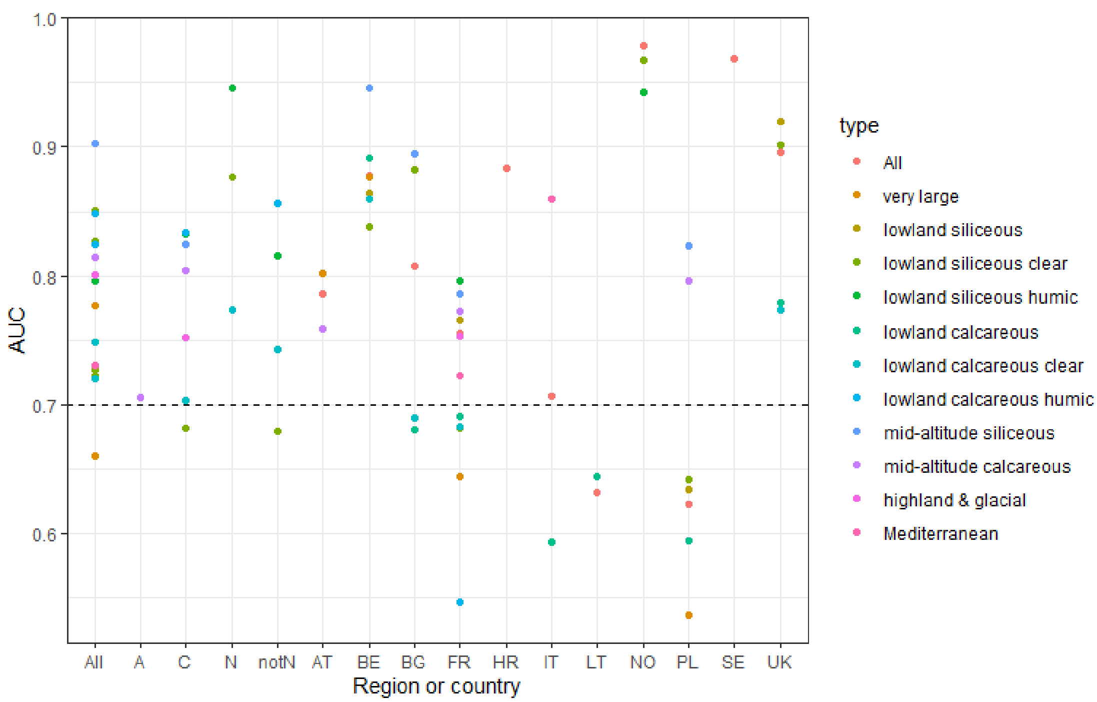

Figure 7.3. AUC values for all river models for different subsets of river data.

7.3.2 Selecting appropriate threshold probability values.

Table 7.1a (at the end of this document) shows the river TP boundary values derived from the BLM models applied to data from all countries, together with details of the mis-classification rates (omission, commission) and model fit parameters. For each river type four potential boundary values are shown, each derived using different threshold probability values (for maximum kappa, omission = commission,commission = 0.2 and commission = 0.1). In all cases selecting the probability threshold that maximised correct classifications (maximum kappa) generated the lowest probability threshold and thus highest boundary value (Figure 7.4ab, Table 7.1a). However, apart from the mid-altitude calcareous rivers, using this threshold generated much higher commission (false +ve) than omission (false -ve) rates (Figure 7.4c). Thus, a boundary predicted using this probability threshold could result in relatively large number of sites being incorrectly classed as good according to TP and thus be less suitable as an indicative pan-European boundary. Thus, a higher probability threshold was selected using the protocol outlined in the introduction and methods document (see folder for this document). This resulted in lower TP boundary concentrations with commission rates ranging from 0.1-0.37 and with one exception kappa values > the suggested threshold of 0.21 (Table 7.1b).

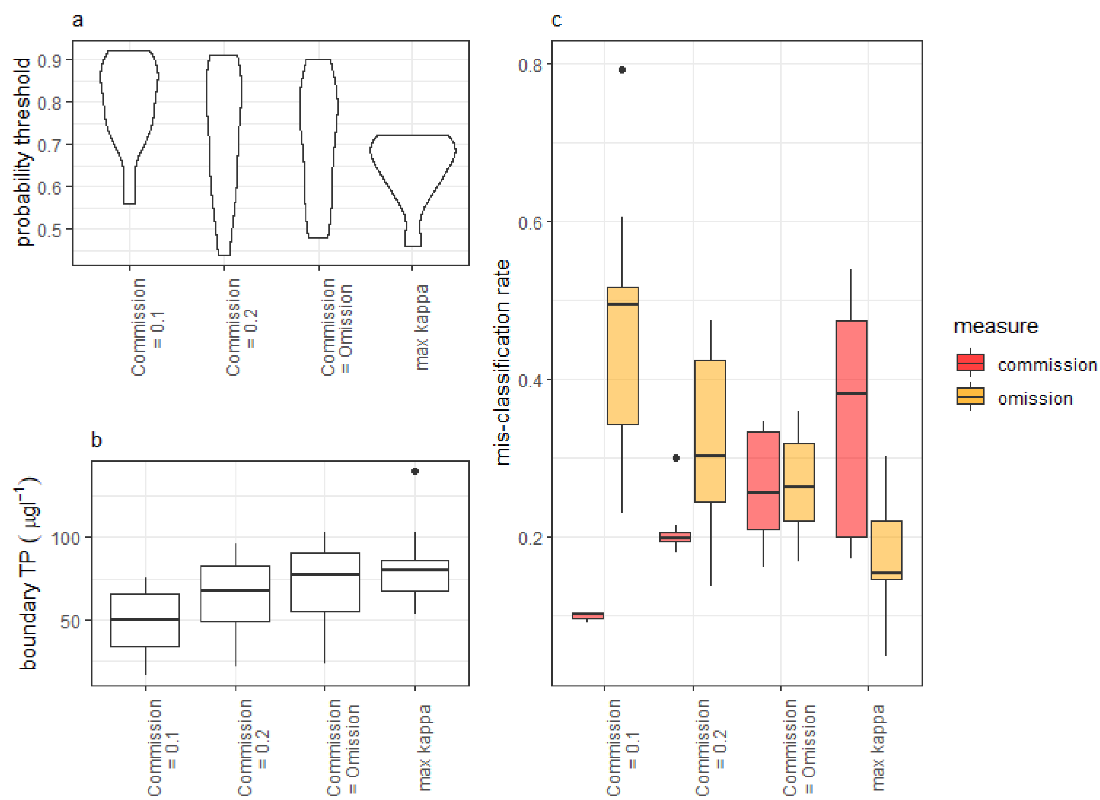

Figure 7.4. Effect of different threshold probability measures on a) the probability threshold, b) predicted TP boundary values, c) mis-classification rates (commission false +ve, omission false -ve).

7.4 *Boundary values*

7.4.1 Pan-European values

The modeled boundary determined using the selected probability thresholds using data from all available countries are shown in Figure 7.5 and in Table 7.1 (further details of mis-classification rates for these models are shown in tables at the end of this chapter). The highest boundary values were found for the very large (69-110 ugL^{-1}) and calcareous rivers (63-73 ugL^{-1}) and the lowest were for highland and glacial rivers (19-40 ugL^{-1}). As expected, the lowland clear siliceous rivers had lower values (40-57 ugL^{-1}) than the calcareous rivers. The siliceous and calcareous humic rivers had higher boundaries (siliceous humic: 68-89 ugL^{-1}, calcareous humic: 75-120 ugL^{-1}) than the clear ones, perhaps due to lower availability of phosphorus. However, there was no significant difference between clear and humic types for siliceous or calcareous mid-altitude rivers. The majority of the predicted boundaries fall within the interquartile ranges of the reported boundary values (Kelly, 2022) (grey shading Figure 7.5), the exceptions being for lowland calcareous humic, highland/glacial and Mediterranean river types. For the latter two types predicted values were lower, but within the range reported by (Nikolaidis, 2021) (yellow shading Figure 7.5), an estimate based on modelling these reported values.

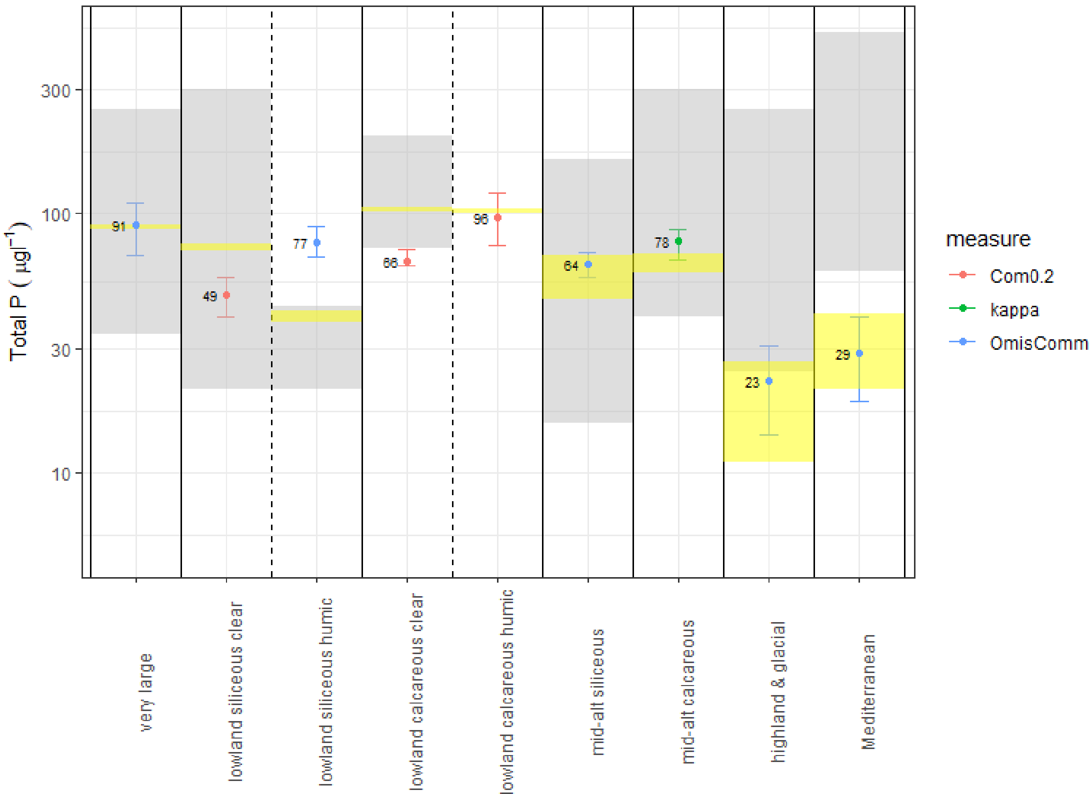

Figure 7.5. Predicted total phosphorus boundary values for rivers by broad type using models fitted to data from all available countries, compared to upper and lower quartiles of MS boundary values (grey shading) and estimated pan-European boundary values reported by (Nikolaidis, 2021) (yellow shading). (Vertical lines show confidence limits).

Table 7.1. Predicted type specific pan-European TP boundary values using selected critical p threshold values using binary logistic models fitted to data from all countries.

river type	boundary	lcl	ucl	p thresh.	measure	AUC	pseudo r2	countries
very large	91	69	110	0.82	OmisComm	0.78	0.27	10
lowland siliceous clear	49	40	57	0.82	Com0.2	0.83	0.31	11
lowland siliceous humic	77	68	89	0.48	OmisComm	0.90	0.60	9
lowland calcareous clear	66	63	73	0.76	Com0.2	0.75	0.24	10
lowland calcareous humic	96	75	120	0.61	Com0.2	0.85	0.47	7
mid-altitude siliceous & organic	64	57	71	0.68	OmisComm	0.90	0.55	12
mid-altitude calcareous & organic	78	66	87	0.61	kappa	0.81	0.36	10
highland & glacial	23	14	31	0.90	OmisComm	0.80	0.26	7
Mediterranean	29	19	40	0.81	OmisComm	0.73	0.21	5

7.4.2 Regional values

There were insufficient data to fit models to regional sub-sets of data for all river types. Predicted boundary values from models that could be fitted are shown in Figure 7.6 and Table 7.2 (at end document) In most cases boundaries from countries in the central region of Europe were higher than those predicted from all countries data, while they were lower for the northern region. Although the extent of the difference depended on the proportions of countries from each region that were included in the *all country* data set. The biggest difference was for lowland siliceous rivers, where the northern region boundary values were much lower (25-41 ugL^{-1}) than those from other regions (52-102 ugL^{-1}). This difference may reflect the relatively high proportion (75%) of norther region records from NO for this river type. The distribution of TP in sites classified by phytobenthos clearly differed between countries and where there were sufficient data models fitted to individual countries data revealed considerable variation, with higher values for BE, BG and PL and low values for NO (Figure 7.7). All these country models had adequate fits (AUC >0.7) and it is clear that there are significant country specific differences. This is investigated in more detail in the following sections.

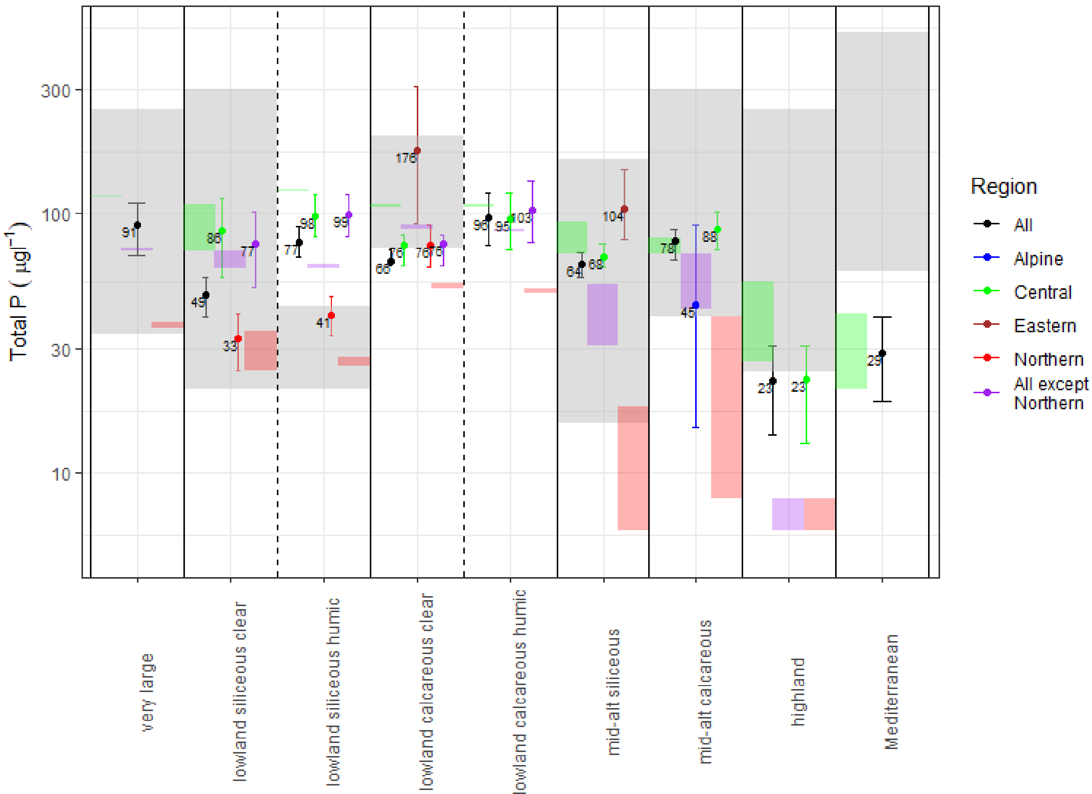

Figure 7.6. Predicted total phosphorus boundary values for rivers by broad type using models fitted to data from all available countries and for different regions, compared to upper and lower quartiles of MS boundary values (grey shading) and estimated regional boundary values reported by (Nikolaidis, 2021) (central=green, Baltic=purple, Scandinavian=red shading). (lines show confidence limits).

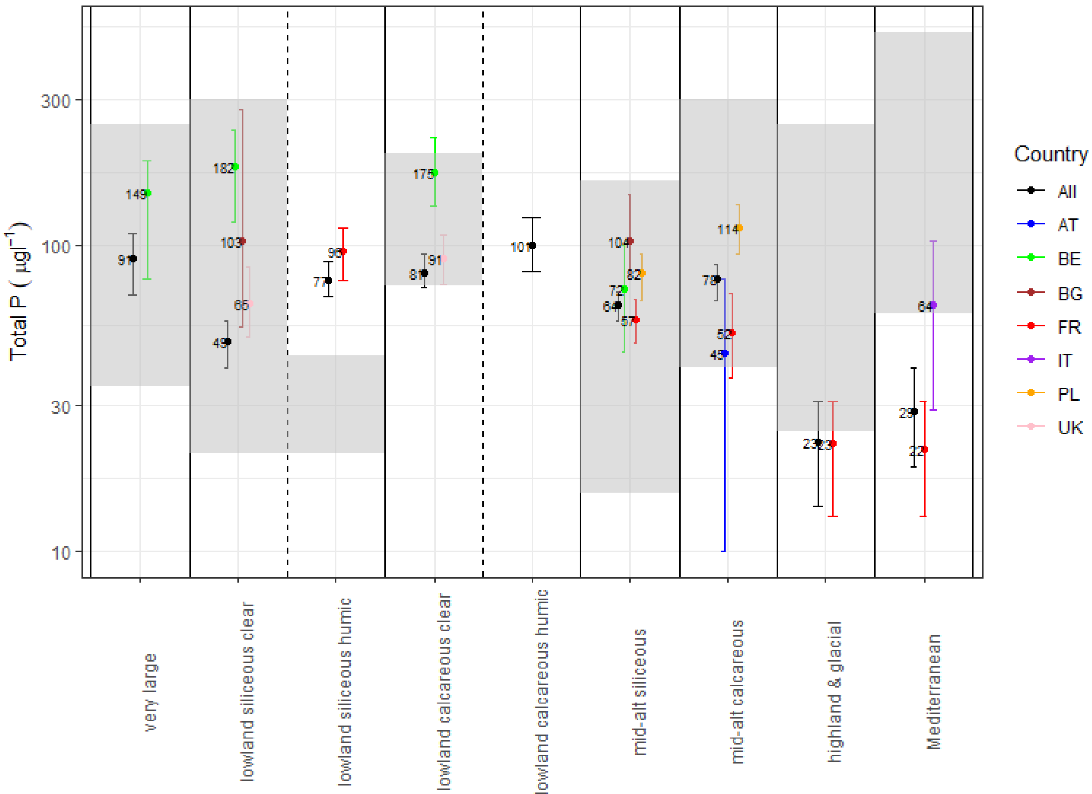

Figure 7.7. Predicted total phosphorus boundary values for rivers by broad type using models fitted to data from all available countries and for individual countries. (Vertical lines show confidence limits).

7.4.2.1 Very large rivers

There were relatively few records from the very large rivers and for this river type the distributions of TP for phytobenthos classified as good and not good overlapped substantially (Figure 7.2). The TP distributions by type (Figure 7.8) provide some indication that TP in sites classified as good were lower in the Northern region, but there were insufficient data to fit regional models. The range of the predicted boundaries taking all the data (69 – 110 ugL^{-1}, pink shading Figure 7.8) is within the interquartile range of reported MS boundary values (grey shading Figure 7.8) and might represent a reasonable estimate of an appropriate boundary range for this river type. However, the lower range of TP in good sites from the Northern region suggest that lower values might be more appropriate in this region. The Scandinavian regional value estimated by (Nikolaidis, 2021) (37 ugL^{-1} red horizontal line Figure 7.8) is close to the upper tail of the data from FI and provides an indication of a potential value.

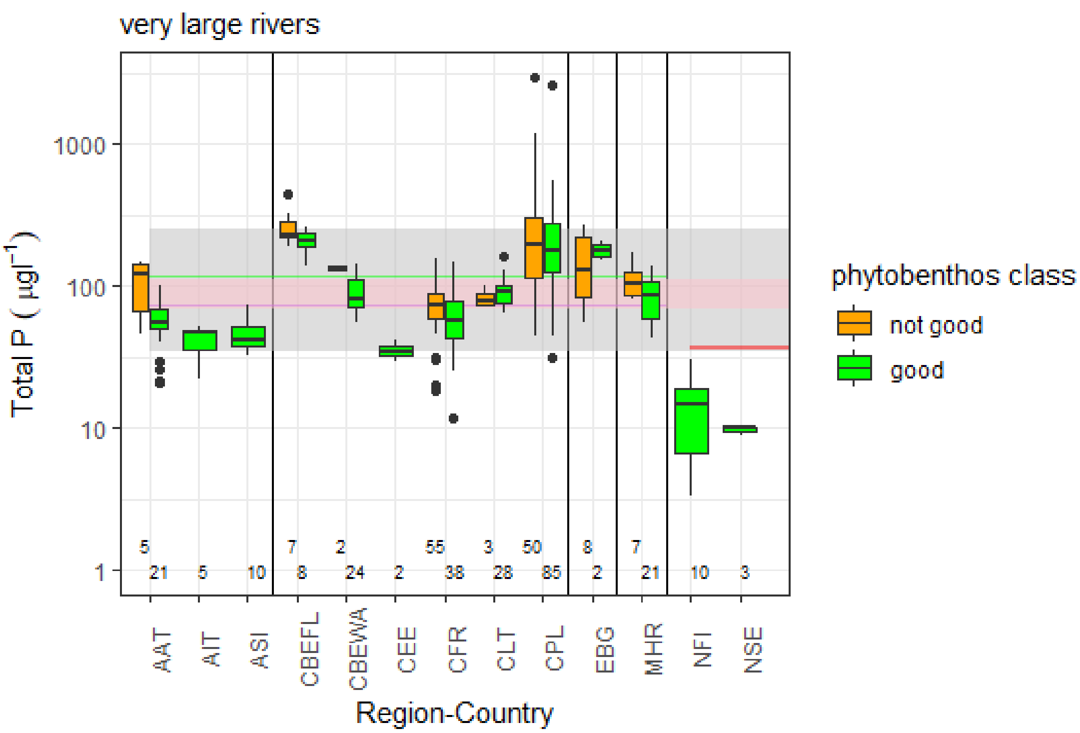

Figure 7.8. Box plots showing range TP values for sites in very large rivers, classified as good or not good using phytobenthos by region and country compared to boundary ranges; grey shading shows interquartile range of MS reported values; coloured shading shows range predicted using BLM fitted to data from all (pink) countries.

7.4.2.2 Lowland siliceous rivers

There was no significant difference between large and small lowland siliceous river models and the distributions of TP by country for all the lowland siliceous clear water rivers suggest lower TP ranges in the Northern region. This was confirmed by significantly different models for the northern region in comparison to countries from all other regions. The range of boundary values predicted using data from all regions other than northern was 52 - 102 ugL^{-1}, (blue shading Figure 7.9a), in comparison to 25 - 41 ugL^{-1}, (red shading Figure 7.9a). Both these values lie within the reported range of MS boundary values (grey shading Figure 7.9a) and are considered reasonable estimates of typical boundary values.

As for the lowland siliceous clear rivers, the TP distributions suggest lower TP in the northern region and this was confirmed by modelling. The range of predicted boundary values for the northern regions were (34 - 48 ugL^{-1} (red shading Figure 7.9b), while those predicted using data from all the other regions was 82 - 118 ugL^{-1} (blue shading Figure 7.9b). These and the pan-European average value (68 - 89 ugL^{-1} pink Figure 7.9b shading) were higher than the interquartile range of reported boundary values (grey shading). However, only 4 countries (FI, PL, NO, SE) reported boundary values, and thus their interquartile range will reflect the northern region rather than a pan-European average.

Thus, for lowland siliceous rivers the available evidence suggests lower TP boundary values for northern rivers and that humic rivers have slightly higher values than clear.

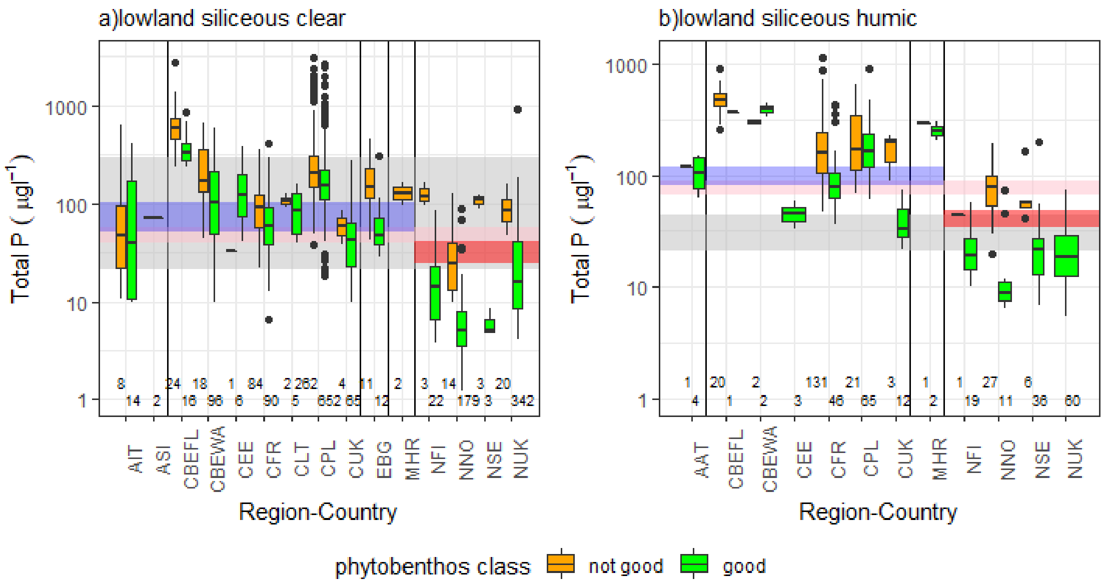

Figure 7.9. Box plots showing range TP values for sites in lowland siliceous a) clear & b) humic rivers, classified as good or not good using phytobenthos by region and country compared to boundary ranges (grey shading shows interquartile range of MS reported values; coloured shading shows range predicted using BLM fitted to data from all (pink), all except northern (blue), northern (red) countries.

7.4.2.3 Lowland calcarerous rivers

As for siliceous rivers there was no significant effect of river size, so the broad type size categories were combined. For the clear water rivers there were sufficient data to model the northern region separately, but unlike the siliceous rivers there was no significant difference in the predicted boundary values. For the northern region the TP boundary range was 62 - 90 ugL^{-1} (Figure 7.10a, red shading), while for the other countries it was 63 - 73 ugL^{-1} (Figure 7.10a, blue shading). Both values were at the lower end of the interquartile range of the reported boundary values (Figure 7.10a, grey shading).

As for the siliceous rivers the calcareous humic rivers had higher predicted TP boundary values, but like the clear rivers no difference between regions. The range of boundary values for the norther region were (77 - 133 ugL^{-1} (Figure 7.10b, red shading) and for the other countries (75 - 120 ugL^{-1} (Figure 7.10b, blue shading).

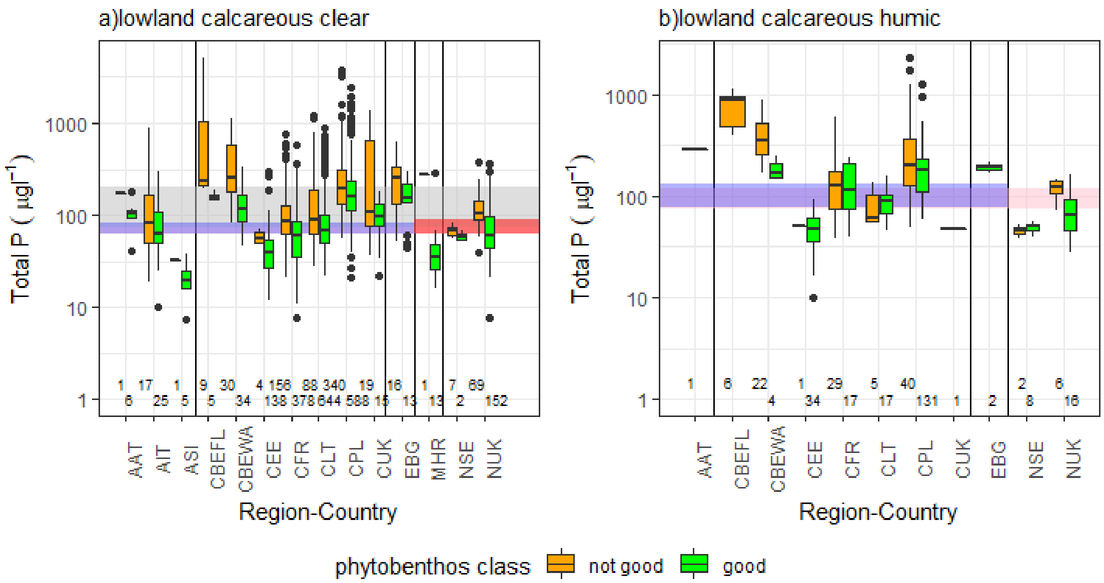

Figure 7.10. Box plots showing range TP values for sites in lowland calcareous a) clear & b) humic rivers, classified as good or not good using phytobenthos by region and country compared to boundary ranges (grey shading shows interquartile range of MS reported values; coloured shading shows range predicted using BLM fitted to data from all (pink), all except northern (blue), northern (red) countries.

7.4.2.4 Mid-altitude rivers

For mid-altitude rivers neither size or humic content were significant predictors in the models and models were fitted to all siliceous and then all calcareous rivers. For both types there were too few impacted sites to fit models to data from the northern region, although the distribution of TP in the good phytobenthos siliceous sites was clearly lower than in those from the other regions (Figure 7.11a). The predicted boundary ranges for sites in the central region were similar for both siliceous and calcareous rivers siliceous: 62 - 76 ugL^{-1} (Figure 7.11a, blue shading). calcareous: 73 - 101 ugL^{-1} (Figure 7.11a, blue shading). There were sufficient data to model the alpine regions separately for the calcareous sites, although the relatively small data set generated a relatively large range, which was similar to the central values: 15 - 90 ugL^{-1} (Figure 7.11a, blue shading). For the central regions (Nikolaidis, 2021) estimated a boundary range of 70-93 like the values derived from the models. Their estimated boundary values for the northern regions were lower 6-18 and might represent a more appropriate value.

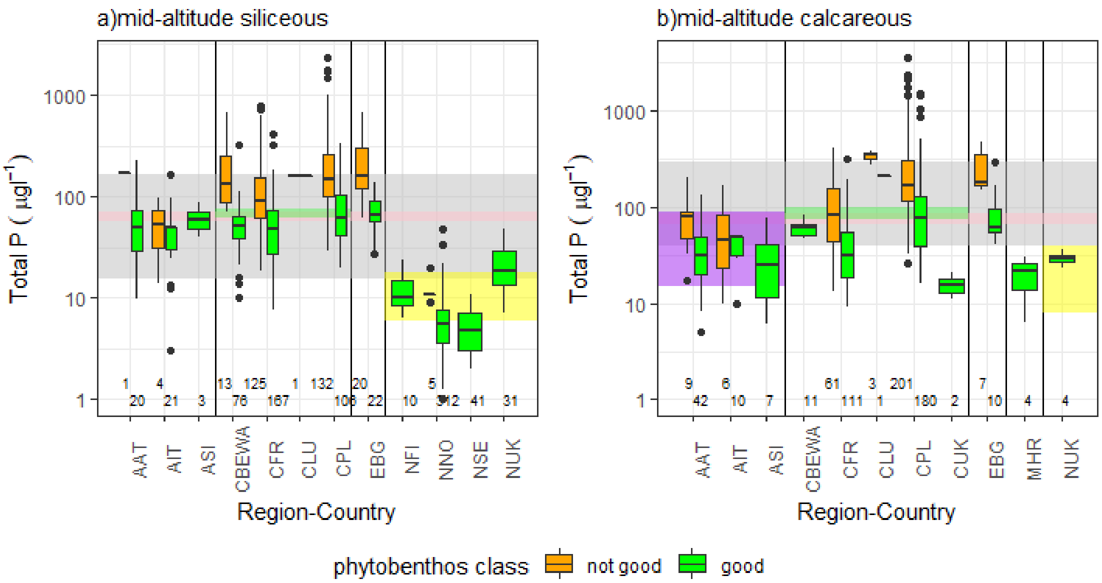

Figure 7.11. Box plots showing range TP values for sites in mid-altitude clear & humic a) siliceous & b) calcareous rivers, classified as good or not good using phytobenthos by region and country compared to boundary ranges (grey shading shows interquartile range of MS reported values; coloured shading shows range predicted using BLM fitted to data from all (pink), central (green), alpine (purple) countries and yellow shading values for the Scandinavian region estimated by (Nikolaidis, 2021).

7.4.2.5 Highland/glacial and Mediterranean rivers

There were too few data to allow for regional models for these river types. For the highland/glacial type the model using all available data was dominated by sites from FR, the predicted boundary was lower than the interquartile range of the reported MS boundaries (Figure 7.12a grey shading) but is potentially representative of sites other than from the northern region siliceous: 14 - 31 ugL^{-1} (Figure 7.12a, pink shading). Sites from NO, the only country with data in the northern region had much lower TP distributions in the good status sites, but the absence of impacted sites prevented modelling. The boundary values reported by (Nikolaidis, 2021) northern (6-8 ugL^{-1}) could be an appropriate estimate, although their values for central were higher than the modeled values (27-55 ugL^{-1})

For the Mediterranean rivers, the predicted boundary range 19 - 40 ugL^{-1} (Figure 7.12b, pink shading) is lower than the interquartile range of the reported MS boundaries (Figure 7.12b grey shading). However, it is similar to the values reported by (Nikolaidis, 2021) (21-41 ugL^{-1}).

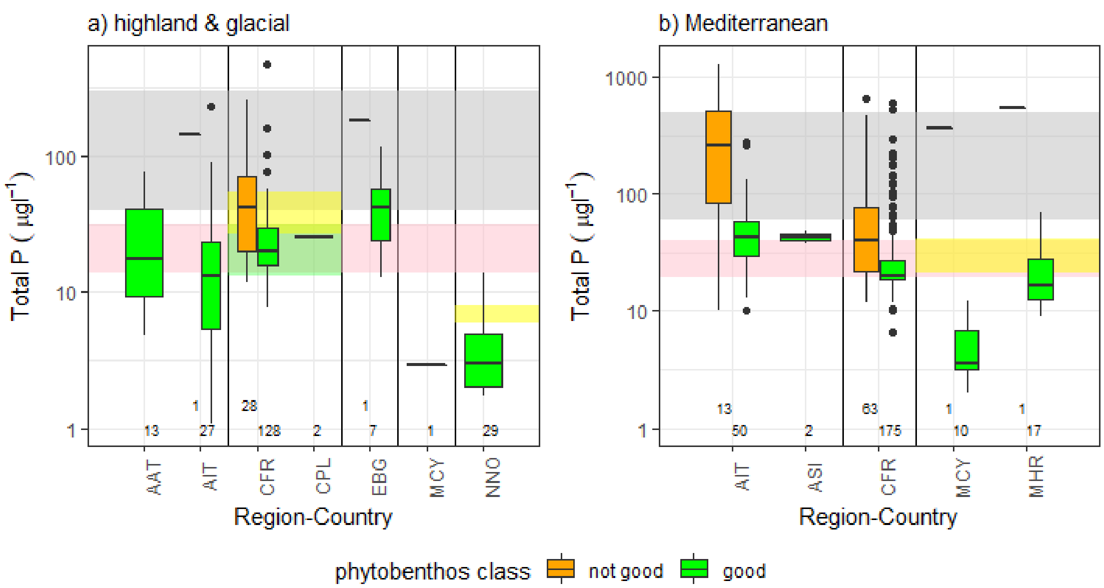

Figure 7.12. Box plots showing range TP values for sites in a) highland/glacial & b) Mediterranean rivers classified as good or not good using phytobenthos by region and country compared to boundary ranges; grey shading shows interquartile range of MS reported values; coloured shading shows range predicted using BLM fitted to data from all (pink) countries and yellow shading values for the Scandinavian region estimated by (Nikolaidis, 2021). (note scale difference between a and b).

7.5 Conclusions

- When using statistical models to establish supporting element boundary values consideration should be given to the overall level of mis-classification and whether it is appropriate to minimise the false positive rate in order to provide for sufficiently precautionary boundaries.
- Using the currently available data it is very challenging to establish objective total phosphorus boundary values. Firstly, substantial country specific differences were found. Given that these differences resulted from models with adequate fits and that they were based on relationships with the status of an inter-calibrated metric (phytobenthos) these differences remain unexplained.
- However, by using data from all available countries (where modelling demonstrated an adequate relationship with pressure) with type specific models boundary values for any river type are most likely to be less than 100 ugL^{-1}.
- There were insufficient data to clarify regional differences, but the indications are that boundaries for Northern regions are likely to be lower than those in other areas of Europe.
- Further work is needed to determine what factors might explain country specific differences.
- A set of our best estimate of type specific boundary ranges derived where possible from the modelling described in this report, supplemented by those reported by (Nikolaidis, 2021) is given in Table 7.2. The values are compared to those reported by MS and summarised in (Kelly, 2022).

Table 7.2. Best available estimate of type specific TP boundary values for rivers, derived where possible from binary logistic models (BLM) reported here, or where insufficient data from GAM models fitted to reported MS boundary values (Nikolaidis, 2021).

river type	boundary	source	boundary northern	source	range reported
very large	69-110	BLM	36-38	(Nikolaidis, 2021)	34-250
lowland siliceous clear	52-102	BLM	25-41	BLM	21-300
lowland siliceous humic	82-118	BLM	34-48	BLM	21-44
lowland calcareous clear	63-83	BLM	62-90	BLM	74-200
lowland calcareous humic	77-133	BLM	49-51	(Nikolaidis, 2021)	
mid-altitude siliceous & organic	62-76	BLM	6-18	(Nikolaidis, 2021)	16-162
mid-altitude calcareous & organic	73-101	BLM	8-40	(Nikolaidis, 2021)	40-300
highland & glacial	13-31	BLM	6-8	(Nikolaidis, 2021)	24-250
Mediterranean	19-40	BLM			60-500

7.6 Tables of results

Table 7.3. Predicted type specific pan-European TP boundary values and with key measures from confusion matrix, using binary logistic models fitted to data from all countries.

river type	boundary	lcl	ucl	measure	p thresh.	commission	omission	prev.	kappa	AUC	pseudo r2	countries
very large	91	69	110	OmisComm	0.82	0.33	0.32	0.80	0.26	0.78	0.27	10
lowland siliceous clear	49	40	57	Com0.2	0.82	0.19	0.30	0.81	0.36	0.83	0.31	11
lowland siliceous humic	77	68	89	OmisComm	0.48	0.16	0.17	0.51	0.67	0.90	0.60	9
lowland calcareous clear	66	63	73	Com0.2	0.76	0.22	0.42	0.70	0.29	0.75	0.24	10
lowland calcareous humic	96	75	120	Com0.2	0.61	0.19	0.24	0.55	0.56	0.85	0.47	7
mid-altitude siliceous & organic	64	57	71	OmisComm	0.68	0.18	0.18	0.73	0.59	0.90	0.55	12
mid-altitude calcareous & organic	78	66	87	kappa	0.61	0.18	0.30	0.57	0.51	0.81	0.36	10
highland & glacial	23	14	31	OmisComm	0.90	0.33	0.31	0.87	0.20	0.80	0.26	7
Mediterranean	29	19	40	OmisComm	0.81	0.35	0.36	0.77	0.23	0.73	0.21	5

Table 7.4. Predicted type specific pan-European TP boundary values using all potential critical p threshold values and key measures from confusion matrix, using binary logistic models fitted to data from all countries.

Select	typology_names_of	Bound	lcl	ucl	measure	threshold	commission	omission	Prev	kappa
0	very large	140	116	184	kappa	0.65	0.538	0.084	0.80	0.41
1	very large	91	69	110	OmisComm	0.82	0.333	0.318	0.80	0.26
0	very large	79	57	96	Com0.2	0.86	0.179	0.474	0.80	0.21
0	very large	75	53	93	Com0.1	0.87	0.103	0.494	0.80	0.24
0	lowland siliceous clear	85	74	102	kappa	0.72	0.381	0.153	0.81	0.42
0	lowland siliceous clear	55	46	69	OmisComm	0.80	0.249	0.263	0.81	0.36
1	lowland siliceous clear	49	40	57	Com0.2	0.82	0.188	0.302	0.81	0.36
0	lowland siliceous clear	33	24	40	Com0.1	0.87	0.091	0.419	0.81	0.29
0	lowland siliceous humic	84	73	98	Com0.2	0.44	0.198	0.138	0.51	0.66
0	lowland siliceous humic	80	71	93	kappa	0.46	0.172	0.148	0.51	0.68
1	lowland siliceous humic	77	68	89	OmisComm	0.48	0.161	0.168	0.51	0.67
0	lowland siliceous humic	66	57	75	Com0.1	0.56	0.104	0.230	0.51	0.67
0	lowland calcareous clear	86	73	94	kappa	0.69	0.388	0.271	0.70	0.32
0	lowland calcareous clear	77	73	83	OmisComm	0.72	0.321	0.340	0.70	0.30
1	lowland calcareous clear	66	63	73	Com0.2	0.76	0.215	0.424	0.70	0.29
0	lowland calcareous clear	50	42	52	Com0.1	0.82	0.103	0.606	0.70	0.21
0	lowland calcareous humic	103	82	129	kappa	0.58	0.209	0.220	0.55	0.57
0	lowland calcareous humic	103	82	129	OmisComm	0.58	0.209	0.220	0.55	0.57
1	lowland calcareous humic	96	75	120	Com0.2	0.61	0.194	0.244	0.55	0.56

Select	typology_names_of	Bound	lcl	ucl	measure	threshold	commission	omission	Prev	kappa
0	lowland calcareous humic	70	50	88	Com0.1	0.74	0.104	0.341	0.55	0.54
0	mid-altitude siliceous & organic	67	62	76	kappa	0.66	0.199	0.156	0.73	0.60
0	mid-altitude siliceous & organic	67	62	76	Com0.2	0.66	0.199	0.156	0.73	0.60
1	mid-altitude siliceous & organic	64	57	71	OmisComm	0.68	0.176	0.175	0.73	0.59
0	mid-altitude siliceous & organic	55	48	62	Com0.1	0.74	0.096	0.243	0.73	0.56
0	mid-altitude calcareous & organic	93	80	109	OmisComm	0.55	0.254	0.246	0.57	0.50
0	mid-altitude calcareous & organic	83	73	94	Com0.2	0.59	0.195	0.291	0.57	0.50
1	mid-altitude calcareous & organic	78	66	87	kappa	0.61	0.178	0.301	0.57	0.51
0	mid-altitude calcareous & organic	49	37	58	Com0.1	0.75	0.101	0.516	0.57	0.36
0	highland & glacial	60	44	100	kappa	0.71	0.533	0.048	0.87	0.46
1	highland & glacial	23	14	31	OmisComm	0.90	0.333	0.309	0.87	0.20
0	highland & glacial	21	12	29	Com0.2	0.91	0.300	0.353	0.87	0.18
0	highland & glacial	19	10	27	Com0.1	0.92	0.100	0.517	0.87	0.15
0	Mediterranean	53	40	76	kappa	0.71	0.474	0.146	0.77	0.38
1	Mediterranean	29	19	40	OmisComm	0.81	0.346	0.358	0.77	0.23
0	Mediterranean	21	14	29	Com0.2	0.85	0.205	0.429	0.77	0.26
0	Mediterranean	16	9	22	Com0.1	0.88	0.090	0.791	0.77	0.06

Table 7.5. Predicted type specific TP boundary values, with with key measures from confusion matrix, for all available regional data sets in comparison to all countries (pan-European).

region	river type	boundary	lcl	ucl	commission	omission	prev.	kappa	AUC	pseudo r2	countries	measure	p thresh.
All	very large	91	69	110	0.33	0.32	0.80	0.26	0.78	0.27	10	OmisComm	0.82
All	lowland siliceous clear	49	40	57	0.19	0.30	0.81	0.36	0.83	0.31	11	Com0.2	0.82
C	lowland siliceous clear	86	57	114	0.32	0.40	0.68	0.25	0.68	0.13	4	kappa	0.70
N	lowland siliceous clear	33	25	41	0.24	0.22	0.92	0.26	0.88	0.32	4	OmisComm	0.90
notN	lowland siliceous clear	77	52	102	0.30	0.41	0.66	0.26	0.68	0.13	8	kappa	0.70
All	lowland siliceous humic	77	68	89	0.16	0.17	0.51	0.67	0.90	0.60	9	OmisComm	0.48
C	lowland siliceous humic	98	82	118	0.20	0.27	0.29	0.50	0.83	0.38	4	Com0.2	0.37
N	lowland siliceous humic	41	34	48	0.09	0.13	0.79	0.69	0.95	0.65	4	Com0.1	0.73
notN	lowland siliceous humic	99	82	118	0.20	0.30	0.31	0.48	0.82	0.35	6	Com0.2	0.39
All	lowland calcareous clear	66	63	73	0.22	0.42	0.70	0.29	0.75	0.24	10	Com0.2	0.76
C	lowland calcareous clear	76	63	83	0.35	0.39	0.80	0.18	0.70	0.16	5	OmisComm	0.82
E	lowland calcareous clear	176	92	308	0.31	0.31	0.45	0.38	0.69	0.08	1	OmisComm	0.44
N	lowland calcareous clear	76	62	90	0.20	0.34	0.67	0.41	0.77	0.24	2	Com0.2	0.71
notN	lowland calcareous clear	76	63	83	0.35	0.34	0.71	0.26	0.74	0.24	9	OmisComm	0.73
All	lowland calcareous humic	96	75	120	0.19	0.24	0.55	0.56	0.85	0.47	7	Com0.2	0.61
C	lowland calcareous humic	95	73	120	0.21	0.32	0.54	0.47	0.83	0.44	5	Com0.2	0.62
notN	lowland calcareous humic	103	77	133	0.19	0.26	0.50	0.56	0.86	0.48	6	kappa	0.57
All	mid-altitude siliceous & organic	64	57	71	0.18	0.18	0.73	0.59	0.90	0.55	12	OmisComm	0.68
C	mid-altitude siliceous & organic	68	62	76	0.21	0.31	0.56	0.48	0.82	0.39	4	Com0.2	0.63
E	mid-altitude siliceous & organic	104	80	147	0.20	0.18	0.52	0.62	0.89	0.59	1	OmisComm	0.50
A	mid-altitude calcareous & organic	45	15	90	0.33	0.37	0.80	0.21	0.71	0.15	3	OmisComm	0.79
All	mid-altitude calcareous & organic	78	66	87	0.18	0.30	0.57	0.51	0.81	0.36	10	kappa	0.61
C	mid-altitude calcareous & organic	88	73	101	0.20	0.31	0.54	0.49	0.80	0.34	5	Com0.2	0.56

region	river type	boundary	lcl	ucl	commission	omission	prev.	kappa	AUC	pseudo r2	countries	measure	p thresh.
All	highland & glacial	23	14	31	0.33	0.31	0.87	0.20	0.80	0.26	7	OmisComm	0.90
C	highland & glacial	23	13	31	0.36	0.35	0.82	0.19	0.75	0.22	2	OmisComm	0.87
All	Mediterranean	29	19	40	0.35	0.36	0.77	0.23	0.73	0.21	5	OmisComm	0.81

Table 7.6. Predicted type specific TP boundary values, with with key measures from confusion matrix, for selected national data sets in comparison to all countries (pan-European).

region	river type	boundary	lcl	ucl	commission	omission	prev.	kappa	AUC	pseudo r2	measure	p thresh.
All	very large	91	69	110	0.33	0.32	0.80	0.26	0.78	0.27	OmisComm	0.82
BE	very large	149	78	190	0.22	0.22	0.78	0.47	0.88	0.48	OmisComm	0.82
BE	lowland siliceous	286	232	351	0.20	0.22	0.64	0.56	0.86	0.45	Com0.2	0.57
BG	lowland siliceous	103	54	278	0.09	0.17	0.52	0.74	0.88	0.49	Com0.1	0.47
FR	lowland siliceous	80	68	93	0.22	0.37	0.39	0.42	0.77	0.26	kappa	0.44
UK	lowland siliceous	83	67	107	0.44	0.05	0.95	0.41	0.92	0.34	kappa	0.83
All	lowland siliceous clear	49	40	57	0.19	0.30	0.81	0.36	0.83	0.31	Com0.2	0.82
BE	lowland siliceous clear	182	119	238	0.21	0.38	0.73	0.32	0.84	0.38	Com0.2	0.79
BG	lowland siliceous clear	103	54	278	0.09	0.17	0.52	0.74	0.88	0.49	Com0.1	0.47
UK	lowland siliceous clear	65	50	85	0.29	0.11	0.94	0.35	0.90	0.29	kappa	0.89
All	lowland siliceous humic	77	68	89	0.16	0.17	0.51	0.67	0.90	0.60	OmisComm	0.48
FR	lowland siliceous humic	96	77	115	0.21	0.30	0.26	0.45	0.80	0.28	Com0.2	0.34
BE	lowland calcareous	191	165	227	0.18	0.19	0.39	0.62	0.89	0.54	Com0.2	0.42
UK	lowland calcareous	83	69	97	0.22	0.34	0.66	0.39	0.78	0.28	Com0.2	0.69
All	lowland calcareous clear	81	73	94	0.37	0.34	0.78	0.23	0.72	0.17	OmisComm	0.79
BE	lowland calcareous clear	175	135	227	0.10	0.18	0.50	0.72	0.86	0.48	Com0.1	0.53
UK	lowland calcareous clear	91	75	108	0.31	0.31	0.65	0.36	0.77	0.27	OmisComm	0.65
All	lowland calcareous humic	101	82	124	0.24	0.23	0.58	0.53	0.82	0.42	OmisComm	0.61
All	mid-altitude siliceous & organic	64	57	71	0.18	0.18	0.73	0.59	0.90	0.55	OmisComm	0.68
BE	mid-altitude siliceous & organic	72	45	100	0.08	0.14	0.85	0.59	0.95	0.47	Com0.1	0.88
BG	mid-altitude siliceous & organic	104	80	147	0.20	0.18	0.52	0.62	0.89	0.59	OmisComm	0.50

region	river type	boundary	lcl	ucl	commission	omission	prev.	kappa	AUC	pseudo r2	measure	p thresh.
FR	mid-altitude siliceous & organic	57	48	67	0.20	0.38	0.57	0.40	0.79	0.32	Com0.2	0.63
PL	mid-altitude siliceous & organic	82	66	94	0.17	0.33	0.45	0.51	0.82	0.41	kappa	0.55
All	mid-altitude calcareous & organic	78	66	87	0.18	0.30	0.57	0.51	0.81	0.36	kappa	0.61
AT	mid-altitude calcareous & organic	45	10	78	0.22	0.33	0.82	0.29	0.76	0.24	Com0.2	0.84
FR	mid-altitude calcareous & organic	52	37	70	0.28	0.28	0.65	0.42	0.77	0.28	OmisComm	0.65
PL	mid-altitude calcareous & organic	114	94	137	0.20	0.34	0.47	0.45	0.80	0.33	Com0.2	0.49
All	highland & glacial	23	14	31	0.33	0.31	0.87	0.20	0.80	0.26	OmisComm	0.90
FR	highland & glacial	23	13	31	0.36	0.34	0.82	0.20	0.75	0.22	OmisComm	0.87
All	Mediterranean	29	19	40	0.35	0.36	0.77	0.23	0.73	0.21	OmisComm	0.81
FR	Mediterranean	22	13	31	0.24	0.32	0.74	0.37	0.72	0.15	Com0.2	0.80
IT	Mediterranean	64	29	104	0.23	0.22	0.79	0.45	0.86	0.47	OmisComm	0.85

References

Fielding, A.H. and Bell, J.F. (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental Conservation 24(1), 38-49.

Kelly, M., Phillips, G., Teixeira, H., Salas Herrero, F., Várbíró, G., Kolada A., Lyche Solheim, A., Poikane, S. (2022). Physico-chemical supporting elements in inland waters under the Water Framework Directive: a review of national standards to support good ecological status. EUR 31040 EN, Publications Office of the European Union, Luxembourg, 2022. ISBN 978-92-76-51133-5, doi:10.2760/470539, JRC127875.

Kelly, M.G., Phillips, G., Teixeira, H., Varbiro, G., Herrero, F.S., Willby, N.J. and Poikane, S. (2021) Establishing ecologically-relevant nutrient thresholds: A tool-kit with guidance on its use. Sci Total Environ, 150977.

Landis, J.R. and Koch, G.G. (1977) The measurement of observer aggreement for categorical data. Biometrics 33(1), 159 - 174.

Lyche Solheim, A., Globevnik, L., Austnes, K., Kristensen, P., Moe, S. J., Persson, J., ... & Birk, S. (2019). A new broad typology for rivers and lakes in Europe: Development and application for large-scale environmental assessments. Science of the Total Environment 697: 134043. https://doi.org/10.1016/j.scitotenv.2019.134043

Nikolaidis, N.P., Phillips, G., Poikane, S., Varbiro, G., Bouraoui, F., Malago, A. and Lilli, M. (2021) River and lake nutrient targets that support ecological status: European scale gap analysis and strategies for the implementation of the Water Framework Directive. Sci Total Environ, 151898.

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

Poikane, S., Birk, S., Böhmer, J., Carvalho, L., de Hoyos, C., Gassner, H., Hellsten, S., Kelly, M., Lyche Solheim, A., Olin, M., Pall, K., Phillips, G., Portielje, R., Ritterbusch, D., Sandin, L., Schartau, A.-K., Solimini, A.G., van den Berg, M., Wolfram, G. and van de Bund, W. (2015) A hitchhiker's guide to European lake ecological assessment and intercalibration. Ecological Indicators 52(0), 533-544.

Ptacnik, R., Andersen, T. and 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.

List of abbreviations and definitions

%sat	Percent oxygen saturation”
AA	Annual Average
AUC	Area Under Curve
ANC	Acid Neutralising Capacity
BOD	Biochemical Oxygen Demand
Boundary	A threshold comprising a numeric value of a parameter (e.g. “1000 μS cm^{-1} conductivity”, or 0.7 probability).
BQE	Biological Quality Element
CCC	Criteria Continuous Concentration
CDOM	Coloured Dissolved Organic Matter
Chla	chlorophyll a
CIS	Common Implementation Strategy
CMC	Criteria Maximum Concentration
Criteria	Criteria comprise an appropriate parameter (e.g. “conductivity”), metric (e.g. “annual mean”) and threshold (e.g. “1000 μS cm-1”)
CW	Coastal Waters
DO	Dissolved Oxygen
DOC	Dissolved Organic Carbon
EA	East Atlantic
ECOSTAT	A working group dedicated to the ecological status of surface water bodies within implementation of the Water Framework Directive that was set up in November 2002.
EEA	European Environment Agency
EQR	Ecological Quality Ratio
EU	European Union
FNU	Formazin Nephelometric Units
G20	An intergovernmental forum comprising 19 sovereign countries, the European Union and the African Union
GES	Good Ecological Status
GIG	Geographical Intercalibration Group
G/M	Good status/Moderate status boundary
Helcom	Helsinki Commission (Baltic marine environment protection commission)
H/G	High status/Good status boundary
IC	Intercalibration
lcl	Lower Confidence Interval
ICP	Waters International Cooperative Programme on assessment and monitoring of the effects of air pollution on rivers and lakes.
MAC	Maximum Allowable Concentration
M/P	Moderate status/Poor status boundary
MSFD	Marine Strategy Framework Directive
NAO	North Atlantic Oscillation

NEA	North-East Atlantic
NTU	Nephelometric Turbidity Units
nEQR	Normalised EQR
OSPAR	Oslo-Paris Convention (Convention for the Protection of the Marine Environment of the North East Atlantic)
P/B	Poor status/Bad status boundary
PAR	Photosynthetically Available Radiation
PSU	Practical Salinity Units
SA	Seasonal Average
ucl	upper confidence interval
SE	Supporting Elements
SoE	State of Environment
SS	Suspended Solids
SSD	Species Sensitivity Distributions
TDS	Total Dissolved Solids
Threshold	A numeric value representing a boundary of a parameter (e.g. "1000 µS cm^{-1} conductivity", or 0.7 probability).
TN	Total Nitrogen
TP	Total Phosphorus
TRAC	Transitional And Coastal Waters
WFD	Water Framework Directive
WHO	World Health Organisation
WISE	Water Information System for Europe (WFD database)

List of figures

List of tables

GETTING IN TOUCH WITH THE EU

In person

All over the European Union there are hundreds of Europe Direct centres. You can find the address of the centre nearest you online (european-union.europa.eu/contact-eu/meet-us_en).

On the phone or in writing

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

— by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),

— at the following standard number: +32 22999696,

— via the following form: european-union.europa.eu/contact-eu/write-us_en.

FINDING INFORMATION ABOUT THE EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website (european-union.europa.eu).

EU publications

You can view or order EU publications at op.europa.eu/en/publications. Multiple copies of free publications can be obtained by contacting Europe Direct or your local documentation centre (european-union.europa.eu/contact-eu/meet-us_en).

EU law and related documents

For access to legal information from the EU, including all EU law since 1951 in all the official language versions, go to EUR-Lex (eur-lex.europa.eu).

Open data from the EU

The portal data.europa.eu provides access to open datasets from the EU institutions, bodies and agencies. These can be downloaded and reused for free, for both commercial and non-commercial purposes. The portal also provides access to a wealth of datasets from European countries.

The European Commission's science and knowledge service

Joint Research Centre

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

EU Science Hub
joint-research-centre.ec.europa.eu

@EU_ScienceHub

EU Science Hub - Joint Research Centre

EU Science, Research and Innovation

EU Science Hub

EU Science

Publications Office
of the European Union