ECOSTAT nutrient meeting (18.-19.11.2015)

Session 2: Analysis of pressure-response relationships and application of nutrient boundaries

Part 1 Methods and results for lakes

Geoff Phillips

Objectives of the work

- To review, or where data are available to analyse, pressure response relationships developed between BQE and nutrients in lakes and rivers during the intercalibration process;
- To provide examples of an appropriate range of boundary values for different lake/river types;
- To compare the above ranges with MS nutrient boundaries for different lake/river types and discuss the extent to which uncertainties in the relationships may influence the development of boundary values which are supporting biological class boundaries;
- To develop approach how to use BQE pressure-response relationships for nutrient boundary development

Draft report available

The use of pressure response relationships between nutrients and biological quality elements as a method for establishing nutrient supporting element boundary values for the Water Framework Directive

Geoff Phillips[1,2], Sebastian Birk[3], Jürgen Böhmer[4], Martyn Kelly[5], Nigel Willby[2]

[1] Marine Science Support, Norwich UK

[2] University of Stirling, Stirling UK

[3] University of Duisburg-Essen, Essen Germany

[4] Bioforum GmbH, Kirchheim/Teck, Germany

[5] Bowburn Consultancy, Newcastle UK

Draft 04 November 2015 for discussion at Berlin workshop (18-19 November 2015)

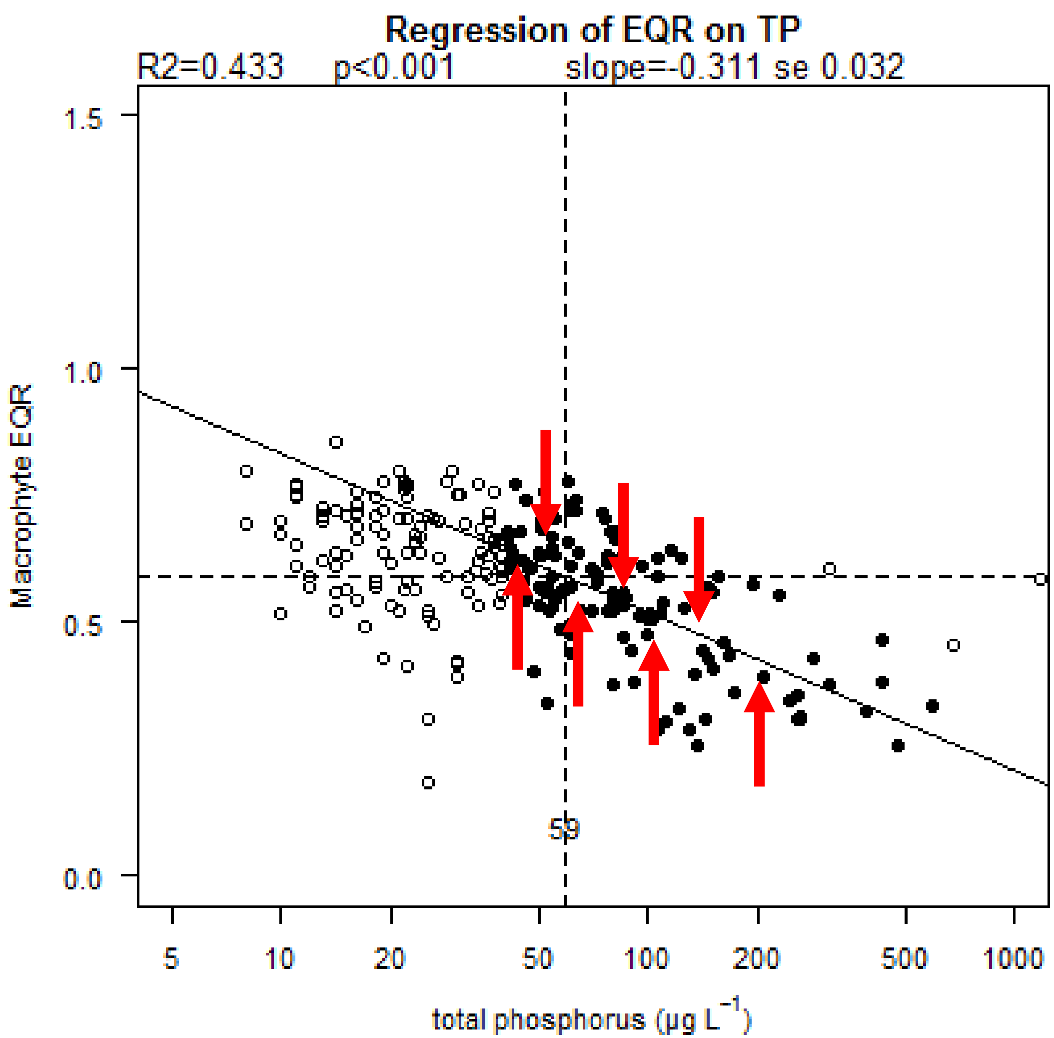
What I will talk about

- Background to the statistics
- Comparison with categorical methods
  - Box plots for classified water bodies
  - Minimisation of mis-match of classified water bodies
- Examples using the intercalibration data sets and the common metrics
- General conclusions

What are we aiming to do when we generate a pressure response relationship?

- Predict the nutrient concentration that occurs at the boundary between Good & Moderate or High & Good
- Use regression model
  - Linear
  - Curved (GAM model)
- Determine the nutrient concentration at the biology boundary using the regression line

Example: Macrophytes v TP (LCB1 lakes)



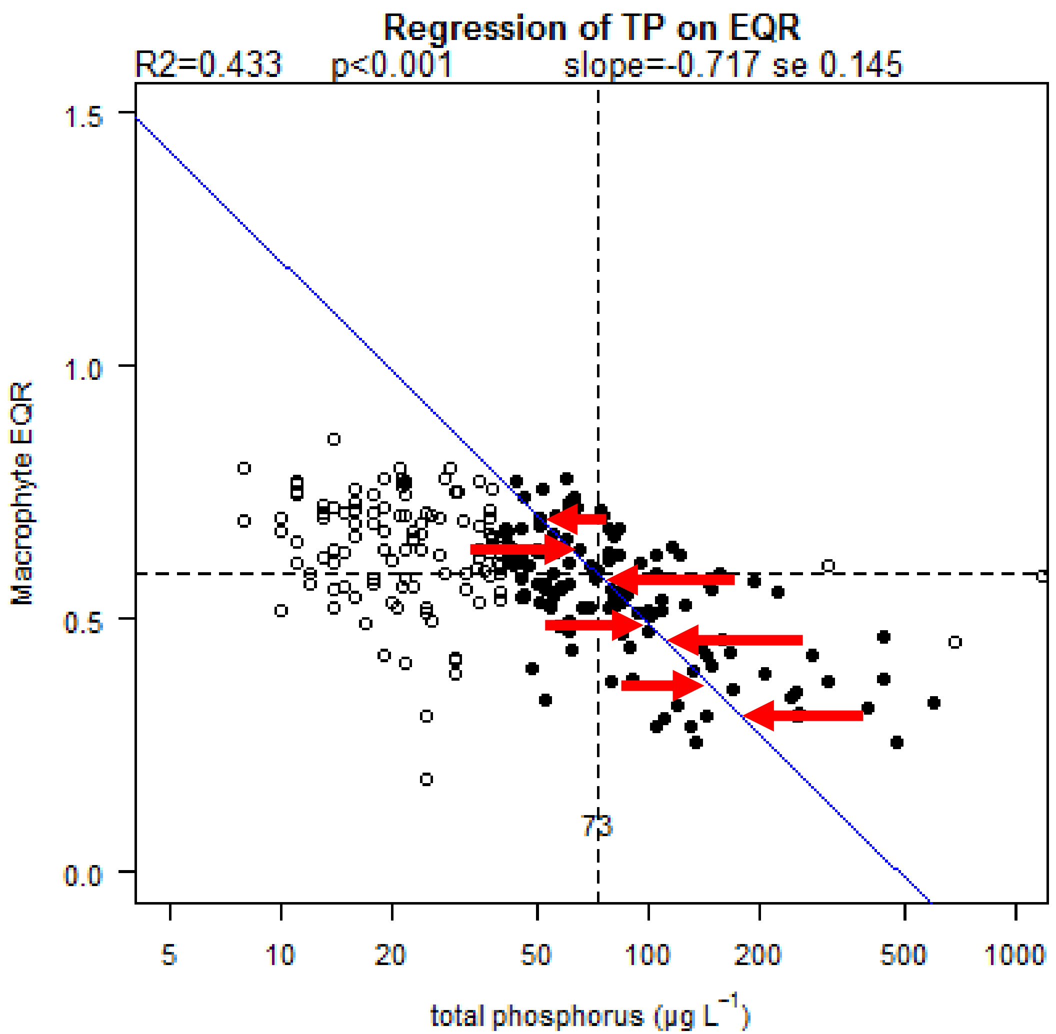
Common Metric EQR boundary 0.59 (average of all MS boundary values on the common metric scale) gives a value of 59 μg/l TP

TP is the independent (X) variable
We minimise the variation in the dependent (Y) variable biology (EQR)

This makes the assumption that the value we use for each TP is precise, measured with no error (such as might be the case in an experimental trial)

We test that the slope is significantly different from 0, otherwise the line effectively represents the mean EQR value of all the data

Example TP v Macrophytes



We could reverse our dependent and independent variables so we predict the value of TP (dependent Y variable) at the EQR boundary value (independent X variable)

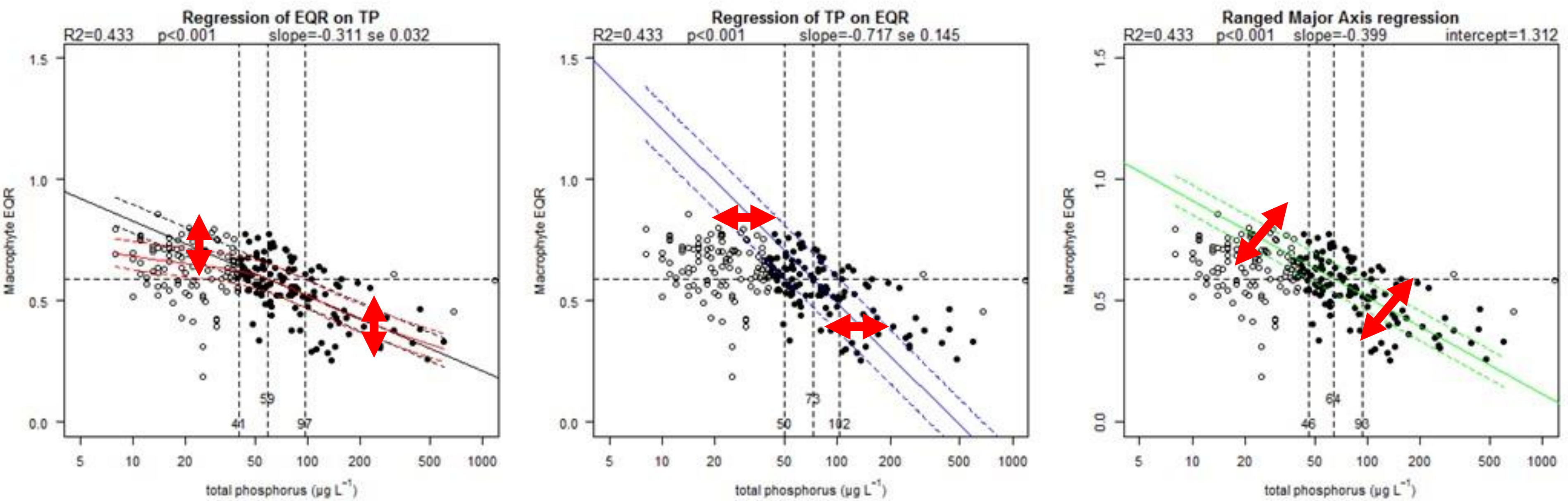
We minimise the variation in TP rather than EQR

Plot shows the result, but with the axis reversed to allow a direct comparison

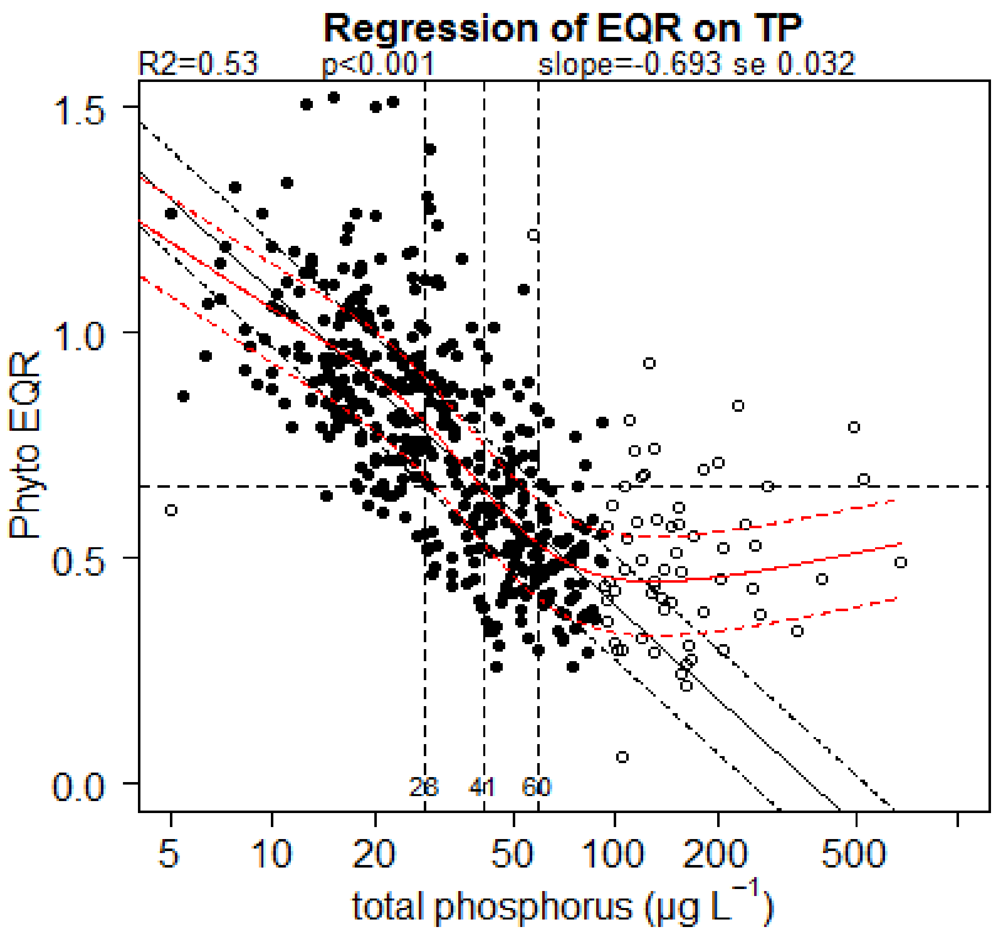
The TP boundary is now 73µg/l TP rather than 59µg/l TP because slope is greater

Comparing regression methods

Where uncertainty is relatively high ($R^2 < 0.6$) a conventional regression is likely to underestimate the true slope, the inverse approach may overestimate the slope
The "true" slope is likely to be between these values
We suggest using a type II regression (Ranged Major Axis regression) when variation in both X & Y are used to establish the line
Provides an intermediate value of slope and thus predicted boundary value



Dealing with uncertainty



Only use data within linear part of response (show using GAM model red line)

Figure shows lines bounding 50% of the data (calculated from upper lower quartiles of regression residuals)

We can predict a range of nutrient concentrations at give boundary value

Represent the range of likely values if we were to use a different (but similar type) data set, due to uncertainty of regression parameters (slope & intercept)

Use of multi-variate regression

Biological status may depend on both nitrogen and phosphorus, so we can produce a multiple regression using both variables

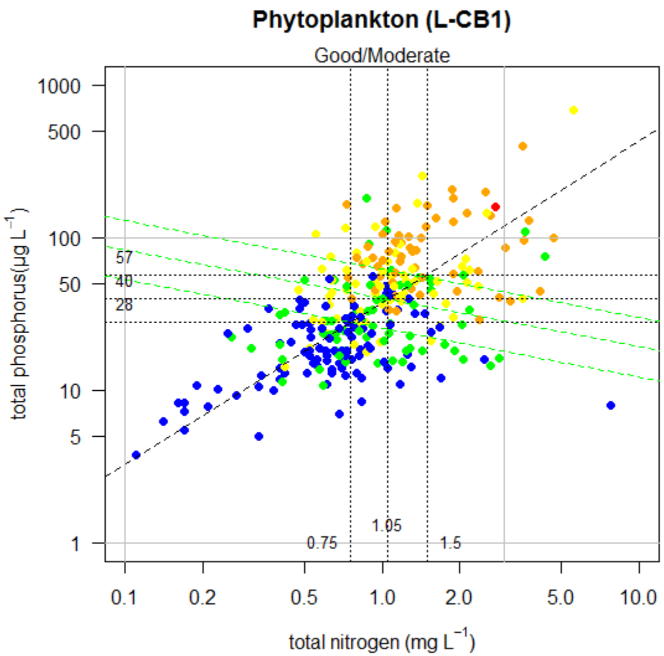
$EQR = aTP + bTN + c + Error$

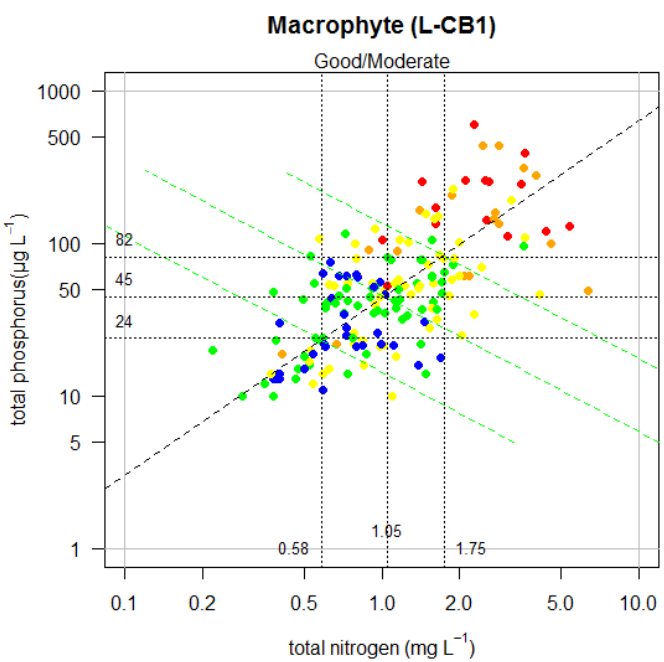
We can use this equation to predict the value of either

1. TP for a fixed value of TN for a given boundary EQR value
2. TN for a fixed value of TP for a given boundary EQR value

Infinite range of boundary values, but we can show these using contour lines by plotting the predicted TP and TN boundary values on a plot of TP v TN for the G/M boundary EQR.

Plot TP v TN for all LCB1 lakes





Points coloured by class for a)Phytoplankton and b)Macrophytes using common metric boundary values

Dotted line shows mean relationship between TP v TN (type II regression)

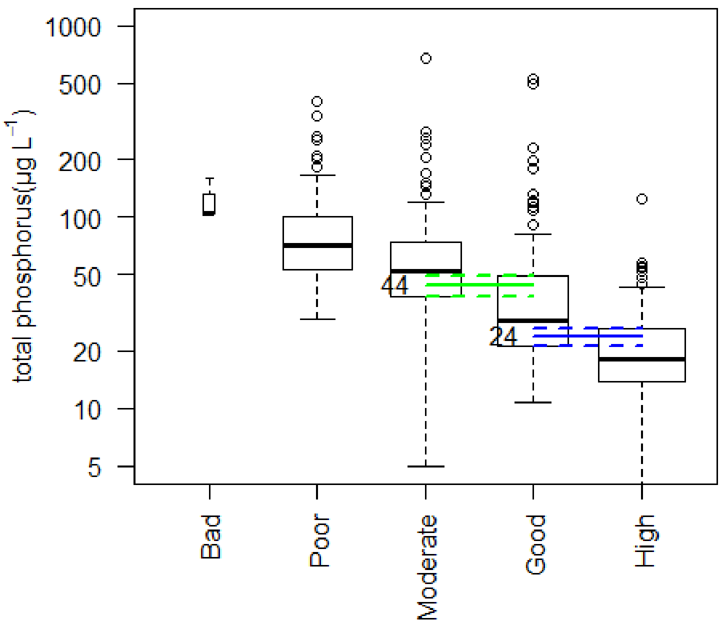
Green dotted lines show contours for GM boundary values for TP and TN, predicted from the multivariate regression

(note that the contours are closer to 45 degrees for macrophytes than they are for phytoplankton)

Where these intersect with the TP v TN regression line we can determine pairs of boundary values for TP and TN

As for the univariate regression we can also determine the uncertainty and thus a range of boundary values for each of TP and TN

Categorical Method



Distribution of TP in LCB1 lakes classified using common metric for phytoplankton

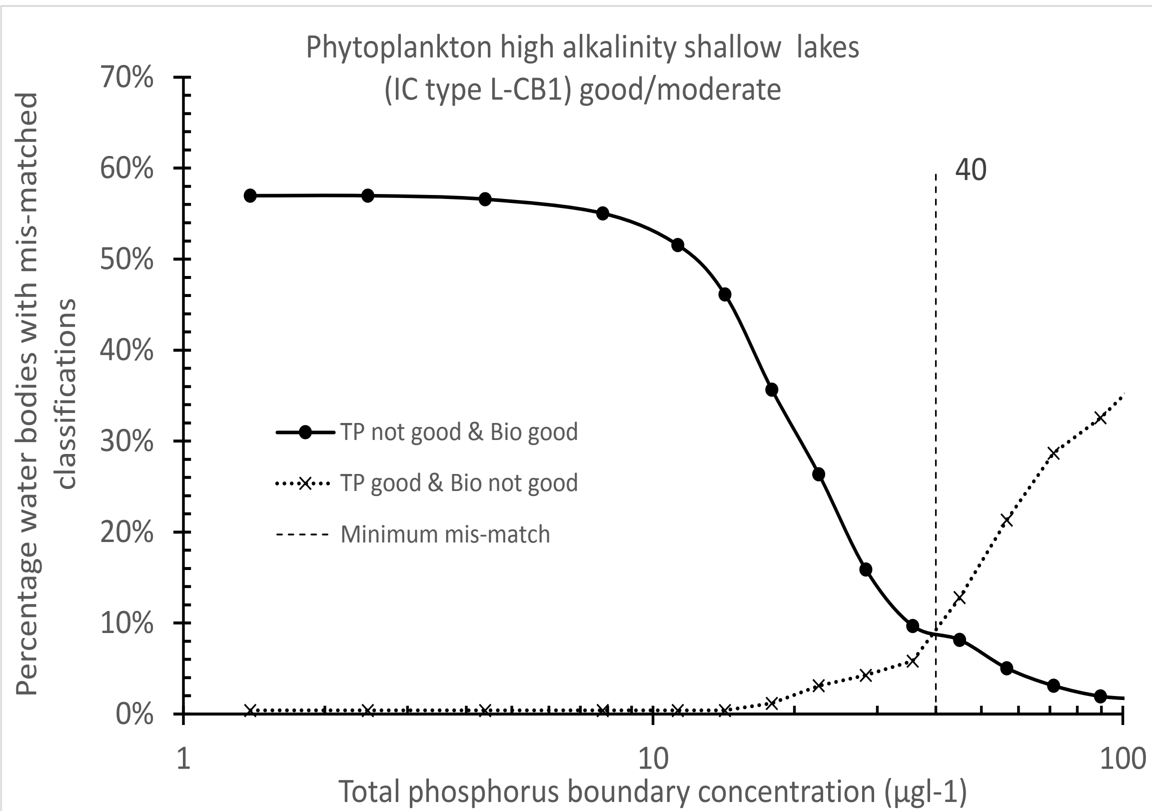
Upper quartile TP for Good (75% of lakes in Good status had TP lower than this value

Lower quartile TP for Moderate (25% of lakes in Moderate status had TP higher than this value)

Average of these values might be a Good/Moderate boundary value

Simple alternative is to average the median value of Good and Moderate and their respective quartiles to provide values half way between these classes (the Good/Moderate boundary)

Minimise mismatch



Compare classifications for biology and for nutrients e.g. phytoplankton and TP using different values for TP boundary (use Excel)

Plot mismatch for each TP boundary value
a)TP not Good, Phyto Good
b)TP Good, Phyto not Good

Intersection shows minimum level of mismatch is 10% and occurs at TP value of 40µg/l

It is possible to estimate uncertainty by sampling the data, but difficult to do in Excel

Summarise results LCB1 lakes

Predicted total phosphorus boundary values for high alkalinity shallow lakes (L-CB1) using regression models and categorical methods

IC Type	Phytoplankton Models	R^2	nutrient range TP $\mu g l^{-1}$	GM TP $\mu g l^{-1}$ Pred	GM 25th	GM 75th	HG TP $\mu g l^{-1}$ Pred	HG 25th	HG 75th
LCB1	EQR v TP + TN (OLS)	0.55	4 - 100	40	28	57	22	15	32
	EQR v TP (OLS)		4 - 91	41	28	60	22	15	32
	TP v EQR (OLS)	0.53	4 - 91	35	26	48	25	18	34
	EQR v TP (RMA)		4 - 91	39	28	51	23	17	31
	Average adjacent quartiles			44			24		
	Average adjacent classes			44	30	61	23	18	37
	Minimise class difference			40			32		

IC Type	Macrophyte Models	R^2	nutrient range TP $\mu g l^{-1}$	GM TP Pred	GM 25th	GM 75th	HG TP Pred	HG 25th	HG 75th
LCB1	EQR v TP + TN (OLS)	0.40	10 - 597	45	24	82	15	8	30
	EQR v TP (OLS)			59	41	97	26	18	43
	TP v EQR (OLS)	0.43	41 - 597	73	50	102	51	35	72
	EQR v TP (RMA)			64	46	93	34	24	50
	Average adjacent quartiles			39			31		
	Average adjacent classes			39	25	68	31	20	44
	Minimise class difference			45			21		

Most likely boundary range using min/max of all predicted values 35 - 44µg/l

Best model predicted 40 and range of values 28 - 57µg/l

Possible range using min/max from all methods 26 - 61µg/l

Repeat this for all BQEs

Summary of predicted total phosphorus boundary values for high alkalinity lakes

IC Type	BQE used		GM TP $\mu g l^{-1}$			HG TP $\mu g l^{-1}$		
			Pred	range		Pred	range	
LCB1	Phytoplankton	most likely boundary		35	44		22	32
		best model R^2 0.55	40	28	57	22	15	32
		possible range		26	61		15	37
	Macrophytes	most likely boundary		39	73		15	51
		best model R^2 0.40	45	24	82	15	8	30
		possible range		24	102		8	72
LCB2	Phytoplankton	most likely boundary		45	66		32	35
		best model R^2 0.68	52	40	75	34	27	42
		possible range		35	122		22	55
	Macrophytes	most likely boundary		66	90		23	53
		best model R^2 0.47	70	36	125	30	16	56
		possible range		25	156		9	87
XGIG LCB1 LCB2	Phytobenthos	most likely boundary		36	47		16	29
		best model R^2 0.50	45	24	83	19	10	35
		possible range		22	96		7	42
	Invertebrates	most likely boundary		41	49		16	27
		best model R^2 0.38	43	22	90	21	11	44
		possible range		15	119		5	48

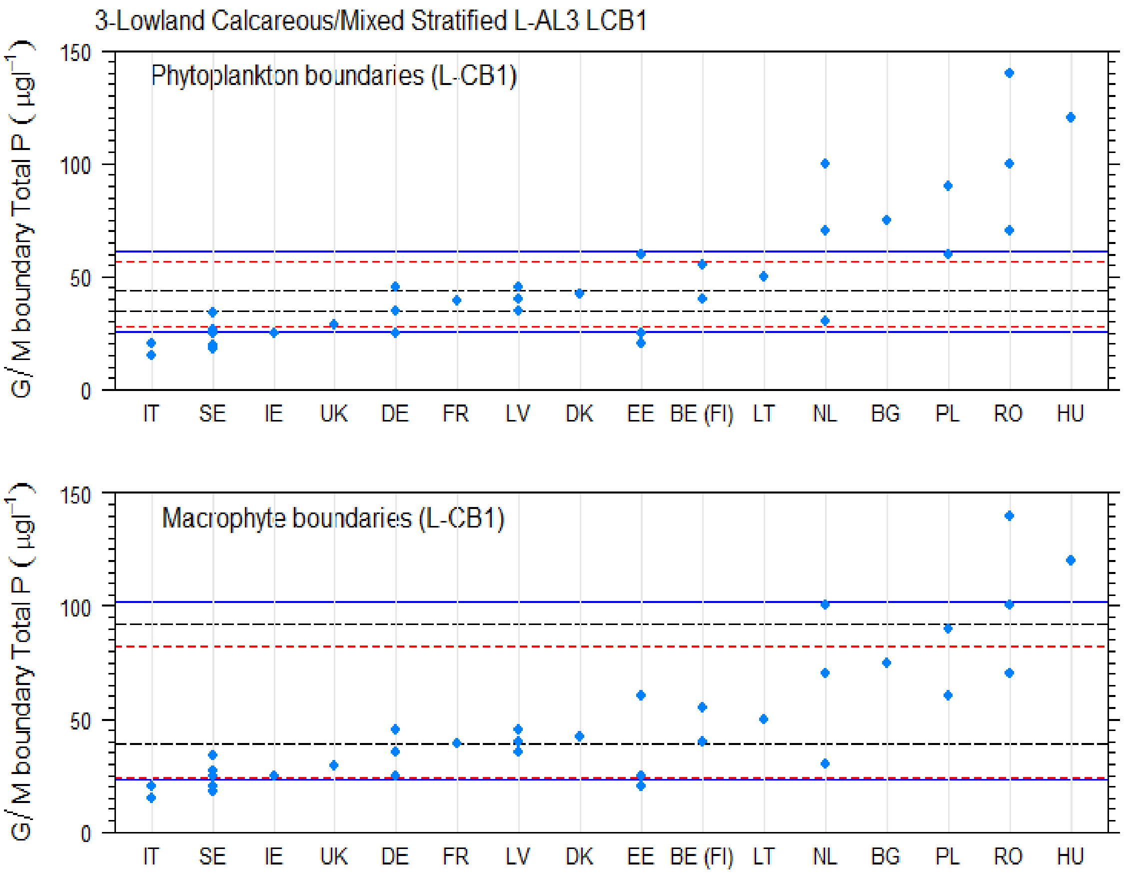
Algal metrics have lowest boundary values

Most sensitive to nutrients

Shallow high alkalinity lakes have higher TP boundaries than very shallow lakes

For lakes all R2 values are significant and most relatively high

Compare predicted values (LCB1) with reported boundary values (Broad Type 3 lowland calcareous/mixed stratified

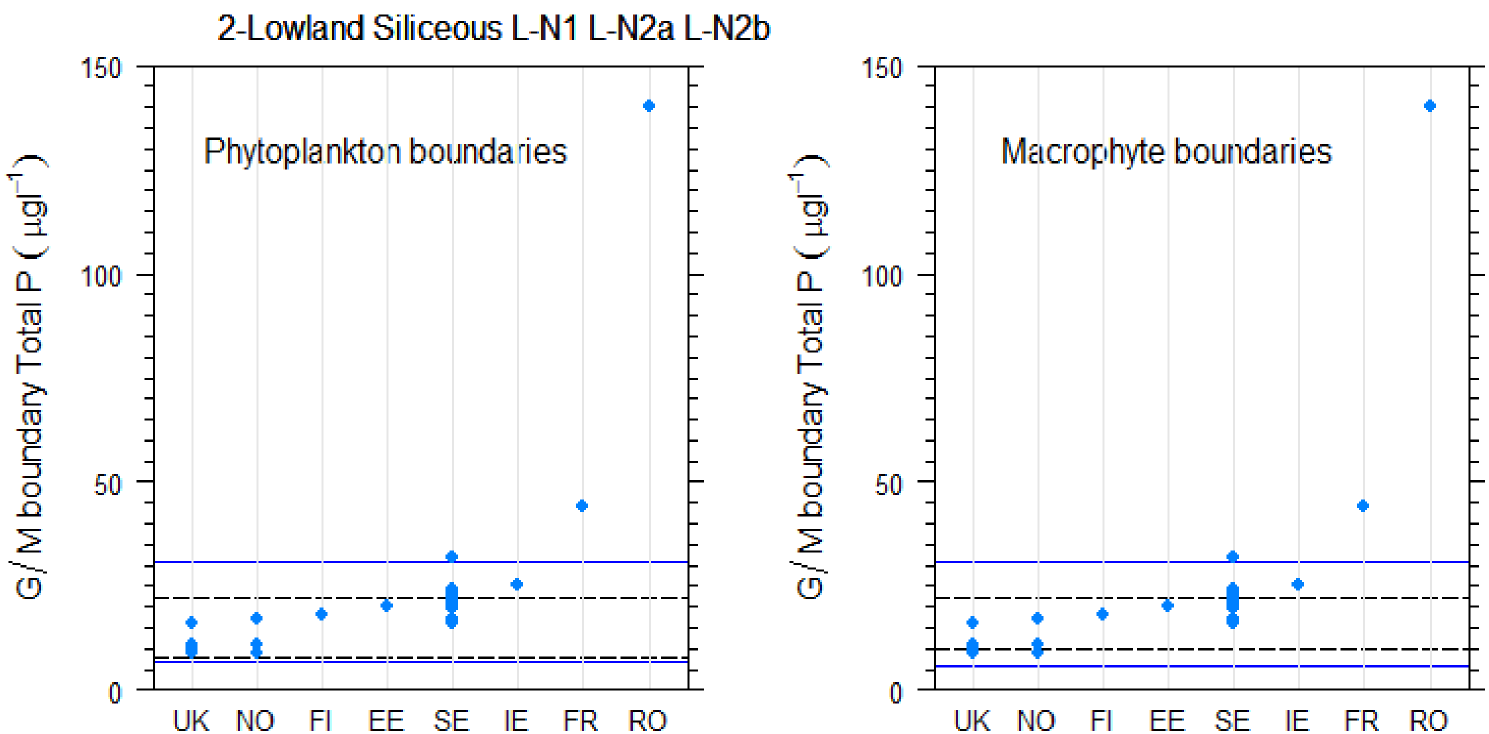


Most likely range (black broken line)

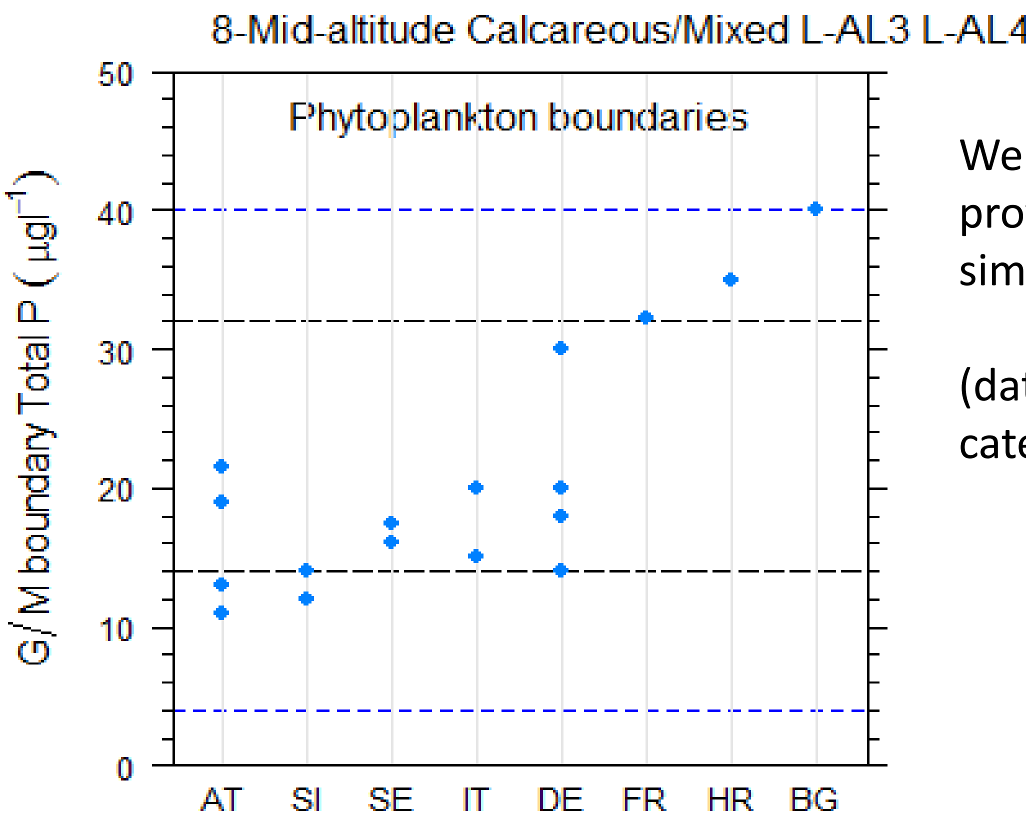
Best model upper/lower quartiles of model residuals (red dotted line)

Possible range (blue solid line)

Broad type 2 Lowland siliceous



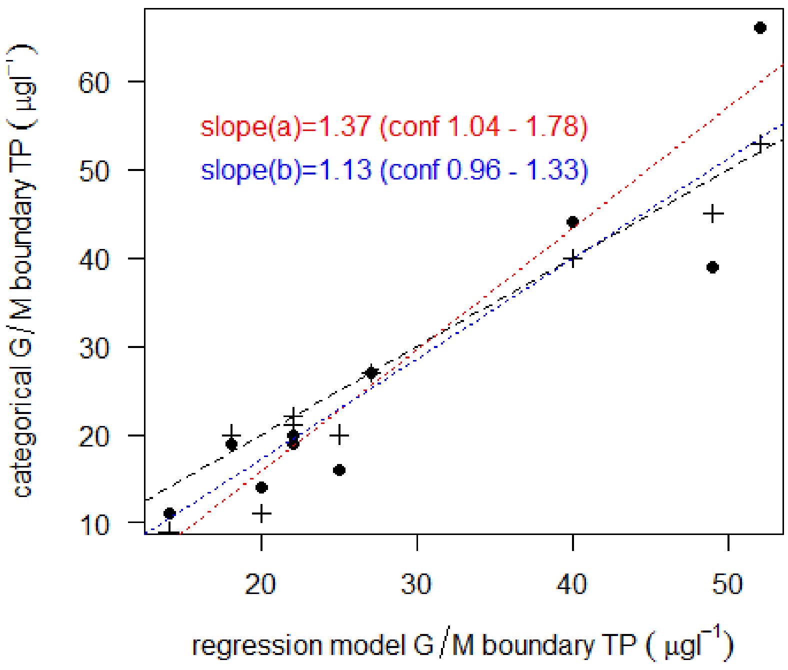
Broad type 8



We used regression equations provided by Alpine GIG to show similar results for Broad type 8 lakes

(data not available so no Type II or categorical methods used)

Comparison of methods



Relationship between good/moderate boundary values predicted from best regression model and

a) minimising mismatch (closed circles)
b) boxplots (cross).

Black dotted line shows 1:1 relationship
Red line RMA regression v mismatch
Blue line RMA regression v box plots.

Slope very close to 1.00 suggesting categorical method is equally likely to generate similar boundary values to regression method

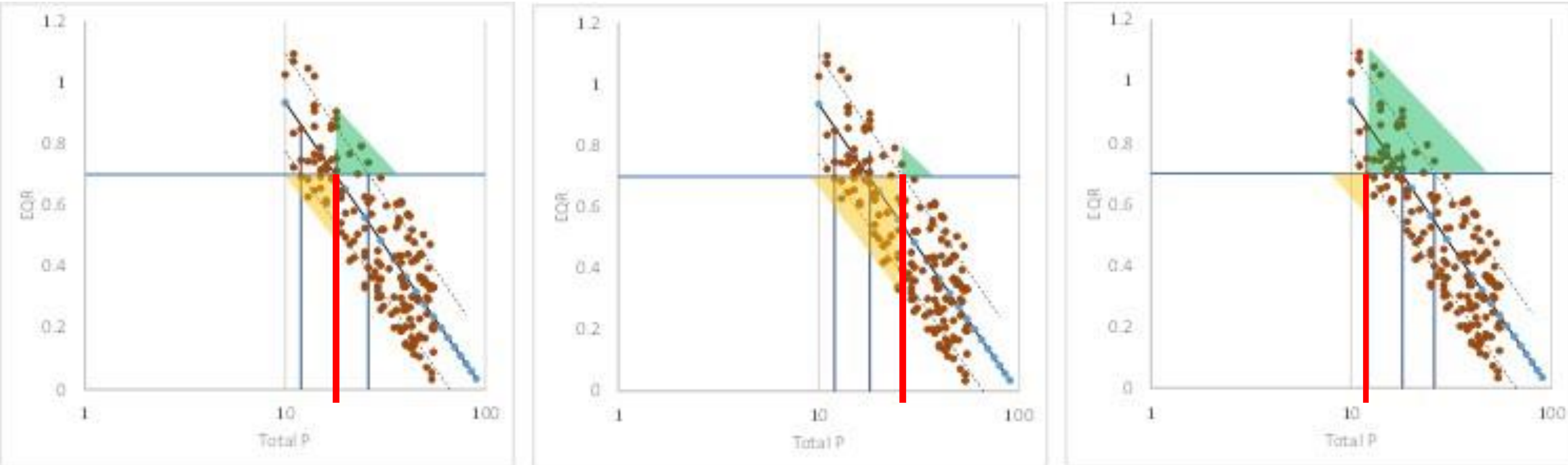
Overall conclusions for lakes

- Relationships for lakes relatively good, particularly for phytoplankton and phytobenthos (macrophytes for TN)
- Uncertainty remains high with a range of boundary values
  - Noise in data (sampling regimes, analysis)
  - Other environmental variables not accounted for by typology
  - Other pressures influencing biology
- Pressure/response and categorical methods of IC data suggest that majority of MS have established boundary value that fall within the most likely or possible ranges

Way forward

- Agreement on best practice
  - Size of data set
  - Which BQE to use for nutrients
  - Method of analysis
  - Interpretation of results

Interpretation of results



Hypothetical relationship between total phosphorus and biological EQR, showing regression line with confidence intervals (dotted lines).

Triangles mark areas where classification mismatches occur, green (biology Good but phosphorus Moderate) and yellow (biology Moderate or worse but phosphorus Good) using 3 different approaches for interpretation.

a)Mismatch occur (as always) but no bias, equal risk of nutrient or biology controlling class
b)Mismatch greater but higher nutrient boundary minimises risk of downgrades by nutrient
c)Mismatch greater but more protective, but with much greater risk of downgrades by by nutrients

Choice of approach depends on how the boundary values are used for management. For classification the first approach seems to be the most sensible and nearest to the current

Summary

- Pressure response or categorical analysis of large data sets provides the foundation for establishing boundaries for nutrient supporting elements
- How MS deal with uncertainty is a critical issue, even for lakes
- For rivers uncertainty is much greater and makes boundary setting more difficult