

Accepted refereed manuscript of: Phillips G, Teixeira H, Poikane S, Salas Herrero F & Kelly MG (2019)

Establishing nutrient thresholds in the face of uncertainty and multiple stressors: A comparison of approaches using simulated datasets. *Science of the Total Environment*, 684, pp. 425-433.

DOI: <https://doi.org/10.1016/j.scitotenv.2019.05.343>

© 2019, Elsevier. Licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0

International <http://creativecommons.org/licenses/by-nc-nd/4.0/>

1 Establishing nutrient thresholds in the face of 2 uncertainty and multiple stressors: a comparison of 3 approaches using simulated data sets

4

5

6 Geoff Philips¹, Heliana Teixeira², Sandra Poikane³, Fuensanta Salas³ and Martyn G. Kelly^{4,5}*

7 ¹Biological and Environmental Sciences, University of Stirling, Stirling FK4 9LA, UK

8 ² Department of Biology & CESAM, University of Aveira, Campus de Santiago, 3810-193 Aveiro,

9 Portugal.

10 ³European Commission Joint Research Centre (JRC), I-21027, Ispra, Italy

⁴ Bowburn Consultancy, 11 Montaigne Drive, Bowburn, Durham DH6 5QB,

UK.

12 ⁵ Department of Geography, Nottingham University, Nottingham NG7 2RD, UK.

13

14 * Corresponding author. *E-mail address*: MGKelly@bowburn-consultancy.co.uk (M.G. Kelly)

15

1 Abstract

2 Various methods have been proposed to identify threshold concentrations of nutrients that would
3 support good ecological status, but the performance of these methods and the influence of other
4 stressors on the underlying models have not been fully evaluated. We used synthetic datasets to
5 compare the performance of ordinary least squares, logistic and quantile regression, as well as,
6 categorical methods based on the distribution of nutrient concentrations categorized by biological
7 status. The synthetic datasets used differed in their levels of variation between explanatory and
8 response variables, and were centered at different position along the stressor (nutrient) gradient. In
9 order to evaluate the performance of methods in “multiple stressor” situations, another set of
10 datasets with two stressors were used. Ordinary least squares and logistic regression methods were
11 the most reliable when predicting the threshold concentration when nutrients were the sole stressor;
12 however, both had a tendency to underestimate the threshold when a second stressor was present.
13 In contrast, threshold concentrations produced by categorical methods were strongly influenced by
14 the level of the stressor (nutrient enrichment, in this case) relative to the threshold they were trying
15 to predict (good/moderate in this instance). Although all the methods tested had limitations in the
16 presence of a second stressor, upper quantiles seemed generally appropriate to establish non-
17 precautionary thresholds. For example, upper quantiles may be appropriate when establishing
18 targets for restoration, but not when seeking to minimise deterioration. Selection of an appropriate
19 threshold concentration should also attend to the regulatory regime (i.e. policy requirements and
20 environmental management context) within which it will be used, and the ease of communicating the
21 principles to managers and stakeholders .

22 Key words

23 eutrophication, phosphorus, nitrogen, threshold, multiple stressors, Water Framework Directive

24

25 1. Introduction

26 Legislation such as the Water Framework Directive (WFD: European Union, 2000) offers significant
27 opportunities to incorporate ecological knowledge into regulatory mechanisms that ensure
28 sustainable water resources. A key assumption behind such legislation is that, if reasons for
29 deterioration of ecosystems can be identified, then appropriate measures can be put in place to
30 remediate and/or protect against future deterioration. To this end, a large number of metrics
31 summarising the response of the aquatic biota have been developed to meet WFD objectives (Birk et
32 al., 2012) and the position of good and high ecological status boundaries have been harmonised
33 between Member States (Birk et al., 2013; Poikane et al., 2014, 2015). In practice, however, the
34 dynamic nature of ecosystems creates uncertainty in relationships between biology and stressors,
35 with the consequence that predictions of the benefits of remediation lack precision (Moe et al., 2015;
36 Prato et al., 2014). This is widely recognised as a major weakness of WFD implementation (Hering et
37 al., 2010; 2015; Carvalho et al., 2018).

38 A good understanding of the relationship between biology and a pressure should, in theory, enable a
39 regulator to set threshold concentrations beyond which ecological degradation is likely to occur;
40 however, relationships with stressors such as nutrient enrichment are often weak and confounded by
41 interactions with other stressors (Page et al., 2012; Harris and Heathwaite, 2012; O'Hare et al., 2018;
42 Munn et al., 2018). Consequently, the process of defining thresholds also needs to account for
43 uncertainty in the relationship between biology and pressure.

44 A number of methods for setting thresholds for nutrient concentrations have been described.

45 Broadly speaking, these fall into two types:

- 46 ● those that assume a continuous response of both explanatory and response variables, from
47 which a threshold concentration can be inferred using linear regression models. This is the case

48 when ecological status (or other measures of biological condition – e.g. Davies & Jackson, 2006) is
49 defined as a position along a metric scale;

50 ● those that assume a categorical response of one of these variables, allowing threshold
51 concentrations to be inferred using a number of approaches, including binomial logistic
52 regression and methods based on the distribution of pressure values within the class(es) of
53 interest. This is the case when ecological status is expressed as one of a number of classes, but
54 would also be relevant if a particular species or habitat required protection. Free et al. (2016), for
55 example, describe the development of nutrient standards to protect the distinctive *Chara*-
56 dominated communities found in shallow, marl lakes in Ireland, protected under the European
57 Union’s Habitats Directive (European Community, 1992).

58 A range of approaches for calculating threshold concentrations, encompassing both of these
59 strategies, have been described in the literature (Dodds & Oakes, 2004; Free et al., 2016; Hausmann
60 et al., 2016; Poikane et al., 2019) but, as far as we know, there has been no attempt to compare the
61 effectiveness of different methods and, importantly, no systematic consideration of the extent to
62 which the values produced may be confounded in the presence of a second stressor. This is
63 important as the biological response to nutrient enrichment in European freshwaters (Phillips et al.
64 2018), transitional and coastal waters (Salas et al., 2019) is often distorted by the presence of other
65 stressors, posing difficulties for setting nutrient thresholds that ensure the integrity of aquatic
66 systems. Recent evidences indicate furthermore that nutrient stress occurs in 71% to 98% of multi-
67 stress situations in Europe’s surface waters (Nöges et al., 2016). The importance of multiple pressures
68 in shaping communities in aquatic systems is now widely acknowledged (Borja et al., 2011; Nöges et
69 al., 2016; Hering et al., 2015; Feld et al., 2016) although there are, as yet, no definitive approaches to
70 setting protective thresholds for constituents of any pressure or stressor “cocktail”. Multiple
71 stressors are likely to interact in different ways and their effects can be difficult to predict, as there is

72 evidence for additive, synergistic and antagonistic effects of multiple stressors in aquatic ecosystems
73 depending on the nature of those stressors and the type of ecosystem (e.g. Jackson et al., 2016;
74 Gieswein et al., 2017; Rodrigues et al., 2018).

75 Although the WFD requires member states (MS) to establish threshold values for physico-chemical
76 metrics that support good status there is no requirement for these values, unlike the biological
77 metrics, to be harmonised. Given that different methods were used and the inherent uncertainty in
78 relationships, it is not surprising that a wide range of values are now in use across Europe (Phillips and
79 Pitt, 2016). To overcome this and facilitate the use of more uniform threshold values, guidance
80 supported by a statistical toolkit has been produced (Phillips et al. 2018). This encourages the use of a
81 variety of approaches and in this paper we apply these approaches to synthetic datasets, designed to
82 resemble stressor-response relationships between nutrient enrichment and biological community
83 changes, in order to draw out some general lessons on the suitability of different approaches in
84 situations where nutrients are the principal stressor shaping biological communities, and also in the
85 presence of a second stressor. It is not our intention to investigate complex effects. Rather, we use
86 simulated datasets to support identification of data patterns and show sensitivity of commonly used
87 statistical methods to the presence of unmeasured stressors.

88 2 Materials and methods

89 2.1 Datasets

90 In order to make comparisons between the different methods, a series of synthetic data sets were
91 produced. Each data set contained 200 random values of total phosphorus (TP) concentration, a
92 simulated observed Ecological Quality Ratio (EQR) representing overall environmental conditions
93 where only this single stressor influences the observed EQR, and a second simulated EQR where the
94 value was determined by a combination of phosphorus and a second unknown stressor that also had
95 a negative effect on the observed EQR. Apart from the negative effect of both stressors, the

96 synthetic data does not explore the nature of the interaction between the two stressors (i.e. additive,
97 synergistic or antagonistic, *sensu* Pigot et al., 2015) as this is also often unknown in real case
98 scenarios. The only assumption is that where a second stressor is suspected different methods will
99 show different sensitivities to its presence. Likewise, the nutrient thresholds derived from the
100 relationship observed may be more or less reliable or approximate to the “true” value of the
101 measured stressor depending on the method.

102 Each data set was generated as follows:

103 1) A normally distributed random set of 200 total phosphorus (TP) concentrations with a known
104 mean and standard deviation was created. The distribution of these data was chosen such that the
105 true EQR would range across the biological gradient from high to moderate status;

106 2) A “true” EQR was generated from these values, using the regression parameters for a
107 relationship between $\log_{10}TP$ and EQR. (parameters taken from the relationship between TP and
108 phytoplankton in lakes used during the Central-Baltic GIG lake intercalibration exercise: Phillips et al.,
109 2014):

$$110 \quad \text{EQR}_{\text{True}} = -0.62(\text{Log}_{10} \text{TP}) + 1.79. \quad \text{equation 1}$$

111 This equation can be re-arranged to determine the “true” TP concentration at the good/moderate
112 boundary EQR, assumed to be 0.6:

$$113 \quad \text{TP} = 10^{((0.6 - 1.79)/-0.62)} = 83 \mu\text{gL}^{-1} \quad \text{equation 2}$$

114 3) A simulated observed EQR was then created from TP by adding a normally distributed
115 random error term (E), which had a mean of 0 and a known standard deviation (Figure 1a).

$$116 \quad \text{EQR}_{\text{SimObs}} = -0.62(\text{Log}_{10} \text{TP}) + 1.79 + E \quad \text{equation 3}$$

117 4) Another normally-distributed set of EQR values (EQR2ndPressure) with a fixed mean and
118 standard deviation) was then generated to represent a hypothetical second stressor together with a
119 random probability (0-1) that this second stressor occurs at a particular site.

120 5) A simulated observed EQR (EQRSimObs 2 pressures) resulting from both TP and the 2nd
121 stressor was calculated by taking the lowest of either the simulated EQR from phosphorus
122 (EQRSimObs) or the 2nd stressor EQR (EQR2ndPressure) where the probability of the second stressor
123 was >0.5 . Where the probability of the 2nd stressor was ≤ 0.5 , the EQR from phosphorus was used.
124 Scatter plots produced from this approach typically had “wedge-shaped” data clouds, an example of
125 which is shown in Figure 1b.

126 To assess the effect of different levels of uncertainty and data that span different levels of stress, ten
127 replicate data sets were generated with the same mean TP and error standard deviation. The process
128 was repeated using ten different mean TP values (40, 50, 60, 70, 80, 90, 100, 110, 120, 130 $\mu\text{g l}^{-1}$) and
129 10 different error standard deviations (0.12 – 0.30), representing increased scatter in the true
130 relationship, to finally produce 1000 data sets for each of the single and two-stressor scenarios, an
131 example with mean TP of 50 μl^{-1} and error standard deviation of 0.15 is shown in figure 1.

132 2.2 Methods for estimating nutrient threshold concentrations

133 The following methods were used to identify threshold concentrations of TP corresponding to the
134 good/moderate status boundary (assumed to be EQR = 0.6):

135 **Ordinary least-squares regression (OLS):** The most obvious approach uses a linear regression
136 between EQR (dependent variable) and log TP (independent variable), with nutrient threshold values
137 determined from the regression parameters.

138 **Logistic regression:** An alternative approach that treats ecological status as a categorical variable
139 where a logistic model is fitted between categorical data using a binary response, “biology moderate
140 or worse” = 1 or “biology good or better” = 0 and log TP. Threshold concentrations are determined to

141 be where the probability of being moderate or worse was 0.5. In the case of two stressors an
142 additional value was determined at probability of 0.75

143 **Categorical methods:** nutrient concentrations associated with a particular ecological status class (e.g.
144 good ecological status) could also be expressed as a distribution from which an upper quantile might
145 be chosen to indicate a nutrient concentration above which good status was very unlikely to be
146 achieved, or a lower quantile below which good status was very likely to be achieved, so long as
147 nutrients were the main drivers of status. However, the variation inherent in biology-nutrient
148 relationships means that there will be many instances where lower concentrations of nutrients are
149 not associated with good status and vice-versa. The risk of misclassification could, therefore, be
150 reduced by also considering the distribution of nutrient concentrations in the adjacent class
151 (moderate, in this case), where a lower quantile could be adopted to indicate the nutrient
152 concentration below which moderate status was unlikely (and good status was likely to be achieved).
153 Three different approaches were included in these comparisons: average of medians of adjacent
154 classes; average of adjacent quartiles (75th percentile of “good status” and 25th percentile of
155 “moderate status”) and the use of the 75th percentile of “good status” alone.

156 **Minimisation of mismatch of classification:** An approach that estimates the nutrient threshold value
157 by minimising the mismatch between status (good or better and moderate or worse) for ecological
158 status and the stressor. The method calculates the proportion of records where the biological status
159 is better than the stressor and where it is worse for incremental values of the nutrient threshold. The
160 nutrient threshold value where these two sets of proportions are equal determines the point at
161 which there is the lowest mismatch of classifications. To determine this value a bootstrap approach
162 was used. For each data set 75% of the data were randomly selected and the proportions of mis-
163 classification determined. A loess model was then fitted to these data to determine the nutrient
164 concentration where the mismatch was equal. This was repeated 50 times and the mean nutrient
165 concentration determined.

166 **Linear quantile regression:** when the nutrient-biology interactions is confounded by other stressors
167 or environmental variables, the variance around the mean of the response variable is also a function
168 of those explanatory variable(s), leading to wedge-shaped distributions. In such cases, the quantile
169 regression allows different rates of change in the response variable to be predicted along the upper
170 (in the presence of stressors) or lower (in the presence of mitigating factors) boundary of the
171 conditional distribution of the data (Cade and Noon, 2003). The choice of an appropriate quantile to
172 use is somewhat arbitrary, though more extreme values will have a greater potential to be influenced
173 by outliers. We have used the 75th percentile as a compromise that enables upper threshold to be
174 modelled with a reasonable degree of precaution and confidence.

175 2.3 Comparison of methods

176 Each of these methods was applied to the synthetic data sets and the predicted good/moderate
177 threshold concentrations (assumed to be EQR = 0.6) for each were recorded for comparison with the
178 "true" threshold concentration defined by equation 2 above. The extent of uncertainty was also
179 recorded using the coefficient of determination (r^2) from the regression between EQRSimObs and TP,
180 these values were categorised into 5 levels (5 ≥ 0.6 , 4 ≥ 0.5 , 3 ≥ 0.4 , 2 ≥ 0.3 , 1 < 0.3) to allow the effect
181 of scatter to be determined. To assess the influence of the data distribution along the stressor
182 gradient, results for each of the data sets categorised by mean TP were compared. When applying
183 the methods to real datasets, the threshold nutrient at the good to moderate EQR boundary would
184 be unknown so it would be impossible to determine how a derived threshold relates to the average
185 TP. In such cases, the mean EQR could be used so we present results for the synthetic datasets using
186 the true EQR value determined from the mean TP of the data set using equation 1.

187 All analyses were performed using R statistical software (R Development Core Team, 2016). Base
188 statistics were used for all methods except linear quantile regression, which was fitted using the rq

189 function from quantreg package (Koenker and Hallock 2001). Original computer code is available
190 from the first author on request.

191 3 Results

192 3.1 Average differences

193 A comparison of the range of predicted TP threshold values for the Good-Moderate boundary shows
194 that ordinary least squares (OLS) regression and binary logistic regression at a probability of 0.5
195 predicted the smallest range of values ($c \pm 5 \mu\text{g TP L}^{-1}$, Figures 2 and 3). The variability of the
196 categorical methods was substantially higher (typically $\pm 15 \mu\text{g TP L}^{-1}$), while the minimisation of
197 mismatch method predicted a range of values that lie between the regression and categorical
198 methods ($\pm 8 \mu\text{g TP L}^{-1}$). When TP was treated as a single stressor all methods, except the 75th
199 percentile of the TP concentration in sites with good- biological status, predicted values that were
200 centred around the true threshold value ($83 \mu\text{g l}^{-1}$ – see equation 2). The 75th percentile predicted
201 significantly higher values than the other methods ($F = 163$ $p < 0.001$, Figure 2).

202 When a second stressor was present (Figure 3) the predicted range of values did not change, but both
203 linear and logistic regression (at a probability of 0.5), underestimated the true threshold value by 36
204 $\mu\text{g TP L}^{-1}$, suggesting that these methods are not appropriate under such circumstances. In contrast
205 the categorical methods were less influenced; the two averaging approaches (median and adjacent
206 quartiles) slightly underestimating ($-5 \mu\text{g TP L}^{-1}$), with the upper 75 percentile closer to the true mean
207 ($+12 \mu\text{g TP L}^{-1}$). The minimisation of mismatch method also underestimated the true threshold,
208 although less so than was the case for the regression methods ($-20 \mu\text{g TP L}^{-1}$). Quantile regression,
209 using the 0.75 quantile and the logistic regression using a probability of 0.75 provided nutrient
210 threshold estimates for the good-moderate boundary that were higher than the true threshold ($+26$
211 $\mu\text{g TP L}^{-1}$ and $+34 \mu\text{g TP L}^{-1}$).

212 3.2 Influence of position of data cloud along stressor gradient and at different levels 213 of variability

214 For a single stressor neither OLS nor logistic regression (using a probability of 0.5) methods were
215 significantly influenced by either their uncertainty or position on the stressor gradient (Figure 4 &
216 Table 1). In contrast, all of the categorical methods and the mismatch approach were significantly
217 influenced by the level of stressor, under-estimating the true threshold at low exposures (i.e.
218 predicting a lower nutrient threshold thus more stringent than necessary) and over-estimating at high
219 pressures (i.e. predicting a higher nutrient threshold thus more relaxed than that required). The
220 average of adjacent quartiles and the 75th percentile of TP in good-moderate status were the most
221 sensitive to uncertainty, the minimisation of mismatch the least, but all had a significant interaction
222 term showing an increasing effect of uncertainty as the stressor level increased. Where a second
223 stressor was present, a similar pattern was seen (Figure 5 and Table 1), although the effects of
224 uncertainty and stressor levels were slightly higher. For example, both OLS and logistic regressions (p
225 = 0.5) predicted boundary values that were significantly affected by both variability and their position
226 on the stressor gradient when two stressors influenced ecological status, although the effect was
227 much smaller than for the other methods. Logistic regression predictions using $p = 0.75$ were
228 particularly influenced by position on the stressor gradient, overpredicting the true threshold at low
229 stressor levels. The predictions using quantile regression ($p = 0.75$) were not significantly influenced
230 by stressor level, but were by variability, with higher predictions at high levels of uncertainty.

231 4 Discussion

232 Where a single stressor dominates the response of biology, linear regression or binary logistic
233 regression are the most reliable approaches. Neither are substantially influenced by the mean of the
234 data set and both are only slightly influenced by scatter in the data. This is unsurprising given that the

235 data were generated with normally distributed errors and thus conform to the requirements of
236 regression models.

237 Any attempt to develop nutrient thresholds in freshwaters or coastal waters, however, also needs to
238 be aware that nutrients rarely act in isolation (Vinebrooke et al., 2004; Wagenhoff et al., 2011;
239 Piggott et al., 2015; Gunderson et al., 2016), particularly in rivers and estuaries, and our analyses
240 indicate how interactions with a second stressor can confound the face-value relationship between
241 biology and the stressor of interest. Consideration of the complex relationships between the
242 ecological response and stressors acting simultaneously is essential to decide management actions,
243 because of non-linear and interactive effects of stressors (Brown et al., 2013)

244 In these situations, neither linear nor logistic regressions are appropriate as the confounding effect of
245 the second stressor will result in the under estimation of nutrient thresholds. Such data show
246 heteroscedasticity, with decreasing variance as stressor levels increase, caused by the 2nd stressor
247 overriding the otherwise low influence of nutrient effects. The categorical approaches initially appear
248 to be less influenced by this problem, as on average they make predictions that are clustered around
249 the true mean. However, unlike the regression methods, they are much more sensitive to the
250 position of the data cloud on the stressor gradient. If the data are clustered around the boundary
251 being predicted (the good/moderate boundary, EQR = 0.6, in our study), they are the least sensitive
252 to the effect of a 2nd pressure. However if the data are centred below or above the boundary of
253 interest, they are likely to under- and overpredict, respectively, with the threshold error increasing as
254 uncertainty increases. The least influenced was the minimisation of mismatch method, but all the
255 approaches, other than those seeking to describe the behaviour of the upper distributions of the
256 data, are likely to underestimate threshold values due to the influence of other stressors.

257 The best solution to this problem would be to develop a more complex model that could account for
258 additional pressures; however, a lack of reliable data and the complexity of modelling make this

259 impractical (Feld et al., 2016 Duarte et al., 2009). Whilst the combined effect of multiple stressors was
260 previously assumed to be additive , this is not always the case in ecological systems, where
261 antagonistic and synergistic interactions may dominate (e.g. Crain et al., 2008; Jackson et al., 2016;
262 Gieswein et al., 2017; Munn et al., 2018; Rodrigues et al., 2018).

263 An alternative approach would be to fit an upper quantile, which identifies an upper surface to the
264 relationship between EQR and nutrient concentration. The problem with this approach is that it
265 needs to consider the uncertainty in the relationship between EQR and phosphorus. Our simulations
266 show that quantile regression predicts higher values as uncertainty increases. As the uncertainty of
267 the true relationship between nutrient and EQR decreases, a clearer upper boundary emerges, with
268 the upper quantile that is modelled to determine a threshold value approaching, or better reflecting,
269 the true effect of the single stressor. On the other hand, it does indicate the highest values of a
270 physicochemical parameter that is consistent with good status (Müller et al., 2017). Beyond this
271 point, nutrients are likely to exert an effect regardless of the presence of other stressors.

272 There is no 'correct' quantile, and one should inspect the distribution of quantiles within the
273 particular range of interest (Koenker and Hallock, 2001). Higher quantiles offer greater chance that
274 the true response to nutrient stressor is being captured but with the risk that the regression line is
275 anchored by fewer, and more extreme, records at any level of pressure (Koenker, 2011). This
276 problem is particularly acute with small datasets. In practice, the 75th percentile offers a balance
277 between precaution and statistical robustness when dealing with medium-size datasets, although
278 our simulations suggest that even this value may be too high, over predicting at all levels of pressure.

279 Choosing an upper probability value with logistic regression is a similar approach, potentially allowing
280 threshold values to be determined when a second pressure is present despite any confounding
281 effects. However, again it is difficult to determine the appropriate probability to use. The selection of

282 probability should obey fit-for-purpose criteria, for which several classification measures exist that
283 can be used as support (Fielding and Bell, 1997).

284

285 Taylor et al. (2018) advocate a combination of spatial, temporal and experimental approaches in
286 order to characterise the response of biota to nutrient enrichment whilst, at the same time,
287 recognising that comprehensive study designs can become prohibitively expensive. Their study was
288 limited to a single group of biota, diatoms, whilst we would advocate the examination of the
289 response of different ecosystem components to enhance the insights (Robertson et al, 2006).

290 Teichert et al. (2016), by contrast, used a random forest algorithms to detect the dominant stressors
291 in estuaries. At the heart of these approaches, however, lie datasets that capture the spatial and/or
292 temporal variation in assemblages along a strong nutrient gradient and it is also important that
293 statistical analysis of such datasets are both robust and easy to communicate to non-specialist
294 managers and stakeholders.

295 In practice, however, such approaches are a necessary element when developing such thresholds
296 because they offer the most straightforward means of capturing the range of uncertainty associated
297 with the water body type under investigation. It is, however, important to validate thresholds using
298 independent sources of evidence. The use of experimental systems (Bowes et al., 2012; McCall et al.,
299 2017; Taylor et al., 2018;) is one means of doing this, but other options are also available (e.g. Free et
300 al., 2016).

301 5 Conclusions

302 Our simulations suggest that, where there is a strong stressor-response relationship between
303 nutrients and ecological status, any of the tested modelling methods, with the exception of the
304 threshold derived from the 75th percentile of nutrient concentration in sites with good ecological
305 status, are likely to give reliable estimates of nutrient concentrations that are associated with the

306 ecological good-moderate boundary. Of these, OLS or logistic regression are the most reliable, while
307 the minimisation of mismatch method is perhaps the easiest to communicate to managers. This is
308 likely to be the situation for lakes where the dominance of the algal response to nutrients is clear.

309 In rivers, estuaries and coastal waters however, multiple stressors are common; here the assumed
310 robust regression approaches may be strongly influenced by stressors other than nutrients and there
311 is a risk that threshold values that are lower than needed may be generated, in effect penalizing
312 nutrients for impacts caused by other stressors. Such situations can be identified from wedge
313 shaped scatter plots and from plots of model residuals and it is important that these are carefully
314 considered before the results of modelling are translated to regulatory regimes.

315 Where there is evidence of multiple stressors, quantile regression or the use of logistic regression
316 with nutrient threshold concentrations are determined using a quantile or a probability greater than
317 0.5 have potential. However, the selection of an appropriate quantile remains an unresolved issue.
318 Supporting chemical element thresholds values determined for different EQR categories are unlikely
319 to be precautionary as, by their nature, they seek to minimise false positives, i.e. effect detection
320 when there is no effect. Such boundaries may be appropriate when establishing targets for
321 restoration, but less so when seeking to minimise deterioration.

322 Eutrophication is a complex issue (Dodds, 2006; O'Hare et al., 2018) but, for strategic planning and
323 high-level overviews, there are still benefits in knowing threshold values beyond which consequences
324 can be expected. Understanding the challenges involved in deriving such targets does, at least,
325 enable regulators to interpret results, and combine various strands of evidence in to make robust
326 decisions.

327 Declaration of interests

328 None

329 Acknowledgements

330 This work was supported by the European Commission (DG-ENV) Joint Research Centre (JRC). We are
331 grateful to two anonymous reviewers for helpful comments on the manuscript

332 References

- 333 Birk, S., Bonne, W., Borja, Á., Brucet, S., Courrat, A., Poikane, S., Solimini, A., Van De Bund, W.,
334 Zampoukas, N., Hering, D., 2012. Three hundred ways to assess Europe's surface waters: An almost
335 complete overview of biological methods to implement the Water Framework Directive. *Ecological*
336 *Indicators* 18, 31–41. doi:10.1016/j.ecolind.2011.10.009
- 337 Birk, S., Willby, N.J., Kelly, M.G., Bonne, W., Borja, A., Poikane, S., van de Bund, W., 2013.
338 Intercalibrating classifications of ecological status: Europe's quest for common management
339 objectives for aquatic ecosystems. *Science of The Total Environment* 454–455, 490–499.
340 doi:10.1016/j.scitotenv.2013.03.037
- 341 Borja, Á., Barbone, E., Basset, A., Borgersen, G., Brkljacic, M., Elliott, M., Garmendia, J.M., Marques,
342 J.C., Mazik, K., Muxika, I., Neto, J.M., Norling, K., Rodríguez, J.G., Rosati, I., Rygg, B., Teixeira, H.,
343 Trayanova, A., 2011. Response of single benthic metrics and multi-metric methods to anthropogenic
344 pressure gradients, in five distinct European coastal and transitional ecosystems. *Marine Pollution*
345 *Bulletin* 62, 499–513. doi:10.1016/j.marpolbul.2010.12.009
- 346 Bowes, M.J., Ings, N.L., McCall, S.J., Warwick, A., Barrett, C., Wickham, H.D., Harman, S.A.,
347 Armstrong, L.K., Scarlett, P.M., Roberts, C., Lehmann, K., Singer, A.C., 2012. Nutrient and light
348 limitation of periphyton in the River Thames: Implications for catchment management. *Science of the*
349 *Total Environment* 434, 201–212. doi:10.1016/j.scitotenv.2011.09.082

350 Brown, C.J., Saunders, M.I., Possingham, H.P., Richardson, A.J., 2013. Managing for interactions
351 between local and global stressors of ecosystems. PLoS One 8, e65765.

352 Cade, B.S., Noon, B.R., 2003. A gentle introduction to quantile regression for ecologists. *Frontiers in*
353 *Ecology and the Environment* 1, 412-420. doi:10.1890/1540-9295(2003)001[0412:AGITQR]2.0.CO;2

354 Carvalho, L., Mackay, E.B., Cardoso, A.C., Baattrup-Pedersen, A., Birk S., Blackstock, K.L., Borics, G.,
355 Borja, A., Feld, C.K., Ferreira, M.T., Globevnik, L., Grizzetti, B., Hendry, S., Hering, D., Kelly, M.,
356 Langaas, S., Meissner, K., Panagopoulos, Y., Penning, E., Rouillard, J., Sabater, S., Schmedtje, U.,
357 Spears, B.M., van de Bund, W., Solheim, A.L., 2018. Protecting and restoring Europe's waters: an
358 analysis of the future development needs of the Water Framework Directive. *Science of the Total*
359 *Environment* (in press) doi.org/10.1016/j.scitotenv.2018.12.255

360 Crain, C.M., Kroeker, K., Halpern, B.S., 2008. Interactive and cumulative effects of multiple human
361 stressors in marine systems. *Ecol. Lett.* 11, 1304–1315

362 Davies, S.P., Jackson, S.K., 2006. The biological condition gradient: a descriptive model for interpreting
363 change in aquatic ecosystems. *Ecological Applications* 16, 1251-1266. doi:10.1890/1051-
364 0761(2006)016[1251:TBCGAD]2.0.CO;2

365 Dodds, W.K., 2006. Eutrophication and trophic state in rivers and streams. *Limnology and*
366 *Oceanography* 51, 671-680. doi:10.4319/lo.2006.51.1_part_2.0671

367 Dodds, W.K., Oakes, R.M., 2004. A technique for establishing reference nutrient concentrations
368 across watersheds affected by humans. *Limnology and Oceanography Methods* 2, 333-341.
369 doi:10.4319/lom.2004.2.333

370 Duarte, C. M., Conley, D. J., Carstensen, J., & Sánchez-Camacho, M., 2009. Return to Neverland:
371 shifting baselines affect eutrophication restoration targets. *Estuaries and Coasts* 32(1), 29-36.
372 doi:10.1007/s12237-008-9111-2

373 European Community (1992). Council Directive 92 /43 /EEC of 21 May 1992 on the conservation of
374 natural habitats and of wild fauna and flora. Official Journal of the European Communities L 206: 7-
375 50.

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

379 Feld, C.K., Segurado, P., Gutiérrez-Cánovas, C., 2016. Analysing the impact of multiple stressors in
380 aquatic biomonitoring data: a cookbook with applications in R. *Science of the Total Environment*
381 573, 1320-1339. doi:10.1016/j.scitotenv.2016.06.243

382 Fielding, A.H., Bell, J.F., 1997. A review of methods for the assessment of prediction errors in
383 conservation presence/absence models. *Environmental Conservation* 24, 38–49.
384 doi:10.1017/S0376892997000088

385 Free, G., Tierney, D., Little, R., Kelly, F., Kennedy, B., Plant, C., Trodd, W., Wynne, C., Caroni, R., Byrne,
386 C., 2016. Lake ecological assessment metrics in Ireland: relationships with phosphorus and typology
387 parameters and the implications for setting nutrient standards. *Biology and Environment:*
388 *Proceedings of the Royal Irish Academy* 116B, 191-204. doi:10.3318/bioe.2016.18

389 Gieswein A., Hering D. & Feld C.K. (2017). Additive effects prevail: The response of biota to multiple
390 stressors in an intensively monitored watershed. *Science of The Total Environment* 593-594, 27-35.
391 <http://dx.doi.org/10.1016/j.scitotenv.2017.03.116>

392 Gunderson, A. R., Armstrong, E. J., Stillman, J. H., 2016. Multiple stressors in a changing world: the
393 need for an improved perspective on physiological responses to the dynamic marine environment.
394 *Annual Review of Marine Science*, 357-378. doi:10.1146/annurev-marine-122414-033953

395 Harris, G.P., Heathwaite, A.L., 2012. Why is achieving good ecological outcomes in rivers so difficult?
396 *Freshwater Biology* 57, 91-107. doi:10.1111/j.1365-2427.2011.02640.x

397 Hausmann, S., Charles, D.F., Gerritsen, J., Belton, T.J., 2016. A diatom-based biological condition
398 gradient (BCG) approach for assessing impairment and developing nutrient criteria for streams.
399 Science of the Total Environment 562, 914-927. doi:10.1016/j.scitotenv.2016.03.173

400 Hering, D., Borja, A., Carstensen, J., Carvalho, L., Elliott, M., Feld, C.K., Heiskanen, A.-S., Johnson, R.K.,
401 Moe, J., Pont, D., Lyche-Solheim, A., v. de Bund, W., 2010. The european water framework directive
402 at the age of 10: A critical review of the achievements with recommendations for the future. Science
403 of The Total Environment 408 (19), 4007-4019. doi:10.1016/j.scitotenv.2010.05.031

404 Hering, D., Carvalho, L., Argillier, C., Beklioglu, M., Borja, A., Cardoso, A.C., Duel, H., Ferreira, T.,
405 Globevnik, L., Hanganu, J., Hellsten, S., Jeppesen, E., Kodeš, V., Solheim, A.L., Nöges, T., Ormerod, S.,
406 Panagopoulos, Y., Schmutz, S., Venohr, M., Birk, S., 2015. Managing aquatic ecosystems and water
407 resources under multiple stress — An introduction to the MARS project. Science of the Total
408 Environment 503-504, 10-21. doi:10.1016/j.scitotenv.2014.06.106

409 Jackson, M. C., Loewen, C. J. G., Vinebrooke, R. D., Chimimba, C. T. (2016). Net effects of multiple
410 stressors in freshwater ecosystems: a meta-analysis. Global Change Biology 22, 180–189,
411 <http://dx.doi.org/10.1111/gcb.13028>

412 Koenker, R., Hallock, K.F., 2001. Quantile regression. Journal of Economic Perspectives 15(4), 143-
413 156. doi:10.1257/jep.15.4.143

414 Koenker, R., 2011. Additive models for quantile regression: Model selection and confidence band-aids.
415 Brazilian Journal of Probability and Statistics 25(3), 239-262. doi:10.1214/10-BJPS131

416 McCall, S.J., Hale, M.S, Smith, J.T., Read, D.S., Bowes, M.J., 2017. Impacts of phosphorus
417 concentration and light intensity on river periphyton biomass and community structure.
418 Hydrobiologia 792: 315-330. doi: 10.1007/s10750-016-3067-1

419 Moe, S.J., Lyche-Solheim, A., Soszka, H., Gołub, M., Hutorowicz, A., Kolada, A., Picinska-Fałtynowicz,

420 J., Białokoz, W., 2015. Integrated assessment of ecological status and misclassification of lakes: The
421 role of uncertainty and index combination rules. *Ecological Indicators* 48, 605-615.
422 doi:10.1016/j.ecolind.2014.08.018

423 Müller, A., Halle, M., Bellack, E., 2017. Schwellenwerte und biologische Indikatoren für
424 physikalischchemische Parameter in Fließgewässern. (Threshold values and biological indicators for
425 physicochemical parameters in running waters). *Wasser und Abfall* 06/2017, 25-30.
426 doi:10.1007/s35152-017-0068-y

427 Munn, M. D., Waite, I., Konrad, C. P. (2018). Assessing the influence of multiple stressors on stream
428 diatom metrics in the upper Midwest, USA. *Ecological Indicators* 85, 1239-1248.
429 <https://doi.org/10.1016/j.ecolind.2017.09.005>

430 Nöges, P., Argillier, C., Borja, Á., Garmendia, J.M., Hanganu, J., Kodeš, V., Pletterbauer, F., Sagouis, A.,
431 Birk, S., 2016. Quantified biotic and abiotic responses to multiple stress in freshwater, marine and
432 ground waters. *Science of the Total Environment* 540, 43-52. doi:10.1016/j.scitotenv.2015.06.045

433 O'Hare, M.T., Baattrup-Pedersen, A., Baumgarte, I., Freeman, A., Gunn, I.D.M., Lázár, A.N., Wade, A.J.,
434 Bowes, M.J., 2018. Responses of aquatic plants to eutrophication in rivers: a revised conceptual
435 model. *Frontiers in Plant Science* 9, 451. doi:10.3389/fpls.2018.00451

436 Page, T., Heathwaite, A.L., Moss, B., Reynolds, C., Beven, K.J., Pope, L., Willows, R., 2012. Managing
437 the impacts of nutrient enrichment on river systems: dealing with complex uncertainties in risk
438 analysis. *Freshwater Biology* 57, 108-123. doi:10.1111/j.1365-2427.2012.02756.x

439 Phillips, G., Pitt, J-A., 2016. A comparison of European freshwater nutrient boundaries used for the
440 Water Framework Directive: A report to WG ECOSTAT. University College London. Available via:
441 <https://circabc.europa.eu/w/browse/58a2363a-c5f1-442f-89aa-5cec96ba52d7>

442 Phillips G, Free G, Karottki I, Laplace-Treyture C, Maileht K, Mischke, U., Ott, I., Pasztaleniec, A.,
443 Portielje, R, Søndergaard, M., Trodd, W. & Van Wichelen, J., 2014. Water Framework Directive
444 Intercalibration Technical Report: Central Baltic Lake Phytoplankton ecological assessment methods.
445 Report EUR 26508 EN. Luxembourg: Publications Office of the European Union. DOI:10.2788/73991

446 Phillips, G., Kelly, M., Teixeira, H., Salas, F., Free, G., Leujak, W., Lyche Solheim, A., Várбірó, G.,
447 Poikane, S., 2018. Best practice for establishing nutrient concentrations to support good ecological
448 status, EUR 29329 EN, Publications Office of the European Union, Luxembourg. doi:10.2760/84425,
449 JRC112667

450 Piggott, J.J., Townsend, C.R., Matthaei, C.D., 2015. Reconceptualizing synergism and antagonism
451 among multiple stressors. *Ecology and Evolution* 5, 1538-1547. doi:10.1002/ece3.1465

452 Poikane, S., Zampoukas, N., Borja, A., Davies, S.P., van de Bund, W., Birk, S., 2014. Intercalibration of
453 aquatic ecological assessment methods in the European Union: Lessons learned and way forward.
454 *Environmental Science & Policy* 44, 237-246. doi:10.1016/j.envsci.2014.08.006

455 Poikane, S., Birk, S., Böhmer, J., Carvalho, L., de Hoyos, C., Gassner, H., Hellsten, S., Kelly, M.,
456 Solheim, A.L., Olin, M., Pall, K., Phillips, G., Portielje, R., Ritterbusch, D., Sandin, L., Schartau, A.-K.,
457 Solimini, A.G., van den Berg, M., Wolfram, G., van de Bund, W., 2015. A hitchhiker's guide to
458 European lake ecological assessment and intercalibration. *Ecological Indicators* 52, 533-544.
459 doi:10.1016/j.ecolind.2015.01.005

460 Poikane, S., Phillips, G., Birk, S., Free, G., Kelly, M., Willby, N., 2019. Deriving nutrient criteria to
461 support "good" ecological status in European lakes: an empirically based approach to linking ecology
462 and management. *Science of the Total Environment* 650, 2074-2084.
463 doi:10.1016/j.scitotenv.2018.09.350

464 Prato, S., La Valle, P., De Luca, E., Lattanzi, L., Migliore, G., Morgana, J.G., Munari, C., Nicoletti, L., Izzo,
465 G., Mistri, M., 2014. The “one-out, all-out” principle entails the risk of imposing unnecessary
466 restoration costs: A study case in two Mediterranean coastal lakes. *Marine Pollution Bulletin* 80, 30-
467 40. doi:10.1016/j.marpolbul.2014.01.054

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

470 Robertson, D. M., Graczyk, D. J., Garrison, P. J., Wang, L., LaLiberte, G., Bannerman, R., 2006.
471 Nutrient concentrations and their relations to the biotic integrity of wadeable streams in Wisconsin.
472 Professional Paper No. 1722, United States Geological Survey, Reston, Virginia.

473 Rodrigues, A. C., Machado, A. L., Bordalo, M. D., Saro, L., Simão, F. C., Rocha, R. J., Golovko O., Žlábek,
474 V., Barata, C., Soares, A.M.V.M., Pestana, J. L. (2018). Invasive Species Mediate Insecticide Effects on
475 Community and Ecosystem Functioning. *Environmental Science & Technology* 52(8), 4889-4900.
476 <http://dx.doi.org/10.1021/acs.est.8b00193>

477 Salas, F. Teixeira, H., Poikane, S. (2019). A novel approach for deriving nutrient criteria to support
478 good ecological status: Application to coastal and transitional waters and indications for use.
479 *Frontiers in Marine Science*. DOI: 10.3389/fmars.2019.00255

480 Taylor, J.M., Back, J.A., Brooks, B.W., King, R.S., 2018. Spatial, temporal, and experimental: three
481 study design cornerstones for establishing defensible numeric criteria for freshwater ecosystems.
482 *Journal of Applied Ecology* 55, 2114-2123. doi:10.1111/1365-2664.13150

483 Teichert, N., Borja, A., Chust, G., Uriarte, A., Lepage, M., 2016. Restoring fish ecological quality in
484 estuaries: Implication of interactive and cumulative effects among anthropogenic stressors. *Science*
485 *of Total Environment* 542: 383-393

486 Vinebrooke, R.D., Cottingham, K.L., Norberg, J., Scheffer, M., Dodson, S.I., Maberly, S.C., Sommer, U.,
487 2004. Impacts of multiple stressors on biodiversity and ecosystem functioning: the role of species co-
488 tolerance. *Oikos* 104, 451-457. doi:10.1111/j.0030-1299.2004.13255.x

489 Wagenhoff, A., Townsend, C.R., Phillips, N., Matthaei, C.D., 2011. Subsidy-stress and multiplestressor
490 effects along gradients of deposited fine sediment and dissolved nutrients in a regional set of streams
491 and rivers. *Freshwater Biology* 56, 1916-1936. doi:10.1111/j.1365-2427.2011.02619.x

1 List of Tables

2 **Table 1 Analysis of variance table showing the influence of variability (r^2 category) and position of**
 3 **data cloud (mean TP) on TP thresholds predicted for good ecological status using the different**
 4 **methods applied to synthetic data set. Significant F values shown in bold.**

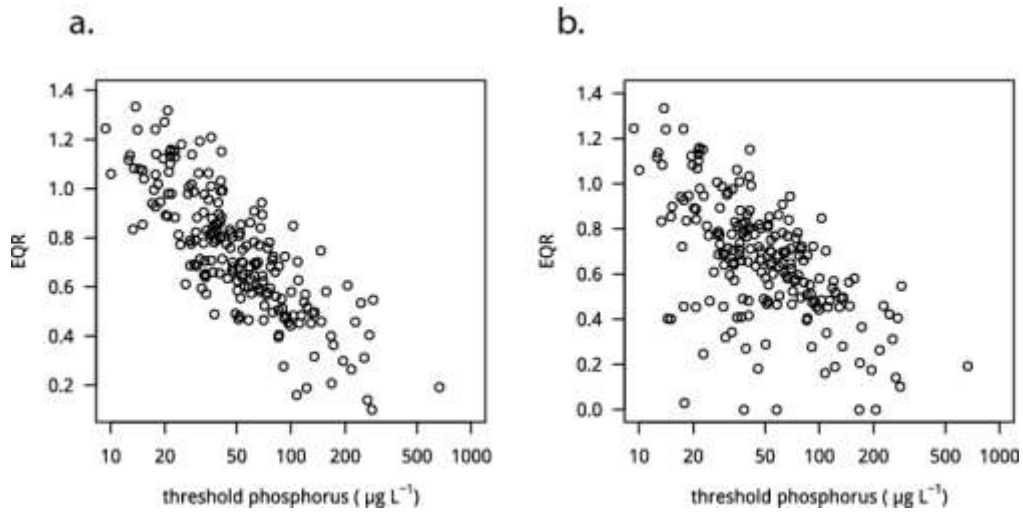
Method	Data set	Variability		Position data cloud (mean TP)		Interaction	
		F	p	F	p	F	p
OLS regression	single stressor	3.5	0.06	0.362	0.548	0.332	0.564
Ave. median	two stressors	50.8	<0.001	29.4	<0.001	2.9	0.09
	single stressor	3.8	0.051	7731	<0.001	312.2	<0.001
Ave. quartile	two stressors	4.1	0.044	8572.1	<0.001	173.7	<0.001
	single stressor	53.5	<0.001	53.5	<0.001	6091.7	<0.001
75th percentile	two stressors	117.6	<0.001	6807.7	<0.001	146.2	<0.001
	single stressor	578.2	<0.001	3149	<0.001	239.7	<0.001
Mismatch	two stressors	512	<0.001	3062.1	<0.001	203.5	<0.001
	single stressor	0.2	0.651	2178.1	<0.001	219.1	<0.001
Logistic regression ($p=0.5$)	two stressors	0.1	0.803	4069.3	<0.001	48.5	<0.001
	single stressor	1.2	0.277	0.059	0.809	0.617	0.432
Logistic regression ($p=0.75$)	two stressors	279.1	<0.001	75.2	<0.001	0.565	0.452
	single stressor	36.5	<0.001	113.5	<0.001	1.6	0.209
Quantile regression ($p=0.75$)	two stressors	272.9	<0.001	1.3	0.251	0.342	0.559

5

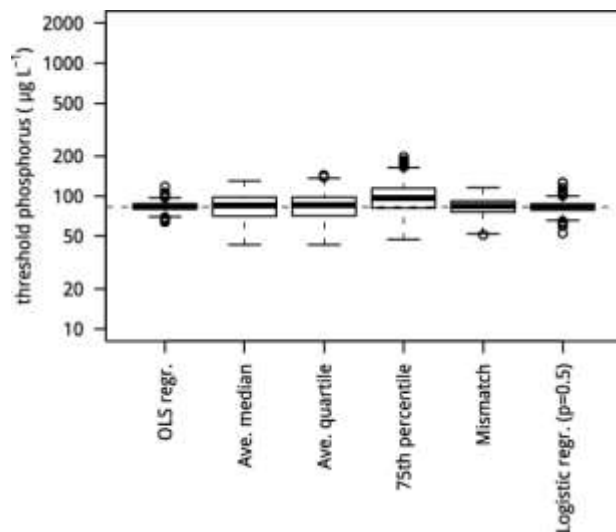
6

7

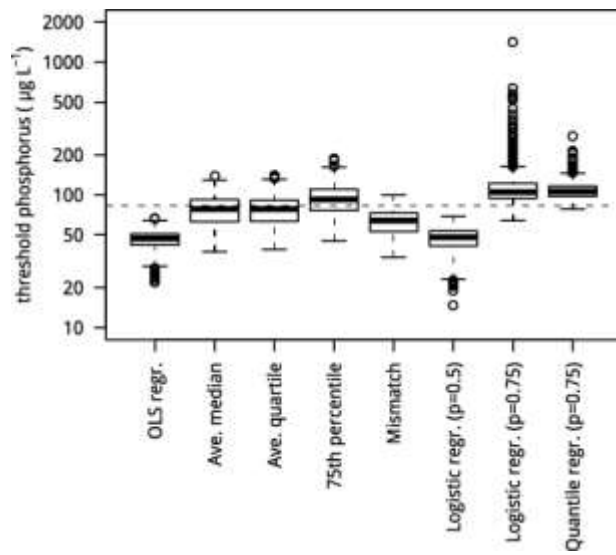
1 **Figures**



3 **Figure 1. Typical scatter plots showing relationships between simulated EQR and total phosphorus**
4 **for a) single stressor gradient of phosphorus, b) combined stressor gradient of phosphorus and a**
5 **second pressure. (Data were for a mean TP of 50 µg⁻¹ and an error standard deviation of 0.15)**



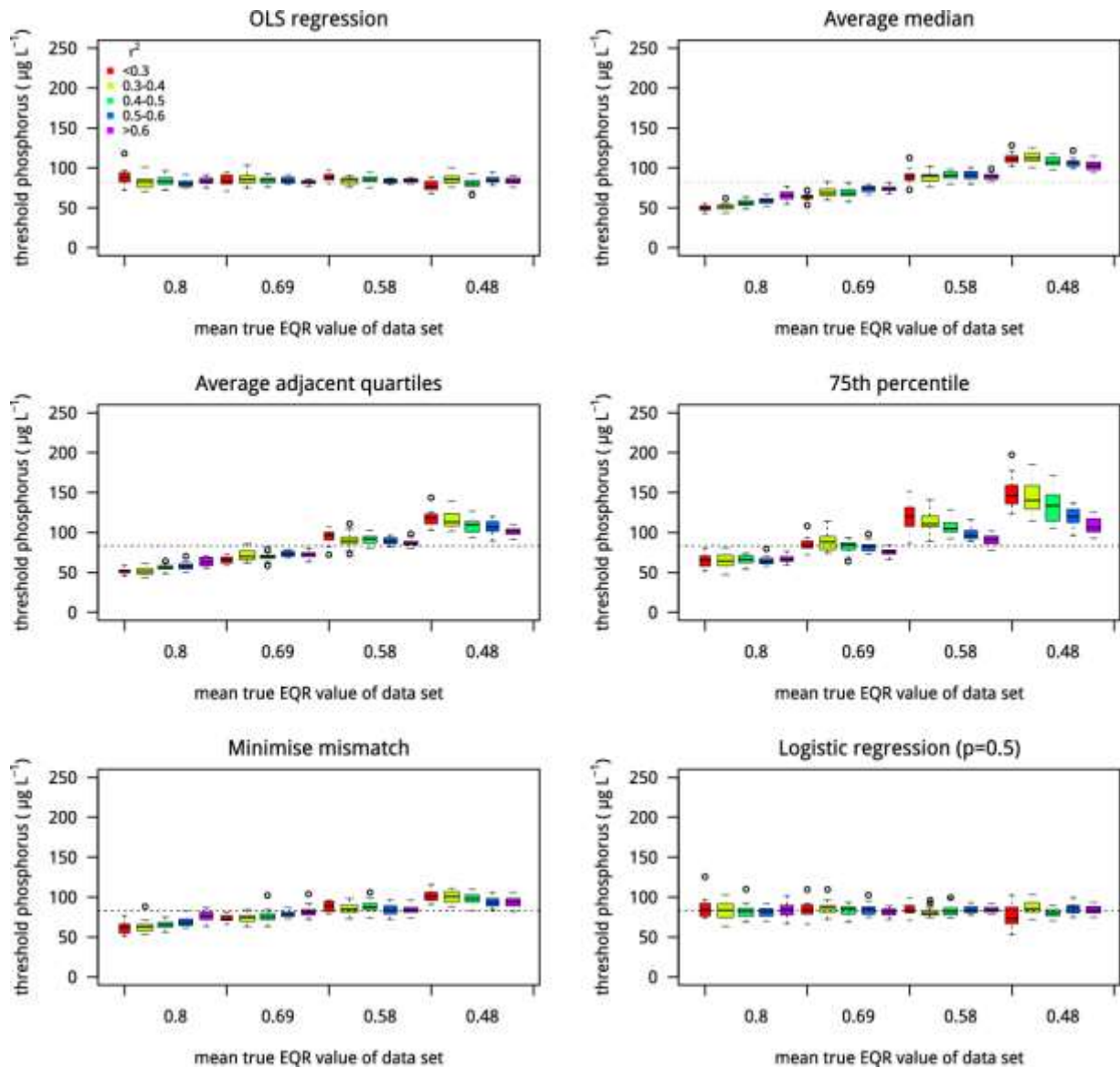
7 **Figure 2. Range of TP concentrations at the good/moderate threshold predicted by the different**
8 **methods. (In this and subsequent figures the dotted line shows the true threshold concentration**
9 **(83ug L⁻¹)).**



10

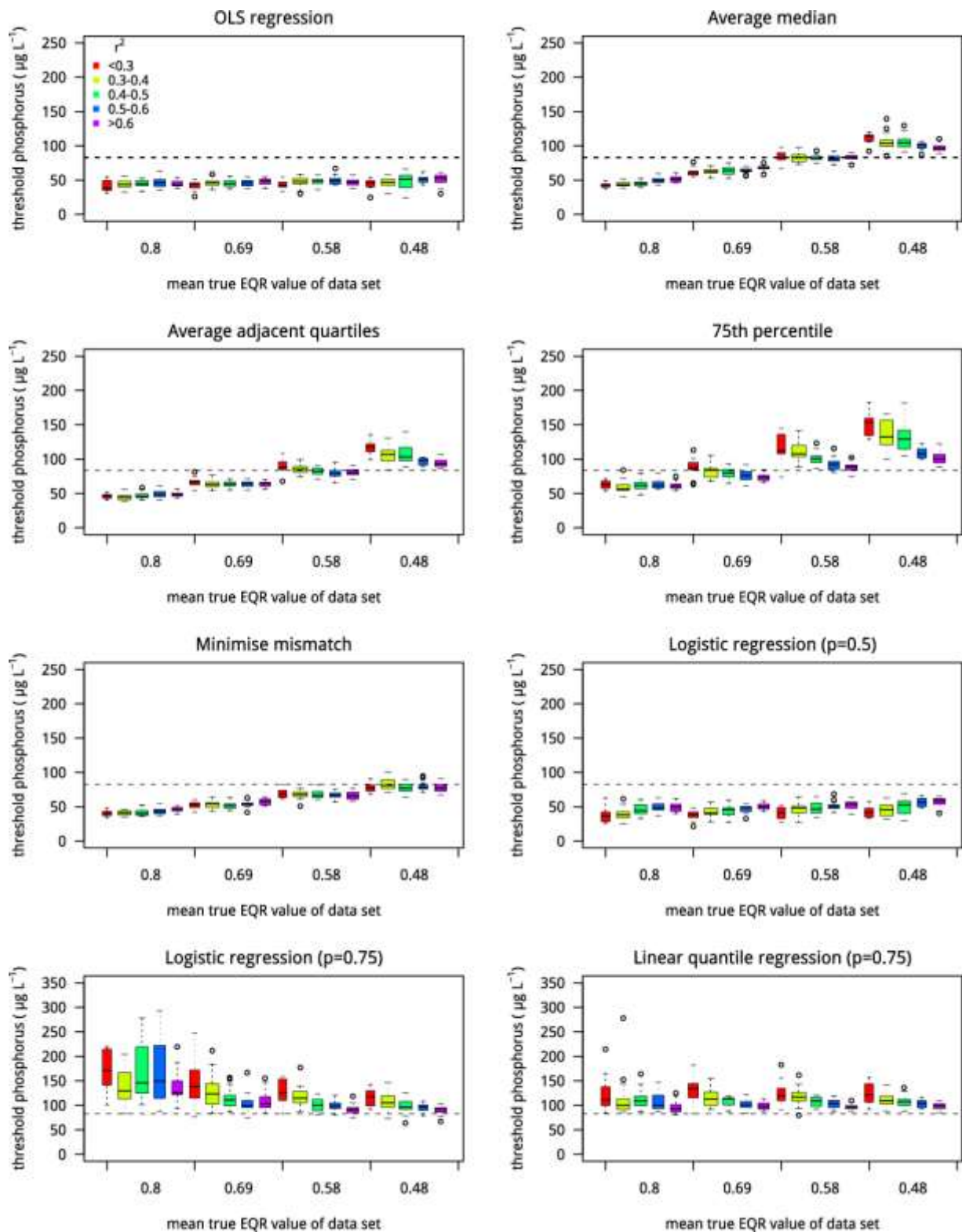
11 **Figure 3. Range of TP concentrations at the good/moderate threshold predicted in the presence of**

12 **a second pressure (“wedge-shaped” data).**



13

14 **Figure 4** Range of TP concentrations at the good/moderate threshold predicted by each of the
 15 methods using simulated data with a single stressor (TP). Boxes grouped by position of the data
 16 cloud, characterised by the true EQR calculated using equation 1 from the mean TP of data set and
 17 arranged (coloured) by variability of the relationship between the simulated TP and EQR (r^2). (for
 18 clarity only 4 of the 10 different stressor levels are shown (40,60,90,130 $\mu\text{g L}^{-1}$)). Dashed line
 19 represents the true mean phosphorus threshold.



20

21 **Figure 5** Range of TP concentrations at the good/moderate threshold predicted by each of the
 22 methods using simulated data with a stressor (TP) and an additional second unknown stressor.

23 **Boxes** grouped by position of the data cloud, characterised by the true EQR calculated using

24 **equation 1 from the mean TP of data set and arranged (coloured) by variability of the relationship**
25 **between the simulated TP and EQR (r^2). (for clarity only 4 of the 10 different pressure categories**
26 **are shown (40,60,90,130 μgL^{-1})). Dashed line represents the true mean phosphorus threshold.**

27

28