Accepted Manuscript

It is recommended to use the published version for citation.

Title: Integrated assessment of ecological status and misclassification of lakes: the role of uncertainty and index combination rules

Article Type: Research Paper

Corresponding Author: Dr. Jannicke Moe, PhD

Corresponding Author's Institution: Norwegian Institute for Water Research

First Author: Jannicke Moe, PhD

Order of Authors: Jannicke Moe, PhD; Anne Lyche Solheim; Hanna Soszka; Małgorzata Gołub; Andrzej Hutorowicz; Agnieszka Kolada; Joanna Picińska-Fałtynowicz

Abstract: The European Water Framework Directive (WFD) requires that the ecological status of waterbodies is assessed using multiple biological quality elements (BQEs) that are combined into a single status class. The recommended combination rule (the "one-out, all-out" rule; OOAO) has been criticized for being unreasonably conservative and for being sensitive to uncertainty. In this study, the objective was to compare the sensitivity to uncertainty of four different combination rules: (1) OOAO, (2) OOAO with exclusion of one element (3) average and (4) weighted average. Index values for 5 BQEs (phytoplankton, phytobenthos, macrophytes, macroinvertebrates and fish) sampled from 10 lakes in the Wel River catchment in Poland were used to classify the lakes according to the OOAO and the three alternative combination rules. Based on the mean and (where possible) standard deviation of these index values, we modelled the risk of misclassification by simulating 10,000 resamples for each BQEs in each lake, classifying each resample and calculating the proportion of misclassified resamples under each combination rule. For individual BQEs, the risk of misclassification increased both with higher uncertainty and with the proximity of the index value to a class boundary. Under the OOAO rule, the risk of misclassification was more biased towards worse status ("underclassification") than towards better status. Furthermore, risk of underclassification was more affected by uncertainty under the OOAO rule compared with the alternative combination rules. This analysis has demonstrated the weaknesses associated with the OOAO rule for integration of BQEs for lake classification. However, the alternative combination rules are associated with other shortcomings, such as the need for subjective judgement, and involve a higher risk of not protecting the most sensitive BQE and thus the whole ecosystem. We recommend that future versions of instructions for WFD implementation consider alternatives to the OOAO combination rule, and provide guidelines for weighting of individual BQEs.

Response to Reviewers: All replies to the reviewers are given in a separate document.

Highlights

- The "one-out, all-out" (OOAO) rule for combining assessment results for different biological quality elements is more prone to underestimation of the "correct" ecological status than alternative combination rules

- The OOAO rule's tendency of underestimation of ecological status increases with the index values' uncertainty

- Analysis of misclassification is complicated by the fact that the rate of misclassification inherently increases with the index values' proximity to status class boundaries

1 **Title:** Integrated assessment of ecological status and misclassification of lakes: the role of

2 uncertainty and index combination rules

3

4 **Authors:**

5 S. Jannicke Moe[a,*], Anne Lyche Solheim[a], Hanna Soszka[b], Małgorzata Gołub[b],

6 Andrzej Hutorowicz[c], Agnieszka Kolada[b], Joanna Picińska-Fałtynowicz[d], Witold Białokoz[e]

7

8 [a] *Norwegian Institute for Water Research, Gaustadalléen 23, 0349 OSLO, Norway*

9 [b] *Institute of Environmental Protection, National Research Institute, Department of*

10 *Freshwater Assessment Methods and Monitoring, Kolektorska 4, 01-692 Warszawa, Poland*

11 [c] *The Stanisław Sakowicz Inland Fisheries Institute, Department of Hydrobiology,*

12 *Oczapowskiego 10, 10-719 Olsztyn, Poland*

13 [d] *Institute of Meteorology and Water Management, National Research Institute, Wrocław*

14 *Branch, Parkowa 30, 52-616 Wrocław, Poland*

15 [e] *The Stanisław Sakowicz Inland Fisheries Institute, Department of Lake Fisheries, Rajska 2,*

16 *11-500 Giżycko, Poland*

17

18 **E-mail addresses:** jmo@niva.no, als@niva.no, hasoszka@ios.edu.pl, mgolub@ios.edu.pl,

19 ahut@infish.com.pl, akolada@ios.edu.pl, joanna.faltynowicz@imgw.pl,

20 wbialokoz@infish.com.pl

21

22 *** Corresponding author:**

23 Jannicke Moe, NIVA, Gaustadalléen 21, NO-0349 Oslo, Norway

24 jmo@niva.no

25 Telephone: +47 908 98 108, Fax: +47 22 18 52 00

26

ABSTRACT

The European Water Framework Directive (WFD) requires that the ecological status of

waterbodies is assessed using multiple biological quality elements (BQEs) that are combined

into a single status class. The recommended combination rule (the "one-out, all-out" rule;

OOAO) has been criticized for being unreasonably conservative and for being sensitive to

uncertainty. In this study, the objective was to compare the sensitivity to uncertainty of four

different combination rules: (1) OOAO, (2) OOAO with exclusion of one element (3) average

and (4) weighted average. Index values for 5 BQEs (phytoplankton, phytobenthos,

macrophytes, macroinvertebrates and fish) sampled from 10 lakes in the Wel River catchment

in Poland were used to classify the lakes according to the OOAO and the three alternative

combination rules. Based on the mean and (where possible) standard deviation of these index

values, we modelled the risk of misclassification by simulating 10,000 resamples for each

BQEs in each lake, classifying each resample and calculating the proportion of misclassified

resamples under each combination rule. For individual BQEs, the risk of misclassification

increased both with higher uncertainty and with the proximity of the index value to a class

boundary. Under the OOAO rule, the risk of misclassification was more biased towards worse

status ("underclassification") than towards better status. Furthermore, risk of

underclassification was more affected by uncertainty under the OOAO rule compared with the

alternative combination rules. This analysis has demonstrated the weaknesses associated with

the OOAO rule for integration of BQEs for lake classification. However, the alternative

combination rules are associated with other shortcomings, such as the need for subjective

judgement, and involve a higher risk of not protecting the most sensitive BQE and thus the

whole ecosystem. We recommend that future versions of instructions for WFD

implementation consider alternatives to the OOAO combination rule, and provide guidelines

for weighting of individual BQEs.

53 *Keywords:*

54 Biological quality element

55 Integrated lake assessment

56 Modeling

57 Uncertainty

58 Water Framework Directive

59 Waterbody classification

60

61 *Abbreviations:*

62 BQE: biological quality element

63 EQR: ecological quality ratio

64 nEQR: normalised ecological quality ratio

65 OOAO: One-out, all-out (combination rule)

66 WFD: Water Framework Directive

67

1. Introduction

The Water Framework Directive (WFD; EC 2000) of the European Union requires that member states must assess the ecological status of their surface waterbodies, including lakes. Across Europe, WFD-compliant national classification systems have been developed and adapted for assigning waterbodies to one of five classes of ecological status (high, good, moderate, poor and bad) (Hering et al., 2010). The WFD further requires that all waterbodies obtain good ecological status by 2015, and consequently all waterbodies found to be in moderate or worse status must be restored. Moreover, the WFD states that estimates of confidence and precision attained by the monitoring system should be provided in river basin management plans (Annex V, Section 1.3.4). Since restoration measures can be expensive, the uncertainty associated with waterbody classification should be of high interest for water resource management (Højberg et al., 2007; Irvine, 2004). If a lake in good or better status is wrongly classified as having less-than-good status ("underclassified"), money may be wasted on restoration measures that were not strictly needed (Prato et al., 2014). On the other hand, if a lake in less-than-good status is wrongly classified as good or better ("overclassified"), the ecosystem quality and services may be compromised.

Classification of ecological status of lakes should be based on a set of biological quality elements (BQEs) representing main ecosystem components, i.e. (1) phytoplankton, (2) macrophytes and phytobenthos, (3) benthic invertebrate fauna (here called "macroinvertebrates") and (4) fish (WFD, Annex V, Section 1.2.2). The WFD states that the policy should be based on the precautionary principle (§11); the idea of this principle is that if at least one component of ecosystem is impaired, this indicates that something is wrong in the ecosystem (waterbody) as a whole. Moreover, the WFD requires that the ecological status

93 class for a waterbody "shall be represented by the lower of the values for the biological and

94 physico-chemical monitoring results for the relevant quality elements" (Annex V, Section

95 1.4.2 (i)). This implies that the status is determined by either the combined biological

96 monitoring result or by the physical-chemical monitoring result (the lower of the two).

97 However, the directive does not specify how to combine the values of multiple BQEs into one

98 biological monitoring result. The guidance on classification provided by the Common

99 Implementation Strategy for the WFD (EC 2005) has recommended the method known as

100 "One-out, all-out" (OOAO): the waterbody status is determined by the BQE with the worst

101 status. However, based on comparison with alternative rules for integrating BQEs, such as

102 (weighted) average, median or other weight-of-evidence approaches, several authors have

103 stated that the OOAO tend to result in a stricter classification than what seems reasonable

104 (Alahuhta et al., 2009; Borja and Rodriguez, 2010; Caroni et al., 2013; Gottardo et al., 2011;

105 Hering et al., 2010; Moss et al., 2003; Nõges et al., 2009; Nõges and Nõges, 2006; Prato et al.,

106 2014; Rask et al., 2010; Sutela et al., 2013; Søndergaard et al., 2005). Another concern with

107 the OOAO method is that higher uncertainty in index values tend to result in even stricter

108 classification (Caroni et al., 2013; EC (European Commission), 2005; Nõges et al., 2009;

109 Sandin, 2005).

110

111 Uncertainty in biological index values results from many sources, including natural temporal

112 and spatial variation and sampling variation (see Clarke, 2013). The quantification of sources

113 of uncertainty in index values and their significance for status classification have been

114 addressed in many studies (Carvalho et al., 2013; Clarke and Hering, 2006; Kelly et al.,

115 2009b; Thackeray et al., 2013). Nevertheless, few studies have investigated the role of joint

116 uncertainty of indices when several BQEs are integrated (but see Caroni et al., 2013). There is

117 therefore a need for more research on how the OOAO and other BQE combination rules

118 perform in waterbody classification based on real data under different levels of sampling

119 uncertainty.

120

121 In our study, we have analysed the effects of joint uncertainty for five BQEs (phytoplankton,

122 phytobenthos, macrophytes, macroinvertebrates and fish) sampled from 10 lakes in Poland.

123 The analysis was based on simulations of index values for all BQEs with three levels of

124 uncertainty (section 3.1), and application of four different combination rules (section 2.3) for

125 the resulting BQE status classes. The objective of this paper was to address the following

126 question: How does increasing levels of uncertainty affect the risk of misclassification of

127 lakes under different BQE combination rules? To answer this question, we also investigated

128 how uncertainty in index values affect the risk of misclassification at the BQE level, and how

129 this risk was transferred to the whole-lake level under the different combination rules.

130

131 **2. Materials and methods**

132

133 *2.1. Data*

134

135 The study area is the catchment of the lowland river Wel in central Poland, with a surface area

136 of 822 km2. Surface waters in the Wel catchment are affected mainly by eutrophication due to

137 agricultural runoff (app. 60% of areas of extensive agriculture in the catchment) and also by a

138 few point sources of organic pollution. Ten lakes with surface area above 0.5 km2 are located

139 in this catchment (Fig. 1, Table 1). The biological data used in this study were collected from

140 all of the ten lakes in 2009 during the Polish-Norwegian project deWELopment (Soszka,

141 2011).

142

2.2. Biological index values and classification system

In this study, each biological quality element (BQE) was represented by one index, as follows.

- Phytoplankton: Phytoplankton Metric for Polish Lakes (Hutorowicz et al., 2011).

- Phytobenthos: Diatom index for lakes (phytobenthos) (Picińska-Fałtynowicz, 2011).

- Macrophytes: Ecological State Macrophyte Index (Kolada et al., 2011)

- Macroinvertebrates: Benthic Quality Index based on Chironomid Pupal Exuviae Technique (macroinvertebrates; based on Ruse, 2010) (Gołub et al., 2011).

- Fish: Lake Fish Index N2 (Białokoz and Chybowski, 2011).

For each index, the sampling method, calculation, the responses to eutrophication pressure gradients as well as classification scheme are described in the given references. For phytoplankton and macrophytes, respectively, a full description of the national assessment methods are given in the Technical Reports from the Intercalibration phase 2 (Phillips et al., 2014; Portielje et al., 2014). Although the WFD defines phytobenthos and macrophytes as one BQE, the two organism groups are treated as two separate BQEs in this paper. The reason is that Poland, like most countries in the Central-Baltic region, has chosen to develop separate assessment methods for macrophytes and phytobenthos (Kelly et al., 2009a), and no integration rules exist at the moment (Portielje et al., 2014, Table 4.4). Moreover, changing environmental conditions may affect macrophytes and phytobenthos indices differently due to the differences in generation time and dispersal rate; therefore these organism groups may provide different information about ecosystem stability (Schneider et al., 2012).

The ecological classification system used in this study (Soszka, 2011) comprises, for each biological index, a *reference condition* representing the index value assumed for lakes undisturbed by anthropogenic impact, and *class boundaries* defining the index values on the

168 borders between the five ecological status classes (high, good, moderate, poor and bad). More

169 information on the methods used for setting reference conditions and class boundaries for the

170 Polish classification system is available in the WISER database on national assessment

171 methods (http://www.wiser.eu/results/method-database; Birk et al., 2012), for all BQEs

172 except macroinvertebrates. The full ecological classification system includes also physico-

173 chemical variables, which were not included here. For each index, as required by the WFD,

174 the ecological quality ratio (EQR) was calculated as the index value divided by the reference

175 condition value. The resulting indices in EQR scale have range 0-1 (Appendix A, Table A.1).

176 Likewise, the class boundaries for each index were converted to EQR scale (Table A.2) by

177 division by the respective reference condition value. Note that the class boundaries are non-

178 evenly spaced for all BQEs except phytoplankton (Table A.2); this is essential in comparison

179 of the BQE classifications. For example, a BQE with narrow class width for good status (e.g.

180 macrophytes, class width = 0.17) may be more susceptible to bias in the assessment of good

181 status compared to a BQE with a wider class (e.g. fish, class width = 24).

182

183 To facilitate comparison of index values for different BQEs, the EQR values (Table A.1) were

184 normalised (nEQR) by a piecewise linear transformation procedure (Caroni et al., 2013). The

185 normalisation is based on the distance from the index value (in EQR scale) to the nearest class

186 boundaries (Eq. 1):

187

$$nEQR = (EQR - lower\_EQR) * \frac{(upper\_nEQR - lower\_nEQR)}{upper\_EQR - lower\_EQR} + lower\_nEQR, \quad \text{(Eq. 1)}$$

190

191 where lower_EQR and upper_EQR are lower and upper class boundaries in EQR scale for the

192 given index (Table A.2), and lower_nEQR and upper_nEQR lower and upper class

193 boundaries in normalised EQR scale (high/good = 0.8, good/moderate = 0.6, moderate/poor= 0.4, poor/bad= 0.2). The transformation to nEQR scale ensures standard class widths and boundaries for all BQEs (see also EC (2011), Fig. 12)[1]. This way, one can infer directly from each nEQR value (Table 2) both the status class and the distance to the nearest class boundaries.

2.3. BQE combination rules

201 Four alternative rules for combining the ecological status of multiple BQEs were applied in this study (Table 2).

204 *Rule 1:"OOAO"* (One-out, all-out). The status of the lake was determined by the lowest status of all the BQEs.

206 *Rule 2: "OOAO-E"* (One-out, all-out after exclusion of one BQE). This combination rule is recommended in cases where one BQE has high variability or is for other reasons associated with low confidence (EC 2005). Here, macroinvertebrates were excluded (see section 3.1).

209 *Rule 3: "Avg"* (Average). Following the WISERBUGS method (Clarke, 2013, see section 3.1.), the status class for each BQE was converted to an integer (H=1, G=2, M=3, P=4, B=5), and the arithmetic average for all BQEs was calculated. This conversion implies that the proximity of an index value to class boundaries is ignored, which is not ideal. We nevertheless chose to base the average on integers instead of the actual nEQR values, to make our results comparable with other studies using WISERBUGS (e.g., Caroni et al., 2013; Kolada et al., 2013). If the average was halfway between two classes, it was assigned to the worse of the two classes.

---

[1] A more detailed illustration of normalisation of EQR values can be found in the European Environment Agency's Data Dictionary for Lakes: http://forum.eionet.europa.eu/nrc-eionet-freshwater/library/wise_reporting_2011/biological_reporting/biologydd_20110617jpg

217 *Rule 4: "Avg-W" (*Weighted average). The status classes were converted to numeric values as

218 for Rule 3, but the class value for each BQE was multiplied by a weight inversely related to

219 the uncertainty assumed for the BQE (see section 3.1). In this study macroinvertebrates were

220 down-weighted (weight = 10%) relative to the other BQEs (each 22.5%).

221 The notation "OOAO(-E)" will be used when the two rules OOAO and OOAO-E are

222 considered jointly, and "Avg(-W)" for the two rules Avg and Avg-W considered jointly.

223

3. Calculation: the WISERBUGS method for estimating risk of misclassification

225

226 For analysing the risk of misclassification, we adopted the WISERBUGS method (Clarke,

227 2013). The method assumes that index values used for ecological classification of a

228 waterbody follow a normal distribution that can be specified by the mean and standard

229 deviation of replicated samples. The standard deviation (SD) then represents the sampling

230 uncertainty of the index. The estimated mean and SD defines a normal probability

231 distribution, from which resamples of the index can be simulated by random drawing. This

232 method recognises that the true status class of a waterbody is unknown, but considers the

233 assessment based on the measured index value (cf. Table 2) as the "correct" class.

234 Misclassification of the simulated index values is defined as assignment to any other class

235 than the "correct" class. The risk of misclassification is thus based on the precision of the

236 index values, which is measured by standard deviation (SD), but does not consider the

237 accuracy (the proximity to the unknown correct status) (Clarke, 2013).

238

3.1. Uncertainty in index values

240

241 Following the WISERBUGS method, characterisation of the probability distribution and

242 estimation of sampling variation for indices should ideally be based on a large number of

243 properly replicated samples, which are not available in most biological studies including ours.

244 However, our aim was not to predict the exact risk of misclassification, but to compare the

245 relative risk of misclassification for different levels of uncertainty. Therefore a pragmatic

246 approach was taken: where possible, the SDs for each BQE was based on multiple samples

247 from the same lake (taken at different stations, in different seasons or by different personnel),

248 and calculated as pooled SD ("SD1") for all lakes. This uncertainty measure may be

249 considered to include spatial and/or temporal variation in addition to sampling variation. For

250 phytoplankton (SD1 = 0.056) and phytobenthos (SD1 = 0.046), the SD1 was calculated from

251 4 lakes with 2-3 stations sampled once in summer and once in autumn, respectively. For

252 macrophytes (SD1 = 0.051), the SD1 was calculated from 10 lakes surveyed by 2-3 different

253 persons, once in the peak of the growing season. The faunal indices had insufficient samples

254 for calculation of SD. However, since macroinvertebrates often had lower nEQR values that

255 the other BQEs, we were particularly interested in how the exclusion or down-weighting of

256 this BQE would affect the overall assessment and risk of misclassification. The

257 macroinvertebrate index is used in national classification but the assessment system has not

258 yet been intercalibrated with the systems of other Central-Baltic countries (Böhmer et al.,

259 2014), therefore this index was associated with lower confidence than the botanical indices.

260 To reflect this lower confidence, we chose as a pragmatic solution to assign higher uncertainty

261 for the macroinvertebrate index (SD1 = 0.10) than for the other BQEs. For fish, for simplicity,

262 the sampling uncertainty was set to the same level as the botanical elements (SD1 = 0.05).

263

264 *3.2. Resampling and probability of misclassification of BQEs and lakes*

265

266 The modelling approach in this study follows the WISERBUGS method (Clarke, 2013):

267 stochastic simulation of biological index values with sampling uncertainty, and calculation of

268 misclassification under different combination rules. Three levels of uncertainty were applied,

269 denoted SD1, SD2 and SD3. Uncertainty level SD1 corresponds to the estimated or assumed

270 SD for the respective indices, as described above. For level SD2, the SD for each BQE was

271 multiplied by $\sqrt{2}$ (i.e., the variance was doubled). In level SD3, correspondingly, the SDs

272 were multiplied by $\sqrt{3}$. We simulated 10,000 resamples for each BQE in each lake, classified

273 each resample and calculated the proportion of misclassified resamples compared with the

274 "correct" class. The simulation routine was programmed in R version 2.14.1 (R Development

275 Core Team, 2011), and can be summarised in the following steps, for each lake.

276

277 1. For each BQE and each SD level, assume that the index values follow a normal probability

278 distribution N ~ (mean, SD) defined by the mean index value for the lake (in EQR scale;

279 Table A.1) and the pooled SD (section 3.1).

280 2. For each BQE and each SD level, simulate 10 000 samples (index values) drawn randomly

281 from their respective probability distributions N ~ (mean, SD).

282 3. For each simulation, assess the status class for each BQE based on their respective index

283 values and class boundaries.

284 4. For each simulation and each BQE combination rule, assess the integrated status class for

285 the lake according to the obtained BQE status classes.

286 5. For each SD level and each combination rule, calculate the proportion of "correct"

287 classification as the number of simulations with the same class as obtained for the input data

288 with the same combination rule (Table 1) divided by the total number of simulations. The

289 remaining proportion of simulations represents the probability of misclassification (under the

290 given combination rule).

12

291

292 For example, for macrophytes in Lake Kiełpińskie, the mean EQR is 0.51 (Table A.1) and the

293 pooled SD1 is 0.051 (section 3.1). Simulation of 10 000 resamples from the normal

294 distribution N ~ (mean=0.51, SD=0.051) resulted in 60.03% resamples in high class and

295 39.96% in good class, as displayed in Fig. 2a (leftmost bar). Since the "correct" class in this

296 case is high (cf. Table 2), the probability of misclassification is 39.96% (Fig. 2b, leftmost

297 bar).

298

299 *3.3. Cross-lake comparisons of risk of misclassification*

300

301 The risk misclassification for the ten lakes combined was analysed by linear models, with the

302 response variable being the number of misclassified simulated resamples for each lake (as

303 described in section 3.2, step 5). Significant difference in misclassification among BQEs was

304 tested both with BQE as a single predictor variable (one-way ANOVA) and with SD as a

305 continuous co-variable (ANCOVA). The dataset used for this test comprised the number of

306 misclassifications in the 10 lakes x 5 BQEs x 3 SD levels (in total 150 records). Likewise,

307 difference in misclassification among combination rules was tested both with combination

308 rule as single predictor variable and with SD level as a co-variable (dataset: 10 lakes x 4

309 combination rules x 3 SD levels; in total 120 records). In addition, the number of over- and

310 underclassifications were also used as alternative response variables. Pairwise comparison of

311 the number of misclassifications between BQEs and between combination rules was

312 performed by Tukey's "honestly significant difference" method (using the R function

313 "TukeyHSD").

314

13

315 Ecological status assessment sensu WFD should in principle be applied to waterbodies (and

316 their components), not to larger geographical levels such as catchment. Nevertheless, to

317 describe more general patterns for the whole catchment in this case study, an "aggregated

318 class" for all lakes combined was assigned to each BQE and each combination rule (i.e., the

319 class with the highest proportion of resamples for all lakes combined). For each BQE and

320 combination rule, the number of misclassifications for all lakes combined was calculated as

321 the total number of resamples for all lakes deviating from the correct "aggregated class".

322

4. Results and Discussion

324

4.1. Cross-lake patterns in status classification: effects of BQE combination rule

326

327 Averaging the status class of individual BQEs generally resulted in higher status than

328 applying the OOAO rule (nine out of ten lakes; Table 2), as expected. Different combination

329 rules for BQE classes have been explored and compared to the OOAO in several other

330 studies, such as average (Caroni et al., 2013; Nõges and Nõges, 2006; Sutela et al., 2013),

331 median (Alahuhta et al., 2009; Caroni et al., 2013; Rask et al., 2010), and weight-of-evidence

332 approaches or decision trees (Borja et al., 2009; Gottardo et al., 2011; Veríssimo et al., 2013).

333 In each case, the alternative rule has given equal or better classification than the OOAO.

334 Many of the authors have expressed concerns that the OOAO seems too conservative,

335 especially when several BQEs are used. For example, the two largest lakes in Estonia

336 (Võrtsjärv and Peipsi) both obtained moderate status, while more subjective expert-based

337 estimates suggest that the status should be good (Nõges and Nõges, 2006).

338

339   Excluding macroinvertebrates (rule OOAO-E) improved the "correct" class compared with

340   the OOAO for only two lakes (Table 2; Lake Grądy and Lake Hartowieckie). In these two

341   lakes, the macroinvertebrates had the lowest status. Correspondingly, down-weighting

342   macroinvertebrates when averaging the BQEs improved the "correct" class for only two lakes

343   (Table 2; Lake Dąbrowa Mała and Lake Grądy). For Lake Dąbrowa Mała there was large

344   disagreement among the BQEs, ranging from high to poor. Therefore, down-weighting one of

345   the two poor BQEs was sufficient to shift the "correct" class from moderate to good.

346

347   *4.2. Misclassification of individual BQEs: effects of uncertainty*

348

349   Higher SD levels generally increased the risk of misclassification at the BQE level (e.g., Lake

350   Kiełpińskie, Fig. 2b), as could be expected. However, the probability distribution across status

351   classes for the simulated samples (Fig. 2a) was also determined by the proximity of the index

352   value to a class boundary. The proximity to a class boundary in normalised EQR scale can be

353   inferred from the normalised EQR values in Table 2. For Lake Zarybinek, for example, the

354   nEQR of phytoplankton and macrophytes (0.39 and 0.37, respectively) were just below the

355   moderate/poor boundary (0.4). This was reflected in the simulated resamples and resulting

356   classification at the BQE level (Fig. 3a): the two mentioned BQEs had almost equal

357   probability of assessment to moderate or poor class. Consequently, the probability of

358   misclassification (Fig. 3b) was high (>40%) for these BQEs. For these BQEs where

359   misclassification was already high due to the proximity to a class boundary, higher SD

360   typically increased this risk only slightly.

361

362   The importance of uncertainty in index values for risk of misclassification at BQE level, as

363   demonstrated here, has also been clearly demonstrated in previous studies (Caroni et al.,

15

364 2013; Clarke et al., 2006; Kelly et al., 2009b; Ruse, 2010; Szoszkiewicz et al., 2007). The

365 estimated or assumed sampling uncertainty for the BQEs in this study (approx. 0.05 - 0.10)

366 were based on few samples, but correspond well to SD levels estimated for index values in

367 other studies (with similar index scale; 0-1). Examples include invertebrates in rivers (SD

368 0.058-0.065; Clarke et al., 2006), invertebrates in lakes (SD 0.032-0.094; Caroni et al., 2013),

369 diatoms in rivers and lakes (temporal variation; SD approx. 0-0.1; Kelly et al., 2009b). The

370 SD levels in this study can therefore be considered to be within a realistic range for sampling

371 uncertainty. For monitoring and classification in practice, index values will also be affected

372 by other sources of uncertainty (e.g. natural temporal variation in the ecosystem). The higher

373 levels of uncertainty used in the simulations (SD2 up to 0.20) might be considered a

374 conservative estimate of other uncertainty sources as well.

375

376 The importance of an index value's proximity to class boundaries for the risk of

377 misclassification of the BQE has also been demonstrated in numerous other studies

378 (Carstensen, 2007; Clarke and Hering, 2006; Kelly et al., 2009b; Kolada et al., 2013; Ruse,

379 2010; Szoszkiewicz et al., 2007). However, although the proximity to a class boundary

380 represents a source of uncertainty for the classification, this factor is not an error that can be

381 reduced. Thus, instead of defining the status as one class (e.g. poor for phytoplankton in Lake

382 Zarybinek, Fig. 3a), one might consider the proximity to class boundaries and describe the

383 status as "poor-to-moderate", or in probabilistic terms (e.g. 60 % poor and 40% moderate). A

384 more advanced approach - a fuzzy inference system - was used by Gottardo et al. (2011): they

385 considered also uncertainty in the class boundaries and assigned the membership of each

386 index to two neighbouring classes, expressed by percentages. If the status of a BQE is defined

387 as belonging to two classes in such a probabilistic way, the very concept of misclassification

388 should be reconsidered.

389

390   *4.3. Misclassification of individual lakes: effects of BQE combination rule and uncertainty*

391

392   As for individual BQEs, the misclassification of whole lakes increased with the level of SD.

393   However, the effects of SD for individual lakes were confounded by the effects of proximity

394   of index values to class boundaries. For Lake Kiełpińskie, the pattern of misclassification

395   under the OOAO(-E) rules (Fig. 2d) reflected the pattern of the worst BQE (macrophytes and

396   macroinvertebrates; Fig. 2b). Under the Avg(-W) rules, in contrast, the risk of

397   misclassification was very low (Fig. 2d), reflecting the fact that most index values were far

398   from the class boundaries (cf. Fig. 2a and Table 2). For Lake Zarybinek, in comparison, the

399   Avg(-W) rules resulted in a high degree of overclassification compared with the OOAO(-E)

400   rules (Fig. 3d); this reflects that three of the BQEs were close to an upper class boundary (as

401   described above; Fig 3c). For Lake Dąbrowa Mała, where several BQEs had almost equal

402   probability of two neighbouring classes (Fig. 4a), and thus high risk of misclassification (Fig.

403   4b), the "correct" lake class was altered by changed weighting in the combination rule (Fig.

404   4c). In this case, the large difference in risk of misclassification for Avg-W vs. Avg (Fig. 4d)

405   was due to the shift in the "correct" class.

406

407   The currently recommended modification of the OOAO rule - excluding the BQE with

408   highest uncertainty (EC 2005) - will automatically reduce the risk of underclassification.

409   However, there is also a risk that the excluded BQE actually is the most vulnerable

410   component, and that excluding this element will result in e.g. good status when moderate

411   status would be more appropriate, and therefore will fail to protect this BQE. More generally,

412   from a scientific point of view, discarding available information (even if uncertain) is not the

413   best means for obtaining a more reliable result. Moreover, if one BQE is routinely excluded

414   from status classification, it is more likely that it will eventually be excluded from monitoring

17

415 programmes (Søndergaard et al., 2005); this loss of information may in the long run increase

416 the risk of inappropriate management decisions.

417

418 The average combination rules may seem favourable from a statistical point of view, because

419 they make better use of all available information, and give more robust and balanced results

420 also under high uncertainty. However, these combination rules will not necessarily ensure the

421 protection of the whole ecosystem, especially in cases where there is large disagreement

422 among the BQEs (e.g. Kiełpińskie, Fig. 2). Down-weighting of BQEs with low confidence

423 may reduce the risk of misclassification (Fig. 6b), but may also fail to protect the most

424 sensitive BQEs (e.g. Dąbrowa Mała, Fig. 4). Moreover, weighted average or other weight-of-

425 evidence approaches (e.g., Gottardo et al., 2011) are not straightforward to implement,

426 because the choices will need to be justified, and there is a risk that the weighting can be

427 manipulated in order to obtain desired results. Guidelines for weighting of different BQEs,

428 e.g. based on uncertainty or other measures of confidence, would therefore be useful.

429

430 *4.4. Cross-lake patterns in misclassification: effects of uncertainty under different*

431 *combination rules*

432

433 The "aggregated class" of individual BQEs for all lakes combined (see section 3.3) ranged

434 from poor to good (Fig. 5a), and was not affected by the uncertainty level in index values

435 (SD1 - SD3). Nevertheless, the uncertainty levels affected the probability distribution across

436 status classes, and hence the risk of misclassification (Fig. 5b).

437

438 The inclusion of uncertainty also revealed a more nuanced picture of the overall effects of

439 combination rules. Although exclusion of macroinvertebrates generally did not alter the

18

"correct" lake class compared to the OOAO combination rule (Table 2), it shifted the overall

aggregated distribution of simulated classes towards higher status (Fig. 5c). Consequently, the

overall rates of misclassification were slightly reduced by this modification of the

combination rule (Fig. 5d).

Analysis of among-lake variation in misclassification for individual BQEs showed significant

effects of both BQE ($F_{4,144} = 10.58$, $p < 0.001$) and SD level ($F_{1,144} = 29.23$, $p < 0.001$). The

Tukey HSD test (Fig. 6a) revealed that on average, misclassification was significantly higher

for macroinvertebrates and macrophytes than for phytobenthos and fish, with phytoplankton

in-between. The extra high SD for assumed macroinvertebrates (approximately twice as high

as for the other BQEs) did not result in a correspondingly high rate of misclassification for

this BQE; this indicates that the proximity to class boundaries is an equally important factor

for the risk of misclassification. The bias towards overclassification (Fig. 6a), especially for

macrophytes, reflects that the index values were often close to upper class boundaries.

The probability of misclassification did not differ significantly among the different

combination rules, according to the ANOVA test (Fig. 6b). However, the combination rules

influenced the numbers of under- or overclassification. The number of underclassifications

were significantly lower under the Avg(-W) rules than under the OOAO rule ($F_{3,116} = 8.68$, p

< 0.001). Conversely, the Avg rule resulted in more overclassifications than OOAO ($F_{3,116} =$

3.30, $p < 0.03$).

The effect of SD level on the number of misclassifications of whole lakes varied among the

combination rules. For the two OOAO(-E) rules, the rate of misclassification increased

significantly with SD ($F_{1,58} = 8.61$, $p < 0.005$). In contrast, under Avg(-W), misclassification

465 was not significantly affected by SD ($F_{1,58} = 0.148$, $p < 0.70$). Correspondingly, the number of

466 underclassifications increased with SD under the OOAO(-E) rules ($F_{1,58} = 3.31$, $p = 0.062$),

467 but was not affected by SD under the Avg(-W) rules ($F_{1,58} = 1.60$, $p = 0.21$). The number of

468 overclassifications was not affected by SD (both $F_{1,58} < 0.73$, $p > 0.39$).

469

470 A similar pattern was found by Caroni et al. (2013), using the WISERBUGS simulation

471 approach for Swedish lakes with data on 2-4 BQEs with SD ranging from 0.00001 to 0.25:

472 The proportion of misclassifications, as well as the bias towards "underclassification",

473 increased more with SD when BQEs were combined by OOAO than when BQE classes were

474 averaged.

475

476 In conclusion, three tendencies can be inferred from the aggregated distribution of status of all

477 lakes (Fig. 5) and from the statistical testing of percentage misclassification among lakes.

478 First, the total number of misclassifications is slightly higher under the OOAO rule than under

479 the other three combination rules. Second, under the OOAO there are considerably more

480 underclassifications than overclassifications; under the other combination rules these two

481 types of misclassifications are more balanced. Third, higher uncertainty (SD) increases the

482 percentage of misclassification, and especially the percentage of underclassification, more

483 under the OOAO than under the other rules. In other words, because the OOAO rule never

484 gives "the benefit of the doubt", higher levels of doubt (uncertainty) will generally lead to

485 stricter assessments.

486

487 *4.5. Implications for water management policy*

488

20

489   The most critical outcome of status classification may be whether a waterbody fails to meet

490   the WFD objective of good ecological status, and therefore will need restoration measures.

491   Classification with the OOAO rule resulted in moderate or worse status for all lakes in the

492   Wel catchment; accordingly all of these lakes need restoration. The average combination rules

493   improved the status from moderate to good in two cases; the weighted average even improved

494   the status from poor to good in one case. In such situations, the choice of a strict combination

495   rule such as the OOAO can determine management decisions in favour lake restoration, and

496   therefore cause considerable economic costs (Prato et al., 2014). Conversely, selecting a more

497   liberal average-based combination rule e.g. for Kiełpińskie would imply that lake restoration

498   is not needed, even though the moderate status of two BQEs indicated that improvement was

499   needed in this case.

500

501 For lake management in practice, the quantification of uncertainty of index values (as

502 required by the WFD) may be difficult, and estimation of the risk of misclassification will

503 therefore be a challenge. Based on this study, it is not possible to conclude for a given lake

504 that an average-based combination rule will give higher or lower risk of misclassification than

505 the OOAO rule. Nevertheless, one can generally expect that the risk of misclassification will

506 be more affected by uncertainty in index values if the OOAO combination rule is used

507 compared with an average-based rule. Moreover, using the OOAO, one can expect a higher

508 risk of underclassification compared with overclassification if the uncertainty is high.

509

510 The OOAO rule for classification of waterbodies was recommended by the EC (2005) as a

511 means implementing the precautionary principle and protecting the whole ecosystem.

512 Moreover, the biological indices based on different taxonomic groups may indicate

513 anthropogenic pressures of different types (e.g., nutrient enrichment vs. habitat degradation)

21

514 or occurring at different spatial and temporal scales (e.g., local habitats vs. watershed-level)

515 (Carlisle et al., 2008; Walters et al., 2009). In this respect, the OOAO rule makes more sense

516 than other, less conservative rules.

517

518 Although the OOAO rule is simple to implement in practice, the consequences of using this

519 rule become more complicated when considering the effects of uncertainty, as shown by our

520 analysis and by Caroni et al. (2013). In future versions of guidelines for WFD

521 implementation, these findings should be considered and alternative combination rules should

522 be discussed. Like other authors (Alahuhta et al., 2009), we will not conclude by

523 recommending one particular combination rule as the most appropriate, but hope that our

524 results contribute to a better understanding of the benefits and shortcomings of different

525 combination rules when applied to different ecosystem.

526

527 We support the statements that more research is needed on combination rules for integrated

528 waterbody assessment under uncertainty (Caroni et al., 2013; Nõges et al., 2009). The existing

529 datasets in this and other cited studies provide opportunities for more investigation using

530 simulation approaches such as WISERBUGS, e.g. with different criteria for weighting BQEs

531 and different uncertainty levels under alternative combination rules. The aim should be to

532 obtain a combination rule that ensures the protection of the whole ecosystem elements under

533 pressure, while the risk of underclassification and "false alarm" for restoration is acceptable

534 for waterbody management in practice.

535

536 **Acknowledgments**

537

22

Supplementary data

Supplementary information associated with this article can be found in the online version.
Figs. S1-7: Uncertainty in classification for each of the remaining seven lakes not shown in
Figs. 2-4.

References

Alahuhta, J., Vuori, K.-M., Hellsten, S., Järvinen, M., Olin, M., Rask, M., Palomäki, A.,

2009. Defining the ecological status of small forest lakes using multiple biological quality

elements and palaeolimnological analysis. Fundamental and Applied Limnology / Archiv für

Hydrobiologie 175, 203-216.

562 Białokoz, W., Chybowski, Ł., 2011. 5. Ichtyofauna, in: Soszka, H. (Ed.), Ecological status

563 assessment of the waters in the Wel River catchment. Guidelines for integrated assessment of

564 ecological status of rivers and lakes to support river basin management plans. Institute of

565 Inland Fisheries, Olsztyn, Poland, pp. 217-234 (in Polish with English summary).

566 Birk, S., Bonne, W., Borja, A., Brucet, S., Courrat, A., Poikane, S., Solimini, A., van de

567 Bund, W., Zampoukas, N., Hering, D., 2012. Three hundred ways to assess Europe's surface

568 waters: An almost complete overview of biological methods to implement the Water

569 Framework Directive. Ecological Indicators 18, 31-41.

570 Borja, A., Bald, J., Franco, J., Larreta, J., Muxika, I., Revilla, M., Rodríguez, J.G., Solaun, O.,

571 Uriarte, A., Valencia, V., 2009. Using multiple ecosystem components, in assessing

572 ecological status in Spanish (Basque Country) Atlantic marine waters. Mar Pollut Bull 59, 54-

573 64.

574 Borja, A., Rodriguez, J.G., 2010. Problems associated with the 'one-out, all-out' principle,

575 when using multiple ecosystem components in assessing the ecological status of marine

576 waters. Mar Pollut Bull 60, 1143-1146.

577 Böhmer, J., Arbaciauskas, K., Benstead, R., Wim Gabriels, Porst, G., Reeze, B., Timm, H.,

578 2014. Central Baltic Lake Benthic invertebrate ecological assessment methods, in: Poikane, S.

579 (Ed.), Water Framework Directive Intercalibration Technical Report. European Commission

580 Joint Research Centre.

581 Carlisle, D.M., Hawkins, C.P., Meador, M.R., Potapova, M., Falcone, J., 2008. Biological

582 assessment of Appalachian streams based on predictive models for fish, macroinvertebrate,

583 and diatom assemblages. Journal of North American Benthological Society 27, 16-37.

584 Caroni, R., Bund, W., Clarke, R.T., Johnson, R.K., 2013. Combination of multiple biological

585 quality elements into waterbody assessment of surface waters. Hydrobiologia 704, 437-451.

586 Carstensen, J., 2007. Statistical principles for ecological status classification of Water

587 Framework Directive monitoring data. Mar Pollut Bull 55, 3-15.

588 Carvalho, L., Poikane, S., Lyche Solheim, A., Phillips, G., Borics, G., Catalan, J., Hoyos, C.,

589 Drakare, S., Dudley, B.J., Järvinen, M., Laplace-Treyture, C., Maileht, K., McDonald, C.,

590 Mischke, U., Moe, J., Morabito, G., Nõges, P., Nõges, T., Ott, I., Pasztaleniec, A., Skjelbred,

591 B., Thackeray, S.J., 2013. Strength and uncertainty of phytoplankton metrics for assessing

592 eutrophication impacts in lakes. Hydrobiologia 704, 127-140.

593 Clarke, R.T., 2013. Estimating confidence of European WFD ecological status class and

594 WISER Bioassessment Uncertainty Guidance Software (WISERBUGS). Hydrobiologia 704,

595 39-56.

596 Clarke, R.T., Hering, D., 2006. Errors and uncertainty in bioassessment methods – major

597 results and conclusions from the STAR project and their application using STARBUGS.

598 Hydrobiologia 566, 433-439.

599 Clarke, R.T., Lorenz, A., Sandin, L., Schmidt-Kloiber, A., Strackbein, J., Kneebone, N.T.,

600 Haase, P., 2006. Effects of sampling and sub-sampling variation using the STAR-AQEM

601 sampling protocol on the precision of macroinvertebrate metrics. Hydrobiologia 566, 441-

602 459.

603 EC (European Commission), 2000. Directive 2000/60/EC of the European Parliament and of

604 the council of 23 October 2000 establishing a framework for Community action in the field of

605 water policy. , Official Journal of the European Communities, Luxembourg.

606 EC (European Commission), 2005. Overall approach to the classification of ecological status

607 and ecological potential, Common Implementation Strategy for the Water Framework

608 Directive (2000/60/CE). Office for Official Publications of the European Communities,

609 Luxembourg.

610 EC (European Commission), 2011. Guidance document on the intercalibration process 2008-

611 2011. Technical Report - 2011 - 045, Common Implementation Strategy for the Water

612 Framework Directive (2000/60/CE). Office for Official Publications of the European

613 Communities, Luxembourg.

614 Gołub, M., Prus, P., Koszałka, J., 2011. 4. Macrozoobenthos, in: Soszka, H. (Ed.), Ecological

615 status assessment of the waters in the Wel River catchment. Guidelines for integrated

616 assessment of ecological status of rivers and lakes to support river basin management plans.

617 Institute of Inland Fisheries, Olsztyn, Poland, pp. 203-216 (in Polish with English summary).

618 Gottardo, S., Semenzin, E., Giove, S., Zabeo, A., Critto, A., de Zwart, D., Ginebreda, A.,

619 Marcomini, A., 2011. Integrated risk assessment for WFD ecological status classification

620 applied to Llobregat river basin (Spain). Part I—Fuzzy approach to aggregate biological

621 indicators. Science of The Total Environment 409, 4701-4712.

622 Hering, D., Borja, A., Carstensen, J., Carvalho, L., Elliott, M., Feld, C.K., Heiskanen, A.-S.,

623 Johnson, R.K., Moe, J., Pont, D., Solheim, A.L., de Bund, W.v., 2010. The European Water

624 Framework Directive at the age of 10: A critical review of the achievements with

625 recommendations for the future. Science of The Total Environment 408, 4007-4019.

626 Hutorowicz, A., Napiórkowska-Krzebietke, A., Pasztaleniec, A., Hutorowicz, J., Lyche

627 Solheim, A., Skjelbred, B., 2011. 1. Phytoplankton, in: Soszka, H. (Ed.), Ecological status

628 assessment of the waters in the Wel River catchment. Guidelines for integrated assessment of

629 ecological status of rivers and lakes to support river basin management plans. Institute of

630 Inland Fisheries, Olsztyn, Poland, pp. 143-168 (in Polish with English summary).

631 Højberg, A.L., Refsgaard, J.C., Geer, F., Jørgensen, L.F., Zsuffa, I., 2007. Use of models to

632 support the monitoring requirements in the Water Framework Directive. Water Resour

633 Manage 21, 1649-1672.

634 Irvine, K., 2004. Classifying ecological status under the European Water Framework

635 Directive: the need for monitoring to account for natural variability. Aquatic Conservation:

636 Marine and Freshwater Ecosystems 14, 107-112.

637 Kelly, M., Bennett, C., Coste, M., Delgado, C., Delmas, F., Denys, L., Ector, L., Fauville, C.,

638 Ferréol, M., Golub, M., Jarlman, A., Kahlert, M., Lucey, J., Ní Chatháin, B., Pardo, I., Pfister,

639 P., Picinska-Faltynowicz, J., Rosebery, J., Schranz, C., Schaumburg, J., van Dam, H.,

640 Vilbaste, S., 2009a. A comparison of national approaches to setting ecological status

641 boundaries in phytobenthos assessment for the European Water Framework Directive: results

642 of an intercalibration exercise. Hydrobiologia 621, 169-182.

643 Kelly, M., Bennion, H., Burgess, A., Ellis, J., Juggins, S., Guthrie, R., Jamieson, J.,

644 Adriaenssens, V., Yallop, M., 2009b. Uncertainty in ecological status assessments of lakes

645 and rivers using diatoms. Hydrobiologia 633, 5-15.

646 Kolada, A., Ciecierska, H., Ruszczyńska, J., Dynowski, P., 2013. Sampling techniques and

647 inter-surveyor variability as sources of uncertainty in Polish macrophyte metric for lake

648 ecological status assessment. Hydrobiologia DOI 10.1007/s10750-013-1591-9.

649 Kolada, A., Ciecierska, H., Ruszczyńska, J., Mjelde, M., Dziedzic, J., Dynowski, P.,

650 Jasnorzewski, A., 2011. 2. Macrophytes, in: Soszka, H. (Ed.), Ecological status assessment of

651 the waters in the Wel River catchment. Guidelines for integrated assessment of ecological

652 status of rivers and lakes to support river basin management plans. Institute of Inland

653 Fisheries, Olsztyn, Poland, pp. 169-186 (in Polish with English summary).

654 Moss, B., Stephen, D., Alvarez, C., Becares, E., Bund, W.V.D., Collings, S.E., Donk, E.V.,

655 Eyto, E.D., Feldmann, T., Fernández-Aláez, C., Fernández-Aláez, M., Franken, R.J.M.,

656 García-Criado, F., Gross, E.M., Gyllström, M., Hansson, L.-A., Irvine, K., Järvalt, A., Jensen,

657 J.-P., Jeppesen, E., Kairesalo, T., Kornijów, R., Krause, T., Künnap, H., Laas, A., Lill, E.,

658 Lorens, B., Luup, H., Rosa Miracle, M., Nõges, P., Nõges, T., Nykänen, M., Ott, I., Peczula,

659 W., Peeters, E.T.H.M., Phillips, G., Romo, S., Russell, V., Salujõe, J., Scheffer, M.,

660 Siewertsen, K., Smal, H., Tesch, C., Timm, H., Tuvikene, L., Tonno, I., Virro, T., Vicente, E.,

661 Wilson, D., 2003. The determination of ecological status in shallow lakes — a tested system

662 (ECOFRAME) for implementation of the European Water Framework Directive. Aquatic

663 Conservation: Marine and Freshwater Ecosystems 13, 507-549.

664 Nõges, P., Bund, W., Cardoso, A.C., Solimini, A.G., Heiskanen, A.-S., 2009. Assessment of

665 the ecological status of European surface waters: a work in progress. Hydrobiologia 633, 197-

666 211.

667 Nõges, P., Nõges, T., 2006. Indicators and criteria to assess ecological status of the large

668 shallow temperate polymictic lakes Peipsi (Estonia/Russia) and Vortsjarv (Estonia). Boreal

669 Environ. Res. 11, 67-80.

670 Phillips, G., Free, G., Karottki, I., Laplace-Treyture, C., Maileht, K., Mischke, U., Ott, I.,

671 Pasztaleniec, A., Portielje, R., Søndergaard, M., Trodd, W., Wichelen, J.V., 2014. Central

672 Baltic Lake Phytoplankton ecological assessment methods, in: Poikane, S. (Ed.), Water

673 Framework Directive Intercalibration Technical Report. European Commission Joint

674 Research Centre.

675 Picińska-Fałtynowicz, J., 2011. 3. Phytobenthos, in: Soszka, H. (Ed.), Ecological status

676 assessment of the waters in the Wel River catchment. Guidelines for integrated assessment of

677 ecological status of rivers and lakes to support river basin management plans. Institute of

678 Inland Fisheries, Olsztyn, Poland, pp. 187-202 (in Polish with English summary).

679 Portielje, R., Bertrin, V., Denys, L., Grinberga, L., Karottki, I., Kolada, A., Krasovskienė, J.,

680 Leiputé, G., Maemets, H., Ott, I., Phillips, G., Pot, R., Schaumburg, J., Schranz, C., Soszka,

681 H., Stelzer, D., Søndergaard, M., Willby, N., 2014. Central Baltic Lake Macrophyte

682 ecological assessment methods, in: Poikane, S. (Ed.), Water Framework Directive

683 Intercalibration Technical Report. European Commission Joint Research Centre.

684 Prato, S., La Valle, P., De Luca, E., Lattanzi, L., Migliore, G., Morgana, J.G., Munari, C.,

685 Nicoletti, L., Izzo, G., Mistri, M., 2014. The "one-out, all-out" principle entails the risk of

686 imposing unnecessary restoration costs: a study case in two Mediterranean coastal lakes. Mar

687 Pollut Bull 80, 30-40.

688 R Development Core Team, 2011. R: A language and environment for statistical computing.

689 R Foundation for Statistical Computing, Vienna, Austria.

690 Rask, M., Vuori, K.-M., Hämäläinen, H., Järvinen, M., Hellsten, S., Mykrä, H., Arvola, L.,

691 Ruuhijärvi, J., Jyväsjärvi, J., Kolari, I., Olin, M., Salonen, E., Valkeajärvi, P., 2010.

692 Ecological classification of large lakes in Finland: comparison of classification approaches

693 using multiple quality elements. Hydrobiologia 660, 37-47.

694 Ruse, L., 2010. Classification of nutrient impact on lakes using the chironomid pupal exuvial

695 technique. Ecological Indicators 10, 594-601.

696 Sandin, L., 2005. Testing the EC Water Framework Directive "one-out, all-out" rule -

697 simulating different levels of assessment errors along a pollution gradient in Swedish streams.

698 Verhandlungen der Internationale Vereinigung für Theoretische und Angewandte Limnologie

699 29, 334-336.

700 Schneider, S.C., Lawniczak, A.E., Picińska-Faltynowicz, J., Szoszkiewicz, K., 2012. Do

701 macrophytes, diatoms and non-diatom benthic algae give redundant information? Results

702 from a case study in Poland. Limnologica - Ecology and Management of Inland Waters 42,

703 204-211.

704 Soszka, H., 2011. Ecological status assessment of the waters in the Wel River catchment.

705 Guidelines for integrated assessment of ecological status of rivers and lakes to support river

706 basin management plans. Institute of Inland Fisheries, Olsztyn, Poland (in Polish with English

707 summary).

708 Sutela, T., Aroviita, J., Keto, A., 2013. Assessing ecological status of regulated lakes with

709 littoral macrophyte, macroinvertebrate and fish assemblages. Ecological Indicators 24, 185-

710 192.

711 Szoszkiewicz, K., Jusik, S., Zgola, T., Czechowska, M., Hryc, B., 2007. Uncertainty of

712 macrophyte-based monitoring for different types of lowland rivers. Belgian Journal of Botany

713 140, 7-16.

714 Søndergaard, M., Jeppesen, E., Peder Jensen, J., Lildal Amsinck, S., 2005. Water Framework

715 Directive: ecological classification of Danish lakes. Journal of Applied Ecology 42, 616-629.

716 Thackeray, S.J., Nõges, P., Dunbar, M.J., Dudley, B.J., Skjelbred, B., Morabito, G., Carvalho,

717 L., Phillips, G., Mischke, U., Catalan, J., de Hoyos, C., Laplace, C., Austoni, M., Padedda,

718 B.M., Maileht, K., Pasztaleniec, A., Järvinen, M., Solheim, A.L., Clarke, R.T., 2013.

719 Quantifying uncertainties in biologically-based water quality assessment: A pan-European

720 analysis of lake phytoplankton community metrics. Ecological Indicators 29, 34-47.

721 Veríssimo, H., Lane, M., Patrício, J., Gamito, S., Marques, J.C., 2013. Trends in water quality

722 and subtidal benthic communities in a temperate estuary: Is the response to restoration efforts

723 hidden by climate variability and the Estuarine Quality Paradox? Ecological Indicators 24, 56-

724 67.

725 Walters, D.M., Roy, A.H., Leigh, D.S., 2009. Environmental indicators of macroinvertebrate

726 and fish assemblage integrity in urbanizing watersheds. Ecological Indicators 9, 1222-1233.

727

728

Appendix A

Table A.1. Mean index values in EQR scale for each BQE.

Lake name	Biological quality element (BQE)[a]				
	PP	**PB**	**MP**	**MI**	**FI**
Dąbrowa Wielka	0.42	0.70	0.67	0.53	0.78
Dąbrowa Mała	0.39	0.82	0.50	0.43	0.94
Rumian	0.31	0.70	0.39	0.38	0.61
Zarybinek	0.39	0.59	0.32	0.36	0.39
Tarczyńskie	0.12	0.56	0.33	0.09	0.08
Grądy	0.30	0.79	0.33	0.06	0.31
Lidzbarskie	0.34	0.67	0.32	0.24	0.36
Kiełpińskie	0.81	0.69	0.44	0.59	0.94
Hartowieckie	0.46	0.78	0.48	0.31	0.33
Zwiniarz	0.18	0.85	0.26	0.28	0.25

Table A.2. Class boundaries in EQR scale used for status classification and for calculation of

normalised EQR values, for each biological quality element (BQE).

736

BQE[a]	Class boundaries[b]			
	H/G	**G/M**	**M/P**	**P/B**
PP	0.8	0.6	0.4	0.2
PB	0.8	0.6	0.4	0.15
MP	0.68	0.51	0.34	0.17
MI	0.9	0.69	0.45	0.21
FI	0.69	0.45	0.25	0.1

737   [a] PP = phytoplankton, PB = phytobenthos, MP = macrophytes, MI = macroinvertebrates, FI =

738   fish

739   [b] H = high, G = good, M = moderate, P = poor, B = bad

740

Figure captions

Fig. 1. Location of the study area. Upper left panel: map of Europe with location of Poland; lower left panel: map of Poland with location of the Wel river catchment; right panel: location of ten lakes within the Wel river catchment in North-Central Poland. The numbers refer to lake names in Table 1.

Fig. 2. Uncertainty in classification of Lake Kiełpińskie for different biological quality elements (BQEs), different BQE combination rules and different uncertainty levels in BQE index values. For more information on the lake, see Table 1. The labels above the bars show the "correct class" (cf. Table 2). For abbreviations and more details, see Table 2. The distribution of status classes based on 10,000 simulated resamples (see section 3.2). (a) Percentage of status classes assessed for each BQE and for each uncertainty level (1, 2, 3). (b) Percentage of resamples of each BQE categorised as underclassification and overclassification, respectively. (c) Percentage of waterbody status classes assessed for each BQE combination rule and for each uncertainty level. (d) Percentage of waterbody status classes categorised as underclassification and overclassification, respectively.

Fig. 3. Uncertainty in classification of Lake Zarybinek for different biological quality elements (BQEs), different BQE combination rules and different uncertainty levels in BQE index values. For abbreviations and more details, see Fig. 2.

Fig. 4. Uncertainty in classification of Lake Dąbrowa Mała for different biological quality elements (BQEs), different BQE combination rules and different uncertainty levels in BQE index values. For abbreviations and more details, see Fig. 2.

Fig. 5. Summarised uncertainty in classification of all ten lakes combined for different biological quality elements (BQEs), different BQE combination rules and different uncertainty levels in BQE index values. In plots (a) and (c), the percentage of resamples for each status class is summed for all lakes. The "correct aggregated status" is the class with the highest proportion of resamples for all lakes combined, for each BQE (a) and each combination rule (c), respectively. For abbreviations and more details, see Fig. 2.

Fig. 6. Outcome of the analysis of variance (ANOVA) in misclassification for the ten lakes by different BQEs, different combination rules and different levels of uncertainty in individual BQE index values. The calculation of misclassification for each simulated resample is described in section 3.2, step 5. The displayed percentage of misclassification represents the average number of misclassified resamples for all lakes. The letters above the bars in plot (a) indicate significant differences between BQEs according to the ANOVA (see method description in secction 3.3): pairs of BQEs with significatntly different percentage of misclassification have no common letters above the bars. For abbreviations and more details, see Fig. 2.
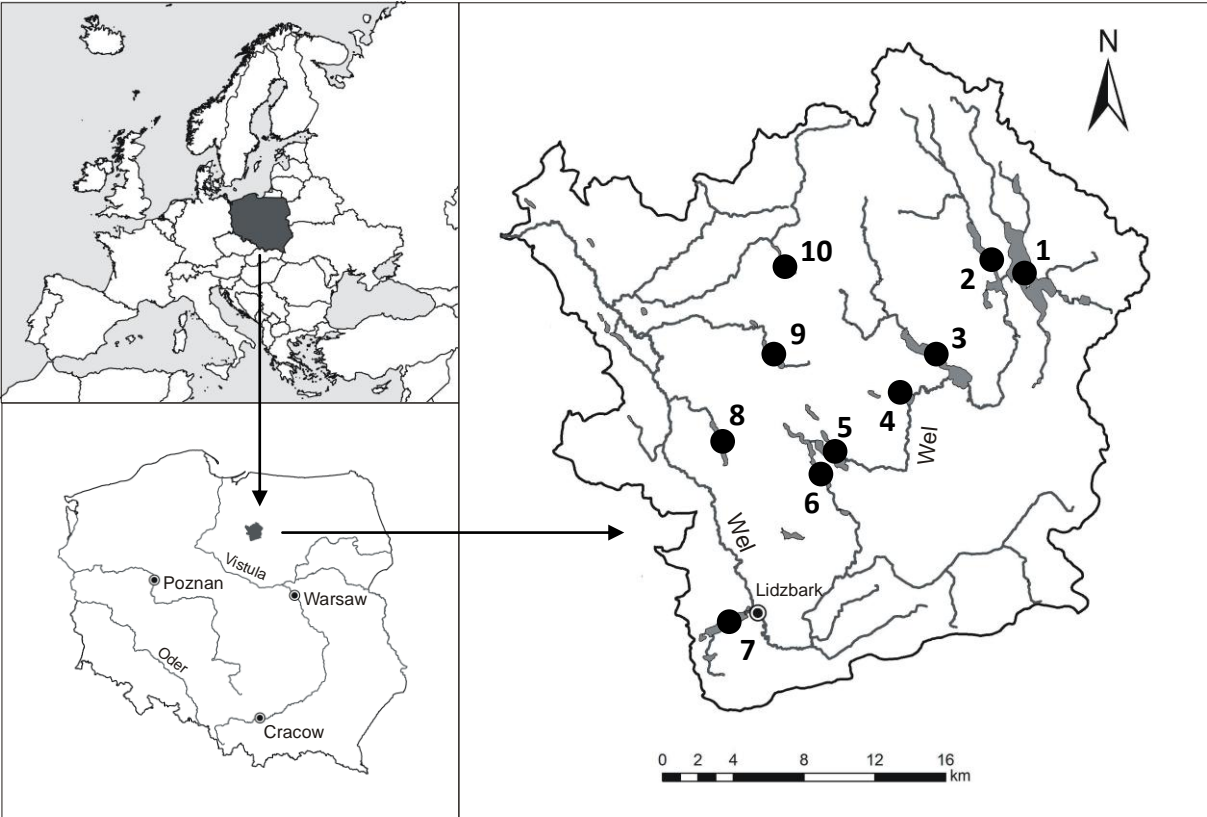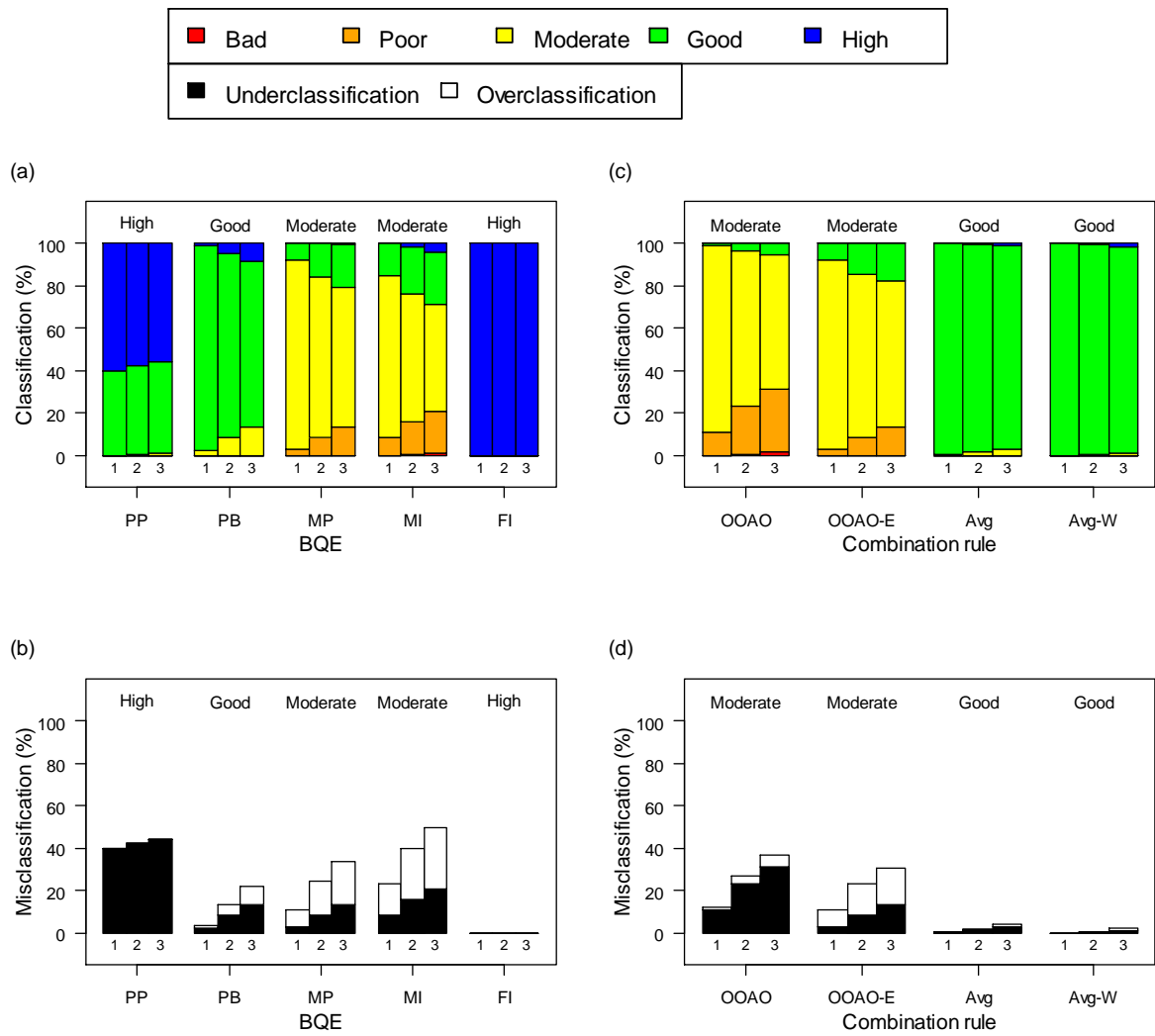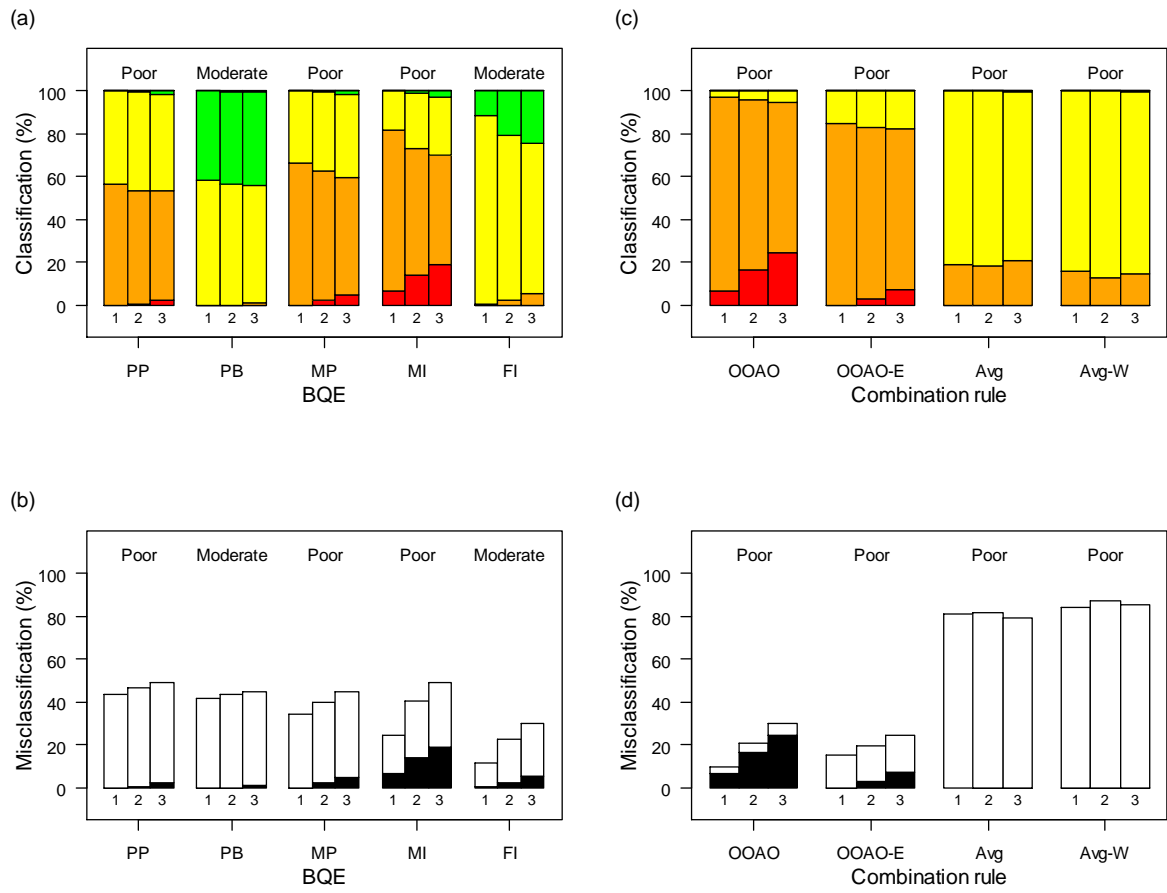
Fig. 1.

Fig. 2.

Fig. 3.

Fig. 4.

(a)

(b)

(c)

(d)

Fig. 5.

(a)

(b)

(c)

(d)

Fig. 6.

(a)

(b)

8

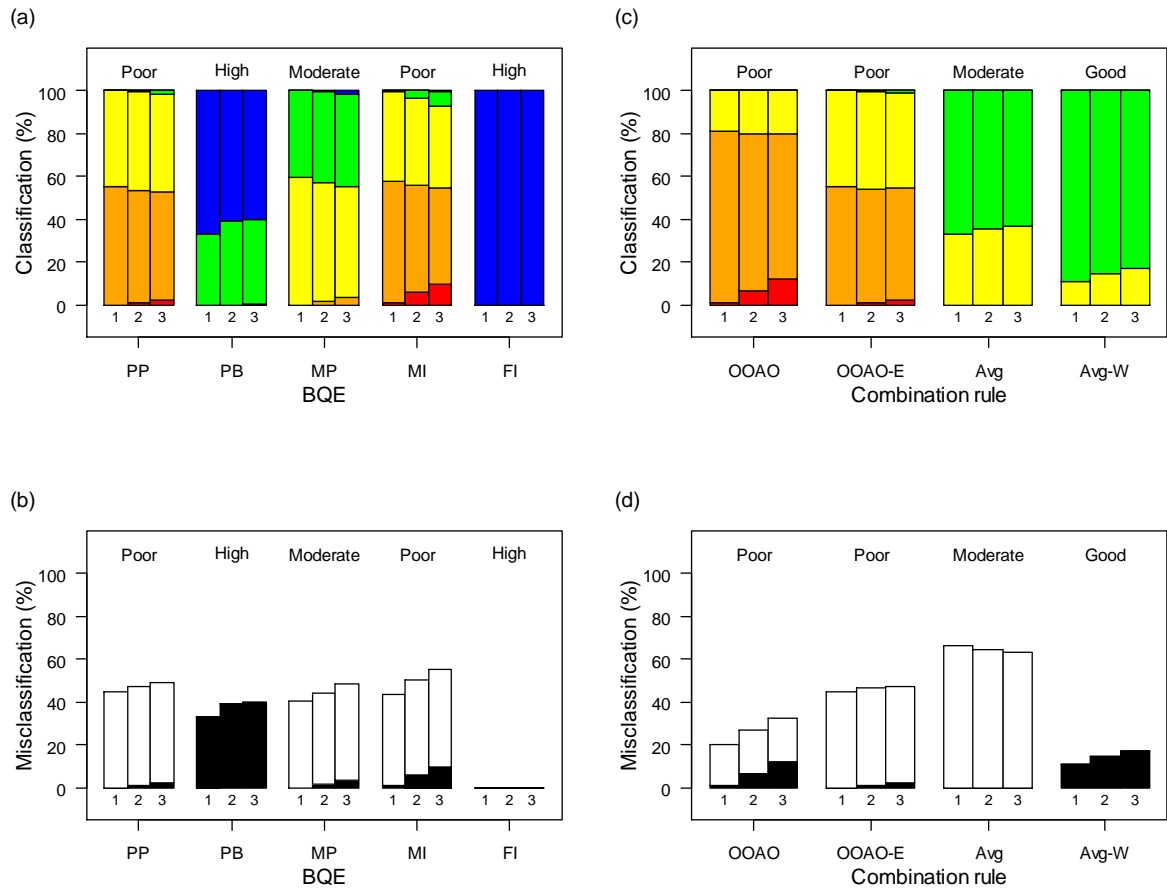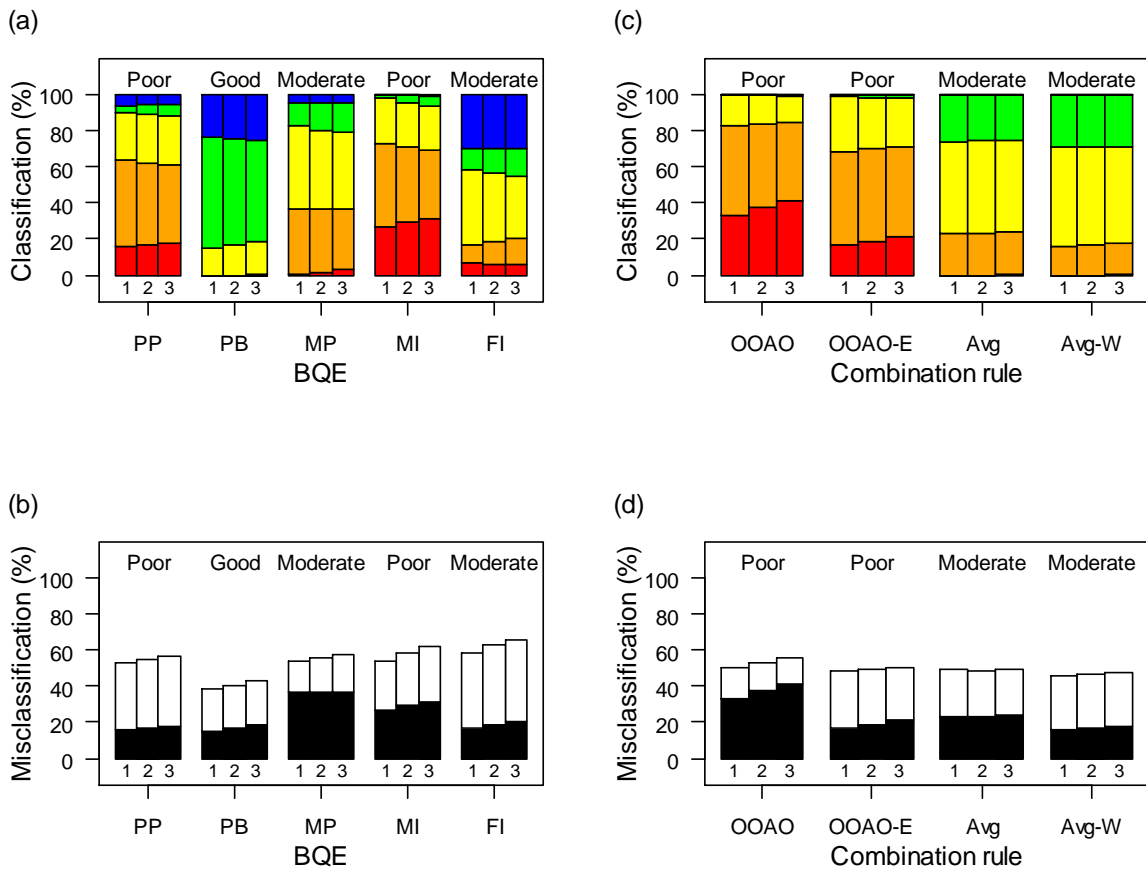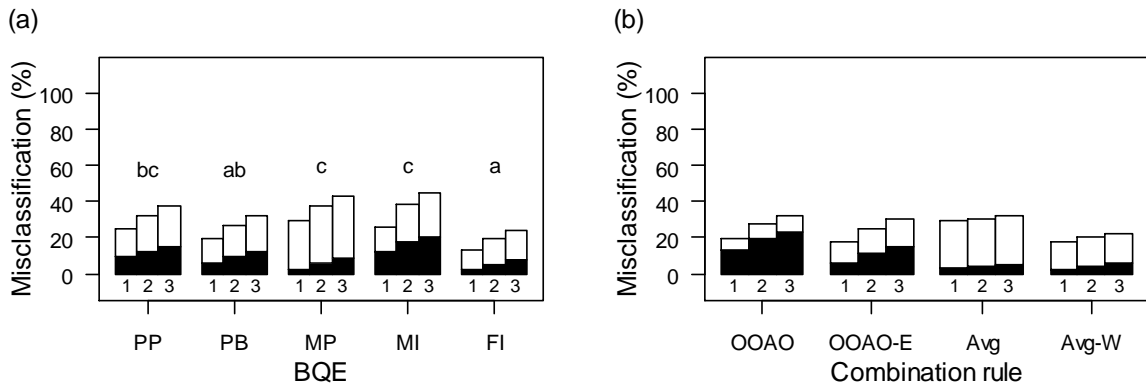**Table 1**

The main characteristics of the ten lakes used in the study. Mixing types: s = stratified, ns = non-stratified. Land use forms: NAT = natural and semi-natural; AGR = agriculture, URB = urban. After Soszka (2011). For geographic location, see Fig. 1.

No.	Lake	Surface area (km^2)	Max depth (m)	Mean depth (m)	Retention time (years)	Mixing type	Catchment area (km^2)	Land use in catchment (%)		
								NAT	AGR	URB
1	Dąbrowa Wielka	6.15	34.7	8.2	2.50	s	95.5	22.6	76.2	1.2
2	Dąbrowa Mala	1.73	34.5	10.0	0.59	s	159.1	25.2	73.8	1.0
3	Rumian	3.06	14.4	6.5	0.36	s	254.4	22.5	76.7	0.8
4	Zarybinek	0.74	7.0	2.4	0.03	ns	270.1	22.7	76.4	0.9
5	Tarczyńskie	1.64	9.2	3.8	0.10	ns	296.2	23.8	75.4	0.8
6	Grądy	1.13	9.1	4.7	0.10	ns	332.9	25.8	73.4	0.8
7	Lidzbarskie	1.22	25.5	10.1	0.12	s	534.3	29.4	69.3	1.3
8	Kiełpińskie	0.61	11.0	6.1	2.50	s	14.3	56.9	42.8	0.3
9	Hartowieckie	0.70	5.2	2.9	1.11	ns	8.7	17.9	82.1	0.0
10	Zwiniarz	0.50	5.8	3.0	1.11	ns	7.9	34.3	65.7	0.0

Table 2

Ecological status of the ten lakes (Table 1) based on data for each biological quality element (BQE) and integration of all BQEs using different combination rules (see section 2.3). The values under BQEs are normalised ecological quality ratios (nEQR; see section 2.2). Ecological status classes correspond to the following intervals of nEQR: 0-0.2 = Bad, 0.2-0.4 = Poor, 0.4-0.6 = Moderate, 0.6-0.8 = Good, 0.8-1 = High.

Lake name	Biological quality element[a]					Combination rule[b]			
	PP	PB	MP	MI	FI	OOAO	OOAO-E	Avg	Avg-W
Dąbrowa Wielka	0.42	0.70	0.79	0.47	0.86	Moderate	Moderate	Good	Good
Dąbrowa Mała	0.39	0.82	0.59	0.38	0.96	Poor	Poor	Moderate	Good
Rumian	0.31	0.70	0.46	0.34	0.73	Poor	Poor	Moderate	Moderate
Zarybinek	0.39	0.59	0.37	0.33	0.54	Poor	Poor	Poor	Poor
Tarczyńskie	0.12	0.56	0.39	0.09	0.16	Bad	Bad	Poor	Poor
Grądy	0.30	0.79	0.38	0.06	0.46	Bad	Poor	Poor	Moderate
Lidzbarskie	0.34	0.67	0.38	0.23	0.51	Poor	Poor	Moderate	Moderate
Kiełpińskie	0.83	0.69	0.51	0.52	0.96	Moderate	Moderate	Good	Good
Hartowieckie	0.46	0.78	0.56	0.28	0.48	Poor	Moderate	Moderate	Moderate
Zwiniarz	0.18	0.85	0.30	0.26	0.40	Bad	Bad	Poor	Poor

[a] PP = phytoplankton, PB = phytobenthos, MP = macrophytes, MI = macroinvertebrates, FI = fish

[b] OOAO = one-out, all-out, OOAO-E = one-out, all-out after exclusion of MI, Avg= average; Avg-W = average with down-weighting of MI
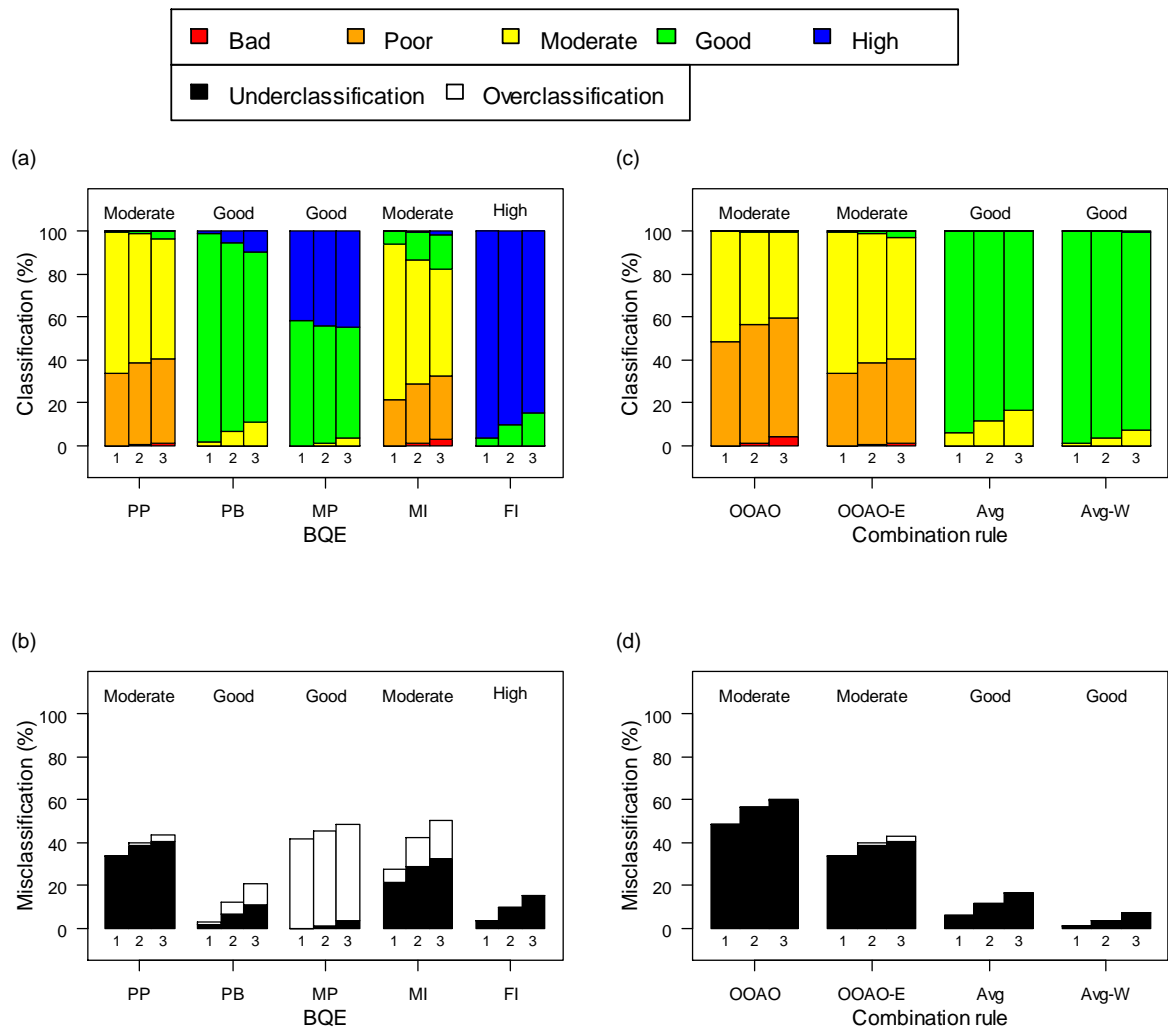
Fig. S1. Uncertainty in classification of Lake Dąbrowa Wielka for different BQEs, different BQE combination rules and different levels of uncertainty in individual BQE index values. For more details, see Fig. 2.
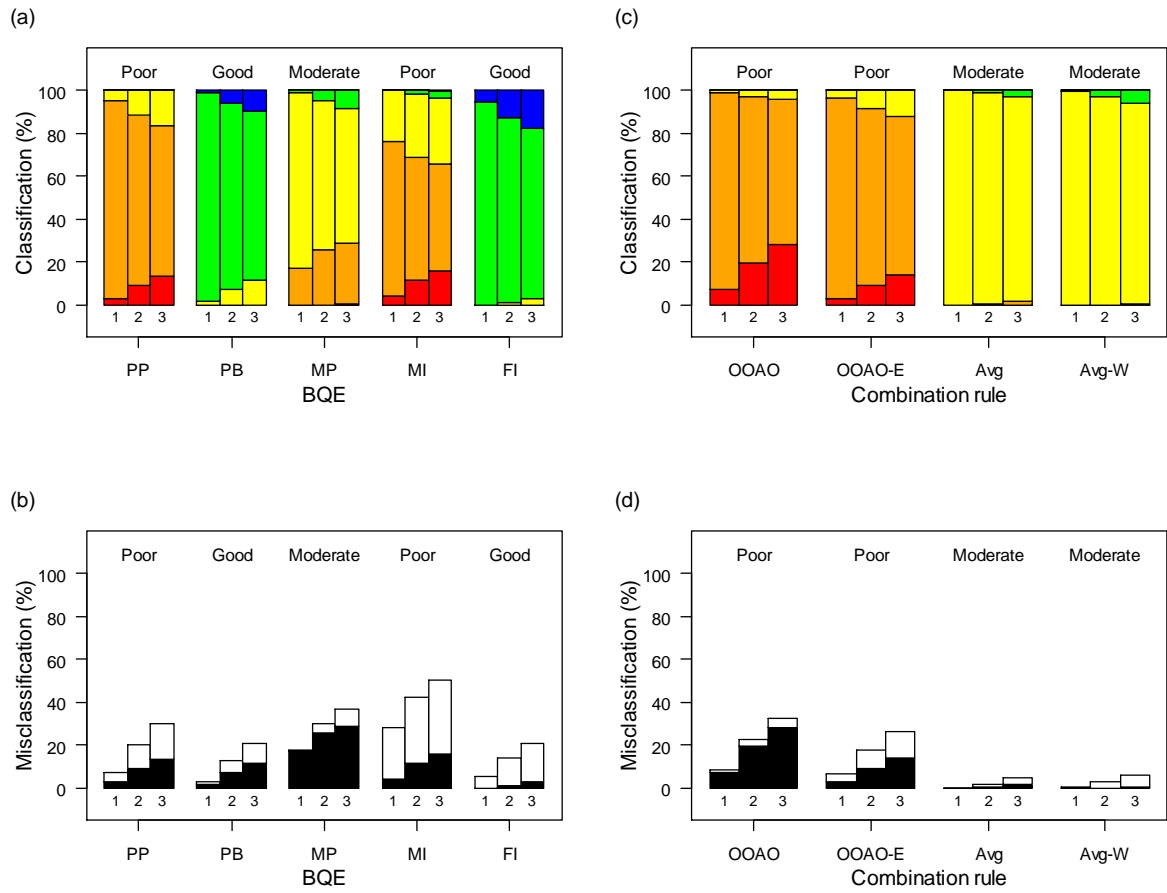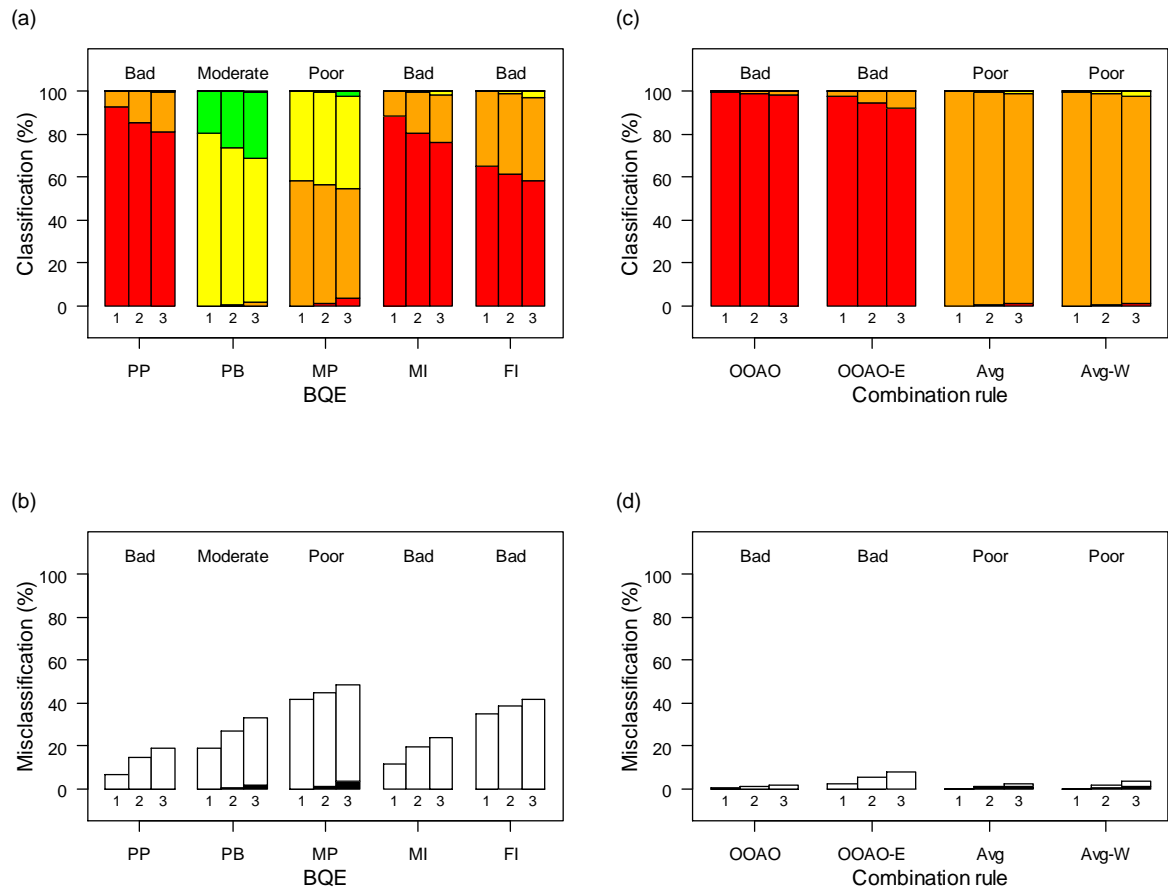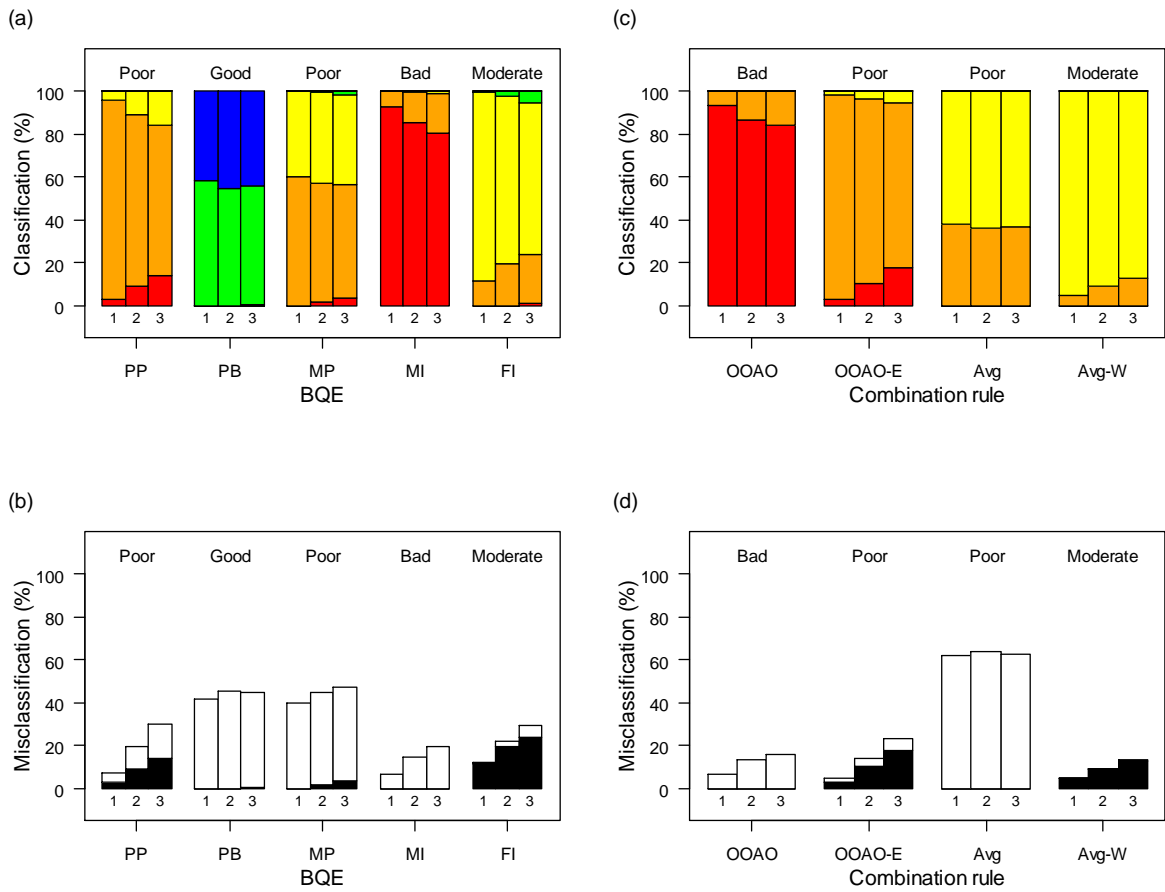
Fig. S2. Uncertainty in classification of Lake Rumian for different BQEs, different BQE combination rules and different levels of uncertainty in individual BQE index values. For more details, see Fig. 2.

Fig. S3. Uncertainty in classification of Lake Tarczyńskie for different BQEs, different BQE combination rules and different levels of uncertainty in individual BQE index values. For more details, see Fig. 2.
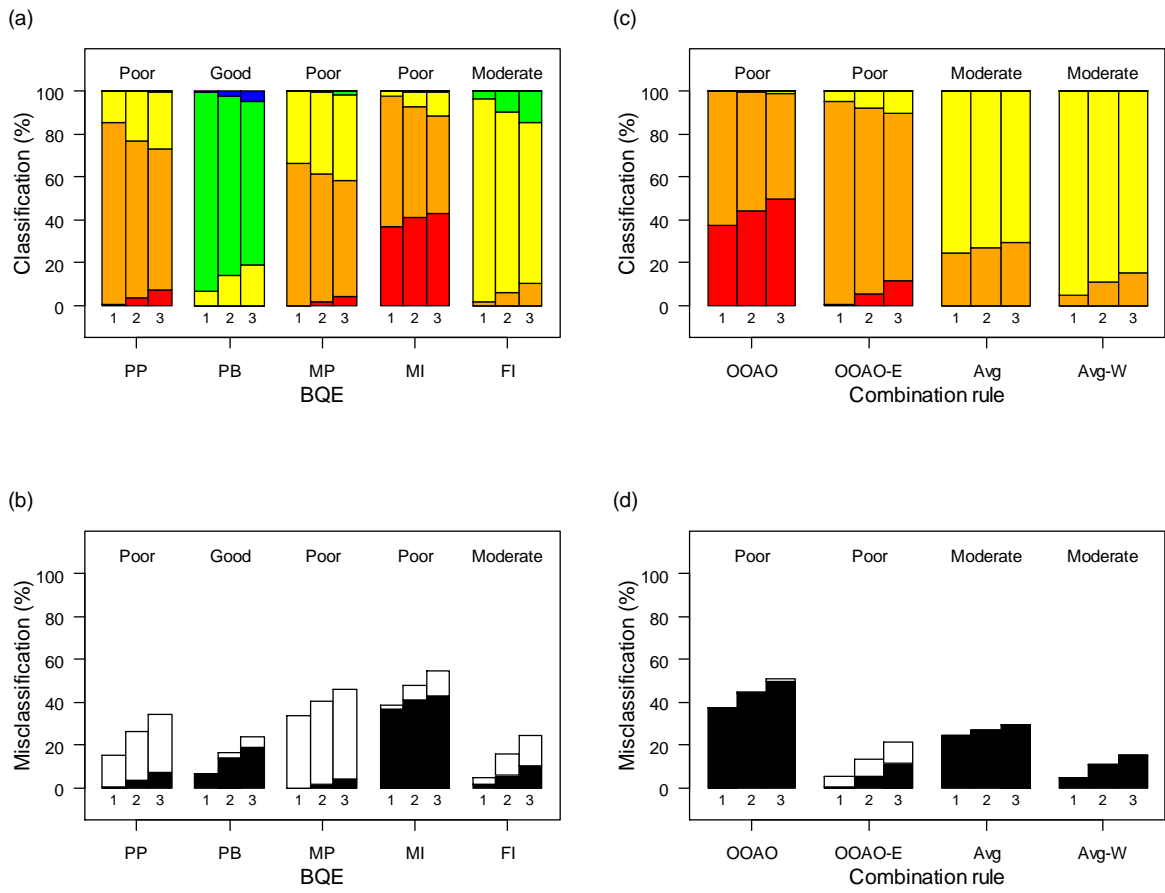
Fig. S4. Uncertainty in classification of Lake Grądy for different BQEs, different BQE combination rules and different levels of uncertainty in individual BQE index values. For more details, see Fig. 2.

Fig. S5. Uncertainty in classification of Lake Lidzbarskie for different BQEs, different BQE combination rules and different uncertainty levels in BQE index values. For more details, see Fig. 2.
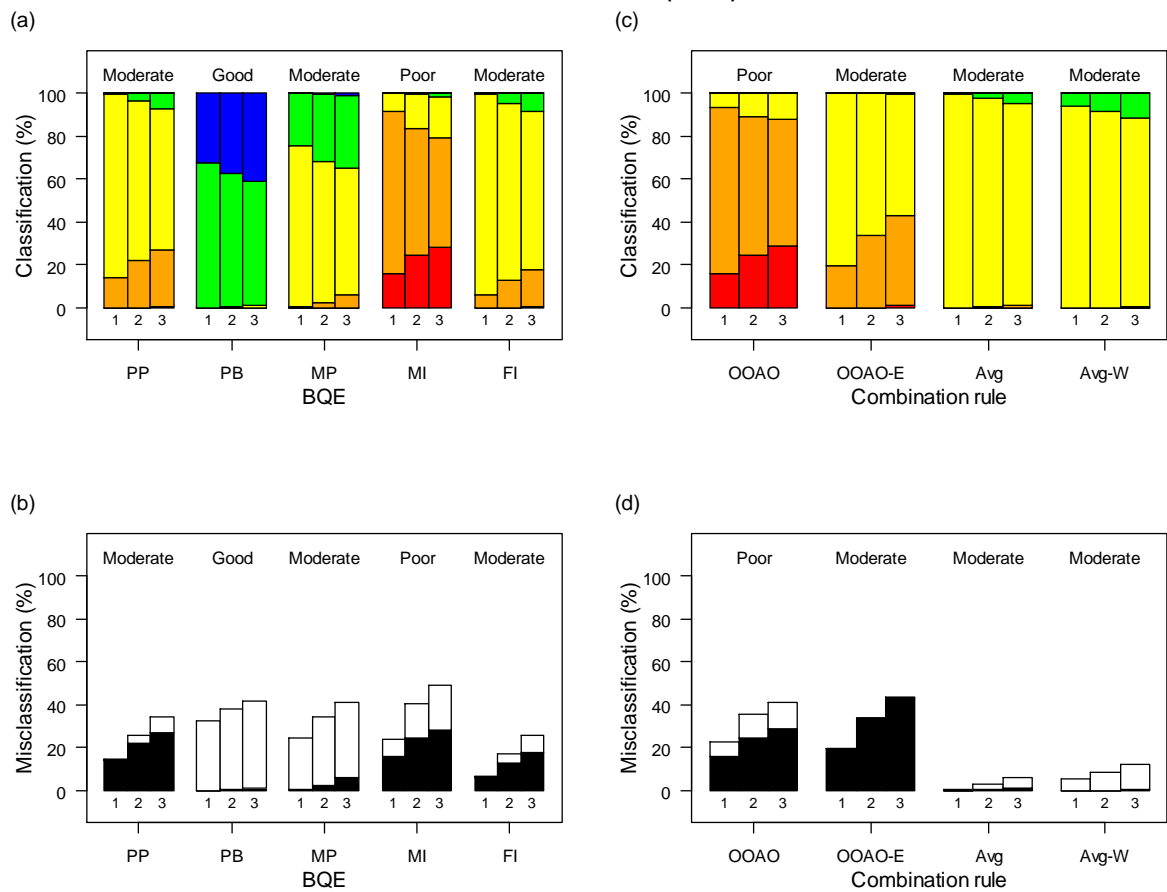
Fig. S6. Uncertainty in classification of Lake Hartowieckie for different BQEs, different BQE combination rules and different levels of uncertainty in individual BQE index values. For more details, see Fig. 2.
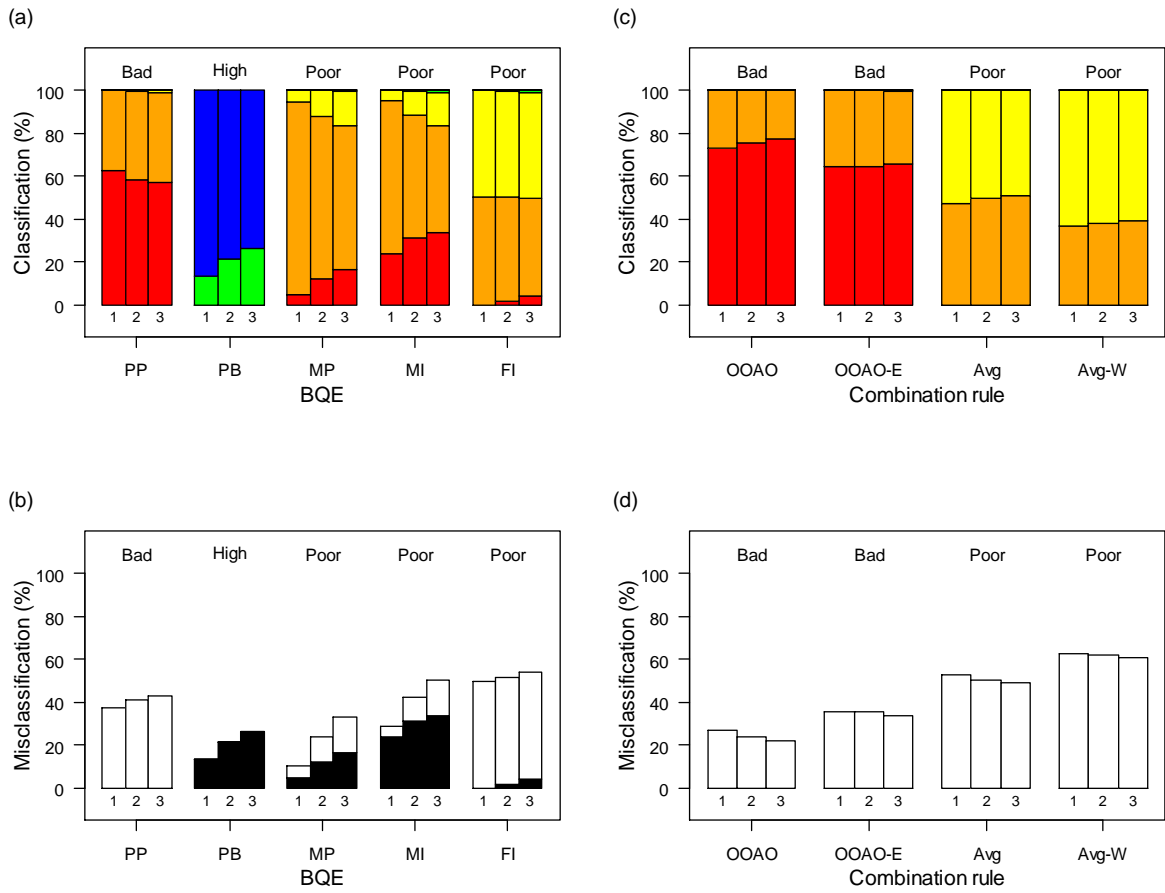
Fig. S7. Uncertainty in classification of Lake Zwiniarz for different BQEs, different BQE combination rules and different levels of uncertainty in individual BQE index values. For more details, see Fig. 2.