S. Jannicke Moe, Anne Lyche Solheim, Hanna Soszka, Małgorzata Gołub, Andrzej Hutorowicz, Agnieszka Kolada, Joanna Picińska-Fałtynowicz, Witold Białokoz. Integrated assessment of ecological status and misclassification of lakes: The role of uncertainty and index combination rules. Ecological Indicators. Volume 48, 2015, pages 605-615, ISSN 1470-160X.

The article has been published in final form by Elsevier at http://dx.doi.org/10.1016/j.ecolind.2014.08.018

Elsevier Editorial System(tm) for Ecological Indicators
Manuscript Draft

Manuscript Number: ECOLIND-3500R1

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

Article Type: Research Paper

Keywords: Biological quality element
Integrated lake assessment
Modeling
Uncertainty
Water Framework Directive
Waterbody classification

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

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

Authors:

S. Jannicke Moe[a,*], Anne Lyche Solheim[a], Hanna Soszka[b], Małgorzata Gołub[b], Andrzej Hutorowicz[c], Agnieszka Kolada[b], Joanna Picińska-Fałtynowicz[d], Witold Białokoz[e]

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

[b] *Institute of Environmental Protection, National Research Institute, Department of Freshwater Assessment Methods and Monitoring, Kolektorska 4, 01-692 Warszawa, Poland*

[c] *The Stanisław Sakowicz Inland Fisheries Institute, Department of Hydrobiology, Oczapowskiego 10, 10-719 Olsztyn, Poland*

[d] *Institute of Meteorology and Water Management, National Research Institute, Wrocław Branch, Parkowa 30, 52-616 Wrocław, Poland*

[e] *The Stanisław Sakowicz Inland Fisheries Institute, Department of Lake Fisheries, Rajska 2, 11-500 Giżycko, Poland*

E-mail addresses: jmo@niva.no, als@niva.no, hasoszka@ios.edu.pl, mgolub@ios.edu.pl, ahut@infish.com.pl, akolada@ios.edu.pl, joanna.faltynowicz@imgw.pl, wbialokoz@infish.com.pl

*** Corresponding author:**

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

jmo@niva.no

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

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.

Keywords:

Biological quality element

Integrated lake assessment

Modeling

Uncertainty

Water Framework Directive

Waterbody classification

Abbreviations:

BQE: biological quality element

EQR: ecological quality ratio

nEQR: normalised ecological quality ratio

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

WFD: Water Framework Directive

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 class for a waterbody "shall be represented by the lower of the values for the biological and physico-chemical monitoring results for the relevant quality elements" (Annex V, Section 1.4.2 (i)). This implies that the status is determined by either the combined biological monitoring result or by the physical-chemical monitoring result (the lower of the two). However, the directive does not specify how to combine the values of multiple BQEs into one biological monitoring result. The guidance on classification provided by the Common Implementation Strategy for the WFD (EC 2005) has recommended the method known as "One-out, all-out" (OOAO): the waterbody status is determined by the BQE with the worst status. However, based on comparison with alternative rules for integrating BQEs, such as (weighted) average, median or other weight-of-evidence approaches, several authors have stated that the OOAO tend to result in a stricter classification than what seems reasonable (Alahuhta et al., 2009; Borja and Rodriguez, 2010; Caroni et al., 2013; Gottardo et al., 2011; Hering et al., 2010; Moss et al., 2003; Nõges et al., 2009; Nõges and Nõges, 2006; Prato et al., 2014; Rask et al., 2010; Sutela et al., 2013; Søndergaard et al., 2005). Another concern with the OOAO method is that higher uncertainty in index values tend to result in even stricter classification (Caroni et al., 2013; EC (European Commission), 2005; Nõges et al., 2009; Sandin, 2005).

Uncertainty in biological index values results from many sources, including natural temporal and spatial variation and sampling variation (see Clarke, 2013). The quantification of sources of uncertainty in index values and their significance for status classification have been addressed in many studies (Carvalho et al., 2013; Clarke and Hering, 2006; Kelly et al., 2009b; Thackeray et al., 2013). Nevertheless, few studies have investigated the role of joint uncertainty of indices when several BQEs are integrated (but see Caroni et al., 2013). There is therefore a need for more research on how the OOAO and other BQE combination rules perform in waterbody classification based on real data under different levels of sampling uncertainty.

In our study, we have analysed the effects of joint uncertainty for five BQEs (phytoplankton, phytobenthos, macrophytes, macroinvertebrates and fish) sampled from 10 lakes in Poland. The analysis was based on simulations of index values for all BQEs with three levels of uncertainty (section 3.1), and application of four different combination rules (section 2.3) for the resulting BQE status classes. The objective of this paper was to address the following question: How does increasing levels of uncertainty affect the risk of misclassification of lakes under different BQE combination rules? To answer this question, we also investigated how uncertainty in index values affect the risk of misclassification at the BQE level, and how this risk was transferred to the whole-lake level under the different combination rules.

2. Materials and methods

2.1. Data

The study area is the catchment of the lowland river Wel in central Poland, with a surface area of 822 km2. Surface waters in the Wel catchment are affected mainly by eutrophication due to agricultural runoff (app. 60% of areas of extensive agriculture in the catchment) and also by a few point sources of organic pollution. Ten lakes with surface area above 0.5 km2 are located in this catchment (Fig. 1, Table 1). The biological data used in this study were collected from all of the ten lakes in 2009 during the Polish-Norwegian project deWELopment (Soszka, 2011).

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 borders between the five ecological status classes (high, good, moderate, poor and bad). More information on the methods used for setting reference conditions and class boundaries for the Polish classification system is available in the WISER database on national assessment methods (http://www.wiser.eu/results/method-database; Birk et al., 2012), for all BQEs except macroinvertebrates. The full ecological classification system includes also physico-chemical variables, which were not included here. For each index, as required by the WFD, the ecological quality ratio (EQR) was calculated as the index value divided by the reference condition value. The resulting indices in EQR scale have range 0-1 (Appendix A, Table A.1). Likewise, the class boundaries for each index were converted to EQR scale (Table A.2) by division by the respective reference condition value. Note that the class boundaries are non-evenly spaced for all BQEs except phytoplankton (Table A.2); this is essential in comparison of the BQE classifications. For example, a BQE with narrow class width for good status (e.g. macrophytes, class width = 0.17) may be more susceptible to bias in the assessment of good status compared to a BQE with a wider class (e.g. fish, class width = 24).

To facilitate comparison of index values for different BQEs, the EQR values (Table A.1) were normalised (nEQR) by a piecewise linear transformation procedure (Caroni et al., 2013). The normalisation is based on the distance from the index value (in EQR scale) to the nearest class boundaries (Eq. 1):

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

where lower_EQR and upper_EQR are lower and upper class boundaries in EQR scale for the given index (Table A.2), and lower_nEQR and upper_nEQR lower and upper class 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

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

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

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).

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

*Rule 4: "Avg-W" (*Weighted average). The status classes were converted to numeric values as for Rule 3, but the class value for each BQE was multiplied by a weight inversely related to the uncertainty assumed for the BQE (see section 3.1). In this study macroinvertebrates were down-weighted (weight = 10%) relative to the other BQEs (each 22.5%).

The notation "OOAO(-E)" will be used when the two rules OOAO and OOAO-E are considered jointly, and "Avg(-W)" for the two rules Avg and Avg-W considered jointly.

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

For analysing the risk of misclassification, we adopted the WISERBUGS method (Clarke, 2013). The method assumes that index values used for ecological classification of a waterbody follow a normal distribution that can be specified by the mean and standard deviation of replicated samples. The standard deviation (SD) then represents the sampling uncertainty of the index. The estimated mean and SD defines a normal probability distribution, from which resamples of the index can be simulated by random drawing. This method recognises that the true status class of a waterbody is unknown, but considers the assessment based on the measured index value (cf. Table 2) as the "correct" class. Misclassification of the simulated index values is defined as assignment to any other class than the "correct" class. The risk of misclassification is thus based on the precision of the index values, which is measured by standard deviation (SD), but does not consider the accuracy (the proximity to the unknown correct status) (Clarke, 2013).

3.1. Uncertainty in index values

Following the WISERBUGS method, characterisation of the probability distribution and estimation of sampling variation for indices should ideally be based on a large number of properly replicated samples, which are not available in most biological studies including ours. However, our aim was not to predict the exact risk of misclassification, but to compare the relative risk of misclassification for different levels of uncertainty. Therefore a pragmatic approach was taken: where possible, the SDs for each BQE was based on multiple samples from the same lake (taken at different stations, in different seasons or by different personnel), and calculated as pooled SD ("SD1") for all lakes. This uncertainty measure may be considered to include spatial and/or temporal variation in addition to sampling variation. For phytoplankton (SD1 = 0.056) and phytobenthos (SD1 = 0.046), the SD1 was calculated from 4 lakes with 2-3 stations sampled once in summer and once in autumn, respectively. For macrophytes (SD1 = 0.051), the SD1 was calculated from 10 lakes surveyed by 2-3 different persons, once in the peak of the growing season. The faunal indices had insufficient samples for calculation of SD. However, since macroinvertebrates often had lower nEQR values that the other BQEs, we were particularly interested in how the exclusion or down-weighting of this BQE would affect the overall assessment and risk of misclassification. The macroinvertebrate index is used in national classification but the assessment system has not yet been intercalibrated with the systems of other Central-Baltic countries (Böhmer et al., 2014), therefore this index was associated with lower confidence than the botanical indices. To reflect this lower confidence, we chose as a pragmatic solution to assign higher uncertainty for the macroinvertebrate index (SD1 = 0.10) than for the other BQEs. For fish, for simplicity, the sampling uncertainty was set to the same level as the botanical elements (SD1 = 0.05).

3.2. Resampling and probability of misclassification of BQEs and lakes

The modelling approach in this study follows the WISERBUGS method (Clarke, 2013): stochastic simulation of biological index values with sampling uncertainty, and calculation of misclassification under different combination rules. Three levels of uncertainty were applied, denoted SD1, SD2 and SD3. Uncertainty level SD1 corresponds to the estimated or assumed SD for the respective indices, as described above. For level SD2, the SD for each BQE was multiplied by √2 (i.e., the variance was doubled). In level SD3, correspondingly, the SDs were multiplied by √3. We simulated 10,000 resamples for each BQE in each lake, classified each resample and calculated the proportion of misclassified resamples compared with the "correct" class. The simulation routine was programmed in R version 2.14.1 (R Development Core Team, 2011), and can be summarised in the following steps, for each lake.

1. For each BQE and each SD level, assume that the index values follow a normal probability distribution N ~ (mean, SD) defined by the mean index value for the lake (in EQR scale; Table A.1) and the pooled SD (section 3.1).
2. For each BQE and each SD level, simulate 10 000 samples (index values) drawn randomly from their respective probability distributions N ~ (mean, SD).
3. For each simulation, assess the status class for each BQE based on their respective index values and class boundaries.
4. For each simulation and each BQE combination rule, assess the integrated status class for the lake according to the obtained BQE status classes.
5. For each SD level and each combination rule, calculate the proportion of "correct" classification as the number of simulations with the same class as obtained for the input data with the same combination rule (Table 1) divided by the total number of simulations. The remaining proportion of simulations represents the probability of misclassification (under the given combination rule).

For example, for macrophytes in Lake Kiełpińskie, the mean EQR is 0.51 (Table A.1) and the pooled SD1 is 0.051 (section 3.1). Simulation of 10 000 resamples from the normal distribution N ~ (mean=0.51, SD=0.051) resulted in 60.03% resamples in high class and 39.96% in good class, as displayed in Fig. 2a (leftmost bar). Since the "correct" class in this case is high (cf. Table 2), the probability of misclassification is 39.96% (Fig. 2b, leftmost bar).

3.3. Cross-lake comparisons of risk of misclassification

The risk misclassification for the ten lakes combined was analysed by linear models, with the response variable being the number of misclassified simulated resamples for each lake (as described in section 3.2, step 5). Significant difference in misclassification among BQEs was tested both with BQE as a single predictor variable (one-way ANOVA) and with SD as a continuous co-variable (ANCOVA). The dataset used for this test comprised the number of misclassifications in the 10 lakes x 5 BQEs x 3 SD levels (in total 150 records). Likewise, difference in misclassification among combination rules was tested both with combination rule as single predictor variable and with SD level as a co-variable (dataset: 10 lakes x 4 combination rules x 3 SD levels; in total 120 records). In addition, the number of over- and underclassifications were also used as alternative response variables. Pairwise comparison of the number of misclassifications between BQEs and between combination rules was performed by Tukey's "honestly significant difference" method (using the R function "TukeyHSD").

Ecological status assessment sensu WFD should in principle be applied to waterbodies (and their components), not to larger geographical levels such as catchment. Nevertheless, to describe more general patterns for the whole catchment in this case study, an "aggregated class" for all lakes combined was assigned to each BQE and each combination rule (i.e., the class with the highest proportion of resamples for all lakes combined). For each BQE and combination rule, the number of misclassifications for all lakes combined was calculated as the total number of resamples for all lakes deviating from the correct "aggregated class".

4. Results and Discussion

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

Averaging the status class of individual BQEs generally resulted in higher status than applying the OOAO rule (nine out of ten lakes; Table 2), as expected. Different combination rules for BQE classes have been explored and compared to the OOAO in several other studies, such as average (Caroni et al., 2013; Nõges and Nõges, 2006; Sutela et al., 2013), median (Alahuhta et al., 2009; Caroni et al., 2013; Rask et al., 2010), and weight-of-evidence approaches or decision trees (Borja et al., 2009; Gottardo et al., 2011; Veríssimo et al., 2013). In each case, the alternative rule has given equal or better classification than the OOAO. Many of the authors have expressed concerns that the OOAO seems too conservative, especially when several BQEs are used. For example, the two largest lakes in Estonia (Võrtsjärv and Peipsi) both obtained moderate status, while more subjective expert-based estimates suggest that the status should be good (Nõges and Nõges, 2006).

Excluding macroinvertebrates (rule OOAO-E) improved the "correct" class compared with the OOAO for only two lakes (Table 2; Lake Grądy and Lake Hartowieckie). In these two lakes, the macroinvertebrates had the lowest status. Correspondingly, down-weighting macroinvertebrates when averaging the BQEs improved the "correct" class for only two lakes (Table 2; Lake Dąbrowa Mała and Lake Grądy). For Lake Dąbrowa Mała there was large disagreement among the BQEs, ranging from high to poor. Therefore, down-weighting one of the two poor BQEs was sufficient to shift the "correct" class from moderate to good.

4.2. Misclassification of individual BQEs: effects of uncertainty

Higher SD levels generally increased the risk of misclassification at the BQE level (e.g., Lake Kiełpińskie, Fig. 2b), as could be expected. However, the probability distribution across status classes for the simulated samples (Fig. 2a) was also determined by the proximity of the index value to a class boundary. The proximity to a class boundary in normalised EQR scale can be inferred from the normalised EQR values in Table 2. For Lake Zarybinek, for example, the nEQR of phytoplankton and macrophytes (0.39 and 0.37, respectively) were just below the moderate/poor boundary (0.4). This was reflected in the simulated resamples and resulting classification at the BQE level (Fig. 3a): the two mentioned BQEs had almost equal probability of assessment to moderate or poor class. Consequently, the probability of misclassification (Fig. 3b) was high (>40%) for these BQEs. For these BQEs where misclassification was already high due to the proximity to a class boundary, higher SD typically increased this risk only slightly.

The importance of uncertainty in index values for risk of misclassification at BQE level, as demonstrated here, has also been clearly demonstrated in previous studies (Caroni et al., 2013; Clarke et al., 2006; Kelly et al., 2009b; Ruse, 2010; Szoszkiewicz et al., 2007). The estimated or assumed sampling uncertainty for the BQEs in this study (approx. 0.05 - 0.10) were based on few samples, but correspond well to SD levels estimated for index values in other studies (with similar index scale; 0-1). Examples include invertebrates in rivers (SD 0.058-0.065; Clarke et al., 2006), invertebrates in lakes (SD 0.032-0.094; Caroni et al., 2013), diatoms in rivers and lakes (temporal variation; SD approx. 0-0.1; Kelly et al., 2009b). The SD levels in this study can therefore be considered to be within a realistic range for sampling uncertainty. For monitoring and classification in practice, index values will also be affected by other sources of uncertainty (e.g. natural temporal variation in the ecosystem). The higher levels of uncertainty used in the simulations (SD2 up to 0.20) might be considered a conservative estimate of other uncertainty sources as well.

The importance of an index value's proximity to class boundaries for the risk of misclassification of the BQE has also been demonstrated in numerous other studies (Carstensen, 2007; Clarke and Hering, 2006; Kelly et al., 2009b; Kolada et al., 2013; Ruse, 2010; Szoszkiewicz et al., 2007). However, although the proximity to a class boundary represents a source of uncertainty for the classification, this factor is not an error that can be reduced. Thus, instead of defining the status as one class (e.g. poor for phytoplankton in Lake Zarybinek, Fig. 3a), one might consider the proximity to class boundaries and describe the status as "poor-to-moderate", or in probabilistic terms (e.g. 60 % poor and 40% moderate). A more advanced approach - a fuzzy inference system - was used by Gottardo et al. (2011): they considered also uncertainty in the class boundaries and assigned the membership of each index to two neighbouring classes, expressed by percentages. If the status of a BQE is defined as belonging to two classes in such a probabilistic way, the very concept of misclassification should be reconsidered.

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

As for individual BQEs, the misclassification of whole lakes increased with the level of SD. However, the effects of SD for individual lakes were confounded by the effects of proximity of index values to class boundaries. For Lake Kiełpińskie, the pattern of misclassification under the OOAO(-E) rules (Fig. 2d) reflected the pattern of the worst BQE (macrophytes and macroinvertebrates; Fig. 2b). Under the Avg(-W) rules, in contrast, the risk of misclassification was very low (Fig. 2d), reflecting the fact that most index values were far from the class boundaries (cf. Fig. 2a and Table 2). For Lake Zarybinek, in comparison, the Avg(-W) rules resulted in a high degree of overclassification compared with the OOAO(-E) rules (Fig. 3d); this reflects that three of the BQEs were close to an upper class boundary (as described above; Fig 3c). For Lake Dąbrowa Mała, where several BQEs had almost equal probability of two neighbouring classes (Fig. 4a), and thus high risk of misclassification (Fig. 4b), the "correct" lake class was altered by changed weighting in the combination rule (Fig. 4c). In this case, the large difference in risk of misclassification for Avg-W vs. Avg (Fig. 4d) was due to the shift in the "correct" class.

The currently recommended modification of the OOAO rule - excluding the BQE with highest uncertainty (EC 2005) - will automatically reduce the risk of underclassification. However, there is also a risk that the excluded BQE actually is the most vulnerable component, and that excluding this element will result in e.g. good status when moderate status would be more appropriate, and therefore will fail to protect this BQE. More generally, from a scientific point of view, discarding available information (even if uncertain) is not the best means for obtaining a more reliable result. Moreover, if one BQE is routinely excluded from status classification, it is more likely that it will eventually be excluded from monitoring programmes (Søndergaard et al., 2005); this loss of information may in the long run increase the risk of inappropriate management decisions.

The average combination rules may seem favourable from a statistical point of view, because they make better use of all available information, and give more robust and balanced results also under high uncertainty. However, these combination rules will not necessarily ensure the protection of the whole ecosystem, especially in cases where there is large disagreement among the BQEs (e.g. Kiełpińskie, Fig. 2). Down-weighting of BQEs with low confidence may reduce the risk of misclassification (Fig. 6b), but may also fail to protect the most sensitive BQEs (e.g. Dąbrowa Mała, Fig. 4). Moreover, weighted average or other weight-of-evidence approaches (e.g., Gottardo et al., 2011) are not straightforward to implement, because the choices will need to be justified, and there is a risk that the weighting can be manipulated in order to obtain desired results. Guidelines for weighting of different BQEs, e.g. based on uncertainty or other measures of confidence, would therefore be useful.

4.4. Cross-lake patterns in misclassification: effects of uncertainty under different combination rules

The "aggregated class" of individual BQEs for all lakes combined (see section 3.3) ranged from poor to good (Fig. 5a), and was not affected by the uncertainty level in index values (SD1 - SD3). Nevertheless, the uncertainty levels affected the probability distribution across status classes, and hence the risk of misclassification (Fig. 5b).

The inclusion of uncertainty also revealed a more nuanced picture of the overall effects of combination rules. Although exclusion of macroinvertebrates generally did not alter the "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 was not significantly affected by SD ($F_{1,58} = 0.148$, $p < 0.70$). Correspondingly, the number of underclassifications increased with SD under the OOAO(-E) rules ($F_{1,58} = 3.31$, $p = 0.062$), but was not affected by SD under the Avg(-W) rules ($F_{1,58} = 1.60$, $p = 0.21$). The number of overclassifications was not affected by SD (both $F_{1,58} < 0.73$, $p > 0.39$).

A similar pattern was found by Caroni et al. (2013), using the WISERBUGS simulation approach for Swedish lakes with data on 2-4 BQEs with SD ranging from 0.00001 to 0.25: The proportion of misclassifications, as well as the bias towards "underclassification", increased more with SD when BQEs were combined by OOAO than when BQE classes were averaged.

In conclusion, three tendencies can be inferred from the aggregated distribution of status of all lakes (Fig. 5) and from the statistical testing of percentage misclassification among lakes. First, the total number of misclassifications is slightly higher under the OOAO rule than under the other three combination rules. Second, under the OOAO there are considerably more underclassifications than overclassifications; under the other combination rules these two types of misclassifications are more balanced. Third, higher uncertainty (SD) increases the percentage of misclassification, and especially the percentage of underclassification, more under the OOAO than under the other rules. In other words, because the OOAO rule never gives "the benefit of the doubt", higher levels of doubt (uncertainty) will generally lead to stricter assessments.

4.5. Implications for water management policy

The most critical outcome of status classification may be whether a waterbody fails to meet the WFD objective of good ecological status, and therefore will need restoration measures. Classification with the OOAO rule resulted in moderate or worse status for all lakes in the Wel catchment; accordingly all of these lakes need restoration. The average combination rules improved the status from moderate to good in two cases; the weighted average even improved the status from poor to good in one case. In such situations, the choice of a strict combination rule such as the OOAO can determine management decisions in favour lake restoration, and therefore cause considerable economic costs (Prato et al., 2014). Conversely, selecting a more liberal average-based combination rule e.g. for Kiełpińskie would imply that lake restoration is not needed, even though the moderate status of two BQEs indicated that improvement was needed in this case.

For lake management in practice, the quantification of uncertainty of index values (as required by the WFD) may be difficult, and estimation of the risk of misclassification will therefore be a challenge. Based on this study, it is not possible to conclude for a given lake that an average-based combination rule will give higher or lower risk of misclassification than the OOAO rule. Nevertheless, one can generally expect that the risk of misclassification will be more affected by uncertainty in index values if the OOAO combination rule is used compared with an average-based rule. Moreover, using the OOAO, one can expect a higher risk of underclassification compared with overclassification if the uncertainty is high.

The OOAO rule for classification of waterbodies was recommended by the EC (2005) as a means implementing the precautionary principle and protecting the whole ecosystem. Moreover, the biological indices based on different taxonomic groups may indicate anthropogenic pressures of different types (e.g., nutrient enrichment vs. habitat degradation) or occurring at different spatial and temporal scales (e.g., local habitats vs. watershed-level) (Carlisle et al., 2008; Walters et al., 2009). In this respect, the OOAO rule makes more sense than other, less conservative rules.

Although the OOAO rule is simple to implement in practice, the consequences of using this rule become more complicated when considering the effects of uncertainty, as shown by our analysis and by Caroni et al. (2013). In future versions of guidelines for WFD implementation, these findings should be considered and alternative combination rules should be discussed. Like other authors (Alahuhta et al., 2009), we will not conclude by recommending one particular combination rule as the most appropriate, but hope that our results contribute to a better understanding of the benefits and shortcomings of different combination rules when applied to different ecosystem.

We support the statements that more research is needed on combination rules for integrated waterbody assessment under uncertainty (Caroni et al., 2013; Nõges et al., 2009). The existing datasets in this and other cited studies provide opportunities for more investigation using simulation approaches such as WISERBUGS, e.g. with different criteria for weighting BQEs and different uncertainty levels under alternative combination rules. The aim should be to obtain a combination rule that ensures the protection of the whole ecosystem elements under pressure, while the risk of underclassification and "false alarm" for restoration is acceptable for waterbody management in practice.

Acknowledgments

The study was financed by a grant from the Norwegian Financial Mechanism (Project No. PNRF – 220 – A I – 1/07) within the project *deWELopment* (Development and validation of methods for integrated assessment of ecological status of rivers and lakes to support river basin management plans). We are grateful to Hanna Ciecierska and Joanna Ruszczyńska (University of Warmia and Mazury) for macrophyte data, Agnieszka Pasztaleniec (Institute of Environmental Protection – NRI), Agnieszka Napiórkowska-Krzebietke, Joanna Hutorowicz (Inland Fisheries Institute) for phytoplankton data, and Łucjan Chybowski (Inland Fisheries Institute) for fish data. Sebastian Kutyła is acknowledged for kindly supporting us in the preparation of Fig. 1. We thank Ralph Clarke and Mike Dunbar for advice on the use of the WISERBUGS method. Anders Gjørwad Hagen and four anonymous reviewers are acknowledged for helpful comments to earlier versions of the manuscript.

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.

Białokoz, W., Chybowski, Ł., 2011. 5. Ichtyofauna, in: Soszka, H. (Ed.), Ecological status assessment of the waters in the Wel River catchment. Guidelines for integrated assessment of ecological status of rivers and lakes to support river basin management plans. Institute of Inland Fisheries, Olsztyn, Poland, pp. 217-234 (in Polish with English summary).

Birk, S., Bonne, W., Borja, A., Brucet, S., Courrat, A., Poikane, S., Solimini, A., van de Bund, W., Zampoukas, N., Hering, D., 2012. Three hundred ways to assess Europe's surface waters: An almost complete overview of biological methods to implement the Water Framework Directive. Ecological Indicators 18, 31-41.

Borja, A., Bald, J., Franco, J., Larreta, J., Muxika, I., Revilla, M., Rodríguez, J.G., Solaun, O., Uriarte, A., Valencia, V., 2009. Using multiple ecosystem components, in assessing ecological status in Spanish (Basque Country) Atlantic marine waters. Mar Pollut Bull 59, 54-64.

Borja, A., Rodriguez, J.G., 2010. Problems associated with the 'one-out, all-out' principle, when using multiple ecosystem components in assessing the ecological status of marine waters. Mar Pollut Bull 60, 1143-1146.

Böhmer, J., Arbaciauskas, K., Benstead, R., Wim Gabriels, Porst, G., Reeze, B., Timm, H., 2014. Central Baltic Lake Benthic invertebrate ecological assessment methods, in: Poikane, S. (Ed.), Water Framework Directive Intercalibration Technical Report. European Commission Joint Research Centre.

Carlisle, D.M., Hawkins, C.P., Meador, M.R., Potapova, M., Falcone, J., 2008. Biological assessment of Appalachian streams based on predictive models for fish, macroinvertebrate, and diatom assemblages. Journal of North American Benthological Society 27, 16-37.

Caroni, R., Bund, W., Clarke, R.T., Johnson, R.K., 2013. Combination of multiple biological quality elements into waterbody assessment of surface waters. Hydrobiologia 704, 437-451.

Carstensen, J., 2007. Statistical principles for ecological status classification of Water Framework Directive monitoring data. Mar Pollut Bull 55, 3-15.

Carvalho, L., Poikane, S., Lyche Solheim, A., Phillips, G., Borics, G., Catalan, J., Hoyos, C., Drakare, S., Dudley, B.J., Järvinen, M., Laplace-Treyture, C., Maileht, K., McDonald, C., Mischke, U., Moe, J., Morabito, G., Nõges, P., Nõges, T., Ott, I., Pasztaleniec, A., Skjelbred, B., Thackeray, S.J., 2013. Strength and uncertainty of phytoplankton metrics for assessing eutrophication impacts in lakes. Hydrobiologia 704, 127-140.

Clarke, R.T., 2013. Estimating confidence of European WFD ecological status class and WISER Bioassessment Uncertainty Guidance Software (WISERBUGS). Hydrobiologia 704, 39-56.

Clarke, R.T., Hering, D., 2006. Errors and uncertainty in bioassessment methods – major results and conclusions from the STAR project and their application using STARBUGS. Hydrobiologia 566, 433-439.

Clarke, R.T., Lorenz, A., Sandin, L., Schmidt-Kloiber, A., Strackbein, J., Kneebone, N.T., Haase, P., 2006. Effects of sampling and sub-sampling variation using the STAR-AQEM sampling protocol on the precision of macroinvertebrate metrics. Hydrobiologia 566, 441-459.

EC (European Commission), 2000. Directive 2000/60/EC of the European Parliament and of the council of 23 October 2000 establishing a framework for Community action in the field of water policy. , Official Journal of the European Communities, Luxembourg.

EC (European Commission), 2005. Overall approach to the classification of ecological status and ecological potential, Common Implementation Strategy for the Water Framework Directive (2000/60/CE). Office for Official Publications of the European Communities, Luxembourg.

EC (European Commission), 2011. Guidance document on the intercalibration process 2008-2011. Technical Report - 2011 - 045, Common Implementation Strategy for the Water Framework Directive (2000/60/CE). Office for Official Publications of the European Communities, Luxembourg.

Gołub, M., Prus, P., Koszałka, J., 2011. 4. Macrozoobenthos, in: Soszka, H. (Ed.), Ecological status assessment of the waters in the Wel River catchment. Guidelines for integrated assessment of ecological status of rivers and lakes to support river basin management plans. Institute of Inland Fisheries, Olsztyn, Poland, pp. 203-216 (in Polish with English summary).

Gottardo, S., Semenzin, E., Giove, S., Zabeo, A., Critto, A., de Zwart, D., Ginebreda, A., Marcomini, A., 2011. Integrated risk assessment for WFD ecological status classification applied to Llobregat river basin (Spain). Part I—Fuzzy approach to aggregate biological indicators. Science of The Total Environment 409, 4701-4712.

Hering, D., Borja, A., Carstensen, J., Carvalho, L., Elliott, M., Feld, C.K., Heiskanen, A.-S., Johnson, R.K., Moe, J., Pont, D., Solheim, A.L., de Bund, W.v., 2010. The European Water Framework Directive at the age of 10: A critical review of the achievements with recommendations for the future. Science of The Total Environment 408, 4007-4019.

Hutorowicz, A., Napiórkowska-Krzebietke, A., Pasztaleniec, A., Hutorowicz, J., Lyche Solheim, A., Skjelbred, B., 2011. 1. Phytoplankton, in: Soszka, H. (Ed.), Ecological status assessment of the waters in the Wel River catchment. Guidelines for integrated assessment of ecological status of rivers and lakes to support river basin management plans. Institute of Inland Fisheries, Olsztyn, Poland, pp. 143-168 (in Polish with English summary).

Højberg, A.L., Refsgaard, J.C., Geer, F., Jørgensen, L.F., Zsuffa, I., 2007. Use of models to support the monitoring requirements in the Water Framework Directive. Water Resour Manage 21, 1649-1672.

Irvine, K., 2004. Classifying ecological status under the European Water Framework Directive: the need for monitoring to account for natural variability. Aquatic Conservation: Marine and Freshwater Ecosystems 14, 107-112.

Kelly, M., Bennett, C., Coste, M., Delgado, C., Delmas, F., Denys, L., Ector, L., Fauville, C., Ferréol, M., Golub, M., Jarlman, A., Kahlert, M., Lucey, J., Ní Chatháin, B., Pardo, I., Pfister, P., Picinska-Faltynowicz, J., Rosebery, J., Schranz, C., Schaumburg, J., van Dam, H., Vilbaste, S., 2009a. A comparison of national approaches to setting ecological status boundaries in phytobenthos assessment for the European Water Framework Directive: results of an intercalibration exercise. Hydrobiologia 621, 169-182.

Kelly, M., Bennion, H., Burgess, A., Ellis, J., Juggins, S., Guthrie, R., Jamieson, J., Adriaenssens, V., Yallop, M., 2009b. Uncertainty in ecological status assessments of lakes and rivers using diatoms. Hydrobiologia 633, 5-15.

Kolada, A., Ciecierska, H., Ruszczyńska, J., Dynowski, P., 2013. Sampling techniques and inter-surveyor variability as sources of uncertainty in Polish macrophyte metric for lake ecological status assessment. Hydrobiologia DOI 10.1007/s10750-013-1591-9.

Kolada, A., Ciecierska, H., Ruszczyńska, J., Mjelde, M., Dziedzic, J., Dynowski, P., Jasnorzewski, A., 2011. 2. Macrophytes, in: Soszka, H. (Ed.), Ecological status assessment of the waters in the Wel River catchment. Guidelines for integrated assessment of ecological status of rivers and lakes to support river basin management plans. Institute of Inland Fisheries, Olsztyn, Poland, pp. 169-186 (in Polish with English summary).

Moss, B., Stephen, D., Alvarez, C., Becares, E., Bund, W.V.D., Collings, S.E., Donk, E.V., Eyto, E.D., Feldmann, T., Fernández-Aláez, C., Fernández-Aláez, M., Franken, R.J.M., García-Criado, F., Gross, E.M., Gyllström, M., Hansson, L.-A., Irvine, K., Järvalt, A., Jensen, J.-P., Jeppesen, E., Kairesalo, T., Kornijów, R., Krause, T., Künnap, H., Laas, A., Lill, E., Lorens, B., Luup, H., Rosa Miracle, M., Nõges, P., Nõges, T., Nykänen, M., Ott, I., Peczula, W., Peeters, E.T.H.M., Phillips, G., Romo, S., Russell, V., Salujõe, J., Scheffer, M., Siewertsen, K., Smal, H., Tesch, C., Timm, H., Tuvikene, L., Tonno, I., Virro, T., Vicente, E., Wilson, D., 2003. The determination of ecological status in shallow lakes — a tested system (ECOFRAME) for implementation of the European Water Framework Directive. Aquatic Conservation: Marine and Freshwater Ecosystems 13, 507-549.

Nõges, P., Bund, W., Cardoso, A.C., Solimini, A.G., Heiskanen, A.-S., 2009. Assessment of the ecological status of European surface waters: a work in progress. Hydrobiologia 633, 197-211.

Nõges, P., Nõges, T., 2006. Indicators and criteria to assess ecological status of the large shallow temperate polymictic lakes Peipsi (Estonia/Russia) and Vortsjarv (Estonia). Boreal Environ. Res. 11, 67-80.

Phillips, G., Free, G., Karottki, I., Laplace-Treyture, C., Maileht, K., Mischke, U., Ott, I., Pasztaleniec, A., Portielje, R., Søndergaard, M., Trodd, W., Wichelen, J.V., 2014. Central Baltic Lake Phytoplankton ecological assessment methods, in: Poikane, S. (Ed.), Water Framework Directive Intercalibration Technical Report. European Commission Joint Research Centre.

Picińska-Fałtynowicz, J., 2011. 3. Phytobenthos, in: Soszka, H. (Ed.), Ecological status assessment of the waters in the Wel River catchment. Guidelines for integrated assessment of ecological status of rivers and lakes to support river basin management plans. Institute of Inland Fisheries, Olsztyn, Poland, pp. 187-202 (in Polish with English summary).

Portielje, R., Bertrin, V., Denys, L., Grinberga, L., Karottki, I., Kolada, A., Krasovskienė, J., Leiputé, G., Maemets, H., Ott, I., Phillips, G., Pot, R., Schaumburg, J., Schranz, C., Soszka, H., Stelzer, D., Søndergaard, M., Willby, N., 2014. Central Baltic Lake Macrophyte ecological assessment methods, in: Poikane, S. (Ed.), Water Framework Directive Intercalibration Technical Report. European Commission Joint Research Centre.

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

R Development Core Team, 2011. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rask, M., Vuori, K.-M., Hämäläinen, H., Järvinen, M., Hellsten, S., Mykrä, H., Arvola, L., Ruuhijärvi, J., Jyväsjärvi, J., Kolari, I., Olin, M., Salonen, E., Valkeajärvi, P., 2010. Ecological classification of large lakes in Finland: comparison of classification approaches using multiple quality elements. Hydrobiologia 660, 37-47.

Ruse, L., 2010. Classification of nutrient impact on lakes using the chironomid pupal exuvial technique. Ecological Indicators 10, 594-601.

Sandin, L., 2005. Testing the EC Water Framework Directive “one-out, all-out” rule - simulating different levels of assessment errors along a pollution gradient in Swedish streams. Verhandlungen der Internationale Vereinigung für Theoretische und Angewandte Limnologie 29, 334-336.

Schneider, S.C., Lawniczak, A.E., Picińska-Faltynowicz, J., Szoszkiewicz, K., 2012. Do macrophytes, diatoms and non-diatom benthic algae give redundant information? Results from a case study in Poland. Limnologica - Ecology and Management of Inland Waters 42, 204-211.

Soszka, H., 2011. Ecological status assessment of the waters in the Wel River catchment. Guidelines for integrated assessment of ecological status of rivers and lakes to support river basin management plans. Institute of Inland Fisheries, Olsztyn, Poland (in Polish with English summary).

Sutela, T., Aroviita, J., Keto, A., 2013. Assessing ecological status of regulated lakes with littoral macrophyte, macroinvertebrate and fish assemblages. Ecological Indicators 24, 185-192.

Szoszkiewicz, K., Jusik, S., Zgola, T., Czechowska, M., Hryc, B., 2007. Uncertainty of macrophyte-based monitoring for different types of lowland rivers. Belgian Journal of Botany 140, 7-16.

Søndergaard, M., Jeppesen, E., Peder Jensen, J., Lildal Amsinck, S., 2005. Water Framework Directive: ecological classification of Danish lakes. Journal of Applied Ecology 42, 616-629.

Thackeray, S.J., Nõges, P., Dunbar, M.J., Dudley, B.J., Skjelbred, B., Morabito, G., Carvalho, L., Phillips, G., Mischke, U., Catalan, J., de Hoyos, C., Laplace, C., Austoni, M., Padedda, B.M., Maileht, K., Pasztaleniec, A., Järvinen, M., Solheim, A.L., Clarke, R.T., 2013. Quantifying uncertainties in biologically-based water quality assessment: A pan-European analysis of lake phytoplankton community metrics. Ecological Indicators 29, 34-47.

Veríssimo, H., Lane, M., Patrício, J., Gamito, S., Marques, J.C., 2013. Trends in water quality and subtidal benthic communities in a temperate estuary: Is the response to restoration efforts hidden by climate variability and the Estuarine Quality Paradox? Ecological Indicators 24, 56-67.

Walters, D.M., Roy, A.H., Leigh, D.S., 2009. Environmental indicators of macroinvertebrate and fish assemblage integrity in urbanizing watersheds. Ecological Indicators 9, 1222-1233.

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).

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

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

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

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.

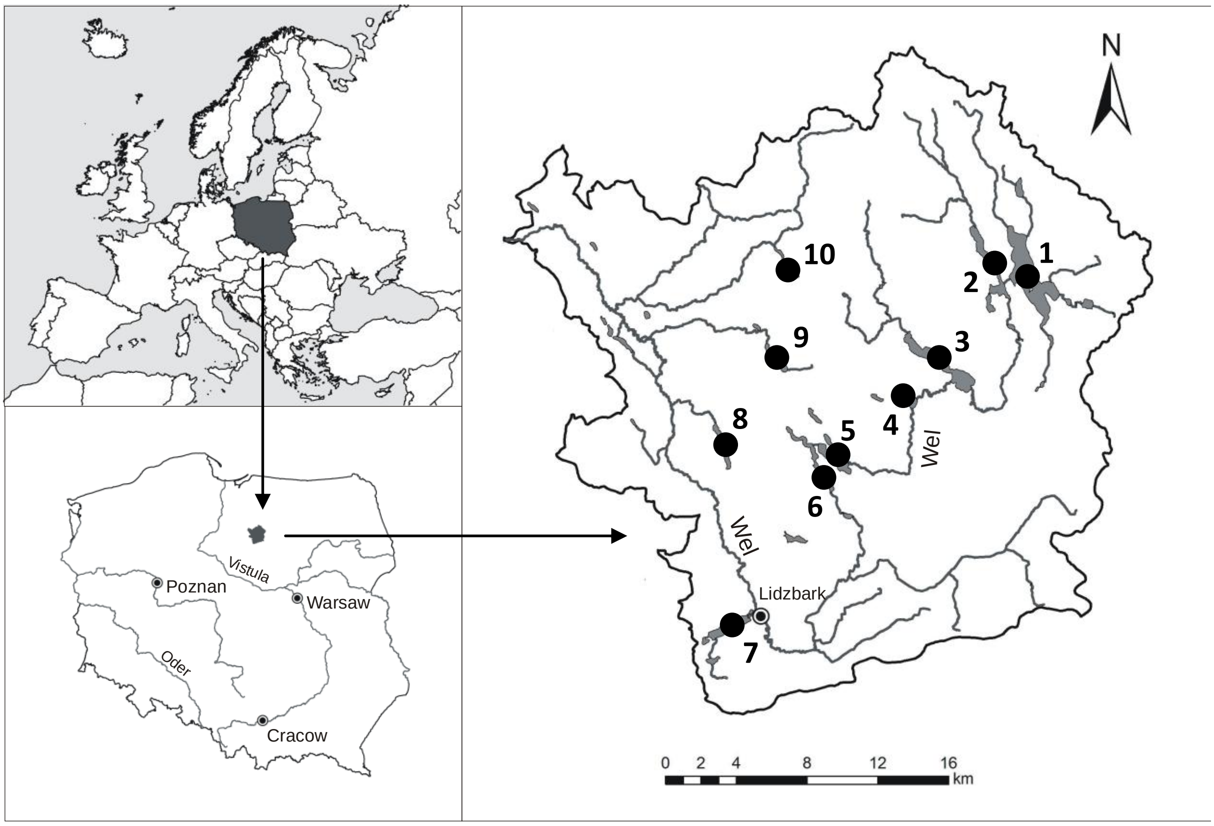
N
10
1
2
9
3
4
8
5
Wel
6
Wel
Lidzbark
7
0 2 4 8 12 16
km
Vistula
Poznan
Warsaw
Oder
Cracow

Fig. 2.

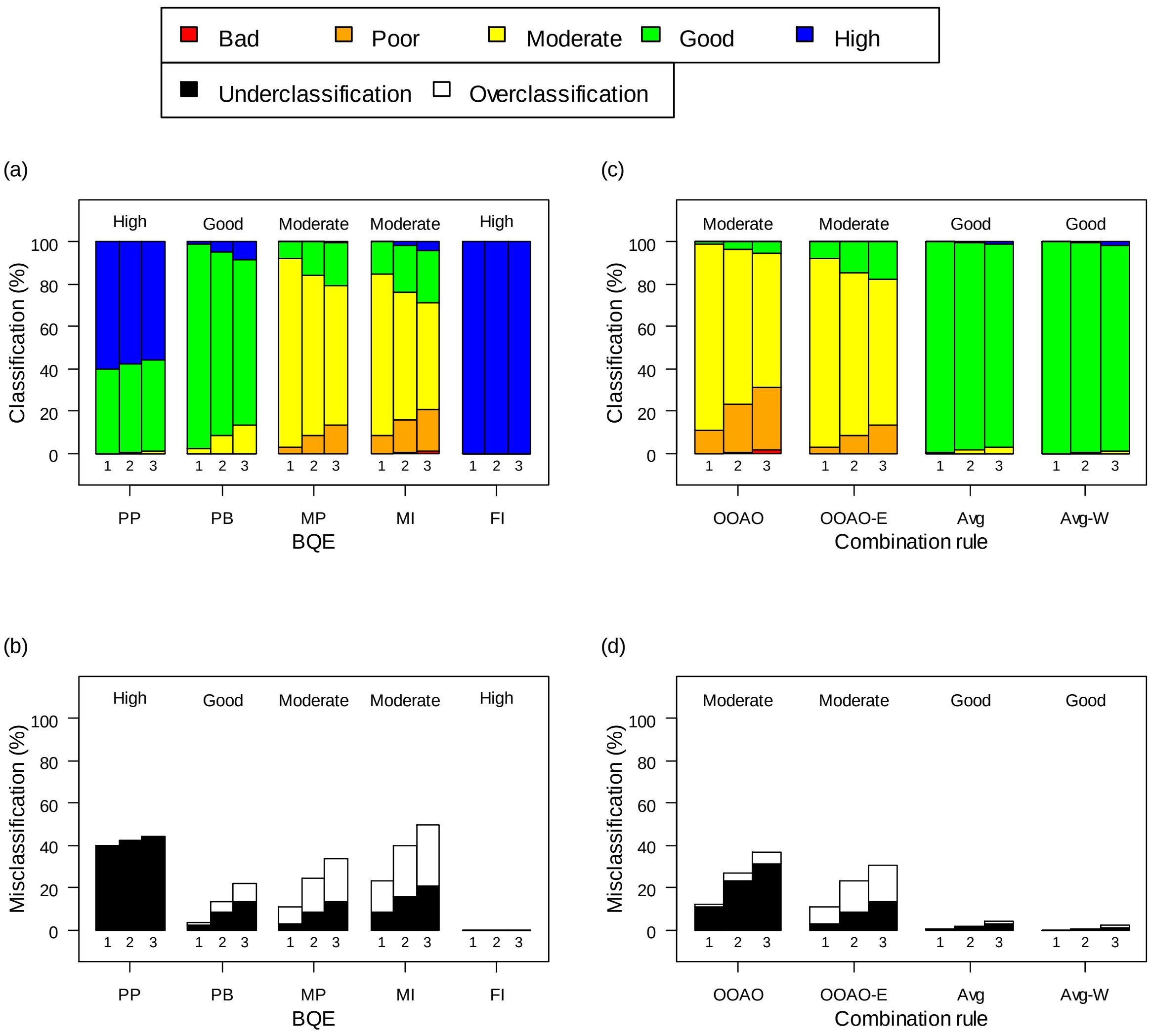
Bad
Poor
Moderate
Good
High
Underclassification
Overclassification
(a)
(b)
(c)
(d)
Classification (%)
Misclassification (%)
High
Good
Moderate
Moderate
High
PP
PB
MP
MI
FI
BQE
OOAO
OOAO-E
Avg
Avg-W
Combination rule

Fig. 3.

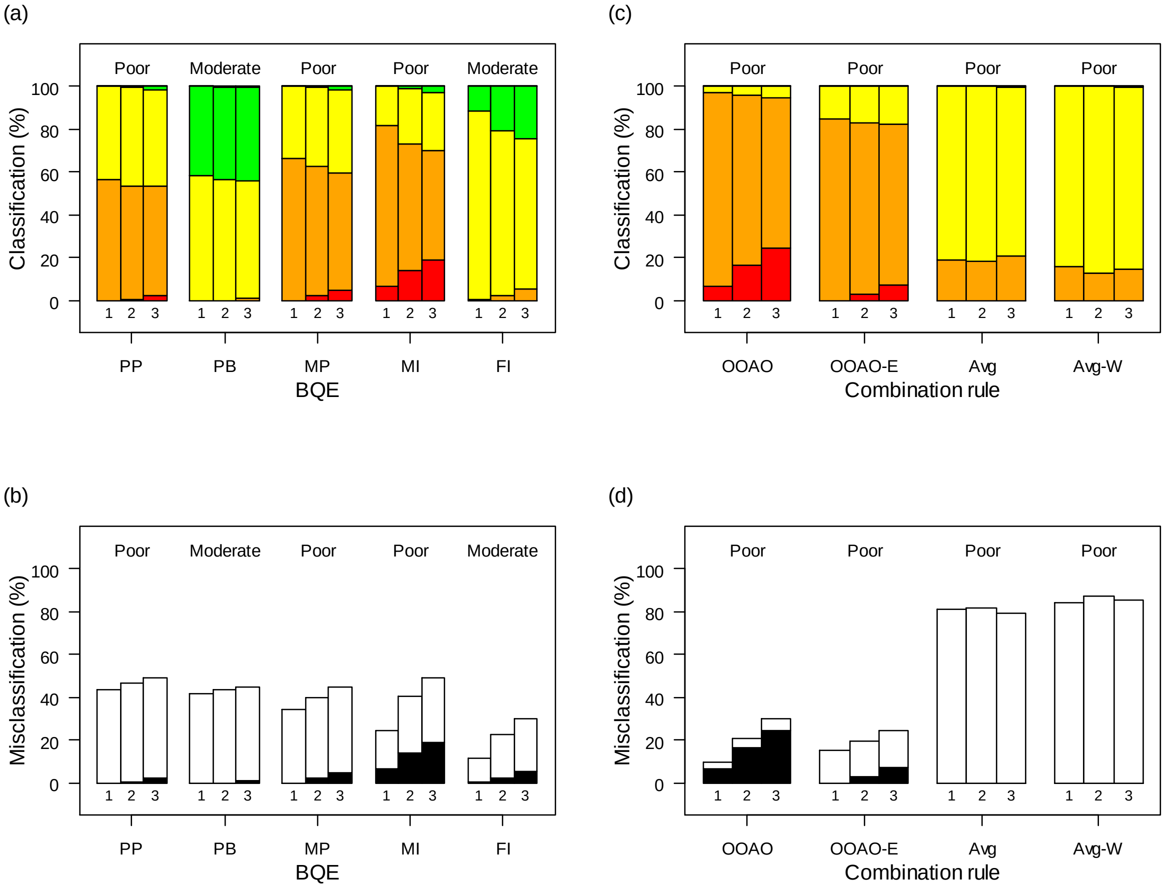
(a)
Poor
Moderate
Poor
Poor
Moderate
Classification (%)
100
80
60
40
20
0
1 2 3
PP
PB
MP
MI
FI
BQE
(c)
Poor
Poor
Poor
Poor
OOAO
OOAO-E
Avg
Avg-W
Combination rule
(b)
Misclassification (%)
PP
PB
MP
MI
FI
BQE
(d)
OOAO
OOAO-E
Avg
Avg-W
Combination rule

Fig. 4.

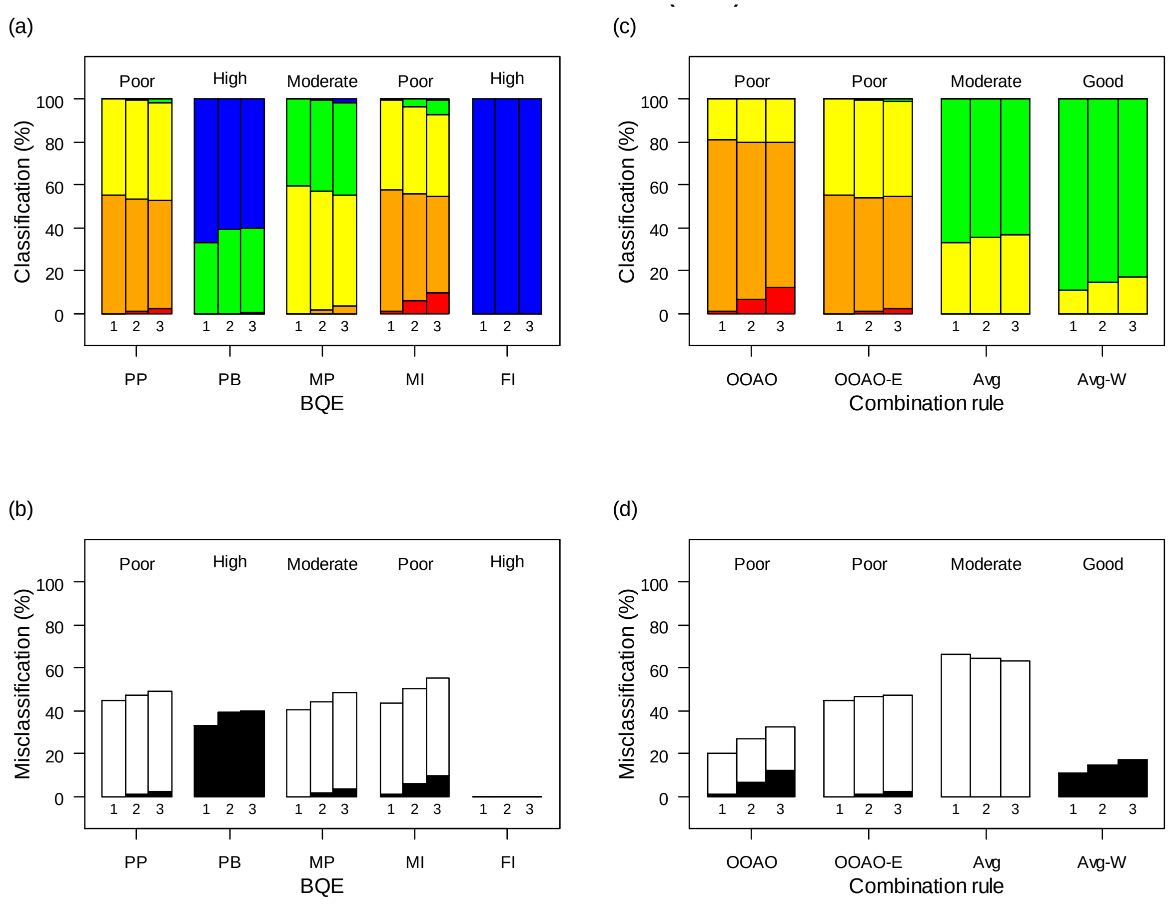
(a)
(c)
(b)
(d)
Poor
High
Moderate
Poor
High
Poor
Poor
Moderate
Good
Classification (%)
Misclassification (%)
0
20
40
60
80
100
1 2 3
PP
PB
MP
MI
FI
BQE
OOAO
OOAO-E
Avg
Avg-W
Combination rule

Fig. 5.

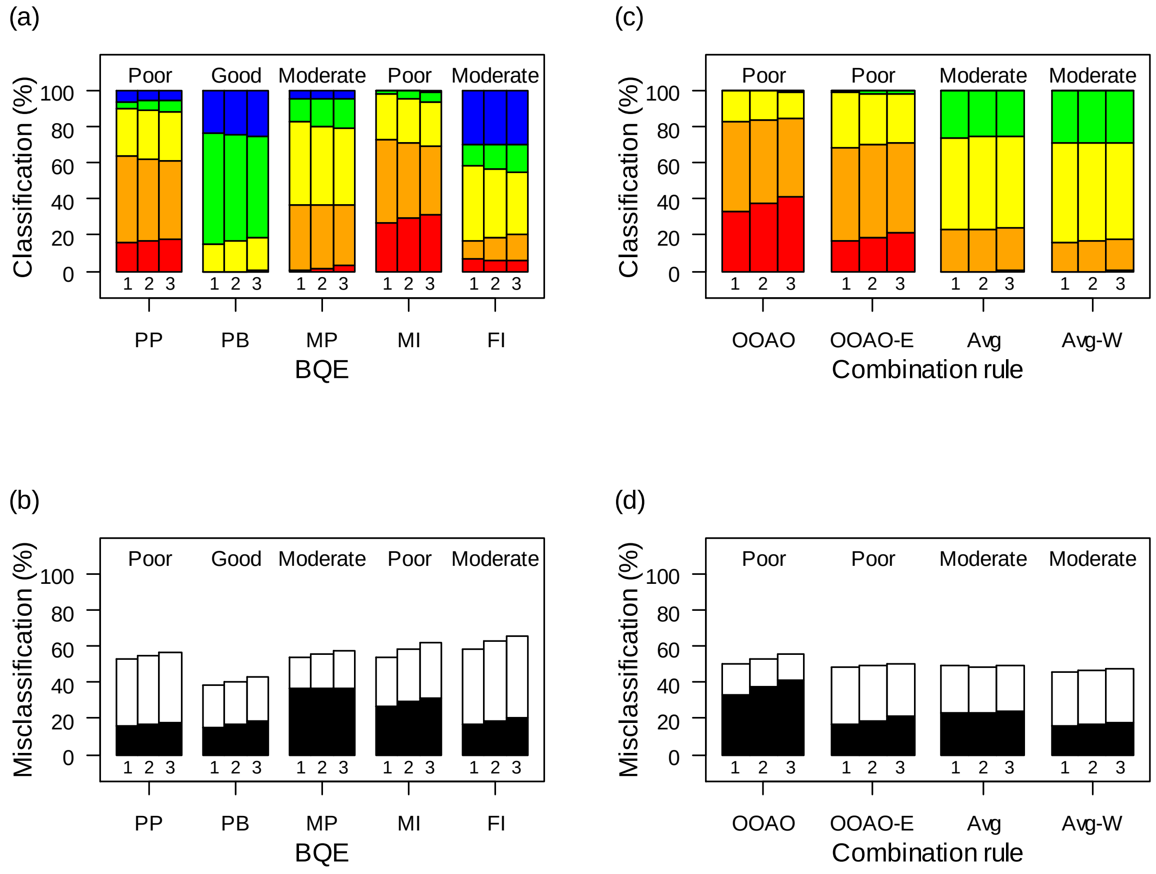
(a)
Poor
Good
Moderate
Poor
Moderate
Classification (%)
100
80
60
40
20
0
1 2 3
PP
PB
MP
MI
FI
BQE
(c)
Poor
Poor
Moderate
Moderate
Classification (%)
OOAO
OOAO-E
Avg
Avg-W
Combination rule
(b)
Misclassification (%)
Poor
Good
Moderate
Poor
Moderate
PP
PB
MP
MI
FI
BQE
(d)
Misclassification (%)
Poor
Poor
Moderate
Moderate
OOAO
OOAO-E
Avg
Avg-W
Combination rule

Fig. 6.

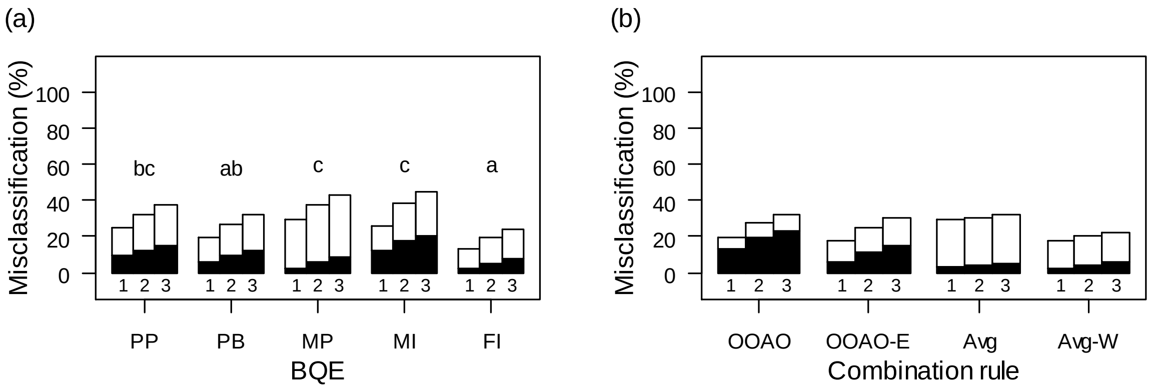
(a)
Misclassification (%)
100
80
60
40
20
0
bc
ab
c
c
a
1 2 3
PP
PB
MP
MI
FI
BQE
(b)
Misclassification (%)
100
80
60
40
20
0
1 2 3
OOAO
OOAO-E
Avg
Avg-W
Combination rule

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

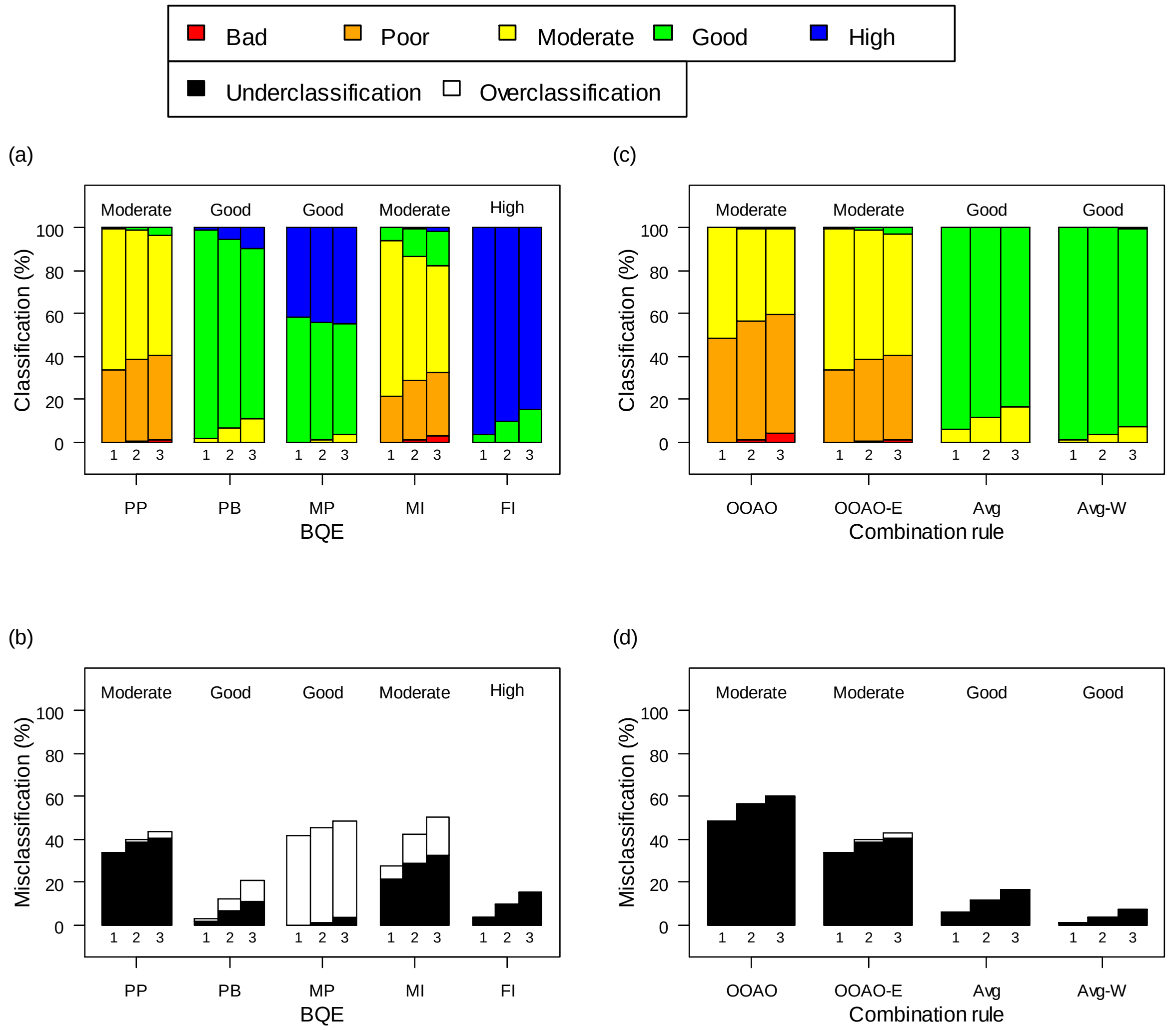

Fig. S1. Uncertainty in classification of Lake Dąbrowa Wielka for different BQEs, different BQE combination rules and different levels of uncertainty in individual BQE index values. For more details, see Fig. 2.

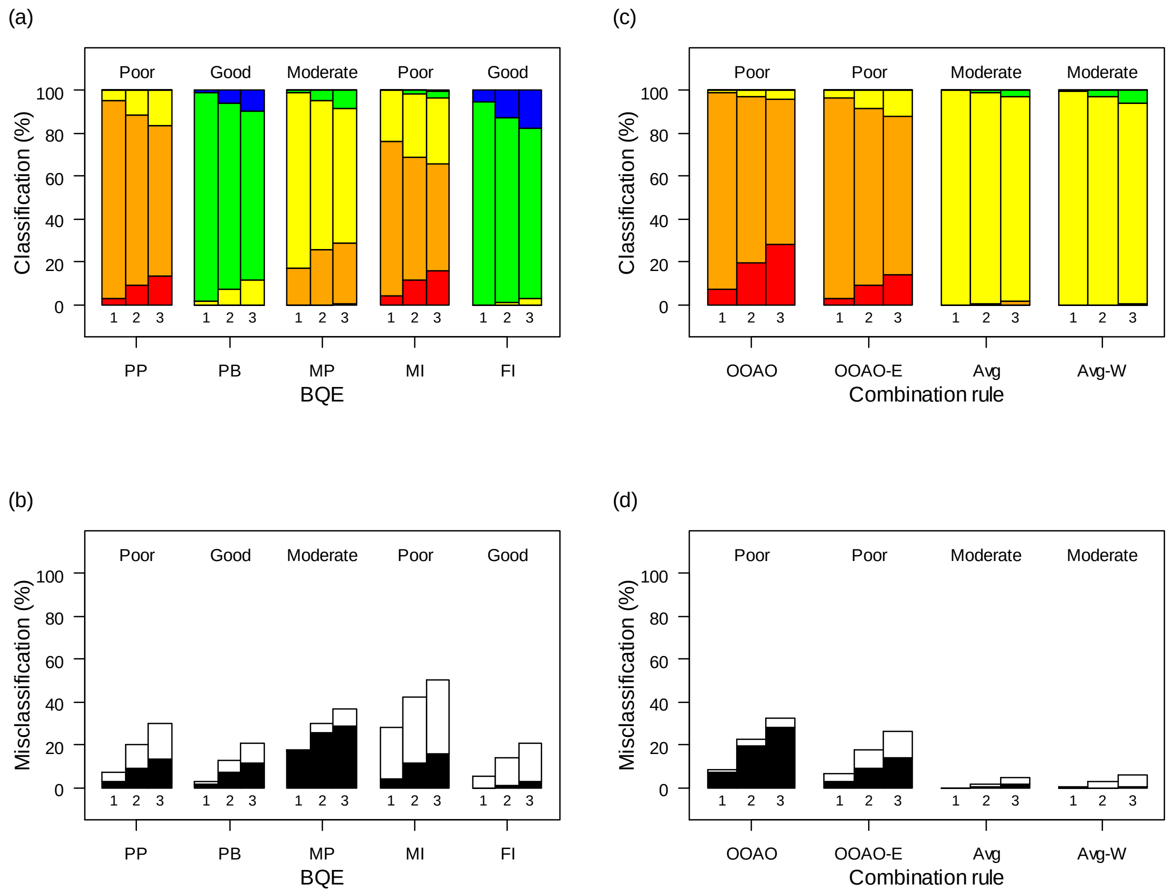

Fig. S2. Uncertainty in classification of Lake Rumian for different BQEs, different BQE combination rules and different levels of uncertainty in individual BQE index values. For more details, see Fig. 2.

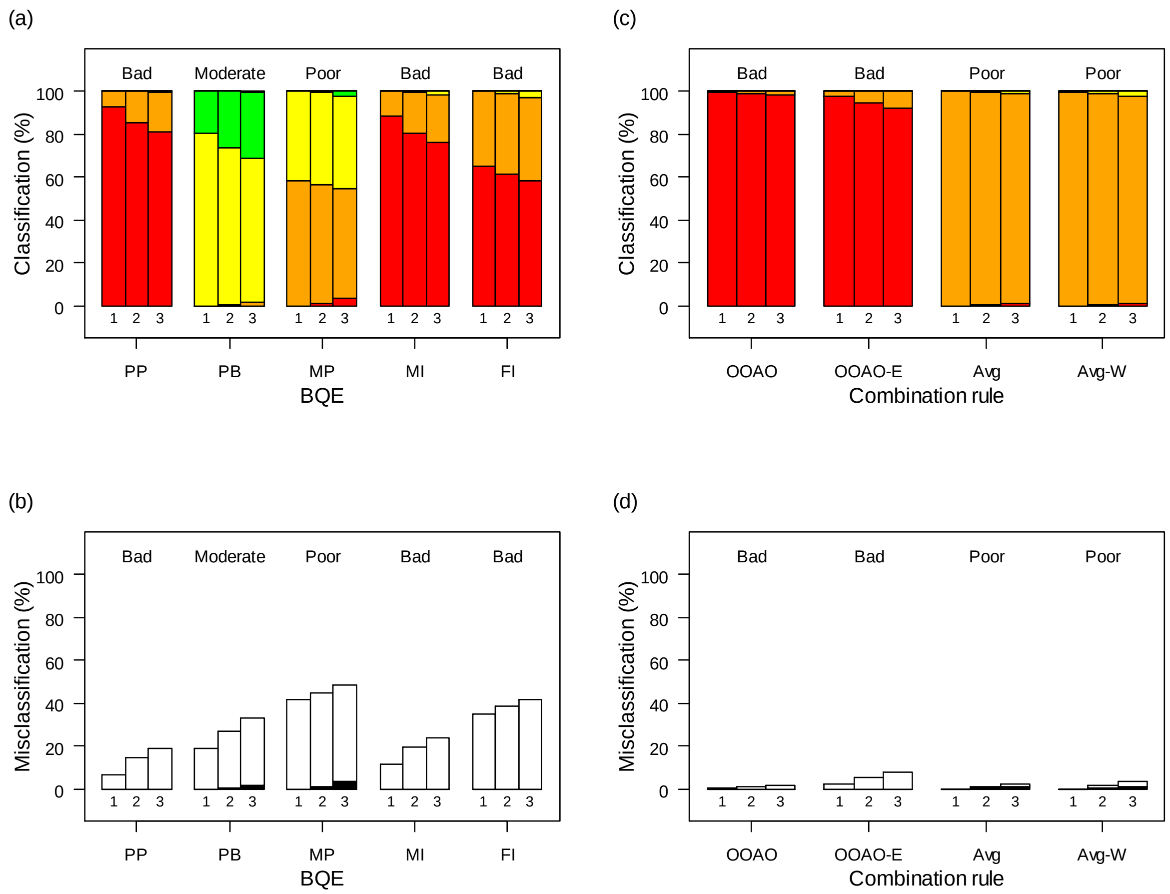

Fig. S3. Uncertainty in classification of Lake Tarczyńskie for different BQEs, different BQE combination rules and different levels of uncertainty in individual BQE index values. For more details, see Fig. 2.

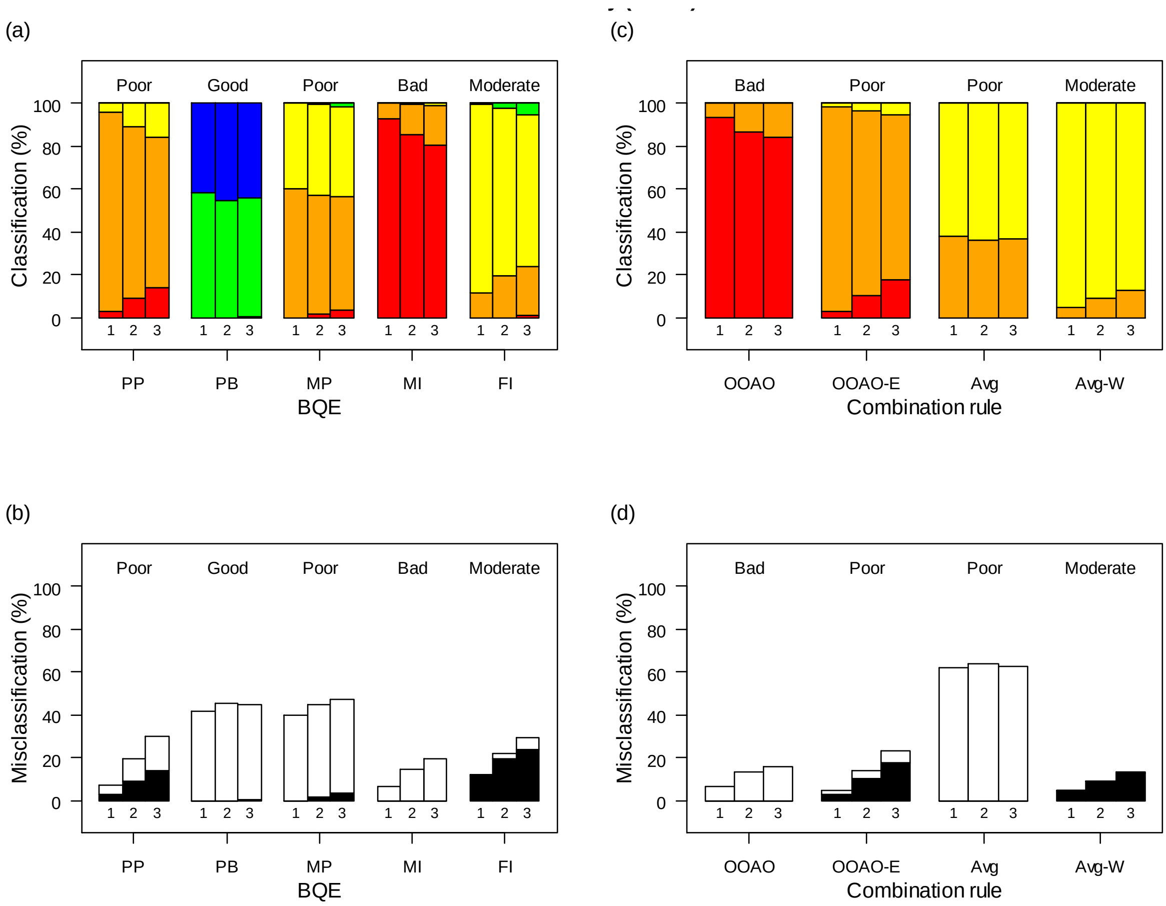

Fig. S4. Uncertainty in classification of Lake Grądy for different BQEs, different BQE combination rules and different levels of uncertainty in individual BQE index values. For more details, see Fig. 2.

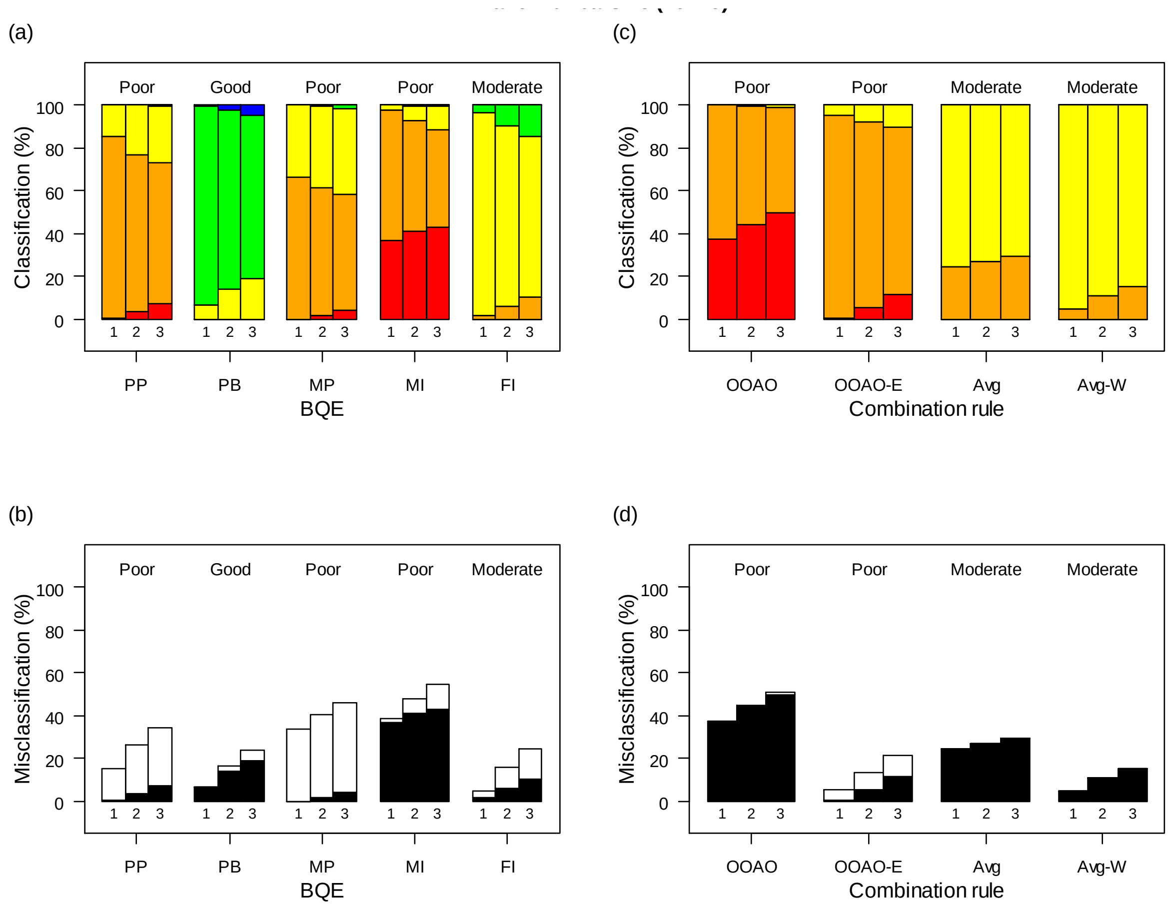

Fig. S5. Uncertainty in classification of Lake Lidzbarskie for different BQEs, different BQE combination rules and different uncertainty levels in BQE index values. For more details, see Fig. 2.

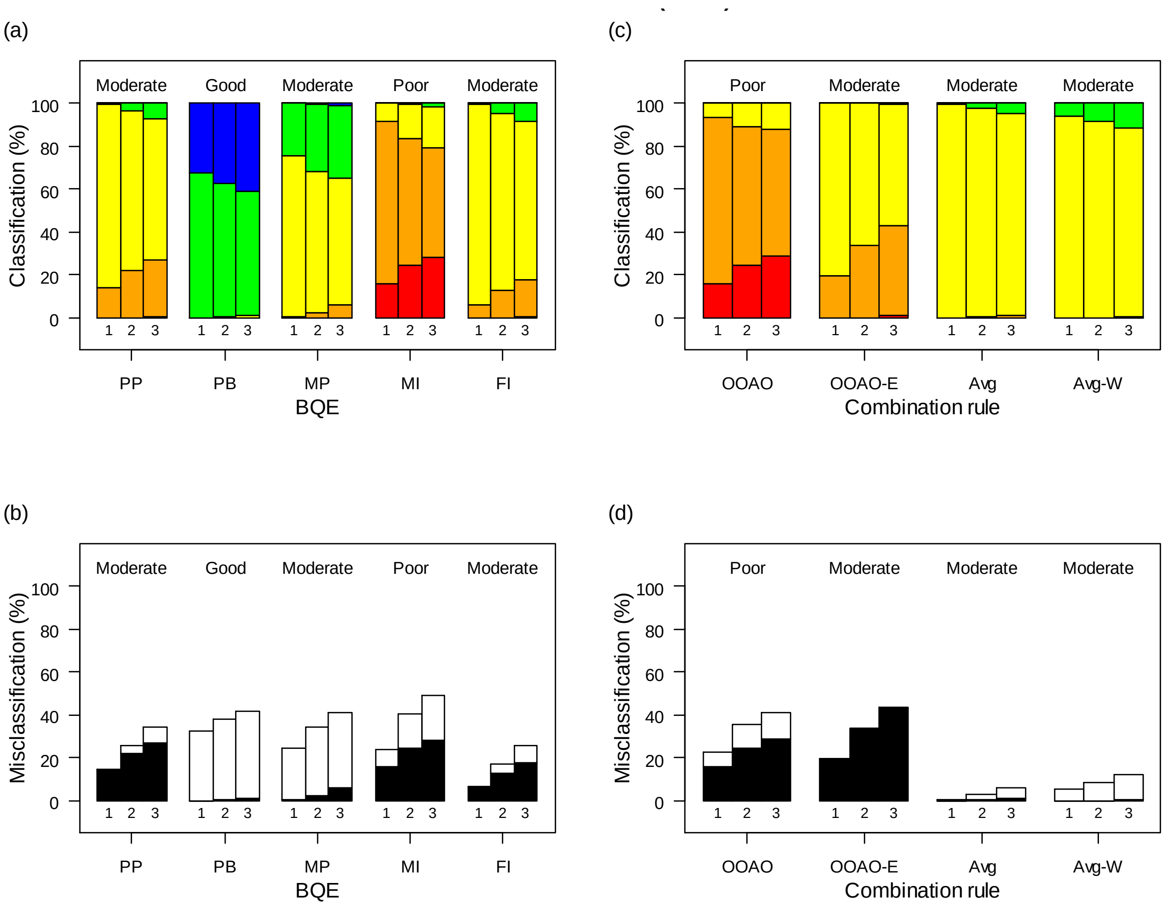

Fig. S6. Uncertainty in classification of Lake Hartowieckie for different BQEs, different BQE combination rules and different levels of uncertainty in individual BQE index values. For more details, see Fig. 2.

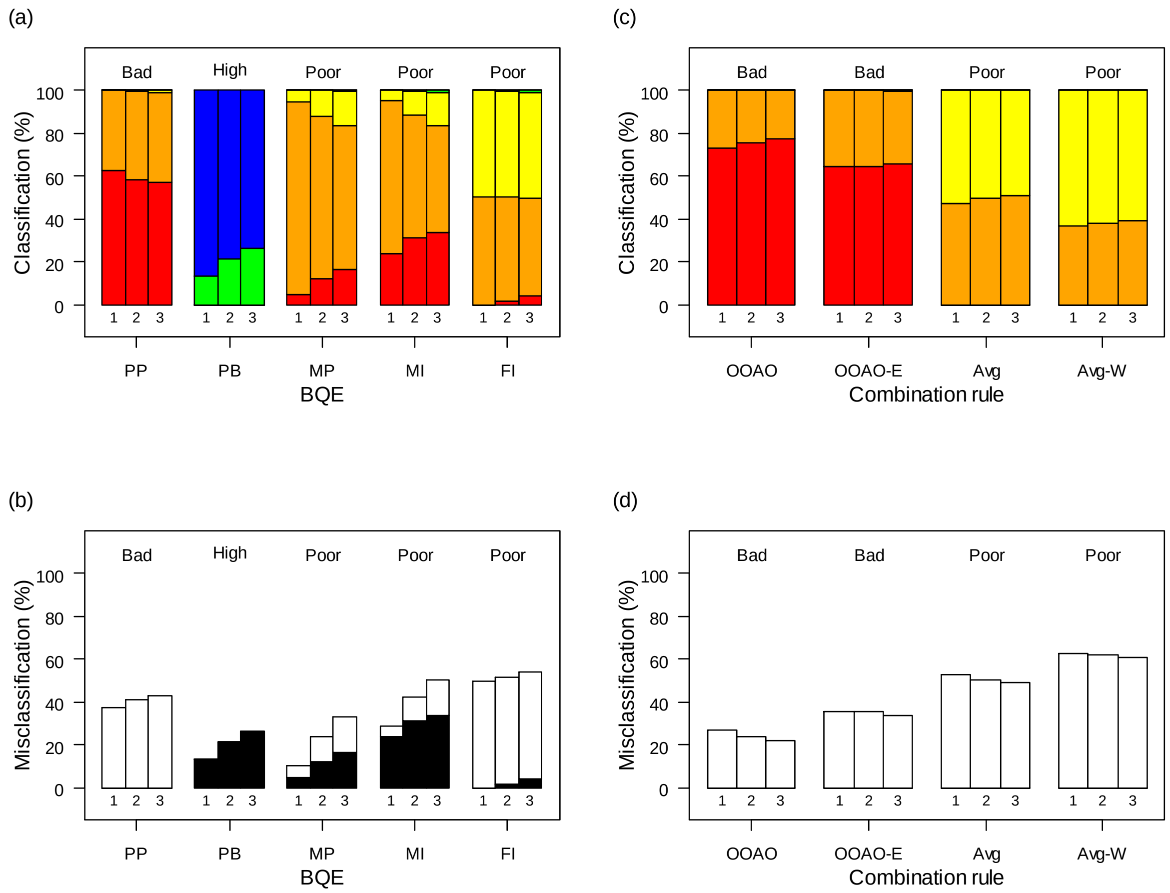

Fig. S7. Uncertainty in classification of Lake Zwiniarz for different BQEs, different BQE combination rules and different levels of uncertainty in individual BQE index values. For more details, see Fig. 2.