sciendo DOI: 10.2478/bile-2018-0015

Biometrical Letters
Vol. 55 (2018), No.2, 215-232

Determining the change point for the error in the Macrophyte Index for Rivers

Anna Budka

Department of Mathematical and Statistical Methods, Poznan University of Life Sciences, Wojska Polskiego 28, 60-637 Poznań, Poland, e-mail: anna.budka@up.poznan.pl

SUMMARY

The consequences of the growing demand for water include a significant deterioration in its quality and a drastic decline in biodiversity, which is a serious threat to the hydrological and biocenotic balance of freshwater ecosystems. A good indicator of aquatic environment quality is macrophytes. Studies on macrophytes are one of the primary elements in the ecological status assessment of surface waters, in accordance with the guidelines of the Water Framework Directive. In Poland, research on the ecological status of rivers with regard to macrophytes has been carried out since 2008, using the Macrophyte Index for Rivers (*MIR*), which takes into account the number and coverage of macrophyte taxa. An analysis of numbers of species that need to be indicated at a site for valid assessment of the ecosystem was conducted on the basis of studies on macrophytes from 2008–2013, at 60 sites in small lowland rivers with a sandy substrate, of which 20 sites were selected on the most diverse watercourses: the least clean (quality class V), moderate (quality class III), and the cleanest (quality class I). The results of the botanical studies served to assess the completeness of the samples (the number of species recorded at a site) used to evaluate the ecological status of a river. The proposed analyses enabled estimation of the approximate number of species required to determine the MIR for rivers in each quality class.

Key words: macrophytes, Macrophyte Index for Rivers (*MIR*), At Most One Change (AMOC)

1. Introduction

The rapid development of civilization entails increasing pressure on the environment. Therefore, various actions to counteract water degradation are of utmost importance. Over the years, a number of methods have been developed to assess and monitor water ecosystems, in order to halt their destruction and evaluate the environmental impact of human activities. At the end of the 19th century, bioindication was already the subject of various research, which focused on assessing the relation between the aquatic environment and organisms living therein (Kolkwitz and Marsson, 1902). Such methods have been used, with local modifications, for more than fifty years (Starmach et al., 1976; Zimny, 2006). It has been noted that various groups of organisms, in particular macrophytes, may be used for the purposes of bioindication and water quality assessment (Dawson, 1999; Zelinka and Marvan, 1961).

Macrophytes (just like other organisms) are continuously exposed to aquatic environmental pressure. Recognition of their sensitivity to a specific contaminant enables determination of the degree of degradation of the aquatic environment resulting from a given pressure acting over a longer timeframe. They are distinct in this respect from other organisms such as phytoplankton and zooplankton, which react very quickly (Wiegleb, 1979; Haslam, 1982; Holmes et al., 1999; Ceschin et al., 2010; Szoszkiewicz et al., 2017).

The quality of flowing waters in Poland is characterized by a biological indicator, the Macrophyte Index for Rivers (MIR). The index was developed on the basis of the occurrence of selected macrophytes serving as indicator species (Szoszkiewicz et al., 2010), and its determination is one of the elements of the Macrophyte Assessment Method for Rivers. Since 2008, it has been one of the indicators used by the Ministry of the Environment for classifying and monitoring water quality (OJ, 2016). The MIR is based on research carried out on a hundred-meter river stretch, and takes into account the occurrence and number of species.

The study includes a description of the cleanest (quality class I), moderate (quality class III), and the dirtiest sites (quality class V) on Polish lowland rivers with a sandy substrate, in terms of the frequency of occurrence and coverage of macrophyte species. Based on simulation analyses, a determination is made of the number of species that need to be identified at a site in order for them to be considered a sufficient source of information in determining the MIR indicator, which aims to reflect the actual state of the aquatic ecosystem.

Given the nature of the research and the associated large costs and workload, it seemed appropriate to use the simulation model as a tool for the preparation of analyses facilitating the making of decisions regarding environmental protection. Significant fluctuations in the examined feature (response variable the standard error of the means) were observed in various ranges of the explanatory variable (missing number of macrophyte species). This means that it was very difficult to formulate a general model to describe the relationship between the response and explanatory variables. Thus, an approach based on multi-phase in particular, two-phase analysis seemed to be appropriate. The merits of the use of two-phase analysis may be observed in the graphic representation of the data in the coordinate system; in particular, when there are two points at which a completely different reaction of an observed feature occurs. Such points will be referred to as changepoints. The aim of many research projects, including the present one, is to localize (obtain the coordinates of) one or a few such points. Apart from the identification of changepoints, great interest is also shown in questions of inference in reaction models (e.g. regression models).

The search for changepoints in various biology-related fields has been undertaken by many researchers (Sprent, 1961; Hudson, 1966; Hinkley, 1969; Bacon and Watts, 1971; Watts and Bacon, 1974; Kirby, 1974; Lerman, 1980; Seber and Wild, 1989; Lavielle, 2005). In particular, Pruska (1996) gives an overview of parameter estimation methods, where the relationship between the examined features is described by means of multi-phase regression.

In its simplest form, changepoint detection has been understood as estimation of the point at which a change in the statistical properties of the observation sequence occurred. Detection of such changes is vital across different fields of application. In particular, such analysis has been widely used in climate science (Reeves et al., 2007), applications of bioinformatics (Erdman and Emerson, 2008), finance (Zeileis et al., 2002; Zeileis et al., 2010), oceanography (Killick et al., 2010), and medicine (Nam et al., 2012; Hirotsu, 2017).

The aim of the present study was to check whether it is possible to provide an approximate number species found at a site which is necessary to obtain the MIR indicator, based on the standard error of the means.

2. Materials and methods

2.1 Experimental data

Physicochemical criteria were considered based on the state environmental monitoring database, for which water samples were collected at monthly intervals (12 monthly samples). The selected sites represent a wide trophic gradient based on the concentration of phosphorus (reactive and total phosphorus) and nitrogen (total nitrogen). Only sites at which nitrogen concentration correlated with the concentration of phosphorus were selected for analysis (Official Journal, 2016).

The analysis used data obtained from research on macrophytes conducted in rivers of the lowland area of Poland. All sites were of the same abiotic type: medium-sized lowland rivers with a sandy bottom, located less than 200 m above sea level. Studies on macrophytes were conducted at water sampling stations on river stretches with a length of 100 m, in the period from 2008 to 2013, from July to the beginning of September. The samples were collected at monthly intervals. A detailed methodology of the experiment is contained in the paper of Szoszkiewicz et al. (2017).

2.2. Statistical methods

For simulation analyses aimed at explaining the patterns occurring during MIR determination, data from 60 sites were used, on rivers representing three classes: 20 sites of class I (the cleanest rivers), 20 sites of class III (medium soil), and 20 sites of class V (the dirtiest rivers). For the set of macrophyte species detected therein, all possible subsets were determined as *k*-element combinations (*k*=1,…, n) from the *n*-element set. The Macrophyte Index (for a site) was calculated from the following formula:

$$MIR = \frac{\sum_{i=1}^{n}\left(L_i \cdot W_i \cdot P_i\right)}{\sum_{i=1}^{n}\left(W_i \cdot P_i\right)} \cdot 10 \qquad (1)$$

where: L_i is the number of the indicative value for the identified species *i* (L_i=1,…,10), where 1 denotes a low indicative value and 10 a high indicative value; W_i is the weight coefficient for the species *i* (W_i=1, 2, 3), where 1 denotes a low indicative value and 3 a high indicative value; P_i is the coverage rate for the species *i*, on a 9-point scale.

For the set of macrophyte species (observations), the mean MIR for all species from the site, and the mean MIR over 1000 selections for individual combinations with respectively one, two, etc. species removed, were calculated. Furthermore, the variance, standard deviation, and standard error of the mean were determined. For this purpose, proprietary software written in C++ was used. A pseudo-random number generator from the Boost library was used to determine the observations in subsequent iterations of the algorithm (Table 2, Figure 2).

At this stage, it was necessary to select a feature for further statistical analysis. Such analysis may be based on the mean MIR or the standard error of that mean (SEM). It should be noticed that the mean MIR is subject to significant fluctuation (variability), which is related to the number of the species observed, but also to a large extent to the weights *W* and *P* assigned to macrophyte indicator species. In this case, the inflection point may be at the point of apparent change in the trend. This problem may be overcome by seeking a changepoint in a feature with a smaller fluctuation (variability). Such a feature is the standard error of the mean MIR. In this case the changepoint will be identified at a place of significant change in the trend.

The results obtained were subjected to Methods for Changepoint Detection analysis to determine the optimal positioning of the standard error of mean sequence for the selected MIR.

It was assumed that there is an ordered sequence of the standard error of mean MIR $y_{1:n} = (y_1, y_2, ..., y_n)$. A changepoint occurs if there exists $\tau \in \{1, ..., n-1\}$ such that the statistical properties of $\{y_1, ..., y_\tau\}$ and $\{y_{\tau+1}, ..., y_n\}$ are different in terms of a specific criterion (Killick and Eckley, 2014).

Let us consider the following hypotheses: the null hypothesis H_0, which claims the changepoint does not exist (m = 0), and the alternative hypothesis H_1, which states that there is a single changepoint ($m = 1$). The null hypothesis test is based on the likelihood ratio test (LR test). It was Hinkley (1970) who first proposed an approach based on the LR test to detect changepoints. He determined an asymptotic distribution of the test statistic of the LR test when the test concerns average observations from a normal population.

The approach based on the LR test was extended to changepoints for variances as part of observations of the normal distribution (Gupta and Tang, 1987; Silva and Teixeira, 2008; Eckley et al., 2011). In accordance with the methodology of the likelihood ratio test, it is essential to determine the maximum likelihood function (we usually calculate the maximum for a logarithm function), for both the null and the alternative hypothesis.

For the null hypothesis we will determine the maximum log-likelihood as:

$$\log\ p(y_{1:n} \mid \hat{\theta}) \tag{2}$$

where $p(\cdot)$ is the probability density function of the feature being tested, and $\hat{\theta}$ is an estimator of the maximum likelihood of parameter θ. In the alternative hypothesis, a model with a changepoint at τ_1 was considered, where $\tau_1 \in \{1,2,\ldots,n-1\}$. Then the maximum logarithm of the likelihood function for a given τ_1 is specified by the formula:

$$ML(\tau_1) = \log\ p(y_{1:\tau_1} \mid \hat{\theta}_1) + \log\ p(y_{(\tau_1+1):n} \mid \hat{\theta}_2) \tag{3}$$

Rejection of the null hypothesis leads to detection of a changepoint. We then take the estimate of τ_1 to be $\hat{\tau}_1$, the value that maximizes $ML(\tau_1)$.

In determining the changepoints for the standard errors of the *MIR* mean, the Changes in Means and Variance function (cpt.meanvar())in the R package was used to find the changepoint in the mean and variance for the data, by means of the test statistic in the test.stat parameter under assumed normal distribution of the tested feature. The changes were detected using a method which can find a single control point: At Most One Change (AMOC). It was assumed that the changepoint occurs with 95% confidence using the Modified Bayes Information Criterion (MBIC) (Chen and Gupta, 2000).

3. Results

Based on the field studies, a number of observations (regarding species of indicator macrophytes) were made for particular sites in respective quality classes (Fig. 1). The total number of indicator species was 85, of which 14 occurred only in quality class I, 4 only in quality class III, and 10 only in quality class V. About 50% of the species found were detected in all of the analyzed quality classes.

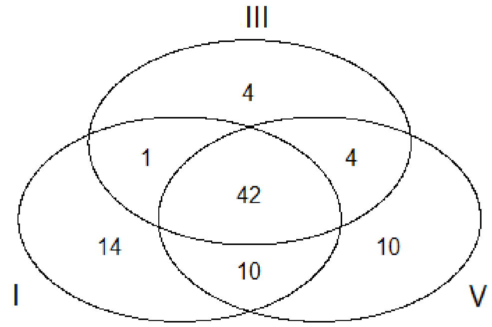

Figure 1. Venn diagram showing numbers of indicator species detected only in quality classes I, III, V, and those common to various quality classes

Indicator species were used to calculate the MIR at particular sites. For each of the examined sites, the MIR value was determined on the basis of formula (1), indicator numbers L and W known from the literature, and the coverage P determined in the field (Table 1). The variance and standard error of the mean were also determined (Table 2).

In quality class I, the highest MIR value (56.11) was obtained at site 32, and the lowest value (35.77) at site 647. The median was 42.98, and the mean was 43.32. The maximum value for quality class III was 46.90 (site 552); the minimum value was 32.41 (site 531); the median was 38.05 and the mean was 37.92. In quality class V, the highest MIR value (38.7) was obtained at site 82, and the lowest value (18.29) at site 437; the median was 29.06 and the mean was 29.01. For each site, mean MIRs were calculated on the basis of simulated deficiencies of one, two, etc. observations in 1000 selections. Standard errors

Table 1. MIR values for sites in the examined quality classes

Quality class	Site no.	MIR	Quality class	Site no.	MIR	Quality class	Site no.	MIR
I	32	56.11	III	3	36.75	V	79	31.88
I	36	44.22	III	14	40.56	V	82	38.07
I	37	42.50	III	58	35.69	V	84	20.00
I	39	42.17	III	66	39.50	V	95	25.97
I	224	46.12	III	109	35.71	V	170	38.70
I	237	45.83	III	141	40.91	V	171	24.12
I	529	46.45	III	210	39.21	V	176	31.25
I	591	46.67	III	226	38.00	V	190	28.43
I	598	45.00	III	236	38.77	V	243	37.80
I	599	41.78	III	435	36.72	V	260	35.87
I	601	42.63	III	439	34.02	V	262	26.18
I	616	38.33	III	451	37.04	V	277	32.31
I	639	43.33	III	462	38.11	V	426	26.68
I	640	39.91	III	469	40.00	V	430	20.39
I	645	42.29	III	531	32.41	V	437	18.29
I	647	35.77	III	552	46.90	V	438	31.50
I	686	44.00	III	634	34.29	V	443	29.69
I	693	38.63	III	694	41.16	V	549	21.70
I	703	43.72	III	697	38.24	V	650	24.41
I	723	40.83	III	707	34.83	V	667	37.03

of the mean were calculated (Table 2, Figure 2) as measures taking account of the diversity of plants and the L_i and W_i values used for MIR determination. It was observed that a small standard error of the mean was associated with a relatively large number of detected species at the site, while a large standard error was associated with a large number of missing taxa.

For example, at site 224, where 19 macrophyte species were detected, the MIR was 46.12 for the full range of species. Next, calculations were made on the basis of 1000 simulations for all combinations of missing species after removing 1, 2, 3, ... macrophyte species respectively. The selection of an equal number of simulations for each site had the aim of ensuring the same precision of the estimators. For all potential combinations of species deficiencies and calculated MIR, the mean MIR at n=1000 and the standard errors of the MIR means were determined.

Table 2. Mean values from simulation for the Macrophyte Index for Rivers ($Mean_{MIR}$) and the standard errors of the MIR means (SEM_{MIR}) at a site of quality class I (no. 224), a site of quality class III (no. 236) and a site of quality class V (no. 82), with given numbers of missing macrophyte species

$No._{missing\ species}$	$Mean_{MIR}$	SEM_{MIR}	$Mean_{MIR}$	SEM_{MIR}	$Mean_{MIR}$	SEM_{MIR}
1	46.12	0.03	38.75	0.019	38.07	0.022
2	46.08	0.05	38.78	0.027	37.99	0.033
3	46.10	0.06	38.75	0.035	37.96	0.042
4	46.16	0.07	38.77	0.042	37.92	0.050
5	46.10	0.08	38.86	0.050	37.78	0.059
6	46.09	0.09	38.80	0.057	37.93	0.067
7	46.02	0.10	38.78	0.063	37.82	0.078
8	46.12	0.11	38.72	0.073	37.75	0.084
9	46.03	0.13	38.58	0.082	37.58	0.097
10	46.05	0.13	38.69	0.093	37.59	0.107
11	45.97	0.15	38.75	0.105	37.46	0.121
12	45.60	0.17	38.72	0.116	37.27	0.139
13	45.60	0.19	38.74	0.133	36.98	0.158
14	45.41	0.20	38.70	0.144	36.73	0.184
15	45.07	0.24	38.92	0.189	36.31	0.212
16	45.26	0.28	39.29	0.226	35.57	0.277
17	44.75	0.30	38.65	0.314	34.21	0.391
18	43.45	0.35	39.69	0.498	-	-

At Most One Change (AMOC) analysis was performed for the values of standard errors of the MIR means at specific sites obtained with zero, one, two, etc. randomly omitted species. Individual changepoints were determined (Table 3).

In quality classes I and III, there were from 6 to 28 indicator species, while in quality class V there were from 6 to 21 species. For the determined MIR mean values and the standard error of the MIR mean, the AMOC analysis proved that for the standard error of the MIR mean, the changepoint occurs at an average of nine missing species in quality classes I and III, or eight species in quality class V. Thus, in the studies conducted, it may be generally assumed that nine different species at the site are necessary for a proper determination of MIR.

Table 3. Values given by At Most One Change (AMOC) analysis for the standard error of the MIR means at a site of quality class I (no. 224), a site of quality class III (no. 23) and a site of quality class V (no. 82)

Quality class I			Quality class III			Quality class V		
$No._{site}$	$No._{species}$	AMOC	$No._{site}$	$No._{species}$	AMOC	$No._{site}$	$No._{species}$	AMOC
32	8	4	3	28	16	79	17	11
36	19	13	14	12	7	82	18	11
37	17	12	58	23	17	84	10	6
39	10	6	66	17	11	95	10	7
224	19	11	109	13	8	170	13	8
237	9	6	141	6	2	171	21	12
529	6	2	210	12	6	176	14	6
591	11	8	226	14	8	190	16	7
598	10	3	236	19	15	243	18	12
599	15	9	435	12	7	260	17	11
601	10	5	439	17	10	262	13	7
616	21	14	451	8	3	277	11	5
639	12	5	462	17	10	426	18	9
640	28	17	469	10	6	430	6	3
645	24	13	531	16	10	437	6	3
647	18	11	552	11	7	438	7	5
686	22	15	634	21	13	443	9	4
693	21	12	694	14	8	549	9	4
703	10	6	697	20	8	650	13	8
723	14	6	707	14	10	667	14	10

The regression of the number of missing species at the site was assessed against the standard errors of the MIR mean. It turned out that linear regression adequately evaluates this relation, as indicated by the high coefficients of determination given in Table 4.

Based on the determined linear regression equations for the analyzed river quality classes, a relationship was determined between the number of observations and the value for which there is a change in the nature of the relationship between the missing observations and the standard errors of the MIR mean. Furthermore, by subtracting from the number of species identified at the site the position of the changepoint, the potential number of species necessary to calculate a reliable MIR was obtained. An adjustment of the linear regression equations was proposed for the obtained value pairs.

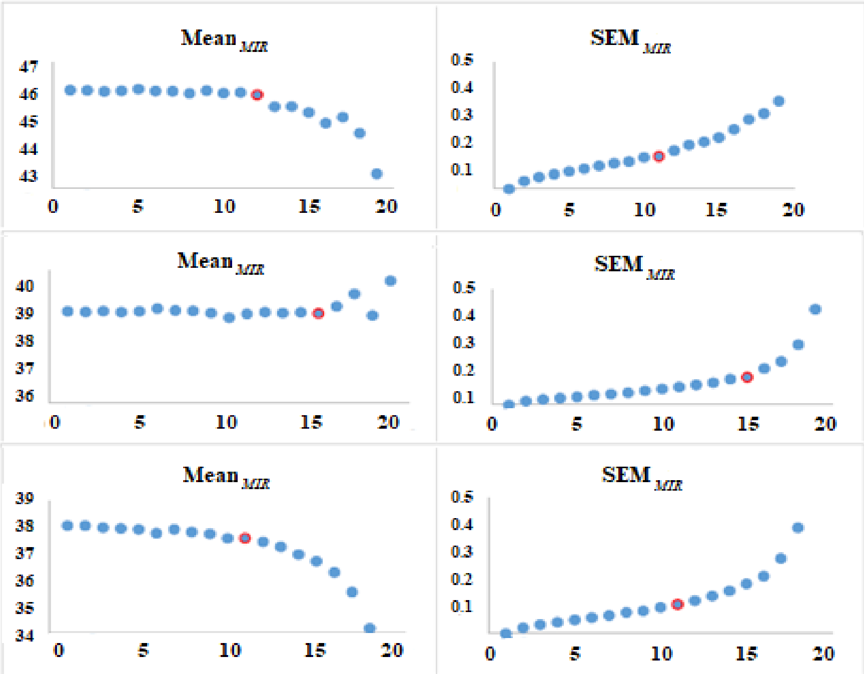

Figure 2. Value of MIR and the standard error of the mean for specific numbers of missing species at a site of quality class I (no. 224), a site of quality class III (no. 236) and a site of quality class V (no. 82)

Table 4. Linear regression for standard errors of the MIR mean, depending on the number of missing observations

Quality class	Regression for number of missing observations	Coefficient of determination
I	y = 0.681x – 1.447	$R^2 = 0.906$
III	y = 0.680x – 1.239	$R^2 = 0.836$
V	y = 0.616x – 0.561	$R^2 = 0.832$

Table 5. Linear regression for the standard errors of the MIR mean, depending on the number of identified macrophyte species at the site

Quality class	Regression for number of identified observations	Coefficient of determination
I	y = 0.319x + 1.447	$R^2 = 0.679$
III	y = 0.3120x +2.239	$R^2 = 0.530$
V	y = 0.384x + 1.561	$R^2 = 0.657$

On the basis of the equations obtained, the study determined the potential number of species that ought to be identified when re-examining the site. Such a relationship is illustrated in Table 6.

Table 6. Predicted changepoints based on linear regression in river quality classes I, III, and V

$No._{species}$	Quality class			$No._{species}$	Quality class			$No._{species}$	Quality class		
	I	III	V		I	III	V		I	III	V
6	3	3	3	14	6	6	6	22	8	8	-
7	4	3	3	15	6	6	6	23	9	9	-
8	4	4	4	16	7	6	7	24	9	9	-
9	4	4	4	17	7	7	7	25	9	9	-
10	5	4	4	18	7	7	7	26	10	10	-
11	5	5	5	19	8	7	8	27	10	10	-
12	5	5	5	20	8	8	8	28	10	10	-
13	6	5	6	21	8	8	-	-	-	-	-

For example, if the actual number (unknown) of species found there was 6, then to estimate a proper MIR index, finding three species would be sufficient, regardless of the cleanliness of the river (Table 6).

Between the sites in the studied quality classes, the differences in individual changepoints predicted with a specific number of observations differ by at most one observation. Therefore, in practice, having a general idea of the level of biodiversity of a given watercourse, it is possible to state approximately how many species would be needed to determine the MIR. It should be noted that in the conducted studies, in some cases, a maximum of 26, 27, or 28 macrophyte species occurred at a site, the majority of which comprised species commonly found in rivers of every quality class. Excluding such sites, Table 6 shows that if there are 9–10 species at any site, one should not be concerned about the full species composition.

4. Discussion

The aim of the study was to check whether it is possible to provide an approximate number of species found at a site which is necessary to obtain the MIR indicator.

Due to the nature of the field studies conducted in medium-sized lowland rivers in Poland, some gaps in the collected data may be expected. Before assessing the quality of waters and consequently taking environmental decisions, it seems necessary to determine the degree of accuracy of the data and to take account of this in further actions. Among the most important sources of uncertainty which significantly affect the variance of indicators and the classification results of various groups of organisms, variability in space and time have already been observed in literature reports (Staniszewski et al., 2006; Carvalno et al., 2013). Among the sources of variability related to macrophytes, significant challenges include the method of data collection in the field (Dudley et al., 2013; Kolada et al., 2010), efficiency in recognizing and classifying the macrophytes, and most importantly, estimation of their coverage. In the source literature, great attention is paid to the unreliability of measurements caused by various errors related to the conduct of the experiment and the other factors mentioned above. However, it is difficult to find unambiguous information that would serve as a practical indication of how many species are necessary to determine the MIR at a site.

This study has shown that the assessment of river plant sampling efforts is difficult to carry out, and advanced statistical and simulation tools may prove very useful. Analyses show that the number of species required for full diversity in each quality class varies between 67 (class I), 66 (class V) and 51 (class III). In total, 85 indicator species were detected in the conducted experiment. This value appears to be relatively small compared with the number of species which constitute the general resources and biodiversity of lowland Poland. It is estimated that the total aquatic flora of watercourses of this type includes about 115 species of vascular plants (Rutkowski, 2008; Bernatowicz and Wolny, 1969).

In addition to typical macrophytes, 63 more terrestrial vascular plants may potentially develop in the waterfront zone (Rutkowski, 2008). This number may be increased by bryophytes, with 10 liverworts and 15 mosses (Jusik, 2012). In addition, 15 surface mosses are regularly recorded in sandy-bottomed lowland rivers (mainly *Bryopsida* and *Mnium* species). Moreover, nine structural algae may potentially occur in waters of this type (*Cladophora*, *Ulva*, *Vaucheria*, *Oedogonium*, *Ulotrix*, *Spirogyra*, *Hildenbrandia*, *Rhizoclonium*, *Stigeoclonuim*). This may indicate that many species could potentially be detected in medium-sized lowland rivers by increasing the sampling effort; at this stage, however, it is difficult to determine how many missing species should be found at a given site. Also, it should be considered whether such a sampling effort is worthwhile. It may be observed that as the number of detected species increases, the standard error of the MIR mean for a given number of missing macrophytes decreases. It may also be seen that due to varying L_i and W_i weight values, different species have different relevance for the determined MIR value.

It has been observed (in agreement with the existing literature) that the cleaner the river, the higher the calculated MIR values (Table 1). The MIR indicator may take a value from 10 (the most degraded) to 100 (the best). In the case of lowland rivers, the highest MIR values do not exceed 60 (Szoszkiewicz et al., 2010). Values consistent with the literature were confirmed in the analyzed experiment and in the simulated data.

The approach used in the study in relation to confidence in changepoint detection was proposed by Hinkley (1970). It originally used a statistical test for the asymptotic distribution of the probability coefficient of the change in mean in the case of a normal distribution; it was later extended to consider the change in variance for a normal distribution, and was further developed by Gupta and Tang (1987), Silva Teixeira (2008), and Eckley et al. (2011) for the Poisson distribution. The analysis revealed the approximate average number of omitted observations which still causes a significant change in the standard errors of the MIR mean. Therefore, it should be taken into account that after exceeding this point, further research will require more and more workload with relatively low cost-effectiveness.

Ultimately, based on linear regression, one may roughly indicate a potential location of the changepoint with a specific number of species. This value, however, should be treated as a guideline. This means that the issues discussed in this study do not fully address the question of the number of macrophyte species necessary to determine the MIR. It is also noteworthy that the changepoints in each of the discussed river quality classes, with a specific number of species detected at the site, may differ. This indicates that importance attaches not only to the number of species detected in a given environment, but also to the specific macrophyte species occurring there, with characteristic L_i and W_i weight values, as well as the various degrees of coverage of macrophytes estimated in the field. Therefore, a necessary further stage of the research will be the assessment of the number of species needed to determine the MIR depending on what kind of species have been identified. This will make it possible to determine the so-called information value provided by specific indicator species of macrophytes to the MIR value in a given river type.

Depending on individual needs, before determining the MIR, it may be assumed that if a site contains nine species of indicator macrophytes, the standard error of the mean MIR will be small. However, this does not negate the fact that perhaps in certain specific situations fewer species would be enough, and in rare cases, finding another species would strengthen the obtained results.

5. Conclusions

- The presented study has shown that the accuracy of the determined MIR depends to a large extent on the completeness of the test sample taken at the site, regardless of the river quality class.
- Simulation analyses have shown the approximate average number of species of indicator plants needed to obtain the MIR value, with a slight error resulting from underestimation of the number of species at the site.

- In order to make more detailed calculations, one should take into account the amount of information provided by each single species with distinguishable *L* and *W* parameters, which may result in the saving of field work or greater precision in the analysis.

REFERENCES

Bacon D.W., Watts D.G. (1971): Estimating the transition between two intersecting straight lines. Biometrika 58: 525-534.

Bernatowicz S., Wolny P. (1969): Fisherman's botany (in Polish). Państwowe Wydawnictwo Rolnicze i Leśne, Warszawa, Poland.

Carvalno L.B., Rojano-Delgado A.M, Costa Aguiar Alves P.L., Prado R. (2013): Differential content of glyphosate and its metabolites in Digitaria insularis biotypes. Communications in Plant Sciences 3(3-4): 17-20.

Ceschin S., Salerno G., Bisceglie S., Kumbaric A. (2010): Temporal floristic variations as indicator of environmental changes in the Tiber River in Rome. Aquatic Ecology 44: 93–100. doi: 10.1007/s10452-009-9292-1.

Chen J., Gupta A.K. (2000): Parametric statistical change point analysis, Birkhauser.

Dawson F.H., Newman J.R., Gravelle M.J., Rouen K.J., Henville P. (1999): Assessment of the Trophic Status of Rivers Using Macrophytes. Evaluation of the Mean Trophic Rank, R&D Technical Report E39, Environment Agency.

Dudley B., Dunbar M., Penning E., Kolada A., Hellsten S., Oggioni A. et al. (2013): Measurements of uncertainty in macrophyte metrics used to assess European lake water quality. Hydrobiologia 704: 179–191. doi: 10.1007/s10750-012-1338-z.

Official Journal (2016). Regulation of the Minister of the Environment of July 21, 2016 on the method of classification of the status of uniform surface water bodies and environmental quality standards for priority substances. Official Journal 2016. item 1187 (in Polish).

Eckley I.A., Fearnhead P., Killick R. (2011): Analysis of Changepoint Models. In Barber D., Cemgil A.T., Chiappa S. (eds.), Bayesian Time Series Models. Cambridge University Press.

Erdman C., Emerson J.W. (2007): An R Package for Performing a Bayesian Analysis of Change Point Problems. Journal of Statistical Software 23(3): 1-13.

Gupta A.K., Tang J. (1987): On Testing Homogeneity of Variances for Gaussian Models. Journal of Statistical Computation and Simulation 27(2): 155-173.

Haslam S.M. (1982): A proposed method for monitoring river pollution using macrophytes. Environmental Technology Letters 3: 19-34. doi: 10.1080/0959333 8209384094

Hinkley D.V. (1969): Inference about the intersection in two-phase regression. Biometrika 56: 495-504.

Hinkley D.V. (1970): Inference about the Change-Point in a Sequence of Random Variables. Biometrika 57(1): 1-17.

Hirotsu C., Harukazu T. (2017): An algorithm for a new method of change-point analysis in the independent Poisson sequence. Biometrical Letters 54(1): 1-24.

Holmes N.T.H., Newman J.R., Chadd S., Rouen K.J., Saint L., Dawson F.H. (1999): Technical report no E38 Environmental Agency of England Wales. Bristol: UK; Mean trophic rank: a user's manual R.

Hudson D.J. (1966): Fitting segmented curves whose joint points have to be estimated. Journal of the American Statistical Association 61: 1097-1129.

Jusik S. (2012): Identification key to mosses and water liverwords required to the ecological status assessment of surface waters in Poland (in Polish). Biblioteka Monitoringu Środowiska, Warszawa, Poland.

Killick R., Eckley I., Haynes K. (2014): Changepoint: An R Package for Changepoint Analysis. R package version 1.1.5, URL http://CRAN.R-project.org/package =changepoint.

Killick R., Eckley I.A., Jonathan P., Ewans K. (2010): Detection of Changes in the Characteristics of Oceanographic Time-Series using Statistical Change Point Analysis. Ocean Engineering 37(13): 1120-1126.

Kirby E.J.M. (1974): Ear development in spring wheat. Journal of Agricultural Science 82: 437-447.

Kolada A. (2010): The use of aquatic vegetation in the lake assessment: testing the sensitivity of macrophyte metrics to anthropogenic pressures and water quality. Hydrobiologia 656: 133-147. doi: 10.1007/s10750-010-0428-z.

Kolkwitz R., Marsson M. (1902): Grundsatze fur die biologische Beurteilung des Wassers nach seiner Flora und Fauna. Mitt. Priifungsanst. Wasserversorg. Abwasserein 1: 33-72.

Lavielle M. (2005): Using Penalized Contrasts for the Change-Point Problem. Signal Processing 85(8): 1501-1510.

Lerman P.M. (1980): Fitting segmented regression model by grid search. Appl. Stat. 29: 77-84.

Nam C.F.H., Aston J.A.D., Johansen A.M. (2012): Quantifying the Uncertainty in Change Points. Journal of Time Series Analysis 33(5): 807-823.

Pruska K. (1996): Metody regresji przełącznikowej i ich zastosowanie (in Polish). Wyd. Uniw. Łódzkiego, Łódź.

Reeves J., Chen J., Wang X.L., Lund R., Lu Q. (2007): A Review and Comparison of Changepoint Detection Techniques for Climate Data. Journal of Applied Meteorology and Climatology 46(6): 900-915.

Rutkowski L. (2008): Identification key to vascular plants of Polish Lowland (in Polish). Wydawnictwo Naukowe PWN, Warszawa, Polska.

Seber G.A.F., Wild C.J. (1989): Nonlinear Regression. Wiley, New York.

Silva E.G., Teixeira A.A.C. (2008): Surveying Structural Change: Seminal Contributions and Bibliometric Account. Structural Change and Economic Dynamics 19(4): 273-300.

Sprent P. (1961): Some hypotheses concerning two phase regression lines. Biometrix 17: 634-645.

Staniszewski R., Szoszkiewicz K., Zbierska J., Leśny J., Jusik S, Clarke R.T. (2006): Assessment of sources of uncertainty in macrophyte surveys and the consequences for river classification. Hydrobiologia 566: 235-246.

Starmach K., Wróbel S., Pasternak K. (1976): Hydrobiologia. Limnologia. Warszawa: PWN.
Szoszkiewicz K., Zbierska J., Jusik S., Zgoła T. (2010): Macrophyte method for river assessment. Poznan: Bogucki Wydawnictwo Naukowe.
Szoszkiewicz K., Budka A., Pietruczuk K., Kayzer D., Gebler D. (2017): Is the macrophyte diversification along the trophic gradient distinct enough for river monitoring? Environtal Monitoring and Assessment 189: 4. doi:10.1007/s10661-016-5710-8.
Watts D.G., Bacon D.W. (1974): Using an hyperbola as a transition model to fit two-regime straight-line data. Technometrics 16: 369-373.
Wiegleb G. (1979): Der Zusammenhang zwischen Gewässergüte und Makrophyten Vegetation in niedersächsischen Fließgewässern. Landschaft + Stadt 11: 32-35.
Zeileis A., Leisch F., Hornik K., Kleiber C. (2002): Strucchange: An R Package for Testing for Structural Change in Linear Regression Models. Journal of Statistical Software 7(2): 1-38. URL http://www.jstatsoft.org/v07/i02/.
Zeileis A., Shah A., Patnaik I. (2010): Testing, Monitoring, and Dating Structural Changes in Exchange Rate Regimes. Computational Statistics & Data Analysis 54(6): 1696-1706.
Zelinka M., Marvan P. (1961): Zur Präzisierung der biologischen klassifikation der Reinheit flieβender Gewässer.-Arch. Hydrobiol. 57: 389-407.
Zimny H. (2006): Ekologiczna ocena stanu środowiska. Bioindykacja i biomonitoring (in Polish). Agencja Reklamowo-Wydawnicza Arkadiusz Grzegorczyk, Warszawa.