JEE Journal of Ecological Engineering

Journal of Ecological Engineering 2024, 25(1), 49–64
https://doi.org/10.12911/22998993/173191
ISSN 2299–8993, License CC-BY 4.0

Received: 2023.08.06
Accepted: 2023.10.03
Published: 2023.12.04

Field Measurements and Machine Learning Algorithms to Monitor Water Quality in Lakes Located in Landscape Parks – A Case Study

Natalia Walczak[1*], Zbigniew Walczak[2], Ireneusz Laks[2]

[1] Department of Hydraulic and Sanitary Engineering, Poznan University of Life Sciences, ul. Wojska Polskiego 28, 60-637 Poznan, Poland

[2] Department of Construction and Geoengineering, Poznan University of Life Sciences, ul. Wojska Polskiego 28, 60-637 Poznan, Poland

* Corresponding author's e-mail: natalia.walczak@up.poznan.pl

ABSTRACT

One of the greatest threats to many lakes is their accelerated eutrophication resulting from anthropogenic pressure, agricultural intensification, and climate change. A very important element of surface water protection in environmentally conserved areas is the proper monitoring of water quality and detection of potential threats by examining the physicochemical properties of water and performing statistical analyses that enable possible exposure of unfavourable trends. The article presents the analyses of the results of measurements made in three lakes located in the Sierakowski Landscape Park. As part of the measurements, water quality indicators i.e., phosphorus, nitrogen, BOD_5 and COD, were determined monthly for a year at the inflows and outflows of the studied lakes. The test results of selected water quality indicators were analysed using machine learning algorithms i.e., PCA and *k*-means. The conducted tests enabled statistical estimation of changes in water quality indicators in the reservoirs and evaluation of their correlation.

Keywords: quality of water in lakes, phosphorus, nitrogen, BOD_5, COD, PCA, *k*-means, ANOVA, Kruskal-Wallis test.

INTRODUCTION

Lakes are one of the most important water resources and generally account for approx. 0.3% of all surface water. Due to the rapid development of economy and agriculture, various environmental problems have arisen, water in lakes has been constantly deteriorating and the process of lacustrine eutrophication has gradually become an urgent problem worldwide [Liu et al., 2011; Nyenje et al., 2010]. Therefore, it is absolutely necessary to properly assess the quality of surface water [Ferahtia et al., 2021].

Water quality in reservoirs is affected by numerous physical, chemical, and biological parameters as well as their mutual influence.

The quality of lake waters depends on the geological structure of the catchment area as well as on anthropogenic activities in their surroundings i.e., construction, landfills, agriculture and other related activities [Mahananda et al., 2010; Mulu and Mehari, 2013]. According to Bhateria and Jain [Bhateria and Jain, 2016], lakes form a unique structure that can serve for studying complex interactions between different water quality parameters and make a one-of-a-kind ecosystem, completely different from land or air.

For some time now, as a result of anthropogenic pressure, the once clean lakes have been losing their aesthetic, utilitarian and recreational values more and more often. Their progressive degradation is closely related to the intensification of eutrophication [Smith, 2003; Withers et al., 2014] due to, among others, the growth of civilisation embodied by technical progress, industrial development, urbanisation and motorisation. Lake ecosystems are more susceptible to pollution than flowing waters [Guz and Doroszkiewicz, 2003]. The very causes of eutrophication are complex as they include various ecological,

social, economic, and many other factors [Álvarez et al., 2017]. Sustainable water management in lake catchment areas requires an understanding of dominants shaping eutrophication processes [Ferencz et al., 2017].

Studies on the loss of social and economic benefits in the context of the deteriorating quality of water in lakes were conducted, among others, by Crase and Gillespie [2008]. They found that when algae bloomed in Lake Hume in Australia, profits (from recreational activities such as sailing, fishing and swimming) decreased by a third.

The Water Quality Index (WQI) is a popular tool for assessing surface water quality [Uddin et al., 2023a, 2021]. Uddin et al. [Uddin et al., 2021] points to up to 20 different WQI models, depending on the type of study area, used in many countries around the world. However, WQIs are usually developed based on region-specific guidelines and are therefore not generic. Moreover, they create uncertainty in converting large amounts of water quality data into a single indicator. The use of mathematical and machine learning techniques, such as principal component analysis and cluster analysis (PCA), can better inform the selection of parameters and their weights, while computer techniques such as fuzzy interface systems and artificial neural networks reduce the uncertainty resulting from the final aggregation process [Uddin et al., 2023a].

An improved water quality index (WQI) model using Cork Harbor as an example was presented by [Uddin et al., 2022a; Uddin et al., 2022b]. The model uses the XGBoost machine learning algorithm to rank and select water quality indicators for inclusion based on relative importance to overall water quality status. The authors ranked the indicators by two seasons (summer and winter) and tested eight sub-indicator aggregation functions (five from existing WQI models and three proposed by their authors). The indices proposed by the authors were identified as the best functions, characterised by less ambiguity than others. Uddin et al. [2023c] also proposed a novel approach for estimating and predicting uncertainty in the water quality index model using machine learning methods and for predicting water status [Uddin et al., 2023b].

Assessments of the impact of land use and land cover on river water quality using a water quality index and remote sensing techniques were made by Gani et al., [2023] based on 12 water samples from the Buriganga, Dhaleshwari, Meghna, and Padma rivers during the 2015 winter season and seven water quality indices.

In Poland, water quality in surface waters is determined by regulation, which includes biological, hydromorphological and physicochemical indicators [Journal of Laws, item 1475, 2021].

Water quality datasets are now often analysed using Machine Learning (ML) algorithms. It is an area of artificial intelligence that focuses on developing the methods, algorithms, and computer models that can "learn" from data and events, without requiring explicit programming [Bishop, 2016].

Machine learning is applied in various aspects of data analysis of water quality [Ahmed et al., 2019; Bui et al., 2020; Chen et al., 2020; Rodríguez-López et al., 2023; Sagan et al., 2020; Zhu et al., 2022], such as water-quality prediction [Rodríguez-López et al., 2023], identification of pollution sources, water quality classification [Dezfooli et al., 2018; Haghiabi et al., 2018; Uddin et al., 2023b], monitoring and alerting.

The machine learning techniques used in data analysis of water quality include classification and regression algorithms, decision trees, neural networks, and data grouping methods. Also, an important aspect is the proper preparation of training data and taking into account the specific features and context of a given ecosystem in analyses [Bishop, 2016]. The results of analyses and predictive models based on ML also require constant verification, primarily through direct tests.

The purpose of this research was to assess the water quality in selected reservoirs located in a partially protected area, which is a landscape park, in order to statistically estimate changes in water quality indicators in these reservoirs and to verify the possibility of using machine learning algorithms in monitoring of water quality based on a small set of data. Three selected water reservoirs, Lake Chrzypskie, Lake Białokoskie and Lake Kuchenne, were located within the Sierakowski Landscape Park (western Poland). Although these water bodies are in close proximity to each other, they differ in terms of land use in the direct catchment area. It was assumed that these methods will indicate the measurement points that differ in the values of physico-chemical water parameters from the other parameter values contained in the measurement dataset. These tests allowed for quickly and easily identifying outliers and comparing the mean values of water quality indicators between the studied lakes at their inflows and outflows, respectively.

Outliers may indicate unfavourable processes in the lake due to the partially different nature of this part of the catchment and/or the location of pollution sources near the lake. This will allow water managers to quickly detect and respond to the changes in water quality within the total catchment.

MATERIAL AND METHODS

Description of the study domain

The Sierakowski Landscape Park with a total area of over 300 km^2 was established in 1991 and is located in the north-western part of the Greater Poland Voivodeship. It is an example of a post-glacial landscape with lake gutters, river valleys, dunes, moraine hills and forest complexes.

The land use structure in the Sierakowski Landscape Park [Central Statistical Office, 2010] includes agricultural lands in 51%, woodlands (mainly mixed) account for 1/3 of the entire park. Lakes and rivers cover 7% and urban areas 8% of the entire park, respectively.

For the purpose of the study three lakes i.e., Chrzypskie, Białokoskie and Kuchenne, the locations of which are shown in Figure 1, were analysed.

Lake Białokoskie (Fig. 1) with an area of 144.2 ha, a maximum length of 3373.4 m, 11279.0 m of shoreline, with an average width of 427.3 m, an average depth of 9.6 m and a volume of 14013.1 thousand m^3 is all surrounded by woodlands. The lake is distinguished by great fishing values due to clean water and good access to the shore. A great advantage of the lake is its varied bottom, with large slopes and numerous deeps. Approx. 90% of the shoreline is covered with submerged vegetation, to a depth of 3 m.

In terms of the balance type, the lake is described as flow-through. The watercourse flowing into and out of Lake Białokoskie is Mianka, and the lake is additionally fed by several other small watercourses. The groundwater level varies between 1-20 m below ground level.

The total catchment area, with 3033 ha, is largely covered with arable lands and woodlands (over 87%). On the other hand, the ratio of the lake area to the direct catchment area, which is 291.8 ha, is almost 50%, the rest is woodlands (31.91%) and arable lands (15.32%) [Bródka and Macias, 2016].

Lake Białokoskie was classified as natural type 3a [WIOŚ in Poznań, 2018], with high catchment impact, stratified, lake was assigned purity class V. It was observed that lake waters are

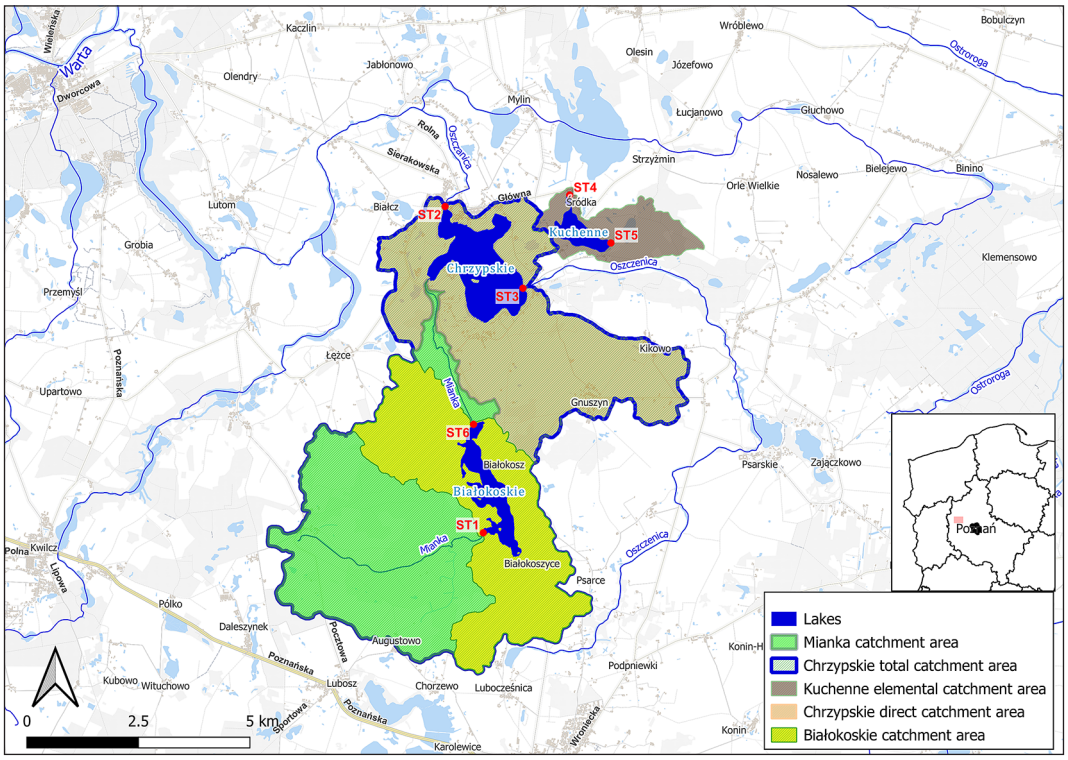

Figure 1. Location of selected lakes

sensitive to eutrophication caused by urban factors and are at risk of failing to meet the environmental and conservation objectives of habitats and species. Water quality tests were repeated after 3 years and class III were obtained [Chief Inspectorate for Environmental Protection, GIOS., 2020].

Administratively, Lake Chrzypskie (Fig. 1) is in the Międzychód poviat in the Chrzypsko Wielkie commune. The reservoir with an area of 300.9 ha and a volume of 18654.0 thous. m^3 is located on gutters with a flat bottom. Its catchment is adjacent to outwash plains, monadnocks and terrace steps. Morphometrically, the lake is characterised by the following maximum parameters: length 2802.9 m, width 1590.9 m, depth 150 m, shoreline length 12077.0 m. The basic and most important function of the lake is recreation (swimming and fishing). Lake Chrzypskie is of a flow-through nature and fed by the Oszczenica River. Also, the Mianka watercourse flows into the lake, connecting it with Lake Białokoskie.

The total catchment area is mostly covered with arable lands. A large part is also occupied by woodlands, meadows, and pastures; these three cover structures account for almost 92% of the total catchment area. The ratio of the lake area to the total catchment area is 2.5%. The total catchment area is 12071.8 ha. The total area of the direct catchment is 560.4 ha, and the ratio of the lake area to the direct catchment area is 53.7%. A large part of the direct catchment is covered by arable lands, woodlands, and meadows or pastures; these three cover structures occupy 41% of the total catchment area.

According to the Voivodship Inspectorate for Environmental Protection, in 2015 the ecological status of lake water bodies was assessed as bad [WIOŚ in Poznań, 2015]. On the other hand, the GIOS tests from 2021 [Chief Inspectorate for Environmental Protection, GIOS., 2021] showed class III. According to Kudelska [1994], the lake was classified as moderately susceptible to degradation, abiotic type 3a, with Schindler's coefficient >2.

Kuchenne Lake is the last analysed reservoir (Fig. 1), with a water table surface of 63.08 ha and one island. The water table is located at 47.8 m a.s.l., and the shoreline is 4.6 km long. The reservoir is located on flat-bottomed gutters, its catchment area is adjacent to outwash plains and monadnock hills.

Kuchenne Lake was classified as type 2a with high calcium content, low catchment impact, stratified. The area around the lake is mainly used for agricultural purposes, development areas and woodlands cover a much smaller part of it. Due to the lack of data on the use of the catchment area of Lake Kuchenne, the division of the land structure was estimated based on available maps.

It was estimated that the direct catchment area is small, 42% of the entire catchment area is the lake area. The largest part is covered by arable lands, which covers approx. 27% of the total catchment area. The forest cover was estimated as 13%, and agricultural wastelands as 10%. The total area of the direct catchment is estimated as 150 ha.

On the basis of GIOS 2020 [Chief Inspectorate for Environmental Protection, GIOS., 2020], Lake Kuchenne was characterised by class III.

None of the analysed lakes has full documentation on water quality tests regarding biological and hydro-morphological parameters. For this reason, the water quality was tested in terms of chemical parameters, extending the methodology with the ones that are not required in accordance with the regulation. In addition, the study included the type of use of the direct catchment area as an element that determines the quality of water.

In the period from 06.2020–06.2021, the sum of monthly rainfall, recorded at the meteorological station in Nojewo (about 5 km from the lake Chrzypsko) did not exceed 80 mm, the maximum rainfall at 21.4 mm. The sum of monthly rainfall and its distribution over the area of the subcatchments of each lake is analogous. The subcatchments occupy a small area and are adjacent to each other.

Research methodology

In order to analyse water quality thoroughly, test samples were taken at the beginning of each month from June 2020 to June 2021. The research was conducted for 13 months, which was justified by the need to take into account different periods of vegetation and post-vegetation.

Water quality analysis was carried out on the basis of water samples taken at the inflow and outflow for one year from Lake Białokoskie, Lake Chrzypskie and Lake Kuchenne (Fig. 1). Sampling locations (Fig. 1) allowed for estimating the impact of the direct catchment area on the obtained results. The same measurement sites were also chosen by Kanclerz et al. [2015], in order to compare the quality of water at two points

located above and below the Stare Miasto reservoir. In turn, Janicka et al. [Janicka et al., 2016] used a similar study period (1 year) to assess the water quality of Lake Raczyńskie.

Water samples allowed for determining the basic parameters [Journal of Laws, item 1475, 2021] related to water quality. In addition to basic determinations P (PN-ISO 8466-1:2003), NH_4^+ (PN ISO 7150-1:2002), Total N (PN-EN ISO 11905-1:2001), NO_2 (PN EN 2677:1999), NO_3 (PN-C-04576-08:1982), chemical oxygen demand COD (PN-ISO 15705:2005) and biochemical oxygen demand BOD_5 (PN-EN 1899-2:2002) were determined. Then, the content of total mineral and organic suspension was determined using the quantitative determination of mass (PN-C-04559-02:1972). Analyses were also performed using ML methods: PCA (Principal Components Analysis) and *k*-means. The results were confirmed by Shapiro-Wilk, ANOVA, Kruskal-Wallis, and Tukey HSD/Tukey Kramer tests.

Statistical analysis and machine learning algorithms

At the planning stage of this experiment, an assumption was made that the conducted field research, in addition to determining the quality of water in selected lakes, would also be used to check whether the commonly known and well-described ML algorithms could be used in the analyses supporting monitoring, despite the small size of the dataset. It was also assumed that the algorithms are available in open-source systems such as the R environment and will not generate additional costs when used for the analysis. These criteria are met by the PCA and *k*-means methods.

Principal Components Analysis (PCA) [Abdi and Williams, 2010; Deutsch and Beinker, 2019; Jolliffe, 2002; Kurita, 2019] is a frequently used tool for exploratory data analysis through the reduction of data dimensionality. As a non-parametric method, it does not require any assumptions on the distribution of the data under study. In this method, the input set of correlated features is replaced by a small number of uncorrelated ones i.e., principal components. Together, they can explain almost all data variability. The first component explains the most of variability. The second component is chosen in such a way that it is not correlated with the first one and explains as much of the remaining variability as possible. If one wants to reduce the dimensionality of data, it must be considered how many components to choose for further analysis. New data can be visualised using one graph, called a biplot. *K*-mean clustering [Hartigan and Wong, 1979] was performed, which was used for initial data analysis, identification of relatively homogeneous groups of observations based on selected characteristics, and verification of the structure of the analysed data. *K*-means is one of the most popular cluster analysis methods in statistics and machine learning. It is a non-hierarchical data clustering algorithm that divides the dataset into *k* clusters, where *k* is a predetermined number.

The article uses two machine learning methods: PCA and *k*-means. The aim was to check whether both methods would give similar results. It was pointed out that PCA is a broader method that not only allows clustering of data but also indicates which data series are correlated with each other. It is important that both methods identified the same measurement stations as different from the others, which may indicate dangerous phenomena occurring nearby that affect water quality.

The experiment assumed the testing of 11 parameters describing water quality at 6 measurement points for 1 year at intervals of 1 month. The total dataset included 792 measurements of physical and chemical parameters monitoring water quality in the aforementioned reservoirs. The appearance of clusters could indicate changes in water quality (beneficial or unfavourable) described by a group or a single parameter. Such an indication allowed for pointing out the positions and parameters that should be analysed in more detail.

Basic statistical analyses (descriptive statistics) were also performed to compare the means between inflows and outflows from the selected lakes for individual indicators i.e., phosphorus, total nitrogen, BOD_5 and COD. The mean equality analysis was performed using ANOVA for normal distribution dataset and the Kruskal-Wallis test for dataset which were not characterised by a normal distribution. The following hypotheses were adopted for the testing, according to the assumptions of the ANOVA/Kruskal-Wallis method, H_0 meaning equality of means between the analysed indicators on the inflows H_0: $\mu_{ST1} = \mu_{ST3} = \mu_{ST5}$ or outflows H_0: $\mu_{ST2} = \mu_{ST4} = \mu_{ST6}$); and H_1 - at least one indicator gives a different mean. Statistical and ML analyses were carried out in the R environment.

RESULTS

Measurements of indicators

The direct catchment area of Lake Białokoskie is located mainly in woodlands and agricultural lands. The water samples taken and then tested showed that the quality of water from the reservoir was unsatisfactory. Compared to other lakes (Table 1 and Figures 2-5), high loads of total nitrogen were recorded at the inflow (annual average 11.89 mgN/l, SD = 5.48), and only slightly higher loads of phosphorus (2.28 mgP/l, SD = 1.32). The concentrations at the outflow, however, remained at similar levels as in other lakes. Slight decreases in phosphorus concentrations were recorded between March and May. On the other hand, between January and May, a significant increase in total nitrogen concentrations was noted. Quite a large unevenness was also noted for COD indicators, both at the inflow and outflow from the lake (average at the inflow 29.23 and SD = 15.90, average at the outflow 37.09 and SD = 13.92). For BOD_5, this unevenness was significantly smaller (average at the inflow 2.42 and SD = 1.04, at the outflow 2.64 and SD 1.34) and it concerned primarily the outflow.

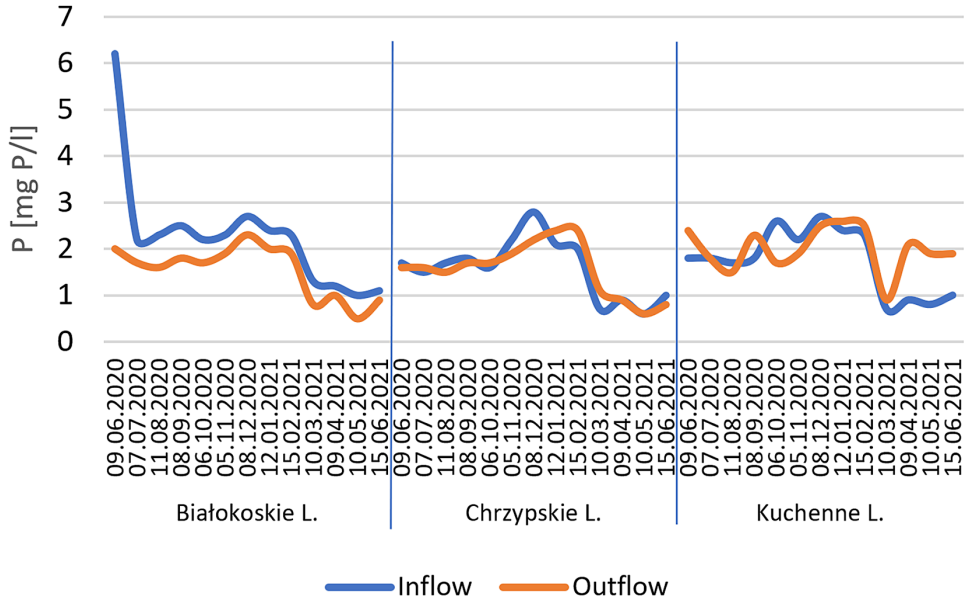

Figure 2. Variability of phosphorus concentration [mg P/l] over time in the studied lakes

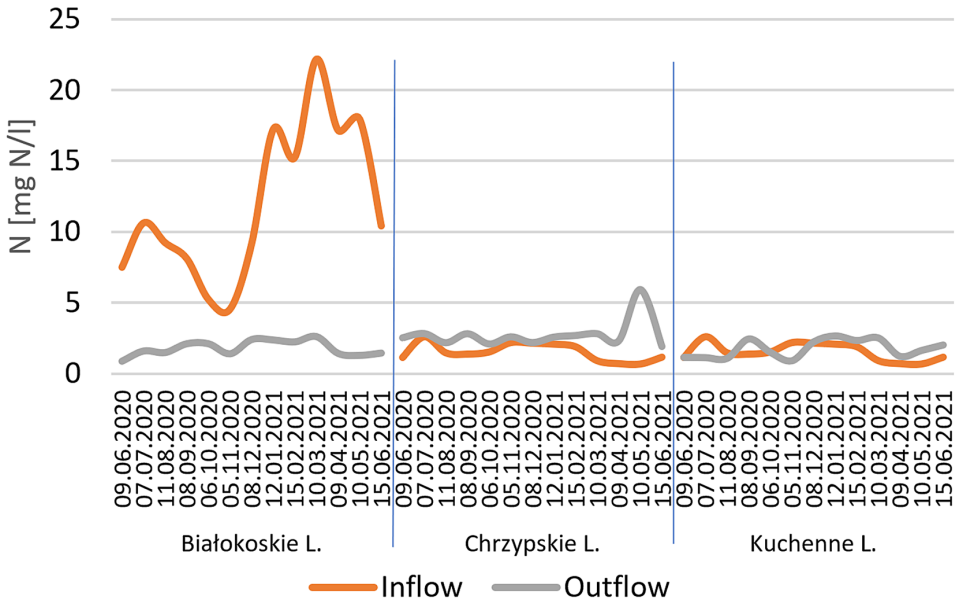

Figure 3. Variability of total nitrogen concentration [mg N/l] over time in the studied lakes

The water quality of Lake Chrzypskie was characterised as average (Table 1 and Figures 2–5). Both loads at the inflow and concentrations at the outflow for nitrogen and phosphorus were recorded at an even level (averages at the inflow 1.52 mgN/l and SD = 0.62, 1.58 mgP/l and SD = 0.64, whereas at the outflow 2.74 mgN/l and SD = 1.01, 1.57 mgP /l and SD = 0.59). However, a significant decrease in phosphorus concentrations was observed between March and May, even more than twofold. For COD and BOD_5, significant variability of indicators was noted both at the inflow and outflow.

The quality of water in Lake Kuchenne, similarly to Lake Chrzypskie, was determined as average (Table 1 and Figures 2–5). Phosphorus concentrations in the analysed period were the same both for the inflow and outflow (average at the inflow 1.75 mgP/l and SD = 0.70, and at the outflow 2.0 mgP/l and SD = 0.48). BOD_5 and COD values fluctuated significantly, more than in other lakes (max-min = 50 for COD and 11 for BOD_5).

In accordance with the current Regulation of the Minister of Infrastructure (Journal of Laws, item 1475, 2021) the limit content of total nitrogen must not exceed 1.0 mg N/l for Class I and 1.4 for Class II. For phosphorus, the limiting value must not exceed 0.04 and 0.06 mg P/l for Class I and II, respectively. If the limit values are exceeded,

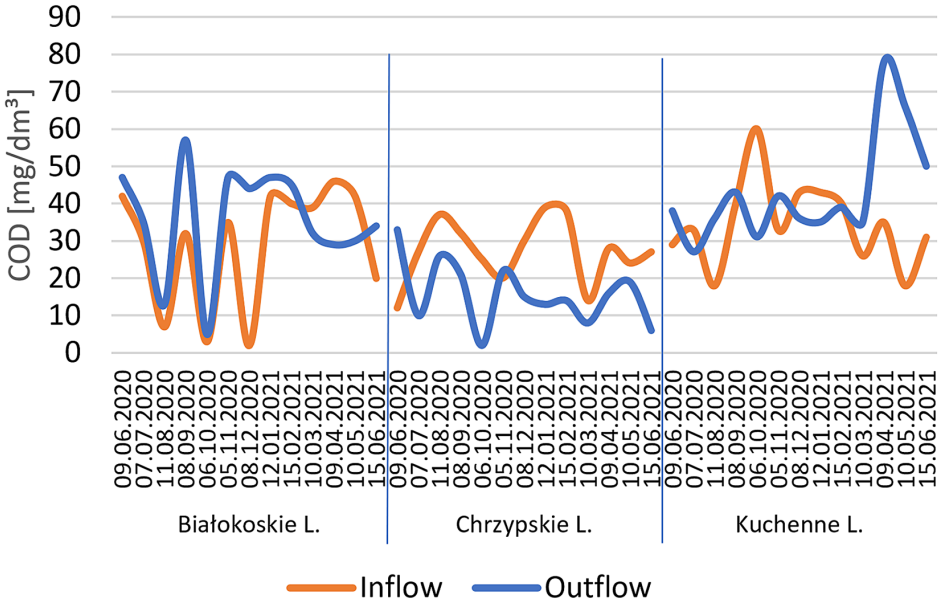

Figure 4. Variability of the COD indicator over time in the studied lakes

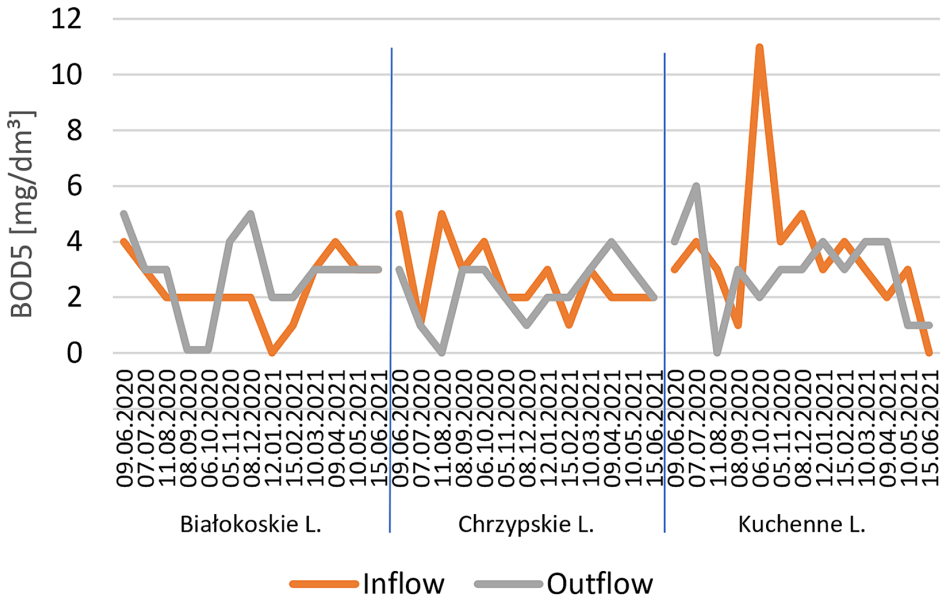

Figure 5. Variability of the BOD_5 indicator over time in the studied lakes

classes below II are not defined. These indicators were exceeded in practically all of the lakes analysed, both at the inflow and outflow, resulting in the adoption of a water quality class below II.

ML analyses

Principal component analysis (PCA) covered 11 parameters describing water quality in the studied lakes. Table 1 shows the dataset used in PCA.

The first stage involved the analysis of the complete dataset. The PCA results for the complete dataset (Figure 6) indicate which parameters have the greatest impact on the principal components PC1 and PC2. The graph clearly shows that NO_2, NO_3, N_Total, pH and P for the Białokoskiein measurement points (ST1) significantly differ from other points and that they have a large impact on the principal component PC2. The ST2 stand (Chrzypskie-out) is also different from the data in terms of the loads of suspensions. These components significantly affect the principal component PC1. However, they were not taken into account for further analysing, because suspension (total, mineral and organic) is an indicator of the quality of flowing waters (watercourses, streams, rivers). The remaining test stands (ST3, ST4, ST5, ST6) are concentrated in one cluster in the 1st and 4th quadrants of the biplot (Figure 6). These analyses also confirm the results obtained by the *k*-means method, where a large correlation of NO_2, NO_3, N_Total and P can be observed.

In the next step, the dataset was reduced to five parameters i.e., pH, P, NO_2, NO_3 and NH_4. N_Total, which reached values very close to NO_3 (the angle between the N_Total and NO_3 vectors indicated a correlation close to 1), was omitted.

PCA was performed again, determining the importance of individual components (Table 2), followed by the analysis of variance and the biplot chart (Figure 7). The analysis shows that for the reduced dataset, the principal components PC1 and PC2 can describe approx. 92% of the reduced dataset.

The biplot clearly shows that the Białokoskiein measurement point differs from other measurement points and significantly affects the principal component PC1.

Table 1. Values of average annual physico-chemical and biological parameters

Location	Index	pH	P	NH_4	NO_2	NO_3	N Total	COD	BOD_5	Total suspension	mineral suspension	Organic suspension
		[-]	[mg P/ dm^3]	[mg NH_4^+/ dm^3]	[mg NO_2^-/ dm^3]	[mg NO_3^-/ dm^3]	[mg N/ dm^3]	[mg O_2/ dm^3]	[mg O_2/ dm^3]	[mg/dm^3]	[mg/dm^3]	[mg/dm^3]
Bialokoskie-in (ST1)	min-max	6.6-8.03	1-6.2	0.07-0.91	0.05-0.61	4.3-22	4.53-22.18	2-46	0.5-4	12.8-57.2	3.6-23.8	5.9-51.4
	Mean	7.38	2.28	0.24	0.11	11.54	11.89	29.23	2.42	30.96	10.34	20.68
	SD	0.35	1.32	0.22	0.15	5.55	5.48	15.90	1.04	12.93	5.58	12.82
Bialokoskie-out (ST6)	min-max	6.88-7.82	0.5-2	0.07-0.35	0.02-0.05	0.6-2.5	0.86-2.62	5-57	0.5-5	9.4-57.3	2.9-23.3	6.3-54.4
	Mean	7.32	1.47	0.19	0.04	1.55	1.77	37.09	2.64	31.31	8.11	23.28
	SD	0.29	0.56	0.09	0.01	0.57	0.55	13.92	1.34	14.87	7.26	15.46
Chrzypskie-in (ST3)	min-max	6.9-7.8	0.6-2.8	0.06-0.72	0.02-0.07	0.3-2.3	0.66-2.61	12-39	1-5	4.4-60.3	2.8-35.4	4.9-50.4
	Mean	7.26	1.58	0.29	0.04	1.19	1.52	27.15	2.69	25.11	9.98	17.89
	SD	0.27	0.64	0.17	0.02	0.53	0.62	8.44	1.32	14.81	9.45	12.90
Chrzypskie-out (ST2)	min-max	6.83-8.14	0.6-2.4	0.06-0.4	0.02-0.05	1.8-5.6	1.93-5.94	2-33	0.5-4	11-78.8	3.8-48.8	5-51.7
	Mean	7.58	1.57	0.16	0.04	2.54	2.74	15.77	2.27	40.30	16.19	24.56
	SD	0.42	0.59	0.09	0.01	0.96	1.01	8.51	1.01	18.47	12.75	14.29
Kuchenne-in (ST5)	min-max	6.9-7.73	0.7-2.7	0.1-0.7	0.02-0.06	0.4-2	0.52-2.15	18-60	0.5-11	17.8-137.4	2.2-60.3	1-77.1
	Mean	7.18	1.75	0.28	0.04	0.95	1.27	34.54	3.58	37.71	13.55	23.59
	SD	0.29	0.70	0.19	0.01	0.38	0.44	11.25	2.55	31.29	16.46	19.94
Kuchenne-out (ST4)	min-max	6.9-7.8	0.9-2.6	0.07-0.92	0.03-0.32	0.7-2.2	0.91-2.64	27-78	0.5-6	7.4-53.8	2.2-27.2	3.3-49.5
	Mean	7.25	2.00	0.43	0.07	1.25	1.75	42.77	2.96	29.37	8.45	20.92
	SD	0.31	0.48	0.28	0.09	0.49	0.62	14.35	1.53	15.67	6.94	14.85

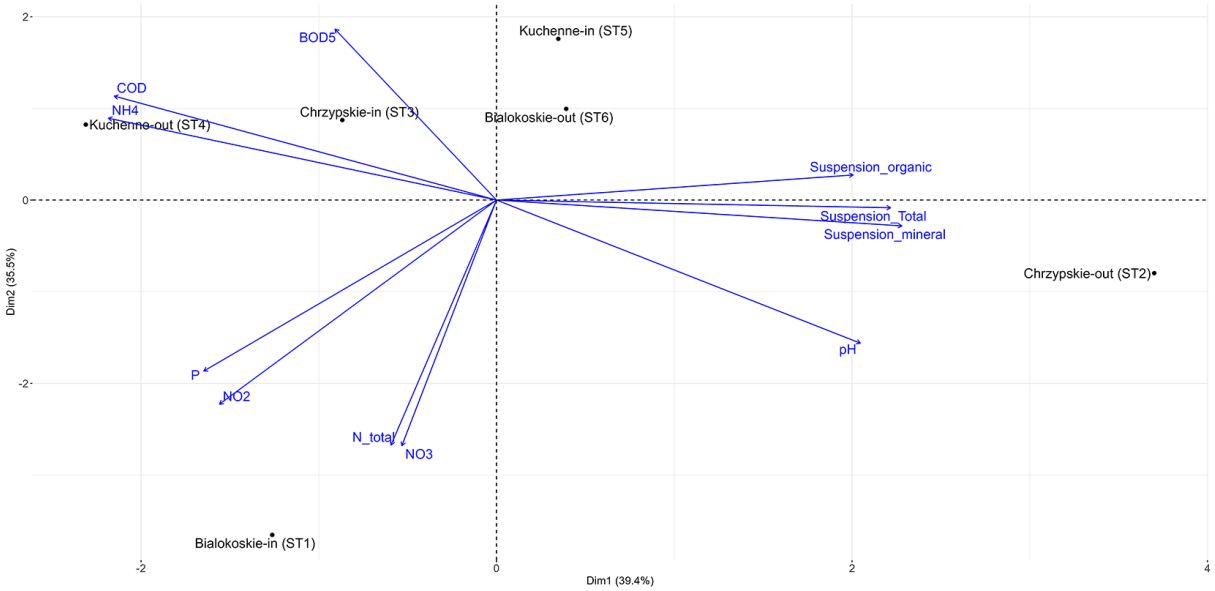

Figure 6. Biplot for principal component analyses for the complete dataset of water quality parameters

Table 2. Importance of components

Index	PC1	PC2	PC3	PC4
Standard deviation	1.66	1.36	0.59	0.17
Proportion of variance [%]	0.55	0.37	0.07	0.01
Cumulative proportion [%]	0.55	0.92	0.99	1.00

One of the most difficult stages of data analysis in the *k*-means method is determining the number of clusters. Three different algorithms were used for this purpose: WSS Plot (Elbow Plot), Silhouette Score, and Gap-stat (Generalized Average Precision).

Individual methods gave *K* values from 1 (GAP-stat method) to 3 (WSS Plot). It was assumed in further calculations that the number of clusters *K* was equal to 2.

Similarly to determining the number of clusters, calculations for the *k*-means method were made using the RStudio environment. For the determined number of clusters K = 2, analyses were performed for individual parameters or groups of parameters. It was noticed that for the group of parameters related to nitrogen content (NO_2, NO_3 and total nitrogen), the *k*-means method in cluster no. 2 grouped measurements primarily for the ST1 stand located at the inflow to Lake Białokoskie (Figure 8), which results from the relatively high loads of these indicators at the

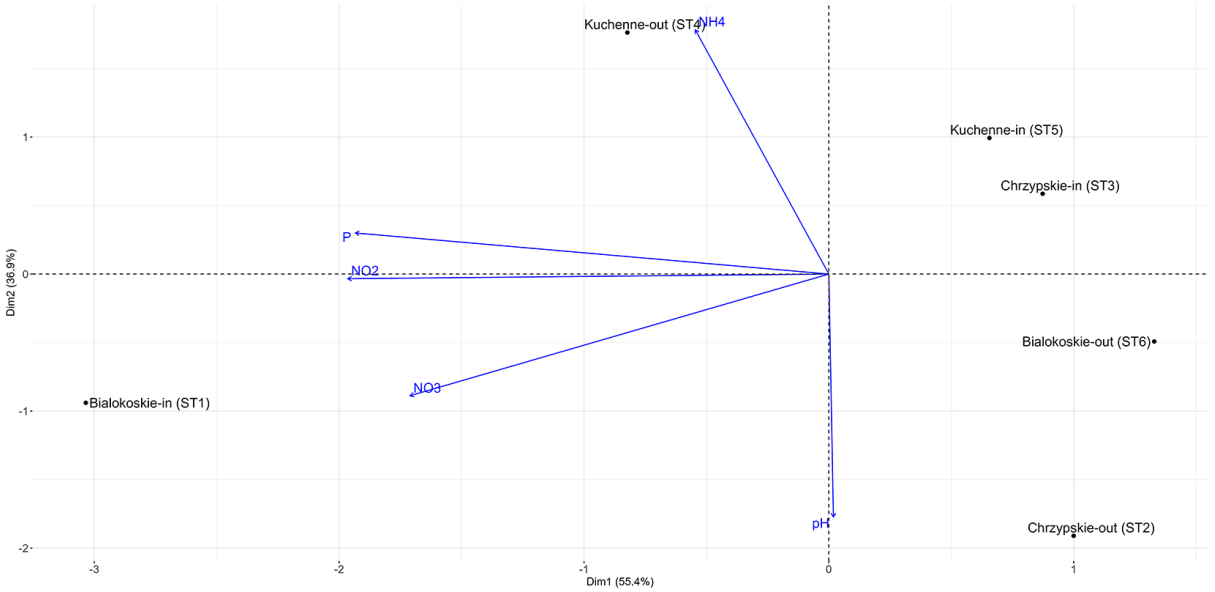

Figure 7. PCA biplot for the reduced dataset

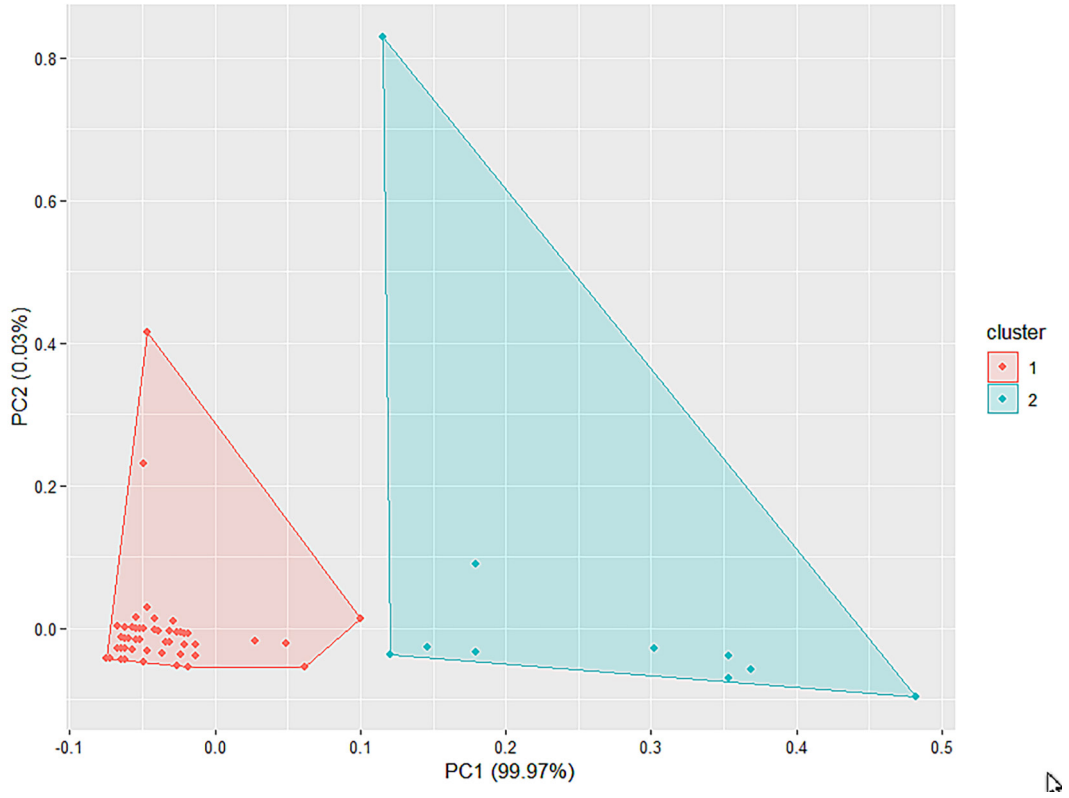

Figure 8. Distribution of measurement observations in individual clusters determined by the *k*-means method for NO_2, NO_3 and total nitrogen measurements

Table 3. Central values for individual clusters for NO_2, NO_3 and total nitrogen measurement data

No. cluster	NO_2	NO_3	N_Total
1	0.046667	1.674242	1.989545
2	0.127	13.35	13.727

inflow, clearly differing from the other studied lakes (Figure 3).

Table 3 summarises the mean values of NO_2, NO_3 and total nitrogen for individual clusters. For cluster no. 2, the mean values are many times greater than for the measurements from cluster no. 1. All measurements in cluster no. 2 come from the same stand, characterised by significantly higher parameters for nitrogen compounds. The test stand is located at the inflow of a small watercourse to Lake Białokoskie, with arable lands located along it, which significantly affected the size of loads.

Statistical analyses of means

Phosphorus

Basic statistical tests were performed, including the Shapiro-Wilk normality tests ($\alpha = 0.05$) for the individual datasets for total nitrogen, phosphorus, BOD_5 and COD indicators.

Considering the analyses of phosphorus content, for all samples, except for the ST1 stand – Lake Białokoskie – inflow (for which the *p*-value is 0.0006739), the data were of normal distribution. Therefore, mean equality analyses were performed using ANOVA ($\alpha = 0.05$) for outflow data and the Kruskal-Wallis test ($\alpha = 0.05$) for inflow data (one of the groups was not normally distributed).

Since in the Kruskal-Wallis test $p = 0.122 > \alpha$, it indicated that the null hypothesis H_0 cannot be rejected. The mean ranks of all groups were assumed to be equal; therefore, the difference between the mean ranks of all groups is not large enough to be statistically significant.

The ANOVA tests for phosphorus outflows with *p*-value = 0.068521 > α, indicated that the null hypothesis H_0 can be accepted, and the means of all groups can be taken equal. Also, the Tukey HSD/Tukey Kramer test indicated that there is no significant difference between the means of any pair.

Nitrogen

Considering the analyses of total nitrogen, normal distributions of measurement data were recorded for all samples, except for the ST2 stand – Lake Chrzypskie – outflow. Therefore, ANOVA

tests (α = 0.05) for the inflows and Kruskal-Wallis tests (α = 0.05) for the outflows were performed.

The ANOVA tests for the inflows indicated that $p = 1.00467e^{-10} < \alpha$ then H_0 must be rejected. The difference between the means of some groups is large enough to be statistically significant. Tukey HSD/Tukey Kramer tests indicated that the means of the following pairs are significantly different: ST1-ST3, ST1-ST5.

The Kruskal-Wallis test for the outflows indicated that there is a significant difference in the means $p = 0.003 < \alpha$, hypothesis H_0 must be rejected. The difference between the mean ranks of some groups is large enough to be statistically significant. The multiple comparisons test indicated that the mean ranks of the following pairs are significantly different: ST2-ST4, ST2-ST6.

BOD_5

The Shapiro-Wilk tests (α = 0.05) for BOD_5 showed that for the ST5 stand – Lake Kuchenne – inflow and ST6 – Lake Białokoskie – outflow, the distribution does not correspond to the normal distribution. Therefore, Kruskal-Wallis tests (α = 0.05) were carried out for both BOD_5 data at the inflows and outflows from individual lakes.

Considering the inflows to individual lakes, $p = 0.176 > \alpha$. Therefore, there is no reason to reject the null hypothesis H_0, the mean ranks of all groups are equal. A pairwise comparison using the Kruskal Wallis test showed that there is no significant difference between the mean ranks of any pair.

Considering the outflows (ST2, ST4 and ST6 stands), the Kruskal-Wallis test showed that $p = 0.138 > \alpha$. Thus, there is no reason to reject the null hypothesis H_0 and the mean ranks of all groups are equal. Also, a pairwise comparison using the Kruskal-Wallis test demonstrated that there is no significant difference between the mean ranks of any pair.

COD

With regard to the COD indicator, Shapiro-Wilk tests (α = 0.05) showed that no normal distributions were observed for the ST1 stand – Lake Białokoskie – inflow and the ST4 stand – Lake Kuchenne – outflow. Therefore, Kruskal-Wallis tests (α = 0.05) were performed for both COD data at the inflows and outflows from individual lakes.

The Kruskal-Wallis test showed that at the inflows, there is an insignificant difference in the dependent variable between different groups, $p = 0.181 > \alpha$, there is no reason to reject H_0, the mean ranks of all groups are equal. Also, the pairwise comparison using the Kruskal Wallis test demonstrated that there is no significant difference between the mean ranks of any pair.

The Kruskal-Wallis H test for the outflows showed that there is a significant difference between the means, $p < 0.001 < \alpha$. The pairwise comparison demonstrated that the mean ranks of the following pairs are significantly different: ST2-ST4, ST2-ST6.

Table 4 summarises the test results of difference significance for individual indicators. Depending on the value of the Shapiro-Wilk test, if the analyzed group of indices had a normal distribution, ANOVA tests were performed. Otherwise, if at least one of the indexes did not have a normal distribution, the Kruskal-Wallis test was performed.

The multiple comparison test for total nitrogen for the outflows indicated that the mean ranks of the following pairs are significantly different: ST2-ST4 and ST2-ST6. Also, for the inflows, Tukey/Kramer tests demonstrated that the means of the following pairs are significantly different: ST1-ST3 and ST1-ST5. Considering COD for the outflows, tests showed that the mean ranks of the following pairs are significantly different: ST2-ST4 and ST2-ST6.

On the basis of seven indicators: average lake depth, ratio of lake volume to shoreline length, percentage of water stratification and water exchange, ratio of active bottom area to epilimnion volume, Schindler coefficient and management of the direct catchment, an assessment of the vulnerability to degradation of Białokoskie, Chrzypskie and Kuchenne lakes was carried out. All lakes were classified into Category III water bodies, not very resistant to the influence of the catchment on their degradation.

Table 4. Summary of ANOVA or Kruskal-Wallis test results

Location	P	N_Total	BZT_5	COD
ST1	+	–	+	+
ST3				
ST5				
ST2	+	–	+	–
ST4				
ST6				

Note: +) there are no significant differences in the means in groups, no grounds to reject the null hypothesis; –) there are significant differences in the means in groups, the null hypothesis must be rejected.

DISCUSSION

As a result of the assessment of morphometric and hydrological parameters of the analysed reservoirs, they were classified in category III of resistance to degradation. The belonging of the objects to this category indicates the poor resistance of the reservoirs to the influence of the catchment area; anthropogenic factors (land use of the catchment area) in conjunction with natural factors had a potential impact on changing water quality. The most disadvantageous parameters, determining the obtained resistance classes, are mainly the low depth of reservoirs and not enough opportunities for water exchange.

In the analysed reservoirs, concentrations of nitrogen and total phosphorus were found to be high, exceeding those permitted for water quality class II. Also, the studies conducted by GIOS for individual lakes in previous years confirm exceedances of permissible values of indicators, although smaller than those recorded recently. Very unfavourable indicator values (especially for NO_3) were recorded for Bialokoskie Lake, on an inflow. High concentrations of NO_3 near the inflow to the lake may have been caused by nutrient runoff from agricultural fields drained by these watercourses. Similar analyses were carried out, for example, by Janicka et al. [2016] indicating for the Solina reservoir that the quality of the reservoir waters is deteriorating from year to year due to the high vulnerability of the reservoir to degradation, among other factors. Liu et al. [2011] indicate that due to rapid agricultural, industrial, and urban development, since the 1980s increasing N deposition in China and China urgently needs to establish national networks for N deposition monitoring. A major and frequent problem in lowland river catchment areas is the pollution from agricultural sources, as well as insufficient sanitation of rural and recreational areas manifesting as a direct discharge of sewage and pollution into the reservoir, which was recorded in an increase in BOD_5 and COD [Grochowska et al., 2019; Kanclerz et al., 2014].

The influence of the catchment area and its land use was also analysed by Smal et al. [2005], who indicated that the threat of lake eutrophication, assessed according to Vollenweider's criteria, was highest for lakes in which forests and grasslands occurred in the smallest proportions in the catchments, and the proportion of fertile, cultivated soils was highest. Kowalczewska et al. [2006], analysing the water quality of Swarzędzkie Lake, also points out that although about 80% of the direct discharge of domestic wastewater into the lake was stopped in 1991, water quality has not improved significantly since then. This is due to intensive internal loading from bottom sediments and external loading from the catchment area, which is dominated by agricultural fields. High concentrations of nitrogen and phosphorus, especially in Bialokoskie Lake, may also be caused by inflows from sewage systems and domestic wastewater treatment plants. According to the report of the Study of Conditions and Directions for Spatial Development of the Municipality of Chrzypsko Wielkie 2022, the municipality of Chrzypsko Wielkie has a municipal wastewater treatment plant located in Chrzypsko Wielkie. The receiver of treated wastewater is Osiecznica, below the Chrzypskie reservoir. The villages of Chrzypsko Wielkie (partially) and Charcice are currently sewered. The remaining villages use individual sewage systems. There are 42 household treatment plants in the municipality. The wastewater from no-outflow tanks is transmitted to the treatment plant in Chrzypsko Wielkie. The Social Welfare Home in Łężeczki has its own wastewater treatment plant.

For instance, the assessment of surface water quality using the water quality index and multivariate statistical analyses was carried out by Kükrer and Mutlu in Lake Saraydüzü in Turkey [Kükrer and Mutlu, 2019]. The conducted research consisted in collecting the samples of water used for drinking and irrigation in monthly cycles for 1 year from 6 measurement points. The obtained results showed that the water from the Saraydüzü dam lake could be classified as "very good" in terms of drinking water quality and was characterised by the lack of nitrogen and phosphates. The quality of water used for irrigation was considered "very good" in terms of COD, BOD_5, nitrates and ammonium, whereas in terms of phosphates it ranged between "very good" and "average". On the other hand, the permissible values for heavy metals (Cu, Zn and Fe) were exceeded, which could be explained by further geological research [Kükrer and Mutlu, 2019].

Studies by Ngoye and Machiwa [2004], who assessed the impact of land use practices in the Ruvu River basin on water quality in the river system, demonstrated the deterioration of river water quality due to anthropogenic activities in the catchment area. The authors found that also agricultural lands contributed significantly to higher nutrient concentrations in the Ruvu river system. The study

used analysis of variance (ANOVA) to compare differences in water quality for different land uses, the post hoc test (Tukey) was used to compare differences between the means. The student's t-test was used to compare the water quality between seasons.

Mahdi and Hamdan [2021] used the ANOVA analysis to investigate the statistically significant spatial and temporal variability of water quality of the Shatt Al-Arab River. The water quality of this river, which is the main source of the water treatment plant (WTP), has been constantly deteriorating due to various industrial, municipal, and agricultural activities taking place along the course of the river. The ANOVA results showed that there were significant spatial differences in water quality between the stations where the impact of wastewater discharge and salinity intrusion from the Arabian Gulf was very distinct and the stations with no source of pollution, stations close to wastewater outflows, as well as stations mainly affected by the salinity intrusion from the Arabian Gulf.

Zhu et al. [2022] indicated in their review of machine learning methods that it is recommended to use ANN (Attention Neural Network), CNN (Correlation Neural Network), SVM (Support Vector Machines) and RF (Random Forest) for monitoring the quality of water. Principal Component Analysis (PCA) was suggested in this work as a method used to select parameters for the Water Quality Index (WQI). Similarly, Radzka, Jankowska and Rymuza [2017] used PCA and cluster analysis to assess the drinking water quality. Cluster analysis was performed there using RF (Random Forest). Similar results were obtained; both methods showed the same number of clusters. The study confirmed the greater universality of PCA, which indicates not only clusters, but also explains the relationships between individual sets of observations and their impact on the entire dataset. Currently, a clear trend, not only in relation to the analysis of water quality presented in the paper, is the dissemination of machine learning methods in analysing various datasets. This is the consequence of their availability in open systems, which provide highly advanced data analysis algorithms at no additional cost.

CONCLUSIONS

The study investigated water quality in three adjacent lakes, located in the protected area of Sierakowski Landscape Park, using several selected indicators, i.e. phosphorus, nitrogen, BOD_5, and COD. Mashing learning algorithms, i.e. PCA and k-means, were used to analyse water quality indicators. The study made it possible to statistically estimate the changes in water quality indicators in the reservoirs and evaluate their correlation between different reservoirs. This made it possible to detect possible indicators statistically different from those noted in other reservoirs. This may suggest the occurrence of unfavorable phenomena negatively affecting the water quality in the reservoir.

The study shows that water quality in the lakes analysed is poor. The lakes are loaded with high nutrient loads, which, combined with the low resistance of the lakes to degradation caused by their catchment area, may result in further deterioration of water quality in these reservoirs. The work confirmed:

- the usefulness of two well-known and described machine learning methods – PCA and *k*-means – for a relatively small dataset of water quality parameters,
- two ML methods correctly indicated the test stands where water quality parameters significantly differed from other parameters measured,
- the application of ML methods significantly extends the possibilities of monitoring changes in parameters and detecting the points or areas that can be the potential sources of pollution. This is important for environmentally protected areas where the quality of lake water is key for various ecosystems,
- the methods allow for narrowing down the search area for the source of pollution as well as they can indicate places where the situation has improved as well as identify the factors behind this improvement.

The analyses using machine learning algorithms that are performed in the first stage of studies allow for focusing on the areas or objects with outlier parameters and on the parameters that require extension of field research or advanced statistical methods. However, neither of the methods provides an answer as to what is the source of biogens and pollutants contributing to the low quality of water in the analysed reservoirs. ML analyses, such as PCA, also enable the estimation of correlations between individual water quality indicators, determined for different lakes, particularly the ones that are interdependent (catchment

area, shared inflow, etc.). The development of such a relationship can facilitate surface water monitoring and allow early response to the emergence of adverse phenomena. Methods like PCA or *k*-means can be useful tools for environmental monitoring and protection teams associated with the ministry of water or environmental protection.

REFERENCES

1. Abdi H., Williams L.J. 2010. Principal component analysis. Wiley Interdisciplinary Reviews: Computational Statistics, 2(4), 433–459.
2. Ahmed U., Mumtaz R., Anwar H., Shah A.A., Irfan R., García-Nieto J. 2019. Efficient Water Quality Prediction Using Supervised Machine Learning. Water, 11(11), 2210.
3. Álvarez X., Valero E., Santos R.M., Varandas S.G.P., Fernandes L.S., Pacheco F.A.L. 2017. Anthropogenic nutrients and eutrophication in multiple land use watersheds: Best management practices and policies for the protection of water resources. Land Use Policy, 69, 1–11.
4. Bhateria R., Jain D. 2016. Water quality assessment of lake water: a review. Sustainable Water Resources Management, 2, 161–173.
5. Bishop C.M. 2016. Pattern Recognition and Machine Learning. Springer, New York, NY.
6. Bródka S., Macias A. 2016. Lakes in the landscape parks of Wielkopolska province: a collective work (in Polish). Bogucki Wydawnictwo Naukowe, Poznań.
7. Bui D.T., Khosravi K., Tiefenbacher J., Nguyen H., Kazakis N. 2020. Improving prediction of water quality indices using novel hybrid machine-learning algorithms. Science of the Total Environment, 721, 137612.
8. Central Statistical Office 2010. Statistical yearbook of the republic of Poland (in Polish). [Data set] https://stat.gov.pl/obszary-tematyczne/roczniki-statystyczne/roczniki-statystyczne/rocznik-statystyczny-rzeczypospolitej-polskiej-2010,2,5.html (accessed 4 May 2023).
9. Chen K., Chen H., Zhou C., Huang Y., Qi X., Shen R., Liu F., Zuo M., Zou X., Wang J. 2020. Comparative analysis of surface water quality prediction performance and identification of key water parameters using different machine learning models based on big data. Water Research, 171, 115454.
10. Chief Inspectorate for Environmental Protection, GIOS. 2020. Classification of indicators and groups of indicators in surface water bodies of lakes for the year 2020 (in Polish). [Data set] https://wody.gios.gov.pl/pjwp/api/publications/media/536 (accessed 9 May 2023).
11. Chief Inspectorate for Environmental Protection, GIOS. 2021. Classification of indicators and groups of indicators in surface water bodies of lakes for the year 2021 (in Polish). [Data set] https://wody.gios.gov.pl/pjwp/api/publications/media/537 (accessed 9 May 2023).
12. Crase L., Gillespie R. 2008. The impact of water quality and water level on the recreation values of Lake Hume. Australasian Journal of Environmental Management, 15(1), 21–29.
13. Deutsch H.-P., Beinker M. W. 2019. Principal Component Analysis. In H.-P. Deutsch & M. W. Beinker (Eds.), Derivatives and Internal Models: Modern Risk Management (pp. 793–804) Springer International Publishing.
14. Dezfooli D., Hosseini-Moghari S.-M., Ebrahimi K., Araghinejad S. 2018. Classification of water quality status based on minimum quality parameters: application of machine learning techniques. Modeling Earth Systems and Environment, 4(1), 311–324.
15. Ferahtia A., Halilat M.T., Mimeche F., Bensaci E. 2021. Surface water quality assessment in semi-arid region (El Hodna watershed, Algeria) based on water quality index (WQI). Studia Universitatis Babes-Bolyai, Chemia, 66(1), 127-142.
16. Ferencz B., Toporowska M., Dawidek J., Sobolewski W. 2017. Hydro-Chemical Conditions of Shaping the Water Quality of Shallow Łęczna-Włodawa Lakes (Eastern Poland). CLEAN–Soil, Air, Water, 45(5), 1600152.
17. Gani M.A., Sajib A.M., Siddik M.A., Md Moniruzzaman. 2023. Assessing the impact of land use and land cover on river water quality using water quality index and remote sensing techniques. Environmental Monitoring and Assessment, 195(4), 449.
18. Grochowska J., Karpienia M., Tandyrak R., Płachta A., Dziczek J., Gołębiewska A.E., Jędrzejewski P., Tomczak M., Turek M., Zaręba F. 2019. Preliminary characteristic of water chemistry of lake Bartąg near Olsztyn and sketch of its protection concept (in Polish). Woda-Środowisko-Obszary Wiejskie, 19(1), 5–18.
19. Guz K., Doroszkiewicz W. 2003. Control and assessment of water quality in the protection of environment and haelth. Ekologia i Technika, 11(4), 22–31.
20. Haghiabi A.H., Nasrolahi A.H., Parsaie A. 2018. Water quality prediction using machine learning methods. Water Quality Research Journal, 53(1), 3–13.
21. Hartigan J.A., Wong M.A. 1979. A k-means clustering algorithm. Applied Statistics, 28(1), 100–108.
22. Janicka E., Kanclerz J., Wiatrowska K., Makowska M. 2016. Biogenic compounds and an eutrophication process of Raczyńskie Lake. Inżynieria Ekologiczna, 49, 124–130.
23. Jolliffe I.T. 2002. Principal component analysis. (2nd edn) Springer, New York, NY.

24. Journal of Laws, item 1475 2021. Regulation of the Minister of Infrastructure of June 25, 2021 on the classification of ecological status, ecological potential and chemical status and the method of classifying the status of surface water bodies, as well as environmental quality standards for priority substances (in Polish) https://isap.sejm.gov.pl/isap.nsf/DocDetails.xsp?id=WDU20210001475.
25. Kanclerz J., Wiatrowska K., Adamska A. 2015. Phosphorous concentration in surface water of Gorzuchowskie lake catchment (in Polish). Polish Journal of Agronomy, 22, 10–17.
26. Kanclerz J., Wicher-Dysarz J., Dysarz T., Sojka M., Dwornikowska Ż. 2014. Influence of the Stare Miasto reservoir on the Powa river water quality (in Polish). Nauka Przyroda Technologie, 8(4) 1–11.
27. Kowalczewska-Madura K., Gołdyn R. 2006. Anthropogenic changes in water quality in the Swarzędzkie Lake (West Poland). Limnological Review, 6, 147–154.
28. Kudelska D., Cydzik D., Soszka H. 1994. Guidelines for basic monitoring of lakes (in Polish). Państwowa Inspekcja Ochrony Środowiska, Warszawa.
29. Kükrer S., Mutlu E. 2019. Assessment of surface water quality using water quality index and multivariate statistical analyses in Saraydüzü Dam Lake, Turkey. Environmental Monitoring and Assessment, 191(2), 71.
30. Kurita T. 2019. Principal Component Analysis (PCA). In Computer Vision: A Reference Guide (pp. 1–4) Springer International Publishing.
31. Liu X., Duan L., Mo J., Du E., Shen J., Lu X., Zhang Y., Zhou X., He C., Zhang F. 2011. Nitrogen deposition and its ecological impact in China: an overview. Environmental Pollution, 159(10), 2251–2264.
32. Mahananda M.R., Mohanty B.P., Behera N.R. 2010. Physico-chemical analysis of surface and ground water of Bargarh District, Orissa, India. International Journal of Research and Reviews in Applied Sciences, 2(3), 284–295.
33. Mahdi Z.H., Hamdan A.N. 2021. Spatial Variations of the Water Quality Parameters in Basra Water Treatment Plants Using ANOVA. Design Engineering, 18, 3684–3703.
34. Mulu B.D., Mehari M.W. 2013. Distribution of trace metals in two commercially important fish species (Tilapia zilli and Oreochromis niloticus) sediment and water from Lake Gubdahri, Eastern Tigrar of Northern Ethiopia. International Journal of Scientific and Research Publications, 3(9), 1–7.
35. Ngoye E., Machiwa J.F. 2004. The influence of land-use patterns in the Ruvu river watershed on water quality in the river system. Physics and Chemistry of the Earth, Parts A/B/C, 29(15–18), 1161–1166.
36. Nyenje P.M., Foppen J.W., Uhlenbrook S., Kulabako R., Muwanga A. 2010. Eutrophication and nutrient release in urban areas of sub-Saharan Africa—a review. Science of the Total Environment, 408(3), 447–455.
37. PN ISO 7150-1:2002, Water quality — Determination of ammonium — Part 1: Manual spectrometric method (in Polish).
38. PN-C-04559-02:1972, Water and wastewater - Suspended solids tests - Determination of total, mineral and volatile suspended solids by weight method (in Polish).
39. PN-C-04576-08:1982, Water and wastewater — Tests for nitrogen compounds — Determination of nitrate nitrogen by colorimetric method with sodium salicylate (in Polish).
40. PN-EN 1899-2:2002, Water quality — Determination of biochemical oxygen demand after 5 days (BOD 5) — Dilution and seeding method (in Polish).
41. PN-EN 26777:1999, Water quality — Determination of nitrite — Molecular absorption spectrometric method (in Polish).
42. PN-EN ISO 11905-1:2001, Water quality — Determination of nitrogen — Part 1: Method using oxidative digestion with peroxodisulfate (in Polish).
43. PN-ISO 8466-1:2003, Water quality — Calibration and evaluation of analytical methods and estimation of performance characteristics — Part 1: Statistical evaluation of the linear calibration function (in Polish).
44. PN-ISO 15705:2005, Water quality — Determination of the chemical oxygen demand index (ST-COD) — Small-scale sealed-tube method (in Polish).
45. Radzka E., Jankowska J., Rymuza K. 2017. Principal component analysis and cluster analysis in multivariate assessment of water quality. Journal of Ecological Engineering, 18(2), 92–96.
46. Rodríguez-López L., Bustos Usta D., Bravo Alvarez L., Duran-Llacer I., Lami A., Martínez-Retureta R., Urrutia R. 2023. Machine Learning Algorithms for the Estimation of Water Quality Parameters in Lake Llanquihue in Southern Chile. Water, 15(11), 1994.
47. Sagan V., Peterson K.T., Maimaitijiang M., Sidike P., Sloan J., Greeling B.A., Maalouf S., Adams C. 2020. Monitoring inland water quality using remote sensing: Potential and limitations of spectral indices, bio-optical simulations, machine learning, and cloud computing. Earth-Science Reviews, 205, 103187.
48. Smal H., Kornijow R., Ligeza S. 2005. The effect of catchment on water quality and eutrophication risk of five shallow lakes (Polesie region, Eastern Poland). Polish Journal of Ecology, 53(3), 313–327.
49. Smith V.H. 2003. Eutrophication of freshwater and coastal marine ecosystems a global problem. Environmental Science and Pollution Research, 10, 126–139.

50. Uddin M.G., Nash S., Mahammad Diganta M.T., Rahman A., Olbert A.I. 2022. Robust machine learning algorithms for predicting coastal water quality index. Journal of Environmental Management, 321, 115923.
51. Uddin M.G., Nash S., Rahman A., Olbert A.I. 2022. A comprehensive method for improvement of water quality index (WQI) models for coastal water quality assessment. Water Research, 219, 118532.
52. Uddin M.G., Nash S., Rahman A., Olbert A.I. 2023a Assessing optimization techniques for improving water quality model. Journal of Cleaner Production, 385, 135671.
53. Uddin M.G., Nash S., Rahman A., Olbert A.I. 2023b Performance analysis of the water quality index model for predicting water state using machine learning techniques. Process Safety and Environmental Protection, 169, 808–828.
54. Uddin M.G., Nash S., Rahman A., Olbert A.I. 2023c A novel approach for estimating and predicting uncertainty in water quality index model using machine learning approaches. Water Research, 229, 119422.
55. Uddin Md. G., Nash S., Olbert A.I. 2021. A review of water quality index models and their use for assessing surface water quality. Ecological Indicators, 122, 107218.
56. WIOŚ in Poznań 2015. Classification of water quality indicators of lakes in Wielkopolska province for the year 2015 - Chrzypskie Lake (in Polish). [Data set] https://poznan.wios.gov.pl/monitoring-srodowiska/wyniki-badan-i-oceny/monitoring-wod-powierzchniowych/jeziora/wyniki-badan-klasyfikacja-wskaznikow-i-oceny-za-rok-2015/klasyfikacja-wskaznikow-jakosci-wod-jezior-w-wojewodztwie-wielkopolskim-za-rok-2015/ (accessed 5 May 2023).
57. WIOŚ in Poznań 2018. Classification of water quality indicators of lakes in Wielkopolska province for the year 2017 - Bialokoskie Lake (in Polish). [Data set] http://poznan.wios.gov.pl/wios/ocena2018/jeziora/Jez_Bialokoskie_2017_klasyfikacja.pdf (accessed 5 May 2023).
58. Withers P.J., Neal C., Jarvie H.P., Doody D.G. 2014. Agriculture and eutrophication: where do we go from here? Sustainability, 6(9), 5853–5875.
59. Zhu M., Wang J., Yang X., Zhang Y., Zhang L., Ren H., Wu B., and Ye L. 2022. A review of the application of machine learning in water quality evaluation. Eco-Environment & Health, 1(2), 107–116.