

## WATER QUALITY ANALYSIS OF LAKE PALIC USING PCA

### АНАЛИЗА КВАЛИТЕТА ВОДЕ ЈЕЗЕРА ПАЛИЋ ПОСТУПКОМ АНАЛИЗЕ ГЛАВНИХ КОМПОНЕНТИ

Zoltan Horvat<sup>1</sup>  
Mirjana Horvat<sup>2</sup>

UDK 556.113/.115(497.113)  
DOI: 10.14415/zbornikGFS37.06  
CC-BY-SA 4.0 license

**Summary:** The principal component analysis (PCA) is a well established methodology used in all research fields to provide a more intelligent understanding of the investigated data. Accordingly, this paper introduces the application of this methodology for the evaluation of the water quality of Lake Palic. As a result of the selected approach we were able to draw meticulous conclusions and direct the impending research in a more rational direction.

**Keywords:** lake water quality, water quality parameters, principal component analysis.

**Резиме:** Анализа главних компоненти (АГК) је опште прихваћена методологија у процесу истраживања која обезбеђује боље разумевање анализираних података. Због тога се ово истраживање фокусира применом ове методе на оцену квалитета воде језера Палић. Примена ове методе нам је омогућила поуздане и свеобухватне закључке, као и рационално усмеравање даљег истраживања.

**Кључне речи:** квалитет воде језера, параметри квалитета воде, анализа главних компоненти.

#### 1. INTRODUCTION

Standard data assessment relies on the investigation of a large number of data, finding various correlations among them, characteristic behaviors, possible clustering and similar tendencies. One of the standard methods often utilized for the reduction of the analyzed sample size is the principal component analysis (PCA), [1,2,3,4,5,6]. In addition to the reduction of the dimensionality, PCA allows enhanced interpretation of the water quality data by helping the identification of spatial and/or temporal trends within the evaluated water quality measurements. Consequently it

#### 1. УВОД

Стандардне процедуре истраживања мерења се често заснивају на анализи великог броја података, покушају идентификације потенцијалних корелација, карактеристичних показатеља, евентуалних груписања података и сличних тенденција. Једна од метода која се често користи као средство за редукују димензионалности посматраног проблема је анализа главних компонената (АГК), [1,2,3,4,5,6]. Поред примене за редукују димензионалности посматране проблематике, ова

<sup>1</sup> Zoltan Horvat, PhD, C.E., University of Novi Sad, Faculty of Civil Engineering Subotica, Kozaracka street 2a, 24000 Subotica, Serbia, e-mail: [horvatz@gf.uns.ac.rs](mailto:horvatz@gf.uns.ac.rs)

<sup>2</sup> Mirjana Horvat, PhD, C.E., University of Novi Sad, Faculty of Civil Engineering Subotica, Kozaracka street 2a, 24000 Subotica, Serbia, e-mail: [isicm@gf.uns.ac.rs](mailto:isicm@gf.uns.ac.rs)

helps to improve the overall interpretation of the gathered information [5]. This method is well established and used by researchers in various fields for the described reasons. This work presents the general steps for employing the PCA focusing on the analysis of water quality data. After utilizing the PCA, we describe the means to draw conclusions from the results. The application of the PCA is demonstrated on measured water quality data gathered from Lake Palic's fourth sector for the time interval from 2011 to 2015.

## 2. MATERIALS AND METHODS

The application of the PCA is investigated on the example of Lake Palic's 4<sup>th</sup> sector. Lake Palic is located in Serbia, roughly 8<sup>th</sup> kilometers from the City of Subotica. The average depth of the fourth sector is about 1.8 meters with sections that measure even 3.4 meters. This sector of the lake receives its water supply mainly though the inflow from the third sector, point 1 on Fig. 1. Although there are other sources such as precipitation, surface runoff and groundwater. The water exists the lake through two submerged pipes connecting it to Lake Omladinsko, marked with 2 on Fig. 1.

The employed measurements are conducted monthly, although there are numerous situations when for various reasons the measuring of all or some of the water quality parameters was omitted. Considering the fact that the evaluation is based on the analysis of five years of data, it is reasonable to expect the extraction of good quality conclusions and results.

The sampling location of the fourth sector, where all of the samples were gathered and further analyzed is located at the outflow from the sector, marked 2 on Fig. 1. The analyzed parameters include pH value, total phosphorus (TP), dissolved oxygen

метода омогућава и боље разумевање података путем идентификације потенцијала за просторно или временско груписање истих. Као резултат тога се обезбеђује боља свеобухватна интерпретација података. Зато се ова метода и примењује често у разним областима истраживања као поступак дубљег схватања расположивих информација.

Овај рад представља основне кораке примене АГК уз нагласак на примену исте методе на анализу параметара квалитета воде. Након спроведеног поступка одређивања главних компоненти се прелази на поступак анализе и тумачења резултата. Примена АГК је приказана на примеру података прикупљених у четвртном сектору језера Палић током периода од 2011 до 2015. године .

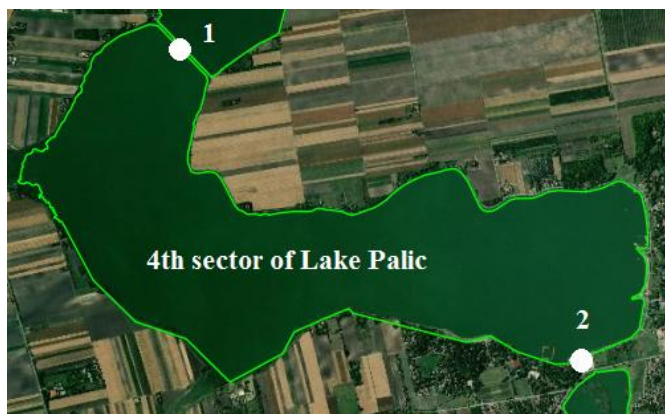
## 2. МАТЕРИЈАЛИ И МЕТОДЕ

Примена АГК се разматра на примеру четвртог сектора језера Палић. Језеро Палић се налази у Србији на око 8км од града Суботице. Просечна дубина језера је 1.8м, са деловима који достижу и дубину од 3.4м. Четврти сектор језера се водом снабдева углавном преко трећег сектора, тачка 1 на слици 1, мада има и других извора прихрањивања језера као што су подземни дотицај, површински отицај и падавине. Вода језеро напушта преко две подводне цеви, означене бројем 2 на слици 1., којима се улива у Онладинско језеро.

Примењени подаци представљају месечна мерења параметара квалитета, мада има доста примера када је из разних разлога дошло до изостанка мерења неких или свих параметара квалитета. Упркос томе се очекују поуздани и разумни резултати анализа јер се ради о испитивању података који су мерени током пет година.

(DO), nitrate nitrogen ( $\text{NO}_3\text{-N}$ ), Chlorophyll-a (Chl-a), total Kjeldahl nitrogen (TNK), electrical conductivity (EC), water temperature (T), chemical oxygen demand (COD), five day biological oxygen demand ( $\text{BOD}_5$ ), ammonium nitrogen ( $\text{NH}_4\text{-N}$ ), suspended matter (SM) and orthophosphate ( $\text{PO}_4^{3-}$ ).

Локација узорковања одакле се сви узорци прикупљају и носе на даља лабораторијска испитивања је означена на слици 1 са бројем 2.



Слика 1 – Мапа анализиране локације  
Figure 1 – Map of the analyzed location

Табела 1 – Анализа података – средња вредност, максимум, медиана, и минималне вредности и стандардна девијација

Table 1 – Data analysis – mean, maximum, minimum, median value and the standard deviation

Variable	Mean	SD	Minimum	Median	Maximum
T	14.430	8.510	1.300	12.750	29.000
pH	8.997	0.397	8.270	9.065	9.930
DO	10.598	3.459	2.340	10.690	16.550
$\text{BOD}_5$	39.220	19.770	5.000	36.000	89.000
COD	110.900	47.660	44.000	98.500	280.500
TP	0.219	0.183	0.047	0.154	0.700
$\text{NO}_3\text{-N}$	0.297	0.381	0.010	0.175	2.270
Chl-a	407.150	266.270	52.000	355.000	1016.000
TNK	13.824	5.937	0.271	12.995	30.740
$\text{NH}_4\text{-N}$	1.564	0.895	0.384	1.363	3.801
$\text{PO}_4^{3-}$	0.035	0.018	0.003	0.031	0.095
EC	937.800	103.000	754.000	911.500	1171.000
SM	32.938	24.724	4.000	27.350	100.000

The computations were carried out by the Analyse-it software. The computed values of mean, standard deviation (SD), minimum, median and maximum values of the analyzed data are given in Tab. 1.

The minimal value of temperature was detected in January 2013 with the value of 1.3°C. The smallest pH value was noticed in December of 2012 and measured 8.27. The minimum value of DO was 2.34 mg/L that occurred in August of 2011. BOD<sub>5</sub> with the smallest value of 5 mg/L was detected in June 2012, while the lowest value of COD was 44 mg/L noticed in November 2013. May 2013 had the lowest values of a couple parameters, TNK=0.271 mg/L, PO<sub>4</sub><sup>3-</sup>=0.003 mg/L and EC=754µS/cm, and April of the same year shows the lowest value of NH<sub>4</sub>-N. The lowest concentration of SM was detected January 2014 and was only 4 mg/L, and for TP it was 0.047 mg/L in February of the same year. Nitrate nitrogen was 0.01 mg/L at its lowest value monitored in August of 2013, while the smallest value of Chl-a was 52mg/m<sup>3</sup> registered in January 2012. The described distribution of variables indicates there isn't any strong clustering of any characteristic in time. A similar situation is noticed when inspecting the highest values, there is no noticeable tendency of the data to have increased values in any particular year or month.

Figures 2 and 3 display a table that includes the examination for every pair combination of the considered quality parameters. The analyzed pair is given in the first column, the second column shows the Pearson's correlation coefficient. The value of the Pearson's coefficient ranges from -1 for absolute negative correlation, to +1 for a full positive correlation, whereas a correlation coefficient of 0 indicates the considered parameters aren't correlated. The next column displays the interval of the Pearson's coefficient for which there is a probability of 95%

Анализирани параметри обухватају рН вредност, укупан фосфор (TP), растворен кисеоник (DO), нитратни азот (NO<sub>3</sub>-N), хлорофил а (Chl-a), укупан азот по Кјелдалху (TNK), електричну проводљивост (EC), температуру воде (T), хемијску потрошњу кисеоника (COD), петодневну биолошку потрошњу кисеоника (BOD<sub>5</sub>), амонијачни азот (NH<sub>4</sub>-N), суспендоване материје (SM) и ортофосфате (PO<sub>4</sub><sup>3-</sup>).

Прорачуни су спроведени применом софтверског пакета Analyse-it, а резултати прорачуна укључују средњу вредност, стандардну девијацију, минималне и максималне вредности и медијану, а приказане су у Таб. 1.

Минимална вредност температуре је уочена у јануару 2013. године, а износила је 1.3°C, најмања рН вредност је измерена у децембру 2012. године и била је 8.27. Најнижа вредност за DO је 2.34 мг/л а примећена је у августу 2011.год. Најнижа вредност за BOD<sub>5</sub> је регистрована у јуну 2012.год. а била је 5мг/л, док је за COD та вредност 44мг/л измерена у новембру 2013. год. За TNK, PO<sub>4</sub><sup>3-</sup> и EC су уочене најниже вредности у мају 2013.год., док је за NH<sub>4</sub>-N најниже мерење регистровано у априлу. У јануару 2014 су измерене најниже вредности SM, а исте године у фебруару је измерена најнижа вредност за TP. У августу 2013 је измерена најмања вредности нитратног азота, а за Chl-a је најмања вредност уочена у јануару 2012. год. Регистровани резултати указују на то да нема неког уочљивог груписања.

Слике 2. и 3. приказују табелу у којој се разматра свака комбинација параметара квалитета. Прва колона приказује разматрани пар параметара, друга колона садржи вредности Пирсоновог коефицијента корелације за разматрану комбинацију, која има вредност од -1 у случају потпуне негативне

that it includes the actual Pearson's coefficient. The following column gives a graphical representation of the interval with the 95% certainty. For example, the interval for the first two parameters T and pH is from 0.483 to 0.836. This is a positive correlation hence the graphical representation is drawn on the positive side from the zero. The second combination of parameters are the T and DO for which we determined a negative correlation in the interval from -0.751 to -0.284.

корелације, +1 у случају потпуне позитивне корелације или 0 уколико нема корелације. Следећа колона даје интервал Пирсоновог коефицијента за који постоји вероватноћа од 95% да се стварна вредност налази у њему. Наредна колона даје графички приказ интервала са 95% вероватноћом појаве. На пример, интервал корелације параметара Т и рН је од 0.483 до 0.836 за који се одмах примећује да указује на позитивну корелацију што значи да је графички приказ ове корелације нацртан са позитивне стране у односу на нулу.

Pair	Pearson's r	95% CI		0		p-value
T, pH	0.700	0.483	to 0.836			<0.0001 <sup>1</sup>
T, DO	-0.561	-0.751	to -0.284			0.0004 <sup>1</sup>
T, BOD5	0.229	-0.107	to 0.519			0.1782 <sup>2</sup>
T, COD	0.131	-0.206	to 0.441			0.4464 <sup>2</sup>
T, TP	0.113	-0.224	to 0.426			0.5126 <sup>2</sup>
T, NO3	-0.343	-0.603	to -0.016			0.0406 <sup>1</sup>
T, Chl-a	0.290	-0.042	to 0.565			0.0858 <sup>2</sup>
T, TNK	-0.105	-0.419	to 0.232			0.5426 <sup>2</sup>
T, NH4	-0.573	-0.759	to -0.301			0.0003 <sup>1</sup>
T, PO4	-0.095	-0.410	to 0.241			0.5827 <sup>2</sup>
T, EC	-0.133	-0.442	to 0.204			0.4389 <sup>2</sup>
T, SM	0.205	-0.132	to 0.500			0.2294 <sup>2</sup>
pH, DO	-0.255	-0.538	to 0.081			0.1338 <sup>2</sup>
pH, BOD5	0.059	-0.275	to 0.380			0.7324 <sup>2</sup>
pH, COD	0.069	-0.265	to 0.389			0.6872 <sup>2</sup>
pH, TP	0.079	-0.256	to 0.397			0.6464 <sup>2</sup>
pH, NO3	-0.415	-0.654	to -0.100			0.0119 <sup>1</sup>
pH, Chl-a	0.265	-0.069	to 0.546			0.1176 <sup>2</sup>
pH, TNK	-0.235	-0.523	to 0.102			0.1684 <sup>2</sup>
pH, NH4	-0.680	-0.824	to -0.452			<0.0001 <sup>1</sup>
pH, PO4	0.039	-0.293	to 0.363			0.8203 <sup>2</sup>
pH, EC	-0.123	-0.434	to 0.214			0.4747 <sup>2</sup>
pH, SM	0.419	0.105	to 0.657			0.0110 <sup>1</sup>
DO, BOD5	-0.274	-0.553	to 0.059			0.1052 <sup>2</sup>
DO, COD	-0.176	-0.477	to 0.162			0.3037 <sup>2</sup>
DO, TP	-0.179	-0.479	to 0.159			0.2973 <sup>2</sup>
DO, NO3	0.128	-0.209	to 0.438			0.4557 <sup>2</sup>
DO, Chl-a	-0.303	-0.575	to 0.028			0.0722 <sup>2</sup>
DO, TNK	-0.286	-0.561	to 0.047			0.0912 <sup>2</sup>
DO, NH4	0.094	-0.242	to 0.410			0.5860 <sup>2</sup>
DO, PO4	-0.177	-0.478	to 0.161			0.3026 <sup>2</sup>
DO, EC	-0.207	-0.502	to 0.130			0.2254 <sup>2</sup>

Слика 2 – Анализа коефицијента корелације  
Figure 2 – Correlation coefficient analysis

Pair	Pearson's r	95% CI	0	p-value
DO, EC	-0.207	-0.502 to 0.130		0.2254 <sup>2</sup>
DO, SM	-0.132	-0.441 to 0.206		0.4445 <sup>2</sup>
BOD5, COD	0.440	0.131 to 0.672		0.0072 <sup>1</sup>
BOD5, TP	0.157	-0.180 to 0.462		0.3593 <sup>2</sup>
BOD5, NO3	-0.182	-0.482 to 0.156		0.2886 <sup>2</sup>
BOD5, Chl-a	0.492	0.195 to 0.706		0.0023 <sup>1</sup>
BOD5, TNK	0.296	-0.035 to 0.570		0.0791 <sup>2</sup>
BOD5, NH4	-0.079	-0.397 to 0.257		0.6486 <sup>2</sup>
BOD5, PO4	0.229	-0.107 to 0.519		0.1788 <sup>2</sup>
BOD5, EC	0.170	-0.168 to 0.472		0.3211 <sup>2</sup>
BOD5, SM	0.152	-0.186 to 0.457		0.3774 <sup>2</sup>
COD, TP	0.092	-0.244 to 0.408		0.5927 <sup>2</sup>
COD, NO3	-0.006	-0.334 to 0.323		0.9730 <sup>2</sup>
COD, Chl-a	0.649	0.407 to 0.805		<0.0001 <sup>1</sup>
COD, TNK	0.512	0.221 to 0.720		0.0014 <sup>1</sup>
COD, NH4	-0.070	-0.390 to 0.264		0.6834 <sup>2</sup>
COD, PO4	0.103	-0.233 to 0.418		0.5481 <sup>2</sup>
COD, EC	0.469	0.166 to 0.691		0.0039 <sup>1</sup>
COD, SM	0.224	-0.113 to 0.515		0.1891 <sup>2</sup>
TP, NO3	-0.161	-0.465 to 0.177		0.3495 <sup>2</sup>
TP, Chl-a	0.165	-0.173 to 0.468		0.3361 <sup>2</sup>
TP, TNK	0.033	-0.299 to 0.358		0.8472 <sup>2</sup>
TP, NH4	-0.003	-0.331 to 0.326		0.9875 <sup>2</sup>
TP, PO4	0.249	-0.087 to 0.534		0.1433 <sup>2</sup>
TP, EC	0.332	0.004 to 0.596		0.0477 <sup>1</sup>
TP, SM	0.138	-0.200 to 0.446		0.4237 <sup>2</sup>
NO3, Chl-a	-0.128	-0.438 to 0.209		0.4553 <sup>2</sup>
NO3, TNK	-0.059	-0.380 to 0.275		0.7317 <sup>2</sup>
NO3, NH4	0.362	0.038 to 0.617		0.0300 <sup>1</sup>
NO3, PO4	0.123	-0.214 to 0.434		0.4730 <sup>2</sup>
NO3, EC	0.142	-0.196 to 0.449		0.4099 <sup>2</sup>
NO3, SM	-0.289	-0.564 to 0.044		0.0879 <sup>2</sup>
Chl-a, TNK	0.344	0.017 to 0.604		0.0401 <sup>1</sup>
Chl-a, NH4	-0.281	-0.558 to 0.052		0.0967 <sup>2</sup>
Chl-a, PO4	0.207	-0.131 to 0.501		0.2268 <sup>2</sup>
Chl-a, EC	0.358	0.033 to 0.614		0.0321 <sup>1</sup>
Chl-a, SM	0.335	0.008 to 0.598		0.0456 <sup>1</sup>
TNK, NH4	0.341	0.014 to 0.602		0.0416 <sup>1</sup>
TNK, PO4	0.055	-0.279 to 0.377		0.7501 <sup>2</sup>
TNK, EC	0.317	-0.012 to 0.585		0.0592 <sup>2</sup>
TNK, SM	0.028	-0.304 to 0.353		0.8722 <sup>2</sup>
NH4, PO4	0.010	-0.320 to 0.337		0.9553 <sup>2</sup>
NH4, EC	0.487	0.188 to 0.703		0.0026 <sup>1</sup>
NH4, SM	-0.041	-0.364 to 0.292		0.8134 <sup>2</sup>
PO4, EC	0.304	-0.027 to 0.575		0.0714 <sup>2</sup>
PO4, SM	0.024	-0.307 to 0.350		0.8909 <sup>2</sup>
EC, SM	0.252	-0.083 to 0.536		0.1381 <sup>2</sup>

Слика 3 – Анализа коефицијената корелације  
Figure 3 – Correlation coefficient analysis

Consequently, the graphical representation shows the line on the left (negative) side of the zero value in the caption. The third considered pair is made from T and BOD<sub>5</sub> for which the interval with the 95% probability reaches from -0.107 to 0.519. This interval has both positive and negative parts hence the graphical representation also stretches on both sides of the zero. Another important thing to notice regarding this case is that it includes the possibility the correlation of these parameters is zero meaning they are not correlated.

The last column displays the values of p for the zero hypothesis which states the correlation coefficient of the bivariate population equals zero. The certainty level for the hypothesis is 5%, meaning that for all p values above that the hypothesis shouldn't be rejected, while for combinations where p is less than 5%, the hypothesis should be rejected in favor of the alternative hypothesis that the correlation coefficient is not equal to zero. Essentially, the last column indicates for which of these parameter combination is it safe to state they have some correlation.

### 3. PRINCIPAL COMPONENT ANALYSIS

The numerical values of the results of the implemented PCA are given in Tab. 2. The first column names the principal component that is considered. The second column shows the variance for the considered component, the third column includes the proportion of the component or the part of the variance contained within the considered principal component. The final column shows the cumulative proportion of each component or the proportion of variance the considered component includes with all of the previous components.

Figure 4 presents the graphical

Наредна комбинација параметара су T и DO за који је идентификована негативна корелација у интервалу од -0.751 до -0.284, тако да је графички приказ корелације нацртан на негативну страну у односу на нулу. Трећа комбинација обухвата T и BOD<sub>5</sub> где се интервал од 95% вероватноће појаве стварне вредности коефицијента корелације протеже од -0.107 до 0.519. Примећује се да овај интервал обухвата и позитивне и негативне вредности коефицијента корелације, због чега се и код графичког приказа уочава његово протезање са обе стране у односу на нулу. Поред тога треба напоменути да код ових случајева анализа указује на то да постоји вероватноћа да је корелација два посматрана параметра нула, дакле да нису корелирани.

Последња колона приказује р вредности за нулту хипотезу којом се тврди да је корелација посматраних парова параметара нула. Праг сигурности за проверу нулте хипотезе је 5%, односно све вредности р које прелазе 5% указују на то да нулту хипотезу не треба одбацити, док у случајевима када је р испод ове вредности нулту хипотезу треба одбацити у односу на алтернативну хипотезу да коефицијент корелације датог пара није нула. Генерално се преко вредности датих у последњој колони може видети за које комбинације параметара је безбедно претпоставити постојање неке корелације.

### 3. АНАЛИЗА ГЛАВНИХ КОМПОНЕНАТА

Бројчане вредности резултата анализе главних компоненти су дате у Таб. 2. где се у првој колони налази разматрана главна компонента, друга колона садржи варијансе компоненти, трећа колона показује

representation of the proportion of variance called the scree plot. By evaluating either, Fig. 4 or the values given in Tab. 2 it becomes obvious that the most significant variance is contained within the first two components, suggesting these two should be included in further analysis, while the remaining ones could be disregarded. Still, considering that they encompass 47.9% of the total variance, it would be justified to include additional components in further examination. For example, by including three additional components, one would account 74% of the total variance.

Figure 5 shows the monoplot that explains which of the investigated quality parameters have the most influence on the first two principal components. It should be pointed out that the arrows that are approaching the circle line are contributing more to the components, e.g. T and pH, while shorter arrows such as PO<sub>4</sub> and TP indicate lesser contribution of the parameters. Another important element of the monoplot is the angle between two arrows. A smaller angle suggests a more pronounced positive correlation between two parameters.

удео варијансе у датој компоненти, док последња колона приказује кумулативну вредност удела варијанси свих дотадашњих компоненти.

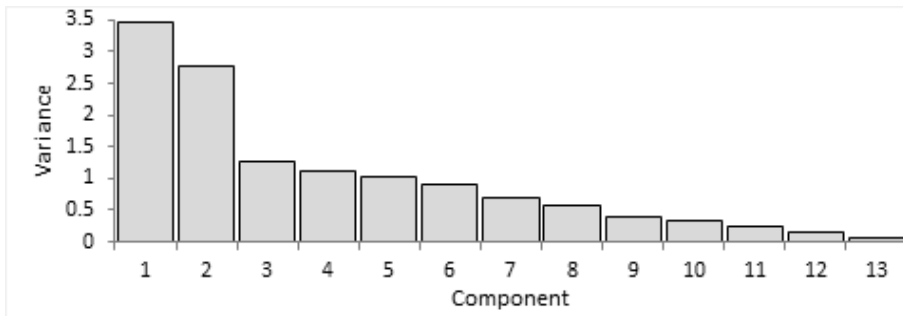
Слика 4. даје графички приказ удела варијансе у компонентама. Анализом података на Сл. 4 и Таб. 2. је јасно да је највећи удео варијанси, 47.9% садржан у прве две компоненте, указујући да би у наставак анализе требало укључити ове две компоненте, док би се остале могле занемарити. С обзиром на сам удео варијансе је такође оправдано и узимање додатних компоненти у даљим анализама. На пример, укључивањем три додатне компоненте у даљу анализу се у обзир узима 74% варијансе.

Слика 5. приказује моноплот који служи да се прикаже који од разматраних параметара квалитета воде имају највећи утицај на прве две главне компоненте. Напомиње се да дуге стрелице указују на веће учешће параметра у компонентама, нпр. Т и рН, док краће стрелице значе да дати параметар има мање учешће, нпр. PO<sub>4</sub> и TP.

Табела 2 – Главне компоненте  
Table 2 – Principal components

Component	Variance	Proportion	Cumulative proportion
1	3.453	0.266	0.266
2	2.771	0.213	0.479
3	1.275	0.098	0.577
4	1.111	0.085	0.662
5	1.017	0.078	0.740
6	0.892	0.069	0.809
7	0.700	0.054	0.863
8	0.585	0.045	0.908
9	0.399	0.031	0.939
10	0.330	0.025	0.964
11	0.240	0.018	0.983
12	0.155	0.012	0.994
13	0.073	0.006	1.000

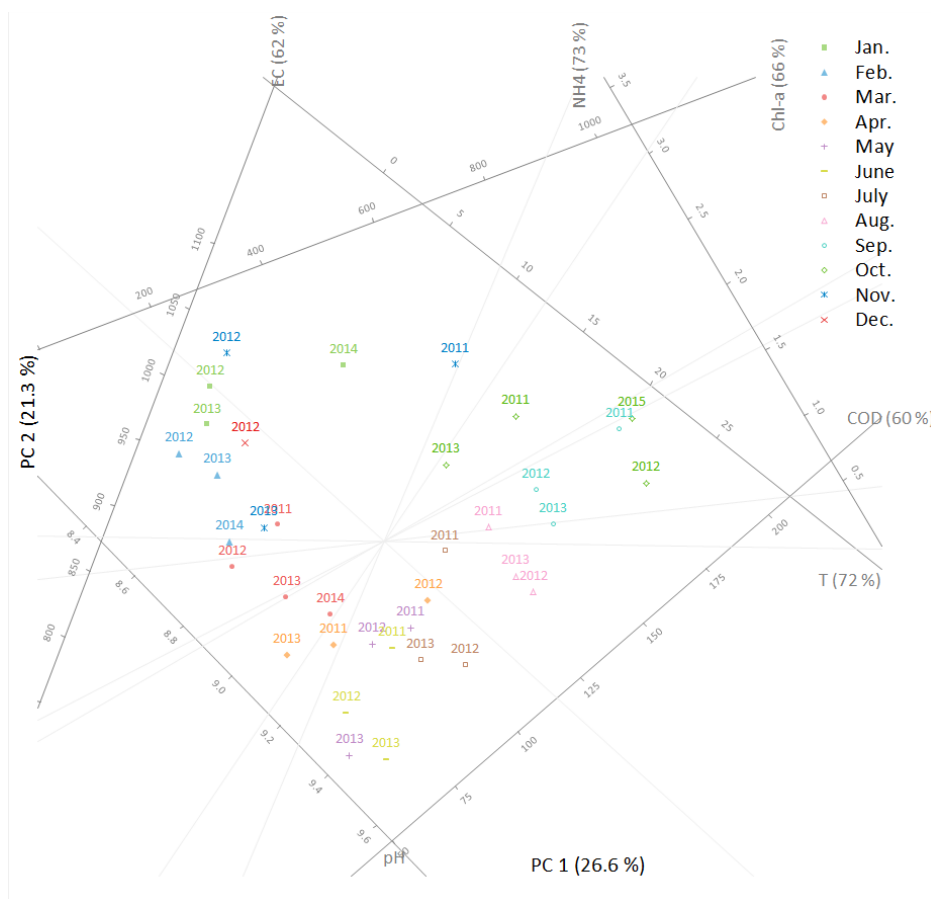




Слика 4 – Удео варијансе у главним компонентама  
Figure 4 – Scree plot



Слика 5 – Моноплот  
Figure 5 – Monoplot



Слика 6 – Биplot  
Figure 6 – Biplot

A strong correlation is noticed between T and pH. A strong negative correlation is presented with angles approaching 180 degrees, such as detected between T and NO<sub>3</sub> while the absence of correlation is depicted with angles nearing 90 degrees, e.g. COD and pH. Figure 6 presents a biplot which is used to identify possible clustering of the data. The biplot was set to show the axes only for the quality parameters whose variation is contained within the two main principal components more than 50%. These include pH, T, COD, EC, NH<sub>4</sub> and Chl-a. The points on the drawing are the measured values of

Друга значајна карактеристика моноплота су углови које посматране стрелице заклапају једна са другом. Мањи угао сугерише већу позитивну корелацију параметара, док угао око 180° сугерише велику негативну корелацију. Одсуство корелације се представља приближно ортогоналним стрелицама. Слика 6. приказује биplot који се користи ради идентификације потенцијалног груписања параметара на основу разних карактеристика. Слика приказује само осе за параметре квалитета које прве две главне компоненте

these parameters which can be determined by drawing an orthogonal line towards the axes of the considered parameter and reading the value it shows. The different symbols and colors indicate different months through years. By carefully analyzing the drawing one can recognize a tendency of the data to group in similar months over the years. For example, we can notice the grouping of the March data (red circles) in one segment and the clustering of the August data (pink triangles) in the other. This indicates there is a similarity of the data acquired in the same months through the years.

#### 4. CONCLUSIONS

This work focused on the implementation of the PCA for the analysis of the fourth sector of Lake Palic. Relying on the utilized method we were able to attain a better understanding of the relationships between various water quality parameters that would under normal circumstances be hard or even impossible to identify.

One important outcome of the conducted analysis was the identification of the clustering tendencies of the measurements in the same (similar) months through the years, rather than the grouping of annual values. This type of conclusion is important since it can enhance further data evaluation. In cases of an upcoming multivariate analysis that would include multivariate regression, such insight can significantly improve the quality of the developed correlations.

#### ACKNOWLEDGEMENTS

This work was funded by the Ministry of Education, Science and Technical Development of the Republic of Serbia.

описују са преко 50%, а обухватају рН, Т, COD, ЕС, NH<sub>4</sub> и Chl-а. Тачке на слици представљају мерене вредности ових параметара а вредности се могу очитати повлачењем управне линије од тачке до осе датог параметра. Различити симболи и боје показују мерења која се односе за различите месеце. Анализом ових резултата се примећује тенденција груписања података везаних за исте месеце током година. На пример, примећује се груписање црвених кругова који се односе на податке за месец март за више година. Слично се примећује и за податке из августа који су означени роза троугловима и груписани су на другој локацији. Са овиме се долази до значајне информације везане за мерења која су сличнија по истим месецима током година него током једне године.

#### 4. ЗАВРШНА РАЗМАТРАЊА

Циљ рада је анализа примене методе АГК на примеру четвртог сектора језера Палић. Применом ове методе је омогућено боље разумевање тенденције груписања података које може да се примени приликом даљих истраживања. Један од значајних закључака је идентификација сезонске карактеристике параметара квалитета што се може применити приликом мултиваријабилне анализе. Код спровођења поступка мултиваријабилне регресије се показало да овакве информације значајно доприносе квалитету развијених функција.

#### ЗАХВАЛНИЦА

Овај рад је финансиран од стране Министарства за образовање, науку и технолошки развој Републике Србије.

## REFERENCES

- [1] Sergeant, C.J., Starkey, E.N., Bartz, K.K., Wilson, M.H., & Mueter, F.J. A practitioners guide for exploring water quality patterns using principal components analysis and Procrustes. *Environmental Monitoring and Assessment*, 2016, Vol. 188, pp. 249.
- [2] Calazans, G.M., Pinto, C.C., da Costa, E.P., Perini, A.F., & Oliveira, S.C. The use of multivariate statistical methods for optimization of the surface water quality network monitoring in the Paraopeba river basin, Brazil. *Environmental Monitoring and Assessment*, 2018, Vol. 190, pp 491.
- [3] Mohamed, I., Othman, F.N., Ibrahim, A.I., Alaa-Eldin, M.E., & Yunus, R.M. Assessment of water quality parameters using multivariate analysis for Klang River basin, Malaysia. *Environmental Monitoring and Assessment*, 2014, Vol.187, pp. 4182.
- [4] Kumari, R., & Sharma, R.C. Assessment of water quality index and multivariate analysis of high altitude sacred Lake Prashar, Himachal Pradesh, India. *International Journal of Environmental Science and Technology*, 2019, Vol. 16, pp. 6125-6134.
- [5] Satheeshkumar, P., & Khan, A.B. Identification of mangrove water quality by multivariate statistical analysis methods in Pondicherry coast, India. *Environmental Monitoring and Assessment*, 2012, Vol. 184, pp. 3761-3774.
- [6] Pastor, K.A., Acanski, M.M., Vujic, Dj.N., Jovanovic, Dj., & Wienkoop, S. Authentication of Cereal Flours by Multivariate Analysis of GC-MS Data. *Chromatographia*, 2016, Vol. 79(19-20), pp. 1387-1393.
- [7] Хорват, М., Хорват, З., Габрић, О.: Транспорт наноса и промена коте дна у мрежи отворених токова, Зборник радова Међународне конференције Савремена достигнућа у грађевинарству 2015, Универзитет у Новом Саду, Грађевински факултет Суботица, 2015., стр. 621 – 626.