

APPLICATION OF MULTIVARIATE ANALYSIS FOR THE EVALUATION OF LAKE WATER QUALITY

ПРИМЕНА МУЛТИВАРИЈАБИЛНЕ АНАЛИЗЕ ЗА ОЦЕНУ КВАЛИТЕТА ВОДЕ ЈЕЗЕРА

Mirjana Horvat¹
Zoltan Horvat²

UDK 556.113/.115+628.113
DOI: 10.14415/zbornikGFS37.03
CC-BY-SA 4.0 license

Summary: Standard monitoring of lake water usually depends on monthly measurements that are affected by financing capabilities and weather conditions, often leading to various data being left out. Aiming to solve this issue, we enrolled on the task of developing a set of fitted equations allowing the researchers to predict the values of various water quality parameters even in situations when not all of the data are available. By implementing a series of tests we were able to establish equations with reasonable adjusted correlation coefficient values suggesting feasible projections of the considered output parameters.

Keywords: lake water quality, water quality parameters, multivariate regression.

Резиме: Стандардни приступ мониторинга језера се обично базира на месечним вредностима мерења која директно зависе од финансијских могућности и временских услова, често резултујући у изостајању појединих параметара. Са циљем решавања овог проблема је уложен напор у развоју сета фитованих једначина које би истраживачима омогућиле разумну предикцију вредности од интереса и у ситуацијама када поједини подаци нису обухваћени мерењима. Применом серије тестова смо дошли до низа зависности са прихватљивим вредностима коригованих коефицијента корелације, што указује на поузданост развијених функција.

Кључне речи: квалитет воде језера, параметри квалитета воде, мултиваријабилна регресија.

1. INTRODUCTION

The appropriate monitoring of water quality is essential for the proper environmental management of lakes and other water bodies. However, performing all the essential measurements to achieve this goal is often time-consuming and costly.

1. УВОД

Одговарајући мониторинг квалитета воде је од кључног значаја при одрживом управљању језерима и другим водним телима. Међутим, спровођење свих неопходних мерења је често временски захтевно и финансијски скупо. Дакле, у оквиру

¹ Mirjana Horvat, PhD, C.E., University of Novi Sad, Faculty of Civil Engineering Subotica, Kozaracka street 2a, 24000 Subotica, Serbia, e-mail: jsicm@gf.uns.ac.rs

² 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

Therefore, the possibility of predicting certain water quality parameters using other (more basic and easier to determine) water quality parameters was investigated. This was done using data from Lake Palic, which is located in northern Vojvodina in Serbia, approximately 8 km east from the city of Subotica. The lake's surface is 4.6 km², with an average depth of 1.8 m (the deepest point can be assumed to be approximately 3.4 m). The lake itself is divided into four sectors, while the first sector consists of three lagoons. There are a number of inflows into the lake, e.g. rainfall, surface runoff and groundwater. However, the main inflow to the first lagoon comes from the Wastewater Treatment Plant Subotica. The water flows downstream through sectors one to four, and finally exits the lake and ends up in Lake Omladinsko. This fact makes the proper monitoring of the lake essential, since it is prone to eutrophication by its nature, however the inflow of treated sewage water makes this issue even more pressing. The goal of this paper was to establish fitted equations (prediction models) for determining two water quality parameters, namely chlorophyll-a and ammonia-nitrogen, using other water quality parameters that can be determined easily (temperature, dissolved oxygen, power of hydrogen, suspended solids and electric conductivity). For this purpose a number of statistical methods have been used throughout this research [1,2,3,4,5,6].

2. MATERIALS AND METHODS

Measurements of water quality parameters were conducted for the time period from 2011 to 2015 at the point of outflow from the fourth sector of Lake Palic. As input parameters water temperature (Temp), electric conductivity (Cond), power of hydrogen (pH), dissolved oxygen (DO) and

овог рада је анализирана могућност предвиђања одговарајућих параметара квалитета воде користећи друге (лакше за мерење) параметре квалитета воде. Ово је урађено користећи податке са језера Палић, које се налази у северној Војводини у Србији, приближно 8 км источно од Суботице. Површина језера је 4.6 км², са просечном дужином од 1.8 м (може се претпоставити да је најдубља тачка језера 3.4 м). Само језеро је подељено на четири сектора, док је први сектор подељен на три лагуне. Постоји низ улива у језеро, као што су падавине, површинско сливање и подземна вода. Међутим, најзначајнији улив потиче из постројења за пречишћавање отпадних вода Суботица. Ова вода потом тече низводно кроз сваки сектор језера Палић, након чега напушта водно тело и завршава у Омладинском језеру. Одговарајући мониторинг језера је јако битан, јер је оно само по себи склоно еутрофикацији, а прерађена отпадна вода само поспешује овај процес. Циљ овог рада је био да се одреде фитоване једначине (тј. модели за предвиђање) за одређивање хлорофила-а и амонијачног азота користећи друге параметре квалитета воде који су лакши за мерење (температура, растворени кисеоник, рН, суспендоване материје и електрична проводљивост). Ради постизања овог циља, коришћен је низ статистичких метода [1,2,3,4,5,6].

2. МАТЕРИЈАЛИ И МЕТОДЕ

Мерења параметара квалитета воде су спроведена од 2011 до 2015 године код места излива воде из четвртог сектора језера Палић. Као улазни параметри олажане су вредности температуре (Temp), електричне проводљивости (Cond), рН вредност (pH), растворени

suspended solids (SS) were measured. While as output parameters, chlorophyll-a (Chl-a) and ammonia-nitrogen (NH₄) were monitored. The general temporal frequency of the gathered data was once a month, although during the overall monitoring period in some months the measurements were omitted entirely.

To make prediction models, multivariate regression was used. In order to determine the quality of the fitted equations three numerical and two graphical methods were used. The first numerical method was the value of the adjusted correlation coefficient R^2_{adj} between predicted and measured values. If this coefficient was below 0.3 there is a negligible correlation, if it is in the range between 0.3 and 0.5 a low correlation can be assumed, the domain between 0.5 and 0.7 indicates a moderate correlation, the domain between 0.7 and 0.9 indicates a high correlation, while if the coefficient is higher than 0.9 a very high correlation is assumed. The second numerical indicator for the fitted equation would be its p-value. This number represents the statistical odds that the null hypothesis (that the predicted value is a constant rather than it is dependent on the input parameters) is the correct one. For example if the fitted equation has a p-value of 0.05, there is a 5% chance that the predicted parameter is actually a constant and it is not dependent on the input parameters as the fitted equation suggests. The final, numerical test would be the p-value of each term in the fitted equation. Namely, if an equation term has a p-value of 0.06, there is a 6% chance that the analyzed term does not contribute to the fitted model. To summarize, a lower p-value, both for the entire equation and for each term in it indicates a better prediction model. The first graphical evaluation technique depicts the predicted and actual values of the output water quality parameter. If the clustering of the actual values is

oxygen (DO) and suspended matter (SS). As output parameters, measured values of chlorophyll-a (Chl-a) and ammonia-nitrogen (NH₄). Measurements were taken once a month, while during the overall monitoring period in some months the measurements were omitted entirely.

To make prediction models, multivariate regression was used. In order to determine the quality of the fitted equations three numerical and two graphical methods were used. The first numerical method was the value of the adjusted correlation coefficient R^2_{adj} between predicted and measured values. If this coefficient was below 0.3 there is a negligible correlation, if it is in the range between 0.3 and 0.5 a low correlation can be assumed, the domain between 0.5 and 0.7 indicates a moderate correlation, the domain between 0.7 and 0.9 indicates a high correlation, while if the coefficient is higher than 0.9 a very high correlation is assumed. The second numerical indicator for the fitted equation would be its p-value. This number represents the statistical odds that the null hypothesis (that the predicted value is a constant rather than it is dependent on the input parameters) is the correct one. For example if the fitted equation has a p-value of 0.05, there is a 5% chance that the predicted parameter is actually a constant and it is not dependent on the input parameters as the fitted equation suggests. The final, numerical test would be the p-value of each term in the fitted equation. Namely, if an equation term has a p-value of 0.06, there is a 6% chance that the analyzed term does not contribute to the fitted model. To summarize, a lower p-value, both for the entire equation and for each term in it indicates a better prediction model. The first graphical evaluation technique depicts the predicted and actual values of the output water quality parameter. If the clustering of the actual values is

closer to the predicted values, the model is more accurate. The second graphical representation of the goodness of the fit is the plotted standardized residuals for the fitted equation. The closer these are to a normal distribution, the more likely it is that the discrepancies between the actual and predicted values are random by nature. The other principle used when constructing the fitted models is based on previous research. By using principal component analysis, it has been determined that there is a clustering of water quality data by seasons. This means that the processes that play out in the lake produce circumstances that are more similar to each other in the same months of different years. Therefore, relatively homogenous time periods have been identified when conditions in the lake are similar. These time periods are from January to March, from April to June, from July to September and from October to December. The proposed fitted models were compiled for each of these relatively homogenous time periods.

3. RESULTS AND DISCUSSION

Chlorophyll-a is the first water quality parameter for which a fitted equation was compiled. As input parameters temperature, dissolved oxygen, power of hydrogen, suspended solids and electric conductivity were used. Table 1 lists the fitted models for the four identified homogenous time periods.

оваквом приказу груписање мерених вредности у близини срачунатих вредности указује на тачност модела. Други начин графичког приказивања резултата је исцртавање стандардизованих остатака фитоване једначине. Што су ове вредности ближе нормалној расподели, то је вероватније да су одступања мерених вредности од срачунатих случајног карактера. Током формулисања фитованих једначина је коришћен и један додатни принцип идентификован током пређашњих истраживања. Коришћењем анализе главних компоненти је установљено да постоји груписање података по годишњим добима. Ово подразумева да су услови који владају у језеру сличнији у истом месецу различитих година него током једне календарске године. Дакле, идентификовани су релативно хомогени временски периоди, и то од Јануара до Марта, од Априла до Јуна, од Јула до Септембра и од Октобра до Децембра. На основу овог сазнања фитоване једначине су формулисане за сваку од ових периода посебно.

3. РЕЗУЛТАТИ И ДИСКУСИЈА

Хлорофил-а је први параметар квалитета воде за који је формулисана фитована једначина. Као улазни параметри су одабрани температура, растворени кисеоник, рН, суспендоване материје и електрична проводљивост. Табела 1 садржи једначине за четири идентификована хомогена временска периода.

Табела 1 – Фитоване једначине за хлорофил-а
Table 1 – Fitted equations for chlorophyll-a

Period:	Equation:
Jan-Mar	$Chl - a = 201 + 262.7 \cdot Temp - 0.2741 \cdot Temp \cdot Cond$
Apr-Jun	$Chl - a = -371.8 + 50.39 \cdot DO + 1.656 \cdot pH \cdot SS - 1.317 \cdot DO \cdot SS$
Jul-Sep	$Chl - a = 950.6 - 0.02515 \cdot Temp \cdot Cond - 0.8188 \cdot Temp \cdot SS + 0.02771 \cdot Cond \cdot SS$
Oct-Dec	$Chl - a = -9974 + 1218 \cdot pH + 94.96 \cdot SS - 11.1 \cdot pH \cdot SS$

The numerical analysis of the fitted equation's quality is summarized in Table 2. For periods Jan-Mar and Oct-Dec the equations have a moderate correlation with the actual values, while for periods Apr-Jun and Jul-Sep there is a high correlation. The highest p-value for the fit is 0.0384, while the highest p-value for the equation terms is 0.0571 (time period Oct-Dec). Since the criteria for the p-value was to stay below 0.05 for the fit itself and below 0.1 for the equation terms, it can be concluded that the fitted models are statistically significant in predicting the values of chlorophyll-a.

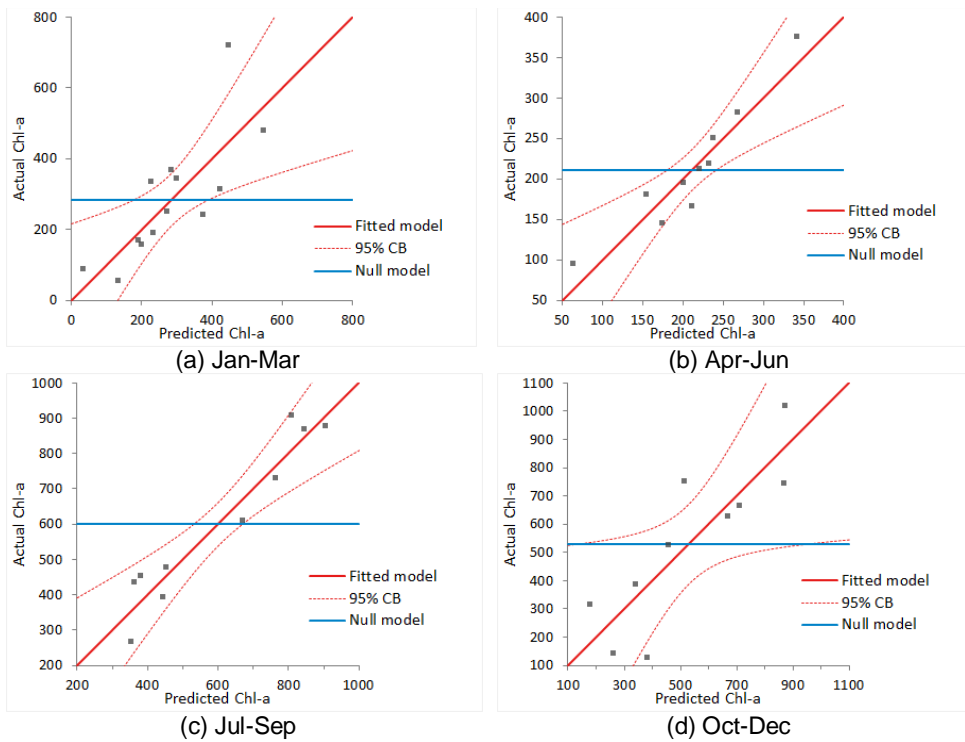
Преглед нумеричке анализе фитованих модела је дата у Табели 2. За периоде од Јануара до Марта и од Октобра до Децембра, фитоване једначине имају умерену корелацију у односу на мерене вредности, док за периоде од Априла до Јуна и од Јула до Септембра корелација се може оценити као висока. Највиша р-вредност за фитоване једначине је 0.0384, док је максимална р-вредност за поједине чланове једначина износила 0.0571 (период од Октобра до Децембра). Пошто је критеријум за р-вредност био да остане испод 0.05 за фитоване једначине, односно испод 0.1 за поједине чланове једначина, може се закључити да су формулисани модели статистички значајни при предвиђању вредности хлорофила-а.

Табела 2 – Анализа фитованих једначина за хлорофил-а
Table 2 – Analysis of fitted equations for chlorophyll-a

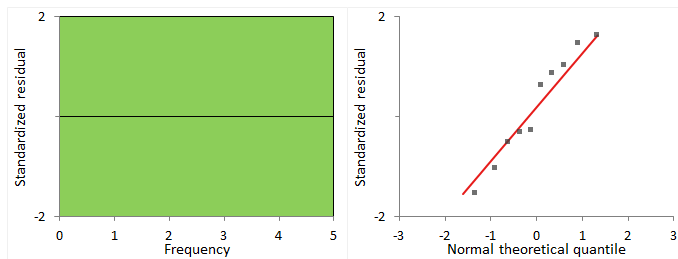
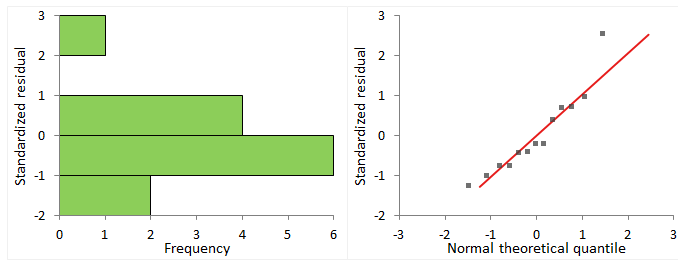
Period	Jan-Mar	Apr-Jun	Jul-Sep	Oct-Dec
R^2_{adj}	0.539	0.825	0.881	0.595
p-value for the fit	0.0084	0.0033	0.0011	0.0384
p-value for equation terms	Temp	0.0075		
	DO		0.0012	
	pH			0.0123
	SS			0.0571
	DO x SS		0.0093	
	pH x SS		0.0106	0.0501
	Temp x SS			0.0017
	Temp x Cond	0.0133		0.0303
Cond x SS			0.0005	

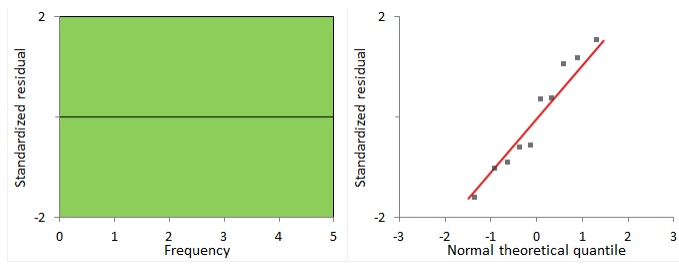
Figure 1 depicts the computed and actual values of chlorophyll-a, while Figure 2 presents the standardized residuals for the fitted equation, both for the four time intervals. It can be stated that the actual values are relatively close to the predicted ones (Fig. 1), while the residuals follow the normal distribution adequately (Fig. 2).

Слика 1 представља срачунате и мерене вредности хлорофила-а, док Слика 2 приказује стандардизоване остатке фитованих једначина за сваки од четири временских интервала посебно. Може се рећи да се мерене вредности предвиђеног параметра налазе у релативној близини срачунатих вредности (Слика 1), док стандардизовани остаци релативно правилно прате нормалну расподелу.

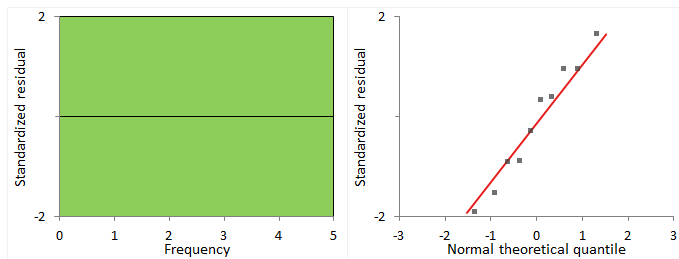


Слика 1 – Срачунате и стварне вредности за хлорофил-а
 Figure 1 – Computed and actual values for chlorophyll-a





(c) Jul-Sep



(d) Oct-Dec

Слика 2 – Стандардизовани остаци фитоване једначине за хлорофил-а
Figure 2 – Standardized residuals for the fitted equation for chlorophyll-a

Ammonium-nitrogen is the second water quality parameter for which a fitted equation was compiled. As input parameters temperature, dissolved oxygen, power of hydrogen, suspended solids and electric conductivity were used. Table 3 lists the fitted models for the four identified time periods.

Амонијачни азот је други параметар квалитета воде за који су формулисане фитоване једначине. Као улазни параметри су коришћени температура, растворени кисеоник, pH, суспендоване материје и електрична проводљивост. Табела 3 садржи једначине за четири идентификована хомогена временска периода.

Табела 3 – Фитоване једначине за амонијачни азот
Table 3 – Fitted equations for ammonium nitrogen

Period:	Equation:
Jan-Mar	$NH_4 = -5.871 + 0.009458 \cdot Cond - 0.0001508 \cdot Temp \cdot Cond$
Apr-Jun	$NH_4 = -392.9 + 67.61 \cdot pH - 0.7972 \cdot Temp \cdot pH - 8.925 \cdot DO + 0.009903 \cdot DO \cdot Cond + 0.3004 \cdot Cond - 0.06024 \cdot pH \cdot Cond + 0.008192 \cdot Temp \cdot Cond$
Jul-Sep	$NH_4 = 85.87 - 3.215 \cdot Temp + 0.1687 \cdot Temp \cdot SS - 0.0005631 \cdot Temp \cdot Cond - 11.69 \cdot pH + 0.01683 \cdot Cond + 0.4377 \cdot Temp \cdot pH$
Oct-Dec	$NH_4 = 53.48 - 0.1823 \cdot Temp - 0.9377 \cdot pH - 4.589 \cdot DO - 0.03385 \cdot Cond + 0.003791 \cdot DO \cdot Cond$

The subsequent numerical analysis of the fitted equation's quality is summarized in Table 4. For all time periods a very high correlation was

Нумеричка анализа формулисаних једначина је приказана у Табели 4. За све временске периоде идентификована је веома висока

detected between the predicted and the actual values. The highest p-value for the fit is 0.0685, while the highest p-value for the equation terms is 0.0701 (time period Jul-Sep). Since the criteria for the p-value was to stay below 0.05 for the fit itself and below 0.1 for the equation terms, it can be concluded that the fitted models are statistically significant only to a certain degree in predicting the values of ammonium-nitrogen.

корелација између предвиђених и осматрених вредности амонијачног азота. Највиша р-вредност за фитоване једначине износи 0.0685, док је максимална р-вредност појединачних чланова једначина 0.0701 (временски период од Јула до Септембра). Пошто је критеријум за р-вредност био да остане испод 0.05 за фитоване једначине, односно испод 0.1 за поједине чланове једначина, може се закључити да су формулисани модели статистички значајни само до неког степена при предвиђању вредности амонијачног азота.

Табела 4 – Анализа фитованих једначина за амонијачни азот
Table 4 – Analysis of fitted equations for ammonium nitrogen

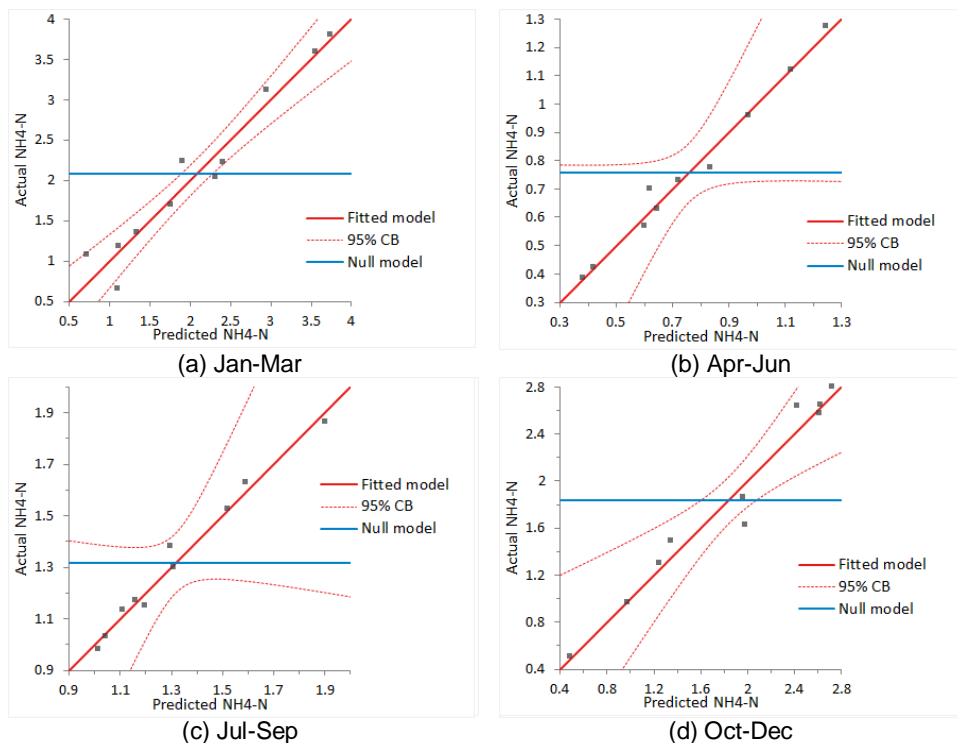
Period	Jan-Mar	Apr-Jun	Jul-Sep	Oct-Dec	
R^2_{adj}	0.933	0.930	0.910	0.919	
p-value for the fit	<0.0001	0.0537	0.0685	0.0054	
p-value for equation terms	Temp		0.0463	0.0229	
	DO		0.0294	0.0179	
	pH		0.0428	0.0379	0.0653
	SS			0.0282	
	Cond	<0.0001	0.0517	0.0521	0.0371
	Temp x Cond	0.0013	0.0401	0.0701	
	Temp x SS			0.0277	
	pH x Cond		0.0432		
	pH x Temp		0.0397	0.0392	
DO x Cond		0.0292		0.0281	

Figure 3 depicts the computed and actual values of ammonium-nitrogen, while Figure 4 presents the standardized residuals for the fitted equation, both for the four time intervals. It can be stated that the actual values are relatively close to the predicted ones, while the residuals follow the normal distribution adequately. However, it should be noted that the 95% confidence boundary on Figures 3(b) and 3(c) is quite wide. This indicates that the fitted model can't predict the values of ammonium-nitrogen with too much certainty. This is in accordance with the data in Table 4, since for these two time

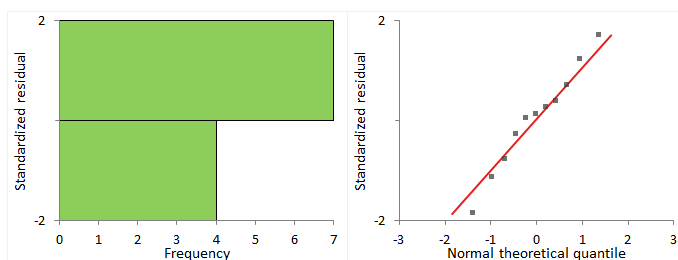
Слика 3 приказује срачунате и мерене вредности амонијачног азота, док се на Слици 4 виде стандардизовани остаци фитованих једначина за сва четири идентификована временска интервала посебно. Може се рећи да се мерене вредности предвиђеног параметра налазе у релативној близини срачунатих вредности (Слика 3), док стандардизовани остаци релативно правилно прате нормалну расподелу (Слика 4). Међутим, такође се примећује да је интервал поверења од 95% прилично широк на Сликама 3(б) и 3(ц). Ово указује на чињеницу да фитоване

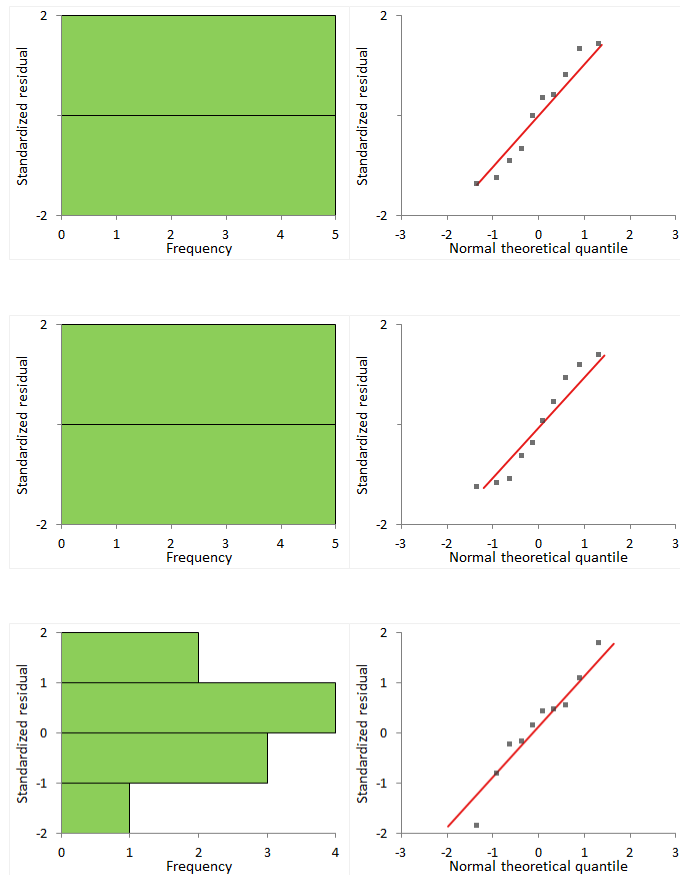
periods (Apr-Jun and Jul-Sep) the p-value of the fitted equations is 0.0537 and 0.0685, which is above the 0.05 criteria.

једначине не могу да предвиде вредности амонијачног азота са великом сигурношћу. Ово је у складу са Табелом 4, пошто је за ова два временска интервала (од Априла до Јуна и од Јула до Септембра) р-вредност фитованих једначина 0.0537 и 0.0685, што је изнад усвојеног критеријума од 0.05.



Слика 3 – Срачунате и стварне вредности за амонијачни азот
Figure 3 – Computed and actual values for ammonium nitrogen





Слика 4 – Стандардизовани остаци фитоване једначине за амонијачни азот
 Figure 4 – Standardized residuals for the fitted equation for ammonium nitrogen

It should be mentioned that although the adjusted correlation coefficient values for the fitted equation in case of ammonium-nitrogen are higher than those for chlorophyll-a, the models for the latter are statistically more significant. This can be clearly seen in the p-values. For chlorophyll-a these range from 0.0011 to 0.0384, while in case of ammonium-nitrogen these values are between <0.0001 and 0.0685. Furthermore, since the p-values for models predicting ammonium-nitrogen for the time periods Apr-Jun and Jul-Sep are higher

Такође је неопходно поменути да иако су кориговане вредности коефицијента корелације за фитоване једначине за амонијачни азот веће од оних за хлорофил-а, формулисани модели за хлорофил-а су статистички значајнији. Ово се може видети у р-вредностима. За хлорофил-а оне износе од 0.0011 до 0.0384, док се у случају за амонијачни азот крећу у интервалу од <0.0001 до 0.0685. А пошто су р-вредности за фитоване једначине за предвиђање вредности амонијачног азота у временском периоду од

than 0.05, they should be used with caution.

4. CONCLUSIONS

The presented research investigated the possibility of establishing fitted models for predicting chlorophyll-a and ammonium-nitrogen using only basic water quality data, namely temperature, dissolved oxygen, electric conductivity, power of hydrogen and dissolved solids. This was done using data from the fourth sector of Lake Palic gathered from 2011 to 2015. The fitted models were compiled for four relatively homogenous time periods using data from every available year, namely Jan-Mar, Apr-Jun, Jul-Sep and Oct-Dec. For the prediction of chlorophyll-a a statistically significant (p -value of the fit <0.05) fitted model was established for every homogenous time period. On the other hand, for the prediction of ammonium-nitrogen statistically significant models (p -value of the fit <0.05) were established for time periods Jan-Mar and Oct-Dec, while for periods Apr-Jun and Jul-Sep the p -values of the fit were 0.0537 and 0.0685. Therefore the fitted models for predicting ammonium-nitrogen for Apr-Jun and Jul-Sep should be used with extreme caution.

The presented research proved the feasibility of the concept that water quality parameters that are more difficult to measure (such are chlorophyll-a and ammonium-nitrogen) can be predicted with a reasonable accuracy with fitted models using only relatively simple (and easy to measure) parameters.

Априла до Јуна и од Јула до Септембра веће од 0.05, оне се могу користити само уз велику резерву.

4. ЗАКЉУЧАК

Приказано истраживање анализира могућност формулисања фитованих једначина за предвиђање вредности хлорофила-а и амонијачног азота користећи основне параметре квалитета воде (температура, растворени кисеоник, електрична проводљивост, рН и суспендоване материје). Ово је урађено за податке из четвртог сектора језера Палић који су сакупљани од 2011 до 2015 године. Фитоване једначине су одређене за четири релативно хомогена временска периода, тј. од Јануара до Марта, од Априла до Јуна, од Јула до Септембра и од Октобра до Децембра. У случају хлорофила-а формулисани су статистички значајни (p -вредност једначине <0.05) модели за сваки хомогени временски период. Са друге стране, у случају амонијачног азота су формулисани статистички значајни (p -вредност једначине <0.05) модели само за периоде од Јануара до Марта и од Октобра до Децембра, док су p -вредности за период од Априла до Јуна и од Јула до Септембра износили редом 0.0537 и 0.0685. Дакле, фитоване једначине за предвиђање вредности амонијачног азота за временске периоде од Априла до Јуна и од Јула до Септембра се могу користити само уз велики опрез.

Сprovedено истраживање је доказало да је могуће предвидети параметре квалитета воде који су тежи за мерење (као што су хлорофил-а и амонијачни азот) користећи само релативно једноставне параметре који су далеко лакши за мерење.

ACKNOWLEDGEMENTS

This work was funded by the Ministry of Education, Science and Technological development of the Republic of Serbia.

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.