Classification and analysis of key parameters in predicting the state of facultative oligotrophs in two different reservoirs

Ivana Radojević[1]*, Nenad Stefanović[2], Aleksandar Ostojić[1]

[1] University of Kragujevac, Faculty of Science, Institute of Biology and Ecology, 34000 Kragujevac, Radoja Domanovića 12, Republic of Serbia; e-mail: ivana.radojevic@pmg.kg.ac.rs
[2] University of Kragujevac, Faculty of Technical Sciences Čačak, Department of Information Technologies, Svetog Save St., 65, 32102 Čačak, Republic of Serbia;

* *Corresponding author*

DOI: 10.46793/ICCBI23.312R

Abstract: This study analyzes the condition, influence and dependence of facultative oligotrophs, with different physico-chemical and biological parameters determined in the routine monitoring of the status of natural water quality. The SeLaR database was used for advanced data analysis - statistical methods and data mining. Data were analyzed for reservoirs of different morphometric qualities, different positions, trophic status and dominant bacterial communities, Gruža and Grošnica. Classification and analysis of key influencing factors were applied in this research. The results indicate that the designed data mining system can successfully analyze the condition and impact of facultative oligotrophs in open waters, thus providing new information and knowledge valuable for research investigation and decision making.

Keywords: data mining, reservoirs, facultative oligotroph, classification, influential parameters

1. Introduction

For a long time, the world has been faced with the fact that there is a contradiction between different needs in drinking water and relatively clean water resources. One way to solve the problem is to build and use reservoirs. That is why there is a need to maintain the condition of the reservoirs at a satisfactory level. Continuous monitoring of water conditions, i.e., water monitoring, includes monitoring of various physical, chemical, and biological parameters. Many microbiological parameters, either from a strictly sanitary or a wider ecological aspect, indicate the nature and intensity of pollution, as well as whether the impact of the pollutant is immediate or permanent.

Various tools and techniques of information technologies, as well as methods of data analysis such as data mining, become an important part in monitoring the state of water quality, enable the prediction of changes and are important in the processes of sustainable management of water resources. Classification, modeling, and interpretation of a large amount of data is an important step when observing water quality [1]. These techniques and methods have been applied and validated as tools for monitoring microbiological quality indicators in aquatic ecosystems [2].

Therefore, the subject of this research is the application of a specialized data mining model of the SeLaR information system in the ecological analysis of the bacterial community of facultative oligotrophs (FO) in reservoirs with different morphometric qualities, positions, trophic status, and dominant bacterial community.

2. Materials and Methods

This research was performed using the SeLaR IS, which combines relevant data on reservoirs in Serbia and enables advanced methods of analysis. For the analysis of data, two reservoirs with different morphometric properties, trophic status, and dominant microbial community Gruža and Grošnica, intended for water supply, were selected. The material in this research are data sets standard routine and broader scientific hydrobiological tests from certain periods selected reservoirs. The data include physico-chemical, biochemical, microbiological and other biological parameters. Gruža - data set 3 (VI 2003. - VII 2009) and Grošnica - data set 2 (V 1997 - X 1998) were analyzed [3]. The units of measurement of all parameters are according to the standards of routine water testing and are given in section 5.1. Basic statistical overview and created models [3].

For the realization of the goal the descriptive methods, classification, and analysis of influencing factors, were used. The methods use different algorithms: a decision tree is used for classification and naïve bayes for key impacts [4].

3. Results and discussion

3.1 Classification and analysis of key influencing factors

Gruža - data set 3 has 1181 objects and Grošnica - data set 2 has a total of 182 objects. Existing values are classified into five classes. The first classes of data are the most common for both reservoirs. The classification, class values, and the number of data with probabilities are given in Table 1.

Table 1. Classification of values of facultative oligotrophs in the reservoir Gruža - data set 3 and Grošnica - data set 2.

Gruža - data set 3			
Class	**Value**	**Number of data**	**Probability**
1.	< 2876	771	64.63%
2.	2876-13912	150	12.75%
3.	13912-27088	26	2.39%
4.	27088-54152	13	1.31%
5.	≥ 54152	2	0.39%
-	-	219	18.52%
Grošnica - data set 2			
1.	< 307	63	33.16%
2.	307-673	42	22.56%
3.	673-1267	31	17.00%
4.	1267-1852	10	6.40%
5.	≥ 1852	8	5.39%
-	-	28	15.49%

The review of the results of the analysis of key influence parameters with impact

over 50, as well as the review of the classes with values is given in Table 2.

Table 2. Key parameters and their impact on the values of the facultative oligotrophs (FO) in the reservoir Gruža - data set 3 and Grošnica - data set 2.

Gruža - data set 3			
Parameter	**Value**	**Class FO**	**Relative influence of the parameter**
heterotroph psihrophile (Hp)	< 2872	< 2876	100
heterotroph mesophile (Hm)	< 1418	< 2876	50
Hp	2872-9363	2876-13912	100
depth	27	2876-13912	50
Hp	9363-20742	13912-27088	100
Hm	1418-6997	13912-27088	57
location	22-27	13912-27088	20
Hp	20742-39809	27088-54152	100
Hp	≥ 39809	27088-54152	88
total coliform	≥ 1500	27088-54152	50
nitrites	0.29-0.52	≥ 54152	100
Hp	≥ 39809	≥ 54152	85
electrical conductivity	≥ 672	≥ 54152	67
Fe	0.39-1.16	≥ 54152	56
Grošnica - data set 2			
electrical conductivity	≥ 458	< 307	100
Hp	< 181	< 307	96
chlorophyll a	< 3.22	< 307	60
month	VI	< 307	52
Phosphatase activity index	1.42-2.94	673-1267	100
Fe	0.05	673-1267	76
Hp	181-475	673-1267	58
Cladocera	306-491	673-1267	50
Phosphatase activity index	< 1.42	1267-1852	100
ammonia	2.61-3.13	1267-1852	85
month	II	1267-1852	72
Copepoda	345-689	≥ 1852	77
Fe	0.08	≥ 1852	51
depth	3	≥ 1852	51

For reservoir Gruža for most classes of facultative oligotrophs, the most influential parameter is heterotroph psychrophile (Hp). The exception is the fifth class, where the

most influential nitrites are in higher values. For this class, high values for electrical conductivity and Fe content have a significant relative influence, but these influences can be questioned due to the extremely small number of this class. Extremely high influence, as well as the phenomenon that low values of Hp follow low values of FO, and vice versa, is consistent with the high degree of correlation between these two parameters in this data set. Among other parameters whose relative influence is greater than 50, the following stand out: lower and medium values of heterotroph mesophile (Hm) for the first and third classes; depths over 25 m for the second class; total coliform in low values for the first class and in high values for the other classes. The analysis shows the importance of the spatial distribution with higher values at the tributaries and lower values from the dam to the bridge.

In the Grošnica reservoir, the first class is most influenced by high values for electrical conductivity and low values of Hp. Average and lower values of chlorophyll a and environmental conditions in June show a high degree of influence. The second class of facultative oligotrophs is characterized by the fact that there is no high influencing parameter. For the third class of FO, there are high influences of average and lower values of phosphatase activity index and Hp, average and higher values of Fe, as well as higher values of Cladocera. In the fourth class of FO, the impacts are relatively highly significant. This class is highly influenced by low phosphatase activity index, high ammonia values and environmental conditions in February. The fifth class is highly influenced by the greater abundance of Copepoda. The highest values of FO occur in February, the lowest in July, and a depth of 3 m. High abundance of FO is conditioned by low ammonia content and vice versa.

Facultative oligotrophs have a spatial distribution in the Gruža reservoir, and a seasonal distribution in the Grošnica reservoir.

3. Conclusions

The results of this research show that developed data mining models for classification and analysis of key influencing parameters can be used in monitoring the microbiological quality of aquatic ecosystems, like reservoirs. The results provide a clear classification of the values for the FO.

Acknowledgment

This research is funded by the Ministry of Education and Ministry of Science, Technological Development and Innovation, Republic of Serbia, Grants: No. 451-03-47/2023-01/ 200122.

References

[1] H. Boyacioglu, H. Boyacioglu, Surface *Water Quality Assessment by Environmetric Methods*. Environmental Monitoring and Assessment, 131 (2007) 371-376.

[2] T.C. Ogwueleka, F.N. Ogwueleka, *Data mining application in predicting Cryptosporidium spp. oocysts and Giardia spp. cysts concentrations in rivers*. Journal of Engineering Science and Technology, 5 (2010) 342-349.

[3] И.Д. Радојевић, Микробиолошки аспекти примене информационих система у мониторингу акумулација и језера, Докторска дисертација, Природно-математички факултет, Универзитет у Крагујевцу, (2012) 218 pp.

[4] D. Harts, Microsoft Office 2007 Bussines Intelligence, McGraw_Hill, (2008). ISBN-13: 978-0071494243; ISBN-10: 00714942