ESTABLISHING THE REPRESENTATIVE PARAMETERS FOR THE CCME WQI COMPUTATION

Zoltan Horvat [1]
Dániel Koch [2]
UDK: 628.161:519.233.5
DOI: 10.14415/konferencijaGFS2019.068
Summary: *This paper represents the steps necessary for the proper establishment of the parameters representative when computing the CCME Water Quality Index, demonstrated on the example of Lake Ludas. In order to avoid misleading representation of the surface water's quality, it is desirable to process the available water quality data and then chose only the parameters that influence the changes in water quality of the considered site. This is usually conducted by computing the Pearson's correlation coefficient for all of the available parameters. The results of this analysis is the ability to better asses which of the accessible data should be included in the computation of the CCME WQI. The method is presented on the water quality data for Lake Ludas for the year 2012.*

Keywords: *water quality, CCME WQI, Pearson's correlation coefficient*

1. INTRODUCTION

The water quality index WQI is a way in which the water quality is represented to a general public in a simpler, more descriptive manner instead of displaying a large number of scientific data. Furthermore, this is a way in which the water body's state is described in all documents emerging from previously conducted research, such as management and monitoring frameworks.
There are various methods for the evaluation of a WQI, from simpler such as the simple WQI method, and those more complex that try to implement more of the influencing parameters into themselves such as the Composite Water Quality Identification Index (CWQII), proposed by Xu [1], or the USA and Canadian water quality index method presented by Lumb [2]. After conducting a comprehensive analysis presented in [2], Lumb et. al. found that the Canadian Council of Ministers of the Environment Water Quality Index CCME WQI is the strictest from a few of the investigated methods for the evaluation of surface water quality. Bilgin [3] used this method for the evaluation of the Coruh River Basin. Almeida [4] evaluated the water quality of the Joanes River basin in Brazil was using the CCME method, where he outlined the flexibility and adequacy to present the real water quality using this method. A detailed representation of the CCME WQI method

[1] Doc. dr Zoltan Horvat, dipl.inž. građ., University of Novi Sad, Faculty of Civil Engineering Subotica, Kozaračka 2a, Subotica, Serbia, tel: +381 24 554 300, e – mail: horvatz@gf.uns.ac.rs
[2] Dániel Koch, MSc, CE, National University of Public Service, Faculty of Water Sciences, Bajcsy-Zsilinszky utca 12-14, H-6500 Baja, Hungary, e – mail: koch.daniel@uni-nke.hu

is given in [5,6]. On the other hand, one must take great care when using measured field data [7].
This paper considers the representation of selecting the representative parameters when engaging the CCME WQI method. The procedure is explained relying on the example of determining these parameters for Lake Ludas in Serbia using measurements for year 2012.

2. THE STUDY AREA

The importance of the water quality status of Lake Ludas in Serbia is outlined by the fact that it represents one of the Ramsar sites in Serbia. A Ramsar site is a wetland site designated to be of international importance under the Ramsar Convention, while the Convention on Wetlands, also known as the Ramsar Convention, is an intergovernmental environmental agreement that was established by UNESCO in 1971. This agreement provides for national action and international cooperation concerning the conservation of wetlands, and sustainable use of their resources. Ramsar identifies wetlands that are of international relevance, including those that serve as waterflow habitats. The List of Ramsar wetlands of international importance contains a record of Ramsar sites. According to the data from the list of Ramsar Wetlands, Lake Ludas occupies the area of around 593 ha. The water in the lake is replenished from two main sources, the Kires channel flowing from the Hungarian side, and the Palic-Ludas channel.

3. PEARSON’S CORRELATION

Keeping in mind that Lake Ludas is a Nature Reserve, it is most important to implement regular monitoring and estimation of it’s water quality. As a result, the Public Health Institute Subotica carries our fairly regular measurements of quality parameters.
The determination of the water quality index requires the researcher to select those quality parameters that will properly represent the state of the investigated water body. Including to many parameters can give biased results in the same way that not including all of the important parameters would. In order to avoid these biased results that would result in steering the monitoring and management of the considered lake in the wrong direction, researches rely on the Pearson’s correlation.
The Pearson correlation coefficient (PCC), also called the bivariate correlation is used to measure the linear correlation between two parameters (variables). The correlation coefficient r has a value between +1 and −1, where 1 stands for a positive linear correlation, 0 suggests no linear correlation, and −1 means a complete negative linear correlation. The Pearson's correlation coefficient is the covariance of the two variables divided by the product of their standard deviations.

$$\mathrm{r} = \frac{N \cdot \sum xy - \left(\sum x\right)\cdot\left(\sum y\right)}{\sqrt{\left[N \cdot \sum x^2 - \left(\sum x\right)^2\right]\cdot\left[N \cdot \sum y^2 - \left(\sum y\right)^2\right]}} \tag{1}$$

where N marks the number of pairs of scores, Σxy is the sum of the products of paired scores, Σx denotes the sum of x scores, Σy is the sum of y scores, Σx^2 stands for the sum of squared x scores, while Σy^2 mark the sum of squared y scores. Using Eq. 1 the Pearson's correlation coefficient is computed that will give an insight into the correlation of the two considered water quality parameters. As a result, one can derive unbiased conclusion on which parameters to include or leave out of the water quality assessment.

Furthermore, due to the large number of the monitored quality parameters, the research was expanded by including the examination of Pearson's correlation between the measured parameters. The goal was to establish a list of seven to ten parameters that are most important and representative for the evaluation of the CCME WQI. The correlation results are given in Tables 1 and 2.

		T	pH	C	DO	COD	BOD_5	Chl-a
		˚C	-	µS/cm	mg/l	mg/l	mg/l	mg/m^3
T	˚C	1						
pH	-	0.620	1					
C	µS/cm	0.220	0.142	1				
DO	mg/l	-0.173	0.161	-0.782	1			
COD	mg/l	0.580	0.610	0.703	-0.453	1		
BOD_5	mg/l	0.364	0.352	0.557	-0.256	0.798	1	
Chl-a	mg/m^3	0.272	0.600	-0.100	0.306	0.441	0.314	1
SS	mg/l	0.455	0.724	0.120	0.107	0.481	0.358	0.692
TP	mg/l	0.298	0.573	0.164	0.084	0.453	0.369	0.199
PO_4	mg/l	-0.192	0.207	0.291	-0.121	0.120	0.014	-0.070
TN	mg/l	0.120	0.510	0.412	-0.037	0.676	0.571	0.517
TK	mg/l	0.283	0.619	0.341	-0.080	0.638	0.418	0.579
NO_2	mg/l	-0.546	-0.226	-0.228	0.365	-0.319	-0.115	-0.103
NO_3	mg/l	-0.409	-0.149	-0.126	0.323	-0.362	-0.200	-0.233
NH_4	mg/l	-0.429	-0.078	0.100	0.068	-0.064	0.086	-0.200

Table 1.Compouted values of the Pearson's correlation coefficients, part 1

The notation used in Tabs. 1 and 2 and on Figs. 1, 2, 3, 4, 5, and 6 are as follows: T marks the temperature of the water, pH is the pH value, C is the conductivity, DO stands for dissolved oxygen, COD is the chemical oxygen demand (bichromate), BOD_5 is the 5 day biochemical oxygen demand, Chl-a marks the content of chlorophill a, SS are the suspended sediments, TP denotes the total phosphorus in the water sample, PO_4 marks the content of orthophosphate, TN is the total nitrogen, while TK marks the total Kjeldalh nitrogen, NO_2 is the nitrite nitrogen, NO_3 the nitrate nitrogen and NH_4 is the ammonium nitrogen in the water.

By setting the limit for the reasonably strong correlation at r=0.5, the following positive correlations are attained: temperature and pH value where r=0.620, COD and Temperature with r=0.580, pH and COD with r=0.610, pH and Chl-a where r=0.600, pH and SS with r=0.724, pH and TP with r=0.573, TN and pH where r=0.510, TK and pH with r=0.619, COD and C where r=0.703, C and BOD_5 with r=0.557, COD and BOD_5 with r=0.798, BOD_5 and TN with r=0.571, SS and Chl-a where r=0.692, Chl-a and TN where r=0.517, Chl-a and TN where r=0.579, SS and TP with r=0.515, SS and TN with 0.614, SS and TK with r=0.625, PO_4 and NO_3 with r=0.604, PO_4 and NH_4 where r=0.506, TN and TK with r=0.8, NO_2 and NO_3 with r=0.735, NO_2 and NH_4 where r=0.701, and NO_3 and NH_4 with the correlation coefficient of r=0.580. These values are outlined in Tabs. 1 and 2.

		SS mg/l	TP mg/l	PO_4 mg/l	TN mg/l	TK mg/l	NO_2 mg/l	NO_3 mg/l	NH_4 mg/l
SS	mg/l	1							
TP	mg/l	0.515	1						
PO_4	mg/l	0.080	0.093	1					
TN	mg/l	0.614	0.477	0.229	1				
TK	mg/l	0.625	0.288	0.216	0.800	1			
NO_2	mg/l	-0.333	-.0326	0.409	0.131	0.157	1		
NO_3	mg/l	-0.261	-0.323	0.604	-0.058	-0.057	0.735	1	
NH_4	mg/l	-0.226	-0.108	0.506	0.119	0.224	0.701	0.580	1

Table 2.Compouted values of the Pearson's correlation coefficients, part 2

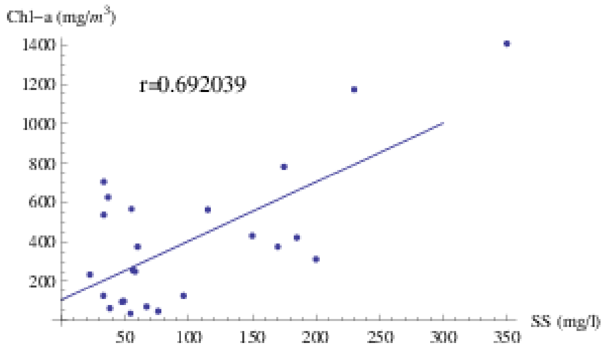

Figure 1. Graphical representation of data and Pearson's correlation of Chl-a and SS

According to these results, the following parameters should be engaged in evaluating the WQI: T, pH, C, COD, BOD_5, Chl-a, SS, TP, PO_4, TN, TK, NO_2, NO_3 and NH_4, which would result in the inclusion of 14 out of the 15 measured quality parameters. Instead, one should also take into consideration the reasons between the high correlation coefficients. For example, COD and BOD_5 both represent the oxygen demand of the system. Consequently including the both would only provide biased results. Instead, the parameter with higher correlation is chosen, in this case COD, while the BOD_5 can be omitted from the determination of WQI.

Similarly, TN and TK show the nitrogen content and shouldn't both be taken into consideration. Since TK gave a greater r coefficient, it should be inserted in further analysis, while TN can be disregarded. Furthermore, since TP only showed a correlation of r=0.573 with pH, while PO_4 had a fair correlation with both NO_3 and NH_4, we could consider excluding it from the WQI computation, although this decision would be left entirely to the researcher.

The graphical representation of some of the attained correlations are given on the following figures. Figure 1. shows the correlation between Chl-a and SS. The dots represent the paired values of the measurements, while the line displays the linear correlation between these data. This correlation is considered a very good correlation between these parameters.

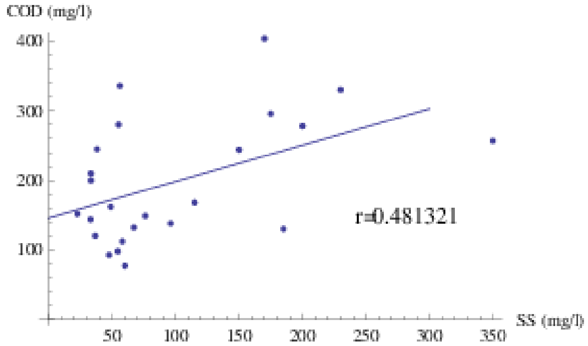

Figure 2. Graphical representation of data and Pearson's correlation of COD and SS

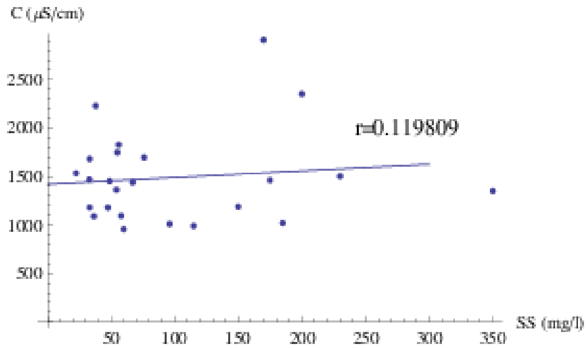

Figure 3. Graphical representation of data and Pearson's correlation of C and SS

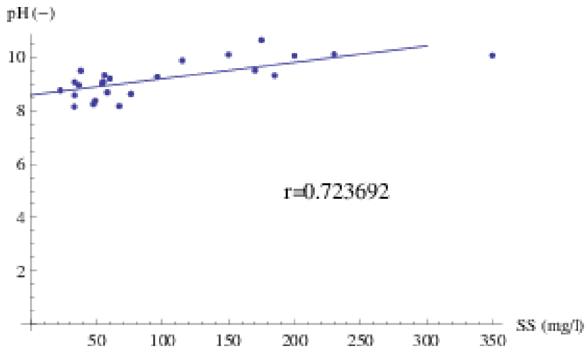

Figure 4. Graphical representation of data and Pearson's correlation of pH and SS

Figure 2. shows the measurements and correlation between COD and SS, with in this case labels a poor correlation of r=0.481.
The correlation and data for C and SS are given on Fig. 3 where it can be seen that these two parameters have a positive, but very poor correlation with the correlation coefficient of only r=0.1198.

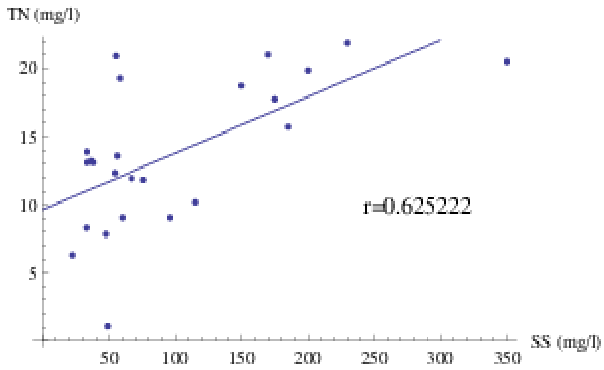

Figure 5. Graphical representation of data and Pearson's correlation of TN and SS

As previously established, the correlation between pH and SS is fairly good with the correlation coefficient of r=0.724, presented on Fig. 4.
Figure 5. displays the correlation between TN and SS where the Pearson's correlation coefficient is r=0.625, while Fig. 5. shows the correlation for TP and SS with r=0.515, also a good correlation between parameters.

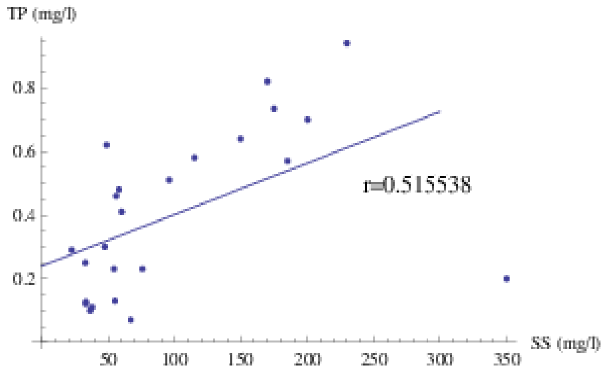

Figure 6. Graphical representation of data and Pearson's correlation of TP and SS

Using this approach, the final suggestion would be to include the following water quality parameters into the evaluation of the WQI: T, pH, C, COD, Chl-a, SS, TP, PO_4, TK, NO_2, NO_3 and NH_4.
Relatively strong negative correlations were attained between DO, NO_2, NO_3, NH_4 and various parameters. These negative correlations could be used as a reason to exclude these parameters from the upcoming WQI analysis, since they are an indicator of their influence on the overall water quality. In these cases, one could conduct further investigations in order to make a final decision. For example, computing the WQI with and without these data would be a one possibility to properly establish their influence on the overall water quality.

4. CONCLUSION

This paper considers the steps that should be taken before conducting a water quality analysis by computing the WQI. Namely, different WQI computation methods all rely on using different water quality parameters measured in situ or determined in the laboratory. Depending on the particular water body at hand, the influence of the same quality parameters on the system's quality can vary. That is why it is most important to conduct a thorough analysis by examining the correlation between these parameters.
This correlation is usually implemented by means of the Person's correlation. The researchers should combine all parameters among themselves and compute the Pearson's correlation coefficient that is a good indicator of the influence certain parameters have on each other, and the evaluated water body as well. The results of these computations are numbers between -1 and +1, where higher numbers indicate better accordance between two parameters.
The results of these investigations can also be displayed graphically, where we can see the measurements and the attained correlation for them.
The presented procedure was in this case implemented to evaluate the representative water quality parameters for the computation of the WQI on Lake Ludas using the measurements from 2012.

Acknowledgments
This work was funded by the Ministry of Education and Science of the Republic of Serbia, Project Number TR 37009.

REFERENCES

[1] Xu Z.X., Composite water quality identification index for environmental quality assessment of surface water, Journal of Tongji University (Natural science), **2005**, Vol. 4(33), pp. 482-488
[2] Lumb A., Sharma T. C., Bibeault J. F. and Klawunn P., A Comparative Study of USA and Canadian Water Quality Index Models, Water Quality, Exposure and Health, **2011**

[3] Bilgin A., Evaluation of surface water quality by using Canadian Council of Ministers of the Environment Water Quality Index (CCME WQI) method and discriminant analysis method: a case study Coruh River Basin, Environmental Monitoring and Assessment, **2018**, Vol. 190(554)
[4] de Almeida G. S. and de Oliveira I. B., Application of the index WQI-CCME with data aggregation per monitoring campaign and per section of the river: case study - Joanes River, Brazil, Environmental Monitoring and Assessment, **2018**, Vol. 190(195)
[5] CCME, Canadian Water Quality Guidelines for the Protection of Aquatic Life, CCME Water Quality index 1.0, User's Manual. CCME, Winnipeg, Manitoba, Canada, **2001**
[6] CCME, Synthesis of Research and Application of the CCME Water Quality Index, CCME Project No. 520-2012, Malahat, British Columbia, **2012**
[7] Majer F. and Horvat M., Validity and errors in hydraulic and suspended sediment measurements, Journal of Faculty of Civil Engineering Subotica, **2018**, Vol. 34, pp. 45-55.

ОДРЕЂИВАЊЕ МЕРОДАВНИХ ПАРАМЕТАРА ЗА ПРОРАЧУН CCME WQI

Резиме: *Овај рад приказује неопходне кораке за одређивање репрезентативних параметара приликом прорачуна CCME WQI индекса квалитета. Да би се избегао непоуздан приказ квалитета воде, пожељно је спровести претходну анализу расположивих параметара квалитета, како би се у даљу анализу укључили само они подаци који су заиста од значаја за опис квалитета површинске воде. Ово се обично спроводи прорачуно Пирсоновог коефицијента корелације за све параметре квалитета који стоје на располагању. На основу добијених резултата се поузданије могу донети одлуке о томе које параметре треба укључити у прорачун CCME WQI. Поступак је приказан на примеру језера Лудаш за 2012. годину.*

Кључне речи: *квалитет вода, CCME WQI, Пирсонов коефицијент корелације*