Statistical analysis of the water pollution indicators pertaining to treated municipal sewage introduced to the river

Dariusz Majerek[1], *Sylwia* Duda[2*], *Roman* Babko[3], and *Marcin K.* Widomski[2]

[1]Lublin University of Technology, Faculty of Fundamentals of Technology, Nadbystrzycka 38, 20-618 Lublin, Poland
[2]Lublin University of Technology, Faculty of Environmental Engineering, Nadbystrzycka 40B, 20-618 Lublin, Poland
[3]Schmalhausen Institute of Zoology, Department Fauna and Systematics of Invertebrates NAS of Ukraine, 01030 Kyiv, Ukraine

Abstract. The assessment of the ratio of self-purification processes in the natural environment, including the water bodies of streams and rivers is possible by the means of the numerical modelling. Nonetheless, the reliability of results of the qualitative numerical calculations may be affected by the quality of the collected input data and efficiency of the model calibration. The commonly required input data include pollutants characteristics and coefficients for empirical equations of their transport and decay, while calibration of a model requires series of measurements of selected pollutants concentration in water. This paper presents studies of concentration of pollutants measured in the treated sanitary wastewater discharged to the Bystrzyca river from municipal wastewater treatment plant, and their concentrations measured directly in the river, above and below the location of discharge point. The performed qualitative measurements of treated sewage discharged from municipal wastewater treatment plant and water of the Bystrzyca river included determination of COD, BOD_5, $N-NH_4$, $N-NO_2$, $N-NNO_3$, TKN, TN, TP and TSS concentration. The presented results of measurements are the first step in development and calibration of the model, allowing to successfully predict the influence of discharged sanitary wastewater on quality of water in the river.

1 Introduction

Discharging wastes – both solid and liquid – to waters was a very convenient solution. Wastewater discharged to rivers via gutters were diluted and underwent self-cleaning, while the river, in certain distance downstream from the waste discharge site, regained its quality. However, at present it is not possible, because, e.g. due to the industrial development, new chemical compounds are being synthesised, which enter the water cycle in the natural environment and may have a negative influence on it as well as on the human health [1-2]. Therefore, due to constantly deteriorating quality of available water resources, only the treated wastewater which meets the requirements stated in valid legal regulations can be discharged to running waters, i.e. rivers [1-4].

Monitoring of receiver's water quality enables to determine the influence of introducing treated wastewater to the environment, as well as the efficiency of self-cleaning process. Due to the fact that rivers may transport pollutants over great distances, control and running water quality assessment, as well as taking the actions aimed at improving their quality, are important. Even a highly-treated medium may carry significant amounts of biogenic and mineral compounds, which may cause unfavourable processes, i.e. eutrophication or increase in water salinity. The treated wastewater, in the case of treatment plant failure, may release, i.a. heavy metals, persistent organic

pollutants, potential carcinogens, as well as bacteria, viruses and parasites into the receiver [5-6]. As a result, numerous publications appeared in recent years, aimed at developing the river condition assessment systems [3, 7-16].

The quality of running waters has been evaluated for a long time on the basis of physicochemical analyses. The main parameters determining the quality of wastewater and receiver waters include BOD_5 (5-day Biochemical Oxygen Demand), COD (Chemical Oxygen Demand), pH, TSS_{tot} (Total Suspended Solids), as well as the content of nitrogen and phosphorus compounds [5-8]. It is also known that a complete assessment of river condition is possible only after introducing the aspects of biological assessment [8-10]; similarly, as in the case of the treated wastewater quality assessment and wastewater treatment plant efficiency.

At present, there are numerous methods for evaluating the condition of river environment, ranging from the organoleptic evaluation of water quality, to highly precise physicochemical methods and biotesting. However, as far as the quality assessment of aquatic ecosystems is concerned, the environmental indices systems are increasingly preferred. They enable to evaluate the quality of aquatic environment as a habitat for organisms – various species of protozoa are commonly used for this purpose [8, 12-16].

Obviously, the best results of river condition control can be achieved by simultaneously applying all three

* Corresponding author: sylwia.m.duda@gmail.com

groups of assessment criteria (hydro-morphological, physicochemical and biological), unfortunately, such an approach is relatively rare [8]. Moreover, studies are usually conducted in terms of presenting new method of evaluation [3,8]. Much more rarely they are the works aimed at determining the sensitivity of these methods and the degree of agreement and complementarity in the final assessment of river conditions [17-20].

In this work, a slightly different approach, based on numerical modelling of various phenomena occurring inside the water body of streams and rivers, allowing to assess the ratio of self-purification processes in the natural environment, was presented. However, reliability of qualitative numerical calculations results depends directly to the gathered input data to modelling and successful calibration of the developed model. Necessary input data usually cover pollutants characteristics and coefficients for empirical equations, while measurements of selected pollutants are required to model calibration.

This work presents the statistical analysis of pollutant concentration in the treated wastewater discharged to the Bystrzyca river from Hajdów municipal wastewater treatment plant, as well as the concentrations in the Bystrzyca river water, above and below the location of discharge point. Numerical modelling of various phenomena occurring in the body of water streams and rivers was used to assess the ratio of self-cleaning processes in the natural environment.

2 Materials and methods

2.1 Materials

The samples were collected from the Bystrzyca river, where water quality changes caused by treated effluent from the wastewater treatment plant could have occurred. The collection was carried out in three locations: 1 – spot above the discharge point; 2 – spot of wastewater discharge to the river, 3 – spot below the wastewater discharge point. In order to ensure appropriate quality of results, the sampling locations were in a certain distance from one another. The treated effluent was subjected to the analysis as well. Samples of water from the river were collected 500 m before the sewage discharge point, at the point of discharge and 1000 m after the discharge point. The collection was carried out using a bucket fastened on a cord from a depth of about 50 cm below the water table to properly prepared containers in accordance with the standard described in the Water and Wastewater Sampling Guidelines for WIOŚ studies in Warsaw [21].

Qualitative measurements of sanitary wastewater discharged from municipal wastewater treatment plant and waters of the Bystrzyca river conducted in our research covered i.a. COD, BOD_5, $N-NH_4$ (ammonium nitrogen), $N-NO_2$ (nitrite nitrogen), $N-NNO_3$ (nitrate nitrogen), TKN, TN (total nitrogen), TP, TSS.

The samples of the wastewater discharge pollutants characteristics took place once for a week for three years, but the data from spot above and below the municipality wastewater treatment plant were conducted only once for a month and only in 2009. All the dates of measurement from less frequent period occurred in set of dates of more frequent measurements. The samples were transported to the laboratory in the Faculty of Environmental Engineering LUT and the qualitative analysis of pollutant indicators was performed. The HACH DR2800 spectrophotometer and special research methods authorized by HACH-Lange company were used to analyse value of the nitrogen (cuvette tests LCK 238 LATON, LCK 340, LCK 341, LCK 303) and phosphorus (LCK 348) compounds, as well as COD (LCK 314) and TSS. On the other hand, the BOD_5 analyses were conducted according to the Polish standards (PN-EN 1899-1:2002).

2.2 Methods

The methods used in this paper were divided into two groups. In the first, the analysis of individual pollutant indicator was conducted, in the sense of the impact of the pollutant concentration level on the water quality in the river below discharge point of municipal wastewater treatment plant. The relationship was estimated by a linear regression

$$C\_below = \beta\_0 + \beta\_1\, C\_above + \beta\_2\, C\_out \qquad (1)$$

where C_above, C_below, C_out mean the level of concentration of particular pollutant measured respectively in the river, above and below the location of discharge point and in the treated sanitary wastewater [22].

The second method was used to judge the influence of the combined pollutant factors on the quality of water in the Bystrzyca river. A hierarchical cluster analysis was conducted with Euclidean metrics and Ward's method of agglomeration in order to compare the groups of observation in view of all nine pollutants [2, 23-30].

All calculations were conducted in statistical programming language R [31]. Cluster analysis was made using R factoextra package [25].

3 Results and discussion

Figure 1 shows a strong relationship between the concentrations of COD measured above and below the Bystrzyca river, but there is much smaller correlation between concentration of COD at the discharge point and in below the river. Most of the pollutants reveal a similar behaviour, but not all, e.g. TN has strong correlation among all three measurements (see Figure 2). The measurement of the size of the effect of the discharge of sewage to the river is shown in the form of the multiple regression parameter (C_out).

All models consider the impact of the condition of the river above the discharge and the level of concentration of the tested compound on its pollution afterwards. If C_out is significant, then it can be inferred that the level of concentration of the pollutant is crucial. Only in two cases the C_out is significant, so these two factors are substantial for assessing C_below. Almost all of the models show that the level of concentration of pollutant depends on C_above (see Table 1).

Table 1. Statistical parameters describing dependences between contamination indices in analysed points.

C_{after}	BOD$_5$ [mgO$_2$ /dm^3]	COD [mgO$_2$ /dm^3]	TSS [mg TSS/dm^3]	N-NH$_4$ [mg/l N-NH$_4$]	N-NO$_2$ [mg/l N-NO$_2$]	N-NO$_3$ [mg/l N-NO$_3$]	TKN [mg/l TKN]	TN [mg/l TN$_b$]	TP [mg/l PO$_4{}^{3-}$]
C_{before}	0.761*	0.849***	1.085	1.872**	3.667**	0.945***	0.787***	0.952***	1.940***
	(0.360)	(0.179)	(0.607)	(0.662)	(1.394)	(0.196)	(0.192)	(0.206)	(0.417)
C_{out}	-0.396	0.168	-1.235	0.041	0.294***	0.073	0.117	0.186***	-0.096
	(0.258)	(0.109)	(1.046)	(0.065)	(0.046)	(0.050)	(0.068)	(0.041)	(0.140)
Constant	3.215	-2.059	19.878	-0.076	-0.110*	0.591	0.261	-0.278	-0.082
	(2.515)	(7.384)	(20.552)	(0.142)	(0.053)	(0.349)	(0.455)	(0.582)	(0.113)
Observations	12	12	12	12	12	12	12	12	12
R^2	0.574	0.773	0.374	0.717	0.837	0.797	0.657	0.885	0.794
Residual Std. Error (df = 9)	1.334	4.960	18.067	0.161	0.036	0.317	0.262	0.311	0.046
F (df = 2; 9)	6.062**	15.340***	2.690	11.379***	23.034***	17.673***	8.632***	34.569***	17.354***

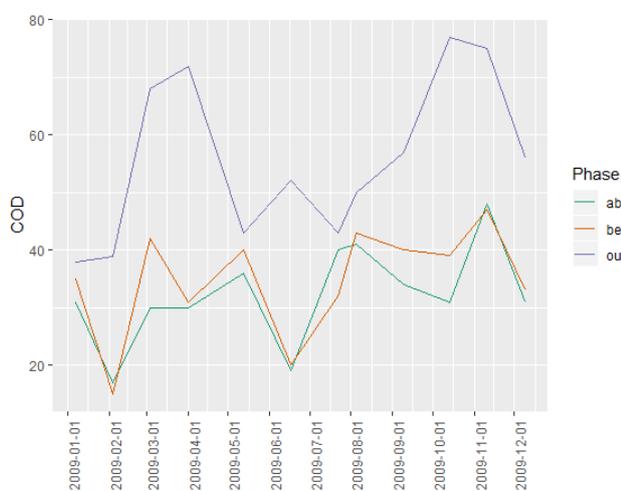*Note:* *p<0.1; **p<0.05; ***p<0.01
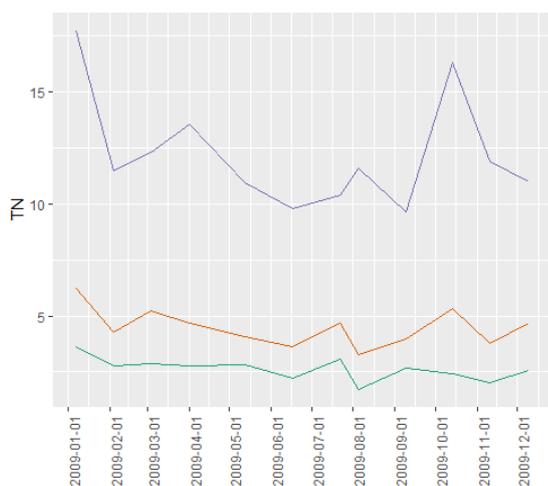

Fig. 1. COD level in water of analysed point.

Fig.2. Total Nitrogen (TN) level in water of analysed point.

The second approach was to evaluate the relationship between C_above, C_below, C_out, which relies on clustering observations into homogenous groups in the sense of concentration of pollutants. On the basis of the hierarchical cluster analysis with Ward's agglomeration method, all observations can be divided into two groups. The division obtained from the analysis is obvious (see Figure 3). The points marked in red are conducted from municipal sewage and blue ones are from the Bystrzyca river. If the impact of pollutants were significant, there would be one more cluster (observations from the river would be divided into two groups – more and less contaminated). The representation of individuals in 2-dimmensional space is quite good (79% of total variance explained), so the map on the right in Figure 4 shows separation between two groups in a good way.
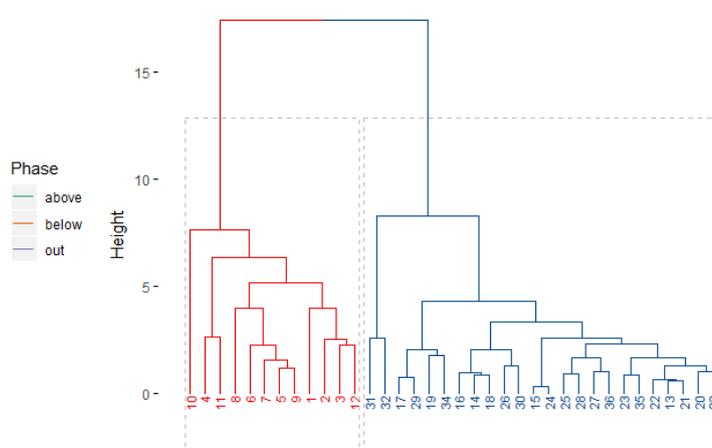

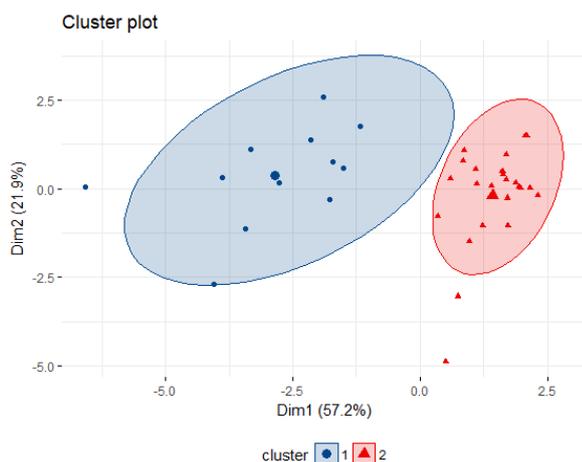**Fig. 3.** Dendrogram of cluster analysis pertained to pollution level in analysed points.

3

Fig. 4. 2-dimmensional PCA projection of individuals.

4 Conclusions

Unsupervised statistical learning method, such as cluster analysis, gives us two-class division based on all measured water pollution indicators. The result was conducted by hierarchical cluster analysis with Ward's agglomeration method and Euclidean metric. All observations in cluster 1 were conducted from municipal sewage. The samples from the Bystrzyca river, both above and below sewage treatment plant are in cluster 2. There is no significant difference between two parts of the river.

The relationships between municipal sewage and the Bystrzyca river below the point of discharge are different than the dependence between the Bystrzyca river above and below the wastewater plant.

The analysis pertaining to the impact of municipal wastewater treatment plant discharge on the Bystrzyca river generally revealed that there is no such effect. Both the analysis of the impact of individual factors and their groups did not give rise to the conclusion that the sewage treatment plant has a negative effect on the river.

References

1. Z.W. Kundzewicz, M., Zalewski, A. Kędziora, E. Pierzgalski, Nauka **4** (2010) (in Polish)

2. J. Kujawska, W. Cel, ITM Web Of Conferences **15**, 1-5 (2017)

3. V. Pliashechnyk, Y. Danko, G. Łagód, J. Drewnowski, T. Kuzmina, R. Babko, E3S Web of Conferences **30** (2018)

4. Regulation of the Minister of the Environment of November 18, 2014 (in Polish)

5. M. Strycharczuk, P. Szostak, A., Kozak, G. Łagód, Monografie Inżynieria Środowiska - Młodym Okiem (2012) (in Polish).

6. P. Szostak, M. Strycharczuk, A. Kozak, G. Łagód, Monografie Inżynieria Środowiska - Młodym Okiem (2012) (in Polish)

7. T. Srebotnjak, G. Carr, A. de Sherbinin, C. Rickwood, Ecol Indic. **17** (2012)

8. M. Gorzel, R. Kornijów, E. Buczyńska, Ecol Chem Eng S. **25** (1) (2018)

9. M. D. Turley, G. S. Bilotta, C. A. Extence, R. E. Brazier, Freshwater Biol. **59** (11) (2014)

10. S. J. Nichols, L. A. Barmuta, B. C. Chessman, P. E. Davies, F. J. Dyer, E. T. Harrison, et al., Mar Freshwater Res. **68 (4)** (2017)

11. A. Bieganowski, G. Łagód, M. Ryżak, A. Montusiewicz, M. Chomczyńska, A. Sochan A., Ecol Chem Eng S. **19** (4) (2012)

12. A.Y. Hoekstra, A. K. Chapagain, P. R. van Oel, Water-Sui. **9** (6) (2017)

13. K.S. Szoszkiewicz, I. Jusik, I. Lewin. J. Czernia-wska-Kusza, M. Kupiec, M. Szostak, Hydrobiologia **808** (1) (2018) (in Polish).

14. K. Jaromin, R. Babko, G. Łagód, Proceedings of ECOpole **4** (2) (2010)

15. R. Babko, T. Kuzmina, Z. Suchorab, M.K. Widomski, M. Franus, Ecol Chem Eng S. **23** (3) (2016)

16. G. Łagód, M. Chomczyńska, A. Montusiewicz, J. Malicki, A. Bieganowski, Ecol Chem Eng S **16** (3) (2009)

17. M. Gorzel, R. Kornijow, Ecohydrol. Hydrobiol. **7** (1) (2007)

18. S. Serrano, L. Arregui, B. Perez-Uz, P. Calvo, A. Guinea, *Guidelines for the identification of ciliates in wastewater treatment plants* (IWA Publishing, London, 2008)

19. W. Foissner, H. Berger, Freshwater Biol. **35**, 313-330 (1996)

20. M.A. Lewis, Ohio J. Sci. **86** (1), 2-8 (1986)

21. WIOŚ w Warszawie Laboratorium, PO-011-04 **12** (2014) (in Polish).

22. G. A.F. Seber, A.J. Lee, *Linear regression analysis*, (Wiley, New Jersey, USA 2012)

23. C. Fraley, A.E. Rafter, J Am Stat Assoc. **97** (458), 611–631 (2002)

24. G. James, *An introduction to statistical learning: with applications in R* (Springer 2015)

25. A. Kassambara, F. Mundt, Extract and visualize the results of multivariate data analyses url: https://CRAN.R-project.org/package=factoextra

26. L. Kaufman, P.J. Rousseeuw, *Finding groups in data: an introduction to cluster analysis* (Wiley, New Jersey, USA 2009)

27. M. Khun, K. Johnson, *Applied Predictive Modelling*, (Springer, 2013)

28. D.C. Montgomery, G.C. Runger, Applied statistics and probability for engineers (Wiley, New Jersey, USA 2011)

29. J. Kujawska, M. Pawłowska, W. Cel, A. Pawłowski, Desalin Water Treat. **3** (57), 1409-1419 (2016)

30. W. Cel, J. Kujawska, A. Czechowska-Kosacka, Rocz. Och. Srod. **2** (17), 1224-1238 (2015)

31. http://www.R-project.org/. R Core Team, A language and environment for statistical computing (2018).