DEVELOPMENT OF WATER QUALITY MATRIX THROUGH SURROGATE MODELING

Laura Kusari

Department of Hydrotechnics, Faculty of Civil Engineering and Architecture,
University of Prishtina "Hasan Prishtina", Prishtina, Kosovo

Corresponding author E-mail: laura.kusari@uni-pr.edu

Abstract

This paper presents the outcome of a research project that was focused on the monitoring of surface water quality through the development of a correlation matrix. The matrix was developed for six main water quality parameters by the use of surrogate relations. The grab sampling was performed at selected sites and the same samples were used in the laboratory for the preparation of subsamples. Those subsamples were examined for Turbidity, Total Suspended Solids (TSS), Chemical Oxygen Demand (COD), Biological Oxygen Demand (BOD), Total Organic Carbon (TOC) and Nitrates. Then, data were analyzed by statistical analyses, using linear regression and the outcome was used for the development of a correlation matrix of main water quality parameters. The analyses revealed that in this study site, TSS has high positive correlation with BOD, COD and NO_3 as well as with turbidity. The highest positive correlation was noticed between turbidity and BOD, NO_3, TSS and COD. On the other hand, only Total Organic Carbon (TOC) was negatively (inversely) correlated with the studied parameters. The correlation matrix developed will help in determining the water quality status by using few parameters and developing water quality and pollution control programs.

Keywords: *water quality, monitoring matrix, river pollution, linear regression, surrogate relationships*

Introduction

The rivers are impaired worldwide and in particular in our new country. The land use changes from rural to urban ones as well as many anthropogenic activities are further contributing to river pollution. Rivers are carrying off municipal and industrial waste waters, as well as runoff from farm land and are one of the most endangered water bodies to pollutants (Sing et al., 2004; Wang et al., 2007). Also, many natural processes such as precipitations and erosion also degrade the surface water quality. In order to design and implement river restoration plans and effective pollution control measures, as a part of watershed management, the monitoring and evaluation of surface water quality is very important. The surface water quality is defined based on the results of physical, chemical and biological

DOI: 10.6092/issn.2281-4485/7735

parameters. Those parameters indicate the pollution load and mostly different water samples will indicate different level of pollution loads, depending on the tested parameters. Unfortunately, until now cost effective and robust methods for the continuous measurement of pollutant concentrations are not yet fully developed (Mingutana et al., 2010). Therefore, over the recent years, the problem of river contamination has led to the need to have evaluation tools that are able to predict the fate of pollutants, either accidentally or intentionally introduced into streams (Boano et al., 2005).

The actual monitoring program is consisted of frequent water sampling at many sites and the laboratory analyses of a large number of samples, for many water quality parameters. This is quite long and ineffective process, since the sampling and the analyses very often miss the precipitation events or a pollution discharge into river. Usually, monitoring of surface waters is within time frames that are previously determined and many times it misses event discharges. Also, investigation of a large number of water quality parameters is time consuming and resource intensive (Mingutana et al., 2010; Kayhanian et al., 2007; Thomson et al., 1997). Therefore, the outcome and the reliability of the monitoring process are very likely to be effected by many factors. As a result, in the last years the need for the connection between main water quality parameters has increased.

As noted by Bertrand – Krajevski (2007), many models are parameterized. Consequently, model calibration and validation and consequently, models outputs significantly depend on the data set used (Mingutana et al., 2010). Since the human, technological and financial resources are often very limited then a more pragmatic approach is used in this research. The key water quality parameters are chosen and the possibility of using surrogate parameters is investigated. Without having to carry out many laboratory examinations these relationships between water quality parameters and their surrogate ones would enable the monitoring of surface water quality. According to previous studies (Settle et al. 2007), these relationships have the potential to enhance rapid generation of vital information from site-based measurements and to reduce the requirements for laboratory-based examinations of indicator concentrations in urban waters. The adoption of a limited number of easy to measure parameters will enable greater quality control in data collection (Mingutana et al., 2010). Furthermore, it will facilitate model application and calibration by contributing to overcoming constraints identified in research literature in relation to water quality models (Lindblom et al., 2007; Wagener and Gupta, 2005; Bertrand- Krajevski et al., 1993). Therefore, the aim of this research is to better understand changes and correlation between water quality parameters through the development of water quality matrix, based on previously prepared subsamples. Surrogate modeling, which is a second level of abstraction, is concerned with developing and utilizing cheaper-to-run surrogates of the original simulation models (Razavi et al., 2012).

Materials and methods

Sitnica Catchment is continuously deteriorated due to anthropogenic activities taking place in it (Fig. 1). This region is facing rapid development, making the Sitnica River pollution a growing concern, which is the main reason that this River is chosen for the research paper.

Figure 1
Map of Kosovo – Sitnica River Basin

The Sitnica Catchment lies on the west – central part of Kosovo and is characterised with medium-continental, with some impact of Aegean- Adriatic climate. The average annual temperature is (10.2-10.4)ºC, minimal temperature is -26 ºC while the maximum temperature reaches up to 37.4ºC (according to Hydro Meteorological Institute of Kosovo). The Sitnica Catchment covers an area of 2931.71 km^2 with the catchment's average slope 4.4 %. The main river is Sitnica, 167 km length and with 13.62 m^3/s average annual flows. Since it has a relatively small longitudinal slope of 0.054%, Sitnica meanders a lot. The minimal and maximal values, as well as average annual flow for Sitnica River, measured at hydrometric station, are as following (Table 1).

Table 1
Annual values of Sitnica River flow rates (m^3/s)

STATION	RIVER	Q_{min} (m^3/s)	Q_{avg} (m^3/s)	Q_{max} (m^3/s)
NEDAKOVC	SITNICA	0.50	13.62	328.0

Sitnica River while flowing through its catchment is subject to many pollution sources. In urban reaches of Sitnica River, the untreated domestic and industrial waste waters are discharged directly into river. On the other hand, agriculture is a non point source contributor to those surface waters as well as direct dumping of solid waste in the vicinity of the river, erosion, mining activities as well as leachate from waste disposal areas. The sampling location within Sitnica River as shown in the Figure 2 is river reach near the Vragoli location, with the coordinates 42.609138° latitude and 21.061295° longitude. Grab sampling was performed to

DOI: 10.6092/issn.2281-4485/7735

obtain needed water samples as it is the method implemented by many water quality monitoring agencies. Water samples from the Sitnica River reach were collected using sampling rod at two cross sections. At each cross section, three sampling verticals were set. The first one near the right river bank, the second one on the middle of the river and the third one near the left river bank. On the each sampling vertical, at each cross section there was only one sampling vertical, approximately 30 cm below the water surface.

Figure 2
Sampling location – Sitnica River reach

It is known that continuous rapid monitoring of water quality variables in rivers is needed to characterise environmentally significant events (Fauvel et al., 2016; Chappell et al., 2017). But, very often the frequency of water sampling and the concentrations of many water quality parameters do not coincide with each other. Therefore, due to the financial constrictions instead of frequent sampling to obtain the needed data, the laboratory subsamples were prepared for this project (Kusari, 2017). From each sampling site, the set of 1000 ml water samples were taken. The process of sub sampling was carried out in a laboratory, based on the procedure described by Earhart, 1984. This procedure was initially designed to ensure compliance of the effluent from dike confined disposal facilities with the total suspended solids violence standards (Kusari, 2017). For this project, two 1000 ml samples of water were taken at the upstream and downstream river sites, while the second one would serve as a dilution for the first water sample. In the laboratory, the sub samplers of 120 ml volume were used for the mixing of water samples and the laboratory examination of the same. From the first 1000 ml water sample, from the upstream site, 100 ml of water were extracted and poured into 120 ml sub sampler. From the second water sample, from the downstream location, 100 ml of water were extracted and added into the first water sample. By this, the first water sample would have again 1000 ml volume, but different concentration level of water quality parameters. This process was continuous, the 100 ml of water extracted from the first water sample, poured into the next sub sampler, and another 100 ml extracted from the second water sample and poured into the first one. The first water sample was diluted continuously and at the same time the concentration of various parameters in sub samples also changed continuously. Those 10 prepared subsamples were representative for the examination of the main water quality parameters, for a chosen sampling site (Kusari, 2017). The prepared subsamples were sent to the Hydrometeorology Institute of Kosovo and examined for Turbidity, Total Suspended Solids (TSS), Chemical Oxygen Demand (COD), Biological Oxygen Demand (BOD) and Nitrate (NO_3). The turbidity level was examined by a portable nephelometer (Hach 2100N Turbid meter) and the turbidity units were reported in Nephelometric Turbidity Units (NTU). This represents a measurement of the light intensity being scattered, when light is transmitted through a water sample. The procedure was repeated for all water subsamples. On the other hand, the concentration of Total Suspended Solids (TSS) was examined with the filtration method, by filtering samples water volume through a membrane filter and weighing the dried residue. The BOD was measured with the dilution method with the dissolved oxygen concentrations measured before and after the incubation period. The TOC was measured using the high temperature combustion method where TOC values were obtained by difference after the removal of inorganic carbon by acidification and the measurement of total carbon. In general, all the subsamples were examined using methods specified by APHA (2005).

The application of both univariate and multivariate statistical data analyses techniques has become a valuable tool in water quality research studies (Goonetilleke et al., 2009). Correlation is a method used to evaluate the degree of interrelation and association between two variables. Correlation coefficients measure the strength of association between two variables (Helsel and Hirsch, 2002). In this study, relationships between variables were developed in the form of linear regression. So, the next step, following the results of laboratory examinations of the above mentioned water quality parameters, was the determination of the correlation between those measured parameters, using linear regression analyses. The correlation between two variables can be explained using statistical indicators such of coefficient of determination (R^2). The coefficient of determination (R^2) is the fraction of variability in the response variable (Y) that is explained by the variability in the predictor variable (X). R^2 can have values from 0 (when there is no variation explained) to 1 (where all the variation is explained). Therefore, to summarize, the good predictive relationship is indicated by high R^2. The scatter plot gives a good estimation of the relationship developed. According to the literature research, the regression line in a scatter plot is often included with ±1 standard error or ±95% confidence limit (Mingutana et.al., 2010). How well the equation describes the data, is expressed through a correlation coefficient R^2 and the closer R^2 is to 1.00 the better the correlation. So an R-value of 1 shows complete correlation and an R value of 0 shows no correlation at all. A negative R-value indicates an inverse correlation so an R value of -1 shows complete inverse correlation. The correlation developed in this research were between turbidity and TSS, COD, BOD, TOC, NO_3; correlation between TSS and COD, BOD, TOC and NO_3; correlation between COD and BOD, TOC, NO_3; correlation between BOD and TOC, NO_3; and finally the correlation between TOC and NO_3. From the linear

DOI: 10.6092/issn.2281-4485/7735

regression analyses, the values of correlation coefficients, for investigated water quality parameters were used for the development of water quality matrix. The correlation coefficients indicate the relationship between investigated parameters and the possibility of using the surrogate ones.

Results and discussions

Estimating the regression relationships between water quality parameters has prove to be an effective approach to amend the deficiency in water quality observations (Xiaoying, et al., 2017). To evaluate the correlations between water quality parameters this research has undergone several phases. Firstly the grab sampling was performed and those water samples were used for the preparation of sub samples. The subsamples were prepared in order to avoid high frequency monitoring, due to the financial restrictions. These subsamples were then examined for 6 main water quality parameters. The obtained results, with the use of linear regression analyses, with Excel's ANOVA analyses tool pack, enabled the determination of possible correlation between such parameters. As the result of these correlations, the coefficients of determination R^2, for the studied water quality parameters are summarized in table 3. This table, consisting of the determination coefficients of main water quality parameters, represents the Correlation Matrix for our study site.

Table 3 *Correlation matrix for six main water quality parameters*

Variable	TTN	TSS	COD	BOD	TOC	NO_3
TTN	1.000					
TSS	0.8687	1.000				
COD	0.7171	0.7958	1.000			
BOD	0.9385	0.8974	0.7679	1.000		
TOC	-0.2095	-0.3413	0.0616	-0.2712	1.000	
NO_3	0.9158	0.8471	0.8673	0.9127	-0.0985	1.000

This matrix will provide a helpful tool when selecting a parameter which can serve as a proxy to another one. Furthermore, from this matrix, we can evaluate how strong is the correlation between two parameters. Many other researchers (Settle et al., 2007; Han et al., 2006) have identified both turbidity and total organic carbon (TOC) as the parameters with the highest potential to act as a surrogate one. As for the turbidity, the correlation between it and total phosphorus concentrations was in the range from (0.78-0.90) in a study conducted in Iowa Rivers (Schilling et al., 2017). Through another study, at Paradise site, the strong correlation between turbidity and total suspended solids TSS (0.95) was investigated. Somehow a weaker correlation (0.84) between the same parameters was noticed in the Mendon site, within the same river (Jones et al., 2011). There are numerous studies from the literature that have documented turbidity to have closer correlation with TSS, with R^2 values in the range (0.46-0.98) (Lannergard, 2016). The correlation between turbidity and total suspended solids in the effluent, shows a strong correlation also, with the determination coefficient R^2=0.979. Many researchers monitoring final effluent had proven that turbidity can serve as a surrogate for an increase of BOD and COD, which requires more complex measurements. Since both analyses, for BOD and COD can be time consuming, insensitive and imprecise, the study conducted by Lee et al. (2016) found that TOC can be considered as an alternative one. This study shows a high correlation between BOD and TOC (R^2=0.75) and a strong correlation between COD and TOC (R^2=0.87), for rivers. As for the correlation between TSS and chemical oxygen demand (COD), a strong correlation (R^2=0.860) is documented (Mucha et al., 2016). Consistent with above mentioned researchers, in this study, turbidity was the one parameter that showed higher correlation with other ones. The highest positive correlation (R^2 value closer to 1) was noticed between turbidity and BOD (0.9385). Turbidity is highly related to NO_3 (0.9158), to TSS (0.8687) and lastly to COD (0.7171). If compared to the literature, the correlation of turbidity to TSS, with the coefficient of determination R^2=0.8687 is within range (0.46-0.98) set by Lannergard, 2016. As it can be seen from the matrix, in our study the turbidity showed rather high positive correlation with all studied parameters, except for the very poor correlation to TOC (-0.2095). As for TSS, this parameter showed a very high positive correlation with BOD (0.8974), COD (0.7958) and NO_3 (0.8471) as well as with turbidity (0.8687). The only negative correlation of TSS was to TOC (-0.3413). The TSS correlation with COD in our study is consistent with other studies, with the determination coefficient 0.7958 similar to 0.860 documented in the study by Mucha et al., 2016. As for the correlation between TSS and NH_3, the determination coefficient in or study (0.8471) shows a very good correlation and is in agreement with other studies as well. The BOD correlation to COD (for this research) with the determination coefficient R^2=0.7679 is very high and is in the agreement with other studies where the correlation between BOD and COD ranged (0.709-0.872) for 3 rivers in Malaysia (Lee et al., 2015). These results are in agreement with several studies that have identified the potential of some parameters to act as surrogate to other water quality parameters (Han et al., 2006; Settle et al., 2007). On the other hand, Total Organic Carbon (TOC) is negatively (inversely) correlated with the studied parameters, except for COD, where it was noticed very poor positive correlation 0.0616. This is the only parameter that showed no correlation or a very poor one with all other water quality parameters and is not consistent with mentioned researchers.

Conclusions

Being the new developing country we are facing the water pollution spread all over our resources. At the same time, water quality is monitored only on previously set up schedule plan and often misses the pollution events. The traditional monitoring

DOI: 10.6092/issn.2281-4485/7735

is even more ineffective in the low order streams as they are subject to anthropogenic disturbances. Therefore, in this paper we have presented a method for using correlations between water quality constituents, in order to predict or estimate the possible changes of water quality. To avoid frequent grab sampling and many uncertainties associated with it, the sub sampling was performed in the laboratory and those subsamples were examined. The final results of the examinations provided the necessary data, which by the use of linear regression enabled the development of relationships between water quality constituents. The regression analyses results were given in the form of a matrix, comprising of the coefficients of determination for various relationships. This approach will amend the deficiency of the water quality monitoring programs and help in planning or designing water quality control measures.

Regression results have shown that there are some positive correlations between water quality constituents. From the matrix table, the highest coefficient of determination is when turbidity is related to BOD (0.9385), NO_3 (0.9158), TSS (0.8687) and COD (0.7171). To summarize, the turbidity has the positive relationship with all the measured parameters except for the TOC (-0.2095). There are some strong correlations too, between the TSS and BOD (0.8974), COD (0.7958) and NO_3 (0.8471). Positive correlation (0.7679) is noticed between BOD and COD, too. Those results obtained in this paper are in a close agreement with those mentioned in the literature. The only case where there is no correlation is when TOC is correlated to other water quality constituents. To summarize, based on the results of this study, turbidity and TSS are the most suitable parameters that can be used as a surrogate one, for other water quality constituents investigated in this case.

This study demonstrated that linear regression analyses could potentially serve as a tool to evaluate water quality and based on an easiest to measure parameter, predict the changes of another water quality constituent. It would enable a shorter list of water quality parameters to be measured. At the same time, the correlation matrix can be used when predicting the pollutants entering the stream, through easy measurements of the surrogate parameters. The surrogate parameters would serve as indicators for the potential pollution entering the stream. This would lead to the development of management strategies to minimize the water quality degradation due to the continuous urbanization of this area. In the future, more efforts should be taken to validate and improve those correlations to gain the best possible results from the same. The use of this correlation matrix in the development of water quality and pollution control programs is immense.

References

APHA (2005). Standard Methods or the Examination of Water and Waste Water. Washington DC. American Public Health Association.

BERTRAND-KRAJEWSKI J. L. (2007) Stormwater Pollutant Loads Modelling: Epistemological Aspects and Case Studies on the Influence of Field Data Sets Calibration and Verification. Water Science and Technology, 55(4):1-17. Doi: 10.2166/wst.2007.090

BERTRAND-KRAJEWSKI J. L., BRIAT P., SCRIVENER O. (1993) Sewer sediment production and transport modelling: a literature review. Journal of Hydraulic Research, 31(4):435-460. DOI: 10.1080/00221689309498869

BOANO F., REVELLI R., RIDOLFI L. (2005) Source identification in river pollution problems: a geostatical approach. Water Resources Research, 41(7). Doi**:** 10.1029/2004 WR003754

CHAPPELL N., JONES T., TYCH W. (2017) Sampling frequency for water quality variables in streams: systems analysis to quantify minimum monitoring rates. Water Research, 123:49-57. Doi: 10.1016/j.watres.2017.06.047

EARHART H. G. (1984) Monitoring total suspended solids by using nephelometry. Environmental Management, 8(1):81-86. Doi**:** 10.1007/BF01867876

FAUVEL H., CAUCHIE H. M., GANTZER C., OGORZALY L. (2016) Contribution of hydrological data to the understanding of the spatio-temporal dynamics of F-specific bacteriophages in river water during rainfall-runoff events. Water Resources, 94:328-340. Doi: 10.1016/j.watres.2016.02.057

GOONETLLIKE A., EGODAWWATTA P., KITCHEN B. (2009) Evaluation of pollutant bulid up and wash off from selected land uses at the Port of Brisbane, Australia. Marine Pollution Bulletin, 58:213-221. Doi:10.1016/j.marpolbul.2008.09.025

JONES A. S., STEVENS D., HORSBURGH J., MESNER N. (2011) Surrogate measures for providing high frequency estimates of total suspended solids and total phosphorus concentrations. Journal of the American Water Resources Association (JAWRA), 47(2): 239-253. Doi: 10.1111/j.1752-1688.2010.00505.x

HAN Y., LAU S., KAYANIAN M., STENSTORM M. (2006) Characteristics of highway stormwater runoff. Water Environment Research, 78(12):2377-2388. Retrieved from http://www.jstor.org/stable/25053644

HELSEL D., HIRSCH R. (2002) Statistical methods in water resources- analyses and interpretations. U.S. Geological Survey, Tecniques of Water Resources Investigations. Retrieved from http://pubs.usgs.gov/twri/twri4a3/

KAYHANIAN M., SUVERKROPP C., RUBY A., TSAY K. (2007) Characterisation and prediction of highway runoff onstituent event mean concentration. Journal of Environment Management, 85:279-295. Doi:10.1016/j.jenvman.2006.09.024

KUSARI L., (2017) Regression model as a tool to predict concentrations of total suspended solids in rivers. EQA- International Journal of Environmental Quality, 23:35-42. Doi: 10.6092/issn.2281-4485/6865

LANNERGARD E., (2016) Potential for using high frequency turbidity as a proxy of total phosphorus in Savjaan. Swedish University o Agricultural Studies. Acculty o Natural Resources and Agricultural Sciences. https://stud.epsilon.slu.se/9475/1/lannergard_e_160908.pdf

LEE A. H., NIKRAZ H. (2015). BOD: COD Ratio as an Indicator for River Pollution. International Proceedings of Chemical, Biological and Environmental Engineering. 88 (2015). Doi: 10.7763/IPCBEE.

LEE J., LEE S., YU S., RHEW D. (2016). Relationhips between water quality parameters in rivers and lakes: BOD5, COD, NBOPs and TOC. Environmental Monitoring and Assessment, 188(4):252. Doi**:** 10.1007/s10661-016-5251-1

DOI: 10.6092/issn.2281-4485/7735

LINDBLOM E., AHLMAN S., MIKKELSEN P. (2007) Uncertainty in model - based prediction of copper loads in stormwater runoff. Novatech, 3:1599-1606.
MINGUTANA N. S., EGODAWATTA P., KOKOT S., GOONETILLEKE A. (2010) Determination of a set of surrogate parameters to asses urbanstormwater quality. Science of the Total Environment, 408(24):6251-6259. Doi: 10.1016/j.scitotenv.2010.09.015.
MUCHA Z., KUULAKOWSKI P. (2016) Turbidity measurements as a tool of monitoring and control of the SBR effluent at the small wastewater treatment plant- preliminary study. Archives of Environmental Protection, 42(3):33-36. Doi 10.1515/aep-2016-0030
RAZAVI S., TOLSON B., BURN D. (2012) Review of surrogate modelling in water resources. Water Resources Research, 48(7)Wo7401:1-32. Doi: 10.1029/2011 WR01 1527
SCHILLING K., KIM S.W., JONES CH. (2017) Use of water quality surrogates to estimate total phosphorus concentrations in Iowa rivers. Journal of Hydrology: Regional Studies, 12:111-121. Doi: 10.1016/j.ejrh.2017.04.006
SETTLE S., GOONETILLIKE A., AYOKO G. (2007) determination of surrogate indicators for phosphorus and solids in urban stormwater. application of multivariete data analyses techniques. Water, Air and Soil Pollution, 182(1-4):149-161. Doi: 10.1007/s112 70 - 006-9328-2
SING K., MALIK A., MOHAN D., SINBA S. (2004) Multivariete statistical techniques or the evaluation of spatial and temporal variations in water quality of Gomi river - case study. Water Resources, 38:3980-3992. Doi: 10.1016/j.watres.2004.06.011
THOMSON N., MCBEAN E., SNODGRASS W., MONSTRENKO I. (1997) Highway stormwater quality: development of surrogate parameter relationship. Water, Air and Soil Pollution, 94:307-347. DOI: 10.1007/BF02406066
WAGENER T., GUPTA H. (2005). Model Identification for Hydrological Forecasting under Uncertainty. Stochastic Environmental Research and Risk Assessment, 19:378-387. DOI: 10.1007/s00477-005-0006-5
WANG X., LU Y., HAN J., HE G.Z., WANG T. (2007). Identification of anthropogenic influences on water quality of rivers in Taihu watershed. Journal of Environmental Science, 19:475-481. Doi: 10.1016/S1001-0742(07)60080-1
XIAOYING Y., QUN L., XINGZHANG L., HENG ZH. (2017). Spatial regression an prediction of water quality in a watershed with complex pollution sources. Scientific Reports 7:8318. Doi: 101038/s41598-017-08254-w