International Proceedings of Chemical, Biological and Environment Engineering
Volume of IPCBEE (2016)
DOI: 10.7763/IPCBEE. 2016. V94. 21

Multivariate Statistical Technique for the Evaluation of the Environmental Factors and the Cyanobacteria Population in Nakdong River, South Korea

Kyeong-deok Park[1], Il-kyu Kim[1+]

[1] Dept. of Environmental Engineering, Pukyong National University, Busan 48513, South Korea

Abstract. In Nakdong River, the eutrophication cause by the cyanobacteria is occurred in summer. Some of cyanobacteria have the various toxic substance, the objectionable taste and troublesome odor. Therefore, using the sampling results (water temperature and the concentration of TN, TP, NO_3-N, NH_3-N, and PO_4-P), the hydrologic data (discharge), and the meteorological data (insolation and sunshine duration), multiple regression analysis was performed for each sampling site. Then, the accuracy of the regression equations was estimated to compare with the estimated values by the equation and the measured values and proved the relationship between the used data with the growth of cyanobacteria for each site. The results showed that the multiple regression equations of each sites seem that can explain the relationship of investigated variables and the cyanobacteria population. And, it seems the similar trends between upstream and downstream, and mainstream and tributaries. The upstream sites is considered that the phosphorus-family nutrients influence the growth of cyanobacteria. In contract, the nitrogen-family nutrient are estimated to effect in the downstream. In the tributary sites, the more variables were needed to approve the relationship with the cyanobacteria. However, the regression equations of the upstream sites showed a limit in the cyanobacteria described investigated variables. It seem to need more study for the new valid variable to use in the analysis.

Keywords: Multivariate statistics, Cyanobacteria, Nakdong River.

1. Introduction

Rivers have always been a source of food like fish and edible aquatic life, and fresh water for drinking. And river water been variously used in agriculture and industry. Many civilizations and cultures like the Indus was built along wide rivers, and many cities was built around rivers. Human life has been intimately related with rivers. But, with the industrial development and population increase, huge loads of waste from industry and domestic sewage have been released into rivers, resulting in deterioration of the water quality [1].

The worsen water quality has caused the cyanobacterial bloom. Cyanobacteria are a group of phytoplankton, and some freshwater cyanobacteria such as *Mycrocystis* and *Anabaena* are known to produce *Mycrocystin* known as haptatoxins, cytotoxins, and neurotoxins [2], [3]. Cyanobacteria also cause the objectionable taste and troublesome odor described as muddy or earthy-musty [4].

Especially, the eutrophication caused by the cyanobacteria is occurred in every summer in Nakdong River, South Korea. The Nakdong River has the gentle bed gradient and slow flow rate. Moreover, because the water of Nakdong River`s mainstream is consumed from the near cities and agricultural land, water flows more gently and water quality has been worse [5], and the cyanobacteria has increased in Summer. This is one of the most serious water quality problems in South Korea.

+ Corresponding author. Tel.: +82-51-629-6528; fax: +82-51-629-6523.
E-mail address: kimilky523@gmail.com.

In present study, the relationship between algae and the nutrients has been investigated with various statistical analysis. Especially, regression analysis has been used to improve the relationship between the nutrients and algae. But, these papers didn`t study the meteorological effect [6], [7], [8]. Recently, the regression analysis has been used the relationship between the nutrients and land-use patterns [9], [10].

In this study, we obtained the concentration of the nutrients, discharge, and metrological data, and performed the multiple regression analysis for each sampling site. Then, the obtained regression equation were evaluated the effectiveness to comparing with the estimated values by the equation and the measured values. In addition, we observed the variables used in the regression equation and proved the characteristics between the valid variables and the growth of cyanobacteria for each site.

2. Materials and Methods

2.1. Monitoring Area

The Nakdong River watershed is located in the South Korea. The Nakdong River are formed by the Taebaek Mountains in the east, and it faces the southern sea area and the total length of the river is 7422.02 km. The administrative districts of the Nakdong River`s watershed include areas from three metropolitan cities (Busan Metropolitan, Ulsan Metropolitan, and Daegu Metropolitan), and five provinces (Gyeongsang Nam and Buk Provinces, Jeolla Nam and Buk Provinces, and Gangwon Province). And the major stream of the Nakdong River`s watershed include 13 national rivers, 10 regional class 1 rivers, and 31 regional class 2 rivers. The Nakdong River watershed is the second largest watershed in South Korea (the watershed`s area is 23,817.3 km^2) [11].

In the present study, the monitoring sites, named Yulgi-gyo (M1), Bakjin-gyo (M2), Hamman-bo (M3), Limhaejin (M4), Chungdeok-gyo (T1), Songdo-gyo (T2), were selected on the Nakdong River as the river network. Four sites (M1, M2, M3, and M4) were located in the main stream of the river between Hapcheon-Changnyeong weir and Changnyeong-Haman weir, and two sites (T1 and T2) were located in the tributaries (Fig. 1).

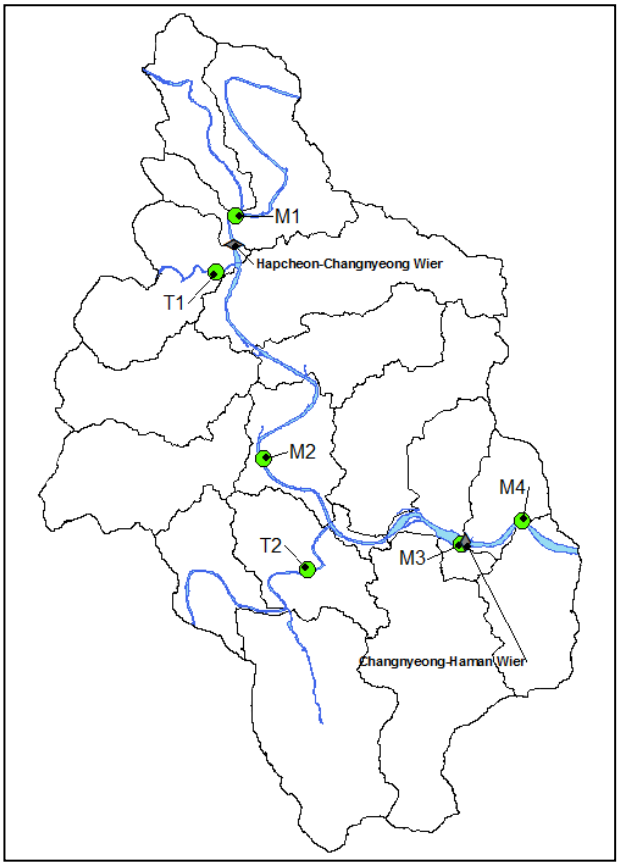

Fig. 1: The map showing the sampling sites on the Nakdong River.

2.2. Monitoring Parameters and Analytical Methods

Water samples were collected from June to October in 2015. Water temperature was measured at sampling sites with multi-parameter measuring instrument (Horiba U-51, Japan). And the collected samples were transported in a shaded and refrigerated state with no head-space. The samples were analysed for the concentration of the nutrients (TN, TP, NO_3-N, NH_3-N, and PO_4-P) and cyanobacteria cell counting with standard method [12]. The hydrologic and meteorological data (discharge, insolation, and sunshine duration) were collected from monitoring gauges close to each sampling sites via the Water Resources Management Information System (WAMIS).

2.3. Statistical Analysis

Statistical analysis for studying the relationship between the growth of cyanobacteria and the obtained data was performed using multiple regression analysis. Multiple regression is performed to examine the relationship between a single dependent variable Y and a set of independent variables. It possible to describe the linear relationship between measured value Y and X variables by using the multiple correlation coefficient as equation (1).

$$\hat{Y} = b_0 + b_1 x_1 + b_2 x_2 + ...b_n x_n \qquad (1)$$

$\hat{Y}$ is the predicted value, and residual e is the difference between the measured value and the predicted value. The degree to which the regression equation fits the data can be assessed by examining a quality called the coefficient of determination, defined as equation (2).

$$\text{coefficient of determination } (R^2) = \frac{\text{sum of squares regression (SSR)}}{\text{sum of squares total (TSS)}} \qquad (2)$$

And, the adjusted multiple correlation coefficient (R^{*2}) was used to estimate the regression equation and each independent values with equation (3).

$$R^{*2} = 1 - \frac{n-1}{n-p-1}(1-R^2) \qquad (3)$$

In equation (3), p is the number of independent values in the equation and n is the size of the sample size [13].

To select the valid independent values, stepwise method were performed. Backward elimination, a kind of stepwise methods, begins with the full set of variables, and increase the R^{*2}, and reduce the significance probability (Sig.) by removing the irrelevant variables. All the statistical analyses were performed using IBM SPSS Statistics 23 for Windows.

3. Results and Discussion

To understand the relationship between the concentration of the nutrients, metrological data, and discharge and the cyanobacteria population, the multiple regression analysis was performed for each sampling site. Table I shows the input variables used in the regression equation for each sampling site.

The results in this study showed that the discharge, insolation, sunshine duration, water temperature, and TP were valuable input variables in most of sampling sites. M1 and M2 are located in upstream of the research area and TP has been investigated a valid input variables in both sites. In addition, these sites were observed that water temperature wasn`t a valid variable, unlike the other sampling sites. It seems to be in the water temperature didn`t not have a relatively large effect on cyanobacteria growth.

The result of the research for M3 and M4, TP wasn`t considered to be a significant variable. The cyanobacteria in these sites is seem to affect the concentration of the nitrogen-family material, rather than phosphorus-family material. Moreover, these cyanobacteria is estimated that different species with the others.

T1 wasn`t investigated to affect by discharge, but was investigated to affect by water temperature and the concentration of TN and TP. Because of the small scale of this site and the small discharge than other sites, discharge isn`t considered a significant effect at T1.

T2 was investigated in which the most effective variables were among the sampling sites. Then, the discharge, weather, and various nutrients is estimated the multiple effects on the growth of cyanobacteria.

Table I: The list of the input variables used in the regression analysis of each sampling site.

Site	Discharge	Insolation	Sunshine Duration	Water Temp.	TP	PO_4-P	TN	NO_3-N	NH_3-N
	[a]	[b]	[c]	[d]	[e]	[f]	[g]	[h]	[i]
M1	○		○		○				
M2	○	○			○				○
M3	○	○		○			○		
M4	○			○					○
T1				○	○		○		
T2	○		○	○	○		○	○	

The multiple regression equations estimate the relationship of the cyanobacteria populations and the valid factors are shown by Table II. The significance probability of the regression equations about each sampling site has been set to less than 0.02.

In the case of M1, only three variables were used to create the regression equation, the R^2 was 0.617, and the Sig. was 0.012. The result means that the equation can explain the growth of cyanobacteria population about 62%, and the null hypothesis is rejected at the significance level of 0.012 degree. And, the regression equation of M2 used 4 variables, the R^2 was 0.710, and the Sig. was 0.009. It can explain the growth of cyanobacteria about 71%, and the null hypothesis is rejected the significance level of 0.01 degree. The R^2 of the regression equation of M3 was 0.758, and the Sig. was 0.004. And, The R^2 of the equation of M4 was 0.673, and the Sig. was 0.005. The result showed that M1 was more difficult to decide the regression equation than the other mainstream sites, and the unknown factors was at M1.

The equation of T1 can explain about 40%, and the sig. 0.020 degree. On the other hand, T2 can explain about 90%, and the Sig. was 0.000 degree. Unlike T2, T1 is estimated that there are more factors which are unknown to account for the growth of cyanobacteria.

Table II: The multiple regression equation for each sampling site

Site	Name	Regression equation	R^2	R^{*2}	Sig.
M1	Yulji-gyo	$\hat{Y} = (-439) \times [a] + (-2702) \times [c] + 468339 \times [e] + 61913$	0.617	0.512	0.012
M2	Bakjin-gyo	$\hat{Y} = 272 \times [a] + (-14374) \times [b] + 1031226 \times [e] + (-849718) \times [i] + 188441$	0.710	0.594	0.009
M3	Hamman-bo	$\hat{Y} = (-298) \times [a] + (-9528) \times [b] + 11963 \times [d] + 185406 \times [g] + (-437567)$	0.758	0.661	0.004
M4	Limhaejin	$\hat{Y} = (-72) \times [a] + (-8588) \times [d] + (-223106) \times [\mathrm{i}] + 291657$	0.673	0.583	0.005
T1	Chungdeok-gyo	$\hat{Y} = (-1623) \times [d] + 270839 \times [e] + (-20153) \times [g] + 71291$	0.576	0.460	0.020
T2	Songdo-gyo	$\hat{Y} = (-619) \times [a] + (-6408) \times [c] + (-3613) \times [d] + (-300539) \times [e] + 22217 \times [g] + 134120 \times [h] + 113600$	0.943	0.900	0.000

The comparison between the actual measured values and the estimated values indicated by the regression analysis are shown graphically by Fig. 2. In general, it seem that each regression equation can estimate the cyanobacteria population. However, the sampling sites located in the upstream sites (M1 and M2) was found to occur the difference between the measured value and estimated value gradually after 11th sampling. 11th is a time that is the beginning of September, it is estimated that a change in the species composition of the cyanobacteria because of the drop of temperature.

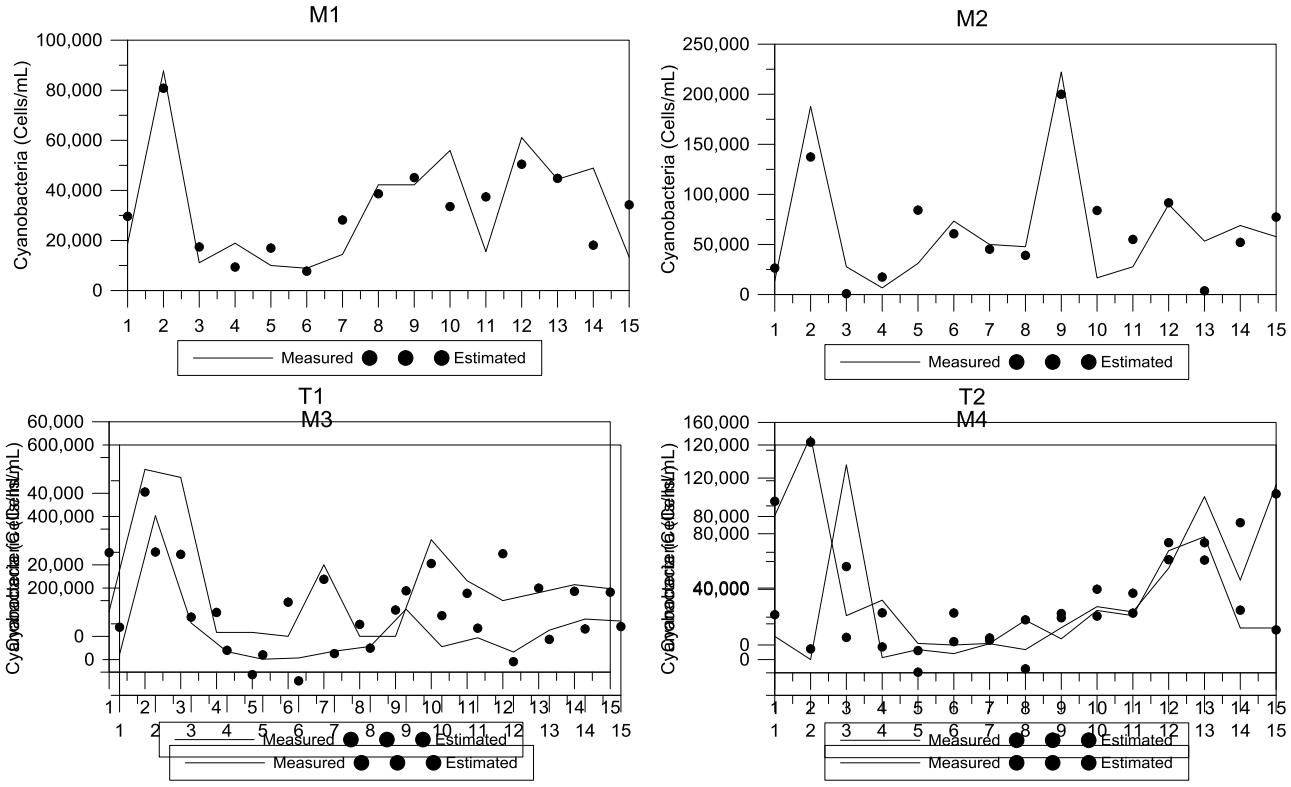

Fig. 2: The comparison of measured data and predicted values obtained by the regression equation (Continued).

4. Conclusions

This study shows that the relationship between the nutrients, discharge, and metrological data and cyanobacteria population for each sampling site. The results were different in the valid factors affecting the growth of cyanobacteria for each site. However, it seems the similar trends between upstream and downstream, and mainstream and tributaries.

In the upstream sites (M1 and M2), the phosphorus-family materials are presumed to influence the growth of cyanobacteria. But, in case of the downstream sites (M3 and M4), the nitrogen-family materials are presumed the valid factor. In the tributary sites (T1 and T2), cyanobacteria population is seem to be affected by both phosphorus and nitrogen materials. And, oobserving the hydrological and meteorological impact of each sites, it was investigated as valid variables in the downstream sites and tributary sites.

The multiple regression equations of each sites seem that can explain the relationship of investigated variables and the cyanobacteria population. However, the regression equations of the upstream sites (M1 and M2) show a limit in the cyanobacteria described investigated variables. It seem to need more study for the new valid variable to use in the analysis.

It is too difficult to investigate the growth of the cyanobacteria because many variables were interacted as a mutual combination. And, to improve the actuary of the statistical analysis, a number of monitoring data. In addition, Understanding topographical environment and the industry around the sampling site, the cyanobacterial characteristics, the movement and reaction of the nutrients, and etc. In order to understand the growth of cyanobacteria, this research need to continue.

5. References

[1] K. Ravindra, Ameena, Meenakshi, Monika, Rani and A. Kaushik. Seasonal variations in physico-chemical characteristics of River Yamuna in Haryana and its ecological best-designated use. *J. Environ. Monit*, 2003, **5**: 419–426.

[2] B. Yuan, J. Qu, and M. Fu. Removal of cyanobacterial microcystin-LR by ferrate oxidation-coagulation. *Toxicon* 2002, **40**: 1129-1134.

[3] I. R. Falconer. An Overview of problems caused by toxic blue–green algae (cyanobacteria) in drinking and recreational water. *Environmental Toxicology* 1999, **14** (1): 5–12.

[4] G. Izaguirre, C. J. Hwang, S. W. Krasner and M. J. McGuire. Geosmin and 2-Methylisoborneol from Cyanobacteria in Three Water Supply Systems. *Appl. Environ. Microbiol*. 1982, **43**(3): 708-714.

[5] W. Heo and B. Kim. Phosphorus and nitrogen loading from the main tributaries into the Nakdong River. *J. of the Korean Environmental Sciences Society* 1995, **4** (3): 187-195.

[6] W. K. Dodds. Eutrophication and trophic state in rivers and streams. *Limnol. Oceanogr*. 2006, **51**(1, part 2): 671–680.

[7] E. E. Van Nieuwenhuyse and J. R. Jones. Phosphorus–chlorophyll relationship in temperate streams and its variation with stream catchment area. *Can. J. Fish. Aquat. Sci*. 1996, **53**: 99–105.

[8] J. Ch ételat, F. R. Pick, A. Morin, and P. B. Hamilton. Periphyton biomass and community composition in rivers of different nutrient status. *Can. J. Fish. Aquat. Sci*. 1999, **56**: 560–569.

[9] J. Ding, Y. Jiang , Q. Liu , Z. Hou, J. Liao, L. Fu, and Q. Peng. Influences of the land use pattern on water quality in low-order streams of the Dongjiang River basin, China: A multi-scale analysis. *Science of the Total Environment* 2016**, 551–552** (1): 205–216.

[10] H. Bu, Y. Zhang, W. Meng, and X. Song. Effects of land-use patterns on in-stream nitrogen in a highly-polluted river basin in Northeast China. *Science of the Total Environment* 2016, **553** (15): 232–242.

[11] K. Y. Jung, K. Lee, T. H. Im, I. J. Lee, S. Kim, K. Han, and J. M. Ahn. Evaluation of water quality for the Nakdong River watershed using multivariate analysis. *Environmental Technology & Innovation* 2016, **5**: 67–82

[12] APHA, AWWA, and WEF. Standard Methods for the Examination of Water and Wastewater 22nd edition, APHA, 2012.

[13] S. Afifi, May, and V. A. Clark. Practical Multivariate Analysis Fifth edition. CRC Press, 2012.