International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 9 (2018) pp. 6723-6732

Assessment of Surface Water Quality in Geum River Basin, Korea using Multivariate Statistical Techniques

Zihan Liu[1], Jin Chul Joo [1,*], Sun Hwa Choi[2] , Namjoo Heo[2] Jungseok Jang[2], Jun Wook Hur [3]

[1]*Civil and Environmental Engineering, Hanbat National University, Republic of Korea.*
[2]*Water Resources & Environment Research Group, Rural Research Institute, Republic of Korea.*
[3]*Bio-Monitoring Center, Daejeon, Republic of Korea.*

Abstract

Water quality analysis for Geum river basin including eight major lakes (e.g., Kanwol lake, Daecheong lake, Seokho lake, Songak lake, Yedang lake, Cheongcheon lake, Chopyung lake, and Geumkang lake) was performed. Then, comprehensive multivariate statistical techniques (i.e., correlation analysis, principal component analysis/factor analysis, and multiple linear regression model) for Kanwol lake as a representative of agricultural reservoir, for Daecheong lake as a representative of drinking water reservoir, and for Geumkang lake as a representative of estuary reservoir were performed. According to the Spearman correlation results, TP values were positively associated with both agriculture and urban lands, and inversely associated with forest land due to the surface runoff of TP from the agriculture practices and impermeable surface of urban land. Based on the Pearson correlation analysis of various lakes, eutrophication from excessive algal growth was a complex function of various water quality parameters. Also, factor rotation lumping of huge and complex water quality parameters into simpler factor structures (i.e., varifactors) was feasible to facilitate the interpretation of complex water quality matrices. Finally, the multiple linear regression models explain approximately up to 58% of the Chl-*a* variations using different water quality variables, indicating that correcting spatial factors would be required to understand each spatial patterns of lakes with different functionalities.

Keywords: Multivariate statistical techniques, Geum river basin, Correlation, Multiple linear regression, Land use, Factor analysis, Water quality, Chl-*a*

INTRODUCTION

Due to the rapid expansion of the economy, South Korea has faced significant challenges in managing the scarce water resources due to the industrialization, the urbanization, and the climate change (Choi et al., 2017). Also, both population and industrial growth have increased the pressures on limited available water resources, and has degraded both water quality and ecosystem because of the municipal, industrial and agricultural pollution (Choi et al., 2017). During these population and industrial growth, the land cover in the river basin has gradually changed from natural vegetation to both agricultural and commercial land dominated by human activities, leading to the degradation of water quality (Mou et al., 2004; Lim et al., 2006; Amiri and Nakane, 2009; Giri and Qiu, 2016).

In South Korea, the contribution of nonpoint source pollution to water pollution was more than 70%, which has become a major pollution source of river basin water quality (Lim et al., 2006; Choi et al., 2017). Especially, the nonpoint source pollution such as the inflow of contaminated water including nitrogen (N) and phosphorus (P) and other nutrients, pesticides, fertilizers, and other organic or inorganic pollutants in farmland planting, livestock and aquaculture, rural life source, and other similar sources has become main causes of water quality degradation (Lim et al., 2006; Choi et al., 2017). Among various nonpoint sources, the load of agricultural nonpoint source pollution has been increased and needs to be urgently controlled.

Although complex water quality monitoring data comprised of a large number of physico-chemical, biological, and hydrological parameters are regularly obtained at numerous sites of river basins in South Korea, the acquisition of accurate water quality information is not easy due to the complex interactions among large numbers of parameters (Kim, 2003; Kazi et al., 2009; Choi et al., 2014). Despite of developments of numerous water quality index in the water quality evaluation, simple analysis of water quality index cannot accurately reflect the overall water quality situation of the river basin. Additionally, the various water quality parameters have certain correlations with parameters and type of land cover and use. Therefore, identifying the crucially main factors in various land cover and use types is important to estimate water quality variation.

Recently, multivariate statistical analysis has been proposed to be an ideal tool to analyze the huge and complex structures in water quality data accounting for both temporal and spatial variations and complex interactions among various parameters caused by natural and anthropogenic factors (Vega et al., 1998; Helena et al., 2000; Singh et al., 2004, 2005; Kuppusamy and Giridhar, 2006; Iscen et al., 2007; Shrestha and Kazama, 2007; Kazi et al., 2009; Zhao and Cui, 2009; Varol et al., 2012; Wang et al., 2013). In this study, water quality analysis for Geum river basin including eight major lakes (e.g., Kanwol lake, Daecheong lake, Seokho lake, Songak lake, Yedang lake, Cheongcheon lake, Chopyung lake, and Geumkang lake) was performed, and then, comprehensive multivariate statistical techniques (i.e., correlation analysis, principal component analysis and factor analysis, and multiple linear regression model) were applied.

The specific objective of this study is (1) to determine the correlation between land cover/use types and water quality parameters, (2) to identify the water quality parameters and covered sources accounting for temporal and spatial variances in water quality, and, (3) to estimate the greater impact factors of water quality parameters for chlorophyll-*a*. From this study, comprehensive multivariate statistical techniques can identify both temporal and spatial variations in complex water quality parameters and offer a valuable site-specific solution for reliable management of water resources.

MATERIALS AND METHODS

Water quality data

In this study, the water quality data of each object lake was extracted from 2000 to 2015 based on the Korean Water Environment Information System. Based on both precision and continuity in measurement, only 13 parameters are selected. The selected parameters include water temperature (WT), dissolved oxygen (DO), biochemical oxygen demand (BOD), chemical oxygen demand (COD), suspended solid (SS), total nitrogen (TN), nitrate nitrogen (NO_3^--N), ammonium nitrogen (NH_4^+-N), total phosphorus (TP), total dissolved phosphorus (TDP), phosphate phosphorous (PO_4^{3-}-P), and Chlorophyll-a (Chl-*a*).

Land cover

Historical land cover from 2000 to 2015 was used to assess changes in land cover, and was obtained from survey of lake environment and ecology in the Geum river system using a Landsat Enhanced Thematic Mapper (ETM) at a 50 m spatial resolution. Land cover was classified and aggregated for seven major land types such as urban (e.g., residential land, commercial land and transport land), agriculture (e.g., farmland and orchards), forest, grassland, unused land, water and wetland.

Statistical analysis

Water quality data set was subjected to four statistical techniques [i.e., correlation analysis (CA), principal component analysis (PCA)/factor analysis (FA), and multiple linear regression analysis]. The basic statistics of the 15-year data set were evaluated to monitor the variation of water quality parameters, and CA was performed to measure the strengths of association between parameters from each lake. Due to the wide differences in data dimensionality, z-scale transformation was applied to eliminate the impact of different units of water quality parameters before the multivariate analysis. Finally, multiple linear regression analysis was performed to estimate the eutrophic state (i.e., Chl-*a*) of each lake. All mathematical and statistical computations were performed using Excel 2016 and SPSS 22.0.

Pearson correlation

Correlation structure between the water quality parameters was evaluated using the Pearson correlation coefficients (*r*) to explain the non-normal distribution of water quality parameters. For Kanwol, Daecheong, and Geumkang lakes, the correlation analysis was applied to withdraw the correlation among water quality parameters, and the correlation of water quality parameters is displayed with Pearson correlation coefficient (*r*).

$$r = \frac{\sum[(Xi-\bar{X})(Yi-\bar{Y})]}{\sqrt{\sum(Xi-\bar{X})^2}\sqrt{\sum(Yi-\bar{Y})^2}} \quad (1)$$

where Xi and Yi refer to water quality in site i, respectively, and, $\bar{x}$ and $\bar{Y}$ are the overall mean water quality, respectively.

Factor analysis/Principal component analysis

Principal components analysis (PCA) was performed to transform the original water quality parameters into new, uncorrelated variables (i.e., principal components, PCs), so that the variations in the water quality can be explained as concisely as possible. Thus, PCs provide the information on the most meaningful water quality parameters, and describe the whole water quality with minimum loss of the water quality information (Vega et al., 1998; Shrestha and Kazama, 2007; Wang et al., 2013). Then, factor analysis (FA) was further performed to reduce the contribution of less significant water quality parameters and to simplify PCs by rotating the axis defined by PCA.

Multiple linear regression analysis

Step-wise regression model in multiple linear regression analysis was performed to estimate the Chl-*a*, and to yield variable F significance probability as the inspection criteria. The optimal model through regression statistical features value (P, R^2) was validated using the standard value is 0.05 and excluded value is 0.10. Independent variables include WT, DO, BOD, COD, SS, TN, NO_3^--N, NH_4^+-N, TP, TDP, and PO_4^{3-}-P for each lake to estimate the Chl-*a*.

The expression form of linear regression equation is as follows:

$$y = \beta x + e \quad (2)$$

where y is the dependent variable, x is the independent variable, β is the regression coefficient, and e is the random error terms.

RESULTS AND DISCUSSION

Relationship between Land Use and Water Quality

As shown in Fig. 1, the land use characteristics of each lake were presented. Land use characteristics of most lakes evaluated are dominated by forest land and agriculture land. Although both forest land and agriculture land accounted for more than 80% of total land area, both agriculture land and urban land have been recently expanded, and these changes in land use would significantly impact water quality of each lake.

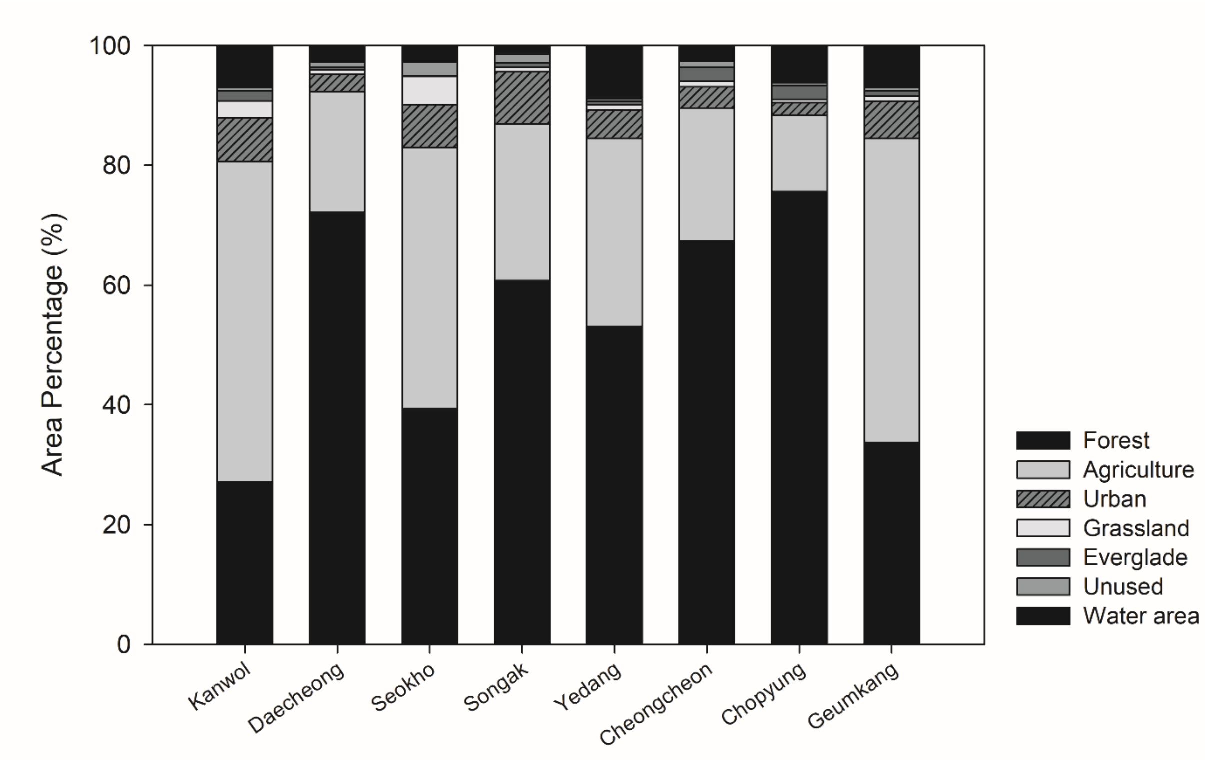


Figure 1. Land use percentage for each lake

According to the Spearman correlation results summarized in Table 1, TP values were positively associated with agriculture land and urban land, and inversely associated with forest land. This result is mainly attributed to the surface runoff of TP from the agriculture practices and the disposal of chemical fertilizers. Also, the increase in impermeable surface of the urban land indeed contribute to the increase in the surface runoff of TP. Similar positive relationships between agriculture land and TP in the watershed have been reported (Amiri and Nakane, 2009; Huang et al., 2013; Giri and Qiu, 2016).

Table 1. Spearman's rank correlation coefficients of land use type and water quality indices

	Temp	pH	DO	TN	TP	TN/TP	NO_3^--N	PO_4^{3-}-P	BOD	COD	Chl-*a*	SS
Urban	0.571	0.084	-0.286	-0.048	0.857**	-0.738*	0.314	0.400	0.886*	0.833*	0.786*	0.810*
Agriculture	0.476	0.467	0.095	0.262	0.857**	-0643	0.429	0.700	0.771	0.833*	0.905**	0.833*
Forest	-0.476	-0.476	-0.095	-0.262	-0.857**	0.643	-0.429	-0.700	-0.771	-0.833*	-0.905**	-0.833*
Grassland	0.214	-0.036	-0.190	0.286	0.024	0.095	0.657	0.900*	0.086	0.238	0.190	-0.024
Everglade	0.262	0.204	0.381	0.024	0.167	-0.238	-0.200	0.400	0.371	0.048	-0.048	0.310
Unused	-0.048	-0.922**	-0.786*	-0.190	-0.238	0.214	0.600	0.200	-0.429	-0.048	-0.310	-0.310
Water area	0.286	0.970**	0.643	0.262	0.571	-0.452	-0.257	-0.200	0.771	0.381	0.643	0.595

* Correlation is significant at the 0.05 level (2-tailed)

** Correlation is significant at the 0.01 level (2-tailed)

However, the forest land was negatively related to TP and Chl-*a*, indicating that the forest land effectively reduced the phosphorus loadings, and improved the water quality. Similarly, both vegetation and soil in the forest land were reported to reduce successfully the nutrient salts brought into the waterbodies by the surface runoff (Amiri and Nakane, 2009, Giri and Qiu, 2016). Considering that the increase in the forest land can reduce the loading amounts of TP and oxygen-consuming organics, and increase the concentration of dissolved oxygen, securing the minimum area of the forest land within the watershed is required to improve the water quality.

A significant positive correlation between BOD_5/COD and urban land has been observed. This increase in the organic matter of certain lake is mainly due to the point sources such as the wastewater treatment plant effluents and the nonpoint sources such as the surface runoff from impermeable surface of urban land. Similar to this study, the increase in impervious surface area negatively impacted stream water quality by increasing oxygen-consuming substances and by decreasing the concentration of dissolved oxygen (Giri and Qiu, 2016).

Temporal and Spatial Variations in Water Quality of Geum River Basin

In this study, water quality analysis for Kanwol lake as a representative of agricultural reservoir, for Daecheong lake as a representative of drinking water reservoir, and for Geumkang lake as a representative of estuary reservoir was performed. Then, comprehensive multivariate statistical techniques (i.e., correlation analysis, principal component analysis and factor analysis, and multiple linear regression model) were applied.

Through the PCA of three representative lakes (i.e., Kanwol lake, Daecheong lake, and Geumkang lake), the water quality parameters were been dimensionally reduced and classified. From the results of PCA, KMO test results were greater than 0.5, and the p value of Bartlett's test was less than 0.05, indicating that all data evaluated are suitable for PCA (Zhao and Cui, 2009).

Kanwol Lake

From water quality statistic descriptions for Kanwol lake summarized in Table 2, both temporal and spatial variations of most water quality parameters are less than 52.2%, especially CV value of temp and pH were 4.34% and 3.68%, respectively. However, both temporal and spatial variations of water quality parameters with P are greater than 81.3%, indicating that both TP and PO_4^{3-}-P are pretty variable in terms of season and space.

Since Chl-*a* is one of the most important water quality parameter to determine the eutrophic status of each lake, Pearson correlation between Chl-*a* and other water quality parameters was determined. According to the results of Pearson correlation analysis of Kanwol lake (Table 3), the correlations between Chl-*a* versus pH, DO, TN/TP, BOD_5, and COD_{mn} were significant at the 0.01 level, and the correlation between Chl-*a* and SS was significant at 0.05 level. These correlations indicated that eutrophication from excessive algal growth is a complex function of various water quality parameters.

Table 2. Water quality statistics of Kanwol lake watershed in 2000-2015

Index	Mean（Average）	SD[a]	CV(%)[b]
Temp(°C)	15.4(13.6-17.0)	0.67	4.34
pH	8.43(7.70-9.10)	0.31	3.68
DO (mg/L)	10.0(8.20-12.9）	1.43	14.3
T-N (mg/L)	2.99(1.39-6.80)	1.45	48.5
T-P (mg/L)	0.16(0.03-0.64)	0.13	81.3
TN/TP (mg/L)	24.9(4.90-61.9)	12.9	52.1
NO_3^--N (mg/L)	1.12(0.28-2.55)	0.54	48.6
PO_4^{3-}-P (mg/L)	0.05(0.003-0.46)	0.08	149.6
BOD_5 (mg/L)	4.57(2.10-8.60)	1.49	32.7
COD_{Mn} (mg/L)	10.8(7.30-18.5)	2.51	23.3
SS (mg/L)	23.3(9.40-81.9)	12.2	52.2
Chl-*a* (mg/L)	46.7(14.6-129)	22.9	48.9

[a] Standard Deviation

[b] Coefficient of Variation

Table 3. Pearson correlation coefficients of Kanwol lake in 2000-2015

	Temp	pH	DO	TN	TP	TN/TP	NO_3^--N	PO_4^{3-}-P	BOD_5	COD_{mn}	SS	Chl-*a*
Temp	1											
pH	0.127	1										
DO	-0.367*	0.127	1									
TN	0.265	-0.169	-0.117	1								
TP	0.182	0.071	0.146	0.754**	1							
TN/TP	-0.109	-0.214	-0.427**	0.314*	-0.634**	1						
NO_3^--N	0.159	-0.282	-0.128	0.859**	0.620**	-0.219	1					
PO_4^{3-}-P	0.064	-0.205	-0.148	0.701**	0.783**	-0.313*	0.628**	1				
BOD_5	0.305*	0.005	0.132	0.609**	0.694**	-0.423**	0.367*	0.629**	1			
COD_{mn}	0.203	0.185	0.131	0.057	0.182	-0.405**	-0.133	0.093	0.503**	1		
SS	0.227	0.039	0.009	0.485**	0.494**	-0.276	0.582**	0.388**	0.480**	0.180	1	
Chl-*a*	0.201	0.461**	0.442**	0.082	0.211	-0.436**	-0.122	-0.123	0.436**	0.673**	0.292*	1

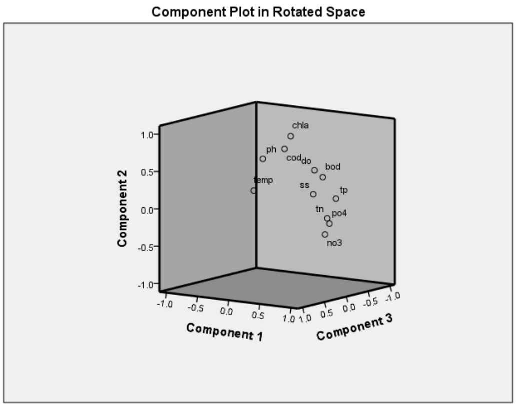
* Correlation is significant at the 0.05 level (2-tailed)

** Correlation is significant at the 0.01 level (2-tailed)

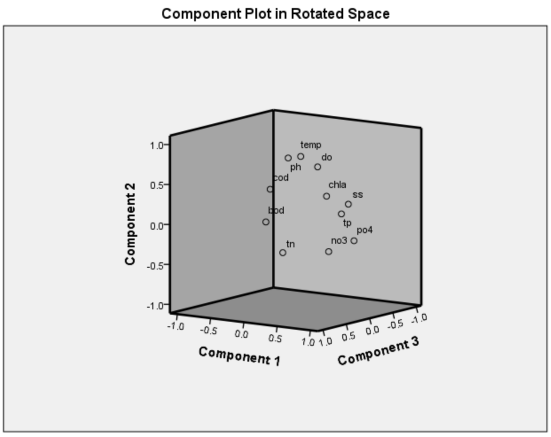
According to the results of FA of Kanwol lake (Table 4 and Fig. 2), the twelve water quality parameters were classified as three groups, and these corresponding three factors were related to the water quality characteristics. The factor 1 includes TN, TP, PO_4^{3-}-P, NO_3^--N, BOD, and SS. Considering that these water quality parameters mainly by composition of the eutrophic factors as N and P, the factor 1 was classified as "Nutrient element factors". The eigenvalue of factor 1 was 4.12, and accounted for 37.4% of the total variance. The factor 2 includes Chl-*a*, COD, and pH, and these water quality parameters all depended on the value of Chl-*a*, and, hence, the factor 2 was called as "Algae factor", and the eigenvalue of factor 2 was 2.44, and accounted for 22.1% of the total variance. The factor 3 includes DO and Temp, and these water quality parameters were related to the natural, and, hence, the factor 3 was called as "Natural factor", and the eigenvalue of factor was 1.42, and accounted for 12.9% of the total variance. The total variance of three factors is 72.5%.

Table 4. Rotated component matrix by factor analysis for Kanwol lake

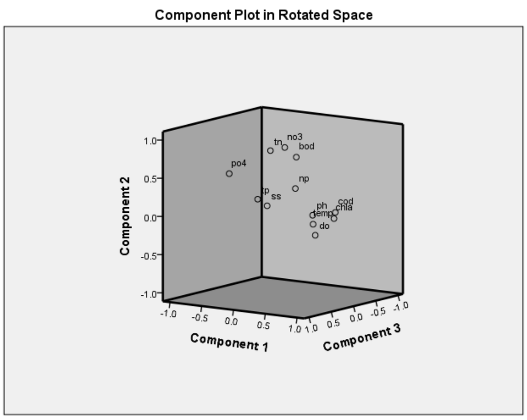
Variable	Component		
	1	2	3
T-N	0.905	-0.063	0.144
T-P	0.883	0.165	-0.088
PO_4^{3-}-P	0.856	-0.156	0.023
NO_3^--N	0.842	-0.291	0.105
BOD_5	0.746	0.454	0.019
SS	0.650	0.228	0.101
Chl-*a*	0.111	0.914	-0.152
COD_{mn}	0.146	0.775	0.040
pH	-0.194	0.609	0.049
DO	0.014	0.349	-0.837
Temp	0.186	0.331	0.801



(a) Kanwol lake



(b) Daecheong lake



(c) Geumkang lake

Figure 2. Component plot in rotated space for each lake

Daecheong Lake

From water quality statistic descriptions for Daecheong lake summarized in Table 5, both temporal and spatial variations of most water quality parameters were less than 50%, except for PO_4^{3-}-P and SS. However, both temporal and spatial variations of PO_4^{3-}-P and SS were great, with the CV value of PO_4^{3-}-P and SS were 69.8% and 55.1%, respectively (see Table 5). For Daecheong lake, the Pearson correlation coefficient between Chl-*a* and TP is significant at 0.05 level whereas correlation coefficients between Chl-*a* and other water quality parameters are 0.20~0.60, indicating that Chl-*a* is moderately correlated with TP (see Table 6). Aforementioned above, the algal growth in the Daecheong lake is phosphorous-limited, and algal growth can be reduced by reducing the P loadings into the waterbodies in terms of water quality management.

Table 5. Water quality statistics of Daecheong lake watershed in 2000-2015

Index	Mean（Average）	SD[a]	CV(%)[b]
Temp (ºC)	14.8(10.2-19.7)	1.83	12.4
pH	7.79(7.20-8.50)	0.26	3.41
DO (mg/L)	9.47(7.70-12.00)	0.78	8.32
T-N (mg/L)	1.67(1.12-2.32)	0.24	14.5
T-P (mg/L)	0.022(0.009-0.05)	0.007	34.8
TN/TP (mg/L)	81.1(38.9-241)	30.3	37.3
NO_3^--N (mg/L)	1.21(0.83-1.76)	0.16	13.8
PO_4^{3-}-P (mg/L)	0.005(0.001-0.025)	0.003	69.8
BOD_5 (mg/L)	1.12(0.70-1.60)	0.17	15.1
COD_{Mn} (mg/L)	3.16(2.70-4.70)	0.33	10.4
SS (mg/L)	3.15(0.70-11.9)	1.73	55.1
Chl-*a* (mg/L)	8.97(3.40-20.2)	3.67	40.9

[a] Standard Deviation

[b] Coefficient of Variation

Table 6. Pearson correlation coefficients of Daecheong lake in 2000-2015

	Temp	pH	DO	TN	TP	TN/TP	NO_3^--N	PO_4^{3-}-P	BOD_5	COD_{mn}	SS	Chl-a
Temp	1											
pH	0.610**	1										
DO	0.444**	0.290**	1									
TN	0.006	-0.164	-0.201*	1								
TP	0.385**	0.023	0.286**	0.441**	1							
TN/TP	-0.412**	-0.140	-0.418**	0.133	-0.734**	1						
NO_3^--N	0.012	-0.228*	0.060	0.576**	0.582**	-0.223*	1					
PO_4^{3-}-P	0.049	-0.220*	0.069	0.303**	0.617**	-0.363**	0.560**	1				
BOD_5	0.294**	0.014	0.026	0.536**	0.366**	-0.077	0.272**	0.098	1			
COD_{mn}	0.319**	0.136	0.248*	0.146	0.090	0.047	-0.027	-0.077	0.182	1		
SS	0.413**	0.116	0.205*	0.255*	0.749**	-0.492**	0.455**	0.585**	0.150	0.166	1	
Chl-a	0.529**	0.190	0.276**	0.294**	0.668**	-0.472**	0.502**	0.331**	0.309**	0.155	0.569**	1

* Correlation is significant at the 0.05 level(2-tailed)

** Correlation is significant at the 0.01 level(2-tailed)

According to the results of FA for Daecheong lake (Table 7 and Fig. 2), the twelve water quality parameters were classified as three groups, and these corresponding three factors were related to the water quality characteristics. Similar to the results of FA for Kanwol lake, the factor 1 includes T-P, PO_4^{3-}-P, NO_3^--N, SS, and Chl-*a*, classified as "Nutrient element factors". The eigenvalue of factor 1 was 3.94, and accounted for 35.8% of the total variance. The factor 2 includes Temp, pH, and DO, and these water quality parameters factor were associated with natural and physical property index, hence, the factor 2 was classified as "Natural factor", and the eigenvalue of factor 2 was 2.28, and accounted for 20.7% of the total variance. Finally, the factor 3 includes BOD_5 and COD_{Mn}, classified as the "Organic factor", and the eigenvalue of the factor 3 was 1.33, and accounted for 12.1% of the total variance. The total variance of three factors is 68.6%.

Table 7. Rotated component matrix by factor analysis for Daecheong lake

Variable	Component		
	1	2	3
T-P	0.858	0.205	0.243
PO_4^{3-}-P	0.838	-0.176	-0.055
SS	0.812	0.292	0.028
NO_3^--N	0.736	-0.264	0.343
Chl-*a*	0.653	0.411	0.269
WT	0.209	0.852	0.190
pH	-0.131	0.767	-0.027
DO	0.199	0.665	-0.178
COD_{mn}	-0.087	0.445	0.423
BOD_5	0.132	0.116	0.827
T-N	0.355	-0.252	0.782

Geumkang Lake

From water quality statistic descriptions for Geumkang lake summarized in Table 8, both temporal and spatial variations of all water quality parameters were less than 50%, indicating that significant variations of water quality were not monitored in terms of time and space. From the Pearson correlation analysis for Guemkang lake, the significant correlation between Chl-*a* versus pH, DO, TN/TP, and COD_{Mn} were observed. For Daecheong lake, the Pearson correlation coefficient between Chl-*a* and TP is significant at 0.05 level whereas correlation coefficients between Chl-*a* and other water quality parameters are 0.20~0.60, indicating that Chl-*a* is moderately correlated with TP (see Table 9). Aforementioned above, the algal growth in the Daecheong lake is phosphorous-limited, and algal growth can be reduced by reducing the P loadings into the waterbodies in terms of water quality management.

Table 8. Water quality statistics of Geumkang lake watershed in 2000-2015

Index	Mean（Average）	SD[a]	CV(%)[b]
Temp (°C)	15.7(13.4-17.5)	1.01	6.47
pH	8.31(7.80-8.9)	0.27	3.34
DO (mg/L)	10.5(8.30-12.9)	1.15	11.0
T-N (mg/L)	3.32(2.40-4.83)	0.49	15.0
T-P (mg/L)	0.103(0.037-0.18)	0.03	36.1
TN/TP (mg/L)	36.8(20.5-95.9)	16.7	45.4
NO_3^--N (mg/L)	1.95(1.53-2.86)	0.25	13.2
PO_4^{3-}-P (mg/L)	0.031(0.005-0.06)	0.01	42.1
BOD_5 (mg/L)	3.25(2.20-4.30)	0.52	16.1
COD_{Mn} (mg/L)	7.42(5.60-9.30)	1.02	13.8
SS (mg/L)	24.9(12.9-61.6)	10.3	41.2
Chl-*a* (mg/L)	39.7(15.7-67.3)	14.6	36.8

[a] Standard Deviation

[b] Coefficient of Variation

Table 9. Pearson correlation coefficients of Geumkang lake in 2000-2015

	Temp	pH	DO	TN	TP	TN/TP	NO_3^--N	PO_4^{3-}-P	BOD_5	COD_{mn}	SS	Chl-*a*
Temp	1											
pH	0.074	1										
DO	0.180	0.735**	1									
TN	-0.415**	-0.046	-0.231	1								
TP	-0.431**	0.254	0.181	0.531**	1							
TN/TP	0.281	-0.243	-0.366*	-0.032	-0.806**	1						
NO_3^--N	-0.346*	-0.240	-0.465*	0.791**	0.238	0.159	1					
PO_4^{3-}-P	-0.273	-0.008	-0.213	0.464**	0.556**	-0.232	0.321*	1				
BOD_5	-0.094	0.395**	0.274	0.505**	0.396**	-0.134	0.424**	0.471**	1			
COD_{mn}	0.111	0.478**	0.552**	-0.084	0.238	-0.462**	0.006	-0.202	0.494**	1		
SS	-0.523**	-0.382**	-0.541**	0.328*	0.203	-0.072	0.445**	0.019	-0.202	-0.223	1	
Chl-*a*	0.112	0.556**	0.523**	0.044	0.350*	0.447**	-0.020	-0.241	0.265	0.700**	-0.143	1

* Correlation is significant at the 0.05 level(2-tailed)

** Correlation is significant at the 0.01 level(2-tailed)

According to the results of FA for Geumkang lake (Table 10 and Fig. 2), the twelve water quality parameters were classified as four groups, and these corresponding four factors were related to the water quality characteristics. The factor 1 includes Chl-a, COD_{Mn}, DO, and pH associated with the algal growth so call the "Algal factor". The factor 2 includes T-N, NO_3^--N, and BOD_5 associated witch with the nitrogen whereas the factor 3 includes T-P, TN/TP, and PO_4^{3-}-P associated with the phosphorus. Both factor 2 and 3 can be classified as "Nutrient element factor". The factor 4 includes Temp and SS associated with the "Natural factor". Four groups accounted for 82.2% of the total contribution.

Table 10. Rotated component matrix by factor analysis of Geumkang lake

Variable	Component			
	1	2	3	4
Chl-*a*	0.891	0.032	0.126	0.021
COD_{mn}	0.872	0.102	0.068	0.138
DO	0.639	-0.203	0.186	0.567
pH	0.600	0.056	0.190	0.514
T-N	-0.039	0.843	0.225	-0.265
NO_3^--N	-0.027	0.841	-0.079	-0.463
BOD_5	0.318	0.783	0.153	0.362
T-P	0.218	0.327	0.878	-0.063
TN/TP	-0.399	0.155	-0.849	0.035
PO_4^{3-}-P	-0.442	0.551	0.576	0.263
Temp	0.099	-0.218	-0.536	0.502
SS	-0.126	0.104	0.175	-0.880

Multiple linear regression analysis

All the explanatory variables were added into the multiple linear regression models to predict Chl-*a*. As summarized in Table 11, COD_{Mn}, DO, pH, and Temp were dominant variables in the multiple linear regression model for Kanwol lake. Whereas T-P, Temp, and NO_3^--N were dominant variables for Daecheong lake whereas BOD_5, pH, COD_{Mn}, and DO were dominant variables for Geumkang lake. From the results, the multiple linear regression analyses using various water quality variables are significant (p<0.05) for all lakes, and identify the explanatory water quality variables for each lakes.

Table 11. Multiple step-wise linear regression models with various water quality parameters

Chl-*a*	Multiple setp-wise linear regression models	R^2	*P*
Kanwol lake	$y = -359.577 + 4.746X_1 + 6.636X_2 + 20.989X_3 + 0.212X_4$	0.682	<0.05
Daecheong lake	$y = -15.054 + 158.946X_5 + 0.822X_4 + 6.799X_6$	0.586	<0.05
Guemkang lake	$y = -233.773 + 5.471X_7 + 23.229X_3 + 0.664X_1 + 0.891X_2$	0.654	<0.05

X_1: COD, X_2:DO, X_3: pH, X_4: Temp, X_5: T-P, X_6: NO_3^--N, X_7: BOD

These multiple linear regression models explain approximately up to 58% of the Chl-*a* variation in water quality, and the relatively low predictive ability of regression models suggests that other factors may not have not been included in the regression models. Also, the Chl-*a* variations for Kanwol lake as a representative of agricultural reservoir, for Daecheong lake as a representative of drinking water reservoir, and for Geumkang lake as a representative of estuary reservoir were quite different, indicating that correcting spatial factors would be required to understand each spatial patterns of functional lakes.

CONCLUSIONS

Water quality analysis for Geum river basin including eight major lakes (e.g., Kanwol lake, Daecheong lake, Seokho lake, Songak lake, Yedang lake, Cheongcheon lake, Chopyung lake, and, and Geumkang lake) was performed, and then, comprehensive multivariate statistical techniques (i.e., correlation analysis, principal component analysis and factor analysis, and multiple linear regression model) for Kanwol lake as a representative of agricultural reservoir, for Daecheong lake as a representative of drinking water reservoir, and for Geumkang lake as a representative of estuary reservoir were performed. According to the Spearman correlation results, TP values were positively associated with agriculture land and urban land, and inversely associated with forest land. And, these results are mainly attributed to the surface runoff of TP from the agriculture practices and impermeable surface of the urban land. According to the results of Pearson correlation analysis of various lakes, eutrophication from excessive algal growth is a complex function of various water quality parameters.

According to the results of FA of Kanwol lake, the twelve water quality parameters were classified as three groups (i.e., nutrient element factors, algae factor, and natural factor), and these corresponding three factors were related to the water quality characteristics accounting for 72.5% of the total variance. Similar to those of Kanwol lake, the twelve water quality parameters for Daecheong lake were classified as three groups (i.e., nutrient element factors, natural factor, and organic factor) accounting for 68.6% of the total variance. For Geumkang lake, the twelve water quality parameters were classified as four groups, and these corresponding four factors were algal factor, nitrogen factor, phosphorus factor, and natural factor accounting for 82.2% of the total contribution. Thus, factor rotation lumping of huge and complex water quality parameters into simpler factor structures (i.e., varifactors) was feasible to facilitate the interpretation of complex water quality matrices. Finally, the multiple linear regression models explain approximately up to 58% of the Chl-*a* variation in water quality using different water quality variables, indicating that correcting spatial factors would be required to understand each spatial patterns of lakes with different functionalities.

ACKNOWLEDGMENT

This research was supported by Geum River Watershed Management Committee, "Investigation of lake environment and ecology around Geum River Basin(201500710002)".

REFERENCES

[1] Choi, I. -C., Shin, H. -J., Nguyen, T. T., Tenhunen, J. Water Policy Reforms in South Korea: A Historical Review and Ongoing Challenges for Sustainable Water Governance and Management. *Water*, 2017, 9, 717-737.

[2] Mou, B., Wang, Q. C., Hershey, A. E. Land-use, stream order and stream water physical and chemical qualities, *Acta Ecologica Sinica*, 2004, 24(7), 1486-1492.

[3] Lim, K. -H, Lee, Y. -S., Kim, L.-H. Estimation of Pollutant Loadings from Watershed into Lakes of Ganwol and Boonam, *Korean Wetlands Society*, 2006, 8(4), 33-40.

[4] Amiri, B. J., Nakane, K. Modeling the linkage between river water quality and landscape metrics in the Chugoku District of Japan, *Water Resources Management*, 2009, 23(5), 931-956.

[5] Giri, S., Qiu, Z. Understanding the relationship of land uses and water quality in twenty first century: A review, *Journal of Environmental Management*, 2016, 173, 41-48.

[6] Kim, Y. J. Evaluation of urban lake water quality using principal component analysis. *Journal of Korean society of environmental administration*, 2003, 9, 197-203.

[7] Kazi, T. G., Arain, M. K., and Jamali, N. J. Assessment of water quality of polluted lake sing multivariate statistical techniques: A case study. *Ecotoxicology and environmental safety*, 2009, 72, 301-309.

[8] Choi, S. H., Kim, H. D., and Kim, D. H. The evaluation of water quality characteristics using a multivariate statistical analysis in the artificial lake. *Journal of Korea commission on irrigation and drainage*, 2014, 21, 87-100.

[9] Vega, M., Pardo, R., and Barrado, E. Assessment of seasonal and polluting effects on the quality of the quality of river water by exploratory data analysis. *Water Research*, 2006, 37, 4119-4124.

[10] Helena, B., Pardo, R., and Vega, M. Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga river, Spain) by principal component analysis. *Water research*, 2000, 34, 807-816.

[11] Singh, K. P., Malik, A., Mohan, D., and, Sinha, S. Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India)-a case study. *Water research*, 2004, 38, 3980-3992.

[12] Singh, K. P., Malik, A. and Sinha, S. Water quality assessment and apportionment of pollution sources of Gomti river (India) using multivariate statistical techniques: a case study, *Analytica. Chimica. Acta.,* 2005, 538, 355-374.

[13] Kuppusamy, M. R., and Giridhar, V. V., Factor analysis of water quality characteristics including trace metal speciation in the coastal environmental system of Chennai Ennore. *Environ Int.* 2006, 32, 174-179.

[14] Shrestha, S., and Kazama, F., Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin, Japan. *Environmental modeling & software*, 2007, 22, 464-475.

[15] Iscen, F. C, Emiroglu, O, Ilhan, S., Arslan, N., Yilmaz, V., Ahiska, S., Application of multicariate statistical tecniques in the assessment of surface water quality in Uluabat Lake. *Journal of environmental science and health, Part A*, 2008, 40, 1775-1790.

[16] Kazi, T. G., Arain, M. B., Jamali, M. K., Jalbani, N., Afridi, H. I., Sarfraz, R. A., Baig, J. A., Shah, A. Q., Assessment of water quality of polluted lake using multivariate statistical techniques: a case study. *Ecotoxicol. Environ. Safety*, 2009, 72, 301-309.

[17] Zhao, Z. W., and, Cui, F. Y. Multivariate Statistical analysis for the surface water quality of Luan River, China. *Journal of Zhejiang University SCIENCE A*, 2009, 10, 142-148.

[18] Varol, M., Gökot, B., and Bekleyen, A. Water quality assessment and apportionment of pollution sources of Tigris river (Turkey) using multivariate statistical techniques – a case study. *River research and applications*, 2012, 28, 1428-1438.

[19] Zhao, P., Xia, B. C., and Qin, J. Q. Multivariate correlation analysis between landscape pattern and water quality. *Acta ecologica sinica*, 2012, 32(8), 2331-2341.

[20] Wang, Y., Wang, P., and Bai, Y. J., Assessment of surface water quality via multivariate statistical techniques: A case study of the Songhua river Harbin region, China. *Journal of hydro-environment research*, 2013, 7, 30-40.