한국환경보건학회지, 제42권 제4호(2016)
J Environ Health Sci. 2016; 42(4): 280-292

pISSN: 1738-4087 eISSN: 2233-8616
http://dx.doi.org/10.5668/JEHS.2016.42.4.280

원 저 Original articles

한강수질 평가를 위한 COD (화학적 산소 요구량) 모델 평가

Chemical Oxygen Demand (COD) Model for the Assessment of Water Quality in the Han River, Korea

Jae Hyoun Kim† and Jinnam Jo*

Department of Health Science, School of Natural Science, Dongduk Women's University, Seoul 136-714 Korea
**Department of Statistics and Information Science, Dongduk Women's University, Seoul 136-714 Korea*

ABSTRACT

Objectives: The objective of this study was to build COD regression models for the Han River and evaluate water quality.

Methods: Water quality data sets for the dry season (as of January) during a four-year period (2012-2015) were collected from the database of the Han River automatic water quality monitoring stations. Statistical techniques, including combined genetic algorithm-multiple linear regression (GA-MLR) were used to build five-descriptor COD models. Multivariate statistical techniques such as principal component analysis (PCA) and cluster analysis (CA) are useful tools for extracting meaningful information.

Results: The r^2 of the best COD models provided significant high values (> 0.8) between 2012 and 2015. Total organic carbon (TOC) was a surrogate indicator for COD (as COD/TOC) with high reliability (r^2=0.63 in 2012, r^2=0.75 for 2013, r^2=0.79 for 2014 and r^2=0.85 for 2015). The ratios of COD/TOC were calculated as 2.08 in 2012, 1.79 in 2013, 1.52 and 1.45 in 2015, indicating that biodegradability in the water body of the Han River was being sustained, thereby further improving water quality. The BOD/COD ratio supported these findings. The cluster analysis revealed higher annual levels of microorganisms and phosphorous at stations along the Hangang-Seoul and Hantangang areas. Nevertheless, the overall water quality over the last four years showed an observable trend toward continuous improvement. These findings also suggest that non-point pollution control strategies should consider the influence of upstreams and downstreams to protect water quality in the Han River.

Conclusion: This data analysis procedure provided an efficient and comprehensive tool to interpret complex water quality data matrices. Results from a trend analysis provided much important information about sources and parameters for Han River water quality management.

Key words: Chemical oxygen demand (COD), genetic algorithm-multiple linear regression (GA-MLR), water quality parameter (WQP)

I. Introduction

The Korea Environmental Protection Agency is responsible for protecting the nation's water quality and drinking water through the monitoring program operated by the Ministry of Environment. Hence, a real-time monitoring system has been established at 2,499 locations across four major river basins in Korea for the implementation of water quality assessments and for the development of guidelines and standards. The monitoring was performed at 20% distance from the top in water depth.

[†] **Corresponding author:** Department of Health Science, Dongduk Women's University, Dongduk Women's University, 60 Hwarang-ro 13-gil, Seongbuk-gu, Seoul, Korea, Tel: +82-2-940-4484, Fax: +82-2-940-4480, E-mail: kjhyon@dongduk.ac.kr
Received: 11 May 2016, Revised: 17 August 2016, Accepted: 17 August 2016

The Han River serves as a main source of drinking water to some 20 million people in Seoul and Kyoungi-do. Waste water from industries and excess nutrients (such as nitrogen and phosphorus) from sewage treatment plants and runoff of agricultural activities increasingly contributes to water pollution of the river.[1,2] Significant increases in algae from eutrophication or climate change harm water quality in rivers or streams.[3] Moreover, a number of large- and medium-sized industries still do discharge without proper treatment at upstream and downstream, as well.

Excessive nutrients and other water pollutants are responsible for serious water quality problems of the Han River aquatic ecosystem particularly under very low flow conditions during the winter dry season.[4-7]

A importance of 2015 data reflect the fact that significant inputs of terrestrially derived dissolved organic matter (DOM) during summer storm events in 2015 significantly affected water quality characteristics of the Han River.[8] As such, streams contain significant amounts of non-biodegradable materials including organics, chlorophyll, solids and toxics even in winter.

Total Organic Carbon (TOC), Biochemical oxygen demand (BOD) and chemical oxygen demand (COD) are indicators to predict the efficiency of removal of a high strength effluents in streams and rivers for regulatory guidance.[9] Furthermore, the advantages of the COD test are that it assesses all chemically oxidizable to industrial wastewater and the results are reproducible. Thus, it provides an index to assess the effect of aquatic system before and after treatment on the receiving stream.[10]

The TOC and BOD or COD correlation procedure has been becoming commonly accepted protocol in water management operations. The coefficient BOD/COD and BOD/TOC ratios are commonly used indicators of biodegradability improvement of the wastewater.[11] If BOD/COD is < 0.3 then it cannot be treated biologically, reflecting the existence of recalcitrant organic compounds.[12-14] The COD should always measure higher than TOC and then BOD.

A recent study indicated that the water quality of COD showed an upward trend at more than 78% of monitoring stations by non-biodegradable organic matter and nutrients.[15] Hence, management for upstream sites is required to safeguard drinking water sources to reduce identified risk parameters.[16]

The water quality index (WQI), a simple numeric expression, has been applied to surface water quality assessment, water quality trends detection, pollution sources identification, and optimization of significant parameters and water quality monitoring stations.[17-21,25] Nonetheless, because the WQI masks specific information, some detail can be lost, even though there are more advantages of WQI than disadvantages.[22]

A simple COD index has been demonstrated to characterize water quality at geographically diverse monitoring stations along the Sumjin River.[23] The study was however unable to quantify a whole spectrum of water quality at diverse sampling sites.

Alternatively, we have conducted to develop a 5-descriptor COD regression model using monitoring data during the dry season (January, 2014) on a combined statistical technique, PCA/CA/GA-MLR.[24] The genetic algorithm-multiple linear regression (GA-MLR) approach simplifies and accelerates considerably the optimization process because the linear parameters are not the fitted ones. This linear model demonstrated a spectrum of water quality contributors, potential pollution sources and specific sites along the Han River basin. These findings led us to investigate and compare the trends and variations of COD models for the past four years (2012-2015).

The primary purpose of this study is to evaluate the inter-annual variation of water quality through COD regression models, and to assess parameters and contaminated sites for solutions of water quality problems along the river.

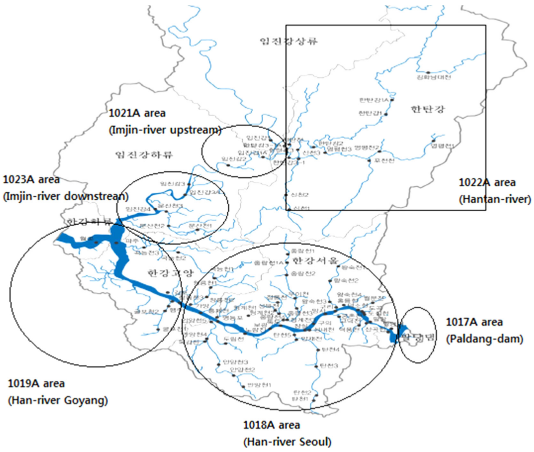

Fig. 1. Map of study area with the six major locations of water quality monitoring stations.[26)]

II. Materials and Methods

1. Study area

Map of study area showing locations of monitoring stations including six major areas (Hantan river-1022A, Paldangdam-1017A, Han River Seoul-1018A, Han River Goyang-1019A, Imjin River downstream-1023A and Imjin River upstream-1021A) in the river basin is depicted in Fig 1. All these areas of the Han River are relatively unstable in their hydraulics and hydrology, and this subsequently affects reservoir systems, resulting in soil erosion due to flooding during the summer months. Low flows/cease-to-flows occur during the winter dry season, causing more concentrated concentration during winter.

2. Statistical methods and data

All the monitoring data was obtained and used from the real-time auto-monitoring database (http://www.koreawqi.go.kr) for this study. The data consisted of monthly (as of January) mean values of 18 parameters on each station for the four years from 2012 to 2015. The dataset was then used to compute min, max, mean and 95th percentiles.

First, we performed the statistical analyses with deletion of any observation with missing values on one or more of analysis variables. The distribution of the COD of each data set fitted the normal distribution, with a goodness of fit p values of in the range of 0.7-0.75. Accordingly, the un-logged form of data was useful for analysis in this study. The final number of testing set (n) were 144 in 2012, 156 in 2013, 198 in 2014 and 234 in 2015 after deletion of missing data. Results of 2014 model were only cited from a previous work for comparison.[24)]

The R-squared of the regression is the fraction of the variation in dependent variable that is accounted for (or predicted by) independent variables. The 95% confidence interval estimates fall within that 95% confidence interval, so if the interval does not contain 0, p value will be 05 or less. The t statistic is the coefficient divided by its standard error. The standard error is an estimate of the standard deviation (SD) of the coefficient. The correlation coefficient r measures the strength and direction of a linear relationship between two variables on a scatterplot

For the GA-MLR model, validation was performed using separation of the data into two independent sets, Y-randomization (Kxy–Kx), cross-validation (Q_{2LOO}) and bootstrap (Q^2_{boot}). Cross-validation of model was performed by Leave-one-out (LOO) method. The best predictive models were selected by maximizing the ΔK ($K_{XY} - K_X$) values; in any case, negative threshold values are not allowed.

To identify the spatial similarity of the sampling stations, hierarchical cluster analysis (CA) was performed on data by means of Ward's method, using Euclidean distances as a measure of high internal similarity and low external similarity. A Principal component analysis PCA analysis was performed to identify which were the most important variables in the CA groups, i.e, those responsible for the spatial variation of water quality recorded in the study area. GA-MLR analysis for variable selection was performed by the MOBYDIGS 1.1 package (TALETE srl-Milano, Italy). The PCA and CA of water quality data sets were made through XLSTAT-Pro 7.5 (Addinsoft) and PAST software package.[24] Details of all statistical techniques are given in the methods section.[26]

III. Results and Discussion

1. Descriptive statistics of data

Descriptive statistics including maximum, minimum, average (arithmetic) and 95th percentiles of WQPs from proposed monitoring stations are given in Table 1. The water quality over the four past years varied depending on chemical, physical, biological, and characteristics of individual parameters.

In 2015, the river presents, slight alkaline pH in average as 7.7 in the range of 5.9 and 8.9. The results showed that DO concentrations ranged from 7.6-19.4 mg/L, the COD varied from 0.6 to 24.4 mg/L, a range of 0.4-17.2 mg/L for BOD, 0¡©18.74 mg/L for NH_3-N, 0-1.07 mg/L for T-P, 0.69-23.38 mg/L for T-N, and 0.4-15.2 mg/L for TOC. The overall average and the highest of F-coli and T-coli were 2710 MPN/100 mL and 61,000 MPN/100 mL, and 18,200 MPN/100 mL and 400,000 MPN/100 mL, respectively. Both parameters showed a large range of variation. The values of Chl were obtained in the range of 0-74.7 mg/L.

There was a positive association between T-N and T-P, with low correlation ($r^2 = 0.31$). High T-P/T-N ratios (>120) were observed at many stations of Namhangang, Bukhangang, Hangang-Seoul, and Hantangang indicating that the wastewater generated by industries, as well as domestic activities, is generally released into streams to increase the T-P concentration. T-N and T-P loadings are regulated by low stream flow in dry winter season.[27] Hence, efficient efforts need to be made to reduce nutrients overload in the particular areas. The U.S. EPA criteria state that T-P as phosphorous (P) should not exceed 0.05 mg/L in any stream at the point where it enters any lake or reservoir.

Mean SS concentrations have sharply been decreasing over the last 4 years from 7.46 in 2012 to 4.32 in 2015. A 95th percentile level, correspondingly, decreased sharply to 13.64 mg/L, which is safely below 30 mg/L, a 95th percentile effluent standard. Anthropogenic, that is, point-source pollution can increase turbidity through the addition of SS. BOD plus SS has traditionally been used to measure of the strength of effluent released by inadequate treatment of sanitary sewage that usually contains large quantities of suspended solids that are mostly organic in nature. However, a negative and weak (r^2= -0.25) association between SS and BOD mean values demonstrates a trend of continuous decrease of SS and increase of BOD in the same period. Input with high BOD and SS levels are presumably from different activities.

The average value of the conductivity, EC, showed an increasing trend during 2012-2014, but a sharp decrease from 2014 to 2015 was shown also in the 95th. The average of 2015 was measured to be 655 µS/cm (ranging from 66 to 31,009 µS/cm). In general, the highest concentrations were seen at a number of downstream targets. A significant increase in conductivity above 444 µS/cm is a useful indicator that polluting discharges have entered the water with various ions including the more dissolved ionic solutes such as nitrate, phosphate, and sodium.[28]

The Chl quantification allows the determination of algal blooms in waters body that may influence the taste and odor of drinking water sources.[29] There was an overall decrease, except a sharp drop to about 4.0 (µg/L) in 2013 which is presumably by N deficiency in water body, from 7.74 (µg/L) to 6.09 in 2015. We could not confirm if chlorophyll was related to input of excess nutrients in the same areas with low r^2 values (logTN-logChl: r^2=0.14; logTP-logChl: r^2= 0.27). Excessive levels of nitrogen and phosphorus in the Hangang-Seoul area

Table 1. Descriptive statistics of 18 water quality parameters for each January in 2012-2015.

	2012					2013					2014					2015				
Parameters	Mean	SD	Min	Max	95th	Mean	SD	Min	Max	95th	Mean	SD	Min	Max	95th	Mean	SD	Min	Max	95th
pH	7.95	0.75	1.6	9.4	8.22	7.75	0.4	6.3	9.9	8	7.63	0.44	6.5	8.8	8.22	7.7	0.49	5.9	8.9	8.4
Temp (°C)	3.33	3	0	13.3	8.72	2.54	2.65	0	11	8.3	2.99	2.43	-0.8	11.8	8.72	3.22	2.53	-0.1	11.6	8.54
DO (mg/L)	14.57	2.05	8.5	19.5	16.6	13.97	2.02	5.8	20.7	16.5	13.81	2.12	4.8	22	14.66	13.86	1.84	7.6	19.4	16.87
BOD (mg/L)	2.18	2.71	0.2	17.5	9.22	2.66	3.77	0	20.4	9.65	2.32	3.14	0.2	18.3	9.22	2.59	3.11	0.2	27.8	7.94
COD (mg/L)	4.61	3.42	1.4	19.9	12.02	4.57	4.76	0.5	32.4	13.25	4.04	3.64	0.6	24.4	12.02	4.35	3.02	0.4	17.2	9.97
TOC (mg/L)	2.28	2.09	0	14.1	7.83	2.91	2.58	0.6	19.1	25.35	2.84	2.87	0	22.6	7.83	2.83	2.27	0.4	15.2	7.5
SS (mg/L)	7.46	23.78	0.2	265.3	23.67	6.51	19.7	0.1	172.2	25.35	5.44	11.85	0.2	102.1	23.67	4.32	7.6	0.1	60.6	13.64
T-N (mg/L)	5.71	4.09	0.98	20.2	14.55	5.14	3.69	0.91	20.42	13.12	5.43	4.4	0.9	28.65	14.55	5.57	3.97	0.69	27.38	13.42
T-P (mg/L)	0.12	0.21	0	1.82	0.53	0.12	0.26	0	2.8	0.46	0.1	0.24	0	2.19	0.53	0.09	0.15	0	1.07	0.38
DTN (mg/L)	5.39	4.01	0.85	19.68	13.68	4.92	3.58	0.78	20.28	11.87	5.18	4.23	0.82	26.21	13.68	5.35	3.82	0.67	26.36	12.76
DTP (mg/L)	0.1	0.2	0	1.632	0.4	0.09	0.24	0	2.708	0.35	0.07	0.21	0	2.07	0.4	0.06	0.12	0	0.973	0.24
NH_3-N (mg/L)	1.4	2.71	0	12.78	8.14	1.38	2.53	0	11.62	6.86	1.32	2.96	0	21.22	8.14	1.33	2.86	0	18.74	4.96
NO_3-N(mg/L)	3.28	1.58	0.8	8.83	6.86	2.93	1.14	0.02	8.41	5.36	3.13	1.64	0.18	11.46	6.86	3.37	1.56	0.47	9.68	6.34
PO_4-P (mg/L)	0.07	0.14	0	0.93	0.35	0.07	0.21	0	2.47	0.294	0.05	0.19	0	2.01	0.35	0.05	0.11	0	0.94	0.19
EC (µS/cm)	397	1073	60	12690	1401	574	2629	47	27904	1205	761	3298	59	29315	1401	655	2611	66	31009	1085
F-Coli (CFU/100 mL)	2628	14529	0	160000	12150	2142	7011	0	73000	7900	2083	7249	0	56000	12150	2710	7999	0	61000	17000
T-Coli (CFU/100 mL)	25692	99413	0	900000	79250	22371	65047	0	620000	122000	13689	40532	0	380000	79250	18200	47967	0	400000	93400
Chl (µg/L)	7.74	10.34	0	62.2	24.92	4.06	4.82	0	36.2	11.59	6.42	11.67	0	107.1	24.92	6.09	10.03	0	74.7	24.1

Observations (N): 2012 (144); 2013 (156); 2014 (198); 2015 (234). *Mean denotes arithmetic average. SD: standard deviation. 95th: 95th percentile levels. Parameters: pH, Temp (Water Temperature), DO (Dissolved Oxygen), BOD (Biochemical Oxygen Demand, BOD_5), COD (Chemical Oxygen Demand), TOC (Total Organic Carbon), SS (Total Suspended Solids), T-N (Total Nitrogen), T-P (Total Phosphorus), DTN (Dissolved Total Nitrogen), DTP (Dissolved total Phosphorus), NH_3-N (Ammonia-Nitrogen), NO_3-N (Nitrate-Nitrogen), PO_4-P (Orthophosphate), EC (Electrical Conductivity), F-Coli (Fecal Coliform), T-Coli (Total Coliform), Chl (Chlorophyll a).

were not from Chl, but from non-point sources such as urban activity, wastewater discharge and sewer.

In the table, we see that the parameters T-N, DTN, NO_3-N, Temp and EC have lines that slow increase over time when comparing means only. A distinct decreasing trend was evident in a P group such as T-P, DTP and PO_4-P. Other parameters remained unchanged. Considering 95th percentiles, a sharp decrease was seen in T-P, DTP, PO_4-P, EC and SS. The P trend showed a gradual decrease in the mean value, but a steep plunge from 2014 to 2015 in the 95th percentile. The upgrade of wastewater treatment plant may have led to decrease P input from domestic wastewater disposal plants and urban sewage to watersheds.30)

The C/N/P composition (as BOD:T-N:T-P) of individual nutrients in wastewater correspond to the needs of the bacteria in the effluents. This is crucial to the effectiveness of the biodegradability. The C:N:P ratio in municipal wastewater is about 100:20:5, and 27–45:8–32:1 in river streams using both arithmetic and discharge-weighted mean concentrations.31) The approximate ratio in 2015 (January) in this study was about 90:28:5 with high carbon content.

The bacterial water quality was separately plotted with the 95th percentile of both T-Coli and F-Coli (Fig. 2). T-Coli showed a markedly high level in 2013, and a 16 percent increase from 2014 to 2015, while F-Coli steadily increased up to 50% from 2013 to 2015.

Microbial contamination in river is extremely serious concern with population increase and their activities.32) Many agencies such as WHO and US EPA have chosen to base criteria for recreational water compliance upon percentage compliance levels, typically 95% compliance levels useful in accounting for precautions on drinking water safety or public health. The mean and 95th percentile levels of T-Coli far exceeded a guideline (grade II) value (T-Coli/F-Coli effluent limit: <5,000 MPN/100 mL). But means of F-Coli were safely below the limit while all 95th percentiles exceeded a recommended effluent limit during the same period (Fig. 2). Corresponding 95th values of F-Coli showed a steady increase above the limit. In 2015, the highest fecal and total coliform concentrations markedly exceeded a limit at stations in Namhangang, Hangang-Seoul and Hangang-Goyang areas. High levels of F-Coli were also observed in the Hangang-Seoul area. Fecal coliform bacteria may occur in human and animal waste, and total coliform group occur in human feces.33) Even though fecal coliforms are below the guideline value, upstream nonpoint sources need careful management for drinking water aesthetics.34)

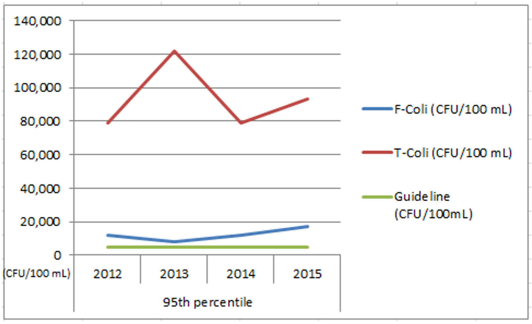

Fig. 2. The 95th percentile values compared to a guideline of T-Coli and F-Coli during 2012-2015.

2. The best GA-MLR model for COD

The results of the best GA-MLR model for COD with the highest Q^2_{LOO} values and the lowest values of PRESS and SE based on 5 descriptors are given in Table 2.

It was apparent that each model [Eq. (1) – Eq. (4)] exhibited robust predictive ability (r^2=0.91, Q^2_{LOO}=89.8 for COD_{2012}; r^2=0.93, Q^2_{LOO}=91.6 for COD_{2013}; r^2=0.93, Q^2_{LOO}=91.9 for COD_{2014}; r^2=0.87, Q^2_{LOO}=86.7 for COD_{2015}) low SDEP (1.01～1.38). Selected parameters included: TOC, DO, T-P, PO_4-P and Chl in 2012; TOC, NH_3-N, NO_3-N, T-Coli, DTN in 2013; TOC, BOD, SS, DTN and Chl in 2014; TOC, Chl, pH, SS and DPT in 2015.

The linear COD models seemed pretty

Table 2. Statistical results of the best GA-MLR models.

	N	r^2	Q^2_{LOO}	Q^2_{boot}	Q^2_{Ext}	Kxy–Kx	SE	SDEP	PRESS	F	Best-Fit model
COD_{2012}	144	0.92	0.90	0.88	0.91	8.76	1.04	1.09	165.5	277.08	COD = – 0.07 DO + 0.67 TOC + 0.90 T-P – 0.49 PO_4-P + 0.10 Chl + 2.76 Eq. (1)
COD_{2013}	156	0.83	0.92	0.91	0.92	5.89	1.3	1.38	292.02	383.69	COD = 0.84 TOC + 0.32 NH_3-N + 0.12 NO_3-N + 0.20 T-Coli – 0.37 DTN + 0.02 Eq (2)
COD_{2014}	198	0.93	0.92	0.91	0.92	9.53	1.06	1.1	236.25	486.67	COD = 0.40 BOD + 0.33 TOC + 0.09 SS + 0.18 DTN + 0.20 Chl + 0.36 Eq (3)*
COD_{2015}	234	0.87 ?	0.87 ?	0.86 ?	0.88 ?	10.32 ?	1.09 ?	1.01 ?	284.36 ?	316.12 ?	COD = 0.79 TOC + 0.1 Chl – 0.08 pH + 0.07 SS + 0.07 DTP + 4.8 Eq (4)

* was cited from ref. (26). coefficient of determination (r^2), Bootstrap (BOOT) validated Q^2 (Q^2_{boot}), Leave-one-out cross-validated Q^2 (Q^2_{LOO}), Collinearity index (***Kxy – Kx***), Externally validated Q^2 (Q^2_{Ext}), Standard Deviation Error in Prediction (SDEP), Standard Deviation Error in Calculation (SDEC), Prediction Sum of Squares (PRESS), Standardized Regression Coefficient (SRC), Standard error of estimate (SE) and the Fisher's (F) value.

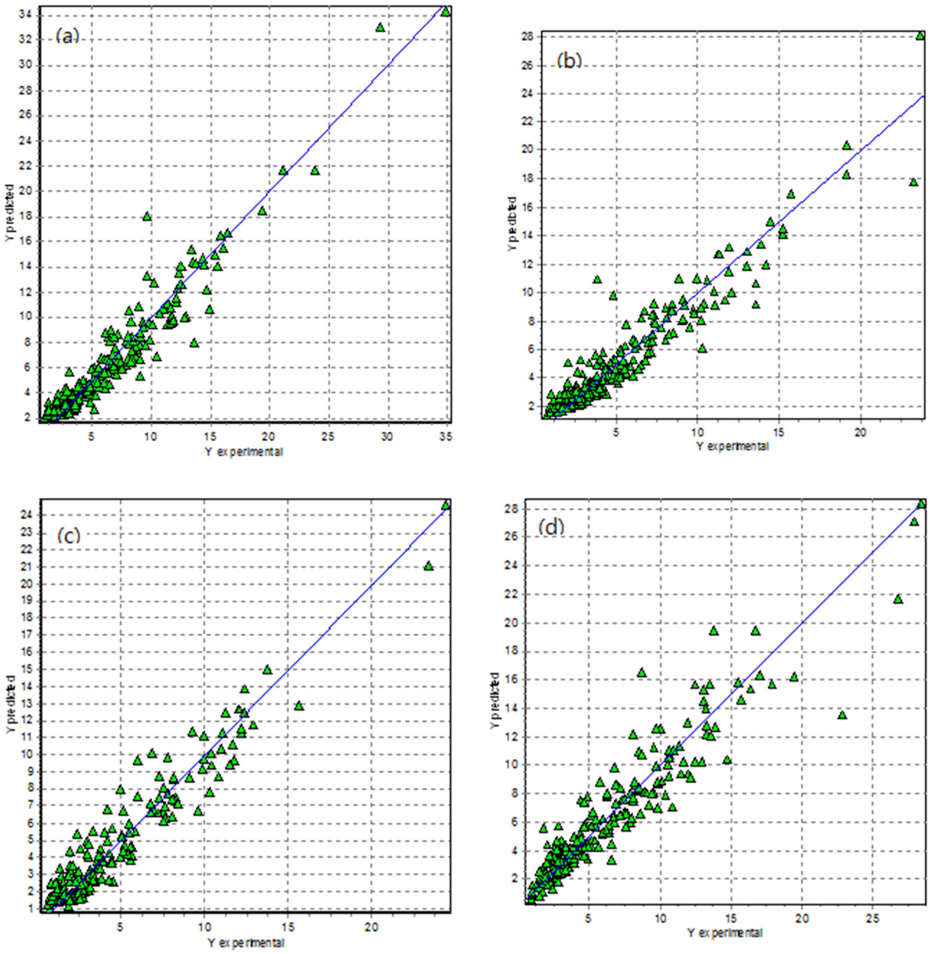

Fig. 3. Measured (x-axis) *vs.* predicted (y-axis) COD models [(a), 2012; (b), 2013; (c), 2014; (d), 2015].

straightforward [Fig. 3, (a)-(d)].

The 1022A45 (Sincheon 1), 1022A50 (Sincheon 2) and 1022A55 (Sincheon 3) sites on 2015 model [Fig. 3-(d)] were identified to have leverage points that had a strong influence on the estimate of regression coefficients.

The COD models showed that TOC, Chl (except in 2013) and SS (except in 2012 and 2013), respectively, were the most commonly used variables in the predictive COD models. TOC and SS measure the total organic carbon concentration of particulate solids as surrogates. The Chl was another important contributor to the river water quality in 2012, 2014 and especially in 2015. A similar finding was observed in the Geum-River, where the COD in the dry season was mainly attributed to Chl and phytoplankton contains most of TP.[35] The Chl is used as a biological indicator in monitoring and assessing the levels of organic pollution.

Water quality degradation was strongly determined by contributing variables with high positive regression coefficient in each COD model: TOC, T-P and PO_4-P in 2012; TOC, NH_3-N and NO_3-N in 2013; TOC and BOD in 2014; TOC and Chl in 2015. Conspicuously, the phenomenal nitrogen export to river waters in 2013 was likely to remain significant in conjunction with T-Coli. High nitrate concentrations in drinking water are dangerous for human health,[36] and that nitrogen enrichment produces eutrophication, which is responsible for toxic algal blooms, water anoxia, fish kills and habitat. Interestingly, high levels of T-Coli and TOC in the 95th percentile for 2013 were seen, possibly by a number of incidents of water pollution due to sewage discharges from municipal wastewater and sewage treatment plants.[40] TOC provides an estimate of the amount of the natural organic matter (NOM) in the water source.

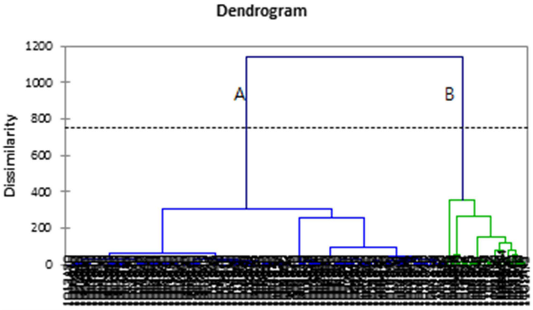

Fig. 4. Two major clusters (A and B) in a dendrogram as of January, 2015.

3. Multivariate analysis

1) Cluster analysis (CA)

The results of CA based on station scores of PCA for the two regions are shown in a dendrogram (Fig. 4).

Table 3. Contamination sites collected from cluster group B (Fig. 4) and extreme sites (bold) obtained from convex hull ordination (Fig. 6).

Area (location)*				
Somgang	Hangang (Seoul)	Hangang (Goyang)	Hantangang	Imjingang (downstream)
1006A55	1018A06, 1018A50	1019A40	1022A35	1023A25
1006A60	1018A12, 1018A64	1019A45	1022A45	1023A40
1007A50	1018A22, 1018A72		1022A50	
	1018A26, 1018A74		1022A55	
	1018A40, 1018A75			
	1018A42, 1018A76			
	1018A46, 1018A80			
	1018A84			

*obtained from 2015 data. Extreme sites are in bold. Number denotes each station number under its Area (location).

Table 4. Correlation matrix of the 18 physicochemical parameters determined for 2015 data.

Variables	pH	Temp	DO	BOD	COD	TOC	SS	T-N	T-P	DTN	DTP	NH_3-N	NO_3-N	PO_4-P	EC	F-Coli	T-Coli	Chl
pH	1	-0.088	0.174	-0.274	-0.367	-0.307	-0.199	-0.383	-0.196	-0.384	-0.175	-0.309	-0.291	-0.179	-0.031	-0.123	-0.165	-0.182
Temp	-0.088	1	-0.584	0.416	0.478	0.504	0.218	0.428	0.440	0.432	0.403	0.433	0.056	0.374	0.159	0.346	0.216	0.115
DO	0.174	-0.584	1	-0.432	-0.416	-0.478	-0.212	-0.491	-0.428	-0.496	-0.384	-0.474	-0.135	-0.368	-0.110	-0.367	-0.183	-0.055
BOD	-0.274	0.416	-0.432	1	0.759	0.763	0.478	0.674	0.470	0.666	0.378	0.659	0.119	0.346	0.139	0.432	0.383	0.435
COD	-0.367	0.478	-0.416	0.759	1	0.921	0.470	0.734	0.605	0.727	0.505	0.658	0.241	0.471	0.073	0.401	0.346	0.484
TOC	-0.307	0.504	-0.478	0.763	0.921	1	0.424	0.772	0.607	0.766	0.488	0.685	0.243	0.454	0.076	0.369	0.321	0.425
SS	-0.199	0.218	-0.212	0.478	0.470	0.424	1	0.346	0.306	0.340	0.237	0.300	0.112	0.226	0.122	0.272	0.287	0.308
T-N	-0.383	0.428	-0.491	0.674	0.734	0.772	0.346	1	0.555	0.998	0.476	0.843	0.507	0.445	0.037	0.321	0.260	0.215
T-P	-0.196	0.440	-0.428	0.470	0.605	0.607	0.306	0.555	1	0.550	0.973	0.296	0.439	0.953	0.069	0.502	0.409	0.240
DTN	-0.384	0.432	-0.496	0.666	0.727	0.766	0.340	0.998	0.550	1	0.474	0.852	0.498	0.442	0.026	0.319	0.263	0.202
DTP	-0.175	0.403	-0.384	0.378	0.505	0.488	0.237	0.476	0.973	0.474	1	0.235	0.430	0.988	0.028	0.517	0.408	0.165
NH_3-N	-0.309	0.433	-0.474	0.659	0.658	0.685	0.300	0.843	0.296	0.852	0.235	1	0.014	0.213	0.022	0.214	0.175	0.149
NO_3-N	-0.291	0.056	-0.135	0.119	0.241	0.243	0.112	0.507	0.439	0.498	0.430	0.014	1	0.409	-0.044	0.226	0.180	0.091
PO_4-P	-0.179	0.374	-0.368	0.346	0.471	0.454	0.226	0.445	0.953	0.442	0.988	0.213	0.409	1	0.021	0.511	0.396	0.132
EC	-0.031	0.159	-0.110	0.139	0.073	0.076	0.122	0.037	0.069	0.026	0.028	0.022	-0.044	0.021	1	0.006	0.011	0.341
F-Coli	-0.123	0.346	-0.367	0.432	0.401	0.369	0.272	0.321	0.502	0.319	0.517	0.214	0.226	0.511	0.006	1	0.634	0.175
T-coli	-0.165	0.216	-0.183	0.383	0.346	0.321	0.287	0.260	0.409	0.263	0.408	0.175	0.180	0.396	0.011	0.634	1	0.303
Chl	-0.182	0.115	-0.055	0.435	0.484	0.425	0.308	0.215	0.240	0.202	0.165	0.149	0.091	0.132	0.341	0.175	0.303	1

Pearson's correlation coefficient (r). Values in bold are significantly different from 0 with a significance level alpha=0.05

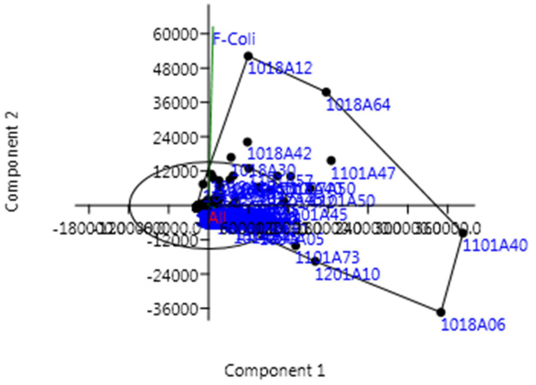

Fig. 5. An ordination graph for the first two axes with 95th percentile confidence limit (circle) including a convex hull polygon.

The observations plot from discriminant analysis (DA) illustrates the clustering of the two classes. The dissimilarity defined by Euclidean distance and the combination of cluster is based on the Ward method. The dendrogram shows that all the monitoring stations may be generally grouped into two clusters (A and B), where cluster B corresponds to regions of relatively high pollution listed in Table 3. Four major influential pollution sources in B group include 1006A (3 stations in Somgang), 1019A (2 stations in Hangang-Goyang), 1018A (15 stations in Hangang-Seoul), 1022A (4 stations in Hantangang) and 1023A (2 stations in Imjingang-downstream).

A convex hull and a 95th confidence limit were drawn in the scatter plot (Fig. 5), in order to show points that were categorized as polluted stations. Five stations on a polygon were shown as extreme sites.

2) Correlation matrix of parameters

Data in Table 4 provides the Pearson's correlation matrix of the water quality parameters obtained from PCA. According to the results of correlation matrices, some clear physico-chemical relationships can be readily inferred.

The river water pH value had relatively weak to fair correlations, i.e., most of the correlation coefficients (r) are less than 0.4 (absolute value)

Table 5. The goodness-of-fit (r^2) and ratios between parameters

Year	r^2			Ratio		
	BOD/COD	COD/TOC	BOD/TOC	BOD/COD	COD/TOC	BOD/TOC
2012	0.73	0.63	0.60	0.40	2.08	1.67
2013	0.82	0.75	065	0.83	1.79	1.47
2014	0.77	0.79	0.59	0.97	1.52	1.48
2015	0.58	0.85	0.58	1.35	1.45	1.80

Arithmetic average values were used for calculation.

with other parameters. There was a positive correlation between pH and DO (0.17); and negative weak correlations existed between pH and COD (-0.37), pH and T-N (-0.38), pH and DTN (-0.38). This indicated that the pH did not affect chemical and biological processes. As expected, DO is negatively correlated with most organic-related parameters. The degradation of organic matter in the water consumes the available DO, leading to the rapid depletion of available DO in water, resulting in high COD, BOD, NH_3-N and TN. There were strong relationships between TN and DTN (1.0), COD and TOC (0.92), TP and DTP (0.97) DTN and NH_3-N (0.85) and moderately relationships between BOD and COD (0.76), TOC and BOD (0.76), TN and COD (0.73), TOC and DTN (0.77), COD and DTN (0.73), F-Coli and T-Coli (0.63) and COD and NH_3-N (0.66). It was shown that significant positive correlations between organic-related parameters (BOD, COD and TOC) and inorganic-nutrients parameters (TN, T-P, DTN, NH_3-N) were also found. However, weak correlations exist in F-Coli (< 0.5) and in E-Coli (0.18 to 0.41) except between F-Coli and T-Coli (0.63), PO_4-P and F-Coli (0.51), DTP and F-Coli (0.52), TP and F-Coli (0.50). Chl also showed weak correlations (0.09-0.48) with other parameters.

As observed in the correlation matrix, strong relationships between TN and DTN (1.0), COD and TOC (0.92), TP and DTP (0.97) DTN and NH_3-N (0.85). These findings suggest that, increase of nutrients (N, P) directly affects to the increase of micro-biological pollution. As such, 1001A75, 1003A48 and 1003A50 were distinctively separated from the all stations which were characterized for having high concentrations of COD, BOD, T-Coli F- coli and T-P, T-N and DTN.

An ordination graph (Fig. 5) additionally identified extreme locations far beyond 95th percentile as 1018A06 (Dogpoongcheon), 1018A12 (Doshimcheon), 1101A40 (Tancheon1) and 1201A10 (Hangang-Seohae). Those sites appear to be influenced by different sources, comparing to places detected in the hierarchical cluster analysis. The result however does not explain what parameters are responsible for the critical site group in a data set.

4. BOD/COD/TOC correlations

The study examined the effectiveness of performance monitoring by comparing their correlations. The goodness-of-fit between BOD_5, COD, and TOC is shown in Table 5. The coefficient of determination (r^2) between COD/TOC gave a high correlation of 0.63 in year 2012, 0.75 in 2013, 0.79 in 2014 and 0.85 in 2015, respectively; 0.73, 0.82, 0.77, 0.58 for BOD/COD; 0.60, 0.65, 0.59 and 0.58, respectively, for BOD/TOC. The ratios of BOD/COD was in the range from 0.4 in 2012 to 1.35 in 2015, from 1.45 to 2.08 for COD/TOC, and 1.48 to 1.80 for BOD/TOC, respectively.

Low BOD/COD values (usually less than 0.1) indicate their resistance to conventional biological treatment. As ratio BOD/COD was >0.4 in this study, the effluent was then easily biodegradable. BOD/TOC ratios of as high as 1.5 during four years indicate since the wastewater in 0.5 or greater is considered to be easily treatable by biological means, and less biodegradable when the ratio is about 0.5 or less. Finally, untreated effluents in the presence of inorganic reducing agents or non-biodegradable organics tend to have COD/TOC ratios (2.6-6).[37] Thus, a strong association and low ratios (from 2.08 in 2012 to 1.45 in 2015) in this study clearly explain that biodegradability of water body of the Han River increased over time, thereby further improving water quality.

Although the present study has yielded some preliminary findings, there are possible limitations that this research has faced: Firstly, our research used unmatched number of observation, due to different amount of missing data dropped out in each model. It is necessary to use a same number to validate and compare predictive COD models for future studies. Secondly, TOC/COD is assumed to be estimated only when the composition of the organic composition in river streams is constant. So stratification or categorization in temporal (such as daily-, weekly-, monthly- or seasonally-adjusted), or spatial (heterogeneity of locations) frames on one station will be essential to derive the best COD model. Finally, the simple CA technique provided information only about the water quality variables to the sources of contamination.[38] But a fully decentralized density-based clustering algorithm[39] to correctly detect noncircular clusters with complex water quality data along population-based geographical areas will verify more details.

The findings above revealed that Namhangang, Bukhangang, Hangang-Seoul and Hantangang were clearly the most contaminated regions in 2015 COD model, of which two leading sources of pollution were Hantangang and Hangang-Seoul. A less number of sites were detected in Hangang (Goyang) and Imjingang (downstream) areas, improving water quality, compared to the 2014 data.

IV. Conclusion

The hybrid GA-MLR was utilized to build the best MLR models. Statistical analyses successfully estimated the degree of contamination due to anthropogenic inputs from point/non-point sources. The COD_{2015} increase in the upper regions and Seoul areas of the Han River is still problematic similar to previous years. And many sites throughout the Hantangang and Hangang-Seoul as well as other sites contained distinguishable levels of microorganisms and nutrients over four past years, causing deterioration of water quality. Nevertheless, the analysis shows there is a trend of improving water quality in most areas of the Han River under investigation. Further evaluation of spatial and seasonal variation over a large geographical area may also be beneficial to forecast water quality of the Han River. The water quality improvement then requires the expansion of controls to nonpoint regulations in the Han River basin, eventually to create drinking water management programs.

Acknowledgements

This study was supported by the sabbatical leave grant of the 2nd semester in 2014, and the 1st semester in 2015 of Dongduk Women's University.

References

1. Han JG, Lee YK, Kim TH, Hwang EJ. Analysis of seasonal water pollution based on rainfall feature at Anyang river basin in Korea. *Environ Geology.* 2005; 28(4): 599-608.
2. Lee LK, Kim JH, Kim J. Monitoring the water quality of the Wangsukcheon river over a two year period. *Toxicol Environ Health Sci*. 2015; 7(1): 91-96.
3. Hur M, Lee I, Tak BM, Lee HJ, Yu JJ, Cheon SU, et al. Temporal shifts in cyanobacterial communities at different sites on the Nakdong River in Korea. *Water Res*. 2013; 47(19): 6973-6982.
4. An KG, Jeon HW, Choi JW. Spatio-temporal water quality variations at various streams of Han-River watershed and empirical models of serial impound-

ment reservoirs. *Kor J Limnol*. 2012; 45(4): 378-391.

5. Kim JK, Shin M, Jang C, Jung S, Kim B. Comparison of TOC and DOC Distribution and the Oxidation Efficiency of BOD and COD in Several Reservoirs and Rivers in the Han River System. *J Kor Soc Water Qual*. 2007; 23(1): 72-80.
6. Lee HW, Choi JH. Temporal Analysis of Trends in Dissolved Organic Matter in Han River Water. *Environ Eng Res*. 2009; 14(4): 256-260.
7. Lee KS, Bong YS, Lee D, Kim Y, Kim K. Tracing the sources of nitrate in the Han River watershed in Korea, using δ^{15}N-NO$_3^-$ and δ^{18}O-NO$_3^-$ values. *Sci Total Environ*. 2008; 395(2-3): 117-124.
8. Nguyen H, VM, Lee MH, Hu J, Schlautman MA. Variations in spectroscopic characteristics and disinfection byproduct formation potentials of dissolved organic matter for two contrasting storm events. *J Hydrol*. 2013; 481: 132-142.
9. Wilson F. Total organic carbon as a predictor of biological wastewater treatment efficiency and kinetic reaction rates. *Water Sci Technol*. 1997; 35(8): 119-126.
10. Gwaski PA, Hati SS, Ndahi NP, Ogugbuaja, VO. Modeling parameters of oxygen demand in the aquatic environment of Lake Chad for depletion estimation. *ARPN J Sci Technol*. 2013; 3(1): 116-123.
11. Lee J, Lee S, Yu S, Rhew D. Relationships between water quality parameters in rivers and lakes: BOD5, COD, NBOPs, and TOC. *Environ Monit Assess*. 2016; 188(4): 252.
12. Lee HW, Choi JH. Temporal Analysis of Trends in dissolved organic matter in Han River water. *Environ Eng Res*. 2009; 14(4): 256-260.
13. Lee JY, Yang JS, Kim DK, Han MY. Relationship between land use and water quality in a small watershed in South Korea. *Water Sci Technol*. 2010; 62(11): 2607-2615.
14. Song ES, Jeon SM, Lee EJ, Park DJ, Shin YS. Long-term trend analysis of chlorophyll a and water quality in the Yeongsan river. *Kor J Limnol*. 2012; 45(3): 302-313.
15. Cho HS, Kim KR, Lim GC, Bae KS, Lee MH. A Study on long-term variations of BOD and COD as indicators of organic matter pollution in the Han River. *Kor J Limnol*. 2012; 45(4): 474-481.
16. Shin MS, Lee JY, Kim BC, Bae YJ. Long-term variations in water quality in the lower Han River. *J Ecol Environ*. 2011; 34(1): 31-37.
17. Venkatramanan S, Chung SY, Lee SY, Park N. (2014). Assessment of river water quality via environmentric multivariate statistical tools and water quality index: a case study of Nakdong River basin, Korea. Carpathian. *J Earth Environ Sci*. 2014; 9: 125-132.
18. Lim, et al. Evaluation of pollutant characteristics in Yeongsan River using multivariate analysis. *Kor J Ecol Environ*. 2012; 45(4): 368-377.
19. Jung, et al. Evaluation of water quality for the Nakdong River watershed using multivariate analysis. *Environ Technol Innovation*. 2016; 5: 67-82.
20. Khan TA. Groundwater quality evaluation using multivariate methods, in parts of Ganga Sot subbasin, Ganga basin, India. *J Water Resource Prot*. 2015; 7: 769-780.
21. Kim YY, Lee SJ. Evaluation of water quality for the Han River tributaries using multivariate analysis. *J KSSE*. 2011; 33(7): 501-510.
22. Berlemann A. Using a water quality index to determine and compare creek water quality. *Am Water Works Assoc*. 2013; 105(6): E291-E298.
23. Park J, Moon M, Lee H, Kim K. A study on characteristics of water quality using multivariate analysis in Sumjin River basin. *J Kor Soc Water Environ*. 2014; 30(2): 119-127.
24. Hammer O, Harper, DAT, Ryan, PD. PAST: Paleontological statistics software package for education and data analysis. *Palaeontologia Electronica*. 2001; 4(1): 9pp.
25. Sun R, Wang ZZ, Chen L, Wang W. Assessment of surface water quality at large watershed scale: Land-use, anthropogenic, and administrative impacts. *J Am Water Res Assoc*. 2013; 49: 741–752.
26. Kim JH. Assessment through statistical methods of water quality parameters (WQPs) in the Han River in Korea. *Kor J Environ Health. Sci*. 2015; 41(2): 90-101.
27. Yuan F, Quellos JA, Fan C. Controls of Phophorus loading and transport in the Cuyahoga River of northeastern Ohio, USA. *Appl Geochem*. 2013; 38: 59-69.
28. Chapman D. 1992. Water quality assessment: a guide of the use of biota, sediments and water in environmental monitoring. University Press, Cambridge, 1992; pp: 585.
29. An KG. Determination of a limiting nutrient regulating algal biomass using in situ experiments of nutrient enrichment bioassay (NEB) and empirical relations of nutrients and chlorophyll-a. *J Environ Biol*. 2003; 24(3): 229-239.
30. Kim JY, An KG. Integrated ecological river health

assessments, based on water chemistry, physical habitat quality and biological integrity. *Water*. 2015; 7: 6378-6403.

31. Ebise S, Inoue T. Change in C:N:P ratios during passage of water areas from rivers to a lake. *Water Res*. 1991; 25(1): 95-100.
32. Páll E, Niculae M, Kiss T, Şandru CD, Spînu M. Human impact on the microbiological water quality of the rivers. *J Med Microbiol*. 2013; 62(Pt 11): 1635-1640.
33. Frenzel SA, Couvillion CS. Fecal-indicator bacteria in streams along a gradient of residential development. *JAWRA*. 2002; 38: 265-273.
34. Kim JY, Lee H, Lee JE, Chung MS, Ko GP. Identification of human and animal fecal contamination after rainfall in the Han River, Korea. *Microbes Environ*. 2013; 28(2): 187-194.
35. Jeong YH, Kim HS, Yang JS. Statistical Analyses of Long-Term Water Quality Variation in the Geumgang-Reservoir: Focused on the TP load by migrating birds excrement. *J Kor Soc Mar Environ Engineer*. 2010; 13(4): 223-233.
36. WHO. 2008. Recommendations. Guidelines for Drinking-Water Quality. 3rd ed., Vol. 1. Geneva: World Health Organization.
37. Aziz A, Tebbutt THY. Significance of COD, BOD and TOC correlations in kinetic models of biological oxidation. *Water Res*. 1980; 14(4): 319-324.
38. Kim YY, Lee SJ. Evaluation of water quality for the Han River tributaries using multivariate analysis. *J KSSE*. 2011; 33(7): 501-510.
39. Tango T, Takahashi K. A flexible spatial scan statistic with a restricted likelihood ratio for detecting disease clusters. *Stat Med*. 2012; 31(30): 4207-4218.
40. Yeon YJ, Kim DH, Lee JL. Water quality modeling for integrated management of urban stream networks. *IJESD*. 2016; 7(12): 928-932.