Desalination and Water Treatment
www.deswater.com

doi: 10.1080/19443994.2015.1049963

57 (2016) 12508–12517
June

Taylor & Francis
Taylor & Francis Group

Enhanced monitoring of water quality variation in Nakdong River downstream using multivariate statistical techniques

Minsoo Kim[a], Yejin Kim[b], Hyosoo Kim[c], Wenhua Piao[a], Changwon Kim[a,*]

[a]*Department of Civil and Environmental Engineering, Pusan National University, Busan 609-735, Republic of Korea, Tel. +82 51 510 2769; emails: daniel@pusan.ac.kr (M. Kim), piaowenhua@pusan.ac.kr (W. Piao), Tel. +82 51 510 2769, +82 51 510 2416; Fax: +82 51 515 5347; email: cwkim@pusan.ac.kr (C. Kim)*
[b]*Department of Environmental Engineering, Catholic University of Pusan, Busan 609-757, Republic of Korea, Tel. +82 51 510 0621; email: yjkim@cup.ac.kr*
[c]*EnvironSoft Co., Ltd. #511 Industry-University Co., Bld., Pusan National University, Busan 609-735, Republic of Korea, Tel. +82 51 583 5351; email: h-sukim@pusan.ac.kr*

Received 10 December 2014; Accepted 5 May 2015

ABSTRACT

The variation in downstream river water quality was investigated using three multivariate statistical techniques: factor analysis (FA), cluster analysis (CA), and discriminant analysis (DA). Four main factors (FA1, FA2, FA3, and FA4) were defined as changes of "organic matter and nitrogen," "suspended solid and climate conditions," "phosphorous and electrical conductivity," and "discharge," respectively. The states of each factor were clustered into *Low*, *Normal* (*Normal_low* and *Normal_high*), and *High* groups using CA. These groups used to summarize water quality data measured as a series of numbers of contaminants for fast evaluation of water quality and enhanced monitoring capability. To set up a procedure for enhanced monitoring of water quality, Fisher's linear discriminant functions were deduced to determine the groups in which newly obtained water quality data should be included. To investigate the effectiveness of the proposed tool for enhanced monitoring of river water quality, a case study was conducted of the data analysis procedures applied to Nakdong River downstream and the monitoring results were examined.

Keywords: River water quality; Multivariate statistical techniques; Factor analysis; Cluster analysis; Discriminant analysis

1. Introduction

River water quality monitoring is so essential for ensuring healthy water usage that the automated monitoring systems or equipments have been distributed widely. For real-time checking of river water quality, many water quality monitoring systems have been continuously evaluated using various water quality databases.

To evaluate the river water quality variation, the pollution index (PI) method [1,2], fuzzy synthetic evaluation [3–5], and neural network method [6,7] have been investigated. However, the PI method suffers the disadvantage that the calculated water quality index often tends to be overestimated because of unnecessary correlations within objective measured parameters. In

*Corresponding author.

fuzzy synthetic evaluation, the weight value is calculated by monitoring data. However, the presence of abnormal monitoring data can prevent the determination of suitable weight values. The neural network method can provide excellent estimation accuracy, but the network structure can be complicated and analyzing the reasons for the deduced results can be difficult.

Multivariate statistical analysis, which is an effective tool to extract some useful information from a historical database or to reduce the dimensions, has been applied in various fields to investigate target characteristics. Various researchers have already used the methodology to assess the water quality of rivers or lakes with high dimensional data [8–14]. The existing cases were focused on investigating the target water system and showed successful data interpretation results. However, few studies have used multivariate statistical analysis to develop an algorithm generating useful target information. Due to its advantages in analyzing relationships between variables, multivariate statistical analysis can be applied as an enhanced monitoring tool that generates results in the form of linguistic information such as "under high nutrient loading," rather than a series of water quality data. Such results may be useful when water quality is checked by inexperience personnel who are not experienced with variations in such water quality data.

Therefore, in this study, the quantitative and qualitative variations of the river water quality were assessed using multivariable statistical techniques. To extract the water quality patterns in river downstream and to develop a tool capable of generating linguistic information for enhanced monitoring, factor analysis (FA), cluster analysis (CA), and discriminant analysis (DA) were used. FA was used to reveal the hidden factors and variable groups having the same factor. Then, each factor was clustered into groups indicating high, normal, or low loading. DA was used to deduce the Fisher's linear discriminant functions giving the final monitoring results of the current state. The water quality in river downstream was classified according to the changes of the organic matter, nitrogen, suspended solid (SS), climate condition, phosphorous, electrical conductivity (EC), and discharge. Based on these classifications, a case study of the enhanced monitoring of river water quality was conducted for Nakdong River downstream.

2. Materials and methods

2.1. Study area

The study area was downstream in the Nakdong River, which is located in the South Korea between 127°–129°E and 35°–37°N, as shown in Fig. 1. The Nakdong River is 506.17 km long with a basin area of 23,384.21 km^2 what comprises 24% of South Korea. The basin consists of 780 streams and 7 dams. Approximately 6.7 million people populate the basin, which is comprised of agricultural (23.52%), industrial (0.58%), commercial (0.24%), and forest (70.34%) areas. The total annual precipitation of the basin is approximately 1,200 mm, 60% of which falls from June to September. The monsoon climate and typhoons in the Korean Peninsula substantially affect the precipitation pattern. In addition, eight weirs were built in the Nakdong River basin to maintain the river capacity from December 2009 to January 2012. Therefore, the variations in the river's water quality were attributed to the changes in its hydraulic features.

The Nakdong River was geographically divided into three large streams [15]. Sixteen and twenty cities are located on the upstream and midstream, respectively. In particular, large-scale industrial estates are sited on the midstream section and the effluents from two wastewater treatment plants (WWTPs) are discharged in midstream. Therefore, high pollutant concentrations in upstream and midstream often flowed downstream and degraded the downstream. Moreover, the accumulated pollutants often caused the eutrophication in an estuary dam located downstream. The presence of these hazards necessitates continuous monitoring of water quality variations.

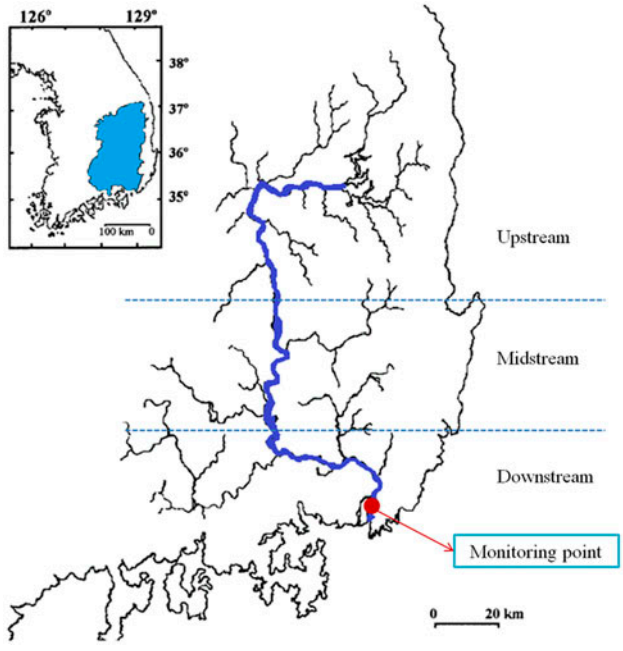

Fig. 1. The Nakdong River basin and the monitoring point, South Korea [27].

Table 1 lists the statistical information of variables in monitoring point from 2004 to 2010. The collected data were monthly average values that were obtained from the water environment information system, South Korea. The selected items were Q (discharge), pH, DO, BOD, COD, SS, T-N, NH_4-N, NOx-N, T-P, EC, *E coli* (coliform count), ST-N, ST-P, PO_4-P, and Chl-a (chlorophyll a), respectively. The RH (relative humidity) and rainfall were the meteorological data related to river discharge.

2.2. Algorithm for enhanced monitoring of the water quality variation

Fig. 2 shows a flowchart for the enhanced monitoring of the water quality variation in Nakdong River downstream. The procedure for generating monitoring results in the form of linguistic information was processed sequentially using FA, CA, and DA.

FA: Before FA was used, the data treatment process was applied for collected datasets. To verify the adequacy of the FA results, the Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy and Bartlett's test were performed. FA is more significant as the KMO value approaches 1 and is not significant if this value is less than 0.5. FA is also suitable when the significance probability is less than 0.05.

FA was used to extract the main factors representative of the entire variables, FA was used to identify the underlying main factors that explain the correlations among a dataset. Varimax rotation was used to prevent multiple variables from being loaded

Table 1
Statistical information of variables from 2004 to 2010

Variables	Units	*K*	Median	Maximum	Average	Standards deviation	Variance	Skewness	Krutosis
Q	m^3/s	289.30	436.68	182.60	743.96	720.32	518,861.11	2.32	4.39
pH	pH unit	6.40	7.60	9.20	7.70	0.69	0.48	0.21	−0.67
DO	mg/l	6.30	9.30	17.10	10.35	2.82	7.95	0.60	−0.90
BOD	mg/l	1.40	2.30	5.20	2.64	0.99	0.99	0.81	−0.39
COD	mg/l	3.80	5.90	9.90	6.09	1.24	1.55	0.78	0.48
SS	mg/l	6.90	14.65	112.50	18.37	15.30	234.18	3.86	18.60
T-N	mg/l	1.95	2.84	4.72	3.04	0.66	0.44	0.60	−0.52
NH_4-N	mg/l	0.01	0.10	0.53	0.13	0.12	0.01	1.77	3.10
NO_X-N	mg/l	0.95	1.74	3.07	1.86	0.48	0.23	0.59	−0.38
T-P	mg/l	0.08	0.13	0.31	0.14	0.04	0.00	1.52	4.63
EC	μmhos/cm	143.00	253.00	581.00	267.65	82.05	6,732.57	1.26	2.37
E coli	MPN[a]100 ml	0.00	9.00	485.00	39.40	89.02	7,925.33	3.68	13.99
ST-N	mg/l	1.41	2.46	3.83	2.60	0.51	0.26	0.44	−0.29
ST-P	mg/l	0.02	0.07	0.26	0.07	0.04	0.00	1.88	5.83
PO_4-P	mg/l	0.00	0.05	0.19	0.05	0.03	0.00	1.50	6.18
Chl-a	mg/m^3	7.40	33.10	182.60	55.15	47.40	2,246.99	1.33	0.58
Rainfall	mm	40.28	8.96	40.28	10.16	7.74	59.96	1.36	2.82
RH	%	37.00	61.10	85.00	61.38	12.72	161.92	0.01	−1.08

Note: where Q, EC, ST-N, ST-P, and RH mean discharge, electrical conductivity, soluble T-N, soluble T-P, relative humidity, respectively.

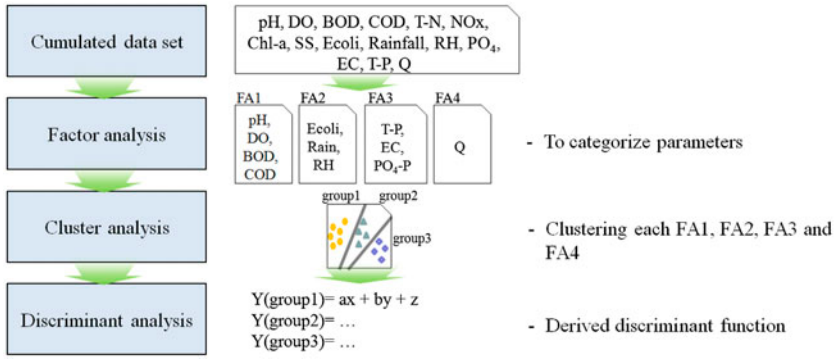

Fig. 2. Flowchart of the proposed tool for enhanced monitoring of water quality.

to a single factor, allowing for easy interpretation of the factor. Moreover, because the unit of each variable was different, the experimental data were standardized [16,17]. The major variables were selected as those with an impact degree value of 0.5 and over [16]. Next, the selected variables were used as the factors for the CA.

CA: To classify the unknown cluster for each factor, the *K*-means CA was used considering nonhierarchical cluster. It is a method for assigning similar clusters by comparing the distance between the center of each cluster and the measured values of selected variables. In this study, the Euclidean distance was calculated as follows:

$$d = \sqrt{\sum_{i=1}^{n} (p_i - q_i)^2} \quad (1)$$

where i is the ith measured value, n is the number of datasets, and p and q are the variables of p and q.

The process of *K*-means clustering was that a number, *K*, of data values was set in the centers and then the data values are allotted to the center whose data values were close to the number of data values. The values allotted in the center form a cluster. Once a cluster was formed with similar data values, the center of each other cluster was moved to another location. This process was repeated until the center remained unchanged using the objective function. This objective function is used for squared error and defined as follows [18]:

$$E = \sum_{i=1}^{k} \sum_{p \in c_i} |p - m_i|^2 \quad (2)$$

where E is the sum of the squared error for all measured values, p is a point in space for variables, and m_i is the center of cluster C_i. It was confirmed by the number of groups belonging in the cluster [19].

DA: DA was used to derive the discriminant function for identifying the groups that each factor belonged to. FA determined the most accurate data representation in a lower dimensional space [20]. However, the directions of maximum variance might be useless for classification. To solve this problem, Fisher's linear DA was used to preserve the directions that were useful for data classification [21]. The discriminant function for each group by Fisher's linear DA is defined as follows [22]:

$$f(G_i) = k_i + \sum_{j=i}^{n} w_{ij}\, p_{ij} \quad (3)$$

where $f(G_i)$ is the discriminant function for group i, k_i is a constant, n is the number of datasets, w_j is the weight value of the jth variable, and p_j is the measured value of the jth variable. The values of w_j and k_i are determined by linear function to maximize the between groups and within group variance. The raw data were used in DA because the coefficient sign of the discriminant was the allotted the weight per variable and could not be interpreted using another variable. The group of the current pollution state was determined using this derived discriminant function. All statistical computations were made using SPSS ver. 18.

2.3. Application of the developed monitoring tool for the new dataset

Fig. 3 shows the application flowchart of the developed enhanced monitoring tool for the new dataset. When the new dataset was obtained, the values calculated with the discriminant function were compared and as a result, the function having the highest value

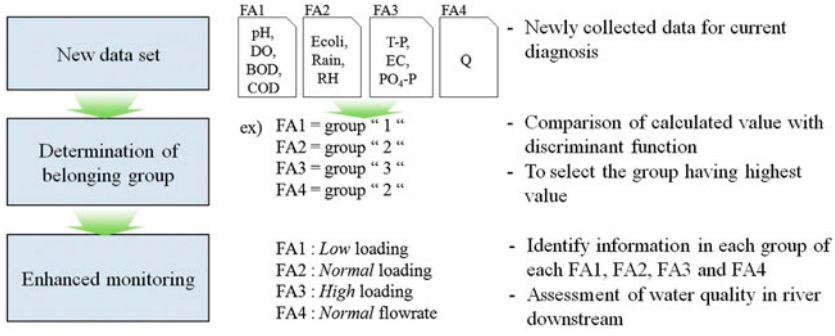

Fig. 3. Application flowchart of the enhanced monitoring tool developed using the new dataset.

in each factor (i.e., FA1, FA2, FA3, and FA4) was determined by the belonging group. Finally, the water quality variation for the new dataset could be accessed using the developed monitoring tool based on the multivariable statistical techniques.

3. Results and discussion

3.1. FA results and definition of classified factors

The four main factors were extracted using FA in terms of the 16 variables. In order to verify the goodness of the applied variables, the KMO measure and Bartlett's test were evaluated. All variables were found to be suitable for FA because the KMO value was 0.81 with a significance probability of 0.00. Table 2 shows the impact degree of each variable on each factor. The selected variables having impact degree of over 0.5 are indicated in bold letters.

Table 2
Impact degrees of each variable derived by FA

Variables	FA1	FA2	FA3	FA4
pH	**.765**	−.383	−.028	−.248
DO	**.887**	−.344	−.091	−.076
BOD	**.912**	−.026	−.081	−.225
COD	**.834**	.279	.186	.034
SS	.004	**.610**	.134	.662
T-N	**.819**	−.086	.287	−.118
NH_4-N	.076	.049	.443	−.546
NO_x-N	**.824**	−.195	.054	.027
T-P	.147	.255	**.876**	.108
EC	.170	−.495	**.647**	−.193
E coli	−.067	**.671**	−.069	.122
PO_4-P	−.541	.139	**.698**	−.165
Chl-a	**.925**	−.095	−.046	−.052
Q	−.195	.037	−.019	**.813**
Rainfall	−.226	**.720**	.211	−.142
RH	−.603	**.643**	.190	.074
Eigenvalue	7.45	2.41	2.12	1.05
(%) Total variance	43.84	14.20	12.47	6.18
Cumulative (%) variance	43.84	58.04	70.51	76.69

The characteristics in the Nakdong River downstream were classified into four factors, which explained 76.7% of the total variance. FA1 was selected as pH, DO, BOD, COD, T-N, NO_X-N, and Chl-a. From this result, the characteristic of FA1 was defined as "Changes of organic matter and nitrogen." By the same definition, the FA2, FA3, and FA4 were extracted by the characteristics of belonging variables and defined as "Changes of SS and climate conditions," "Changes of phosphorous and EC," and "Change of discharge," respectively. Table 3 shows the definition of each factor and belonging variables.

3.2. Group classification for the each factor using CA

Based on the variables that were selected using the FA, the pollution state of each factor was grouped using CA. As shown in Table 4, the each factor was divided into the groups of qualitative pollution levels. The four belonging groups for FA1 were defined as *Low*, *Normal_Low*, *Normal_High*, and *High* loading, respectively. The three belonging groups for each of FA2 and FA3 were defined as *Low*, *Normal*, and *High* loading. The groups for FA4 were defined as *Low*, *Normal*, and *High* flowrate due to the single variable.

3.3. Discriminant functions of the belonging groups for each FA

The Fisher's linear DA was applied to the derived discriminant function. When the newly obtained dataset was used for generating linguistic information of the water quality variation, each discriminant function was firstly calculated. Then, the group having the highest value among the calculated values was selected as the defined pollution state for each FA.

FA1: Changes of organic matter and nitrogen

$$\text{Group 1 } (Low) = 58.14 \times \text{pH} - 5.15 \times \text{DO} - 16.01 \times \text{BOD} + 16.41 \times \text{COD} + 5.15 \times \text{T-N} + 19.43 \times NO_x\text{–N} - 0.36 \times \text{Chl-a} - 239.10$$

Table 3
Definitions and belonging variables extracted using FA

	FA1	FA2	FA3	FA4
Definition	Changes of organic matter and nitrogen	Changes of SS and climate conditions	Changes of phosphorous and electrical conductivity	Change of discharge
Variables included	pH, DO, BOD, COD, T-N, NO_x-N, Chl-a	SS, *E coli*, Rainfall, RH	T-P, EC, PO_4-P	Q

Table 4
Information of pollution state based on cluster center value in each factor

FA1: Changes of organic matter and nitrogen	pH	DO	BOD	COD	T-N	NO_X-N	Chl-a
Group 1 (*Low* loading)	7.30	8.48	1.96	5.49	2.71	1.59	23.46
Group 2 (*Normal_Low* loading)	7.74	10.71	2.79	5.82	2.96	1.91	49.88
Group 3 (*Normal_High* loading)	8.45	13.64	3.83	7.31	3.79	2.30	109.95
Group 4 (*High* loading)	8.81	15.08	4.34	8.26	3.96	2.57	163.74

FA2: Changes of SS and climate conditions	SS	*E coli*	Rainfall	RH
Group 1 (*Low* loading)	13.95	4.26	6.90	52.58
Group 2 (*Normal* loading)	12.65	21.50	5.69	56.22
Group 3 (*High* loading)	24.33	6.13	14.08	74.55

FA3: Changes of phosphorous and electrical conductivity	T-P	EC	PO_4-P
Group 1 (*Low* loading)	0.15	165.50	0.06
Group 2 (*Normal* loading)	0.12	226.91	0.04
Group 3 (*High* loading)	0.14	282.10	0.04

FA4: Change of discharge	Q
Group 1 (*Low* flowrate)	339.82
Group 2 (*Normal* flowrate)	420.70
Group 3(*High* flowrate)	536.49

Group 2 (*Normal_Low*) = 57.93 × pH − 5.20 × DO − 13.66 × BOD + 15.80 × COD + 5.00 × T-N + 20.07 × NO_x−N − 0.19 × Chl-a − 246.35

Group 3 (*Normal_High*) = 59.63 × pH − 6.26 × DO − 14.66 × BOD + 16.37 × COD + 9.92 × T-N + 16.16 × NO_x−N + 0.34 × Chl-a − 298.58

Group 4 (*High*) = 60.52 × pH − 7.68 × DO − 16.20 × BOD + 16.72 × COD + 11.29 × T-N + 15.13 × NO_x−N + 0.87 × Chl-a − 356.91

FA2: Changes of SS and climate conditions

Group 1 (*Low*) = 0.22 × SS + 0.34 × Ecoli − 0.21 × Rainfall + 0.94 × RH − 27.34

Group 2 (*Normal*) = 0.12 × SS + 1.00 × Ecoli − 0.37 × Rainfall + 1.11 × RH − 42.80

Group 3 (*High*) = 0.39 × SS + 0.44 × Ecoli − 1.29 × Rainfall + 1.30 × RH − 54.59

FA3: Changes of phosphorous and electrical conductivity

Group 1 (*Low*) = 110.51 × T-P + 0.64 × EC + 91.14 × PO_4−P − 65.10

Group 2 (*Normal*) = 87.42 × T-P + 0.87 × EC + 79.77 × PO_4−P − 107.06

Group 3 (*High*) = 103.61 × T-P + 1.08 × EC + 79.27 × PO_4−P − 162.80

FA4: Change of discharge

Group 1 (*Low*) = 0.45 × Q − 76.89

Group 2 (*Normal*) = 0.55 × Q − 117.25

Group 3 (*High*) = 0.70 × Q − 189.99

3.4. Enhanced monitoring of the water quality using the accumulated dataset

Fig. 4 shows the group variation in each FA for the cumulated dataset. There were 45, 19, 12, and 8 cases of the four groups *Low*, *Normal_Low*, *Normal_High* loading in FA1, respectively. For the three groups *Low*, *Normal*, and *High*, there were 35, 33, and 16 cases in FA2, 12, 33, and 39 cases in FA3, respectively. Finally, the cases of group *Low*, *Normal*, *High* flowrate in FA4 were 20, 32, and 32, respectively. This result confirmed that the cases of each group in each FA were properly divided for the cumulated dataset.

FA1 was appeared as a characteristic of the periodic variation for every year, because the belonging variables were affected sensitively by seasonal condition. Similarly, the low loading was observed by dilution effect according to heavy rain during summer. In addition, the pollutants in the upstream and midstream were flowed into the Nakdong River downstream. Consequently, the high loading in FA1 was appeared during the dry season from November to March. In the case of FA2, the RH ratio and SS con-

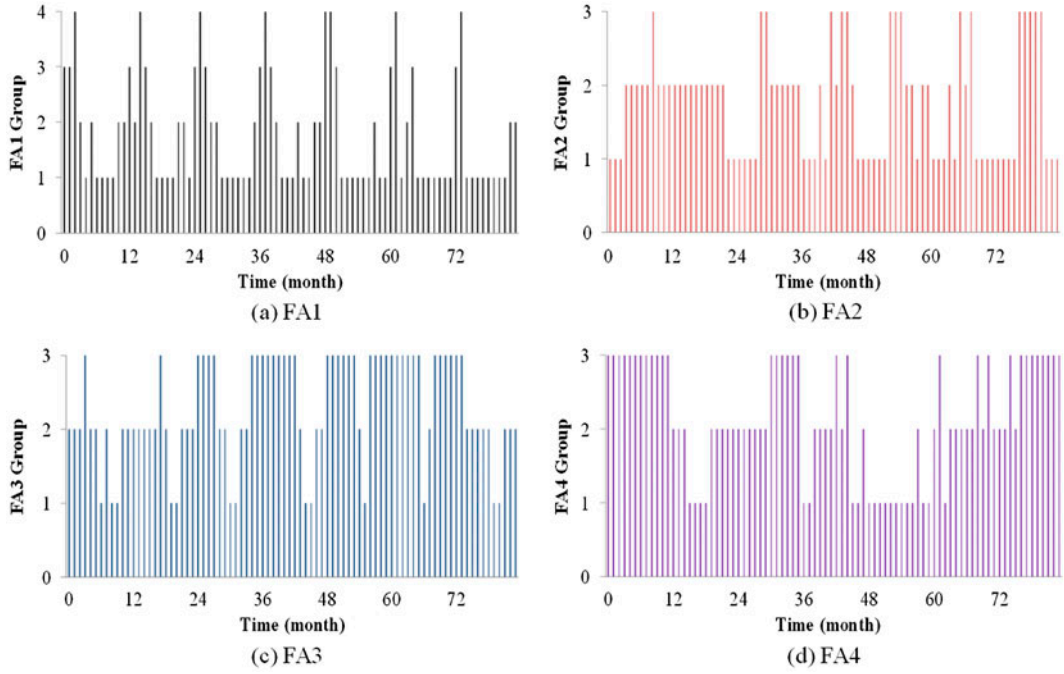

Fig. 4. Group variations for the cumulated dataset of 2004–2010 (for FA1, 1 = Low; 2 = Normal_Low; 3 = Normal_High; 4 = High loading and for FA2, FA3, and FA4, 1 = Low; 2 = Normal; 3 = High).

centration were increased by heavy rainfall during the summer season, so that the group 3 which defined as high loading, appeared repeatedly. Moreover, the high loading was also affected by the growth of *E coli* due to the presence of the stagnant water and air temperature. The belonging variables in FA3 were T-P, EC, and PO_4-P, respectively, and the T-P was positive correlation to SS. Therefore, the high loading was appeared during the summer season. In addition, the EC value was increased by the backflow of saltwater because of the current density and tidal distribution due to the gate operation at Nakdong estuary during the summer season. Therefore, the pollution state in FA3 was affected by the EC variation [23,24]. Finally, the pollution state in FA4 was affected by changes of climate condition and river capacity.

3.5. Application results of the developed enhanced monitoring tool for the new dataset

Fig. 5 shows the application results of the developed monitoring tool for the newly obtained dataset. The groups for the each FA were confirmed using the dataset obtained in 2011 and 2012. In FA1, group 2 (*Normal_Low*), which was defined as the moderately low loading for the organic matter and nitrogen, appeared during January to March in 2011, because the pollutants flowed into the river downstream from the upstream and midstream sections during the dry season with low flowrate. After March, the pollution state was continuously identified as low loading during April to December. A similar pattern was identified in 2012. When the group variations between the cumulated dataset (Fig. 4(a)) and new dataset (Fig. 5(a)) were compared in the FA1, the annual patterns of group in 2010–2012 were different to those in 2004–2009, because of a river restoration project and the construction of weirs in the Nakdong River basin from December 2009 to January 2012. When the project and the construction were finished, the river capacity was increased, which diluted the pollutants in the river downstream. On the other hand, as shown in Fig. 5(a), the high loading was appeared at the beginning of the year upon completion of the project in 2012. However, the low loading was continuously observed due to stabilization of the river and improved self-purification during April to December in 2012.

For FA2, the SS concentration was increased by the construction work of the river restoration project in 2011. As a result, the group 3 (*High*), which was defined as the high solids loading, appeared in throughout 2011, as shown in Fig. 5(b). However, when the river restoration project was finished in January 2012, the settling of SS at the bottom of the river downstream led to the appearance of group 2,

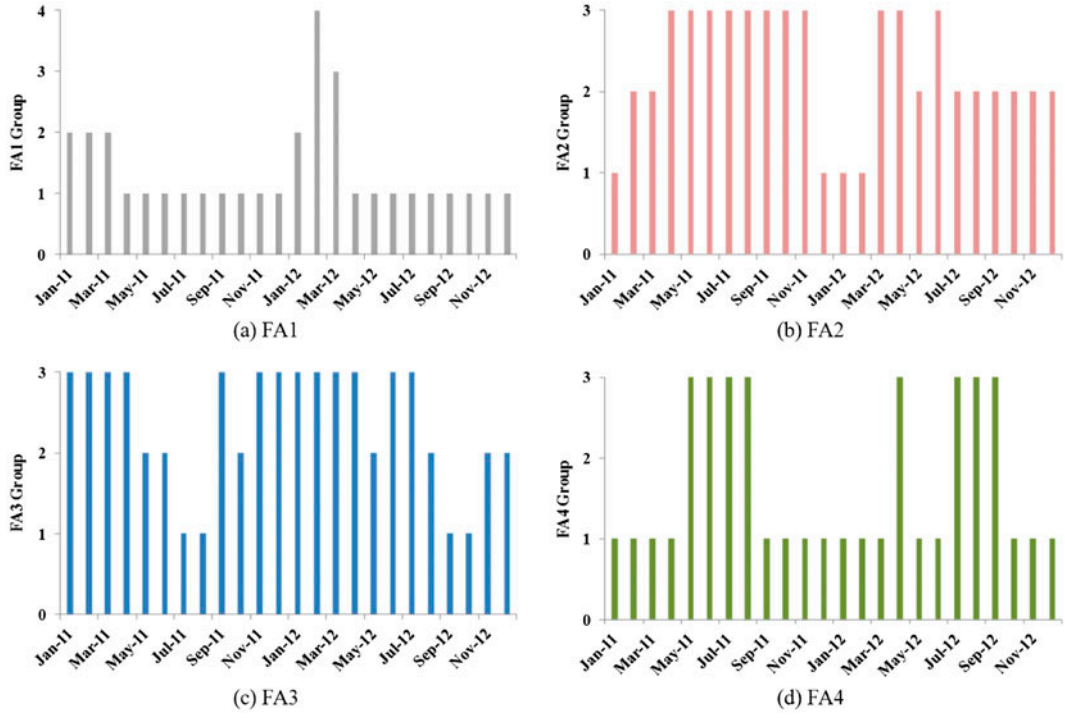

Fig. 5. Group variations for the newly obtained data in 2011 and 2012 (for FA1, 1 = Low; 2 = Normal_Low; 3 = Normal_High; 4 = High loading and for FA2, FA3, and FA4, 1 = Low; 2 = Normal; 3 = High).

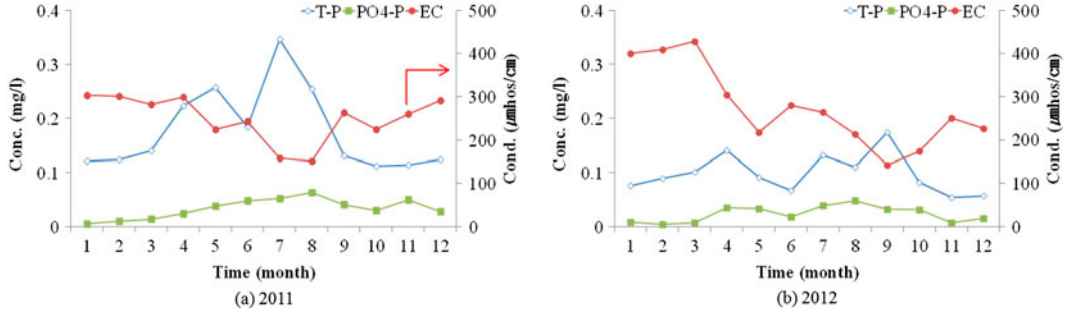

Fig. 6. Water quality variation on T-P, PO_4-P, EC in 2011 and 2012.

which was defined as the *normal* loading was appeared in 2012.

For FA3, group 3 (*High*), which was defined as the *high* loading of phosphorous and EC, appeared mostly in 2011. As shown in Fig. 6, the pollution state of FA3 was strongly influenced by EC because the other variables, T-P, and PO_4-P concentration, were observed at low concentrations. As mentioned above, the backflow of saltwater was flowed into the river downstream during the summer season, which increased the EC concentration. However, the *normal* and *low* loading appeared from August 2012 due to low EC concentration because of the stabilization of the upstream river water level by the completed weirs [25]. The range of fluctuation of water level was decreased by the weirs after the barrage construction in the Nakdong River basin [26]. As a result, the gate operation of estuary was more reduced than in 2011 due to the stabilization of water level and the backflow of saltwater was not diffused by the monitoring point.

Finally, FA4 was influenced by the change of discharge. Therefore, the high flowrate was appeared in summer due to the heavy rainfall and the low flowrate was observed for the rest of the year, as shown in Fig. 5(d). FA4 exhibited a different pattern between the accumulated datasets and the new datasets. The discharge pattern of the accumulated datasets was irregular, whereas that of the new datasets appeared repeatedly as *high* flowrate and *low* flowrate during two years. This was because FA4 was affected by increasing

the river cross section and discharge rates due to the completion of the river restoration project.

4. Conclusion

In this study, three multivariate statistical techniques, i.e., FA, CA, and DA were used for enhanced monitoring of the water quality variation in the Nakdong River downstream. The four main factors were extracted and defined as "Changes of organic matter and nitrogen," "Changes of SS and climate conditions," "Changes of phosphorous and EC," and "Changes of discharge," respectively. The pollution states of each factor were grouped as *low*, *normal* (*normal_low* and *normal_high*), and *high* loading using CA. The discriminant function was derived by DA to decide the belonging group for each factor.

In the water quality data collected in 2011 and 2012 with the new characteristics, the current pollution state of the river downstream could be successfully assessed using the developed tool based on the multivariable statistical techniques. The current pollution states of FA1 and FA2 were influenced by the river restoration project, which affected the river capacity and SS concentration in the Nakdong River basin. The pollution state of FA3 was influenced by the change of the EC, which was affected by the backflow of saltwater in the river downstream. In FA4, the difference of the discharge patterns between the cumulated dataset and the new dataset was influenced by the river restoration project.

An enhanced monitoring tool based on multivariate statistical techniques was developed in this study to assess the water quality variation in the Nakdong River downstream and to generate some useful linguistic information about water quality level for personnel who are not familiar with the relevant data. The developed tool can assist the decision-making of gate operators and water resource managers. The operators of WWTPs can use the information provided by this enhanced monitoring system to optimize their plant operational condition. Moreover, the WWTPs discharging their effluent into the river water being monitored by this algorithm can help to recognize the relationship between their effluent water quality and that of the river.

The monitoring results, obtained by application of the newly measured data, were interpreted by considering the geographical and topographical characteristics in the Nakdong River downstream. Application of the tool developed in this study to other river basin for assessing the water quality variations will facilitate an assessment of the entire river basin using various monitoring datasets. Additional variables and dataset may also be required for investigating the characteristic of the target river basin more effectively. Moreover, when the accumulated datasets are used for a long-term calibration and validation of the developed tool, the water quality variation in the target river basin can be assessed more accurately.

Acknowledgments

This research was supported by the Korea Ministry of Environment as "The Eco-innovation project." This work was also financially supported by the second stage of the Brain Korea 21 Project in 2013.

References

[1] S.G. Nives, Water quality evaluation by index in Dalama, Water Res. 33(16) (1999) 3423–3440.
[2] S.F. Pesce, D.A. Wunderlin, Use of water quality indices to verify the impact of Córdoba City (Argentina) on Suquía River, Water Res. 34(11) (2000) 2915–2926.
[3] N.B. Chang, H.W. Chen, S.K. Ning, Identification of river water quality using the fuzzy synthetic evaluation approach, J. Environ. Manage. 63(3) (2001) 293–305.
[4] R.S. Lu, S.L. Lo, Diagnosing reservoir water quality using selforganizing maps and fuzzy theory, Water Res. 36(9) (2002) 2265–2274.
[5] Z.H. Zou, Y. Yuan, J.N. Sun, Entropy method for determination of weight of evaluating indicators in fuzzy synthetic evaluation for water quality assessment, J. Environ. Sci. 18(5) (2006) 1020–1023.
[6] G.R. Shetty, H. Malki, S. Chellam, Predicting contaminant removal during municipal drinking water nanofiltration using artificial neural networks, J. Membr. Sci. 212(1–2) (2003) 99–112.
[7] P. Chaves, T. Kojiri, Deriving reservoir operational strategies considering water quantity and quality objectives by stochastic fuzzy neural networks, Adv. Water Res. 30(5) (2007) 1329–1341.
[8] W. Petersen, L. Bertino, U. Callies, E. Zorita, Process identification by principal component analysis of river water-quaity data, Ecol. Modell 138 (2001) 193–213.
[9] K.W. Chau, N. Muttil, Data mining and multivariate statistical analysis for ecological system in coastal waters, J. Hydroinformatics. 9(4) (2007) 305–317.
[10] S.K. Singh, C.K. Singh, K.S. Kumar, R. Gupta, S. Mukherjee, Spatial temporal monitoring of groundwater using multivariate statistical techniques in Bareilly District of Uttar Pradesh, India, J. Hydrol. Hydromech. 57(1) (2009) 45–54.
[11] T.G. Kazi, M.B. Arain, M.K. Jamali, N. Jalbani, H.I. Afridi, R.A. Sarfraz, J.A. Baig, A.Q. Shah, Assessment of water quality of polluted lake using multivariate statistical techniques : A case study, Ecotoxicol. Environ. Saf. 72 (2009) 301–309.
[12] W.C. Liu, H.L. Yu, C.E. Chung, Assessment of water quality in a subtropical alpine lake using multivariate

statistical techniques and geostatistical mapping: A case study, Int J. Environ. Res. Publ. Health 8(4) (2011) 1126–1140.
[13] R. Koklu, B. Sengorur, B. Topal, Water quality assessment using multivariate statistical methods—A case study: Melen river system (Turkey), Water Resour. Manage. 24 (2010) 959–978.
[14] V. Gvozdic, J. Brana, N. Malatesti, D. Roland, Principal component analysis of surface water quality data of the River Drava in eastern Croatia (24 year survey), J. Hydroinformatics. 14(4) (2012) 1051–1060.
[15] J.W. Noh, J.C. Kim, J.H. Park, Turbidity control in downstream of the reservoir: The Nakdong River in Korea, Environ. Earth Sci. 71(4) (2014) 1871–1880.
[16] C.W. Liu, K.H. Lin, Y.M. Kuo, Application of factor analysis in the assessment of groundwater quality in a blackfoot disease area in Taiwan, Sci. Total Environ. 313 (2003) 77–89.
[17] V. Simeonov, J.A. Stratis, C. Samara, G. Zachariadis, D. Voutsa, A. Anthemidis, M. Sofoniou, T. Kouimtzis, Assessment of the surface water quality in Northern Greece, Water Res. 37 (2003) 4119–4124.
[18] A.K. Jain, Data clustering: 50 years beyond K-means, Pattern Recognit. Lett. 31(8) (2010) 651–666.
[19] H.S. Kim, T.S. Moon, Y.J. Kim, M.S. Kim, W.H. Piao, S.J. Kim, C.W. Kim, Evaluation of rule-based control strategies according to process state diagnosis in A2/O process, Chem. Eng. J. 222 (2013) 391–400.
[20] S. Shrestha, F. Kazama, Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin, Jpn Environ. Model. Softw. 22(4) (2007) 464–475.
[21] T.S. Moon, Y.J. Kim, J.R. Kim, J.H. Cha, D.H. Kim, C.W. Kim, Identification of process operating state with operational map in municipal wastewater treatment plant, J. Environ. Manage. 90(2) (2009) 772–778.
[22] K.P. Singh, A. Malik, D. Mohan, S. Sinha, Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India): A case study, Water Res. 38 (2004) 3980–3992.
[23] C.S. Han, S.K. Park, S.W. Jung, T.Y. Roh, The Study of Salinity Distribution at Nakdong River Estuary, J. Korean Soc. Coastal Ocean Eng. 23(1) (2011) 101–108.
[24] J.I. Song, B.U. Yoon, J.W. Kim, C.W. Lim, S.B. Woo, Spatial and temporal variability of residual current and salinity distribution, J. Korean Soc. Coastal Ocean Eng. 26(3) (2014) 184–195.
[25] G.B. Kim, E.J. Cha, H.G. Jeong, K.H. Shin, Comparison of time series of alluvial groundwater levels before and after Barrage construction on the lower Nakdong River, J. Eng. Geol. 23(2) (2013) 105–115.
[26] M.J. Kim, K.T. Min, K.S. Jun, Operation of Estuary Barrage and Weirs in the Nakdong River during the Flood Period, J. Korean Soc. Hazard. Mitigation 14(4) (2014) 289–299.
[27] K. Ha, H.-W. Kim, G.-J. Joo, The phytoplankton succession in the lower part of hypertrophic Nakdong River (Mulgum), South Korea, Hydrobiologia 369–370 (1998) 217–227, doi: 10.1023/A:1017067809089.