Environmental Technology & Innovation 5 (2016) 67–82

Contents lists available at ScienceDirect

Environmental Technology & Innovation

journal homepage: www.elsevier.com/locate/eti

Evaluation of water quality for the Nakdong River watershed using multivariate analysis

Kang Young Jung [a], Kyung-Lak Lee [a], Toe Hyo Im [a], In Jung Lee [a], Shin Kim [a], Kun-Yeun Han [b], Jung Min Ahn [a,*]

[a] *Nakdong River Environment Research Center, 24, Pyeongri-1gil, Dasan-myeon, Goryeong-gun, Gyeongsangbuk-do 717-873, Republic of Korea*

[b] *Department of Civil Engineering, Kyungpook National University, Republic of Korea*

HIGHLIGHTS

- The water quality characteristics were evaluated by conducting multivariate statistical analysis.
- The first factors were shown to be nutrient salts (TN, TP) and organic pollutants (TOC, COD).
- The priorities of required water quality were determined according to the water quality pollution level.

ARTICLE INFO

Article history:
Received 8 May 2015
Received in revised form 30 November 2015
Accepted 16 December 2015
Available online 21 December 2015

Keywords:
The Nakdong River
Factor analysis
Cluster analysis
Principal component analysis
Water quality
Water quality pollution level

ABSTRACT

Water quality observation data were collected from 20 representative monitoring sites located in the main stream of the Nakdong River and its major tributaries between 2008 and 2012. Based upon these data, the water quality distribution and characteristics of each river were evaluated by conducting multivariate statistical analysis for 12 pollution indicators using SPSS-17.0. Monitoring site T8 passing through the city of Daegu and monitoring site M5 located downstream of the Nakdong River exhibited high-concentration tendencies. The monitoring sites located near the city and midstream and downstream of the Nakdong River exhibited high pollution levels in the investigation. To analyze the spatial and temporal variations in the water quality at 20 major monitoring sites in the Nakdong River watershed, principal component and factor analyses were conducted by separating the average water quality data based upon (a) monitoring site, and (b) season. As a result, three factors were obtained for (a), and (b), respectively. In the Nakdong River watershed, the first factor was shown to be nutrient salts (total nitrogen and total phosphorus) and organic pollutants (total organic carbon and chemical oxygen demand), and as a result of cluster analysis, two statistically significant groups were classified. The results of multivariate statistical analysis indicated that the monitoring sites with high levels of pollution were mostly those sites going through the heart of the city or the sites affected by residential sewage directly, as well as the sites located midstream and downstream of the Nakdong River. The water quality pollution level was calculated based upon the above study results, and the priorities for water quality improvement items required in future watershed management were determined in order to facilitate efficient water quality management.

* Corresponding author. Tel.: +82 54 950 9780; fax: +82 54 950 9775.
E-mail addresses: happy3313@korea.kr (K.Y. Jung), micow1022@korea.kr (K. Lee), imtoehyo@korea.kr (T.H. Im), ijlee5344@korea.kr (I.J. Lee), sinima@korea.kr (S. Kim), kshanj@knu.ac.kr (K.-Y. Han), jahn@korea.kr (J.M. Ahn).

http://dx.doi.org/10.1016/j.eti.2015.12.001

1. Introduction

A river is a system comprising both the main course and the tributaries, carrying the one-way flow of a significant load of matter in dissolved and particulate phases from both natural and anthropogenic sources (Shrestha and Kazama, 2007). Since rivers are the most important inland water resources for human consumption, it is imperative to have reliable information on characteristics and trends of water quality for effective water management. The quality of a river at any point reflects several major influences, including the lithology of the basin, atmospheric inputs, climatic conditions and anthropogenic inputs (Bricker and Jones, 1995).

There are also several types of pollution indicators derived from these influences. However, up to now studies have only examined the pollution levels of plain watersheds based upon subjective pollution evaluations and mathematical water quality analyses regarding various pieces of information. There are many pollution sources affecting water quality, and there is a methodology that can provide basic data to accurately support this view. However, this methodology has not been utilized, and in most cases, the analysis and generalization were influenced by the subjectivity and experience of the researchers. Therefore, the effective, long-term management of rivers requires a fundamental understanding of watershed characteristics through multivariate analysis for the water quality monitoring data. Due to spatial and temporal variations in water quality, a monitoring program, providing a representative and reliable estimation of the quality of surface waters, is necessary (Dixon and Chiswell, 1996).

The application of different multivariate statistical techniques, such as cluster analysis (CA), principal component analysis (PCA), factor analysis (FA) and discriminant analysis (DA), helps in the interpretation of complex data matrices to better understand the water quality and ecological status of the studied systems, allows the identification of possible factors/sources that influence watershed, and offers a valuable tool for reliable management of water resources as well as rapid solution to pollution problems (Vega et al., 1998; Lee et al., 2001; Adams et al., 2001; Wunderlin et al., 2001; Reghunath et al., 2002; Simeonova et al., 2003; Simeonov et al., 2004). Multivariate statistical techniques has been applied to characterize and evaluate surface and freshwater quality, and it is useful in verifying temporal and spatial variations caused by natural and anthropogenic factors linked to seasonality (Helena et al., 2000; Singh et al., 2004). Abahams (1972) conducted a factor analysis to identify the characteristics of drainage watersheds, and Karim and Taha (2003) used principal component analysis to identify the commonality between the spatial and temporal factors influencing the water quality in New Jersey's Passaic River. To understand the nutrition state of lakes, Bernard et al. (2004) used principal component analysis for the evaluation and management of relevant water quality items. Fang et al. (2010) used a statistical method for spatial fluctuation characteristics of water pollution in Qiantang River, and Homa et al. (2010) conducted a study on spatial and temporal variability by using a pattern recognition method.

The water quality of the river is determined by interactions of pollution loads from the tributary flowing in, hydrological characteristics, sediment and metabolism in the water, and seasonal factors adjusted spatially and temporally. Therefore, to identify the characteristics of the Nakdong River watersheds, this study analyzed the spatial water quality distribution characteristics using the position and the changing characteristics of water quality according to time by season.

In this way, this study aims to analyze the major factors recently affecting the water quality spatially and temporally in the watersheds of the Nakdong River, and to provide data for more scientific and rational water quality management in response to the increases in the causes of environmental pollution and the water demand due to the effects of developments performed based on more complex and broader surrounding environments.

2. Materials and methods

2.1. Study area

The Nakdong River watershed is located between the longitude 127° 29′19″ and 129°18′00″, and the latitude 34°59′41″ and 37°12′52″. The Nakdong River faces the watersheds of Han River to the north and Geum River and Seomjin River to the west, the watersheds of the East Sea coast are formed by the Taebaek Mountains in the east, and it faces the southern sea area of the Nakdong River in the south (Jung, 2009).

The administrative districts include areas from three metropolitan cities and five provinces, including Busan Metropolitan City, Daegu Metropolitan City, Ulsan Metropolitan City, Gyeongsang Nam and Buk Provinces, Jeolla Nam and Buk Provinces, and Gangwon Province. The Nakdong River watershed is the second largest watershed in South Korea. The watershed's area is 23,817.3 km^2 (which is about 1/4 of the area of South Korea), the length of the channels is 521.5 km, the circumference of the watershed is 1097.13 km, the average width is 46.03 km, the average elevation is 291.2 m, and the average slope is 32.26%. The watershed consists of a total of 803 streams, including the watershed of West Nakdong River (758 Nakdong River watershed and 18 West Nakdong River watershed), and the total length of the river is 7422.02 km.

The major streams of the Nakdong River's watershed include thirteen national rivers (Nakdong River, Naeseong Creek, Gam Creek, Geumho River, Hwang River, Nam River, Deokcheon River, Miryang River, Yangsan Creek, West Nakdong River, Pyeonggang Creek, and Maekdo River), ten regional class 1 rivers (Nakdong River, Banbyeon Creek, Byeonseng Creek, Naeseong Creek, We Creek, Shin Creek, Hwea Creek, Geochangwi Creek, Nam River, and Hamyangwi Creek), and 31 regional

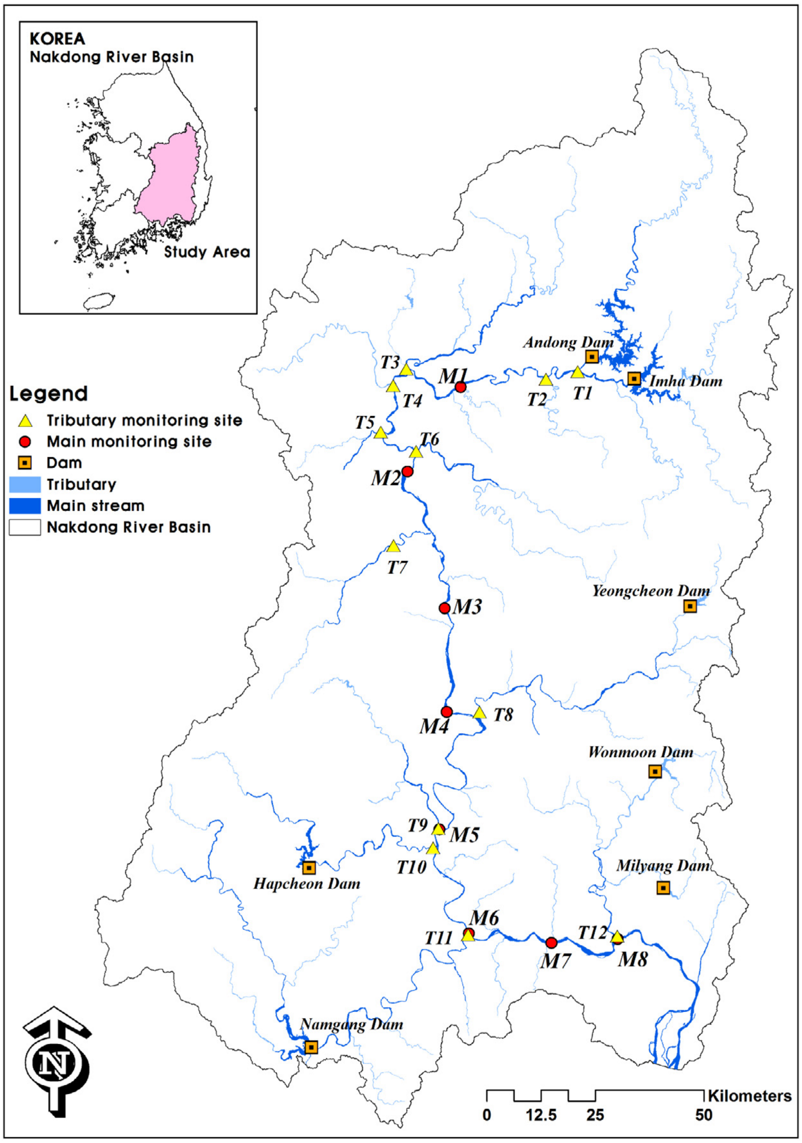


Fig. 1. Monitoring locations for water quality and discharge monitoring in the Nakdong River watershed, Korea.

class 2 rivers with a watershed area of 100 km^2, which join with the national rivers and regional class 1 rivers. The annual average precipitation of the Nakdong River watershed from 2000 to 2010 was 1326.2 mm, which is similar to the annual average precipitation of South Korea (1971–2000), and the coefficient of the river regime is 1:372, which shows a large runoff characteristic compared to the rivers of other countries.

The soil is mainly composed of sedimentary rocks; metamorphic and igneous rocks are partially distributed. The land usage includes an agricultural land area of 4372 km^2, which accounts for 18.4% of the total watershed area. From this, farm fields account for 7.0%, rice paddies for 11.4%, land and industrial complex area 1.9%, and woodland 69.6% (Kim et al., 2012). The upstream area of the Nakdong River is comprised mostly of rural towns. Inland of the midstream area, Andong, Gumi, Miryang, Daegu, etc. have a population of 3.5 million, and industries dealing with electronics, textile, and automobile parts are being operated. Located in the downstream area are Gimhae, Miryang, and Yangsan, as well as Busan, the second largest city of South Korea. This area has a population of 5 million, and it is a center for logistics as well as the machine industry. Thus, this area requires continuous water quality management and monitoring because it is directly affected by various forms of industrial and residential waste water (Park, 2010).

As shown in Fig. 1, the water quality and river flow were monitored for the representative monitoring sites located at the main stream of the Nakdong River and its major tributaries, and multivariate statistical analysis was conducted.

Table 1
River (Stream) water monitoring sites investigated.

No.	Code	Stream (River) name	Monitoring site name	Municipality
1	M1	Nakdong River	Andong 5	
2	M2	Nakdong River	Sangju 3	
3	M3	Nakdong River	Gumi	Gumi city
4	M4	Nakdong River	Yongam	
5	M5	Nakdong River	Daeam-1	
6	M6	Nakdong River	Yongsan	
7	M7	Nakdong River	Bukmyeon	
8	M8	Nakdong River	Samlangjin A	
9	T1	Banbyeon stream	Banbyeon stream 2-1	
10	T2	Mi stream	Mi stream	
11	T3	Naeseong stream	Naeseong stream 3-1	Yeongju city
12	T4	Young River	Young River 2-1	Mungyeong city
13	T5	Byeongseong stream	Byeongseong stream -1	Sangju city
14	T6	We stream	We stream 6	
15	T7	Gam stream	Gam stream 1A	Gimcheon city
16	T8	Geumho River	Geumho River 6	Daegu metropolitan city
17	T9	Hoi stream	Hoi stream 2-1	
18	T10	Hwang River	Hwang River 5	
19	T11	Nam River	Nam River 4-1	Jinju city
20	T12	Miryang River	Milyang River 3	Miryang city

2.2. Monitored parameters and analytical methods

In this study, water quality was monitored for a total of five years (from 2008 to 2012) at the end of a tributary (12 monitoring sites) and the main stream (8 monitoring sites) in the Nakdong River watershed, and the acquired data were used in multivariate analysis. The measurement of river flow followed the "River Flow Measurement Guidelines", and the measurement of water quality followed the "Water Quality Pollution Process Test Criteria" (SWRRC, 2004; MOE, 2008). It is possible to continuously measure river flow, and a commonly used vertical-axis current meter (Price Type) was used to calculate it according to the flow velocity-area method (mid-section method). The water quality was measured at the sites using a multi-parameter measuring instrument (YSI-556 mps), which was corrected in the laboratory. In addition, the hydrogen ion concentration (pH), dissolved oxygen (DO), electrical conductivity (EC), and water temperature were also measured. The collected water samples were transported in a shaded and refrigerated state with no head space. Twelve analysis items were used in this study: biological oxygen demand (BOD), chemical oxygen demand (COD), total organic carbon (TOC), total nitrogen (TN), total phosphorus (TP), chlorophyll-a, water temperature, pH, DO, conductivity, discharge, and suspended solids. Table 1 presents the details by code at 20 monitoring sites of the major tributaries and the main stream in the Nakdong River watershed.

2.3. Statistical analysis

Creating a water quality monitoring system with appropriate efficiency is one difficulty in evaluating surface water quality; and to measure variables that would express water quality changes as much as possible. To achieve this goal, a multivariate statistical method such as correlation Analysis, principal component analysis (PCA) and cluster analysis (CA) can be utilized. The river flow and water quality analysis data were analyzed statistically using a statistical analysis program, SPSS (ver. 17.0). The water quality characteristics of the Nakdong River watershed were analyzed through component analysis and cluster analysis by using correlation analysis and multivariate analysis. Multivariate analysis refers to all statistical techniques that can be used to analyze two or more variables simultaneously, and it is very complex to analyze the multivariate data containing several variables (Kim et al., 2013).

2.3.1. Principal component analysis

Principal component analysis is a multivariate analysis technique that can be used to find new variables represented by a linear combination of variables having correlations via the variance–covariance matrix of several multivariate variables; it explains most of the total variations with some important principal components (Jung et al., 2013). The new axes lie along the directions of maximum variance. PCA provides an objective way of finding indices of this type so that the variation in the data can be accounted for as concisely as possible (Sarbu and Pop, 2005). Factor analysis is a data reduction technique that can be used for easy identification of the complex structure, which is indicated by original variables, by obtaining m (where $m < p$) common factors from the covariance or correlation matrix, which shows the interdependency structure of p (arbitrary number of variables) dimensional variables (Park et al., 2001; Sakamoto et al., 1988). To make the analysis easier by simplifying the structure of the factor patterns, the factor axis is rotated. The main methods of rotating the factor axis include orthogonal rotations and oblique rotations. The orthogonal rotations include varimax, quartimax, and equimax rotations; in this study, the varimax perpendicular rotation method was used. PC provides information on the most meaningful parameters, which describe a whole data set affording data reduction with minimum loss of original information (Helena et al., 2000). The principal component (PC) can be expressed as:

$$Z_{ij} = a_{i1}x_{1j} + a_{i2}x_{2j} + a_{i3}x_{3j} + \cdots + a_{im}x_{mj} \quad (1)$$

where z is the component score, a is the component loading, x is the measured value of the variable, i is the component number, j is the sample number, and m is the total number of variables.

2.3.2. Cluster analysis

A cluster analysis was conducted to analyze the similarity of water quality variation tendencies between the investigated target sites. Cluster analysis is a group of multivariate techniques whose primary purpose is to assemble objects based on the characteristics they possess. Cluster analysis classifies objects, so that each object is similar to the others in the cluster with respect to a predetermined selection criterion. The resulting clusters of objects should then exhibit high internal (within-cluster) homogeneity and high external (between cluster) heterogeneity. Hierarchical agglomerative clustering, which is the most common approach, provides intuitive similarity relationships between any one sample and the entire data set, and it is typically illustrated by a dendrogram (tree diagram) (McKenna, 2003). The Euclidean distance usually gives the similarity between two samples, and a distance can be represented by the difference between analytical values from the samples (Otto, 1998).

Cluster analysis is a multivariate analysis method that divides all the data into several clusters based upon the distances or similarities between the observed values for parent groups; it is divided into hierarchical and non-hierarchical methods. Out of these, by using the hierarchical method, the squared Euclidean distance method of Eq. (2), which determines the distance between the target clusters by aggregating the square of differences of all variables, was applied; the Ward method was used for the cluster combination method (Shrestha and Kazama, 2007; Singh et al., 2004),

$$\mathrm{Distance}(Q_i, Q_j) = \sum_{j=1}^{n}(X_{1i} - X_{2j})^2 \quad (2)$$

where Q_i is the ith object and X_{ij} is the value of the jth variable of the ith object.

3. Results and discussion

3.1. Nakdong river watershed data set

The box plot of Fig. 2 show the analysis results of BOD, COD, TOC, TN, TP, and chlorophyll-a, which were investigated at 20 monitoring sites of the Nakdong River watershed by monitoring each site from the upstream to downstream direction. Table 2 shows evaluation of data for water quality standard of living environments criteria, and Table 3 shows water temperature, pH, DO, conductivity, discharge and suspended solid.

The box plot is a figure showing the distribution or populated shape of data with maximum value, third quartile, median value, first quartile, and minimum value. In the box plot, the scattering level of data values, i.e., the median location of data observation values and the location where the observation values are gathered, can be identified; the existence of an outlier, i.e., excessively large or small values compared to other values in the observation values, can be identified.

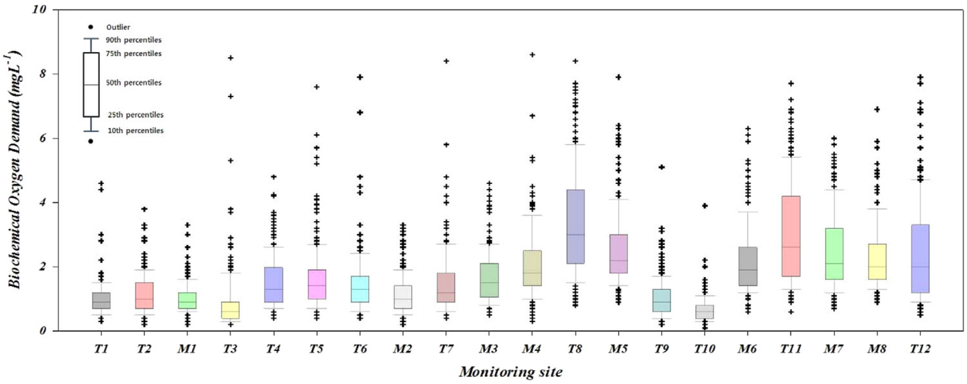
The median values of BOD were 0.9–2.2 mg/L, the first quartiles were 0.7–1.8 mg/L, and the third quartiles were 1.2–3.2 mg/L. Up to the monitoring site M3 (*Gumi*), good water qualities of 2.0 mg/L or lower were shown, and the distribution concentrated on the median value of concentration distribution was shown. At M5 *(Daeam-1)*, the median value was 2.2 mg/L, and upon approaching the midstream and downstream areas, the concentration increased and the concentration fluctuation also exhibited a tendency to become large. At T8 (*Geumho River 6*), a high concentration of 3.0 mg/L was shown, and the range of quartiles was 0.8–4.8 mg/L, also indicating large concentration fluctuations.

The median COD values were 3.9–6.3 mg/L, the first quartiles were 3.1–6.3 mg/L, and the third quartiles were 4.5–7.5 mg/L. T8 (*Geumho River 6*) was shown to have the highest concentration of 8.5 mg/L, and at T1 (*Banbyeon stream 2-1)*, T5 (*Byeongseong stream-1),* and T6 (*We stream 6)*, the concentrations were 4.9, 4.6, and 4.8 mg/L, respectively, indicating that the concentration was high in the upstream sites.

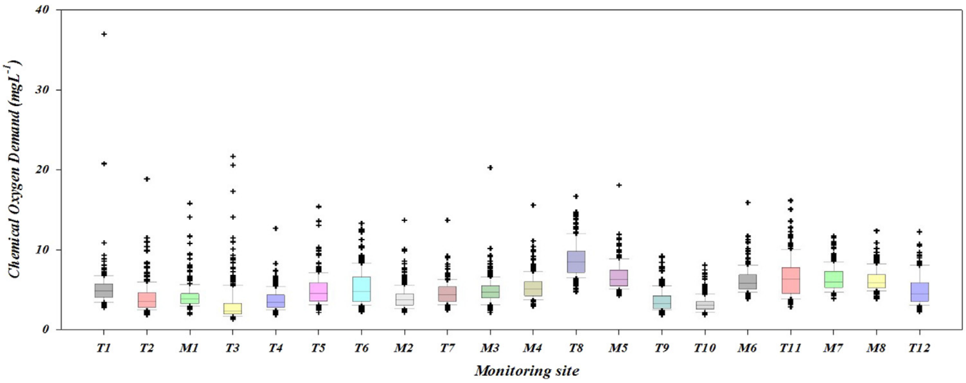
The median TOC values were 2.5–4.4 mg/L, the first quartiles were 2.0–3.8 mg/L, and the third quartiles were 3.2–5.1 mg/L. At T8 (*Geumho River 6*), the concentration was 6.0 mg/L, which was very high.

The median TN values were 2.236–3.677 mg/L, the first quartiles were 1.894–3.114 mg/L, and the third quartiles were 2.724–4.572 mg/L. At T8 (*Geumho River 6*), the median value was 6.709 mg/L, which indicated the tendency of being higher compared to other monitoring sites (in the range of 1.483–3.859 mg/L).

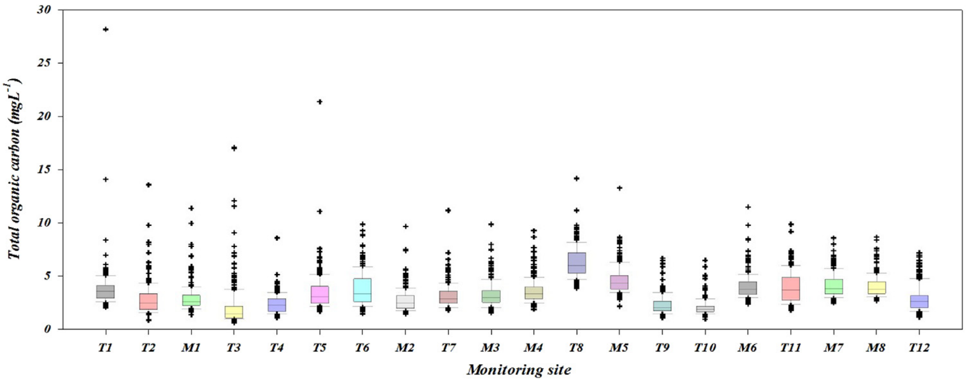
The median TP values were 0.34–0.190 mg/L, the first quartiles were 0.023–0.142 mg/L, and the third quartiles were 0.047–0.251 mg/L. At M5 *(Daeam-1)*, the median value showed a large increase to 0.190 mg/L, and the range of quartiles also showed a large scattering level based on the median values of 0.023–0.251 mg/L. The sectors after M5 (*Daeam-1*) showed a tendency to decrease gradually. In particular, at T8 (*Geumho River 6*), the median value was 0.485 mg/L, which means that its pollution level was measured to be higher compared to the concentration range of other sites (0.018–0.172 mg/L).



(a) BOD.



(b) COD.

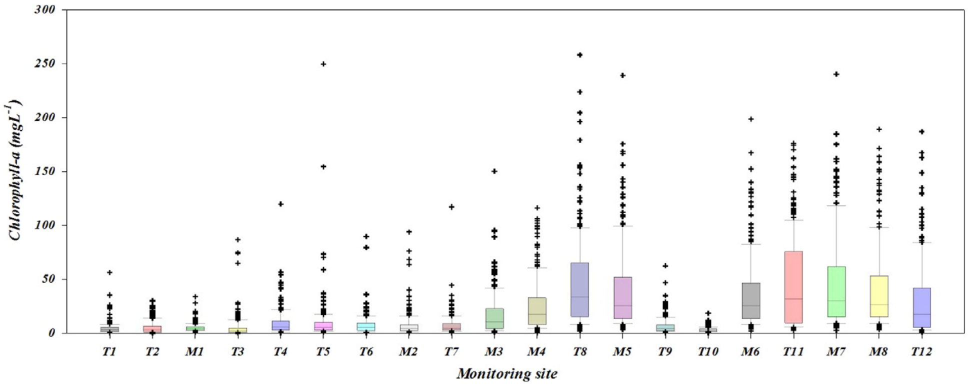


(c) TOC.

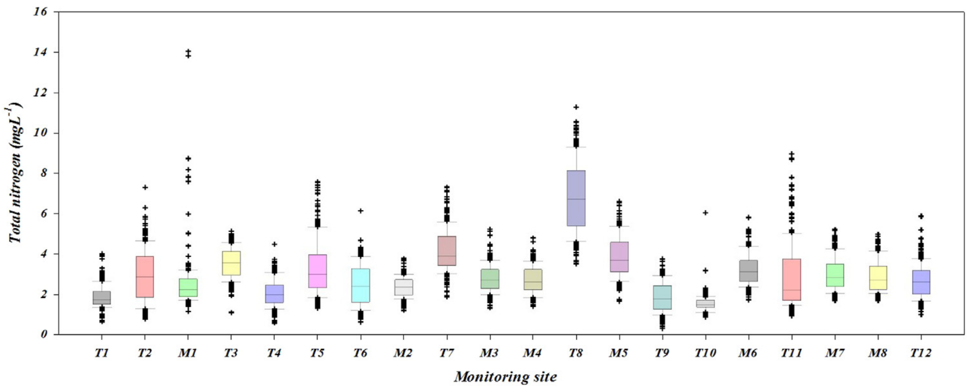
Fig. 2. Box plot of water quality values of each monitoring sites from 2008 to 2012.

The median values of Chl-a were 4.3–30.4 mg/L, with a first quartile range of 3.3–15.9 mg/L and a third quartile range of 6.3–61.8 mg/L, and the annual concentration range of Chl-a was 2.4–240.4 mg/m^3, which showed a very large fluctuation range. Furthermore, at the monitoring sites T8 (*Geumho River 6*), T11 (*Nam River 4-1*), and T12 (*Milyang River 3*), where organic matter and nutrient salt concentrations were high, the Chl-a concentration was also high, and it was investigated that the sites located near the city and midstream and downstream of Nakdong River had high pollution levels.

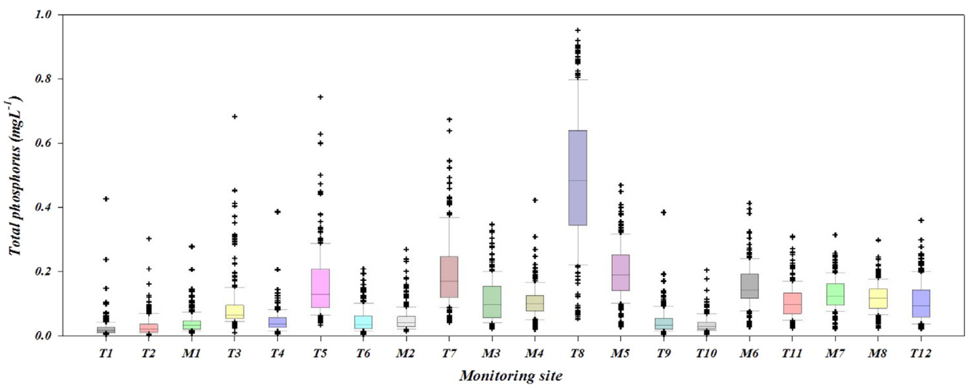
Overall, the water quality pollution level was high at the monitoring site T8, which goes through the center of Daegu City and is influenced by the discharged water from the sewage treatment plant and the residential sewage. At the midstream and downstream of Nakdong River watershed, because small cities and industrial areas are located there, the generation of pollutants was considerably high; due to the gentle river bed slope, large amounts of nutrient salts, such as organic pollutants, nitrogen, and phosphorus, were added, and thus severe flotation occurred throughout the year, agreeing with



(d) Chlorophyll-a.



(e) TN.



(f) TP.

Fig. 2. (*continued*)

the study result of Yoon et al. (2006). The water quality pollution from the upstream to the downstream showed a tendency for the concentration to increase gradually from the site M5 (*Daeam-1*) after the inflow of river flow at T8, and a tendency for small decreases when passing through M5 (*Samlangjin A*) after inflow at T8, which maintained a clean water quality. However, as the influx of flow from the site T11, i.e., the river flow of the Nam River watershed, the water quality pollution level showed a tendency to become high again.

3.2. Correlation analysis

To evaluate the relationships between river flow and water quality factors at major monitoring sites of the Nakdong River watershed, correlation analysis was conducted for 12 items by monitoring (a) site, (b) season; results are shown in Table 4. The correlations between the BOD, which is organic pollution indicator, and other water quality items at the monitoring sites were, respectively, 0.904 ($p < 0.01$) with COD, 0.805 ($p < 0.01$) with TOC, and 0.940 ($p < 0.01$) with Chl-a, showing

Table 2
Summary of monitoring sites result in Nakdong River watershed, Korea.

Site	Monitoring site	BOD[a]		COD[a]		TOC[a]		TN[b]		TP[a]		Chl-a[b]	
		mg/L	Class	mg/L	Class	mg/L	Class	mg/L	Class	mg/L	Class	ug/L	Class
M1	Andong 5	1.0	I a	4.3	II	2.9	I b	2.541	V	0.043	II	5.5	I b
M2	Sangju 3	1.1	I b	4.0	I b	2.8	I b	2.372	V	0.053	II	7.7	I b
M3	Gumi	1.7	I b	5.0	II	3.3	II	2.915	V	0.120	III	17.5	III
M4	Yongam	2.0	I b	5.4	III	3.6	II	2.774	V	0.112	III	24.6	IV
M5	Daeam-1	2.5	II	6.8	IV	4.6	III	3.920	V	0.205	IV	39.1	V
M6	Yongsan	2.2	II	6.2	IV	4.1	III	3.257	V	0.158	III	36.7	V
M7	Bukmyeon	2.5	II	6.4	IV	4.1	III	3.050	V	0.134	III	47.9	V
M8	Samlangjin A	2.4	II	6.4	IV	4.1	III	2.938	V	0.123	III	44.3	V
T1	Banbyeonstream	1.0	I a	5.2	III	3.8	II	1.885	V	0.025	I b	4.9	I a
T2	Mistream	1.2	I b	4.1	II	2.9	I b	2.998	V	0.033	I b	5.1	I b
T3	Naeseongstream	0.8	I a	3.3	I b	2.1	I b	3.556	V	0.092	II	5.6	I b
T4	Young River 2	1.5	I b	3.8	I b	2.4	I b	2.101	V	0.049	II	9.3	II
T5	Byeongseongstream-1	1.6	I b	5.0	II	3.5	II	3.402	V	0.164	III	10.1	II
T6	Westream 6	1.4	I b	5.3	III	3.8	II	2.500	V	0.051	I b	8.0	I b
T7	Gamstream 1A	1.5	I b	4.8	II	3.2	II	4.204	V	0.207	IV	7.6	I b
T8	Geumho River 6	3.4	III	8.8	V	6.3	IV	6.793	V	0.494	V	46.2	V
T9	Hoistream 2-1	1.0	I a	3.7	I b	2.3	I b	1.923	V	0.047	II	6.7	I b
T10	Hwang River 5	0.7	I a	3.3	I b	2.1	I b	1.545	V	0.038	I b	3.4	I a
T11	Nam River 4-1	3.0	III	6.6	IV	4.0	III	2.880	V	0.106	III	45.8	V
T12	Miryang River 3	2.4	II	4.9	II	2.8	I b	2.679	V	0.109	III	30.4	IV

[a] River (Stream) environmental standard for water quality.
[b] Lake environmental standard for water quality. In score: Excellent (I a), Good (I b), Above average (II), Fair (III), Poor (IV), Very poor (V).

high correlations. COD was 0.976 ($p < 0.01$) with TOC and 0.854 ($p < 0.01$) with Chl-a, also showing high correlations. Chl-a exhibited a relatively high correlation of 0.778 ($p < 0.01$) with pH. EC also showed high correlations with the nutrient salts: 0.922 ($p < 0.01$) with TP and 0.912 with TN, indicating that it acts as an electrolyte along with organic materials in the water (Kim et al., 2013). In addition, SS exhibited 0.811 ($p < 0.01$) with discharge, indicating high significance.

By season, discharge had 0.668 ($p < 0.01$) with COD and 0.681 ($p < 0.01$) with TOC, showing the relationship of relatively high amounts, and 0.861 ($p < 0.01$) with SS, which was very high. The reason for the high correlation with discharge overall was the influence of inflow along with the soil erosion in rain. The overall correlation with discharge was high as the correlation with SS was 0.861 ($p < 0.01$), which is very high, because of the effects of soil run-off and specific point inflow during periods of rain (Yoon et al., 2006).

The overall correlation with discharge was high as the correlation with SS was 0.861 ($p < 0.01$), which is very high, because of the effects of soil run-off and specific point inflow during periods of rain (Yoon et al., 2006).

The correlation of Chl-a and BOD, which is an organic matter indicator, was 0.781 ($p < 0.01$), and with T-P, which is a nutrient salt, a high correlation of 0.729 ($p < 0.01$) was shown. Thus, the increases in algae having inorganic nutrient salts as food led to the increases in organic matter, and this result is similar to the previous research results that reported them as a naturally grown BOD supply source. (Park et al., 2001; Song et al., 2012) Furthermore, Chl-a shows a relatively high correlation of 0.599 ($p < 0.01$) with pH, showing that the photosynthesis of algae can have an effect on pH increases (Park et al., 2008).

3.3. Principal component and factor analysis

To understand the water quality characteristics of the investigated target sites, the five-year observation data were divided into average data based upon monitoring site, and season. After standardizing the original data, the analysis was conducted. For the analysis, the data matrix by monitoring site was 20 (Site) $\times$ 12 (Water quality), the data matrix by season was 60 (Season) $\times$ 12 (Water quality). Using the principal component analysis, the data were extracted and the contribution rate was calculated for the eigenvalues and the water quality by monitoring site. Here, the methods of determining the extracted number of dimensions include a method based on the degree of explanation provided by a factor in the total variances, and a method in which a researcher determines the number of factors in advance. In this study, however, it was calculated through the graph (Fig. 3) by using the method for extracting the factor with an eigenvalue of 1.0 or higher and considering the number of dimensions immediately before the variance becomes dramatically small with the screen plot, since one factor explains the variance of one or more variables when the eigenvalue is higher than 1, and one factor cannot explain the variance of one variable when the eigenvalue is less than 1.

To determine the number of principal components, the axis of principal components having an eigenvalue of 1.0 or higher was considered, and three principal components were obtained in total by monitoring site, and season.

To find out the validity of the analysis before conducting the factor analysis, the Kaiser–Meyer–Olkin (KMO) test and Bartlett's test were carried out. The KMO test is a scale showing how well the correlations of variables are explained by other variables; a value closer to 1 indicates high validity, while less than 0.5 means that the analysis is not valid. The sphericity

Table 3
Observed water quality and discharge of major tributaries in(at) Nakdong-River monitoring site from 2008 to 2012.

Site	Water temperature (°C)					pH					DO (mg/L)				
code	N	Max	Min	Mean	SD	N	Max	Min	Mean	SD	N	Max	Min	Mean	SD
M1	201	30.0	0.0	14.4	7.70	201	8.9	6.1	7.7	0.38	201	17.6	6.9	11.0	2.24
M2	239	31.0	0.0	14.9	8.86	239	9.2	6.9	7.8	0.47	239	19.9	7.2	10.9	2.33
M3	199	30.3	−2.0	16.3	8.20	199	9.2	6.5	7.9	0.47	199	17.3	7.3	11.1	2.12
M4	196	32.0	0.8	16.6	8.34	196	9.7	6.7	8.0	0.52	196	18.2	6.7	10.7	2.49
M5	199	31.0	0.0	16.4	8.15	199	9.6	7.0	8.1	0.55	199	17.2	6.2	10.7	2.69
M6	199	29.7	0.0	16.3	7.92	199	9.4	6.7	7.9	0.49	199	18.0	6.4	10.4	2.31
M7	199	31.0	0.8	16.7	8.00	199	10.0	6.6	8.1	0.63	199	18.9	6.2	10.8	2.64
M8	198	30.0	−1.0	16.4	8.21	198	9.9	6.4	8.1	0.61	198	23.0	6.0	10.7	2.91
T1	241	31.0	0.0	14.6	8.42	241	9.1	5.9	7.8	0.43	241	16.2	6.6	10.8	2.21
T2	200	30.0	0.0	15.9	8.01	200	9.5	6.2	7.7	0.44	200	16.1	7.1	11.1	2.23
T3	240	32.4	−1.0	14.9	9.54	240	9.5	6.6	7.7	0.42	240	18.2	6.0	10.8	2.52
T4	200	32.4	1.0	16.4	8.36	200	9.5	6.7	7.9	0.52	200	17.8	7.0	11.2	2.36
T5	200	33.4	0.0	16.6	8.35	200	9.2	6.6	7.8	0.51	200	15.8	6.0	10.7	2.02
T6	199	31.0	0.3	15.6	8.74	199	8.7	6.9	7.7	0.44	199	18.5	7.1	10.6	2.43
T7	194	32.9	0.0	17.7	8.18	194	9.5	6.4	7.8	0.51	194	16.7	7.0	10.1	1.83
T8	239	31.0	−1.0	16.7	8.23	239	9.3	6.4	7.9	0.52	239	18.1	6.0	10.7	2.62
T9	199	32.2	0.0	16.9	8.74	199	9.2	6.4	7.8	0.46	199	16.0	5.4	10.2	2.21
T10	200	31.5	1.0	16.0	7.45	200	8.8	6.1	7.8	0.45	200	16.5	7.0	10.3	1.95
T11	238	32.4	1.0	16.8	8.73	238	9.5	6.4	8.0	0.62	238	16.4	6.5	10.7	2.23
T12	200	30.6	1.0	16.8	7.91	200	9.5	6.4	8.0	0.63	200	19.6	6.1	11.0	2.52

Site	SS (mg/L)					Discharge (m^3/s)					EC (μm hos/cm)				
code	N	Max	Min	Mean	SD	N	Max	Min	Mean	SD	N	Max	Min	Mean	SD
M1	201	97.6	1.6	13.4	15.25	201	1247.863	11.924	67.888	112.91	201	352.0	90.0	195.4	35.40
M2	239	313.0	1.2	19.5	28.96	239	1646.863	11.710	133.329	235.22	239	521.0	103.0	222.5	50.31
M3	199	560.0	1.4	27.4	46.53	199	2685.713	9.425	175.020	353.23	199	680.0	104.0	279.7	95.87
M4	196	550.0	1.6	27.8	49.75	196	3760.616	4.075	223.883	510.75	196	412.0	116.0	270.2	68.27
M5	199	436.0	2.4	35.5	46.46	199	3894.387	5.157	243.419	504.63	199	860.0	120.0	438.4	157.52
M6	199	356.0	3.6	31.4	42.61	199	3613.877	14.886	286.708	523.61	199	674.0	105.0	357.2	117.66
M7	199	259.0	2.4	25.7	29.64	199	3777.120	25.864	325.093	572.03	199	615.0	116.0	322.2	121.42
M8	198	330.0	2.2	24.1	31.15	198	4265.801	20.505	351.972	591.72	198	576.0	98.0	338.3	112.66
T1	241	414.0	0.4	8.3	29.4	241	711.249	0.043	15.879	55.98	241	1040.0	80.0	189.1	61.67
T2	200	122.0	0.2	7.3	11.3	200	234.209	0.024	4.314	18.95	200	548.0	50.0	306.2	61.17
T3	240	498.0	0.4	23.0	58.2	240	608.264	1.580	31.565	71.92	240	455.0	70.0	246.3	51.23
T4	200	283.0	0.5	8.8	23.8	200	300.054	0.129	14.135	34.92	200	491.0	111.0	297.5	85.86
T5	200	356.0	0.9	14.2	36.6	200	76.027	0.182	5.650	11.82	200	1382.0	125.0	366.4	205.47
T6	199	179.1	0.8	18.8	26.5	199	1212.300	0.756	19.301	90.82	199	581.0	126.0	306.3	84.21
T7	194	158.0	1.6	13.3	18.0	194	703.482	0.314	12.025	52.37	194	611.0	91.0	334.2	106.44
T8	239	221.0	2.0	16.5	19.7	239	1569.063	1.260	36.040	109.48	239	1340.0	194.0	773.6	239.92
T9	199	153.6	1.3	10.7	14.1	199	666.459	0.226	18.458	60.56	199	605.0	54.0	235.2	86.25
T10	200	114.0	1.2	15.4	14.3	200	149.513	1.653	28.240	23.73	200	268.0	59.0	126.9	36.64
T11	238	178.0	2.6	19.9	19.4	238	835.277	0.280	76.326	141.10	238	626.0	64.0	236.9	119.97
T12	200	189.3	1.8	11.6	18.3	200	999.166	0.308	25.485	86.88	200	438.0	84.0	217.1	68.66

Site	BOD (mg/L)					COD (mg/L)					TOC (mg/L)				
code	N	Max	Min	Mean	SD	N	Max	Min	Mean	SD	N	Max	Min	Mean	SD
M1	201	3.3	0.2	1.0	0.46	201	15.8	2.0	4.3	1.77	201	11.4	1.4	3.0	1.29
M2	239	3.3	0.2	1.1	0.57	239	13.7	2.2	4.0	1.43	239	9.7	1.5	2.8	1.07
M3	199	4.6	0.5	1.7	0.78	199	20.3	2.2	4.9	1.78	199	9.9	1.6	3.3	1.18
M4	196	8.6	0.3	2.1	1.09	196	15.6	3.0	5.4	1.56	196	9.3	1.9	3.6	1.12
M5	199	7.9	0.9	2.6	1.22	199	18.1	4.3	6.8	1.78	199	13.3	2.2	4.7	1.34
M6	199	6.3	0.6	2.2	1.07	199	15.9	3.9	6.2	1.61	199	11.5	2.4	4.1	1.20
M7	199	6.0	0.7	2.5	1.24	199	11.7	3.9	6.4	1.56	199	8.6	2.5	4.1	1.10
M8	198	6.9	0.9	2.3	1.09	198	12.4	3.9	6.3	1.43	198	8.7	2.7	4.0	1.01
T1	241	7.0	0.3	1.0	0.63	241	37.0	2.8	5.2	2.62	241	28.2	2.1	3.8	1.96
T2	200	3.8	0.2	1.2	0.61	200	18.9	1.9	4.1	1.94	200	13.6	0.9	2.9	1.53
T3	240	8.5	0.2	0.9	0.93	240	21.7	1.3	3.3	2.80	240	17.1	0.7	2.1	2.06
T4	200	4.8	0.4	1.5	0.82	200	12.7	1.9	3.8	1.33	200	8.6	1.1	2.4	0.89
T5	200	7.6	0.4	1.6	1.03	200	15.4	2.2	5.0	1.92	200	21.4	1.7	3.5	1.80
T6	199	7.9	0.4	1.5	0.95	199	13.3	2.3	5.3	2.16	199	9.9	1.5	3.8	1.59
T7	194	8.4	0.4	1.5	0.99	194	13.7	2.5	4.7	1.44	194	11.2	1.8	3.1	1.10
T8	239	8.4	0.8	3.4	1.57	239	16.7	4.8	8.8	2.16	239	14.2	3.9	6.3	1.44
T9	199	5.1	0.2	1.0	0.63	199	9.2	1.9	3.7	1.36	199	6.7	1.1	2.4	0.95
T10	200	3.9	0.1	0.7	0.41	200	8.1	1.9	3.3	1.03	200	6.5	1.0	2.1	0.76

(continued on next page)

Table 3 (*continued*)

T11	238	7.7	0.6	3.1	1.61	238	16.3	2.9	6.6	2.39	238	9.9	1.8	4.0	1.44
T12	200	7.9	0.5	2.4	1.54	200	12.3	2.3	5.0	1.90	200	7.2	1.2	2.9	1.21
Site	TN (mg/L)					TP (mg/L)					Chl-a (mg/m^3)				
code	N	Max	Min	Mean	SD	N	Max	Min	Mean	SD	N	Max	Min	Mean	SD
M1	201	14.042	1.161	2.565	1.53	201	0.279	0.010	0.042	0.03	201	34.1	1.0	5.5	3.96
M2	239	3.767	1.210	2.371	0.49	239	0.270	0.015	0.053	0.04	239	94.2	0.8	7.7	11.01
M3	199	5.242	1.333	2.798	0.69	199	0.348	0.024	0.113	0.07	199	150.6	1.0	17.7	19.95
M4	196	4.801	1.419	2.714	0.67	196	0.423	0.022	0.108	0.05	196	116.5	0.7	25.9	23.95
M5	199	6.596	1.667	3.856	1.04	199	0.471	0.029	0.204	0.09	199	239.2	3.5	40.1	39.11
M6	199	5.814	1.714	3.219	0.77	199	0.413	0.030	0.156	0.07	199	198.7	2.4	37.5	34.05
M7	199	5.214	1.704	2.976	0.82	199	0.315	0.023	0.131	0.05	199	240.4	2.9	47.4	43.64
M8	198	4.977	1.684	2.882	0.78	198	0.299	0.026	0.121	0.05	198	189.2	3.4	41.4	39.04
T1	241	3.997	0.643	1.889	0.55	241	0.427	0.006	0.025	0.03	241	56.3	0.6	4.9	5.32
T2	200	7.307	0.787	2.932	1.28	200	0.304	0.004	0.032	0.04	200	30.2	0.3	5.3	5.75
T3	240	5.141	1.106	3.551	0.75	240	0.684	0.012	0.092	0.08	240	87.0	0.3	5.7	10.46
T4	200	4.486	0.582	2.056	0.66	200	0.387	0.009	0.047	0.04	200	120.0	0.7	9.7	12.82
T5	200	7.579	1.320	3.293	1.39	200	0.745	0.035	0.161	0.11	200	249.9	0.9	10.4	22.14
T6	199	6.149	0.639	2.462	1.01	199	0.210	0.007	0.050	0.04	199	89.9	0.5	8.0	9.80
T7	194	7.320	1.884	4.152	1.07	194	0.675	0.043	0.201	0.12	194	117.2	0.7	8.0	10.41
T8	239	11.292	3.509	6.788	1.72	239	0.953	0.054	0.494	0.21	239	258.3	2.4	46.4	42.83
T9	199	3.750	0.313	1.848	0.76	199	0.385	0.006	0.046	0.04	199	62.6	0.7	7.1	7.68
T10	200	6.038	0.868	1.549	0.44	200	0.206	0.006	0.038	0.03	200	18.6	0.7	3.4	2.14
T11	238	8.977	0.948	2.868	1.65	238	0.312	0.026	0.106	0.05	238	176.4	2.9	46.1	41.93
T12	200	5.879	0.991	2.654	0.82	200	0.360	0.023	0.109	0.06	200	187.2	1.0	31.2	35.80

SD : Standard deviation.

Table 4a
Pearson correlation matrix of total monitoring sites.

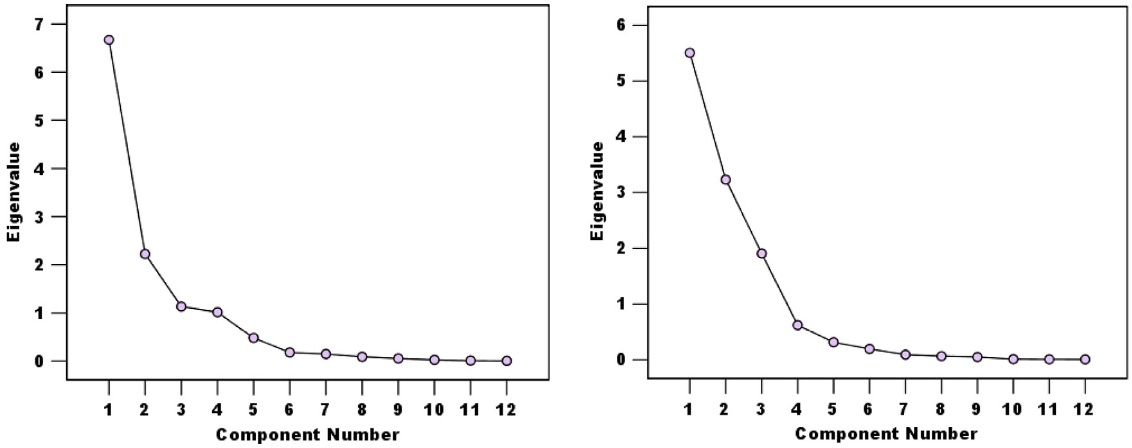
Variable	EC	Water temperature	DO	pH	BOD	COD	TOC	SS	Chl-a	Discharge	TN	TP
EC	1.000	0.439	−0.040	0.241	0.666**	0.766**	0.816**	0.214	0.523*	0.135	0.912**	0.922**
Water temperature		1.000	−0.297	0.569**	0.596**	0.418	0.325	0.063	0.459*	0.027	0.441	0.548*
DO			1.000	0.208	0.110	−0.025	−0.043	−0.003	0.089	0.129	−0.137	−0.199
pH				1.000	0.712**	0.541*	0.408	0.514*	0.778**	0.650**	0.143	0.307
BOD					1.000	0.904**	0.805**	0.417	0.940**	0.453*	0.620**	0.710**
COD						1.000	0.976**	0.434	0.854**	0.458*	0.695**	0.773**
TOC							1.000	0.381	0.733**	0.381	0.728**	0.787**
SS								1.000	0.562**	0.811**	0.232	0.269
Chl-a									1.000	0.671**	0.479**	0.577**
Discharge										1.000	0.063	0.133
TN											1.000	0.948**
TP												1.000

* Correlation is significant at the 0.05 level (2-tailed).
** Correlation is significant at the 0.01 level (2-tailed).

Table 4b
Pearson correlation matrix of season.

Variable	EC	Water temperature	DO	pH	BOD	COD	TOC	SS	Chl-a	Discharge	TN	TP
EC	1.000	0.770**	0.823**	0.012	0.249	−0.285	−0.318	−0.570	0.490*	−0.765**	0.656**	0.428
Water temperature		1.000	0.983**	0.378	0.093	0.538*	0.566**	0.494*	−0.009	0.697**	0.938**	0.010
DO			1.000	−0.296	0.021	0.482*	−0.535*	−0.522*	0.115	−0.702**	0.933**	0.072
pH				1.000	0.511*	0.081	−0.034	−0.212	0.599**	−0.164	−0.468*	0.280
BOD					1.000	0.560*	0.384	0.011	0.781**	−0.054	0.021	0.465*
COD						1.000	0.941**	0.726**	0.279	0.668**	−0.312	0.387
TOC							1.000	0.748**	0.209	0.681**	−0.372	0.312
SS								1.000	−0.267	0.861**	−0.310	0.068
Chl-a									1.000	−0.316	−0.013	0.729**
Discharge										1.000	−0.508*	0.002
TN											1.000	−0.020
TP												1.000

* Correlation is significant at the 0.05 level (2-tailed).
** Correlation is significant at the 0.01 level (2-tailed).



(a) Total monitoring sites. (b) Season.

Fig. 3. Screen plot.

Table 5
Initial eigenvalues and selected factor loading after Varimax rotation (Total monitoring sites (a), season (b)).

	Component	Initial eigenvalues			Rotation sums of squared loadings		
		Total	% of variance	% Cumulative	Total	% of variance	% Cumulative
	1	6.672	55.602	55.602	4.844	40.364	40.364
(a)	2	2.223	18.527	74.129	2.771	23.093	63.457
	3	1.131	9.428	83.557	2.203	18.360	**81.817**
	1	5.504	45.866	45.866	3.791	31.588	31.588
(b)	2	3.231	26.926	72.791	3.629	30.240	61.828
	3	1.908	15.900	88.692	3.224	26.864	**88.692**

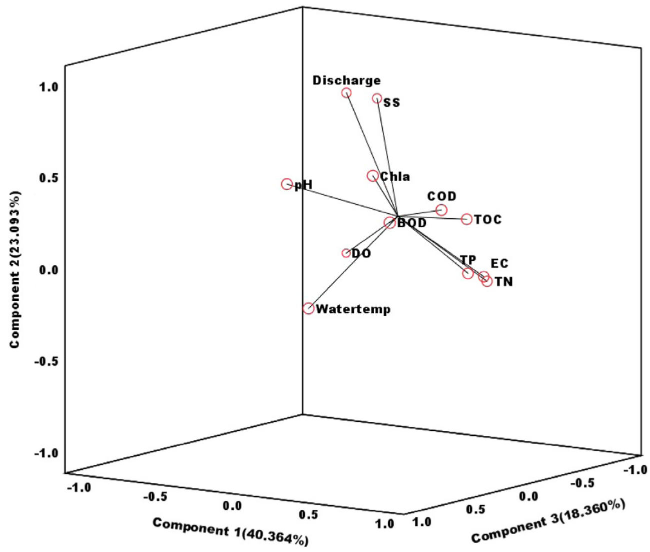
Table 6
Rotated factor matrix extracted from principal component analysis.

Variable	(a)			Variable	(b)		
	Com 1	Com 2	Com 3		Com 1	Com 2	Com 3
EC	**0.949**	0.028	0.131	TN	**0.953**	−0.133	−0.049
TN	**0.939**	−0.005	0.086	Water temperature	**−0.915**	0.375	0.053
TP	**0.922**	0.059	0.240	DO	**0.912**	−0.374	0.065
TOC	**0.861**	0.337	0.164	EC	**0.728**	−0.384	0.468
COD	**0.801**	0.403	0.317	SS	−0.217	**0.899**	−0.146
BOD	0.655	0.359	0.592	TOC	−0.229	**0.873**	0.304
Discharge	0.057	**0.943**	0.145	COD	−0.208	**0.856**	0.421
SS	0.171	**0.906**	0.020	Discharge	−0.448	**0.811**	−0.220
Chla	0.504	0.591	0.535	Chla	0.023	−0.114	**0.965**
Water temperature	0.311	−0.104	**0.862**	BOD	0.004	0.171	**0.856**
pH	0.104	0.539	**0.769**	TP	0.134	0.205	**0.768**
DO	−0.065	0.030	−0.027	pH	−0.567	−0.371	0.635

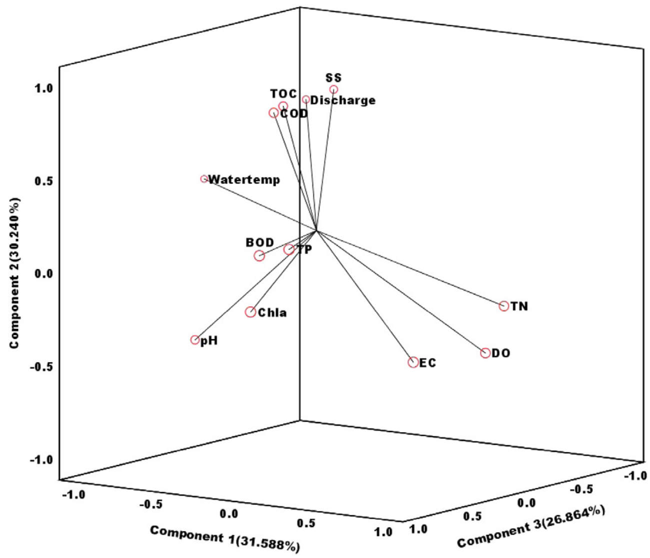
verification value of Bartlett's test is said to show a significant relationship when each variable shows that correlation exists with others; a value closer to 0 indicates higher significance (Bernard et al., 2004; Kim et al., 2007a).

The result of the KMO test in this study was analyzed to be 0.583 by monitoring site (a), and 0.590 by season (c). In Bartlett's test, all were 0.000 ($p < 0.05$), indicating that correlations exist between the variables by dismissing the idea that the correlation matrix is a unit matrix. Therefore, both analyses satisfied the conditions and explained that factor analysis is possible. The purpose of factor analysis is to find the hidden common factor affecting the observed variables, and it defines each influencing factor more clearly after orthogonal rotation. As a result of the analysis, all identical numbers of axes were calculated after rotation, and it was found to be valid to consider up to three axes by monitoring site, and season. Table 5 shows the eigenvalue after orthogonal rotation and cumulative share.

By monitoring site, the principal component's total water quality variation of 81.817% is explained, and the first factor contributes 40.364%, the second factor 23.093%, and the third factor 18.360%. By season, the principal component's total water quality variation of 88.692% is explained, and the first factor contributes 31.588%, the second factor 30.240%, and the third factor 26.864%.



(a) Total monitoring sites.



(b) Season.

Fig. 4. Component plot.

Fig. 4 is a spatial component graph by monitoring site, and season for 12 variables used in the factor analysis. Fig. 4 by position (a), in the first factor axis, most of the indicators corresponding to the nutrient salts (TN, TP) and EC and organic contaminants (TOC, COD) of their subsequent electrolyte effect are converged; and discharge and SS, which were affected by precipitation, are dispersed in the second factor axis.

In Fig. 4(b) (Month, along with Discharge and SS), the organic pollutants (TOC, COD) are concentrated together on the first factor axis, showing that the physical variables are influencing the increase in organic pollutants. In addition, the water temperature and DO, which are seasonally affected, are dispersed with negative relationships on the second factor axis. Furthermore, the major variables of biological metabolism, Chl-a, BOD, and pH are shown together on the third factor axis. In Fig. 4(c) (Season), the season factors and nutrient salts are on the first and second factor axis, respectively, and the major variables of biological metabolism are dispersed on the third factor axis.

By season (b), Watertemp, which has an inverse relationship with DO and TN, is a first factor, and on the second factor axis, Discharge and SS are converged along with the organic contaminants TOC and COD. Thus, it is confirmed that physical variables affect increases of organic contaminants and that the major variables of biological metabolism, namely Chl-a, BOD, and pH, dispersed together on the third factor axis.

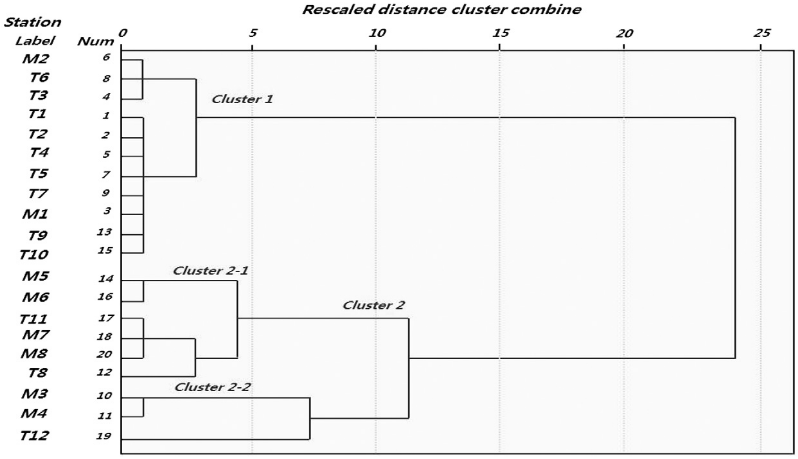


Fig. 5. Dendrogram of the 20 monitoring sites using hierarchical cluster analysis based on water quality of the Nakdong River, Korea.

In South Korea, summer is a wet season, during which precipitation is abundant, and winter and spring are dry seasons. During the dry seasons (winter, early spring), the Nakdong River is greatly affected by physical and hydrological factors such as low water temperature (5–10 °C) and small flow rate along with abundant nutrient salts (Kim et al., 2007b). Table 6 shows that point pollution sources have the largest influence on the temporal and spatial distributions of TN concentration. Moreover, the peak precipitation and flow rate annually repeated from July to September show a typical type of monsoon climate, thereby indicating that climate and hydrological characteristics are important factors controlled as a result of the specific point inflow effect caused by a large inflow of unprocessed water overflowing from a livestock system, a residential system, and industrial park zones, along with summer discharge and SS. Furthermore, the occurrence ratio of phosphorous caused by the surface run-off of phosphorous from land soil can further increase (Mainstone and Parr, 2002). Based on this, it is explained that, in the watersheds of the Nakdong River overall, the contamination by nutrient salts and organic contaminants is the largest, followed by the seasonal factor and biological metabolisms in that order.

3.4. Cluster analysis

The results of conducting cluster analysis based on the result obtained from factor analysis are shown in the dendrogram of Fig. 5 to analyze the similarities of water quality variation tendencies between the target monitoring sites. A dendrogram is easy to use for analyzing the similarities of water quality variation between the major investigation target tributaries and the main stream. Also, it helps to classify the clusters easily and provides visual information (Kim et al., 2007a). For classification, which is clearer than monthly average values that change over time, a cluster analysis was conducted by using the annual average values as representative values to identify the spatial (watershed) factors. Two statistically significant groups were classified by comprehensively considering the characteristics of the sources of pollution for river watersheds and the seasonal characteristics.

For Cluster 1, the upstream and major tributaries of the Nakdong River, which have low pollution levels, were classified. For Cluster 2, the midstream and downstream sites were classified, which were influenced by pollution sources that flowed in due to high share of the agricultural lands and small and large cities in the vicinity; two sub-clusters were classified. Among them, Cluster 2-1 includes the representative tributaries of the Nakdong River, i.e., Geumho River and Nam River, which have high pollution levels due to high influences of residential and industrial waste water. These two rivers have large environmental changes in the watershed, and as the rivers pass through large cities, the organic matter concentration and nutrient salt concentration are very high due to direct influences of residential sewage, and they have a large effect on the downstream area of the Nakdong River. Furthermore, the main stream positions downstream of the Nakdong River are also included, in which the concentration is high due to an increase of pollution load and a slowdown of current speed caused by gentle hydraulic characteristics in addition to the influences of two rivers. For Cluster 2–2, the sites located midstream of the Nakdong River and the sites of Miryang River, which have a higher pollution level than the upstream area of the Nakdong River and a lower level than the downstream area, were classified.

According to Singh et al. (2004), depending on the water quality pollution level of the investigated target sites, the groups are classified as a result of cluster analysis. Shrestha and Kazama (2007) also reported that the cluster analysis results reflected the influences of land use of the target river watershed and the location of a sewage processing plant well, and groups are classified according to pollution sources in the vicinity, such as residential sewage and agricultural activities. The reason that water quality pollution characteristics are distinguished in this way is deeply related to each region's industry characteristics, population density, and land usage characteristics. The results of this study showed that the groups were classified according to the characteristics of pollution sources in the vicinity or the water quality pollution

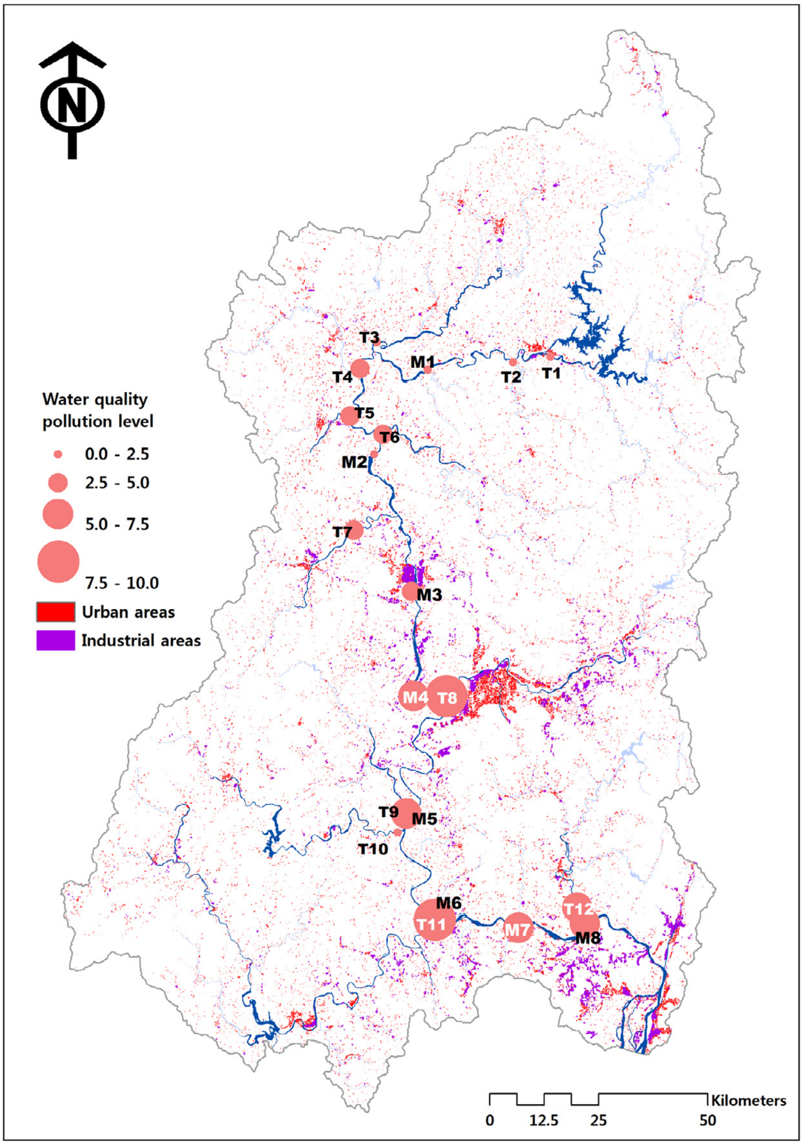


Fig. 6. GIS Mapping results from main stream and tributaries water quality pollution level.

level of the river. Cluster analysis is very useful in classifying rivers and in distinguishing the influences of classifications or tributaries of the river having similar water quality characteristics in a large river. Furthermore, when establishing water quality management measures for the Nakdong River, it can be used as a method of establishing a more efficient spatial sampling plan by classifying watershed groups showing similar water quality characteristics.

3.5. Water quality pollution level

Fig. 6 illustrates the water quality pollution level of the river according to quantitative indicators by reflecting various factors between major tributaries and the main stream, which are the targets being investigated in this study. The water quality pollution level of the river is shown as GIS by calculating the ratio, which divides the sum of the average concentrations at each monitoring site for four items (TN, TP, COD, and TOC) classified as the first factors as a result of principal component analysis and factor analysis targeting all investigated sites. It is found that relatively high water quality pollution levels are shown at T8 and T11. T8 is a point where Geumho River joins, and T11 is a point where Nam River joins.

Park (2003) carried out a study on the effect of the environmental characteristics of Geumho River watershed on river water quality. He stated that more than 30% of the area in the Geumho River watershed is residential, commercial, and road areas, and the industrial areas account for more than 15%. Furthermore, the correlations between water quality indicators and the residential, commercial, industrial, and road area ratios were evaluated to be positive. In the case of Nam River watershed, upon analyzing the land cover status from 1975 to 2005, Park and Lee (2012) reported that the urbanized areas increased radically from 9.4 to 119 km^2. Geumho River and Nam River pass through large cities, and in these areas, urbanization is continuously in progress and industries continuously discharge industrial water and industrial complex waste water. Therefore, the risk of water quality pollution is relatively high, and strict management of water quality must be a priority for these tributaries.

4. Conclusion

In this study, the following results were obtained from the water quality distribution characteristics and basic correlation analysis as well as the multivariate statistical analysis for a total of 12 pollution indicators based on the water quality observation data of 20 monitoring sites from 2008 to 2012 to evaluate the water quality characteristics of the Nakdong River's watershed.

(1) The average BOD of the investigated target sites was 0.9–2.2 mg/L, and COD showed a range of 3.9–6.3 mg/L. T8, which passes through the city of Daegu, and M5, which is located downstream of the Nakdong River, tended to show high values, while Chl-a had a concentration range of 2.4–240.4 mg/m^3 year-round, showing a very large fluctuation range. Therefore, the city rivers and the monitoring sites located midstream and downstream of the Nakdong River were shown to have high pollution levels.

(2) The observation values of 20 major sites of the Nakdong River's watershed were divided into average data by monitoring (a) site, and (b) season for five years, and the correlation was analyzed between river flow and water quality factors. In the results by monitoring site, BOD showed high amounts of correlations with COD, TOC, and Chl-a; and EC and the nutrient salts (TP, TN) showed a high correlation. In addition, SS showed a high significance with discharge.

(3) Summarizing the results of researches previously carried out and the analyzed data, the primary factors in the Nakdong River watershed were nutrient salts (TN, TP) and organic pollutants (TOC, COD). As the secondary factors, seasonally affected Discharge and SS were classified; and, biological metabolism (Chl-a, BOD, pH) was assesses as the tertiary factor. Furthermore, as a result of cluster analysis, two statistically significant groups were classified. Groups were classified according to the characteristics of pollution sources in the vicinity or the water quality pollution level. In particular, the tributaries with a high pollution level and the downstream site of the main stream affected by it were classified as two sub-clusters.

(4) According to the study result of multivariate statistical analysis, the places showing high pollution levels were mostly the sites passing through the city or directly affected by residential sewage in the city river, as well as the sites located midstream and downstream of the Nakdong River. Based on the above study results, by determining the priorities necessary depending on the water quality pollution level of the river and year-round management of the polluted areas(temporal/spatial distribution (winter and spring, mid/downstream)), it is expected that a contribution will be made to data usage for efficient water quality management in terms of future watershed management.

References

Abahams, A.D., 1972. Factor analysis of drainage basin properties: Evidence for stream abstraction accompanying the degradation of relief. Water Resour. Res. 8 (3), 624–633.
Adams, S., Titus, R., Pietesen, K., Tredoux, G., Harris, C., 2001. Hydrochemical characteristic of aquifers near Sutherland in the Western Karoo, South Africa. J. Hydrol. 241, 91–103.
Bernard, P., Antoine, L., Bernard, L., 2004. Principal component analysis an appropriate tool for water quality evaluation and management application to a tropical lake system. Ecol. Model. 178, 295–311.
Bricker, O.P., Jones, B.F., 1995. Main factors affecting the composition of natural waters. In: Salbu, B., Steinnes, E. (Eds.), Trace Elements in Natural Waters. CRC Press, Boca Raton, FL, pp. 1–5.
Dixon, W., Chiswell, B., 1996. Review of aquatic monitoring program design. Water Res. 30, 1935–1948.
Fang, H., Xiaoquan, W., Liping, L., Zhiqing, Z., Jiaping, W., 2010. Spatial variation and source apportionment of water pollution in Qiantang River (China) using statistical techniques. Water Res. 44, 1562–1572.
Helena, B., Pardo, R., Vega, M., Barrado, E., Ferna'ndez, J.M., Ferna'ndez, L., 2000. Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga river, Spain) by principal component analysis. Water Res. 34, 807–816.
Homa, R., Ahmad, A., Atoob, T., 2010. Evaluation of spatial and temporal variation in water quality by pattern recognition techniques: A case study on Jajrood River (Tehran, Iran). J. Environ. Manag. 91, 852–860.
Jung, K.Y., 2009. Advanced control strategieso for water quality management in Nakdong river (Master's Thesis), Kyungpok National University, pp. 15–17.
Jung, S.J., Kim, K.S., Seo, D.J., Kim, J.H., Lim, B.J., 2013. Evaluation of water quality characteristics and grade classification of Yeongsan River tributaries. J. Korean Soc. Water Environ. 29 (4), 504–513.
Karim, B., Taha, F.M., 2003. Using principal component analysis to monitor spatial and temporal changes in water quality. J. Hazardous Mater. B 100, 179–195.
Kim, G.H., Jung, K.Y., Yoon, J.S., Cheon, S.U., 2013. Temporal and spatial analysis of water quality data observed in lower watershed of Nam river dam. J. KOSHAM 13 (6), 429–437.
Kim, M.A., Kang, T.G., Lee, H., Shin, Y.N., Kim, K.H., 2012. Analysis of pollutant characteristics in Nakdong river using confirmatory factor modeling. J. Korean Soc. Water Environ. 28 (1), 84–93.
Kim, G.A., Kim, Y.J., Song, M.J., Ji, K.W., Yu, P.J., Kim, C.W., 2007a. Evaluation of water quality characteristics in the Nakdong river using multivariate analysis. J. Korean Soc. Water Qual. 23 (6), 814–821.
Kim, M.A., Lee, J.K., Zoh, K.D., 2007b. Evaluation of the Geum River by multivariate analysis: Principal component analysis and factor analysis. J. Korean Soc. Water Qual. 23 (1), 161–168.
Lee, J.Y., Cheon, J.Y., Lee, K.K., Lee, S.Y., Lee, M.H., 2001. Statistical evaluation of geochemical parameter distribution in a ground water system contaminated with petroleum hydrocarbons. J. Environ. Qual. 30, 1548–1563.
Mainstone, C.P., Parr, W., 2002. Phosphorus in rivers—ecology and management. Sci. Total Environ. 282–283, 25–47.

McKenna Jr., J.E., 2003. An enhanced cluster analysis program with bootstrap significance testing for ecological community analysis. Environ. Model. Softw. 18 (3), 205–220.
Otto, M., 1998. Multivariate methods. In: Kellner, R., Mermet, J.M., Otto, M., Widmer, H.M. (Eds.), Analytical Chemistry. Wiley-VCH, Weinheim.
Park, J.H., 2010. A study of waterbody health diagnosis method using load duration curve (Master's Thesis), Kyungpok National University, pp. 33–31.
Park,, 2003. The impact of environmental characteristics in the Geumho river watershed on stream water quality. J. Korean Assoc. Geogr. Inform. Stud. 6 (4), 85–98.
Park, H.K., Byeon, M.S., Choi, M.J., Kim, Y.J., 2008. The effect factors on the growth of phytoplankton and the sources of organic matters in downstream of South-Han River. J. Korean Soc. Water Qual. 24 (5), 556–562.
Park, J., Lee, Y., 2012. A Study on the Application of the Sustainable Environmental System for Gyeongnam Province – Water Resources Security by Establishing of Water Cycling System. Gyeong Nam Development Institute.
Park, H.S., Park, C.K., Song, M.K., Baek, K.H., Shin, S.K., 2001. Evaluation of water quality characteristics using factor analysis in the Nakdong river. J. Korean Soc. Water Qual. 17 (6), 693–701.
Reghunath, R., Murthy, T.R.S., Raghavan, B.R., 2002. The utility of multivariate statistical techniques in hydrogeochemical studies: an example from Karnataka, India. Water Res. 36, 2437–2442.
Sakamoto, H., Hamaguchi, R., Iwase, M., 1988. Evaluation of river water pollution in the Chanai district of Eastern Hokkaido by the multivariate analysis method. Water and Waste 30 (3), 239–245.
Sarbu, C., Pop, H.F., 2005. Principal component analysis versus fuzzy principal component analysis. A case study: the quality of Danube water (1985–1996) Talanta 65, 1215–1220.
Shrestha, S., Kazama, F., 2007. Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river Basin, Japan. Environ. Model. Softw. 22, 464–475.
Simeonova, P., Simeonov, V., Andreev, G., 2003. Environmetric analysis of the Struma river water quality. Cent. Eur. J. Chem. 2, 121–126.
Simeonov, V., Simeonova, P., Tsitouridou, R., 2004. Chemometric quality assessment of surface waters: two case studies. Chem. Eng. Ecol. 11 (6), 449–469.
Singh, K.P., Malik, A., Mohan, D., Sinha, S., 2004. Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti river (India)—a case study. Water Res. 38, 3980–3992.
Song, E.S., Jeon, S.M., Lee, E.J., Park, D.J., Shin, Y.S., 2012. Long-term trend analysis of chlorophyll a and water quality in the Yeongsan River. Korean J. Limnol. 45 (3), 302–313.
Vega, M., Pardo, R., Barrado, E., Deban, L., 1998. Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. Water Res. 32, 3581–3592.
Wunderlin, D.A., Diaz, M.P., Ame, M.V., Pesce, S.F., Hued, A.C., Bistoni, M.A., 2001. Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality. A case study: Suquia river basin (Cordoba, Argentina). Water Research 35, 2881–2894.
Yoon, Y.S., Yu, J.J., Kim, M.S., Lee, H.J., 2006. Computation and assessment of delivery pollutant loads for the streams in the Nakdong river basin. J. Korean Soc. Water Qual. 22 (2), 277–287.