

## 낙동강 주요지점에서 유량-수질의 관련특성에 관한 연구

조 현 경\*

영남이공대학 토목과

(2010년 12월 2일 접수; 2010년 12월 26일 수정; 2011년 1월 31일 채택)

### A Study on the Related Characteristics of Discharge-Water Quality in Nakdong River

Hyeon-Kyeong Cho\*

Dept. of Civil Engineering, Yeungnam College of Science & Technology, Daegu 705-703, Korea

(Manuscript received 2 December, 2010; revised 26 December, 2010; accepted 31 January, 2011)

#### Abstract

This study aims at the examination of the relative characteristics of discharge and water quality in river basins using statistical methods. For it, water quality and discharge data was collected in observed stations of Nakdong river and carried out correlation analysis, regression analysis, factor analysis and cluster analysis. And it was investigated the applicability of water quality prediction using Nearest-neighbor method. As a result, it grasped a trend characteristics and mutual relations between discharge and water quality data.

Therefore, this results were suggested the comprehensive data and methods for a management of water quality, effective operation and policy development in Nakdong river basin.

**Key Words** : Discharge and water quality, Factor analysis, Cluster analysis

#### 1. 서 론

요즘들어 고도경제 산업사회로의 국가정책에 따라 국민의 생활수준은 크게 향상되고 있으나 이에 따른 부작용으로 환경오염은 점차 심화되고 있다. 특히 각종 산업폐수 및 생활하수 등의 방류로 오염물질이 하천으로 유입됨으로써 하천수질의 악화가 심화되고 있어 상수원 오염의 가능성이 높아지고 있다. 또한 하천 유역은 시가지의 개발 확산과 하천의 인위적인 변경에 따라 자정능력을 점점 잃어가고 있어 수질오염을

가속화시키고 있는 실정이다.

이에 따라 최근들어 국가나 지방자치단체들이 환경정책 중에서 수질오염 방지 및 개선에 우선순위를 두고 정책을 시행하고 있고, 이러한 지금까지의 수질 개선 노력이 큰 성과를 거두고 있는 것 또한 사실이다. 그러나 현재의 각 수계의 수질은 아직도 많은 문제점을 안고 있는 형편에 있다.

특히 영남지역의 대부분을 차지하고 있는 낙동강 유역은 하천을 따라 개발과 도시화가 진행되어 수질 오염의 가능성이 어느 지역보다 높다. 따라서 이러한 상황에 놓여있는 하천수질의 관리를 위한 계획을 수립하기 위해서는 정확한 수질자료의 획득과 자료분석이 필수적이다. 낙동강 유역에는 하천 오염문제를 해결하기 위한 목적으로 많은 수의 수질 및 수문 관측망

\*Corresponding author : Hyeon-Kyeong Cho, Dept. of Civil Engineering, Yeungnam College of Science & Technology, Daegu 705-703, Korea  
Phone: +82-53-650-9316  
E-mail: chohk@ync.ac.kr

이 설치되어 있어 많은 양의 데이터가 수집되어 관리되고 있으나 자료의 통계학적인 특성을 파악하고 이를 하천관리 및 운영, 수질예측 등에 응용하고자 하는 노력은 상대적으로 미흡한 실정이다.

최근까지 수질자료의 특성에 관한 연구를 수행한 국내의 연구문헌들을 살펴보면, 이 등(1988)은 인자분석법에 의해 강우와 하천수질변동 특성에 관한 연구를 수행하였으며, 권(1992)은 하천 유량변동에 따른 수질변화에 관한 연구를 수행하였다. 그리고 유(1993)는 수문인자에 의한 하천수질변화에 관한 연구를 발표하였고, 양(1994)은 하천유량 변화에 따른 수질오염도 분석을 실시한 바 있다. 백(1995)은 한강하류부에서 통계학적 방법에 의한 수질분석을 수행한 바 있으며, 이(1996)는 중소하천의 유량과 수질변동 특성에 관한 연구를 수행하였다. 또한 김(1998)은 낙동강에서 수질오염 특성조사를 통한 수질개선 방안에 대해 연구하였으며, 임 등(1999)은 낙동강수계의 수질현황에 대해 연구한 바 있다. 그리고 정 등(1999)은 다변량분석법을 이용하여 하천수질을 평가한 바가 있으며, 정 등(2000)은 금강지류 유역에서 유출량과 오염부하량의 상관관계를 분석한 바가 있다. 김(2002)은 통계분석 기법을 이용한 금강수계의 수질을 평가한 바가 있으며, 이 등(2007)은 남한강수계의 저수기 때 수질변동 특성에 관한 연구를 수행하였다.

그리고 국외의 연구문헌들을 살펴보면, Abahams (1972)은 배수유역 특성을 파악하기 위해 요인분석을 실시하였으며, Mehta 등(1975)은 Passaic강 수질의 추계학적 변화를 연구한 바 있다. 그리고 Thomann (1976)이 수질자료의 시계열분석을 실시하였고, Shrestha (2007)는 다변량 통계방법을 이용하여 수질을 평가한 바가 있다. 또한 Fang 등(2010)은 Qiantang강의 수질오염의 공간적인 변동특성을 통계학적인 방법을 이용하여 연구하였으며, Homa 등(2010)은 패턴인식 방법을 이용하여 수질의 시공간 변동성을 연구한 바가 있다.

그런데 위의 국내외 연구 성과들을 살펴보면 단편적인 통계학적 방법을 이용하여 결과를 얻은 것이 대부분이었고 자료의 특성, 상관성, 더 나아가서 예측까지 종합적으로 결과를 도출하여 이를 하천관리 및 운영 등에 필요한 자료와 방법으로 제시한 것은 드물었다.

따라서 본 연구에서는 낙동강분류 주요지점에서

유량과 수질자료를 수집하고 통계학적 방법을 이용하여 유량을 고려하였을 때의 수질자료의 특성, 항목간의 상관성을 파악하였으며, 또한 수질예측 방법의 적용성을 조사하여 향후 낙동강의 수질관리와 효율적인 운영 및 관련 정책개발에 필요한 자료와 방법을 제시하는데 그 목적을 두었다.

## 2. 자료 및 방법

### 2.1. 대상유역 및 분석자료

본 연구에서 낙동강 분류에서 유량과 수질항목들의 관련특성을 파악하기 위해 선정한 대상하천으로는 한반도의 동남단에 위치한 낙동강을 선택하였다(Fig. 1). 낙동강은 강원도 태백 함백산에서 발원하여 경상북도와 경상남도를 관류하고 부산광역시 사하구 하단지점에서 남해로 유입하는 우리나라 제 2의 하천이다. 낙동강은 총연장 506.17 km, 유역면적 23,384.21 km<sup>2</sup>에 825 개의 소하천으로 구성되어 있어 다양한 오염원에 의한 수질오염사고가 발생할 가능성이 매우 높은 유역이다.

그리고 낙동강 분류에서 유량과 수질의 특성을 파악하기 위해 선정한 대상지점으로는 낙동강 상류수질을 보여주는 안동지점, 중류수질을 나타내는 구미, 왜관, 고령, 현풍지점, 하류수질을 나타내는 임해진, 삼랑진, 구포지점 등 총 8개 지점을 선택하였다.

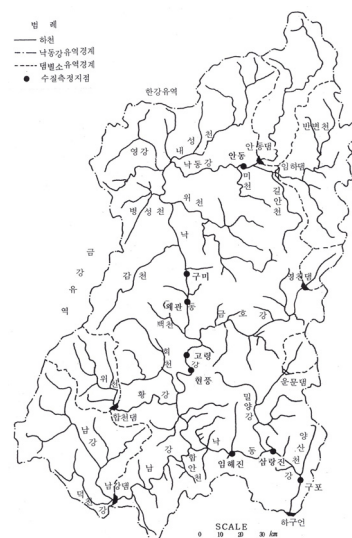


Fig. 1. Study river.

본 연구에 필요한 자료는 유량 및 수질자료이며, 대상 지점별로 2003년에서 2009년까지의 자료를 수집하여 이용하였다. 여기서, 유량과 수질자료는 물환경정보시스템(2003~2009)의 월평균 자료를 이용하였으며, 특히 수질자료는 DO, BOD, COD, SS, T-N, T-P, 수온을 선택하였다. Fig. 2는 상·중·하류의 대표지점으로 선정된 안동, 왜관, 삼랑진의 유량과 주요 수질자료(DO, BOD, COD)를 도시한 것이다.

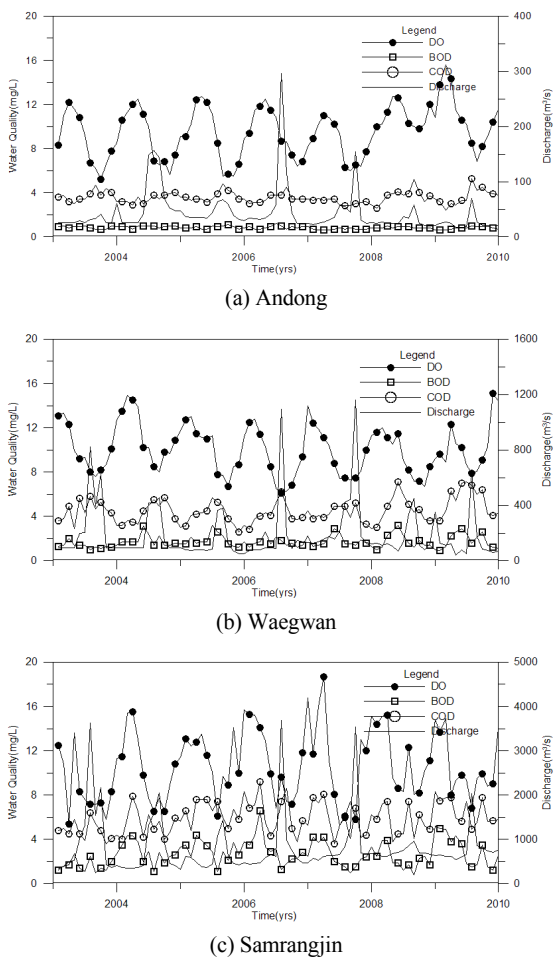


Fig. 2. Study data.

## 2.2. 분석방법

### 2.2.1. 상관분석(Correlation analysis)

상관분석은 연구하고자 하는 변수들 간의 관련성

을 분석하기 위해 사용하는 통계방법이다. 상관분석은 두 변수간의 선형적인 상관관계를 파악해보는 것으로 한 변수가 다른 변수와의 관련성이 있는지 여부와 관련성이 있다면 어느 정도의 관련성이 있는지를 알고자 할 때 특히 유용한 분석기법이다.

여기서, 두 변수간의 관련성을 나타내기 위해서는 요약된 지수를 계산하여 관련성의 정도를 계량화하는 것이 일반적이다. 이 때 자주 이용되는 척도로는 Pearson 상관계수(Pearson Correlation Coefficient,  $r$ )가 많이 이용되며,  $-1.0 \leq r \leq 1.0$ 의 값을 갖는다.

### 2.2.2. 회귀분석(Regression analysis)

회귀분석이란 변수들 중 한 개를 종속변수로 나머지를 독립변수로 하여 이들 변수들이 서로 상관관계를 가질 때 이러한 관계를 규명하고자 하는 방법으로 가장 많이 이용되는 통계기법이다. 대부분의 회귀분석에서는 종속변수와 독립변수의 관계가 선형의 관계에 있다고 가정하고 주로 1차식으로 나타낸다.

이 때 모집단의 회귀모형은 다음 식(1)과 같이 나타낼 수 있다.

$$Y_i = \alpha + \beta X_i + \epsilon_i \quad (1)$$

여기서,  $\alpha$ 와  $\beta$ 를 회귀계수(Coefficient of regression)라고 한다.

그러나 모집단 전체를 대상으로 조사하여 모수의  $\alpha$ 와  $\beta$ 를 찾는 것은 실제로는 거의 불가능하므로 이들을 추정하기 위해서는 표본으로부터 구한 자료를 사용해야 하며, 표본의 회귀식은 다음 식(2)와 같이 나타낼 수 있다.

$$\hat{Y}_i = a + bX_i \quad (2)$$

여기서,  $a$ 와  $b$ 는 모집단회귀식의 회귀계수  $\alpha$ 와  $\beta$ 의 추정치이다.

또한 종속변수와 여러 개의 독립변수들 사이의 관계를 규명하고자 할 때 사용되는 통계학적 기법으로 중회귀분석이 있다. 그러나 중회귀분석은 독립변수의 수만 많아진 것이며, 그 이론은 기본적으로 단순회귀분석의 경우와 같다.

표본 중회귀식은 다음 식(3)과 같이 나타낼 수 있다.

$$\hat{Y}_i = a + b_1 X_{1i} + b_2 X_{2i} + b_3 X_{3i} + \dots + b_k X_{ki} \quad (3)$$

여기서,  $\hat{Y}_i$ 는  $\mu_{Y: X_1, X_2, \dots, X_k}$ 의 추정치,  $a$ 는 절편의 추정치,  $b_1, \dots, b_k$ 는 편회계수  $\beta_1, \dots, \beta_k$ 의 추정치이다.

그리고 표본회귀식의 적합도를 나타내는 것으로써 가장 많이 사용되는 것이 결정계수( $r^2$ )이다. 결정계수는 표본회귀식에 의하여 설명된 변동이 총변동에서 차지하는 상대적 크기를 나타내며, 결정계수가 1이라는 것은 잔차가 전혀 없어 모든 관찰치가 표본회귀식으로 설명된다는 의미이고, 반대로 0이라는 것은 회귀식이 변동을 전혀 설명하지 못하고 있다는 뜻이다.

2.2.3. 요인분석(Factor analysis)

요인분석은 여러 변수들 간의 상관관계를 기초로 하여 변수 속에 내재되어 있는 공통 차원을 찾아냄으로써 변수를 적은 수로 축약하기 위해 사용하는 통계 기법이다.

요인분석의 이론을 처음으로 확립한 Spearman은 어떤 자료의 표본상관행렬에서 발견되는 체계적인 효과에 주목하고, 다음 식(4)의 단일 인자모형과 같이 두 부분의 결합으로 구성되었다면 이와 같은 체계적 효과를 설명할 수 있다고 주장하였다.

$$X_i - \mu_i = \lambda_i F + \epsilon_i, \quad i=1, 2, \dots \quad (4)$$

여기서,  $F$ 는 모든  $X$ 에 걸쳐 공통적인 어떤 확률변수,  $\epsilon_i$ 는  $X_i$ 에만 관계되는 특수한 확률변수,  $\mu_i$ 는  $X_i$ 의 기대값이다.

위의 단일 인자모형을 일반화시켜서 여러 개의 인자를 고려한 Thurstone의 다중 인자모형에서  $i$ 번째 변수  $X_i$ 는 다음 식(5)와 같이 나타낼 수 있다.

$$X_i - \mu_i = \sum_{k=1}^m \lambda_{ik} F_k + \epsilon_i, \quad i=1, 2, \dots, p \quad (5)$$

또한 행렬을 이용하여 이들을 표현하면 다음 식(6)과 같다.

$$X - \mu = \Lambda F + \epsilon \quad (6)$$

$(p \times 1) \quad (p \times 1) \quad (p \times m) \quad (m \times 1) \quad (p \times 1)$

그리고 인자분석에서 보유되는 인자의 갯수를 결정하기 위해서는 여러 가지 판정기준들을 생각할 수 있는데, 고유값의 크기, 인자의 공헌도,  $\chi^2$  적합도검정, Scree 도형 등이 있다.

2.2.4. 군집분석(Cluster analysis)

군집분석은 실험이나 조사에서 얻어진 결과를 어떤 성질에 따라 몇 개의 군집으로 나누어 분류하여 집단간의 차이를 명확하게 하는데 이용되는 통계기법이다. 군집분석은 군집의 형태와 사용되는 거리의 척도에 따라 다양한 방법들이 있으며, 군집분석을 수행하기 위해서는 우선 군집되는 각 객체들 간의 상사성 또는 비상사성의 정도를 측정하는 척도가 필요하다. 이때 대부분의 군집방법들은 주어진 객체들 간의 상사나 비상사의 관계를 그 객체들 간의 거리라는 개념을 통해 나타낸다.

어떤 한 쌍의 벡터  $X'_i = (X_{i1}, X_{i2}, \dots, X_{ip})$ 와  $X'_j = (X_{j1}, X_{j2}, \dots, X_{jp})$  ( $i, j=1, \dots, N$ ) 사이의 비상사성을 측정하는 가장 기본적인 방법은  $X'_i$ 와  $X'_j$  간의 거리  $d_{ij} = d(X_i, X_j)$  측정하여 모든 객체에 대해 다음과 같은 크기( $N \times N$ )의 거리행렬  $D$

$$D = \begin{bmatrix} d_{11} & d_{12} & \dots & d_{1j} & \dots & d_{1N} \\ d_{21} & d_{22} & \dots & d_{2j} & \dots & d_{2N} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ d_{i1} & d_{i2} & \dots & d_{ij} & \dots & d_{iN} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ d_{M1} & d_{M2} & \dots & d_{Nj} & \dots & d_{NN} \end{bmatrix}$$

를 얻는 것이다. 여기서, 두 객체  $i$ 와  $j$  사이의 거리  $d_{ij}$ 는  $X_i$ 와  $X_j$ 에 대한 함수이다.

2.2.5. 최근린법(Nearest-neighbor method)

많은 데이터에서 숨겨진 패턴과 상관관계를 찾아내어 이를 의사결정에 이용하는 데이터 마이닝(Data mining)으로 자주 이용되는 방법으로써 신경망(Neural network), 카오스 이론(Chaos theory) 등이 있다. 본 연구에서는 그 중에서도 최근에 널리 관심을 받고있는 카오스 이론 중의 최근린법을 이용하여 시계열 예

측을 수행하였다.

카오스 이론이란 겉으로 보기에 한없이 무질서하고 불규칙해 보이면서도 나름대로 어떤 질서와 규칙성을 가지고 있는 현상들을 설명하려는 이론이다. 특히 최근린법은 이러한 규칙성이 없어 보이는 데이터로부터 어떤 규칙성을 파악하는데 매우 유용한 방법이다.

최근린법은 상태공간에서의 비선형 시계열 분석에 기초를 두고 있으며 예측을 위하여 공간적인 상관관계를 이용한다. 이 때 주어진 과거 자료주변의 벡터의 거동을 분석함으로써 예측을 수행한다. 다음 식(7)은 최근린법을 이용한 상태공간 모형을 나타낸다(권 등 (2005)).

$$v_{t+T} = a_0 + \sum_{j=1}^d a_j v_{t-(j-1)\tau} + \epsilon_t \quad (7)$$

여기서,  $d$ 는 Embedding Dimension,  $\tau$ 는 지체시간이다. 이 때 예측은 각 상태공간에 대한 매개변수들을 추정함으로써 가능하며, 이를 계산하기 위한 전 과정은 다음 Fig. 3과 같다.

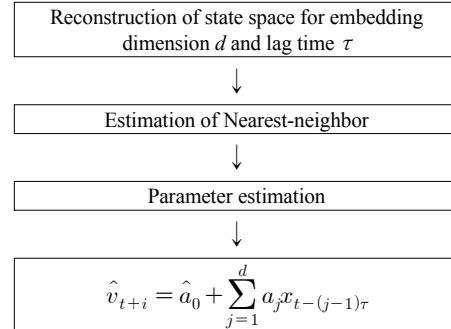


Fig. 3. Logic diagram of Nearest-neighbor method.

여기서, Nearest-neighbor의 거리(Distance)를 추정하기 위해서 가장 일반적인 방법으로 Euclidean 거리를 이용하였다.

### 3. 결과 및 고찰

#### 3.1. 자료의 통계적 특성치

본 연구의 대상하천인 낙동강 본류의 주요 지점인 안동, 왜관, 삼랑진 지점을 대상으로 기본적인 통계특

Table 1. Statisti of Andong

Division	Minimum	Maximum	Average	Standard deviation	Skewness
DO(mg/L)	5.2	15.6	9.530	2.339	-0.006
BOD(mg/L)	0.6	1.2	0.831	0.128	0.172
COD(mg/L)	2.4	5.3	3.548	0.537	0.905
SS(mg/L)	0.8	51.2	5.127	7.075	4.046
T-N(mg/L)	1.158	2.926	1.643	0.307	1.183
T-P(mg/L)	0.006	0.185	0.040	0.039	1.926
TEMP(°C)	5	24	13.17	5.631	0.009
Discharge(m <sup>3</sup> /s)	16.99	296.52	43.09	40.57	3.962

Table 2. Statisti of Waegwan

Division	Minimum	Maximum	Average	Standard deviation	Skewness
DO(mg/L)	6.1	15.1	10.045	2.249	0.325
BOD(mg/L)	0.9	3.2	1.726	0.554	1.006
COD(mg/L)	2.6	7.1	4.465	1.056	0.612
SS(mg/L)	1.4	76.3	15.049	14.650	1.941
T-N(mg/L)	1.848	4.604	2.980	0.601	0.391
T-P(mg/L)	0.031	0.245	0.106	0.044	0.792
TEMP(°C)	2	28	15.80	8.275	-0.163
Discharge(m <sup>3</sup> /s)	41.28	1161.55	189.80	209.31	2.872



**Table 6.** Correlation coefficient of Samrangjin

Division	DO	BOD	COD	SS	T-N	T-P	TEMP	Discharge
DO(mg/L)	1	0.679	0.477	-0.213	0.497	-0.015	-0.702	-0.415
BOD(mg/L)		1	0.632	-0.265	0.458	-0.023	-0.507	-0.353
COD(mg/L)			1	0.214	0.333	-0.001	-0.252	0.023
SS(mg/L)				1	-0.033	0.015	0.248	0.626
T-N(mg/L)					1	-0.012	-0.638	-0.218
T-P(mg/L)						1	0.049	0.014
TEMP(℃)							1	0.403
Discharge(m <sup>3</sup> /s)								1

그러나 이러한 상관분석 결과로부터 전절에서 언급한 바와 같이 일부 항목간에 상관성이 높은 경우도 있지만 대체로 유량과 수질항목들의 상관성은 하나의 변수와의 관계라기보다는 여러 항목들간의 복합적인 영향으로 지점마다 그 특성이 상이한 결과를 보여주고 있다고 판단된다.

**3.3. 회귀분석 결과**

위의 상관분석 결과로부터 여러 항목들간의 복합적인 경향을 파악하기 위해 주요 지점별 주요 수질항목인 DO, BOD, COD의 다중회귀식을 유도하였으며, 그 결과는 다음 Table 7~9와 같다. 그 결과를 분석해보면 각 수질항목별 다중회귀식의 상관계수가 상당히 높아지는 것으로 나타나 여러 항목들간의 복합적인 경향이 각각의 수질에 영향을 미치는 것을 확인할 수

있다.

**3.4. 요인분석 결과**

유량에 의한 수질의 변동특성은 비선형적인 관계를 가지므로 원자료에 대해 요인분석을 실시하였다. 여기서, 인자 개수의 결정을 위해서 고유값이 1 이상인 인자를 선정하였으며(Fig. 4~6), 주요 지점에 대한 결과는 다음 Table 10~12와 같았다.

먼저, 안동지점에서는 인자의 개수가 2개로 결정되었으며, 제1인자 그룹에서는 T-N, TEMP, 유량, BOD, 제2인자 그룹에서는 DO, T-P, SS, COD 항목이 포함되었다. 그리고 왜관지점에서는 인자의 개수가 3개로 결정되었으며, 제1인자 그룹에서는 TEMP, 제2인자 그룹에서는 T-N, T-P, SS, 유량, 제3인자 그룹에서는 DO, BOD, COD 항목이 포함되었다. 또한 삼랑진 지

**Table 7.** Multiple regression equation of water quality data in Andong

Station	Multiple regression equation	R
Andong	DO=13.948-0.673BOD+1.063COD+0.008SS-2.134TN-12.851TP-0.286TEMP+0.003Q	0.795
	BOD=0.426-0.004DO+0.103COD-0.007SS+0.055TN+0.191TP-0.001TEMP+0.001Q	0.486
	COD=0.632+0.066DO+1.020BOD+0.024SS+0.509TN+2.774TP+0.036TEMP-0.003Q	0.754

**Table 8.** Multiple regression equation of water quality data in Waegwan

Station	Multiple regression equation	R
Waegwan	DO=12.312+0.952BOD+0.215COD+0.007SS-0.344TN+4.067TP-0.262TEMP-0.001Q	0.841
	BOD=-1.461+0.089DO+0.330COD-0.012SS+0.205TN-1.763TP+0.043TEMP	0.741
	COD=2.530+0.053DO+0.883BOD+0.023SS-0.445TN+4.132TP+0.022TEMP	0.818

**Table 9.** Multiple regression equation of water quality data in Samrangjin

Station	Multiple regression equation	R
Samrangjin	DO=10.009+0.814BOD+0.396COD+0.007SS-0.162TN+0.192TP-0.183TEMP-0.001Q	0.812
	BOD=-1.156+0.108DO+0.434COD-0.016SS+0.192TN-0.048TP-0.002TEMP	0.818
	COD=1.481+0.094DO+0.777BOD+0.023SS+0.089TN+0.023TP+0.024TEMP	0.758

점에서는 인자의 개수가 3개로 결정되었으며, 제1인자 그룹에서는 DO, BOD, COD, T-N, 제2인자 그룹에서는 TEMP, SS, 유량, 제3인자 그룹에서는 T-P 항목이 포함되었다.

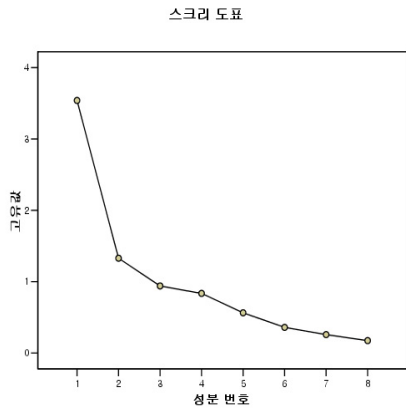


Fig. 4. Scree Diagram of Andong.

Table 10. Eigenvalue of Andong

Division	Raw data	
	Factor 1	Factor 2
T-N	0.810	0.219
TEMP	0.724	0.422
DO	-0.681	-0.369
Discharge	0.593	0.064
BOD	0.588	-0.095
T-P	-0.094	0.869
SS	0.223	0.866
COD	0.418	0.644

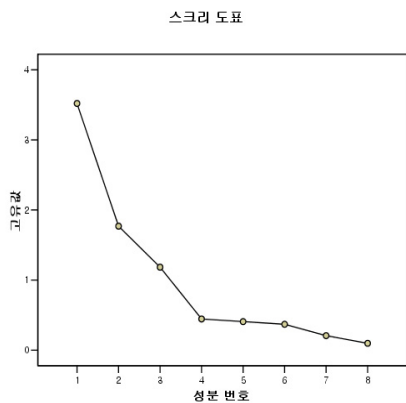


Fig. 5. Scree Diagram of Waegwan.

Table 11. Eigenvalue of Waegwan

Division	Raw data		
	Factor 1	Factor 2	Factor 3
DO	-0.876	-0.199	0.068
TEMP	0.851	0.233	0.355
T-N	-0.811	0.156	-0.328
T-P	-0.257	0.857	0.018
SS	0.343	0.784	0.101
Discharge	0.480	0.657	-0.144
BOD	0.050	-0.163	0.922
COD	0.363	0.413	0.747

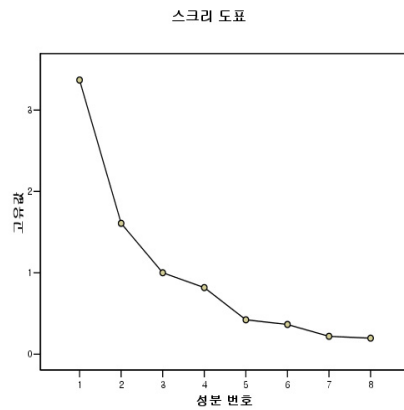


Fig. 6. Scree Diagram of Samrangjin.

Table 12. Eigenvalue of Samrangjin

Division	Raw data		
	Factor 1	Factor 2	Factor 3
DO	0.837	-0.256	0.001
BOD	0.830	-0.171	0.015
COD	0.751	0.390	0.043
TEMP	-0.740	0.349	0.071
T-N	0.730	-0.063	-0.028
SS	-0.030	0.895	0.002
Discharge	-0.276	0.831	-0.003
T-P	-0.017	0.000	0.997

위의 결과를 이용하여 수질 및 유량자료의 관계식을 구성한다면, 축약된 인자를 사용할 수 있으므로 의사결정시 좀더 신속하고 정확한 결과를 도출할 수 있다고 판단된다.

3.5. 군집분석 결과

본 연구에서는 다변량 자료를 각 특성의 유사성을 바탕으로 동질적인 몇 개의 군집으로 분류하기 위해



계층적 군집방법(Hierarchical clustering method)을 이용하여 분석하였다. 여기서, 군집들이 병합 또는 분리되는 과정을 덴드로그램(Dendrogram)을 이용하여 표현하였으며, 2003~2009년까지 7년치 자료의 평균을 계절별로 나누어서 군집분석을 실행하였고, 그 결과는 Fig. 7과 같다. 이 때 3~5월 까지를 봄, 6~8월 까지를 여름, 9~11월까지를 가을, 12~2월까지를 겨울로 간주하였다.

분석 결과 봄에는 구미, 왜관, 현풍, 임해진 지점이 제 1군집, 안동, 고령지점이 제 2군집, 삼랑진, 구포 지점이 각각 제 3, 4군집으로 나타났으며, 여름에는 구미, 고령, 현풍, 왜관지점이 제 1군집, 임해진, 구포지점이 제 2군집, 안동, 삼랑진 지점이 각각 제 3, 4군집으로 형성되었다. 그리고 가을에는 구미, 현풍, 왜관, 고령지점이 제 1군집, 안동, 임해진, 삼랑진, 구포지점이 각각 제 2, 3, 4군집으로 나타났고, 겨울에는 구미,

임해진, 현풍, 왜관, 안동, 고령지점이 제 1군집, 삼랑진, 구포지점이 제 2군집으로 분석되었다.

위의 결과를 이용한다면 이후 낙동강의 수질관리와 운영 및 관련 정책개발을 수행할 때 지점별로 개별적인 정책 제시보다는 유역을 유사한 성격의 그룹으로 나누어 정책을 제시할 수 있어 보다 효율적인 의사결정을 하는데 많은 도움이 될 것으로 판단된다.

3.6. 최근린법에 의한 수질예측 결과

3.6.1. DO

최근린법을 수질예측에 적용하기 위하여 주요 지점별 DO를 대상으로 2008년부터 2009년까지의 월별 자료를 이용하여 수질예측을 수행하였으며, 수질자료에 대해 계산치와 관측치를 비교하였다(Fig. 8). 분석 결과 대체적으로 각 지점에서 실측치와 계산치 사이에 근접하는 예측결과를 나타내었고 상관성이 비교적

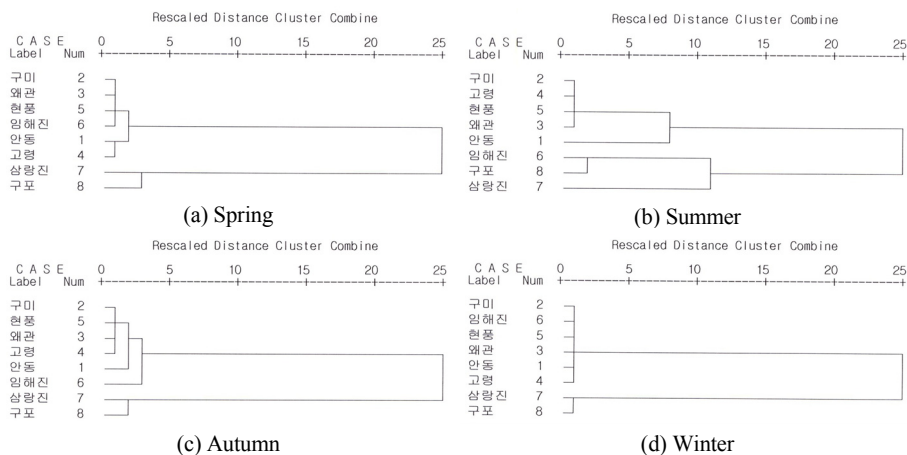


Fig. 7. Results of Cluster Analysis by each season.

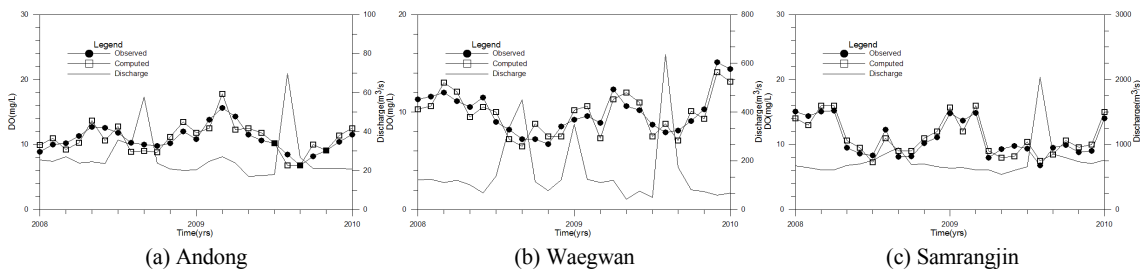


Fig. 8. Results of forecasting of DO by each Station.

높은 것으로 파악되었다.

3.6.2. BOD

최근린법을 수질예측에 적용하기 위하여 주요 지점별 BOD를 대상으로 2008년부터 2009년까지의 월별 자료를 이용하여 수질예측을 수행하였으며, 수질 자료에 대해 계산치와 관측치를 비교하였다(Fig. 9). 분석 결과 대체적으로 각 지점에서 실측치와 계산치 사이에 근접하는 예측결과를 나타내었고 상관성이 높은 것으로 파악되었다.

3.6.3. 적합성 검증

예측결과의 적합성을 판정하기 위하여 통계적인 검정을 실시하였다. 전질의 계산치와 관측치를 비교하기 위하여 모델의 효율도(Model Efficient, ME), 절대오차의 평균(Mean Square Error, MSE), 상대오차의 평균을 나타내는 Bias 및 체적오차의 백분율(Volume Error, VER)을 검토하였다. 그리고 이들 적합성 검증에 사용되는 방법들을 식으로 표시하면 아래 식(8)~(11)과 같다.

$$ME = 1 - \frac{\sum(Q_{obs} - Q_{est})^2}{\sum(Q_{obs} - Q_{est})^2} \quad (8)$$

$$MSE = \left[ \frac{\sum(Q_{obs} - Q_{est})^2}{n} \right]^{1/2} \quad (9)$$

$$Bias = \frac{\sum(Q_{obs} - Q_{est})}{n} \quad (10)$$

$$VER = \frac{\sum Q_{obs} - \sum Q_{est}}{\sum Q_{obs}} \times 100 \quad (11)$$

여기서, n은 자료의 수, Q<sub>obs</sub>는 관측치, Q<sub>est</sub>는 계산

치이다.

Table 13. Results of the statistical test

Factor	Station	ME	MSE	Bias	VER (%)
DO	Andong	0.871	1.263	1.596	-0.877
	Waegwan	0.841	1.076	1.158	0.967
	Samrangjin	0.844	1.057	1.117	-2.317
BOD	Andong	0.888	0.127	0.016	-5.670
	Waegwan	0.860	0.428	0.183	-5.172
	Samrangjin	0.917	0.380	0.144	-6.173

그리고 결과의 적합성을 판정하기 위하여 검정을 실시한 결과 Table 13과 같이 나타났다. 먼저, DO에 대하여 살펴보면 ME는 0.841~0.871, MSE는 1.057~1.263, Bias는 1.117~1.596, VER는 -2.317~0.967%를 나타내었다. 다음으로 BOD를 살펴보면 ME는 0.860~0.917, MSE는 0.127~0.428, Bias는 0.016~0.183, VER는 -6.173~-5.172%를 나타내었다.

이와 같이 Fig. 8~9 및 Table 13에 나타난 결과로 볼 때 본 연구에서 제시한 최근린법은 대체적으로 계산치와 관측치가 근접한 결과를 나타내어 예측방법으로써 적용성이 높음을 알 수 있었다.

4. 결론

본 연구에서는 낙동강의 수질관리와 효율적인 운영 및 관련 정책개발에 필요한 자료와 방법을 제시하기 위하여 본류 주요지점에서 유량과 수질자료를 수집하고 통계학적 방법을 이용하여 유량을 고려하였을 때의 수질자료의 특성 및 항목간의 상관성을 파악하

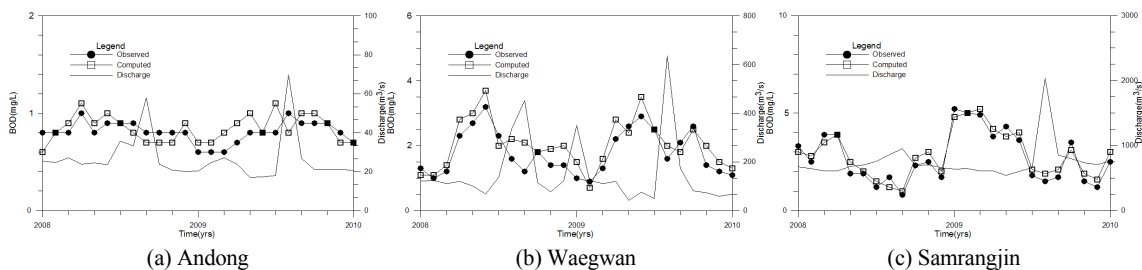


Fig. 9. Results of forecasting of BOD by each Station.

였으며, 또한 수질예측 방법의 적용성을 파악하는 과정에서 다음과 같은 결론을 얻었다.

1) 유량과 수질항목간의 상관성을 분석하기 위하여 상관분석을 실시하였다. 그 결과 유량과 수질항목들의 상관성은 하나의 변수와의 관계라기보다는 여러 항목들간의 복합적인 영향을 갖는 것으로 파악되었다.

2) 상관분석 결과로부터 주요 지점별 주요 수질항목인 DO, BOD, COD의 다중회귀식을 유도하였다. 그 결과 각 수질항목별 다중회귀식의 상관계수가 상당히 높아지는 것으로 나타나 여러 항목들간의 복합적인 경향이 각각의 수질에 영향을 미치는 것을 알 수 있다.

3) 유량에 의한 수질의 변동특성은 편차가 매우 큰 비선형적 관계를 가지므로 원자료에 대해 인자분석을 실시하였다. 인자 개수의 결정을 위해서 고유값이 1 이상인 인자를 선정하였으며, 변수를 적은 수의 구조로 축약하였다.

4) 군집분석을 위해서 계층적 군집방법(Hierarchical Clustering Method)을 이용하였으며, 여기서 군집들이 병합 또는 분리되는 과정을 덴드로그램(Dendrogram)을 이용하여 표현하였다. 2003~2009년까지 7년치 자료의 평균을 계절별로 나누어 군집분석을 실행한 결과, 수질 및 유량 자료의 상관성이 높은 군집으로 지점들을 분류하였다.

5) 최근린법을 수질예측에 적용하기 위하여 주요 지점별 주요 수질항목을 대상으로 2008년부터 2009년까지의 월별 자료를 이용하여 수질예측을 수행하였으며, 수질자료에 대해 계산치와 관측치를 비교하였다. 분석 결과 계산치와 관측치가 근접한 결과를 나타내어 수질예측의 상관성이 높음을 알 수 있었다.

이상과 같이 낙동강의 수질관리와 효율적인 운영 및 관련 정책개발에 필요한 자료를 얻기 위하여 통계적인 분석을 실시한 결과 그 효용성을 확인할 수 있었다. 이후 분석지점을 분류 뿐만 아니라 지류 등으로 확대하고, 다른 다양한 수질자료와 비점오염원에 영향을 미치는 강우량 등을 추가하여 분석을 실시한다면 더욱 신뢰성 높은 결과를 얻을 수 있을 것으로 판단된다.

### 감사의 글

이 논문은 2008학년도 영남이공대학 연구조성비 지원에 의한 것입니다.

### 참고 문헌

권혁윤, 1992, 하천 유량변동에 따른 수질변화에 관한 연구, 석사학위논문, 연세대학교 보건대학원.

권현한, 문영일, 2005, 상태-공간 모형과 Nearest Neighbor 방법을 통한 수문시계열 예측에 관한 연구, 대한토목학회논문집, 25(4B), 275-283.

김중구, 2002, 통계분석 기법을 이용한 금강수계의 수질 평가, 한국환경과학회지, 11(12), 1281-1289.

김준근, 1998, 낙동강 수질오염 특성조사를 통한 수질 개선 방안 모색, 석사학위논문, 경북대학교.

물환경정보시스템, 2003-2009, <http://water.nier.go.kr>, 환경부.

백경원, 1995, 한강하류부에서의 확정론 및 통계학적 기법에 의한 수질분석, 박사학위논문, 홍익대학교.

양우석, 1994, 하천유량 변화에 따른 수질오염도 분석, 석사학위논문, 수원대학교 산업경영대학원.

유희정, 1993, 수문인자에 의한 하천수질변화에 관한 연구, 한국수문학회지, 26(3), 75-85.

이상도, 1996, 중소하천의 유량과 수질변동 특성에 관한 연구, 석사학위논문, 홍익대학교.

이용환, 백경원, 한건연, 송재우, 1988, 인자분석법에 의한 강우와 하천수질변동 특성에 관한 연구, 대한토목학회논문집, 18(II-3), 263-276.

이형진, 공동수, 김상훈, 신기석, 박지형, 김병익, 김성미, 장승현, 천세억, 2007, 남한강수계 저수기 수질변동 특성에 관한 연구, 한국물환경학회지, 23(6), 889-896.

임영성, 조주식, 이홍재, 이영환, 허종수, 1999, 낙동강 수계의 수질현황, 한국환경농학회지, 18(2), 126-134.

정상만, 임경호, 최정현, 2000, 금강지류 유역에서의 유출량과 오염부하량의 상관관계 분석, 한국수자원학회지, 33(5), 527-536.

정용태, 심순섭, 이용환, 1999, 다변량분석법을 이용한 하천수질의 평가, 석사학위논문, 부경대학교.

Abahams, A. D., 1972, Factor Analysis of drainage basin properties: evidence for stream abstraction accompanying the degradation of relief, Water Resources Research, 8(3), 624-633.

Fang, H., Xiaoquan, W., Liping, L., Zhiqing, Z., Jiaping, W., 2010, Evaluation of spatial variation and source

- apportionment of water pollution in Qiantang River (China) using statistical techniques, *Water Research*.
- 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), *Journal of Environment Management*.
- Mehta, B. M., Ahlert, R. C., Yu, S. L., 1975, Stochastic variation of water quality of the Passaic River, *Water Resources Research*, 11(2), 300-308.
- Shrestha, F. K., 2007, Assessment of surface water quality using multivariate statistical technique: A case study of the Fuji river basin, Japan, *Environment Modelling & Software*.
- Thomann, R. V., 1976, Time series analysis of water quality data, *J. Sanit. Eng. Div., ASCE*, 93(1), 1-23.