

## 패턴분류를 위한 위락시설지역의 비점오염원 유출특성분석

### Analysis of Non-Point Source Pollution Discharge Characteristics in Leisure Facilities Areas for Pattern Classification

김 용 구\* / 진 영 훈\*\* / 박 성 천\*\*\* / 김 정 민\*\*\*\*

Kim, Yong Gu / Jin, Young Hoon / Park, Sung Chun / Kim, Jung Min

#### Abstract

In meteorology Korea has 2/3 of rain of annual total rainfall at the month of Jun through Sept and it has possibility to have serious flood damage because geographically it is composed of mountainous area with steep slope which account for 70% of its country. Also, the increase of impervious layer due to industrialization and urbanization causes direct runoff, which deteriorates contamination of rivers by moving the contaminated material on the surface at the beginning of rain. In particular, the area of leisure facilities needs the management of water quality absolutely because dense population requires space of park function and place to relax and increases moving capability of non-point pollution source. For disposition of rainfall & runoff, the standard of initial rainfall, which is to be used for the computation of disposition volume, is significant factors for the runoff study of non-point pollution source, Until now, a great deal of study has been done by many researchers. However, it is the current reality that the characteristics of runoff varies according to land protection comprising river basin and the standard of initial rainfall by each researcher is not clearly defined yet. Therefore, in this research, it is suggested that, with the introduction of SOM (Self-Organizing Map), the standard of initial rainfall be determined after analyzing each sectional data by executing pattern classification about runoff and water quality data measured at the test river basin for this research.

**Keywords** : area of leisure facilities, SOM, pattern classification, initial rainfall

#### 요 지

우리나라는 기상학적으로 연중 총강수량의 약 2/3가 6~9월에 편중해서 내리고 있고, 지형적으로 국토의 70% 이상이 산지로 구성되어 경사가 급해 수해를 입을 가능성이 매우 크다. 또한 산업화 및 도시화로 인해 불투수층의 증가로 강수량의 대부분이 직접유출로 기여해 강우초기에 노면상의 오염물질을 급속히 하천으로 이동시켜 오염을 가중시키고 있다. 특히 위락시설지역은 도시화에 따른 인구집중으로 지역민들에게 휴식공간 및 공원기능의 제공으로 비점오염원의 이동능력도 증가하여 수질관리가 절실히 필요하다. 따라서 본 연구는 위락시설지역의 비점오염원 유출특성분석을 위해 SOM (Self-Organizing Map) 이론을 도입하여 시험유역에서 측정된 유출 및 BOD, TOC자료에 대해 패턴분류를 수행하여 분

\* 동신대학교 토목공학과 연구원 (e-mail: kyg8987@paran.com)

Researcher, Department of Civil Engineering, Dongshin University, Naju 520-714, Korea

\*\* 교신저자, 수도대학동경 도시환경과학연구과 (e-mail: nmrdjin@gmail.com)

Corresponding Author, Researcher, Department of Urban Environmental Sciences, Tokyo Metropolitan University

\*\*\* 동신대학교 토목공학과 교수 (e-mail: psc@dsu.ac.kr)

Professor, Department of Civil Engineering, Dongshin University, Naju 520-714, Korea

\*\*\*\* 동신대학교 공과대학 토목공학과 박사과정 (e-mail: sin600866@hanmail.net)

Ph.D. student, Dept. of Civil Engineering, Dongshin University, Naju 520-714, Korea

할구역별 자료의 특성을 분석하였다. 그 결과 BOD 농도는 4개 TOC 농도는 각각 4개, 6개 분할구역으로 구분되었으며 BOD 농도의 경우 Cluster1과 Cluster2가 TOC 농도의 경우 Cluster4와 Cluster5가 비점오염원 유출 연구에 있어서 매우 중요한 요소인 초기강우 양상을 나타내었다. 또한 유출량 자료 수집과정에서 불가피하게 발생하는 이상치 자료를 정상적으로 측정된 수위와 유출량관계를 통해 추정된 회귀식을 이용하여 보완하여 이상치 자료를 활용하였다.

**핵심용어** : 위락시설지역, SOM, 패턴분류, 초기강우

## 1. 서 론

오늘날의 도시는 고도의 산업화와 더불어 인구의 도시 집중 현상 및 도시화로 인하여 대부분의 수질개선을 위한 정책이 점오염원의 처리에 편중되어 추진되어왔다. 그러나 20세기 중반이후 토지이용의 고도화에 따른 비점오염원의 증가로 인해 점 오염원의 처리만으로 해당 수계의 목표수질을 달성할 수 없다는 사실이 인식되면서 단일 수계내에서 오염총량관리를 지역에 따라 단계적으로 시행하고 있다. 오염총량관리제도에 의하면 배출농도규제만으로 오염부하의 양적증가를 통제할 수 없어 수질개선에 한계가 있으므로 점오염원의 규제와 더불어 비점오염원의 관리에도 중점을 두어야한다.

2004년 4월 오염총량관리제 실시에 있어서 비점오염원의 관리 없이는 목표수질 달성을 기대할 수 없음을 인식하여 오염총량관리제와 함께 2004년에 관계기관 합동으로 물관리 종합대책의 추진강화를 위한 「4대강 비점오염원관리 종합대책」을 발표하였고, 2005년 3월에 강우-유출수 관리에 대한 내용을 수질환경보전법에 추가함으로써 비점오염원 관리를 위한 연구 및 시범사업이 계획 및 진행 중이다(김이형과 이선화, 2005a; b).

비점오염원은 대부분이 강우 시에 유출되고 오염원의 경로가 명확하지 않으므로 오염원을 정량화하기에는 수많은 노력과 비용이 소요된다. 강우 시 유출되는 비점오염원의 영향은 삶의 질이 향상되고, 경제활동수준이 증가하고, 토지이용이 고밀도로 이용될수록 상대적으로 높아지고 있는 특성을 보이나 국내에서는 현재까지도 비점오염원에 대한 기초자료 확보가 미흡한 실정이다. 따라서 강우-유출수 처리용량 산정 등 비점오염원의 최적관리방안을 위해서는 토지피복별로 발생하는 오염물질의 특성 파악 및 초기강우 기준이 제시되어야 한다.

본 연구에서 적용한 SOM의 국외의 적용 예를 살펴보면 Hsu et al. (2002)은 Self-Organizing Linear Output map (SOLO)을 이용하여 강우-유출패턴을 구분하였고, Garcia and Gonzalez (2004)은 입력패턴의 분류를 위해 K-means 알고리즘을 이용하여 하수처리모니터링에 적용하였다. 국내의 적용 사례는 김용구 등 (2006)은 강우-유출과정의

양상에 따른 분류와 그에 따른 분석을 위해 Kohonen 네트워크 이론에 의한 SOM 이론을 적용하였고, 박성천 등 (2006)은 SOM에 의해 강우-유출관계를 분류하고, SOM에 의한 분류에 따라 각각의 예측모형을 구성하여 과거의 인공신경망의 일반적인 적용에 의한 결과보다 더 나은 예측능력을 보여주었고, 더불어 유출량의 과소 및 과대추정과 지속성과 같은 문제점을 해소한 연구결과를 발표하였다. 그러나 김용구 등 (2006)의 결과에 의하면 SOM 훈련을 위한 지도크기 및 배열의 결정은 SOM 성능에 영향을 미치는 것으로 보고되고 있으나, 지도크기 결정시 지도의 중방향 크기 및 횡방향 크기를 결정할 수 있는 확정론적인 방법이나 이론식이 없고, 지도배열은 일반적으로 육각형배열을 주로 이용하고 있는 것으로 보고되고 있다.

따라서 본 연구에서는 패턴분류 성능을 지닌 SOM 이론을 적용하여 본 연구의 시험유역에서 측정된 유출량 및 수질자료에 대한 패턴을 구분하여 분할구역별 자료의 특성분석을 통해 위락시설지역의 유출특성분석 및 초기강우 기준 제안을 위한 기초자료로 활용하고자 한다.

## 2. 시험유역 선정 및 모니터링 방법

### 2.1 시험유역선정

본 연구의 목적을 위한 시험유역 선정은 측정의 용이성, 유역면적, 우수 관 계통, 장기적 관측의 안전성 등 평가항목 및 평가기준을 작성하고 도 상작업에 의해 후보지를 선정하였고 현지조사를 통해 최종시험유역을 선정하였다.

이와 같은 과정에 의해 선정된 본 연구의 시험유역은 광주광역시 광산구 쌍암동 쌍암공원으로 전체면적은 28,332.72 m<sup>2</sup>이며, 투수면적은 26,819.66 m<sup>2</sup>, 불투수면적은 1,513.06 m<sup>2</sup>으로 대부분이 초지인 투수지역으로 구성되어 있고, 행사를 위한 야외무대와 놀이시설의 약 5% 정도의 불투수 지역으로 구성되어있다. 하수배제방식은 분류식으로 우수관거는 300, 450 mm의 관경으로 설치되어 있다.

### 2.2 모니터링 방법

비점오염원 유출특성분석을 위한 자료는 강우-유출 및 수질자료이다. 강우자료는 KDC-S13-R1-5의 전도 mass

방식의 우량계를 이용하였고, 유출량 자료는 비만관식 전 자식 유량계/Flo-Tote 3을 이용하여 실측에 의한 자료를 이용하여 분석하였다.

분석에 이용되는 수질자료는 15분, 30분, 1시간, 2시간 간격으로 채취한 시료를 공정시험법에 의해 분석한 BOD,

TOC 자료이다.

위락시설지역의 비점오염원 유출특성분석을 위해 2008년 4월 9일~2008년 10월 22일까지의 총 7회의 실측에 의한 자료를 이용하였으며, 측정자료에 대한 기본적인 통계 값과 유출률, 선행무강우일수 등을 Table 1에 나타내었다.

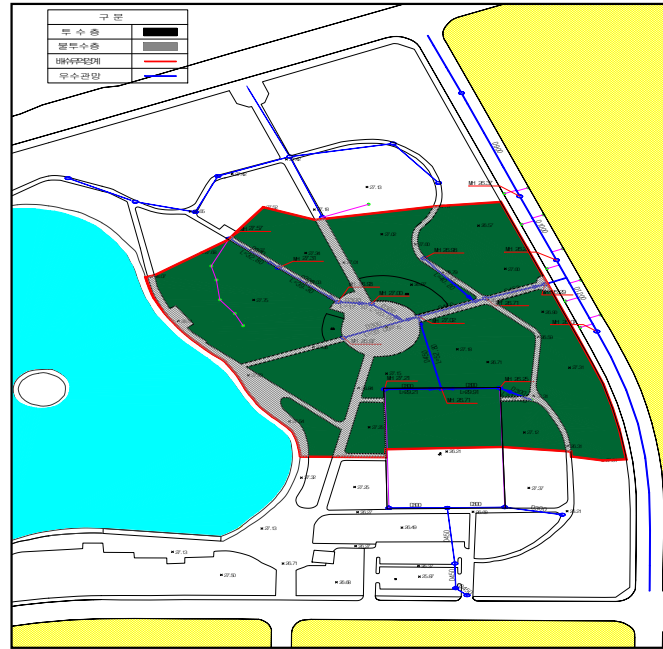


Fig. 1. Study Basin

Table 1. Result of Data Analysis Each Event

Part		Measured Date	2008/4/9	2008/5/28	2008/6/4	2008/6/18	2008/6/28	2008/7/25	2008/8/22
		Number of no rainy days (day)	2	9	6	8	6	3	9
Runoff rate (%)		26.7	49.6	37.0	67.7	40.9	49.1	21.9	
B O D	Max. (mg/L)	53.550	22.520	16.430	26.909	18.429	30.503	52.872	
	Min. (mg/L)	25.270	3.070	6.410	9.966	6.143	10.168	14.235	
	Average (mg/L)	36.275	9.292	10.237	17.504	11.813	17.709	36.604	
	standard deviation	9.632	4.936	2.897	4.841	3.562	5.792	14.338	
	Degree of distortion	0.816	1.504	1.043	0.229	0.080	0.738	-0.569	
	Kurtosis	-0.662	3.990	0.763	-0.683	-0.775	0.847	-1.216	
	Number of data	11	13	11	16	13	12	8	
T O C	Max. (mg/L)	19.480	3.790	5.930	9.291	5.844	3.804	4.837	
	Min. (mg/L)	2.790	1.410	2.120	1.660	1.560	1.629	1.835	
	Average (mg/L)	7.851	2.330	4.059	5.590	3.217	2.762	3.674	
	Standard deviation	5.757	0.814	1.290	2.560	1.405	0.672	1.044	
	Degree of distortion	1.478	0.728	-0.064	-0.421	0.396	-0.126	-0.747	
	Kurtosis	1.137	-0.865	-1.510	-1.238	-0.993	-0.631	-0.441	
	Number of data	11	13	11	16	13	12	8	

### 2.3 유량자료의 이상치 보완

본 시험유역의 본격적인 모니터링에 앞서 2007년 12월 예비조사를 통해 모니터링에 따른 문제점을 파악하여 기술하고자 한다.

Fig. 2는 본 시험유역의 측정의 개념도를 나타낸 그림으로 유출량 자료 수집에 있어 Fig. 3에 나타낸바와 같이 짧은 시간동안에 유량자료의 이상치가 발생하고 있다. 이와 같은 이상치는 와류의 영향과 특정지점에 설치된 구조물의 영향에 의한 것이고, 또한 유량의 급증으로 관로나 노면의 부유물이 유량계의 센서에 충격을 가하거나 흐름

에 방해줄 경우에도 유량자료의 이상치가 발생한다. 각 지점특성에 따라 발생하는 와류나 구조물 영향에 따른 이상치들은 다수의 시행착오를 거쳐 정상적인 유량자료를 확보할 수 있으나, 부유물의 영향으로 발생하는 이상치의 영향은 불가피하다.

따라서 짧은 시간간격에서 발생하는 유량자료의 이상치의 보정을 위해 Fig. 4에 나타낸 바와 같이 정상적으로 측정된 수위와 유량관계를 상승부와 하강부 자료로 각각 구분하여 회귀분석을 통해 추정된 회귀식을 이용하여 유량자료의 이상치를 보완하였다. 이와 같은 일련의 과정으로 Fig. 3과 같이 정상적인 수문곡선으로 보정하였다.

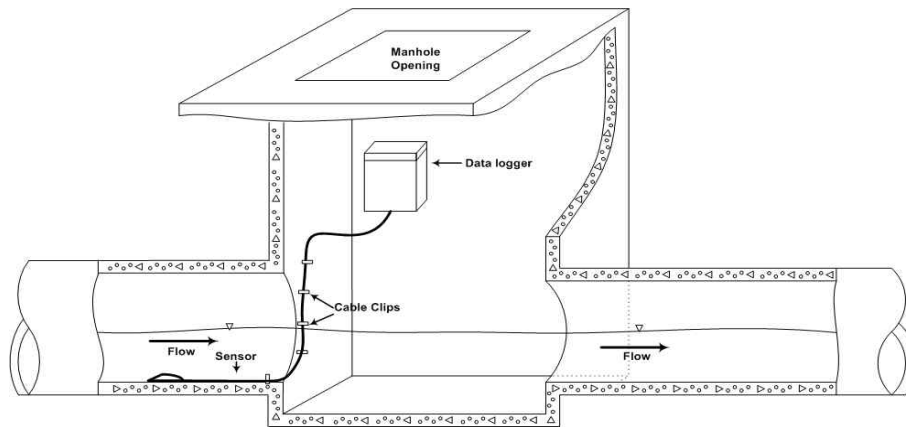


Fig. 2. Key Map of Measured Basin

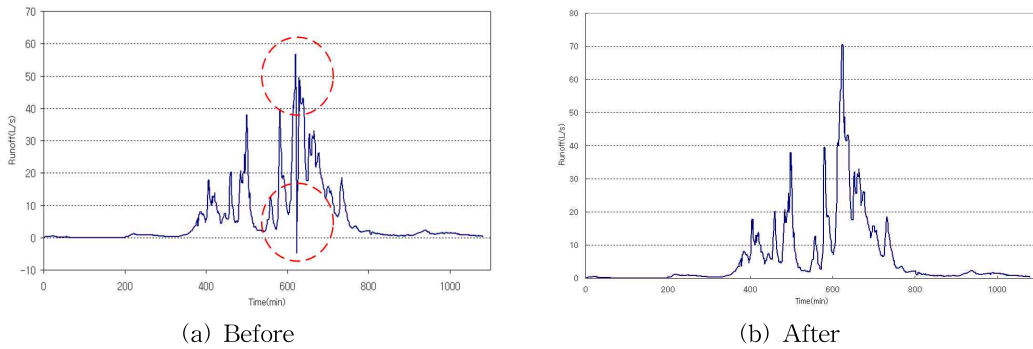


Fig. 3. Correctional Result of Abnormal Hydrograph

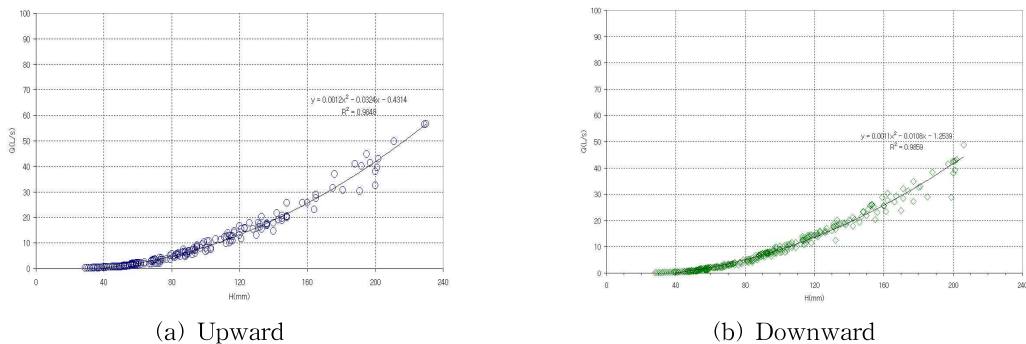


Fig. 4. Regression Analysis Result of Water Level-Discharge for Correcting Abnormal Value

### 3. SOM의 기본이론

#### 3.1 개요

SOM은 다차원의 자료를 2차원으로 사상(mapping)시켜주는 신경회로망의 한가지로서, 클러스터링을 위한 방법으로 많이 사용된다. 특히 SOM은 복잡한 다차원 자료의 클러스터링에 그 적용성이 뛰어난 방법으로 알려져 있으며, 자료의 가시화가 쉽고, 클러스터링 결과의 구조를 미리 지정해줄 수 있는 장점을 가지고 있다. 또한 입력 자료의 수가 많아도 빠른 시간 내에 양질의 결과를 얻을 수 있다.

SOM은 비교사 학습방법(unsupervised learning algorithm)의 일종으로 스스로 n차원의 입력 자료들을 클러스터링하여 2차원으로 사상시켜주며 Fig. 5에 SOM의 일반적인 구조를 나타내었다. 그림에 도시되어 있는 SOM 구조는 n차원의 입력자료를 표현하는 n개의 입력 노드들과 k개의 분류영역을 표현하기 위한 k개의 출력노드로 구성되어 있다. 모든 입력노드들은 모든 출력노드들과 연결되어 있고 각각의 연결강도(weight)를 가진다. 일반적으로 입력노드는 입력자료를 네트워크로 전달하는 기능을 하며, 출력노드는 전달된 입력벡터와 입·출력 노드들을 연결하는 연결강도벡터를 이용하여 두 벡터간의 거리를 계산하는 기능을 수행한다. Fig. 6은 입력노드 i와 출력노드 j를 연결하는 연결강도  $w_{ij}$ 들의 행렬을 보여준다.

이러한 과정에서 각 노드들은 학습할 수 있는 특권을 부여받기 위해 서로 경쟁하며, 거리가 가장 가까운 노드가 승리하게 된다. 결국 이 승자 노드만이 출력신호를 보

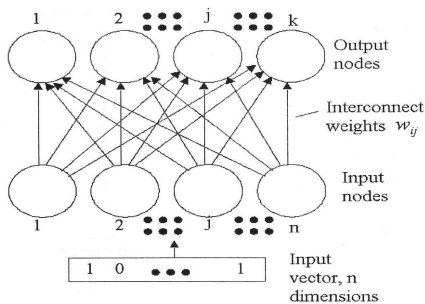


Fig. 5. General Structure of SOM

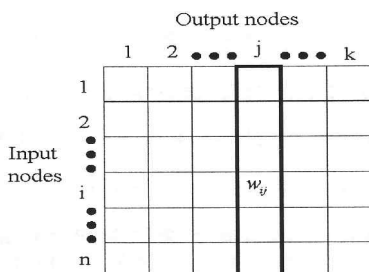


Fig. 6. Weight Matrix

낼 수 있는 유일한 노드가 된다. 승자 노드와 이에 인접한 이웃 노드들만이 제시된 입력벡터에 대하여 학습이 허용된다. 이것은 학습에 있어서 전혀 새로운 접근방식이며 이러한 모형이 제안되기 이전에는 네트워크에 있는 모든 노드들이 반복학습과정인 훈련을 통해 연결강도를 조정하는 방법을 이용해 왔다. 그러나 Kohonen 네트워크의 학습 철학은 승자만이 출력을 할 수 있는 ‘승자독점’ 개념이며 승자와 그의 이웃들만이 그들의 연결강도를 조정할 수 있다. 이를 위해 먼저 노드의 연결강도 벡터가 임의의 값을 가지면서 적합하게 초기화되어야 하며 각 노드는 다음 세 가지 단계의 중요한 처리 과정 3.2 경쟁과정, 3.3 근접반경의 조정과정, 3.4 적응학습 과정이 진행되는 동안 연결강도를 조정하게 된다.

#### 3.2 경쟁과정 (Competitive Process)

경쟁과정에서 각 노드들은 학습할 수 있는 특권을 갖기 위해 서로 경쟁하며 연결강도 벡터와 입력벡터의 거리가 가장 가까운 노드가 승자로 판정된다. n개의 입력을 가진 입력패턴과 노드 j의 연결강도 벡터를 다음과 같이 정의한다.

$$X = [x_1, x_2, \dots, x_n]^T \quad (1)$$

$$W_j = [w_{j1}, w_{j2}, \dots, w_{jn}]^T, \quad j = 1, 2, \dots, k \quad (2)$$

여기서, j은 전체 노드 수이다.

출력 노드 중의 승자 노드( $i(X)$ )은 다음의 조건에 의해 결정된다.

$$i(X) = \arg \min_j \|X - W_j\|, \quad j = 1, 2, \dots, k \quad (3)$$

위의 식은 입력벡터(X)와 연결강도벡터(W)간의 거리가 최소 일 때의 노드의 위치를 결정하여 i라는 변수를 나타내는 수학적 표현이다.

결국, 승자 노드의 선택은 입력벡터의 패턴과 가장 유사한 연결강도 벡터를 선정하는 것이며, 유사도(similarity matching) 측정을 위해 유클리드 거리(Euclidean distance)를 이용한다.

#### 3.3 근접반경의 조정과정 (Cooperative Process)

코호넨의 시스템은 생물학적 모델에서 보여지는 바와 유사하게 이웃한 노드와의 조정과정(cooperative process)을 통한 측면제어(lateral inhibition)를 사용한다. 즉 경쟁과정에서 승자가 된 노드와 함께 그에 인접한 노드들에게도 제시된 입력벡터에 대한 학습이 허용되며 인접노드를 결정하는 반경은 학습이 진행됨에 따라 서서히 줄어들어 점점 적은 개수의 노드들이 학습을 하게 된다. 최종적으로

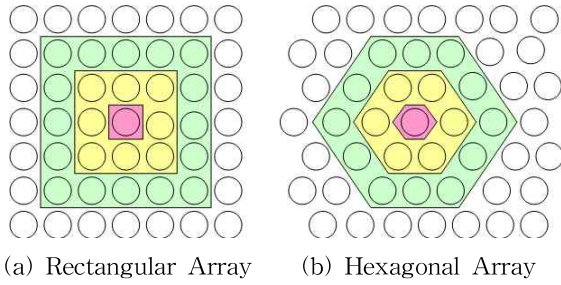


Fig. 7. Type of Map Array

는 단지 승자 노드만이 그것의 연결강도를 조정하게 된다.

이러한 과정을 위해 출력층의 노드는 1차원으로 배열하는 방법과 2차원으로 배열하는 방법이 있으며, 특히 2차원 배열에는 Fig. 7에서와 같이 사각형배열 (rectangular array)과 육각형배열 (hexagonal array)의 2가지 형태가 있다.

기하학적 반경을 조정하는 과정에 있어 대칭성과 수렴 특성을 지닌 가우시안 함수 (Gaussian function)를 이용하였다. 기하학적 이웃 반경을 정의하는  $h_{j,i}(x)$ 는 이웃한 노드와의 거리를 나타내는  $d_{j,i}$ 와 함께 다음과 같이 나타낼 수 있다.

$$h_{j,i}(x)(t) = \exp\left(-\frac{d_{j,i}^2(x)}{2\sigma^2(t)}\right), \quad t = 0, 1, 2, \dots, \quad (4)$$

여기서,  $d_{j,i}(x)$ 는 거리 벡터  $r_j$ 와 승자노드  $i$ 에 의해 정의되는  $r_i$ 에 의해 다음과 같이 정의된다.

$$d_{j,i}^2(x) = \|r_j - r_i\|^2 \quad (5)$$

그리고 일반적으로 지수적 감쇠를 위한  $\sigma$ 는 다음과 같이 선택할 수 있다.

$$\sigma(t) = \sigma_0 \exp\left(-\frac{t}{\tau_1}\right), \quad t = 0, 1, 2, \dots, \quad (6)$$

Eq. (6)에서  $\sigma_0$ 는 SOM 알고리즘의 초기화에서 갖는  $\sigma$

의 값이다.

### 3.4 적응학습 과정 (Adaptive Process)

이상과 같은 단계의 처리가 끝나면 마지막으로 시냅스의 적응과정과 같은 실제 연결강도의 조정이 이루어진다. 조정되기 이전의 연결강도 벡터를  $W_j(t)$ , 조정된 후의 새로운 연결강도 벡터를  $W_j(t+1)$ 와 같이 이산 시간  $t$ 에 대하여 정의할 때 조정을 위한 규칙은 다음 식으로 표현된다.

$$W_j(t+1) = W_j(t) + \eta(t) \cdot h_{j,i}(x)(t) \cdot [X - W_j(t)] \quad (7)$$

여기서,  $\eta$ 는 시간  $t$ 이 증가함에 따라 서서히 감소하는 학습률을 나타내는 매개변수로서 초기값  $\eta_0$ 와 지수적 감쇠를 만족시킬 수 있도록 다음과 같은 식으로 나타낼 수 있다.

$$\eta(t) = \eta_0 \exp\left(-\frac{t}{\tau_2}\right), \quad t = 0, 1, 2, \dots, \quad (8)$$

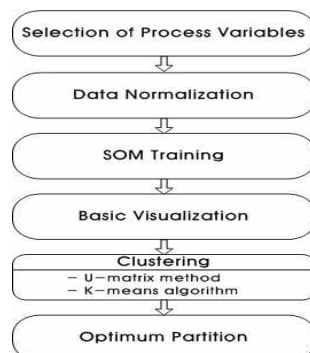
## 4. 결과 및 고찰

### 4.1 SOM 모형의 적용

본 연구에서는 위락시설지역에 대해 유량에 대한 수질의 변동 특성을 분석하기 위하여 SOM이론을 적용하였다. SOM의 적용은 Fig. 8의 SOM 적용 흐름도와 같고, 모형의 입력자료는 Event별 측정자료에 대해 순위를 부여하고, 실측에 의한 유출량 및 수질자료의 총 84개 자료를 이용하여 입력자료를 구축하였으며 Eq. (9)와 같다.

본 분석에서 측정자료에 대한 순위부여의 의미는 1~4번은 유출이 시작하여 15분 간격으로 채수한 시료를 나타낸 것이고, 5~7번은 유출이 시작하여 1시간이 지난 후 30분 간격으로 측정된 시료이며, 8번 이후는 유출 시작 후 2시간 30분이 지난 후 1시간 간격으로 측정된 시료를 나타낸다.

선택된 자료의 정규화 과정을 거친 후 SOM 훈련을 위



$$\begin{aligned} x(t) &= [Q, N \text{ of } D, \text{ BOD}] \\ x(t) &= [Q, N \text{ of } D, \text{ TOC}] \end{aligned} \quad (9)$$

여기서, Q는 유출량, N of D는 시료채취 순번

Fig. 8. SOM Flow Chart

해 지도크기를 결정하는데 기존의 연구에서는 지도를 구성하는 단위구조의 총 수(M)의 결정은 Garcia et al. (2004)에 의해 연구 보고된 식인  $M = 5\sqrt{N}$  식을 이용하였다. 여기서, M은 지도를 구성하는 단위구조(unit)의 총 수이며, N은 훈련 자료의 수이다. 본 연구에서 수집된 자료에 대해 SOM 적용을 위한 지도의 구성은 9×5의 Map 구조로 결정하였다.

#### 4.2 적용결과

위락시설지역의 Map 구조로 결정된 9×5의 지도구조에 대해 각각의 입력자료를 K-means 알고리즘 방법을 적용하여 구분된 분할영역을 Fig. 9에 도시하였다. 그 결과 BOD의 경우는 4개의 분할영역으로 구분되었고, TOC는 6개의 분할영역으로 구분되었다.

본 연구의 목적인 초기강우 기준을 검토하기 위해 수질 항목별 분할구역에 대한 자료의 box plot 결과와 기본적인 통계치를 토대로 검토하였다.

##### 4.2.1 BOD

수질자료 BOD 농도에 대한 분할구역별 자료의 통계치와 box plot의 결과는 Table 2와 같다. 분할구역별 특성을 살펴보면 Cluster1은 유량의 범위는 0.16~22.45 l/sec, BOD 농도의 범위는 8.19~34.75 mg/L로 구분되었고, 자료순번으로 유출이 발생한지 1시간 이전의 자료들로 구성되어 있다. Cluster 2의 경우는 유량의 범위는 0.02~22.73 l/sec, BOD 농도의 범위는 25.27~53.55 mg/L로 가장 큰 BOD 농도의 분포를 보이고 있다. Cluster1과 Cluster2는 유출량의 범위는 유사하나 BOD 농도의 범위로 서로 다

른 분할구역으로 구분되어짐을 확인할 수 있다. 일반적으로 초기강우에 의해 높은 수질농도를 나타내는 도시지역의 비점오염원 유출특성과는 다른 일정시간이 지난 후에 BOD 최대 농도를 나타내는 유출특성을 나타내므로 본 시험구역의 BOD농도의 초기강우 기준을 Cluster1과 Cluster2까지의 자료로 확대해야 할 것으로 판단된다.

Cluster3의 경우는 유량의 범위는 25.10~76.22 l/sec, BOD 농도의 범위는 3.07~35.59 mg/L로 가장 큰 유출량의 범위를 나타내고 있으며, 시료채취 시간으로는 유출발생 후 2시간 30분 후의 자료들로 구성되어 있다. Cluster4의 경우는 유량의 범위는 0.26~11.21 l/sec, BOD 농도의 범위는 3.07~24.40 mg/L로 강우가 종료되어 유출이 종료되기 전의 자료들로 구성되어 있다. 이와 같은 결과로 유량과 BOD농도 변화의 양상을 유추해 보면 Cluster1→Cluster2→Cluster3→Cluster4의 양상으로 확연히 구분되어진다.

##### 4.2.2 TOC

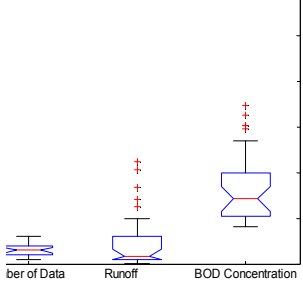
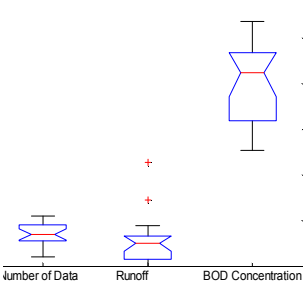
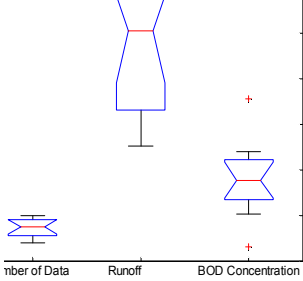
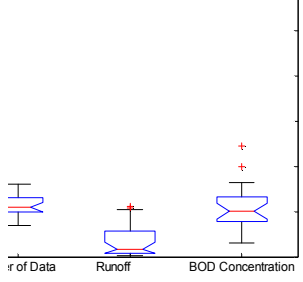
유량과 TOC의 분할구역별은 6개로 구분되어졌으며 분할구역별 자료의 통계치와 box plot의 결과는 Table 3과 같으며, 분할구역별 특성을 살펴보면 Cluster1은 유량의 범위는 0.26~25.10 l/sec, TOC 농도의 범위는 1.43~5.31 mg/L로 구분되었고, Cluster2의 경우는 유량의 범위는 0.02~11.21 l/sec, TOC 농도의 범위는 3.46~8.18 mg/L로 강우 종료 후에 측정된 후반기에 측정된 자료들로 구분되어있다. Cluster3의 경우는 유량의 범위는 25.25~76.22 l/sec로 첨두부를 나타내고 있으며, TOC 농도의 범위는 1.84~8.25 mg/L로 구분되었다. Cluster4의 경우

(a) BOD

(b) TOC

Fig. 9. K-means Algorithm Results

Table 2. Analysis Result of Water Quality BOD

	<table border="1"> <thead> <tr> <th>Part</th> <th>Data No.</th> <th>Discharge ( ℓ/s)</th> <th>BOD (mg/L)</th> </tr> </thead> <tbody> <tr> <td>Min.</td> <td>1.00</td> <td>0.16</td> <td>8.19</td> </tr> <tr> <td>Max.</td> <td>6.00</td> <td>22.45</td> <td>34.57</td> </tr> <tr> <td>Average</td> <td>2.97</td> <td>4.85</td> <td>17.40</td> </tr> <tr> <td>Median</td> <td>3.00</td> <td>1.71</td> <td>14.28</td> </tr> <tr> <td>Standard Deviation</td> <td>1.60</td> <td>6.27</td> <td>7.70</td> </tr> <tr> <td>Number of Data</td> <td>30</td> <td>30</td> <td>30</td> </tr> </tbody> </table>	Part	Data No.	Discharge ( ℓ/s)	BOD (mg/L)	Min.	1.00	0.16	8.19	Max.	6.00	22.45	34.57	Average	2.97	4.85	17.40	Median	3.00	1.71	14.28	Standard Deviation	1.60	6.27	7.70	Number of Data	30	30	30
Part	Data No.	Discharge ( ℓ/s)	BOD (mg/L)																										
Min.	1.00	0.16	8.19																										
Max.	6.00	22.45	34.57																										
Average	2.97	4.85	17.40																										
Median	3.00	1.71	14.28																										
Standard Deviation	1.60	6.27	7.70																										
Number of Data	30	30	30																										
Cluster1																													
	<table border="1"> <thead> <tr> <th>구분</th> <th>Data No.</th> <th>Discharge ( ℓ/s)</th> <th>BOD (mg/L)</th> </tr> </thead> <tbody> <tr> <td>Min.</td> <td>2.00</td> <td>0.02</td> <td>25.27</td> </tr> <tr> <td>Max.</td> <td>11.00</td> <td>22.73</td> <td>53.55</td> </tr> <tr> <td>Average</td> <td>7.15</td> <td>5.48</td> <td>40.26</td> </tr> <tr> <td>Median</td> <td>7.00</td> <td>5.09</td> <td>42.40</td> </tr> <tr> <td>Standard Deviation</td> <td>2.39</td> <td>5.17</td> <td>8.66</td> </tr> <tr> <td>Number of Data</td> <td>20</td> <td>20</td> <td>20</td> </tr> </tbody> </table>	구분	Data No.	Discharge ( ℓ/s)	BOD (mg/L)	Min.	2.00	0.02	25.27	Max.	11.00	22.73	53.55	Average	7.15	5.48	40.26	Median	7.00	5.09	42.40	Standard Deviation	2.39	5.17	8.66	Number of Data	20	20	20
구분	Data No.	Discharge ( ℓ/s)	BOD (mg/L)																										
Min.	2.00	0.02	25.27																										
Max.	11.00	22.73	53.55																										
Average	7.15	5.48	40.26																										
Median	7.00	5.09	42.40																										
Standard Deviation	2.39	5.17	8.66																										
Number of Data	20	20	20																										
Cluster2																													
	<table border="1"> <thead> <tr> <th>구분</th> <th>Data No.</th> <th>Discharge ( ℓ/s)</th> <th>BOD (mg/L)</th> </tr> </thead> <tbody> <tr> <td>Min.</td> <td>4.00</td> <td>25.10</td> <td>3.07</td> </tr> <tr> <td>Max.</td> <td>10.00</td> <td>76.22</td> <td>35.59</td> </tr> <tr> <td>Average</td> <td>7.25</td> <td>48.24</td> <td>17.81</td> </tr> <tr> <td>Median</td> <td>7.50</td> <td>50.39</td> <td>17.70</td> </tr> <tr> <td>Standard Deviation</td> <td>1.83</td> <td>16.93</td> <td>7.76</td> </tr> <tr> <td>Number of Data</td> <td>12</td> <td>12</td> <td>12</td> </tr> </tbody> </table>	구분	Data No.	Discharge ( ℓ/s)	BOD (mg/L)	Min.	4.00	25.10	3.07	Max.	10.00	76.22	35.59	Average	7.25	48.24	17.81	Median	7.50	50.39	17.70	Standard Deviation	1.83	16.93	7.76	Number of Data	12	12	12
구분	Data No.	Discharge ( ℓ/s)	BOD (mg/L)																										
Min.	4.00	25.10	3.07																										
Max.	10.00	76.22	35.59																										
Average	7.25	48.24	17.81																										
Median	7.50	50.39	17.70																										
Standard Deviation	1.83	16.93	7.76																										
Number of Data	12	12	12																										
Cluster3																													
	<table border="1"> <thead> <tr> <th>구분</th> <th>Data No.</th> <th>Discharge ( ℓ/s)</th> <th>BOD (mg/L)</th> </tr> </thead> <tbody> <tr> <td>Min.</td> <td>7.00</td> <td>0.26</td> <td>3.07</td> </tr> <tr> <td>Max.</td> <td>16.00</td> <td>11.21</td> <td>24.40</td> </tr> <tr> <td>Average</td> <td>11.14</td> <td>3.70</td> <td>11.06</td> </tr> <tr> <td>Median</td> <td>11.00</td> <td>1.62</td> <td>10.10</td> </tr> <tr> <td>Standard Deviation</td> <td>2.36</td> <td>3.69</td> <td>4.82</td> </tr> <tr> <td>Number of Data</td> <td>22</td> <td>22</td> <td>22</td> </tr> </tbody> </table>	구분	Data No.	Discharge ( ℓ/s)	BOD (mg/L)	Min.	7.00	0.26	3.07	Max.	16.00	11.21	24.40	Average	11.14	3.70	11.06	Median	11.00	1.62	10.10	Standard Deviation	2.36	3.69	4.82	Number of Data	22	22	22
구분	Data No.	Discharge ( ℓ/s)	BOD (mg/L)																										
Min.	7.00	0.26	3.07																										
Max.	16.00	11.21	24.40																										
Average	11.14	3.70	11.06																										
Median	11.00	1.62	10.10																										
Standard Deviation	2.36	3.69	4.82																										
Number of Data	22	22	22																										
Cluster4																													

는 유량의 범위는 0.16~9.66 ℓ/sec, TOC 농도의 범위는 1.60~7.22 mg/L로 유출이 발생하지 1시간 이전의 자료들로 구성되어 있다. Cluster5의 경우는 유량의 범위는 9.96~22.45 ℓ/sec, TOC 농도의 범위는 1.41~3.94 mg/L로 자료순번으로 볼 때 Cluster4와 유사하게 유출이 발생하지 1시간 이전의 유사한 자료들로 구성되어 있으나 강우특성에 지배를 받는 유출의 양상에 의해 Cluster4와 Cluster5

가 구분되어짐을 알 수 있다. Cluster6의 경우는 유량의 범위는 1.42~8.92 ℓ/sec, TOC 농도의 범위는 8.16~19.48 mg/L로 구분되었다. TOC 농도 변화의 양상은 BOD 농도와 달리 다소 혼재된 양상을 보이거나 Cluster4, Cluster5 → Cluster3 → Cluster1, Cluster6 → Cluster2의 양상으로 구분되어진다.



Table 3. Analysis Result about Water Quality Items TOC

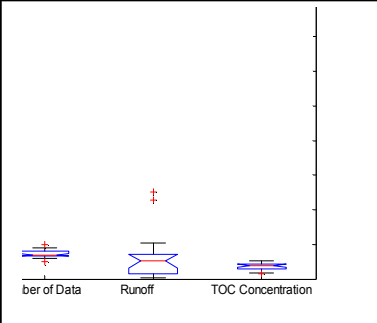
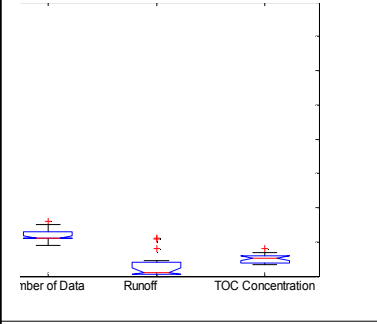
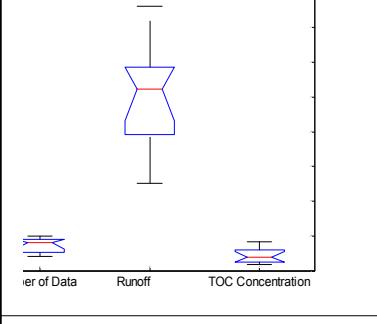
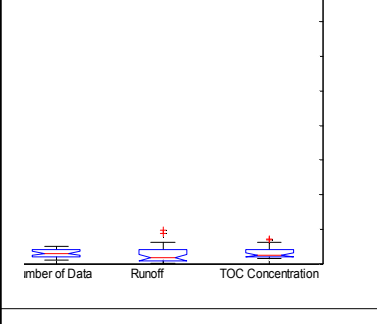
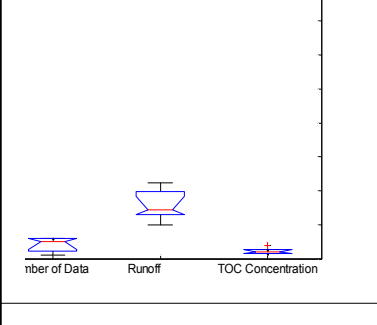
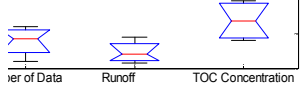
	<table border="1"> <thead> <tr> <th>Part</th> <th>Data No.</th> <th>Discharge (ℓ/s)</th> <th>TOC (mg/L)</th> </tr> </thead> <tbody> <tr> <td>Min.</td> <td>5.00</td> <td>0.26</td> <td>1.43</td> </tr> <tr> <td>Max.</td> <td>10.00</td> <td>25.10</td> <td>5.31</td> </tr> <tr> <td>Average</td> <td>7.47</td> <td>6.62</td> <td>3.66</td> </tr> <tr> <td>Median</td> <td>7.00</td> <td>5.26</td> <td>3.93</td> </tr> <tr> <td>Standard Deviation</td> <td>1.33</td> <td>6.89</td> <td>1.03</td> </tr> <tr> <td>Number of Data</td> <td>17</td> <td>17</td> <td>17</td> </tr> </tbody> </table>	Part	Data No.	Discharge (ℓ/s)	TOC (mg/L)	Min.	5.00	0.26	1.43	Max.	10.00	25.10	5.31	Average	7.47	6.62	3.66	Median	7.00	5.26	3.93	Standard Deviation	1.33	6.89	1.03	Number of Data	17	17	17
Part	Data No.	Discharge (ℓ/s)	TOC (mg/L)																										
Min.	5.00	0.26	1.43																										
Max.	10.00	25.10	5.31																										
Average	7.47	6.62	3.66																										
Median	7.00	5.26	3.93																										
Standard Deviation	1.33	6.89	1.03																										
Number of Data	17	17	17																										
Cluster1																													
	<table border="1"> <thead> <tr> <th>Part</th> <th>Data No.</th> <th>Discharge (ℓ/s)</th> <th>TOC (mg/L)</th> </tr> </thead> <tbody> <tr> <td>Min.</td> <td>9.00</td> <td>0.02</td> <td>3.46</td> </tr> <tr> <td>Max.</td> <td>16.00</td> <td>11.21</td> <td>8.18</td> </tr> <tr> <td>Average</td> <td>11.79</td> <td>2.87</td> <td>5.25</td> </tr> <tr> <td>Median</td> <td>11.00</td> <td>1.16</td> <td>5.24</td> </tr> <tr> <td>Standard Deviation</td> <td>1.82</td> <td>3.39</td> <td>1.45</td> </tr> <tr> <td>Number of Data</td> <td>19</td> <td>19</td> <td>19</td> </tr> </tbody> </table>	Part	Data No.	Discharge (ℓ/s)	TOC (mg/L)	Min.	9.00	0.02	3.46	Max.	16.00	11.21	8.18	Average	11.79	2.87	5.25	Median	11.00	1.16	5.24	Standard Deviation	1.82	3.39	1.45	Number of Data	19	19	19
Part	Data No.	Discharge (ℓ/s)	TOC (mg/L)																										
Min.	9.00	0.02	3.46																										
Max.	16.00	11.21	8.18																										
Average	11.79	2.87	5.25																										
Median	11.00	1.16	5.24																										
Standard Deviation	1.82	3.39	1.45																										
Number of Data	19	19	19																										
Cluster2																													
	<table border="1"> <thead> <tr> <th>Part</th> <th>Data No.</th> <th>Discharge (ℓ/s)</th> <th>TOC (mg/L)</th> </tr> </thead> <tbody> <tr> <td>Min.</td> <td>4.00</td> <td>25.25</td> <td>1.84</td> </tr> <tr> <td>Max.</td> <td>10.00</td> <td>76.22</td> <td>8.25</td> </tr> <tr> <td>Average</td> <td>7.27</td> <td>50.35</td> <td>4.30</td> </tr> <tr> <td>Median</td> <td>8.00</td> <td>52.30</td> <td>3.83</td> </tr> <tr> <td>Standard Deviation</td> <td>1.91</td> <td>16.11</td> <td>2.10</td> </tr> <tr> <td>Number of Data</td> <td>11</td> <td>11</td> <td>11</td> </tr> </tbody> </table>	Part	Data No.	Discharge (ℓ/s)	TOC (mg/L)	Min.	4.00	25.25	1.84	Max.	10.00	76.22	8.25	Average	7.27	50.35	4.30	Median	8.00	52.30	3.83	Standard Deviation	1.91	16.11	2.10	Number of Data	11	11	11
Part	Data No.	Discharge (ℓ/s)	TOC (mg/L)																										
Min.	4.00	25.25	1.84																										
Max.	10.00	76.22	8.25																										
Average	7.27	50.35	4.30																										
Median	8.00	52.30	3.83																										
Standard Deviation	1.91	16.11	2.10																										
Number of Data	11	11	11																										
Cluster3																													
	<table border="1"> <thead> <tr> <th>Part</th> <th>Data No.</th> <th>Discharge (ℓ/s)</th> <th>TOC (mg/L)</th> </tr> </thead> <tbody> <tr> <td>Min.</td> <td>1.00</td> <td>0.16</td> <td>1.60</td> </tr> <tr> <td>Max.</td> <td>5.00</td> <td>9.66</td> <td>7.22</td> </tr> <tr> <td>Average</td> <td>2.81</td> <td>2.49</td> <td>3.19</td> </tr> <tr> <td>Median</td> <td>3.00</td> <td>1.64</td> <td>2.48</td> </tr> <tr> <td>Standard Deviation</td> <td>1.33</td> <td>2.51</td> <td>1.76</td> </tr> <tr> <td>Number of Data</td> <td>26</td> <td>26</td> <td>26</td> </tr> </tbody> </table>	Part	Data No.	Discharge (ℓ/s)	TOC (mg/L)	Min.	1.00	0.16	1.60	Max.	5.00	9.66	7.22	Average	2.81	2.49	3.19	Median	3.00	1.64	2.48	Standard Deviation	1.33	2.51	1.76	Number of Data	26	26	26
Part	Data No.	Discharge (ℓ/s)	TOC (mg/L)																										
Min.	1.00	0.16	1.60																										
Max.	5.00	9.66	7.22																										
Average	2.81	2.49	3.19																										
Median	3.00	1.64	2.48																										
Standard Deviation	1.33	2.51	1.76																										
Number of Data	26	26	26																										
Cluster4																													
	<table border="1"> <thead> <tr> <th>Part</th> <th>Data No.</th> <th>Discharge (ℓ/s)</th> <th>TOC (mg/L)</th> </tr> </thead> <tbody> <tr> <td>Min.</td> <td>1.00</td> <td>9.96</td> <td>1.41</td> </tr> <tr> <td>Max.</td> <td>6.00</td> <td>22.45</td> <td>3.94</td> </tr> <tr> <td>Average</td> <td>4.14</td> <td>15.85</td> <td>2.22</td> </tr> <tr> <td>Median</td> <td>5.00</td> <td>14.39</td> <td>2.02</td> </tr> <tr> <td>Standard Deviation</td> <td>1.96</td> <td>4.11</td> <td>0.84</td> </tr> <tr> <td>Number of Data</td> <td>7</td> <td>7</td> <td>7</td> </tr> </tbody> </table>	Part	Data No.	Discharge (ℓ/s)	TOC (mg/L)	Min.	1.00	9.96	1.41	Max.	6.00	22.45	3.94	Average	4.14	15.85	2.22	Median	5.00	14.39	2.02	Standard Deviation	1.96	4.11	0.84	Number of Data	7	7	7
Part	Data No.	Discharge (ℓ/s)	TOC (mg/L)																										
Min.	1.00	9.96	1.41																										
Max.	6.00	22.45	3.94																										
Average	4.14	15.85	2.22																										
Median	5.00	14.39	2.02																										
Standard Deviation	1.96	4.11	0.84																										
Number of Data	7	7	7																										
Cluster5																													

Table 3. Analysis Result about Water Quality Items TOC (Continued)

	Part	Data No.	Discharge (ℓ/s)	TOC (mg/L)
	Min.	2.00	1.42	8.16
	Max.	12.00	8.92	19.48
	Average	7.75	4.67	13.80
	Median	8.50	4.17	13.77
	Standard Deviation	3.77	2.84	5.10
	Number of Data	4	4	4
Cluster6				

### 5. 결론

본 연구는 비점오염원 유출 연구에 있어서 매우 중요한 요소인 초기강우 기준을 제안하기 위해 시험유역에서 측정된 유출 및 수질자료에 대해 SOM 이론을 도입하여 패턴분류를 수행하여 다음과 같은 결론을 얻었다.

- 1) 유출량 자료 수집에 있어 부유물의 영향으로 발생하는 짧은 시간동안의 유량자료의 이상치는 정상적으로 측정된 수위와 유량관계를 상승부와 하강부 자료로 구분하여 회귀분석을 통해 추정된 회귀식을 이용하여 이상치를 보완하였다.
- 2) 수집된 유출량 및 수질자료에 대해 SOM 이론을 도입하여 패턴분류를 수행한 결과 BOD농도의 경우 4개 분할구역, TOC 농도의 경우 6개 분할구역으로 구분되어졌다.
- 3) 초기강우의 경향을 보인 각 수질항목별 분할구역자료들은 BOD 농도의 경우 Cluster1과 Cluster2, TOC 농도의 경우 Cluster4와 Cluster5의 자료가 초기강우의 경향을 보였으며, 본 시험유역인 위락시설 지역은 일반적으로 초기강우에 의해 높은 수질농도를 나타내는 도시지역의 비점오염원 유출특성과는 다른 일정시간이 지난 후에 수질농도가 최대가 되는 유출특성을 나타내었다.

이상의 결과로 SOM 이론을 도입하여 측정된 유출량 및 수질자료에 대한 패턴분류를 수행하여 분할구역별 자료의 특성분석을 통해 초기강우 특성이 구분되어짐을 확인할 수 있었다. 그러나 현재 축적된 자료수의 한계로 인해 명확한 구분이 이루어지지 않는 항목도 있었으나 향후 지속적인 모니터링을 통해 충분한 자료수가 축적되면 초기강우 기준 제시를 위한 새로운 접근방법으로 제시될 수 있을 것으로 기대된다.

### 감사의 글

본 연구는 국립환경과학원 영산강물환경연구소의 연구비 지원에 의해 수행된 연구입니다.

### 참고문헌

김용구, 진영훈, 박성천 (2006). “강우-유출특성 분석을 위한 자기조직화방법의 적용.” **대한토목학회논문집**, 대한토목학회, 제 26권, 제1B호, pp. 61-67.

김이형, 이선하 (2005a). “강우시 주차장 및 교량에서 유출되는 비점오염물질의 특성 및 동적 EMCs.” **한국물환경학회지**, 한국물환경학회, 제21권, 제4호, pp. 248-255

김이형, 이선하 (2005b). “주차장 및 교량 강우유출수의 중금속 오염물질 특성과 동적 EMCs.” **한국물환경학회지**, 한국물환경학회, 제 21권, 제4호, pp. 385-392.

박성천, 진영훈, 김용구 (2006). “강우-유출 예측모형 개발을 위한 자기조직화 이론의 적용.” **대한토목학회 논문집**, 대한토목학회, 제 26권, 제4B호, pp. 389-398.

Garcia, H. L., and Gonzalez, I. M. (2004) Self-organizing map and clustering for wastewater treatment monitoring. *Engineering Applications of Artificial Intelligence*, Vol. 17, pp. 215-225.

Hsu, K. L., Gupta, H. V., Gao, X., Soroochian, S., and Imam, B. (2002). “Self-organizing linear output map (SOLO): An artificial neural network suitable for hydrologic modeling and analysis.” *Water Resources Research*, Vol. 38, No. 12, pp. 38-1-38-17.

논문번호: 10-044	접수: 2010.05.31
수정일자: 2010.10.04/11.11	심사완료: 2010.11.11