

호소수의 강우-저류량 및 TOC변동 특성분석을 위한 자기조직화 방법의 적용

김용구 · 진영훈[†] · 정우철* · 박성천

동신대학교 토목공학과

*동신기술개발(주)

Application of Self - Organizing Map for the Characteristics Analysis of Rainfall - Storage and TOC Variation in a Lake

Yong Gu Kim · Young Hoon Jin[†] · Woo Cheol Jung* · Sung Chun Park

Department of Civil Engineering, Dongshin University

*Dongshin Engineering & Consultants

(Received 9 June 2008, Accepted 26 August 2008)

Abstract

It is necessary to analysis the data characteristics of discharge and water quality for efficient water resources management, aggressive alternatives to inundation by flood and various water pollution accidents, the basic information to manage water quality in lakes and to make environmental policy. Therefore, the present study applied Self-Organizing Map (SOM) showing excellent performance in classifying patterns with weights estimated by self-organization. The result revealed five patterns and TOC versus rainfall-storage data according to the respective patterns were depicted in two-dimensional plots. The visualization presented better understanding of data distribution pattern. The result in the present study might be expected to contribute to the modeling procedure for data prediction in the future.

keywords : Artificial neural networks (ANN), Map size, Pattern classification, Self-organizing map (SOM), Total organic carbon (TOC)

1. 서론

비선형이 강한 자료에 대한 정도 있는 수문 및 환경자료 생성과 이에 따른 보다 정확한 관측 자료들의 분포적 양상을 파악하고 그 패턴을 분석하는 것은 수자원의 효율적인 관리 측면에서 볼 때 중요한 연구과제이다.

수문과정의 비선형 거동은 단위유량도와 같은 전통적인 선형모형을 사용해 오는 동안 계속 제기되어 오던 문제이며, 비선형 모형은 선형모형과는 달리 수학적 일반해가 없어 수문학적인 거동의 설명 및 해석에 어려움이 있다. 수문과정의 모형화는 수문자료에 포함된 많은 불확실성과 물리적 현상의 비정규성에 따른 애매함 때문에 결과에 대한 신뢰성에도 많은 문제가 따른다. 즉, 불확실한 강우입력 자료, 불확실한 선행조건, 비합리적인 시스템의 초기상태, 그리고 비정규성 및 비선형성에 기인한 불확실한 매개변수 산정 등의 문제로 인해 수문순환과정의 수학적 모형화는 아직도 완벽하게 규명되지 못하고 있는 실정이다. 이처럼 시공간적 변동성이 크며 비선형성이 강한 자료에는 입 · 출력간의 패턴에 의존하는 지능형 모형이 적합하다 할 수 있다. 따라서 본 연구는 인공지능망 기법의 일종인 자기조직

화 이론을 이용하여 섬진강 유역의 주암호 수질자동측정망 지점에서 측정되고 있는 수질자료 중 전체 유기물의 양을 나타내며 난분해성 물질에 대한 해석이 가능하고 재현성이 탁월한 TOC와 저류량 자료의 분포적 양상과 특징을 분석하여 예측을 위한 모형화 과정에 기여하고자 한다. 본 연구에서 적용한 코호넨 네트워크 이론은 생물들에서 일어나는 매우 중요한 현상을 보여주며, 복잡한 패턴인식 시스템에서 매우 중요한 부분을 차지하며, 신경 생리학적 시스템을 본따서 모델링했기 때문에 백프로퍼게이션 모델과 같은 지도학습 모델보다는 인간의 두뇌에 가깝다고 볼 수 있다 (김대수, 1996).

자기조직화 이론의 국외의 적용 예를 살펴보면 Hsu 등 (2002)은 Self-Organizing Liner Output map(SOLO)을 이용하여 강우-유출패턴을 구분하였고, Garcia and Gonzalez (2003)은 입력패턴의 분류를 위해 K-means 알고리즘을 이용하여 하수처리모니터링에 적용하였다. 국내의 적용 사례로는 김용구(2006)와 김용구 등(2006)은 강우-유출과정의 양상에 따른 분류와 그에 따른 분석을 위해 Kohonen 네트워크 이론에 의한 자기조직화 방법을 적용하였고, 박성천 등(2006)은 SOM에 의해 강우-유출관계를 분류하고, SOM에 의한 분류에 따라 각각의 예측모형을 구성하여 과거의 인공지능망의 일반적인 적용에 의한 결과보다 더 나은 예

[†] To whom correspondence should be addressed.
nmdrjin@gmail.com

측능력을 보여주었고, 더불어 유출량의 과소 및 과대추정과 Persistence 현상과 같은 문제점을 해소한 연구결과를 발표하였다.

본 연구는 자기조직화 과정에 의한 패턴별 분류를 수행한 후 이를 통해 저류량에 따른 TOC자료를 2차원의 지도상에 시각적으로 가시화하여 각각의 분류된 분할구역의 특징을 분석하고자 한다.

2. 연구방법

2.1. SOM(Self-Organizing Map)의 기본이론

2.1.1. 개요

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

SOM은 비교사 학습방법(unsupervised learning algorithm)의 일종으로 스스로 n차원의 입력 자료들을 클러스터링하여 2차원으로 사상시켜주며 Fig. 1에 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 연결강도 벡터를 이용하여 두 벡터간의 거리(distance)를 계산하는 기능을 수행한다. Fig. 2는 입력노드 i와 출력노드 j를 연결하는 연결강도 w_{ij} 들의 행렬을 보여준다.

이러한 과정에서 각 노드들은 학습할 수 있는 특권을 부여받기 위해 서로 경쟁(competitive)하며, 거리가 가장 가까운 노드가 승리하게 된다. 결국 이 승자 노드만이 출력신호를 보낼 수 있는 유일한 노드가 된다. 승자 노드와 이에

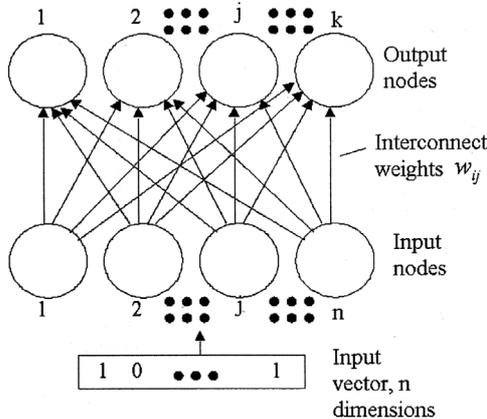


Fig. 1. General structure of the SOM.

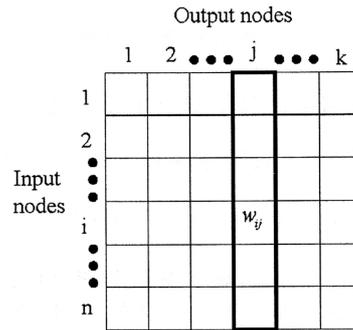


Fig. 2. Weight matrix.

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 코호넨 네트워크의 학습 철학은 승자만이 출력을 할 수 있는 ‘승자독점(winner take all)’ 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 2.1.2. 경쟁과정, 2.1.3. 근접반경의 조정과정, 2.1.4. 적응학습 과정이 진행되는 동안 연결강도를 조정하게 된다.

2.1.2. 경쟁과정(Competitive Process)

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

$$X = [x_1, x_2, \dots, x_m]^T \tag{1}$$

$$W_j = [w_{j1}, w_{j2}, \dots, w_{jm}]^T, \quad j = 1, 2, \dots, l \tag{2}$$

여기서 l은 전체 노드 수이다.

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

$$i(X) = \underset{j}{\operatorname{arg\,min}} \|X - W_j\|, \quad j = 1, 2, \dots, l \tag{3}$$

위의 식은 입력벡터(X)와 연결강도벡터(W)간의 거리가 최소일 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.

2.1.3. 근접반경의 조정과정(Cooperative Process)

코호넨의 시스템은 생물학적 모델에서 보여지는 바와 유사하게 이웃한 노드와의 연계과정(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. 최종적으로는 단지 승자 노드만이 그것의 연결강도를 조정하게 된다.

이러한 과정을 위해 출력층의 노드는 1차원으로 배열하는 방법과 2차원으로 배열하는 방법이 있으며, 특히 2차원 배열에는 Fig. 3에서와 같이 사각형배열(rectangular array)과 육각형배열(hexagonal array)의 2가지 형태가 있다.

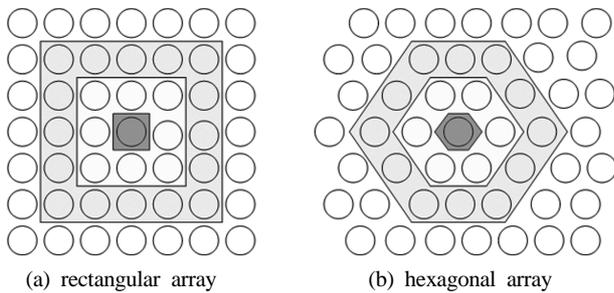


Fig. 3. Types of map array.

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

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

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

$$d_{ji}^2 = \|r_j - r_i\|^2 \quad (5)$$

그리고, 일반적으로 지수적 감쇠를 위한 σ 는 다음과 같이 선택할 수 있다.

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

식 (6)에서 σ_0 는 SOM 알고리즘의 초기화에서 갖는 σ 의 값이며, τ_1 은 시간상수이다.

2.1.4. 적응학습 과정(Adaptive Process)

이상과 같은 단계의 처리가 끝나면 마지막으로는 시냅스의 적응과정과 같은 실제 연결강도의 조정이 이루어진다. 조정되기 이전의 연결강도 벡터를 $W_j(n)$, 조정된 후의 새

로운 연결강도 벡터를 $W_j(n+1)$ 와 같이 이산 시간 t에 대하여 정의할 때 조정을 위한 규칙은 다음 식으로 표현된다.

$$W_j(n+1) = W_j(n) + \eta(n) \cdot h_{ji}(x) \cdot [X - W_j(n)] \quad (7)$$

여기서, η 는 시간 n 이 증가함에 따라 서서히 감소하는 학습율을 나타내는 매개변수로서 초기값 η_0 와 지수적 감쇠를 만족시킬 수 있도록 다음과 같은 식으로 나타낼 수 있다.

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

τ_2 는 SOM 알고리즘의 다른 시간상수 이다.

2.2. 대상지점 및 입력자료

본 연구의 대상유역인 점진강 유역은 우리나라 4대강 유역의 하나로서 우리나라 남부지역의 중서부에 위치하고 있으며, 총 유역면적은 4,896.5 km²이며 본류의 유로연장은 212.3 km이다.

유역의 최남단은 북위 34° 40' 26", 최북단은 북위 35° 50' 00"이며, 최동단은 동경 127° 53' 05", 최서단은 동경 126° 51' 50"로 본 유역의 동쪽에 낙동강 유역, 서쪽에는 영산강 유역과 동진강 유역, 북쪽에는 금강 유역과 만경강 유역이 접하고 있으며 유역 내에 4개 댐이 건설 운영 중에 있다.

본 연구의 대상지점인 주암호는 1992년에 Rock Fill형식으로 준공된 댐으로서 유역면적 1,010 km², 저류량 4억 5700만t로 보성군·순천시·화순군의 3개 시군에 걸쳐 있다. 또한 광주광역시·나주시·목포시·화순군 등 전라남도 서부권과 여수시·순천시 등에 하루 118만t의 생활용수 및 농·공업용수와 전력을 공급하고 있다. Fig. 4는 주암댐 유역도를 나타내고 있으며, 주암호 수질관측망은 전라남도 순천시 대광면 주암리에 위치해 있다. 본 연구의 대상 자료는 저

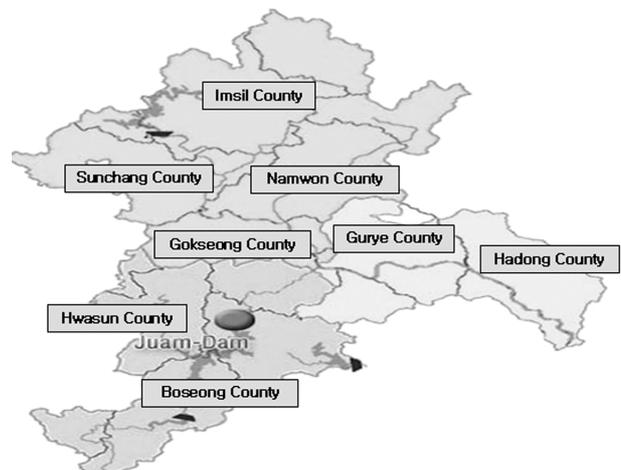


Fig. 4. Geographical location of the Juam lake and Seomjin river basin.

류량과 TOC 그리고 강우 자료이다. 본 연구에서 사용된 TOC 수질자료는 섬진강수계의 수질오염 조기경보체계 구축을 위해 영산강유역환경청에서 관리 보관 중인 자료를 이용하였으며, 주암호 지점으로부터 2004년 1월 1일 01시부터 2004년 10월 24일 17시까지 총 7,145개의 자료를 이용하였다. 강우량 자료는 주암댐지점의 기상청 자료, 저류량 자료는 주암댐 관리단의 자료를 이용하였으며, 대상기간은 TOC 자료와 동일하다.

SOM 훈련 데이터 구성은 강우-저류량 및 TOC 변동 양상의 관계를 파악하기 위해서는 긴 시간의 많은 자료가 필요하다고 판단되나 본 연구에서 자료의 검토결과 강우-저류량 및 TOC 관계를 파악하기 위한 최소한의 시간이 다섯 개의 연속자료를 이용하면 그 특성 파악이 가능하다고 판단되고, 자료의 크기를 최소화하기 위하여 강우-저류량 및 TOC 변동 양상을 구분하기 위하여 5시간 동안의 자료를 SOM의 입력자료로 이용하였다. 따라서 입력벡터 $x(t)$ 는 식 (9)와 같은 15차원의 구조가 된다.

$$x(t) = (P_{(t-4)}, P_{(t-3)}, P_{(t-2)}, P_{(t-1)}, P_{(t)}, S_{(t-4)}, S_{(t-3)}, S_{(t-2)}, S_{(t-1)}, S_{(t)}, T_{(t-4)}, T_{(t-3)}, T_{(t-2)}, T_{(t-1)}, T_{(t)}) \quad (9)$$

여기서, P 는 강우량, S 는 저류량, T 는 TOC이다.

3. 결과 및 고찰

3.1. SOM 모형의 적용

본 연구에서는 SOM를 이용한 패턴 분류를 위해 주암댐 지점의 저류량과 TOC, 강우량에 대한 시자료를 이용하였고, 대상자료는 2004년 1월 1일 01시부터 2004년 10월 24일 17시까지 총 7,145개 자료이며, SOM 적용은 Fig. 5와 같다. 강우량과 TOC, 저류량의 총 7,145개 자료에 대하여 시차를 적용하였으며 총 $7,145 \times 15$ 행렬의 SOM 입력 자료를 구축하였다. 선택된 자료의 정규화 과정을 거친 후 SOM 훈련을 위해 Map 크기를 결정하는데 우선, Map을 구성하는 단위구조의 총 수(M)의 결정은 Garcia and Gonzalez(2003)에 의해 연구 보고된 식을 이용하였다(식 (10)).

$$M = 5\sqrt{N} \quad (10)$$

여기서 M은 map을 구성하는 단위구조(unit)의 총 수이며, N은 훈련 자료의 수이다.

본 연구에서의 훈련자료는 7,145쌍이며, map 크기는 식 (10)에 의해 산정된 422의 근사치로부터 22×19 로 결정하였다. 결정된 map 크기에 대해 최적의 분할구역 구분은 U-matrix method와 K-means algorithm의 방법을 적용하여 구분할 수 있다. U-matrix method에 의해 구분된 결과를 Fig. 6(a)에 나타내었으며, K-means algorithm 결과 최적의 분할구역으로 구분된 결과는 Fig. 6(b)와 같이 5개 분할구역으로 구분되었다.

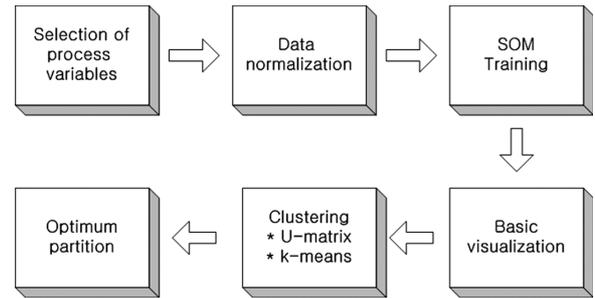
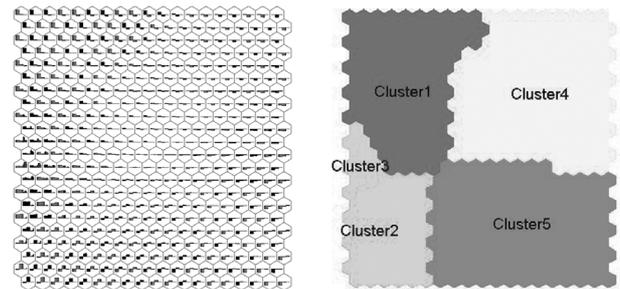


Fig. 5. SOM flow chart.



(a) result of u-matrix method (b) result of k-means algorithm

Fig. 6. Clustering of 2D input data using SOM.

3.2. SOM 적용 결과

SOM에 의해 구분된 강우-저류량 및 TOC 관계를 규명하기 위해 Fig. 7과 같이 강우-저류량 및 TOC 관계 분석결과 도시하였고, 명확한 결과의 도시를 위하여 Fig. 8과 같이 각 분할구역별 강우-저류량 및 TOC 관계와 box plot를 도시하였다. 본 연구에서는 강우-저류량 및 TOC 관계 분석에는 각 분할구역별 자료를 추출하여 분할구역별 자료의 통계치(평균)로 그 증감 양상을 분석하였다.

그 결과 Cluster1은 저류량은 미미하게 증가하는 양상을 보이고, TOC 농도는 평균 1.61 mg/L로 일정하며 강우는 작은 강우가 발생하며 변화양상은 미미한 경향으로 구분되었으며, Cluster1은 강우종료단계시의 저류량은 큰 분포를 나타내는 분할구역 특성을 보인다.

Cluster2는 강우량이 급격히 증가함에 따라 저류량도 일정하게 증가하였고, TOC 농도는 강우가 계속됨에 따라 감소하는 경향으로 구분되었으나, 5개 분할구역 중 TOC 농도는 가장 큰 분포를 보이고 있는 분할구역 특성을 보이고 있다.

Cluster3은 가장 큰 강우가 발생하여 점차 감소하지만 저류량은 계속해서 증가하였고 TOC 농도는 유량의 증가로 평균 1.8 mg/L의 농도값으로 일정하게 유지되는 경향으로 구분되었고, Cluster4는 대부분의 자료가 무강우 기간으로 저류량과 TOC농도는 변화가 없는 일정한 양상으로 TOC 농도값이 가장 낮은 분포를 보이고 있다.

마지막으로 Cluster5는 Cluster4와 유사한 경향을 보이고 있으나 Cluster4와 비교해보면, 저류량 규모와 패턴은 작고 감소하는 반면에 TOC 농도는 더 높은 값을 나타내는 경향으로 구분되었다.

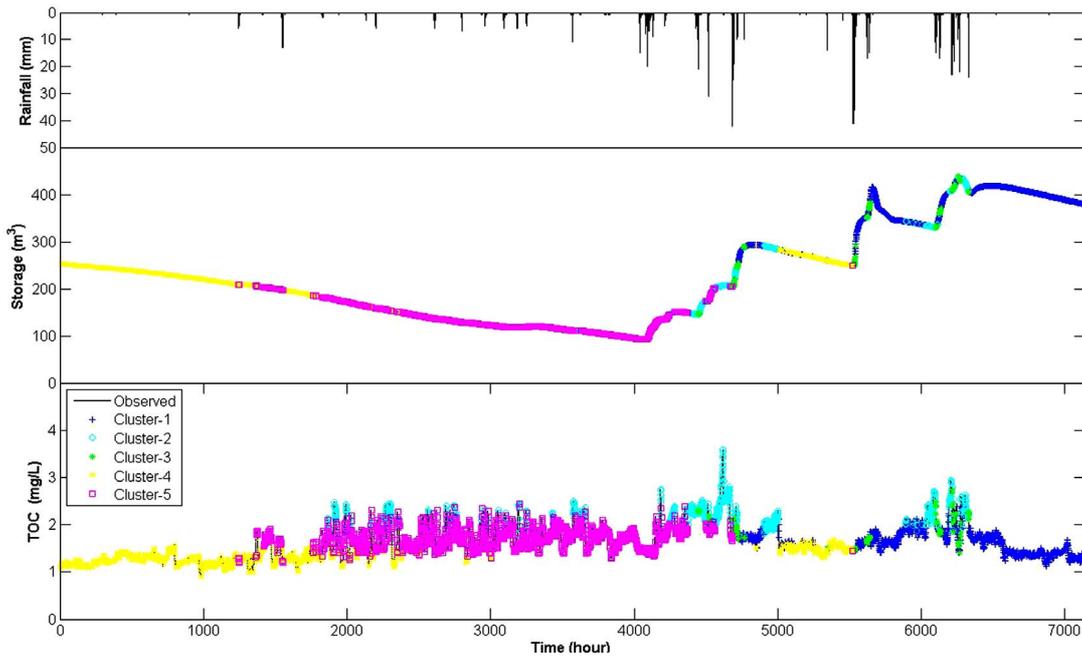
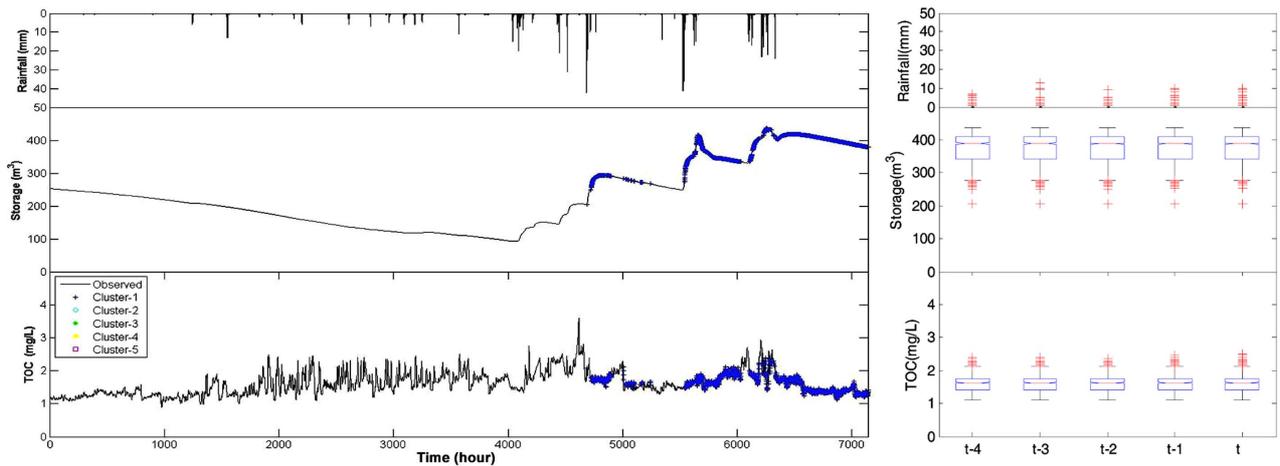
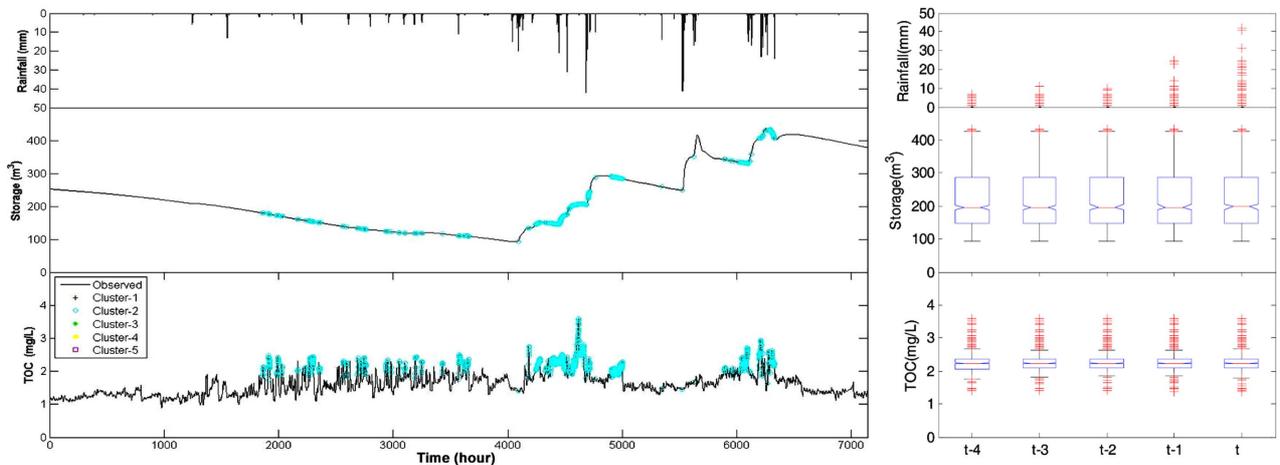


Fig. 7. Result of rainfall-storage and TOC variation characteristics analysis.

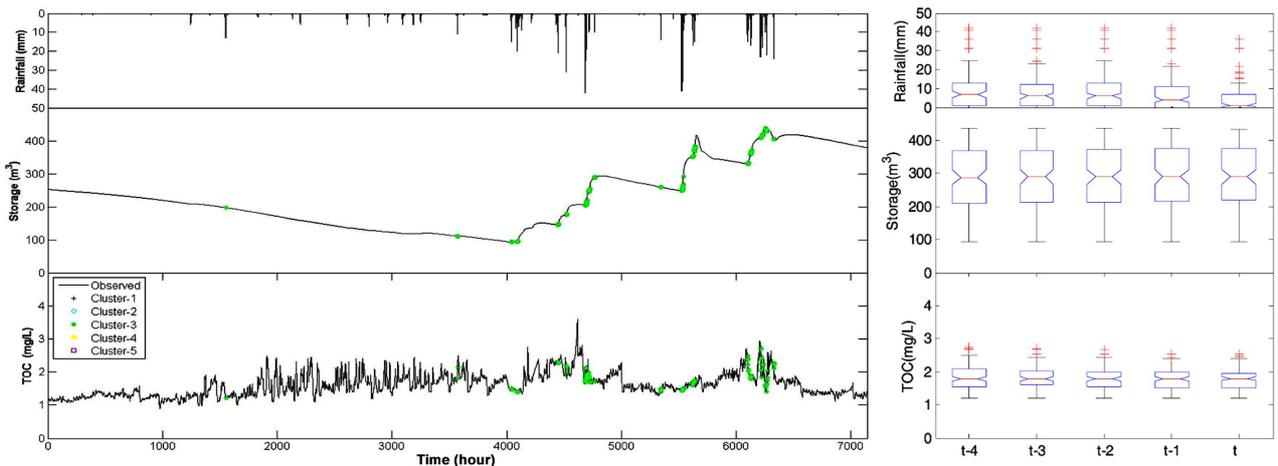


(a) Rainfall-Storage and TOC relationship and box plot of cluster1

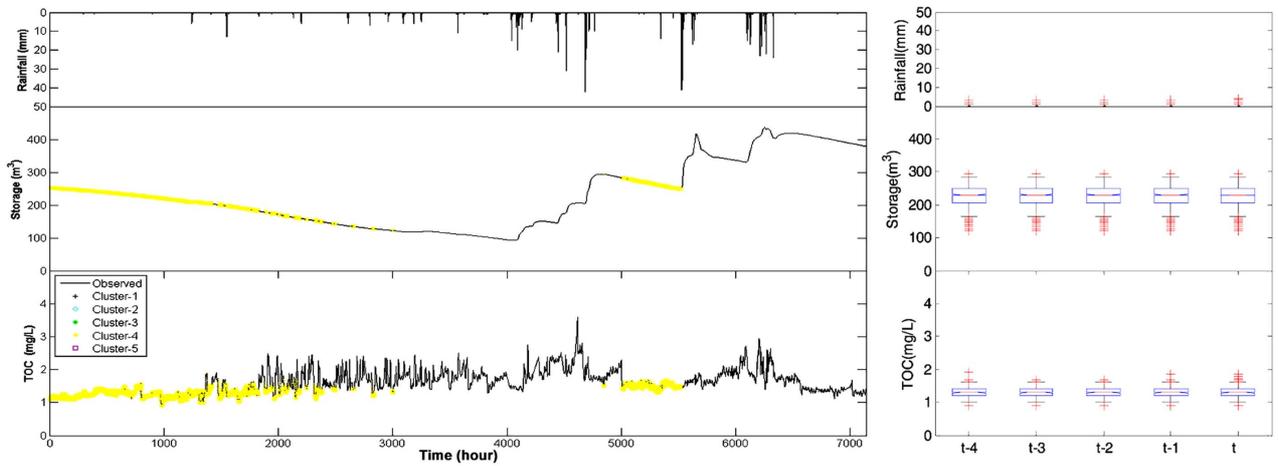


(b) Rainfall-Storage and TOC relationship and box plot of cluster2

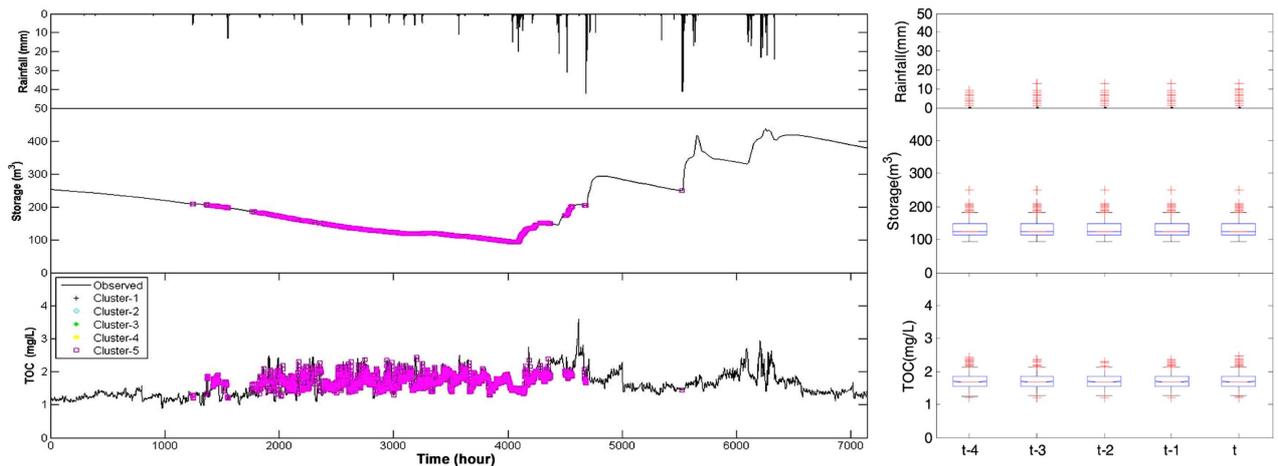
Fig. 8. Rainfall-Storage and TOC relationship and box plot of each cluster.



(c) Rainfall-Storage and TOC relationship and box plot of cluster3



(d) Rainfall-Storage and TOC relationship and box plot of cluster4



(e) Rainfall-Storage and TOC relationship and box plot of cluster5

Fig. 8. Rainfall-Storage and TOC relationship and box plot of each cluster. (Continued)

이상의 결과와 같이 강한 비선형성을 가지고 있는 강우-저류량 및 TOC 농도가 SOM으로 패턴분류가 가능함을 확인할 수 있었고, Cluster4와 Cluster5의 경우에서처럼 각 변수의 규모와 증감에 따른 양상에 의해서도 구분되

는 것을 확인할 수 있는데 이는 본 연구에서 적용하는 SOM 방법은 강우-저류량 및 TOC의 패턴과 규모를 동시에 복합적으로 고려하여 그 양상을 구분하는 것으로 판단된다.

4. 결론

본 연구는 비선형성이 강한 자료에 대한 정도 있는 수문 및 환경자료 생성과 이에 따른 보다 정확한 관측 자료들의 분포적 양상을 파악하기 위해 인공신경망의 기법인 자기조직화 이론을 이용하여 강우-저류량 및 TOC관계의 특성을 분석한 결과 다음과 같은 결론을 얻었다.

- 1) 강한 비선형성을 가지고 있는 강우-저류량 및 TOC관계가 SOM으로 패턴분류가 가능함을 확인할 수 있었다.
- 2) SOM으로 구분된 분할구역별 특성을 검토한 결과 Cluster1은 저류량은 미미하게 증가하는 양상을 보이고, TOC 농도는 일정하며 강우는 작은 강우가 발생하며 변화양상은 미미한 경향으로 구분되었다. Cluster2는 강우량이 급격히 증가함에 따라 저류량도 일정하게 증가하였고, TOC 농도는 강우가 계속됨에 따라 감소하는 경향으로 구분되었고, Cluster3은 가장 큰 강우가 발생하여 점차 감소하지만 저류량은 계속해서 증가하고 TOC 농도는 유량의 증가로 낮은 농도값으로 일정하게 유지되는 경향으로 구분되었다. Cluster4는 대부분의 자료가 무강우 기간으로 저류량과 TOC농도는 변화가 없는 일정한 양상으로 TOC 농도값이 가장 낮은 분포를 보이고 있다. 마지막으로 Cluster5는 Cluster4와 유사한 경향을 보이고 있으나 Cluster4와 비교해보면, 저류량 규모와 패턴은 작고 감소하는 반면에 TOC 농도는 더 높은 값을 나타나는 경향으로 구분되었다.
- 3) Cluster4와 Cluster5의 경우에서처럼 각 변수의 규모와 증감에 따른 양상에 의해서도 구분되는 것을 확인할 수 있는데 이는 본 연구에서 적용하는 SOM 방법은 강우-

저류량 및 TOC의 패턴과 규모를 동시에 복합적으로 고려하여 그 양상을 구분하는 것으로 판단된다.

- 4) 이상의 연구결과를 토대로 비선형적인 경향이 강한자료의 분포적 양상을 이해하는데 큰 도움이 되며, 향후 이를 통해 예측을 위한 모형화 과정에도 크게 도움을 줄 것으로 기대된다. 또한, 강우자료 또는 저류량 자료만을 이용한 단일변량의 패턴분류를 위해 SOM의 적용이 가능할 것으로 판단되며, 이는 각 변량의 본질적인 특성을 파악할 수 있을 것으로 기대된다.

참고문헌

- 김대수(1996). *신경망 이론과 응용(I)*, 하이테크정보.
- 김용구(2006). 자기조직화 이론을 적용한 강우-유출 예측모형 개발. 박사학위논문, 동신대학교.
- 김용구, 진영훈, 박성천(2006). 강우-유출특성 분석을 위한 자기조직화방법의 적용. *대한토목학회 논문집*, **26(1B)**, pp. 61-67.
- 박성천, 진영훈, 김용구(2006). 강우-유출 예측모형 개발을 위한 자기조직화 이론의 적용. *대한토목학회 논문집*, **26(4B)**, pp. 389-398.
- Garcia, H. L. and Gonzalez, I. M. (2003). Self-organizing map and clustering for wastewater treatment monitoring. *Engineering Applications of Artificial Intelligence*, **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*, **38(12)**, pp. 38-1-38-17.