

# 나주시점의 강우-유출 해석을 위한 최적의 SOM 구조 결정

## Determination of the Optimized Structure of Self-Organizing Map for the Rainfall-Runoff Analysis in Naju

김용구\* / 진영훈\*\* / 박성천\*\*\* / 정천리\*\*\*\*

Kim, Yong Gu / Jin, Young Hoon / Park, Sung Chun / Jeong, Choen Lee

### Abstract

Studies on modeling the rainfall-runoff relationship which shows nonlinear trend strongly use artificial neural networks theory not only for the prediction but also for the characteristics analysis of the data used by pattern classification. For the pattern classification, the results from Self-Organizing Map (SOM) mention that the map size and array for the SOM training have significantly influenced on the SOM performance. Since there is no deterministic method or theoretical equation to determine the number of rows and columns for the map size, hexagonal array is generally used for the map array. Therefore, this study present a determination of the optimized map structure for the rainfall-runoff analysis in Naju station considering the map size and array simultaneously which can represent the classified characterization of rainfall-runoff relationship. The result showed that the map size of 20×16 hexagonal array with 8-clustered patterns was selected as an appropriate map structure for rainfall-runoff analysis in Naju station.

**keywords** : rainfall- runoff, Artificial Neural Networks (ANNs), Self-Organizing Map (SOM), pattern classification, map size, map array

### 요 지

인공신경망 이론을 이용하여 강한 비선형성의 경향을 보이고 있는 강우-유출간의 관계를 모형화하기 위한 연구들은 예측뿐만 아니라 대상자료들의 양상을 분류하여 그 특성을 분석하는 데에도 이용되고 있다. 이와 같은 패턴 분류를 위한 SOM(Self-Organizing Map: SOM)의 연구 결과를 검토해보면 SOM 훈련을 위한 지도크기 및 배열의 결정은 SOM 성능에 큰 영향을 미치는 것으로 보고되고 있으나 지도크기 결정시 지도의 종방향 크기와 횡방향 크기를 결정할 수 있는 확정론적인 방법이나 이론식이 없고, 지도배열은 주로 육각형 배열(hexagonal array)을 이용하여 적용하고 있다. 따라서 본 연구에서는 영산강 나주시점을 대상으로 강우-유출관계의 분할특성을 나타내는 지도크기와 배열을 복합적으로 검토하여 나주시점의 강우-유출 해석을 위한 적절한 지도구조를 결정하였다. 그 결과

\* 동신대학교 토목공학과 연구원  
Researcher, Dept. of Civil Engineering, Dongshin University, Naju, 520-714, Korea  
(e-mail: kyg8987@paran.com)

\*\* 교신저자, 동신대학교 연구교수  
Research Professor, Dongshin University, Naju, 520-714, Korea  
(e-mail: nmrdjin@gamil.com)

\*\*\* 동신대학교 토목공학과 교수  
Professor, Dept. of Civil Engineering, Dongshin University, Naju, 520-714, Korea

\*\*\*\* 동신대학교 공과대학 토목공학과 박사과정  
Ph.D. student, Dept. of Civil Engineering, Dongshin University, Naju, 520-714, Korea

8개의 패턴으로 구분된 지도크기 20×16의 육각형배열 구조가 나주지점의 강우-유출해석을 위한 적절한 지도구조로 결정되었다.

**핵심용어** : 강우-유출, 인공신경망, 자기조직화, 패턴분류, 지도크기, 지도배열

## 1. 서론

강한 비선형성의 경향을 보이고 있는 강우-유출간의 관계를 모형화하기 위한 연구는 다양한 방법론으로 적용되어 활발히 연구되고 있으며 그 중에서 인공신경망을 이용하여 강우-유출간의 관계를 모형화하기 위한 대부분의 연구들은 역전파 학습 알고리즘을 이용하였으며, 이는 강한 비선형성을 나타내는 입·출력간의 관계를 나타내는데 탁월한 성능을 보이고 있는 것으로 알려져 있고, 또한 자료들의 급격한 변화에 대한 뛰어난 적응성을 보여주고 있다. 이러한 인공신경망의 장점들은 다른 모형들과의 비교에서 우수한 예측력을 보여주는 것으로 평가받고 있다. 그러나 지금까지 적용되어진 기존의 인공신경망을 이용한 대부분의 연구들은 강우-유출해석에 있어 다양한 강우사상을 반영하는데 미진할 뿐만 아니라, 예측의 대상인 유출량의 경우 저·갈수기의 자료에 대해서는 과대평가의 경향이 나타나며, 홍수기의 자료에 대해서는 과소평가되는 경향이 있다. 또한 인공신경망 모형의 구축에 있어 예측 대상자료인 유출량의 선행값을 이용할 경우 그 예측값이 선행 유출량의 지속성을 갖는 지속성이 유발된다. 또한 강우-유출관계를 규명하기 위한 대부분의 연구들이 홍수기의 강우 및 유출 자료를 이용하기 때문에 예측모형이 자료의 크기에 따라 한정되는 한계를 보여주었다. 이러한 문제점들을 해결하기 위해 패턴분류 성능을 지닌 자기조직화(Self-Organizing Map: SOM) 이론을 도입하여 자료의 전처리 과정으로 이용함으로써 기존의 인공신경망의 일반적인 적용을 개선한 연구들이 수행되어지고 있다. 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 백프로퍼레이션 모델과 같은 지도학습 모델보다는 인간의 두뇌에 가깝다고 볼 수 있다(김대수, 1996).

본 연구에서 적용한 자기조직화 이론의 국외의 적용 예를 살펴보면 Hsu *et al.*(2002)은 Self-Organizing Linear Output map(SOLO)을 이용하여 강우-유출패턴을 구분하였고, Garcia *et al.*(2004)은 입력패턴의 분류를 위해 K-means 알고리즘을 이용하여 하수처리모니

터링에 적용하였다. 국내의 적용 사례는 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 인공신경망의 일반적인 적용에 의한 결과보다 더 나은 예측능력을 보여주었고, 더불어 유출량의 과소 및 과대추정과 지속성과 같은 문제점을 해소한 연구결과를 발표하였다.

그러나 김용구(2006)의 결과에 의하면 SOM 훈련을 위한 지도크기 및 배열의 결정은 SOM 성능에 영향을 미치는 것으로 보고되고 있으나, 지도크기 결정시 지도의 종방향 크기 및 횡방향 크기를 결정할 수 있는 확정론적인 방법이나 이론식이 없고, 지도배열은 일반적으로 육각형배열을 주로 이용하고 있는 것으로 보고되고 있다.

따라서 본 연구에서는 일반적으로 적용되어지고 있는 지도크기 및 배열을 보다 구체적인 접근방법으로 적절한 지도구조를 결정하고자 한다.

## 2. SOM의 기본이론

### 2.1 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 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 연결강도벡터를 이용하여 두 벡터간의 거리(distance)를 계산하는 기능을 수행한다. Fig. 2는 입력노드  $i$ 와 출력노드  $j$ 를 연결하는 연결강도  $w_{ij}$ 들의 행렬을 보여준다.

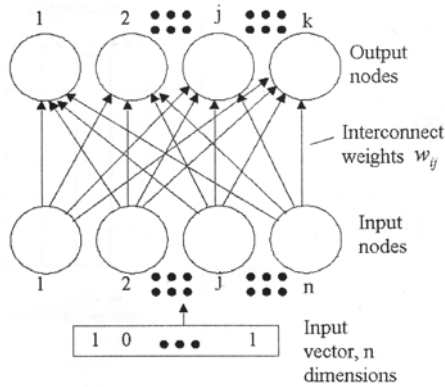


Fig. 1. General Structure of SOM

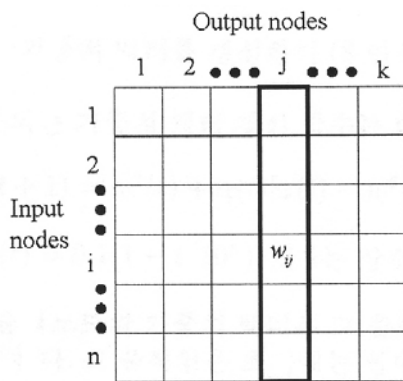


Fig. 2. Weight Matrix

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 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 takes 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 2.2 경쟁과정, 2.3 근접반경의 조정과정, 2.4 적응학습 과정이 진행되는 동안 연결강도를 조정하게 된다.

## 2.2 경쟁과정(Competitive Process)

경쟁과정에서 각 노드들은 학습할 수 있는 특권을 갖기 위해 서로 경쟁하며 연결강도 벡터와 입력벡터의 거리가 가장 가까운 노드가 승자로 판정된다.  $n$ 개의 입력을 가진 입력패턴과 노드  $j$ 의 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  $k$ 는 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  $i$ 라는 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.

## 2.3 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.

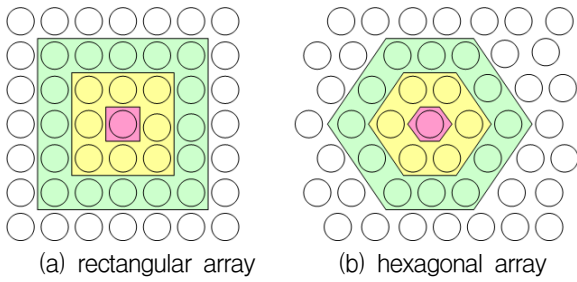


Fig. 3. Type of Map Array

기하학적 반경을 조정하는 과정에 있어 대칭성과 수렴특성을 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.

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

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

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

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.

#### 2.4 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.

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

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

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

### 3. 대상지점 및 입력자료

본 연구의 대상지점인 영산강 유역은 우리나라 서남부에 위치하며, 총 유역면적은 3,455km<sup>2</sup>이고, 본류의 유로연장은 129.5km이며, 본 연구에서는 강우의 시·공간적 분포의 비선형적 변동성을 고려한 강우패턴을 분류하고 강우-유출관계의 특성을 분석하기 위해 Fig. 4에 나타내었듯이 영산강 본류의 대표지점인 나주수위관측소와 그에 영향을 미치는 인근의 상류부에 위치한 광주, 동곡, 능주 강우관측소를 선정하였다. 수위 및 강우 자료는 영산강홍수통제소의 30분 자료를 이용하였고, 나주지점의 수위자료를 유출량자료로 환산하기 위해 유량조사보고서(건설교통부, 2007)의 나주지점의 수위-유량관계곡선을 이용하였다.



Fig. 4. Study Area

모형의 입력자료 구축을 위해 광주, 동곡, 능주 강우 관측소의 강우량이 나주지점 유출에 기여하는 시차를 산정하기 위해 무강우를 제외한 각 강우관측소의 강우 자료와 그에 대한 지체시간을 적용한 나주지점의 유출 자료 사이의 상관분석을 실시한 결과 Fig. 5와 같이 광주 강우관측소는 14시간 30분, 동곡 강우관측소는 16시간, 능주 강우관측소는 17시간 30분 후가 가장 높은 상관성을 갖는 값으로 분석되었다. 이러한 일련의 과정을 통해 나주지점의 유출과 인근 강우관측소의 강우량과의 강우-유출관계 특성 파악을 위해 나주 유출에 대응하는 각 강우관측소의 지체시간을 고려한 강우량자료를 산출 평균하고, 1시간 30분의 시차를 고려하여 Eq. (9)와 같은 SOM 훈련 데이터를 구성하였다.

## 4. 적응방법 및 결과

### 4.1 SOM 모형의 적용

SOM의 적용을 위해 Fig. 6의 SOM 적용 흐름도와 같이 광주, 동곡, 능주강우관측소의 각 지체시간을 고려한 평균 강우량과 나주 수위자료의 30분 자료로부터 2007년 6월 1일부터 2007년 10월 12일까지의 6,400쌍의 자료를 선택하여 1시간 30분의 시차를 적용하여 6,400×6 행렬의 SOM 입력자료를 구축하였다. 적용된 자료의 범위는 강우자료의 최소값은 0 mm, 최대값은 40 mm 이고, 유출량자료의 최소값은 11.95 m<sup>3</sup>/sec, 최대값은 2242.49 m<sup>3</sup>/sec이다.

본 연구에서는 얼마나 많은 노드를 움직일 것인가를 결정하는 학습율 감소인자는 0.2로 최종 학습률은 0.18, 학습회수는 1000회이다. 선택된 자료의 정규화 과정을 거친 후 SOM 훈련을 위해 지도크기를 결정하는데 기존의 연구에서는 지도를 구성하는 단위구조의 총 수(M)의 결정은 Garcia *et al.*(2003)에 의해 연구 보고된 식인  $M = 5\sqrt{N}$  식을 이용하였다. 여기서 M은 지도를 구성하는 단위구조(unit)의 총 수이며, N은 훈련 자료의 수이다.

그러나 본 연구에서는 SOM 성능에 영향을 미치는 지도크기 결정의 이전 과정인 지도를 구성하는 단위구조의 총 수(M)의 결정에 있어서 기존의 식을 다소 변형한  $M = 4\sqrt{N}$ ,  $M = 6\sqrt{N}$  에 대해서도 검토하였다. 지도크기 결정시 지도의 종방향 및 횡방향의 크기를 결정 할 수 있는 확정론적인 방법이나 이론식이 없어 지도를 구성하는 단위구조의 총수의 값을 만족하는 지도크기를 정사각형 구조를 기준으로 종·횡방향 크기를 다르게 적용하는 방법으로 Table 1과 같이 검토하였다. 또한 지도배열 또한 사각형 배열과 육각형 배열에 대해 복합적으로 검토하였다. 그 결과 지도구조에 따른 분할구역구분 결과를 Table 1에 나타내었다.

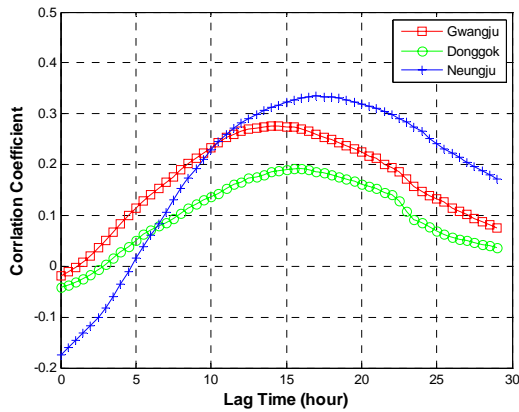


Fig. 5. Cross-correlation Analysis

동곡 강우관측소는 다른 강우관측소와 비교하여 나주 수위관측소와 매우 가까운 거리에 위치하나, 동곡 강우관측소에 해당하는 지배면적 대부분이 농경지와 산지가 산재되어있고, 대부분의 유출이 지표면 유출로 기여하는 지형적인 영향으로 인해 근접한 거리에도 불구하고 다소 긴 지체시간을 나타내는 것으로 판단된다. 또한 1시간 30분의 시차를 적용한 이유는 SOM 훈련 데이터 구성은 강우의 양상과 수위 사이의 관계를 파악하기 위해서는 긴 시간의 많은 자료가 필요하다고 판단되나 본 연구에서 자료의 검토결과 강우-유출관계를 파악하기 위한 최소한의 시간이 세 개의 연속자료를 나타내는 1시간 30분으로도 그 특성 파악이 가능하다고 판단되고, 자료의 크기를 최소화하기 위하여 강우-유출의 양상을 구분하기 위하여 1시간 30분 동안의 자료를 SOM의 입력자료로 이용하였다. 또한, Eq. (9)에서 P는 현재, 30분전 및 1시간 전 강우자료가 입력변수로 사용되었으며, Q는 현재, 30분전 및 1시간 전 유출량자료를 입력변수로 사용하였다.

$$x(t) = (P_{(t-2)}, P_{(t-1)}, P_{(t)}, Q_{(t-2)}, Q_{(t-1)}, Q_{(t)}) \quad (9)$$

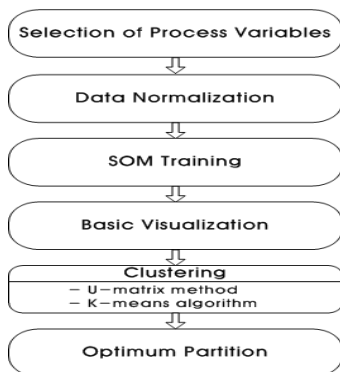


Fig. 6. SOM Flow Chart

Table 1. Classified Clusters According to Different Equations and Map Size

Equation	M	Map size	Number of classified clusters	
			Hexagonal array	Rectangular array
$M = 4\sqrt{N}$	320	18×17	3	3
		20×16	8	6
		15×20	3	3
$M = 5\sqrt{N}$	400	20×20	2	2
		25×16	3	3
		16×25	7	3
$M = 6\sqrt{N}$	480	22×22	2	2
		24×20	5	5
		20×24	2	2

본 연구의 목적을 위해 Fig. 7의 강우주상도와 유출수문곡선으로 Table 1의 검토결과에 대해 검토하고자 하는 지도구조를 1차적으로 선별하고자 한다. 일반적으로 강우가 발생하여 유역에 물을 공급하여 발생하는 유출순환과정은 무강우단계, 강우초기 단계, 강우지속단계, 강우중만단계, 강우 종료단계의 다섯 단계로 구분될 수 있으며, 이로 인한 유출수문곡선의 양상은 Fig. 7과 같이 Region 1의 기저유출부분, Region 2의 수문곡선의 상승기점 부근, Region 3의 수문곡선의 상승부, Region 4의 첨두부, Region 5의 수문곡선의 감수부로 구분할 수 있다. 따라서 보다 설득력 있는 강우-유출관계의 분할 특성이 최소 5개 이상으로 구분되어야 할 것으로 판단된다.

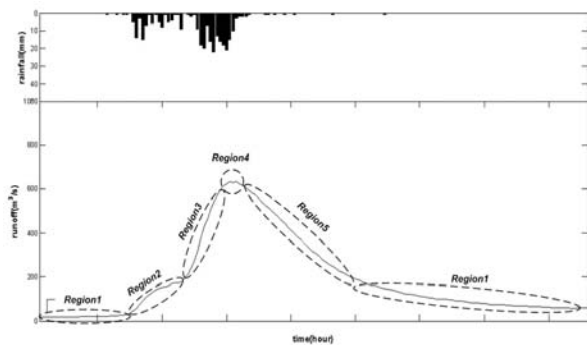


Fig. 7. Relationship between Five Classified Regions and Rainfall-Runoff Process

Table 1의 결과에서 보듯이 지도를 구성하는 단위구조의 총 수 결정을 위한 식을  $M=4\sqrt{N}$ 과  $M=6\sqrt{N}$ 을 이용할 경우에는 종방향 크기가 횡방향 크기보다 큰 경우가 다양한 강우-유출관계의 분할 특성을 보여주고 있으나  $M=5\sqrt{N}$  식을 이용할 경우에는  $M=4\sqrt{N}$ 과  $M=6\sqrt{N}$ 의 결과와는 반대되는 결과를 보이고 있고, 또한 지도배열별 검토에서도 사각형배열에서 다양한 강우-유출관계를 잘 반영해주지 못하는 결과를 보여

$M=5\sqrt{N}$  식은 본 연구에서 적용한 자료에 대해 적절한 결과를 나타내지 못할 것으로 판단되어 검토에서 제외하였다.

따라서 본 연구에서는 강우-유출간의 관계가 지도배열별로 동일하게 5개 이상의 분할 특성으로 구분되어지는  $M=4\sqrt{N}$  식을 적용한 지도크기  $20 \times 16$ 과  $M=6\sqrt{N}$  식을 적용한 지도크기  $24 \times 20$ 에 대해 기본적인 통계값과 분할구역별 자료의 분산정도 및 도식적인 기준을 이용하여 본 연구에서 적용한 자료에 대해 나주 지점의 강우-유출해석을 위한 적절한 지도구조를 결정하고자 한다.

#### 4.2 SOM 적용결과

1차적으로 선별된  $20 \times 16$  지도크기의 육각형배열과 사각배열,  $24 \times 20$  지도크기의 육각형배열과 사각배열의 지도구조에 대해 Fig. 8과 같이 K-means 알고리즘 방법을 적용하여 각 지도구조에 대한 분할영역 구분을 도하였다. Fig. 8의 결과에서 보듯이 지도크기  $20 \times 16$ 의 사각배열과  $24 \times 20$ 의 육각형배열의 경우는 Cluster3의 분할구역이 동일영역임에도 불구하고 근접한 영역에 위치하지 못하는 결과를 보여주고 있다. 이는 근접한 영역의 조정과정에서 오류가 있는 것으로 판단되며 본 연구의 목적에 부합되는 지도구조로는 부적절하다고 판단되어 2차적으로 검토할 지도구조에서 제외하였다.

이와 같이 K-means 알고리즘 결과를 이용하여 검토할 지도구조에 대한 2차선별을 실시하였고, 그 결과 지도크기  $20 \times 16$ 의 육각형배열 구조와  $24 \times 20$ 의 사각배열 구조 중 어떤 구조가 본 연구의 대상지점인 나주 지점의 강우-유출관계를 잘 반영해 주는지를 검토하기 위해 Figs. 9~12의 각 지도구조별 강우-유출관계분석 결과 및 각 분할구역별 자료의 도식적 기준 및 box plot 결과와 Tables 2 and 3의 분할구역별 자료의 기본적인 통계치를 토대로 검토하였다.

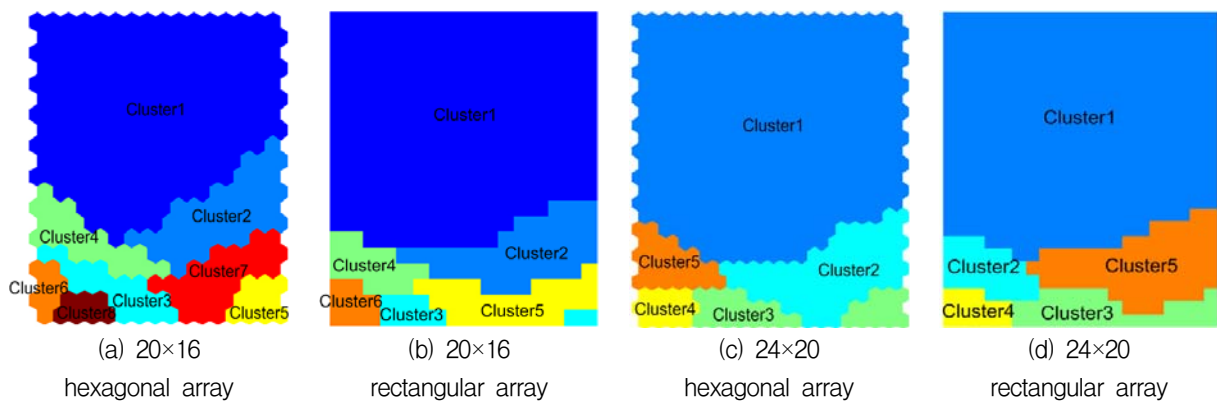


Fig. 8. K-means Algorithm Results According to the Respective Map Structure



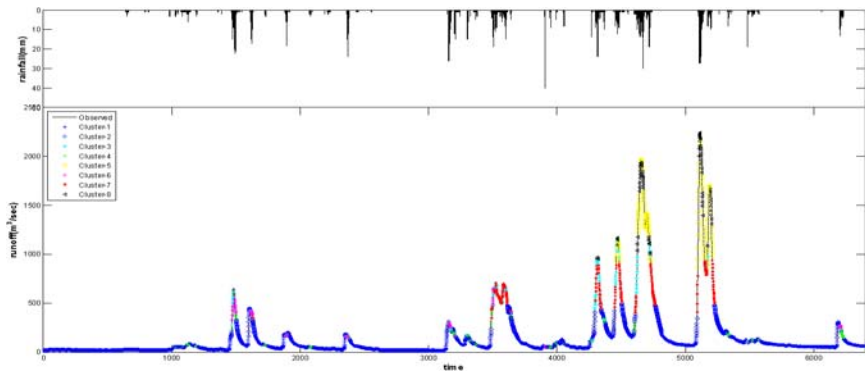
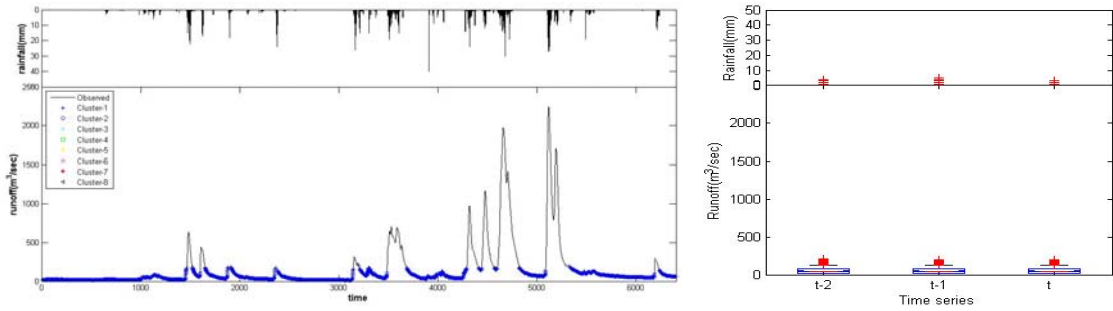
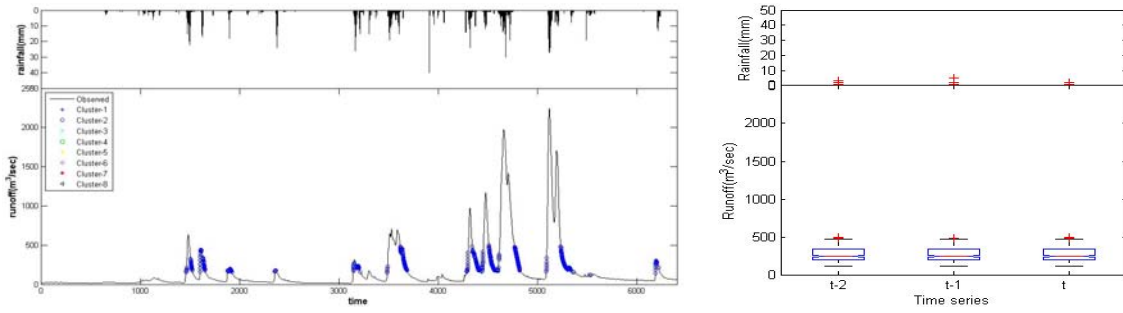


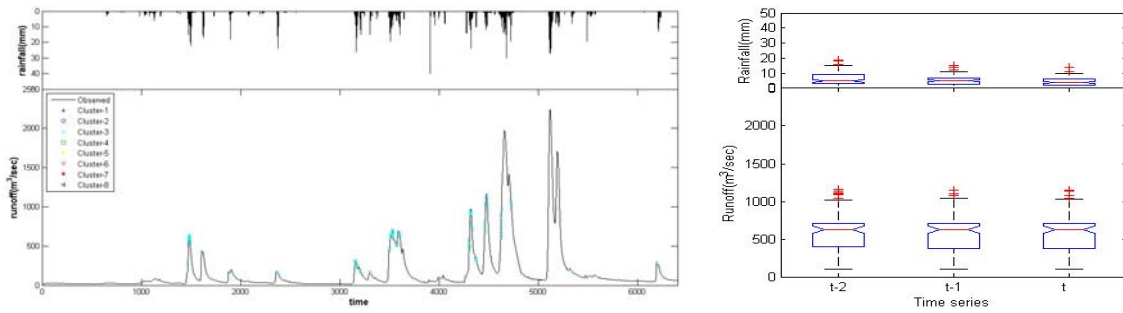
Fig. 9. Result of Rainfall-Runoff Relationship Analysis(20×16 Hexagonal Array)



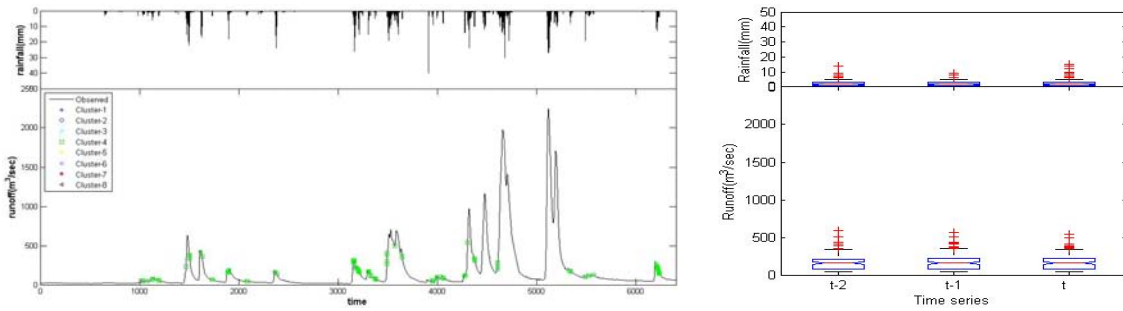
(a) Rainfall-Runoff relationship and Box Plot of cluster1



(b) Rainfall-Runoff relationship and Box Plot of cluster2

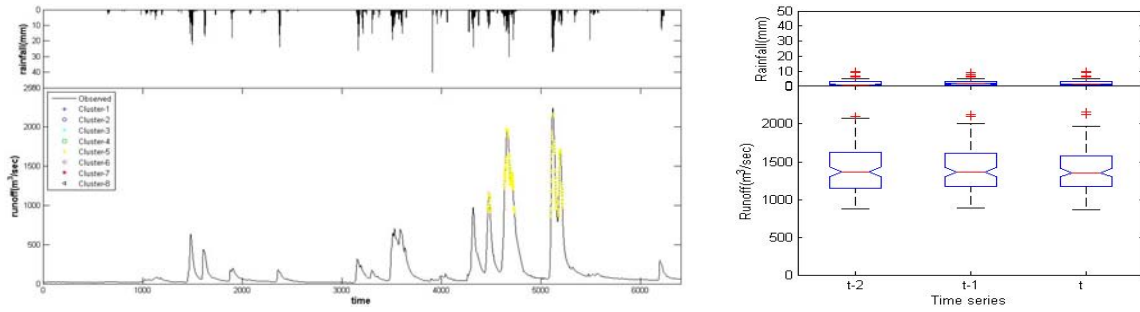


(c) Rainfall-Runoff relationship and Box Plot of cluster3

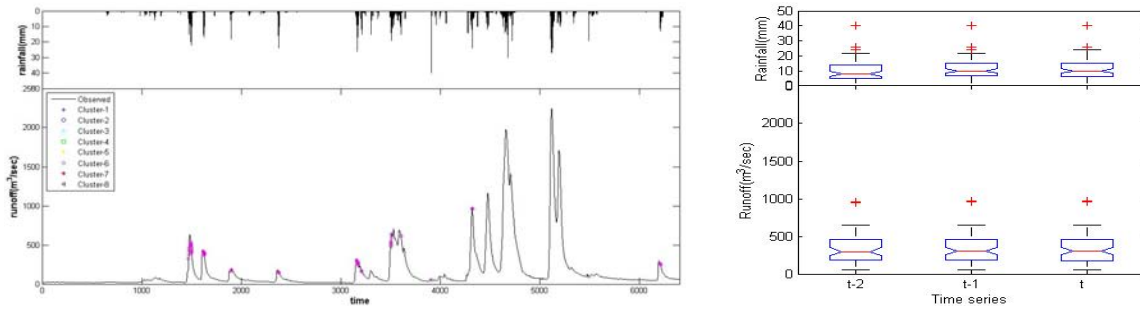


(d) Rainfall-Runoff relationship and Box Plot of cluster4

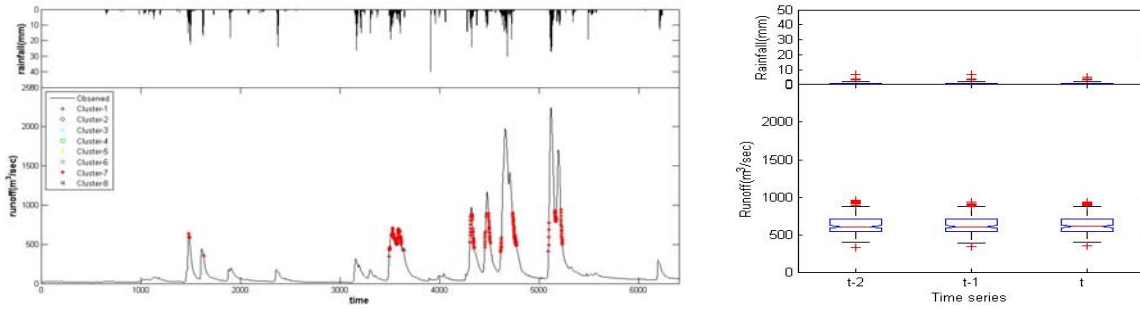
Fig. 10. Rainfall-runoff Relationship and Box Plot of 20×16 Hexagonal Array



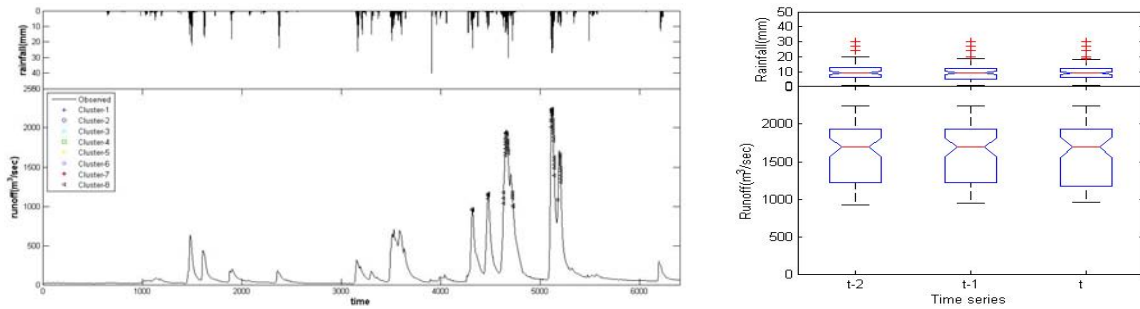
(e) Rainfall-Runoff relationship and Box Plot of cluster5



(f) Rainfall-Runoff relationship and Box Plot of cluster6



(g) Rainfall-Runoff relationship and Box Plot of cluster7



(h) Rainfall-Runoff relationship and Box Plot of cluster8

Fig. 10. Rainfall-Runoff Relationship and Box Plot of 20×16 Hexagonal Array(Continued)

각각의 지도구조의 분할구역별 특성을 살펴보면, 지도 크기 20×16의 육각형배열 구조의 분할구역별 특성의 경우 Cluster1은 대부분의 자료가 무강우기간에 해당하며 유출량은 평균 54.5 m<sup>3</sup>/sec로 시간구간별로 유출량의 변화는 미미한 수문곡선의 기저유출 부분으로 구분되었고, Cluster2의 경우는 작은 강우가 발생하거나 혹은 종료될 때 유출량은 점진적으로 증가하는 수문곡선의 상승기점과 그와 유사한 유출량에 대한 수문곡선의 감수부 유출부분으로 구분되고 유출량의 규모는 116.9 m<sup>3</sup>/sec~498.4 m<sup>3</sup>/sec의 범위로 나타났다.

Cluster3과 Cluster4는 강우가 감소함에 따라 유출량도 점진적으로 감소하는 경향으로 비슷한 양상을 보이나, 강우와 유출량의 규모에 따라 각각 다른 분할구역으로 구분되어지는 결과를 보여주고 있다. Cluster3의 강우량과 유출량 자료의 범위는 시간구간별 평균값으로 나타내면 강우량은 4.1 mm~5.8 mm, 유출량은 580.9 m<sup>3</sup>/sec~593.4 m<sup>3</sup>/sec의 범위로 나타났으며, Cluster4의 경우에는 강우량은 2.1 mm~2.9 mm, 유출량은 169.4 m<sup>3</sup>/sec~170.0 m<sup>3</sup>/sec로 Cluster3과 Cluster4는 강우와 유출량이 변화되는 경향은 유사하나 Cluster3이 강우와



Table 2. Basic Statistics for the Classified Regions of 20×16 Hexagonal Array

Cluster		P(t-2)	P(t-1)	P(t)	Q(t-2)	Q(t-1)	Q(t)
Cluster_1	Minimum	0.0	0.0	0.0	12.0	12.0	12.0
	Maximum	4.0	5.0	3.0	210.0	191.7	196.9
	Mean	0.0	0.0	0.0	54.6	54.5	54.5
	Median	0.0	0.0	0.0	46.9	46.9	46.9
	Variance	0.1	0.1	0.1	1397.0	1387.4	1389.6
	Standard Deviation	0.2	0.3	0.2	37.4	37.2	37.3
	Number of Data	5101	5101	5101	5101	5101	5101
Cluster_2	Minimum	0.0	0.0	0.0	114.6	116.9	116.9
	Maximum	3.0	5.0	2.0	498.4	485.0	498.4
	Mean	0.1	0.1	0.1	273.4	273.6	273.9
	Median	0.0	0.0	0.0	245.1	245.1	245.1
	Variance	0.2	0.2	0.1	7831.7	7815.7	8065.9
	Standard Deviation	0.4	0.4	0.4	88.5	88.4	89.8
	Number of Data	524	524	524	524	524	524
Cluster_3	Minimum	0.0	0.0	0.0	99.4	99.4	101.5
	Maximum	19.0	15.0	14.0	1157.2	1144.7	1153.1
	Mean	5.8	5.0	4.1	593.4	587.0	580.9
	Median	5.0	5.0	4.0	627.9	626.1	622.5
	Variance	18.0	11.2	8.1	73039	70105	68139
	Standard Deviation	4.2	3.4	2.9	270.3	264.8	261.0
	Number of Data	112	112	112	112	112	112
Cluster_4	Minimum	0.0	0.0	0.0	46.9	46.9	46.9
	Maximum	14.0	9.0	15.0	588.5	563.9	542.9
	Mean	2.5	2.1	2.9	170.0	169.7	169.4
	Median	2.0	2.0	2.0	156.6	159.1	159.1
	Variance	4.5	2.9	8.4	11449	10990	10596
	Standard Deviation	2.1	1.7	2.9	107.0	104.8	102.9
	Number of Data	147	147	147	147	147	147
Cluster_5	Minimum	0.0	0.0	0.0	869.2	889.5	861.1
	Maximum	10.0	9.0	10.0	2103.8	2129.0	2162.6
	Mean	2.0	2.2	2.1	1390.6	1387.2	1383.7
	Median	1.0	2.0	1.5	1367.3	1365.1	1352.5
	Variance	4.0	4.8	4.3	96412	92773	91096
	Standard Deviation	2.0	2.2	2.1	310.5	304.6	301.8
	Number of Data	142	142	142	142	142	142
Cluster_6	Minimum	0.0	0.0	0.0	50.4	52.2	52.2
	Maximum	40.0	40.0	40.0	954.8	967.1	967.1
	Mean	9.9	11.6	11.4	343.3	344.0	343.8
	Median	8.0	10.0	10.0	296.3	302.1	305.1
	Variance	47.8	42.9	46.0	34927	34510	33918
	Standard Deviation	6.9	6.5	6.8	186.9	185.8	184.2
	Number of Data	75	75	75	75	75	75
Cluster_7	Minimum	0.0	0.0	0.0	331.9	340.9	347.0
	Maximum	7.0	7.0	5.0	963.0	934.3	938.4
	Mean	0.7	0.7	0.7	633.7	635.9	638.2
	Median	0.0	0.0	0.0	602.8	602.8	609.9
	Variance	1.3	1.1	1.4	16534	16292	17014
	Standard Deviation	1.1	1.1	1.2	128.6	127.6	130.4
	Number of Data	221	221	221	221	221	221
Cluster_8	Minimum	1.0	1.0	1.0	926.2	942.5	954.8
	Maximum	30.0	30.0	30.0	2242.5	2242.5	2242.5
	Mean	10.2	9.8	10.2	1627.3	1638.0	1644.7
	Median	9.0	9.0	9.0	1691.2	1693.2	1693.2
	Variance	35.8	35.2	32.1	157480	155333	154837
	Standard Deviation	6.0	5.9	5.7	396.8	394.1	393.5
	Number of Data	78	78	78	78	78	78

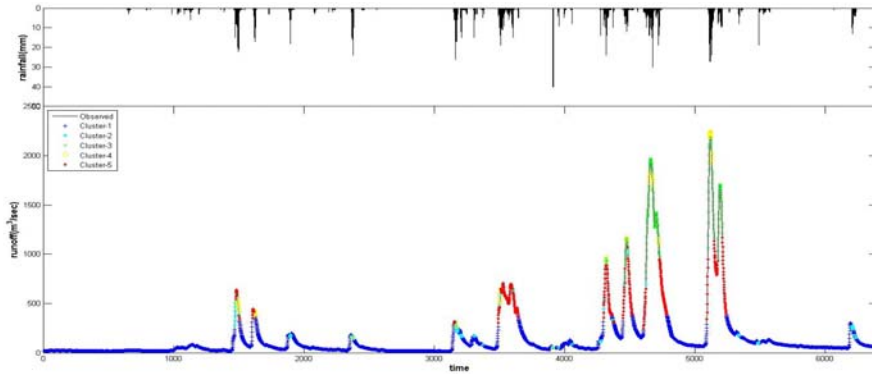


Fig. 11. Result of Rainfall-Runoff Relationship Analysis(24×20 Rectangular Array)

유출량의 규모면에서 Cluster4보다 큼을 알 수 있다.

Cluster5는 강우가 소강상태나 종료단계에서 변화가 미미한 경우로 유출량의 규모는 평균 1,380 m<sup>3</sup>/sec 이상으로 중규모 이상의 수문곡선의 상승부와 감수부의 유출로 구분되었으며, Cluster6은 강우자료의 범위가 가장 큰 분할구역으로 단기간에 집중해서 내리는 강우에 의해 발생하는 작은 규모의 수문곡선 첨두부를 나타낸다. Cluster7은 강우가 종료되거나 약한 강도의 강우가 발생할 때의 유출량은 347.03 m<sup>3</sup>/sec~938.41 m<sup>3</sup>/sec 범위로 수문곡선의 상승 기점과 고수위 유량을 구분 지어주는 수문곡선의 상승부와 감수부의 유출로 구분되어졌고, Cluster8의 경우는 유출량의 규모는 Cluster5의 경우와 유사하나 강우량은 Cluster5의 경우보다 더 큰 강우가 발생할 때 수문곡선의 첨두부와 중규모 이상의 상승부와 감수부의 유출로 구분되어졌다.

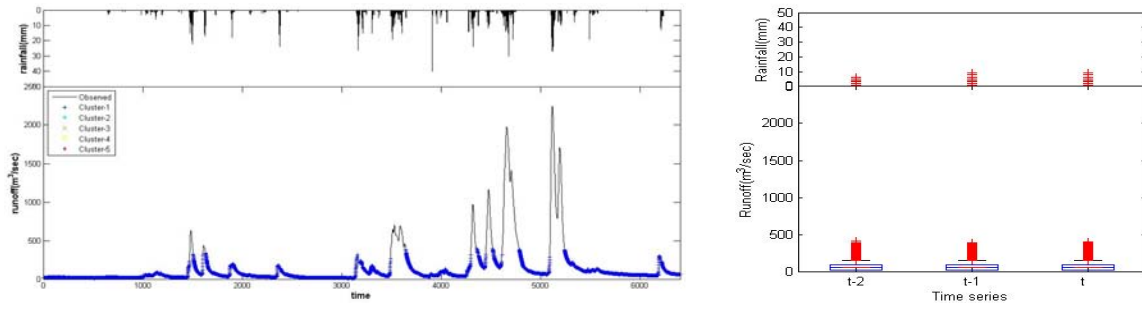
지도크기 24×20의 사각배열 구조의 분할구역별 특성을 살펴보면, Cluster1은 무강우기간의 수문곡선의 기저 유출 부분과 강우가 시작됨에 따라 유출량은 점진적으로 증가하는 수문곡선의 상승기점과 유사한 유출량에 대한 수문곡선의 감수부 유출부분으로 구분되었고, 지도크기 20×16의 육각형배열 구조와 비교해 보면 Cluster1과 Cluster2가 여기에 포함된다.

Cluster2는 평균값으로 보면 강우가 감소함에 따라 유출량도 점진적으로 감소하는 경향으로 구분되었고, Cluster3은 강우가 소강상태나 종료단계에서 유출량은 변화는 미미한 경우로 구분되었으며, 20×16의 육각형배열 구조와 비교해보면 Cluster5와 유사하다. 그러나 강우와 유출량 자료의 범위를 살펴보면 20×16의 육각형배열 구조의 Cluster5가 강우량과 유출량의 자료 범위 면에서 더 세분되게 구분되어짐을 알 수 있다. Cluster4는 다른 분할구역의 강우와 비교할 때 가장 강한 강도의 강우가 발생할 경우 유출은 유출량의 규모와는 무관하게 모든 수문곡선의 첨두부를 포함하고 있다. Cluster4의 강우량과 유출량 자료의 범위를 살펴보면

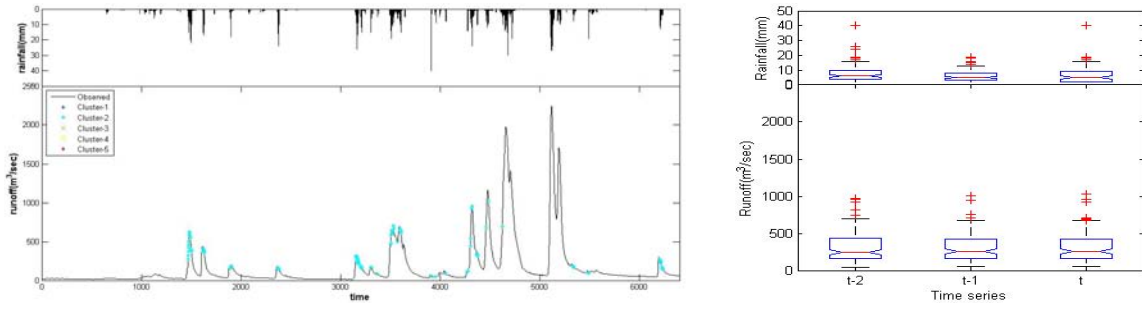
강우량은 무강우에서 최대 40 mm, 유출량은 52.5 m<sup>3</sup>/sec~2,242.5 m<sup>3</sup>/sec의 범위로 자료의 분산정도를 나타내는 표준편차가 가장 큰 분할구역으로 강우-유출관계를 세분적으로 구분하지 못하는 결과를 보인 분할구역이다. Cluster5는 20×16의 육각형배열 구조의 Cluster7과 유사한 분할구역으로 수문곡선의 상승 기점과 고수위 유량을 구분 지어주는 수문곡선의 상승부와 감수부의 유출로 구분되어졌다. 그러나 20×16의 육각형배열 구조의 Cluster7의 자료의 범위와 비교해 보면 강우량은 평균 1.0 mm 정도로 유사하나, 유출량은 24×20의 사각배열 구조의 Cluster5는 최소 245.1 m<sup>3</sup>/sec에서 최대 1,190.6 m<sup>3</sup>/sec로 20×16의 육각형배열 구조인 Cluster7의 331.9 m<sup>3</sup>/sec~963.0 m<sup>3</sup>/sec보다 더 큰 유출량의 범위로 구분되어짐을 알 수 있다.

이상의 결과를 토대로 나누지점의 강우-유출해석을 위한 적절한 지도구조를 결정하기 위해 지도크기 20×16의 육각형배열 구조와 지도크기 24×20의 사각배열 구조를 비교·검토해 보면 다음과 같다.

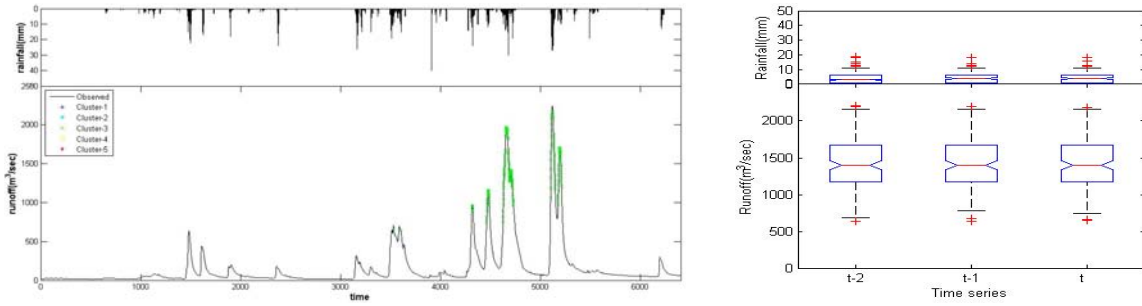
지도크기 20×16의 육각형배열 구조는 Cluster1의 기저유출부와 Cluster2의 수문곡선 상승기점 부근의 분류가 확인한 반면에 지도크기 24×20의 사각배열 구조는 기저유출부와 상승기점이 분류되지 못하고 하나의 분할 영역으로 구분되었다. 또한 Table. 2의 Cluster1과 Cluster2의 box plot의 결과에서 보듯이 자료의 이상치와 범위가 Table. 3의 Cluster1로 구분된 경우보다 자료의 이상치와 범위가 현저하게 작은 양호한 결과를 보여주었다. 또한 수문곡선의 상승기점과 고수위 유량을 구분지어주는 20×16의 육각형배열의 Cluster7과 24×20의 사각배열 구조의 Cluster5의 유출량 범위를 비교해보면 20×16의 육각형배열의 Cluster7은 331.9 m<sup>3</sup>/sec~963.0 m<sup>3</sup>/sec로 구분되었고, 24×20의 사각배열 구조의 Cluster5는 245.1 m<sup>3</sup>/sec~1190.6 m<sup>3</sup>/sec로 20×16의 육각형배열이 더 세분적으로 강우-유출관계를 구분짓는 결과를 보여주고 있다.



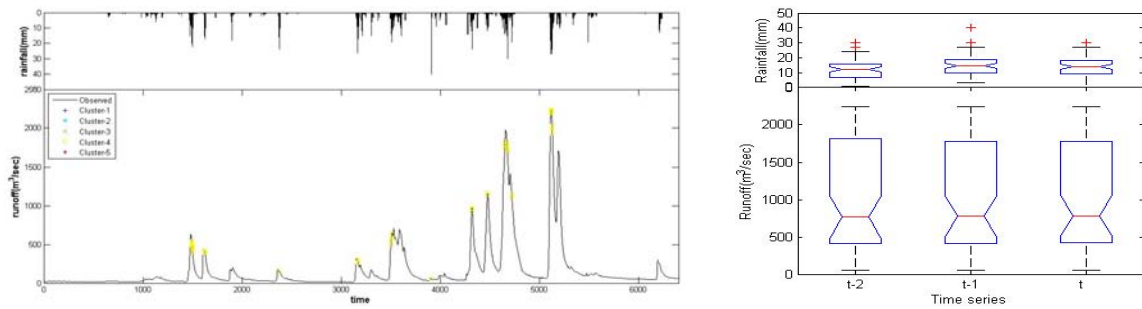
(a) Rainfall-Runoff relationship and Box Plot of cluster1



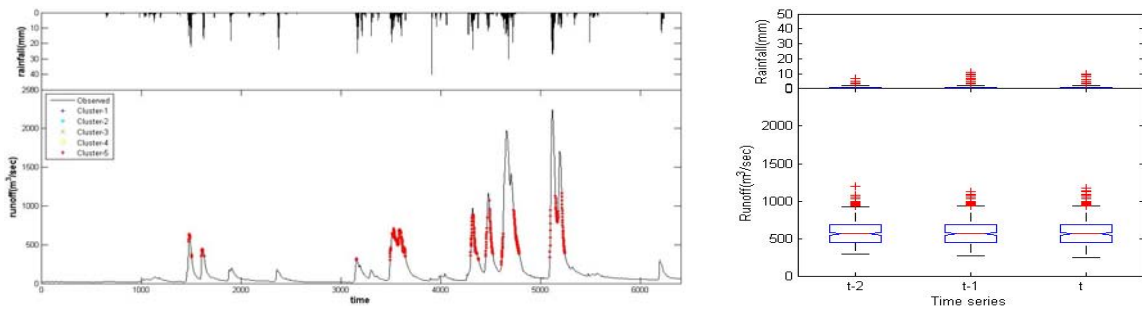
(b) Rainfall-Runoff relationship and Box Plot of cluster2



(c) Rainfall-Runoff relationship and Box Plot of cluster3



(d) Rainfall-Runoff relationship and Box Plot of cluster4



(e) Rainfall-Runoff relationship and Box Plot of cluster5

Fig. 12. Rainfall-runoff Relationship and Box Plot of 24×20 Rectangular Array

Table 3. Basic Statistics for the Classified Regions of 24×20 Rectangular Array

Cluster		P(t-2)	P(t-1)	P(t)	Q(t-2)	Q(t-1)	Q(t)
Cluster_1	Minimum	0.0	0.0	0.0	12.0	12.0	12.0
	Maximum	6.0	9.0	9.0	409.5	387.3	396.8
	Mean	0.1	0.1	0.1	69.8	69.8	69.8
	Median	0.0	0.0	0.0	50.4	50.4	50.4
	Variance	0.2	0.2	0.2	4039.3	4021.3	4040.5
	Standard Deviation	0.4	0.4	0.5	63.6	63.4	63.6
	Number of Data	5625	5625	5625	5625	5625	5625
Cluster_2	Minimum	0.0	0.0	0.0	46.9	48.7	48.7
	Maximum	40.0	19.0	40.0	971.2	1004.0	1028.7
	Mean	7.5	5.7	6.2	319.0	318.2	317.0
	Median	6.0	5.0	5.0	247.9	253.5	256.3
	Variance	31.8	17.7	30.5	42710	41661	40740
	Standard Deviation	5.6	4.2	5.5	206.7	204.1	201.8
	Number of Data	129	129	129	129	129	129
Cluster_3	Minimum	0.0	0.0	0.0	635.1	638.8	646.0
	Maximum	19.0	18.0	18.0	2213.0	2192.0	2179.4
	Mean	4.2	4.2	4.2	1435.9	1432.8	1428.0
	Median	3.0	4.0	4.0	1396.9	1396.9	1396.9
	Variance	13.6	12.7	11.7	114963	115644	118484
	Standard Deviation	3.7	3.6	3.4	339.1	340.1	344.2
	Number of Data	193	193	193	193	193	193
Cluster_4	Minimum	1.0	3.0	0.0	52.2	52.2	52.2
	Maximum	30.0	40.0	30.0	2242.5	2242.5	2242.5
	Mean	12.6	15.0	14.1	1046.6	1054.6	1060.4
	Median	12.0	14.5	14.0	765.7	775.9	784.3
	Variance	44.4	50.7	42.8	521487	525986	528992
	Standard Deviation	6.67	7.12	6.55	722.14	725.25	727.32
	Number of Data	58	58	58	58	58	58
Cluster_5	Minimum	0.0	0.0	0.0	290.5	270.4	245.1
	Maximum	7.0	11.0	10.0	1190.6	1119.8	1169.7
	Mean	0.8	1.0	0.9	586.2	587.4	589.0
	Median	0.0	0.0	0.0	560.4	563.9	563.9
	Variance	1.5	3.4	2.5	29631	29899	31424.
	Standard Deviation	1.2	1.8	1.6	172.1	172.9	177.3
	Number of Data	395	395	395	395	395	395

첨두부를 나타내는 20×16의 육각형배열 구조의 Cluster5와 Cluster8 그리고 24×20의 사각배열 구조의 Cluster3과 Cluster4를 비교해보면 20×16의 육각형배열 구조의 경우 유사한 규모의 유출량에 대해 강우량의 대소에 따라 첨두부가 2개의 분할구역으로 구분된 반면에 24×20의 사각배열 구조의 경우는 강우와 유출량 자료가 혼재되어 있으며, 표준편차의 통계치의 비교에서도 더 큰 값을 나타내고 있다.

이와 같이 분할구역별 자료의 분석을 통해 분할구역별 자료가 세분하게 구분되어지고, 기본적인 통계치의 비교에서도 양호한 결과를 보인 8개의 분할구역으로 강우-유출관계가 구분된 지도크기 20×16의 육각형배열 구조가 나주지점의 강우-유출해석을 위한 적절한 지도

구조라고 판단된다.

### 5. 결 론

본 연구는 기존에 적용한 인공지능망 이론들은 몇몇 문제점들을 해결하기 위해 패턴분류 성능을 지닌 자기조직화(Self-Organizing Map; SOM) 이론을 도입하여 자료의 전처리 과정으로 이용함으로써 기존의 인공지능망의 일반적인 적용을 개선한 연구들이 수행되어지고 있는데 SOM 훈련을 위한 지도크기 및 배열의 결정은 SOM 성능에 큰 영향을 미치나 지도크기 결정시 지도의 종방향 크기와 횡방향 크기를 결정할 수 있는 확정론적인 방법이나 이론식이 없고 지도배열 또한 주로 유

각형 배열(Hexagonal Array)을 이용하여 적용하고 있다. 따라서 본 연구에서는 영산강 나주지점을 대상으로 강우-유출 해석을 위한 적절한 Map 구조를 결정하기 위해 다양한 방법으로 검토를 실시하였고, 그 결과는 다음과 같다.

- 1) 적절한 지도구조 결정을 위한 1차선별은 지도를 구성하는 단위구조의 총 수(M)를 이용하여 기존에 연구 보고된 식인  $M=5\sqrt{N}$ 을 포함한 다소 변형한  $M=4\sqrt{N}$ ,  $M=6\sqrt{N}$ 에 대해서도 검토한 결과 강우-유출간의 관계가 지도배열별로 동일하게 5개 이상의 분할 특성으로 구분되어지는  $M=4\sqrt{N}$  식을 적용한 지도크기 20×16과  $M=6\sqrt{N}$  식을 적용한 지도크기 24×20이 1차 검토대상으로 선정되었다.
- 2) 선정된 지도구조에 대해 K-means 알고리즘의 결과를 이용하여 2차선별을 실시한 결과 근접반경의 조정과정에 문제가 되는 지도구조를 제외하고 지도크기 20×16의 육각형배열 구조와 24×20의 사각배열구조가 2차 검토대상으로 선정되었다.
- 3) 2차 검토대상으로 선정된 지도구조의 비교·분석 결과 분할구역별 자료가 세분하게 구분되어지고, 기본적인 통계치 및 도식적인 비교에서도 양호한 결과를 보인 8개의 분할구역으로 강우-유출관계가 구분된 지도크기 20×16의 육각형배열 구조가 나주지점의 강우-유출해석을 위한 적절한 지도구조라고 판단된다.

이와 같은 결과를 토대로 보다 합리적이고 구체적인 SOM 적용이 가능할 것으로 판단되며, 강우-유출 예측

모형의 개발을 위한 선행과정으로 활용되어질 것으로 기대된다.

## 참 고 문 헌

- 건설교통부(2007). **유량조사 보고서**, pp. 596-599.
- 김대수(1996). **신경망 이론과 응용(I)**, 하이테크정보, pp. 243-282.
- 김용구(2006). 자기조직화 이론을 적용한 강우-유출 예측모형 개발. 박사학위논문, 동신대학교.
- 김용구, 진영훈, 박성천(2006). “강우-유출특성 분석을 위한 자기조직화방법의 적용.” **대한토목학회 논문집**, 대한토목학회, 제26권, 제1B호, pp. 61-67.
- 박성천, 진영훈, 김용구(2006). “강우-유출 예측모형 개발을 위한 자기조직화 이론의 적용.” **대한토목학회 논문집**, 대한토목학회, 제26권, 제4B호, 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*, 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:08-84/접수:2008.07.29/심사완료:2008.08.18)