

SOM 적용을 위한 Map Size와 Array의 변화에 따른 강우-유출 및 TOC관계 분석

Analysis of Classification Characteristics for Rainfall-runoff and TOC Variation according to the Change of Map Size and Array using SOM

박성천*, 김용구**, 노경범***, 이한민****

Sung Chun Park, Yong Gu Kim, Kyong Bum Roh, Han Min Lee

요 지

본 연구는 인공신경망(Artificial Neural Networks: ANNs)기법의 일종인 자기조직화(Self Organizing Map: SOM) 이론을 이용한다. 자기조직화 특성을 이용하여 스스로 학습이 가능하고, 구조상 수행이 빨라 학습 단계에 소요되는 시간을 줄일 수 있는 장점을 가진 자기조직화 이론을 도입하고, 수질자료 중 전체 유기물의 양을 나타내며 난분해성 물질에 대한 해석이 가능하고 재현성이 탁월한 TOC 와 강우-유출량 자료의 분포적 양상과 특징을 분석하여 예측을 위한 모형화 과정에 기여하고자 한다. 최적의 Map Size와 Map Array 결정을 위해 수집된 강우와 유출량자료 및 TOC 자료에 대해 Garcia의 경험식을 이용하여 Map을 구성하는 단위구조의 총 수(M)를 산정하여 M값에 따른 중방향 및 횡방향 크기를 결정하는 다수의 Map 크기를 검토하고, 또한 Map 배열은 2차원 배열의 사각형배열(Rectangular array)과 육각형배열(Hexagonal array)에 대해서도 복합적으로 검토하여 최적의 특성조건을 결정하여 강우-유출 및 TOC 관계의 분할특성을 분석한다.

핵심용어 : 인공신경망, 자기조직화 이론, Map Size, Map Array

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 수학적인 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 연구에서는 강한 비선형성을 나타내는 입·출력간의 관계를 나타내는데 탁월한 성능을 보이고, 자료들의 급격한 변화에 대해 뛰어난 적응성을 보여주는 것으로 알려져 있는 SOM 적용을 위한 최적의 Map Size와 Map Array를 결정하고자 한다. Map Size와 Map Array 결정을 위해 수집된 강우와 유출량자료

* 정회원·동신대학교 토목공학과 교수·E-mail : psc@dsu.ac.kr

** 정회원·동신대학교 토목공학과 연구원·E-mail : kyg8987@paran.com

*** 정회원·동신대학교 토목공학과 연구원·E-mail : kbyj3711@naver.com

**** 정회원·광주광역시청 종합건설본부 토목지원과·E-mail : inseoi@naver.com

및 TOC 자료에 대해 Garcia의 경험식을 이용하여 Map을 구성하는 단위구조의 총 수(M)를 산정하여 M값에 따른 종방향 및 횡방향 크기를 결정하는 다수의 Map 크기를 검토하고, 또한 Map 배열은 2차원 배열의 사각형배열(Rectangular array)과 육각형배열(Hexagonal array)에 대해서도 복합적으로 검토하여 최적의 특성조건을 결정하여 강우-유출 및 TOC 관계의 분할특성을 분석한다.

2. 대상지점 및 입력자료

본 연구의 대상지점인 영산강 유역은 우리나라 서남부에 위치하며, 총 유역면적은 3,455km²이고, 본류의 유로연장은 129.5km이며, 본 연구에서는 강우의 시-공간적 분포의 비선형적 변동성을 고려한 강우패턴을 분류하고 강우-유출 및 TOC 관계 분석을 위해 그림 1과 같이 영산강 본류의 대표지점인 나주수위관측소와 그에 영향을 미치는 상류의 광산 강우관측소를 선정하였다. TOC 수질자료는 수질오염 조기경보체계 구축을 위해 영산강유역환경청에서 관리·보관 중인 자료를 이용하였으며 수위 및 강우자료는 영산강홍수통제소의 시자료를 이용하였고, 나주지점의 수위자료를 유출량자료로 환산하기 위해 『영산강 유량측정 보고서』(영산강홍수통제소, 2004)의 나주지점의 수위-유량관계곡선을 이용하였다.

SOM 훈련 데이터 구성은 강우-유출 및 TOC 사이의 관계를 파악하기 위해서는 긴 시간의 다수의 자료가 필요하나 본 연구에서 자료의 검토결과 강우-유출 및 TOC 관계를 파악하기 위한 최소한의 시간이 세 개의 연속자료를 나타내는 3시간으로도 그 특성 파악이 가능하다고 판단되고, 자료의 크기를 최소화하기 위하여 강우-유출 및 TOC 특성을 구분하기 위하여 3시간 자료를 SOM의 입력자료로 이용하였다.

그러므로 입력벡터 $x(t)$ 는 식 1과 같은 9차원 구조의 SOM 훈련 데이터의 구조가 된다.

$$x(t) = (P_{(t-2)}, T_{(t-2)}, Q_{(t-2)}, P_{(t-1)}, T_{(t-1)}, Q_{(t-1)}, P_{(t)}, T_{(t)}, Q_{(t)}) \dots \dots \dots (1)$$

여기서, P는 강우량, T는 TOC, Q는 유출량

3. 적용방법 및 결과



그림 1. 대상지점 유역도

3.1 적용방법

SOM의 적용을 위해 강우-유출 및 TOC 자료에 대해 2003년 3월 16일 ~ 2003년 6월 10일, 2004년 2월 15일 ~ 2004년 6월 5일까지의 4774개의 자료를 선택하여 3시간의 시차를 적용한 4774×9 행렬의 SOM 입력자료를 구축하였다. 선택된 자료의 정규화 과정을 거친 후 SOM 훈련을 위해 Map Size와 Map array를 결정하는데 우선, Map을 구성하는 단위구조의 총 수(M)의 결정은 Garcia 등(2003)에 의해 연구 보고된 식 2를 이용하였다.

$$M = 5\sqrt{N} \dots \dots \dots (2)$$

여기서, M은 Map을 구성하는 단위구조(unit)의 총 수, N은 훈련 자료의 수이다.

Map 크기 결정시 Map의 종(세로)·횡(가로)방향의 크기를 결정 할 수 있는 확정론적인 방법이나 이론식이 없어 식 2에 의해 산정된 345(Map을 구성하는 단위구조의 총 수; M)의 근사치로부터 다양한 Map Size를 선택하기

위하여 Map을 구성하는 단위구조의 총 수 M의 값에 따른 Map Size를 결정하는 다수의 경우를 상호·비교하였고, 사각형배열(Rectangular Array)과 육각형배열(Hexagonal Array)에 대해서도 복합적으로 검토하였다.

그 결과 표 1과 같이 단위구조의 수에 따른 Map Size 및 Map Array에 따른 분할구역을 각각 구분하였고, 본 연구에서 구축한 SOM 훈련자료를 식 2에 적용하여 산정된 단위구조의 총 수 345와 동일한 23×15의 Map Size를 각각의 Map array별로 구분하여 분할구역별 특성을 검토하였다. 또한 각각의 Map Size 및 Map Array에 따른 분할구역으로 구분된 결과는 그림 2와 같다.

표. 1 단위구조의 수에 따른 Map Size 및 Map array에 따른 분할구역 구분

육각형 배열(Hexagonal Array)				사각형 배열(Rectangular Array)			
Map Size		단위구조의 수	분할구역 구분	Map Size		단위구조의 수	분할구역 구분
종방향	횡방향			종방향	횡방향		
20	17	340	5	20	17	340	5
21	16	336	5	21	16	336	7
22	16	352	8	22	16	352	9
23	15	345	7	23	15	345	8
24	14	336	3	24	14	336	9
25	14	350	8	25	14	350	9

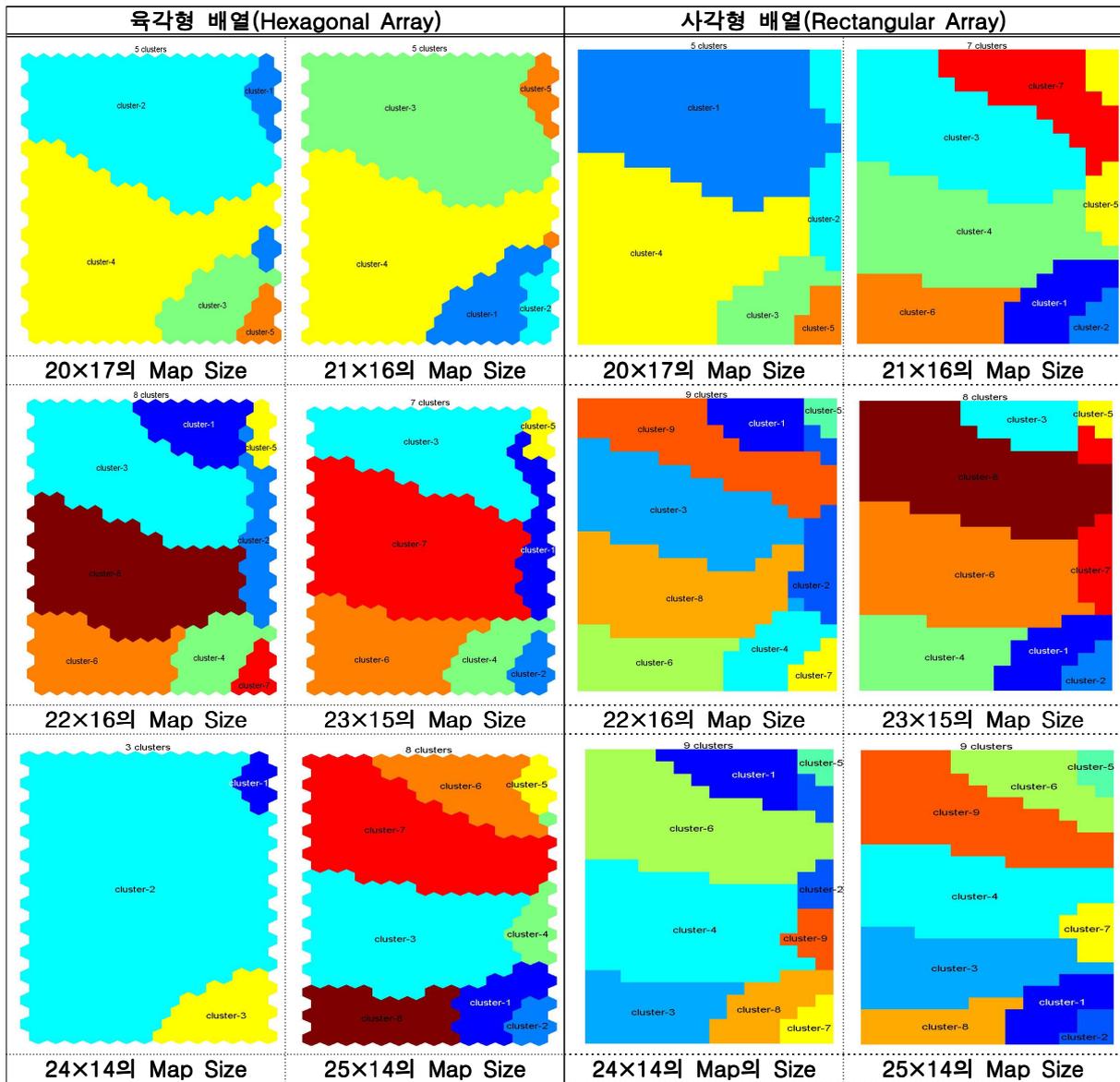
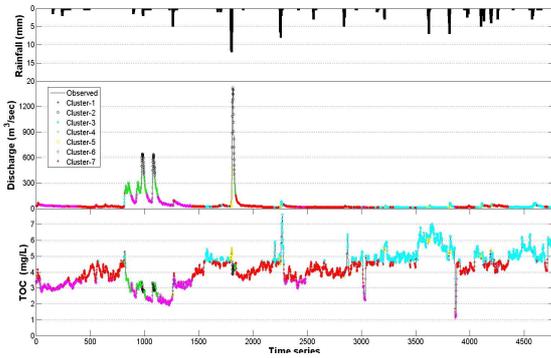


그림 2. k-means algorithm 결과

육각형 배열(Hexagonal Array)



사각형 배열(Rectangular Array)

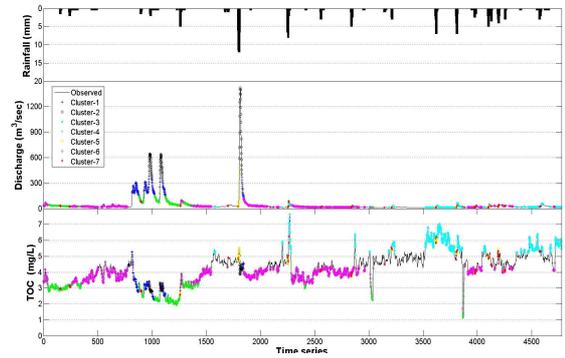


그림 2. k-means algorithm 결과

본 연구에서 적용한 입력자료로 산정된 단위구조의 총 수 345와 동일한 23×15의 Map Size의 육각형 배열의 분할구역별 특성을 살펴보면, Cluster-1은 강우가 시작되는 단계에 해당하며 유출량과 TOC가 점진적으로 상승하는 경향을 나타내고 있으며, Cluster-2는 강우가 종료되는 단계에서 유출량은 대규모로 발생하며 수문곡선의 첨두부를 포함한 상승부와 하강부로 구분되었으며, TOC농도는 유출량의 증가로 그 값이 감소하는 경향으로 나타났다. Cluster-3은 규모가 작은 강우가 발생하며 유출량은 점진적으로 증가하고, TOC농도는 증가하는 경향을 나타내고 있으며, Cluster-4는 강우량은 감소하고 유출량은 수문곡선에 변곡점 아래의 상승부와 하강부의 중규모 유출량의 양상으로 구분되었고, TOC농도는 약 3.28~3.31mg/L에서 감소하는 부분으로 분할되었다. Cluster-5는 강우량이 증가함에 따라 유출량은 급격히 증가하고, TOC농도가 비교적 5.03mg/L의 높은값에서 서서히 감소를 하는 양상으로 구분되었으며, Cluster-6은 강우가 종료되고 유출량은 감소하며, TOC농도는 2.98mg/L의 저농도에서 차차 증가하는 경향으로 구분 되었다. 마지막으로 Cluster-7은 무강우 기간에 해당되며, 유출량은 점진적으로 감소하거나 변화가 없는 반면에 TOC농도는 증가하는 경향으로 구분되었다. 사각형 배열의 경우는 하나의 분할구역이 세분되어 분할구역별로 구분된 원자료와 Map 경향은 다르지만 분할구역별 특성은 육각형 배열의 경우와 유사하였다.

4. 결론

본 연구는 Map 크기 및 배열 결정에 관한 확정론적인 방법이나 이론식이 없어 Map Size 결정시 Garcia의 경험식을 이용하여 Map을 구성하는 단위구조의 총 수(M)를 산정하여 M값에 따른 크기를 결정하는 6가지의 Map Size를 검토하였다. 또한 동일한 Map Size에 대해 사각형배열(Rectangular Array)과 육각형 배열(Hexagonal Array)의 Map array에 대해서도 복합적으로 검토한 결과 사각형 배열의 경우는 5~9개의 분할구역으로 육각형 배열의 경우는 3~8의 분할구역으로 구분되었다. 또한 횡방향의 크기와 종방향의 크기를 다르게 하여 검토한 결과 횡방향 크기보다 종방향 크기를 크게 하였을 때 다양한 패턴으로 구분되었다.

향후 SOM 적용을 위한 최적의 Map Size 및 Map array의 결정을 위해서는 분할구역별로 구분된 자료를 예측모형에 적용하여 통계적인 기법과 도식적인 기준을 비교·검토하는 연구가 진행되어야 할 것으로 판단된다.

참 고 문 헌

- 1 김용구(2006) 자기조직화 이론을 적용한 강우-유출 예측모형 개발, 박사학위논문, 동신대학교.
- 2 건설교통부 영산강홍수통제소(2004) 영산강 유량측정 보고서.
- 3 김용구, 진영훈, 박성천(2006) 강우-유출특성 분석을 위한 자기조직화방법의 적용, 대한토목학회 논문집, 제26권, 제1B호, pp.61-67
- 4 박성천, 진영훈, 김용구(2006) 강우-유출 예측모형 개발을 위한 자기조직화 이론의 적용, 대한토목학회 논문집, 제26권, 제4B호, pp.389-398

- 5 이관수, 박성천, 이한민, 진영훈(2000) 인공신경망의 이론의 B.P.알고리즘을 적용한 영산강의 유출량 예측에 관한 연구, **대한토목학회 논문집**, 제 20권, 제5B호, pp. 679-688.
- 6 Coulibaly, P., Anctil, F. and Bobee, B.(2000) Daily reservoir inflow forecasting using artificial neural networks with stopped training approach, *Journal of Hydrology*, Vol. 230, pp. 224-257.
- 7 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.
- 8 Hsu, K. L., Gupta, H. V., Gao, X., Sorochian, 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.
- 9 Luk, K. C., Ball, J. E. and Sharma, A.(2000) A Study of optimal model lag and spatial inputs to artificial neural network for rainfall forecasting, *Journal of Hydrology*, Vol. 227, pp. 56-65.
- 10 Kohonen, T.(1982) Self-organized formation of topologically correct feature maps, *Biological Cybernetics*, Vol. 43, pp. 59-69.