

# SOFM의 적용에 의한 영산강 수질 및 유량자료의 시·공간적 패턴분류 특성

## Characteristics of Spatio-temporal Pattern Classification for Water Quality and Runoff Data in the Yeongsan River by the Application of SOFM

박성천\*, 송자섭\*\*, 진영훈\*\*\*, 노경범\*\*\*\*

Sung-Chun Park, Ja-Seob Song, Young-Hoon Jin, Kyong-Bum Roh

.....  
유역관리 및 수질 향상을 위해 다양한 환경정책이 시행되고 있으며, 최근 수질오염총량관리제의 시행으로 인해 보다 집중적인 유역관리와 수질 향상을 위한 노력이 배가되고 있다.

이러한 노력의 일환으로 현재 환경부 국립환경과학원에서는 수질오염총량관리를 위하여 단위유역의 말단지점에서 수질 및 유량자료에 대한 정기적인 측정을 8일 간격으로 시행하고 있으며, 데이터 베이스 및 웹시스템을 통하여 자료를 공개하고 있다(이호열, 2009).

이와 같은 자료의 측정과 축적은 그 분석을 통해 수질 개선을 위한 우선 관리 대상지점의 파악 등과 같이 수질오염총량관리제의 시행과 평가를 위해 사용될 수 있을 뿐만 아니라 새로운 환경정책의 수립에도 활용될 수 있을 것으로 기대된다.

그러나 현재 소수의 연구들에서만 상기의 자료를 단순히 활용한 결과를 찾을 수 있으며, 특히 측정된 수질 및 유량자료를 분석하여 발표한 연구결과 역시 소수에 지나지 않는다(김철겸 등, 2009). 측정 자료에 대한 분석 및 이에 따른 자료의 활용성 제고를 위해서 다양한 자료 분석 기법의 개발과 적용이 절실하다.

이러한 자료 분석 기법의 개발 및 적용에 관한 연구의 일환으로 최근 패턴분류를 위해 다양한 분야에서 활용되고 있는 자기조직화 특성 지도(Self Organizing Feature Map: SOFM)를 상기의 측정 자료에 적용한 연구 결과가 보고된 바 있다(진영훈 등, 2009; 2010).

본 연구에서는 수질오염총량관리제를 위해 측정되고 있는 수질 및 유량자료를 수집하여 자료에 내재되어 있는 시·공간적 특성을 분석하고자 하였다. 영산강 유역을 대상으로 하여, 본 유역 내의 단위유역들 중 황룡\_A, 지석\_A, 영본\_A, 영본\_B, 영본\_C, 영본\_D의 말단지점에서 측정되고 있는 BOD (Biochemical Oxygen Demand), TOC (Total Organic Carbon), T-N (Total Nitrogen), T-P (Total Phosphorus), SS (Suspended Solids) 수질농도 및 유량자료를 대상으로 연구를 진행하였다.

**핵심용어 : 수질 및 유량자료, 시·공간적 특성을 분석, 자기조직화 특성 지도**

---

\* 정회원 · 동신대학교 토목공학과 교수 · E-mail : [psc@dsu.ac.kr](mailto:psc@dsu.ac.kr)  
\*\* 정회원 · 전라남도청 · E-mail : [jaseobsong@hanmail.net](mailto:jaseobsong@hanmail.net)  
\*\*\* 정회원 · 수도대학동경 대학원 도시환경과학연구과 연구교수 · E-mail : [nmdrjin@gmail.com](mailto:nmdrjin@gmail.com)  
\*\*\*\* 정회원 · 전남발전연구원 환경해양연구팀 연구원 · E-mail : [kbyj3711@naver.com](mailto:kbyj3711@naver.com)

## 1. 서론

본 연구는 수질오염총량관리제를 위해 단위유역 말단지점에 8일 간격으로 측정되어 축적되고 있는 수질 및 유량자료의 활용성을 제고하고, 수질오염총량관리제와 같은 환경정책의 지속적인 추진과 향후 새로운 환경정책의 수립에 대한 기초적인 정보 제공을 위하여 자료 분석 기법을 개발하여 적용하고자 하였다. 자료 분석 기법으로는 최근 패턴분류를 위해 전 세계적으로 그 적용이 활발히 이루어지고 있는 SOFM을 활용하였다.

먼저 연구를 위한 대상지점을 영산강 유역 내에 존재하는 단위유역의 말단지점인 황룡\_A, 지석\_A, 영본\_A, 영본\_B, 영본\_C 및 영본\_D를 선정하여 수질 및 유량자료를 수집하였다. 현재 측정되고 있는 다양한 수질자료들 중 수질오염총량관리제를 위해 그 중요도가 높은 항목들을 선별하여 유량자료와 함께 본 연구를 위하여 적용하였다.

## 2. 대상 지점

본 연구에서는 수질오염총량관리제가 시행되고 있는 영산강 유역 내에 존재하는 단위유역 말단지점 중 하구언의 조작에 따른 유량자료에 인위적인 간섭이 현저한 영본\_E 지점을 제외한 6개 지점을 대상지점으로 선정하였다.

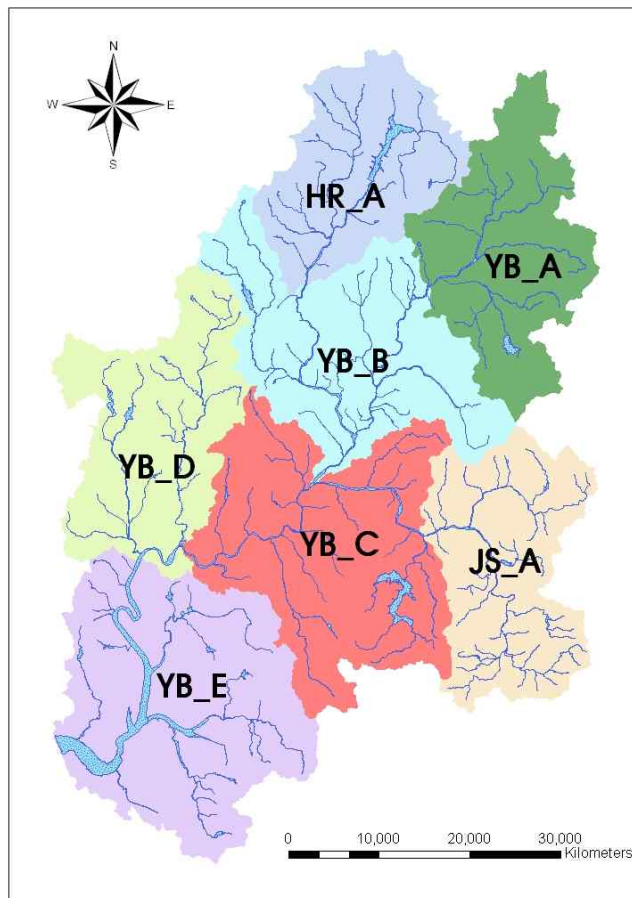


Fig. 2.1 Study area and location of unit watersheds

## 2.1 수질 및 유량자료의 분포

수집된 수질 및 유량자료에 대한 분포를 알아보기 위해 각 지점에 대하여 최소값, 1사분위값, 중앙값, 3사분위값 및 최대값을 나타내는 박스 플롯을 도시하였다. 황룡\_A와 지식\_A의 경우 유량자료의 분포가 수질자료의 분포에 비해 넓게 퍼져 있음을 알 수 있다.

영본\_A의 경우에는 BOD, TOC 및 유량자료의 분포가 다른 수질항목들의 분포에 비해 상당히 넓게 분포하고 있으며, 영본\_B, 영본\_C 및 영본\_D 지점의 자료에 대한 박스 플롯을 보면, 전반적으로 수질항목의 중앙값이 크며 유량자료 역시 큰 값들을 나타내고 있다.

본 연구에 사용된 전 지점에 대한 전체자료를 대상으로 박스 플롯을 Fig. 2.3에 도시하였으며, 그 결과 유량자료가 넓게 분포되어 있음을 재확인할 수 있었다. 특히 SS의 경우 높은 왜곡도를 나타내고 있었다.

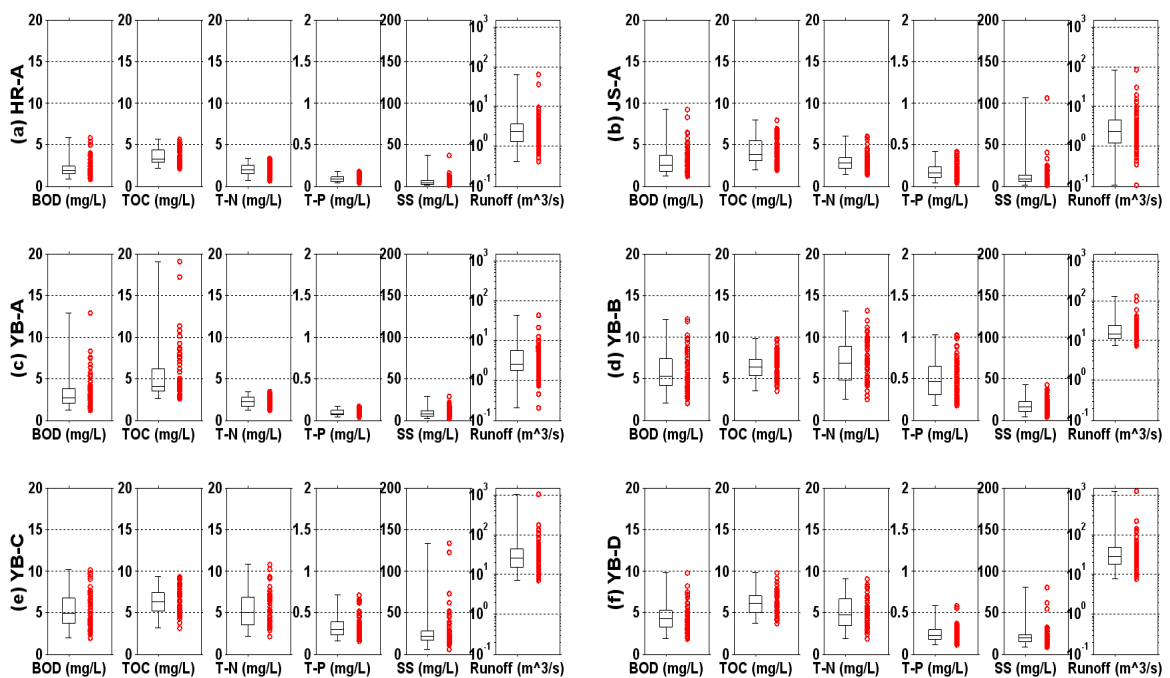


Fig. 2.2 Box-whisker plot for water quality and runoff data at the respective stations:

(a) HR\_A, (b) JS\_A, (c) YB\_A, (d) YB\_B, (e) YB\_C, (f) YB\_D

## 3. 자기조직화 특성 지도

Helsinki대학의 Kohonen(1982)은 사용자의 개입없이 스스로 입력 자료를 조직화할 수 있는 시스템을 소개하였으며, 이 시스템은 사용자의 도움없이 외부의 공간에 위치하는 입력자료를 시스템이 정의한 내부의 공간으로 사상시킨다. Kohonen은 이 시스템을 Self-Organizing Feature Map (SOFM) 또는 Adaptive Vector Quantizer (AVQ)라고 명명하였다. 그 이후로 다양한 종류의 SOFM이 나타나게 되었고, 많은 발전이 있었다.

SOFM은 다차원의 자료를 2차원으로 사상(mapping)시켜주는 인공신경망 이론의 일종이며, 클러스터링을 위한 방법으로 최근 다양한 분야에서 그 적용이 활발히 이루어지고 있다. 특히 SOFM은 복잡한 다차원 자료의 패턴분류 또는 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.

SOFM은 비감독 학습방법(unsupervised learning algorithm)의 일종으로 스스로 n차원의 입력

자료들을 클러스터링하여 2차원으로 사상시켜주며, 이러한 SOFM의 일반적인 구조다. SOFM 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 입력 벡터와 입·출력 노드들을 연결하는 연결강도벡터를 이용하여 두 벡터간의 거리(distance)를 계산하는 기능을 수행한다. 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. 이 승자 노드만이 출력신호를 보낼 수 있는 유일한 노드가 되며, 이 노드와 이와 인접한 이웃 노드들만이 제시된 입력벡터에 대하여 학습이 허용된다. 이것은 학습에 있어서 새로운 접근방식이며, 이러한 모형이 제안되기 이전에는 네트워크에 있는 모든 노드들이 반복학습과정인 훈련을 통해 연결강도를 조정하는 방법을 이용해 왔다. Kohonen 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 과정인 경쟁과정, 근접반경의 조정과정, 적응학습 과정이 진행되는 동안 연결강도를 조정하게 된다.

#### 4.SOFM 패턴분류

본 연구에 사용된 SOFM의 구조를 살펴보면 자료의 총 수(392개)에 대하여 99개의 노드를 사용할 수 있으며, 주성분 분석에 의한 제 1 고유치와 제 2 고유치의 비율에 의해 가로 및 세로축에 배치될 노드의 수를 산정하였다. 결과적으로 16(세로축) × 6(가로축) 배열의 SOFM을 구축할 수 있었으며, 훈련을 통해 생성된 U-matrix를 Fig. 4.1에 나타내었다.

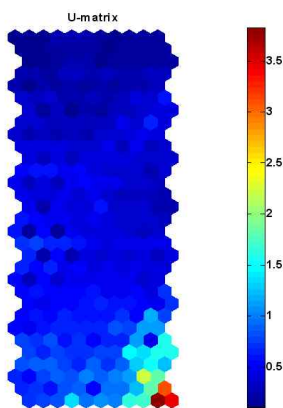


Fig. 4.1 U-matrix generated by SOFM for the study

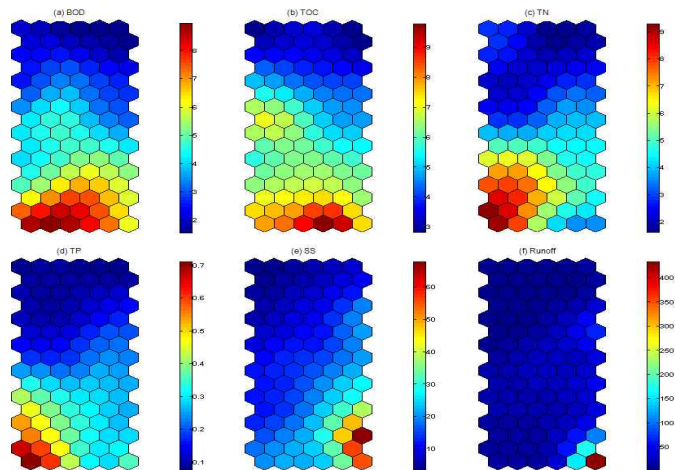


Fig. 4.2 Component planes for the six variables after denormalization

본 연구를 위해 사용된 각 변수들에 대한 각각의 패턴분류 결과를 Fig. 4.2에 component plane을 이용하여 나타내었다. Fig. 4.2 (a) - (d)에서 보는 바와 같이 BOD, TOC, T-N 및 T-P에 대한 component plane은 유사한 변화 양상을 보이고 있으며, SS(Fig. 4.2 (e))와 유량(Fig. 4.2 (f))은 이와는 상이한 변화 패턴을 보이고 있다. component plane의 변화 양상에 따른 결과에 의해

BOD, TOC, T-N 및 T-P가 강한 상관성을 가지고 있으며, SS와 유량자료가 또 다른 강한 상관성을 가지고 있음을 알 수 있다.

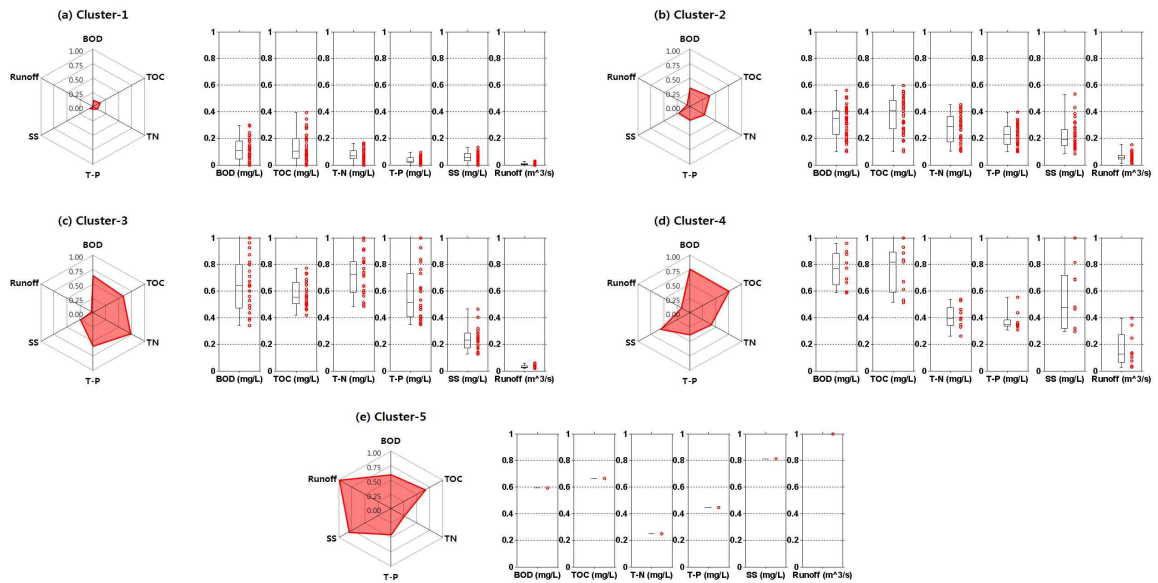


Fig. 4.3 Radar chart for the standardized mean values of the weights classified into the respective clusters and box-whisker plots for the values

각 클러스터별 패턴분류 특성을 나타내기 위하여 각 클러스터로 분류된 노드들의 가중치를 이용하여 Fig. 4.5와 같은 레이다 차트를 도시하였다

## 5. 결론

수질오염총량관리제를 위해 측정되고 있는 수질 및 유량자료의 활용성 제고를 위한 분석 기법의 개발을 위해 SOFM과 Ward 기법의 일련적 적용이 타당한 것으로 판단되며, 이에 따른 패턴분류 결과의 보다 심도있는 분석을 통해 측정된 자료의 시·공간적 분포를 파악할 수 있었다. 이러한 결과는 향후 환경정책의 입안 및 지속적인 추진을 위한 기초적인 정보를 제공해 줄 수 있을 것으로 기대된다

## 참고문헌

1. 광주광역시 (2009), 제2단계 광주광역시 영산강 수질오염총량관리 기본계획.
2. 김용구, 진영훈, 박성천 (2006), 강우-유출특성 분석을 위한 자기조직화방법의 적용, 대한토목학회논문집, 26(1B), pp. 61-67
3. 김철겸, 김남원 (2009), 환경부 8일 간격 유량·수질 관측자료와 분포형 모형을 이용한 연속오염부하곡선의 유도, 한국물환경학회논문집, 25(1), pp. 125-135.
4. Aguilera, P. A., Frenich, A. G., Torres, J. A., Castro, H., Vidal, J. L. M. and Canton, M. (2001), Application of the Kohonen neural network in coastal water management: Methodological development for the assessment and prediction of water quality, Water Research, 35, pp. 4053-4062.