

## 패턴분류 방법 적용에 의한 장성호 수문·수질자료의 특성파악

박성천·진영훈<sup>†</sup>·노경범\*·김종오\*\*·유호규\*\*\*

동신대학교 토목공학과  
\*목포대학교 창업지원단  
\*\*목포대학교 환경교육과  
\*\*\*전라남도청

## Characteristics Detection of Hydrological and Water Quality Data in Jangseong Reservoir by Application of Pattern Classification Method

Sung-Chun Park · Young-Hoon Jin<sup>†</sup> · Kyong-Bum Roh<sup>\*</sup> · Jongho Kim<sup>\*\*</sup> · Ho-Gyu Yu<sup>\*\*\*</sup>

*Department of Civil Engineering, Dongshin University*

*\*Startup Assistance Foundation, Mokpo National University*

*\*\*Department of Environmental Education, Mokpo National University*

*\*\*\*Jeollanamdo Provincial Office*

*(Received 15 July 2011, Revised 22 September 2011, Accepted 23 September 2011)*

### Abstract

Self Organizing Map (SOM) was applied for pattern classification of hydrological and water quality data measured at Jangseong Reservoir on a monthly basis. The primary objective of the present study is to understand better data characteristics and relationship between the data. For the purpose, two SOMs were configured by a methodologically systematic approach with appropriate methods for data transformation, determination of map size and side lengths of the map. The SOMs constructed at the respective measurement stations for water quality data (JSD1 and JSD2) commonly classified the respective datasets into five clusters by Davies-Bouldin Index (DBI). The trained SOMs were fine-tuned by Ward's method of a hierarchical cluster analysis. On the one hand, the patterns with high values of standardized reference vectors for hydrological variables revealed the high possibility of eutrophication by TN or TP in the reservoir, in general. On the other hand, the clusters with low values of standardized reference vectors for hydrological variables showed the patterns with high COD concentration. In particular, Cluster1 at JSD1 and Cluster5 at JSD2 represented the worst condition of water quality with high reference vectors for rainfall and storage in the reservoir. Consequently, SOM is applicable to identify the patterns of potential eutrophication in reservoirs according to the better understanding of data characteristics and their relationship.

**keywords** : Data characteristics, Hydrological data, Pattern classification, Reservoir, Self Organizing Map (SOM), Water quality

## 1. 서론

최근 홍수 및 가뭄에 대비하고 갈수기에 하천유지 유량을 확보하기 위하여 기존 저수지 등의 증고 사업이 진행되고 있다(농림수산식품부, 한국농어촌공사, 2010). 이와 같은 사업의 진행은 향후 저수지와 같은 정체수역의 수자원을 보다 적극적으로 활용하겠다는 의지가 포함되어 있는 것으로 판단된다. 상술한 저수지 등 높이기사업의 목적들 중 하나인 하천유지 유량의 확보를 고려하여 보면, 향후 저수지의 수질이 하천의 수질에 지금까지 미친 영향보다 더 큰 영향을 미치게 된다. 즉, 주로 농업용수로 사용된 저수지 수체의 수질은 하천수질에 영향을 미치지 않거나 토양을

통과하여 하천으로 유입되는 간접적인 영향을 미쳐왔으나, 저수지의 저수량이 하천유지 용수로 활용될 경우에는 저수지의 방류량이 하천으로 유입되기 때문에 저수지와 하천의 수질패턴이 직접적인 연관성을 지니게 된다.

따라서 저수지의 수질에 영향을 미치는 강우량 및 저수량과 같은 수문학적 변수와 수질현황에 대한 보다 심도 있는 분석이 필요한 실정이다. 특히 저수지의 수질개선 및 생태계의 건강성 확보를 위해서는 수문변수 및 각종 수질자료들이 나타내는 복합적 패턴에 대한 보다 깊은 이해와 이를 바탕으로 하여 부영양화 현상에 대한 보다 다양한 연구가 필요하다.

이와 관련된 국내연구 현황을 살펴보면, 김범철 등(2007)이 하수처리장 방류수의 인 농도 기준 강화의 필요성 제시를 위해 호수의 부영양화 제한영양소에 대하여 고찰한 바 있다. 이현준과 안광국(2009)은 국내 인공호의 영양상태를

<sup>†</sup> To whom correspondence should be addressed.  
nmdrjin@gmail.com

평가하기 위한 호소 평가 기법을 개발하여 연구대상 호소에 적용하여 모형의 평가와 검증을 수행하였으며, 박유미 등(2009)은 대상 저수지에서 강우에 따른 수질특성 변이 및 경향성과 부영양화 특성을 파악하였다.

호소수 수질의 평가 및 분석에 관한 외국의 사례를 살펴보면, 다변량 통계기법을 이용하여 오염된 호소에 대하여 수질 측정지점들 간의 유사성을 평가하고 수질변수에 영향을 미치는 오염원을 파악한 연구(Kazi et al., 2009)와 유량과 비점 오염원이 호수의 화학적 특성에 미치는 영향에 관한 연구(Ikem and Adisa, 2011)가 있다. 다양한 수질항목을 고려하기 위해 상술한 다변량 통계분석 기법의 적용과 함께 최근에는 자기조직화지도(Self Organizing Map: SOM)를 이용하여 수질 및 수처리 분야에서 다양한 수질자료에 대한 패턴분류와 분석을 수행하고 있다.

김용구 등(2008)은 주암호를 대상으로 수문변수와 TOC (Total Organic Carbon)에 대한 변동 특성파악을 위해 SOM을 적용한 바 있으며, 진영훈 등(2009)은 영산강·섬진강·탐진강 수계의 수질오염총량관리제를 위한 단위유역 말단지점에서 측정된 수질 및 유량자료의 통계량에 대한 SOM 적용 결과를 제시하였다. 또한, 농업용수 수질개선사업을 위한 저수지의 개략적인 수질변화 예측을 목적으로 SOM을 적용하여 저수지를 물리화학적 및 수질특성에 따라 분류한 후, 각 유형에 적용 가능한 모형을 개발한 연구가 수행된 바 있다(한국농어촌공사, 2010).

수질자료의 분석을 위한 외국의 SOM 적용 사례로는 해안의 수질평가 및 예측을 위해 SOM을 적용하였으며(Aguilera et al., 2001), Alvarez-Gurerra 등(2008)은 수질의 분류 및 분석을 위한 SOM의 타당성을 검토한 바 있다. 또한, 하천의 수질에 영향을 미칠 수 있는 인자들에 대한 평가를 위하여 SOM을 적용한 연구(Tobiszewski et al., 2010) 및 온라인 수질자료에 대한 이상현상 및 변동을 파악하는데 SOM을 활용한 사례(Mustonen et al., 2008)와 수질자료의 시·공간 분포를 파악하기 위한 SOM의 적용이 있다(Su et al., 2010).

이와 같이 최근 수질자료의 평가 및 분석, 예측을 위해 SOM의 적용이 활발히 이루어지고 있으며, 많은 연구결과에서 그 적용 타당성을 제시하고 있다. 그러나 많은 연구들이 대상 자료에 대한 통계적 특성을 고려하지 않은 자료의 전처리로 인해 부적합한 SOM의 적용을 수행하였으며, 수질자료에 대한 SOM의 패턴분류에 관한 연구들은 수량을 배제한 결과들을 제시함으로써 수량과 수질에 대한 관계를 고려하지 않았다. 따라서 본 연구는 수량과 수질을 동시에 포함하는 다변량 자료가 나타내는 복합적 패턴을 SOM의 적용에 의해 분류하고자 하였다. 다양한 패턴이 혼재되어 있는 측정자료에 대한 전체적인 특성을 파악하기 어려우므로, 측정 자료에 대한 패턴분류를 통해 부분별 특성을 파악하기 위해 SOM을 적용하였다. 이를 위해 적절한 자료의 전처리 방법 및 SOM의 크기와 종·횡축에 대한 노드의 수 결정 방법과 같은 체계적인 방법론을 적용하였다. 그 결과를 분석하여 저수지의 수문·수질자료의 특성 및 상호간의 연관성을 파악하였다.

## 2. 연구방법

### 2.1. 자기조직화 방법(Self Organizing Map: SOM)

SOM은 다차원의 정보를 2차원의 사각형 또는 육각형 배열을 사용하는 저차원의 공간으로 사상시킬 수 있는 인공 신경회로망 이론의 한 종류이며, 일반적으로 육각형 배열을 사용하여 패턴분류 결과를 쉽게 이해할 수 있게 가시화할 수 있는 능력을 가지고 있다. 그러나 SOM을 적용하는 궁극적인 목적은 대상자료에 대한 SOM의 적용에 산출되는 참조벡터(reference vector)를 얻는 데 있다. 이러한 참조벡터는 가중치(weight), 연결가중치(connection weight), 일반화된 중앙값(generalized median) 또는 코드북(codebook)으로 알려져 있다(Bedoya et al., 2009; Hsu et al., 2002; Kalteh et al., 2008; Lu et al., 2006). 참조벡터를 얻기 위해서는 노드들 사이의 경쟁과정, 승자노드의 선택 및 참조벡터의 갱신과 같은 일련의 훈련과정을 반복적으로 수행하여야 한다.

만약 입력자료가  $n$ 차원의 변수로 구성되어 있고, SOM에 의해  $k$ 개의 패턴으로 구분된다고 가정하면, 해당 SOM은  $n$ 개의 입력노드와  $k$ 개의 출력노드를 가지게 된다. 각 입력노드들은 모든 출력노드들과 연결되며, 이렇게 연결된 값들을 참조벡터라 한다. 각 노드들은 경쟁과정을 통해 학습할 수 있는 기회를 부여받기 위해 서로 경쟁하며, 입력자료와 거리가 가장 가까운 참조벡터를 갖는 노드가 승리하게 된다. 이러한 승자노드의 선택은 유사정도(similarity matching) 측정을 위해 유클리드 거리(Euclidean distance)를 이용한다. 참조벡터의 갱신과정에서는 선정된 승자노드와 가우시안 함수(Gaussian function)에 의해 결정되는 이웃 노드들만이 제시된 입력자료에 대하여 학습이 허용된다(진영훈 등, 2009).

상기의 SOM의 훈련을 위해서는 입력자료에 대한 적절한 자료변환이 필요하다. 이는 본 연구의 대상자료와 같이 분산이 크거나 왜곡이 심한 원자료를 그대로 이용할 경우, 큰 값들에 의해 패턴분류 과정 전체가 영향을 받아 작은 값들에 대한 부적절한 분류 결과를 도출할 수 있기 때문이다. 따라서 원자료의 분포를 파악한 후 그에 상응하는 적절한 변환방법을 이용하여 처리된 자료를 SOM에 적용한다.

자료의 전처리 후 SOM의 반복학습을 위해 먼저 그 구조를 결정하며, 이를 위해 총 노드의 수 및 종·횡축 방향의 노드 배열의 수를 결정한다. 총 노드의 수는 일반적으로  $m = 5\sqrt{n}$  ( $m$ : 총 노드의 수;  $n$ : 입력자료의 수)이라는 경험식을 사용하며, 산정된 수( $m$ )를 종·횡축으로 분배하기 위해 입력자료에 대한 가장 큰 두 개의 고유치(eigenvalue)의 비를 구하여 이용한다(Alvarez-Guerra et al., 2008; Garcia and González, 2004; Hentati et al., 2010; Vesanto et al., 2000).

구조가 결정된 SOM은 참조벡터의 초기화가 필요하며, 이를 위해서는 일반적으로 난수 초기화와 선형 초기화 방법이 있다. 최근에는 대규모의 자료를 요구하는 난수 초기화 방법보다는 훈련 속도를 개선시킬 수 있는 선형 초기화 방법이 널리 이용되고 있다(Jeong et al., 2010). 이러한 선

형 초기화 방법은 입력자료의 고유치(eigenvalue)와 고유벡터(eigenvector)를 이용하여 진행되므로 이는 SOM의 초기화 과정에 이미 입력자료의 정보를 제공함으로써 훈련을 보다 효율적으로 수행할 수 있도록 도와준다(Vesanto et al., 2000).

이상의 전처리과정에 의해 변환된 자료를 SOM에 입력하여 훈련을 통해 패턴분류를 수행한다. 최적의 패턴분류의 수를 결정하기 위해 패턴분류의 타당성을 나타내는 Davies-Bouldin Index (DBI)를 Garcia and González (2004)에 의해 제안된 다음의 식 (1)-(3)을 이용하여 산정한다(진영훈 등, 2009).

$$DBI = \frac{1}{K} \sum_{j=1}^K R_j, \quad R_j = \max_k \left\{ \frac{(S_j + S_k)}{D_{jk}} \right\} \quad (j \neq k) \quad (1)$$

$$S_j = \frac{1}{N_j} \sum_{m_i \in Q_j} tm_i - M_j t \quad (j = 1, \dots, K) \quad (2)$$

$$D_{jk} = tM_j - M_k t \quad (3)$$

여기서  $M_j$ 는 클러스터-j의  $N_j$ 개의 노드로부터 계산된 평균 참조벡터이며,  $S_j$ 는 클러스터-j로 구분된 노드들 사이의 유사성(similarity)을 나타낸다. 또한  $D_{jk}$ 는 두 클러스터( $j, k$ ) 사이의 이질성(dissimilarity)을 나타낸다. 여기서  $k$ 는 클러스터-j와 서로 다른 클러스터를 나타내는 첨자이다.

이러한 과정에 의해 생성된 SOM의 참조벡터를 계층적 클러스터 분석(hierarchical cluster analysis)의 일종인 Ward 기법에 적용하여 보다 타당한 패턴분류 결과를 유도한다(Faggiano et al., 2010; Hentati et al., 2010; Park et al., 2007; Song et al., 2007). Ward 기법은 계층적 응집형 클러스터 분석방법으로 각 자료들 사이의 거리를 계산하여 그 거리에 따라 가까운 자료들로 구성되는 새로운 클러스터를 형성한다. 이러한 과정이 반복적으로 적용되어 최종적으로는 하나의 클러스터를 나타내게 된다(Hentati et al., 2010).

따라서 Ward 기법의 적용 초기 단계에서는 SOM의 각 노드들에 대한 참조벡터들이 개별적인 클러스터를 형성하며, 각 노드의 참조벡터를 활용하여 거리를 산정한 후 가까운 거리의 노드들이 보다 큰 규모의 클러스터를 형성한다. 이러한 과정의 반복 적용에 의해 대상 자료를 적정한 수의 클러스터로 분류할 수 있으며, 각 클러스터는 대상 자료에 대한 패턴으로 간주된다. Ward 기법의 적용에 따른 결과는 dendrogram이라고 하는 계층적 클러스터 트리(hierarchical cluster tree)에 의해 각 클러스터의 형성과정을 확인할 수 있으며, 적정한 수의 패턴 분류를 위한 가시적인 기준으로 이용될 수 있다(송자섭, 2011).

## 2.2. 대상지점

본 연구는 저수지의 수문·수질자료의 특성을 파악하기 위한 대상지점으로 전라남도 장성군에 위치한 장성호를 선

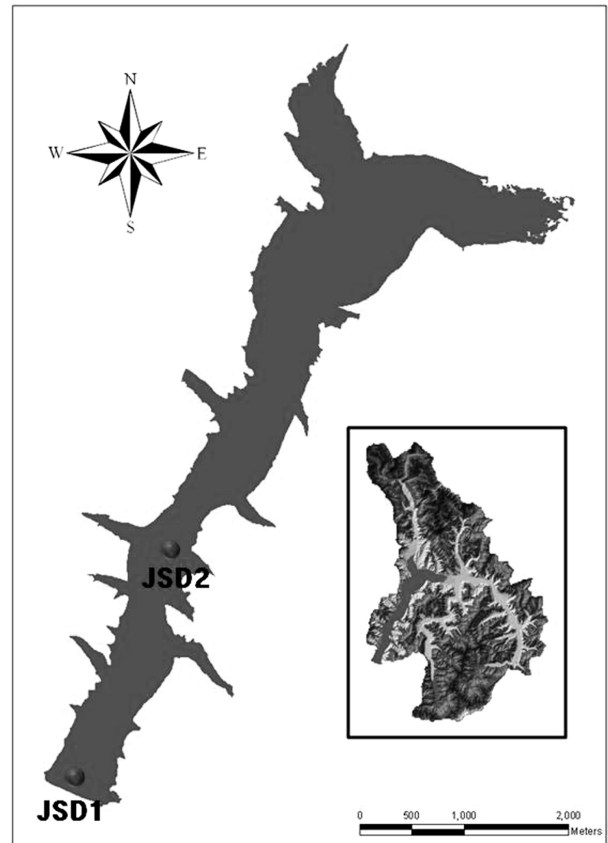


Fig. 1. Map showing study area and measurement stations in Jangseong Reservoir.

정하였다. 장성호는 전라남도의 지방하천 중 하나인 황룡강의 상류에 위치해 있으며, Fig. 1에서 보는 바와 같이 저수지 내의 2개소에서 수질을 측정하고 있다. JSD1(장성댐1) 지점은 전라남도 장성군 장성읍 봉덕리에 위치하고 있으며, JSD2(장성댐2)는 전라남도 장성군 장성읍 용강리에 위치해 있다. 두 지점의 수질은 영산강유역환경청에서 측정을 진행하고 있으며, 측정자료는 환경부 산하 국립환경과학원의 물환경정보시스템(<http://water.nier.go.kr>)을 통해 수집하였다.

## 2.3. 기초통계분석

다양한 수질자료들 중 저수지의 수질특성을 잘 반영하는 것으로 알려진 수질항목을 선정하여 장성댐 관측소에서 측정된 강우량 및 저수량 자료와 같은 수문자료와 함께 수집·정리하였다. 대상 수질항목으로는 DO (Dissolved Oxygen), COD (Chemical Oxygen Demand), TN (Total Nitrogen) 및 TP (Total Phosphorus) 자료를 선정하였으며, 최근 자료의 특성을 보다 심도 있게 연구하기 위해 2006년부터 2010년까지의 최근 5년간의 월자료를 이용하였다. Table 1은 장성호 내 2개의 수질측정지점에서 측정된 각 수질자료와 강우량 및 저수량에 대한 기초 통계량을 나타내고 있다.

강우량 자료의 경우 표준편차가 가장 큰 것으로 나타나 그 분산의 정도가 큰 것으로 나타났다. 수질자료들 중 TP의 경우 가장 큰 왜도 및 첨도를 나타내어 비대칭성이 강하고 자료들이 일정 범위에서 집중적으로 측정되었음을 알

**Table 1.** Basic statistics of hydrological and water quality data at two stations within Jangseong Reservoir

Basic statistics	No. Data: 60		Station: JSD1				Station: JSD2			
	Rainfall (mm)	Storage (Mm <sup>3</sup> )	DO (mg/L)	COD (mg/L)	TN (mg/L)	TP (mg/L)	DO (mg/L)	COD (mg/L)	TN (mg/L)	TP (mg/L)
Mean	111.43	69.68	8.21	3.43	1.00	0.03	9.13	3.60	1.01	0.03
Median	57.50	72.03	7.75	3.60	0.97	0.02	8.65	3.70	1.04	0.02
Standard deviation	129.39	14.76	3.01	0.93	0.33	0.03	2.82	1.06	0.33	0.03
Skewness	1.86	-0.49	0.25	-0.36	1.01	3.29	0.46	-0.37	0.60	3.90
Kurtosis	5.65	2.27	1.81	2.53	4.62	14.60	2.64	2.55	3.79	21.23

**Table 2.** Correlation coefficients between stations and variables at two stations within Jangseong Reservoir marked by \* for 1% and \*\* for 5% significance level

Variables	Between stations	No. Data: 60		Station: JSD1				Station: JSD2			
		Rainfall	Storage	DO	COD	TN	TP	DO	COD	TN	TP
Rainfall	1.000	1.000	0.135	-0.251	-0.068	0.291**	-0.064	-0.313**	0.015	0.289**	-0.051
Storage	1.000		1.000	0.133	-0.233	0.401*	-0.121	-0.019	-0.406*	0.416*	-0.168
DO	0.900*			1.000	-0.051	-0.139	-0.021	1.000	-0.116	-0.378*	-0.020
COD	0.783*				1.000	-0.197	0.012		1.000	-0.453*	0.099
TN	0.812*					1.000	0.218			1.000	-0.058
TP	0.192						1.000				1.000

수 있다. 이와 같이 분산이 크며 왜곡이 심한 자료들은 패턴분류를 위한 SOM을 적용할 때 적합한 자료의 전처리 과정이 필요하다. 자료의 전처리 과정은 소수의 극한값들에 의해 패턴이 부적절하게 분류될 수 있으므로 SOM의 적용에 앞서 반드시 수행되어야 한다.

Table 2는 두 측정지점 사이의 변수별 상관관계와 각 지점에서 변수들 사이의 상관관계를 나타내고 있다. 먼저 측정지점 사이의 수질자료에 대한 상관관계를 살펴보면, DO, COD 및 TN은 1% 유의수준에서 매우 높은 상관성을 보이고 있으나, TP 항목에 대한 상관성은 통계학적으로 유의하지 못한 수준의 매우 낮은 상관성을 나타내고 있다. 따라서 본 연구에서는 대상 자료의 이질성으로 인하여 SOM의 적용을 각 지점별로 수행하였다.

다음으로 각 지점별 변수들 사이의 1% 유의수준을 갖는 상관관계를 살펴보면, JSD1에서 저수량과 TN 사이에 0.401의 상관성을 보였으며, JSD2에서도 저수량과 TN 사이에 0.416의 상관성을 보여 두 지점 모두에서 저수량의 증가와 함께 TN이 증가하는 유사한 양상의 상관관계를 나타내고 있다. 또한, JSD2에서는 저수량과 COD, DO와 TN, COD와 TN이 각각 -0.406, -0.378, -0.453의 상관계수를 나타내어 역상관을 보이고 있다. 이와 같은 동일변수에 대한 JSD1 지점의 상관관계는 통계학적으로는 유의하지 못한 수준이나 JSD2 지점과 마찬가지로 역상관을 보이고 있다.

또한, 각 지점별 변수들 사이의 5% 유의수준을 갖는 상관관계를 살펴보면, 강우량과 JSD1 지점의 TN, JSD2 지점의 DO 및 TN이 각각 0.291, -0.313 및 0.289의 상관성을 보이고 있다. 강우량과 TN 사이의 상관관계는 두 지점 모두에서 통계학적으로 유의한 수준에서 매우 유사하게 나타났으며, 강우량의 증가에 따른 DO의 감소를 나타내는 양상 역시 두 지점에서 공통적으로 나타났다.

이상의 상관관계 분석을 통해 수문학적 변수인 강우량과 저수량의 증가에 따라 TN은 증가하며, 강우량의 증가에 따른 DO의 감소 및 저수량의 증가에 따른 COD의 감소 양상을 파악하였다. 또한, DO 및 COD의 증가는 공통적으로 TN의 감소를 나타내는 매우 유의한 수준의 상관관계를 보였다. TN을 중심으로 다른 변수들과의 상관관계를 정리하여 보면, TN은 수문학적 변수들과는 정상관의 양상을 나타내며, DO 및 COD와 같은 수질항목과는 역상관의 상관성을 보이고 있다. TP의 경우 두 지점 모두에서 다른 변수들과 유의한 수준의 상관관계를 보이지 않았다.

따라서 본 연구는 상술한 바와 같이 변수들 사이의 상관관계의 방향성 및 그 정도가 다양하여 측정 자료에 의해 대표되는 수문·수질자료의 특성이 전체적으로 파악되기 곤란하므로, 자료 내에 혼재되어 있는 패턴을 분류하여 각각의 특성을 파악하고자 하였다.

### 3. 결과 및 고찰

#### 3.1. SOM 적용

본 연구는 SOM의 적용을 위해 체계적인 방법을 이용하였으며, 이를 전처리 과정, SOM 적용 과정, 계층적 클러스터 분석 과정 및 후처리 과정의 4단계로 구분하여 Fig. 2에 흐름도(flow chart)로 나타내었다. Fig. 2에 제시된 바와 같이 SOM의 구조 결정을 위해 선정되는 고유치와 고유벡터는 입력자료에 따라 다르게 되므로, 결과적으로는 원자료에 대한 변환방법에 따라 SOM의 결과가 민감하게 반응하게 된다. 따라서 원자료에 대한 적절한 변환방법을 선정하는 것이 체계적인 SOM의 적용을 위해 필수적인 요소가 된다. 본 연구는 두 측정지점에 대한 자료의 전처리를 위해 다른 자료들과 비교하여 큰 양의 왜곡도를 나타내는 강

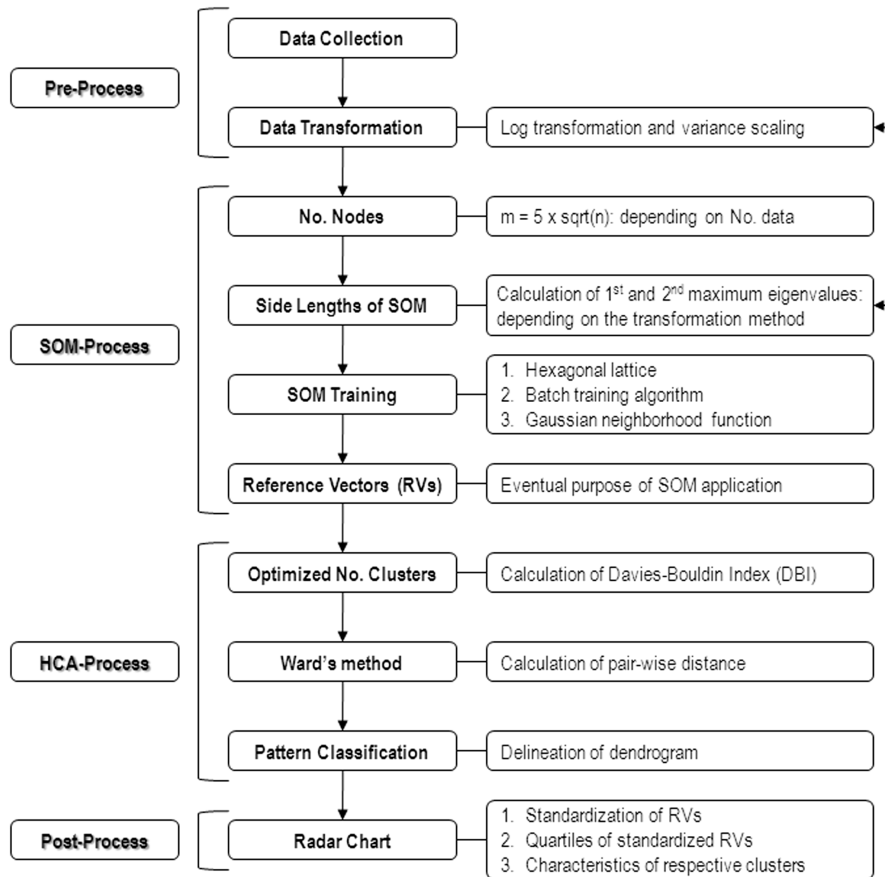


Fig. 2. Flow chart for methodologically systematic application of SOM.

우량과 TP 자료에 로그변환(log transformation)을 적용한 후 모든 자료에 공통적으로 평균과 표준편차를 이용하여 자료의 표준화를 수행하였다.

SOM의 구조는 일반적으로 사용되는 육각형 배열을 사용하였으며, 훈련방법으로는 입력자료를 일괄적으로 처리하는 방식(batch mode)을 적용하였다. 구축된 SOM에 의한 최적의 패턴분류의 수를 결정하기 위해 DBI를 산정하였으며, Ward 기법을 적용하여 보다 타당한 패턴분류 결과를 유도하였다. 각 클러스터로 분류된 노드들의 참조벡터에 대한 1~3사분위 값을 산정하여 패턴별 특성을 알아보기 위한 가시화 방법으로 레이더 차트(radar chart)를 활용하였다.

먼저 각 지점별 자료를 전처리과정에 의해 변환한 후 SOM에 입력하여 패턴분류를 수행하였으며, SOM의 구조를 살펴보면, JSD1에 대하여 총 노드의 수가 42개로 결정되었으며 종·횡축에 대한 노드의 수는 7×6의 배열을 갖는 것으로 나타났다. 반면, JSD2 지점에 대해서는 40개의 노드를 포함하는 8×5의 배열을 갖는 SOM 구조가 결정되었다. 이와 같이 결정된 SOM 구조를 통해 패턴분류를 수행한 결과를 각 지점별로 아래와 같이 기술하였다.

3.2. JSD1(장성댐1) 지점의 패턴분류 결과

JSD1 지점에 대한 변환자료를 이용하여 DBI를 산정한 결과 Fig. 3(a)에 도시된 바와 같이 5개의 패턴으로 분류할 때 가장 최적의 결과를 나타내었다. 이에 따른 패턴분류

결과에 의해 도출된 참조벡터를 상술한 바와 계층적 클러스터 분석방법인 Ward 기법에 적용하여 보다 타당한 패턴분류 결과를 유도하였다(Fig. 3(b)). 최종 패턴분류 결과를 Fig. 3(c)에 나타내었으며, 각 노드 내에 표기된 숫자는 각 노드로 분류된 자료의 수를 의미한다.

패턴분류가 완료된 SOM의 참조벡터를 각 변수별로 [0, 1] 사이로 표준화하여 각 클러스터별 패턴을 비교하였다. 이를 Fig. 4에 레이더 차트를 이용하여 나타내었으며, 각 차트에 1사분위값, 중앙값(2사분위값) 및 3사분위값을 도시하였으며, 이들 중 각 변수에 해당하는 중앙값을 Table 3에 제시하였다. 수질자료에 대한 표준화된 참조벡터는 DO가 '0'에 가까울수록, COD, TN 및 TP가 '1'에 가까울수록 그 오염의 정도가 심각한 상태를 나타낸다. Table 4는 각 클러스터로 분류된 자료의 수와 분류자료의 원자료에 대한 평균값을 각 변수별로 제시하고 있다.

Table 3. Median values of standardized reference vectors for the respective variables at JSD1 station

Median	Rainfall	Storage	DO	COD	TN	TP
Cluster1	0.61	0.66	0.12	0.49	0.71	0.46
Cluster2	0.16	0.72	0.31	0.50	0.35	0.09
Cluster3	0.39	0.94	0.48	0.15	0.54	0.07
Cluster4	0.26	0.21	0.10	0.78	0.23	0.19
Cluster5	0.12	0.30	0.89	0.71	0.09	0.15

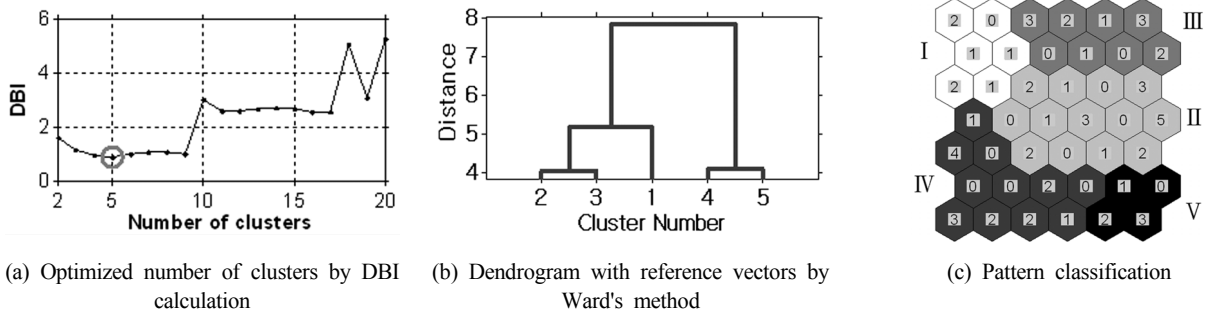


Fig. 3. Determination of optimized number of clusters by DBI calculation, dendrogram by application of Ward's method and pattern classification result at JSD1 station.

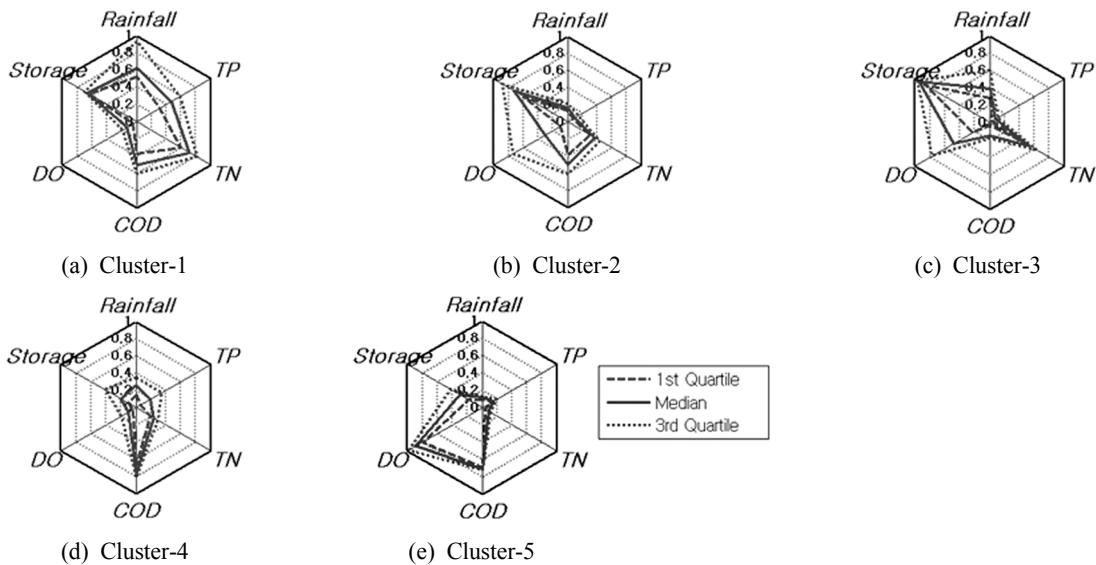


Fig. 4. Radar chart using standardized reference vectors for each cluster at JSD1 station.

Table 4. Number of data and mean values of raw data for the respective variables classified into each cluster at JSD1 station

Mean values	No. Data	Rainfall (mm)	Storage (Mm <sup>3</sup> )	DO (mg/L)	COD (mg/L)	TN (mg/L)	TP (mg/L)
Cluster1	7	198.00	71.63	6.74	3.63	1.46	0.06
Cluster2	20	63.50	76.73	8.87	3.49	0.98	0.02
Cluster3	12	184.50	82.60	8.71	2.21	1.17	0.02
Cluster4	15	104.40	55.40	6.10	4.08	0.84	0.03
Cluster5	6	41.67	53.77	11.98	3.83	0.55	0.03

먼저 Cluster1을 보면, 7개의 자료가 분류되었으며, 다른 패턴들과 비교하여 수문학적 변수인 강우량과 저수량이 큰 양상을 나타내고 있다(Fig. 4(a)). Fig. 3(b)의 계통도(dendrogram)에 도시된 바와 같이 Cluster1은 다른 Cluster들에 비해 Cluster2와 3에 보다 근접한 양상을 보이고 있으나 전반적으로 다른 패턴들로부터 독립적으로 분류되어 있다. Table 3에 나타난 바와 같이 Cluster1로 분류된 노드들의 표준화 참조벡터의 중앙값을 패턴들과 비교할 때, 상당히 큰 강우량을 나타내는 패턴을 포함하고 있기 때문에 판단된다. 수질변수인 TN 및 TP 역시 상당히 높은 양상을 보이고 있는 반면 DO는 낮은 패턴을 나타내었다. Table 4에 제시된 Cluster1로 분류된 강우량, TN 및 TP의 원자료

에 대한 평균값들이 전체 평균(Table 1)과 비교하여 상당히 높은 값들을 보이고 있다. 결과적으로, Cluster1로 분류된 자료들은 수문학적 변수의 값이 상대적으로 큰 것으로 나타났으며, 수질항목들을 고려할 때 다른 패턴들과 비교하여 전반적으로 그 오염의 정도가 심한 것으로 나타났다. 이는 이현준과 안광국(2009)에 의해 기술된 바와 같이 수체 내 수질 특성이 강우 특성에 큰 영향을 받는다는 내용에 부합된다.

Cluster2와 3은 다른 패턴들에 비해 서로 유사한 양상을 나타내고 있다(Fig. 3(b)). Cluster2의 경우(Fig. 4(b)), 20개의 자료가 분류되었으며, 강우량은 작은 양상을 보이고 있으나 저수량과 DO는 비교적 큰 패턴을 나타내고 있다. 또

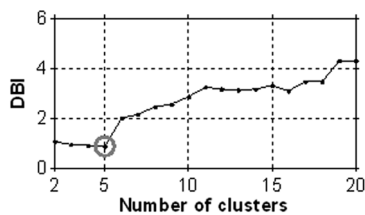
한, COD, TN 및 TP는 낮은 값을 갖는 패턴을 나타내었다. 특히 DO에 대한 1~3사분위값의 분포는 분류된 자료들이 크게 분산되어 있음을 보여주고 있다. 이는 Table 3에 제시된 바와 같이 분류된 노드들의 DO에 대한 표준화 참조벡터의 중앙값이 다른 패턴들에 비해 현저히 높은 값을 나타내지 않고 있어 Cluster2로 분류된 DO가 크게 분산되어 있음을 재확인하고 있다. 따라서 Cluster2는 강우량이 적은 시기에 저수량이 큰 값을 나타내고 있으며, 수질항목들은 높은 DO값과 전반적으로 낮은 값의 오염물질을 포함하고 있는 것으로 나타났다.

Cluster3에는 12개의 자료가 분류되었으며, 다른 패턴들과 비교하여 큰 강우량과 가장 큰 패턴의 저수량 자료를 포함하고 있다. 또한, 상대적으로 높은 값의 DO와 TN, 가장 낮은 값의 COD와 TP를 갖는 패턴으로 분류되었다(Fig. 3(c)). 따라서 Cluster3은 다른 Cluster들과 비교하여 TN을 제외한 수질항목을 고려할 때 양호한 상태의 수질을 나타내고 있다.

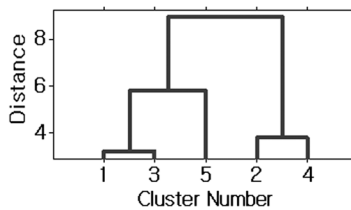
Fig. 4(d)와 (e) 및 Table 4에 제시된 바와 같이 Cluster4와 5는 전체 자료에 대한 평균 이하인 값의 수문학적 변수와 TN 및 TP, 그리고 높은 값의 COD를 포함하는 유사한 패턴을 보이고 있다. 15개의 자료가 분류된 Cluster4에서는 노드들의 표준화 참조벡터의 중앙값(Table 3) 및 원자료에 대한 평균값(Table 4)에 대한 DO의 값이 낮은 반면 6개의 자료가 분류된 Cluster5는 DO의 값이 다른 패턴들과 비교하여 가장 높은 양상을 포함하고 있다. 이에 따라 Cluster4와 Cluster5로 분류된 자료들의 발생 시기를 조사한 결과, Cluster4는 주로 여름철에 측정된 자료가 많이 포함되었으며, Cluster5에는 겨울철에 측정된 자료가 분류된 것으로 나타났다. 즉, 계절에 따른 온도의 영향이 두 패턴을 나누는 분류요인으로 작용한 것으로 판단된다. 결과적으로, Cluster4와 5는 수문학적 변수가 작은 값을 나타낼 때 COD를 제외한 수질항목에서 상당히 양호한 상태의 수질을 보이고 있다.

3.3. JSD2(장성댐2) 지점의 패턴분류 결과

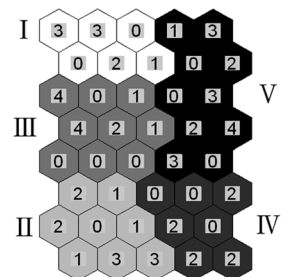
Fig. 5(a) 도시된 바와 같이 JSD2 지점에 대한 SOM 적용 결과, JSD1 지점과 마찬가지로 5개의 패턴으로 분류되었다.



(a) DBI variation



(b) Dendrogram with reference vectors by Ward's method



(c) Pattern classification

Ward 기법의 적용에 의한 계통도(Fig. 5(b))를 보면, Cluster1과 3이 서로 유사한 패턴을 포함하고 있으며, Cluster2와 4가 또 다른 유사 패턴을 보이고 있음을 알 수 있다. Cluster5는 Cluster1과 3에 보다 근접한 양상을 보이거나 전반적으로 다른 패턴들로부터 독립적으로 분류되어 있다.

먼저 Table 6에 제시된 바와 같이 9개의 자료가 분류된 Cluster1은 큰 양상의 수문학적 변수와 가장 작은 패턴의 COD 및 TP를 포함하고 있으며, DO와 TN은 다소 큰 양상을 보이고 있다(Fig. 6(a)). 특히, Cluster1로 분류된 노드들의 표준화 참조벡터의 중앙값을 보면(Table 5), 가장 높은 값의 저수량과 TN 및 가장 낮은 값의 COD와 TP를 포함하고 있다. 따라서 Cluster1은 다른 Cluster들과 비교하여 TN을 제외한 수질항목을 고려할 때 비교적 양호한 상태의 수질을 나타내고 있으며, JSD1 지점의 Cluster3과 유사한 양상을 보여주고 있다.

Cluster3의 경우(Fig. 6(c)), 12개의 자료가 분류되었으며, 대체적으로 Cluster1과 유사한 양상을 보이고 있으나, 강우량은 Cluster1과 비교하여 더 작은 패턴을 나타내고 있는 반면 COD는 더 큰 양상을 포함하고 있으며, 다른 패턴들과 비교하여 큰 값의 DO 자료들을 포함하고 있다. 이는 Table 5에 제시한 표준화 참조벡터의 중앙값과 Table 6의 원자료에 대한 평균값으로부터 수치적으로 확인할 수 있다. 따라서 Cluster3의 패턴은 강우량이 적은 시기에 저수량이 크며, 수질항목을 고려할 때 전반적으로 양호한 상태의 수질을 나타내고 있다.

Cluster2와 Cluster4는 다른 패턴들과 비교하여 수문학적 변수의 값이 작은 것으로 나타났다(Table 5). Fig. 6(b)에 도시된 바와 같이 13개의 자료가 분류된 Cluster2는 가장

Table 5. Number of data and mean values of raw data for the respective variables classified into each cluster at JSD1 station

Median	Rainfall	Storage	DO	COD	TN	TP
Cluster1	0.54	0.95	0.59	0.15	0.63	0.18
Cluster2	0.08	0.37	0.74	0.81	0.07	0.51
Cluster3	0.28	0.87	0.76	0.50	0.48	0.25
Cluster4	0.29	0.25	0.31	0.93	0.23	0.64
Cluster5	0.72	0.66	0.13	0.65	0.62	0.48

Fig. 5. Determination of optimized number of clusters by DBI calculation, dendrogram by application of Ward's method and pattern classification result at JSD2 station.

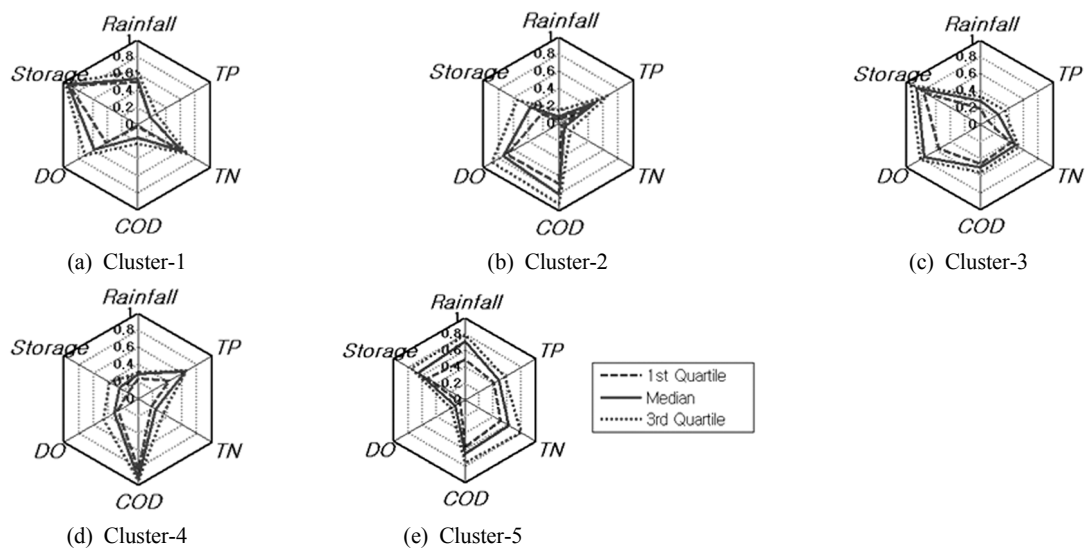


Fig. 6. Radar chart using standardized reference vectors for each cluster at JSD2 station.

Table 6. Number of data and mean values of raw data for the respective variables classified into each cluster at JSD2 station

Mean Values	No. Data	Rainfall (mm)	Storage (Mm <sup>3</sup> )	DO (mg/L)	COD (mg/L)	TN (mg/L)	TP (mg/L)
Cluster1	9	125.11	82.25	9.67	1.93	1.21	0.02
Cluster2	13	30.38	60.75	11.28	4.12	0.64	0.03
Cluster3	12	73.92	79.92	11.58	3.33	1.06	0.02
Cluster4	8	112.88	55.18	7.19	4.58	0.80	0.05
Cluster5	18	187.50	69.46	6.54	3.80	1.24	0.03

작은 양상의 강우량 및 TN, 상당히 높은 양상의 DO, COD 및 TP를 포함하고 있다. Cluster4에는 8개의 자료가 분류되었으며, 가장 작은 패턴의 저수량과 상대적으로 낮은 패턴의 DO 및 TN자료가 포함된 것으로 나타났다(Fig. 6(d)). 또한, Table 5에 제시된 표준화 참조벡터의 중앙값과 Table 6에 제시된 원자료의 평균값을 보면 가장 높은 COD 및 TP의 평균을 나타내고 있다. 결과적으로, Cluster2와 4는 COD 및 TN이 상대적으로 높은 값을 나타내는 패턴으로 분류되어 다른 패턴들과 비교하여 다소 악화된 상태의 수질을 보이고 있다. 두 패턴의 차이는 DO항목에서 명확하게 드러나며, Cluster2는 주로 겨울에 측정된 자료를 포함하고 있으며, Cluster4는 주로 여름에 측정된 자료를 포함하고 있는 측정시기의 계절적 차이에 의해 두 패턴이 분류된 것으로 판단된다.

마지막으로, 18개의 자료가 분류된 Cluster5는 수문학적 변수가 다른 패턴들에 비해 큰 것으로 나타났으며(Fig. 6(e)), 특히, 강우량의 경우 Table 5(표준화 참조벡터의 중앙값)와 Table 6(원자료에 대한 평균값)에 제시된 바와 같이 가장 큰 강우량 평균을 포함하고 있다. 반면에 DO는 가장 낮으나, TN과 TP는 현저히 높은 것으로 나타났으며, COD 역시 상대적으로 높은 값을 나타내고 있다. 따라서 JSD2 지점의 Cluster5의 경우, JSD1 지점의 Cluster1과 같이 다른 패턴들과 비교하여 높은 값의 수문학적 변수를 갖는 패턴이며, 수질항목들을 고려할 때 전반적으로 그 오염의 정도가 심한 것으로 분류되었다. 이는 이현준과 안광국(2009)에

의해 기술된 바와 같이 강우와 같은 수문학적 변수의 특성이 수체 내 수질에 영향을 미친다는 내용과 일치한다.

#### 4. 결론

본 연구는 장성호에 대한 수문학적 변수인 강우량과 저수량 및 호내의 2개 지점(JSD1, JSD2)에서 측정되고 있는 수질항목인 DO, COD, TN 및 TP자료가 복합적으로 나타내는 특성을 파악하기 위해 각 자료에 적합한 전처리 과정을 적용하여 변환한 후 SOM을 적용하여 패턴분류를 수행하였다.

먼저, JSD1 지점에 대한 SOM의 적용에 따른 패턴분류 결과를 보면, 총 5개의 패턴분류 결과를 나타내었으며, 전반적으로 수문학적 변수에 의해 크게 두 개의 그룹으로 분류되었다. Cluster1~3은 강우량 또는 저수량이 큰 패턴을 나타내었으나, Cluster4와 5는 해당 변수가 작은 패턴을 보였다. 특히, 영양염류인 TN 및 TP의 패턴을 고려할 때, Cluster1의 경우 두 영양염류에 의한 부영양화의 발생 가능성이 가장 높은 패턴으로 나타났으며, Cluster3의 경우는 TN에 의한 부영양화 발생 가능성이 높은 패턴으로 나타났다. 즉, JSD1 지점에서는 TN 및 TP에 의한 부영양화 발생 가능성이 높은 자료들이 강우량과 저수량이 큰 값을 갖는 Cluster1과 3으로 분류되었음을 알 수 있다. 또한, 수문학적 변수의 패턴이 낮으며, COD가 높은 패턴은 Cluster4와 5로 분류되었다.

JSD2 지점에 대한 SOM의 적용 결과 역시 총 5개의 패



턴으로 분류되었으며, 전반적으로 수문학적 변수에 의해 크게 두 개의 그룹으로 분류되었다. Cluster1과 3 및 5는 강우량 또는 저수량이 큰 패턴을 나타내었으나, Cluster2와 4는 해당 변수가 작은 패턴을 보였다. 특히 Cluster5의 경우 영양염류인 TN 및 TP에 의한 부영양화의 발생 가능성이 가장 높은 패턴으로 나타났으며, Cluster1의 경우는 TN에 의한 부영양화 발생 가능성이 높은 패턴으로 나타났다. JSD2 지점 역시 강우량과 저수량이 큰 값을 갖는 Cluster1과 5에 두 영양염류에 의한 부영양화 발생 가능성이 높은 자료들이 포함되었다. 수문학적 변수의 패턴이 낮은 양상을 보이는 Cluster2와 4의 경우, 오염물질인 COD에 대한 표준화 참조벡터가 큰 값을 나타내고 있다는 점에서 JSD1 지점과 유사한 결과를 보이고 있으나, TP에 대한 높은 값을 보인다는 점에서 상이한 결과를 제시하고 있다.

즉, JSD1지점에서 분류된 패턴들을 보면, 두 개의 클러스터(Cluster1과 3)에서 영양염류인 TN과 TP가 상대적으로 높은 값을 갖는 패턴을 나타내었으나, JSD2 지점에서는 보다 많은 클러스터(Cluster1, 2, 4, 5)에서 영양염류에 의한 오염의 정도가 심한 것으로 나타났다. 이와 같은 영양염류에 대한 패턴의 공간적 이질성은 측정지점의 위치에 기인한 것으로 판단된다. 따라서 저수지 내의 수체에 대한 수질평가 및 수질과 관련된 심도있는 연구를 위해 보다 많은 지점에 대한 수질측정이 필요할 것으로 판단된다.

특히, JSD1 지점의 Cluster1과 JSD2 지점의 Cluster5의 경우, 수문학적 변수가 크며 오염의 정도가 매우 심한 상태의 유사한 양상을 나타내고 있다. 이는 우기동안 유역 내의 오염물질이 수체 내로 유입되어 발생하는 현상으로 판단되므로, 이러한 패턴을 나타내는 자료들이 측정될 경우 부영양화 및 수체의 오염에 대해 보다 깊은 관심이 필요할 것으로 보인다.

본 연구의 대상 지점인 장성호와 같은 농업용 저수지는 주로 중소규모의 댐에 의해 조성되어 있어 유역면적이 상대적으로 작으며 얕은 수심을 갖는 지형적 특성을 가지고 있다. 따라서 강우의 발생에 따른 저수지로의 유입량이 단시간 내에 발생할 수 있으며, 유입된 유량에 포함된 오염물질이 수체 전체에 영향을 미칠 수 있는 공간적 특성을 지니고 있다. 따라서 시·공간적으로 수질오염에 대한 취약성이 높다고 할 수 있으므로 이에 대한 대안의 수립이 필요하며, 이를 위한 기초자료 분석이 반드시 선행되어야 한다.

결론적으로 SOM의 적용 결과, 측정 자료에 의해 대표되는 저수지의 수문·수질자료에 대한 전체적인 특성을 패턴 분류를 통해 분할하였으며, 각 변수에 대한 패턴별 상대성을 고려한 클러스터별 특성 및 지점들 사이의 이질성을 파악할 수 있었다. 이는 측정 자료에 대한 분석 기법으로서 SOM의 적용성을 나타내고 있는 것으로 판단되며, 향후 저수지에서 측정되고 있는 보다 많은 수질항목을 포함한 패턴분류를 위해 SOM의 적용이 가능할 것으로 판단된다.

또한 상술한 바와 같이 수문학적 변수에 대한 표준화 참조벡터가 큰 값을 나타낼 때 부영양화 발생 가능성이 높다는 점을 고려하여, 향후 심화연구를 위해 수문학적 변수의

변화에 따른 오염물질에 대한 예측모형의 개발이 필요하며, 이를 위해 SOM의 또 다른 응용이 가능할 것으로 판단된다.

## 사 사

본 연구과제는 환경부지정 전남녹색환경지원센터의 연구비 지원에 의해 수행한 연구과제입니다.

## 참고문헌

- 김범철, 사승환, 김문숙, 이윤경, 김재구(2007). 국내 호수의 제한영양소와 하수처리장 방류수 인 기준 강화의 필요성. *수질보전 한국물환경학회지*, **23**(4), pp. 512-517.
- 김용구, 진영훈, 정우철, 박성천(2008). 호소수의 강우, 저수량 및 TOC변동 특성분석을 위한 자기조직화 방법의 적용. *수질보전 한국물환경학회지*, **24**(5), pp. 611-617.
- 농림수산식품부, 한국농어촌공사(2010). *농업용저수지 독 높이 기사사업 사전환경성검토서* (영산강수계: 장성호).
- 박유미, 이의행, 이상재, 안광국(2009). 탐정저수지의 부영양화 특성 및 주요 변수 간의 상호관계. *한국하천호수학회지*, **42**(3), pp. 382-393.
- 송자섭(2011). SOFM의 적용에 의한 영산강 수질 및 유량자료의 시·공간적 패턴분류 특성. 박사학위논문, 동신대학교.
- 이현준, 안광국(2009). 우리나라 인공호 관리를 위한 다변수 수질평가 모델의 개발 및 적용. *한국하천호수학회지*, **42**(2), pp. 242-252.
- 진영훈, 김용구, 노경범, 박성천(2009). 수질 및 유량자료의 기초통계량 분석에 따른 공간분포 파악을 위한 SOM의 적용. *수질보전 한국물환경학회지*, **25**(5), pp. 735-741.
- 한국농어촌공사(2010). *농업용저수지 유형별 수질예측모델 적용방안 연구*.
- 환경부 국립환경과학원 물환경정보시스템(2011). <http://water.nier.go.kr/>.
- 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.
- Alvarez-Guerra, M., González-Piñuela, C., Andrés, A., Galán, B., and Viguri, J. R. (2008). Assessment of Self-Organizing Map artificial neural networks for the classification of sediment quality. *Environmental International*, **34**, pp. 782-790.
- Bedoya, D., Novotny, V., and Manolagos, E. S. (2009). Instream and offstream environmental conditions and stream biotic integrity: Importance of scale and site similarities for learning and prediction. *Ecological Modelling*, **220**, pp. 2393-2406.
- Faggiano, L., Zwart, D., García-Berthou, E., Lek, S., and Gevrey, M. (2010). Patterning ecological risk of pesticide contamination at the river basin scale. *Science of the Total Environment*, **408**, pp. 2319-2326.
- Garcia, H. L. and González, I. M. (2004). Self-organizing map and clustering for wastewater treatment monitoring. *Engineering Applications of Artificial Intelligence*, **17**, pp. 215-225.
- Hentati, A., Kawamura, A., Amaguchi, H., and Iseri, Y. (2010). Evaluation of sedimentation vulnerability at small hillside

- reservoirs in the semi-arid region of Tunisia using the Self-Organizing Map. *Geomorphology*, **122**, pp. 56-64.
- Hsu, K. L., Gupta, H. V., Sorooshian, S., and Imam, B. (2002). SOLO-An artificial neural network suitable for hydrologic modeling and analysis. *Water Resources Research*, **38**, pp. 1-38.
- Ikem, A. and Adisa, S. (2011). Runoff effect on eutrophic lake water quality and heavy metal distribution in recent littoral sediment. *Chemosphere*, **82**, pp. 259-267.
- Jeong, K. S., Hong, D. G., Byeon, M. S., Jeong, J. C., Kim, H. G., Kim, D. K., and Joo, G. J. (2010). Stream modification patterns in a river basin: Field survey and self-organizing map (SOM) application, *Ecological Informatics*, DOI: 10.1016/j.ecoinf.2010.04.005.
- Kalteh, A. M., Hjorth, P., and Berndtsson, R. (2008). Review of the self-organizing map (SOM) approach in water resources: Analysis, modelling and application. *Environmental Modelling and Software*, **23**, pp. 835-845.
- Kazi, T. G., Arain, M. B., Jamali, M. K., Jalbani, N., Afridi, H. I., Sarfraz, R. A., Baig, J. A., and Shah, A. Q. (2009). Assessment of water quality of polluted lake using multivariate statistical techniques: A case study. *Ecotoxicology and Environmental Safety*, **72**, pp. 301-309.
- Lu, H. C., Hsieh, J. C., and Chang, T. S. (2006). Prediction of daily maximum ozone concentrations from meteorological conditions using a two-stage neural network. *Atmospheric Research*, **81**, pp. 124-139.
- Mustonen, S. M., Tissari, S., Huikko, L., Kolehmainen, M., Lehtola, M. J., and Hirvonen, A. (2008). Evaluating online data of water quality changes in a pilot drinking water distribution system with multivariate data exploration methods. *Water Research*, **42**, pp. 2421-2430.
- Park, Y. S., Song, M. Y., Park, Y. C., Oh, K. H., Cho, E., and Chon, T. S. (2007). Community patterns of benthic macroinvertebrates collected on the national scale in Korea. *Ecological Modelling*, **203**, pp. 26-33.
- Song, M. Y., Hwang, H. J., Kwak, I. S., Ji, C. W., Oh, Y. N., Youn, B. J., and Chon, T. S. (2007). Self-organizing mapping of benthic macroinvertebrate communities implemented to community assessment and water quality evaluation. *Ecological Modelling*, **203**, pp. 18-25.
- Su, S., Zhi, J., Lou, L., Huang, F., Chen, X., and Wu, J. (2010). Spatio-temporal patterns and source apportionment of pollution in Qiantang River (China) using neural-based modeling and multivariate statistical techniques. *Physics and Chemistry of the Earth*, DOI: 10.1016/j.pce.2010.03.021.
- Tobiszewski, M., Tsakovski, S., Simeonov, V., and Namiesnik, J. (2010). Surface water quality assessment by the use of combination of multivariate statistical classification and expert information. *Chemosphere*, **80**, pp. 740-746.
- Vesanto, J., Himberg, J., Alhoniemi, E., and Parhankangas, J. (2000). *SOMToolbox for Matlab5*, Helsinki University of Technology Report A57.