

수질 및 유량자료의 기초통계량 분석에 따른 공간분포 파악을 위한 SOM의 적용

진영훈[†] · 김용구^{*} · 노경범^{*} · 박성천^{*}

동신대학교 공업기술연구소

^{*}동신대학교 토목공학과

Application of SOM for the Detection of Spatial Distribution considering the Analysis of Basic Statistics for Water Quality and Runoff Data

Young-Hoon Jin[†] · Yong-Gu Kim^{*} · Kyong-Bum Roh^{*} · Sung-Chun Park^{*}

Institute of Industrial Research and Technology, Dongshin University

^{}Department of Civil Engineering, Dongshin University*

(Received 9 June 2009, Revised 28 July 2009, Accepted 12 August 2009)

Abstract

In order to support the basic information for planning and performing the environment management such as Total Maximum Daily Loads (TMDLs), it is highly recommended to understand the spatial distribution of water quality and runoff data in the unit watersheds. Therefore, in the present study, we applied Self-Organizing Map (SOM) to detect the characteristics of spatial distribution of Biological Oxygen Demand (BOD) concentration and runoff data which have been measured in the Yeongsan, Seomjin, and Tamjin River basins. For the purpose, the input dataset for SOM was constructed with the mean, standard deviation, skewness, and kurtosis values of the respective data measured from the stations of 22-subbasins in the rivers. The results showed that the 4×4 array structure of SOM was selected by the trial and error method and the best performance was revealed when it classified the stations into three clusters according to the basic statistics. The cluster-1 and 2 were classified primarily by the skewness and kurtosis of runoff data and the cluster-3 including the basic statistics of YB_B, YB_C, and YB_D stations was clearly decomposed by the mean value of BOD concentration showing the worst condition of water quality among the three clusters. Consequently, the methodology based on the SOM proposed in the present study can be considered that it is highly applicable to detect the spatial distribution of BOD concentration and runoff data and it can be used effectively for the further utilization using different water quality items as a data analysis tool.

keywords : BOD, Runoff, Self-organizing map, Spatial distribution, Statistics, Water quality

1. 서론

정부합동으로 1998~2000년 동안 수립한 「4대강 수계별 물 관리종합대책」의 일환으로 도입된 수질오염총량관리제도에 따라 각 수계별 기본계획 및 시행계획이 수립되고 이에 대한 이행평가가 수행되고 있으며, 이는 낙동강·금강·영산강 등 3대강의 경우 수계별 법령에 따라 의무제로 관리되고 있다(강복규, 2007). 이와 같은 환경관리를 위한 제도의 원활한 시행을 위해서 하천유역에 대하여 다양하고 정확한 자료의 축적과 그에 대한 분석이 필요하다. 이를 위해 현재 환경부의 국립환경과학원에서는 수질 및 유량측정망을 통하여 자료를 측정하고 있으며, 이에 관한 DB 및 웹시스템(<http://smat.nier.go.kr>)을 운영하여 자료를 공개하고 있다(이호열, 2009).

이러한 양질의 자료에 대한 활용성을 높이기 위해서 다양한 자료분석기법의 개발이 필요하며, 이는 앞서 언급한

바와 같이 수질오염총량관리제도의 원활한 수행을 위한 기반을 이룰 수 있다. 또한 수질과 수량에 대한 통합적 분석 기법을 적용할 경우 향후 환경정책의 입안을 위해서도 기초적인 정보를 제공할 수 있을 것으로 기대된다. 특히 수질오염총량관리제도와 같이 수계별로 수행되는 제도에서 수질 및 유량자료의 공간적 분포를 파악하는 것이 수계내의 각 단위유역별 특성을 파악하는데 도움이 될 수 있다.

일정기간 측정된 자료를 이용하여 공간적 분포를 파악하기 위해서는 일반적으로 패턴분류 기법이 사용될 수 있으며, 이러한 기법들 중 최근 자기조직화 지도(Self Organizing Map, SOM) 기법의 적용이 널리 이루어지고 있다. 이러한 SOM의 하천 유량자료에 대한 적용 예로는 Hsu 등(2002), Jain and Srinivasulu(2006), Srinivasulu and Jain(2006), 김용구 등(2006), 박성천 등(2006) 및 김용구 등(2008a)이 있으며, 수질 및 수처리 분야에서는 김용구 등(2008b)과 López and Machón(2004)가 SOM을 적용하였다. 또한 기상분야에서는 Nishiyama 등(2007)이 SOM을 적용하여 우기 시 폭우를 유발하는 패턴을 파악하였다.

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

본 연구는 국립환경과학원에서 공개한 자료들 중 영산강·섬진강·탐진강 수계의 단위유역별 수질 및 유량자료를 이용하여 그 공간적 분포를 파악하고자 하였으며, 이를 위한 분석기법으로 앞서 언급한 바와 같이 패턴분류에 유용한 SOM을 적용하였다.

2. 연구방법

2.1. 대상지점 및 자료

본 연구는 수질오염총량관리제도의 대상 수계들 중 영산강·섬진강·탐진강 수계의 단위유역 말단지점을 대상으로 하였으며, 현재 환경부의 국립환경과학원에서 공개하고 있는 수질 및 유량자료를 활용하였다.

대상수계 내에는 총 28개의 단위유역 말단지점들이 있으며, 이들 중 25개 지점을 Fig. 1에 나타내었다. 본 연구에서는 이들 중 활용 가능한 자료가 존재하지 않는 지점과 측정횟수가 적은 지점을 제외한 22개의 지점들에서 측정된 자료를 이용하였다.

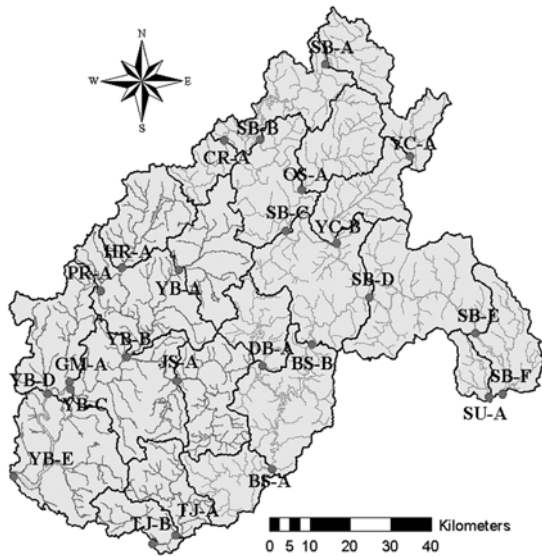


Fig. 1. Study area and stations.

현재 공개된 수질자료는 수운을 포함하여 총 10개 항목이 있으나, 본 연구는 수질항목 중 BOD 농도자료와 유량자료를 SOM에 적용하였다. 선택된 자료에 대한 공간적 분포를 파악하기 위해 각 지점별 측정자료에 대한 평균, 표준편차, 왜도 및 첨도를 계산하여 SOM의 입력자료로 활용하였다. Table 1에 나타난 바와 같이 각 지점별 기초통계량 자료를 이용하여 입력자료를 구성하였으며, 모든 지점이 121회 이상의 측정자료를 보유하고 있다.

2.2. 자기조직화 기법(SOM)

SOM은 다차원의 자료를 2차원으로 사상시킬 수 있는 인공신경회로망 이론의 한 종류이며, 클러스터링을 위한 방법으로 널리 이용되고 있다. 또한 SOM은 훈련과정에 목표값

을 갖지 않는 비교사 학습방법(unsupervised learning algorithm)의 일종으로 다차원의 입력자료들을 분류하여 2차원으로 사상시킨다. SOM의 장점으로는 먼저 복잡한 다차원 자료의 클러스터링에 그 적용성이 뛰어나다는 점이다. 또한 자료의 가시화가 쉽고, 이에 따라 자료의 특성 파악을 위한 자료분석도구로 활용될 수 있으며, 클러스터링 결과의 구조를 미리 지정해줄 수 있는 장점을 가지고 있다.

SOM 구조는 n 차원의 입력자료를 사용할 경우 n 개의 입력노드를 갖게 되며, 입력된 자료를 k 개의 영역으로 분류하고자 할 경우 k 개의 출력노드를 갖게 된다. 각 입력노드들은 모든 출력노드들과 연결되고 각각의 연결강도를 갖는다. 일반적으로 입력노드는 입력자료를 네트워크로 전달하는 기능을 하며, 출력노드는 전달된 입력자료와 입·출력노드 사이의 연결강도를 이용하여 거리를 계산하는 기능을 수행한다.

이러한 과정에서 각 노드들은 학습할 수 있는 기회를 부여받기 위해 서로 경쟁하며, 입력자료와의 거리가 가장 가까운 연결강도를 갖는 노드가 승리하게 된다. 이를 경쟁과정이라 하며, 다음의 식 (1)과 같은 m 차원의 입력자료(X)와 식 (2)와 같은 출력노드 j 의 연결강도(W_j)에 대하여 출력노드 중의 승자노드($i(X)$)는 식 (3)에 의해 결정한다. 즉, 승자노드의 선택은 입력자료의 패턴과 가장 유사한 연결강도를 선정하는 것이며, 유사정도(similarity matching) 측정을 위해 유클리드 거리(Euclidean distance)를 이용한다.

$$X = [x_1, x_2, \dots, x_m]^T \quad (1)$$

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

$$i(X) = \arg \min_j \|X - W_j\| \quad (3)$$

여기서 T 는 전치행렬을 의미하며, l 은 전체 출력노드의 수이다.

상기의 과정에 의해 결정된 승자노드만이 출력신호를 보낼 수 있는 유일한 노드가 되며, 승자노드와 이에 인접한 이웃 노드들만이 제시된 입력자료에 대한 학습이 허용된다. 인접노드를 결정하는 반경에 따라 학습이 진행되는 노드의 수가 결정되며, 이 반경은 학습이 진행됨에 따라 서서히 줄어들어 점점 적은 개수의 노드들이 학습을 하게 된다. 일반적으로 기하학적 이웃반경의 조정을 위해서 대칭성과 수렴특성을 지닌 가우시안 함수(Gaussian function)를 이용한다. 이를 근접반경 조정과정이라 하며, 최종적으로 단지 승자노드만이 그것의 연결강도를 조정하게 된다.

이상과 같은 단계의 처리가 끝나면 마지막으로 적응학습 과정에 의해 실제 연결강도의 조정이 이루어진다. 조정되기 이전의 연결강도를 $W_j(n)$, 조정된 후의 새로운 연결강도를 $W_j(n+1)$ 이라 할 때, 이산적인 시간간격 t 에 대한 조정규칙은 다음 식 (4)와 같이 표현된다.

$$W_j(n+1) = W_j(n) + \eta(n) \cdot h_{ji(X)}(n) \cdot [X - W_j(n)] \quad (4)$$

Table 1. Basic statistics of BOD concentration and runoff used for input data of SOM

Station	Data period	No. data	BOD (mg/L)				Runoff (m ³ /s)			
			Mean	Standard deviation	Skewness	Kurtosis	Mean	Standard deviation	Skewness	Kurtosis
GM_A GoMak_A	Apr.03.2006 - Dec.18.2008	121	3.289	1.648	0.990	2.097	4.620	12.407	8.394	79.560
DB_A DongBok_A	Sep.17.2004 - Dec.02.2008	176	1.076	0.492	1.105	1.769	1.675	5.826	10.224	119.092
BS_A BoSeong_A	Aug.03.2004 - Dec.12.2007	139	2.163	1.015	1.457	3.102	8.726	32.398	5.454	32.318
BS_B BoSeong_B	Sep.17.2004 - Dec.02.2008	176	1.024	0.633	5.535	48.710	5.080	16.692	9.752	102.111
SB_A SeomBon_A	Jan.19.2005 - Dec.01.2008	163	0.848	0.375	0.704	0.906	3.454	4.854	2.775	8.502
SB_B SeomBon_B	Aug.08.2004 - Dec.13.2007	142	0.854	0.367	0.823	0.745	3.470	20.926	8.641	76.385
SB_C SeomBon_C	Jan.28.2005 - Dec.01.2008	158	1.465	0.802	1.221	2.356	14.764	27.092	5.798	40.846
SB_D SeomBon_D	Sep.17.2004 - Dec.02.2008	171	1.356	0.657	1.077	1.530	33.526	51.102	4.421	23.651
SB_E SeomBon_E	Sep.17.2004 - Dec.02.2008	174	1.339	0.763	1.149	1.059	42.808	61.813	3.813	16.643
YB_A YeongBon_A	Sep.14.2004 - Dec.18.2008	186	2.749	1.616	1.381	2.327	8.715	18.437	6.563	55.496
YB_B YeongBon_B	Sep.14.2004 - Apr.29.2008	153	8.032	3.116	0.523	-0.027	33.522	46.171	3.202	10.301
YB_C YeongBon_C	Sep.15.2004 - Dec.04.2008	170	6.405	3.450	1.352	3.090	58.043	125.304	7.671	71.330
YB_D YeongBon_D	Sep.15.2004 - Dec.04.2008	170	4.805	2.030	0.741	1.250	66.347	172.827	6.490	48.141
OS_A OSu_A	Jan.19.2005 - Dec.01.2008	165	1.508	1.087	1.996	5.397	6.516	13.409	5.551	36.006
YC_A YoCheon_A	Jan.28.2005 - Dec.01.2008	159	0.777	0.646	6.625	62.150	2.178	4.626	8.427	87.115
YC_B YoCheon_B	Jan.28.2005 - May.01.2008	130	1.576	0.971	1.404	3.531	6.955	15.939	8.290	80.773
JS_A JiSeok_A	Sep.14.2004 - Dec.18.2008	184	2.377	1.321	1.204	1.460	7.124	16.218	8.007	81.756
CR_A ChuRyung_A	Jan.19.2005 - Dec.01.2008	165	0.798	0.377	0.531	-0.440	3.767	15.990	11.079	131.816
TJ_A TamJin_A	Sep.16.2004 - Dec.18.2008	179	1.831	1.109	1.900	4.415	5.377	9.631	5.150	29.032
TJ_B TamJin_B	Sep.16.2004 - Dec.18.2008	177	1.659	1.013	2.415	9.585	9.173	24.125	8.798	88.255
PR_A PyungRim_A	Jan.22.2005 - Dec.18.2008	171	2.002	1.571	4.237	29.738	1.487	2.244	4.166	26.734
HR_A HwangRyong_A	Sep.14.2004 - Dec.18.2008	184	2.460	1.366	0.824	-0.093	6.739	15.589	5.395	33.636

여기서 η 는 시간 n 이 증가함에 따라 서서히 감소하는 학습율을 나타내는 매개변수이며, $h_{j,i}(x)$ 는 근접반경 조정과정의 기하학적 이웃반경을 나타낸다.

3. 결과 및 고찰

SOM에 의한 패턴분류를 수행하기 위해 영산강, 섬진강 및 탐진강 유역의 수질오염총량관리제도의 단위유역 말단

지점들 중에서 22개 지점에 대한 BOD 농도 및 유량자료에 대한 기초통계량인 평균, 표준편차, 왜도 및 첨도를 산정하여 사용하였다. 이에 따라 SOM에 대한 입력자료는 총 8차원으로 구성된 22개 자료로 구축되었다. SOM의 배열은 육각형배열(hexagonal array)을 적용하였으며, 그 크기는 시행착오법에 의해 Fig. 2에 도시한 바와 같이 4×4의 육각형 배열을 갖는 구조로 결정하였다.

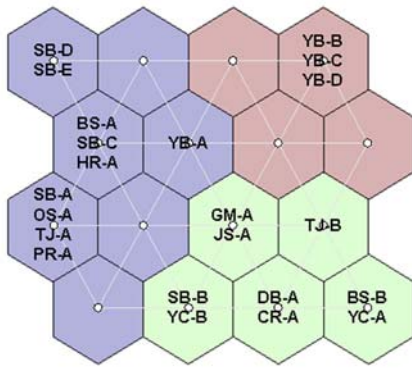


Fig. 2. Pattern classification by SOM using the basic statistics of BOD concentration and runoff.

3.1. Davies-Bouldin Index (DBI)

이러한 구조를 갖는 SOM에 대하여 분류 가능한 최소 및 최대 패턴의 수를 2개에서 16개까지 적용하였다. 이에 따른 최적의 패턴 수를 결정하기 위해 패턴분류의 타당성을 나타내는 DBI(Davies-Bouldin Index)를 López and Machón (2004)에 의해 제안된 다음의 식 (5)~(7)을 적용하여 각 클러스터의 수에 따라 산정하여 Fig. 3에 나타내었다.

$$DBI = \frac{1}{K} \sum_{j=1}^K R_j, R_j = \max_k \left\{ \frac{(S_j + S_k)}{D_{jk}} \right\} \quad (j \neq k) \quad (5)$$

$$S_j = \frac{1}{N_j} \sum_{m_i \in Q_j} tm_i - M_j t \quad (j = 1, \dots, K) \quad (6)$$

$$D_{jk} = tM_j - M_k t \quad (7)$$

여기서 M_j 는 클러스터- j 의 N_j 개의 노드로부터 계산된 평균 연결강도 벡터이며, S_j 는 클러스터- j 로 구분된 노드들 사이의 동질성의 정도를 나타낸다. 또한 D_{jk} 는 두 클러스터 (j, k) 사이의 차이의 정도를 나타내며, 여기서 k 는 클러스터- j 와 다른 클러스터를 나타내는 첨자이다. S_j 의 경우, 그 값이 작을수록 높은 동질성을 의미하며, D_{jk} 는 그 값이 클

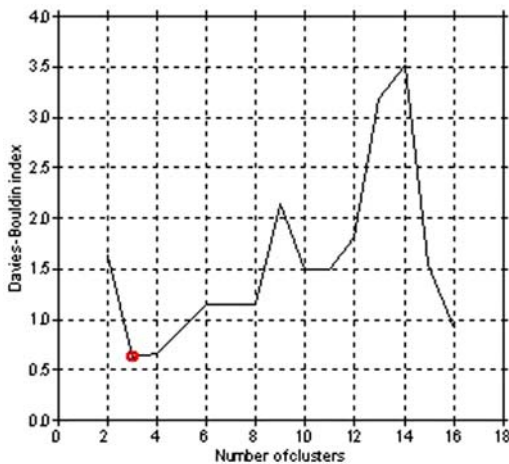


Fig. 3. Results of Davies-Bouldin index.

수록 두 평균 연결강도 벡터들 사이의 유사성이 적어짐을 나타낸다(Nishiyama et al., 2007). 이러한 과정에 따라 계산된 DBI에 의해 총 클러스터의 수가 3개로 분류될 때 최적의 결과를 보이고 있음을 Fig. 3에서 확인할 수 있다.

따라서 4x4의 육각형배열을 갖는 SOM 구조에 의해 총 3개의 클러스터로 구분된 결과를 Fig. 2에 나타내었다. 즉, 이는 8차원을 갖는 22개의 입력자료를 이용한 SOM의 패턴분류 결과를 보여주고 있다. 첫 번째 클러스터(Cluster-1)에는 10개의 지점이 분류되었으며, 두 번째 클러스터(Cluster-2)에는 9개 지점이, 마지막으로 세 번째 클러스터(Cluster-3)에는 3개 지점이 분류되었다.

3.2. BOD 패턴특성 분석결과

각 클러스터별 특징을 알아보기 위해 Fig. 4와 Fig. 5에 BOD 농도 및 유량자료에 대한 Box whisker plot을 도시하였다. Fig. 4에는 BOD 농도에 대한 각 클러스터별 평균, 표준편차, 왜도 및 첨도에 대하여 나타내었으며, 각각의 통계치에 대하여 최소값, 1사분위값, 중앙값, 3사분위값, 최대값 및 이상치($\{1\text{사분위값} - 1.5 \times (3\text{사분위값} - 1\text{사분위값})\}$ 보다 작거나 $\{3\text{사분위값} + 1.5 \times (3\text{사분위값} - 1\text{사분위값})\}$ 보다 큰 값)를 도시하였다. 이와 마찬가지로 Fig. 5에는 유량자료에 대하여 동일하게 나타내었다.

먼저 Fig. 4(a)에 도시한 BOD 농도의 평균값에 대한 패턴분류 결과를 살펴보면, Cluster-1과 Cluster-2는 비슷한 분포를 나타내어 그 구분이 어려우나, Cluster-3의 경우 나머지 두 클러스터와 명확한 차이를 보이고 있다. 이는 BOD 농도의 표준편차가 나타내고 있는 패턴분류 결과에서도 유사한 양상을 나타내고 있다(Fig. 4(b)). 또한 BOD 농도에 대한 왜도와 첨도의 분포를 살펴보면(Fig. 4(c), (d)), Cluster-2에서 다른 클러스터에 비하여 넓게 분포된 양상을 발견할 수 있으나, 클러스터간의 명확한 구분에는 어려움이 있었다. 따라서 Cluster-3은 BOD 농도 자료의 평균 및 표준편차에서 Cluster-1과 Cluster-2와 비교하여 현저한 상이성에 의해 분류되었다.

Cluster-3에 포함된 단위유역 말단지점은 YB_B, YB_C 및 YB_D이며, 이들 세 지점은 영산강 분류에 위치하고 있으며, 광주광역시시의 출구지점인 YB_B와 같은 대도시의 영향을 받는 지점과 그 하류에 위치한 지점들(YB_C, YB_D)로 이루어져 있다. 따라서 다른 지점들에 비해 오염의 정도가 매우 심한 편이며, 이러한 대도시의 하수처리정과 같은 점오염원의 영향을 강하게 받는 물리적인 현상이 SOM의 적용결과에 반영되었다.

3.3. 유량 패턴특성 분석결과

유량자료에 대한 SOM의 패턴분류 결과를 살펴보면, Fig. 5(a)에 22개 대상지점의 평균 유량값 분포를 각 클러스터별로 도시하였으며, 그 양상이 Fig. 4(a)의 BOD 농도의 평균값에 대한 결과에 유사하였다. 또한 유량의 표준편차를 이용한 Fig. 5(b)의 결과 역시 Fig. 4(b)와 매우 유사한 양상을 보여주고 있다. 그러나 평균 유량값 및 표준편차에

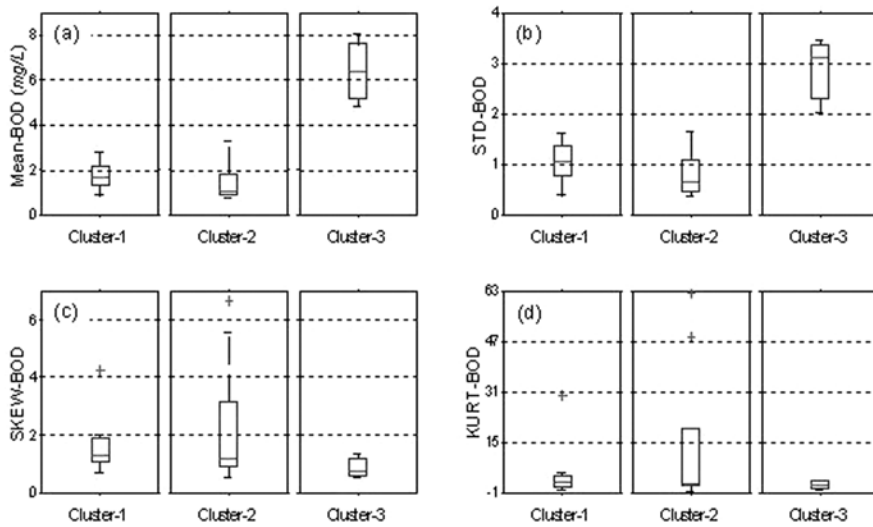


Fig. 4. Box-whisker plots for (a) mean, (b) standard deviation, (c) skewness, and (d) kurtosis of BOD concentration data according to the respective clusters.

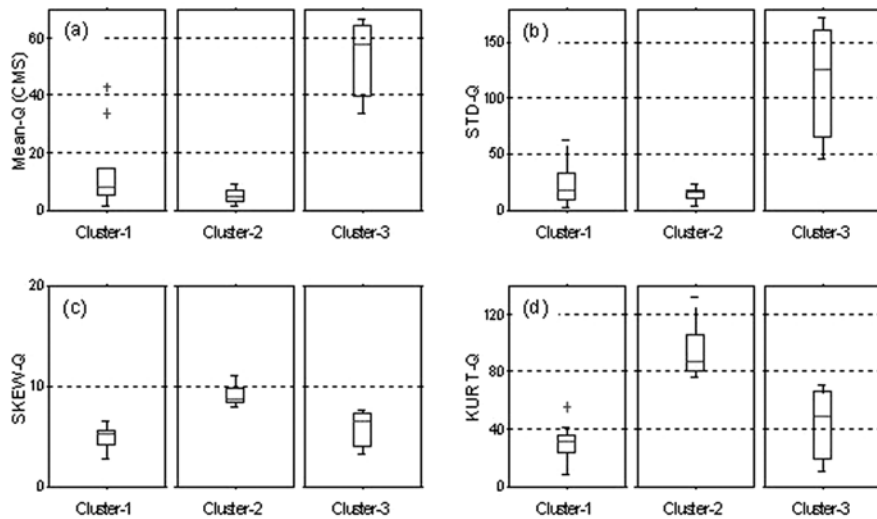


Fig. 5. Box-whisker plots for (a) mean, (b) standard deviation, (c) skewness, and (d) kurtosis of runoff data according to the respective clusters.

대한 Cluster-1의 높은 범위의 값들과 Cluster-3의 낮은 범위의 값들이 중복되어 있는 결과를 보여주고 있으며, 이에 따라 BOD 농도의 평균과 표준편차에 대한 결과와 비교하여 클러스터별 구분이 명확하지 않았다.

BOD 농도 자료에 대한 평균, 표준편차, 왜도, 첨도(Fig. 4(a)~(d)) 및 유량자료의 평균과 표준편차(Fig. 5(a)~(b))에 의한 패턴분류 결과는 Cluster-1과 Cluster-2가 중복된 범위를 보여주고 있는 반면에, 유량자료의 왜도(Fig. 5(c))와 첨도(Fig. 5(d))에 대한 결과는 두 클러스터가 확연하게 구분되고 있는 양상을 보여주고 있다. 따라서 Cluster-1과 Cluster-2는 유량자료의 왜도와 첨도의 차이에 의해 구분되었다.

두 클러스터 사이의 차이를 살펴보면, Cluster-2에 포함된 지점들의 왜도와 첨도가 Cluster-1으로 분류된 지점들의 왜도와 첨도에 비해 상당히 큰 값들을 나타내고 있다. 이러한 결과를 재확인하기 위해 Fig. 6에 측정된 유량자료의 측

정횡수 및 그 크기가 비슷한 지점을 각 클러스터별로 선정하여 각각에 대한 히스토그램을 도시하였다. Fig. 6(a)는 Cluster-1로 분류된 10개의 지점들 중 BS_A 지점의 유량자료 분포를 보여주고 있으며, Fig. 6(b)는 Cluster-2에 포함된 9개의 지점들 중에서 SB_B 지점의 유량자료에 대한 히스토그램을 나타내고 있다.

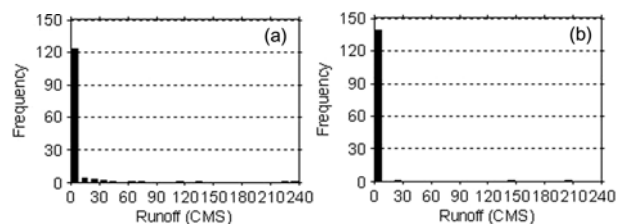


Fig. 6. Histograms for (a) BS_A classified into the Cluster-1 and (b) SB_B classified into the Cluster-2.

두 지점 모두 유량자료의 0~10 CMS 범위에 다수의 자료가 분포하고 있는 것으로 파악된다. 그러나 BS_A 지점(Cluster-1)의 유량자료의 분포가 SB_B 지점(Cluster-2)의 분포에 비하여 각 계급구간에 걸쳐 고르게 존재하고 있음을 알 수 있다. 이는 BS_A의 유량자료는 상류의 인공구조물에 의한 간섭이 없는 자연유량이며, SB_B의 유량자료는 섬진강댐에 의해 조절되는 간섭유량이라는 차이에 기인한 것으로 판단된다. 결국 상기에서 언급한 바와 같이 Cluster-1과 Cluster-2는 측정된 유량자료의 왜도와 첨도의 크기에 따라 구분된 것으로 파악되며, Cluster-2에 포함된 지점들의 유량자료의 분포가 Cluster-1의 지점들의 자료들보다 더욱 비대칭적이며, 뾰족한 형태의 분포형태를 보이는 것으로 나타났다.

4. 결론

본 연구에서 최근 수질오염총량관리제도와 같은 환경관리를 위한 정책의 기초적 정보의 제공을 위해 국립환경과학원에서 측정하여 공개한 자료들 중 영산강·섬진강·탐진강 수계의 단위유역별 수질 및 유량자료를 이용하여 그 공간적 분포를 파악하고자 하였으며, 이를 위한 분석기법으로는 패턴분류에 탁월한 성능을 보이고 있는 SOM을 적용하였다.

영산강·섬진강·탐진강 수계의 단위유역들 중에서 충분한 횡수의 측정자료를 보유하고 있는 총 22개 지점의 BOD 농도와 그에 해당하는 유량자료를 수집하고, 각 지점의 자료에 대한 통계적 특성을 나타내기 위하여 각 자료에 대한 평균, 표준편차, 왜도 및 첨도를 산정하였다. 따라서 SOM을 위한 입력자료는 BOD 농도와 유량에 대한 상기의 4가지 통계적 특성치를 포함하여 총 8차원으로 구성되었다.

SOM의 구조 결정을 위하여 우선 시행착오법에 의해 4×4의 육각형 배열을 선택하였으며, 이러한 구조가 나타낼 수 있는 최소 2개의 패턴에서 최대 16개의 패턴까지 분류하여 최적의 패턴 수를 결정하였다. 이를 나타내는 지수인 DBI에 의해 본 연구에서 결정된 SOM이 3개의 패턴으로 분류될 때 최적의 결과를 나타내었으며, 이에 따른 각 클러스터별 특성을 분석하였다.

먼저, 각 자료의 통계적 특성치의 분포를 알아보기 위해 Box-whisker plot을 도시하였으며, 이에 의해 SOM에 의해 분류된 각 클러스터별 특징을 파악할 수 있었다. BOD 농도의 평균과 표준편차에 의한 결과를 보면, Cluster-3은 Cluster-1과 Cluster-2에 비하여 현저한 차이를 보이고 있다. 따라서 이러한 결과에 의해 Cluster-3이 다른 클러스터들과 구분된 것으로 판단되며, Cluster-3로 분류된 지점들인 YB_B, YB_C, 및 YB_D는 광주광역시에 대한 단위유역의 말단지점과 그 직하류의 지점들을 나타내고 있다.

또한 Cluster-1과 Cluster-2는 측정된 유량자료의 왜도와 첨도에 의해 분류되었으며, Cluster-1로 분류된 지점들의 왜도와 첨도가 Cluster-1의 지점들의 왜도와 첨도보다 큰 값들을 나타내었다. 이러한 결과를 확인하기 위하여 각 클러

스터에서 자료측정 횟수 및 그 범위가 유사한 두 지점을 선정하여 유량자료에 대한 히스토그램을 도시하였다. Cluster-1의 지점들 중 BS_A와 Cluster-2의 지점들 중 SB_B의 유량자료에 대한 히스토그램의 비교에서 나타낸 바와 같이 BS_A의 유량자료가 SB_B의 자료에 비해 각 계급구간별로 보다 고르게 분포하고 있음을 파악하였다. 즉, Cluster-1의 지점들에 비하여 Cluster-2로 분류된 지점들의 유량자료가 보다 큰 비대칭성을 나타내었으며, 보다 뾰족한 형태의 분포양상을 보였다.

결론적으로, 본 연구에서 제안한 SOM의 적용에 의해 수질 및 유량자료에 대한 공간분포 파악이 가능한 것으로 나타났다. 이에 대한 각 클러스터별 분류 기준을 파악할 수 있었다. 즉, Cluster-3는 BOD 농도의 평균 및 표준편차의 크기에 의해 명확히 구분되었으며, 나머지 패턴들인 Cluster-1과 Cluster-2는 유량자료의 왜도와 첨도에 의해 분류되었다. 이러한 연구결과를 바탕으로 하여 본 연구에서 수질 및 유량자료의 공간분포 파악을 위해 제안한 SOM이 현재 측정 및 측정되고 있는 다양한 자료에 대하여 적용 가능할 것으로 판단되며, 향후 다양한 자료에 대하여 분석의 제고를 위한 기초분석 도구로 활용될 수 있을 것으로 기대된다.

사 사

이 논문은 2008년 정부(교육과학기술부)의 재원으로 한국학술진흥재단의 지원을 받아 수행된 연구임. [KRF-2008-359-D00020]

참고문헌

- 강복규(2007). 수질보전 패러다임의 전환 - 4대강수계 수질 오염총량관리제도. *한국수자원학회지*, **40**(3), pp. 22-27.
- 김용구, 진영훈, 박성천(2006). 강우-유출특성 분석을 위한 자기조직화방법의 적용. *대한토목학회논문집*, **26**(1B), pp. 61-67.
- 김용구, 진영훈, 박성천, 정천리(2008a). 나주지점의 강우-유출 해석을 위한 최적의 SOM 구조 결정. *한국수자원학회논문집*, **41**(10), pp. 995-1007.
- 김용구, 진영훈, 정우철, 박성천(2008b). 호소수의 강우, 저류량 및 TOC변동 특성분석을 위한 자기조직화 방법의 적용. *수질보전 한국물환경학회지*, **24**(5), pp. 611-617.
- 박성천, 진영훈, 김용구(2006). 강우-유출 예측모형 개발을 위한 자기조직화 이론의 적용. *대한토목학회논문집*, **26**(4B), pp. 389-398.
- 이호열(2009). 영산강유역 수질측정망 운영 및 활용 현황. *한국수자원학회지*, **42**(3), pp. 17-23.
- 환경부 국립환경과학원 수질·유량측정망 등 DB 및 웹시스템(2008). <http://smat.nier.go.kr/>.
- Hsu, K. L., Gupta, H. V., Gao, X., Sorooshian, 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. 1302 (doi:10.1029/2001WR000795).

- Jain, A. and Srinivasulu, S. (2006). Integrated approach to model decomposed flow hydrograph using artificial neural network and conceptual techniques. *Journal of Hydrology*, **317**, pp. 291-306.
- López, H. and Machón, I. (2004). Self-organizing map and clustering for wastewater treatment monitoring. *Engineering Applications of Artificial Intelligence*, **17**, pp. 215-225.
- Nishiyama, K., Endo, S., Jinno, K., Uvo, C. B., Olsson, J., and Berndtsson, R. (2007). Identification of typical synoptic patterns causing heavy rainfall in the rainy season in Japan by a Self-Organizing Map. *Atmospheric Research*, **83**, pp. 185-200.
- Srinivasulu, S. and Jain, A. (2006). A comparative analysis of training methods for artificial neural network rainfall-runoff models. *Applied Soft Computing*, **6**, pp. 295-306.