



Assessing applicability of self-organizing map for regional rainfall frequency analysis in South Korea

Ahn, Hyunjun^a · Shin, Ju-Young^a · Jeong, Changsam^b · Heo, Jun-Haeng^{a*}

^aDepartment of Civil and Environmental Engineering, Yonsei University

^bDepartment of Civil Engineering, Induk University

Paper number: 18-004

Received: 15 January 2018; Revised: 2 February 2018; Accepted: 2 February 2018

Abstract

The regional frequency analysis is the method which uses not only sample of target station but also sample of neighborhood stations in which are classified as hydrological homogeneous regions. Consequently, identification of homogeneous regions is a very important process in regional frequency analysis. In this study, homogeneous regions for regional frequency analysis of precipitation were identified by the self-organizing map (SOM) which is one of the artificial neural network. Geographical information and hourly rainfall data set were used in order to perform the SOM. Quantization error and topographic error were computed for identifying the optimal SOM map. As a result, the SOM model organized by 7×6 array with 42 nodes was selected and the selected stations were classified into 6 clusters for rainfall regional frequency analysis. According to results of the heterogeneity measure, all 6 clusters were identified as homogeneous regions and showed more homogeneous regions compared with the result of previous study.

Keywords: Self-organizing map, cluster analysis, Regional frequency analysis, Heterogeneity measure

Self-organizing map을 이용한 강우 지역빈도해석의 지역구분 및 적용성 검토

안현준^a · 신주영^a · 정창삼^b · 허준행^{a*}

^a연세대학교 토목환경공학과, ^b인덕대학교 토목공학과

요 지

지역빈도해석은 대상 지점과 수문학적 동질성을 만족하는 주변 지점을 하나의 지역으로 보고 빈도해석을 수행하는 방법이다. 따라서 동질한 지역의 구분은 지역빈도해석에 있어서 가장 중요한 가정이라고 할 수 있다. 이에 본 연구에서는 인공신경망 기법중 하나인 자기조직화지도(self-organizing map, SOM) 기법을 활용하여 강우 지역빈도해석을 위한 동질 강수 지역을 구분하였다. 지역구분 인자로는 지형 정보와 시 단위 강우 자료를 활용하였다. 최적 SOM 지도 구성을 위해 정량적 오차와 위상관계 오차를 활용하였다. 그 결과 7×6 배열의 42개의 노드를 갖는 모형을 선정하였고 최종적으로 강우 지역빈도해석을 위해 6개의 군집으로 구분하였다. 동질성 검토 결과 6개의 군집 모두 동질한 지역으로 나타났으며 기존의 유사하게 구분된 지역들과 이질성 척도를 비교하였을 때 좀 더 안정적인 지역 구분결과를 나타내는 것을 확인하였다.

핵심용어: 자기조직화지도, 군집분석, 지역빈도해석, 이질성 척도

1. 서 론

집중호우나 홍수와 같은 물 관련 재난은 우리나라에서 발생하고 있는 가장 크고 대표적인 자연재난이다. 확률수문량

산정은 그러한 물 관련 재난에 대비하여 방재계획을 수립하는데 가장 기초적이고 근간이 되는 절차이다.

확률수문량 산정 방법 중 통계적 기법을 기반으로 하는 빈도해석 방법에는 지점빈도해석과 지역빈도해석이 있다. 그 중 지역빈도해석은 지점빈도해석이 가지고 있는 표본 크기에 따라 높은 재현 기간에 해당하는 확률수문량 산정 시 발생하는 불확실성과 같은 단점을 보완할 수 있는 방법이다. 따라서

*Corresponding Author. Tel: +82-2-2123-2805
E-mail: jhheo@yonsei.ac.kr (J.-H. Heo)

확률수문량 산정 시 지점빈도해석보다 좀 더 안정적이고 신뢰성 있는 추정 가능성이 있는 것으로 알려져 있다(Cunnane, 1989; Lettenmaier and Potter, 1985; Heo *et al.*, 1990; Hosking and Wallis, 1997; Heo, 2016).

이러한 장점을 가진 지역빈도해석은 처음 Darlymple (1960)에 의해 홍수량 자료에 대해서 홍수지수법(index flood method)의 형태로 제시된 이후로 많은 연구가 계속되어 현재는 홍수량뿐만 아니라 강우 자료에 이르기까지 다양하게 적용되며 널리 활용되고 있다(Greis and Wood, 1981; Stedinger, 1983; Hosking *et al.*, 1985; Lee *et al.*, 2007; Heo *et al.*, 2007).

지역빈도해석이 지점빈도해석과 가장 뚜렷하게 구분되는 차이점은 지역빈도해석은 대상 지점의 수문 자료만을 사용하는 것이 아닌 수문학적 동질지역(homogeneous region)으로 구분된 지점들의 모든 자료를 이용하여 확률수문량을 추정한다는 점이다. 따라서 지역빈도해석을 수행하기 위해서는 수문학적 동질지역을 구분하는 과정이 필요하며 이 과정이 지역빈도해석이 지점빈도해석과 구분되는 과정이며 가장 중요한 절차라고 할 수 있다.

그러나 동질지역의 구분에 있어서 정해진 절차와 기준은 따로 마련되지 않았다. 이는 기법의 부재가 아닌 군집의 대상(자료)과 목적에 따라 적용 가능한 기법과 자료가 다르기 때문이다. 따라서 동질지역 구분을 위해서 다양한 기법과 인자들을 활용한 연구가 진행되어 왔다. Mallants and Feyen (1990)과 Guttman (1993)은 일 강우 자료, 지형 인자, 그리고 기상인자 등을 각각 활용하여 수문학적 동질지역 구분을 시도한 바 있다. 또한, 동질지역 구분의 연구가 진행됨에 따라 단일 변량에 의한 구분이 아닌 다양한 인자를 활용하여 지역을 구분하는 다변량 분석 기법을 통해 좀 더 심층적인 지역 구분에 대한 연구가 진행되었다. Barring (1988)은 요인 분석(factor analysis)을 통해 지역 구분을 수행하였고, Burn (1989)과 Smithers and Schulze (2001)는 각각 K-means와 Ward 기법을 이용하여 지역빈도해석을 수행하였다. Dinpashoh *et al.* (2004)와 Nam *et al.* (2015)은 동질지역 구분의 효율성 향상을 위해 주성분 분석(principal component analysis)과 프로크루스테스 기법(procrustes analysis)을 적용한 바 있다.

하지만 이러한 주성분 분석과 같은 방법들은 지금처럼 컴퓨터 기술이 발달하기 이전에 다차원 대용량 자료에 대해 소량의 자료들을 활용하여 효율적으로 군집을 이루어 낼 수 있다는 장점이 있었지만 기본적으로 기법자체가 가지고 있는 선형적 데이터 축소기술로 인해 정보의 손실이 존재하고 비선형적 형태를 보이는 분석 대상에는 적합하지 않은 단점이 존재한다(Liu *et al.*, 2006; Reusch *et al.*, 2007).

한편 최근 컴퓨터 기술의 발전으로 인해 데이터 마이닝이나 인공지능망 기법이 발달함에 따라 대용량의 데이터를 신속히 처리하거나 과거에는 해석 하지 못했던 데이터 간의 관계와 패턴 분석이 가능함에 따라 다양한 방법을 활용한 군집 분석방법이 연구되고 있다(Fayyad *et al.*, 1996; Jain *et al.*, 1999; Lu *et al.*, 2006; Kohonen, 2007; Arribas-Bel *et al.*, 2011). 그 중 자기조직화지도(self-organizing map, SOM)는 비교사학습(unsupervised learning)을 활용하는 인공지능망 기법 중 하나로 기존의 주성분 분석의 대안으로 제시되고 있다(Kohonen, 2001; Jun and Choi, 2013). SOM은 다차원의 대용량 데이터를 인간이 직관적으로 판단하기 쉽게 2차원으로 시각화 할 수 있다는 장점을 가지며, 자료에 숨겨진 의미 있는 패턴이나 입력 요인과 자료간의 상관관계를 분석할 수 있는 것으로 알려져 있다(Hewitson and Crane, 2002; Arribas-Bel *et al.*, 2011; Kim *et al.*, 2017). 또한 Mangiameli *et al.* (1996)에 의하면 SOM은 기존의 활용되고 있는 7가지의 계층적 군집 분석 방법 중 가장 뛰어난 성능을 보이는 것으로 알려져 있다.

따라서 본 연구에서는 강우 지역빈도해석을 위해 SOM을 활용하여 수문학적 동질 지역을 구분하였으며, 구분된 지역의 동질성 검토를 통해서 구분된 지역이 강우 지역빈도해석을 수행하는데 있어 적합한지 검토하였다. 또 기존의 다변량 요인 분석을 통해 구분된 지역과의 비교 검토를 수행하였다. 이를 통해 강우 지역빈도 해석을 위한 동질지역 구분의 기초 자료를 제시하고자 한다.

2. 연구 방법론

2.1 자기조직화지도

SOM 기법은 훈련(training) 과정과 지도화(mapping) 과정으로 나누어 볼 수 있다. 훈련 과정에서는 다차원의 입력 데이터를 이용하여 경쟁 학습(competitive learning)을 수행하고 지도화 과정에서는 훈련 과정에서 얻은 결과로 시각화를 극대화하여 새로운 형태(일반적으로 2차원)로 표출한다.

일반적인 인공지능망 기법에서는 명확한 목표 값과 학습을 통해 추정된 결과 값의 차이를 이용하여 역전파(back-propagation) 알고리즘과 같은 오차 보정 학습(error-correction learning)을 적용하여 목표 값을 찾지만, SOM은 주어진 입력 데이터 안에서 가중치 벡터들의 경쟁 학습을 통해 승자독식(winner-take-all) 형태로 입력 벡터에 가장 근접한 가중치 벡터(best matching unit, BMU)를 찾는다. 또한 이웃 함수(neighborhood function)를 활용하여 BMU 뿐만 아니라 주변

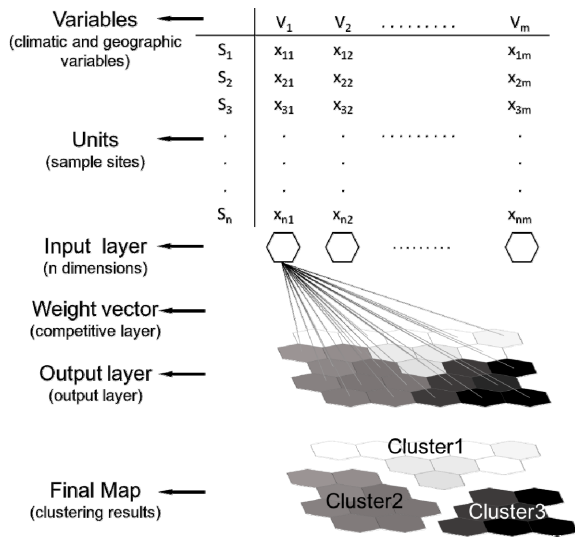


Fig. 1. Structure of self-organizing map

뉴런의 가중치 벡터까지 그 영향을 반영시키는 조정과정 (adaptation)을 거쳐 학습되는 점이 다른 인공신경망 기법과 구분되는 가장 큰 SOM의 특징이다(Moon, 2006; Jun and Choi, 2013).

SOM의 기본구조는 Fig. 1과 같이 입력층(input layer)과 출력층(output layer)으로 구성된다. 입력층은 표본 자료로부터 얻어진 다차원 입력자료로 구성되고 출력층은 일반적으로 육각형 모양의 격자 단위로 구성되며, 각 격자 단위에는 가중치 벡터가 초기화 방법에 따라 임의의 값으로 입력된다. 가중치 벡터는 SOM의 훈련과정을 통해 갱신되며, 최종적으로 갱신된 가중치 벡터는 각 격자 단위에 포함되는 입력 벡터를 대표하게 된다. SOM의 상세한 방법론과 가중치 벡터 계산 및 학습을 통한 갱신에 대한 상세한 설정은 Kohonen (1998)과 Kohonen and Hauta-Kasari (2008)가 제안한 방법을 참고하였다.

2.2 SOM의 초기화 및 평가 방법

SOM의 학습이 이루어지기 전 입력층의 입력 벡터들은 일반적으로 정규화 과정을 거쳐 0과 1 사이의 값으로 변환되며, 가중치 벡터는 임의의 값으로 초기화된다.

가중치 벡터 초기화 방법에는 선형 초기화(linear initialization)와 무작위 초기화(random initialization) 방법이 있다. 무작위 초기화의 경우 입력 데이터의 범위 내에서 무작위로 값을 뽑아내서 가중치 벡터로 선정하여 초기화하는 방법으로 인공신경망 기법에서 일반적으로 널리 활용되는 가중치 초기화 방법이다. 하지만 SOM에서 무작위 초기화 방법은 구성된 네트워크의 가장자리 또는 그 부근의 노드에서 경계 효과 (boundary effect) 현상이 나타나는 단점이 있다(Kiang *et al.*,

1997, 2001; Su *et al.*, 2002).

선형 초기화 방법은 주성분 분석으로 얻어진 고유벡터 중 고유치가 가장 큰 두 개의 값을 선형 내삽하는 방식으로 이루어지며, 일반적으로 무작위 초기화에 비해 학습 속도가 빠르며 작은 표본 자료에 대하여 좀 더 적절한 패턴을 구분할 수 있다고 알려져 있다(Jeong *et al.*, 2011; Jun and Choi, 2013). 그러나 선형 초기화 방법은 학습이 시작되기 전에 동일한 가중치가 고정되기 때문에 최적의 노드 구성방식(topological network)을 가질 수 없다는 단점이 있다(Jeong *et al.*, 2010). 본 연구에서는 선형 초기화 방식으로 가중치를 초기화 하는 방식을 선택하였으며 선형 초기화 방식의 단점을 보완하기 위하여 양자화 오차(quantization error, QE)와 위상관계 오차(topographic error, TE)를 고려하여 SOM 모형을 구성하였다.

2.3 이질성 척도

지역빈도해석에서 지역적 동질성의 의미는 지형학적인 동질 구분이란 수문학적 특성의 유사성을 의미한다. 따라서 지역빈도해석에서는 군집분석을 통해 하나의 지역으로 설정된 지점들에 대해서 수문학적 동질성 여부를 판단하기 위하여 마련된 기준들이 있다. 본 연구에서는 Hosking and Wallis (1997)가 제안한 이질성 척도(heterogeneity measure, H)를 활용하여 군집된 지역의 동질성을 검토하였으며 이질성 척도의 식은 Eq. (1)과 같다.

$$H = \frac{V - u_v}{s_v} \tag{1}$$

여기서, V 는 지점의 L-변동계수들의 가중표준편차로 Eq. (2)에 의해 산정되며, u_v 와 s_v 는 각각 V 의 평균과 표준편차의 미한다.

$$V = \left\{ \frac{\sum_{i=1}^N n_i (t^{(i)} - t^R)^2}{\sum_{i=1}^N n_i} \right\}^{1/2} \tag{2}$$

여기서, n_i 는 지점 i 의 자료 기간이고 $t^{(i)}$ 는 i 번째 지점의 L-변동계수이고 t^R 은 지역의 L-변동계수로서 Eq. (3)처럼 지점의 자료 기간에 비례하여 가중평균하여 구한다.

$$t^R = \frac{\sum_{i=1}^N n_i t^{(i)}}{\sum_{i=1}^N n_i} \tag{3}$$

Hosking and Wallis (1997)는 H 의 값의 따른 동질지역 여부를 Table 1과 같이 제시하였다. H 의 값이 $H < 1$ 일 경우, 동질

Table 1. Identification of regions by heterogeneity measure (Hosking and Wallis, 1997)

Heterogeneity measure (H)	Results
$H < 1$	Acceptably homogeneous
$1 \leq H < 2$	Possibly heterogeneous
$H \geq 2$	Definitely heterogeneous

지역(acceptably homogeneous), $1 \leq H \leq 2$ 일 경우 이질 가능성 지역(possibly heterogeneous), 그리고 $H > 2$ 일 경우 이질 지역(definitely heterogeneous)으로 구분하였다.

Table 2. Stations code, latitude, and longitude of selected stations

No.	Name	Station code	Latitude	Longitude	No.	Name	Station code	Latitude	Longitude
1	Daegwallyeong	100	128.718	37.680	33	Yeongdeok	277	129.41	36.53
2	Jecheon	221	128.194	37.160	34	Pohang	138	129.38	36.03
3	Chungju	127	127.953	36.970	35	Namhae	295	127.93	34.82
4	Wonju	114	127.947	37.340	36	Geoje	294	128.61	34.89
5	Yangpyeong	202	127.494	37.490	37	Masan	155	128.58	35.17
6	Icheon	203	127.484	37.260	38	Tongyeong	162	128.44	34.85
7	Inje	211	128.167	38.060	39	Geumsan	238	127.48	36.11
8	Chuncheon	101	127.736	37.900	40	Chupungnyeong	135	127.99	36.22
9	Hongcheon	212	127.880	37.680	41	Boeun	226	127.73	36.49
10	Seoul	108	126.966	37.570	42	Daejeon	133	127.37	36.37
11	Suwon	119	126.988	37.270	43	Cheongju	131	127.44	36.64
12	Incheon	112	126.625	37.480	44	Buyeo	236	126.92	36.27
13	Ganghwa	201	126.447	37.710	45	Cheonan	232	127.12	36.78
14	Sokcho	090	128.565	38.250	46	Seosan	129	126.50	36.77
15	Gangneung	105	128.891	37.750	47	Gunsan	140	126.76	36.00
16	Taebaek	216	128.989	37.170	48	Boryeong	235	126.56	36.33
17	Andong	136	128.707	36.570	49	Jeonju	146	127.16	35.82
18	Yeongju	272	128.517	36.870	50	Jeongeup	245	126.87	35.56
19	Mungyeong	273	128.149	36.630	51	Buan	243	126.72	35.73
20	Uiseong	278	128.689	36.360	52	Imsil	244	127.29	35.61
21	Gumi	279	128.321	36.130	53	Namwon	247	127.33	35.41
22	Daegu	143	128.619	35.890	54	Wando	170	126.70	34.40
23	Yeongcheon	281	128.951	35.980	55	Suncheon	174	127.37	35.02
24	Geochang	284	127.911	35.670	56	Goheung	262	127.28	34.62
25	Hapcheon	285	128.170	35.570	57	Yoesu	168	127.74	34.74
26	Sancheong	289	127.879	35.410	58	Gwangju	156	126.89	35.17
27	Jinju	192	128.040	35.160	59	Jangheung	260	126.92	34.69
28	Miryang	288	128.744	35.490	60	Haenam	261	126.57	34.55
29	Ulsan	152	129.320	35.560	61	Mokpo	165	126.38	34.82
30	Busan	159	129.032	35.100	62	Jeju	184	126.53	33.51
31	Ulleungdo	115	130.899	37.480	63	Seogwipo	189	126.57	33.25
32	Uljin	130	129.413	36.990	64	Seongsan	188	126.88	33.39

3. 적용 지점 및 자료

본 연구에서는 강우 지역빈도해석을 위해서 우리나라의 기상청 지점 64개소의 위·경도와 고도와 같은 지형정보 및 시 단위 강우 자료를 활용하였다. 64개의 지점은 우리나라 남한 전역에 비교적 고르게 분포되어 있으며 울릉도와 제주도서 지역의 지점도 포함 시켰다. 각 지점별 자료의 보유 기간은 30~56년이다. Table 2 and Fig. 2 에 관측소 지점명, 지점 번호, 그리고 위치 정보 등을 나타내었다.

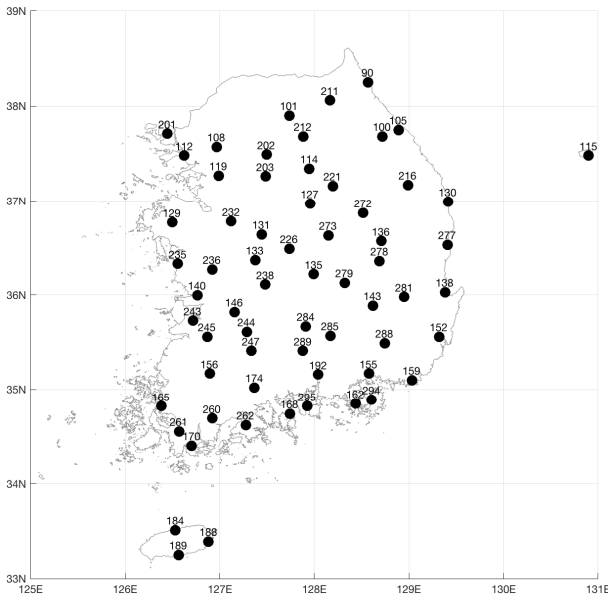


Fig. 2. Locations of 64 stations

SOM을 이용하여 군집분석을 수행하기 위하여 기상청 64개 지점에 대해서 Table 3처럼 위·경도, 고도, 월평균 강우량, 연강우량의 평균과 표준편차 등 총 17가지의 변수를 활용하였다. 17가지의 변량의 경우 각 변량의 상대적 크기가 다르기 때문에 SOM 학습을 위하여 Table 3의 변환작업을 거쳐 입력 변수를 0과 1 사이의 값으로 변환하였다.

4. 결과 및 토의

4.1 SOM 군집분석

4.1.1 최적 노드 수 및 지도배열 결정

안정적인 훈련과정을 위해서 노드 배열은 비대칭을 택했으며, 선형 초기화 방법의 단점을 보완하고자 배열은 QE와 TE를 활용하여 Tian *et al.* (2014)이 제안한 노드 구성의 수를 만족하는 범위 내에서 선정하였다.

Table 4. Results of quantization error and topographic error for each SOM map

M	Row	Column	Node	QE	TE
40	4	10	40	0.285	0.047
	4	9	36	0.295	0.031
	5	8	40	0.295	0.016
	6	7	42	0.289	0.031
	7	6	42	0.285	0.016
	8	5	40	0.287	0.016
	9	4	36	0.298	0.016
	10	4	40	0.294	0.016

우선 학습을 위한 노드 구성을 위해서 대략적인 지도의 크기를 선정하기 위하여 Tian *et al.* (2014)이 제안한 Eq. (4)를 사용하였다.

$$M \approx 5\sqrt{N} \tag{4}$$

여기서, N 은 관측대상의 수이고 M 은 우변의 값에 가까운 정수인 뉴런의 수이다.

Tian *et al.* (2014)이 제시한 Eq. (4)로 산정된 40 ($= 5\sqrt{64}$)을 만족하는 배열 조합 8가지를 선정하고 각각 QE와 TE를 산정하였다. 계산된 결과는 Table 4와 같다. 그 중 가로 열이 7이고 세로 열이 6인 노드 42개의 배열 지도의 경우, QE와 TE의 값이 각각 0.285와 0.016으로 8가지 종류의 배열 중에서 가장 안정적인 노드 배열로 추정되었다. 따라서 본 연구에서는 7×6의 배열의 육각형 격자 단위로 구성된 SOM을 이용하여 군집화를 수행하였다. 분석된 최종 결과는 Fig. 3과 같다.

Fig. 3(a)는 U-matrix 결과이고 Fig. 3(b)는 Fig. 3(a)의 결과를 바탕으로 7개의 군집으로 구분된 군집분석의 결과이다. Fig. 3(c)는 입력 인자별 SOM의 패턴을 분석한 결과이다. 여기서, U-matrix는 입력 데이터 차원의 공간에서 뉴런 간의 거리를 시각적으로 나타낸 지도이다.

Fig. 3(a)를 살펴보면 울릉도(115)와 대관령(100)의 경우

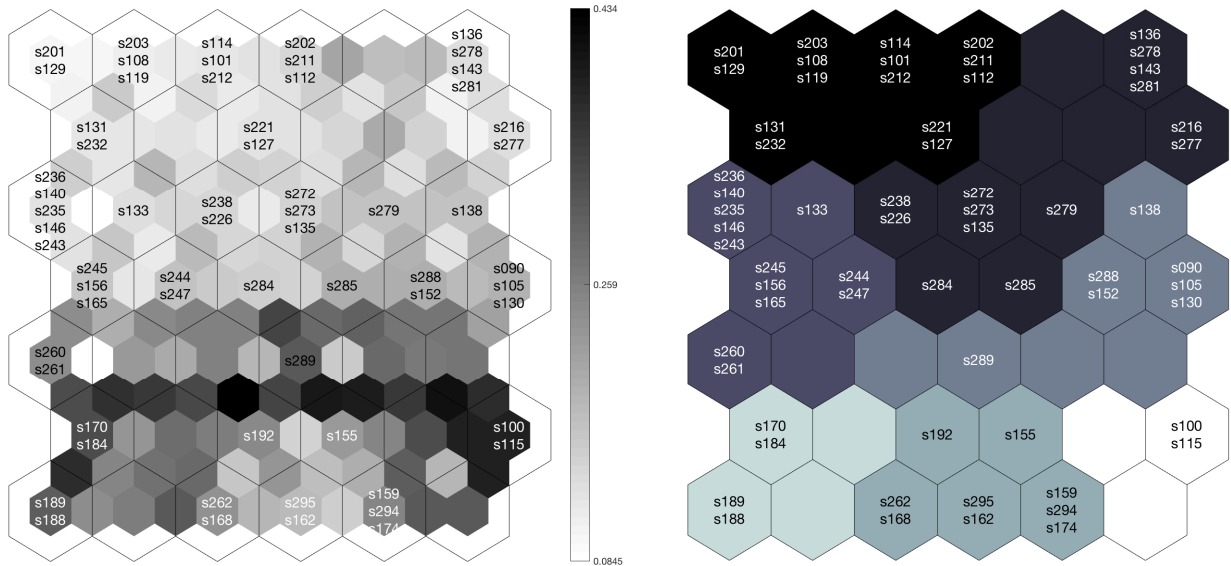
Table 3. Variables and transformations of station characteristics

Variable X	Description	Transformation of variable Y
LAT (°)	Latitude	$Y = (X-Xmin)/(Xmax-Xmin)$
LONG (°)	Longitude	$Y = (X-Xmin)/(Xmax-Xmin)$
EL (m)	Elevation	$Y = X/Xmax$
MAP (mm)	Mean annual rainfall	$Y = X/Xmax$
APM (mm)	Average precipitation in a month	$Y = X/Xmax$
STD (mm)	Standard deviation of annual rainfall	$Y = X/Xmax$

U-matrix 상에서 주변의 다른 지역들과 비교하여 짙은 색으로 둘러싸인 것을 확인할 수 있다. 이는 다른 지점들이 포함된 노드들과 상당히 먼 거리로 떨어져 있음을 의미한다. 이를 통하여 울릉도와 대관령의 경우 지역 구분을 위한 인자들을 검토해볼 필요성이 있고 적절한 지역 구분을 위해서 임의적인 군집이동이 필요한 것으로 판단된다.

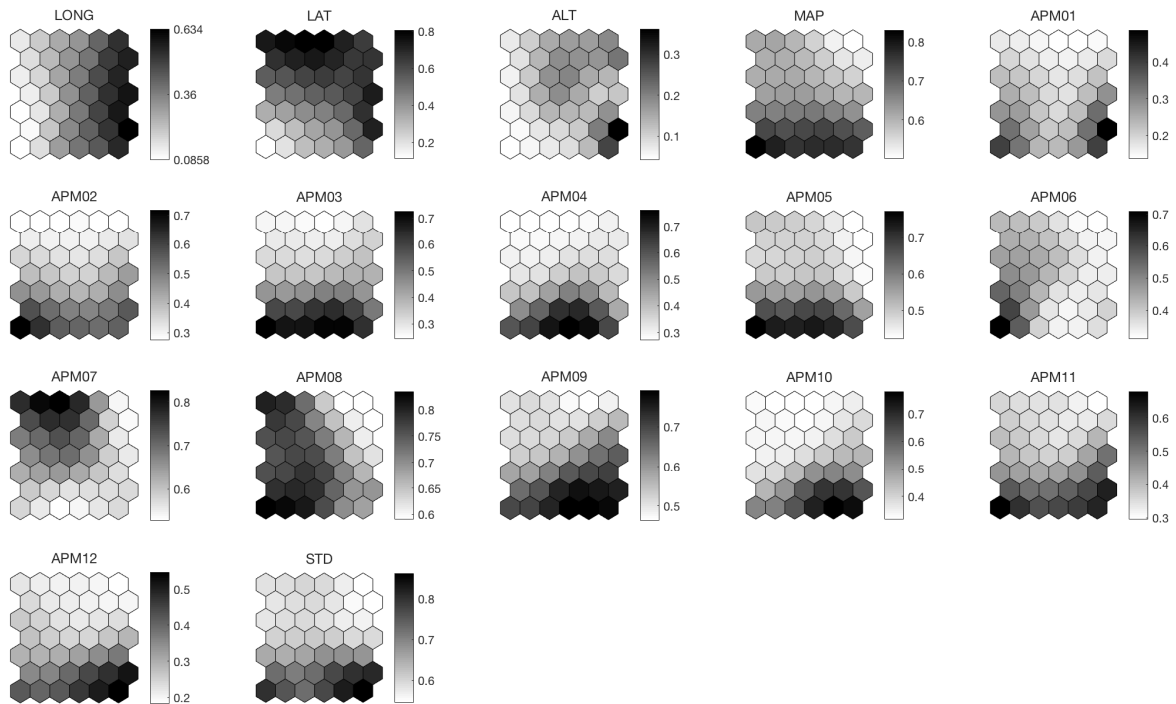
4.1.2 입력 인자별 SOM 패턴 분석

SOM 분석 결과의 타당성을 검토하기 위하여 Fig. 3(c)에서 대상 지점의 물리적인 위치(위도, 경도)를 통해 패턴 분석을 수행해보면, 첫 번째 경도와 관련한 입력 값으로 구분된 지역을 보면 패턴 지도의 우측의 짙은 검은색 부분의 노드들에는 상대적으로 경도 값이 큰, 속초, 강릉, 그리고 태백 등이 포함되



(a) U-matrix results

(b) Clustering results on the SOM map



(c) Results of the patterns for geophysical and precipitation index on the SOM map

Fig. 3. SOM results for 7 × 6 SOM map

어 있는 것을 알 수 있으며, 패턴 지도의 좌측 하단 부분의 밝은 흰색 부분에는 상대적으로 경도 값이 작은 해남, 목포, 그리고 제주도서 지역이 포함되어 있는 것을 확인할 수 있었다. 두 번째 패턴 지도에서는 위도에 따른 지역 구분 결과를 확인한 결과, 좌측 상단의 부분과 우측 하단 부분의 짙은 검은색 부분에서는 강화, 이천, 그리고 속초 등 위도가 높은 지역이 군집화되었고, 좌측 하단의 밝은 흰색 부분에는 제주도서 지역이 군집화됨을 알 수 있다. 이와 같은 결과를 통하여 입력 변수에 따른 패턴 결과가 적절히 이루어졌음을 판단할 수 있다.

패턴 분석 결과 중 특이한 점은 8월부터 10월까지의 월별 강수량의 평균을 살펴보면 각각 좌측에서 우측 방향으로 상대적으로 큰 값을 갖는 지역이 옮겨감을 확인할 수 있다. 이는 8월부터 10월 사이의 한반도의 나타난 강우 현상에 태풍이 영향을 미친 것으로 판단이 된다. Ahn *et al.* (2008)에 의하면 한반도에 영향을 미친 태풍은 7월에서 10월까지 시간에 흐름에 따라 서해와 동해에 발생하는 월별 영향빈도가 극명하게 변하는 것을 확인할 수 있다. 서해의 경우 7월에 영향을 미친 태풍의 발생빈도가 가장 높고 10월로 갈수록 월별 영향빈도가 감소하는 추세를 보인다. 반면, 동해의 경우 6월과 7월에는 영향을 미치는 태풍의 빈도가 적은 반면 8월과 9월로 갈수록 영향 빈도가 증가하는 것을 확인할 수 있다. 이는 본 연구의 결과인 Fig. 3(c)에서 7, 8월의 평균 강수량이 높은 지점들이 SOM 지도상에서 좌측과 좌측 하단에 위치한 서해안과 제주도서 지역에서 두드러지게 나타나며 8월에서 9월로 갈수록 SOM 지도상에서 하단과 우측 하단에 위치한 남해안과 동해안 쪽으로 월별 강수량 큰 지점들이 나타나는 모습을 통해 알 수 있다.

또한 Lee and Choi (2013)의 연구결과인 태풍 내습 빈도 및 일 강수량 80 mm 이상의 극한 강수 현상의 시·공간적 분포가 본 연구에서 나타난 패턴 분석과 매우 흡사하였다. 특히 한반도 남부지역 상륙형 태풍의 경우 7월보다는 8월과 9월에 극명하게 발생 빈도가 증가한 것으로 분석되었다. Fig. 3(c)에서 7월부터 9월까지의 SOM 패턴 분석 지도를 살펴보면, 지도의 하단 부분에 해당하는 남해안 지역의 값이 7월보다 8월과 9월에 집중되어 있음을 확인할 수 있다.

4.2 최종 SOM 군집결과 및 비교분석

Hosking and Wallis (1997)에 의하면 군집분석의 결과는 지역빈도해석을 위해 최종적인 결과로 반드시 적용되는 것보다 분석자의 주관적인 판단에 의해 일부 조정하여 향상된 값을 산정할 것을 추천하고 있다. SOM의 분석결과 및 패턴 분석을 통한 지역적 조정을 통해 결정된 최종 지역 구분 결과는 Fig. 4와 같다.

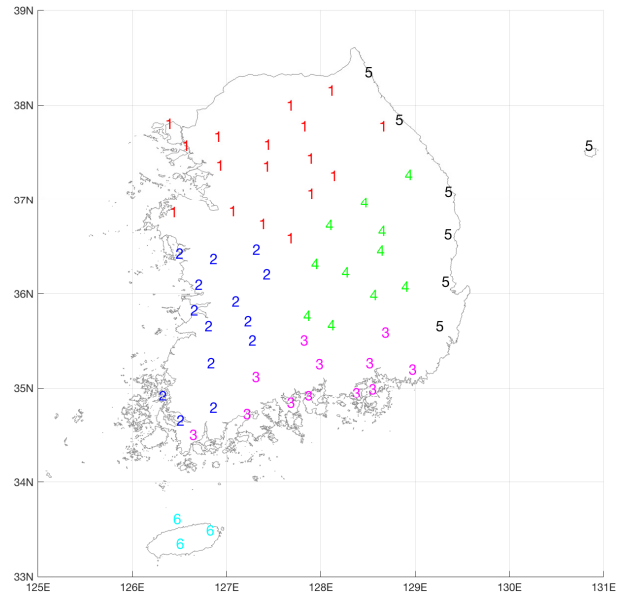


Fig. 4. Cluster result from SOM

최종 구분된 군집은 총 6개의 지역으로 구분되었으며, 서울 지역을 중심으로 한 한강 수계 권역과 서해안 경계를 포함하고 일부 서쪽 남부 내륙지역의 지점을 포함하고 있는 서해안·내륙지역, 남해지역을 포괄하고 있는 남해안 지역, 울릉도를 포함하며 동쪽 해안지역을 포함하는 동해안 지역, 경상도 내륙 지역을 중심으로 군집된 경상도 내륙 지역, 그리고 마지막으로 제주도서 지역이 있다. 이러한 최종 구분된 지역은 Heo *et al.* (2007)과 Nam *et al.* (2008)의 결과와 흡사했다.

Fig. 5는 Heo *et al.* (2007)과 Nam *et al.* (2008)의 지역 구분 결과이다. Heo *et al.* (2007)은 최초 전국을 5대강(한강, 낙동강, 금강, 영산강, 섬진강) 유역과 2개의 해안지역(동해안, 남해안) 지역으로 나누었으며 이후 20개 이상의 많은 지점을 보유하고 있는 지역에 대해서는 지점 간 종속성으로 인한 빈도 해석의 성능 저하를 방지하기 위하여 5개의 유역에 해당하는 지역에 대해서는 다시 군집해석을 수행하여 지역을 좀 더 세분화하였다. 하지만 최초 구분된 유역 중심의 지역 구분 및 도서 지역을 포함한 해안지역의 구분은 본 연구 결과와 큰 차이가 없는 것으로 확인되었다. 또한 Nam *et al.* (2008)이 제시한 지역 구분 결과와는 일부 지점이 종속되는 지역의 차이만 있었을 뿐 매우 흡사한 결과를 확인할 수 있었다. 본 연구에서 최종적으로 구분된 각 지역의 지점 수와 지점명을 정리한 결과는 Table 5와 같다.

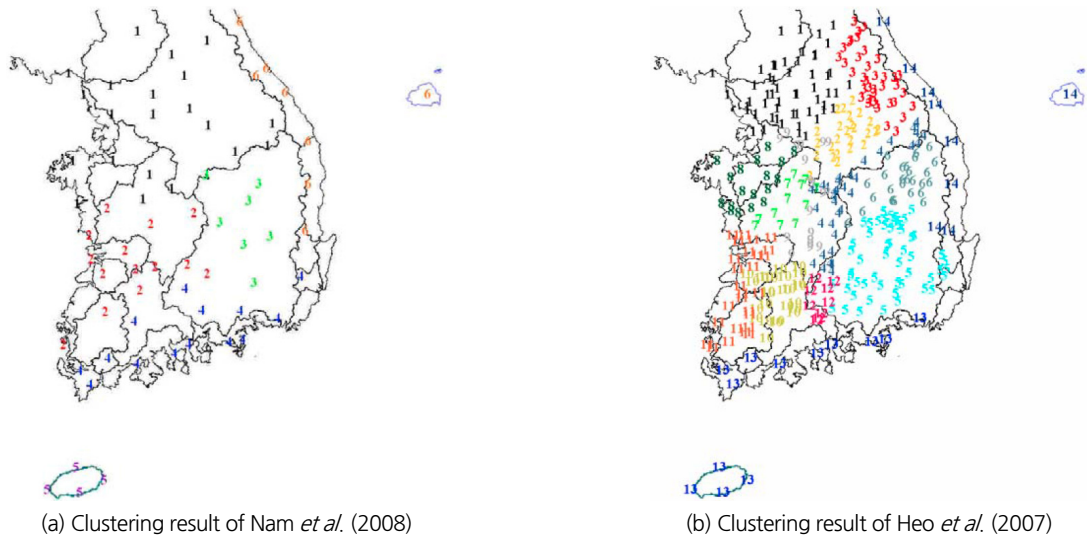


Fig. 5. Identification of homogeneous regions for precipitation of South Korea

Table 5. The number and name of stations for each region

Region	R1	R2	R3	R4	R5	R6
No. of station	17	14	12	11	7	3
Station	Chuncheon Seoul Incheon Wonju Suwon Chungju Seosan Cheongju Ganghwa Yangpyeong Icheon Inje Hongcheon Jecheon Cheonan Boeun Daegwallyeong	Daejeon Gunsan Jeonju Gwangju Mokpo Boryeong Buyeo Geumsan Buan Imsil Jeongeup Namwon Jangheung Haenam	Masan Busan Tongyeong Yoesu Wando Suncheon Jinju Goheung Miryang Sancheong Geoje Namhae	Chupungnyeong Andong Daegu Taebaek Yeongju Mungyeong Uiseong Gumi Yeongcheon Geochang Hapcheon	Sokcho Gangneung Ulleungdo Uljin Pohang Ulsan Yeongdeok	Jeju Seongsan Seogwipo

4.3 동질성 검토

SOM을 활용한 지역 구분 결과를 가지고 강우지역빈도해석의 적용성 검토를 수행하기 위하여 최종 구분된 지역의 동질성을 판단하였다. 적용 자료는 각 지점에서 지속기간 24시간의 연 최대치 강우 자료를 활용하였으며 각 지역별 이질성 척도 검사의 결과는 Table 6과 같다. Table 6의 결과를 보면 총 6개의 지역에서 모두 $H < 1$ 인 경우로 모든 지역에 대해서 동질한 지역으로 판단되었다.

Table 7은 본 연구의 결과를 Nam *et al.* (2008)이 제시한 지역 결과에 맞춰서 정리한 것이다. Nam *et al.* (2008)이 제시한

Table 6. Results of heterogeneity tests for SOM clustering regions (duration: 24-hr)

Cluster		H	
Region	Number of station	Value	Result
1	17	-0.48	Homogeneous
2	14	0.54	Homogeneous
3	12	-0.14	Homogeneous
4	11	-0.02	Homogeneous
5	7	-0.50	Homogeneous
6	3	0.85	Homogeneous

Table 7. Comparison of results between Nam *et al.* (2008) and this study (duration: 24-hr)

Nam <i>et al.</i> (2008)				This study			
Cluster		<i>H</i>		Cluster		<i>H</i>	
Region	Number of station	Value	Result	Region	Number of station	Value	Result
1	21	-1.99	Homogeneous	1	17	-0.48	Homogeneous
2	14	-0.04	Homogeneous	2	14	0.54	Homogeneous
3	7	1.42	Possibly heterogeneous	4	11	-0.02	Homogeneous
4	14	-0.78	Homogeneous	3	12	-0.14	Homogeneous
5	4	-0.40	Homogeneous	6	3	0.85	Homogeneous
6	8	-0.19	Homogeneous	5	7	-0.50	Homogeneous

Table 8. Comparison of heterogeneous test results between Nam *et al.* (2008) and SOM results

Region	Region 3 of Nam <i>et al.</i> (2008)	Region 3 of Nam <i>et al.</i> (2008) with additional data until 2016	Region 4 of SOM
No. of station	7		11
Station	Andong Daegu Yeongju Mungyeong Uiseong Gumi Yeongcheon Miryang		Yeongcheon Daegu Gumi Andong Uiseong Mungyeong Taebaek Yeongju Geochang Hapcheon Chupungnyeong
<i>H</i>	1.42	1.05	-0.02
Results	Possibly heterogeneous		Homogeneous

지역 구분 결과의 경우 전체적으로 동질한 지역으로 구분된 바 있으나, Fig. 5에서 경상도 내륙지역에 해당하는 3번 지역에서는 *H*값이 1.42로 이질 가능성이 있는 것으로 나타났다. 그러나 본 연구에서 구분된 경상도 내륙 지역에 해당하는 4번 지역은 *H*값이 -0.02로 동질지역으로 구분되었다. 하지만 두 연구의 지역 구분에 따른 동질성 검토 결과는 표본자료의 크기가 다르므로 동일한 비교라고 할 수 없다. 동질성 검토 결과의 차이가 각 지점별 표본자료의 추가에 의한 결과인지 알아보기 위하여 Nam *et al.* (2008)에 의하여 구분된 지역 3을 대상으로 2016년까지의 표본을 추가하여 동질성 척도 검사를 수행하였다. 비교된 결과는 Table 8과 같다. 2016년까지의 표본을 더했을 경우 값이 1.42에서 1.05로 낮아지긴 하였으나 여전히 $H > 1$ 이므로 이질 가능성이 있는 지역으로 나타났다. Nam *et al.* (2008)에 의해 제시된 지역과 본 연구에서 구분된 지역의 동질성 척도 검사 결과를 비교하여 본 결과 표본자료의 길이

보다는 구분된 지역 내의 지점들과 연관이 있다는 것을 확인하였다.

5. 결론

지역구분은 지역빈도해석을 수행하는 데 있어서 가장 중요한 절차 중 하나이다. 본 연구는 지역 구분을 하기 위하여 SOM 기법을 적용하여 군집분석을 수행하고, 구분된 지역을 바탕으로 강우 지역빈도해석을 수행할 수 있는지 알아보기 위하여 동질성 검토를 통해 SOM 기법의 적용성을 검토하였다. 분석된 주요 연구 결과를 요약하면 다음과 같다.

- 1) 강우 지역빈도해석을 위하여 SOM을 활용하여 군집분석을 수행할 경우 안정적인 지역 구분 결과를 얻을 수 있을

뿐만 아니라 지역 구분 인자와 지역 구성과의 관계를 파악할 수 있었다.

- 2) 입력 인자에 따른 패턴 분석을 살펴보면, 물리적인 지형정보와 월별 강우량의 영향이 지역 구분에 큰 영향을 미치는 것으로 나타났다. 또한 SOM 패턴 분석을 통하여 입력자료가 각 지역별로 어떠한 영향을 미치는지 파악할 수 있었고 이를 통하여 강우 현상에 대한 시·공간적 분석이 가능함을 확인하였다.
- 3) SOM을 이용하여 군집분석을 수행한 결과, 선행 연구와 비슷한 지역 구분 결과를 얻을 수 있었다. 하지만 이질성 척도를 활용하여 동질성 검토를 해본 결과 SOM으로 구분된 지역이 선행 연구로 구분된 지역보다 좀 더 동질한 지역으로 군집됨을 확인하였다.
- 4) 선행 연구와 유사하게 구분된 지역에 대해서 지역 동질성 평가가 다르게 나왔을 경우 표본의 크기에 따른 영향보다는 군집분석 기법에 따라 구분된 지역을 구성하는 지점 구성의 영향이 큰 것으로 나타났다.

향후 추가적인 연구 수행 시 한반도 강우현상에 크게 영향을 미치는 태풍과 장마와 관련된 입력 인자를 추가한다면 남한 지역의 강우 지역 구분과 관련하여 좀 더 유의미한 결과를 얻을 수 있을 것으로 판단된다. 마지막으로 본 연구를 통하여 지역빈도해석 수행 시, 다양한 방법을 이용한 지역 구분의 필요성을 확인하였으며, 이를 위해서는 수문 자료에 대한 군집 분석 연구가 지속되어야 할 것이다.

감사의 글

본 연구는 정부(행정안전부)의 재원으로 재난안전기술개발사업단의 지원을 받아 수행된 연구임[MOIS-재난-2015-03].

References

- Ahn, S.-H., Kim, B.-J., Lee, S.-L., and Kim, H.-K. (2008). "The characteristics of disaster by track of typhoon affecting the Korean peninsula." *Journal of The Korean Society of Hazard Mitigation*, Vol. 8, No. 3, pp. 29-36.
- Arribas-Bel, D., Nijkamp, P., and Scholten, H. (2011). "Multi-dimensional urban sprawl in Europe: a self-organizing map approach." *Computers, Environment and Urban Systems*, Vol. 35, No. 4, pp. 263-275.
- Bärring, L. (1988). "Regionalization of daily rainfall in Kenya by means of common factor analysis." *International Journal of Climatology*, Vol. 8, No. 4, pp. 371-389.
- Burn, D. H. (1989). "Cluster analysis as applied to regional flood frequency." *Journal of Water Resources Planning and Management*, Vol. 115, No. 5, pp. 567-582.
- Cunnane, C. (1989). *Statistical distributions for flood frequency analysis*. Operational Hydrology Report, No. 33.
- Darlymple, T. (1960). "Flood frequency analyses." *Water Supply Paper*, U.S. Geological Survey, Reston, VA., p. 80.
- Dinpashoh, Y., Fakheri-Fard, A., Moghaddam, M., Jahanbakhsh, S., and Mirmia, M. (2004). "Selection of variables for the purpose of regionalization of Iran's precipitation climate using multivariate methods." *Journal of Hydrology*, Vol. 297, No. 1, pp. 109-123.
- Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. (1996). "From data mining to knowledge discovery in databases." *AI Magazine*, Vol. 17, No. 3, p. 37.
- Greis, N. P., and Wood, E. F. (1981). "Regional flood frequency estimation and network design." *Water Resources Research*, Vol. 17, No. 4, pp. 1167-1177.
- Guttman, N. B. (1993). "The use of L-moments in the determination of regional precipitation climates." *Journal of Climate*, Vol. 6, No. 12, pp. 2309-2325.
- Heo, J. H. (2016). *Statistical hydrology*. Gumi Press, Seoul, Korea.
- Heo, J. H., Boes, D. C., and Salas, J. D. (1990). *Regional flood frequency modeling and estimation*. Water Resources Papers, No.101, Colorado State University, Fort Collins, CO.
- Heo, J.-H., Lee, Y. S., Shin, H., and Kim, K.-D. (2007). "Application of regional rainfall frequency analysis in South Korea (I): rainfall quantile estimation." *Journal of the Korean Society of Civil Engineers B*, Vol. 27, No. 2B, pp. 101-111.
- Hewitson, B., and Crane, R. (2002). "Self-organizing maps: applications to synoptic climatology." *Climate Research*, Vol. 22, No. 1, pp. 13-26.
- Hosking, J., and Wallis, J., (1997). *Regional frequency analysis: an approach based on L-moments*. Cambridge University Press, NY.
- Hosking, J., Wallis, J. R., and Wood, E. F. (1985). "An appraisal of the regional flood frequency procedure in the UK flood studies report." *Hydrological Sciences Journal*, Vol. 30, No. 1, pp. 85-109.
- Jain, A. K., Murty, M. N., and Flynn, P. J. (1999). "Data clustering: a review." *ACM Computing Surveys (CSUR)*, Vol. 31, No. 3, pp. 264-323.
- 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*, Vol. 5, No. 4, pp. 293-303.
- Jeong, K.-S., Kim, D.-K., Shin, H.-S., Yoon, J.-D., Kim, H.-W., and Joo, G.-J. (2011). "Impact of summer rainfall on the seasonal water quality variation (chlorophyll a) in the regulated Nakdong river." *KSCE Journal of Civil Engineering*, Vol. 15, No. 6, pp. 983-994.

- Jun, S., and Choi, Y. (2013). "The applicability of the self-organizing map method for synoptic climatology studies over the Republic of Korea." *Journal of Climate Research*, Vol. 8, No. 1, pp. 1-11.
- Kiang, M. Y., Kulkarni, U. R., Goul, M., Philippakis, A., Chi, R., and Turban, E. (1997). "Improving the effectiveness of self-organizing map networks using a circular Kohonen layer, System Sciences, 1997." *Proceedings of the Thirtieth Hawaii International Conference on. IEEE*, pp. 521-529.
- Kiang, M., Kulkarni, U., and St Louis, R. (2001). "Circular/wrap-around self-organizing map networks: an empirical study in clustering and classification." *Journal of the Operational Research Society*, Vol. 52, No. 1, pp. 93-101.
- Kim, H.-U., Shim, J.-K., and Choi, B.-C. (2017). "Analysis on rainfall pattern during typhoons using self-organizing map." *Korean Society of Hazard Mitigation*, Vol. 17, No. 4, pp. 55-64.
- Kohonen, O., and Hauta-Kasari, M. (2008). "Distance measures in the training phase of self-organizing map for color histogram generation in spectral image retrieval." *Journal of Imaging Science and Technology*, Vol. 52, No. 2, pp. 20201-1 - 20201-11.
- Kohonen, T. (1998). "The self-organizing map." *Neurocomputing*, Vol. 21, No. 1, pp. 1-6.
- Kohonen, T. (2001). *Self-organizing maps*. 3rd Edition, Springer, Berlin.
- Kohonen, T. (2007). "The self-organizing map." *Proceedings of the IEEE*, Vol. 9, pp. 1464-1479.
- Lee, S. W., and Choi, G. Y. (2013). "Spatio-temporal patterns of extreme precipitation events by typhoons across the Republic of Korea." *Journal of The Korean Association of Regional Geographers*, Vol. 19, No. 3, pp. 384-400.
- Lee, Y. S., Heo, J.-H., Nam, W. S., and Kim, K.-D. (2007). "Application of regional rainfall frequency analysis in South Korea (II): Monte Carlo simulation and determination of appropriate method." *Journal of the Korean Society of Civil Engineers B*, Vol. 27, No. 2B, pp. 113-123.
- Lettenmaier, D. P., and Potter, K. W. (1985). "Testing flood frequency estimation methods using a regional flood generation model." *Water Resources Research*, Vol. 21, No. 12, pp. 1903-1914.
- Liu, Y., Weisberg, R. H., and Mooers, C. N. (2006). "Performance evaluation of the self-organizing map for feature extraction." *Journal of Geophysical Research: Oceans*, Vol. 111, No. C5.
- Lu, H.-C., Chang, C.-L., and Hsieh, J.-C. (2006). "Classification of PM 10 distributions in Taiwan." *Atmospheric Environment*, Vol. 40, No. 8, pp. 1452-1463.
- Mallants, D., and Feyen, J. (1990). "Defining homogeneous precipitation regions by means of principal components analysis." *Journal of Applied Meteorology*, Vol. 29, No. 9, pp. 892-901.
- Mangiameli, P., Chen, S. K., and West, D. (1996). "A comparison of SOM neural network and hierarchical clustering methods." *European Journal of Operational Research*, Vol. 93, No. 2, pp. 402-417.
- Moon, S. W. (2006). *A study on the performance of clustering methods using neural networks*. Master dissertation, Yonsei University, Seoul, South Korea.
- Nam, W. S., Kim, T. S., Shin, J.-Y., and Heo, J.-H. (2008). "Regional rainfall frequency analysis by multivariate techniques." *Journal of Korea Water Resources Association*, Vol. 41, No. 5, pp. 517-525.
- Nam, W., Shin, H., Jung, Y., Joo, K., and Heo, J. H. (2015). "Delineation of the climatic rainfall regions of South Korea based on a multivariate analysis and regional rainfall frequency analyses." *International Journal of Climatology*, Vol. 35, No. 5, pp. 777-793.
- Reusch, D. B., Alley, R. B., and Hewitson, B. C. (2007). "North atlantic climate variability from a self-organizing map perspective." *Journal of Geophysical Research: Atmospheres*, Vol. 112, No. D2.
- Smithers, J., and Schulze, R. (2001). "A methodology for the estimation of short duration design storms in South Africa using a regional approach based on L-moments." *Journal of Hydrology*, Vol. 241, No. 1, pp. 42-52.
- Stedinger, J. R. (1983). "Estimating a regional flood frequency distribution." *Water Resources Research*, Vol. 19, No. 2, pp. 503-510.
- Su, M.-C., Liu, T.-K., and Chang, H.-T. (2002). "Improving the self-organizing feature map algorithm using an efficient initialization scheme." *Tamkang Journal of Science and Engineering*, Vol. 5, No. 1, pp. 35-48.
- Tian, J., Azarian, M. H., and Pecht, M. (2014). "Anomaly detection using self-organizing maps-based K-nearest neighbor algorithm." *Proceedings of the European Conference of the Prognostics and Health Management Society*.