

# 다변량 분석을 통한 우리나라 가뭄의 지역적 특성 분석

## Regional Characterization Analysis of Drought in Korea Using Multivariate Analyses

유지영\*, 최민하\*\*, 김태웅\*\*\*

Ji Young Yoo, Minha Choi, Tae-Woong Kim

### 요 지

우리나라 가뭄의 지역적 특성은 수문학적으로 동질한 지역의 구분 결과에 따라 달라진다. 지역의 구분에는 가뭄에 영향을 미치는 다양한 변수들이 사용될 수 있다. 가뭄을 특징짓는 요소로서 지속기간, 심도, 이외의 통계적 특성들이 있으며, 이 변수들을 정보화하여 변수의 유형을 구분지어 모든 변수들을 요약된 정보로 활용하여 가뭄의 특성을 구분할 수 있다. 본 연구에서는 우리나라 기상청 강우자료 75개 관측지점 중 30년 미만의 강우기록이 있는 17개의 지점을 제외한 58개 강우 관측 지점을 대상으로 가뭄지수(SPI)를 산정하여 가뭄사상의 특성을 정량화 과정으로 남한지역 가뭄특성을 분류하였다. SPSS를 활용한 다변량 분석기법인 주성분 분석(principal component analysis)을 통해 가뭄특성인자의 상관관계가 높은 변수들을 조합하여 그 변수들 중 가뭄정보를 가능한 많이 함축하고 있는 새로운 특성 변수를 만들어 내었으며, 선정된 변수들을 바탕으로 요인분석(factor analysis)의 직각회전 방식(Varimax)을 이용하여 변수들의 표준화를 통해 가뭄특성요인을 찾아내었다. 이를 통해 지역간 동질성을 파악하여 K-means기법을 적용하여 군집해석(clustering analysis)을 실시하였다.

**핵심용어:** 표준강수지수(SPI), 가뭄특성분석, 다변량분석, 군집분석

## 1. 서론

우리나라 가뭄은 최근 들어 자연재해로서 이슈화 되고 있으며, 가뭄으로 인한 피해도 빈번해지고 있어 기존 가뭄해석을 바탕으로 한 지속적인 연구가 필요함을 알 수 있다. 가뭄은 일반적으로 RUN의 개념으로(Yevjevich, 1967) 이용하여, 주어진 절단수준(truncation level)에 대한 가뭄의 지속기간, 심도, 발생간격 등을 분석하고, 이와 관련하여 가뭄재현기간 및 지속특성을 분석하다.

그러나 우리나라는 수문자료의 관측기간이 짧고, 지점 내 충분한 수의 자료를 확보하기가 어렵다는 문제가 있어 우리나라 강우량의 자료만으로 가뭄의 특성을 파악하기에는 한계가 있다. 이와 같은 문제를 해결하기 위해서 본 연구에서는 가뭄의 특성을 다변량 분석 기법을 활용하여 가뭄요인을 분류하고, 이를 통해 동일한 가뭄특성을 지닌 지역을 군집화 하여 지역해석을 실시하였다. 이는 향후 가뭄빈도해석과 지역해석을 동시에 고려하는 이변수해석을 위한 사전연구 단계로서 기존의 가뭄해석보다 신뢰성이 높은 가뭄평가를 가능하게 하는데 목적이 있다.

## 2. 가뭄특성인자를 이용한 다변량 분석

\* 정회원 · 한양대학교 대학원 토목공학과 석박사통합과정 · E-mail : [7924pooh@hanyang.ac.kr](mailto:7924pooh@hanyang.ac.kr)

\*\* 정회원 · 한양대학교 건설환경공학과 조교수 · E-mail : [minha1218@gmail.com](mailto:minha1218@gmail.com)

\*\*\* 정회원 · 교신저자 · 한양대학교 건설환경시스템공학전공 조교수 · E-mail : [twkim72@hanyang.ac.kr](mailto:twkim72@hanyang.ac.kr)

## 2.1 가뭄특성인자 구축

### 2.1.1 가뭄지수 산정

본 연구에서는 전국의 기상청 산하 72개 강우 관측지점의 강우자료 중 30년 이하의 자료를 제외한 58개 지점의 강우자료를 이용하여 3, 6, 9, 12개월 이동평균강우량을 바탕으로 한 표준강수지수(SPI)를 산정하였다. SPI는 Makee 등(1993)이 가뭄이 상대적으로 물 수요에 비해 물의 부족을 유발하는 강수량의 감소에 의하여 시작된다는 것에 착안하여 가뭄을 정의하기 위해 개발되었다. 그림 1은 강릉지점 3, 6, 9, 12개월 표준강수지수(SPI)이며, 가뭄의 심도를 SPI를 이용하여 나타내기 위해 산정하였다.

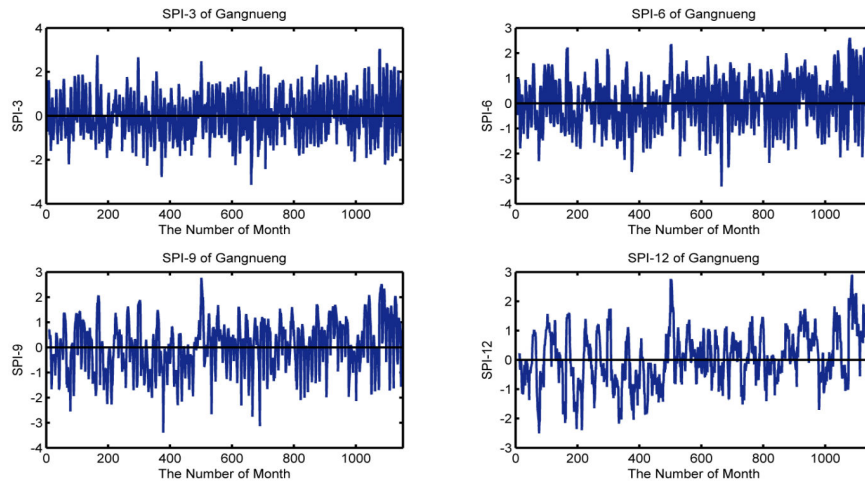


그림 1. 강릉지점 3, 6, 9, 12개월 SPI

### 2.1.2 가뭄특성인자 산정

본 연구에서 Yevjevich(1967)에 의해 제안된 연속이론과 Mckee 등(1993)의 제안에 의하여 SPI의 -1이하의 연속된 사건을 가뭄사상으로 정의하였다. 이러한 각각의 가뭄 사상을 통해 가뭄의 두 가지 주요 특성인 지속기간(duration)과 심도(severity)를 구할 수 있다. 가뭄의 지속기간은 절단수준인 -1이하의 SPI가 지속된 기간을 의미하며, 심도는 가뭄사상의 SPI의 합으로 산정되었다. 그림 2는 강릉지점의 가뭄지속기간, 그림 3은 강릉지점의 가뭄심도이다.

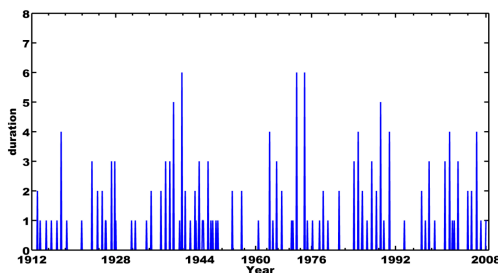


그림 2. 강릉지점 가뭄지속기간

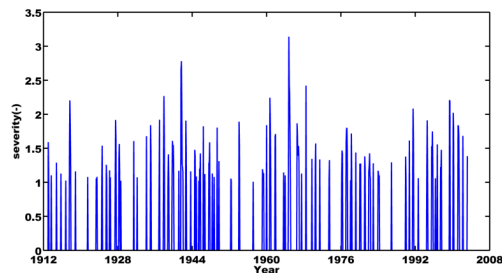


그림 3. 강릉지점 가뭄심도

본 연구에서는 58개 강우 관측지점별 3, 6, 9, 12개월 월 평균 SPI와 가뭄특성(가뭄지속기간, 심도, 평균심도)을 추출하여 가뭄사상의 특성을 정량화하였으며, 총 60개의 가뭄특성인자를 생성하였다.

## 2.2 주성분 분석기법을 통한 요인분석

주성분분석(principal component analysis)은 요인분석(factor analysis)의 예비단계로서 다수의 지표를 통합한 종합적인 지표를 작성하는 경우 사용하며, 이는 변수의 수가  $p$ 개( $x_1, x_2, \dots, x_p$ ), 관측대상의 수가  $n$ 개인 다변량 데이터가 있을 경우, 이 데이터를  $p$ 개보다 적은  $m$ 개의 새로운 변수  $z_1, z_2, \dots, z_m$ 을 만들어내서 원래의 변수와 새로운 변수 사이의 상관성을 이용하여 요인(factor)이라 불리는 변수 내의 새로운 변수를 도출함으로써 이 요인들이 가지고 있는 특성은 전체 자료의 특성을 최대한 설명하는 기법이다(남우성 등, 2008).

본 연구에서 기초통계분석인 SPSS를 활용하여 요인분석을 실시하였으며, 가뭄특성인자의 수(60개)와 같이 변수가 많은 경우 군집해석의 효율성 저하 문제를 해결하기 위해 수행하였다. 요인분석 수행에 필요한 회전 방법으로는 기후자료에 대해 가장 적합한 회전방법으로 알려진 varimax를 사용하였으며, 표 1은 주성분 분석을 통해 7개의 주성분요소를 추출한 결과이며, 이 7가지 요소는 전체 60개의 변량을 95.429% 설명해주는 새로운 변수로 생성되었다.

표 1. 주성분분석 요소추출 결과

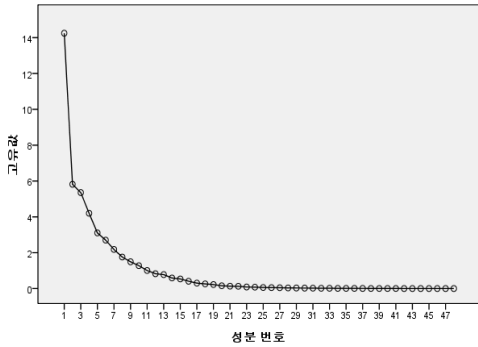


그림 4. 고유치(값)의 변화과정

성분	초기고유값			회전제곱합 적재값		
	변수(수)	분산(%)	누적(%)	변수(수)	분산(%)	누적(%)
1	22.869	38.115	38.115	19.591	32.652	32.652
2	16.624	27.707	65.822	14.515	24.192	56.844
3	7.294	12.157	77.979	10.041	16.736	73.580
4	3.600	5.999	83.978	3.881	6.468	80.047
5	2.781	4.635	88.613	3.612	6.021	86.068
6	2.694	4.490	93.103	3.429	5.716	91.784
7	1.396	2.326	95.429	2.187	3.646	95.429

그림 4는 요인별로 고유치가 변하는 과정이며, 이를 살펴본 후 요인이 8개인 지점부터 고유치의 변화폭이 작아지는 것으로 판단되어 최종적으로 주요인의 개수는 7개로 결정했다. 또한, 일반적으로 주성분의 채택과정에 있어서는 누적기여율이 70%~80% 이상이면 채택이 가능하다(노형진, 2001).

각 요인은 어떤 변수의 특성을 대표할 수 있으며, 주요인은 어떠한 가뭄특성변수에 영향을 받았는지 분류하였다. 요인 1은 3, 6, 9, 12개월 이동평균강우량을 바탕으로 생성된 월평균 SPI-6, SPI-7, SPI-8, SPI-9로서 여름철의 SPI변수 값에 영향을 받은 요인이며, 요소 2는 SPI-1, SPI-2, SPI-11, SPI-12로 겨울철의 SPI변수 값에 영향을 받은 요인이라고 말할 수 있다. 요소 3은 12개월 이동평균강우량으로 생성한 SPI변수 값에 관련된 요인으로 나타나며, 요소 4는 3개월 이동평균강우량으로 분석된 가뭄의 특성치와 관련된 요인으로 나타났다.

### 3. 다변량 분석기법에 의한 가뭄지역 구분

#### 3.1 가뭄 동질지역 분석

우리나라 58개 지점의 가뭄의 특성인자(60개 변량)를 요인분석 실시 후, 7개의 주요인으로 선택하여 지역별 가뭄의 특성을 분석하기 위해 군집분석을 실시하였다. 군집분석은 Hosking and Wallis(1997)에 의하면 확률수문량 추정 시 정확도가 높아지다가 지역 내 지점수가 늘어나 20개 정도가 넘어서면 지점간에 종속성 생겨 정확도가 저하하는 현상이 생긴다고 한다. 그러므로 확률수문량을 추정하기 전에 지역의 군집화를 통해 군집 내 동질성을 파악하고 군집간의 이질성을 파악할 수 있다면 신뢰성이 높은 연구를 수행할 수 있을 것

이다.

본 연구에서는 K-means 기법을 이용한 군집분석을 실시하였으며, 이는 비계층적 군집분석법으로 이전의 계층적 군집분석보다 군집화 속도가 빨라서 군집화할 대상이 다수인 경우에 신속하게 처리할 수 있는 방법이다(노형진, 2001). 한 번의 군집이 결합될 때마다 각 군집별로 그 군집의 평균을 중심으로 군집 내 관측대상들간의 유클리디안 거리의 합을 구하는데, 이 값을 군집화 과정에서 발생하는 오류라고 할 수 있다(노형진, 2001). 이 값이 낮을수록 군집화에 따르는 오류가 낮은 것이고, 그 결과 관측대상들이 보다 타당성 있게 군집화되었다고 한다. K-means 군집분석법에서는 각 군집화 과정에서 발생하는 오류를 최소화하는 방향으로 군집화를 계속하게 하며, 결국 오류가 발생하지 않는 군집화 단계에서 군집은 종료하게 된다.

### 3.2 군집분석 결과

58개의 강우지점을 대상으로 가뭄요인분석 결과를 이용한 군집분석은 군집 값의 변화가 6번 만에 수렴하여 군집화가 종료되었으며, 최종적으로 7개의 군집을 형성하였다. 표 2와 그림 5는 지점별로 군집된 결과를 나타내며, 총 58개 지점이 7개로 분류되었다. 이 중 울릉도는 개별분류가 되어 군집의 수는 6개로 나누었으며, 표 3은 군집별로 각각 군집중심으로 부터의 거리를 계산한 기초통계량결과로 군집 내의 통계적 특성을 파악할 수 있다.

표 2. 가뭄특성에 따른 지역별 군집분석 결과

Region	1	2	3	4	5	6
No. of Site	15	15	9	8	5	5
Sites	Cheongju Chuncheon Chungju Chupungnyeong Ganghwa Hongcheon Icheon Incheon Inje Jecheon Mungyeong Seoul Suwon Wonju Yangpyeong	Boeun Boryeong Buan Buyeo Cheonan Daejeon Geochang Geumsan Haenam Jangheung Jeongeup Mokpo Seosan Suncheon Wando	Daegu Gunsan Gwangju Hapcheon Imsil Jeonju Namwon Sancheong Uiseong	Busan Geoje Goheung Jinju Miryang Namhae Tongyeong Yeosu	Gumi Pohang Ulsan Yeongcheon Yeongju	Daegwallyeong Gangneung Sokcho Ulsin Yeongdeok

### 4. 결 론

본 연구에서는 전국 58개 강우 관측지점의 자료를 이용하여 표준강수지수(SPI)를 산정하여 가뭄을 분석하였고, 이를 통해 얻은 가뭄특성치(60개)를 이용하여 다변량 분석 기법을 적용하여 가뭄지역을 구분하였다. 다변량 분석으로 사용한 요인분석은 회전방법 중 varimax를 사용하여 주성분분석을 실시하였고, 이를 통해 자료의 정보를 최소화 하면서 상관관계분석 후 고유치를 이용하여 주성분요인을 추출하였다. 선정된 7개의 가뭄주성분요인은 K-means 군집분석법에서 지역을 구분하는 척도로 사용되었으며, 그 결과는 허준행 등(2007)의 지역 구분 결과와 유사한 형태를 보여주었다. 가뭄지역해석을 기반으로 향후에는 재현기간에 따른

가뭄분석을 통하여, 가뭄빈도해석과 지역해석을 동시에 고려하는 이변수해석을 수행할 것이고, 그 결과는 기존의 가뭄해석보다 신뢰성이 높은 가뭄평가를 가능하게 할 것이다.

표 3. 군집별 기초통계량 분석 결과

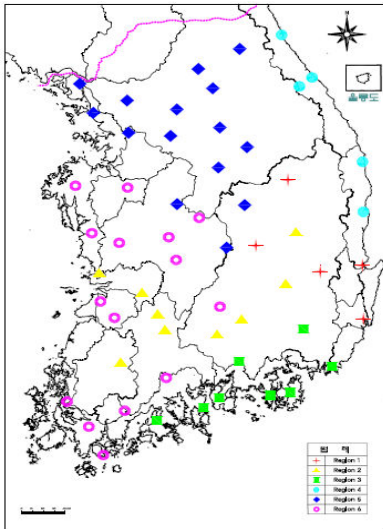


그림 5. 남한지역의 가뭄특성 지역분류

Region	Mean	No. of Site	Standard dev.	Variance
1	1.408	15	0.510	0.260
2	1.442	15	0.637	0.406
3	1.343	9	0.290	0.084
4	1.460	8	0.633	0.400
5	1.420	5	0.557	0.31
6	1.362	5	0.523	0.274
Total	1.39	57	0.549	0.301

### 감 사 의 글

본 연구는 2009년 국가가뭄재해정보시스템 시범개발의 재원으로 소방방재청의 지원을 받아 수행되었습니다.

### 참 고 문 헌

1. 남우성, 김태순, 신주영, 허준행(2008), 다변량 분석 기법을 활용한 강우 지역빈도해석, *한국수자원학회 논문집*, 제41권 제5호, pp. 517-525.
2. 노형진(2001), *한글 SPSS 10.0에 의한 조사방법 및 통계분석*, 형설출판사.
3. 허준행, 이영석, 신흥준, 김경덕 (2007), 우리나라 강우자료의 지역빈도해석 적용성 연구(I): 확률강우량 산정, *대한토목학회 논문집*, 제27권 제2B호, pp. 101-111.
4. Hosking, J.R.M. and Wallis, J.R. (1997). *Regional Frequency Analysis*, Cambridge University Press.
5. Mckee, T. B., Doesken, N. J., and Klest J. (1993). *The Relationship of Drought Frequency and Duration of Time scales*. Department of Atmospheric Science, Colorado State University, Fort Collins, Colorado.
6. Yevjevich, V. M. (1967). "An objective approach to definitions and investigations of continental hydrologic droughts" *Hydrology Paper 23*, Colorado State Univ., Fort Collins, Colorado.