

군집분석을 이용한 우리나라 가뭄특성의 공간적 분석

Spatial Analysis of Drought Characteristics in Korea Using Cluster Analysis

유지영* / 최민하** / 김태웅***
Yoo, Ji Young / Choi, Minha / Kim, Tae-Woong

Abstract

Regional frequency analysis is often used to overcome the limitation of point frequency analysis to estimate probability rainfall depths. However, point frequency analysis is still used in drought analyses. This study proposed a practical method to categorize the homogeneous regions of drought characteristics for the analyses of regional characteristics of droughts in Korea. Using rainfall data from 58 observation stations managed by the Korea Meteorological Administration, this study calculated drought attributes, i.e., mean drought indices for various durations using the Standardized Precipitation Index (SPI) and drought severities expressed by durations, depth, and intensity. The drought attributes provided useful information for categorizing stations into the hydrological homogeneous regions. This study introduced a cluster analysis with K-means techniques to group observation stations. The cluster analysis grouped observation stations into 6 regions in Korea. The data in the hydrological homogeneous region would be used in spatial analysis of drought characteristics and drought regional frequency analysis.

Keywords : Standardized Precipitation Index, Drought Characteristics, Clustering Analysis

요 지

최근에는 확률강우량을 산정할 경우 지점빈도해석의 단점을 보완한 지역빈도해석법이 자주 실무에 적용되고 있으나, 가뭄에 관련한 연구에서는 대부분 아직까지 지점자료를 이용한 가뭄분석을 실시하고 있다. 본 연구에서는 가뭄의 지역적 특성 분석을 실시하기 위하여 필요한 동질한 가뭄특성을 지닌 지역을 구분하는 연구를 수행하였다. 본 연구에서는 기상청 강우관측 지점자료 중 30년 이상의 강우자료를 보유한 58개의 관측지점을 대상으로 표준강수지수(SPD)를 산정하여 가뭄의 심도, 지속기간, 강도, 발생빈도 등과 같은 가뭄특성인자를 생성하였다. 가뭄특성인자는 수문학적으로 동질한 특성을 지닌 지역을 구분하는데 중요한 정보를 제공한다. 본 연구에서는 다양한 가뭄특성인자를 효율적으로 활용하여 K-means 기법을 적용한 군집분석을 실시하여 동질한 가뭄특성을 지닌 지역을 6개 지역으로 구분하였다. 이러한 지역구분은 가뭄 특성의 공간적 해석을 가능하게 할 수 있고, 지점빈도 해석의 단점을 보완하는 지역빈도 해석도 가능하게 할 수 있다.

핵심용어 : 표준강수지수, 가뭄특성, 군집분석

* 정회원, 한양대학교 대학원 건설환경공학과 석·박사과정
Graduate Student, Dept. of Civil and Environmental Engineering, Hanyang University, Seoul 133-791, Korea
** 정회원, 한양대학교 건설환경공학과 조교수
Assistant Professor, Dept. of Civil and Environmental Engineering, Hanyang University, Seoul 133-791, Korea
*** 교신저자, 정회원, 한양대학교 건설환경시스템공학전공 조교수
Corresponding Author, Assistant Professor, Dept. of Civil and Environmental System Engineering, Hanyang University, Ansan 426-791, Korea
(e-mail: twkim72@hanyang.ac.kr)

1. 서 론

현재 전 세계적으로 기후변화로 인한 온난화 현상은 극심해 지고 있으며, 이러한 온난화 현상은 우리나라에서 가뭄의 발생 가능성을 높이고 있다. 국립기상연구소(2009)에 의하면 우리나라 연평균 기온, 연평균 최저기온, 연평균 최저기온의 변화율은 각각 $0.24^{\circ}\text{C}/10\text{년}$, $0.11^{\circ}\text{C}/10\text{년}$, $0.37^{\circ}\text{C}/10\text{년}$ 으로 연평균 최저기온의 상승경향이 뚜렷하다. 이와 같은 온난화 현상의 영향으로 우리나라는 강수량의 감소와 더불어 증발량도 크게 증가하게 되었으며, 결국 가뭄과 관련한 피해도 더욱 가중될 것이다.

일반적으로 가뭄에 대한 연구를 수행할 경우, 관측지점별 자료를 이용하여 가뭄현상을 정량화하고, 이를 통해 가뭄특성을 평가한다. 즉 평균수준(혹은 절단수준)을 선정하고 가뭄의 지속기간, 심도, 발생간격 등을 정의한 후 이를 시계열 분석하여 가뭄의 특성을 분석하는 것이다(Yevjevich, 1967). 하지만, 가뭄은 홍수와 달리 진행속도가 느리고, 시·공간적으로 매우 비균질하게 발생하므로(Salas *et al.*, 2005), 가뭄을 분석하고 평가하는 데 있어 지역적 특성을 고려하는 분석이 필요하다.

경기개발연구원(2004)과 김보경 등(2006)은 EOF 해석 및 Kriging 기법을 이용하여 가뭄의 공간적인 분포 특성을 고려하여 가뭄심도-면적-지속기간 곡선을 유도하였으며, 유철상 등(2004)은 포아송 과정(Poisson Process)을 이용한 지점분석을 공간화한 가뭄공간분석을 수행하였으며, 김휘린 등(2004)은 자료계열 특성에 따른 지역가뭄빈도해석을 실시하여 강우지속기간-재현기간-갈수우량 관계를 고려하여 2001년 가뭄의 심도와 지역적 범위를 평가하였다. 또한, 경민수 등(2007)은 표준강수지수(SPI)의 지점간 유사성을 기준으로 하여 군집분석을 실시하였으며, EOF 분석 결과를 공간으로 확장하기 위하여 Kriging 기법을 적용하였다. 이 결과 가뭄심도-가뭄면적-가뭄지속기간 곡선을 작성하였다. 이외에도 가뭄의 공간분석에 대한 다양한 연구들이 존재하나, 공통적인 연구의 한계점을 지니고 있다. 먼저 현재까지 대부분의 가뭄의 공간분포 해석에 관련한 연구에서는 가뭄의 시간적인 패턴을 파악할 수 없으며, 또한 각 지점에서 극한사상이 동시에 발생할 확률은 0에 가깝다는 점과 가뭄해석에서 지속기간이 길어지면 사상의 발생이 점점 사라지게 되어 매개변수를 추정할 수 없게 된다는 약점을 갖는다(유철상 등, 2004).

본 연구에서는 기존의 연구결과를 바탕으로 지점별 강우자료를 이용하여 가뭄속성을 정량화하였으며, 이러한 지점별 가뭄속성자료를 이용하여 동질한 가뭄특성을

지닌 지역을 구분하였다. Mallants and Feyen (1990)는 일 강우량을 이용하여 주성분 분석을 실시하여 동질한 강우 지역을 구분한 바 있으며, Zhang and Hall (2004)은 다양한 군집해석 기법을 지역홍수빈도해석에 적용한 바 있다. 또한 Dinpashoh *et al.* (2004)은 강우의 특성이 동질한 지역을 주성분 분석과 요인분석을 활용하여 구분한 바 있다. 본 연구에서는 국외선행연구에서 다양하게 이용되어진 다변량분석기법 중 주성분 분석(Principal Component Analysis)과 요인분석(Factor Analysis)을 우리나라 가뭄분석에 활용하여 군집분석을 수행하였다. 여기서 생성된 각각의 군집 내 지점들은 상호간의 가뭄특성이 동질하다는 가정이 통계적으로 유의함으로, 군집으로 형성되어진 지역 내 모든 지점의 가뭄특성자료(가뭄지속기간, 심도, 발생간격 등)를 동시에 이용하여 지역별 가뭄분석을 실시할 경우, 하나의 지점에 해당하는 강우자료를 이용한 가뭄해석보다 비교적 훨씬 많은 가뭄특성자료를 활용할 수 있다.

2. 가뭄지수를 이용한 가뭄특성인자 구축

2.1 가뭄지수 선정

가뭄을 평가하기 위해 주로 사용하는 방법은 가뭄지수를 선정하는 것이다. 가뭄지수는 강우, 용설, 하천유량, 지하수 등 물공급 관련 자료를 이용하여 가뭄 상황을 파악할 수 있는 하나의 지표라 할 수 있다. 일반적으로 널리 이용되는 지수는 표준강수지수(Standardized Precipitation Index, SPI), 팔머가뭄지수(Palmer Drought Severity Index, PDSI), 지표수공급지수(Surface Water Supply Index, SWSI) 등이 있다. 이들 지수들은 적용 대상 지역의 수문 및 기상 특성, 수자원 공급 시설들을 고려하여 각 지수들의 적용성을 파악하여 적절한 지수를 이용하거나 통합하여 이용한다. 본 연구에서는 다양한 가뭄지수 중 SPI(표준강수지수)를 이용하였다. SPI는 가뭄지수를 선정하는 데 있어 다른 지수들과 달리 강우자료만을 이용하기에 비교적 제약조건이 적으며, 또한 각 시간단위에 따라 강수 부족량을 측정하여 개개의 강수량이 가뭄에 미치는 영향을 산정할 수 있기 때문이다. 이렇게 특정 시간단위로 산정된 SPI는 각 시간단위에 따라 여러 분야에 사용될 수 있다. 예를 들면, 단기간의 시간 축척은 농업 분야에 사용되며, 비교적 장기간의 시간 축척은 수자원 공급관리 등에서 사용되어진다(장연규 등, 2006).

2.1.1 SPI 이론

Mckee *et al.* (1993)는 강수량이 부족하면 용수 공급

원인 지하수량, 적설량, 저수지 저류량, 토양함유 수분량, 하천유출량 등에 영향을 미치므로, 강수량의 부족분을 이용하여 가뭄을 정량화하여 지수로 표현하였다. 즉, SPI는 특정한 시간에 대한 계산 단위를 3, 6, 9, 12개월 등과 같이 설정하고, 시간단위별로 강수 부족량을 계산하여 개개의 용수공급원이 가뭄에 미치는 영향을 산정하는 방법이다. SPI는 선형적으로 갈수부족에 비례하고 확률, 누가 강수 평균 백분율, 그리고 누가 강수 부족량을 고려한다. 이렇게 산정된 SPI는 확률분포함수를 도입하였기 때문에 현재의 가뭄뿐만 아니라 가뭄이 끝나기 위해 필요한 강수의 확률을 결정하는 것을 가능하게 한다.

2.1.2 SPI 산정 방법 및 결과

SPI를 산정하기 위해서는 우선 시간단위별 누가강수 시계열을 구성하여야 하며, 이는 이동누가에 의한 방법으로 월 강수량을 시간단위에 따라 연속적으로 중첩하여 산정한다. 즉, 각 월을 기준으로 하여 시간단위에 해당하는 누가강수 시계열을 구성하여, 이를 누가된 월수로 나누어 이동 평균강수계열을 산정한다. 월별 평균강수계열의 확률분포함수를 추정하여, 누가확률값에 해당

하는 표준정규변량을 추정하면, 이것이 각 지속기간에 해당하는 SPI가 된다.

본 연구에서는 Gamma 분포(Eq. (1))를 이용하여 기상청 강우관측지점의 58개 지점에 해당하는 SPI를 산정하였으며, Fig. 1은 58개 지점 (Fig. 6 참조) 중 부산 지점의 3, 6, 9, 12개월 이동평균강수에 따른 표준강수 지수(SPI)를 예시적으로 나타낸 것이다. 대부분 지점의 SPI는 -3에서 3까지의 범위를 갖고 있으며, 산정된 SPI를 이용하여 가뭄속성인자를 산정하였다.

$$\text{확률밀도함수: } f(x) = \frac{1}{\alpha\Gamma(\beta)} \left(\frac{x}{\alpha}\right)^{\beta-1} \exp\left(-\frac{x}{\alpha}\right) \quad (1a)$$

누적분포함수:

$$F(x) = \int_0^x f(x)dx = \int_0^x \frac{1}{\alpha\Gamma(\beta)} \left(\frac{x}{\alpha}\right)^{\beta-1} \exp\left(-\frac{x}{\alpha}\right) dx \quad (1b)$$

여기서, α 는 축척매개변수(scale parameter), β 는 형상매개변수(shape parameter), $\Gamma(\beta)$ 는 감마함수(Gamma function)이다.

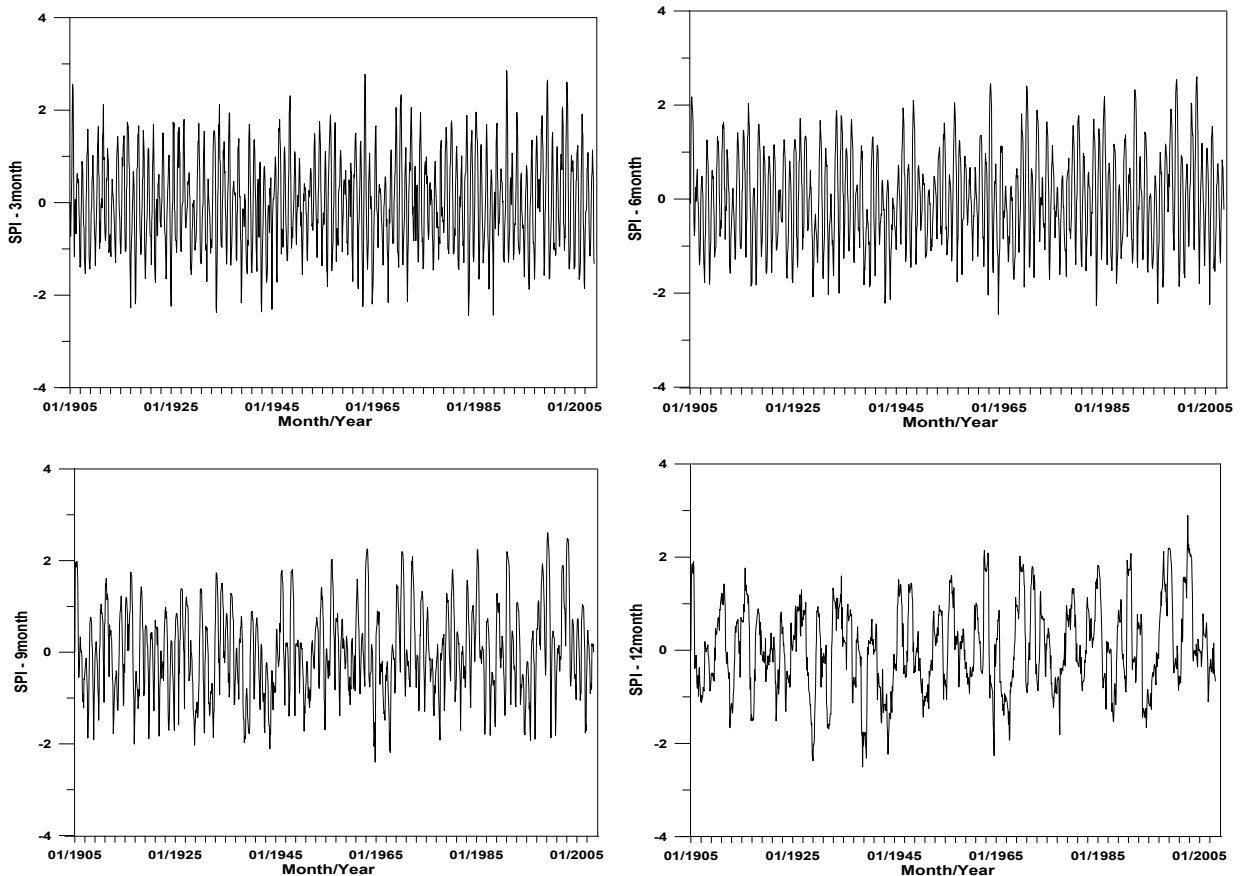


Fig. 1. The Standardized Precipitation Index at Busan Station (1905–2008)

Table 1. SPI and Moisture Categories

| SPI Values | Drought Category |
|----------------|------------------|
| ≥ 2.00 | Extreme wet |
| 1.99 to 1.50 | Severe wet |
| 1.49 to 1.00 | Moderate wet |
| 0.99 to -0.99 | Normal |
| -1.00 to -1.49 | Moderate drought |
| -1.50 to -1.99 | Severe drought |
| ≤ -2.00 | Extreme drought |

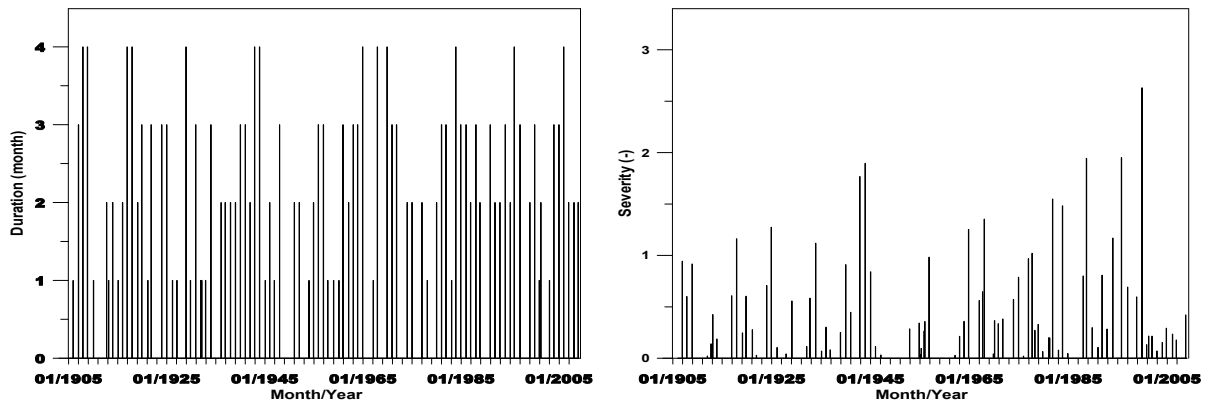


Fig. 2. Drought Duration and Severity (3-month) at Busan Station (1905-2008)

2.2 가뭄속성인자 산정

McKee *et al.* (1993)은 SPI의 결과로부터 얻어지는 습윤 및 가뭄정도를 정의하기 위한 Table 1과 같은 SPI 가뭄 분류를 이용하여 임의의 시간규모에 적용될 수 있는 가뭄사상에 대한 기준을 명시하였다.

가뭄지수가 특정 평균수준(혹은 절단수준) 이하의 값이 지속적으로 나타날 때 가뭄이 발생한다. 그러므로 각 가뭄사상은 시작과 끝으로 확정되는 지속기간(Duration)을 가지며, 이때 지속되는 월에 대한 가뭄심도(Severity)를 갖게 된다. 본 연구에서는 Yevjevich (1967)에 의해 제안된 연속이론(Run Theory)과 McKee *et al.* (1993)의 제안에 의한 절단수준(SPI = -1)에 의해 가뭄사상을 정의하였다. 이러한 각각의 가뭄 사상을 통해 가뭄의 두 가지 주요 특성인 지속기간(Duration)과 심도(Severity)를 산정하였다. 가뭄의 지속기간은 절단수준인 -1 이하의 SPI가 지속된 기간을 의미하며, 심도는 가뭄사상의 SPI의 합으로 산정되었다. Fig. 2는 부산지점의 3개월 표준강수지수(SPI-3month)로 가뭄사상을 추출한 결과를 예시적으로 보여주는 것으로 왼쪽 그림은 가뭄지속기간이며, 오른쪽 그림은 가뭄심도

를 나타낸다.

Sen (1976)과 Mathier *et al.* (1992)는 일반적으로 관측된 가뭄 지속기간과 심도를 각각 분리하여 분포시켜 가뭄의 특성을 얻어내었다. 본 연구에서는 이러한 가뭄의 특성치(지속기간, 심도)가 가뭄의 특성을 정량화 하는데 있어서, 상호간에 어느 정도의 상관성을 갖고 있는지 알아보기 위해, Eq. (2)를 이용하여 가뭄 지속기간과 심도의 상관계수를 산정하였다.

$$\rho_k = \frac{COV(X_t, Y_t)}{\sqrt{VAR(X_t)VAR(Y_t)}} \quad (2)$$

여기서, $COV(X_t, Y_t)$ 는 두 변수간의 공분산이며, $VAR(X_t)$ 와 $VAR(Y_t)$ 는 각 변수의 분산이다. 가뭄의 지속기간과 심도의 상관계수($\rho_{D,S}$)는 부산 0.9381, 목포 0.9728, 인천 0.9683 등으로 우리나라 대부분의 지점에서 높은 상관관계를 갖고 있음을 확인할 수 있다.

본 연구에서는 58개 강우 관측지점별 SPI(3, 6, 9, 12month)의 월별 평균값과 가뭄특성치(가뭄지속기간, 심도, 강도)의 평균값을 산정하여, 각 지점별 60개의 가뭄속성인자(Fig. 3)를 생성하였다.

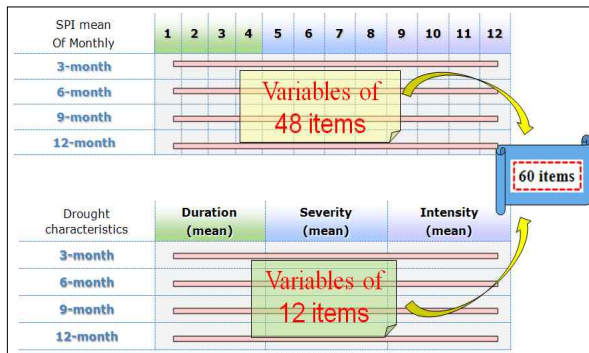


Fig. 3. Drought Attributes of Total 60 Items

3. 다변량 분석기법을 활용한 가뭄 지역 분석

우리나라는 강우관측기간이 100년 이하로 강우관측 지점별 관측자료 수가 매우 적으며, 관측지점마다 차이가 심하다. 그러므로 지점별 강우자료를 이용하여 수문학적 해석을 실시할 경우 불확실성은 커지게 된다. 지역빈도해석은 위와 같은 단점을 갖는 지점빈도해석을 보완한 방법이다. 지역빈도해석에서 가장 중요한 단계는 동질지역을 구분하는 것이며, 동질지역을 구분하는 방법으로는 정해진 기준이 따로 존재하는 것이 아니기 때문에 다양한 기법과 변수들을 활용되어지고 있다. 본 연구에서는 Dinpashoh *et al.* (2004)이 제안한 다변량 분석기법 중 주성분 분석과 요인분석을 활용하여 동질 지역을 구분하였다.

본 연구에서는 지역별 가뭄의 특성을 한 요소로만 구분할 수 없으며, 월별 SPI 특성과 가뭄지속기간, 가뭄 심도 등과 같은 복합적인 구분요소가 존재하므로 상호 의존적인 관계를 고려하여 가뭄지역의 동질성을 구분해야 한다. 따라서 본 연구에서는 통계 소프트웨어인 SPSS (Statistical Package for the Social Science)를 활용하여 다변량분석(Multivariate Analysis of Variance)을 수행하였다. 앞서 산정한 가뭄속성인자들 간의 상관관계가 높은 변수들을 조합하여, 그 변수들 중 가뭄정보를 가능한 많이 함축하고 있는 새로운 특성의 변수를 산출하였다. 이는 앞에서 산정한 가뭄지수로 만들어진 60개의 가뭄속성인자를 주성분분석(Principal Component Analysis, PCA) 기법을 적용하여 48개 변수로 함축한 후, 48개의 함축된 가뭄특성인자를 대상으로 요인분석(Factor Analysis)을 실시하여 전체 가뭄속성변동의 95.429 %를 설명해 주는 새로운 가뭄특성요소 7개를 추출한 것이다. 이렇게 추출된 7개의 가뭄특성요소를 바탕으로 K-means 기법을 이용한 군집분석(Clustering Analysis)을 실시하여 우리나라 가뭄의 지

역간 동질성을 파악하였다.

3.1 주성분분석을 통한 요인분석

3.1.1 주성분분석 이론 및 결과

주성분분석(PCA)은 변수간의 상관행렬로부터 상관관계가 높은 여러 개의 변수들을 조합하여 그 변수들의 정보를 가능한 많이 함축하고 있는 새로운 변수를 만들어내기 위한 다변량 통계분석 기법이다. 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 만들어서 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). 주성분분석의 일반적인 절차는 Fig. 4와 같으며, 본 연구에서는 주성분분석을 통해 처음 60개의 가뭄속성인자를 설명할 수 있는 48개의 변수로 집약되었으며, 이렇게 새롭게 집약된 48개의 변수는 요인분석에서 입력변수로 바로 활용되어진다.

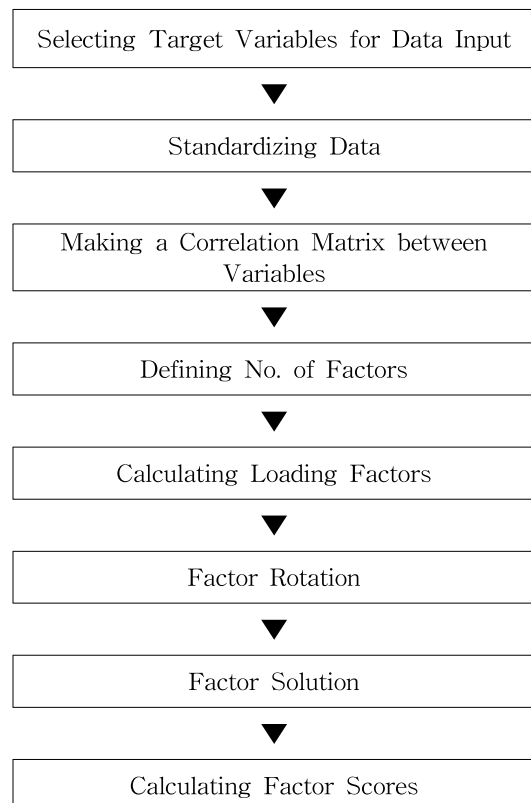


Fig. 4. General Procedure of PCA

3.1.2 요인분석 이론 및 결과

주성분분석을 통해 얻은 48개의 속성변수를 이용하여 요인분석(Factor Analysis)을 실시하였으며, 요인분석을 실시함으로써 변수가 많아 군집분석에서 발생할 수 있는 효율성 문제를 해결할 수 있다. 요인분석은 기존의 변수들의 상관성을 이용하여 변수 내의 공통적인 새로운 변수를 도출하여 요인(Factor)이라고 불리는 공통적인 새로운 변수를 생성하게 된다.

본 연구에서는 SPSS를 활용하여 요인분석을 실시하였으며, 특정변수간의 상호관계를 통해서 잠재적인 구조를 탐색하는 과정에 있어 요인부하량 행렬이나 요인 구조 행렬의 추정치 등으로부터 요인분석이 실시되었다. 그러나 변수들이 여러 요인에 대하여 비슷한 요인부하량을 나타낼 경우에 변수들이 어느 요인에 속하는지를 분류하기가 어렵다. 따라서 변수들의 요인부하량이 어느 한 요인에 높게 나타나도록 하기 위해서 요인축을 회전시킨다(노형진, 2001). 본 연구에서는 기후자료에 대해 가장 적합한 회전방법으로 알려진 Varimax Normal Rotation (Puvaneswaran, 1990; White *et al.*, 1991)를 사용하여 요인축을 회전시켰다. 주성분분석으로 생성된 48개의 새로운 가뭄속성변수가 요인분석을 실시함으로써 Fig. 5와 같은 요인별 고유치가 생성되었다. 이 각 요인에 따르는 고유치의 수를 살펴보면 8번째 요인에서부터 고유치 수의 변화폭이 급격히 줄어들었음을 확인할 수 있다. 결국 8번째 요인부터는 가뭄의 특성인자를 설명해주는 요인으로 판단하기가 어렵기 때문에 Table 2과 같이 가뭄속성인자 48개의 변량을 95.4 % 설명해주는 7개의 요인을 주요인으로 채택하였다.

일반적으로 요인분석 후 주요인의 채택과정에 있어

서는 누적기여율이 70~80 % 이상이면 채택이 가능하며(노형진, 2001), 여기서 추출된 주요인들은 어떤 가뭄 특성변수에서 영향을 받아 나타난 결과인지를 설명해 준다. 요인 1은 3, 6, 9, 12-month의 SPI 중 6월, 7월, 8월, 9월의 SPI 값으로서, 시간단위(단·장기) 구분이 없는 여름철의 SPI 특성과 관련이 있는 요인이며, 요인 2는 3, 6, 9, 12-month의 SPI 중 1월, 2월, 11월, 12월의 SPI 값으로서, 시간단위(단·장기) 구분이 없는 겨울철의 SPI 특성과 관련이 있는 요인이라고 할 수 있다. 요인 3은 12-month의 SPI 값으로서 시간단위가 장기간인 가뭄과 관련된 요인으로 나타나며, 요인 4는 3-month의 SPI 값으로서 시간단위가 단기간인 가뭄의 특성치(지속기간, 심도, 강도)와 관련된 요인으로 나타났다. 요인 5는 3, 9-month의 SPI 중 4월, 5월, 10월의 SPI 값으로서 여름철이 다가오기 전·후의 달의 SPI의 특성을 나

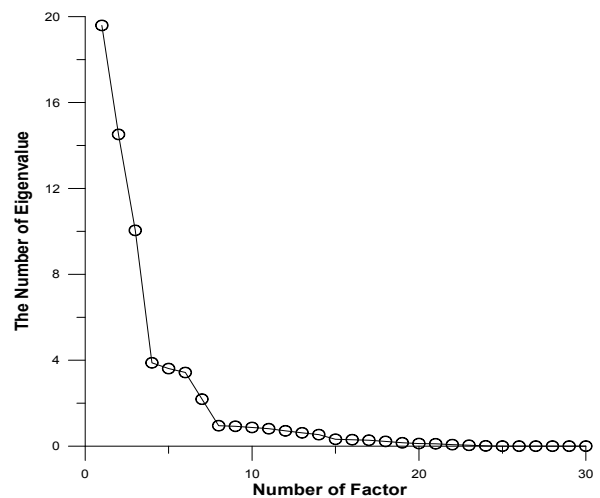


Fig. 5. Transition Process of Eigenvalue

Table 2. Results from Factor Analysis

| Factor | Initial Eigenvalue | | | Factor Loading Eigenvalue of Verimax | | |
|--------|---|-----------------|-------------------|---|-----------------|-------------------|
| | Eigenvalue (the number of articles) | Variance (%) | Cumulation (%) | Eigenvalue (the number of articles) | Variance (%) | Cumulation (%) |
| 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 |

타내는 요인으로 나타나고, 요인 6은 9, 12-month의 SPI 값으로 추출된 가뭄사상의 특성치 중 가뭄심도와 관련된 요인으로 나타나며, 요인 7은 장기적으로 지속되는 가뭄인 12-month의 SPI 값 중 9, 10, 11, 12월의 겨울철 SPI 값과 관련된 요인이다.

3.2 우리나라 가뭄의 지역적 특성분석

3.2.1 군집분석 이론 및 적용

우리나라 58개 강우관측지점(Fig. 6 참고)의 가뭄의 속성인자를 주성분 분석과 요인분석을 실시하여, 가뭄 특성의 주요인 7개를 채택하였으며, 동일한 가뭄특성을 나타내는 지역을 구분하기 위하여 가뭄특성의 주요인 7개를 사용하여 군집분석을 실시하였다. Hosking and Wallis (1997)에 의하면 확률수문량 추정 시 정확도가 높아지다가 지역 내 지점수가 늘어나 20개 정도가 넘어서면 지점 간에 종속성이 생겨서 정확도가 저하되는 현상이 생긴다. 그러므로 확률수문량을 추정하기 전에 지역의 군집화를 통해 군집 내 동질성을 파악하고 군집간의 이질성을 파악할 수 있다면 신뢰성이 높은 연구를 수행할 수 있을 것이다.

남우성 등(2008)은 지역빈도해석을 통한 확률강우량을 산정하기 위해 Fuzzy c-mean 기법을 이용하여 군집화를 실시하였고, 오태석 등(2007)은 강우관측소의 위도, 경도, 및 지속기간의 연최대 시간강수량의 평균자료를 이용하여 지역빈도해석을 하기 위해 계층적 군집화 방법 중 하나인 Centroid Neighbor 기법을 이용한 군집분석을 실시하였다. 차석빈 등(2008)에 의하면 계층적 군집화 방법과 비계층적 군집화 방법은 두 방법 모두 장단점이 있기 때문에 우월함을 따질 수는 없으나, 계층적 군집화 방법은 비계층적 군집화 방법에 비해 표본수의 크기에 제한이 있어 극단치(Outliers)의 영향을 많이 받는다는 단점이 있다.

본 연구에서는 극단치의 영향이 적은 비계층적 군집 방법 중 하나인 K-means 기법을 이용하여 군집분석을 실시하였다. 여기서 K는 미리 정하는 군집의 수이며, K-means 기법은 계층적인 군집화의 결과에 의하여 미리 군집의 수를 정해야 하며, 군집의 중심(Cluster Center)도 정해야 한다. 군집의 중심을 잘 정해야 정확한 군집의 결과를 얻을 수 있으며, 여기서 한번 군집이 결합될 때마다 각 군집별로 그 군집의 평균을 중심으로 군집 내 관측대상들 간의 유클리드 거리의 합을 구하는데, 이 값을 군집화 과정에서 발생하는 오류라고 할 수 있다. 이 값이 낮아질수록 군집화에 따르는 오류가 낮은 것이고, 그 결과 관측대상들이 보다 타당성 있게

군집화 되었다고 한다(박우창 등, 2003). K-means 군집 분석법에서는 각 군집화 과정에서 발생하는 오류를 최소화하는 방향으로 군집화를 계속하게 하며, 결국 오류가 발생하지 않는 군집화 단계에서 군집은 종료하게 된다. 본 연구에서는 6번의 반복계산 후 군집이 종료되었으며, 최종 7개의 군집을 형성하였다.

3.2.2 우리나라 가뭄의 지역구분 결과

58개의 강우지점을 대상으로 가뭄요인분석 결과를 이용하여 군집분석을 실시한 결과 최종적으로 7개의 군집을 형성하였다. Fig. 6은 본 연구의 군집결과로서 동일한 가뭄의 특성을 지닌 지역을 구분한 것이다.

군집1은 충주, 춘천, 청주, 강화, 홍천, 인천 등 총 15개의 지점으로 이루어졌으며, 지리적으로는 태백산맥을 기준으로 서쪽부분에 위치하여 경기도와 충청북도 북쪽부분, 강원도 일부를 포함한다. 대체적으로 한강유역에 포함되어 있다. 군집2는 보은, 보령, 부안, 부여, 천안, 대전 등 총 15개의 지점으로 이루어 졌으며, 지리적으로 충청남도과 전라도의 서해안 지역에 해당되고, 유역으로는 금강, 금강서해, 삼교천, 동진강에 해당된다.

군집3은 대구, 군산, 광주, 합천, 임실, 전주 등 총 9개

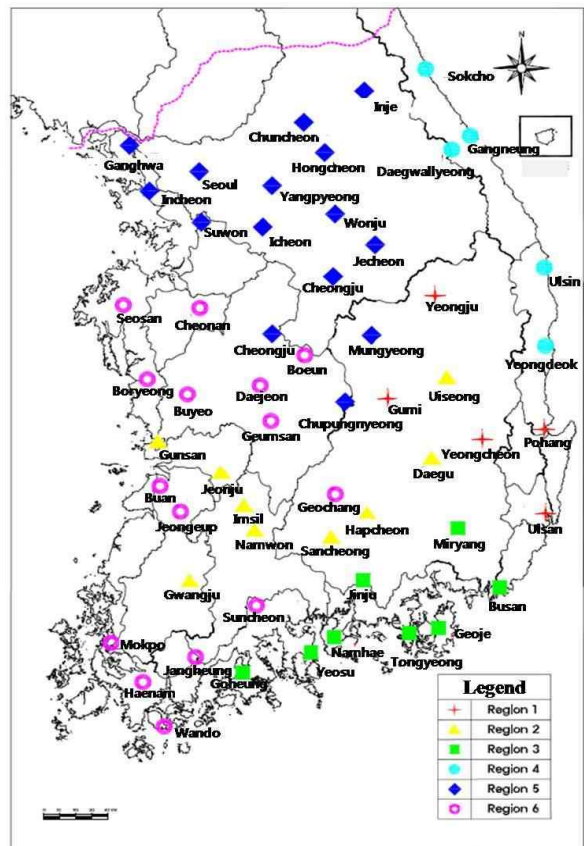


Fig. 6. Cluster Result of Drought Homogeneous Regions

의 지점으로 나타났으며, 지리적으로는 전라도와 경상도의 일부에 걸쳐져 있고, 대유역으로는 구분짓기가 어려우며, 군집4는 부산, 거제, 고흥, 진주, 밀양, 남해 등 총 8개의 지점으로 구성되었으며, 지리적으로는 남해 해안선에 인접한 지역으로 경상남도, 전라남도의 해안 지역에 해당되고, 유역으로는 낙동강 남해와 섬진강 남해, 회야·수영유역에 포함된다. 군집5는 구미, 포항, 울산, 영천, 영주 등 5개의 지점으로 이루어 졌으며, 지리적으로는 태백산맥을 기준으로 서쪽부분의 경상도 일부를 포함하고 있으며, 낙동강 유역의 일부와 형산강유역에 포함된다. 군집6은 대관령, 강릉, 속초, 울진, 영덕으로 군집이 형성되었고, 지리적으로는 태백산맥을 기준으로 동쪽에 해당하는 해안지역으로 강원도와 경상도 일부에 해당되며, 유역으로는 한강동해와 낙동강동해에 포함된다. 반면 울릉도는 개별분류가 되어 최종 군집의 수는 6개로 결정하였다.

허준행 등(2007)은 우리나라의 지역별 강수의 특성을 크게 5대강을 중심으로 한 지역과 남해안, 동해안으로 구분하였고, 5대강을 세분화하여 14개의 지역으로 구분하였다. 본 연구의 가뭄의 지역구분 결과인 Fig. 6은 허준행 등(2007)의 강수특성에 따른 지역구분결과와 유사한 형태로 나타났음을 확인할 수 있다. 그 결과 지점별 강수자료를 이용하여 우리나라 가뭄의 특성을 파악하고

그에 따른 군집을 실시한 결과가 신뢰성이 있음을 확인할 수 있었다.

Table 3은 군집별로 각각 군집중심으로 부터의 거리를 계산한 기초통계량으로서 군집 내의 통계적 특성을 파악할 수 있다. 최종 군집결과인 6개 군집의 SPI 특성은 Table 4와 같은 기초통계량을 가지고 있다. 군집4는 SPI 평균값이 가장 작게 나타났으며, 표준편차와 분산도 다른 군집에 비해 크게 산출되어 상대적으로 가뭄발생 가능성이 가장 크게 해석할 수 있으며, 반면 군집2는 SPI 평균값이 가장 크게 나타나고, 표준편차와 분산도 다른 군집에 비해 작게 산출되어 상대적 가뭄발생 가능성이 가장 적은 지역으로 해석하였다.

또한, 시계열 값의 군집별 시간에 대한 의존성을 살펴보기 위해 마코프 체인 모형을 이용하여 상태전환확률(State Transition Probability)을 산정하였다. 본 연구에서는 지체-2 마코프 체인 모형을 사용하였으며, Run 이론을 이용하여 시간 t에 대하여 SPI 값을 가뭄(절단수준 아래)과 가뭄 상태가 아닌 습윤(절단수준 위)의 2가지 상태로 나누어 Eq. (3)에 의해 상태전환확률을 산정하였다.

$$P_{DD} = P(SPI_{t+1} \leq -1 | SPI_t \leq -1) \quad (3a)$$

Table 3. Elementary Statistics (Distance from Group Centroid) for Each Group

| Group | Number of stations | Mean | Standard dev. | Variance |
|-------|--------------------|-------|---------------|----------|
| 1 | 15 | 1.408 | 0.510 | 0.260 |
| 2 | 15 | 1.442 | 0.637 | 0.406 |
| 3 | 9 | 1.343 | 0.290 | 0.084 |
| 4 | 8 | 1.460 | 0.633 | 0.400 |
| 5 | 5 | 1.420 | 0.557 | 0.31 |
| 6 | 5 | 1.362 | 0.523 | 0.274 |
| Total | 57 | 1.39 | 0.549 | 0.301 |

Table 4. Elementary Statistics of SPI (3month) in Each Group

| Group | Mean | Standard dev. | Variance | Minimum value |
|-------|-------|---------------|----------|---------------|
| 1 | 0.016 | 0.960 | 0.921 | -2.389 |
| 2 | 0.026 | 0.942 | 0.887 | -2.153 |
| 3 | 0.019 | 0.956 | 0.914 | -2.879 |
| 4 | 0.000 | 1.000 | 1.001 | -2.814 |
| 5 | 0.006 | 0.988 | 0.976 | -3.234 |
| 6 | 0.015 | 0.971 | 0.942 | -3.140 |

Table 5. Probability of Drought Duration in Each Group

| Group | Probability by Duration | | | | | | | |
|-------|-------------------------|---------|---------|---------|---------|---------|---------|---------|
| | 1-month | 2-month | 3-month | 4-month | 5-month | 6-month | 7-month | 8-month |
| 1 | 0.468 | 0.249 | 0.133 | 0.071 | 0.038 | 0.020 | 0.011 | 0.006 |
| 2 | 0.502 | 0.250 | 0.125 | 0.062 | 0.031 | 0.015 | 0.008 | 0.004 |
| 3 | 0.472 | 0.249 | 0.132 | 0.070 | 0.037 | 0.019 | 0.010 | 0.005 |
| 4 | 0.421 | 0.244 | 0.141 | 0.082 | 0.047 | 0.027 | 0.016 | 0.009 |
| 5 | 0.417 | 0.243 | 0.142 | 0.083 | 0.048 | 0.028 | 0.016 | 0.010 |
| 6 | 0.467 | 0.249 | 0.133 | 0.071 | 0.038 | 0.020 | 0.011 | 0.006 |

$$P_{DW} = P(SPI_{t+1} > -1 | SPI_t \leq -1) \quad (3b)$$

상태전환확률의 합은 1.0이 되어야 하므로 $P_{DD} + P_{DW} = 1$ 이 된다. 군집별 가뭄 지속기간의 분포 값 Eq. (4)에 의해 산정되어 Table 5와 같이 산정되었으며, 그 결과 가뭄이 1, 2개월 지속될 경우에는 군집2에서 가장 큰 확률 값이 산정되었고, 3개월 이상의 가뭄 지속기간에서는 군집5에서 가장 큰 확률값이 산정되었다.

$$f_D(d) = P_{DW} \cdot P_{DD}^{d-1}; \quad d = 1, 2, 3, \dots \quad (4)$$

4. 결 론

본 연구에서는 강우지점자료를 이용하여 가뭄을 지역적 특성을 분석하기 위하여, 동질한 가뭄의 특성을 지닌 지역을 구분하기 위한 연구를 실시하였다. 전국 58개 강우 관측지점의 자료를 이용하여 표준강수지수(SPI)를 산정하여 60개의 가뭄속성인자 생성하였다. 60개의 가뭄속성인자를 대상으로 다변량 분석 기법 중 주성분분석과 요인분석을 실시하여 가뭄속성인자의 변동 중 95.4% 설명해 주는 주요인 7개를 추출하였다. 선정된 7개의 가뭄주성분요인은 K-means 군집분석에서 지역을 구분하는 척도로 사용되었다. 이러한 연구과정은 가뭄을 정량적으로 표현해주는 특성자료를 바탕으로 동질한 지역을 구분하기 위한 기준척도를 결정해 주는 과정으로 군집분석 시 중요하게 고려되어야 할 사항이다. 또한 군집분석을 통해 가뭄특성이 동질한 지역을 구분하였으며, 이러한 동질지역은 지리적 특성과 유역별 특성을 공통적으로 내포한 것으로 강우자료를 이용하여 우리나라를 수문학적으로 지역구분을 한 허준행 등(2007)의 연구결과와 유사한 결과를 보여주었다.

또한 본 연구에서는 동질한 가뭄특성을 갖고 있는 지역의 구분을 바탕으로 군집별 표준강수지수(SPI)의 특성과 공간적으로 가뭄의 지속기간에 따라 가뭄과 습윤상태가 어떠한 확률적 거동을 갖고 있는지 파악하였다. 이는 지점분석을 실시하였을 경우의 가뭄사상의 자료수 보다 가뭄특성이 동질한 다지점의 가뭄사상을 모두 고려하여 분석한 통계적 특성임으로, 표본 자료의 수가 부족할 수 있는 지점분석의 단점을 보완할 수 있는 방안으로 판단된다. 이는 가뭄특성분석시 강우관측기간이 짧은 지점자료를 이용하기 때문에 발생할 수 있는 불확실성을 감소할 수 있을 것이다.

이러한 연구결과는 지역적 가뭄의 공간분석을 바탕으로 동질한 가뭄특성을 지닌 지역의 가뭄 사상의 재현기간 산정 및 재현기간에 따른 가뭄심도 등 가뭄의 확률적 특성 분석에 관한 연구를 수행하는데 유용하게 사용되어질 것이다.

감사의 글

이 연구는 소방방재청 자연재해저감기술개발사업 [NEMA-08-NH-05] 연구비 지원으로 수행되었으며 이에 감사드립니다.

참 고 문 헌

- 경기개발연구원 (2004). 경기북부 농업가뭄현황 감시기법 연구. 기본연구 2004-06, pp. 20-24.
- 경민수, 김상단, 김보경, 김형수 (2007). “군집분석을 통한 수문학적 가뭄의 가뭄심도-가뭄면적-가뭄지속기간곡선의 작성.” **대한토목학회논문집**, 대한토목학회, 제27권, 제3B호, pp. 267-276.
- 국립기상연구소 (2009). 기후변화 이해하기 III - 서울의 기후변화. 국립기상연구소, pp. 8.

- 김보경, 김상단, 이재수, 김형수 (2006). “가뭄의 시공간적 분포 특성 연구: 가뭄심도-가뭄면적-가뭄지속기간 곡선의 작성.” **대한토목학회논문집**, 대한토목학회, 제26권, 제1B호, pp. 69-78.
- 김휘린, 박무중, 윤용남 (2004). “자료계열 특성에 따른 지역가뭄빈도 해석.” **한국수자원학회 학술발표회 논문집**, 한국수자원학회, pp. 1027-1031.
- 남우성, 김태순, 신주영, 허준행 (2008). “다변량 분석 기법을 활용한 강우 지역빈도해석.” **한국수자원학회 논문집**, 한국수자원학회, 제41권 제5호, pp. 517-525.
- 노형진 (2001). 한글 SPSS 10.0에 의한 조사방법 및 통계분석, 형설출판사. pp. 437-476.
- 박우창, 승현우, 용환승, 최기현 (2003). 데이터 마이닝 개념 및 기법, 자유아카데미. pp. 413-434.
- 오태석, 오근택, 문영일, 이일주 (2007). “군집분석과 비매개적 지역빈도 해석을 이용한 확률강우량 산정.” **대한토목학회 정기학술대회 논문집**, 대한토목학회, pp. 1942-1945.
- 유철상, 안재현, 류소라 (2004). “포아송과정을 이용한 가뭄의 공간분포 분석.” **한국수자원학회 논문집**, 한국수자원학회, 제37권, 제10호, pp. 813-822.
- 장연규, 김상단, 최계운 (2006). “SPI 가뭄지수의 EOF 분석을 이용한 가뭄의 시공간적인 특성 연구.” **한국수자원학회 논문집**, 한국수자원학회, 제39권, 제8호, pp. 691-702.
- 차석빈, 김홍범, 오홍철, 윤지환, 김우곤 (2008). 사례를 통해 본 다변량 분석의 이해, 백산출판사. pp. 257-271.
- 허준행, 이영석, 신흥준, 김경덕 (2007). “우리나라 강우 자료의 지역빈도해석 적용성 연구(I): 확률강우량 산정.” **대한토목학회 논문집**, 대한토목학회, 제27권 제2B호, pp. 101-111.
- Dinpashoh, Y., Fakheri-Fard, A. Moghaddam, M., Jahanbakhsh, S., Mirnia, M. (2004). “Selection of variables for the purpose of regionalization of Iran’s precipitation climate using multivariate methods.” *Journal of Hydrology*, Vol. 297, pp. 109-123.
- Hosking, J.R.M., and Wallis, J.R. (1997). *Regional Frequency Analysis: an Approach Based on L-moments*, Cambridge University Press.
- Mallants, D., and Feyen, J. (1990). “Defining homogeneous precipitation regions by means of principal component analysis.” *Journal of Applied Meteorology*, Vol. 29, pp. 892-901.
- Mathier, L., Perreault, L., Bobe, B., and Ashkar, F. (1992). “The use of Geometric and Gamma-related distribution for frequency analysis of water deficit.” *Stochastic Environmental Research and Risk Assessment*, Vol. 6, pp. 239-254.
- McKee, T.B., Doesken, N.J., and Kleist, J. (1993). “The Relationship of Drought Frequency and Duration of Time scales.” *Proceedings Eighth Conference on Applied Climatology*, Anaheim, C.A., pp. 179-187.
- Puvaneswaran, M. (1990). “Climatic classification for Queensland using multivariate statistical techniques.” *International Journal of Climatology*, Vol. 10, pp. 591-608.
- Salas, J.D., Fu, C., Cancelliere, A., Dustin, D., Bode, D., Pineda, A., and Vincent, E. (2005). “Characterizing the severity and risk of drought in the Poudre River, Colorado.” *Journal of Water Resources Planning and Management*, Vol. 131, No. 5, pp. 383-393.
- Sen, Z. (1976). “Wet and dry periods of annual flow series.” *Journal of Hydraulics Division, ASCE*, Vol. 102, pp. 1503-1514.
- White, D., Richman, M., and Yanel, B. (1991). “Climate regionalization and rotation of principal components.” *International Journal of Climatology*, Vol. 11, pp. 1-25.
- Yevjevich, V.M. (1967). “An objective approach to definitions and investigations of continental hydrologic droughts.” *Hydrology Paper 23*, Colorado State Univ., Fort Collins, Colorado.
- Zhang, J., and Hall, M.J. (2004). “Regional flood frequency analysis for the Gan-Ming River basin in China.” *Journal of Hydrology*, Vol. 296, pp. 98-117.

| | |
|------------------------|------------------|
| 논문번호: 09-089 | 접수: 2009.07.25 |
| 수정일자: 2009.10.07/11.02 | 심사완료: 2009.11.02 |