

# 극치강수량의 시공간적 특성 분석 및 지역화에 관한 연구

## Analysis on Spatio-Temporal Pattern and Regionalization of Extreme Rainfall Data

이정주\* · 권현한\*\*

Lee, Jeong-Ju · Kwon, Hyun-Han

### Abstract

The spatio-temporal pattern in precipitation is a significant element in defining characteristics of precipitation. In this study, a new scheme on regionalization utilizing temporal information was introduced on the basis of existing approaches that is mainly based on simple moments of data and geographical information. Given the identified spatio-temporal pattern, this study was extended to characterize regional pattern of annual maximum rainfall over Korea. We have used circular statistics to characterize the temporal distribution on the precipitation, and the circular statistics allow us to effectively assess changes in timing of the extreme rainfall in detail. In this study, a modified K-means method was incorporated with derived temporal characteristics of extreme rainfall in order to better characterize hydrologic pattern for regional frequency analysis. The extreme rainfall was reasonably separated into five categories that considered most attributes in both quantitative and temporal changes in extremes. The results showed that the proposed approach is a promising approach for regionalization in term of physical understanding of extreme rainfall.

**Keywords :** spatio-temporal variation, regionalization, circular statistics, precipitation

### 요 지

강수량은 시공간적으로 변화하는 수문변량으로서 강수량의 시간적인 특성 또한 강수량의 특성을 정의하는데 매우 중요한 요소이다. 본 연구에서는 극치강수량의 지역빈도해석을 위한 범주화 과정에서 기존의 강수량과 관측소 위치좌표만을 이용한 범주화를 통해 해결할 수 없는 강우 발생의 계절적 변화와 집중에 대한 고려를 반영하기 위하여, 기존의 양적 범주화 과정에 시간적인 영향을 고려할 수 있는 요소로서 극치강수량 발생 시기 통계치를 활용할 수 있는 범주화 과정을 제시하였다. 본 연구에서는 극치강수량의 발생 시기에 대한 정량적인 분석이 가능한 순환통계기법을 이용하여 관측 지점별 시간 통계량을 산정하고, 이를 극치강수량과 결합하여 시·공간적인 특성자료를 생성한 후 수정 K-means 방법을 이용하여 군집화 해석을 수행하였으며, 전국을 다섯 개의 군집으로 분류하였다. 기존의 양적 범주화 결과와의 비교를 통해 발생 시간 통계치를 결합한 범주화 결과가 지형 및 권역을 반영하는 결과를 보임을 확인하였다.

**핵심용어 :** 시공간적 변동성, 지역화, 순환 통계치, 강수

### 1. 서 론

수공구조물의 설계, 수자원 관리계획의 수립, 재해영향 검토 등을 수행할 때, 필요한 수문사상은 재현기간에 따른 확률강수량, 홍수량 및 유입량 등을 산정하여 사용하게 되며, 보통 대상지역 인근의 장기 수문관측 자료를 이용하여 수문사상의 확률분포를 산정한 후 재현기간을 연장하여 원하는 설계빈도에 해당하는 양을 추정하게 된다. 특히 유량의 경우는 관측자료의 기간이 통계적 유의성을 갖기에는 짧은 경우가 대부분이고, 미계측 지역이 많은 이유로 강우자료의 빈도 분석을 통해 확률강수량을 산정하고 이를 강우-유출모형 등

을 이용하여 유량으로 변환한 자료를 동일한 재현기간을 갖는 유량으로 간주하고 사용하는 경우가 대부분이다. 즉 우리나라 수자원 관리계획의 기초자료로 가장 빈번히 사용되는 것이 바로 강우량이라 할 수 있다. 강우량의 빈도분석은 지점 빈도분석과 지역 빈도분석으로 나눌 수 있으며, 원하는 지역의 관측자료 유무에 따라 각각 사용될 수 있다. 이중 지역 빈도분석은 자료연수가 상대적으로 작은 유역에 대해서 자료를 확충하고 신뢰성 있는 빈도분석 결과를 추정하기 위해서 국내외적으로 많은 적용이 이루어지고 있다.

지역빈도해석을 위해서는 강우자료들의 동질성을 파악하는 것이 가장 기본적인 과정이 되며 이를 위해 통계학적인 범

\*정현원 · 전북대학교 토목공학과 박사후연구원 (E-mail : julee@jbnu.ac.kr)

\*\*정현원 · 교신저자 · 전북대학교 토목공학과 조교수 (E-mail : hkwon@jbnu.ac.kr)

주회분석이 선행되어야 한다. 국내에서는 강수량의 범주화에 관한 연구와 이를 통한 지역빈도분석 결과에 대한 영향을 검토한 연구가 이루어져 왔다. 강수량의 범주화에 대한 논문으로서 윤용남과 박무중(1997)은 전국을 47개 소유역으로 분할한 후 월 강우량계열로부터 작성된 기뿔우량계열의 지역빈도분석을 하였다. L-Moment법을 사용하여 확률분포의 매개변수를 결정하였으며 지속기간별 확률기뿔우량을 작성하였다. 박상우 등(2003)은 기상청 산하 60개 강우관측소에 대한 32개의 강우특성자료를 추출하였으며, 공간적 특성인 경위도와 강우의 양적개념의 연평균, 지속기간별 평균들, 지속기간에 따른 초과발생확률 등의 특성 값들을 주성분분석을 통해 3개의 중요주성분으로 추출하였으며, 주성분점수를 이용하여 군집 분석한 결과 전국을 5개의 권역과 2개의 기타지역으로 분류하였다. 오태석 등(2008)은 관측소별 위치자료와 지속기간별 최대시간강우량의 평균을 이용해 군집분석을 실시하고, 지역빈도분석을 위해 관측연수 10년 이상인 70개 지점의 강우자료와 지점빈도분석을 위한 관측연수 30년 이상인 61개 지점을 대상으로 빈도분석을 실시하여 확률강우량을 비교하였다. 김상욱과 이길성(2008)은 K-means 알고리즘을 이용해 저수량에 대한 수문학적 동질영역을 분류하고, Bayesian 다중회귀분석을 이용한 회귀계수 검정을 통해 저수량 지역 빈도분석을 수행하였다.

강수빈도분석을 위한 범주화 연구 이외에도 기뿔분석을 위한 범주화 적용사례가 있다. 신현석과 박무중(1999)은 전국 39개 강우관측소의 연강우량 자료를 기초로 공간 해석 신경망(SANN)을 이용하여 전 지역에 대한 기뿔확률을 산출하고 Bayesian 기뿔 심도 지수(BDSI)를 통하여 극심, 심, 경심, 비 기뿔지역으로 분류하였다. 경민수 등(2007)은 기상청 산하 58개 강우관측소를 대상으로 구한 표준강수지수를 이용하여 계층적 군집분석을 수행하였다. 우선 강우관측소의 위치자료를 이용하여 공간적 군집분석을 통해 6개의 군집으로 분류하고 각 군집별 표준강수지수(SPI)에 대해 경험적 직교함수(EOF) 분석결과를 크리깅(Kriging)을 적용하여 범주화하였다.

강수량의 지역 빈도분석은 보통 지점 빈도분석 결과의 수문학적 동질성을 L-moment 방법, 군집분석 등을 통해 판별하고, 그 결과를 회귀분석, 통계적 지표를 이용한 그래프, 크리깅과 같은 공간 보간법을 이용하여 적용하게 된다(김상욱과 이길성, 2008).

강수량의 지역 빈도분석에 있어서 문제가 되는 것이 강우관측소의 각기 다른 위치관계와 관측 자료기간의 불규칙성으로 인한 자료의 규격화이며, 이는 표준화 또는 정규화를 통해 해소는 가능하지만(신현석과 박무중, 1999) 강우 발생의 계절적 변화와 집중에 대한 고려를 어떻게 반영할 것인가 하는 것은 기존의 강수량과 관측소 위치좌표만을 이용한 범주화에서 해결할 수 없는 문제이다. 즉, 기존 지역빈도분석을 위한 범주화 과정에서는 강수지점의 공간적인 좌표 및 강수량의 양적인 범주화에 초점을 맞추고 있으나 강수량의 시간적인 발생특성은 고려되지 않는 것이 일반적이다. 그러나 강수량은 시공간적으로 변화하는 수문변량으로서 강수량의 시간적인 특성 또한 강수량의 특성을 정의하는데 매우 중요한 요소이다. 이러한 점에서 본 연구에서는 기존 범주화

과정에 시간적인 영향을 고려할 수 있는 요소들을 결정하고 이를 활용할 수 있는 범주화 과정을 제시하고자 한다. 즉, 본 연구에서는 극치강수량의 발생 시기에 대한 정량적인 분석이 가능한 순환통계기법을 이용하여 관측 지점별 시간 통계량을 산정하고, 이를 극치강수량과 결합하여 시·공간적인 특성자료를 생성한 후 이를 이용한 군집화 해석 모형을 개발하는데 연구의 목적이 있다.

본 연구의 진행 절차는 다음과 같다. 첫째, 순환통계해석 방법에 대한 특성을 평가하고 모의자료를 통해 해석과정을 설명한다. 둘째, 기존, 양적인 개념에 근거한 수문학적 동질성 해석 방법에 대한 문제점을 평가하고 극치강수량의 시공간적 특성을 반영한 수문학적 동질성 해석 방법을 제시한다. 셋째, 우리나라 강수지점들의 강수량에 대해서 수문학적 동질성 평가를 실시한다.

## 2. 지역화를 위한 시공간적 분석방법

강우의 지역빈도분석을 위해서는 우선 각 관측지점의 자료에 대한 수문학적 동질성 판별이 필수적이며, 지형적인 분할, 행정구역에 따른 분할, L-moment 방법에 의한 분할, 군집분석(cluster analysis)에 의한 분할 등을 적용한다(김상욱과 이길성, 2008). 본 연구에서는 극치강수량의 시간속성을 순환통계기법을 이용해 정량화하고, K-means 알고리즘을 이용해 강수량과 시간속성을 결합하여 군집분석을 수행한 후, L-moment 방법으로 지역화 결과에 대한 검증을 수행하였다.

### 2.1 순환통계치

강수량 집중시기의 변화를 분석하기 위해 순환통계기법(Circular Statistics)을 적용하였으며, 순환통계는 월최대강수량의 발생지점, 일최대강수량의 발생지점 등 시간의 속성을 가지는 자료계열의 특성을 평가하는데 매우 유리하다(권현환과 이정주, 2010). 즉 시간의 속성을 각도(angular)의 형태로 변환하여 자료가 가지는 발생 패턴을 효과적으로 분석할 수 있다. 단순히 자료간의 유클리드거리(euclidean distance)를 측정하는 것이 아니라 자료의 방향 통계치(directional statistics)를 고려함으로써 시간의 속성을 갖는 발생특성을 정량화하기에 유리하다. 단적인 예로 각도로 측정된 풍향의 경우 359도 방향과 0도의 풍향은 거의 같은 방향의 의미를 갖지만, 일반적인 거리개념으로 비교할 경우 올바르게 분석될 수 없다. 또한, 연중 일자와 같은 시간의 속성을 갖는 수문 시계열의 경우에도 발생특성을 분석하는데 있어서 순환 통계치가 유리하게 적용될 수 있다(Magilligan과 Graber, 1996; Berens, 2009). Fig. 1은 월강수량의 최대값이 발생한 시점을 대상으로 추정된 예를 나타낸다. Fig. 1(b)에서 히스토그램은 발생빈도를 나타내며 직선은 방향각의 평균을 의미한다. 즉, 월강수량의 최대값이 나타나는 시기는 평균적으로 7월 중순과 하순 사이임을 알 수 있다.

본 논문에서는 순환통계에 대한 일반적인 사항을 권현환과 이정주(2010)가 제시한 방법론을 참조하여 요약하였다. 순환통계 계산을 위해서는 원 자료계열을 방향각으로 나타내야 한다. 자료계열을 Radian으로 환산한 방향값  $\alpha$ 는 아래와 같이 계산할 수 있다.

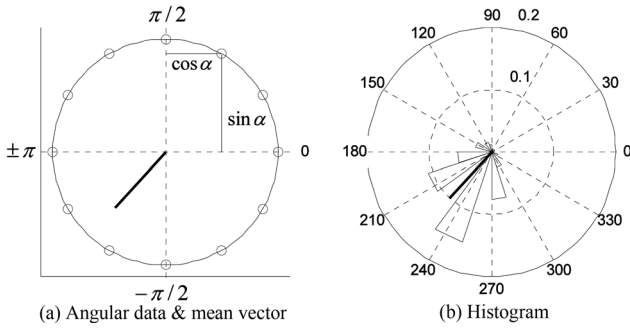


Fig. 1 An example showing angular histograms for the timing of the maximum monthly precipitation on the unit circle

$$\alpha = \frac{2\pi x}{k} \quad (1)$$

여기서  $x$ 는 원 데이터 값이며,  $k$ 는 원 데이터의 계급값이다. 원자료에서 추출한 방향값을 이용하여 이동평균 등을 구하고자 할 때, angular scale에서의 평균은 방향값을 단순히 산술평균하는 것이 아니라  $\alpha$ 에 대한  $x, y$ 방향의 단위벡터  $r_i$ 를 구한 후 벡터의 평균을 구해야 한다.

$$r_i = \begin{pmatrix} \cos \alpha_i \\ \sin \alpha_i \end{pmatrix} \quad (2)$$

$$\bar{r} = \frac{1}{N} \sum_i r_i \quad (3)$$

여기서,  $\bar{r}$ 는 평균합성벡터(mean resultant vector)이고,  $N$ 은 자료의 개수이다. 이로부터 평균벡터의 길이  $R$ 을  $|\bar{r}|$ 로부터 구할 수 있다. Angular scale에서의 편차는 0부터  $\sqrt{2}$  사이의 값을 가지며, 아래와 같이 정의된다.

$$s = \sqrt{2(1-R)} \quad (4)$$

또한 원형좌표의 표준편차는 아래와 같이 구한다.

$$s_0 = \sqrt{-2 \ln R} \quad (5)$$

자료의 대칭정도를 구하기 위한 왜곡도(skewness)는 원형좌표계에서 아래와 같이 구해진다(Pewsey, 2004). 왜곡도가 0에 가깝다는 것은 평균벡터 방향을 주위로 대칭적인 분포를 보인다는 의미이다.

$$b = \frac{1}{N} \sum_{i=1}^N \sin 2(\alpha_i - \bar{\alpha}) \quad (6)$$

여기서  $\bar{\alpha}$ 는  $\alpha$ 값들의 평균이다. 표준화된 왜곡도 계산은 아래 식과 같이 정의된다(Fisher, 1995).

$$b_0 = \frac{R_2 \sin(\bar{\alpha}_2 - 2\bar{\alpha})}{(1-R)^{3/2}} \quad (7)$$

여기서  $R_2$ 는 2차모멘트의 평균 벡터길이,  $\bar{\alpha}_2$ 는 2차모멘트의 평균 각을 의미한다. 마지막으로 두 개의 방향성을 갖는 자료간의 상관성을 평가하기 위해서 다음 식 (8)을 이용하여 상관계수를 추정하였다(Jammalamadaka와 Sengupta, 2001).

$$\rho = \frac{\sum_i \sin(\alpha_i - \bar{\alpha}) \sin(\beta_i - \bar{\beta})}{\sqrt{\sum_i \sin^2(\alpha_i - \bar{\alpha}) \sum_i \sin^2(\beta_i - \bar{\beta})}} \quad (8)$$

여기서  $\bar{\beta}$ 는  $\beta$ 값들의 평균이다.

## 2.2 시공간적 해석기법

연최대강수량과 최대강수 발생시점에 대하여 시·공간적 특성을 분석하기 위해서, K-means 알고리즘을 이용한 군집 분석을 수행하였다. K-means 알고리즘은 임의의 초기 군집 중심점을 설정하고, 중심점으로부터 표본까지의 거리를 제공하는 합을 최소로 하는 표본들과 중심점을 반복 계산을 통해 찾아가는 과정이다(Hartigan, 1975). K-means 알고리즘의 수행과정을 요약하면 다음과 같다.

Step 1: K개의 초기 군집 중심  $Z_1(1), Z_2(1), Z_3(1), \dots, Z_K(1)$ 을 선택한다. 이 초기값들은 임의로 선택될 수 있으며, 일반적으로 처음 K개의 표본을 선택한다.

Step 2: K번째 반복 단계에서는 다음의 관계식을 이용하여 표본(K)를 각 K개의 군집에 분배한다.

$$x \in S_j(k) \text{ if } \|x - z_j\| < \|x - z_i\| \text{ for all}$$

$$i = 1, 2, 3, \dots, K, i \neq j$$

Step 3: Step 2의 결과로부터, 모든  $j=1, 2, 3, \dots, K$ 에 대하여  $S_j(k)$ 에 포함된 모든 점들로부터 군집 중심까지의 거리에 제공하는 합을 최소로 하는 새로운 군집 중심  $Z_j(k+1)$ 을 계산한다. 즉 다음과 같은 성능 지표를 최소화하도록 새로운 군집 중심  $Z_j(k+1)$ 을 계산한다.

$$J_j = \sum_{x \in S_j(k)} \|x - z_j(k+1)\|^2$$

$$j = 1, 2, 3, \dots, K$$

이 성능지표를 최소화하는  $Z_j(k+1)$ 은 항상  $S_j(k)$ 의 표본에 대한 평균이 된다. 따라서  $N_j$ 가  $S_j(k)$ 에 속한 표본의 개수일 때, 새로운 군집은 다음과 같이 나타낼 수 있다.

$$Z_j(k+1) = \frac{1}{N_j} \sum_{x \in S_j(k)} x$$

$$j = 1, 2, \dots, K$$

Step 4: 모든 군집들에 대해 중심의 변화가 없을 때까지 Step2와 Step3을 반복 수행한다.

$Z_j(k+1) = Z_j(k)$ 를 만족하면 알고리즘은 수렴하며 종료된다.

본 연구에서는 K-means 방법에 의한 군집화 결과를 평가하기 위해 L-moments 방법에 의한 이질성척도(Heterogeneity Measure)를 이용하였다. L-moments 방법에 의한 매개변수 추정량은 아래 식 (9)~(12)와 같이 나타낼 수 있다(Hosking과 Wallis, 1997).

$$\xi_1 = E(X) \quad (9)$$

$$\lambda_2 = 1/2 E(X_{(2,2)} - X_{(1,2)}) \quad (10)$$

$$\lambda_3 = 1/3 E(X_{(3,3)} - 2X_{(2,3)} + X_{(1,3)}) \quad (11)$$

$$\lambda_4 = 1/4 E(X_{(4,4)} - 3X_{(3,4)} + 3X_{(2,4)} - X_{(1,4)}) \quad (12)$$

여기서,  $E(\cdot)$ 는 기대값이고,  $X(i, n)$ 는 크기가  $n$ 인 표본 중에서  $i$ 번째로 작은 통계량이다.

L-moment 비는 다음 식 (13)~(15)와 같다.

$$\tau_2 = \lambda_2 / \lambda_1 \Rightarrow L-CV(L\text{-coefficient of variation}) \quad (13)$$

$$\tau_2 = \lambda_3 / \lambda_2 \Rightarrow L\text{-skewness} \quad (14)$$

$$\tau_4 = \lambda_4 / \lambda_2 \Rightarrow L\text{-kurtosis} \quad (15)$$

소속된 지점의 자료계열의 수문학적 동질성을 파악하기 위해 다음 식 (16)과 같이 이질성척도(Heterogeneity Measure)  $H$ 를 사용한다.

$$H = \frac{V - \mu_v}{\sigma_v} \quad (16)$$

$$V = \frac{\sum_{i=1}^N n_i (t^i - \bar{t})^2}{\sum_{i=1}^N n_i} \quad (17)$$

여기서, 표본자료 집단의 평균 L-moments인  $1, t^R, t_3^R, t_4^R$ 를 Kappa 분포형에 적합시킨 다음, 충분한 횟수의 모의 발생을 실시하여  $V$ 를 계산한 후 평균값과 표준편차를 계산하여  $H$ 를 구한다. Hosking과 Wallis(1997)는  $H < 1$ 이면 동질성 수용 가능 지역,  $1 \leq H < 2$ 이면 이질성 가능 지역,  $H > 2$ 이면 이질성 지역으로 구분하였다.

### 2.3 모의실험

모의실험을 통해 시공간적 특성을 함께 고려한 군집화의 효과를 검증하였다. 우선 일정한 간격으로 배열된 100개의 관측소 위치좌표를 생성하였으며, 10×10의 형태로 배열하였다. 관측소는 모의데이터를 부여하는 그룹을 네 개로 구분하였다.

관측소별로 30개의 극치강수량 자료와 발생시간을 랜덤 발생하였다.  $x=5.5, y=5.5$ 인 중앙선을 기준으로 사분면을 구분

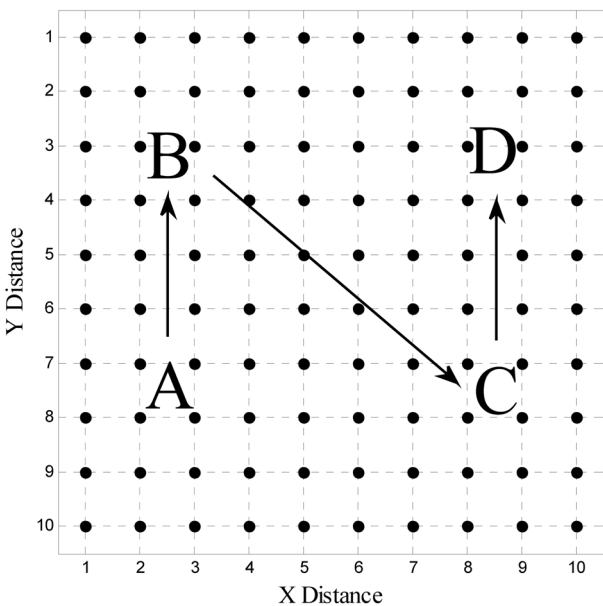


Fig. 2 Graphical representation of the case study for the Experiment

한 후 각 사분면에 평균값을 차등 지정하는 방법으로 구역의 양적 차이를 부여하였다. 100개의 각 셀을 하나의 관측소로 간주하고 각 셀에 30개의 극치강수량을 모의 발생시켰으며, 모의 발생시킨 극치강수 계열은 정규분포이다. 매개변수는 평균(location parameter)과 분산(scale parameter)을 이용하여 큰 네 개의 구역에 대해 차등을 두었다. 자료 모의발생에 사용된 확률분포함수는 MATLAB의 기본함수를 이용하였다. 랜덤발생 값의 범위는 표준편차로 지정하였으며, (a)Maximum Rainfall의 각 셀에 나타난 음영 값은 해당 셀(관측소)에 생성된 30개 극값 계열의 평균을 이용해 도시하였다. 사분면 범위에 표기한 숫자는 해당 사분면의 랜덤 값 생성에 사용된 평균이다. 각 사분면에 속한 25개의 관측소들은 Table 1에 정리한 것과 같이 극치강수량(R)에 대하여 동일한 평균을 기준으로 표준편차 범위에서 랜덤하게 자료를 발생시켰다. 최대강우 발생시간은 100(관측소 개수)×30(극치강우자료 연수)개의 각 자료가 발생된 시간을 월(month)값으로 지정하였다. 그룹별 모의자료 발생에 사용된 극치강수량의 평균값(R)과 발생시간(T) 및 표준편차(STD)는 Table 1과 같다.

Case 1은 강수량과 시간이 동일한 경계로 구분되는 경우이다. 두 경우 모두 예상된 구역을 경계로 군집화가 이루어지는 것을 볼 수 있다.

Case 2는 강수량이 경계에 따라 차이가 있으며, 시간은 세 구역이 동일한 값을 갖는 경우로, 전 지역에 걸쳐 동일한 시기에 최대 강우가 발생하나 지형적 영향 등으로 위치에 따른 강수량의 차이가 발생하는 경우를 가정한 모의자료이다. 그 결과는 강수량의 차이에 의해 동일한 시간속성을 갖는 구역이 다르게 나뉜 것을 볼 수 있다.

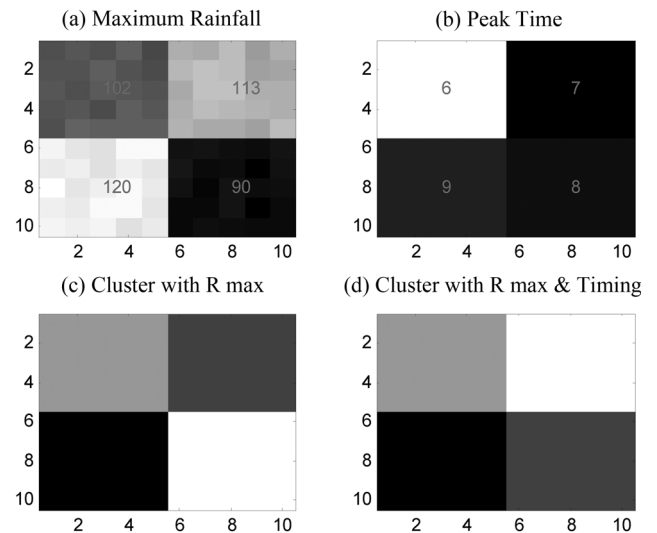


Fig. 3 Synthetic Test Result of Case 1

Table 1. Mean Value for Synthetic Data

Synthetic Data	Group A		Group B		Group C		Group D		STD of R
	R	T	R	T	R	T	R	T	
Case 1	120	9	102	6	90	8	113	7	5
Case 2	120	7	102	6	90	7	113	7	5
Case 3	104	9	102	6	100	8	103	7	5
Case 4	118	6	116	8	105	6	107	9	5

Case 3은 강수량의 구역별 편차가 거의 없지만 시간은 구역별로 차이가 존재하는 경우로, 지역별로 각기 다른 시기에 최대 강우가 발생하였으나 그 양은 비슷한 경우를 가정하였다. 강수량에 의한 클러스터링이 심하게 오류가 발생하는 것

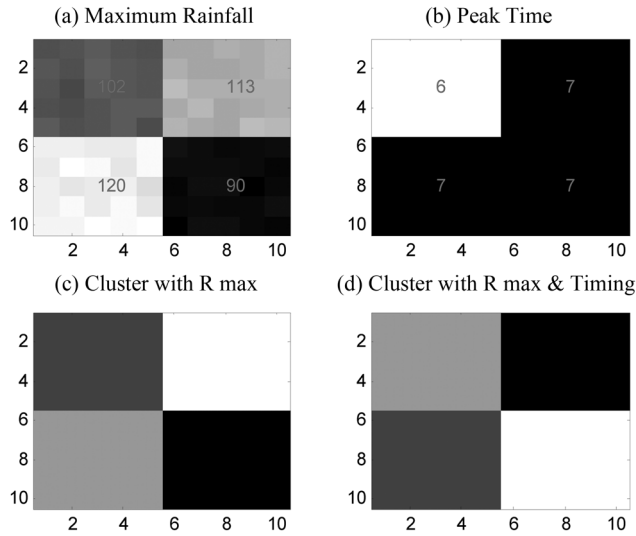


Fig. 4 Synthetic Test Result of Case 2

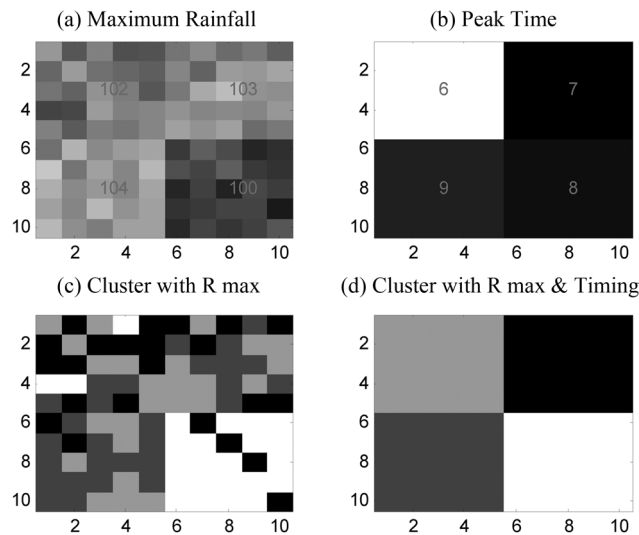


Fig. 5 Synthetic Test Result of Case 3

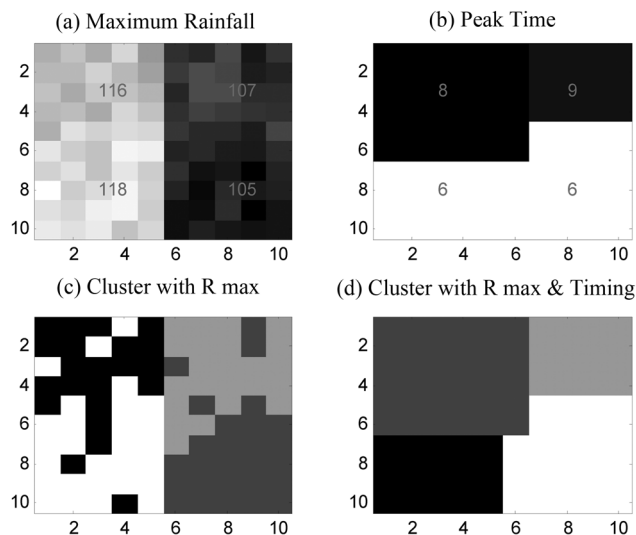


Fig. 6 Synthetic Test Result of Case 4

에 비해 시간속성을 이용하여 보정되는 효과를 볼 수 있다.

Case 4는 강수량과 시간의 분포 구역이 다른 경우이다. 전선형 강우의 공간이동 효과를 고려한 테스트이며, 시간요소를 고려한 경우 정확한 군집화가 이루어지는 것을 볼 수 있다.

### 3. 극치강수량의 시공간적 분석

#### 3.1 강수량자료

본 연구에서는 기상청에서 관리하는 관측소 중 장기간의 관측 자료를 보유한 58개 관측소를 대상으로 분석을 실시하였다. 자료의 연한은 관측소에 따라 상이하며 전체적으로 35년 이상의 자료를 보유하고 있다. Fig. 7은 본 연구에서 사용된 강수지점의 공간적 위치를 나타낸다.

#### 3.2 극치강수량의 범주화 분석

모의실험을 통하여 시·공간자료의 결합에 의한 군집화는 매우 양호하게 이루어지는 것을 확인할 수 있으나, 두 가지 문제점을 고려할 필요가 있는 것으로 나타났다. 첫째, 같은 데이터에 대해 반복적으로 군집분석을 수행하는 경우라도, 매번 임의로 지정되는 초기 Centroid(무계중심점)의 위치에 따라 분석 결과가 다르게 나올 수 있는 문제가 있다(Khan과 Ahmad, 2004). 둘째, K-means 방법에 의해 군집화를 수행할 경우 군집별 분류번호가 불규칙하게 생성되는 문제가 있다. Fig. 3과 Fig. 4의 군집화결과 (c)와 (d)에서 볼 수 있듯이, 1~4까지의 군집번호에 따라 지정된 색인명암(index contrast)의 위치가 다른 것을 볼 수 있다. 즉, K-means 방법에 의해 군집을 분석할 경우 결과적으로는 동일한 경계로 군집이 나뉘더라도 각 군집의 번호는 임의로 부여되는 것으로 인해 군집번호간의 물리적 크기 순서를 정할 수 없는 문제가 발생한다. 이 문제는 K-means 방법에 의해 해석된 관측소별 군집번호를 이용해 지역화를 수행할 경우 등고선 개념의 지역구분선이 잘못 그려지는 원인이 될 수 있다. 이상의 문제점을 바로잡기 위해서 자료의 예비해석 단계와 군집간의 크기순서에 따른 번호부여 과정을 추가하였

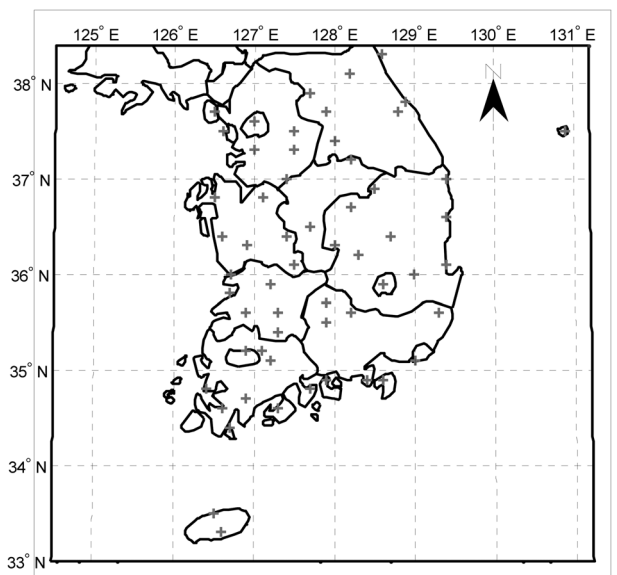


Fig. 7 A map showing the stations for this study

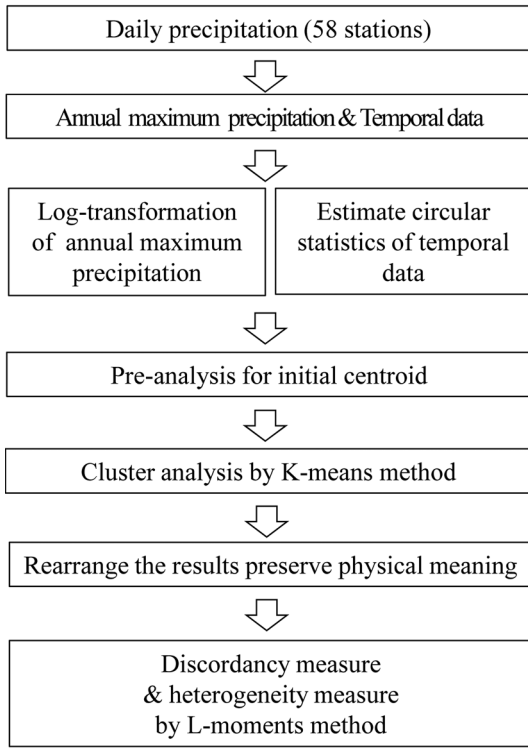


Fig. 8 Work Flow

다. 극치강수량의 시공간적 분석 과정은 Fig. 8과 같다.

분석에 사용될 데이터는 각 강우관측소의 일자료를 이용한 극치강수량과 그 발생일자이다. 극치강수량은 Log transform을 적용하였으며 발생일자는 순환통계분석 결과를 이용하였다. 군집화에 사용된 두 속성자료 중 극치강수량(mm)은 일강수량으로서 그 크기가 일반적으로 두 자리 또는 세 자리의 실수이며, 발생시간은 Circular scale에서 각도로 환산된 (발생일/365) Radian 값으로 최대값이  $\pm\pi$  이기 때문에 Log transform을 통해 강수량을 환산하지 않을 경우 그 양적인 차이로 인해 발생시간 속성에 의한 군집 분류의 차별성을 갖기 어렵다. 예비해석 단계에서 100회의 군집분석을 통해 평균적인 Centroid를 산정하고, 해당 값을 본 해석의 초기 Centroid로 지정하여, 변동적인 군집화 경향을 안정화시켜 해석이 반복됨에 따라 군집화 결과가 달라지는 오류를 방지하였다. 또한 K-means 방법으로 계산된 군집별 공간거리 합의 크기에 따라 군집번호를 부여함으로써 군집의 번호순서대로 물리적인 연관성이 인접하도록 설정하였으며, 군집간의 경계

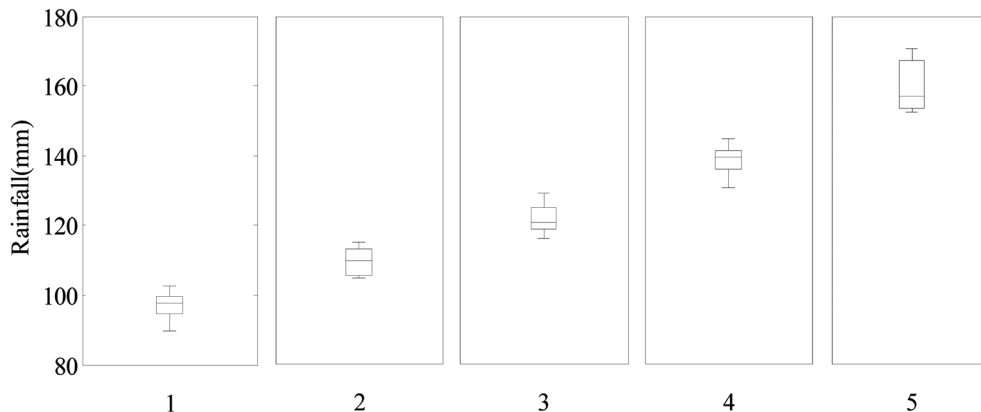


Fig. 9 Mean distribution of extreme rainfall according to each category. The regionalization was only based on first-order moment

Table 2. Heterogeneity Measure (Amounts)

Cluster	Standardized test value H(1)	Standardized test value H(2)	Standardized test value H(3)
Cluster 1	-0.0673	-0.7421	-1.1206
Cluster 2	-0.0428	-0.8407	-0.679
Cluster 3	-0.7237	-0.6315	-1.0308
Cluster 4	-0.2694	0.2077	0.999
Cluster 5	-1.6455	-0.9874	-0.8402

Table 3. Heterogeneity Measure (Amounts & Timing)

Cluster	Standardized test value H(1)	Standardized test value H(2)	Standardized test value H(3)
Cluster 1	-0.2341	-0.3945	-0.2357
Cluster 2	-0.1647	-0.7431	-1.2322
Cluster 3	-0.1578	-1.0511	-1.3078
Cluster 4	-1.1207	-0.1936	0.9584
Cluster 5	-1.1153	-0.0418	-0.2415

선을 추출할 때 발생할 수 있는 오류를 방지하였다.

시공간 결합모형은 네 개의 속성자료로 구성되어 있으며, 이 자료들은 극대치강수량의 평균과 표준편차, 발생시간의 평균벡터와 표준편차이고 이들 값을 모형 입력 자료로 사용하여 분석하였다. 울릉도와 제주도(제주, 서귀포)의 극대치강수량과 그 발생 시기는 인근 육지와 다른 경향을 보이며, 강릉과 대관령의 경우도 다른 지역과는 매우 다른 강우형태를 보인다. 이들 지점은 별도의 군집으로 분류하기는 관측소의 개수가 적고, 다른 지점으로부터의 공간거리는 매우 커서 군집화 성능지표와 지역화를 위한 보간법 적용에 미치는 영향이 크기 때문에 모형 입력 자료에서 제외하였다. 전체 58개 관측소 중 울릉도, 제주, 서귀포, 강릉 및 대관령을 제외한 53개 기상관측소에서 관측된 극치강수량의 시공간적 특성을 위와 같은 분석과정을 통해 분석하였다. Table 2는 극치강수량의 양적인 속성만을 이용하여 분류된 군집별 이질성척도이며, Table 3은 양과 시간속성을 함께 이용하여 분류된 군집별 이질성척도이다.

이질성척도만을 통해 비교한다면 두 경우 모두 1이하의 값을 갖는 동질성지역으로 군집화된 것을 볼 수 있다. 하지만 각 군집별 속성자료들의 분포를 보여주는 Fig. 9와 Fig. 10을 통해 분류된 군집의 특성을 자세히 살펴보면, 몇 가지

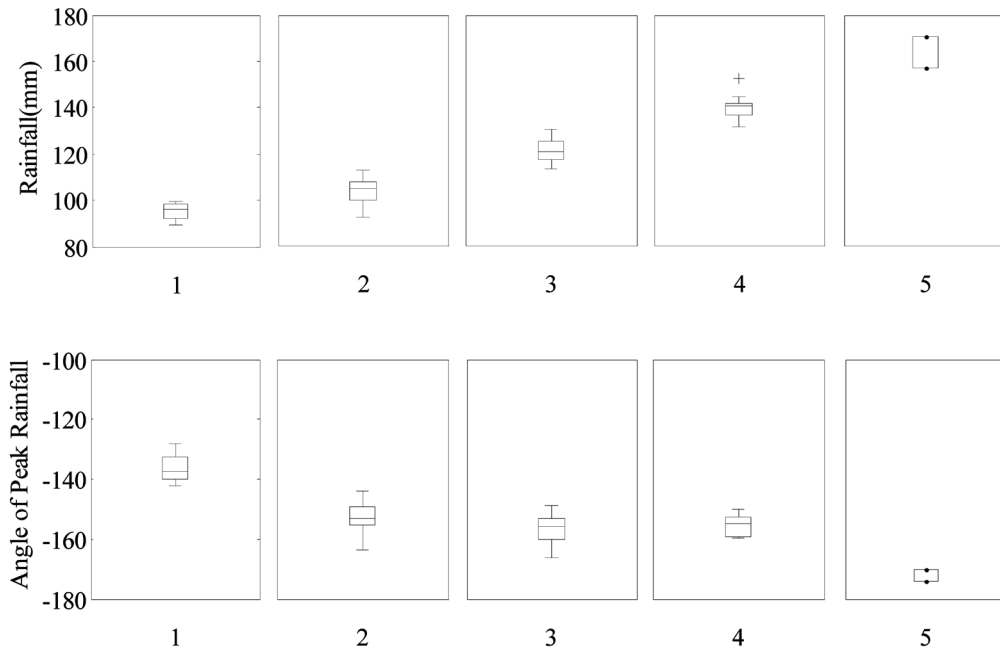


Fig. 10 Mean distribution of extreme rainfall(upper panel) and mean angle distribution of peak rainfall according to each category. The regionalization was based on first-order moment and circular statistics

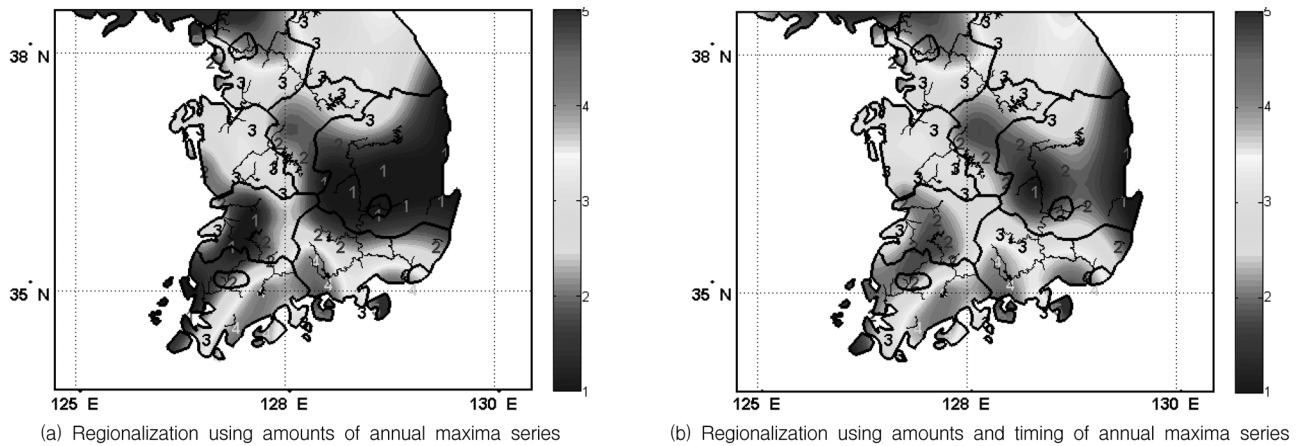


Fig. 11 Regionalization Result by spatio-temporal analysis

흥미로운 차이점들을 발견할 수 있다. 차이점에 대한 서술에 앞서 우선 그림을 설명하면, Fig. 9는 극치강수량 자료만을 이용하여 분류된 5개의 군집별 극치강수량자료의 분포를 보여주고 있으며, 1번 군집으로부터 5번 군집으로 갈수록 극치강수량의 평균이 큰 것을 볼 수 있다. Fig. 10은 극치강수량과 발생시간 속성을 이용하여 분류된 군집별 속성자료 분포를 보여주고 있다. Fig. 10의 상단은 군집별 극치강수량의 분포이며, 하단의 그림은 극치강수량 발생시간에 해당하는 방향각이다. 발생 시기 평균을 통해 1번 군집보다 5번 군집이 극치강수의 발생시기가 더 빠른 것을 알 수 있다. 각 그림에서 상자안의 가로선은 자료들의 평균이고, 상자 윗선분과 아래선분은 자료들의 상위 75%, 하위 25%에 해당하는 값이며, 상자 밖의 선분 끝은 95%와 5% 값을 나타내고 있다. Fig. 9의 1번과 2번 군집의 양적 차이가 Fig. 10의 1번과 2번 군집의 양적 차이보다 더 크다는 사실로 극치강수량만을 이용한 군집 분류가 더 명확하게 나뉜 것으로 판단할 수는 없다. 그 이유는 모의실험 Case3에 해당하는 경우로 Fig. 5의 결과에서 볼 수 있듯이 극치강수량이 양적으로 동질적이지만, 그 발생 시기가 분명히 다른, 즉 동질지역이

라 할 수 없는 관측소들이 같은 군집으로 잘못 분류되는 경우인 (c)를 통해 설명할 수 있다. Fig. 10을 보면, 새로 분류된 1번과 2번 군집의 양적 차이는 앞의 경우보다 줄어들었으나, 발생시기의 평균이 다른 군집들에 비해 확연히 차이가 나는 것을 확인할 수 있다. 즉, 양적인 속성만을 이용한 군집 분류의 오류 가능성을 발생 시기 속성을 이용하여 보완한 결과로 해석할 수 있다. 또한 3번과 4번 군집의 경우는 발생 시기의 차이는 크지 않으나 그 양적 차이로 인해 군집이 분류된 모의실험 Case 2의 결과인 Fig. 4와 같은 효과를 보인 것으로 해석할 수 있다. 군집화 방법별 결과를 비교한 Fig. 11을 보면, 결과적으로 1번 군집과 2번 군집의 변화가 가장 큰 것을 확인할 수 있다. 극치강수량만을 이용한 분류결과인 Fig. 11의 (a)에서는 1번 군집과 2번 군집에 속하는 관측소의 개수가 각각 11개와 12개였으나, 발생 시기 속성을 포함한 (b)의 경우는 각각 4개와 16개로 달라졌으며, 대구 경북지역 대부분과 전북지역 일부 관측소가 모두 1번 군집으로 분류되었던 것이 영산강권역, 낙동강권역, 낙동강동해권역으로 지형 및 유역별 구분이 보다 명확해진 것을 통해 물리적 의미가 반영된 결과로 판단할 수 있다. 극치강

Table 4. Comparison of Regional Frequency Analysis Results

		No. of Stations	2year	5year	10year	20year	25year	30year	40year	50year	80year	100year	150year	200year
Amounts	Cluster 1	11	97.1	135.5	163.5	192.4	202.1	210.2	223.1	233.3	255.7	266.7	287.1	302.5
	Cluster 2	12	108.5	153.2	186.8	222.6	234.8	245	261.4	274.7	303.8	318.3	345.5	366.3
	Cluster 3	17	121.1	167.1	200.1	233.8	244.9	254.2	269	280.7	306	318.3	341.1	358.2
	Cluster 4	10	140.3	198.3	239.7	281.9	295.7	307.3	325.6	340.1	371.5	386.7	414.8	435.9
	Cluster 5	3	159.2	224.9	274.6	327.5	345.5	360.6	385	404.6	447.9	469.4	509.8	540.7
Amounts & Timing	Cluster 1	4	95.3	137.2	171	209	222.4	233.8	252.5	267.8	302.5	320.2	354.1	380.8
	Cluster 2	16	102.8	143.2	172.8	203.6	213.9	222.5	236.2	247.2	271.1	282.9	304.7	321.2
	Cluster 3	21	121.3	168.8	203	237.9	249.5	259.2	274.5	286.7	313.1	326	349.7	367.6
	Cluster 4	10	140.9	198.1	238.8	280.2	293.8	305.1	323.1	337.3	368	382.9	410.3	430.8
	Cluster 5	2	164	231.7	283	337.7	356.3	372	397.3	417.6	462.5	484.9	526.9	559.1

수량을 이용한 지역빈도해석과 본 연구를 통해 제시한 극치 강수량과 극치강수 발생시간 통계치를 동시에 고려한 지역 빈도해석 결과를 Table 4에 비교하였다. 앞에서 설명했듯이 시간정보를 포함한 경우 기존 방법과 비교해 군집 분류가 달라지는 관측소들이 일부 발생하기 때문에 동일한 조건의 맞비교는 아니지만, 극치강수량의 발생시간특성을 고려했을 때 달라지는 군집번호의 위치를 도시한 Fig. 11에서 각 군집에 해당하는 관측소의 재현기간별 확률강수량의 변화량을 통해 지역적인 확률강수량의 증감을 알 수 있다.

#### 4. 결 론

본 연구에서는 극치강수량의 발생 시기에 대한 정량적인 분석이 가능한 순환통계기법을 이용하여 관측 지점별 시간 통계량을 산정하고, 이를 극치강수량과 결합하여 시·공간적인 특성자료를 생성한 후 시공간적인 특성자료를 이용한 군집화 해석을 수행하였다. 그 과정에서 극치강수량 발생시기 통계량을 이용한 군집화 성능을 모의실험을 통해 확인하였으며, 극치강수량과 관측소 공간정보만을 이용한 군집화에 비해 극치강수 발생시기의 동질성을 고려한 분류가 가능한 것으로 분석되었다. 또한 K-means 방법에 의한 군집화 과정에서 초기 Centroid의 위치에 따라 군집화 결과가 조금씩 달라질 수 있는 오류를 방지하기 위해 반복적인 사전해석 단계에서 구해진 Centroid의 평균위치를 초기 Centroid 위치로 이용하였으며, 분류된 군집별 공간거리 합의 크기에 따른 군집번호의 부여로 군집간의 물리적 연관성이 인접하도록 설정하였다. 위와 같은 과정을 통해 본 연구에서는 우리나라 전체 지역을 5개의 군집으로 분류하였으며, 극치강수의 양적 특성만을 이용한 기존의 군집화 결과와의 비교를 통해 극치강수 발생 시기 속성을 이용한 군집화 결과가 지형 및 대권역에 따른 극치강수의 물리적 차이를 보다 합리적으로 반영한 결과를 도출하는 것으로 판단할 수 있었다.

#### 참고문헌

경민수, 김상단, 김보경, 김형수(2007) 군집분석을 통한 기쁨의

기쁨심도-기쁨면적-기쁨지속기간곡선의 작성, **대한토목학회 논문집**, 대한토목학회, 제27권, 제3B호, pp. 267-276.

권현한, 이정주(2010) 순환통계 분석을 통한 강수량 시계열의 시공간적 변동성 분석, **대한토목학회논문집**, 대한토목학회, 제30권, 제2B호, pp. 191-198.

김상욱, 이길성(2008) Bayesian 다중회귀분석을 이용한 저수량 (Low flow) 지역 빈도분석, **한국수자원학회논문집**, 한국수자원학회, 제41권, 제3호, pp. 325-340.

박상우, 전병호, 장석환(2003) 다변량 분석기법에 의한 지점강수의 권역화 연구, **한국수자원학회논문집**, 한국수자원학회, 제36권, 제5호, pp. 879-892.

신현석, 박무중(1999) 신경망을 이용한 우리나라의 시공간적 기쁨의 해석, **한국수자원학회논문집**, 한국수자원학회, 제32권, 제1호, pp. 15-29.

오테석, 문영일, 오근택(2008) 군집분석과 변동행렬도합수를 이용한 지역빈도해석의 확률강수량 산정, **대한토목학회논문집**, 대한토목학회, 제28권, 제2B호, pp. 225-236.

윤용남, 박무중(1997) L-Moment 법을 이용한 월 강수량 자료의 지역기쁨빈도 해석, **한국수자원학회논문집**, 한국수자원학회, 제30권, 제1호, pp. 55-62.

Berens, P. (2009) CircStat: A MATLAB Toolbox for Circular Statistics, *Journal of Statistical Software*, Vol. 31, Issue 10. ISSN: 1548-7660.

Fisher, N.I. (1995) *Statistical Analysis of Circular Data. Revised edition*, Cambridge University Press.

Hartigan, J.A. (1975) *Clustering Algorithms*, Wiley. MR0405726. ISBN: 0-471-35645-X.

Hosking, J.R.M. and Wallis, J.R. (1997) *Regional Frequency Analysis*, Cambridge University Press.

Jammalamadaka, S.R. and Sengupta, A. (2001) *Topics in Circular Statistics*, World Scientific.

Khan, S.S. and Ahmad, A. (2004) Cluster center initialization algorithm for K-means clustering, *Pattern Recognition Letters*, Vol. 25, No. 11, pp. 1293-1302.

Magilligan, F.J. and Graber, B.E. (1996) Hydroclimatological and geomorphic controls on the timing and spatial variability of floods in New England, USA, *Journal of Hydrologic Engineering*, Vol. 178, pp. 159-180.

Pewsey, A. (2004) The Large-Sample Joint Distribution of Key Circular Statistics, *Metrika*, Vol. 60, No. 1, pp. 25-32.

(접수일: 2010.8.2/심사일: 2010.9.14/심사완료일: 2010.11.16)