극치강수량의 시공간적 특성을 이용한 지역빈도분석

Regionalization of Extreme Rainfall with Spatio-Temporal Pattern

이정주*, 권현한**, 김병식***, 윤석영****

Jeong-Ju Lee, Hyun-Han Kwon, Byung-Sik Kim, Seok-Yeong Yoon

요 지

수공구조물의 설계, 수자원 관리계획의 수립, 재해영향 검토 등을 수행할 때, 재현기간에 따른 확률개념의 강우량, 홍수량, 저수량 등을 산정하여 사용하게 되며, 보통 대상지역의 장기 수문관측 자료를 이용하여 수문사상의 확률분포를 산정한 후 재현기간을 연장하여 원하는 설계빈도에 해당하는 양을 추정하게 된다. 미계측지역 또는 관측자료의 보유기간이 짧은 지역의 경우는 지역빈도 분석 결과를 이용하게 된다. 지역빈도해석을 위해서는 강우자료들의 동질성을 파악하는 것이 가장 기본적인 과정이 되며 이를 위해 통계학적인 범주화분석이 선행되어야 한다. 지점 빈도분석의 수문학적 동질성 판별을 위해 L-moment 방법, K-means 방법에 의한 군집분석 등이 주로 사용되며 관측소 위치좌표를 이용한 공간보간법을 적용하여 시각화하고 있다. 강수량은 시공간적으로 변하는 수문변량으로서 강수량의 시간적인 특성 또한 강수량의 특성을 정의하는데 매우 중요한 요소이다. 이러한 점에서 본 연구를 통해 강수지점의 공간적인 좌표 및 강수량의 양적인 범주화에 초점을 맞춘 기존 지역빈도분석의 범주화 과정에 덧붙여 시간적인 영향을 고려할 수 있는 요소들을 결정하고 이를 활용할 수 있는 범주화 과정을 제시하고자 한다. 즉, 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. 분석 과정에 있어서 시간속성의 정량화 및 일반화는 순환통계기법을 사용하였으며, 극치강수량과 발생시점의 속성자료는 각각의 평균과 표준편차를 이용하였다. K-means 알고리즘을 이용해 결합자료를 군집화 하고, L-moment 방법으로 지역화 결과에 대한 검증을 수행하였다. 속성 결합 자료의 군집화 효과는 모의데이터 실험을 통해 확인하였으며, 우리나라의 58개 기상관측소 자료를 이용하여 분석을 수행하였다. 예비해석 단계에서 100회의 군집분석을 통해 평균적인 centroid를 산정하고, 해당 값을 본 해석의 초기 centroid로 지정하여, 변동적인 클러스터링 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 경계선을 추출할 때 발생할 수 있는 오류를 방지하였다. 지역빈도분석 결과는 3차원 Spline 기법으로 도시하였다.

핵심용어 : 시공간적 변동성, 지역화, 순환 통계치, 강수

* 정회원・전북대학교 토목공학과 박사수료・E-mail : julee@jbnu.ac.kr
** 정회원・전북대학교 토목공학과 조교수・E-mail : hkwon@jbnu.ac.kr
*** 정회원・한국건설기술연구원 수자원연구실 선임연구원・E-mail : hydrokbs@kict.re.kr
**** 정회원・한국건설기술연구원 수자원연구실 연구위원・E-mail : syyoon@kict.re.kr

1. 서 론

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 빈도해석 결과를 추정하기 위해서 국내외적으로 많은 적용이 이루어지고 있다. 지역빈도해석을 위해서는 강우자료들의 동질성을 파악하는 것이 가장 기본적인 과정이 되며 이를 위해 통계학적인 범주화분석이 선행되어야 한다. 강우 발생의 계절적 변화와 집중에 대한 고려를 어떻게 반영할 것인가 하는 것은 기존의 강수량과 관측소 위치좌표만을 이용한 범주화에서 해결할 수 없는 문제이다. 즉, 기존 지역빈도해석을 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.

2. 모의실험

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

Table 1. Mean Value for Synthetic Data

Synthetic Data	Group A		Group B		Group C		Group D		S.D. 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

관측소별로 30개의 극치강수량 자료와 발생시간을 랜덤 발생하였다. x=5.5, y=5.5인 중앙선을 기준으로 사분면을 구분한 후 각 사분면에 평균값을 차등 지정하는 방법으로 구역의 양적 차이를 부여하였다. 랜덤발생 값의 범위는 표준편차로 지정하였으며, (a)Maximum Rainfall의 각 셀에 나타난 음영 값은 해당 셀(관측소)에 생성된 30개 극값 계열의 평균을 이용해 도시하였다.

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

Case 2는 강수량이 경계에 따라 차이가 있으며, 발생시간은 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.

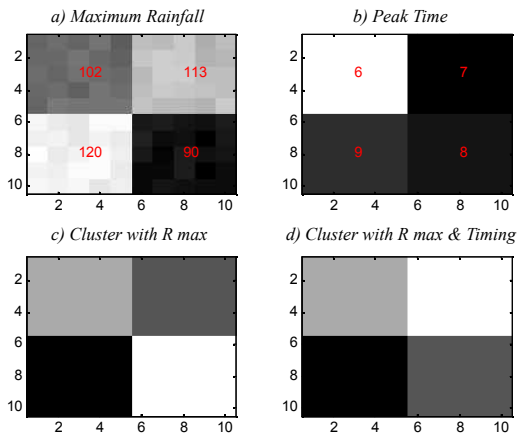

Figure 1. Synthetic Test Result of Case 1

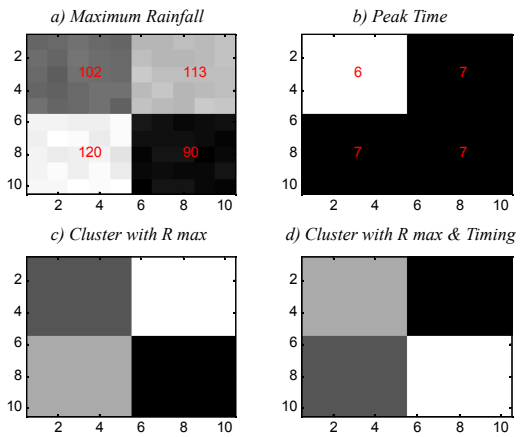

Figure 2. Synthetic Test Result of Case 2

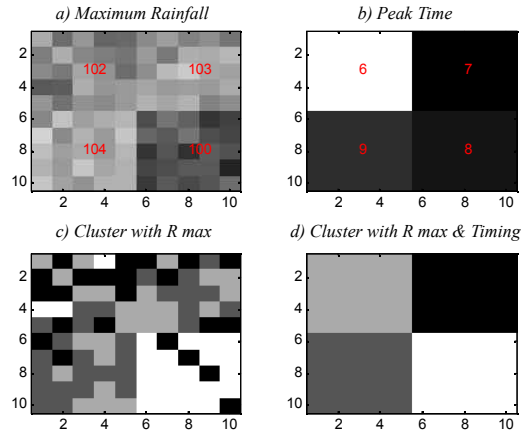

Figure 3. Synthetic Test Result of Case 3

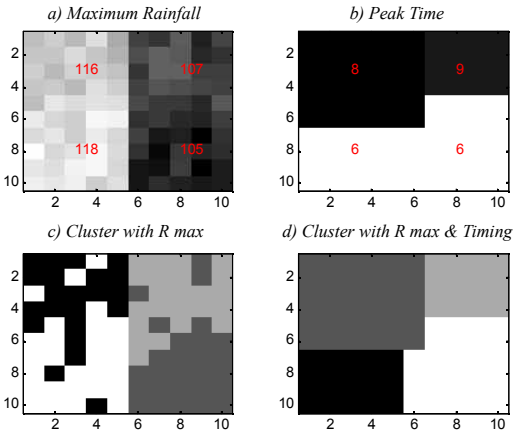

Figure 4. Synthetic Test Result of Case 4

Case 3은 강수량의 구역별 편차가 거의 없지만 발생시간은 구역별로 차이가 존재하는 경우로, 지역별로 각기 다른 시기에 최대 강우가 발생하였으나 그 양은 비슷한 경우를 가정하였다. 강수량에 의한 클러스터링이 심하게 오류가 발생하는 것에 비해 시간속성을 이용하여 보정되는 효과를 볼 수 있다.

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

3. 적용 및 결과

시공간속성 자료의 생성은 순환통계기법을 이용하였으며(Berens, 2009), 군집화는 K-means방법을 이용하였다(Hartigan, 1975; Hosking and Wallis, 1997). 순환통계 계산을 위해서는 원 자료

계열을 방향각으로 나타내야 한다. 자료계열을 Radian으로 환산한 방향값 α는 아래와 같이 계산할 수 있다.

$$\alpha = \frac{2\pi x}{k} \tag{1}$$

여기서 x는 원 데이터 값이며, k는 원 데이터의 계급값이다. 원자료에서 추출한 방향값을 이용하여 이동평균 등을 구하고자 할 때, angular scale에서의 평균은 방향값을 단순히 산술평균하는 것이 아니라 α에 대한 x, y방향의 단위벡터 r_i를 구한 후 벡터의 평균을 구해야 한다.

$$r_i = \begin{pmatrix} \cos\alpha_i \\ \sin\alpha_i \end{pmatrix} \tag{2}$$

$$\bar{r} = \frac{1}{N}\sum_i r_i \tag{3}$$

이로부터 평균벡터의 길이 R을 $\| \bar{r} \|$로부터 구할 수 있다(Fisher, 1995).

본 연구에서는 기상청에서 관리하는 관측소 중 장기간의 관측자료를 보유하는 58개 관측소를 대상으로 분석을 실시하였다. 자료의 연한은 관측소에 따라 상이하며 전체적으로 35년 이상의 자료를 보유하고 있다. Figure 5는 본 연구에서 사용된 강수지점의 공간적 위치를 나타낸다.

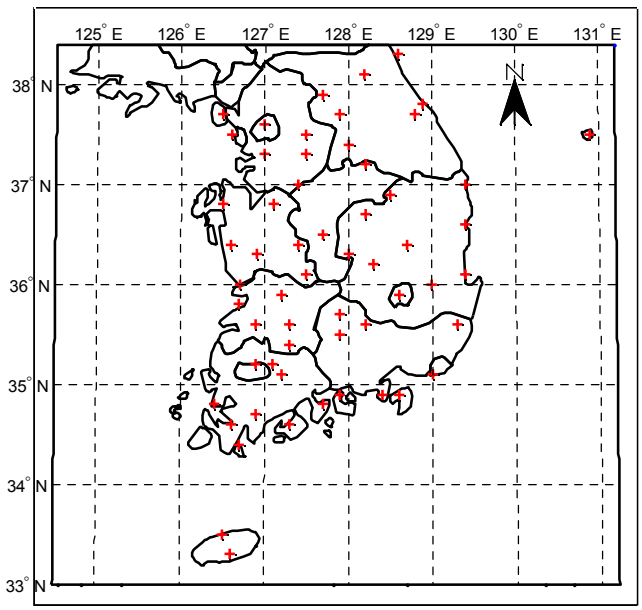

Figure 5. A map showing the stations for this study

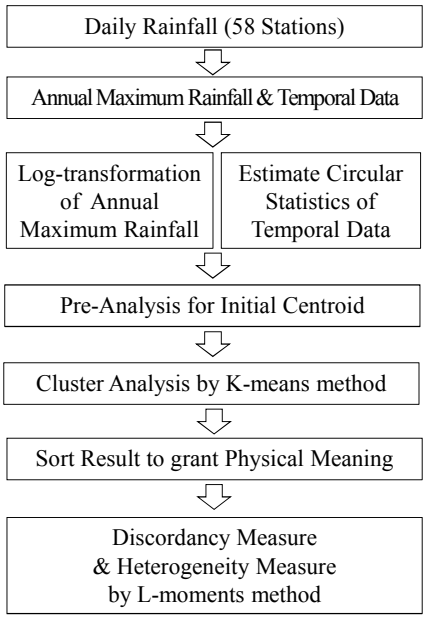

Figure 6. Work Flow

분석에 사용한 데이터는 각 강우관측소의 일자료를 이용한 극치강수량과 그 발생일자이다. 극치강수량은 Log Transform을 적용하였으며 발생일자는 순환통계분석 결과를 이용하였다. 예비해석 단계에서 100회의 군집분석을 통해 평균적인 Centroid를 산정하고, 해당 값을 본 해석의 초기 Centroid로 지정하여, 변동적인 클러스터링 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 경계선을 추출할 때 발생할 수 있는 오류를 방지하였다.

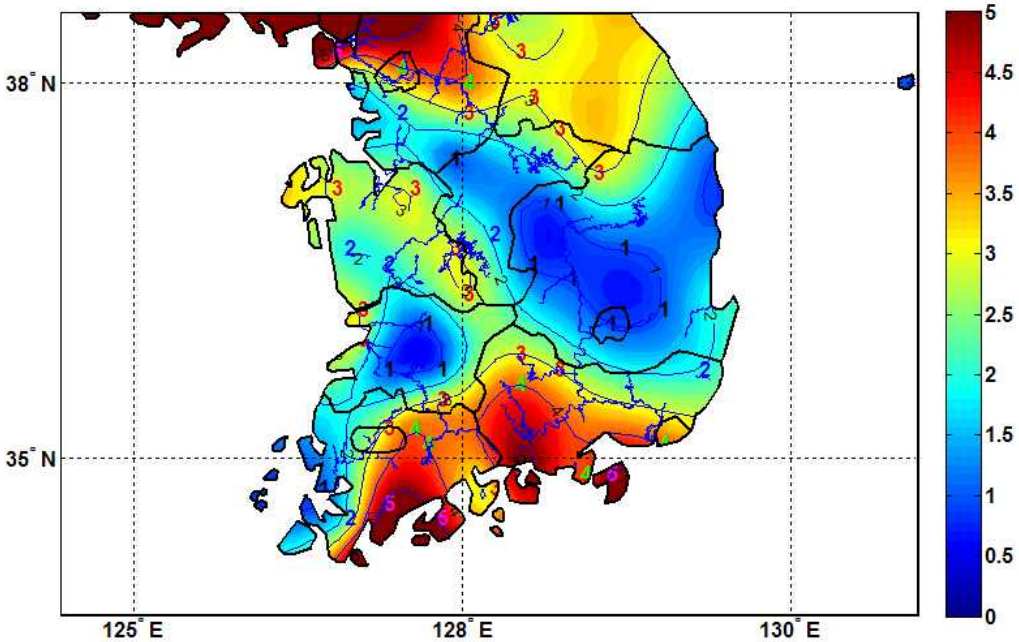

Figure 7. Reginalization with spatio-temporal attributes

본 연구에서는 전체 지역을 5개의 군집으로 분류하였으며, 서울, 대구, 광주의 경우 강우량 관측값이 인근지역과는 다른 양상을 보이는 경향이 반영된 것을 볼 수 있다. 강릉, 속초의 경우 넓은 지역에 너무 적은 관측소가 배치되어 있고, 강우의 양상도 다른 지역과는 크게 다르다. 수도권을 중심으로 하는 중부지역과 중부 이남의 서해안지역, 남해안지역과 낙동강 권역의 남부, 북부지역이 각각 다른 극치강수량의 분포 특성을 보이는 것을 볼 수 있다.

4. 결 론

본 연구에서는 극치강수량 자료와 극치강수 발생시점의 순환통계 특성치를 결합하여 K-means방법으로 우리나라 58개 관측소의 일최대강수계열에 대한 지역화를 수행하였으며, L-moment법을 이용한 군집화 결과의 검증을 통해 양호한 군집화 결과를 확인할 수 있었다.

참 고 문 헌

1. Berens, P., (2009) CircStat: A MATLAB Toolbox for Circular Statistics, Journal of Statistical Software, Vol. 31, Issue 10.
2. Fisher, N.I., (1995) Statistical Analysis of Circular Data. Revised edition. Cambridge University Press.
3. Hartigan, J. A. (1975). Clustering Algorithms. Wiley. MR0405726. ISBN 0-471-35645-X.
4. Hosking, J. R. M. and Wallis, J. R., (1997) Regional Frequency Analysis, Cambridge University Press