

통계적 방법에 의한 지점강우의 권역화 연구

A Study on the Regionalization of Point Rainfall by Statistical Methods

이정식* · 신창동** · 김영욱***1)

Lee, Jung Sik · Shin, Chang Dong · Kim, Young Wook

Abstract

The objective of this study is to analyze the regionalization of point rainfall by statistical methods for regional frequency analysis of the rainfall. The rainfall data used in this study are annual maximum rainfall at 57 stations during the period of more than 30 years for 12 durations(10min, 1, 2, 3, 4, 5, 6, 8, 10, 12, 18, 24hr) in Korea. The Mann-Whitney U test, Kruskal-Wallis one-way analysis of variance of nonparametric test, the principal component and the cluster analysis have been performed to analyze the regionalization of rainfall. The results of this study are as follows; (1) The region which hydrological homogeneous is accepted does not exist for whole duration in Korea. (2) The result of nonpametric test shows that hydrological homogeneous regions of point rainfall are divided by 5 regions. (3) In case of cluster analysis hydrological homogeneous regions of point rainfall are divided by 6 regions and 4 other areas.

key words : regionalization, regional frequency analysis, nonparametric test, cluster analysis

1. 서론

수공구조물의 설계시 설계기준으로 널리 이용되는 것은 확률강우량으로서 실제 구조물의 설계에 필요한 것은 설계홍수량이지만 이는 계획 대상지점에 충분한 기간동안 관측된 유량자료가 있어야만 산정이 가능하며, 대부분의 경우 유량계측지점이 적기 때문에 구조물의 설계기준으로 확률강우량이 많이 이용되고 있다. 따라서 장단기 수자원계획과 수공구조물의 설계를 위하여 확률강우량을 산정하는 일은 대단히 중요하며, 그동안 홍수량이 측정되어 있지 않은 미계측 지역에서의 수공구조물의 계획규모 결정은 일반적으로 강우량의 통계적 처리에 의하여 확률강우량을 산정, 이용하여 왔다. 확률강우량을 산정하는 빈도해석에 있어 지역빈도 해석은 자료기간이 짧은 지점이나 미계측 지점의 확률수문량을 결정하기 위해 제안된 방법으로, 우리나라와 같이 강우자료의 기록년수가 100년 미만인 경우에 지점빈도해석을 통해 산정된 재현기간 100년 이상인 확률강우량의 신뢰도가 떨어지는 단점을 보완할 수 있다(Institute of Hydrology, 1999). 지역빈도해석은 통계학적으로 동질한 지역에 적용하므로, 동질지역의 구분은 지역빈도해석 수행의 중요한 단계라 할 수 있다. 따라서 본 연구에서는 우리나라 기상청의 강우자료를 수집하여 각 지점의 강우와 강우관련 및 지형특성들을 분석 추출하고, 추출된 특성들의 지점간 상관성을 통계학적으로 분석할 수 있는 비모수적 검정과 군집분석을 실시하여 수문학적 동질성을 갖는 지점강우의 권역화를 수행하고자 한다.

2. 대상지역 및 기본자료

강우량과 같은 수문량을 수문학적으로 해석하는 경우 자료의 수집이 가장 중요한 문제이다. 수문자료의 확

* 정회원 · 금오공과대학교 토목환경공학부 · 교수 · E-mail : jungsik@kumoh.ac.kr

** 정회원 · 금오공과대학교 토목공학과 · 박사과정

*** 금오공과대학교 토목공학과 · 석사과정

물빈도를 통계적으로 해석하기 위해서는 자료들이 독립사상이어야 하며, 전체적으로 동질성을 가져야 한다. 강우량 기록년수는 일반적으로 30년 이상으로 하여야 하기에 본 연구에서는 국내의 수문관측 실정을 감안하고 관측소의 분포를 고려하여 최소 30년 이상의 기록년수를 가지고 있는 57개 지점의 지속기간별 매년 최대치계열 강우 및 적설자료 등의 기상자료와, 위도 및 경도, 해발고도와 같은 지형자료를 이용하였다. 섬 지역인 제주도와 울릉도는 자료 년수는 충분하나 강우특성이 내륙 및 내륙에 인접한 도서지역과 너무 상이하여 분석에서 제외하였다.

지속기간 10분 강우자료는 기상청의 기상예보를 이용하여 추출하였으며, 지속기간 1, 2, 3, 4, 5, 6, 8, 10, 12, 18, 24시간의 강우자료는 기상청 산하 각 관측소에서 매시간 측정된 강우자료를 이용하여 연간 최대치를 추출하였다.

3. 동질성검정

3.1 Mann-Whitney 검정

이 방법은 두 모집단간의 동질성을 비교하기 위한 비모수적 검정중의 하나로서, 모수적 검정인 t -검정의 대안으로 사용되고 있으며, 귀무가설 H_0 와 대립가설 H_A 는 다음과 같다.

H_0 ; 두개의 모집단은 동일하다.

H_A ; 두개의 모집단은 동일하지 않다.

표본의 크기 N 을 올림차순으로 정리한 후, 식 (1)과 같은 검정통계량에 의해 검정된다.

$$Z = \frac{U - (n_1 \cdot n_2 / 2)}{S} \quad (1)$$

자유도가 ∞ 인 t 분포는 정규분포와 동일하므로 임계치 Z_α 는 $t_{\alpha, \infty}$ 의 임계치와 같다. 따라서 양측검정(two-tailed test)에서 $|Z_\alpha| > t_{\alpha, \infty}$ 이면 귀무가설은 기각된다.

3.2 Kruskal-Wallis 검정

Kruskal-Wallis 검정방법은 k 개 모집단의 동질성검정에 사용되며, 귀무가설과 대립가설은 다음과 같다.

H_0 ; 연속인 k 의 모집단은 동일하다.

H_A ; k 개의 모집단이 모두 동일하지는 않다.

통계량 H 는 식(2)와 같다.

$$H = \frac{12}{N(N+1)} \sum_{j=1}^k \frac{R_j^2}{n_j} - 3(N+1) \quad (2)$$

여기서 k 는 표본의 개수, R_j 는 j 번째 표본의 순위합, n_j 는 j 번째 표본의 관측치의 수이며, N 은 총 관측치의 수($= \sum_{j=1}^k n_j$)이다.

3.3 군집분석

군집분석(cluster analysis)은 자료의 구조를 아는 유효한 수단으로서, 그 자료를 구성하고 있는 관측대상(혹은 속성)을 어떠한 기준에 의하여 분류하는 방법을 말한다. 그 종류는 무엇을 기준으로 해서 자료를 분류하느냐에 따라서 두 가지로 구분할 수 있다. 첫째, 관측대상간(혹은 속성간)에 정해지는 유사성(또는 거리)을 기초로 하나의 그룹으로, 전체를 몇 개의 그룹으로 분할하는 것이다. 둘째는 자료 및 그룹내의 분산개념을 기초로, 그룹간의 분리의 정도를 기준으로 해서 구분하는 것이다. 본 연구에서는 보다 실용적이라 통용되고

있는 전자의 방법으로 분석을 실시한다.

군집분석은 회귀분석 등과 달리 의미 없는 변수들을 제외시키는 기능이 없어 중요한 설명변수가 제외되거나 중복된 정보를 갖는 변수가 추가되는 경우 유사성거리가 제대로 측정되지 못한다. 즉, 군집분석에서는 선정된 변수들이 모두 동일한 비중으로 유사성 거리의 측정에 반영되므로, 본 연구에서는 다변량으로 이루어진 자료로부터 소수의 특징적인 변수를 찾아서 결합한 후에 자료를 분석하는 기법인 주성분분석(principal component analysis)을 선행하고 군집분석을 적용하였다.

개체의 군집화를 위한 유사성 거리의 측정방법에는 주로 유클리드 거리 또는 유클리드 제곱거리가 이용된다. 두 개체의 변수값 차이를 제곱하여 합산한 값으로써 다차원 공간에서 통계적 직선 최단거리를 나타내는 유클리드 제곱거리는 다음과 같다.

$$d(Q_1, Q_2) = \sum_{j=1}^m (X_{1j} - X_{2j})^2 \quad (3)$$

4. 적용 및 분석

4.1 Mann-Whitney 검정

본 연구에서는 하나의 지점을 기준으로 그 지점과 강우동질성이 있는 임의의 한 지점을 검정하는 Mann-Whitney 검정을 실시하여 강우 동질성이 존재하는 지점을 검정하였다.

전체적인 검정 결과를 보면, 지속기간이 변하여도 항상 높은 동질성을 가지고 있는 지점들이 보이는 반면 계속 낮은 동질성을 보이는 지점들도 나타났으며, 대표적으로 서산, 전주, 광주, 통영지점은 동질성을 보이는 지점이 상당히 많았으며, 포항, 추풍령, 강릉지점은 상대적으로 동질성을 보이는 지점이 적었다. 지속기간별로 보면 지속기간이 길어질수록 강우동질성이 존재하는 지점의 수가 늘어났다.

4.2 Kruskal-Wallis 검정

기준지점과 여러 지점 ($k=3\sim 57$)에 대한 강우동질성을 검정하기 위하여 Kruskal-Wallis 검정을 실시하였다. 먼저 전 지점에 대한 Kruskal-Wallis 검정을 실시하였을 때 전 지점에 대한 각 지속기간별 검정통계량은 비교 임계치와 큰 차이를 보여 우리나라 전역을 일률적으로 동일 분포로 하여 최적분포형을 결정할 수는 없었으나, Mann-Whitney 검정의 결과에 의해 나타나는 강우동질성 존재 지점들로 범위를 축소했을 경우 Kruskal-Wallis 검정을 실시한 결과를 보면 지속기간별로 그림 1과 2와 같이 몇 개의 권역으로 구분됨을 알 수 있었다.

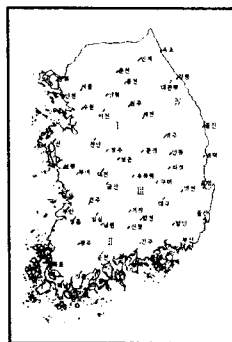


그림 1. 비모수적 검정을 통한 권역도(지속기간 10분)

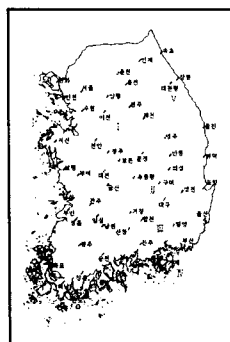


그림 2. 비모수적 검정을 통한 권역도(지속기간 1~24시간)

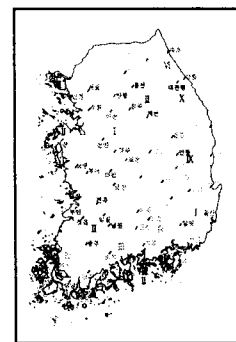


그림 3. 군집분석에 의한 권역구분도

그림 1과 2를 살펴보면 권역구분이 다른 형태를 보임을 알 수 있는데, 이는 강우의 지속기간이 짧은 강우의 경우 국지적인 호우의 형태가 많고 산발적이기 때문에 지속기간 1~24시간과는 다른 형태의 권역을 보였다. 따

라서 본 연구에서는 지속기간 10분을 제외한 모든 지속기간에서 권역의 구분이 유사한 형태를 보였기 때문에 권역설정시 지속기간 1~24시간에서는 동일한 권역도를 채택하였다.

4.3 군집분석

각 지점들의 수문학적 특성을 반영하는 인자들의 권역화 작업을 위하여 서로 상관관계가 높은 여러 개의 변수들을 조합하여 그 변수들의 정보를 가능한 많이 함축하고 있는 새로운 인위적 변수를 만들어내는 통계적 기법인 주성분분석을 채택하였다.

본 연구에서 적용한 주성분분석의 변수는 지속기간별 연 최대강우자료 등의 강우자료 14개와 강설자료 2개, 강우의 시간적 집중특성 4개와 위도, 경도, 해발 등의 지형자료 3개로서 총 23개이다. 표준화된 원변수의 상관계수행렬을 이용하여 원변수의 수만큼 주성분을 산출하였고, 중요주성분의 선택은 분산비율법에 의한 누가분산비가 80%를 넘는 성분까지 취한다는 경험적 기준을 적용하여 5개의 중요주성분이 선택되었다. 산출된 5개의 주성분점수를 이용하여 군집분석의 그룹화 작업을 수행하여 관측지점간의 상관성을 유클리디안 거리행렬로 나타내었으며, 거리행렬의 계수는 상이성의 크기를 나타내기 때문에 숫자가 작을수록 유사성이 높다고 해석할 수 있다. 수원을 기준으로 인천과는 0.436으로 유사성이 높게, 대관령은 44.476으로 상대적으로 유사성이 낮게 발생하였고, 전체적으로는 의성과 구미가 0.239로 가장 높게, 속초와 대관령이 46.478로 가장 낮게 발생하였다. 이러한 과정을 거쳐 군집분석에 의한 각 지점의 권역구분은 그림 3과 같다. 그림 3을 살펴보면 권역구분은 크게 6개 지역으로 나누어지며, 강화, 안동, 남해, 대관령 지점은 수문학적 강우특성이 인근 지역과 상당한 차이를 보여 기타 독립지점으로 구분하였다.

5. 결론

본 연구는 강우의 지역빈도해석에서 반드시 선행되어야 하는 수문학적 동질성을 갖는 지점강우의 권역화를 위해 우리나라 전역에 걸쳐 강우기록 년수 30년 이상의 기상청 산하 57개 지점을 선정하고, 각 지점에서 추출한 23개의 강우 및 지형자료를 이용하여 통계학적 분석방법인 비모수적 검정과 군집분석을 통해 권역을 구분한 결과, 다음과 같은 결론을 얻을 수 있었다.

- (1) 강우자료의 동질성검정에 의하여 전 지점에 대하여 강우동질성이 성립되지 않았으므로 동질성검정의 범위를 축소하여 검정을 해본 결과, 비모수적 검정으로 10분 지속기간에서는 4개의 권역, 지속기간 1~24시간에서는 5개의 권역으로 나눌 수 있었다.
- (2) 주성분분석에서 산출된 지점별 주성분 점수를 이용한 군집분석은 강우지점간의 수문학적 동질성 파악을 위한 지점간 강우정보의 유사성을 합리적이고 효율적으로 분석할 수 있게 하였으며, 군집분석의 결과로부터 지형학적 영향을 고려한 지점강우의 권역화는 우리나라의 강우지역을 6개의 권역과 4개의 기타지역으로 분류되었다.
- (3) 많은 수의 강우자료에 대한 주성분분석은 원자료의 정보와 특성을 극대화시킨 소수의 독립적인 주성분으로 함축할 수 있었으며, 부하행렬의 분석으로부터 각 주성분의 의미와 정보량의 함축정도를 파악할 수 있었다.

감사의 글

본 연구는 건설교통부가 출연하고 한국건설교통기술평가원에서 위탁 시행한 2003년도 건설핵심 기술연구개발사업(03산학연C03-01)에 의한 도시홍수 재해관리 기술연구사업단의 연구성과입니다.

참고문헌

1. 기상청 (1961-2005). 기상연보
2. 허준행, 이정식(2006). 도시지역의 확률강우량 산정, 도시홍수 재해관리 기술보고서 FFC05-07.
3. Institute of Hydrology (1999). Flood Estimation Handbook, Institute of Hydrology, Wallingford, UK.