

군집분석과 지역빈도해석을 이용한 확률강우량 추정에 대한 연구

Study of Rainfall Quantile Estimation using Cluster Analysis and Regional Frequency Analysis

정영훈*, 정창삼**, 남우성***, 허준행****

Younghun Jung, Chang-Sam Jeong, Woo Sung Nam, Jun-Haeng Heo

요 지

본 연구에서는 한강유역 109개 지점의 강우관측소에서 관측된 지속기간별 연최대강우량을 산정하고 지역빈도해석을 적용하기 위하여 한강유역에 대하여 지역구분을 실시하였다. 지역구분은 군집분석 방법인 Ward 방법, 평균연결법, Fuzzy-c means 방법, Two-Step 방법을 적용하였으며 군집분석을 수행하기 위해서 한강유역의 지점별 기상학적 인자와 지형학적 인자를 이용하여 군집분석을 수행하였다. 그 중 Fuzzy-c means 방법을 이용한 지역구분이 적합한 것으로 나타났다. 또한 모든 지속기간에 대하여 적합성 척도를 산정한 결과 GLO 분포형이 적정분포형으로 나타났으며, 지역빈도해석 방법인 지수홍수법을 이용하여 산정한 확률강우량과 지점빈도해석으로 산정한 확률강우량과 비교하여 적용성을 판단하였다.

핵심용어 : 군집분석, 지역빈도해석, 적정분포형, 확률강우량

1. 서 론

설계강우량을 산정하기 위한 가장 일반적인 방법은 빈도해석을 이용한 방법이며 그 중 지역빈도해석은 관측자료가 충분하지 않은 경우 자료의 불확실성의 오차를 감소하여 확률강우량을 산정할 수 있는 방법이다. 지역빈도해석은 Dalrymple(1960)이 지수홍수법(index flood method)을 처음으로 적용하였으며 꾸준한 연구가 진행되고 있다. 미국의 경우 홍수빈도해석 시 지역화된 왜곡도계수를 이용한 log-Pearson type III 분포를 추천하고 있으며, 영국의 경우는 적정분포형으로 Generalized Logistic(GLO) 분포를 추천하고 있다(Institute of Hydrology, 1999). 그러나 지역빈도해석을 수행하기 위해서는 모든 지점들이 수문학적 동질성을 가정하고 있으나 실제자료들은 지점들 간의 공간적 특성으로 지점간의 이질성이 증가하여 확률강우량 산정 시 오차가 증가되는 단점을 가지고 있다. 그렇기 때문에 지역빈도해석 시 동질 지역의 구분은 무엇보다 중요하다. 현재 동질 지역을 구분하기 위한 뚜렷한 기준이 제시되지 않았으며 다양한 방법들이 연구되고 있는 실정이다. 본 연구에서는 한강유역을 지역화 하기 위하여 4가지 군집 분석(Ward, 평균연결법(Average Liked Methodm ALM), Fuzzy-c means, Two-Step 방법)을 적용하였다. 그 중 Fuzzy-c means 방법을 이용한 지역구분이 적합하였으며 지점 및 지역빈도해석을 이용하여 확률강우량을 산정하였다.

* 정회원 · 연세대학교 대학원 사회환경시스템공학부 토목공학과 박사과정 · E-mail: yhjung2000@yonsei.ac.kr

** 정회원 · 인덕대학 건설환경설계학과 교수 · 공학박사 · E-mail: csjeong@induk.ac.kr

*** 정회원 · 연세대학교 대학원 사회환경시스템공학부 토목공학과 박사과정 · E-mail: nws77@yonsei.ac.kr

**** 정회원 · 연세대학교 사회환경시스템공학부 토목환경공학과 교수 · E-mail: jhheo@yonsei.ac.kr

2. 군집분석

군집 분석(Cluster analysis)은 개체들이 지니고 있는 다양한 속성의 유사성을 동질적인 집단으로 군집화하는 방법을 말한다. 군집분석 기본원리는 분석하고자 하는 여러 특성들을 유사성(similarity) 거리(distance)로 환산하고 거리가 상대적으로 가까운 개체들을 동질적으로 군집화 하는 것이다(박상우 et al., 2003). 지역빈도해석을 수행하기 위한 지역 구분 방법은 여러 가지가 있으나 비교적 객관적인 방법이라 할 수 있는 군집해석 방법이 활용된다. 군집해석 방법은 크게 계층적(hierarchical) 방법과 비계층적(nonhierarchical) 방법으로 구분되고, 계층적 방법에는 평균연결법, Ward 방법 등이 있으며, 비계층적 방법에는 K-means, Adaptive K-means, K-medoids, Fuzzy c-means 방법 등이 있다.

3. 지역빈도해석의 적용

3.1 대상구역의 지역구분

본 연구에서는 전체 한강유역 중 고미탄천, 금강산댐, 임진강 상류, 임진강 하류, 평화의 댐, 한탄강 유역을 제외한 유역을 대상지역으로 선정하였다. 대상지역의 지역구분은 기상학적 인자와 지형학적 인자를 이용하여 군집분석을 실시하였으며 군집 수를 결정하기 위하여 덴드로그램(dendrogram)을 이용하였다. 군집분석에서는 군집들의 유사성을 거리차로 환산하여 나타내기 때문에 거리 차에 따라 군집 수가 달라지므로 개체의 특성에 맞게 군집 수를 결정해야 한다.

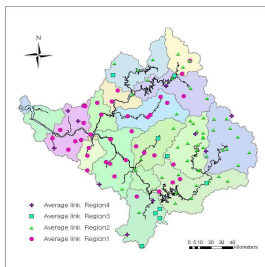


그림 1. 평균연결법을 이용한 지역구분

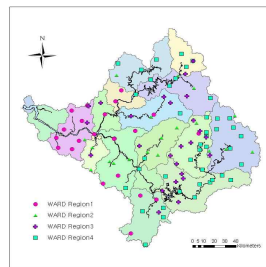


그림 2. Ward 방법을 이용한 지역구분

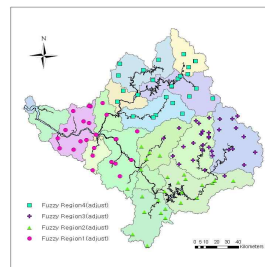


그림 3. Fuzzy-c means 방법을 이용한 지역구분

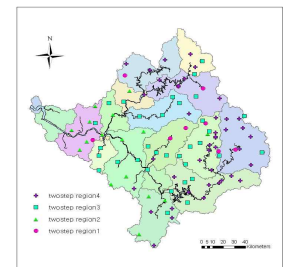


그림 4. Two-Step 방법을 이용한 지역구분

Hosking and Wallis(1997)가 제안한 군집내 최소 지점수 7개를 고려할 경우 한강유역은 4개 지역으로 구분 하는 것이 적절 하다고 판단되며, 이러한 군집해석 결과를 나타내면 Figs 1과 2와 같다. 비계층적 방법인 Fuzzy-c means 방법을 수행하기 위해서는 임의의 군집수를 정하거나, 계층적 방법에서 정한 군집 수를 목표 군집 수로 규정하여야 하는데 본 연구에서는 계층적 방법에서 정한 군집수 4개를 적용하여 분석하였다. 비계층적 군집방법은 하나의 지역에 속한 강우 관측소가 다른 지역과 중복 집단이 허용되므로 중복된 지점들에 대하여 소유역을 고려하여 군집결과를 수정하여 나타내면 Fig. 3과 같다. 계층·비계층적 방법인 Two-Step 방법은 계층적 군집분석과 비계층적 군집분석방법으로 얻어진 군집분석결과를 검토하기 위하여 각 개체들의 변수 값을 하나의 독립 변수로 가정하여 변수마다 개별 정규(Gaussian)분포를 가지도록 하고 거리측도는 로그우도함수를 사용하였다. 또한 Akaike 정보기준(AIC)을 사용하여 군집 수를 결정하여 분석하였으며 결과는 Fig. 6과 같다. 비계층적 방법인 Fuzzy-c means인 경우 처음단계의 군집결과가 좋지 못할 경우 다음 단계에서 적합한 다른 군집에 포함하여 다른 군집방법보다 명확하고 효과적인 군집결과를 보이는 것으로 나타났다. 따라서 본 연구의 대상지역인 한강유역에 적합한 군집분석방법으로 비계층적 방법인 Fuzzy-c means방법이 가장 적합하다고 판단된다.

3.2 적정분포형 선정

Gabriele와 Arenell(1991)에 의하면 지역화 과정에서 L-CV와 L-skewness의 값이 일정하도록 군집분석을 수행하였기 때문에 L-CV와 L-skewness의 변동폭이 작은 값을 가진다는 것은 지역구분이 잘된 것으로 볼 수 있으며 계층적 방법에 유리하다고 볼 수 있다. 적정분포형을 선정하기 위해 대상지역내 지점들의 불일치 척도(D_i)를 산정하였으며, 전반적으로 지역 구분된 자료가 동질하다고 판단된다. 군집화한 지역의 이질성 척도 H_1, H_2, H_3 을 산정한 결과 지역Ⅳ의 지속기간 1시간을 제외하면 모두 동질한 것으로 나타났다. 본 연구에서는 Generalized Extreme Value(GEV), Lognormal, Pearson Type III(PTⅢ), Generalized Pareto(GPA), Generalized Logistic(GLO) 등의 5가지 확률분포형을 적용하여 매개변수를 추정하고 적합도 검정을 수행하였다. 각 분포형은 적합성 척도가 절대값 1.64을 기준으로 판단하였다.

표 1. 지속기간별 적합성 척도를 만족하는 개수

지역	분포형				
	GLO	GEV	Lognormal	PTⅢ	GPA
I	5	5	4	1	0
Ⅱ	7	2	2	0	0
Ⅲ	5	1	0	0	0
Ⅳ	7	3	3	0	0
Total(Case)	24/28	11/28	9/28	1/28	0/28
Accept(%)	85.71	39.28	32.14	3.57	0.00

Table 1 은 전체 지속기간(1, 2, 3, 6, 9, 12, 24시간)에 대하여 분포형별 적합성 척도를 만족하는 지속기간의 개수를 나타낸 것이며 이를 총 분포형 개수로 나누어 적합성 척도 통과율(%)을 나타낸 것이다. 따라서 GLO 분포형을 적정분포형으로 선정하는 것이 적합하다고 판단된다.

3.3 확률강우량 산정

Figs 5~8은 분포형별 적합성 척도에 따라 적정분포형으로 선정된 GLO분포형에 대하여 지속기간 24시간의 확률강우량을 산정한 것이다. 4개 지역으로 구분된 지점에 대하여 지점빈도해석과 비교하여 확률강우량을 산정하였으며 실선으로 나타낸 것은 지점빈도해석의 결과이고 점선으로 나타낸 것은 지역빈도해석(지수홍수법)으로 산정한 확률강우량의 결과이다.

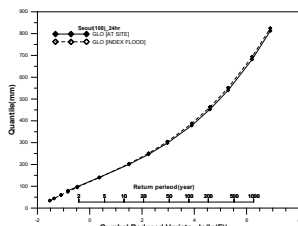


그림 5. 서울지점의 성장곡선(지역Ⅰ)

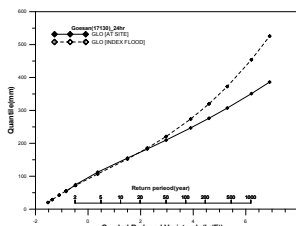


그림 6. 괴산지점의 성장곡선(지역Ⅱ)

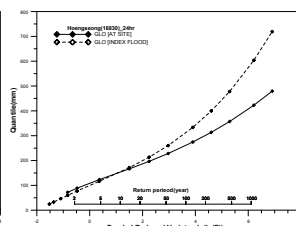


그림 7. 황성지점의 성장곡선(지역Ⅲ)

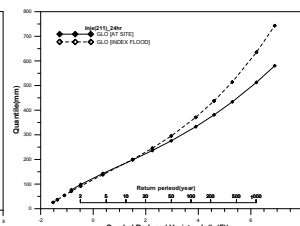


그림 8. 인제지점의 성장곡선(지역Ⅳ)

Fig. 5의 서울지점의 경우 지점빈도해석과 지역빈도해석으로 산정한 확률강우량의 결과와 큰 차이가 없는 것으로 나타났으며, Figs 6~8의 괴산, 황성, 인제지점의 경우 지역빈도해석으로 산정한 확률강우량이 비교적 높게 산정된 것을 알 수 있다. Table 2는 Figs 5~8에 나타낸 서울, 괴산, 황성, 인제지점에 대하여 지점빈도해석과 지역빈도해석으로 산정한 확률강우량을 정량적으로 나타낸 것이다.

표 2. GLO분포형에 대한 확률강우량(지속기간24시간)

Region	Station	Method	Nonexceedance probability							
			0.010	0.100	0.500	0.900	0.950	0.990	0.995	0.999
I	Seoul	At-site	42.17	79.54	140.45	246.33	297.90	452.86	540.42	812.61
		Index Flood	34.40	75.14	140.19	251.10	304.46	463.02	551.73	824.54
II	Goesan	At-site	1.31	53.85	112.33	182.77	209.87	276.18	307.14	386.10
		Index Flood	20.17	55.73	106.79	185.31	220.71	319.95	372.61	525.74
III	Hoengseong	At-site	33.57	71.95	123.16	196.61	228.26	313.59	357.25	479.49
		Index Flood	24.46	59.79	116.47	213.53	260.36	399.85	478.06	719.19
IV	Inje	At-site	24.95	76.29	143.01	236.28	275.82	380.92	433.98	580.52
		Index Flood	25.50	70.63	138.01	245.63	295.32	437.63	514.61	743.12

4. 결 론

본 연구에서는 한강 유역 109개 강우 관측소의 시우량 자료 등을 이용하여 지속기간별 자료를 구축하고 4가지 군집해석 방법(Ward, 평균연결법, Fuzzy-c means, Two-Step 방법)을 이용하여 지역구분을 하였다. 또한 설정된 지역에 대하여 동질성 검토 및 적합성 척도를 산정한 결과 비계층적 방법인 Fuzzy-c means 방법의 비교적 좋은 결과를 보이는 것으로 나타났으며, Fuzzy-c means 방법을 이용하여 구분된 4개 지역의 지속기간별 적합도 척도를 산정한 결과 GLO분포형이 적정분포형인 것으로 나타났다. 또한 4개 지역으로 구분된 4개 지점(서울, 괴산, 횡성, 인제 지점)에 대하여 지점빈도해석과 지역빈도해석으로 산정한 확률강우량과의 비교를 통하여 비교적 자료기간이 긴 서울지점의 경우 2가지 빈도해석의 방법으로 산정한 결과와 유사하게 나타났으며 그 외 지점(괴산, 횡성, 인제 지점)의 경우 지역빈도해석으로 산정한 확률강우량이 지점빈도해석으로 산정한 결과보다 크게 산정됨을 알 수 있었다.

감 사 의 글

본 연구는 21세기 프론티어연구개발사업인 수자원의 지속적 확보기술개발사업단의 연구비지원(과제번호:1-6-3)에 의해 수행되었습니다.

참 고 문 헌

1. 박상우, 전병호, 장석환(2003). 다변량 분석기법에 의한 지점강우의 권역화 연구. 한국수자원학회논문집, 한국수자원학회, 제36권 제5호, pp.879-892.
2. Darlymple, T.(1960). Flood-Frequency Analyses, Water Supply Paper 1543-A, U. S. Geological Survey, Reston, USA.
3. Gabriele, S.G. and Arnell, N.(1991). A Hierarchical Approach to Regional Flood Frequency Analysis, Water Resources Research, Vol. 27, No. 6, pp. 1281-1289.
4. Hosking, J.R.M. and Wallis, J.R.(1997). Regional Frequency Analysis. Cambridge University Press.
5. Institute of Hydrology(1999). Flood estimation handbook, Institute of Hydrology, Wallingford, UK.