

다변량 분석 기법을 활용한 강우 지역빈도해석

Regional Rainfall Frequency Analysis by Multivariate Techniques

남 우 성* / 김 태 순** / 신 주 영*** / 허 준 행****

Nam, Woo Sung / Kim, Tae Soon / Shin, Ju-Young / Heo, Jun-Haeng

Abstract

Regional rainfall quantile depends on the identification of hydrologically homogeneous regions. Various variables relevant to precipitation can be used to form regions. Since the type and number of variables may lead to improve the efficiency of partitioning, it is important to select those precipitation related variables, which represent most of the information from all candidate variables. Multivariate analysis techniques can be used for this purpose. Procrustes analysis which can decrease the dimension of variables based on their correlations, are applied in this study. 42 rainfall related variables are decreased into 21 ones by Procrustes analysis. Factor analysis is applied to those selected variables and then 5 factors are extracted. Fuzzy-c means technique classifies 68 stations into 6 regions. As a result, the GEV distributions are fitted to 6 regions while the lognormal and generalized logistic distributions are fitted to 5 regions. For the comparison purpose with previous results, rainfall quantiles based on generalized logistic distribution are estimated by at-site frequency analysis, index flood method, and regional shape estimation method.

keywords : Procrustes analysis, factor analysis, fuzzy-c means, regional frequency analysis

요 지

지역빈도해석을 통한 확률강우량 산정 결과는 수문학적으로 동질한 지역의 구분 결과에 따라 달라진다. 지역을 구분할 때에는 강우에 영향을 미치는 다양한 변수들이 사용될 수 있다. 변수의 유형과 개수가 지역 구분의 효율성을 좌우하기 때문에 활용 가능한 모든 변수들의 정보를 요약할 수 있는 변수들을 선택하는 것이 지역 구분의 효율성 면에서 유리하다고 할 수 있다. 이런 면에서 지역 구분의 효율성을 증대시킬 목적으로 다변량 분석 기법이 활용될 수 있다. 본 연구에서는 변수들 간의 상관관계를 바탕으로 모든 변수가 표현하는 정보를 대표할 수 있는 더 적은 수의 변수를 선정하는 기법으로 Procrustes analysis를 활용하였다. 이 기법을 활용하여 42개의 강우 관련 변수들을 21개로 줄일 수 있었다. 선정된 변수들을 바탕으로 요인분석을 수행하여 5개의 요인을 추출하였고, 이를 근거로 군집해석

* 연세대학교 대학원 토목공학과 박사과정

Ph. D. Candidate, Dept. of Civil Engrg., Yonsei University, Seoul 120-749, Korea
(e-mail: nws77@yonsei.ac.kr)

** 연세대학교 사회환경시스템공학부 BK21 연구교수

BK21 Researcher, School of Civil & Environmental Engrg., Yonsei University, Seoul 120-749, Korea
(e-mail: chaucer@yonsei.ac.kr)

*** 연세대학교 대학원 토목공학과 석사과정

M.D. Candidate, Dept. of Civil Engrg., Yonsei University, Seoul 120-749, Korea
(e-mail: ausran@yonsei.ac.kr)

**** 연세대학교 사회환경시스템공학부 교수

Professor, School of Civil & Environmental Engrg., Yonsei University, Seoul 120-749, Korea
(e-mail: jhheo@yonsei.ac.kr)

기법인 fuzzy-c means 기법을 활용하여 지역을 구분하였다. 68개 강우 관측 지점을 대상으로 지역을 구분한 결과 6개의 지역으로 구분되었다. 6개의 지역에서 GEV 분포가 적합한 것으로 나타났고, 3변수 대수정규 분포와 generalized logistic 분포가 5개 지역에서 적합한 것으로 나타났다. 기존 연구 결과와의 비교를 위해 generalized logistic 분포를 바탕으로 지점빈도해석, 홍수지수법, 지역형상추정법을 적용하여 확률강우량을 산정하였다.

핵심용어 : Procrustes analysis, 요인분석, fuzzy-c means, 지역빈도해석

1. 서론

지역빈도해석은 자료 기간이 짧은 경우 확률수문량의 불확실성이 커지는 단점을 가진 지점빈도해석을 보완한 방법이다. 우리나라와 같이 강우 관측 기간이 100년 이하로 짧은 지역에서는 100년 이상의 재현기간에 대한 확률강우량 산정에 지역빈도해석을 적용함으로써 확률강우량 산정의 신뢰도를 향상시킬 수 있다.

지역빈도해석은 Darlymple(1960)이 홍수지수법(index flood method)을 적용한 이래 다양한 연구가 진행되어 왔다. 이러한 연구를 통해 지역빈도해석이 지점빈도해석 보다 더 정확한 설계 수문량 산정을 가능하게 한다는 사실이 밝혀졌다(Lettenmaier and Potter, 1985; Potter, 1987; Cunnane, 1989; Hosking and Wallis, 1997). Cunnane(1989)과 Hosking and Wallis(1997)는 심지어 이질성이 강한 지역에서도 지역빈도해석이 지점빈도해석 보다 더 정확함을 보였다.

지역빈도해석은 대개 홍수량 자료에 적용되어 왔지만, 최근에는 강우 빈도해석에도 널리 활용되고 있다. Smithers and Schulze(2001)는 Hosking and Wallis(1997)가 제안한 지역 L-모멘트 알고리즘을 남아프리카의 강우 자료에 적용했고, Sveinsson et al.(2002)은 홍수지수법을 적용하여 미국 콜라라도 일부 지역에 대한 강우 빈도곡선을 제안했다. Christopher et al.(2005)은 강우 지역빈도해석을 통해 미국 미시간 주의 IDF 곡선을 보완하였다.

지역빈도해석에서 가장 중요한 단계는 동질 지역의 구분이라 할 수 있다. 동질 지역 구분을 위한 정해진 기준은 없기 때문에 이를 위한 다양한 기법과 변수들이 활용되어 왔다. 일례로, Mallants and Feyen(1990)은 일강우량 자료만을 활용하여 지역을 구분하였고, Guttman(1993)은 7개의 지형 및 기후와 관련된 변수를 바탕으로 지역을 구분하였다. Zhang and Hall(2004)은 몇 가지 군집해석 기법을 활용해서 지역을 구분했고, Dinpashoh et al.(2004)은 주성분 분석, Procrustes analysis, 요인분석을 활용하여 지역을 구분함으로써 동질 지역 구분의 효율성 향상에 대해 연구하였다.

본 연구에서는 동질 지역 구분에 활용되는 구분인자들을 좀 더 효율적으로 선택하기 위해 Procrustes analysis를 통해서 원래 변수의 성질을 충분히 설명할 수 있는 최소 개수 변수를 선정하고, 요인분석(factor analysis)과 군집해석(cluster analysis)을 적용하여 구분된 동질 지역을 바탕으로 확률강우량을 산정하여 허준행 등(2007)의 결과와 비교함으로써 본 연구에서 활용한 방법의 국내 적용성을 검토하는데 그 목적이 있다.

2. 다변량 분석 기법을 활용한 지역구분

2.1 Procrustes Analysis

Procrustes analysis (Krzanowski, 1987)는 다변량 해석을 위해서 변수를 선택하는 과정을 모의하는 방법의 하나로서 원래 설정된 p 개의 변수를 이용해서 구한 주성분분석 점수(principal component score) 중에서 k 차원에 해당하는 점수와 설정된 변수 중에서 적절한 개수를 제외한 q 개의 변수를 이용해서 구한 주성분분석 점수 중 k 차원에 해당하는 점수를 비교해서, 두 개 차원의 점수의 차이를 최소화시키는 q 개의 변수를 찾아내는 기법이다.

Fig. 1은 Procrustes analysis의 절차를 나타낸 것으로서 모든 변수를 가지고 있는 원래의 행렬을 $X_{(n \times p)}$ 라고 할 때, 원래 변수의 성질을 최대한 반영하는 최소한

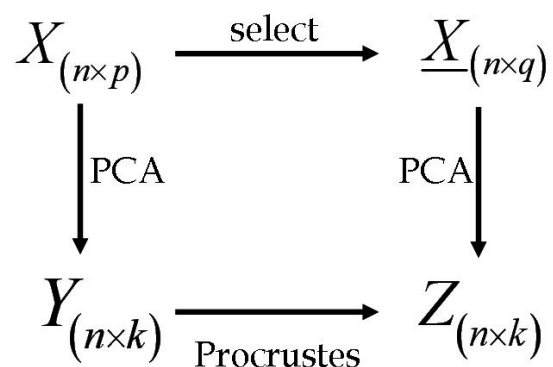


Fig. 1. Variable Selection by Procrustes Analysis (Krzanowski, 1987)

의 변수를 가지는 행렬을 $X_{(n \times q)}$ 라 하고, 원래 변수로부터 구한 주성분분석 점수 행렬 중 k 차원의 행렬 $Z_{(n \times k)}$ 와 $Y_{(n \times k)}$ 을 비교해서 그 차이가 최소가 되도록 만드는 것을 의미한다.

$Z_{(n \times k)}$ 와 $Y_{(n \times k)}$ 을 비교하기 위해서는 아래의 M^2 으로 정의된 제곱오차합(sum of squared differences)이 최소가 되도록 하는 차원수를 구하면 된다.

$$M^2 = \text{Trace}\{YY' + ZZ' - 2\Sigma\} \quad (1)$$

여기서, '은 전치(transpose)를 의미하고, Σ 는 행렬 $Z'Y$ 의 SVD (singular value decomposition)를 통해서 구한 대각행렬(diagonal matrix)을 가리킨다. $Z'Y$ 는 아래와 같이 정의된다(Bunch and Nielson, 1978).

$$Z'Y = U\Sigma V' \quad (2)$$

여기서, U 는 $k \times k$ 행렬로서 $UU' = I_k$ 을 만족시키는 행렬이다. V 는 $k \times k$ 행렬로서 $V'V = VV' = I_k$ 를 만족시키는 행렬이다. 여기서, I_k 는 $k \times k$ 단위행렬(identity matrix)을 의미한다.

본 연구에서는 이러한 이론을 바탕으로 Dinpashoh et al. (2004)이 활용한 Procrustes analysis 알고리즘을 근거로 변수를 선정하였다.

2.2 요인분석

요인분석은 변수가 많은 경우에 군집해석의 효율성 저하 문제를 해결하기 위해 주로 사용되는 방법이다. 요인분석은 기존 변수들의 상관성을 이용하여 요인(factor)이라 불리는 변수 내의 공통적인 새로운 변수를 도출함으로써 이 요인들이 가지고 있는 특성으로 전체 자료의 특성을 최대한 설명하는 기법이다.

본 연구에서는 Procrustes analysis에 의해 선정된 변수를 바탕으로 통계 패키지인 SAS(조인호, 2004)를 활용하여 요인을 선정하였으며, 요인분석 수행에 필요한 회전 방법은 varimax normalized rotation (Overall and Klett, 1973)을 적용하였다. 이 회전 방법은 기후 자료에 대해 가장 적합한 회전방법으로 알려져 있다 (Puvaneswaran, 1990; White et al., 1991).

2.3 군집해석

Hosking and Wallis(1997)에 의하면 기존의 연구에 의하면 지역 내 지점수가 증가할수록 확률수문량 추정 의 정확도가 높아지다가 20개 정도의 지점수를 기점으로

로 지점간 종속성으로 인하여 정확도 저하가 나타나는 것으로 알려져 있다. 따라서 해석 초기에 지역 내 지점수를 결정할 수 있는 장점이 있고, 통계적으로 최적값을 도출하는 것으로 알려져 있는 fuzzy-c means 기법을 적용하였다(Bezdek, 1981). Fuzzy-c means 기법은 fuzzy 기법 중 가장 널리 사용되는 방법으로, Dunn (1973)에 의해 개발되었고, Bezdek(1981)이 Dunn의 알고리즘에서 사용된 fuzzy 목적함수를 가중 성분을 도입하여 일반화 하였다.

Fuzzy 군집해석에서는 한 지점이 하나의 군집에만 포함되는 것이 아니라 정도의 차이를 가지고 여러 개의 군집에 포함된다. 따라서 어떤 군집의 경계에 위치하는 지점은 군집의 중심에 있는 지점보다 더 작은 영향력으로 그 군집에 속하게 된다. 임의의 지점 x 에 대해 k 번째 군집에 속하는 정도를 나타내는 계수를 $u_k(x)$ 라 하면, 이러한 계수들의 합은 보통 1이 된다. 따라서 $u_k(x)$ 는 임의의 군집에 포함되는 확률을 뜻하게 된다.

Fuzzy c-means 기법은 보통 다음의 절차를 따른다.

- (1) 군집의 개수를 결정한다.
- (2) 각 지점에 무작위로 각 군집에 대한 계수를 할당한다.
- (3) 각 군집에 대한 도심(centroid)을 계산한다.
- (4) 각 지점에 대해 각 군집에 대한 계수를 계산한다.
- (5) 계수의 변화가 일정한 범위에 들어올 때까지 (3), (4)의 과정을 반복한다.

3. 다변량 분석 기법에 의한 지역빈도해석 적용

3.1 대상유역 및 자료

본 연구에서는 전국의 기상청 산하 68개 강우 관측 지점의 강우자료를 바탕으로 분석을 수행하였다. 68개 지점은 우리나라 전역에 분포되어 있으며, 자료 기간은 10~88년에 이른다. 빈도해석에 활용되는 자료는 연최대 강우량 자료이므로 68개 지점에서 지속기간별 연최대 자료를 추출하였다. 또한 동질 지역 구분을 위해 Table 1과 같이 강우에 영향을 미치는 42개의 변수를 68개 강우 관측 지점에 대해 수집하였다.

3.2 동질 지역 구분

앞서 설명한 것처럼 Dinpashoh et al.(2004)이 제안한 Procrustes analysis 알고리즘과 요인분석 및 군집해석을 활용하여 지역 구분을 위한 변수를 선정하고 그 변수를 바탕으로 동질 지역을 구분하였다.

Table 1에서 제시한 변수들 간의 상관성을 파악하여

Table 1. Climatological and Geographical Candidate Variables Related with Precipitation

Variable	Description
LAT	LATITUDE
LONG	LONGITUDE
ALT	ALTITUDE
MAP	Mean Annual Precipitation
DayP	Number of Days with Precipitation in a year
APM _i , i = 1, 2, ..., 12	Average Precipitation in a Month
DP _i , i = 1, 2, ..., 12	Number of Days with Precipitation in a month
MDP _i , i = 1, 2, ..., 12	Average Maximum Daily Precipitation in a month
AMaxMDP	Average Maximum of Maximum Daily Precipitation in a month

Table 2. Factor Loadings for the 21 Variables Selected by Procrustes Analysis

	F1	F2	F3	F4	F5
LONG	-0.0696	-0.0532	-0.5601	0.2492	0.6597
LAT	-0.7715	-0.0965	0.2231	0.1765	0.3812
ALT	-0.1270	0.1554	0.1862	0.0394	0.8109
MAP	0.7068	0.3484	0.5046	0.3055	0.0916
DayP	0.1320	0.9187	0.2047	-0.0958	0.1336
APM2	0.3696	0.8245	-0.2905	0.1945	0.0597
APM3	0.8245	0.4669	-0.1045	0.2167	-0.0013
APM4	0.9549	0.0740	0.1087	0.0838	-0.0930
APM5	0.8995	0.0407	0.2574	0.1331	-0.0350
APM6	0.9068	-0.0695	0.2199	0.0155	-0.0191
APM7	0.1486	-0.2174	0.8996	-0.1321	0.1468
APM8	0.0006	-0.0554	0.9079	0.2639	0.0415
APM11	0.1874	0.8281	-0.0055	0.4230	-0.0768
APM12	0.0235	0.9552	-0.1620	0.0327	-0.0368
DP2	0.1217	0.9409	-0.2282	-0.0178	0.0289
DP4	0.7510	0.4949	-0.0339	-0.1579	0.1144
DP07	-0.0944	-0.2682	0.7487	-0.2071	0.4469
MDP3	0.8480	0.1925	0.1223	0.2913	-0.0913
MDP7	0.2216	-0.1085	0.8783	-0.0365	-0.0635
MDP8	0.1895	0.1012	0.7519	0.4164	-0.1963
MDP9	0.2073	0.1233	0.0689	0.8676	0.1621
Cumulative variance	32%	58%	82%	91%	100%

42개 변수의 특성을 표현할 수 있는 변수 21개를 Procrustes analysis에 의해 선정하였으며 선정된 변수는 Table 2와 같다. 선정된 변수를 바탕으로 군집해석의 입력 자료로 사용할 5개의 인자를 요인분석을 활용하여 Table 2와 같이 선정하였다. 요인분석은 SAS(조인호, 2005)를 활용하였고, SAS에서의 요인분석 수행시 선정된 21개 변수들은 SAS 자체적으로 표준화 과정을

거치게 되므로 따로 변수들의 표준화 과정을 수행하지는 않았다. Table 2에서 확인할 수 있는 것처럼 요인 F1은 APM3, APM4, APM5, DP4, MDP3에 큰 인자적재(factor loading)가 부여되어 있으므로, 이 요인은 3, 4, 5월경의 강우와 관련되어 있다고 할 수 있다. 요인 F2는 DayP, APM2, APM11, APM12, DP2에 큰 인자적재가 부여되어 있어, F2가 강수일수와 겨울철 강우와

관련된 요인임을 알 수 있고, 요인 F3는 MAP, APM7, APM8, MDP7에 큰 인자적재가 부여되었으므로, 연평균강수량과 여름철 강우에 관련된 요인이라 할 수 있다. 요인 F4는 MDP9에 큰 인자적재가 부여되어 9월의 평균 최대 일강우량에 관련된 요인임을 알 수 있고, 요인 F5는 LONG과 ALT에 큰 인자적재가 부여되었으므로 관측소의 위치와 관련된 요인임을 알 수 있다. 이처럼 요인분석은 의해 상관성이 높은 변수들을 묶어 하나의 요인을 구성함으로써 변수들 간의 상관관계 파악에 도움이 되고, 많은 변수를 처리하는 데 들어가는 시간과 노력을 줄여주며, 이어서 수행되는 군집해석의 효율성 측면에서도 도움을 준다고 할 수 있다.

Fuzzy-c means 기법을 적용하여 다양한 개수로 지역을 구분해 본 결과, Table 3과 같이 6개로 구분했을 때 거리상으로 근접한 지점끼리 지역이 형성되고, 각 지역에도 고루 지점들이 분포되었다. 구분된 지역을 지도상에 표시하면 Fig. 2와 같고, 이는 허준행 등(2007)에서 구분한 지역인 Fig. 3과 유사했다. 허준행 등(2007)에서는 크게 5대강을 중심으로 한 지역과 남해안, 동해안으로 구분하고, 5대강 지역을 세분하여 14개의 지역으로 구분하였다. 본 연구에서의 지역 구분도 한강, 금강, 낙동강, 남해안, 동해안, 제주도 지역으로 구분하여 전체적으로 허준행 등(2007)과 유사한 지역 구분을 보였다. 다만 허준행 등(2007)에서는 제주도가 남해안

지역에 포함되었고, 섬진강 지역이 있었지만, 본 연구에서는 제주도 자체로 하나의 지역이 형성되었고, 섬진강 지역은 남해안 지역에 포함되는 차이를 보였다. 이는 본 연구에서 활용한 자료가 기상청 관측소에 국한되어 있어 허준행 등(2007)에 비해 지점수의 차이가 크기 때문인 것으로 판단된다.

구분된 지역에 Hosking and Wallis (1997)가 제안한 이질성 척도(heterogeneity measure)를 적용하였다. 이질성 척도 H 는 $H < 1$ 일 경우 "acceptably homogeneous", $1 \leq H < 2$ 일 경우 "possibly heterogeneous", $H \geq 2$ 일 경우 "definitely heterogeneous"로 제안하고 있다. 이러한 기준을 근거로 Table 4의 결과를 분석하면, 지역 3을 제외하고 5개 지역이 동질하고, 지역 3은 이질한 지역일 가능성이 있는 지역임을 알 수 있다.

3.3 확률강우량

6개의 구분된 지역에 Table 5와 같이 Gumbel, GEV, 3변수 대수정규(log normal), Pearson type III, generalized Pareto, generalized logistic 분포를 적용하였다. Hosking and Wallis (1997)가 제안한 적합성 척도를 적용한 결과 유의수준 5%에서는 GEV 분포가 6개 지역에서 통과되어 적정 분포형인 것으로 나타났으며, 3변수 대수정규 분포와 generalized logistic 분포

Table 3. Precipitation Regions

Region	1	2	3	4	5	6
No. of sites	21	14	7	14	4	8
Sites	Icheon Hongcheon Wonju Suwon Yangpyeong Chuncheon Seoul Jecheon Incheon Cheolwon Boeun Ganghwa Boryeong Cheongju Cheonan Chunyang Yeongju Seosan Taebaek Daejeon Inje	Namwon Jeongeup Jeonju Imsil Gunsan Buan Geumsan Gwangju Chupungryeong Jangsu Mokpo Geochang Buyeo Hapcheon	Yeongcheon Daegu Gumi Andong Uiseong Mungyeong Milyang	Tongyeong Goheung Jangheung Busan Masan Jinngju Yeosu Wando Suncheon Haenam Sancheong Ulsan Namhae Geojae	Seogwipo Seongsanpo Jeju Jejugocheung	Sokcho Gangneung Ulleungdo Daegwallyeong Yeongdeok Donghae Uljin Pohang

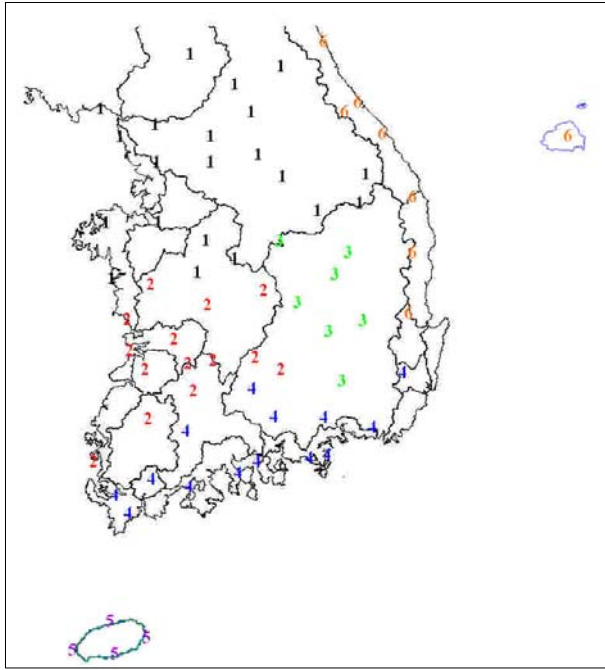


Fig. 2. Precipitation Regions

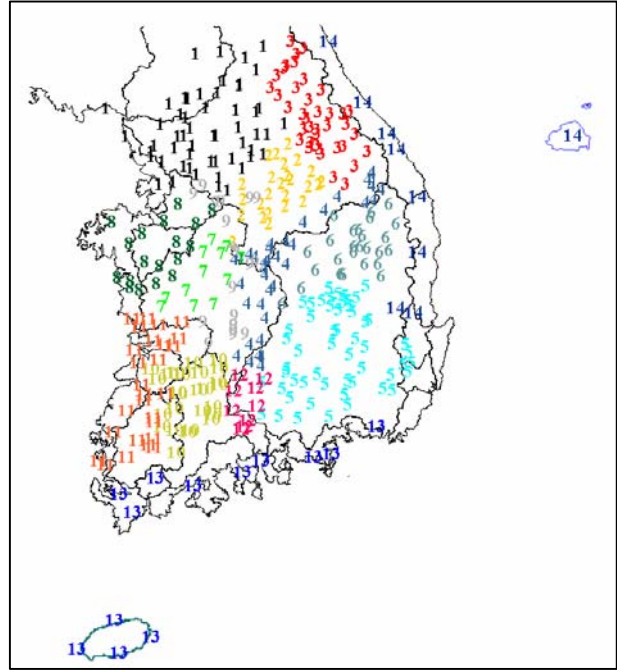


Fig. 3. Precipitation Regions (허준행 등, 2007)

Table 4. Heterogeneity Measure for Each Region (Duration: 24hr)

Region	Heterogeneity Measure	H	
		Value	Heterogeneity
1		-1.99	Homogeneous
2		-0.04	Homogeneous
3		1.42	Possibly Heterogeneous
4		-0.78	Homogeneous
5		-0.40	Homogeneous
6		-0.19	Homogeneous

Table 5. Goodness-of-fit Measure for Each Region

Region	Gumbel		GEV		Log normal-3		PT III		GPA		GLO	
	Z	Result	Z	Result	Z	Result	Z	Result	Z	Result	Z	Result
1	-2.25	N.G.	-0.33	O.K.	-1.12	O.K.	-2.55	N.G.	-4.54	N.G.	1.38	O.K.
2	-1.68	N.G.	-1.06	O.K.	-1.50	O.K.	-2.41	N.G.	-5.38	N.G.	0.83	O.K.
3	-0.89	O.K.	-1.45	O.K.	-1.45	O.K.	-1.80	N.G.	-5.23	N.G.	0.37	O.K.
4	-2.60	N.G.	-0.91	O.K.	-1.59	O.K.	-2.79	N.G.	-4.29	N.G.	0.44	O.K.
5	1.19	O.K.	0.88	O.K.	0.88	O.K.	0.68	O.K.	-1.28	O.K.	1.92	N.G.
6	-2.45	N.G.	-1.25	O.K.	-1.72	N.G.	-2.55	N.G.	-3.51	N.G.	-0.35	O.K.
No. of accepted	2		6		5		1		1		5	

도 5개 지역에서 적합한 것으로 나타났으며, 참고로 유의수준 2.5%에서는 3변수 대수정규 분포와 gene-

ralized logistic 분포도 6개 지역 전체에 대해서 적합한 것으로 나타났다.

허준행 등(2007)에서는 generalized logistic 분포가 적정 분포형으로 선정된 바 있다. 본 연구에서는 허준행 등(2007)의 결과와 비교를 위해 허준행 등(2007)에서 주요 지점으로 선정되었던 지점 중 3개 지점의 확률강우량을 generalized logistic 분포를 적용하고, 지점빈도 해석, 홍수지수법, 지역형상추정법(regional shape estimation method)으로 산정·비교하였다.

3개 지점의 확률강우량을 허준행 등(2007)과 비교하면 Table 6과 같고, 이를 도시하면 Figs. 4 ~ 6과 같다. 그림에서 x축은 Gumbel reduced variate로서 Eq. (3)에 의해 계산되고, y축은 확률강우량을 나타낸다.

$$y = -\ln(-\ln(F)) \quad (3)$$

여기서, y 는 Gumbel reduced variate, F 는 비초과확률을 의미한다. 또한 그림에서 GLO [AT SITE]는 generalized logistic 분포를 적정 확률분포형으로 한 지점빈도해석(at-site frequency analysis) 결과이고, GLO [RSE]와 GLO [INDEX FLOOD]는 generalized logistic 분포를 적정 확률분포형으로 하여 지역형상추정법과 홍수지수법을 적용한 결과이며, PRE는 이전 연구인 허준

행 등(2007)의 확률강우량 산정 결과이다.

Table 6에서 허준행 등(2007)의 홍수지수법 적용 결과와 본 연구의 홍수지수법 적용 결과를 비교해 보면, 전체적으로 본 연구 결과가 다소 작은 값을 나타내고 있다. 이는 세 지점이 포함되는 지역이 변경되었기 때문인데, 허준행 등(2007)에서는 기상청 지점뿐만 아니라 건교부와 수공의 강우 관측 지점 자료도 포함되어 본 연구에서 활용된 자료보다 한 지역에 더 많은 지점이 포함되게 된다. 강릉의 경우 두 연구 결과가 거의 유사한 값을 보이는데, 이는 허준행 등(2007)에서 강릉이 포함된 지역과 본 연구에서 강릉이 포함된 지역이 유사하게 구성되어 그 지역에 포함된 지점이 거의 유사하기 때문이다.

서울 지점의 경우(Fig. 4), 허준행 등(2007)에서 서울이 포함된 지역과 유사한 지역에 포함되어 있어 두 연구의 결과가 유사하게 나타났다. 또한 서울은 지점빈도 해석과 홍수지수법 및 지역형상추정법에 의한 확률강우량 산정 결과가 큰 차이가 없는 것으로 나타났는데, 이는 서울 지점의 자료 기간이 75년으로 비교적 긴 자료 기간을 보유하고 있기 때문인 것으로 판단된다. 이는 자료 기간이 길 경우 지점빈도해석에 의한 결과의 신뢰

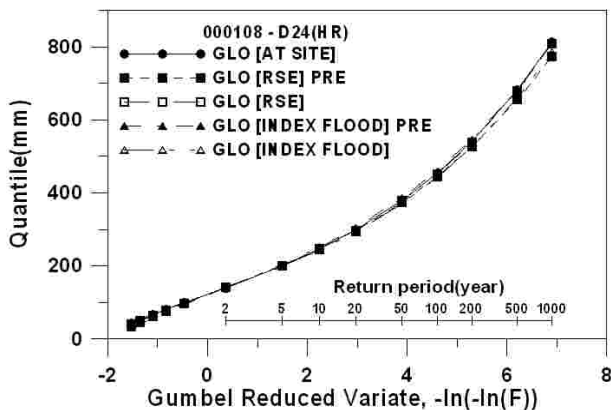


Fig. 4. Estimated Rainfall Quantiles of Seoul

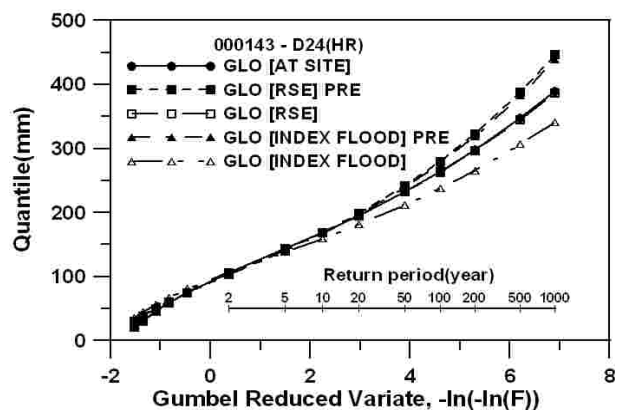


Fig. 5. Estimated Rainfall Quantiles of Daegu

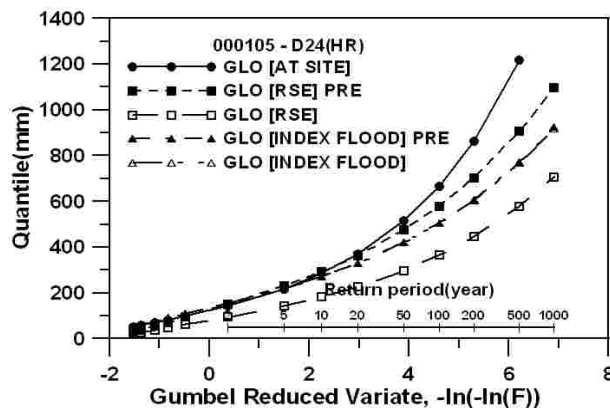


Fig. 6. Estimated Rainfall Quantiles of Gangneung

Table 6. Quantiles by Frequency Analysis Methods

Station	Return Period (yr) Methods	10	20	50	100	200	500
		At-site	246.3	297.9	378.9	452.9	540.4
Seoul	IF (Heo et al., 2007)	251.7	304.4	386.2	459.7	545.7	682.8
	IF	243.2	292.9	370.7	441.6	525.5	660.8
	At-site	169.3	195.6	232.9	263.9	297.8	347.7
Daegu	IF (Heo et al., 2007)	169.3	198.5	242.1	279.9	322.8	388.6
	IF	159.8	181.5	212.4	238.0	266.0	306.9
	At-site	286.2	369.9	517.6	668.1	863.9	1217.2
Gangneung	IF (Heo et al., 2007)	272.1	329.7	421.4	506.0	607.2	772.6
	IF	273.1	330.7	421.8	505.4	605.1	767.3

도가 향상됨을 의미하는 것이다. 대구 지점의 경우(Fig. 5), 자료 기간이 88년으로 서울보다 길지만, 이질 가능성이 있는 지역으로서 이질성 있는 지역에 더 적합한 것으로 알려진 지역형상추정법(Hosking and Wallis, 1997)에 의한 결과와 홍수지수법에 의한 결과가 긴 자료기간에도 불구하고 차이를 나타내고 있고, 지역형상 추정법과 지점빈도해석 결과가 유사하게 나타남을 확인할 수 있다. 강릉 지점의 경우(Fig. 6), 지점빈도해석 결과가 지역빈도해석 결과보다 크게 산정되었는데, 이는 2002년 태풍 “루사”에 의한 큰 호우 사상이 지점빈도해석 결과에 반영된 것이라 할 수 있다. 이는 지역빈도해석이 자료에 포함된 극한값의 영향을 크게 받지 않아 안정적인(robust) 값을 산출함을 보여주는 결과라 할 수 있다.

4. 결 론

본 연구에서는 전국 68개 강우 관측지점의 자료에 다변량 분석 기법을 적용하여 지역을 구분하고, 지역빈도해석을 수행하였다. Dinpashoh et al. (2004)이 제안한 Procrustes analysis 알고리즘을 활용하여 강우에 영향을 미치는 42개 자료의 정보 손실을 최소화 하면서 변수 개수를 21개로 줄일 수 있었다. 다변량 분석 수행에 사용되는 자료의 개수를 줄이는 것은 그 분석의 효율성 향상 측면에서 의미 있는 작업이라 할 수 있다. 따라서 본 연구에서 변수의 개수를 절반으로 줄인 것은

요인분석과 군집해석 수행의 효율을 향상시켰다고 볼 수 있다. 이러한 결과는 동질 지역 구분을 위한 초기 변수 개수가 많을수록 본 연구에서 활용된 다변량 분석 기법들이 더욱 유용하게 활용될 수 있음을 의미한다.

선정된 21개의 변수로부터 요인분석을 통해 5개의 요인을 추출하고, fuzzy-c means 기법으로 지역을 구분한 결과 6개의 지역으로 구분되었다. 이 구분 결과는 허준행 등(2007)의 지역 구분 결과와 유사한 형태를 보였다. 구분된 지역에 적합한 확률분포형으로는 GEV, generalized logistic, 3변수 대수정규 분포가 선정되었다.

기존 연구 결과와의 비교를 위해 generalized logistic 분포를 적정 분포형으로 선정하고, 지점빈도해석과 홍수지수법 및 지역형상추정법을 적용하였다. 그 결과 자료 기간이 긴 경우 지점빈도해석에 의한 확률강우량의 신뢰도가 향상됨을 확인할 수 있었고, 지역빈도해석을 적용할 경우 극한값의 영향을 상대적으로 덜 받는 확률강우량 산정이 가능함을 알 수 있었다.

감사의 글

이 논문은 2005년 정부(교육인적자원부)의 재원으로 한국학술진흥재단의 지원을 받아 수행된 연구임(KRF-2005-041-D00808).

참 고 문 헌

- 조인호 (2005). SAS 강좌와 통계컨설팅 2nd Ed. 영진닷컴.
- 허준행, 이영석, 신흥준, 김정덕 (2007). "우리 나라 강우자료의 지역빈도해석 적용성 연구(I): 확률강우량 산정", **대한토목학회논문집**, 대한토목학회 제27권, 제2B호, pp. 101-111.
- Bezdek, J.C. (1981). *Pattern recognition with fuzzy objective function algorithms*, Plenum Press, New York.
- Bunch, J.R., and Nielsen, C.P. (1978). "Updating the singular value decomposition." *Numerische Mathematik* Vol. 31, pp. 111-129.
- Christopher, M.T., David, W.W.Jr., and Dennis J. (2005). "Regional Rainfall Frequency Analysis for the State of Michigan." *Journal of Hydrologic Engineering*, ASCE, Vol. 10, No. 6, pp. 437-449.
- Cunnane, C. (1989). *Statistical distributions for flood frequency analysis*, Hydrol. Rep. No. 33, WMO Publ. No. 718, Geneva.
- Darlymple, T. (1960). *Flood frequency analyses*. Water Supply Paper 1543-A, U.S. Geological Survey, Reston, Va.
- 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*, Elsevier Science, Vol. 297, pp. 109-123.
- Dunn, J. C. (1973) "A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters." *Journal of Cybernetics*, Springer, Vol. 3, pp. 32-57.
- Guttman, N.B. (1993). "The use of L-Moments in the determination of regional precipitation climates." *Journal of Climate*, American Meteorological Society, Vol. 6, pp. 2309-2325.
- Hosking, J.R.M. and Wallis, J.R. (1997). *Regional frequency analysis: an approach based on L-moments*, Cambridge University Press.
- Krzanowski, W.J. (1987). "Selection of variables to preserve multivariate data structure using principal components." *Applied Statistics*, Royal Statistical Society [etc.], Vol. 36, No. 1, pp. 22-33.
- Lettenmaier, D.P. and Potter, K.W. (1985). "Testing flood frequency estimation methods using a regional floodgeneration model." *Water Resources Research*, American Geophysical Union, Vol. 21, pp. 1903-1914.
- Mallants, D., Feyen, J. (1990). "Defining homogeneous precipitation regions by means of principal component analysis." *Journal of Applied Meteorology*, The Society, Vol. 29, pp. 892-901.
- Overall, J.E. and Klett, C.J. (1973). *Applied multivariate analysis*, McGraw-Hill, New York.
- Potter, K.W. (1987). "Research on flood frequency analysis, 1983-1986." *Reviews of Geophysics*, American Geophysical Union, Vol. 25, No. 2, pp. 113-118.
- Puvaneswaran, M. (1990). "Climatic classification for Queensland using multivariate statistical techniques." *Int. J. Climatol.*, John Wiley & Sons, Vol. 10, pp. 591-608.
- Smithers, J.C. and Schulze, R.E. (2001). "A Methodology for the Estimation of Short Duration Design Storms in South Africa Using a Regional Approach Based on L-moments." *Journal of Hydrology*, Elsevier Science, Vol. 241, pp. 42-52.
- Sveinsson, O.G.B, Salas, J.D., and Boes, D.C. (2002). "Regional frequency analysis of extreme precipitation in Northeastern Colorado and Fort Collins flood of 1997." *Journal of Hydrologic Engineering*, ASCE, Vol. 7, No. 1, pp. 49-63.
- White, D., Richman, M., and Yanel, B. (1991). "Climate regionalization and rotation of principal components." *Int. J. Climatol.*, John Wiley & Sons, Vol. 11, pp. 1-25.
- Zhang Jingyi, M.J. Hall (2004). "Regional flood frequency analysis for the Gan-Ming River basin in China." *Journal of Hydrology*, Elsevier Science, Vol. 296, pp. 98-117.

(논문번호:07-141/접수:2007.12.07/심사완료:2008.03.26)