이상치검정을 고려한 국내 확률강우량 산정

Estimation of Probability Rainfall considering Test for Outliers in Korea

이정식*, 신창동**, 강미순***

Jung Sik Lee, Chang Dong Shin, Mi Soon Kang

요 지

본 연구에서는 확률강우량 산정을 위한 빈도해석 시 보다 타당하고, 신뢰도 높은 강우자료를 선정하여 수공구조물 설계 및 재난방재계획에 합리적인 기준을 제시하고자 이상치 적용에 따른 확률강우량을 산정·비교하였다. 이상치검정의 비교를 위하여 Grubbs and Beck 방법과 미국 수자원평의회 방법을 적용하였으며, 적용결과에 따라 강우자료를 조정하여 확률강우량을 산정·비교하여 타당성을 검토하였다. 본 연구의 수행으로 이상치검정에 따라 강우자료 년수에 변화가 있었으며, 기존의 방법과 확률강우량을 비교하였을 때, 증감이 나타났다.

핵심용어 : 이상치 검정, Grubbs and Beck 방법, 미국 수자원평의회 방법

1. 서론

일반적으로 홍수량을 산정하기 위해서 주로 사용되는 수문자료는 강우자료로서, 수리구조물의 설계에 필요한 확률강우량을 산정하기 위하여 대상지점의 자료를 수집하고 일련의 통계적인 절차를 거쳐 빈도해석을 수행하는 방법인 지점빈도해석 기법이 널리 이용되고 있다. 빈도해석을 하기 위하여 수집된 강우자료는 극치계열의 연최대치 계열자료로서, 자료수집과정에서 장비활용에 따른 기계적 오류 및 기술자의 표독오차와 같은 인적오류, 또는 이상기후에 따라 본래의 값에서 벗어나는 극치의 값 등이 존재할 수 있다. 이와 같은 값들은 분석시 자료 본연의 특성을 왜곡시켜 편향된 결과를 산정할 수 있으므로 예비적해석 절차를 수행하여 자료의 적정성을 확인하여야 한다. 예비적해석으로는 이상치에 대한 이상치검정과 계열내 자료들이 서로 상관성이 없는 무작위성을 가져야 한다는 무작위성검정, 평균이나 분산 등의 크기가 증가하거나 감소하는 등의 경향성을 가지지 않아야 하는 경향성검정 등이 있다. 이중에서 이상치검정은 고려해야하는 결과로서, 이상치는 전체 관측자료의 경향에서 현저하게 벗어난 자료를 의미한다. 이러한 이상치를 수정하거나 기각하는 등의 처리에 따라 적합한 확률분포형이 변화되며 이에 따른 빈도해석 결과가 상당히 변화하게 된다. 현재 실무의 경우, 강우빈도해석시 예비적해석을 수행하고 있지만, 형식적인 참고자료로 활용하고 최종결과인 확률강우량 산정에는 고려하지 않고 있는 실정이다.

따라서, 본 연구에서는 국내 주요 도시의 강우자료로 이상치검정을 수행하여, 검정 결과에 따라 강우자료의 기각정도와 기각 유·무 자료군의 확률강우량 변화를 검토하여 검정에 따른 결과의 변화를 분석한다. 이를 위하여 이상치검정 방법으로 널리 쓰이고 있는 Grubbs and Beck 방법(1972)과 수문자료의 이상치검정방법으로 특화된 미국 수자원평의회 방법(1981)을 적용하여 검정방법간의 차이와 검정결과를 비교하였다.

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

** 정회원·금오공과대학교 건설기술연구소 연구원·E-mail : sichado88@kumoh.ac.kr

*** 금오공과대학교 산업대학원 토목환경및건축공학과 석사과정·E-mail : younione@nate.com

또한, 검정결과에 따라 이상치로 판명된 자료를 삭제한 자료군과 삭제전 원래의 자료군을 이용하여 확률강우량을 산정·비교하였다. 본 연구의 수행으로 확률강우량 산정을 위한 빈도해석시 보다 타당하고, 신뢰도 높은 강우자료를 선정하여 수공구조물 설계 및 재난방재계획에 합리적인 기준을 제시할 것으로 판단된다.

2. 기본이론

강우자료의 이상치검정을 위하여 확률강우량을 산정하기로 하며, 확률분포형은 Gumbel 분포형, 모수산정 방법으로는 Hosking(1990)이 제안한 L-moment 방법을 적용하였다. 이상치검정 방법으로는 Grubbs and Beck 방법과 미국 수자원평의회 방법을 적용하여 검정방법간의 차이와 검정결과를 비교하였다.

2.1 Gumbel 분포형

Type- I 극치분포로 GEV분포형에서 $\beta = 0$이며, GEV-1분포로도 알려져 있는 GUM 분포는 자료의 극치자료중에서도 최대치에 해당하는 연 최대홍수량 및 강우량자료의 분석에 많이 이용되고 있다. 국내에서도 건설교통부(2000)의 발표 이후로, 국내 강우의 빈도해석에서 널리 사용되고 있는 분포형이다. GUM 분포의 누가분포함수와 확률밀도함수는 식 (1)과 (2)와 같이 정의된다.

$$F(x) = \exp\left\{-\exp\left[-\frac{(x-x_0)}{\alpha}\right]\right\} \tag{1}$$

$$f(x) = \frac{1}{\alpha} exp\left\{-\frac{(x-x_0)}{\alpha} - \exp\left[-\frac{(x-x_0)}{\alpha}\right]\right\}, \quad -\infty < x < \infty \tag{2}$$

여기서 α는 축척 모수, x_0는 위치 모수이다.

2.2 Grubbs and Beck 방법

Grubbs and Beck 방법은 모집단이 정규분포를 이룬다고 가정하고, 최고치 또는 최저치가 유의수준 10% 내에 있는지 여부를 검정하는 방법이다.

$$y_H = \bar{y} + K_n \sigma_y, \qquad y_L = \bar{y} - K_n \sigma_y \tag{3}$$

$$K_n = -3.62201 + 6.28446 \cdot n^{1/4} - 2.49836 \cdot n^{1/2} + 0.491436 \cdot n^{3/4} - 0.037911 \cdot n \tag{4}$$

여기서 y는 $\ln x$, $\bar{y}$는 평균, σ_y는 표준편차, y_H는 고이상치 기준, y_L은 저이상치 기준, K_n은 표본수에 따른 이상치 검정 빈도계수(10% 유의수준), n은 자료수이다.

2.3 미국 수자원평의회 방법

왜곡도가 0.4보다 크면 고이상치에 대한 검정, 왜곡도가 -0.4보다 작으면 저이상치에 대한 검정을 실시하고, 왜곡도가 -0.4~0.4의 범위에 있으면 고이상치와 저이상치 모두에 대한 검정을 실시하는 방법이다. 고이상치와 저이상치 한계값은 다음 식에 의하여 결정된다.

$$X_H = \overline{X} + K_N \cdot S \tag{5}$$

$$X_H = \overline{X} + K_N \cdot S \tag{6}$$

여기서 X_H는 로그단위의 상한치, $\overline{X}$는 평균, S는 표준편차, K_N은 표본수에 따른 빈도계수이다.

3. 적용 및 분석

3.1 대상지역

강우량 기록년수는 일반적으로 30년 이상으로 하고 있으며, 본 연구에서는 서울특별시와 광역시인 부산. 인천, 대구, 대전, 광주, 울산 지점을 대상으로 기상청의 지속기간별 매년 최대치계열 강우자료를 이용하였

다. 지속기간 10분 강우자료는 기상청의 기상연보를 이용하여 수집하였으며, 지속기간 1, 2, 3, 4, 5, 6, 9, 12, 15, 18, 24시간 강우자료는 관측소에서 매시간 측정한 자료를 이용하였다. 수집된 강우자료의 기록연수는 서울, 대구, 부산, 광주, 울산, 인천지점의 1961～2015년까지 55년이며, 대전지점의 경우 1970～2015년까지 46년이다.

3.2 이상치 분석

이상치검정 방법으로는 Grubbs and Beck 방법과 미국 수자원평의회 방법을 적용하여 검정방법간의 차이와 검정결과를 비교하였다. Grubbs and Beck 방법은 모집단이 정규분포를 이룬다고 가정하고, 최고치 또는 최저치가 유의수준 10%내에 있는지 여부를 검정하는 방법이며, 미국 수자원평의회 방법은 왜곡도가 0.4보다 크면 고이상치에 대한 검정, 왜곡도가 -0.4보다 작으면 저이상치에 대한 검정을 실시하고, 왜곡도가 -0.4～0.4의 범위에 있으면 고이상치와 저이상치 모두에 대한 검정을 실시하는 방법이다. Grubbs and Beck 방법과 미국 수자원평의회 방법을 적용하여 이상치검정을 실시한 결과는 표 1과 같다. 표 1을 살펴보면, Grubbs and Beck 방법에서는 서울, 부산, 광주, 울산지점에서 총 14개의 이상치, 미국 수자원평의회 방법에서는 부산, 광주, 울산지점에서 총 8개의 이상치가 나타났다. Grubbs and Beck 방법으로 이상치가 나타난 자료는 서울지점의 경우 1962년과 1981년이며 지속기간은 10분과 1시간, 부산지점의 경우에는 1966년과 1984년이며, 지속기간은 10분이다. 광주지점의 경우에는 1967년과 1989년, 2004년이며 지속기간은 1, 2, 4, 9, 12, 24시간, 울산지점의 경우에는 1969년과 1991년으로, 지속기가는 15, 18, 24시간이다. 미국 수자원평의회 방법으로 이상치가 나타난 자료는 부산지점의 경우 1981년의 지속기간 10분, 광주지점의 경우에는 1989년과 2004년으로 지속기간 9, 12, 24시간이다.

표 1. 이상치검정 결과

지점명	Grubbs and Beck 방법		미국 수자원평의회 방법	
	기각자료수	지속기간/발생연도	기각자료수	지속기간/발생연도
서울	2	10분/1962 1시간/1981	-	-
부산	2	10분/1966, 1981	1	10분/1981
인천	-	-	-	-
대구	-	-	-	-
대전	-	-	-	-
광주	6	1, 2, 4시간/1967 9, 12시간/2004 24시간/1989	3	9, 12시간/2004 24시간/1989
울산	4	15, 18시간/1969 24시간/1969, 1991	4	15, 18시간/1969 24시간/1969, 1991

각 지점별 Grubbs and Beck 방법에 의한 상·하한임계치를 나타내면 그림 1～4와 같다. 그림 1～4를 살펴보면, 전체적으로 저이상치 값에 비하여 고이상치 값이 많이 발생하고 있음을 알 수 있다. 미국 수자원평의회 방법에 의한 상·하한 임계치를 나타내면 그림 5～8과 같으며, 그림 5～8을 살펴보면 전체적으로 고이상치의 값이 발생하고 있음을 알 수 있다.

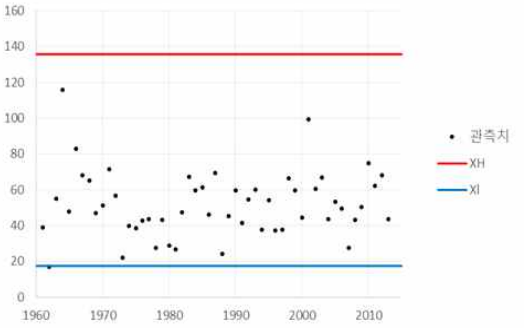

그림 1. Grubbs and Beck 방법을 이용한 이상치검정(서울, 10분)

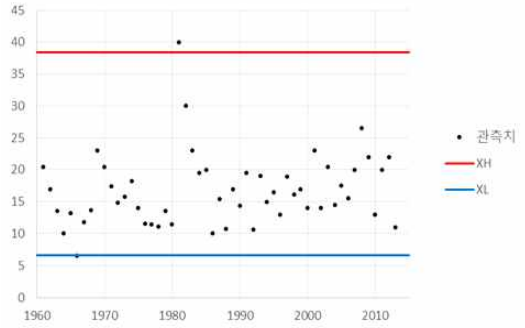

그림 2. Grubbs and Beck 방법을 이용한 이상치검정(부산, 10분)

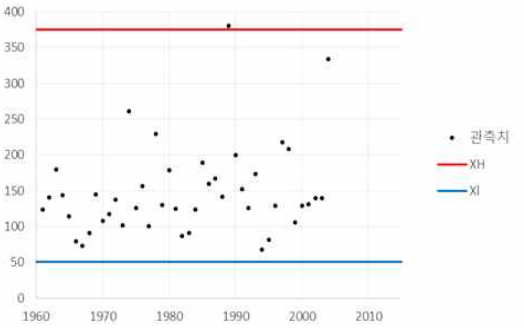

그림 3. Grubbs and Beck 방법을 이용한 이상치검정(광주, 24시간)

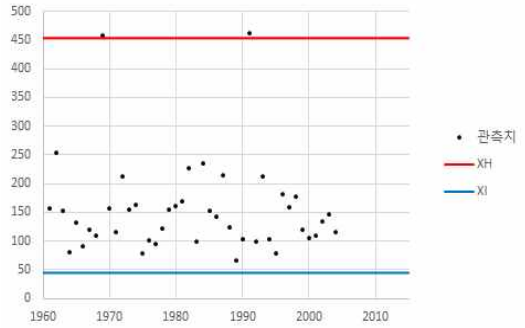

그림 4. Grubbs and Beck 방법을 이용한 이상치검정(울산, 24시간)

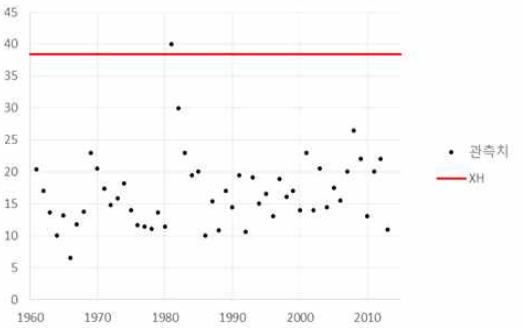

그림 5. 미국 수자원평의회 방법을 이용한 이상치 검정(부산, 10분)

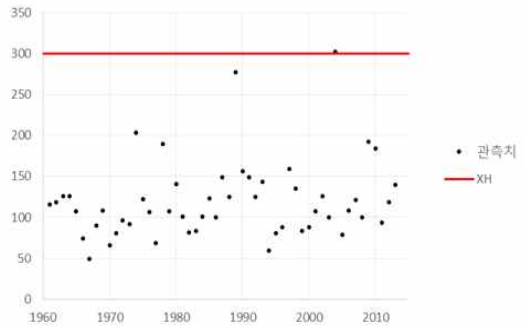

그림 6. 미국 수자원평의회 방법을 이용한 이상치 검정(광주, 12시간)

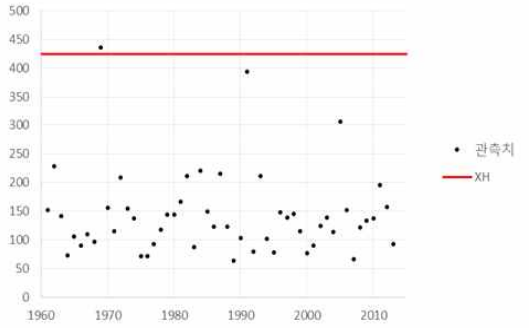

그림 7. 미국 수자원평의회 방법을 이용한 이상치 검정(울산, 18시간)

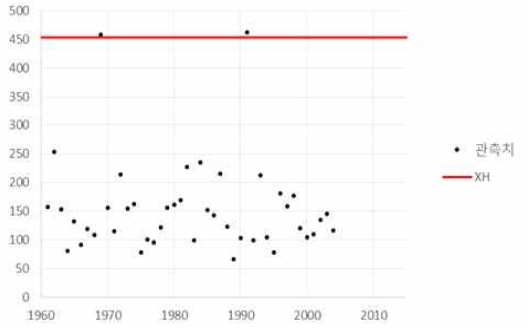

그림 8. 미국 수자원평의회 방법을 이용한 이상치 검정(울산, 24시간)

이상치검정의 결과에 따른 확률강우량을 산정하여 비교하였으며, 비교결과는 표 2와 같다. 표2를 살펴보면, 이상치의 하한치와 상한치결과에 따라 산정된 확률강우량의 차이가 발생하였으며, 서울과 부산지점의 경우에는 검정후의 결과가 기존의 값에 비하여 증가하였고 광주와 울산지점의 경우에는 확률강우량의 값이 기존의 값에 비하여 감소하였다.

표 2. 이상치검정을 적용한 확률강우량 비교

지점명	지속기간	구분	10년	20년	30년	50년	100년
서울	10분	검정전	24.6	27.6	29.3	31.4	34.3
		G-B검정	24.8	27.9	29.7	31.9	34.9
부산	10분	검정전	23.9	27.0	28.7	30.9	33.9
		G-B검정	22.6	25.3	26.9	28.8	31.5
		W검정	22.8	25.7	27.4	29.5	32.4
광주	1시간	검정전	63.6	71.7	76.4	82.2	90.1
		G-B검정	63.6	71.4	75.9	81.6	89.2
	2시간	검정전	88.2	100.1	107.0	115.5	127.1
		G-B검정	88.8	101.1	108.2	117.0	129.0
	4시간	검정전	116.6	132.4	141.4	152.7	168.0
		G-B검정	116.6	131.9	140.7	151.7	166.5
	9시간	검정전	164.9	188.4	201.9	218.7	241.5
		G/W검정	155.0	175.7	187.6	202.5	222.5
	12시간	검정전	178.6	203.8	218.2	236.3	260.7
		G/W검정	168.6	191.0	203.8	219.9	241.5
	24시간	검정전	221.9	253.8	272.1	295.1	326.1
		G/W검정	209.1	237.4	253.7	274.1	301.6
울산	15시간	검정전	214.6	248.8	268.4	293.0	326.1
		G/W검정	199.3	229.2	246.3	267.8	296.7
	18시간	검정전	230.0	267.1	288.3	315.0	350.9
		G/W검정	213.8	246.2	264.8	288.1	319.6
	24시간	검정전	248.6	288.3	311.1	339.7	378.2
		G/W검정	213.6	243.4	260.5	282.0	310.9

주) 검정전 : 기각전의 전체 자료, G-B검정 : Grubbs and Beck, W검정 : 미국 수자원평의회 방법

4. 결론

본 연구에서는 국내 주요 도시의 강우자료를 이용하여 이상치검정을 수행하여 검정 결과에 따라 강우자료의 기각정도와 기각에 따른 확률강우량의 변화, 기수립된 자료와의 비교를 수행하여 검정에 따른 결과의 변화를 분석한다. 이를 위하여 이상치검정 방법으로는 Grubbs and Beck 방법과 미국 수자원평의회 방법을 적용하여 검정방법간의 차이와 검정결과를 비교하였다. 또한, 검정결과에 따라 이상치로 판명된 자료를 삭제한 자료군을 조성하여 삭제전의 자료군을 이용하여 확률강우량을 산정·비교하였다. 본 연구의 수행으로 얻어진 결과를 요약하면 다음과 같다.

(1) 이상치는 서울과 부산, 광주, 울산지점에서 발생하였으며, 지속기간은 10분, 9, 12, 15, 18, 24시간에서 발생하였다.

(2) Grubbs and Beck 방법에서는 14개의 이상치, 미국 수자원평의회 방법에서는 8개 이상치가 발생하였다.

(3) 이상치검정을 고려한 강우자료로부터 산정한 확률강우량은 기존의 값과 차이가 발생되었으며, 추후 지점 및 자료의 보완이 수행되면 활용성을 증대시킬 수 있을 것으로 판단된다.

참 고 문 헌

1. 건설교통부 (2000), 한국 확률강우량도 작성, 1999년도 수자원관리 기법개발연구조사 보고서, 제1권.
2. Grubbs, F. and G. Beck(1972), “Extension of Sample Sizes and Percentage Points for Significance Tests of Outlying Observations”, Technometrics, Vol.14, No.4, pp.847-854.
3. Hosking J.R.M.(1990), L-moments-Analysis and estimation of distribution using linear combinations of order statics, Journal Royal Statistical Society B, v.52, no.1, pp.105～124.
4. U.S. water Resources Council(1981), “Guidelines for determining flood flow frequency”, Bulletin 17B(revised), Hydrology Committee, Water Resources Research Council, Washington.