

## 표본자료의 왜곡도 영향을 고려한 GEV 분포의 확률도시 상관계수 검정방법 비교 검토

### Comparison on Probability Plot Correlation Coefficient Test Considering Skewness of Sample for the GEV Distribution

안 현 준\* / 신 흥 준\*\* / 김 수 영\*\*\* / 허 준 행\*\*\*\*

Ahn, Hyunjun / Shin, Hongjoon / Kim, Sooyoung / Heo, Jun-Haeng

#### Abstract

It is important to estimate an appropriate quantile for design of hydraulic structure. For this purpose, it is necessary to find the appropriate probability distribution which can represent the sample data well. Probability plot correlation coefficient test as one of goodness-of-fit test, is recently developed and has been known as a simple and powerful method. In this study, probability plot correlation coefficient test statistics using the plotting position considering the coefficients of skewness for the GEV distribution is derived, and represented by the regression equation. Monte-Carlo method is also performed to compare the rejection power between each method. As the results, the probability plot correlation coefficient test which is derived in this study is better than the others. In particular, when sample size is small and distribution has the shape parameter, rejection power of probability plot correlation coefficient test considering the coefficients of skewness is bigger than the others.

**Keywords** : probability plot correlation coefficient, GEV, plotting position, coefficients of skewness

#### 요 지

수공구조물의 설계 시 적절한 확률수문량을 추정하는 것은 매우 중요하며, 이러한 확률수문량을 추정하기 위해서는 표본으로서의 수문자료를 잘 표현할 수 있는 확률분포형을 찾아야 한다. 이와 같이 수문자료에 통계적 특성을 잘 표현할 수 있는 확률분포형을 찾기 위해서 적합도 검정을 실시하며, 적합도 검정 중 하나인 확률도시 상관계수 검정은 비교적 최근에 개발되어 그 사용법이 간단하며 높은 기각능력을 갖는다고 알려져 있다. 본 연구에서는 왜곡도 계수의 영향을 고려할 수 있는 도시위치공식을 이용하여 확률도시 상관계수 검정통계량을 유도하고 그 기각능력을 검토하였으며, 그 결과를 기존에 왜곡도 계수를 고려하지 않은 확률도시 상관계수 검정 방법과 비교해보았다. 그 결과 본 연구에서 유도된 확률도시 상관계수 검정에 의한 기각능력이 기존의 검정 방법들 보다 뛰어났으며, 특히 표본 크기가 작을수록, 발생 분포 형이 형상 매개변수를 가질 경우 기각능력이 높게 나타나는 것으로 나타났다.

\* 연세대학교 대학원 토목환경공학과 통합과정 (e-mail: kamjakang@yonsei.ac.kr)

Graduate Student, School of Civil and Environmental Engineering, Yonsei Univ., Seoul 120-749, Korea.

\*\* 연세대학교 대학원 토목환경공학과 박사후과정, 공학박사 (e-mail: sinong@yonsei.ac.kr)

Post-Doctoral Fellow, School of Civil and Environmental Engineering, Yonsei Univ., Seoul 120-749, Korea.

\*\*\* 연세대학교 대학원 토목환경공학과 박사후과정, 공학박사 (e-mail: sykim79@yonsei.ac.kr)

Post-Doctoral Fellow, School of Civil and Environmental Engineering, Yonsei Univ., Seoul 120-749, Korea.

\*\*\*\* 교신저자, 연세대학교 사회환경공학부 토목환경공학과 교수, 공학박사 (e-mail: jhheo@yonsei.ac.kr, Tel: 02-2123-2805)

Corresponding Author, Professor, School of Civil and Environmental Engineering, Yonsei Univ., Seoul 120-749, Korea.

## 1. 서론

수문학에서 확률수문량을 얻기 위해서는 표본으로 추출된 수문자료에 대하여 여러 확률분포형을 적용해 보고, 대상 자료를 적절하게 나타낼 수 있는 확률분포형을 선정해야 한다. 선정된 확률분포형에 따라 확률수문량이 결정되고, 이는 곧 수공구조물의 설계로 이어진다. 하천의 범람이나 제방의 붕괴 등 잘못된 수공구조물의 설계로 인한 사고 발생 시, 사회적 비용측면 뿐만 아니라 인명 및 재산 피해와도 직결 되는 문제가 발생할 수 있다. 때문에 정확한 확률수문량을 산정하기 위하여 표본으로 추출된 수문 자료를 적절하게 표현할 수 있는 확률분포형을 찾아내는 일은 매우 중요한 일이라고 할 수 있다. 이처럼 주어진 표본이 특정 분포형에 얼마나 적합한지를 판별하는 방법을 적합도 검정(goodness-of-fit test)이라 하는데, 이러한 적합도 검정에는  $\chi^2$ -검정, Kolmogorov-Smirnov (KS) 검정, Cramer von-Mises (CVM) 검정, 확률도시 상관계수(probability plot correlation coefficient; PPCC) 검정과 같은 방법들이 있다. 이 방법들 중 PPCC 검정은 Filliben (1975)에 의해 정규분포에서 자료의 정규성을 알아보기 위한 새로운 검정 방법으로 처음 제시되었다. 이후 여러 학자들이 연구를 통해 Gumbel과 GEV, log-Pearson type III (LP3) 분포형 등 다양한 분포형에 대하여 PPCC 검정을 이용한 검정통계량 값들을 제시하였으며(Vogel, 1986; Vogel and McMartin, 1991; Chowdhury et al., 1991), Heo et al. (2008)은 다양한 분포형에 대한 PPCC 검정통계량 값을 회귀식으로 나타내었다. 다른 검정방법들 보다 비교적 최근에 개발된 PPCC 검정은 도시위치(plotting position) 공식을 기반으로 하여 검정을 수행하므로 정성적인 도시적 해석방법을 정량적으로 나타낼 수 있다는 점에서 의미하는 바가 크다고 할 수 있으며(Heo et al., 2001), 그 원리가 단순하여 이해하기 쉽고 사용하기 편리할 뿐 아니라 기각능력 또한 뛰어나다(Vogel, 1986).

한편, PPCC 검정을 수행하기 위해서는 각 분포형마다 적합한 도시위치공식이 필요하다. PPCC 검정은 표본의 순서 통계량(order statistics)을 고려한 도시위치공식을 통해 가정된 확률분포형으로부터 얻은 자료와 실제 자료의 상관성을 통해 적합도 검정을 수행하기 때문이다. 이러한 도시위치공식은 처음 Hazen (1914)에 의해 제안된 이후 최근까

지도 연구가 지속되고 있다(Kimball, 1946; Blom, 1958; Gringorten, 1963; Filliben, 1969; Cunnane, 1978; Arnell et al., 1986; In-na and Nguyen, 1989; Goel and De, 1993; Kim et al., 2012). 여러 분포형에 대한 도시위치공식의 연구 결과, GEV 분포에서는 Cunnane (1978)이 제시한 방법이, Gumbel 분포형에서는 Gringorten (1963)이 제시한 공식이 가장 적절하다고 알려져 있다(Cunnane, 1978; Guo, 1990a). 하지만 GEV 분포처럼 형상 매개변수(shape parameter)를 갖는 분포형과 관련해서는 Cunnane (1978)이 제시한 방법처럼 자료의 순서 통계량만 고려하는 것과는 달리, 분포형이 가지고 있는 형상 매개변수와 높은 상관성을 갖는 왜곡도 계수를 고려할 수 있는 새로운 도시위치공식에 대한 연구가 진행되어 왔다(In-na and Nguyen, 1989; Goel and De, 1993; Kim et al., 2012). 그 중 Kim et al. (2012)이 제시한 왜곡도 계수를 고려할 수 있는 GEV 분포에 대한 도시위치공식은 기존의 다른 도시위치공식들과 GEV의 이론적 축소변량(reduced variate)에 대하여 RMSE와 RBIAS 값을 종합적으로 비교해 보았을 때 가장 유용하게 사용할 수 있다고 하였다(Kim et al., 2012). 따라서 형상 매개변수와 밀접한 연관이 있는 왜곡도 계수를 고려할 수 있는 도시위치공식을 사용한다면 PPCC 검정의 기각능력을 보다 높일 수 있을 것으로 예상된다.

본 연구에서는 수문극치분포를 다룰 때 널리 사용되며 형상 매개변수를 가지고 있는 GEV 분포를 대상으로 왜곡도 계수를 고려할 수 있는 도시위치공식을 적용하여 검정통계량을 유도하고 이를 회귀식으로 제시하였다. 또한 기존의 Blom (1958), Gringorten (1963), Cunnane (1978)이 제시한 도시위치공식들을 이용한 PPCC 검정들과의 기각능력을 비교하기 위해서 모의실험을 수행하였다.

## 2. 적용 이론

### 2.1 GEV 분포

극치 수문사상에 대한 빈도해석에 주로 사용되는 분포형 중 하나인 GEV 분포형은 Gumbel 분포형과 더불어 우리나라에서 가장 많이 사용되고 있는 분포형이다. 특히 GEV 분포형은 Gumbel 분포형에 비해 형상 매개변수라는 매개변수가 하나 더 존재하기 때문에 극치수문자료에 대하여 적용 시키기에 좀 더 유연(flexible)하다고 볼 수 있다. GEV 분포

의 누가분포함수는 다음과 같다(Jenkinson, 1955).

$$F(x) = \exp[-\{1 - \beta(x - x_0)/\alpha\}^{1/\beta}], \quad \beta \neq 0 \quad (1a)$$

$$= \exp[-\exp\{-(x - x_0)/\alpha\}], \quad \beta = 0 \quad (1b)$$

여기서,  $x_0$ ,  $\alpha$ ,  $\beta$ 는 각각 위치(location), 규모(scale), 형상(shape) 매개변수를 나타내고, 형상 매개변수  $\beta$ 가 0일 경우 Gumbel 분포, 음수일 경우 log-Gumbel 분포, 양수일 경우는 Weibull 분포를 나타낸다.

## 2.2 PPCC 검정

Filliben (1975)에 의해 처음 제시된 방법인 PPCC 검정은 정규분포에서 자료의 정규성을 검사하기 위해 제시된 통계적 기법이다. 이 방법은 두 자료 간의 선형성을 측정하는 방법으로, 여기서 말하는 두 자료는 표본으로 얻어진 자료를 순차적으로 정렬한 것과 가정한 확률분포형의 누가분포함수의 역함수로부터 얻어지는 자료를 순차 정렬한 것을 의미한다. 또한 선형성의 측정 방법은 Eq. (2)와 같은 적모멘트 상관계수를 이용한다.

$$r_c = \frac{\sum_{i=1}^n (X_i - \bar{X})(M_i - \bar{M})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (M_i - \bar{M})^2}} \quad (2)$$

여기서,  $M_i = \Phi^{-1}(m_i)$ 이고,  $\Phi^{-1}(\cdot)$ 는 가정한 각 확률분포형의 누가분포함수의 역함수를 의미하고  $m_i$ 는 누가분포함수의 중간값(median)이다. Filliben (1975)은 평균(mean)이 아닌 중간값(median)을 이용한 이유에 대해 순서통계량에서 평균은 계산하기 어렵거나 추정하는데 많은 시간이 소요되고 분포형에 따라서 항상 정의되지 않기 때문이라고 하였다. Filliben (1975)은 중간값  $m_i$ 를 산정하는 식을 다음과 같이 제시하였다.

$$\begin{aligned} m_i &= 1 - (0.5)^{1/n} \quad i = 1 \\ m_i &= \frac{(i - 0.3175)}{(n + 0.365)} \quad i = 2, \dots, n - 1 \\ m_i &= (0.5)^{1/n} \quad i = n \end{aligned} \quad (3)$$

## 2.3 도시위치공식

도시위치공식은 수집된 표본 자료가 이론적 분포형으로부터 각각 어떠한 누가확률을 갖는지 알 수 있는 방법으로 수문학에서 확률도시법(probability plotting technique)이

나 적합도 검정 등에 널리 이용되고 있다. 특히 PPCC 검정에서는 확률분포형과 그 확률분포형이 갖는 형상 매개변수에 따라 PPCC 검정통계량 값이 달라지기 때문에 분포형별로 적절한 누가확률을 산정해주는 도시위치공식의 선정이 중요하다고 할 수 있다. 이러한 도시위치공식은 Hazen (1914)에 의해 처음 제시된 공식으로써 이후 많은 연구자들에 의해 다양한 확률분포형에 맞는 도시위치공식이 개발되었다(Gumbel, 1958; Blom, 1958; Gringorten, 1963; Cunnane, 1978). 일반적으로 정규분포형(normal distribution)에서는 Blom (1958)이 제시한 도시위치공식이, Gumbel 분포형에서는 Gringorten (1963)이 제시한 방법이, GEV 분포형에서는 Cunnane (1978)이 제시한 도시위치공식이 적절하다고 알려져 있다(Looney and Gulledege, 1985; Gringorten, 1963; Guo, 1990b; Stedinger et al., 1993). 특히 Cunnane (1978)은 기존에 Filliben (1975)이 제시한 중간값(median)과는 달리 평균(mean)으로 접근하는 도시위치공식을 주장하였다. 한편, 앞서 연구된 도시위치공식과는 달리 자료의 순서 통계량뿐만 아니라 왜곡도 계수를 고려함으로써 분포형의 형상 매개변수를 고려할 수 있는 도시위치공식에 대한 연구도 활발히 이어지고 있다(Arnell et al., 1986; In-na and Nguten, 1989; Goel and De, 1993; Kim et al., 2012). 특히 Kim et al. (2012)이 제시한 도시위치공식은 앞서 언급한 방법들 중 비교적 가장 최근에 개발된 공식으로써 기 개발된 도시위치공식들과 비교를 해보면, GEV 분포의 형상 매개변수가 -0.2에서 +0.2의 범위에 있을 때 GEV 분포의 이론적 축소변량과 가장 적은 오차를 갖는다. 위에서 언급한 도시위치공식을 정리해보면 Table 1과 같다.

## 3. 도시위치공식별 PPCC 검정통계량 유도

앞서 2.2 PPCC 검정에서 설명했던 것처럼 PPCC 검정은 실제 관측자료로부터 얻은 표본과 이론적 분포형으로부터 얻은 확률수문량 값과의 선형적 상관성을 통해서 적합도 검정을 수행하게 된다. 일반적으로 이론적 분포형의 확률수문량 값은 누적분포함수(cumulative distribution function, CDF)의 역함수에 누가발생확률값을 대입함으로써 구할 수 있는데, 이때 CDF의 역함수에 대입하는 확률값을 도시위치공식을 통해 산정하게 되는 것이다. 결국 PPCC 검정은 확률분포형에 대한 적절한 도시위치공식에 따라 그 기각능력이 좌우된다고 볼 수 있다. 따라서 본 연구에서는 GEV분포에 대해서 PPCC 검정통계량을 유도하는데 왜곡도계수를 고려할 수 있는 도시위치공식을 비롯한 다양한 도시위치공식을 이용하여 도시위치공식별 PPCC

Table 1. Plotting Position Formulas

| Class                   | Plotting position formular                                    | Recommended probability distribution            |
|-------------------------|---|---|
| Blom (1958)             | $P_i = \frac{i-3/8}{n+1/4}$                                   | normal, gamma, log normal, log-Pearson type III |
| Gringorten (1963)       | $P_i = \frac{i-0.44}{n+0.12}$                                 | Gumbel, Weibull                                 |
| Cunnane (1978)          | $P_i = \frac{i-0.4}{n+0.2}$                                   | GEV, log-Gumbel                                 |
| Arnell et al. (1986)    | $P_i = \frac{i-\alpha}{n+\beta}$                              | GEV / $\alpha, \beta$ are suggested in paper    |
| In-na and Nguyen (1989) | $P_i = \frac{n-i+0.05\gamma+0.65}{n-0.08\gamma+0.38}$         | GEV   |
| Goel and De (1993)      | $P_i = \frac{i-0.02\gamma-0.32}{n-0.04\gamma+0.36}$           | GEV   |
| Kim et al. (2012)       | $P_i = \frac{i-0.3200}{n+0.0149\gamma^2-0.1364\gamma+0.3225}$ | GEV   |

검정통계량을 유도하고 이를 회귀식으로 제시하였다.

### 3.1 PPCC 검정통계량 유도 방법

PPCC 검정통계량을 유도하기 위하여 Vogel and McMartin (1991)은 다음과 같은 방법을 제시하였다.

- ① 확률분포형에 따른 위치 매개변수와 규모 매개변수는 각각 상수 값으로 고정하여 놓고 형상 매개변수와 표본 크기별 자료를 100,000개씩 발생시킨다.
- ② 확률분포형에 따라 적절한 도시위치공식을 적용하여 도시위치를 구한 뒤, 누가분포함수의 역함수를 이용하여  $M_i$ 를 구한다.
- ③ ①번 과정에서 발생시킨 표본 자료와 ②번 과정에서 구한  $M_i$ 를 Eq. (2)을 통해서 PPCC 검정통계량  $r_c$ 을 산정한다.
- ④ 이때 유의수준  $\alpha$ 에 대한  $r_\alpha$ 는  $r_c$ 중에서  $100,000 \times \alpha$  번째로 작은 값과 같고 Eq. (4)와 같이 표시할 수 있다.

$$r_\alpha = r_{(100,000\alpha)} \quad (4)$$

Vogel and McMartin (1991)이 제시한 방법에 따라서 PPCC 검정통계량은 확률분포형과 형상 매개변수, 표본 크기 그리고 유의수준에 따라 각기 다른 검정통계량을 갖게 된다.

### 3.2 도시위치공식별 PPCC 검정통계량 유도

GEV 분포형에서 PPCC 검정통계량을 유도하기 위하여 앞서 언급된 Vogel and McMartin (1991)이 제시한 방법을 이용하였고, 도시위치공식으로는 Cunnane, Blom,

Gringorten이 제시한 방법과 GEV 분포형에 적합하며 왜곡도 계수를 고려할 수 있는 Kim et al. (2012)의 방법을 이용하였다. 모의 발생시 사용된 표본 크기는 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 120, 150, 200, 300, 500으로 총 15가지의 경우로 나누었고, 형상 매개변수는  $-0.25$ 부터  $+0.20$ 까지 0.05 간격으로(0.00 제외) 9개의 범위를 사용하였으며, 각각의 표본 크기와 형상 매개변수에 따라 100,000개의 표본 자료를 발생시켰다. 유의 수준은 1%, 5%, 10%에 대하여 검정통계량을 산정하였다.

본 연구에서 유도한 도시위치공식별 PPCC 검정통계량 값(유의수준 5%)과 이를 회귀식 곡선으로 나타낸 결과는 Fig. 1과 같다.

### 3.3 도시위치공식별 PPCC 검정통계량 회귀식 산정

표본 크기, 형상 매개변수 및 유의수준에 따라 PPCC 검정통계량이 각각 다르기 때문에 이를 사용하기 간편하도록 하나의 회귀식으로 표현하였다. 회귀식의 형태는 Heo et al. (2008)이 발표한 논문을 참고로 하였다. 앞서 유도한 PPCC 검정통계량의 형상 매개변수 범위와 회귀식의 형상 매개변수에 대한 범위가 유의수준 0.01과 0.05에서 각각 다른데, 이는 유의수준별로  $\alpha = 0.01$ 일 때 형상 매개변수  $-0.25$ 에서  $-0.10$ 까지,  $\alpha = 0.05$ 일 때 형상 매개변수  $-0.25$ 에서 유도된 검정통계량 값이 표본 크기에 증가에 따라 매끄러운 증가 추세를 나타내지 못하여서 회귀식으로 표현하기에 적절하지 않음으로 제외하였다. PPCC 검정통계량 회귀식의 기본식과 유의수준에 따른 회귀식의 범위는 Eq. (5)와 같고, 도시위치공식별 유의수준에 따른

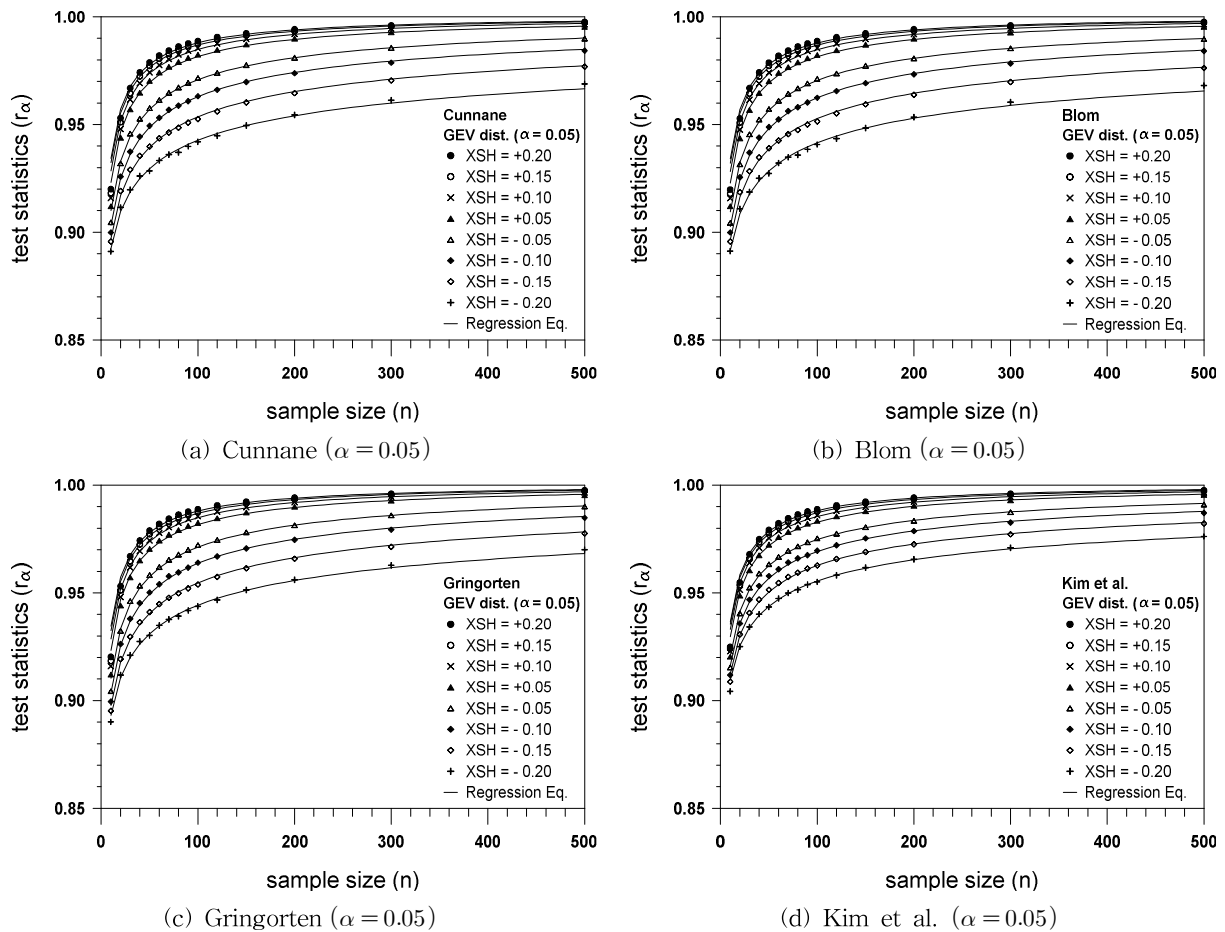


Fig. 1. PPCC Test Statistics (significance level 0.05) for Each Plotting Position Formula

매개변수  $a_1, a_2, a_3, a_4, a_5, a_6, a_7$ 은 Table 2와 같다.

$$\ln\left(\frac{1}{1-r_\alpha}\right) = (a_1 + a_2\beta + a_3\beta^2 + a_4\beta^3) \cdot n^{(a_5 + a_6\beta + a_7\beta^2)} \quad (5)$$

$$\begin{aligned} \alpha: 0.01, & \quad -0.05 \leq \beta \leq 0.20 \\ \alpha: 0.05, & \quad -0.20 \leq \beta \leq 0.20 \\ \alpha: 0.10, & \quad -0.25 \leq \beta \leq 0.20 \end{aligned}$$

여기서,  $r_\alpha$ =유의수준  $\alpha$ 일 때 PPCC 검정통계량,  $\beta$ =형상 매개변수,  $n$ =표본 크기

PPCC 검정통계량 회귀식이 유도한 검정통계량 값을 얼마나 잘 나타내는지에 대하여 결정계수로써 검증하였고, 가장 낮은 결정계수는 Gringorten 공식에서 유의수준 1%의 형상 매개변수 0.20에서의 값으로 0.9859였다. 전체적으로 유도된 검정통계량을 회귀식이 잘 나타내고 있음을 확인하였으며, Kim et al. (2012)의 회귀식이 전반적으로 가장 높은 결정계수 값을 보이는 것으로 나타났다. 각 도시위치공식별 회귀식에 따른 결정계수는 Table 3과 같다.

#### 4. 도시위치공식별 기각능력 검토

본 연구에서 유도한 검정통계량을 바탕으로 만든 PPCC 검정 회귀식을 이용하여 각 도시위치공식별 PPCC 검정의 기각능력을 검토하기 위해 모의수행을 실시하였다. 모의수행을 위하여 난수(random number)발생 분포형은 총 6개의 분포형을 사용하였고, 그 중 형상 매개변수를 갖는 분포형은 각각 8가지의 형상 매개변수를 고려하였으며 분포형별로 9가지의 표본 크기로 10,000번씩 반복하여 발생시켰다. 또한 모의수행 뿐만 아니라 국내 수문관측자료에 대해서도 연구 결과를 적용하여 각 방법들 간의 기각능력을 검토하였다. 본 연구에서 적용한 관측 자료로는 기상청에서 제공하는 시 강우자료를 이용하였다.

##### 4.1 모의수행조건

도시위치공식별 기각능력 검토를 위한 모의수행으로써 몬테카를로 방법(Monte-Carlo method)을 사용하였

Table 2. Parameters of Regression Equation for Each Plotting Position Formula

| PP formular | $\alpha$ | $a_1$  | $a_2$   | $a_3$  | $a_4$   | $a_5$  | $a_6$  | $a_7$   |
|-------------|----------|--------|---------|--------|---------|--------|--------|---------|
| Cunnane     | 0.01     | 1.3001 | 0.0459  | 3.1018 | -2.6898 | 0.1966 | 0.4242 | -1.2773 |
|             | 0.05     | 1.6407 | 0.0912  | 1.7126 | -6.1241 | 0.1806 | 0.2557 | -0.5414 |
|             | 0.10     | 1.8302 | 0.1546  | 0.3641 | -5.1527 | 0.1724 | 0.1939 | -0.2892 |
| Blom        | 0.01     | 1.2928 | 0.0972  | 3.1157 | -3.7332 | 0.1967 | 0.4194 | -1.2427 |
|             | 0.05     | 1.6383 | 0.0868  | 1.7469 | -6.2631 | 0.1804 | 0.2603 | -0.5524 |
|             | 0.10     | 1.8308 | 0.1330  | 0.4195 | -5.2459 | 0.1721 | 0.1978 | -0.2963 |
| Gringorten  | 0.01     | 1.3128 | 0.0259  | 2.9326 | -2.0677 | 0.1960 | 0.4186 | -1.2535 |
|             | 0.05     | 1.6394 | 0.1171  | 1.6989 | -6.0434 | 0.1815 | 0.2471 | -0.5305 |
|             | 0.10     | 1.8255 | 0.1878  | 0.3250 | -5.0151 | 0.1732 | 0.1885 | -0.2844 |
| Kim et al.  | 0.01     | 1.4425 | -0.6581 | 4.6684 | -6.2737 | 0.1841 | 0.4407 | -1.1519 |
|             | 0.05     | 1.7381 | -0.1448 | 1.1499 | -5.3779 | 0.1741 | 0.2363 | -0.3671 |
|             | 0.10     | 1.9058 | -0.1466 | 0.5922 | -3.1413 | 0.1675 | 0.2005 | -0.2971 |

Table 3.  $R^2$  of Regression Equation for Each Plotting Position

| PP formular | $\alpha$ | -0.25  | -0.20  | -0.15  | -0.10  | -0.05  | 0.05   | 0.10   | 0.15   | 0.20   |
|-------------|----------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| Cunnane     | 0.01     | -      | -      | -      | -      | 0.9970 | 0.9975 | 0.9936 | 0.9890 | 0.9960 |
|             | 0.05     | -      | 0.9941 | 0.9981 | 0.9983 | 0.9985 | 0.9951 | 0.9937 | 0.9915 | 0.9909 |
|             | 0.10     | 0.9949 | 0.9985 | 0.9969 | 0.9974 | 0.9973 | 0.9953 | 0.9939 | 0.9938 | 0.9919 |
| Blom        | 0.01     | -      | -      | -      | -      | 0.9963 | 0.9976 | 0.9934 | 0.9892 | 0.9860 |
|             | 0.05     | -      | 0.9924 | 0.9976 | 0.9986 | 0.9985 | 0.9955 | 0.9938 | 0.9916 | 0.9910 |
|             | 0.10     | 0.9941 | 0.9985 | 0.9970 | 0.9975 | 0.9973 | 0.9953 | 0.9942 | 0.9939 | 0.9919 |
| Gringorten  | 0.01     | -      | -      | -      | -      | 0.9978 | 0.9971 | 0.9929 | 0.9886 | 0.9859 |
|             | 0.05     | -      | 0.9959 | 0.9981 | 0.9979 | 0.9979 | 0.9953 | 0.9942 | 0.9913 | 0.9907 |
|             | 0.10     | 0.9958 | 0.9985 | 0.9969 | 0.9974 | 0.9970 | 0.9950 | 0.9946 | 0.9937 | 0.9919 |
| Kim et al.  | 0.01     | -      | -      | -      | -      | 0.9987 | 0.9976 | 0.9942 | 0.9913 | 0.9880 |
|             | 0.05     | -      | 0.9976 | 0.9983 | 0.9972 | 0.9977 | 0.9959 | 0.9949 | 0.9929 | 0.9925 |
|             | 0.10     | 0.9978 | 0.9985 | 0.9974 | 0.9975 | 0.9975 | 0.9963 | 0.9954 | 0.9948 | 0.9935 |

다. 난수 발생 분포형은 좌우 대칭인 분포(symmetric distribution)와 왜곡된 분포(skewed distribution)로 나누어 각각 3개씩 총 6개의 분포형을 사용하였다. 대칭인 분포형으로는 normal 분포형, cauchy 분포형, logistic 분포형을 사용하였고 왜곡된 분포형으로는 GEV 분포형, generalized logistic 분포형, generalized pareto 분포형을 사용하였다. 그 중 왜곡된 분포형은 각각 8가지의 형상 매개변수를 고려하였으며 각 분포형별로 왜곡도 계수 범위 0.25에서 3.60에 해당하는 형상 매개변수를 사용하였다. 또한 표본 크기로는 10, 20, 30, 40, 50, 80, 100, 200, 300으로 총 9가지의 경우를 고려하였다. 이를 바탕으로 분포형

에 따라 형상 매개변수와 표본 크기 별로 각각 10,000번씩 반복하여 난수를 발생시켜 유의수준 5%에 대하여 적합도 검정을 수행 하였다. 기각능력 검토를 위한 몬테카를로 방법의 조건은 Table 4와 같다.

#### 4.2 도시위치공식별 기각능력 결과

몬테카를로 방법을 이용하여 도시위치공식별 PPCC 검정을 통해 기각능력 비교를 수행한 결과는 Table 5와 같다. Table 5에서 NOR는 Normal 분포를 모집단으로 하여 4.1절의 모의수행 조건을 바탕으로 각 도시위치공식별 PPCC 검정을 수행한 결과이다. 이와 마찬가지로 CAU

**Table 4. Conditions of Power Test**

|                                 |   |  |  |
|---------------------------------|---|--|--|
| Significance level ( $\alpha$ ) | 0.05  |  |  |
| Sample size ( $n$ )             | 10, 20, 30, 40, 50, 80, 100, 200, 300         |  |  |
| Shape parameter ( $\gamma$ )    | normal = cauchy = logistic = 0.00 (symmetric) |  |  |
|                                 | GEV   | = -0.20, -0.15, -0.10, -0.05, +0.05, +0.10, +0.15, +0.20 |  |
|                                 | generalized logistic                          | = -0.24, -0.21, -0.18, -0.15, -0.10, -0.08, -0.06, -0.03 |  |
|                                 | generalized pareto                            | = -0.15, -0.08, +0.01, +0.10, +0.35, +0.45, +0.60, +0.75 |  |
| Iteration                       | 10,000  |  |  |

**Table 5. Result of Power Test**

NOR

| n      | 10     | 20     | 30     | 40     | 50     | 80     | 100    | 200    | 300    |
|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| PPCC-C | 0.0029 | 0.0127 | 0.0191 | 0.0243 | 0.0295 | 0.0401 | 0.0497 | 0.1024 | 0.1724 |
| PPCC-B | 0.0029 | 0.0125 | 0.0187 | 0.0237 | 0.0291 | 0.0397 | 0.0489 | 0.1017 | 0.1715 |
| PPCC-G | 0.0029 | 0.0130 | 0.0196 | 0.0245 | 0.0296 | 0.0402 | 0.0506 | 0.1030 | 0.1728 |
| PPCC-K | 0.0063 | 0.0165 | 0.0224 | 0.0286 | 0.0329 | 0.0434 | 0.0545 | 0.1064 | 0.1745 |

CAU

| n      | 10     | 20     | 30     | 40     | 50     | 80     | 100    | 200    | 300    |
|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| PPCC-C | 0.2386 | 0.4007 | 0.6148 | 0.7521 | 0.8470 | 0.9593 | 0.9814 | 0.9999 | 1.0000 |
| PPCC-B | 0.2357 | 0.3975 | 0.6116 | 0.7484 | 0.8443 | 0.9583 | 0.9809 | 0.9999 | 1.0000 |
| PPCC-G | 0.2472 | 0.4055 | 0.6189 | 0.7570 | 0.8502 | 0.9611 | 0.9823 | 0.9999 | 1.0000 |
| PPCC-K | 0.3317 | 0.4725 | 0.6775 | 0.8034 | 0.8849 | 0.9732 | 0.9896 | 1.0000 | 1.0000 |

LGS

| n      | 10     | 20     | 30     | 40     | 50     | 80     | 100    | 200    | 300    |
|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| PPCC-C | 0.0068 | 0.0484 | 0.0883 | 0.1336 | 0.1716 | 0.2904 | 0.3561 | 0.6759 | 0.8669 |
| PPCC-B | 0.0066 | 0.0473 | 0.0878 | 0.1322 | 0.1707 | 0.2891 | 0.3550 | 0.6745 | 0.8665 |
| PPCC-G | 0.0069 | 0.0490 | 0.0897 | 0.1346 | 0.1734 | 0.2920 | 0.3591 | 0.6771 | 0.8674 |
| PPCC-K | 0.0124 | 0.0612 | 0.1034 | 0.1492 | 0.1889 | 0.3053 | 0.3722 | 0.6822 | 0.8671 |

GEV

| n      | 10     | 20     | 30     | 40     | 50     | 80     | 100    | 200    | 300    |
|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| PPCC-C | 0.0053 | 0.0083 | 0.0121 | 0.0141 | 0.0157 | 0.0181 | 0.0198 | 0.0255 | 0.0324 |
| PPCC-B | 0.0051 | 0.0082 | 0.0118 | 0.0137 | 0.0154 | 0.0178 | 0.0193 | 0.0249 | 0.0316 |
| PPCC-G | 0.0058 | 0.0086 | 0.0124 | 0.0146 | 0.0163 | 0.0186 | 0.0205 | 0.0267 | 0.0335 |
| PPCC-K | 0.0146 | 0.0156 | 0.0179 | 0.0207 | 0.0227 | 0.0249 | 0.0272 | 0.0332 | 0.0400 |

GLO

| n      | 10     | 20     | 30     | 40     | 50     | 80     | 100    | 200    | 300    |
|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| PPCC-C | 0.0074 | 0.0310 | 0.0582 | 0.0853 | 0.1084 | 0.1753 | 0.2170 | 0.4046 | 0.5465 |
| PPCC-B | 0.0072 | 0.0305 | 0.0572 | 0.0841 | 0.1071 | 0.1736 | 0.2150 | 0.4018 | 0.5433 |
| PPCC-G | 0.0080 | 0.0318 | 0.0593 | 0.0867 | 0.1102 | 0.1776 | 0.2197 | 0.4092 | 0.5516 |
| PPCC-K | 0.0200 | 0.0476 | 0.0763 | 0.1046 | 0.1297 | 0.1990 | 0.2405 | 0.4303 | 0.5717 |

GPA

| n      | 10     | 20     | 30     | 40     | 50     | 80     | 100    | 200    | 300    |
|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| PPCC-C | 0.0059 | 0.0016 | 0.0024 | 0.0041 | 0.0084 | 0.0416 | 0.0863 | 0.3038 | 0.4079 |
| PPCC-B | 0.0057 | 0.0015 | 0.0023 | 0.0039 | 0.0081 | 0.0408 | 0.0847 | 0.3011 | 0.4046 |
| PPCC-G | 0.0067 | 0.0017 | 0.0024 | 0.0043 | 0.0087 | 0.0431 | 0.0887 | 0.3082 | 0.4137 |
| PPCC-K | 0.0174 | 0.0048 | 0.0051 | 0.0075 | 0.0140 | 0.0553 | 0.1050 | 0.3326 | 0.4390 |

는 Cauchy 분포형, LGS는 logistic 분포형, GEV는 GEV 분포형, GLO는 generalized logistic 분포형, GPA는 generalized pareto 분포형을 각각 모집단으로 한 모의수행결과를 의미한다. 또 PPCC-C는 Cunnane의 도시위치공식을 이용한 PPCC 검정, PPCC-B는 Blom의 도시위치공식을 이용한 PPCC 검정, PPCC-G는 Gringorten의 도시위치공식을 이용한 PPCC 검정, PPCC-K는 Kim et al.의 도시위치공식을 이용한 PPCC 검정을 의미한다. 또한 표본 크기별, 검정 방법별로 가장 높은 기각률에 대해서는 회색으로 음영 표시하였다. GEV 분포에서는 도시위치공식에 관계없이 표본 크기가 증가함에 따라 유의수준 5%에 근접하였고, 자료의 왜곡도를 고려할 수 있는 도시위치공식을 이용한 PPCC 검정이 logistic 분포형에서 표본이 300일 때를 제외한 모든 경우에서 기존에 순서 통계량만을 고려하는 도시위치공식을 이용한 PPCC 검정 보다 기각능력이 높았으며, 특히 표본 크기가 작을수록 좌우 대칭인 분포형보다 왜곡된 분포형에서 기각능력 증가 비율이 더 컸다. 이를 통해 PPCC 검정에서 자료의 왜곡도 계수를 고려한 도시위치공식을 이용하였을 때 기각능력이 높아지며, 특히 형상 매개변수를 갖는 분포형에 대한 기각능력이 기존 방법에 비해 높아졌음을 확인하였다.

### 4.3 국내수문자료에 적용결과

본 연구에서는 모의실험뿐만 아니라 우리나라의 강우 자료에 대하여 연구 결과를 적용하여 실제 적용성을 검토하였다. 기상청 관측 지점 중 40개 지점을 대상으로 지속 시간 1, 2, 3, 4, 5, 6, 9, 12, 15, 18, 24, 48시간에 대하여 검토를 실시하였으며 매개변수 추정 방법으로는 모멘트 법(Method of Moments)을 이용하였다. 예로 문산(099)과

포항(138) 지역의 결과를 Table 6에 제시하였다. 문산지점의 지속시간 9, 12시간의 경우 기존 방법인 PPCC-C, PPCC-B, PPCC-G는 적합도 검정을 통과하였으나 이번에 새롭게 유도한 PPCC-K 방법으로는 기각이 되는 것을 확인하였다. 또한 포항지점의 지속시간 5, 6, 12시간의 경우 같은 결과를 보였다. 따라서 순서통계량만을 이용한 도시위치공식 대신에 왜곡도 계수를 고려할 수 있는 도시위치공식을 사용 할 경우 기존 GEV 분포에서 PPCC 검정을 통한 적합도 검정의 기각능력이 향상됨을 확인하였다.

## 5. 결론

본 연구에서는 표본 자료의 왜곡도 계수를 고려할 수 있는 GEV에 대한 PPCC 검정 방법에 대하여 비교 검토해보았으며 결론은 다음과 같다.

- 1) 도시위치공식으로 순서통계량만을 고려한 Cunnane, Blom, Gringorten이 제시한 방법과 자료의 왜곡도 계수를 고려할 수 있는 Kim et al.이 제시한 방법을 이용하여 확률도시 상관계수 검정통계량 값을 유도하였고, 이를 사용하기 간편하도록 회귀식으로 제시하였다.
- 2) 각각의 방법별로 기각능력을 검토하기 위하여 확률 분포형, 형상 매개변수, 표본 크기들을 다양하게 변화시켜가면서 모의실험을 실시하였고 그 결과 전체적으로 기존의 방법보다 본 연구에서 왜곡도를 고려할 수 있는 도시위치공식을 이용하여 새로 유도한 PPCC-K에서 높은 기각능력을 보였다.
- 3) 특히 일반적인 수문자료에 해당하는 비대칭 분포에 대해서 기존에 방법과 비교하였을 때 높은 기각능력

Table 6. Results of the PPCC Test (Munsan and Pohang)

| SITE   | Duration (hr) | Skew. Coeff. | XSH    | C.C.  | PPCC-C |        | PPCC-B |        | PPCC-G |        | PPCC-K |        |
|--------|---------------|--------------|--------|-------|--------|--------|--------|--------|--------|--------|--------|--------|
|        |               |              |        | Com.  | Tab.   | Check  | Tab.   | Check  | Tab.   | Check  | Tab.   | Check  |
| Munsan | 5             | 0.201        | 0.215  | 0.938 | 0.935  | Accept | 0.934  | Accept | 0.935  | Accept | 0.936  | Accept |
|        | 6             | 0.831        | 0.058  | 0.933 | 0.924  | Accept | 0.924  | Accept | 0.824  | Accept | 0.930  | Accept |
|        | 9             | 1.574        | -0.062 | 0.910 | 0.908  | Accept | 0.908  | Accept | 0.909  | Accept | 0.920  | Reject |
|        | 12            | 1.665        | -0.073 | 0.910 | 0.907  | Accept | 0.906  | Accept | 0.907  | Accept | 0.919  | Reject |
| Pohang | 5             | 3.096        | -0.182 | 0.944 | 0.933  | Accept | 0.932  | Accept | 0.934  | Accept | 0.946  | Reject |
|        | 6             | 3.036        | -0.179 | 0.946 | 0.934  | Accept | 0.933  | Accept | 0.935  | Accept | 0.947  | Reject |
|        | 9             | 2.935        | -0.174 | 0.950 | 0.935  | Accept | 0.934  | Accept | 0.936  | Accept | 0.947  | Accept |
|        | 12            | 3.148        | -0.184 | 0.943 | 0.933  | Accept | 0.932  | Accept | 0.934  | Accept | 0.946  | Reject |

XSH=Shape parameter, C.C.=Correlation Coefficient.



을 보였으며 또한 표본의 크기가 작을수록 더 높은 기각능력을 보였다.

- 4) 또한 연구 결과를 바탕으로 실제 강우자료에 적용하여 본 결과 기존 PPCC 검정 결과보다 새로 제시한 PPCC-K 방법이 더 높은 기각능력을 갖는 것으로 나타났다.

이를 통하여 일반적인 수문자료에 본 연구에서 유도한 PPCC-K 방법을 이용한다면 좀 더 높은 기각능력을 얻을 수 있을 것으로 판단된다.

## 감사의 글

본 연구는 소방방재청 자연재해저감기술개발사업인 '3차원 BIM 기술을 활용한 수방시설의 능동형 재난관리체계 구축' [NEMA-12-자연-57]과제의 일환으로 이루어졌습니다. 이에 감사드립니다.

## References

- Arnell, N.W., Beran, M., and Hosking, J.R.M. (1986). "Unbiased plotting positions for the general extreme value distribution." *Journal of Hydrology*, Vol. 86, pp. 59-69.
- Blom, G. (1958). *Statistical estimates and transformed beta variables*. John Wiley and Sons, New York.
- Chowdhury, J.D., Stedinger, J.R., and Lu, L.H. (1991). "Goodness-of-fit tests for regional generalized extreme value flood distributions." *Water Resources Research*, Vol. 27, No. 7, pp. 1765-1776.
- Cunnane, C. (1978). "Unbiased plotting positions-A review." *Journal of Hydrology*, Vol. 37, No. 3/4, pp. 205-222.
- Filliben, J.J. (1969). *Simple and robust linear estimation of the location parameter of a symmetric distribution*. Unpublished Ph.D. dissertation, Princeton university, Princeton, New Jersey.
- Filliben, J.J. (1975). "The Probability Plot Correlation Coefficient Test for Normality." *Technometrics*, Vol. 17, No. 1, pp. 111-117.
- Goel, N.K., and De, M. (1993). "Development of unbiased plotting position formula for General Extreme Value distribution." *Stochastic Environmental Research and Risk Assessment*, Vol. 7, pp. 1-13.
- Gringorten, I.I. (1963). "A plotting rule for extreme probability paper." *Journal of Geophysical Research*, Vol. 68, No. 3, pp. 813-814.
- Gumbel, E.J. (1958). *Statistics of Extremes*. Columbia University Press, New York, NY.
- Guo, S.L. (1990a). "A discussion on unbiased plotting positions for the general extreme value distribution." *Journal of Hydrology*, Vol. 121, pp. 33-44.
- Guo, S.L. (1990b). "Unbiased plotting position formulae for historical floods." *Journal of Hydrology*, Vol. 121, pp. 45-61.
- Hazen, A. (1914). "Storage to be provided in impounding reservoirs for municipal water supply." *Transactions American society of Civil Engineers*, Vol. 1308, No. 77, pp. 1547-1550.
- Heo, J.-H., Kho, Y., Shin, H., Kim, S., and Kim, T. (2008). "Regression equations of probability plot correlation coefficient test statistics from several probability distributions." *Journal of Hydrology*, Vol. 355, pp. 1-15.
- Heo, J.-H., Kho, Y.-W., and Kim, K.-D. (2001). "Test statistics derivation and power test for probability plot correlation coefficient goodness of fit test." *Journal of Korean Society of Civil Engineering*, Vol. 21, No. 2-b, pp. 85-92.
- In-na, N., and Nguyen, V-T-V. (1989). "An unbiased plotting position formula for the generalized extreme value distribution." *Journal of Hydrology*, Vol. 106, pp. 193-209.
- Jenkinson, A.F. (1955). "The frequency distribution of the annual maximum (or minimum) values of meteorological elements." *Quarterly Journal of the Royal Meteorological Society*, Vol. 87, pp. 158-171.
- Kim, S., Shin, H., Joo, K., and Heo, J.-H. (2012). "Development of plotting position for the general extreme value distribution." *Journal of Hydrology*, Vol. 475, pp. 259-269.
- Kimball, B.F. (1946). "Assignment of frequencies to a completely ordered set of sample data." *Transaction on the American Geophysical Union*, Vol 27. pp. 843-846.
- Looney, S.W., and Gullidge, T.R. (1985). "Use the correlation coefficient with normal probability plots." *The American Statistician*, Vol. 39, No. 1, pp. 75-79.

Stedinger, J.R., Vogel, R.M., Foufoula-Georgiou, E. (1993). *Frequency analysis of extreme events. Handbook of Hydrology*, D.R. Maidment, de., McGraw-Hill, New York, N.Y., pp. 18.24-18.26

Vogel, R.M. (1986). "The probability plot correlation coefficient test for the normal, lognormal, and Gumbel distributional hypotheses." *Water Resources Research*, Vol. 22, No. 4, pp. 587-590.

Vogel, R.M., and McMartin, D.E. (1991). "Probability plot goodness-of-fit and skewness estimation procedures for the Pearson type III distribution." *Water Resources Research*, Vol. 27, No. 12, pp. 3149-3158.

|                  |                  |
|------------------|------------------|
| 논문번호: 13-097     | 접수: 2013.12.04   |
| 수정일자: 2014.01.06 | 심사완료: 2014.01.06 |