韓國水資源學會論文集
第42卷 第5號 · 2009年 5月
pp. 365~374

DOI: 10.3741/JKWRA.2009.42.5.365

Gumbel 분포에 대한 도시위치공식의 비교

Comparison of Plotting Position Formulas for Gumbel Distribution

김 수 영* / 허 준 행** / 신 홍 준*** / 고 연 우****
Kim, Sooyoung / Heo, Jun-Haeng / Shin, Hongjoon / Kho, YounWoo

Abstract

Probability plotting positions are used for the graphical display of annual maximum rainfall or flood series and the estimation of exceedance probability of those values. In addition, plotting positions allow a visual examination of the fitness of probability distribution provided by frequency analysis for a given data. Therefore, the graphical approach using plotting position has been applied to many fields of hydrology and water resources planning. In this study, the plotting position formula for the Gumbel distribution is derived by using the order statistics and the probability weight moment of the Gumbel distribution for various sample sizes. And then, the parameters of plotting position formula for the Gumbel distribution are estimated by using genetic algorithm. The appropriate plotting position formulas for the Gumbel distribution are examined by the comparison of root mean square errors and biases between theoretical reduced Gumbel variates and those calculated from derived and existing plotting position formulas. As the results, Gringorten's plotting position formula has the smaller root mean square errors and biases than any other formulas.

keywords : Plotting position formula, Order statistic, Probability weigh moment, Genetic algorithm, Gumbel distribution

요 지

확률도시위치는 주로 도시적 해석을 통한 연최대홍수량 또는 연최대강우량의 초과확률의 추정치 산정에 사용되며 빈도해석을 통해 선정된 적정 확률분포형과 표본자료의 적합도를 도시적으로 파악할 수 있도록 해주기 때문에 오래 전부터 수문 및 수자원 분야에 널리 이용되어 왔다. 본 연구에서는 Gumbel 분포에 적합한 도시위치공식을 추정하고자 Gumbel 분포의 순서통계량과 확률가중모멘트를 이용하여 다양한 표본크기에 대한 도시위치공식의 기본식을 유도하였고, 최적화 기법 중 하나인 유전자 알고리즘을 이용하여 도시위치공식의 매개변수를 추정하였다. 또한 Gumbel 분포에 적합한 도시위치공식을 검토하고자 Gumbel 분포의 이론적인 축소변량과 본 연구에서 추정한 도시위치공식과 기존의 도시위치공식에 의해 계산된 축소변량 간의 평균제곱근오차와 편의를 비교하였다. 그 결과, Gringorten이 제안한 도시위치공식을 적용한 경우의 축소변량간 평균제곱근오차와 순서별 편의가 가장 작은 것으로

* 연세대학교 사회환경시스템공학부 박사과정
Ph.D student, School of Civil and Environmental Engineering, Yonsei University, Seoul 120-749, Korea
** 교신저자, 연세대학교 사회환경시스템공학부 교수
Corresponding Author, Professor, School of Civil and Environmental Engineering, Yonsei University, Seoul 120-749, Korea
(e-mail: jhheo@yonsei.ac.kr)
*** 연세대학교 사회환경시스템공학부 박사과정
Ph.D student, School of Civil and Environmental Engineering, Yonsei University, Seoul 120-749, Korea
**** 연세대학교 사회환경시스템공학부 박사과정
Ph.D student, School of Civil and Environmental Engineering, Yonsei University, Seoul 120-749, Korea

분석되었다.

핵심용어 : 도시위치공식, 순서통계량, 확률가중모멘트, 유전자 알고리즘, Gumbel 분포

1. 서 론

확률도시위치(probability plotting position)는 연최대홍수량 자료나 연최대강우량 자료의 도시적 해석에 이용되는데, 주로 연최대홍수량 또는 연최대강우량의 초과확률의 추정에 사용된다. 또한 확률도시위치는 빈도해석을 통해 선정된 적정 확률분포형과 표본자료의 개략적인 적합도를 도시적으로 파악할 수 있도록 해주며, 주어진 확률분포형에 대한 비모수적 평균을 추정할 수 있도록 하는 역할을 수행하여 오래 전부터 수문·수자원 분야에서 널리 이용되어 왔다.

도시위치공식은 다양한 확률분포형에 대해 연구되어 왔는데, Hazen(1914)은 확률지(probability paper)에 자료를 도시하기 위한 도시위치공식을 제안하였으나 매우 긴 기간의 관측치에 대해 재현기간의 확률도시가 부분적으로 어렵다는 단점이 있었다(Gumbel, 1958). Gumbel(1958)은 Hazen이 제안한 도시위치공식의 이러한 단점을 보완할 수 있는 도시위치공식을 새롭게 제안하였다.

도시위치공식은 대부분 비편의(unbiased)된 수문량(quantile)에 적용할 수 있는 것으로 Weibull(1939)도 비편의된 수문량에 적용할 수 있도록 정렬된 관측자료에 대한 평균 초과확률(average exceedance probability)로 나타내는 도시위치를 제안하기도 하였다. Blom(1958)은 정규분포를 따르는 비편의 수문량에 적용가능한 도시위치공식을 연구한 바 있고, Gringorten(1963)은 Gumbel 분포 계열의 장기간 관측자료에 적용가능한 도시위치공식을 제안하였다. 또한 Cunnane(1978)은 도시위치가 자료의 평균치에 부여되어야 한다고 주장하며 자료의 평균에 부합되는 새로운 도시위치를 정의하고, 여러 가지 확률분포형에 적용가능한 도시위치공식의 일반 형태를 유도하였으며 대부분의 확률분포형에 적용이 가능하도록 일반형 도시위치공식의 매개변수를 제안하였다.

Adamowski(1981)는 평균제곱오차를 이용한 새로운 형태의 도시위치공식을 제안한 바 있고, Xuewu *et al.*(1984)은 Monte Carlo 모의를 이용하여 0과 2 사이의 왜곡도계수를 고려할 수 있는 도시위치공식을 추정하기도 하였으며, Guo(1990b)는 관측된 홍수자료에 대해 기존의 도시위치공식들을 비교하여 Cunnane(1978)에 의해 개발된 도시위치공식이 홍수자료에 적합하다는 결론을 내린 바 있다. 다양한 방법으로 유도된 도시위치공식 중 실제로 가장 널리 사용되고 있는 것은 Gringorten(1963)과 Cunnane(1978)에 의해 개발된 도시위치공식이나, 그 이후에도 도시위치공식에 대한 연구는 꾸준히 지속되고 있다.

그와 같은 연구 중의 하나로는 확률가중모멘트(probability weight moment)를 이용하여 비편의 도시위치공식을 추정하는 방법을 들 수 있다. Arnell *et al.*(1986)은 확률가중모멘트를 이용하여 GEV 분포에 대한 도시위치공식을 추정하였고, In-na and Nguyen (1989)은 Arnell *et al.*(1986)의 방법론을 사용함과 동시에 형상 매개변수(shape parameter)를 고려할 수 있는 도시위치공식을 GEV 분포에 대해 유도하였다. 또한 Nguyen *et al.*(1989)은 확률가중모멘트를 이용하는 방법을 Pearson type Ⅲ 분포에 대해 적용하여 분포형의 특성을 잘 반영하면서도 비교적 간단한 형태로 표현되는 새로운 도시위치공식을 제안하였다. Nguyen and In-na(1992)는 Pearson type Ⅲ 분포에 대한 도시위치공식을 확률가중모멘트를 이용하여 유도하고 홍수자료에 적용하여 적용성을 평가하였으며, Goel and De (1993)는 확률가중모멘트를 적용하여 GEV 분포의 형상 매개변수를 고려할 수 있는 도시위치공식을 유도하였다. Haktanir and Bozduman(1995)은 GEV 분포와 Log-Pearson type Ⅲ 분포에는 확률가중모멘트를 이용한 도시위치공식의 유도가 최적이라는 결론을 내린 바 있으며, De(2000)는 Cunnane(1978)이 제시한 도시위치공식의 일반식 형태를 변형하여 Gumbel 분포에 대한 도시위치공식을 유도하기도 하였다.

본 연구에서는 우리나라의 강우빈도해석에 널리 적용되고 있는 Gumbel 분포에 적합한 도시위치공식을 추정하기 위해 Gumbel 분포의 순서통계량과 확률가중모멘트를 이용하여 다양한 표본크기에 대한 도시위치공식의 기본식을 유도하고, 최적화 기법 중 하나인 유전자 알고리즘(genetic algorithm)을 이용하여 유도된 도시위치공식의 매개변수를 추정하였다. 또한 Gumbel 분포에 더 적합한 도시위치공식을 알아보기 위해 이론적으로 유도된 Gumbel 분포에 대한 축소변량(NERC, 1975)을 기준으로 본 연구에서 추정한 도시위치공식과 기존에 사용되고 있던 도시위치공식에 의해 계산되는 축소변량 사이의 평균제곱근오차(Root Mean Square Error, RMSE)와 편의(bias)를 비교하였다.

2. Gumbel 분포에 대한 도시위치공식의 유도

2.1 순서통계량과 확률가중모멘트를 이용한 도시위치 기본식의 유도

오름차순으로 정렬된 표본크기 n의 관측자료 $y_1 \le y_2 \le y_3 \le \cdots \le y_n$의 m번째 순서통계량(order statistic) y_m에 대한 확률밀도함수 $g(y_m)$은 다음과 같다(Arnell *et al.*, 1986).

$$g(y_m) = \frac{n!}{(m-1)!(n-m)!}\{F(y_m)\}^{m-1}\{1-F(y_m)\}^{n-m}f(y_m) \tag{1}$$

여기서, $F(y_m)$와 $f(y_m)$은 표본으로부터 유도된 y_m에 대한 누가분포함수와 확률밀도함수를 나타낸다.

m번째 관측자료 y_m의 평균 $E[y_m]$은 다음과 같다.

$$E[y_m] = \frac{n!}{(m-1)!(n-m)!}\int_{-\infty}^{\infty} y_m\{F(y_m)\}^{m-1}\{1-F(y_m)\}^{n-m}f(y_m)dy_m \tag{2}$$

만약 y_m의 누가분포함수 $F(y_m)$을 F_m라고 한다면, $y_m = F^{-1}(F_m)$, $f(y_m)dy_m = dF_m$으로 나타낼 수 있으며 $-\infty \to 0$, $\infty \to 1$로 수렴한다. 따라서 Eq. (2)는 다음과 같이 나타낼 수 있다.

$$E[y_m] = \frac{n!}{(m-1)!(n-m)!}\int_0^1 F^{-1}(F_m)F_m^{m-1}\{1-F_m\}^{n-m}dF_m \tag{3}$$

Gumbel 분포의 m번째 축소변량(reduced variate)은 Eq. (4)와 같고, 이를 Eq. (3)에 대입하여 정리하면 Eq. (5)와 같다.

$$y_m = -\ln\{-\ln(F_m)\} \tag{4}$$

$$E[y_m] = \frac{n!}{(m-1)!(n-m)!}\int_0^1 -\ln\{-\ln(F_m)\}F_m^{m-1}\{1-F_m\}^{n-m}dF_m \tag{5}$$

Gumbel 분포의 도시위치공식을 유도하기 위해 Greenwood *et al.*(1979)이 제시한 확률가중모멘트(probability weighted moment)를 이용하며, 확률가중모멘트의 일반적인 형태는 Eq. (6)과 같다.

$$M_{p,r,s} = \int_0^1 x^p F^r (1-F)^s dF \tag{6}$$

여기서, $n = r+s+1$ 및 $p=1$에 대한 Eq. (6)을 Eq. (5)에 대입하여 정리하면 m번째 축소변량 y_m에 대한 평균은 Eq. (7)과 같고, 이를 다시 정리하여 나타내면 Eq. (8)과 같이 나타낼 수 있다.

$$E[y_m] = \frac{n!}{(m-1)!(n-m)!}\sum_{s=0}^{n-m}\binom{n-m}{s}(-1)^s M_{1,m+s-1,0} \tag{7}$$

$$E[y_m] = \frac{n!}{(m-1)!(n-m)!}\sum_{s=0}^{n-m}\binom{n-m}{s}(-1)^s[\gamma+\ln(m+s)](m+s)^{-1} \tag{8}$$

여기서, $\gamma = 0.5772$로 Euler의 상수를 뜻한다. 또한 Eq. (8)을 가장 큰 표본크기($m=n$)에 대한 순서통계량에 대해 나타내면 다음과 같다.

$$E[y_n] = \gamma + \ln(n) \tag{9}$$

본 연구를 통해 유도하고자 하는 Gumbel 분포에 적합한 도시위치공식의 기본 형태를 Eq. (10)과 같이 가정하였다.

$$P_m = \frac{m+a}{n+b} \tag{10}$$

여기서 a와 b는 도시위치공식의 매개변수를 나타내고, n은 표본크기, m은 오름차순으로 정렬되었을 때의 순서를 나타낸다. 이때, Eq. (10)에서 도시위치를 나타내는 P_m은 Eq. (4)의 역함수 형태인 Eq. (11)에 대입하여 산정한다.

$$P_m = F_m = \exp\{-\exp(-y_m)\} \tag{11}$$

2.2 유전자 알고리즘을 이용한 도시위치공식의 매개변수 추정

Gumbel 분포에 대한 도시위치공식의 매개변수를 추정하기 위해 본 연구에서는 유전자 알고리즘을 적용하였다. 유전자 알고리즘은 1960년대에 Holland(1975)와 그의 동료들에 의해서 창안되었으며 다윈(Darwin)이 제안한 적자생존(survival of the fittest)을 컴퓨터 기법화한 최적화 방법이다. Goldberg(1989)는 유전자 알고리즘을 자연 선택(natural selection)과 유전자의 응용을 기반으로 하는 기법으로 정의한 바 있다. 유전자 알고리즘은 초기에 무작위로 초기 모집단(population)을 형성하고 이들을 부모세대로 하여 선택, 교배(crossover), 돌연변이(mutation) 등의 연산과정을 거쳐 부모세대보

다 진화한 새로운 자식세대를 생성하게 되며, 적합도를 평가하여 적합한 개체를 생성시킨다.

본 연구에서는 유전자 알고리즘의 한 종류인 real-coded genetic algorithm(Deb and Beyer, 2001; Beyer and Deb, 2001)을 적용하였고, 유전자 알고리즘의 목적함수(objective function)는 Eq. (10)의 좌변과 우변간의 평균제곱근오차로 설정하여 평균제곱근오차가 최소가 되는 매개변수를 추정토록 하였다. 유전자 알고리즘의 초기조건 중 모집단 수는 1,000개, 전체 세대수(generation)는 500세대, 교배 확률은 0.8, 돌연변이 확률은 0.01로 설정하였다. 또한, seed number의 영향을 알아보기 위해 초기 설정값 0.123을 포함하여 각기 다른 10개의 seed number에 대해 모의를 수행하였고, 매개변수 추정에 사용된 표본크기는 $n = 2(1)30(5)100(10)200(50)500$이다. 단, 식(8)과 같은 근사식은 표본크기가 35이상일 경우에는 반올림오차에 의해 실제 축소변량과 상당한 차이가 발생하는 것으로 알려져 있으므로(Guo, 1990a), 도시위치공식의 매개변수 추정 과정에서 반올림 오차에 의한 오차를 줄이기 위해 표본크기 $2 \le n \le 30$의 범위에서는 각각의 순서(order)별 축소변량까지 고려할 수 있도록 하였으며, $n > 30$의 범위에서는 최대 표본크기 n에 대한 축소변량을 Eq. (9)를 이용하여 보다 많은 표본크기를 고려할 수 있도록 하였다.

유전자 알고리즘의 적용 결과를 살펴보면, 모의회수에 관계없이 도시위치공식의 매개변수는 $a = -0.296102$, $b = 0.222769$로 추정되었는데, 이때의 최소 평균제곱근오차는 0.02466인 것으로 나타났다. 또한 유전자 알고리즘을 이용한 매개변수의 추정에 있어 seed number에 따른 매개변수 추정 결과가 동일하게 나타나 매개변수 추정과정에서의 seed number의 영향은 미미한 것으로 판단되었다. 이에 따른 결과를 정리하면 Eq. (12)와 같다.

$$P_m = \frac{m-0.30}{n+0.22} \tag{12}$$

본 연구에서는 금회 추정된 도시위치공식의 실제 적용성을 알아보기 위해 Eq. (12)와 함께 기존에 사용되고 있는 도시위치공식에 의해 추정되는 비초과확률(q)과 그에 대한 재현기간(RT)을 산정하여 비교하였다. 기존에 사용되고 있는 도시위치공식은 Blom(1958), Gringorten(1963), Cunnane(1978), De(2000)에 의해 제시된 도시위치공식으로 형태는 Table 1과 같다. 금회 추정된 도시위치공식과 Table 1의 도시위치공식을 이용하여 산정한 결과 중 표본크기 10, 25에 대한 비초과확률과 재현기간 결과를 나타내면 Table 2와 같다. Table 2의 결과를 살펴보면, Eq. (12)와 De(2000)에 의해 산정된 재현기간은 기존의 다른 도시위치공식과 비교하였을 때 동일 순서에 대한 재현기간이 크게 산정되었으나 초과확률과 비초과확률의 합이 1을 넘는 오류를 범하고 있는 것으로 나타났다. 표본크기 10에 대해 예를 들어 살펴보면, Table 2에서 보는 바와 같이 Eq. (12)에 의해서는 첫 번째 순서에서 계산된 비초과확률(q)은 0.068이고, 마지막 순서에서 계산된 비초과확률(q)은 0.949로 이 값은 첫 번째 순서의 초과확률과 같으며 결과적으로 비초과확률과 초과확률의 합이 1.017로 1보다 큰 값을 보인다. De(2000)의 경우에도 첫 번째 순서에서 계산된 비초과확률(q)은 0.070이고 마지막 순서에서 계산된 비초과확률(q)이 0.946으로 이들의 합이 1.016이 되어 1보다 큰 것을 알 수 있다. 이와 같은 결과는 Eq. (12)의 매개변수 추정에 있어 표본크기 30 이후의 표본크기에 대해서는 가장 큰 표본크기에 대한 결과만을 적용함에 따라 입력 자료가 가장 큰 표본크기에 대해 편향됨과 동시에 비초과확률과 초과확률간의 관계를 고려하지 않아 나타난 것으로 판단된다.

Table 1. Applied Plotting Position Formulas in This Study

Type	Plotting position formula
Blom(1958)	$P_m = \frac{m-0.375}{n+0.25}$
Gringorten(1963)	$P_m = \frac{m-0.44}{n+0.12}$
Cunnane(1978)	$P_m = \frac{m-0.40}{n+0.20}$
De(2000)	$P_m = \frac{m-0.28}{n+0.28}$

따라서 본 연구에서는 비초과확률과 초과확률의 관계를 고려하는 제한조건($2a - b + 1 = 0$)을 적용한 후, Eq. (12)와 같이 가장 큰 표본크기에 대해 편향된 입력자료를 사용한 경우(Case 1)와 Flood Studied Report (NERC, 1975)에서 제시하고 있는 Gumbel 분포의 축소변량만을 입력 자료로 사용한 경우(Case 2)로 구분하여 매개변수를 추정하였고, 그 결과를 정리하면 Table 3과 같다. Case 1의 RMSE는 0.09512로 Case 2의 0.07767에 비해 상대적으로 높은 것으로 나타남에 따라 본 연구에서는 Case 2의 결과를 Gumbel 분포에 대한 도시위치공식의 매개변수로 선정하였고, 이를 정리하면 다음과 같다.

$$P_m = \frac{m-0.42}{n+0.16} \tag{13}$$

Table 2. The Comparison of Nonexceedance Probability and Return Period Calculated by Various Plotting Position Formulas

Sample size = 10										
Order	Derived		Gringorten(1963)		Cunnane(1978)		Blom(1958)		De(2000)	
	q	RT (year)	q	RT (year)	q	RT (year)	q	RT (year)	q	RT (year)
1	0.068	1.074	0.055	1.059	0.059	1.062	0.061	1.065	0.070	1.075
2	0.166	1.200	0.154	1.182	0.157	1.186	0.159	1.188	0.167	1.201
3	0.264	1.359	0.253	1.339	0.255	1.342	0.256	1.344	0.265	1.360
4	0.362	1.567	0.352	1.543	0.353	1.545	0.354	1.547	0.362	1.567
5	0.460	1.851	0.451	1.820	0.451	1.821	0.451	1.822	0.459	1.849
6	0.558	2.261	0.549	2.219	0.549	2.217	0.549	2.216	0.556	2.254
7	0.656	2.903	0.648	2.843	0.647	2.833	0.646	2.828	0.654	2.888
8	0.753	4.056	0.747	3.953	0.745	3.923	0.744	3.905	0.751	4.016
9	0.851	6.724	0.846	6.487	0.843	6.375	0.841	6.308	0.848	6.590
10	0.949	19.654	0.945	18.071	0.941	17.000	0.939	16.400	0.946	18.357
Sample size = 25										
Order	Derived		Gringorten(1963)		Cunnane(1978)		Blom(1958)		De(2000)	
	q	RT (year)	q	RT (year)	q	RT (year)	q	RT (year)	q	RT (year)
1	0.028	1.029	0.022	1.023	0.024	1.024	0.025	1.025	0.028	1.029
2	0.067	1.072	0.062	1.066	0.063	1.068	0.064	1.069	0.068	1.073
3	0.107	1.120	0.102	1.113	0.103	1.115	0.104	1.116	0.108	1.121
4	0.147	1.172	0.142	1.165	0.143	1.167	0.144	1.168	0.147	1.173
5	0.186	1.229	0.182	1.222	0.183	1.223	0.183	1.224	0.187	1.230
6	0.226	1.292	0.221	1.284	0.222	1.286	0.223	1.287	0.226	1.292
7	0.266	1.362	0.261	1.353	0.262	1.355	0.262	1.356	0.266	1.362
8	0.305	1.439	0.301	1.431	0.302	1.432	0.302	1.433	0.305	1.440
9	0.345	1.527	0.341	1.517	0.341	1.518	0.342	1.519	0.345	1.527
10	0.385	1.625	0.381	1.614	0.381	1.615	0.381	1.616	0.384	1.625
11	0.424	1.737	0.420	1.725	0.421	1.726	0.421	1.726	0.424	1.736
12	0.464	1.865	0.460	1.853	0.460	1.853	0.460	1.853	0.464	1.864
13	0.504	2.014	0.500	2.000	0.500	2.000	0.500	2.000	0.503	2.013
14	0.543	2.189	0.540	2.173	0.540	2.172	0.540	2.172	0.543	2.187
15	0.583	2.397	0.580	2.379	0.579	2.377	0.579	2.376	0.582	2.394
16	0.623	2.649	0.619	2.628	0.619	2.625	0.619	2.623	0.622	2.644
17	0.662	2.960	0.659	2.935	0.659	2.930	0.658	2.928	0.661	2.953
18	0.702	3.354	0.699	3.323	0.698	3.316	0.698	3.311	0.701	3.344
19	0.741	3.868	0.739	3.829	0.738	3.818	0.738	3.811	0.741	3.854
20	0.781	4.569	0.779	4.518	0.778	4.500	0.777	4.489	0.780	4.547
21	0.821	5.580	0.818	5.509	0.817	5.478	0.817	5.459	0.820	5.544
22	0.860	7.165	0.858	7.056	0.857	7.000	0.856	6.966	0.859	7.101
23	0.900	10.008	0.898	9.813	0.897	9.692	0.896	9.619	0.899	9.875
24	0.940	16.592	0.938	16.103	0.937	15.750	0.936	15.538	0.938	16.205
25	0.979	48.500	0.978	44.857	0.976	42.000	0.975	40.400	0.978	45.143

Table 3. The Comparison of Parameters Estimated by Genetic Algorithm

Case	a	b	RMSE
Case 1	−0.406706	0.186588	0.09512
Case 2	−0.419224	0.161552	0.07767

3. 도시위치공식의 비교 및 검토

본 연구에서는 Gumbel 분포 계열의 자료에 가장 적합한 도시위치공식을 검토하기 위해 Eq. (13)과 같이 본 연구에서 추정된 도시위치공식과 기존에 주로 이용되어오던 도시위치공식의 정확도를 비교하였다. 본 연구에서 추정한 도시위치공식과의 비교를 위해 선정한 도시위치공식은 Blom(1958), Gringorten(1963), Cunnane(1978)에 의해 유도된 도시위치공식이다. 정확도 측정은 도시위치공식을 이용하여 산정되는 축소변량과 Flood Studied Report(NERC, 1975)에서 제시하고 있는 이론적인 Gumbel 분포의 축소변량간의 오차를 산정하여 수행되었다. 여기에서 산정한 오차는 표본크기에 대한 평균제곱근오차와 순서별 편의이며, Flood Studied Report(NERC, 1975)에 제시되고 있는 Gumbel 분포의 축소변량이 표본크기 50까지임을 고려하여 정확도 비교를 수행하였다.

Table 4는 표본크기 5, 10, 15, 20, 25, 30, 35, 40, 45, 50에 대해서 전체 표본크기에 대한 평균제곱근오차를 계산하여 나타낸 결과이다. Table 4에 의하면 모든 표본크기에 대해 Blom(1958)에 의해 제안된 도시위치공식을 적용했을 경우에는 가장 큰 평균제곱근오차가 산정되었고, Gringorten(1963)의 도시위치공식을 적용한 경우에 가장 작은 평균제곱근오차가 산정된 것으로 나타났다. Eq. (13)과 같이 본 연구에서 제안한 도시위치공식을 적용한 경우에 계산된 평균제곱근오차는 Cunnane(1978)의 도시위치공식을 적용한 경우보다는 작으나 Gringorten(1963)의 경우보다는 큰 것을 알 수 있다.

전체 표본크기에 대한 평균제곱근오차와 함께 표본크기별 순서에 대한 오차를 살펴보기 위해 2부터 50까지의 표본크기 중 25, 40에 대해 순서별 축소변량간의 편의를 나타내면 Table 5와 같다. Table 5의 결과에서는 표본크기에 대해 순서가 상대적으로 낮거나 높은 경우 도시위치공식별 편의의 차가 큰 편이지만, 중간 순서의 경우에는 도시위치공식별 편의의 차가 크지 않은 것으로 나타났다. 또한, 대부분의 도시위치공식은 높은 순서에서 편의가 점차 커지는데 반해 Gringorten(1963)의 도시위치공식은 높은 순서에서 최소 편의값이 발생하는 것으로 나타났다. Eq. (13)의 경우 편의가 Gringorten(1963)에 이어 두 번째로 작은 것으로 나타났으나, 결과적으로 Gringorten(1963)에 의해 제안된 도시위치공식이 상대적으로 Gumbel 분포의 도시위치를 가장 정확하게 나타내고 있는 것으로 분석되었다.

본 연구에서 제안된 도시위치공식과 Blom(1958), Gringorten(1963), Cunnane(1978)의 도시위치공식에 의한 비초과확률(q)과 그에 대한 재현기간을 재산정하여 비교하면 Table 6과 같다. Table 6을 살펴보면, 순서가 낮은 경우 산정되는 재현기간은 도시위치공식별로 큰 차이가 없는 것으로 나타났으나 순서가 높아질수록 도시위치공식별 차이가 증가하는 것을 알 수 있다. 가장 높은 순서의 경우에 대해 살펴보면 표본크기가 25인 경우 Gringorten(1963)에 의한 재현기간은 약 44.9년, 본 연구에서 추정된 도시위치공식의 경우 약 43.4년으로 산정되었고, Blom(1958)과 Cunnane(1978)의 도시위치공식의 경우는 이보다 더 작은 것으로 나타났다. 결과적으로 Gringorten(1963)에 의한 도시위치공식을 적용하는 경우 순서가 높아질수록 산정된 재현기간이 가장 큰 것으로 나타났다.

Table 4. The Root Mean Square Errors Between the Reduced Variates Calculated by Various Plotting Position Formulas and the Reduced Variates Provided by NERC(1975)

Sample size	Root Mean Square Errors			
	New	Blom(1958)	Gringorten(1963)	Cunnane(1978)
5	0.0692	0.0801	0.0684	0.0728
10	0.0420	0.0536	0.0406	0.0460
15	0.0301	0.0420	0.0282	0.0345
20	0.0249	0.0356	0.0234	0.0289
25	0.0212	0.0322	0.0190	0.0255
30	0.0175	0.0279	0.0156	0.0215
35	0.0145	0.0243	0.0131	0.0183
40	0.0062	0.0190	0.0029	0.0119
45	0.0062	0.0184	0.0026	0.0116
50	0.0061	0.0175	0.0029	0.0112

Table 5. The Bias Between the Reduced Variates Calculated by Various Plotting Position Formulas and the Reduced Variates Provided by NERC(1975)

Order	Sample size = 25				Sample size = 40			
	New	Blom (1958)	Gringorten(1963)	Cunnane (1978)	New	Blom (1958)	Gringorten (1963)	Cunnane (1978)
1	-0.04	-0.02	-0.05	-0.03	0.006	0.023	-0.002	0.014
2	-0.02	-0.01	-0.02	-0.01	0.006	0.014	0.002	0.009
3	-0.01	-0.01	-0.02	-0.01	0.000	0.006	-0.002	0.003
4	-0.02	-0.01	-0.02	-0.02	-0.003	0.002	-0.005	-0.001
5	-0.01	-0.01	-0.01	-0.01	0.005	0.008	0.003	0.006
6	-0.01	-0.01	-0.01	-0.01	0.000	0.003	-0.001	0.001
7	-0.01	-0.01	-0.01	-0.01	-0.003	0.000	-0.004	-0.002
8	-0.01	-0.01	-0.01	-0.01	-0.001	0.001	-0.002	0.000
9	-0.01	-0.01	-0.01	-0.01	-0.004	-0.002	-0.005	-0.003
10	-0.01	-0.01	-0.02	-0.01	0.000	0.002	-0.001	0.001
11	-0.02	-0.02	-0.02	-0.02	0.002	0.003	0.001	0.003
12	-0.02	-0.02	-0.02	-0.02	0.002	0.003	0.001	0.003
13	-0.01	-0.01	-0.01	-0.01	0.001	0.002	0.000	0.001
14	-0.02	-0.02	-0.02	-0.02	-0.001	0.000	-0.001	0.000
15	-0.01	-0.02	-0.01	-0.01	-0.003	-0.002	-0.004	-0.003
16	-0.01	-0.02	-0.01	-0.02	0.005	0.005	0.004	0.005
17	-0.02	-0.02	-0.01	-0.02	0.003	0.003	0.002	0.003
18	-0.01	-0.02	-0.01	-0.02	0.001	0.001	0.001	0.001
19	-0.02	-0.02	-0.01	-0.02	0.000	0.001	0.000	0.000
20	-0.02	-0.02	-0.01	-0.02	0.001	0.001	0.001	0.001
21	-0.03	-0.03	-0.02	-0.03	0.003	0.003	0.003	0.003
22	-0.03	-0.04	-0.02	-0.03	-0.004	-0.004	-0.004	-0.004
23	-0.03	-0.04	-0.02	-0.03	0.002	0.002	0.002	0.002
24	-0.04	-0.07	-0.03	-0.06	0.000	0.000	0.001	0.000
25	-0.04	-0.11	-0.01	-0.07	0.001	0.001	0.002	0.001
26					-0.004	-0.005	-0.003	-0.004
27					0.005	0.004	0.006	0.004
28					-0.001	-0.003	-0.001	-0.002
29					-0.002	-0.004	-0.001	-0.003
30					-0.005	-0.007	-0.004	-0.006
31					0.000	-0.003	0.001	-0.001
32					-0.004	-0.008	-0.003	-0.006
33					-0.005	-0.010	-0.004	-0.007
34					0.001	-0.004	0.003	-0.002
35					0.000	-0.006	0.003	-0.003
36					0.001	-0.007	0.005	-0.002
37					-0.009	-0.020	-0.004	-0.014
38					-0.008	-0.024	-0.001	-0.015
39					-0.015	-0.041	-0.003	-0.026
40					-0.030	-0.103	0.005	-0.063

Table 6. The Comparison of Nonexceedance Probability and Return Period Calculated by Derived Plotting Position Formula and Various Plotting Position Formulas

Sample size = 25								
Order	New		Gringorten(1963)		Cunnane(1978)		Blom(1958)	
	q	RT (year)	q	RT (year)	q	RT (year)	q	RT (year)
1	0.023	1.024	0.022	1.023	0.024	1.024	0.025	1.025
2	0.063	1.067	0.062	1.066	0.063	1.068	0.064	1.069
3	0.103	1.114	0.102	1.113	0.103	1.115	0.104	1.116
4	0.142	1.166	0.142	1.165	0.143	1.167	0.144	1.168
5	0.182	1.223	0.182	1.222	0.183	1.223	0.183	1.224
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
21	0.818	5.493	0.818	5.509	0.817	5.478	0.817	5.459
22	0.858	7.028	0.858	7.056	0.857	7.000	0.856	6.966
23	0.897	9.752	0.898	9.813	0.897	9.692	0.896	9.619
24	0.937	15.924	0.938	16.103	0.937	15.750	0.936	15.538
25	0.977	43.379	0.978	44.857	0.976	42.000	0.975	40.400
Sample size = 100								
Order	New		Gringorten(1963)		Cunnane(1978)		Blom(1958)	
	q	RT (year)	q	RT (year)	q	RT (year)	q	RT (year)
1	0.006	1.006	0.006	1.006	0.006	1.006	0.006	1.006
2	0.016	1.016	0.016	1.016	0.016	1.016	0.016	1.016
3	0.026	1.026	0.026	1.026	0.026	1.027	0.026	1.027
4	0.036	1.037	0.036	1.037	0.036	1.037	0.036	1.038
5	0.046	1.048	0.046	1.048	0.046	1.048	0.046	1.048
6	0.056	1.059	0.056	1.059	0.056	1.059	0.056	1.059
7	0.066	1.070	0.066	1.070	0.066	1.071	0.066	1.071
8	0.076	1.082	0.076	1.082	0.076	1.082	0.076	1.082
9	0.086	1.094	0.085	1.093	0.086	1.094	0.086	1.094
10	0.096	1.106	0.095	1.106	0.096	1.106	0.096	1.106
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
91	0.904	10.455	0.905	10.473	0.904	10.437	0.904	10.416
92	0.914	11.674	0.915	11.696	0.914	11.651	0.914	11.623
93	0.924	13.214	0.924	13.243	0.924	13.184	0.924	13.148
94	0.934	15.222	0.934	15.262	0.934	15.182	0.934	15.132
95	0.944	17.950	0.944	18.007	0.944	17.893	0.944	17.822
96	0.954	21.869	0.954	21.956	0.954	21.783	0.954	21.676
97	0.964	27.978	0.964	28.124	0.964	27.833	0.964	27.655
98	0.974	38.822	0.974	39.109	0.974	38.538	0.974	38.190
99	0.984	63.392	0.984	64.179	0.984	62.625	0.984	61.692
100	0.994	172.690	0.994	178.786	0.994	167.000	0.994	160.400

4. 결 론

본 연구에서는 Gumbel 분포에 적합한 도시위치공식을 새롭게 추정하기 위해 Gumbel 분포의 순서통계량과 확률가중모멘트를 이용하여 다양한 표본크기에 대한 도시위치공식의 기본식을 유도하였고, 도시위치공식의 매개변수를 추정하기 위해 유전자 알고리즘을 이용하였다. 또한 금회 추정된 도시위치공식을 포함하여

Gumbel 분포에 대해 가장 적합한 도시위치공식을 알아보기 위해 축소변량간의 오차를 산정하여 정확도를 비교한 결과, 아래와 같은 결론을 얻었다.

1) 유도된 기본식에 유전자 알고리즘을 적용하여 Gumbel 분포에 대한 도시위치공식을 유도하는 것이 가능한 것으로 나타났다. 그러나 제한조건이 적용되지 않은 경우 유전자 알고리즘에 의해 추정된 도시위치공식에 의해 산정되는 비초과확률과 초과확률의 합이 1을 초과하는 오류를 포함하고 있는 것으로 나타났다.
2) 초과확률과 비초과확률의 관계를 고려하여 본 연구에서 재추정한 도시위치공식은 Gringorten의 도시위치공식보다는 축소변량간 평균제곱근오차와 순서별 편의가 상대적으로 크게 산정되었으나, Cunnane과 Blom의 도시위치공식보다는 작게 산정되는 것으로 분석되었다.
3) 본 연구에서 제안한 공식을 포함해서 기존에 Gumbel 분포에 적용되고 있는 공식중에서는 Gringorten 공식이 가장 좋은 결과를 보였다.

감사의 글

본 연구는 국토해양부가 출연하고 한국건설교통기술평가원에서 위탁시행한 건설기술혁신사업(08기술혁신 F01)에 의한 차세대홍수방어기술개발연구단의 연구비 지원에 의해 수행되었습니다.

참 고 문 헌

Adamowski, K. (1981). "Plotting position formula for flood frequency." *Water Resource Bulletin*, Vol. 17, No. 2, pp. 197-201.

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.

Beyer, H.-G. and Deb, K. (2001). "On self-adaptive features in real-parameter evolutionary algorithms." *IEEE Transactions on Evolutionary Computation*, Vol. 5, No. 3, pp. 250-270.

Blom, G. (1958). *Statistical Estimates and Transformed Beta Variables*. Wiley, New York, N.Y..

Cunnane, C. (1978). "Unbiased plotting positions - A review." *Journal of Hydrology*, Vol. 37, No. 3/4, pp. 205-222.

De, M. (2000). "A new unbiased plotting position formula for Gumbel distribution." *Stochastic Environmental Research and Risk Assessment*, Vol. 14, pp. 1-7.

Deb, K. and Beyer, H.-G. (2001). "Self-adaptive genetic algorithms with simulated binary crossover." *Evolutionary Computation Journal*, Vol. 9, No. 2, pp. 197-221.

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.

Goldberg, D. E. (1989). *Genetic algorithms in search, optimization & machine learning*. Addison Wesley, Massachusetts.

Greenwood, J. A., Landwehr, J. M., Matalas, N. C., and Wallis, J. R. (1979). "Probability weighted moments : definition and relation to parameters of several distributions expressible in inverse Form." *Water Resources Research*, Vol. 15, No. 5, pp. 1049-1054.

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, N.Y., pp.28-34.

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.

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.

Haktanir, T. and Bozduman, A. (1995). "A study on sensitivity of the probability-weighted moments method on the choice of the plotting position formula." *Journal of Hydrology*, Vol. 168, pp. 265-281.

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.

Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems*. University of Michigan Press.

Natural Environment Research Council (1975). *Flood Studies Report*. Vol. 1, London.

Nguyen, V-T-V., In-na, N., and Bobee, B. (1989). "New plotting-position formula for Pearson type Ⅲ distribution." *Journal of Hydraulic Engineering*, Vol. 115, No. 6, pp. 709-730.

Nguyen, V-T-V. and In-na, N. (1992). "Plotting formula for Pearson type Ⅲ distribution considering historical information." *Environmental Monitoring and Assessment*, Vol. 23, pp. 137-152.

Weibull, W. (1939). "A statistical theory of strength of materials." *Ing. Vetenskaps Akad. Handl*, No. 151, Generalstabens Litografiska Anstals Forlag, Stockholm.

Xuewu, J., Jing, D., Shen, H. W., and Salas, J. D. (1984). "Probability plots for Pearson type Ⅲ distribution." *Journal of Hydrology*, Vol. 74, pp. 1-29.

(논문번호:08-82/접수:2008.07.22/심사완료:2009.03.31)