

Gumbel 분포형의 수정 Anderson-Darling 검정통계량 유도 및 기각력 검토

Derivation of Modified Anderson-Darling Test Statistics and Power Test for the Gumbel Distribution

신 흥 준* / 성 경 민** / 허 준 행***

Shin, Hongjoon / Sung, Kyungmin / Heo, Jun-Haeng

Abstract

An important problem in frequency analysis is the estimation of the quantile for a certain return period. In frequency analysis an assumed probability distribution is fitted to the observed sample data to estimate the quantile at the upper tail corresponding to return periods which are usually much larger than the record length. In most cases, the selection of an appropriate probability distribution is based on goodness of fit tests. The goodness of fit test method can be described as a method for examining how well sample data agrees with an assumed probability distribution as its population. However it gives generally equal weight to differences between empirical and theoretical distribution functions corresponding to all the observations. In this study, the modified Anderson-Darling (AD) test statistics are provided using simulation and the power study are performed to compare the efficiency of other goodness of fit tests. The power test results indicate that the modified AD test has better rejection performances than the traditional tests. In addition, the applications to real world data are discussed and shows that the modified AD test may be a powerful test for selecting an appropriate distribution for frequency analysis when extreme cases are considered.

Keywords : goodness-of-fit test, modified Anderson-Darling test, power test

요 지

빈도해석에 있어서 중요한 문제는 특정 재현기간에 대한 수문량의 크기를 산정하는 것으로, 빈도해석에서는 일반적으로 관측기간보다 긴 재현기간에 해당하는 수문량의 크기를 산정하기 위해 가정된 확률분포형을 표본 자료에 적합시키게 된다. 따라서 적절한 확률분포형의 선정이 무엇보다 중요하며 이는 일반적으로 대상 자료로부터 얻어지는 경험적 빈도분포와 가정한 확률분포의 일치 정도를 판단하는 적합도 검정 방법을 이용하게 된다. 일반적으로 많이 사용되는 적합도 검정 방법들은 모든 표본 자료들의 적합 정도를 동일하게 고려하기 때문에 극치사상의 크기 증가에 따른 영향은 반영하기 힘든 방법들이다. 따라서 본 연구에서는 모의실험을 통해 극치사상에 대하여 가중치를 주는 modified Anderson-Darling (AD) 검정 방법의 Gumbel 분포형에 대한 검정 통계량 한계값을 제시하였으며, 기존의 여러 적합도 검정 방법과의 기각력을 비교해 보고, 이를 실제 자료에 적용하여 그 결과를 살펴보았다. 그 결과 modified AD 검정 방법이 기존의 여러 가지 적합도 검정 방법보다도 기각력이 더 우수한 것으로 나타났으며, 기존의 적합도 검정 방법으로는 부족한 분포

* 연세대학교 대학원 토목공학과 박사후과정 (e-mail: sinong@yonsei.ac.kr)

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

** 연세대학교 대학원 토목공학과 석사과정 (e-mail: kmini33@yonsei.ac.kr)

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

*** 교신저자, 연세대학교 공과대학 토목공학과 교수 (e-mail: jhheo@yonsei.ac.kr)

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

형 선정 기준의 부족한 부분을 어느 정도 보완해 줄 수 있을 것으로 판단되었다.

핵심용어 : 적합도 검정, modified Anderson-Darling 검정 방법, 기각력 검토

1. 서 론

수문통계에 있어서 가장 중요한 문제 중 하나는 표본이 주어졌을 때 표본을 이용하여 모집단의 형태를 적절하게 유추하는 것이라고 할 수 있다. 모집단에 대해서 누구도 알지 못하는 상황에서 주어진 표본만을 이용하여 모집단을 추정한다는 것은 어려운 일이나, 주어진 자료가 한정되어 있는 현실에서 확률분포형에 최대한 적합하도록 노력하는 것이 한계일 것이다. 수문설계에 있어서 설계수문량을 산정하기 위해서는 대상 자료에 대한 무작위성 검토 후 확률분포형을 적용하고 매개변수 추정값의 검토를 마친 후 적절한 분포형을 선정하여 설계에 필요한 재현기간에 대한 확률수문량을 산정하게 된다. 이러한 확률수문량 산정시 적절한 확률분포형을 선정하기 위해서는 일반적으로 대상 자료로부터 얻어지는 경험적 빈도분포와 가정한 확률분포가 얼마나 잘 일치하는지를 판단하는 적합도 검정 방법을 이용하게 된다. 이 적합도 검정 방법에는 도시적인 방법과 통계적인 방법이 있는데, 도시적인 방법의 경우에는 주관적인 경우가 많으므로 보다 객관적인 방법으로 통계적인 방법을 많이 이용하고 있다. 통계적인 적합도 검정 방법으로는 χ^2 분포 기반의 χ^2 (CS) 검정, 경험적 분포 기반의 Kolmogorov-Smirnov (KS) 검정, Cramer von Mises (CVM) 검정, 상관성 기반의 확률도시 상관계수 (probability plot correlation coefficient; PPCC) 검정 등의 방법이 있으나, 이 방법들은 가정한 확률밀도함수와 경험적 확률밀도함수 또는 누가분포함수와 경험적 누가분포함수를 비교하거나, 순서통계량과 가정한 누가분포함수의 역함수간의 상관계수를 이용하여 확률지 상에서 선형 여부에 따른 적합도를 판단하는 방법이다. 이 적합도 검정 방법들은 그동안 수문설계에 있어서 우수한 적합도 검정 방법으로 알려져 있으나, 표본 자료들의 적합 정도를 크기에 관계없이 전체적으로 동일하게 고려하기 때문에 실제 수문설계에서 관심 있는 확률분포형의 오른쪽 꼬리 (tail) 부분보다는 전체적인 양상에 따른 적합도를 판정하는 단점을 가지고 있다. 즉, 과거의 표본 자료를 확률분포형에 적합시키는 것은 비교적 정확하게 할 수 있으나 최근의 기상이변으로 인한 극치사상의 크기 증가에 따른 영향은 반영하기 힘든 방법들이다. 따라서 수문설계에 있어서는 표본의 전체적인 양상을 고려하는 것도 중요하지만 발생빈도가 작을 수밖에 없는 극치사상에 대한 적합여부

가 더 중요하다고 할 수 있다.

Anderson-Darling (AD) 검정 방법은 Anderson and Darling (1952)에 의해 추계학적 과정에서의 오차를 감소시키기 위한 기준으로 처음 제시된 적합도 검정 방법이다. Stephens (1974, 1976)는 광범위한 모집단에서 여러 적합도 검정 방법들보다 AD검정 방법의 높은 적용성을 확인하였으나 이 방법 역시 경험적 분포함수와 이론적 분포함수 간에 동일한 가중치를 두는 단점을 가지고 있다. 한편, 이러한 단점을 극복하고자 Ahmad et al. (1988)은 기존의 AD 검정 방법을 수정하여 오른쪽 꼬리부분 또는 왼쪽 꼬리부분의 관심영역에 좀 더 높은 가중치를 두는 modified Anderson-Darling 검정 방법을 제시하였으며, generalized extreme value (GEV) 분포와 generalized logistic (GLO) 분포형에 대하여 모의실험을 통해 확률가중모멘트법에 의해 매개변수를 추정하고 이를 바탕으로 modified AD 검정을 적용하여 각 분포에 대한 한계값을 나타내었다. 하지만 이들에 의해 추정된 한계값은 형상매개변수의 고려범위가 $-0.2 \sim +0.2$ 로 제한적이어서 실제보다 과소 추정된 경향이 있는 것으로 판단된다. 또한 Sinclair et al. (1990)은 modified AD 검정 방법의 왼쪽 꼬리부분과 오른쪽 꼬리부분의 민감도 분석을 실시하여 기존의 AD 검정 방법보다 modified AD 검정 방법이 더 우수함을 보인 바 있다. 한편, Onoz and Bayazit (1995)는 빈도해석과정에서 적절한 분포형 선정을 위한 여러 가지 적합도 검정 방법을 비교하였다. 기존의 CS 검정, KS 검정, PPCC 검정, AD 검정, L-모멘트비도 등의 검정 방법을 이용하여 여러 가지 분포형에 실제 자료를 적용해 본 결과 일반적으로 χ^2 검정과 PPCC 검정의 결과가 좋게 나타났으며, 극치사상의 적합도가 중요할 경우에는 AD 검정의 효과가 매우 좋음을 보인바 있다. 또한 Crown (2000)은 modified AD 검정 방법을 이용하여 normal 분포와 exponential 분포에 대하여 적합도 검정의 기준값을 제시하였으며 그에 대한 기각력 검토를 실시한 바 있다. Arshad et al. (2003)은 확률가중모멘트법으로 매개변수를 추정한 후 generalized Pareto 분포형에 대하여 AD 검정 방법과 modified AD 검정 방법을 적용하여 기준값들을 제시하였으며, Sultan and Khaleel (2005)는 CVM 검정과 AD 검정 방법을 이용하여 normal type 2 분포에 대하여 검정 통계량 한계값을 제시하고 그 기각력을 비교하였다. 또한 Heo et al. (2008)은 3변수 gamma, GEV, gumbel, normal, 3변수 Weibull 분

포에 대하여 PPCC 검정의 검정 통계량 및 회귀식을 제시하고 기존의 적합도 검정 방법들과의 기각력에 대한 비교를 수행한 바 있다. 그 외 여러 가지 적합도 검정 방법의 한계값을 산정하기 위한 연구로, Lilliefors (1967, 1969)는 Monte Carlo 모의를 통해 normal 분포의 평균과 분산이 추정되고 exponential 분포의 평균이 추정되었을 때 KS 검정의 검정 통계량 한계값을 산정하였으며, 이 결과를 바탕으로 Stephens (1970, 1974), Durbin (1975), Green and Hegazy (1976)은 normal과 exponential의 두 분포형의 매개변수를 알지 못하는 경우에 대한 적합도 검정에 관한 연구로 확장시켰다. 또한 Chandra et al. (1981)은 extreme value 분포와 Weibull 분포에 대하여 KS 검정의 검정 통계량에 대한 연구를 수행하였으며, 그 후에도 Weibull 분포에 대한 여러 가지 적합도 검정에 대한 연구가 Littell et al. (1979), Bush et al. (1983), Aho et al. (1985), Murthy et al. (2004) 등에 의해 계속 이루어졌다.

따라서 본 연구에서는 우리나라 지점빈도해석에서 가장 많이 적용되는 Gumbel 분포에 대하여 확률가중모멘트 방법으로 추정된 매개변수를 이용하여 modified AD 검정의 검정통계량 한계값과 회귀식을 제시하고, 여러 분포형에 대하여 modified AD 검정과 기존의 여러 적합도 검정 방법과의 기각력을 비교해 보았으며 이를 실제 자료에 적용하여 그 결과를 살펴보았다.

2. 확률분포형 및 적합도 검정

수문설계 시 필요한 설계수문량을 산정하기 위하여 주로 사용하는 수문자료해석 방법은 일반적으로 확률분포형을 이용하는 빈도해석 방법이다. 대상 자료가 가정된 모집단의 특성을 나타내기 위하여 미지의 매개변수를 가지는 확률분포형을 가정하여 분석하는 것이 빈도해석의 일반적인 방법이다. 모집단의 성질을 간단한 확률분포형으로 완벽하게 나타낼 수 없으며, 추출된 자료 또한 모집단의 성격을 대표하는 값인지 알 수가 없으므로 여러 가지 적합도 검정 방법을 통하여 최대한 표본 자료가 모집단의 성질을 가지는지 여부를 판별하게 된다.

2.1 Gumbel 분포

수문자료 가운데서 강우 자료에 적합한 분포형으로 자주 선정되는 분포형은 Gumbel과 GEV 분포형이다. Gumbel 분포형은 GEV 분포형의 위치 매개변수가 0인 경우에 해당하는 2변수 확률분포형으로서 한국 확률강우량도(건설교통부, 2000)에서는 우리나라 강우자료에 적합한 확률분포형으로서 Gumbel 분포형을 채택하고 있다. Gumbel 분포형은 매개변수가 2개이기 때문에 대상 자료의 분포 양

상을 표현하는 데 있어서 매개변수가 3개인 GEV 분포형에 비해 그 성능이 다소 떨어지는 단점이 있으나 한 두개의 극값에 지나치게 영향을 받지 않아 비교적 안정적인 (robust) 추세를 나타내는 장점이 있다. Gumbel 분포의 누가분포함수와 확률밀도함수는 다음 식과 같이 주어진다 (Gumbel, 1958).

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

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

여기서, α 는 크기매개변수, x_0 는 위치매개변수이며, Gumbel 분포의 왜곡도계수 (coefficient of skewness)는 1.1396으로 고정된 상수값을 갖는다.

2.2 적합도 검정

적합도 검정은 대상 자료로부터 얻어지는 경험적 빈도 분포와 가정한 확률분포가 얼마나 잘 일치하는가를 판단하는 방법으로 도식적인 방법과 통계적인 방법에 의해 이루어진다. 도식적인 방법의 대표적인 예는 확률지 (정규 확률지, 대수정규확률지, 극치확률지, 대수극치확률지 등)를 이용하는 방법이 있는데, 만약 대상 표본자료의 경험적 누가분포가 정규확률지 상에서 직선으로 나타나면 이 자료는 정규분포를 갖는다고 할 수 있다. 그러나 이러한 방법은 주관적인 경우가 많으므로 보다 객관적인 방법으로 통계적 방법이 널리 사용되고 있다. 모집단의 확률분포형을 알지 못하면 기존의 확률분포형으로 모집단의 성질을 정확히 나타내기 어려우므로 대상 확률분포형 가운데 적정 확률분포형을 선정하는 방법 또한 매우 어렵다고 할 수 있으므로 다양한 기법을 통해서 많은 정보를 이용하여 분포형을 선정하는 것이 합리적이라 할 수 있다.

2.2.1 Anderson-Darling 검정 방법

Anderson and Darling (1952)은 다음 Eq. (3)과 같은 검정통계량을 제시하였다.

$$Q_n = n \int_{-\infty}^{+\infty} [F_n(x) - F(x)]^2 \psi(x) dF(x) \quad (3)$$

여기서, $\psi(x)$ 는 가중치 함수로, 특히 가중치 함수 $\psi(x) = 1$ 일 경우 Eq. (3)의 검정통계량은 CVM 검정의 검정통계량과 같으며, $\psi(x) = [F(x)(1-F(x))]^{-1}$ 인 경우 AD 검정통계량이라고 하며 일반적으로 Eq. (4)와 같이 A_n^2 으로 표시한다. 이 통계량은 홍수 또는 가뭄의 빈도해석에 해당하는 오른쪽, 왼쪽 꼬리부분 양쪽 모두에 가중치를 주어 표본의 중간값 보다는 표본의 극치값에 대한 적합 정

도를 강조하는 검정 방법으로, 일반적으로 수치적인 계산의 편의를 위해 Eq. (4) 대신 Eq. (5)을 사용한다.

$$A_n^2 = n \int_{-\infty}^{+\infty} \frac{[F_n(x) - F(x)]^2}{F(x)\{1 - F(x)\}} dF(x) \quad (4)$$

$$A_n^2 \approx -n - \frac{1}{n} \sum (2i-1) [\log\{F(x_{(i)})\} + \log\{1 - F(x_{(n+1-i)})\}] \quad (5)$$

여기서, $F(\cdot)$ 는 추정된 매개변수를 포함하는 누가분포함수, $F_n(\cdot)$ 은 경험적 분포함수를 나타낸다.

2.2.2 Modified Anderson-Darling 검정 방법

Anderon-Darling (AD) 검정 방법은 확률분포형의 오른쪽, 왼쪽 꼬리부분을 동일하게 가중치를 둔 방법이나, 수문설계자들은 모든 표본 자료들 중에서 홍수 또는 가뭄에 따른 극치값에 가중치를 둔 것을 더 중요하게 생각하며, 실제 일반적인 수문설계에서도 주로 재현기간 50년보다 큰 재현기간에 대한 설계수문량을 필요로 하고 있다. 이런 경우 확률밀도함수의 전체적인 적합정도 보다는 오른쪽 꼬리부분에 대한 적합정도가 더 중요하다고 할 수 있다. 따라서 Ahmad et al. (1988)은 기존의 Anderson-Darling (AD) 검정 방법에서 가중치 함수를 수정한 modified Anderson-Darling 검정 방법을 제안하였다. Modified AD 검정 방법은 오른쪽 꼬리부분에 $\psi(x) = [1 - F(x)]^{-1}$ 와 같은 가중치를 주어 왼쪽 꼬리부분보다 더 큰 가중치를 주는 것이며, 이의 검정 통계량은 AU_n^2 으로 나타내며 Eq. (6)과 같이 표현된다. 또한 이와는 반대로 왼쪽 꼬리부분의 적합도가 더 중요하다면 가중치함수를 $\psi(x) = [F(x)]^{-1}$ 와 같이 변경하여 Eq. (7)과 같이 검정 통계량 AL_n^2 으로 나타낼 수 있다.

$$AU_n^2 = n \int_{-\infty}^{+\infty} \frac{[F_n(x) - F(x)]^2}{1 - F(x)} dF(x) \quad (6)$$

$$AL_n^2 = n \int_{-\infty}^{+\infty} \frac{[F_n(x) - F(x)]^2}{F(x)} dF(x) \quad (7)$$

계산의 편의를 위해 Eqs. (6) and (7)은 다음과 같이 나타낼 수 있다.

$$AU_n^2 \approx \frac{n}{2} - 2 \sum_{j=1}^n F(x_{(j)}) - \sum_{j=1}^n \left\{ 2 - \frac{(2j-1)}{n} \right\} \log\{1 - F(x_{(j)})\} \quad (8)$$

$$AL_n^2 \approx -\frac{3n}{2} + 2 \sum_{j=1}^n F(x_{(j)}) - \sum_{j=1}^n \frac{(2j-1)}{n} \log\{F(x_{(j)})\} \quad (9)$$

여기서, $x(1) \leq x(2) \leq \dots \leq x(n)$ 은 순서통계량이며, 결과적으로 Eqs. (8) and (9)를 통해 AU_n^2 과 AL_n^2 의 합은 가중치함수 $\psi(x) = [1 - F(x)]^{-1}$ 과 $\psi(x) = [F(x)]^{-1}$ 의 합과 마찬가지로 AD 검정의 검정 통계량 및 가중치와 동일한 값을 가지는 것을 확인할 수 있다.

3. 근사적 한계값 산정

본 연구에서는 Gumbel 분포에 대한 modified AD 검정의 검정 통계량 한계값을 산정하기 위해 Monte Carlo 모의실험을 수행하였다. 모의실험은 Gumbel 분포형으로부터 일관성의 유지를 위해 위치 및 크기매개변수를 각각 0과 1로 고정시켰으며, 분포로부터 임의의 난수를 발생시켜 표본 크기 10, 15, 20, 25, 30, 35, 40, 45, 50, 100에 대한 총 10가지 경우에 대하여 각각 10,000개의 표본을 발생시키고 각각의 경우에 대하여 검정통계량 값을 계산하였다. 표본 크기에 따른 총 100,000개의 표본을 통하여 유의 수준 0.2, 0.15, 0.10, 0.05, 0.01에 해당하는, 즉 크기순으로 80%, 85%, 90%, 95%, 99%에 해당하는 값을 검정 통계량의 한계값으로 산정하였다. 다음의 Table 1은 Gumbel 분

Table 1. Modified Anderson-Darling Test Statistics for the Gumbel Distribution

n	Significance levels				
	0.20	0.15	0.10	0.05	0.01
10	0.23776	0.26402	0.30100	0.36506	0.51315
15	0.24446	0.27169	0.31046	0.37811	0.53763
20	0.24880	0.27679	0.31695	0.38420	0.54631
25	0.25003	0.27757	0.31668	0.38437	0.54946
30	0.25166	0.27975	0.31984	0.38914	0.55170
35	0.25371	0.28236	0.32168	0.39243	0.55974
40	0.25403	0.28233	0.32242	0.39153	0.55263
50	0.25531	0.28408	0.32396	0.39411	0.56675
100	0.25678	0.28532	0.32608	0.39658	0.56636

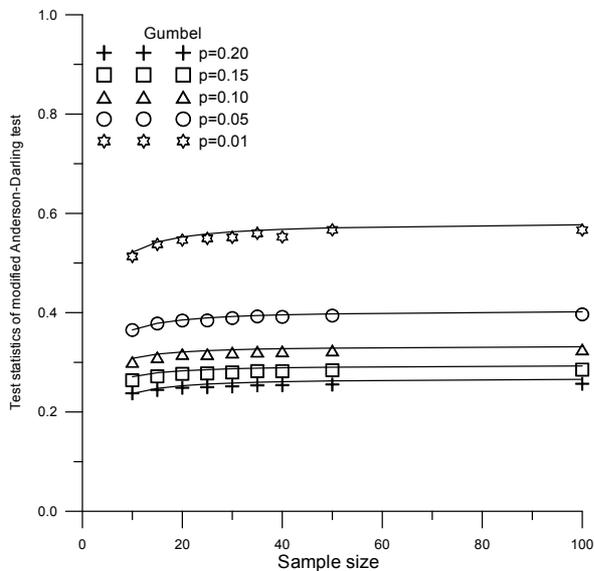


Fig. 1. The Regression Equations of Modified Anderson-Darling Test Statistics for the Gumbel Distribution

포형의 모의실험을 통해 얻어진 modified AD 검정 방법의 검정 통계량 한계값을 나타낸 표이다.

또한 위에서 산정된 각 분포형의 검정 방법에 따른 한계값을 하나의 식으로 표현하기 위해 다음 식과 같은 비선형 회귀모형을 적용하였으며, 회귀를 통한 계수는 다음 Eq. (10)과 같고 결정계수 (determination coefficient)는 0.9964로 매우 높은 정확도를 보이고 있으며, 이를 그림으로 나타내면 Fig. 1과 같다.

$$AU = \frac{1}{1 + \exp\{1.71546 + 0.44533 \ln(p) + (2.73188 - 25.58061p + 101.9063p^2)/n\}} \quad (10)$$

4. 기각력 검토

Gumbel 분포에 대한 modified Anderson-Darling (AD) 검정의 기각력을 살펴보기 위해, 기존의 χ^2 (CS) 검정, Kolmogorov-Smirnov (KS) 검정, Cramer von Mises (CVM) 검정 및 PPCC (probability plot correlation coefficient) 검정 결과와 비교해 보았다. 먼저 기각력 검토에서 표본 발생을 위한 모집단을 3변수 분포인 generalized extreme value (GEV), generalized logistic (GLO), gamma (GAM), Weibull (WBU), lognormal (LOG) 분포와 2변수 분포인 Gumbel (GUM) 분포를 고려하여 표본을 발생시켰다. 유의수준 20%, 15%, 10%, 5%, 1%의 5가지 경우 및 표본 크기 10, 20, 30, 40, 50, 100, 200의 7가지 경우에 대하여 각각 10,000번의 실험을 반복 수행하였으며, 여기서 각 분포들의 매개변수는 매개변수의 변화에 따른 영향을

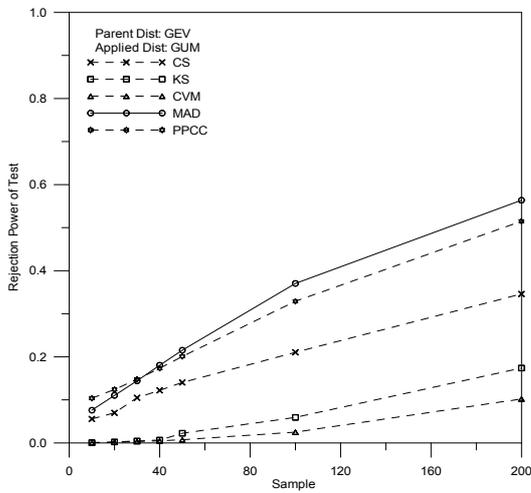
없애기 위해 위치매개변수는 0, 크기매개변수는 1로 고정하였고, 3변수 분포의 경우 형상매개변수를 왜곡도 계수에 따라 각 분포형별로 10가지 경우를 고려하였다. 앞에서 언급된 각각의 경우에 대하여 표본 자료를 발생시킨 후, 발생된 표본을 Gumbel 분포에 적합시켜 표본에 대한 매개변수를 구하고 각 적합도 검정 방법에 적용하여 기각 여부를 판단하여 총 100,000개의 표본 중 기각된 표본의 개수를 각 검정 방법의 기각력으로 정의하였다.

다음의 Fig. 2는 각각 모집단 분포를 (a)GEV, (b)GLO, (c)WBU, (d)GAM, (e)GUM, (f)LOG 분포로 가정하여 표본을 발생시키고 Gumbel 분포에 적합시켰을 때 여러 가지 적합도 검정 방법에 따른 기각력 정도를 나타낸 그림으로 여기서는 유의수준 5%에 대한 경우만 나타내었다. 대부분의 경우 표본의 크기에 따라 기각력 역시 증가하는 것을 확인할 수 있으며, 일반적으로 표본크기 10인 경우를 제외하고는 modified AD (MAD) 검정에 의한 기각력이 가장 높고, PPCC, CS, KS, CVM 순으로 기각력이 높은 것으로 나타났다. 이는 기존 연구들에서 알려진 바와 같이 PPCC 검정이 CS, KS, CVM 등 기존의 타 검정 방법보다 기각력이 월등히 높음을 확인할 수 있었으며, KS 및 CVM 검정은 모분포에 따라 매우 낮은 정도의 기각력을 보이는 것을 확인하였다. 또한 GUM 분포를 발생시켜 GUM 분포에 적합시킨 (e)의 경우에는 CS, PPCC, modified AD 검정 방법에서는 발생 및 적용한 분포가 동일한 분포이기 때문에 유의수준 5%에 근접하는 기각력을 보이고 있으나 KS 및 CVM 검정에서는 거의 기각을 시

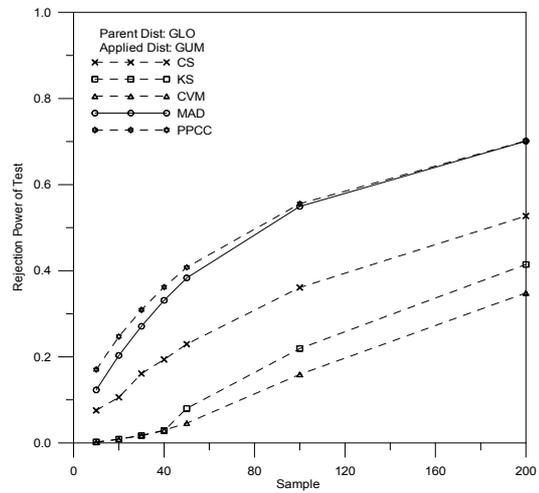
키지 못하는 것으로 나타났다. 한편 GLO 분포의 경우 나머지 GEV, WBU, GAM 분포와는 달리 modified AD 검정보다 PPCC 검정에 의한 기각력이 약간 높은 것을 볼 수 있으나 큰 차이는 없으며 표본 크기가 100보다 클 경우에는 거의 동일한 기각력을 가지는 것을 볼 수 있다. 따라서 본 연구에서 제시한 modified AD 검정 방법은 기존의 통계적 적합도 검정 방법에 있어 가장 기각력이 우수하다고 알려진 PPCC 검정 방법보다도 기각력이 더 뛰어난 것으로 나타나 향후 빈도해석의 분포형 선정에 대한 판단기준으로서 역할을 할 수 있을 것으로 판단된다.

5. 적용

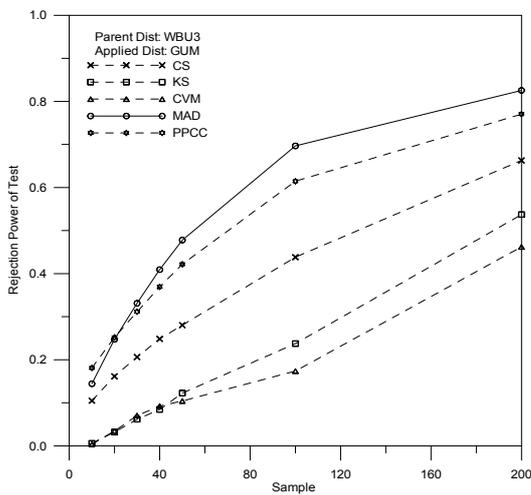
이상의 결과들을 바탕으로 우리나라 강우 자료에 modified AD 검정 방법을 실제 적용해 보았다. 이를 위해



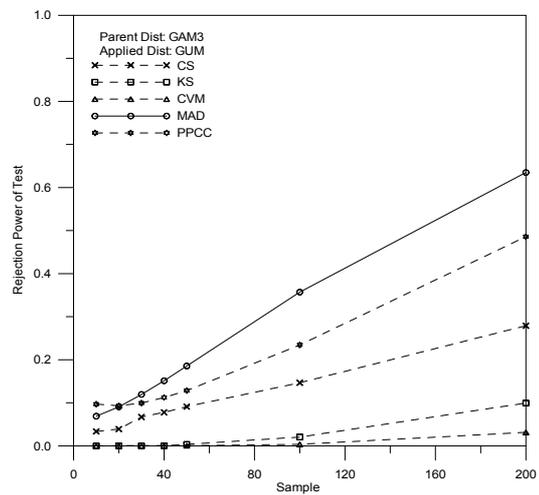
(a) Parent model: GEV



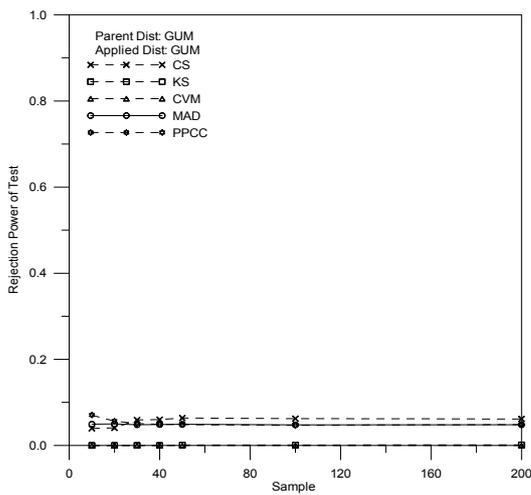
(b) Parent model: GLO



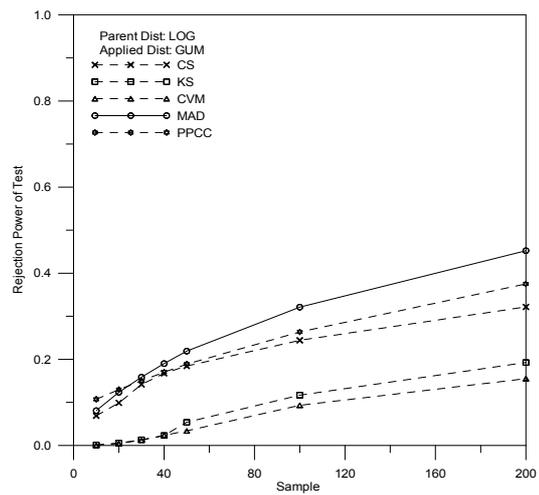
(c) Parent model: WBU



(d) Parent model: GAM



(e) Parent model: GUM



(f) Parent model: LOG

Fig. 2. Power Test of Goodness-of-Fit Tests

국내 기상청 지점 중 관측기간이 10년 미만의 짧은 지점을 제외하고 총 50개 지점에 대하여 적용하였다. 적용 지

점의 지점명 및 자료개수는 다음의 Table 2와 같으며, 지속시간 1, 2, 3, 6, 9, 12, 15, 18, 24, 48시간의 자료에 대해

Table 2. Summary of the Site Characteristics

Site		Number of Data	Site		Number of Data
Number	Name		Number	Name	
090	Sokcho	36	170	Wan-do	33
100	Daegwallyeong	30	192	Jinju	35
101	Chuncheon	37	201	Ganghwa	33
105	Gangneung	46	202	Yangpyeong	32
108	Seoul	75	203	Icheon	33
112	Incheon	51	211	Inje	32
114	Wonju	32	212	Hongcheon	32
115	Ulleung-do	53	221	Jecheon	33
119	Suwon	40	226	Boeun	33
129	Seosan	36	232	Asan	33
130	Uljin	33	235	Boryeong	32
131	Chungju	37	236	Buyeo	33
133	Daejeon	35	238	Geumsan	32
135	Chupungnyeong	49	243	Buan	31
136	Andong	30	244	Imsil	32
138	Pohang	50	245	Jeong-eup	32
140	Gunsan	36	247	Namwon	32
143	Daegu	88	256	Suncheon	32
146	Jeonju	63	260	Jangheung	33
152	Ulsan	57	261	Haenam	33
156	Gwangju	65	262	Goheung	32
159	Busan	61	272	Yeongju	32
162	Tongyeong	36	273	Mungyeong	32
165	Mokpo	81	277	Yeongdeok	32
168	Yeosu	61	278	Uiseong	32

Table 3. Number of Rejection for Each Site with the Goodness-of-Fit Tests

Site	090	100	101	105	108	112	114	115	119	129
CS	1	3	1	18	0	1	0	0	3	0
KS	0	0	0	3	0	0	0	0	0	0
CVM	0	0	0	1	0	0	0	0	0	0
PPCC	6	0	0	27	0	0	0	0	6	0
MAD	5	4	0	25	0	1	0	0	10	0
Site	130	131	133	135	136	138	140	143	146	152
CS	7	9	7	2	2	2	6	8	0	3
KS	0	0	0	0	0	0	0	0	0	0
CVM	0	0	0	0	0	0	0	0	0	0
PPCC	0	3	0	0	3	0	3	0	0	9
MAD	0	7	1	0	4	1	8	0	0	2
Site	156	159	162	165	168	170	192	201	202	203
CS	2	2	3	8	0	16	0	0	2	0
KS	0	0	0	0	0	0	0	0	0	0
CVM	0	0	0	0	0	0	0	0	0	0
PPCC	0	0	0	15	0	0	0	0	0	0
MAD	0	1	1	8	0	12	0	0	1	0
Site	211	212	221	226	232	235	236	238	243	244
CS	5	2	0	14	2	2	21	0	2	0
KS	0	0	0	0	0	0	0	0	0	0
CVM	0	0	0	0	0	0	3	0	0	0
PPCC	9	0	0	24	0	3	27	3	0	0
MAD	6	3	0	23	6	0	27	0	0	0
Site	245	247	256	260	261	262	272	273	277	278
CS	3	3	2	9	9	1	0	7	4	1
KS	0	0	0	0	0	0	0	0	0	0
CVM	0	0	0	0	0	0	0	0	0	0
PPCC	6	0	0	15	21	15	0	0	6	0
MAD	3	0	0	6	20	13	0	0	6	0

Table 4. Results of Goodness-of-Fit Test (Chungju, 12hr)

Dist.	Test	MOM			ML			PWM		
		Com.	Tab.	Check	Com.	Tab.	Check	Com.	Tab.	Check
Gumbel	CS	6.622	7.810	Accept	5.649	7.810	Accept	5.000	7.810	Accept
	KS	0.130	0.218	Accept	0.118	0.218	Accept	0.132	0.218	Accept
	CVM	0.140	0.461	Accept	0.083	0.461	Accept	0.121	0.461	Accept
	PPCC	0.966	0.958	Accept	0.966	0.958	Accept	0.966	0.958	Accept
	MAD	0.433	0.395	<i>Reject</i>	0.433	0.395	<i>Reject</i>	0.428	0.395	<i>Reject</i>

Table 5. Results of Goodness-of-Fit Test (Andong, 1hr)

Dist.	Test	MOM			ML			PWM		
		Com.	Tab.	Check	Com.	Tab.	Check	Com.	Tab.	Check
Gumbel	CS	1.667	5.990	Accept	2.333	5.990	Accept	1.667	5.990	Accept
	KS	0.089	0.242	Accept	0.110	0.242	Accept	0.100	0.242	Accept
	CVM	0.121	0.461	Accept	0.105	0.461	Accept	0.121	0.461	Accept
	PPCC	0.954	0.952	Accept	0.954	0.952	Accept	0.954	0.952	Accept
	MAD	0.344	0.392	Accept	0.394	0.392	<i>Reject</i>	0.343	0.392	Accept

여 모멘트법, 최우도법, 확률가중모멘트법에 의해 매개변수를 추정하였으며, CS, KS, CVM, PPCC 및 modified AD 적합도 검정을 수행하였다.

다음의 Table 3은 각 지점별로 10개 지속기간의 자료에 대하여 적합도 검정을 수행하였을 때 각각의 횟수를 나타낸 것으로, 모든 지점, 지속기간, 매개변수 추정방법, 적합도 검정 방법에 대한 결과를 한 지면상에 나타내기는 어려우므로, 각 지점 및 각 적합도 검정 방법별로 모든 지속기간, 모든 매개변수 추정방법에 의한 자료의 각각의 횟수를 합한 것을 나타내었다. 예를 들어 090 지점의 경우 10개 지속기간 및 3개 매개변수 추정방법에 대해서 CS 검정으로 각각의 횟수는 총 1회를 나타내며, KS 및 CVM 검정에서는 0회, PPCC 및 modified AD 검정에서는 각각 6회와 5회 각각의 횟수를 나타낸다. 50개 총 지점의 각 적합도 검정 방법별로 총 각각의 횟수는 CS 검정 193회, KS 검정 3회, CVM 검정 4회, PPCC 검정 201회 및 modified AD 검정 204회로 나타나, modified AD 검정에 의해 가장 많이 각각의 횟수를 알 수 있다. 여기서 CS 및 PPCC 검정과 modified AD (MAD) 검정 간의 각각의 횟수의 차이가 거의 없는데, 이는 4장 각각의 검정의 모의실험 결과에서 살펴보았듯이 자료 개수가 작을 경우에는 각각의 차이가 크지 않는 것을 확인할 수 있다.

위의 모든 결과를 모두 제시하기는 어려우므로, 본 논문에서는 대상 지점으로 modified AD 검정에 대해 설명하

기 용이한 결과가 나타난 Chungju 지점과 Andong 지점만 나타내었다. 두 지점의 관측 년수는 Chungju 37년, Andong 30년으로, 다음의 Table 4는 Chungju 지점의 지속시간 12시간의 결과, Table 5는 Andong 지점의 지속시간 6시간의 적합도 검정 결과를 나타낸 것이다. 결과를 살펴보면 Chungju 지점의 경우 Gumbel 분포에서 매개변수 추정방법에 상관없이 매개변수를 추정하고 modified AD 검정으로 적합도 검정을 수행하였을 경우 표본이 모집단 분포와 맞지 않다고 각각의 횟수를 알 수 있다. 기존에 국내에서 많이 사용되었던 적합도 검정 방법 (CS, KS, CVM, PPCC 방법)으로는 두 지점 모두 매개변수 추정방법에 상관없이 Gumbel 분포형이 사용 가능한 것으로 판단되었지만, 새롭게 제시된 modified AD 검정에 의해 Chungju 지점에서는 모든 매개변수 추정방법, Andong 지점에서는 최우도법에 의한 결과가 각각의 횟수 및 매개변수 추정방법 결정에 있어 주의를 기울일 수 있다고 판단되었다.

한편, 이 결과를 바탕으로 확률수문량을 도시하면 다음의 Fig. 3과 같다. 각각 Chungju 지점 및 Andong 지점의 결과로 각각의 매개변수 추정방법에 따른 확률강우량과 도시위치 공식을 이용하여 관측 자료를 도시해 본 것이다. 관측 자료는 점으로 표시하였으며, 각각의 매개변수 추정방법에 따른 확률강우량은 실선 및 점선으로 표시하였다. 두

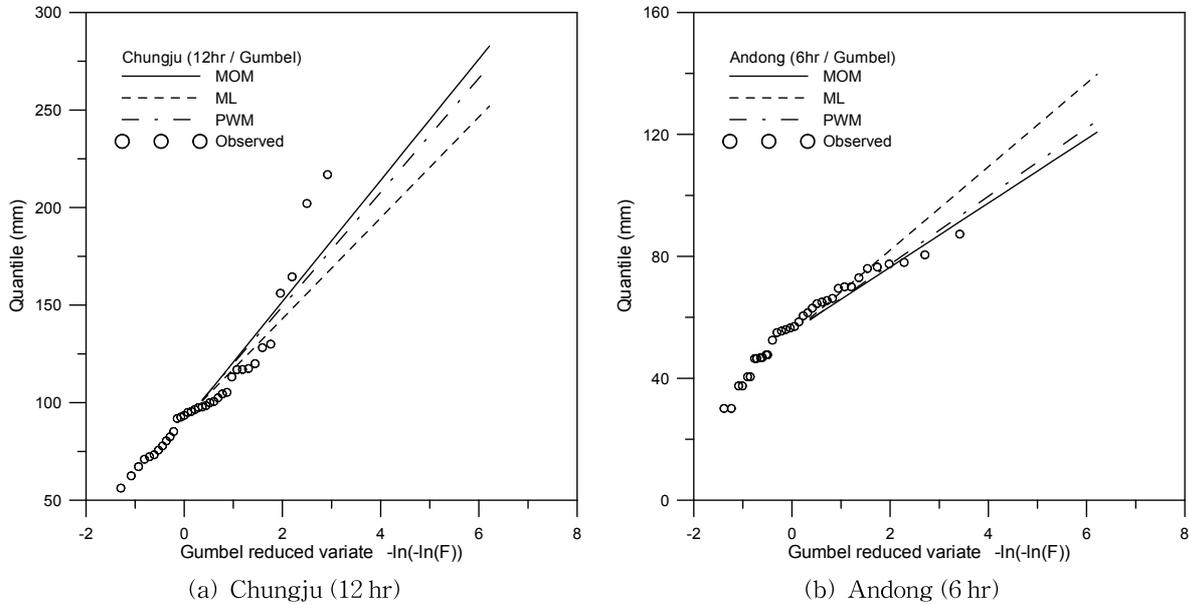


Fig. 3. Probability Distributions Fitted to Observed Rainfall Data for Each Site

그림의 경우 모두 각각의 매개변수 추정방법에 따른 결과를 눈으로 살펴볼 때는 큰 차이가 없어 보이기 때문에 분포형 선정에 있어 고민의 여지가 있으나, 위에서 제시된 표의 결과를 통해 분포형 선정에 있어 도움이 될 것으로 판단된다.

6. 결 론

본 연구에서는 적합도 검정 방법 중의 하나인 modified Anderson-Darling 검정의 검정통계량 한계값 산정 및 기각력 검토에 대하여 연구하였다. 우리나라의 빈도해석 과정에서 많이 사용되어온 기존의 적합도 검정 방법 역시 충분히 효과적인 검정 방법이나 수문설계에 있어서 중요한 부분인 극치값에 중점을 둔다는 점에서 Ahmad et al. (1988)에 의해 개발된 modified Anderson-Darling 검정 방법이 매우 유용할 것으로 판단된다. 본 연구에서는 우리나라 지점빈도해석에서 대표적으로 많이 적용되는 Gumbel 분포의 modified Anderson-Darling 검정의 검정 통계량 한계값을 Monte Carlo 모의실험을 통해 제시하였으며, 이를 표본크기와 유의수준에 대한 다변량 비선형 회귀식에 적용하여 근사적인 계수값을 계산하였으며 비교적 정확하게 회귀되는 것을 알 수 있었다. 또한 Gumbel 분포형에 대한 modified Anderson-Darling 검정과 기존의 여러 통계적 적합도 검정 방법과의 기각력을 비교해 보고, 이를 실제 자료에 적용하여 그 결과를 살펴보았다. 기각력 검토 결과 기존의 여러 가지 적합도 검정 방법보다도 기각력이 더 우수한 것으로 나타났으며,

이를 실제 우리나라 강우 자료에 적용해 본 결과, 기존의 적합도 검정 방법으로는 부족한 분포형 선정 기준의 부족한 부분을 어느 정도 보완해 줄 수 있을 것으로 판단되었다.

감사의 글

본 연구는 국토해양부가 출연하고 한국건설교통기술평가원에서 위탁시행한 건설기술혁신사업 (08기술혁신F01)에 의한 차세대홍수방어기술개발연구단의 연구비 지원에 의해 수행되었습니다.

참고문헌

- 건설교통부 (2000). **수자원관리기법개발연구조사 보고서: 제1권 한국확률강우량도 작성.**
- Ahmad, M.I., Sinclair, C.D., and Spurr, B.D. (1988). "Assessment of flood frequency models using empirical distribution function statistics." *Water Resources Research*, Vol. 24, No. 8, pp. 1323-1328.
- Aho, M., Bain, L.J., and Engelhardt, M. (1985). "Goodness-of-fit tests for the Weibull distribution with unknown parameters and heavy censoring." *J. Stat. Comp. Simul.*, Vol. 21, pp. 213-225.
- Anderson, T.W., and Darling, D.A. (1952). "Asymptotic theory of certain "goodness-of-fit" criteria based on stochastic processes." *Ann. Math. Stat.*, Vol. 23, pp.

- 193-212.
- Arshad M., Rasool, M.T., and Ahmad, M.I. (2003). "Anderson Darling and modified Anderson Darling tests for generalized Pareto distribution." *Pakistan Journal of Applied Sciences*, Vol. 3, No. 2, pp. 85-88.
- Bush, J.G., Woodruff, B.W., Moore, A.H., and Dunne, E.J. (1983). "Modified Cramer von Mises and Anderson-Darling tests for Weibull distributions with unknown location and scale parameters." *Comm. Stat. Theory Meth*, Vol. 12, No. 21, pp. 2465-2476.
- Chandra, M., Singpurwalla, N.D., and Stephens, M.A. (1981). "Kolmogorov statistics for tests of fit the extreme value and Weibull distributions." *Journal of the American Statistical Association*, Vol. 76, No. 375, pp. 729-731.
- Crown, J.S. (2000). "Percentage points for directional Anderson-Darling goodness-of-fit tests." *Commun. Statist.-Simula.*, Vol. 29, No. 2, pp. 523-532.
- Durbin, J. (1975). "Kolmogorov-Smirnov tests when parameters are estimated with applications to tests of exponentially and test on spacings." *Biometrika*, Vol. 62, No. 1, pp. 5-22.
- Green, J.R., and Hegazy, Y.A.S. (1976). "Powerful modified EDF goodness-of-fit tests." *Journal of the American Statistical Association*, Vol. 71, pp. 204-209.
- Gunbel, E.J. (1958). *Statistics of Extremes*. Columbia University Press, New York, NY.
- Heo, J., Kho, Y.W., 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.
- Lilliefors, H.W. (1967). "On the Kolmogorov-Smirnov test for normality with mean and variance unknown." *Journal of the American Statistical Association*, Vol. 62, pp. 399-402.
- Lilliefors, H.W. (1969). "On the Kolmogorov-Smirnov test for the exponential distribution with mean unknown." *Journal of the American Statistical Association*, Vol. 64, pp. 387-389.
- Littell, R.D., McClave, J.T., and Offen, W.W. (1979). "Goodness-of-fit tests for the two parameter Weibull distribution." *Comm. Stat. Simul. Comp.*, Vol. 8, No. 3, pp. 257-269.
- Murthy, P.D.N., Xie, M., and Jiang, R. (2004). *Weibull Models*, Wiley, NJ, USA.
- Onoz, B., and Bayazit, M. (1995). "Best-fit distributions of largest available flood samples." *Journal of Hydrology*, Vol. 167, pp. 195-208.
- Sinclair, C.D., Spurr, B.D., and Ahmad, M.I. (1990). "Modified Anderson Darling test." *Commun. Statist.-Theory Meth*, Vol. 19, No. 10, pp. 3677-3686.
- Stephens, M.A. (1970). "Use of the Kolmogorov-Smirnov, Cramer von Mises and related statistics without extensive tables." *J. Roy. Stat. Soc.*, Vol. B, No. 32, pp. 115-122.
- Stephens, M.A. (1974). "EDF statistics for goodness of fit and some comparisons." *J. Amer. Stat. Ass.*, Vol. 69, pp. 730-737.
- Stephens, M.A. (1976). "Asymptotic power of EDF statistics for exponentiality against gamma and Weibull alternatives." Tech report No. 297, Dept. Stat. Stanford Univ.
- Sultan, A.M., and Khaleel, H.M. (2005). "A new modified goodness of fit tests for type 2 censored sample from normal population." *Mathematical and Computational Applications*, Vol. 10, No. 2, pp. 165-177.

논문번호: 10-054	접수: 2010.07.06
수정일자: 2010.08.18/08.25	심사완료: 2010.08.25