

## 3변수 Weibull 분포형의 형상매개변수 및 극치값 가중치를 고려한 EDF 검정에 대한 연구

### A Study on Empirical Distribution Function with Unknown Shape Parameter and Extreme Value Weight for Three Parameter Weibull Distribution

김 태 립\* / 신 흥 준\*\* / 허 준 행\*\*\*

Kim, Taereem / Shin, Hongjoon / Heo, Jun-Haeng

---

#### Abstract

The most important procedure in frequency analysis is to determine the appropriate probability distribution and to estimate quantiles for a given return period. To perform the frequency analysis, the goodness-of-fit tests should be carried out for judging fitness between obtained data from empirical probability distribution and assumed probability distribution. The previous goodness-of-fit could not consider enough extreme events from the recent climate change. In this study, the critical values of the modified Anderson-Darling test statistics were derived for 3-parameter Weibull distribution and power test was performed to evaluate the performance of the suggested test. Finally, this method was applied to 50 sites in South Korea. The result shows that the power of modified Anderson-Darling test has better than other existing goodness-of-fit tests. Thus, modified Anderson-Darling test will be able to act as a reference of goodness-of-fit test for 3-parameter Weibull model.

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

---

#### 요 지

적절한 확률분포형을 결정하고 그에 따른 확률수문량을 산정하는 것은 빈도해석에서 가장 중요한 절차이며, 이를 수행하기 위해서는 경험적 확률분포에서 얻어지는 자료와 가정한 확률분포에서 얻어지는 자료의 일치 정도를 판별하는 적합도 검정을 거쳐야 한다. 지금까지 일반적으로 적용된 적합도 검정 방법은 분포형의 전체적인 적합정도를 판별하여 최근의 기상이변으로 인한 극치 사상에 대하여는 충분히 고려하지 못하고 있다. 따라서 본 연구에서는 분포형의 극치 사상에 가중치를 주는 modified Anderson-Darling(AD) 검정 방법을 3변수 Weibull 분포형에 적용하여 검정통계량 한계값과 기각력을 살펴보았으며 이를 실제 자료에 적용한 결과, modified AD 검정 방법이 다른 기존의 적합도 검정보다 더 우수한 기각력을 가지고 있음을 확인하였다. 이는 앞으로 3변수 Weibull 분포형을 이용한 극치 수문량 선정에 있어 modified AD 방법이 하나의 기준으로 작용할 수 있을 것이라 판단된다.

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

---

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

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

\*\* 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.

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

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

## 1. 서 론

수문설계를 하기 위해서는 수문자료 계열을 가지고 적절한 통계분석을 실시하여 모집단의 특성을 파악한 후, 이를 확률분포형으로 나타내어 확률수문량을 얻는 과정이 필요하다. 여러가지 확률분포형 중에서 모집단 전체를 완벽하게 나타내는 확률분포형은 없으므로 보다 적절한 분포형을 선택하는 것이 중요시되고 있다. 최적 확률분포형을 선정하기 위해서는 적합도 검정을 거치게 되는데 이 방법은 대상 자료로부터 얻어지는 경험적 빈도분포와 가정한 확률분포의 일치 정도를 판단하는 방법으로, 관측 수문 자료로부터 검정통계치를 산정한 후 이를 표본자료의 수와 유의수준에 의해 결정되는 검정방법별 검정한계치와 비교하여 가정한 분포형과의 적합여부를 판단한 후 확률분포형을 결정하게 된다. 적합도 검정 방법에는 확률지를 사용하는 도시적인 방법과 확률분포를 이용하는 통계적인 방법이 있는데, 도시적인 방법보다 객관적인 방법인 통계적 적합도 검정을 많이 이용하고 있다. 통계적 적합도 검정 방법으로는  $\chi^2$  분포 기반의  $\chi^2$  (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 기존에 주로 사용되고 있으며 우수한 기각력을 보이고 있지만, 표본 자료들의 적합 정도를 전체적으로 분포형에 동일하게 고려하기 때문에 확률분포형의 극치사상에 대한 적합도가 전체적인 양상의 적합도보다 떨어지는 단점이 있다. 따라서 전체적인 표본의 양상을 고려하는 것도 중요하지만, 최근 강우량 및 홍수량의 발생 경향은 기상이변에 따른 극치사상의 발생이 증가하고 있는 추세이므로 적합도 검정 역시 극치사상에 대한 적합여부를 더 고려하여 추정해야 할 것으로 판단된다.

Anderson-Darling (AD) 검정 방법은 Anderson and Darling (1952)에 의해 분포의 양쪽 꼬리에 가중치를 줌으로써 분포형의 극치 사상의 적합정도를 향상시키기 위하여 제안되었다. 또한 Ahmad et al. (1988)은 AD 검정 방법을 수정하여 오른쪽 또는 왼쪽으로 뻗어있는 꼬리부분의 한쪽의 극치사상에 가중치를 두는 modified AD 검정 방법을 제시하였다. 이를 generalized extreme value (GEV) 분포와 generalized logistic (GLO) 분포에 대해 모의실험을 하고 확률가중모멘트법으로 매개변수를 추정한 뒤 극치값에 대한 가중치를 고려하는 적합도 검정 방법인 modified AD 검정 방법에 적용하여 각 분포에 대한 한계값을 추정하였다. 또한 Sinclair et al. (1990)은 modified AD 검정 방

법의 오른쪽 및 왼쪽꼬리부분의 민감도 분석을 수행하여 AD 검정 방법보다 modified AD 검정 방법이 더 나은 방법임을 검증하였고, Porter III et al. (1992)은 KS, AD 및 CVM 검정 방법을 Pareto 분포형에서 발생시킨 표본에 적용한 결과 검정통계량과 형상 매개변수가 관련이 있음을 밝혔다. Evans et al. (1989)은 2변수 Weibull 분포형과 3변수 Weibull 분포형의 적합도 검정 비교에서 2변수 Weibull 분포형으로부터 발생시킨 변수에 적합도 검정을 실시하였을 때 AD 검정 방법이 다른 적합도 검정 방법보다 우수함을 보였고, 3변수 Weibull 분포형으로부터 발생시킨 변수들의 적합도 검정 결과의 상관계수가 더 높음을 확인하였다. 또한 기각값은 표본 크기의 범위와 관련이 있음을 보였고 기각력은 분포형으로부터의 자료에 가장 영향을 많이 받음을 밝힌 바 있다. 한편 Choulakian and Stephens (2001)은 CVM 검정과 AD 검정 방법으로 generalized Pareto (GPA) 분포에 대하여 적합도 검정을 수행하였고, Luceno (2006)는 경험적 분포함수를 기반으로 한 KS, CVM, AD 검정 방법과 modified AD 검정 방법을 오른쪽 꼬리, 왼쪽 꼬리, 2차 오른쪽 꼬리, 2차 왼쪽 꼬리, 2차 AD 방법의 총 다섯 가지 방법으로 분류하여 normal 분포와 GPD 분포에 적용한 후 비교하였다. 한편 Tolikas and Heravi (2008)은 확률가중모멘트법을 이용하여 3변수 log-normal 분포의 매개변수를 추정하고 AD 검정을 통한 적합도 검정을 수행하였고, Shin et al. (2010)은 Gumbel 분포형에 대하여 modified AD 검정 방법을 이용하여 검정통계량 한계값을 유도하여 제시한 후 기각력을 검토하였다. 이후 Pakyari와 Balakrishnan (2012)은 2변수 분포인 normal, Gumbel 및 log-normal 분포형으로부터 발생시킨 변수를 KS, CVM, AD 검정 방법으로 Gumbel과 Weibull 분포형에 적용하여 비교하였다. 이와 같은 연구는 Lewis et al. (1961), Stephens (1974, 1977), Littell et al. (1979), Arshad et al. (2003), Laio (2004)에 의해 현재까지도 꾸준히 여러 방법과 적용을 거듭하며 이루어지고 있으며, 따라서 본 연구에서는 지역빈도해석에 관한 연구에서 많이 활용된 바 있는 Weibull 분포형에 대하여 Monte Carlo 모의 실험으로부터 표본을 발생시키고, 극치값 가중치를 고려한 modified AD 검정으로 적합도 검정을 실시하여 유의수준별, 표본크기별로 구분한 뒤 Weibull 분포의 매개변수 조건에 따른 왜곡도 계수 해당 범위 내에서 임의의 형상 매개변수를 지정하여 변화시켜가며 그에 따른 검정통계량 한계값과 회귀식을 제시한 후 여러 분포형에 대하여 극치값 가중치를 고려한 검정 방법 및 기존 적합도 검정 방법들과의 기각력을 비교하였다. 또한 앞의 결과들을 바탕으

로 실제 자료에 적용하여 그 결과를 살펴보았다.

## 2. 확률분포형 및 적합도 검정

수문자료의 빈도분석에서는 무엇보다도 표본 자료에 가장 알맞은 최적 확률분포형을 채택하여 이를 통해 알맞은 확률강우량을 산정하는 것이 중요하다. 이를 시행하기 위해서는 표본 자료의 기본적 통계값 계산 후 예비해석(randomness test), 매개변수 추정 및 매개변수 적합성을 검토한 후 적합도 검정을 실시하여 확률분포형에 따른 확률강우량을 산정 등의 절차를 거쳐야 한다. 적합도 검정은 확률분포형을 결정하는 가장 중요한 절차 중 하나로서 한정된 자료 안에서 다양한 기법을 통하여 최적의 방법으로 결과를 도출하는 것이 합리적이라 할 수 있다.

### 2.1 Weibull 분포

Weibull 분포는 일반적으로 연 최대치 계열을 나타내는 데 효과적인 분포형으로 generalized extreme value type 3(GEV-3) 분포로도 알려져 있으며 우리나라에서는 지역 빈도해석에 관한 연구에서 활용된 바 있다. 3변수 Weibull 분포의 누가분포함수(CDF)와 확률밀도함수(PDF)는 다음 Eqs. (1) and (2)와 같다(Johnson and Kotz, 1970).

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

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

$$x_0 \leq x < \infty, \quad \alpha > 0, \quad \beta > 0$$

여기서,  $\alpha$ 는 크기매개변수,  $\beta$ 는 형상매개변수,  $x_0$ 는 위치매개변수를 나타낸다.

### 2.2 적합도 검정

적합도 검정은 관측 수문자료로부터 얻어지는 경험적 확률분포와 가정한 이론 확률분포형이 적합하다는 가설을 검정하기 위한 방법으로 적절한 확률분포형을 선정하는데 중요한 역할을 한다. 이러한 적합도 검정 방법에는 도식적인 방법과 통계적인 방법이 있는데 대표적인 도식적 방법은 확률지를 이용하는 방법으로 관측 수문자료의 경험적 누가분포를 정규확률지 상에 도시하여 나타내는 방법이지만 이러한 방법은 주관적일 가능성이 크므로 일반적으로는 객관적인 통계적 방법을 주로 사용하고 있다. 여러 확률분포형 중에서 모집단을 가장 비슷하게 나타내

는 확률분포형을 정확하게 선정하는 것은 매우 어려우므로 다양한 기법으로 검정을 수행하여 비교한 후 최적 확률분포형을 선정하는 것이 중요하다고 할 수 있다.

### 2.2.1 Anderson-Darling 검정 방법

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

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

여기서,  $\psi(x)$ 는 가중치 함수로,  $\psi(x)=1$ 일 경우 위의 식은 CVM의 검정통계량과 같다.  $\psi(x) = [F(x)(1-F(x))]^{-1}$ 일 경우, 이를 AD 검정통계량이라 하며 이를 식으로 나타내면 Eq. (4)와 같다.

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

여기서,  $F(x)$ 는 추정된 매개변수를 포함하는 누가분포함수,  $F_n(x)$ 는 경험적 누가분포함수를 나타낸다. 제시된 AD함수는 오른쪽과 왼쪽 꼬리부분 양쪽 모두 가중치를 줌으로써 일반 분포보다 극치값의 적합 정도에 가중을 주어 계산하는 방법으로 극한 수문사상의 빈도해석에 적절한 방법이라 할 수 있다. 또한, 일반적인 계산의 편의를 위해 Eq. (4)는 Eq. (5)로 변환시켜 사용한다.

$$A_n^2 \approx -n - \frac{1}{n} \sum_{i=1}^n (2i-1) [\log F(x_i) + \log 1 - F(x_{n+1-i})]$$

$$\approx -n - \frac{1}{n} \sum_{i=1}^n (2i-1) \log F(x_i) - \frac{1}{n} \sum_{i=1}^n (2n-2i+1) \log F(x_i) \quad (5)$$

### 2.2.2 Modified Anderson-Darling 검정 방법

Anderson-Darling 검정 방법은 분포의 양쪽 꼬리에 동일한 가중치를 주는 방법이다. 최근 기후변화에 따른 홍수 또는 가뭄 등의 극한 사상이 증가됨에 따라 극치값의 사상이 중요시 여겨지고 있다. 따라서 이러한 극한수문사상의 경우 확률밀도함수의 꼬리부분에 가중을 두어 적합 정도를 판단하는 것이 더 적절하다고 판단된다. 이와 같은 상황을 반영하여 Ahmad et al. (1988)은 기존의 AD 검정 방법에서 가중치 함수를 수정하여 modified AD 검정 방법을 제안하였다. Modified AD 검정 방법은 한쪽 꼬리부분에 가중치를 주어 적합 정도를 판별하는 방법으

로 가중치함수  $\psi(x)$ 에 따라 오른쪽 또는 왼쪽 꼬리부분의 적합정도에 가중치를 주어 계산할 수 있다. 홍수사상의 경우에는 오른쪽 꼬리에 가중을 주고 이는 가중치함수  $\psi(x) = [1 - F(x)]^{-1}$ 과 같이 계산하고, 반대로 가뭄사상일 경우에는 왼쪽 꼬리부분에 가중치를 주어 가중치함수  $\psi(x) = [F(x)]^{-1}$ 과 같이 계산하여 나타내며 이는 Eqs. (6) and (7)과 같다.

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

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

계산의 편의를 위해 Eqs. (6) and (7)은 다음과 같이 변환시켜 사용한다.

$$AU_n^2 \approx \frac{n}{2} - 2 \sum_{i=1}^n F(x_i) - \sum_{i=1}^n (2 - \frac{2i-1}{n}) \log\{1 - F(x_i)\} \quad (8)$$

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

여기서, Eqs. (8) and (9)의 합으로 나온 검정통계량과 가중치함수의 합은 AD 검정 방법의 검정통계량과 가중치함수의 합과 같음을 알 수 있다. 즉,  $A_n^2 = AU_n^2 + AL_n^2$ 으로 표현할 수 있다.

### 3. 근사적 한계값 산정

본 연구에서는 Monte Carlo 모의실험을 수행하여 3변수 Weibull 분포에 대한 modified AD 검정의 검정통계량 한계값을 산정하였다. 모의실험은 표본 크기를 10, 15, 20, 25, 30, 35, 40, 50, 100의 총 9가지 경우로 정하고, 일관성 유지를 위하여 Weibull 분포의 위치매개변수 및 크기매개변수를 각각 0과 1로 고정시킨 후 형상매개변수를 1.5, 2.0, 2.2, 2.4, 2.6, 2.8, 3.0, 4.0, 5.0, 9.0으로 변화시키면서 표본크기별로 각각 10,000개의 임의의 난수를 발생시켰다. 여기서 형상매개변수  $\beta$ 는 0보다 커야하는 Weibull 분포의 매개변수 조건에 따라 왜곡도 계수에 해당하는 범위 내에서 변화시켜 나타내었으며 이에 해당하는 식은 Eq. (10)과 같다(Heo et al., 2001).

$$\begin{aligned} \beta = & -0.729268 - 0.338679 C_S \\ & + 4.96077 (C_S + 1.14)^{-1.0422} \\ & + 0.683609 [\ln(C_S + 1.14)]^2 \end{aligned} \quad (10)$$

여기서,  $C_S$ 는 왜곡도 계수를 나타내며 범위는  $-1.08 \leq C_S \leq 6.0$ 과 같다.

이에 따라 발생된 표본을 오름차순으로 정리한 후 크기 순으로 80%, 85%, 90%, 95%, 99%에 해당하는 즉, 유의수준 0.2, 0.15, 0.10, 0.05, 0.01의 값을 검정통계량 한계값으로 산정하였다. 다음의 Table 1은 Weibull 분포형의 modified AD 검정방법을 통한 검정통계량 한계값을 나타낸 표이다. 여기서는 표본 크기 10, 30, 50, 100에 대한 값만 나타내었다.

또한 위에서 산정된 한계값을 하나의 식으로 나타내기 위해 유의수준( $p$ ), 형상매개변수( $\beta$ ), 표본크기( $n$ )의 함수로 표현되는 다중 비선형회귀방정식을 적용하였다. 적용된 식은 다음과 같다.

$$Y = a + b/n + c/n^2 + d\beta + e\beta^2 \quad (11)$$

Table 2는 회귀식 적용 결과 산정된 회귀계수를 나타내며, 각 유의수준별 결정계수(determination coefficient)는 0.99154, 0.98980, 0.98836, 0.98134, 0.96818로 회귀식이 높은 정확도를 가짐을 보였다.

### 4. 기각력 검토

Modified Anderson-Darling (AD) 검정 방법의 적합성과 기각력을 검토하기 위하여 Weibull 분포에 기존의 적합도 검정 방법인  $x^2$ (CS) 검정, Kolmogorov-Smirnov (KS) 검정, Cramer von Mises (CVM) 및 Anderson-Darling (AD) 검정 결과와 비교 검토해 보았다.

표본 크기 10, 15, 20, 25, 30, 35, 40, 50, 100, 200의 경우에 대하여 normal (NOR), Cauchy (CAU), Logistic (LGS), generalized extreme value (GEV), generalized logistic (GLO), generalized Pareto (GPA), exponential (EXP), Weibull (WBU) 등 각각 8가지 분포형을 모집단으로 하는 표본을 유의수준 0.2, 0.15, 0.10, 0.05, 0.01을 고려하여 발생시켰다. 10개의 표본 크기에 대하여 각각 10,000번을 반복 수행하였으며, 각 분포의 매개변수는 일관성 유지를 위하여 위치매개변수 및 크기매개변수는 각각 0, 1로 고정하였고 형상매개변수는 왜곡도 계수에 따라 각 분포형별 허용 범위 내에서 10가지 경우를 고려하였다. 분포형별로 발생된 표본을 Weibull 분포에 적합시켜 매개변수를 구하고 앞에서 언급한 기존의 적합도 검정 방법에 적용하여 기각여부를 판단하였다.

다음의 Table 3은 위의 8가지 분포형을 모집단으로 하여 발생한 표본을 3변수 Weibull 분포에 적합시켰을 때

**Table 1. Modified Anderson–Darling Test Statistics for the Weibull Distribution**

sample size (n)	shape parameter ( $\beta$ )	significance levels ( $p$ )				
		0.2	0.15	0.1	0.05	0.01
10	1.5	0.15817	0.17197	0.19181	0.22267	0.30632
10	2.0	0.15771	0.17239	0.19212	0.22329	0.30759
10	2.2	0.15840	0.17298	0.19249	0.22417	0.30764
10	2.4	0.15877	0.17351	0.19313	0.22470	0.30954
10	2.6	0.15889	0.17353	0.19324	0.22500	0.30790
10	2.8	0.15897	0.17387	0.19356	0.22529	0.31239
10	3.0	0.15876	0.17367	0.19359	0.22535	0.31219
10	4.0	0.15942	0.17433	0.19449	0.22696	0.31075
10	5.0	0.16042	0.17450	0.19457	0.22709	0.30854
10	9.0	0.16135	0.17558	0.19565	0.22904	0.31305
30	1.5	0.17116	0.18765	0.21090	0.25173	0.34611
30	2.0	0.17346	0.19007	0.21467	0.25648	0.35480
30	2.2	0.17422	0.19123	0.21555	0.25836	0.35505
30	2.4	0.17462	0.19136	0.21605	0.25901	0.35710
30	2.6	0.17544	0.19234	0.21672	0.26030	0.35992
30	2.8	0.17593	0.19277	0.21813	0.26067	0.36279
30	3.0	0.17638	0.19348	0.21884	0.26285	0.36635
30	4.0	0.17804	0.19587	0.22139	0.26687	0.36869
30	5.0	0.17862	0.19634	0.22183	0.26768	0.37057
30	9.0	0.17742	0.19538	0.22094	0.26582	0.36917
50	1.5	0.17429	0.19045	0.21513	0.25393	0.34926
50	2.0	0.17705	0.19390	0.21801	0.25804	0.35441
50	2.2	0.17832	0.19566	0.21896	0.25938	0.35365
50	2.4	0.17885	0.19644	0.21948	0.26017	0.35531
50	2.6	0.17929	0.19764	0.22041	0.26143	0.35872
50	2.8	0.17957	0.19792	0.22090	0.26191	0.36000
50	3.0	0.17985	0.19829	0.22134	0.26290	0.36215
50	4.0	0.18218	0.20103	0.22509	0.26808	0.37278
50	5.0	0.18334	0.20260	0.22702	0.26995	0.37568
50	9.0	0.18279	0.20232	0.22625	0.26854	0.37438
100	1.5	0.17808	0.19531	0.21717	0.25826	0.34595
100	2.0	0.17819	0.19548	0.21750	0.25835	0.34432
100	2.2	0.17929	0.19652	0.22013	0.25953	0.34966
100	2.4	0.17972	0.19687	0.22161	0.26086	0.35393
100	2.6	0.18070	0.19736	0.22257	0.26213	0.36061
100	2.8	0.18119	0.19761	0.22286	0.26308	0.36021
100	3.0	0.18178	0.19844	0.22422	0.26540	0.36143
100	4.0	0.18474	0.20125	0.22765	0.26956	0.36812
100	5.0	0.18693	0.20346	0.23056	0.27292	0.37426
100	9.0	0.18882	0.20579	0.23183	0.27378	0.37725

**Table 2. Regression Coefficient and Determination Coefficient**

Coefficient	Significance levels ( <i>p</i> )				
	0.2	0.15	0.1	0.05	0.01
a	0.17742	0.19350	0.21669	0.25348	0.33112
b	-0.34385	-0.34412	-0.37891	-0.32226	0.15638
c	0.75994	0.46665	0.20319	-1.35596	-6.88932
d	0.00361	0.00428	0.00533	0.00730	0.01392
e	-0.00026	-0.00031	-0.00039	-0.00054	-0.00108
<i>R</i> <sup>2</sup>	0.99154	0.98980	0.98836	0.98134	0.96818

**Table 3. Power of Goodness-of-Fit Test**

▪ Applied distribution : Weibull

Alternative Distribution	n=10					n=30				
	CS	KS	CVM	AD	mAD	CS	KS	CVM	AD	mAD
NOR	0.016	0.000	0.000	0.056	0.059	0.037	0.000	0.000	0.056	0.079
CAU	0.106	0.000	0.000	0.222	0.224	0.539	0.072	0.097	0.793	0.740
LGS	0.013	0.000	0.000	0.055	0.067	0.039	0.000	0.000	0.113	0.120
GEV	0.010	0.000	0.000	0.052	0.070	0.036	0.000	0.000	0.049	0.072
GLO	0.011	0.000	0.000	0.052	0.066	0.036	0.000	0.000	0.067	0.092
GPA	0.014	0.000	0.000	0.068	0.104	0.044	0.000	0.000	0.111	0.193
EXP	0.014	0.000	0.000	0.070	0.117	0.052	0.000	0.000	0.111	0.204
WBU	0.010	0.000	0.000	0.052	0.067	0.033	0.000	0.000	0.056	0.084

Alternative Distribution	n=50					n=100				
	CS	KS	CVM	AD	mAD	CS	KS	CVM	AD	mAD
NOR	0.042	0.000	0.000	0.058	0.087	0.040	0.000	0.000	0.072	0.099
CAU	0.833	0.316	0.416	0.955	0.911	0.997	0.900	0.944	0.974	0.945
LGS	0.045	0.000	0.000	0.155	0.164	0.066	0.000	0.000	0.297	0.310
GEV	0.037	0.000	0.000	0.049	0.071	0.039	0.000	0.000	0.050	0.072
GLO	0.040	0.000	0.000	0.085	0.108	0.045	0.000	0.000	0.119	0.148
GPA	0.052	0.000	0.000	0.148	0.262	0.062	0.000	0.000	0.223	0.384
EXP	0.055	0.000	0.000	0.142	0.263	0.058	0.000	0.000	0.193	0.345
WBU	0.036	0.000	0.000	0.060	0.098	0.039	0.000	0.000	0.066	0.115

여러 적합도 검정 방법에 따른 기각력 정도를 나타낸 그림이다. 기각력은 총 100,000개의 표본 중에서 기각된 표본의 개수의 비로 나타내었고, 여기서는 표본 크기 10, 30, 50, 100에 해당하는 경우와 일반적으로 많이 쓰이는 유의 수준 5%에 대한 결과를 표현하였다.

검토 결과 기각력은 modified AD, AD, CS, CVM, KS 순으로 높음을 보였고, modified AD에 의한 기각력이 다

른 방법에 의한 기각력보다 높음을 확인할 수 있었다. 특히 기존의 적합도 방법으로 널리 쓰이고 있는 KS, CVM 검정 방법에 비하여 월등히 높은 기각력을 보이고 있었다. 또한 모든 검정 방법의 경우 표본 크기가 증가할수록 기각력 역시 증가하는 것으로 나타났다. 한편 Cauchy 분포형의 경우 AD 검정 방법이 modified AD 검정 방법보다 약간 높은 기각력을 보였지만 큰 차이가 없으며 표본

크기가 커질수록 차이도 줄어들어 거의 동일해짐을 보였다. 따라서 본 연구에서 제시한 modified AD 검정은 기존의 여러 가지 통계적 적합도 검정 방법보다 더 우수한 기각력을 가지고 있으며 기후변화에 따른 극치사상이 중요해짐에 따라 향후 극치사상에 더 적합한 분포형을 선정하는 기준으로의 역할을 할 수 있을 것으로 판단된다.

## 5. 적용

이상의 결과를 바탕으로 국내 기상청 지점 중 관측기간이 10년 미만의 지점을 제외한 총 50개 지점을 선택하여 modified Anderson-Darling 검정 방법을 실제 우리나라 강우 자료에 적용하였으며 적용 지점은 Table 4에 나타내었다.

각 지점별 강우지속시간 1, 2, 3, 6, 9, 12, 15, 18, 24, 48시간에 대하여 모멘트법(MM), 최우도법(ML), 확률가중모

멘트법(PWM)을 이용하여 매개변수를 추정하였으며 CS, KS, CVM, AD 및 modified AD 검정 방법을 통해 적합도 검정을 수행한 결과를 비교하였다. 전체적인 기각력을 살펴보기 위하여 각 지점의 강우지속시간과 매개변수 추정 방법별로 기각시킨 횟수를 합하여 나타낸 결과 modified AD 검정 63회, CS 검정 52회, AD 검정 50회, KS 및 CVM 검정 각각 0회로, 앞에서 모의한 결과와는 CS 검정과 AD 검정의 순위에 차이가 있지만 modified AD 검정은 변함없이 가장 높은 기각력을 가지고 있음을 보였고, 이를 Table 5에 나타내었다. 이 중에서 modified AD 검정만 유일하게 기각을 시키는 경우가 나타났으며 AD 및 modified AD 검정 방법이 함께 기각력을 보이는 경우도 있었다. 본 논문에서는 그 중에서도 결과가 비교적 잘 표현된 6지점의 자료를 선택하여 Table 6에 나타내었다.

Table 5 (a), (b), (c), (d), (e), (f)는 각각 원주, 울진, 이천, 보령, 금산, 부안의 지속시간 3시간, 9시간, 2시간, 6시

Table 4. Summary of the Site Characteristics

Site Name	Number of Data	Site Name	Number of Data
Sokcho	36	Wan-do	33
Daegwallyeong	30	Jinju	35
Chuncheon	37	Ganghwa	33
Gangneung	46	Yangpyeong	32
Seoul	75	Icheon	33
Incheon	51	Inje	32
Wonju	32	Hongcheon	32
Ulleung-do	53	Jecheon	33
Suwon	40	Boeun	33
Seosan	36	Asan	33
Ulsan	33	Boryeong	32
Chungju	37	Buyeo	33
Daejeon	35	Geumsan	32
Chupungnyeong	49	Buan	31
Andong	30	Imsil	32
Pohang	50	Jeong-eup	32
Gunsan	36	Namwon	32
Daegu	88	Suncheon	32
Jeonju	63	Jangheung	33
Ulsan	57	Haenam	33
Gwangju	65	Goheung	32
Busan	61	Yeongju	32
Tongyeong	36	Mungyeong	32
Mokpo	81	Yeongdeok	32
Yeosu	61	Uiseong	32

Table 5. Number of Rejection for Each Site

Site Name	CS	KS	CVM	AD	mAD	Site Name	CS	KS	CVM	AD	mAD
Sokcho	0	0	0	0	0	Wan-do	0	0	0	0	0
Daegwallyeong	1	0	0	2	4	Jinju	0	0	0	1	1
Chuncheon	0	0	0	0	2	Ganghwa	0	0	0	0	0
Gangneung	0	0	0	0	2	Yangpyeong	0	0	0	0	0
Seoul	0	0	0	0	0	Icheon	0	0	0	0	0
Incheon	0	0	0	0	0	Inje	2	0	0	0	0
Wonju	0	0	0	0	1	Hongcheon	0	0	0	0	0
Ulleung-do	4	0	0	6	6	Jecheon	2	0	0	1	2
Suwon	1	0	0	0	0	Boeun	0	0	0	0	0
Seosan	0	0	0	2	2	Asan	2	0	0	0	0
Uljin	4	0	0	3	3	Boryeong	0	0	0	2	2
Chungju	2	0	0	2	2	Buyeo	0	0	0	0	0
Daejeon	2	0	0	0	0	Geumsan	0	0	0	0	1
Chupungnyeong	0	0	0	0	0	Buan	3	0	0	5	7
Andong	5	0	0	1	1	Imsil	0	0	0	0	0
Pohang	1	0	0	1	1	Jeong-eup	2	0	0	3	5
Gunsan	1	0	0	0	0	Namwon	3	0	0	5	4
Daegu	4	0	0	2	0	Suncheon	0	0	0	0	0
Jeonju	1	0	0	0	0	Jangheung	1	0	0	2	2
Ulsan	0	0	0	0	0	Haenam	0	0	0	0	1
Gwangju	0	0	0	0	0	Goheung	1	0	0	0	1
Busan	2	0	0	3	4	Yeongju	0	0	0	0	1
Tongyeong	2	0	0	2	2	Mungyeong	2	0	0	4	2
Mokpo	2	0	0	0	0	Yeongdeok	0	0	0	0	0
Yeosu	1	0	0	0	0	Uiseong	1	0	0	3	4

Table 6. Results of Goodness-of-Fit Test

(a) Wonju, 3hr

Test	ML		
	COM.	TAB.	CHECK
CS	3.625	5.990	Accept
KS	0.130	0.234	Accept
CVM	0.066	0.461	Accept
AD	0.413	0.513	Accept
mAD	0.264	0.259	<b>Reject</b>

(b) Uljin, 9hr

Test	ML		
	COM.	TAB.	CHECK
CS	2.091	5.99	Accept
KS	0.150	0.231	Accept
CVM	0.107	0.461	Accept
AD	0.644	0.515	<b>Reject</b>
mAD	0.312	0.263	<b>Reject</b>

(c) Icheon, 2hr

Test	ML		
	COM.	TAB.	CHECK
CS	5.000	5.990	Accept
KS	0.112	0.231	Accept
CVM	0.089	0.461	Accept
AD	0.525	0.512	<b>Reject</b>
mAD	0.275	0.256	<b>Reject</b>

(d) Boryeong, 6hr

Test	ML		
	COM.	TAB.	CHECK
CS	2.500	5.990	Accept
KS	0.118	0.234	Accept
CVM	0.107	0.461	Accept
AD	0.664	0.515	<b>Reject</b>
mAD	0.308	0.263	<b>Reject</b>

(e) Geumsan, 2hr

Test	PWM		
	COM.	TAB.	CHECK
CS	4.750	5.990	Accept
KS	0.096	0.234	Accept
CVM	0.073	0.461	Accept
AD	0.484	0.513	Accept
mAD	0.262	0.259	<b>Reject</b>

(f) Buan, 15hr

Test	ML		
	COM.	TAB.	CHECK
CS	3.677	3.840	Accept
KS	0.121	0.238	Accept
CVM	0.071	0.461	Accept
AD	0.466	0.508	Accept
mAD	0.288	0.252	<b>Reject</b>



간, 2시간, 15시간에 해당하는 적합도 검정 방법의 기각 여부를 나타낸 결과로써 (a) 원주, (b) 울진, (c) 이천, (d) 보령, (f) 부안의 경우, 해당되는 지속시간에 따라 매개변수 추정방법 중 최우도법에 의한 결과를 나타내었고, (e) 금산의 경우는 확률가중모멘트법에 의한 결과를 나타내었다. (b) 울진, (c) 이천, (d) 보령의 경우는 AD 검정 방법 및 modified AD 검정 방법이 함께 기각시키는 양상을 보였고, (a) 원주, (e) 금산, (f) 부안의 경우는 5가지 검정 방

법 중 유일하게 modified AD 검정만 기각을 시키는 결과를 나타냈다. 이는 기존의 통계적 적합도 검정 방법에서는 사용 가능하다고 판단했던 분포형을 modified AD 검정 방법이 기각시킴으로서 기존의 방법과는 또 다른 기준을 제시하고 있음을 보였다. 다음과 같은 결과를 바탕으로 도시위치 공식(plotting position)을 이용하여 확률수문량을 도시하면 Fig. 1과 같다. 각각 6개 지점의 관측 자료를 점으로 표시하였으며, 매개변수에 따른 확률강우량은

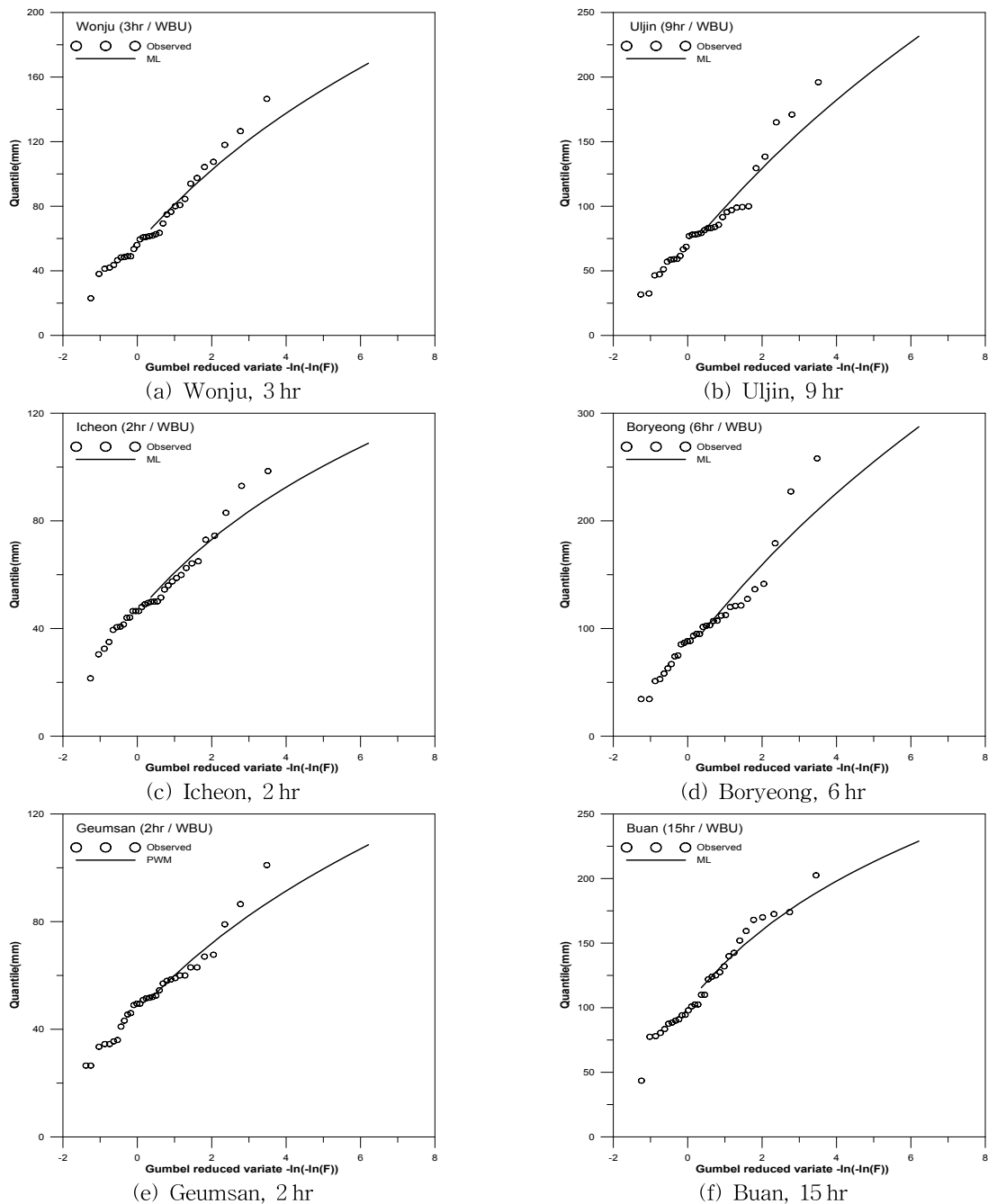


Fig. 1. Probability Distributions Fitted to Observed Rainfall Data for Each Site

실선으로 표시하였다. 대부분의 관측값이 처음에는 매개 변수에 따른 확률강우량을 잘 따라가는 듯 보였으나 극치 값에 가까워질수록 차이가 나타나는 결과를 보였다. 이는 극치값을 표현함에 있어서 modified AD 검정 방법이 더 정확한 적합도 검정으로 높은 기각력을 가지고 있음을 잘 나타내고 있다.

## 6. 결 론

본 연구에서 적합도 검정 방법 중의 하나인 Ahmad et al. (1988)에 의해 개발된 modified AD 검정 방법을 3변수 Weibull 분포에 적용하여 형상매개변수 및 극치값 가중치를 고려한 검정통계량 한계값 산정과 기각력 검토에 대하여 연구한 결과 다음과 같은 결론을 도출하였다.

- 1) Monte Carlo 모의실험을 통하여 표본크기, 유의수준, Weibull 분포의 매개변수 조건과 왜곡도 계수를 고려한 형상매개변수별로 3변수 Weibull 분포에 대한 modified AD 검정통계량 한계값을 산정하였다.
- 2) 유도된 검정통계량값을 이용하여 유의수준별로 표본크기 및 형상매개변수를 포함하는 비선형회귀방정식을 유도하였다. 유도된 비선형회귀방정식의 결정계수를 통하여 높은 정확도를 가지고 있음을 확인하였다.
- 3) 또한 기각력을 검토하기 위해 기존에 일반적으로 쓰이던 통계적 적합도 검정 방법과 modified AD 검정 방법을 8가지 분포형으로부터 난수를 발생시킨 값으로 기각력을 비교하여 modified AD 검정 방법이 기존의 통계적 적합도 검정 방법보다 우수한 기각력을 가지고 있음을 확인하였다.
- 4) 이러한 결과를 바탕으로 실제 우리나라 우량관측지점에 적용한 결과 역시 기존의 적합도 검정 방법보다 modified AD 검정 방법이 더 우수한 기각력을 나타내고 있음을 확인하였다.

이상의 결과를 통하여 기존의 적합도 검정 방법보다 제시된 modified AD 검정 방법이 더 우수한 기각력을 보이므로 향후 Weibull 분포형을 이용한 극치수문량을 산정시 유용하게 사용될 수 있을 것으로 판단된다.

## 감사의 글

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

졌습니다. 이에 감사드립니다.

## REFERENCES

- 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.
- 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 Science*, 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.
- Choulakian, V., and Stephens, M.A. (2001). "Goodness-of-Fit Tests for the Generalized Pareto Distribution." *American Statistical Association and American Society for Quality*, Vol. 43, No. 4, pp. 478-484.
- Evans, J.W., Johnson, R.A., and Green, D.W. (1989). Two-and Three-Parameter Weibull Goodness-of-Fit Tests. United States Department of Agriculture, Forest Products Laboratory, Research Paper FPL-RP-493.
- Heo, J.H., Salas, J.D., and Kim, K.D. (2001). "Estimation of confidence intervals of quantiles for the Weibull distribution." *Stochastic Environmental Research and Risk Assessment*, Vol. 15, pp. 284-309.
- Johnson, N.L., and Kotz, S. (1970). Continuous Univariate Distributions-1. John Wiley & Sons.
- Laio, F. (2004). "Cramer-von Mises and Anderson-Darling goodness of fit tests for extreme value distributions with unknown parameters." *Water Resources Research*, Vol. 40, No. 9, W09308.
- Lewis, P.A.W. (1961). "Distribution of the Anderson-Darling Statistic." *Institute of Mathematical Statistics*, Vol. 32, No. 4, pp. 1118-1124.
- Littell, R.C., McClave, J.T., and Offen, W.W. (1979).

- “Goodness-of-Fit Tests for the two parameter Weibull distribution.” *Commun. Stat. Simul. Comp.*, Vol. 8, No. 3, pp. 257-269.
- Luceno, A. (2006). “Fitting the generalized Pareto distribution to data using maximum goodness-of-fit estimators.” *Computational Statistics & Data Analysis*, Vol. 51, pp. 904-917.
- Pakyari, R., and Balakrishnan, N. (2012). “A General Purpose Approximate Goodness-of-Fit Test for Progressively Type-II Censored Data.” *IEEE Trans. Reliability*, Vol. 61, No. 1, pp. 238-244.
- Porter III, J.E., Coleman, J.W., and Moore, A.H. (1992). “Modified KS, AD, and C-vM Tests for the Pareto Distribution with Unknown Location & Scale Parameters.” *IEEE Trans. Reliability*, Vol. 41, pp. 112- 117.
- Shin, H.J., Sung, K.M., and Heo, J.H. (2010). “Derivation of Modified Anderson-Darling Test Statistics and Power Test for the Gumbel Distribution.” *Journal of Korea Water Resources Association*, Vol. 43, No. 9, pp. 813-822.
- 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. (1974). “EDF statistics for goodness of fit and some comparisons.” *J. Amer. Stat. Ass.*, Vol. 69, pp. 730-737.
- Stephens, M.A. (1977). “Goodness of fit for the extreme value distribution.” *Biometrika*, Vol. 64, pp. 583-588.
- Tolikas, K., and Heravi, S. (2008). “The Anderson-Darling Goodness-of-Fit Test Statistic for the Three-Parameter Lognormal Distribution.” *Comm. Statist.-Theory Meth*, Vol. 37, pp. 3135-3143.

논문번호: 13-024	접수: 2013.02.26
수정일자: 2013.03.22	심사완료: 2013.03.22