

낙동강 유역의 일 유출량에 대한 적정 확률분포형 선정

백선욱 · 탁우현* · 어 규* · 윤정환* · 김형수***

인하대학교 스마트시티공학과

*인하대학교 토목공학과

**인하대학교 사회인프라공학과

Identification of appropriate probability distribution for daily streamflow in the Nakdong river basin

Seonuk Baek · Woohyun Tak* · Gyu Eo* · Junghwan Yun* · Hung Soo Kim**

Department of Smart City Engineering, Inha University, Korea

**Department of Civil Engineering, Inha University, Korea*

(Received : 21 March 2025, Revised : 06 April 2025, Accepted : 06 April 2025)

요약

최근 기후변화로 인해 홍수 및 가뭄 피해가 증가하면서 극한사상에 대한 수문설계의 중요성이 더욱 강조되고 있다. 수문설계에 있어 중요한 설계홍수량은 수문자료(강우 및 유출량)의 빈도분석을 통해 산정된다. 빈도분석을 수행하기 위해서는 먼저 해당 수문자료의 특성을 명확하게 파악하여야 하며, 이를 통해 적정 확률분포모형을 선정해야 한다. 이에 본 연구에서는 낙동강 유역의 8개 유량 관측소를 대상으로 최근 20년간의 일 단위 유출량 자료를 수집하였으며, 연 최대, 최소, 평균유량에 대한 기본통계량을 분석하였다. 또한, 6가지 확률분포모형에 최우도법을 적용하여 매개변수를 추정하였으며, 적합도 검정과 정보기준모형을 활용하여 적정 확률분포형을 선정하였다. 분석 결과, 낙동강 유역의 일 유출량 자료에 대한 적정 확률분포형으로 Lognormal, Normal, Weibull, Gamma, Gumbel 분포가 선정되었다. 본 연구의 결과는 지속적인 하천관리와 자연재해 저감 방안을 수립하는 과정에서 설계수문량 산정을 위한 기초자료로 활용할 수 있을 것으로 판단된다.

핵심용어 : 설계수문량, 빈도분석, 적합도 검정, 정보기준모형, 최우도법

Abstract

Recently, the impact of floods and droughts has intensified due to climate change, highlighting the growing importance of hydrological design for extreme events. In hydrological design, design flood is estimated through frequency analysis of hydrological data, including precipitation and streamflow. To conduct frequency analysis, it is essential to first clearly identify the characteristics of the hydrological data and then select an appropriate probability distribution model. In this study, daily streamflow data from the past 20 years were collected from eight hydrological stations in the Nakdong river basin. Basic statistical analyses were performed on the annual maximum, minimum, and mean streamflow values. Additionally, parameters for six probability distribution models were estimated using the maximum likelihood method, and the most suitable probability distribution models were identified based on goodness-of-fit tests and information criteria. As a result, Lognormal, Normal, Weibull, Gamma, and Gumbel distributions were determined to be the most appropriate probability distribution models for the Nakdong River basin. The findings of this study are expected to serve as fundamental data to estimate hydrologic design quantity for continuous river management and the development of strategies to mitigate natural disasters.

Key words : Design hydrologic quantity, Frequency analysis, Goodness-of-fit test, Information criteria model, Maximum Likelihood Method

*To whom correspondence should be addressed.

Department of Civil Engineering, Inha university, Korea

E-mail : sookim@inha.ac.kr

• Seonuk Baek Department of Smart City Engineering, Inha University, Korea / Ph.D. Candidate (uk9509@nate.com)

• Woohyun Tak Department of Civil Engineering, Inha University, Korea / Master (wjndfsawa@naver.com)

• Gyu Eo Department of Civil Engineering, Inha University, Korea / Ph.D. Candidate (good3437@hotmail.com)

• Junghwan Yun Department of Civil Engineering, Inha University, Korea / Ph.D. Candidate (lenablue12@gmail.com)

• Hung Soo Kim Department of Civil Engineering, Inha University, Korea / Professor (sookim@inha.ac.kr)



This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. 서론

수문자료(강우량, 유출량 등)는 물과 관련된 분석에서 가장 기초적인 자료이며, 수자원 관리와 자연재해 대응을 위한 필수적인 요소이다(Walker, 2000; Choe and Kim, 2008; Stewart, 2015; Shin and Lee, 2020). 특히, 홍수와 같은 자연재난을 효과적으로 예방 및 대응하기 위해서는 수문자료의 빈도분석이 선행되어야 한다(Kim et al., 2022). 또한, 수문 시계열의 변화를 고려하여 가뭄과 같은 장기적인 물부족 문제에 대한 적절한 대책이 마련되어야 한다(Kim et al., 2000). 결과적으로, 하천의 안정적인 치수 및 이수 관리를 위해서는 설계홍수량과 수문자료의 시·공간적 변화 특성을 분석해야 하며, 이를 위해 기초통계량 분석을 통한 적정 확률분포형 선정이 반드시 수행되어야 한다.

수문자료 확률분포형에 대한 국내 선행 연구사례를 살펴 보면, Kim et al.(2013)은 3변수 Weibull 분포에 modified Anderson-Darling 검정 방법을 적용하여 적합도 검정의 성능을 평가하였다. Monte Carlo 모의실험을 통해 표본 크기, 형상 매개변수, 유의수준에 따른 검정통계량 한계값을 산정하였으며, 기존 적합도 검정 방법과의 기각역을 비교 및 분석하였다. 이를 통해 modified AD 검정 방법이 극치값을 더욱 효과적으로 반영한다는 것을 확인하였다. Kim and Lee (2015)은 비정상성을 고려한 홍수 빈도분석을 수행하고자 Generalized Extreme Value(GEV) 분포를 활용하여 정상성 및 비정상성 모형을 구축하였다. 비정상성 모형에서는 위치 모수(location parameter), 규모 모수(scale parameter), 두 모수 모두가 시간에 따라 선형적으로 변화하는 세 가지 경우를 고려하였다. 또한, 상대적 우도비 검정(likelihood ratio test)과 Akaike Information Criterion(AIC)을 사용하여 적합한 모형을 선정하였다.

Kim et al.(2017)은 이변량 가뭄빈도해석을 위해 Bayesian Copula 모형을 개발하여 매개변수의 불확실성을 정량적으로 평가하였다. Monte Carlo 모의실험을 통해 Bayesian 기법이 기존 Copula 모형의 매개변수 추정 과정에서 효과적으로 작용함을 확인하였으며, 이를 한강유역의 2013~2015년 가뭄 사상에 적용하여 결합재현기간을 분석하였다. 대부분의 관측소에서 결합재현기간이 100년을 초과하는 것으로 나타났으며, 일부 지역에서는 1,000년 이상의 빈도를 기록하는 극한가뭄으로 평가되었다. 해당 연구 결과를 통해 Bayesian Copula 모형이 가뭄빈도해석의 불확실성을 정량화하는 데 유용하다는 것을 확인하였다. Joo et al.(2019)는 수문자료 빈도해석에서 이변량 확률분포형의 매개변수 추정방법을 개선하기 위해 Copula 모형을 활용하였다. 기존 의사최우도법(Maximum Pseudo-Likelihood Method, MPL)의 한계를 보완하기 위해 수정된 Modified Maximum Pseudo-Likelihood Method(MMPL)을 제안하였으며, Weibull 도시위치공식 대신 Cunnane 도시위치공식을 적용하여 확률변수의 왜곡도를 고려하였다. 또한, Monte Carlo 시뮬레이션을 통해 기존 방법과의 성능을 비교하였으며, 매개변수 추정의 정확도를 평가

하였다. 결과적으로, MMPL 방법이 기존 의사최우도법보다 매개변수 추정에 있어 정확도가 높다는 것을 확인하였으며, 극치 수문자료 분석에서 더욱 신뢰성 있는 결과가 도출될 수 있음을 확인하였다. Kwak et al.(2016a, 2016b)은 나이트 자료에 의해 생성한 유출량 자료에 대해 이변량 가뭄분석을 수행하였으며, 기후변화를 고려한 기상학적 가뭄을 평가하였다. 또한 Yoo et al.(2016)은 이변량 가뭄 빈도곡선에 대한 신뢰 구간 분석을 수행하였다.

수문자료 확률분포형에 대한 국외 선행 연구사례를 살펴 보면, Olofinloye et al.(2009)은 기존 확률분포형 비교 방법의 한계를 보완하고자 다양한 적합도 검정을 수행하고 여러 확률분포형을 비교하였다. 그 결과, Log-Pearson Type III 분포가 50%, Pearson Type III 분포가 40%, Log-Gumbel 분포가 10%에서 가장 적합한 것으로 나타났다. 이를 통해 극치값(강우량) 예측을 위한 확률분포형 선정 시, Log-Pearson Type III 분포가 가장 신뢰할 수 있음을 확인하였다. Khosravi et al.(2012)은 연평균 및 최대유량의 확률분포를 결정하고자 다양한 확률분포형을 비교 및 분석하였다. 기존 확률분포 평가 방법의 한계를 보완하기 위해 Residual Sum of Squares(RSS)을 활용한 평가 방법을 적용하였으며, SMADA 소프트웨어를 활용하여 확률분포의 적합성을 분석하였다. 또한, Log Pearson Type III, Pearson Type III, Log-Normal 등의 확률분포를 비교하였으며, 관측값과 예측값 간의 적합도를 평가하였다. 이를 통해 Log Pearson Type III 분포가 연평균 및 최대유량을 추정하는 데 있어 가장 적합하다는 것을 확인하였다. Gheidari(2013)는 지역 홍수 빈도분석에 있어 최적의 확률분포를 선정하고자 L-모멘트 및 LH-모멘트 방법을 비교하였다. 기존 L-모멘트 방법의 한계를 보완하기 위해 LH-모멘트(LH-moments) 개념을 적용하였으며, GEV, Generalized Logistic(GLO), Generalized Pareto(GPA)를 활용하여 홍수 자료의 극치값을 분석하였다. 또한, RRMSE 및 L-kurtosis difference 기준을 통해 확률분포의 적합성을 평가하였다. 결과적으로, GEV 분포가 극치값을 추정하는 데 가장 적합한 것으로 나타났으며, GLO 분포는 전체 데이터 시리즈에서 우수한 성능이 나타났음을 확인하였다. 해당 선행 연구사례 외에도 가뭄 및 홍수분석을 수행하는 과정에서 최적의 확률분포형을 활용한 연구가 수행되었다(Kim et al., 2006; Kyoung et al., 2007; Kim et al., 2012).

이처럼 확률분포형에 대한 연구가 활발히 진행되었으며, 대부분의 선행연구에서 하천의 치수 및 이수의 관리를 중심으로 적정 확률분포형을 선정하였다. 한편, 기후변화로 인해 수해(홍수 및 침수)의 피해가 급증하는 현재 본류뿐만 아니라 소하천과 지류에서도 수해가 더욱 빈번하게 발생하고 있으며, 이에 따라 하천정비 지역이 본류에서 소하천과 지류로 확대되고 있다. 그럼에도 불구하고, 지류 및 소하천을 대상으로 최적의 확률분포형을 선정한 연구는 미흡한 실정이다. 따라서 본 연구에서는 지류 및 소하천을 대상으로 기초통계량을 활용하여 유출량의 특성을 파악하였다. 또한, 적정 확률분포형을 선정하고자 연도별 수문시계열의 유출 특성을 분석

하였으며, 이를 통해 일 최대유량, 최소유량, 평균유량에 적합한 분포모형을 파악하였다. 이후, 최우도법(maximum likelihood estimation, MLE)을 활용하여 매개변수를 추정하였으며, K-S, CVM, AD 검정을 통해 적합도를 평가를 수행하였다. 또한, AIC와 Bayesian Information Criterion(BIC)를 활용하여 모형을 비교·분석하였다.

2. 연구방법

2.1 확률분포형 및 매개변수 추정

최적의 확률분포형을 선정하려면 사전에 수문자료를 기반으로 각 모형의 매개변수를 추정해야 한다(Kim et al., 2017). 매개변수 추정방법에는 모멘트법(Moment of Method, MOM), 최우도법(Maximum Likelihood Estimation, MLE), 확률가중 모멘트법(Probability-Weighted Moments, PWM), L-모멘트법(L-moment Method) 등이 있다. 이 중 모멘트법은 이상치에 민감하고 변동성이 커 실용성이 떨어지는 경우가 많다. 반면, 최우도법과 확률가중모멘트법은 모멘트법보다 이상치에 덜 민감하여 매개변수 추정 방법으로 널리 활용되고 있다. 또한, 확률가중모멘트법은 자료의 가장 큰 값에 많은 가중치를 부여한다는 특징으로 인해 재현기간이 클수록 모멘트법이나 최우도법보다 지나치게 확률강우량이 높게 산정될 수 있다. 따라서 본 연구에서는 일 단위 유출량에 대한 매개변수를 추정하고자 확률분포모형에 최우도법을 적용하였다.

2.1.1 확률분포형

수문자료 분석에 있어 확률분포형은 이산형 확률분포와

연속형 확률분포로 구성되어 있다(Kim et al., 2014; Joo et al., 2016). 이산형 확률분포는 시간 또는 공간 내에서 개별적인 사상의 발생 횟수를 설명하는데 사용되며, 강우 발생 횟수, 홍수 발생 횟수 등의 분석에 활용된다. 연속형 확률분포는 유량 및 강우자료, 연 최대값, 연 최소값 등 연속자료 분석에 활용된다. 따라서 본 연구에서는 연속자료(유출량)에 대한 적정 확률분포형을 선정하고자 수문자료 분석에서 널리 사용되는 연속형 확률분포형을 활용하였다(Table 1).

2.1.2 최우도법

최우도법(MLE)이란 관측자료 또는 표본자료로부터 발생 가능성이 가장 높은 우도(likelihood) 또는 결합 확률(joint probability)이 가장 큰 매개변수를 추정하는 방법이다(Park and Yoo, 2012). 일반적으로 계산의 편리성을 위해 우도함수(likelihood function)보다는 대수 우도함수(log-likelihood function)를 주로 사용하며, 매개변수 θ_i 에 대해 미분하여 0이 되는 지점을 찾아 최적의 매개변수를 추정한다(Ryu., 2010).

$$\frac{\partial \ln L(\theta_i)}{\partial \theta_i} = 0 \quad (1)$$

여기서, θ_i 는 확률분포형의 매개변수를, $\ln L(\theta_i)$ 는 대수 우도함수를 의미한다.

2.2 확률분포모형의 적합도 검정 방법

적합도 검정이란 자료의 계급 구간별 빈도를 계산하여 상대대수분포도를 작성한 후, 이를 이론적 분포의 확률밀도

Table 1. General description of the probability distribution function and its range

Distribution model	Probability Distribution Function (PDF)	Range
Normal	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$	$-\infty < x < \infty$
Lognormal	$f(x) = \frac{1}{\sigma_y \sqrt{2\pi}} \frac{1}{x} \exp\left[-\frac{1}{2} \left(\frac{\ln x - \mu_y}{\sigma_y}\right)^2\right]$	$0 \leq x < \infty$
Gamma	$f(x) = \frac{1}{\alpha \Gamma(\beta)} \left(\frac{x}{\alpha}\right)^{\beta-1} e^{-\frac{x}{\alpha}}$	$0 \leq x < \infty$
GEV	$f(x) = \frac{1}{\alpha} \left[1 - \beta \left(\frac{x-x_0}{\alpha}\right)\right]^{\frac{1}{\beta}-1} \exp\left[-\left[1 - \beta \left(\frac{x-x_0}{\alpha}\right)\right]^{\frac{1}{\beta}}\right]$	$\beta < 0, \beta > 0, \beta = 0$
Gumbel	$f(x) = \frac{1}{\alpha} \exp\left[-\frac{x-x_0}{\alpha} - \exp\left(-\frac{x-x_0}{\alpha}\right)\right]$	$-\infty < x < \infty$
Weibull	$f(x) = \frac{\beta}{\alpha} \left(\frac{x}{\alpha}\right)^{\beta-1} \exp\left[-\left(\frac{x}{\alpha}\right)^\beta\right]$	$x \geq 0$
Exponential	$f(x) = \frac{1}{\alpha} \exp\left[-\left(\frac{x-x_0}{\alpha}\right)\right]$	$x_0 \leq x < \infty$
Beta	$f(y) = \frac{1}{B(\alpha, \beta)} \frac{(y-a)^{\alpha-1} (b-y)^{\beta-1}}{(b-a)^{\alpha+\beta-1}}$	$a \leq y \leq b$
Uniform	$F(x) = 1 - \exp\left\{-\frac{x}{\alpha}\right\}$	-
Logistic	$f(x; \mu, \sigma) = \frac{\pi}{\sigma\sqrt{3}} \frac{e^{-\pi(x-\mu)/\sigma\sqrt{3}}}{\{1 + e^{-\pi(x-\mu)/\sigma\sqrt{3}}\}^2}$	$-\infty < x < \infty,$ $-\infty < \mu < \infty, \sigma > 0$

함수(probability density function, PDF)와 정량적으로 비교하는 것을 의미한다(You et al., 2017). 적합성 검정은 크게 절대적 검정과 상대적 선택 방법으로 나눌 수 있다. 절대적 검정 방법은 개별 확률분포의 적합도를 직접 평가하는 방식으로, 대표적으로 Chi-square 검정, Kolmogorov-Smirnov (K-S) 검정, Anderson-Darling(AD) 검정, Cramér-von Mises(CVM) 검정 등이 있다. 반면, 상대적 선택 방법은 여러 확률분포 중 최적의 분포를 비교 및 선택하는 방식으로, AIC, BIC, Bayesian 방법이 존재한다. 본 연구에서는 다양한 적합도 검정을 활용하여 유출량 데이터에 적합한 확률분포형을 평가하였다. 이를 위해 Chi-square 검정, Anderson-Darling 검정, Kolmogorov-Smirnov 검정, Cramér-von Mises 검정을 수행하여 개별 분포의 적합성을 분석하였다. 또한, AIC 및 BIC를 적용하여 상대적인 모델 선택을 수행함으로써 최적의 확률분포형을 결정하였다.

2.2.1 Chi-square test

카이제곱 검정은 분석 목적에 따라 관측된 자료가 특정 확률분포와 얼마나 잘 일치하는지를 검정하는 통계적 방법이다(Im et al., 2020). 크기가 n 인 표본을 k 개의 구간으로 분류할 때, 각 구간의 관측도수, $f(i) = n_i$, $i = 1, 2, \dots, k$ 또는 상대도수 $f_s(x_i) = n_i/n$ 이라 한다. 각 구간에 대응되는 모집단 또는 이론적인 비율(상대도수)을 $p(x_i)$ 라 하면 귀무가설은 각 $p(x_i)$ 들에 어떤 이론에서 제시하는 값, $F(x_i) - F(x_{i-1})$ 를 따르는 형태를 가진다. 즉, 귀무가설은 다음 Eq. (2)와 같이 표현할 수 있다.

$$H_0 : p(x_i) = F(x_i) - F(x_{i-1}) \quad (2)$$

귀무가설 H_0 하에서 구간 i 의 기대도수(expected frequency)는 표본의 크기 n 에 무가설에서 제시한 구간 i 의 비율을 곱하여 $np(x_i)$ 가 된다. 이와 같이 관측도수와 기대도수의 차이를 측정하는 도구로 Karl Pearson이 제안한 검정통계량(χ_c^2)은 Eq. (3)과 같이 나타낼 수 있다.

$$\chi_c^2 = \sum_{i=1}^k \frac{n[f_s(x_i) - p(x_i)]^2}{p(x_i)} = \sum_{i=1}^k \frac{(O - E)^2}{E} \quad (3)$$

여기서, $nf_s(x_i) = n_i$ 는 관측도수(또는 O)이며, 이에 대응하는 기대도수는 $np(x_i)$ (또는 E)이다.

2.2.2 Anderson-Darling test

AD 검정방법은 주어진 데이터가 특정 확률분포를 따르는지 확인하는 데 사용되는 적합도 검정방법이다(Shin et al., 2010). 해당 방법은 K-S 검정보다 분포의 꼬리를 더 강조하므로 이상값을 감지하는데 유용하다. 해당 방법에 대한 공식은 다음 Eq. (4)와 같다.

$$Q_n = n \int_{-\infty}^{\infty} [F_n(x) - F(x)]^2 \psi(x) dF(x) \quad (4)$$

여기서, 가중치 함수($\psi(x)$)가 1일 경우, 검정통계량은 CVM 검정의 검정통계량과 같다. 또한 $\psi(x) = [F(x)(1 - F(x))]^{-1}$ 인 경우를 AD 검정통계량이라고 하며, 이는 Eq. (5)와 같이 A_n^2 로 표시된다. 해당 A_n^2 는 홍수 또는 가뭄의 극한 사상의 빈도해석에서 좌우측 꼬리 부분에 가중치를 부여하여 표본의 중간값보다 극치값에 대한 적합도를 더욱 강조하는 통계량 값이다.

$$A_n^2 = n \int_{-\infty}^{\infty} \frac{[F_n(x) - F(x)]^2}{F(x)\{1 - F(x)\}} dF(x) \quad (5)$$

일반적으로 수치적인 계산의 편의를 위해 Eq. (5) 대신 Eq. (6)을 사용하며, AD 통계량이 95% 유의수준($\alpha = 0.05$)에서 임계값 2.5018보다 크면 검정 가설이 기각된다.

$$A_n^2 = -N - \frac{1}{N} \sum_{i=1}^N (2i-1) \times [\ln F_c(Q_i) + (\ln(1 - F_D(Q_i)))] \quad (6)$$

여기서, A_n^2 는 AD 검정통계량이며, $F(x)$ 는 주어진 확률분포모형의 누적분포함수로 추정된 매개변수를 포함한 이론적 분포를, $F_n(x)$ 는 표본 자료를 기반으로 계산된 경험적 누적분포함로 관측값의 누적 비율을 나타낸다.

2.2.3 Kolmogorov-Smirnov Test

Kolmogorov-Smirnov(1933)가 처음으로 K-S 검정의 원리를 제안하였으며, 이후 Smirnov(1948)가 검정 절차를 개발하였다. 해당 방법의 기본적인 절차는 표본자료의 누가 확률분포와 가정된 이론 확률분포의 누가확률분포를 비교하는 것으로, 양자의 최대편차가 표본의 크기와 유의수준에 따라 결정되는 한계편차보다 크면 분포는 기각된다. 표본자료의 누가확률분포를 산정하기 위해서는 우선 n 개의 자료를 오름차순으로 정렬한 후 Eq. (7)을 활용하여 각 자료값의 누가 확률을 계산해야 한다.

$$S_n(x) = \frac{k}{n}, x_k \leq x < x_{k+1} \quad (7)$$

여기서, $S_n(x)$ 는 오름차순으로 정렬했을 때의 k 번째 자료값의 누가 발생 확률이며, n 은 자료의 총수이다.

2.2.4 Cramér Von Mises test

CVM 검정방법은 K-S 검정방법과 마찬가지로 누가확률 분포함수를 이용하여 검정하는 방법이다(Hong and Kim, 2009). 해당 검정방법은 표본자료계열($X_1, X_2, \dots, X_n$)이 누가확률분포함수($F(X; \theta)$)로 정의되는 확률분포 모형을 모집단으로 한다는 가정을 검정한다. 이를 위해 다음과 같은 검정통계량(W)을 산정하여 검정을 수행한다. 또한, AD 검정

방법, K-S 검정방법과 달리 CVM 검정은 관찰된 수문 시계열을 오름차순으로 고려한다. 검정통계량(W)가 95% 유의수준($\alpha=0.05$)에서 임계값 0.221보다 크면 검정 가설이 기각된다.

$$W = \frac{1}{12N} + \sum_{i=1}^N \left[F(X_i; \hat{\theta}) - \frac{2i-1}{2n} \right]^2 \quad (8)$$

여기서, n 은 표본자료의 수, $\hat{\theta}$ 는 표본자료로부터 추정한 대상 확률분포의 매개변수이며, $F(X_i; \hat{\theta})$ 는 표본자료를 오름차순으로 정렬했을 때 X_i 위치에서의 누적확률이다. Eq. (8)로 구한 검정통계치(W)가 유의수준 α 에 대한 검정통계치($W_{1-\alpha}(n)$)보다 작으면 표본자료는 검정하고자 하는 확률분포모형을 따르는 것으로 판단된다. 여기서, CVM 검정의 검정한계치 $W_{1-\alpha}(n)$ 은 유의수준(α)와 표본자료의 수 n 의 함수로 정의되며, 표본의 크기가 크면 α 의 함수로 수렴한다.

2.2.5 Akaike Information and Bayesian Information criteria

AIC(Akaike Information Criteria)은 가장 최소의 정보 손실을 갖는 모형이 가장 데이터와 가장 적합한 모형으로 선택된다(Jeong et al., 2022). 해당 방법은 첫 번째 항에서 최우도법을 사용하고, 두 번째 항에서 파라미터 개수에 대한 보정을 수행함으로써 자유도가 높은 모형의 AIC 값에 페널티(penalty)를 부여하여 과적합을 방지할 수 있다. 그러나 AIC는 데이터 개수에 대한 보정이 없어 표본 크기가 작은 경우 정확도가 낮아질 수 있다.

$$AIC = -2\ln(L) + 2k \quad (9)$$

BIC(Bayesian Information Criteria)이란 Bayesian 이론에서 우도함수와 사전분포(prior probability distribution)를 활용하여 산정된 사후분포(posterior probability distribution)를 근사화하여 유도된 통계량이다(Kim et al., 2016). 앞의 방법과 유사하게 최소의 BIC 값을 갖는 모형이 가장 적합한 모형을 선정한다.

$$BIC = -2\ln(L) + k\ln(n) \quad (10)$$

2.3 정규성 검정

Quantile-Quantile Plot(Q-Q Plot)이나 Probability-Probability Plot(P-P Plot)을 활용하여 정규성을 검정할 경우, 데이터가 정규분포를 따르는지를 판단할 때 연구자의 주관성이 반영될 수 있다. 그러나 편차를 수치적으로 표현하여 정규성 검정의 척도로 활용한다면 주관적인 판단을 배제할 수 있다(Lee and Park, 2020).

2.3.1 Probability-Probability Plot

P-P Plot은 확률표본 $X_1, X_2, \dots, X_n$ 이라 할 때 확률표본이 정규성일 경우에는 확률분포함수는 다음 Eq. (11)과 같다.

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}[(x-\mu)/\sigma]^2} \quad (11)$$

여기서, 확률표본의 순서통계량을 $X_{1:n}, X_{2:n}, \dots, X_{n:n}$ 이라 하며, 이를 표준화된 확률변수 X_i 가 정규분포를 따를 경우, 확률표본의 누적분포함수로 정의된다. 또한, $z_i = \Phi^{-1}\left(\frac{i-0.375}{i+0.25}\right)$ 로 표현할 때, P-P plot은 (x, y) 좌표 평면에서 (z_i, x_i) 를 나타낸다. 따라서, 확률표본이 정규분포를 따른다면 P-P plot에서 기대되는 직선은 $y = \mu + \sigma_x$ 이며, 데이터가 이론적으로 기대되는 직선과 얼마나 일치하는지를 확인함으로써 정규성 여부를 판단할 수 있다.

2.3.2 Quantile-Quantile Plot

Q-Q Plot은 확률표본 $X_1, X_2, \dots, X_n$ 가 주어졌을 때, 해당 확률변수 X 의 누적분포함수를 $\Phi(x)$ 라고 하고, 그 역함수를 $\Phi^{-1}(\cdot)$ 로 정의한다. Q-Q plot은 (x, y) 좌표 평면에서 분포 Φ 의 (i/n) 분위수와 표본의 (i/n) 분위수를 대응시킨 그래프이다. 이때, 분포 Φ 의 (i/n) 분위수는 $\Phi^{-1}\left(\frac{i}{n}\right)$ 로 표현되며, 표본의 (i/n) 분위수는 $X_{i:n}$ 가 된다. 따라서 Q-Q plot은 좌표 평면에서 $\left(\Phi^{-1}\left(\frac{i}{n}\right), X_{i:n}\right)$ 을 나타낸다. 또한, Uniform 분포를 따르는 확률변수 X 의 누적분포함수(CDF)를 $U_i = \Phi(X_i)$ 로 두고, $U_1, U_2, \dots, U_n$ 의 순서 통계량을 $U_{1:n}, U_{2:n}, \dots, U_{n:n}$ 이라 정의한다. 이를 고려하면 $U_{i:n} = \Phi(X_{i:n})$ 이며, 이는 $BETA(i, n-i+1)$ 분포를 따른다. 따라서, $U_{i:n}$ 의 확률밀도함수(PDF) $b_i(u_i, n)$ 는 다음과 같으며, 분포의 파라미터는 각각 $\alpha = i, \beta = n-i+1$ 로 정의된다. $BETA$ 분포의 확률밀도함수(PDF) 공식은 Eq. (12)와 같다.

$$b_i(u_{i:n}) = \frac{\Gamma(n+1)}{\Gamma(i)\Gamma(n-i+1)} (u_{i:n})^{i-1} (1-u_{i:n})^{n-i} \quad (12)$$

3. 적용 및 결과

3.1 대상지역 선정 및 자료 수집

낙동강 유역은 한반도 동남부에 위치하는 지역으로, 하절기 남쪽에서 북상하는 남태평양 고기압과 동쪽에서 북상하는 북태평양 고기압으로 인해 계절성 집중호우와 홍수가 빈번하게 발생한다. 이와 같이 낙동강 유역은 여름철 국지적이고 불규칙한 강우 패턴과 함께 불안정한 유출 특성이 자주 발생하며, 강우-유출 반응이 매우 민감하여 유출량을 산정할 때 확률론적 접근이 요구된다. 또한, 해당 유역은 도시화로 인한 불투수면 증가와 함께 토양 침투율이 낮아 단시간 내 급격한 유출량이 발생한다. 특히, 낙동강 지류 및 소하천 지역은

하천 경사와 유역 내 하천 밀도가 높아 도달시간이 짧고, 침식과 유사의 영향도 크게 작용하여 수문 반응이 매우 빠르게 나타난다. 이러한 지형적, 기상학적, 수문학적 특성을 종합적으로 고려하여 본 연구에서는 낙동강 유역의 지류와 소하천을 대상으로 일 유출량에 대한 최적의 확률분포형을 선정하였다. 이를 위해 낙동강 유역의 유출 특성을 분석하고자 국가수자원관리종합정보시스템(WAMIS)의 8개 유량관측소에서 제공하는 최근 20년(2000~2019년)의 일 단위 유출량 자료를 수집하였다. 이때, 유출량 자료는 수위 보정이 가능하고, 수위-유량 관계곡선식이 제공되어 정확도가 높은 유량을 산정할 수 있는 관측소의 데이터를 사용하였으며, 결측으로 인해 유량 보정 및 환산이 불가능한 관측소는 분석에서 제외하였다. 본

연구에서는 8개 관측소의 유출량에 대해 모두 분석을 수행하였으며, 04번의 안동시(목계교) 관측소에 대한 유출량을 예시로 분석한 내용을 설명하였다.

3.2 수문시계열 특성 분석 및 기초통계량 산정

3.2.1 일 단위 유출량의 특성 분석

2000년부터 2019년까지의 일 단위 유출량 자료를 보유하고 있는 8개 유량관측소로부터 연 최대, 최소, 평균유량에 대한 평균값을 산정하여 Table 2에 나타내었다. 해당 유량 값을 살펴보면, 낙동강 유역의 최대, 최소, 평균 유량의 평균값은 각각 약 $28,961\text{m}^3/\text{sec}$, $135\text{m}^3/\text{sec}$, $1,454\text{m}^3/\text{sec}$ 으로

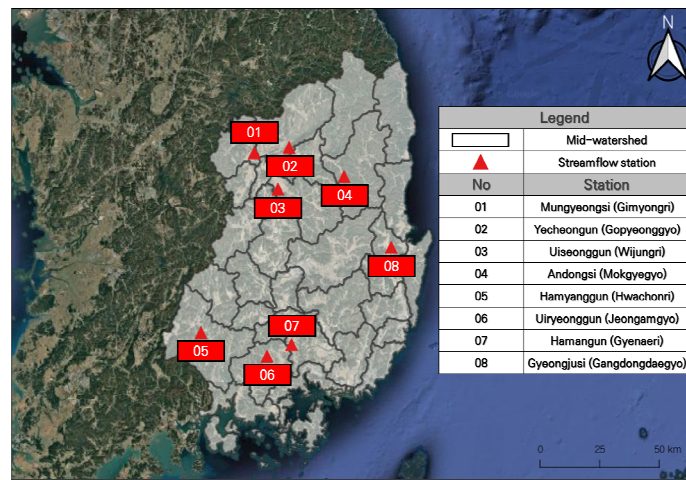


Fig. 1. Nakdong river basin and locations of 8 streamflow stations

Table 2. Annual maximum, minimum, and mean streamflows (m^3/s)

Year	Maximum streamflow (m^3/s)	Minimum streamflow (m^3/s)	Mean streamflow (m^3/s)
2000	1985.64	7.56	104.38
2001	1146.04	5.08	50.24
2002	2414.67	6.52	117.11
2003	2237.48	15.27	178.48
2004	1715.19	11.73	108.82
2005	965.90	8.05	59.58
2006	2307.37	7.06	87.55
2007	1670.05	9.83	86.26
2008	842.79	6.99	38.14
2009	1035.42	4.96	41.04
2010	1094.43	6.42	68.57
2011	2330.83	2.46	89.01
2012	2401.64	1.40	78.08
2013	980.75	11.11	52.74
2014	1066.82	4.69	55.85
2015	425.87	10.07	34.46
2016	960.71	1.77	55.76
2017	336.98	2.82	31.55
2018	1263.46	6.75	64.19
2019	1788.98	4.06	51.10
Average	1448.55	6.73	72.65

산정되었다. 해당 평균값을 비교한 결과, 연 최대유량은 2002년이 가장 높은 것으로, 연 최소유량은 2012년이 가장 낮은 것으로 나타났다. 또한 평균유량은 2015년에 상대적으로 낮은 유출량 값이 발생하였다. 이를 통해 대상지역의 유출량은 하상계수가 매우 크고, 유량의 변동이 불안정하다는 것을 파악하였다.

3.2.2 기초통계량 분석

수문자료(유출량)에 대한 분포 특성, 대칭성, 정규분포와의 차이를 파악하고자 기초통계량(최대, 최소, 중간값, 평균 등)을 산정하였다. 기초통계량 결과를 통해 해당 수문자료의 집중도, 평균값에 대한 분포 특성, 대칭성, 정규분포와의 비교 등을 파악하였다. Table 3에 제시된 바와 같이 최대, 최소, 평균 유량의 왜도(skewness) 값이 0보다 크기 때문에 오른쪽으로 왜곡된 분포(양의 왜도)가 형성될 것으로 판단하였다. 또한, 평균유량의 첨도(kurtosis) 값이 3보다 커 정규분포보다 뾰족한 형태로 나타났으며, 이를 통해 극치값이 존재할 가능성이 높을 것으로 판단하였다. 반면, 최대 및 최소유량의 첨도 값은 3보다 작아 정규분포보다 완만한 형태를 보일 것으로 판단하였다. 이러한 결과를 종합한다면, 최대, 최소, 평균유량의 분포가 우측으로 왜곡되어 있으며, 이를 통해 극한사상의 영향을 받을 가능성이 높을 것으로 판단하였다. 또한, 평균유량의 첨도 값이 3보다 커 이상치(극치값)가 존재할 가능성이 높을 것으로 판단하였다. 한편, Table 3에 기재된 기초통계량 결과는 안동시 유량관측소를 대상으로 수행된 결과이며, 낙동강 유역 전체 8개 관측소에 대한 종합적인 분석 결과는 본문에서 추가적으로 논의하였다.

3.3 적정 확률분포형 선정 및 매개변수 추정

3.3.1 Cullen과 Frey 그래프 및 PDF에 의한 적정 확률분포모형 평가

Cullen and Frey(1999)는 첨도와 왜도를 활용하여 확률 분포의 형태를 추정할 수 있는 방법을 제안하였다. 또한, Bootstrap 방법은 Efron(1979)에 의해 제시되었으며, 주어진 표본을 기반으로 통계량의 성질을 분석하는 기법으로, 복잡한 문제를 포함한 다양한 통계 문제에 적용되어 왔다(Kwak., 2012). 특히, 확률분포의 적합도를 평가할 때 Bootstrap 기법을 적용한다면 표본의 불확실성이 반영된 첨도-왜도 분포를 얻을 수 있어 신뢰성 있는 분석이 가능하다. 본 연구에서는 확률

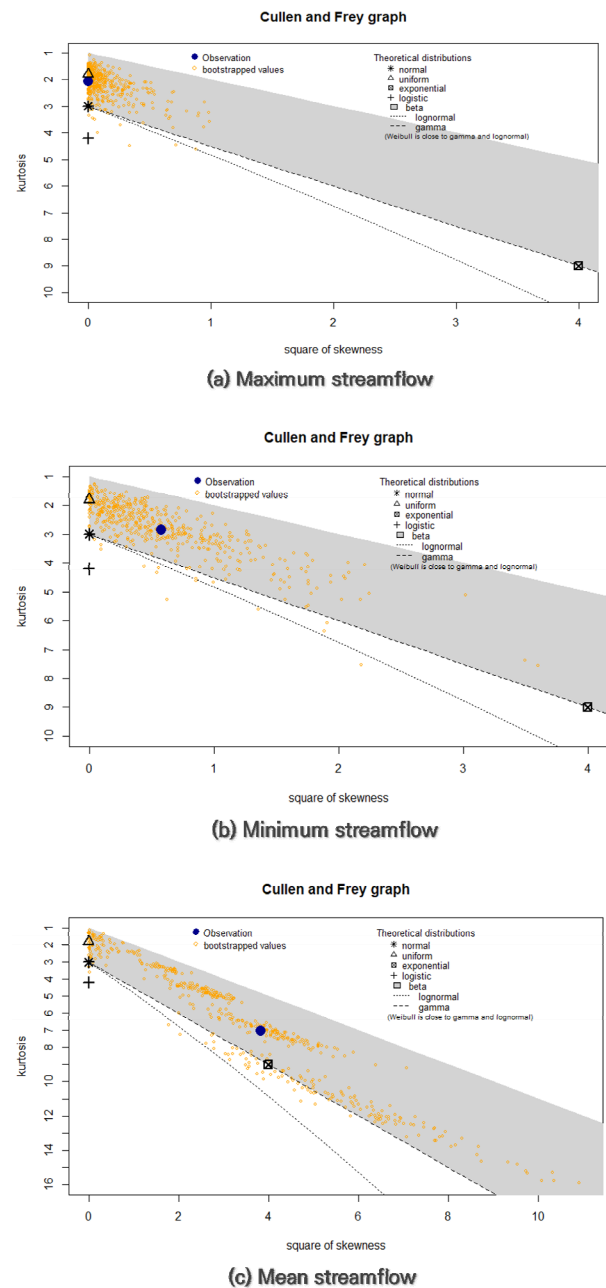


Fig. 2. Description of (a) maximum streamflow, (b) minimum streamflow, (c) mean streamflow samples from distribution with uncertainty on skewness and kurtosis estimated by bootstrap

Table 3. Basic statistics of daily streamflow (2000–2019) for Andongsi city station

Statistic	Maximum streamflow	Minimum streamflow	Mean streamflow
Minimum (m^3/s)	48.86	0.01	1.88
Maximum (m^3/s)	454.45	0.44	18.71
Median (m^3/s)	242.02	0.14	5.43
Mean (m^3/s)	254.37	0.15	6.24
Standard deviation (m^3/s)	115.94	0.13	4.25
Skewness	0.05	0.67	1.81
Kurtosis	2.07	2.58	5.80

Table 4. Possible probability distributions for daily maximum, minimum, and mean streamflows in Cullen and Frey graphs for Andongsi city station

Classification	Possible probability distribution
Maximum Streamflow	Gamma, Lognormal, Weibull
Minimum Streamflow	Gamma, Lognormal, Weibull
Mean Streamflow	Gamma, Lognormal, Weibull

Table 5. Selection of appropriate probability distribution by PDFs of daily maximum, minimum, and mean streamflows for Andongsi city station

Classification	Appropriate probability distribution
Maximum Streamflow	Lognormal, Gamma, Weibull
Minimum Streamflow	Lognormal, Normal, Gumbel
Mean Streamflow	Lognormal, Gamma, GEV

분포형의 적합성을 평가하고자 각 유출량의 첨도 및 왜도 값을 활용하여 Cullen and Frey 그래프를 도시하였다(Fig 2). 또한, 표본의 변동성을 고려하기 위해 Bootstrap 방법을 적용하였다.

Normal, Uniform, Exponential, Logistic 분포는 각각 하나의 점으로, Lognormal, Gamma, Weibull 분포는 선으로, Beta 분포는 더 넓은 영역으로 표시하였다. 또한, 자료의 첨도 및 왜도의 제곱은 청색 점으로 시각화하였다. 이와 함께, Bootstrap 기법을 활용하여 일 유출량의 관측값을 포함한 500개의 임의의 관측값을 생성하였으며, 이를 바탕으로 왜도와 첨도를 평가하였다. 해당 그림에 나타난 바와 같이 모든 분포에서 왜도 및 첨도 값의 변동성이 크게 나타났으며, 이를 통해 유출량의 변동성이 크고 강우-유출 반응이 일정하지 않다는 것을 확인하였다. 이러한 분석 결과를 바탕으로, 단일 확률 분포만으로 유출량을 설명하는 데에는 한계가 있으며, 복수의 분포모형을 고려하는 것이 필요하다고 판단하였다. 이에 본 연구에서는 일 유출량 자료를 최대유량, 최소유량, 평균유량으로 구분하여 각 유출량에 적합한 확률분포형을 Table 4와 같이 제시하였다.

이후 Fig 3과 같이 2000년부터 2019년까지의 일 유출량 자료와 최우도법을 활용하여 6개의 분포모형(Normal, Lognormal, Gamma, Weibull, GEV, Gumbel)에 대한 확률밀도함수(probability density functions, PDFs)를 비교하였다. 즉, 계산된 경험적 분포(데이터)의 히스토그램과 이론적 확률밀도함수를 중첩하여 최대, 최소, 평균유량에 대한 분포를 비교하였다. 해당 그림을 살펴보면, 최대유량, 최소유량, 평균유량 모두 우측으로 왜곡된 형태를 보이며, 왜도가 0보다 커 데이터의 분포가 완벽한 대칭을 이루지 않는 것을 파악하였다. 또한, 최대유량과 평균유량에서 대부분의 확률분포함수가 유사한 형태로 나타난 것과 최소유량에서 GEV 분포를 제외한 다른 분포 모형들이 흡사한 형태로 표현된 것을 확인하였다. 이와 함께, 경험적 분포(히스토그램)와 이론적 확률밀도함수 곡선이 얼마나 잘 일치하는지를 확인하였다. 여기서, 특정 분포의 확률분포함수가 히스토그램의 형태를 가장 잘 따라가고 있다면, 해당 분포가 데이터에 적합하다고 볼 수 있다. 각 유출량의 분포모형을 비교한 결과, 최대유량에서는 Lognormal, Gamma, Weibull 분포가 적합한 것으로, 최소유량에서는

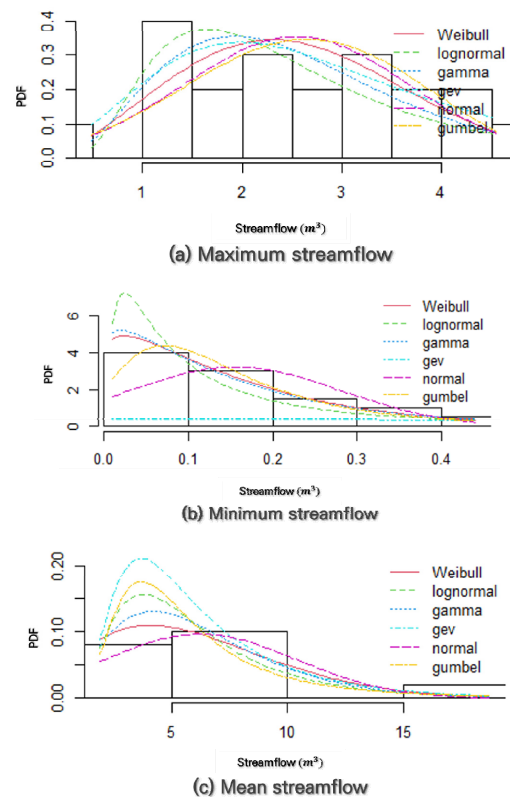


Fig. 3. Empirical and theoretical PDFs for daily (a) maximum streamflow, (b) minimum streamflow, and (c) mean streamflow at Andongsi city station

Lognormal, Normal, Gumbel 분포가 적합한 것으로, 평균 유량에서는 Lognormal, Gamma, GEV 분포가 적합한 것으로 평가되었다(Table 5).

3.3.2 P-P 및 Q-Q plot에 의한 정규성 검토 및 확률분포형 검토

유출량 데이터의 특정 분포모형에 대한 적합성을 평가하고자 P-P Plot을 활용하였으며, 이론적 확률(x축)에 대한 경험적 확률(y축)에서 평가된 경험적 분포 함수를 나타내었다. 또한, Q-Q Plot을 활용하여 선택한 분포모형의 적합도를 시각적

으로 평가하였으며, 데이터가 6개의 선택된 이론적 분포와 얼마나 유사한지 테스트하였다. Q-Q Plot에서는 분위수(x 축)에 대한 경험적 분위수(y 축)의 관계를 나타 내었다. 여기서, 데이터가 정규분포를 따르는 경우, P-P Plot과 Q-Q Plot은 1:1이 된다. 결과적으로, 6개 이론적 분포에 대해 최대, 최소, 평균유량을 기준으로 P-P 및 Q-Q Plot을 Fig. 3과 같이 도출하였다.

P-P Plot 분석한 결과, 대체적으로 정규분포를 따르는 것을 파악하였다. 최대유량의 경우 Lognormal 분포가 Weibull, Normal, Gumbel, Gamma 분포보다 적합선에 가장 근접한

것으로 나타났으나, 이들 분포 간 차이는 크지 않았다. 한편, 최소유량의 경우 Normal 분포가 다른 분포모형보다 적합선에 더욱 가깝게 나타났으나, 이들 간의 차이는 미미하였다. 평균 유량에서는 Weibull 분포가 Lognormal, Normal, Gumbel, Gamma 분포보다 상대적으로 적합선을 가장 잘 따르는 것으로 나타났지만, 네 가지 분포와의 차이는 크지 않았다. 또한, Q-Q Plot에서는 최대유량 분포모형에서 적합선과의 간격이 작게 나타났으나, 최소유량 및 평균유량의 경우 상대적으로 간격이 크게 형성되어 특정 분포에 대한 적합성이 낮을 것으로 판단하였다. 이에 따라, 관측된 데이터 계열을 실제 분포모형

Table 6. The obtained probability distribution by P-P and Q-Q plots

Classification	P-P Plot	Q-Q Plot
Maximum Streamflow	Lognormal	Lognormal
Minimum Streamflow	Normal	Normal
Mean Streamflow	Weibull	Gamma

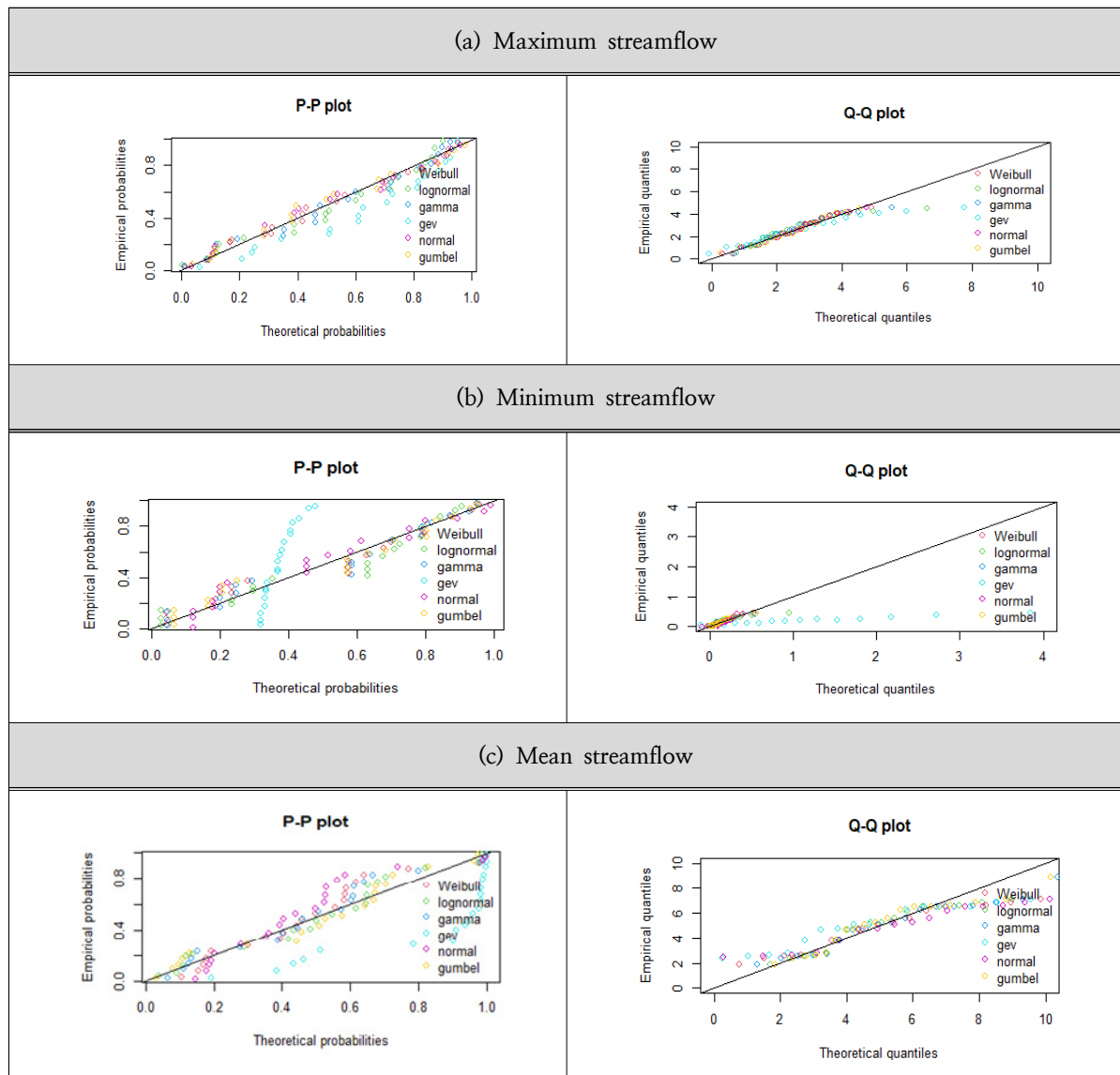


Fig. 4. P-P and Q-Q plots of daily streamflows

으로 간주하고, 이를 기반으로 추정한 분위수를 이론 분포의 분위수와 비교하는 방식이 적절하다고 판단하였으며, 각각 Lognormal, Normal, Gamma 분포를 선정하였다. 반면, 최대, 최소, 평균유량 모두에서 GEV 분포는 최적 적합선에서 가장 멀리 위치하였으며, 다른 분포들과 비교했을 때 상대적으로 큰 편차가 발생하였다. 이는 GEV 분포가 유출량 데이터의 특성을 반영하는데 적합하지 않다는 것을 의미한다. 따라서, Lognormal, Normal, Weibull, Gamma 분포모형이 데이터의 특성을 보다 잘 반영하여 적합한 결과를 제공할 수 있을 것으로 판단하였다.

3.3.2 최우도법을 통한 확률분포형의 매개변수 추정

분포모형별 매개변수 추정치는 Table 7에 제시하였으며, 적합한 분포모형의 경험적 및 이론적 누적분포함수는 Fig 4에 제시하였다. 시뮬레이션된 경험적 데이터와 적합한 분포모형의 이론적 CDF를 비교한 결과, 모든 데이터가 적합선과 매우 근접하게 일치하는 것으로 나타났으며, 그 결과는 Table 7의 상관행렬로 제시되었다. Fig 4에 따르면, 최대, 최소, 평균 유량의 6가지 분포에 대해 약 80% 이상 가장자리로 형성되어 높은 적합도가 나타났다. 또한, 일 유출량에 대한 6가지 분포 모형의 누적분포함수 적합도 결과에 따르면, 최대유량은 Gumbel

Table 7. Distribution sample estimates, standard errors, and correlation matrix of parameter shape and scale values estimated using the MLE method at Andongsi city station

Maximum streamflow					
Distribution	Standard error			Correlation matrix	
	Sample estimate	Standard error		Shape	Scale
Gamma (Pearson type 3)	shape	4.07	1.24	1.00	0.94
	scale	1.60	0.51	0.94	1.00
Lognormal	shape	0.81	0.12	1.00	0.00
	scale	0.55	0.09	0.00	1.00
Weibull	shape	2.43	0.44	1.00	0.31
	scale	2.87	0.28	0.31	1.00
GEV	shape	2.19	0.30	1.00	0.10
	scale	1.14	0.23	0.10	1.00
Gumbel	shape	1.97	0.28	1.00	0.00
	scale	1.57	0.35	0.00	1.00
Normal	shape	2.54	0.25	1.00	0.00
	scale	1.13	0.18	0.00	1.00
Minimum streamflow					
Distribution	Standard error			Correlation matrix	
	Sample estimate	Standard error		Shape	Scale
Gamma (Pearson type 3)	shape	1.14	0.32	1.00	0.80
	scale	7.33	2.57	0.80	1.00
Lognormal	shape	- 2.36	0.26	1.00	0.00
	scale	1.18	0.19	0.00	1.00
Weibull	shape	1.13	0.21	1.00	0.31
	scale	0.16	0.03	0.31	1.00
GEV	shape	0.09	0.03	1.00	0.73
	scale	0.09	0.02	0.73	1.00
Gumbel	shape	0.15	0.22	1.00	0.00
	scale	1.01	0.23	0.00	1.00
Normal	shape	0.16	0.03	1.00	0.00
	scale	0.12	0.02	0.00	1.00
Mean streamflow					
Distribution	Standard error			Correlation matrix	
	Sample estimate	Standard error		Shape	Scale
Gamma (Pearson type 3)	shape	3.02	0.91	1.00	0.92
	scale	0.48	0.16	0.92	1.00
Lognormal	shape	1.66	0.13	1.00	0.00
	scale	0.58	0.09	0.00	1.00
Weibull	shape	1.66	0.26	1.00	0.34
	scale	7.05	1.00	0.34	1.00
GEV	shape	4.21	0.58	1.00	0.65
	scale	2.18	0.48	0.65	1.00
Gumbel	shape	3.65	0.42	1.00	0.00
	scale	3.58	0.80	0.00	1.00
Normal	shape	6.24	0.93	1.00	0.00
	scale	4.15	0.66	0.00	1.00

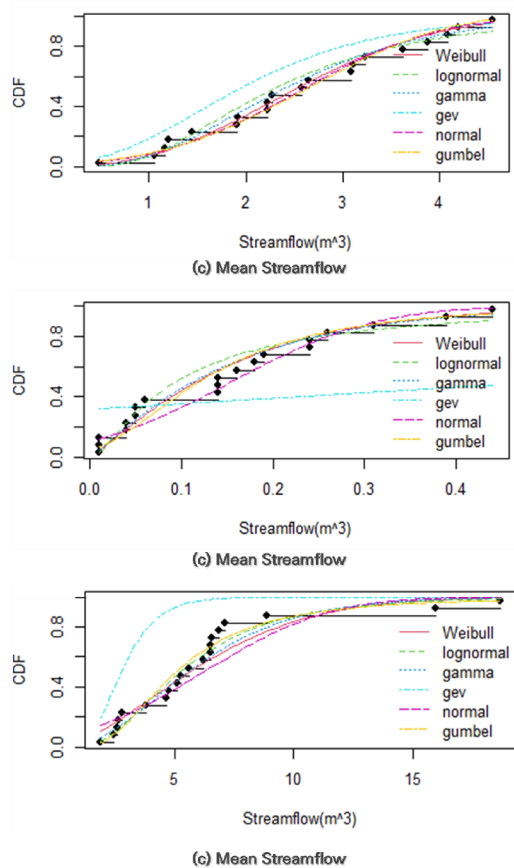


Fig. 4. Cumulative distribution functions (CDF) of the six selected distribution models: (a) maximum streamflow, (b) minimum streamflow, and (c) mean streamflow at Andongsi city station

분포가 다른 분포와 비교하여 최적의 곡선에 더 가깝게 형성되었으며, 최소유량은 Normal 분포가 최적의 곡선에 더 가깝게 형성되었다. 또한, 평균유량은 Weibull 분포가 최적의 곡선에 더 가깝게 형성되었다. 이를 통해 각 유출량 범주(최대, 최소, 평균)에 따라 최적의 적합 분포가 달라지는 것을 확인였다.

3.4 적합도 검정 및 정보기준모델 적용

3.4.1 적합도 검정의 적용 및 적정 확률분포형 선정

본 연구에서는 통계적 분석을 위해 3개의 3-매개변수 분포 (Gamma, GEV, Lognormal)와 3개의 2-매개변수 분포 (Weibull, Normal, Gumbel)를 고려하였다. K-S 검정, CVM 검정, AD 검정을 적용하여 95% 유의수준($\alpha = 0.05$)에서 6개의 확률 분포모형을 평가하였다(Table 8). 여기서, 각 분포모형에는 1에서 6까지의 순위를 지정하였으며, 1은 가장 적합한 분포, 6은 가장 적합하지 않은 분포로 표시하였다. Table 8에 제시된 적합도 검정 결과에 따르면, K-S 검정, CVM 검정, AD 검정의 모든 통계값이 95% 유의수준에서 허용 가능한 범위 내에 위치하였다. 즉, 계산된 검정 통계값이 선택한 유의수준에서 6개의 확률 분포모형에 대한 임계값보다 낮아, 모든 분포모형이 적합성을 충족하는 것으로 나타났다.

검정 결과를 구체적으로 살펴보면, K-S 검정에서는 최대 유량의 경우 임계값 0.20517보다 작아 Normal 분포가 가장 적합한 것으로 나타났으며, 최소 및 평균유량에서는 GEV 분포가 가장 적합한 것으로 분석되었다. CVM 검정에서는 최대유량의 경우 Weibull 분포가, 최소유량은 Lognormal 분포가, 평균유량은 Weibull 분포가 가장 적합한 것으로

Table 8. Fitted distributions to streamflow data by MLE and comparison of goodness-of-fit statistics

Maximum streamflow		
Distribution	Kolmogorov-Smirnov (Critical value at 0.05 = 0.20517)	
	Statistic	Rank
Gamma (Pearson type 3)	0.1153	4
Lognormal	0.1448	5
Weibull	0.0977	2
GEV	0.1051	3
Gumbel	0.2604	6
Normal	0.0965	1
Distribution	Cramer-von Mises (Critical value at 0.05 = 0.221)	
	Statistic	Rank
Gamma (Pearson type 3)	0.0443	4
Lognormal	0.0738	5
Weibull	0.0296	1
GEV	0.0355	3
Gumbel	0.4037	6
Normal	0.0331	2
Distribution	Anderson-Darling (Critical value at 0.05 = 2.5018)	
	Statistic	Rank
Gamma (Pearson type 3)	0.3110	4
Lognormal	0.4975	5
Weibull	0.2218	1
GEV	0.2458	3
Gumbel	2.0647	6
Normal	0.2321	2

나타났다. AD 검정에서는 최대유량이 Weibull 분포, 최소 유량이 Lognormal 분포, 평균유량이 Gamma 분포와 가장 높은 적합도가 나타났다. 이와 같이, 적합한 확률분포 모형은

검정 방법에 따라 다소 차이가 있으며, 각 유량 범주(최대, 최소, 평균)에 따라 서로 다른 분포가 가장 적합한 것으로 도출되었다. 따라서, 유출량 데이터를 확률적으로 모델링할

Table 8. Continued

Minimum streamflow		
Distribution	Kolmogorov-Smirnov (Critical value at 0.05 = 0.20517)	
	Statistic	Rank
Gamma (Pearson type 3)	0.1844	3
Lognormal	0.1644	2
Weibull	0.2110	4
GEV	0.1412	1
Gumbel	0.6029	6
Normal	0.2652	5
Distribution	Cramer-von Mises (Critical value at 0.05 = 0.221)	
	Statistic	Rank
Gamma (Pearson type 3)	0.0905	3
Lognormal	0.0706	1
Weibull	0.1311	4
GEV	0.0797	2
Gumbel	2.4944	6
Normal	0.2513	5
Distribution	Anderson-Darling (Critical value at 0.05 = 2.5018)	
	Statistic	Rank
Gamma (Pearson type 3)	0.6353	3
Lognormal	0.4626	1
Weibull	0.8540	4
GEV	0.4698	2
Gumbel	23.2647	6
Normal	1.5706	5
Mean streamflow		
Distribution	Kolmogorov-Smirnov (Critical value at 0.05 = 0.20517)	
	Statistic	Rank
Gamma (Pearson type 3)	0.1839	4
Lognormal	0.2324	5
Weibull	0.1735	2
GEV	0.1703	1
Gumbel	0.5233	6
Normal	0.1776	3
Distribution	Cramer-von Mises (Critical value at 0.05 = 0.221)	
	Statistic	Rank
Gamma (Pearson type 3)	0.0799	2
Lognormal	0.1314	5
Weibull	0.0779	1
GEV	0.0952	4
Gumbel	1.5258	6
Normal	0.0836	3
Distribution	Anderson-Darling (Critical value at 0.05 = 2.5018)	
	Statistic	Rank
Gamma (Pearson type 3)	0.4840	1
Lognormal	0.8553	5
Weibull	0.4905	2
GEV	0.5516	3
Gumbel	7.1054	6
Normal	0.5919	4

Table 9. Goodness of fit information criteria (AIC, BIC)

Distribution	Maximum Streamflow	
	AIC	BIC
Gamma (Pearson type 3)	66.6	68.6
Lognormal	69.3	71.3
Weibull	64.9	66.9
GEV	66.9	69.9
Gumbel	62.1	64.1
Normal	65.6	67.6
Distribution	Minimum Streamflow	
	AIC	BIC
Gamma (Pearson type 3)	-30.8	-28.8
Lognormal	-27.2	-25.2
Weibull	-30.9	-28.9
GEV	-24.3	-21.3
Gumbel	44.3	46.3
Normal	-22.6	-20.6
Distribution	Mean Streamflow	
	AIC	BIC
Gamma (Pearson type 3)	107.1	109.1
Lognormal	104.9	106.9
Weibull	109.6	111.6
GEV	106.8	109.8
Gumbel	95.1	97.1
Normal	117.6	119.6

Table 10. Selection of appropriate distribution by AIC and BIC for daily streamflows from 8 stations

Station	Maximum Streamflow	Minimum Streamflow	Mean Streamflow
Andongsi (Mokgyegyo)	Gumbel.	Weibull	Gumbel.
Gyeongjusi (Gangdongdaegyo)	Gumbel.	Weibull	Gumbel.
Mungyeonggi (Gimyongri)	Gumbel.	Gamma, Weibull	Gumbel.
Yecheongun (Gopyeonggyo)	Gumbel.	Gamma	Weibull
Uiryeonggun (Jeongamgyo)	Gumbel.	Gumbel.	Weibull
Uiseonggun (Wijungri)	Gumbel.	Lognormal	Gumbel.
Hamangun (Gyenaeri)	Gumbel.	Lognormal	Gumbel.
Hamyanggun (Hwachonri)	Gumbel.	Lognormal	Gumbel.

때 단일 검정 방법에 의존하기보다는 여러 검정을 종합적으로 고려하여 최적의 분포모형을 선정하는 것이 필요하다고 판단하였다.

3.4.2 AIC와 BIC의 적용 및 적정 확률분포형 선정

본 연구에서는 적합한 분포모형 선정하고자 정보기준 모델(AIC 및 BIC)을 활용하였으며, 그 결과를 Table 9에 제시하였다. AIC는 후보 분포모형 간 비교를 위해 동일한 변수를 사용하였으며, 정보기준모델과 귀무가설 테스트를 혼용하지 않았다. 여기서, AIC 값을 기준으로 값이 작을 수록 최적의 분포모형으로 간주하였다. 또한, BIC는 우도 함수를 기반으로 분포모형을 선택하는 기준이며, 마찬가지로 더 작은 값을 가지는 분포가 데이터에 가장 적합한 것으로 판단하였다. 정보기준모델 산정 결과를 살펴보면, 최대유량에서 Gumbel 분포가 가장 적합한 것으로 나타났

으며, 최소유량에서는 Weibull 분포가 최적의 분포로 분석되었다. 또한, 평균유량에서는 Gumbel 분포가 가장 적합한 것으로 나타났다. 이후 각 관측소 및 유량별 적정 확률분포형은 Table 10에 제시하였다.

4. 결 론

본 연구에서는 수문시계열의 유출 특성을 분석하고자 낙동강 지류 및 소하천에 위치한 8개 유량관측소를 대상으로 최근 20년(2000~2019년)의 일 단위 유출량 자료를 구축하였다. 이후 연 최대, 최소, 평균유량의 기초통계량을 산정하였으며, 최우도법(MLE)을 적용하여 확률분포 모형의 매개변수를 추정하였다. 또한, 적합도 검정(K-S 검정, CVM 검정, AD 검정)과 정보기준모델(AIC, BIC)을 활용하여 각 분포모형을 비교 및 평가하였다.

기초통계량 분석 결과를 살펴보면, 8개 관측소의 최대유량, 최소유량, 평균유량 모두 왜도 값이 0보다 커 우측으로 치우친 분포가 나타났다. 또한, 모든 유량(최대, 최소, 평균)의 왜도가 양수로 나타났으며, 특정 유량 범주에서 상대적으로 높은 왜도 값이 산정되었다. 한편, 평균유량의 첨도 값이 3보다 크게 발생하여 극치값이 존재할 가능성이 높다고 판단하였다. 결과적으로, 기초통계량을 통해 최대, 최소, 평균유량의 분포가 모두 우측으로 왜곡되어 있어 극한사상의 영향을 받을 가능성이 높을 것으로 판단하였다. 이후, 안동시 관측소를 포함한 8개 관측소의 일 유출량 데이터를 활용하여 최우도법(MLE)으로 6개의 확률분포 모형(Gamma, Weibull, GEV, Lognormal, Normal, Gumbel)의 매개변수를 추정하였다. 또한, 경험적 분포(히스토그램)와 이론적 확률밀도함수 PDF를 비교하여 각 분포의 적합성을 평가하였다. 최대유량에서는 Lognormal, Gamma, Weibull 분포가, 최소유량에서는 Lognormal, Normal, Gumbel 분포가, 평균유량에서는 Lognormal, Gamma, GEV 분포가 적합한 것으로 나타났다. 또한 최대유량과 평균유량에서는 대부분의 확률분포 모형이 유사한 형태를 보였으나, 최소유량에서는 GEV 분포가 다른 분포와 차이를 보였다. 이러한 결과를 통해 유량의 범주(최대, 최소, 평균)에 따라 적합한 확률분포 모형이 달라질 수 있으며, 단일 확률분포로 모든 유량 특성을 설명하는데 한계가 있을 것으로 판단하였다.

적합도 검정(K-S, CVM, AD) 결과를 살펴보면, 검정 방법에 따라 적합한 분포가 다소 차이가 발생하였다. 최대유량에서 Normal 분포가 가장 적합한 것으로, 최소 및 평균유량에서는 GEV 분포가 가장 적합한 것으로 나타났다. CVM 검정에서는 최대유량에서 Weibull 분포가, 최소유량에서 Lognormal 분포가, 평균유량에서 Weibull 분포가 가장 적합한 것으로 나타났다. 또한, AD 검정에서는 최대유량에서 Weibull 분포가, 최소유량에서 Lognormal 분포가, 평균유량에서 Gamma 분포가 가장 적합한 분포로 나타났다. 이러한 분석 결과는 8개 관측소에서 유사한 경향을 보였으며, 단일 검정 방법만으로는 최적의 분포를 선정하는데 한계가 있을 것으로 판단되었다. 따라서 유출량 데이터를 확률적으로 모델링할 때는 여러 검정 방법을 종합적으로 고려하여 최적의 분포 모형을 선정하는 것이 필요하다. 정보기준모델(AIC 및 BIC)을 활용하여 각 분포모형을 비교한 결과, 최대유량에서는 Gumbel과 Weibull 분포가 가장 적합한 것으로, 최소유량에서는 Gumbel 분포가 최적의 분포로 나타났다. 또한, 평균유량에서는 Gumbel 분포가 가장 높은 적합도를 보였다. 이러한 결과는 안동시뿐만 아니라 8개 관측소 전반에서 일관되게 나타났으며, 이를 종합하여 낙동강 유역의 최적 확률분포형으로 Gumbel, Weibull, Lognormal, Normal, Gamma 분포를 선정하였다.

해당 연구 결과를 통해 낙동강 유역에서 유출량을 확률적으로 분석할 때, 단일 분포모형만을 적용하는 것보다 유량의 범주(최대, 최소, 평균)에 따라 서로 다른 최적의 확률분포 모형을 고려하는 것이 필요하다고 판단하였다. 또한, 정보기준모델과 적합도 검정을 종합적으로 활용할 경우, 보다 신뢰성 높은

확률분포형을 선정할 수 있을 것으로 판단하였다. 마지막으로, 본 연구에서 선정한 확률분포형은 향후 기후변화에 따른 강우-유출 특성 변화, 홍수 및 가뭄 발생 패턴 분석, 유출량 예측 모델 구축 등의 연구에 기초자료로 활용될 수 있을 것으로 판단된다. 추가적으로, 「하천기본계획」 수립 시 설계홍수량 산정 및 유역의 유출 특성 분석과 「재해영향성검토」의 유출량 특성과 재해 영향 분석을 위한 기초자료로 활용될 수 있을 것으로 사료된다.

감사의 글

이 논문은 2022년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2022R1A2C2091773).

References

- Choe, HJ, Kim, JY (2008). Quality Management and Improvement of Hydrological Data at the Geum River Flood Control Office. *Water for future*, 42(2), pp. 49–54. [Korean Literature]
- Cullen, A, Frey, H (1999). Probabilistic Techniques in Exposure Assessment: A Handbook for Dealing with Variability and Uncertainty in Models and Inputs. *Springer Science & Business Media*, pp. 1–335.
- Efron, B (1979). Bootstrap methods : another look at the jackknife. *Annals of Statistics*, 17, pp. 1–26.
- Gheidari, M (2013). Comparisons of the L- and LH- moments in the selection of the best distribution for regional flood frequency analysis in Lake Urmia Basin. *Civil Engineering and Environmental Systems*, 30(1), pp. 72–84. [DOI: <https://doi.org/10.1080/10286608.2012.749870>]
- Hong, CS, Kim, JH (2009). Nonparametric homogeneity tests of two distributions for credit rating model validation. *Journal of the Korean data & information science society*, 20(2), pp. 261–272. [Korean Literature]
- Im, CE, Park, WJ, Jang, KH (2020). Analysis of Crime Characteristics Based on Chi-Square Using Crime Statistics Data. *Proceedings of the 2020 Conference of the Korean Institute of Communications and Information Sciences*, Korean Institute of Communications and Information Sciences, Seoul, Korea. [Korean Literature]
- Jeong, HY, Hong, SH, Jeon, JS, Lim, SC, Kim, JC, Park, HW, Park, CY (2022). Nonparametric homogeneity tests of two distributions for credit rating model validation. *Journal of Korea Multimedia Society*, 25(1), pp. 82–91. [Korean Literature] [DOI: <https://doi.org/10.9717/kms.2022.25.1.082>]

- Joo, HJ, Kim, DH, Kim, JW, Choi, CH, Han, DG, Lee, JH, Kim HS (2016). A Study on Optimal Stage Gauge Network Considering Correlation of Individual Stage Gauge Station. *Journal of Wetlands Research*, 18(4), pp. 404–412. [Korean Literature] [DOI: <http://dx.doi.org/10.17663/JWR.2016.18.4.404>]
- Joo, KW, Ahn, HJ, Kim, TR, Heo, JH (2019). AImprovement of the Parameter Estimation Method of Copula Model for Hydrological Data. *Proceedings of the 2019 Conference of the KSCE Journal of Civil Engineering*, KSCE Journal of Civil Engineering, Seoul, Korea. pp. 212–213. [Korean Literature]
- Khosravi, G, Majidi, A, Nohegar, A (2012). Determination of suitable probability distribution for annual mean and peak discharges estimation (case study: Minab river–bar antin gage, iran). *Int. J. Probab. Stat*, 1, pp. 160–163. [DOI: <https://doi.org/10.5923/j.ijps.20120105.03>]
- Kim, BK, Kim, SD, Lee, JS, Kim, HS (2006). Spatio-Temporal Characteristics of Droughts in Korea: Construction of Drought Severity–Area–Duration Curves. *KSCE Journal of Civil and Environmental Engineering Research B*, 26(1B), pp. 69–78. [Korean Literature]
- Kim, HB, Kim, TR, Shin, HJ, Heo, JH (2017). A study on a tendency of parameters for nonstationary distribution using ensemble empirical mode decomposition method. *Journal of Korea Water Resources Association*, 50(4), pp. 253–261. [Korean Literature] [DOI: <https://doi.org/10.3741/JKWRA.2017.50.4.253>]
- Kim, HS, Moon, JW, Kim, JH, Kim, JH (2000). A Study on Drought Trend in Han River Basin. *Journal of Korea Water Resources Association*, 33(4), pp. 437–446. [Korean Literature]
- Kim, JY, Kim, JG, Cho, YH, Kwon, HH (2017). A development of Bayesian Copula model for a bivariate drought frequency analysis. *Journal of Korea Water Resources Association*, 50(11), pp. 745–758. [Korean Literature] [DOI: <https://doi.org/10.3741/JKWRA.2017.50.11.745>]
- Kim, MJ, Jeon, MH, Sung, KI, Kim, YJ (2016). Bayesian structural equation modeling for analysis of climate effect on whole crop barley yield. *The Korean Journal of Applied Statistics*, 29(2), pp. 331–344. [Korean Literature] [DOI: <http://dx.doi.org/10.5351/KJAS.2016.29.2.331>]
- Kim, MJ, Park, KW, Kim, SW, Kim, TW (2022). Estimating design floods based on bivariate rainfall frequency analysis and rainfall–runoff model. *Journal of Korea Water Resources Association*, 55(10), pp. 737–748. [Korean Literature] [DOI: <https://doi.org/10.3741/JKWRA.2022.55.10.737>]
- Kim, SJ, Joo, HJ, Lee, JH, Jun, HD, Kim, HS (2014). Evaluation of Stream Gauge Network Considered Discharge Characteristics Between Upstream and Downstream of the River. *Journal of The Korean Society of Hazard Mitigation*, 14(4), pp. 309–319. [Korean Literature] [DOI: <https://doi.org/10.9798/kosham.2014.14.4.309>]
- Kim, SU, Lee, YS (2015). Flood Frequency Analysis Considering Probability Distribution and Return Period under Non-stationary Condition. *Journal of Korea Water Resources Association*, 48(7), pp. 567–579. [Korean Literature] [DOI: <https://doi.org/10.3741/JKWRA.2015.48.7.567>]
- Kim, TR, Shin, HJ, Heo, JH (2013). A Study on Empirical Distribution Function with Unknown Shape Parameter and Extreme Value Weight for Three Parameter Weibull Distribution. *Journal of Korea Water Resources Association*, 46(6), pp. 643–653. [Korean Literature] [DOI: <http://dx.doi.org/10.3741/JKWRA.2013.46.6.643>]
- Kim, YS, Kim, SJ, Kwak, JW, Kim, HS (2012). Estimation of Frequency Based Snowfall Depth Considering Climate Change. *Proceedings of the 2012 Conference of the Korea Water Resources Association*, Korea Water Resources Association, Seoul, Korea. 196. [Korean Literature]
- Kim, YU (2017). *Uncertainty analysis of probability distributions for the design rainfall*. Master's Thesis, Kumoh National Institute of Technology, Gumi, Korea.
- Kolmogorov, A (1933). Sulla determinazione empirica di una legge di distribuzione. *Giorn Dell'inst Ital Degli Att*, 4, pp. 89–91.
- Kwak, JW (2012). *Copula Theory Based Drought Analysis and Impact of Climate Change on Droughts*. Ph.D Dissertation, Inha University, Incheon, Korea.
- Kwak, JW, Kim, SJ, Jung, JW, Singh, V, Lee, DR, Kim, HS (2016). Assessment of meteorological drought in Korea under climate change. *Advances in Meteorology*, 2016(1), 1879024. [DOI: <https://doi.org/10.1155/2016/1879024>]
- Kwak, JW, Kim, SJ, Kim, GH, Singh, V, Park, JS, Kim, HS (2016). Bivariate drought analysis using streamflow reconstruction with tree ring indices in the Sacramento Basin, California, USA. *Water*, 8(4), 122. [DOI: <https://doi.org/10.3390/w8040122>]
- Kyoung, MS, Kim, SD, Lee, BK, Kim, HS (2007). Construction of Hydrological Drought Severity–Area–Duration Curves Using Cluster Analysis. *KSCE Journal of Civil and Environmental Engineering Research B*, 27(3B), pp. 267–276. [Korean Literature]
- Lee, KJ, Park, CK (2020). Goodness-of-fit Test for Normality Based on Decile Dispersion Ratio. *Journal of the Korean Data Analysis Society*, 22(3), pp. 1023–1032. [Korean Literature] [DOI: <http://dx.doi.org/10.37727/jkdas.2020.22.3.1023>]

- Olofintoye, O, Sule, B, Salami, A (2009). Best-fit Probability distribution model for peak daily rainfall of selected Cities in Nigeria. *New York Science Journal*, 2(3), pp. 1-12.
- Park, CS, Yoo, CS (2012). Review of Parameter Estimation Procedure of Freund Bivariate Exponential Distribution. *Journal of Korea Water Resources Association*, 45(2), pp. 191-201. [Korean Literature] [DOI: <https://doi.org/10.3741/JKWRA.2012.45.2.191>]
- Ryu, YY (2010). *Comparative Study of Flood Frequency Analysis Using Observed Flow and Frequency Based Rainfall*. Master's Thesis, Inha University, Incheon, Korea.
- Shin, HJ, Lee, TH (2020). A Study on the Estimation of Missing Hydrological Data Using Adaptive Network-based Fuzzy Inference System (ANFIS). *Proceedings of the 2020 Conference of the Korea Water Resources Association*, Korea Water Resources Association, Seoul, Korea. 264. [Korean Literature]
- Shin, HJ, Sung, KM, Heo, JH (2010). Derivation of Modified Anderson-Darling Test Statistics and Power Test for the Gumbel Distribution. *Journal of Korea Water Resources Association*, 43(9), pp. 813-822. [Korean Literature] [DOI: <https://doi.org/10.3741/jkwra.2010.43.9.813>]
- Smirnov, N (1948). Table for estimating the goodness of fit of empirical distributions. *The annals of mathematical statistics*, 19(2), pp. 279-281. [DOI: <https://doi.org/10.1214/aoms/1177730256>]
- Stewart, B (2015). Measuring what we manage—the importance of hydrological data to water resources management. *Proceedings of the International Association of Hydrological Sciences*, 366, pp. 80-85.
- Walker, S (2000). The value of hydrometric information in water resources management and flood control. *Meteorological Applications*, 7(4), pp. 387-397. [DOI: <https://doi.org/10.1017/S1350482700001626>]
- Yoo, JY, Kim, DH, Kim, HS, Kim, TW (2016). Application of copula functions to construct confidence intervals of bivariate drought frequency curve. *Journal of Hydro-environment Research*, 11, pp. 113-122. [DOI: <https://doi.org/10.1016/j.jher.2014.10.002>]
- You, KT, Kim, JR, Yun, ZH, Pak, GJ (2017). Development of Application Method of Influent Wastewater Generation and Activated Sludge Process Design Based on Probability Density Function. *Journal of Korean Society on Water Environment*, 33(2), pp. 140-148. [Korean Literature] [DOI: <https://doi.org/10.15681/KSWE.2017.33.2.140>]