

Bayesian 기법을 이용한 혼합 Gumbel 분포 매개변수 추정 및 강우빈도해석 기법 개발

최홍근* · 오랑치맥숨아** · 김용탁*** · 권현한****

Choi, Hong-Geun*, Uranchimeg, Sumiya**, Kim, Yong-Tak***, Kwon, Hyun-Han****

A Bayesian Approach to Gumbel Mixture Distribution for the Estimation of Parameter and its use to the Rainfall Frequency Analysis

ABSTRACT

More than half of annual rainfall occurs in summer season in Korea due to its climate condition and geographical location. A frequency analysis is mostly adopted for designing hydraulic structure under the such concentrated rainfall condition. Among the various distributions, univariate Gumbel distribution has been routinely used for rainfall frequency analysis in Korea. However, the distributional changes in extreme rainfall have been globally observed including Korea. More specifically, the univariate Gumbel distribution based rainfall frequency analysis is often fail to describe multimodal behaviors which are mainly influenced by distinct climate conditions during the wet season. In this context, we purposed a Gumbel mixture distribution based rainfall frequency analysis with a Bayesian framework, and further the results were compared to that of the univariate. It was found that the proposed model showed better performance in describing underlying distributions, leading to the lower Bayesian information criterion (BIC) values. The mixed Gumbel distribution was more robust for describing the upper tail of the distribution which plays a crucial role in estimating more reliable estimates of design rainfall uncertainty occurred by peak of upper tail than single Gumbel distribution. Therefore, it can be concluded that the mixed Gumbel distribution is more compatible for extreme frequency analysis rainfall data with two or more peaks on its distribution.

Key words : Bayesian method, BIC, Frequency analysis, Mixed Gumbel distribution

초록

우리나라의 기후 지형적 특성에 따라 연강수량의 50% 이상이 여름철에 내린다. 이러한 짧은 기간에 집중적으로 내리는 강수량 조건하에 수공구조물을 설계할 경우 대부분 극치빈도분석을 활용한다. 특히 우리나라의 경우 Gumbel 분포를 활용한 극치빈도분석을 많이 이용한다. 하지만, 최근 이상기후로 인하여 전세계적으로 강수량의 특징이 급격히 변하고 있으며, 우리나라 연강수량 특징도 바뀌고 있다. 즉, 기존의 단일 분포형으로 재현이 가능했던 수문기상 자료들이 혼합분포형의 특징을 가지게 되었으며 이러한 변화를 고려할 수 있는 극치빈도분석 개발이 요구되고 있는 실정이다. 본 연구에서는 두 개 이상의 첨두를 가지는 형태의 극치강수량 자료에 대해서 기존의 단일 Gumbel 분포형 기반 극치빈도분석과 혼합 Gumbel 분포형 기반의 극치빈도분석 결과를 비교하였다. 확률분포의 매개변수 산정시 우도함수를 Bayesian 기법을 통해 산정하여 각 분

* 정회원 · 전북대학교 토목공학과 석사과정 (Chonbuk National University · chlghdms12@jbnu.ac.kr)

** 전북대학교 토목공학과 박사수료 (Chonbuk National University · sumya963@jbnu.ac.kr)

*** 전북대학교 토목공학과 박사과정 (Chonbuk National University · dydxkr@jbnu.ac.kr)

**** 종신회원 · 교신저자 · 전북대학교 토목공학과 부교수, 공학박사 (Corresponding Author · Chonbuk National University · hkwon@jbnu.ac.kr)

Received November 29, 2017/ revised January 3, 2018/ accepted January 8, 2018

포형의 Bayesian information criterion (BIC) 값을 비교하였다. 분석한 결과, 앞서 제안된 혼합 Gumbel 분포형은 하나의 첨두를 가지는 단일 Gumbel 분포형에서 반영되지 못한 꼬리(tail)부분의 이중첨두 부분의 거동을 효과적으로 모의하는 것을 확인할 수 있었다. 결과적으로 설계강수량을 추정할 때 보다 신뢰성있는 접근이 가능하였다. 이러한 점에서 우리나라 극치강우자료 분석시 기존 단일분포기반의 빈도해석기법에 대안으로 적용이 가능할 것으로 판단된다.

검색어 : 베이지안 기법, BIC, 빈도해석, 혼합 Gumbel 분포형

1. 서론

우리나라의 강수 현상은 계절적으로 여름철에 집중되며 이 기간의 대표적인 강수발생 메커니즘으로는 장마와 태풍이 있다. 최근 들어 여름철 강수특성이 과거와 많이 달라지고 있다는 연구들이 많아지고 있다(Park et al., 2008). 장마와 태풍으로 인한 강수량은 강우지속시간 및 강우강도 측면에서 서로 다른 특징을 가지고 있으며(Kwon et al., 2008b), 이러한 특징을 고려한 강우분석과 관련된 연구가 다수 진행되었다. 즉, 우리나라의 여름철 극치강수량은 이러한 특징으로 인해 지속시간이 길어지는 경우 이중첨두(multi mode)를 가지는 경우가 많으며(Ho et al., 2003), 단일 확률분포를 이용한 일반적인 강우빈도해석 시 이러한 이중첨두(multi mode)의 특징을 고려하는데 한계가 있다.

우리나라는 실무적인 관점에서 확률강우량도 작성을 위해 Gumbel 분포형을 추천하고 있으며 매개변수 추정방법으로는 확률기중모멘트 방법이 일반적으로 사용되고 있다. 우리나라에서 강우빈도해석 시 확률분포형 선정에 대한 연구가 다수 진행되었다. Heo et al.(1999)은 전국 22개 관측지점의 연 최대치계열 자료를 이용하여 우리나라의 강수 확률분포형을 분석한 결과, GEV 분포를 우리나라 연최대 강우자료에 가장 적합한 확률분포형으로 선정하였다. Lee et al.(2000)은 21개 관측지점의 연 최대치계열 자료를 이용하여 우리나라의 강수 확률분포형을 분석한 결과, Gumbel 분포를 대표 확률분포형으로 선정하였다. Lee et al.(2009)은 모멘트법, 최우도법, 확률기중모멘트법을 이용하여 대표확률분포형 후보군을 선정한 후, resampling 방법인 Jackknife 기법을 적용하여 Gumbel 분포가 지속기간 12시간, 24시간에 대해 가장 적합한 것으로 평가했다.

매개변수적 확률분포형과는 다르게 핵밀도함수(kernel density function)를 이용한 비매개변수적 방법에 대한 연구도 다수 진행되었다(Kwon et al., 2007). 댐위험도 분석시 분포형 선정이 요구되지 않는 비매개변수적 방법을 활용하여 다중첨두를 가지는 경우에도 유연하게 확률분포 추정이 가능하다고 제시했고, 기존 매개변수적 방법에서 댐안전성이 비현실적으로 추정되는 반면, 비매개변수적 방법은 주요 매개변수들의 합리적인 댐안전성 평가를 제공하였다(Kwon and Moon, 2006). 특히, 자료의 기지점 내부의 내삽에 있어서는 우수한 성능을 발휘하는 장점이 있다. 핵밀도함수 방법은

여러 장점이 있는 반면 꼬리(tail)부분의 외삽의 경우 상대적으로 빨리 수렴하는 단점이 있어 빈도해석 시 특정빈도 이상에서는 확률강수량의 변화가 미미한 경우가 있으며, 이러한 문제점을 해결하고자 핵밀도함수를 개발하는 연구도 진행되었다(Kwon and Moon, 2006). 비매개변수적 방법의 또 다른 문제점은 Quantile 함수가 존재하지 않으며, 수치적인 접근만이 가능하다는 점이다. 이러한 점에서 2개 이상의 매개변수적 확률분포가 혼합된 혼합분포의 적용에 대한 연구가 필요하다.

통계적으로 단일한 모집단이라는 가정 하에서 우리나라의 경우 Gumbel 분포가 전반적으로 최적 분포형으로 선정되고 있지만, 많은 경우에 있어서 혼합분포의 형태를 갖는 경우가 있으며 이로 인해 강우빈도해석 시 과대/과소 추정되는 문제가 발생할 수 있다. 이러한 혼합분포의 발생원인은 여러 가지가 있을 수 있다. 첫째, 강우의 표본오차(sampling error)로 기인할 수 있다. 즉, 빈도해석을 위한 자료가 충분치 않아서 나타나는 일시적인 현상일 수 있다. 둘째, 강우의 특성이 다른 두 개 이상의 발생원이 서로 혼재되어 있는 경우로 가정할 수 있다. 혼합분포형과 관련된 국내연구 사례에서는 강우자료를 모집단별로 분류(특정사건 또는 기간별로 역추적하여 장마와 태풍으로 분류)해서 해석하는 방법(Yoon et al., 2012)과 특정한 분류 없이 혼합분포를 적용하여 해석하는 방법으로 구분된다(Shin and Lee, 2014). 혼합분포를 이용한 빈도해석 연구는 매개변수 추정에 대한 어려움으로 제한적으로 수행되고 있으며, 최우도법 또는 EM (expectation-maximization) 알고리즘이 주로 제안되고 있으며, 특히 EM 기법이 가장 대표적인 혼합분포 매개변수 추정방법이라 할 수 있다. 그러나 EM 기법의 경우 초기치에 매우 민감하며 수치해석상의 불안정성으로 인해 자료가 부족한 경우 매개변수 추정에 어려움이 크다(Redner and Walker, 1984). 이러한 점에서 본 연구의 주요 목적은 Bayesian 기법을 이용한 혼합분포의 매개변수 추정방법을 개발하고자 하며, 이를 통해 해석상의 안정성을 도모하고 표본오차에 대한 영향을 불확실성으로 표현하고자 한다. 본 연구에서는 강우자료를 특정사건 또는 기간을 조사하여 모집단을 분리하지 않았으며, Bayesian MCMC (Markov Chain Monte Carlo) 기법을 적용하여 매개변수 추정 및 강우빈도해석을 수행하였다. 모형의 적합성을 평가하기 위하여 모의실험을 통하여 매개변수 추정 결과를 비교 검토하였다. 최종적으로 자료가

간이 비교적 긴 전주지점을 대상으로 앞서 단일 Gumbel 분포모형과 혼합 Gumbel 분포모형을 구현하여 주요 지속시간에 대하여 상호비교를 실시하였다.

2. 대상자료

본 연구에서는 자료기간이 비교적 긴 전주지점의 1961년부터 2016년까지 56년의 기상청 강우자료를 활용하였으며, 8개의 지속시간별 연최대강우량을 추출하여 사용하였다. 지속시간별 강우자료의 핵밀도함수와 Gumbel 확률지(probability plot)를 이용해

Fig. 1에 제시하였다. 대부분의 지속시간에서 다중침투의 형태를 띄며, 단일분포보다는 혼합분포의 적용이 자료의 분포특성을 묘사하는데 유리할 것으로 판단된다. 확률지를 통해 보더라도 1시간 지속시간의 경우 이론적인 Gumbel 분포를 나타내는 빨간 실선을 벗어나는 자료가 2번째 모드에서 나타나고 있으며, 6시간 지속시간의 경우도 2번째 모드에서 이론적인 분포에서 벗어나고 있음을 확인할 수 있다. Fig. 1에서 지속시간 1, 6, 9, 12시간의 자료에서 서로 다른 침투가 육안으로 식별되었다. 따라서 전주 지점의 강우자료는 서로 다른 2개 이상의 모집단으로 이루어졌다고 판단할 수 있다.

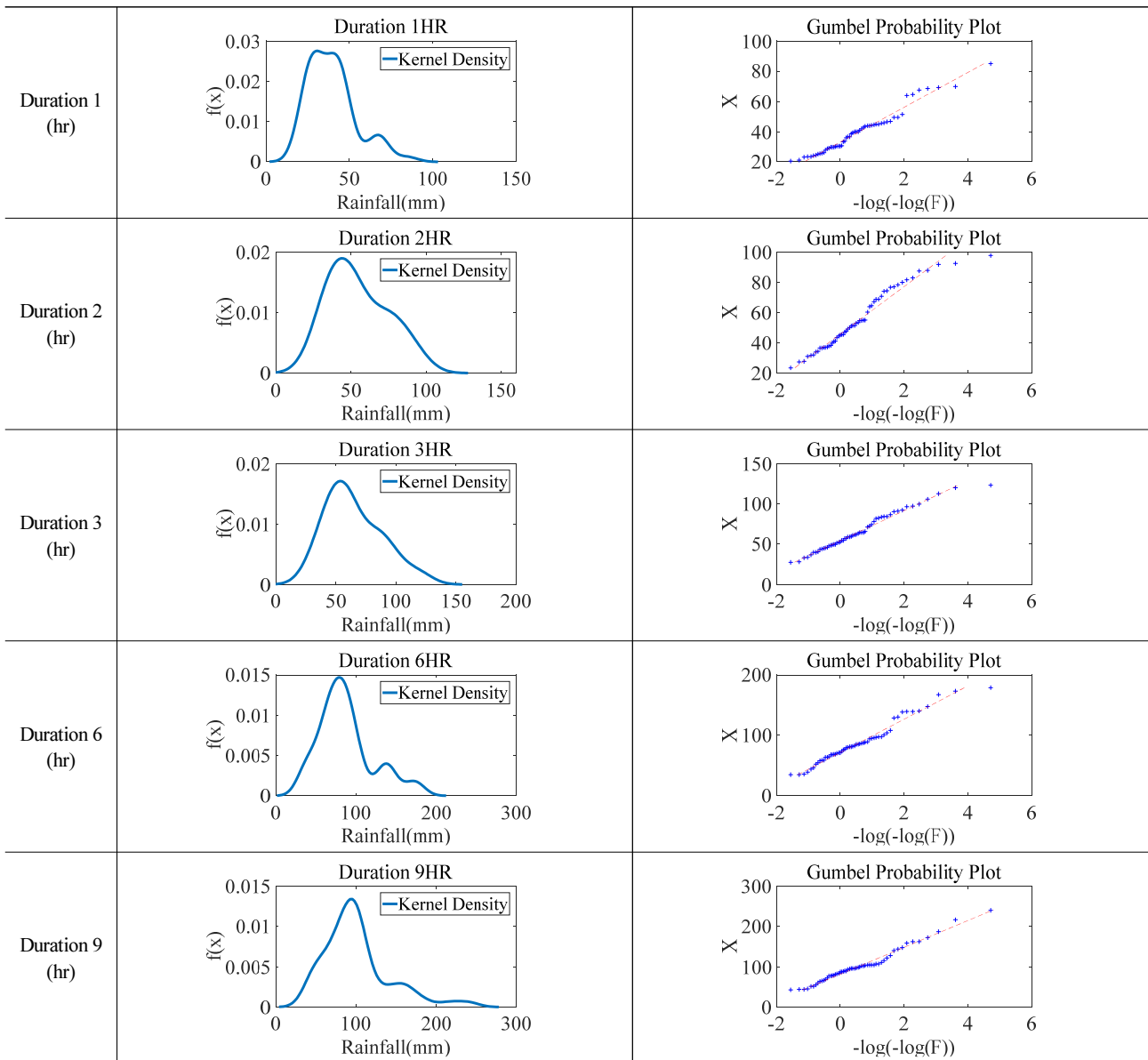


Fig. 1. A Graphical Representation of Annual Maximum Rainfall at Jeonju Station. Kernel Density Functions are Displayed in the Left Panel While Gumbel Probability Plots are Illustrated in the Right Panel for Different Durations

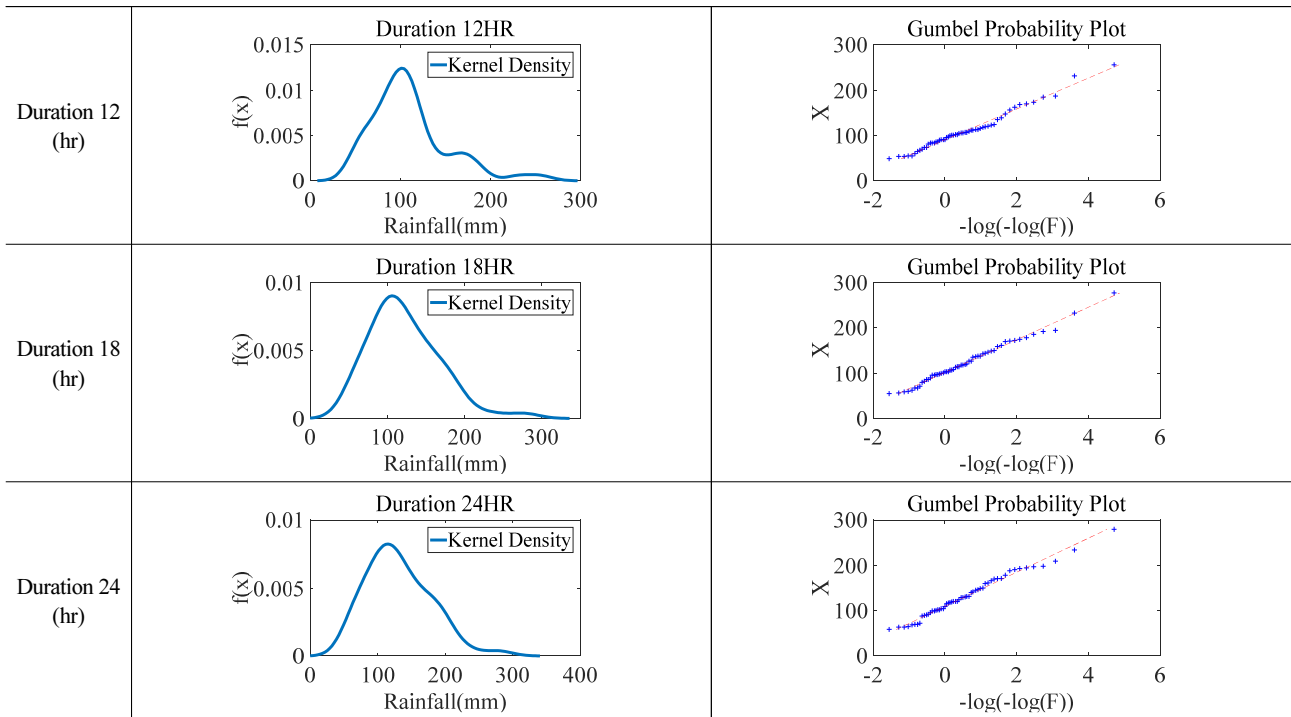


Fig. 1. A Graphical Representation of Annual Maximum Rainfall at Jeonju Station. Kernel Density Functions are Displayed in the Left Panel While Gumbel Probability Plots are Illustrated in the Right Panel for Different Durations (Continued)

3. Bayesian 기법 기반의 혼합 Gumbel 분포 매개변수 추정

본 절에서는 Bayesian 기법 기반의 매개변수 추정과 불확실성을 평가하는 방법에 대해서 서술하였다. 이와 더불어, 본 연구에서 제안한 Bayesian MCMC 기법을 모의실험을 통하여 검증하였다.

3.1 Bayesian 기법을 이용한 혼합 확률분포 매개변수 추정 방법

Bayesian 기법을 통한 매개변수 추정기법은 기존 방법들(최우도법, 모멘트법, 확률가중모멘트법)과는 다르게 매개변수를 하나의 확률변수로 취급한다. 즉, 매개변수가 단일 값이 아닌 확률분포의 형태로 부여되며 최종적으로 매개변수의 사후 분포(posterior distribution)를 추정하는데 목적을 두며, Bayes' 정리(Eq. (1))를 기반으로 한다.

$$p(\theta | x) = \frac{p(\theta, x)}{p(x)} \propto p(x | \theta) \cdot p(\theta) \quad (1)$$

여기서, $p(\theta|x)$ 은 사후 분포로, 2장에서 제시한 강우자료의 각 모집단들이 Gumbel 분포($Gumbel(\mu_1, \sigma_1)$), ($Gumbel(\mu_2, \sigma_2)$)를 따른다고 가정하며, $\theta = [\mu_1, \sigma_1, \mu_2, \sigma_2, \rho]$ 는 Gumbel 분포에 대한 전체매개변수들의 집합을 나타낸다. $p(x)$ 는 관측자료 x 의

주변분포(marginal distribution), $p(\theta)$ 는 매개변수들의 사전분포를, $p(x | \theta)$ 는 극치자료 x 의 우도함수(likelihood function)를 의미하므로 Eq. (2)와 같이 나타낼 수 있다.

$$p(\mathbf{x} | \theta) = \rho \prod_{n=1}^N Gumbel(x_n | \mu_1, \sigma_1) + (1 - \rho) \prod_{n=1}^N Gumbel(x_n | \mu_2, \sigma_2) \quad (2)$$

여기서, N 은 강수시계열의 자료연수를 나타낸다. 매개변수들의 사전분포($p(\theta)$)는 사전정보를 기반으로 하는 Informative 사전분포와 사전정보에 크게 의존하지 않는 Noninformative 사전분포가 있으며, Noninformative 사전분포로는 정규분포, 균일분포, 지수분포가 대표적으로 활용된다(Gelman et al., 2004). 사전분포를 결정하는데 있어서 공액분포(conjugate distribution), 즉 사전분포와 사후분포가 동일하게 결정되는 사전분포를 활용하는 것이 매개변수 추정에 안정성을 도모할 수 있으나, 본 연구와 같이 공액분포 추정이 어려운 경우, 비공액분포를 사용하는 것이 일반적이다(Gelman et al., 2004). 이러한 점에서 본 연구에서는 사전분포로 각각의 매개변수를 정규분포와 Gamma 분포, Dirichlet 분포로 가정하였다. 다시 말해, 위치매개변수에 대한 사전분포로서 분산이

큰 정규분포를 활용하였으며, 규모매개변수에 대해서는 음의 값을 방지하기 위하여 Gamma 분포를 활용하였다. 또한, 각 모집단별 가중치(weight)매개변수는 총합이 1인 것을 이용하여 Dirichlet 분포를 활용하였다. 본 연구에서 추정이 필요한 5개의 매개변수 수에 비해 전주시점 56년의 자료 연수가 충분히 큰 점을 고려할 때, Noninformative 사전분포를 통한 추정에도 큰 무리가 없을 것으로 판단하였다.

$$p(\mu_1) \sim Normal(\mu_{\mu_1}, \sigma_{\mu_1}^2) \quad (3a)$$

$$p(\mu_2) \sim Normal(\mu_{\mu_2}, \sigma_{\mu_2}^2) \quad (3b)$$

$$p(\sigma_1) \sim Gamma(k_{\sigma_1}, s_{\sigma_1}) \quad (3c)$$

$$p(\sigma_2) \sim Gamma(k_{\sigma_2}, s_{\sigma_2}) \quad (3d)$$

$$p(\rho) \sim Dirichlet(K, \alpha_{i=1:K}) \quad (3e)$$

Eqs. (3a)~(3e)에서 위치매개변수 μ_1, μ_2 는 각각 $\mu_{\mu_1}, \sigma_{\mu_1}^2$ 과 $\mu_{\mu_2}, \sigma_{\mu_2}^2$ 를 갖는 정규분포를 따르고, 규모매개변수 σ_1, σ_2 는 각각 $k_{\sigma_1}, s_{\sigma_1}$ 와 $k_{\sigma_2}, s_{\sigma_2}$ 를 갖는 Gamma 분포를 따르며, 가중치매개변수 ρ 는 Dirichlet 분포를 따른다. 본 연구에서는 각 모집단에서 가중치 ρ 와 각 Gumbel 분포의 2개의 매개변수가 모두 불확실성을 가지고 있다는 가정 아래 확률적 추론(statistical inference)을 수행했으며, 5개의 매개변수에 대한 결합 확률밀도함수(joint distribution)는 Eq. (4)와 같다.

$$p(\mu_1, \sigma_1, \mu_2, \sigma_2, \rho) \propto 1 \quad (4)$$

Eq. (2)~(4)에서 정의되는 매개변수들의 사전분포들을 Eq. (1)에 대입시킴으로써 매개변수들의 사후분포를 Eq. (5)와 같이 추정할 수 있다.

$$p(\theta | x) \propto [\rho \prod_{n=1}^N Gumbel(x_n | \mu_1, \sigma_1) \cdot Normal(\mu_{\mu_1}, \sigma_{\mu_1}^2) \cdot \Gamma(k_{\sigma_1}, s_{\sigma_1}) + (1-\rho) \prod_{n=1}^N Gumbel(x_n | \mu_2, \sigma_2) \cdot Normal(\mu_{\mu_2}, \sigma_{\mu_2}^2) \cdot \Gamma(k_{\sigma_2}, s_{\sigma_2})] \quad (5)$$

Eq. (5)에서 모든 매개변수에 대한 적분을 통해 직접적으로 추정하는 것은 불가능하다. 따라서, 본 연구에서는 앞서 언급한 MCMC방법을 도입하여 깃스표본법(gibbs sampling)을 이용하여

매개변수들의 사후분포를 추정하였다.

3.2 Bayesian Markov Chain Monte Carlo 모의

Bayesian MCMC기법은 사후분포를 추정하는데 있어서 다변량에 대한 복잡한 적분을 위해서 적용되는 수치해석 기법으로 매개변수를 추정하는 수단으로 활용되고 있다(Kwon et al., 2008a). 일반적인 Monte Carlo기법은 확률변수들 간의 독립성을 가정으로 이루어지는 샘플링 방법이라면, MCMC기법은 다변량에 대해서 종속성을 기준으로 조건부 샘플링이 가능한 방법이라 할 수 있다. 즉, 주어진 다변량 확률분포가 복잡하고 iid (independent identically distributed) 난수 대신 Markov Chain에 근거한 난수를 추출하는데 적합하다. Markov Chain을 통해 다변량 확률분포를 따르는 난수를 재현하기 위해서는 반복 시행을 통해 분포에 수렴시키는 과정이 필요하다. 따라서 MCMC기법은 Bayesian 통계 기법에서 사후분포의 추론을 위해서 활용될 수 있으며 매개변수들 간의 결합 확률 특성을 고려한 해석을 가능하게 한다.

MCMC기법의 대표적인 방법으로 메트로폴리스-헤스팅 알고리즘(Metropolis-Hastings algorithm)과 깃스표본법 등이 있으며, 본 연구에서는 깃스표본법을 이용하였다. 깃스표본법은 원하는 다변량 확률분포에서 iid 표본을 추출하는 것이 복잡할 때 사용하는 방법으로써 2개의 변수를 갖는 다변량 확률분포를 이용하여 설명하면 다음과 같다. 2개의 변수를 갖는 다변량 확률밀도함수를 $f(\mu, \sigma)$ 라고 정의한다. 깃스표본법은 확률밀도함수로부터 직접 표본을 추출할 수는 없으나 각각의 변수들의 대해서 다른 두 변수들이 주어졌을 때의 조건부 분포가 알려져 있고 이로부터의 표본추출이 가능한 경우에 사용할 수 있다(Kwon et al., 2012). 알고리즘을 간단히 정리하면 다음과 같다.

[1] Gumbel 분포의 두 매개변수에 대한 초기 값($\mu^{(0)}, \sigma^{(0)}$)을 부여한다.

[2] i 번째 난수 벡터 $(\mu^{(i)}, \sigma^{(i)})$ 가 주어졌을 때 $i+1$ 번째 난수 벡터를 다음과 같은 조건부 분포에서 추출한다.

$$(1) \mu^{i+1} \sim f(\mu | \sigma^{(i)})$$

$$(2) \sigma^{i+1} \sim f(\sigma | \mu^{(i+1)})$$

[3] 위의 과정을 충분히 반복한 후 초기의 일정부분 난수를 제거한 이후의 난수들을 이용한다. 이러한 제거과정을 Burning이라고 하며 Bayesian 해석에서 일반적으로 요구되는 단계이다(George and McCulloch, 1993).

위의 방법에서 깃스표본법은 현재시점의 매개변수들의 값은

정확하게 바로 직전의 추출된 값들이 사용되게 되며, 따라서 조건부 분포에서 추출된 난수들이 안정 상태에 도달하는 것이 주어진 다변량 확률분포를 정확히 따르는 난수가 되는 척도가 되며 깃스표 본법을 구현하는데 가장 중요한 부분이 된다(Kwon et al., 2012)

3.3 모의실험

Bayesian 기법과 연계한 혼합 Gumbel 분포의 매개변수를 실제 자료에 적용하기 전에 모의실험으로 검증하는 과정이 필요하다. 본 연구에서는 강우자료의 발생 원인을 두 가지로 판단하여 두 개의 모집단으로 설정하였다. 두 모집단으로 혼합분포형을 표현하기 위해서는 두 개의 위치매개변수, 두 개의 규모매개변수, 한 개의 혼합비(mixing ratio) 등 총 다섯 개의 매개변수가 필요하다. 모의실험 순서는 다음과 같다.

- [1] 기지의 매개변수($\mu_1, \mu_2, \sigma_1, \sigma_2, \rho$)를 설정하고 발생비율에 맞게 각 모집단별로 자료를 모의발생시켜 하나의 강우자료로 합성한다.
- [2] 합성한 강우자료를 앞서 제안한 방법론을 통해 두 개의 모집단으로 분류하고 매개변수를 추정한다.
- [3] 추정된 매개변수로 모의발생시 기존 모의발생자료와의 상관성 및 경향성을 통하여 방법론을 검증한다.

두 개 모집단의 혼합정도를 나타내는 매개변수 ρ 는 전체자료에서의 모집단의 비율을 의미하며 다른 매개변수와 다르게 하나의 값으로 두 모집단의 비율이 정해진다. 혼합비를 세 개의 경우로

분류한 모의실험 결과는 다음과 같다. CaseI($\rho=0.5$), CaseII($\rho=0.3$), CaseIII($\rho=0.7$)이다.

위의 순서에 따른 Case별 모의실험결과는 Table 1과 Fig. 2에 제시하였으며, 높은 상관성 및 경향성을 보여 제안한 방법론을 실제자료에 적용할 수 있다고 판단하였다.

4. 분포형에 따른 매개변수 추정 및 빈도해석 평가 결과

확률밀도함수를 추정하는 방법에는 매개변수적(parametric)방법과 비매개변수적(non-parametric)방법으로 구분된다. 매개변수적 방법은 사전에 특정 확률밀도함수에 대한 모델을 정해놓고 자료들로부터 모델의 매개변수만 추정하는 방식이다. 본 연구에서 제안하고 있는 혼합확률분포함수가 이에 해당한다. 그러나 현실적으로 특정 확률분포를 사전에 인지하는 것은 쉽지 않다. 따라서 사전 정보나 지식 없이 관측된 자료만으로 확률밀도함수를 추정하는 비매개변수적 핵밀도추정(Kernel density estimation, KDE) 방법을 시각적인 검토 측면에서 대조군으로 사용하였다. 즉, 매개변수적 방법으로 얻은 확률밀도함수와 KDE를 통해 얻은 확률밀도함수가 일치하는 정도를 가지고 간접적으로 분포추정의 정도를 판단하였다. 전주지점의 연최대강우량에 대해 지속시간별로 적용한 결과는 다음과 같다.

우선 지속시간별로 추정된 매개변수를 비교하여 Table 2에 정리하여 나타내었다. 모든 지속시간에 대해서 첫 번째 Gumbel 분포의 규모매개변수가 두 번째 Gumbel 분포에 비해 상대적으로 작게 추정되고 있으며, 단일 Gumbel 분포와 비교해보면 두 번째 Gumbel

Table 1. Estimated Parameters and Their Credible Intervals of Synthetic AMRs for Three Different Mixing Ratios

Known parameter		Case I ($\rho=0.5$)			Case II ($\rho=0.3$)			Case III ($\rho=0.7$)		
		25%	median	75%	25%	median	75%	25%	median	75%
μ_1	30	30.05	30.23	30.40	29.11	29.32	29.54	30.09	30.24	30.40
μ_2	60	60.06	60.31	60.56	59.88	60.07	60.25	59.61	59.96	60.31
σ_1	5	5.11	5.25	5.40	4.63	4.81	4.99	5.26	5.39	5.52
σ_2	7	6.77	6.95	7.15	6.58	6.72	5.87	6.70	6.96	7.22
ρ (Case I, II, III)		0.51	0.52	0.53	0.28	0.29	0.30	0.70	0.71	0.72

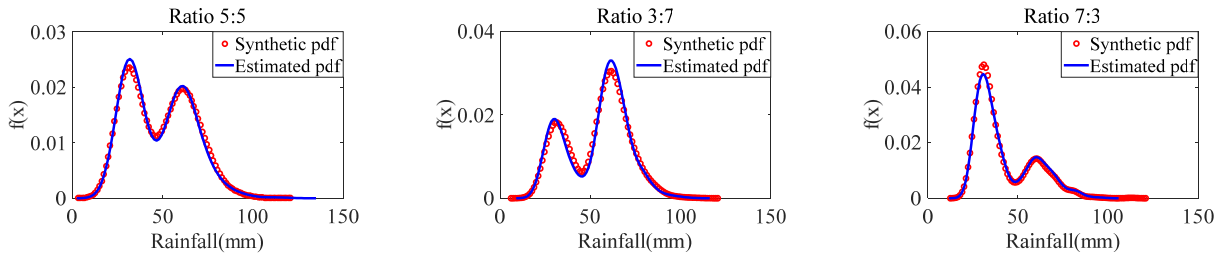


Fig. 2. A graphical Representation of Synthetic Annual Maximum Series for Different Mixing Ratios

Table 2. Comparison of the Estimated Parameters between Univariate and Mixed Gumbel Distributions

Duration	Quartiles & Mean	Mixed Gumbel				Univariate Gumbel		ρ
		Gumbel-1		Gumbel-2		μ_3	σ_3	
		μ_1	σ_1	μ_2	σ_2			
1	25%	26.54	5.75	36.30	9.94	31.94	9.04	0.37
	median	29.22	7.86	40.44	11.59	32.84	9.64	0.56
	75	31.52	9.53	45.42	14.17	33.74	10.31	0.79
	mean	28.99	7.98	42.09	12.98	32.84	9.71	0.56
2	25%	35.26	8.78	50.77	9.70	43.34	12.91	0.35
	median	38.67	10.71	60.26	13.57	44.67	13.72	0.60
	75	41.17	12.84	73.86	16.04	45.98	14.63	0.75
	mean	37.79	11.00	61.84	13.15	44.66	13.81	0.54
3	25%	33.66	8.99	55.49	15.99	51.18	15.21	0.12
	median	41.26	12.83	58.96	17.80	52.69	16.17	0.28
	75	47.49	16.44	64.26	19.80	54.22	17.14	0.54
	mean	40.25	12.87	61.20	18.06	52.69	16.23	0.34
6	25%	34.99	1.69	73.15	21.86	66.57	20.88	0.06
	median	36.08	4.39	76.10	23.80	68.73	22.07	0.09
	75	40.59	11.71	79.16	25.98	70.84	23.32	0.15
	mean	39.46	7.87	76.11	24.17	68.71	22.16	0.14
9	25%	42.81	3.22	84.18	25.20	76.06	23.30	0.07
	median	44.84	5.95	87.69	27.54	78.52	24.61	0.11
	75	48.42	11.19	91.34	30.02	80.86	26.00	0.16
	mean	46.28	8.43	87.80	27.77	78.51	24.71	0.13
12	25%	50.46	4.54	92.57	25.98	84.48	23.80	0.07
	median	53.29	7.29	96.42	28.27	86.98	25.17	0.12
	75	56.91	12.98	100.20	30.91	89.43	26.62	0.17
	mean	54.14	10.01	96.22	28.67	86.96	25.27	0.15
18	25%	55.68	3.53	103.40	28.82	93.46	26.43	0.07
	median	57.74	5.67	107.40	31.34	96.26	27.87	0.11
	75	60.36	9.76	111.60	34.14	98.87	29.45	0.16
	mean	58.40	8.11	107.10	31.75	96.18	27.98	0.13
24	25%	60.79	4.65	111.80	30.40	99.88	28.00	0.09
	median	63.25	6.44	116.40	33.11	102.70	29.53	0.13
	75	65.64	9.98	120.80	36.15	105.60	31.16	0.17
	mean	62.80	8.64	116.20	33.44	102.70	29.62	0.14

분포가 전반적으로 유사성이 크다. 그러나 지속시간이 짧은 1-3시간의 경우에는 단일 Gumbel 분포와 차이가 크게 나타나고 있으며 이는 혼합비 ρ 와도 연관성이 크다. 또한, 지속시간 6시간에서의 혼합비와 첫 번째 Gumbel 분포의 위치, 규모매개변수가 급감하는 것으로 나타나고 있다. 이는 첫 번째 Gumbel 분포에 속하는 자료들의 상대적인 크기와 개수가 줄어들면서 추정의 불확실성도 줄어드는 것으로 판단할 수 있다. 즉, 지속시간이 짧은 1~3시간의 경우

혼합비 ρ 가 34% 이상으로 추정되었으며, 이는 2개의 Gumbel 분포가 거의 대등하게 가중치를 갖는 것으로 볼 수 있다. 이러한 점에서 단일 Gumbel 분포와는 분포의 형태 및 매개변수 측면에서 상이성이 크다고 할 수 있다. 이러한 특징은 Fig. 3에서 나타난 확률밀도함수의 비교 그림에서도 확인될 수 있다. 즉, 지속시간이 짧은 1-3시간의 경우 두 번째 첨두의 위치가 첫 번째 첨두와 상대적으로 가까운 반면 지속시간이 긴 경우에는 두 번째 첨두가 Upper

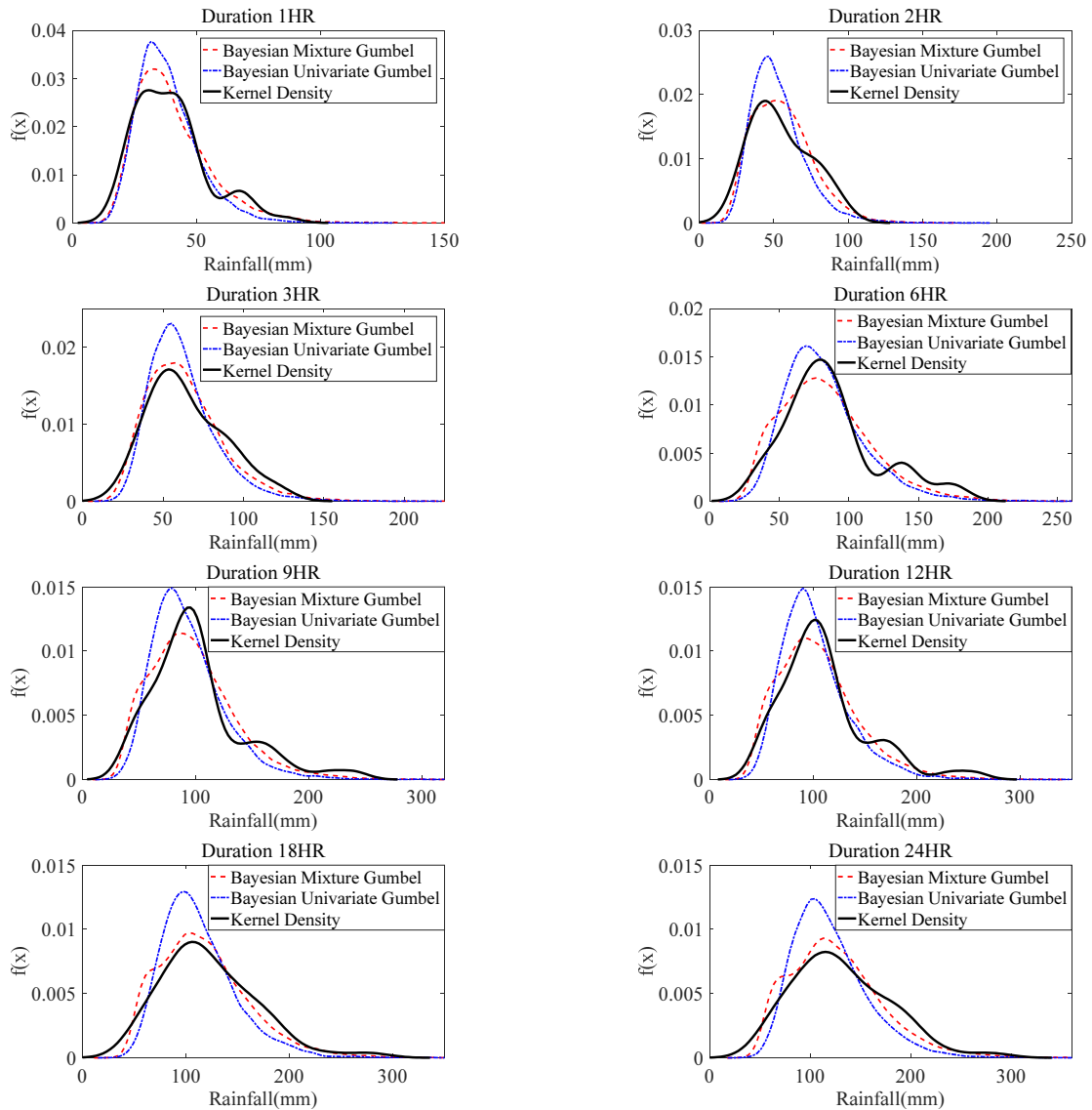


Fig. 3. Comparison of Probability Density Functions between Univariate and Mixed Gumbel Distributions at Jeonju Station. All the Parameters for the Univariate and Mixed Gumbel Distribution were Estimated Within a Bayesian Framework

Tail 쪽에 위치하고 있는 점을 확인할 수 있다.

지속시간별로 단일 Gumbel 분포와 혼합 Gumbel 분포의 특성을 자세히 살펴보면 다음과 같다. 지속시간 1, 6, 9, 12시간에 대한 자료특성이 두 개의 첨두가 뚜렷하고, 지속시간 2, 3, 18, 24시간에 대한 자료특성에서는 전자에 비해 뚜렷하지 않다. 모든 지속시간에 대한 자료특성들이 꼬리부분에서의 첨두 형태를 가지고, 단일 Gumbel 분포 특성상 꼬리부분의 확률거동을 표현하기 힘들다. 그 결과 빈도해석시 혼합 Gumbel에 비해 과소추정되는 경향을 보인다. 중앙부분 첨두, 꼬리부분 첨두로 구분하여 분석해 볼 때 지속시간 6, 9, 12시간에 대한 중앙첨두의 확률이 혼합 Gumbel 분포형에서는 실제자료보다 다소 작게 추정되고, 단일 Gumbel

분포형에서는 다소 크게 추정되었다. 이외의 모든 지속시간에 대해서는 혼합 Gumbel 분포형이 자료의 분포 특성을 보다 실질적으로 묘사하고 있는 것으로 확인할 수 있다.

최적분포형을 선정하는 기준에는 AIC (Akaike Information Criterion), BIC (Bayesian Information Criterion), DIC (Deviance Information Criterion) 등이 있다. 본 연구에서는 자료의 개수(n), 매개변수의 개수(k), 우도(\hat{L})를 고려하여 우선순위를 선정하는 BIC, 즉 Eq. (6)를 사용하였다. \hat{L} 은 모델의 최우도값(the maximized value of the likelihood function)이며, Eq. (7)로 나타내며, $\hat{\theta}$ 는 최우도함수의 매개변수를 나타낸다.

Table 3. Estimations of Design Rainfalls of Univariate and Mixed Gumbel Distribution and Their Comparison Corresponding to Different Durations

Duration	1-Hour		2-Hour		3-Hour		6-Hour		9-Hour		12-Hour		18-Hour		24-Hour	
	Mixed (mm)	Univ. (mm)	Mixed (mm)	Univ. (mm)	Mixed (mm)	Univ. (mm)	Mixed (mm)	Univ. (mm)	Mixed (mm)	Univ. (mm)	Mixed (mm)	Univ. (mm)	Mixed (mm)	Univ. (mm)	Mixed (mm)	Univ. (mm)
2	37	36	54	50	60	59	80	77	92	88	100	96	112	106	121	114
5	52	47	71	65	81	77	108	102	125	116	134	125	150	138	161	147
10	61	55	82	76	95	89	127	119	146	134	156	144	174	159	186	169
30	76	66	97	91	115	108	154	144	178	162	189	172	210	191	224	203
50	82	71	104	99	124	116	167	155	192	175	204	186	226	205	242	218
100	91	78	113	108	137	127	184	171	212	192	224	203	249	225	265	239
200	100	84	122	118	149	139	201	186	231	209	244	221	271	244	288	260
500	112	93	134	130	166	154	223	206	256	232	270	244	300	270	319	287
BIC	445	454	477	496	512	516	543	559	560	576	564	580	571	593	574	601

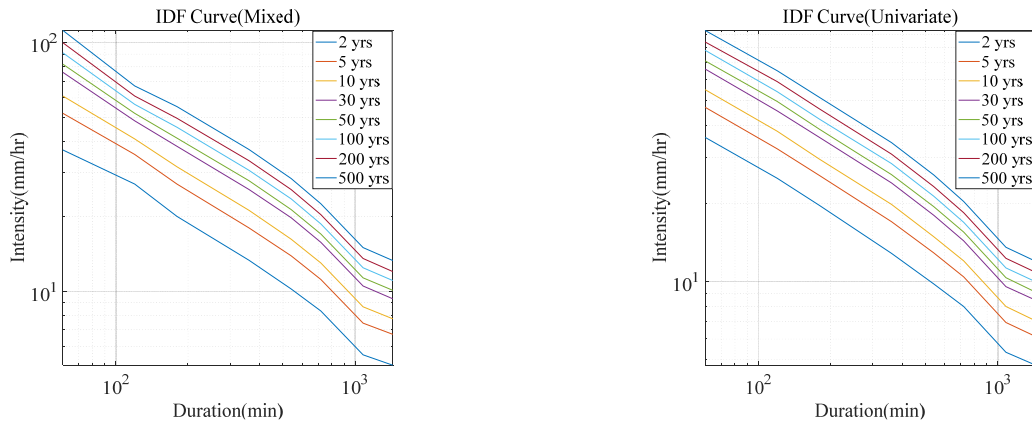


Fig. 4. IDF Curves of Univariate and Mixed Gumbel Distributions

$$BIC = \ln(n)k - 2\ln(\hat{L}) \tag{6}$$

$$\hat{L} = p(x|\hat{\theta}) \tag{7}$$

적은 매개변수와 큰 우도값을 토대로 BIC값이 작게 산정될 때 통계적으로 우수한 분포형이라 판단할 수 있다. 특히 과적합 (over-fitting) 문제에 있어서 매개변수의 개수에 대한 벌점 (Penalty)이 크게 작용하므로 혼합분포형이 전반적으로 불리하다. 그러나 Table 3에 정리한 것처럼 혼합분포형의 우도(likelihood)가 단일분포형에 비해 개선되는 효과가 매개변수 개수의 증가에 비해 크게 나타나고 있다. 따라서 모든 경우에 있어서 혼합 Gumbel 분포의 BIC값이 더 작게 산정되고 있다. 즉, 통계적 측면에서 혼합 Gumbel 분포형이 더 적합하다고 평가될 수 있다. 또한, 동일 지속시간에서 낮은 재현기간일 경우 강우강도가 혼합분포와 단일분포간의 차이가 작지만, 높은 재현기간으로 갈수록 그 차이가 커지는

것을 Table 3의 재현기간별 확률강수량과 Fig. 4의 Intensity-Duration-Frequency (IDF) 곡선에서 확인할 수 있다.

5. 결론

우리나라에서는 빈도해석 시 일반적으로 확률가중모멘트법을 적용한 Gumbel 분포형을 사용하고 있지만, 여러 발생요인으로 이루어진 자료의 모집단추정은 혼합분포형이 적합하다는 많은 연구사례가 있다. 본 연구에서는 불확실성에 대한 신뢰성을 확보하는데 우수한 성능을 발휘하는 Bayesian 기법과 연계한 빈도해석 기법을 개발하였다. 본 연구에서는 전주지역 강우자료에 대해서 모형의 적합성을 평가하였으며, BIC를 통해 기존 방법론과 정량적인 비교 검토를 수행하였다. 본 연구를 통하여 도출된 결론은 다음과 같다.

- (1) 일반적인 강우빈도해석법은 연최대치강우가 단일 모집단을 이룬다고 가정하지만 전주지역의 강우자료를 평가해 보면 하나의 특성을 가진 모집단이라고 보기 어렵다. 즉, 발생 비율면에서 차이가 있지만, 최소 2개 이상의 특성을 가진 모집단으로 구분되고 있다. 이러한 자료를 통계적으로 하나의 분포형으로만 적합시키는데 한계가 있으며 본 연구에서 제시한 혼합 Gumbel 분포형은 신뢰할 만한 추정치를 제공해주고 있다.
- (2) 혼합 Gumbel 혼합분포형이 단일 Gumbel 분포형에 비해 매개변수가 많아 BIC를 비교할 때 불리하게 작용한다. 그러나 혼합분포를 통한 자료의 적합 특성을 나타내는 우도값의 증가가 매개변수 증가로 인한 벌점(Penalty)보다 상대적인 우위를 보이고 있어 모든 경우에 있어 혼합 Gumbel 분포형의 적합성이 우수한 것으로 평가되었다. 즉, 단일 Gumbel 분포형보다 극치 강우량 자료를 나타내는데 적합하다고 판단할 수 있다.
- (3) 모든 지속시간에 대한 극치빈도해석에서 혼합분포가 단일분포보다 큰 값을 가진다. 이는 현재의 설계기준에 의한 설계 시 극치 확률강수량의 과소추정으로 이어져 수공구조물의 안전도 측면에서 불리할 것으로 판단된다. 제안된 혼합 Gumbel 분포형은 단일 Gumbel 분포형이 반영하지 못하는 꼬리부분의 신뢰성 있는 추정이 가능하였으며, 이러한 점에서 설계강수량 추정 시 보다 신뢰성 있는 접근이 가능할 것으로 판단된다.

본 연구에 적용된 Bayesian 기법과 연계한 혼합분포 매개변수 추정 및 빈도해석 기법은 단일 분포형에 반영되지 못하는 꼬리(tail) 부분의 이중침투 거동을 효과적으로 모의할 수 있는 방법으로서 본 연구에서는 장마와 태풍으로 인한 우리나라 여름철 강우의 강우지속시간 및 강우강도의 변화에 따른 이중침투 특징을 고려한 빈도해석 시 신뢰성있는 결과를 도출하는데 활용하였다. 결과적으로 기존의 단일분포를 통한 설계강수량 추정 시 보다 신뢰성있는 접근이 가능하게 됐으며, 우리나라 극치강우자료 분석 시 기존 단일 분포기반의 빈도해석기법 대안으로 적용이 가능할 것으로 판단된다. 이러한 점에서 향후 연구로써 서로 다른 분포형을 모형에 적용하여 최적화된 모형을 구축하고 강우의 시간에 따른 변동성을 고려할 수 있는 비정상성 빈도분석을 수행하여 과거부터 현재에 이르는 여름철 장마 및 태풍에 따른 강우변동성을 고려한 보다 신뢰성있는 설계강수량 추정에 활용할 수 있을 것으로 판단된다.

감사의 글

본 연구는 국토교통부 물관리연구사업의 연구비지원(17AWMP-B079625-04)에 의해 수행되었습니다.

References

Gelman, A., Chew, G. L. and Shnaidman, M. (2004). "Bayesian analysis of serial dilution assays." *Biometrics*, Vol. 60, No. 2, pp. 407-417.

George, E. I. and McCulloch, R. E. (1993). "Variable selection via Gibbs sampling." *Journal of the American Statistical Association*, Vol. 88, No. 423, pp. 881-889.

Heo, J. H., Kim, G. D. and Han, J. H. (1999). "Derivation of rainfall Intensity-Duration-Frequency equation based on the appropriate probability distribution." *Journal of Korea Water Resources Association*, Vol. 32, No. 3, pp. 247-254 (in Korean).

Ho, C. H., Lee, J. Y., Ahn, M. H. and Lee, H. S. (2003). "A sudden change in summer rainfall characteristics in Korea during the late 1970s." *International Journal of Climatology*, Vol. 23, No. 1, pp. 117-128.

Kwon, H. H. and Moon, Y. I. (2006). "Improvement of overtopping risk evaluations using probabilistic concepts for existing dams." *Stochastic Environmental Research and Risk Assessment*, Vol. 20, No. 4, p. 223.

Kwon, H. H., Brown, C. and Lall, U. (2008a). "Climate informed flood frequency analysis and prediction in Montana using hierarchical Bayesian modeling." *Geophysical Research Letters*, Vol. 35, No. 5.

Kwon, H. H., Khalil, A. F. and Siegfried, T. (2008b). "Analysis of extreme summer rainfall using climate teleconnections and typhoon characteristics in South Korea." *JAWRA Journal of the American Water Resources Association*, Vol. 44, No. 2, pp. 436-448.

Kwon, H. H., Kim, J. G., Lee, J. S. and Na, B. K. (2012). "Uncertainty assessment of single event rainfall-runoff model using bayesian model." *Journal of Korea Water Resources Association*, Vol. 45, No. 5, pp. 505-516 (in Korean).

Kwon, H. H., Moon, Y. I. and Khalil, A. F. (2007). "Nonparametric Monte Carlo simulation for flood frequency curve derivation: an application to a Korean watershed." *JAWRA Journal of the American Water Resources Association*, Vol. 43, No. 5, pp. 1316-1328.

Lee, J. J., Lee, J. S., Kim, B. I. and Park, J. Y. (2000). "Derivation of probable rainfall intensity formula of individual zone based on the representative probability distribution." *Journal of Korea Water Resources Association*, Vol. 33, No. S1, pp. 124-129 (in Korean).

Lee, J. J., Lee, S. W. and Kwak, C. J. (2009). "Application of jackknife method for determination of representative probability distribution of annual maximum rainfall." *Journal of Korea Water Resources Association*, Vol. 42, No. 10, pp. 857-866 (in Korean).

Park, C. Y., Moon, J. Y., Cha, E. J., Yun, W. T. and Choi, Y. E. (2008). "Recent changes in summer precipitation characteristics over South Korea." *Journal of the Korean Geographical Society*, Vol. 43, No. 3, pp. 324-336.

Redner, R. A. and Walker, H. F. (1984). "Mixture densities, maximum likelihood and the EM algorithm." *SIAM Review*, Vol. 26, No. 2, pp. 195-239.

Shin, J. Y. and Lee, T. (2014). "Parameter estimation of the mixture

normal distribution for hydro-meteorological variables using meta-heuristic maximum likelihood.” *Journal of Korean Society of Hazard Mitigation*, Vol. 14, No. 4, pp. 93-99 (in Korean).

Yoon, P. Y., Kim, T. W., Yang, J. S. and Lee, S. O. (2012). “Estimating

quantiles of extreme rainfall using a mixed Gumbel distribution model.” *Journal of Korea Water Resources Association*, Vol. 45, No. 3, pp. 263-274 (in Korean).