



## A development of nonstationary rainfall frequency analysis model based on mixture distribution

Choi, Hong-Geun<sup>a</sup> · Kwon, Hyun-Han<sup>b</sup> · Park, Moon-Hyung<sup>c\*</sup>

<sup>a</sup>Ph.D Candidate, Department of Civil and Environmental Engineering, Sejong University, Seoul, Korea

<sup>b</sup>Professor, Department of Civil and Environmental Engineering, Sejong University, Seoul, Korea

<sup>c</sup>Senior Researcher, Korea Institute of Construction Technology, Goyang, Korea

Paper number: 19-062

Received: 7 August 2019; Revised: 7 October 2019; Accepted: 7 October 2019

### Abstract

It has been well recognized that extreme rainfall process often features a nonstationary behavior, which may not be effectively modeled within a stationary frequency modeling framework. Moreover, extreme rainfall events are often described by a two (or more)-component mixture distribution which can be attributed to the distinct rainfall patterns associated with summer monsoons and tropical cyclones. In this perspective, this study explores a Mixture Distribution based Nonstationary Frequency (MDNF) model in a changing rainfall patterns within a Bayesian framework. Subsequently, the MDNF model can effectively account for the time-varying moments (e.g. location parameter) of the Gumbel distribution in a two (or more)-component mixture distribution. The performance of the MDNF model was evaluated by various statistical measures, compared with frequency model based on both stationary and nonstationary mixture distributions. A comparison of the results highlighted that the MDNF model substantially improved the overall performance, confirming the assumption that the extreme rainfall patterns might have a distinct nonstationarity.

**Keywords:** Two (or more)-component mixture distribution, Bayesian framework, Time-varying moments

## 혼합분포 기반 비정상성 강우 빈도해석 기법 개발

최홍근<sup>a</sup> · 권현한<sup>b</sup> · 박문형<sup>c\*</sup>

<sup>a</sup>세종대학교 건설환경공학과 박사과정, <sup>b</sup>세종대학교 건설환경공학과 부교수, <sup>c</sup>한국건설기술연구원 수석연구원

### 요 지

극치 강우 자료는 정상성 빈도모델에서 효과적으로 구현되지 않는 비정상성 거동을 종종 보인다. 또한, 극치 사상의 확률밀도함수는 여름 장마와 태풍 등의 서로 다른 강우 패턴에 의해 2개 이상의 첨두를 가지는 혼합분포형태이다. 이러한 강우 패턴의 변화에 대해 Bayesian 이론을 활용한 비정상성 혼합분포(mixture distribution based nonstationary frequency, MDNF)모델을 제안하였다. 2개의 Gumbel 분포형이 혼합된 MDNF 모델은 Gumbel 분포형 매개변수 중 하나인 위치매개변수의 시변성을 효과적으로 설명한다. 제안한 모델의 성능평가를 위해 정상성 혼합분포모델과의 다양한 통계치 결과를 비교하였다. 정상성 혼합분포모델보다 전반적으로 향상된 성능을 보여주는 MDNF 모델을 통해 극치 강우 패턴이 비정상성을 보인다는 가정을 확인할 수 있다.

**핵심용어:** 2요소 혼합분포형, 베이시안 이론, 시변성

\*Corresponding Author. Tel: +82-31-995-0868  
E-mail: moon@kict.re.kr (M.-H. Park)

## 1. 서론

우리나라의 강우 발생은 계절적으로 여름철에 집중되며 극치강우는 이 기간의 장마와 태풍에 의한 것이 대부분이다. 최근 들어 우리나라 여름철 강수특성이 과거와 많이 달라지고 있다는 연구결과들이 제시되고 있다(Lee and Kwon, 2004; Park *et al.*, 2008). 장마와 태풍으로 인한 강수량은 강우지속 시간 및 강우강도 측면에서 서로 다른 특징을 가지고 있으며(Kwon *et al.*, 2008b), 이러한 특징을 고려한 강우분석과 관련된 연구가 다수 진행되었다. 강우빈도분석 시 자료가 통계적으로 단일한 모집단이고 정상성이라는 가정에 따라 이루어진다. 우리나라의 여름철 극치강수량은 지속시간이 길어지는 경우 이중첨두(multi mode)를 가지는 경우가 빈번하여(Ho *et al.*, 2003), 단일첨두의 분포형으로 표현하는데 한계를 보인다.

우리나라에서 강우 빈도분석 시 확률분포형 선정 연구가 다수 진행되었고, 실무적인 관점에서 확률강우량도 작성을 위해 매개변수 추정방법으로 확률가중모멘트법을 채택하며 Gumbel 분포형을 추천하고 있다(Ministry of Construction & Transportation, 2000). 단일분포형을 사용하는 국내연구 사례는 다음과 같다. 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 확률 분포형으로 선정하였다. 그러나 극치강우자료의 확률밀도함수는 혼합분포의 형태를 보이는 경우가 빈번하고 이로 인해 강우빈도분석 시 과대/과소 추정되는 문제가 발생할 수 있다(El Adlouni *et al.*, 2008; Smith, 1987). 앞서 언급한 매개변수적 확률분포형을 사용하는 방법과는 다르게 핵밀도함수(kernel density function)를 이용한 비매개변수적 방법에 대한 연구도 다수 진행되었다(Kwon *et al.*, 2007). 댐위험도 분석 시 분포형 선정이 요구되지 않는 비매개변수적 방법을 활용하여 다중첨두를 갖는 경우에도 유연하게 확률분포 추정이 가능하다고 제시했고, 기존 매개변수적 방법에서 댐 위험도가 비현실적으로 추정되는 반면, 비매개변수적 방법은 합리적인 댐 위험도 평가 결과를 제공하였다(Kwon and Moon, 2006). 그러나 핵밀도함수를 이용한 방법은 분포형의 꼬리(tail)부분 외삽 시 상대적으로 수렴이 빠른 문제로 빈도분석 시 특정빈도 이상에서는 확률강수량의 변화가 미미하고 Quantile 함수가 존재하지 않아 수치적인 접근만이 가능하다는 단점이 있다. 이러한 점에서 2개 이상의 매개변수적 확률 분포형이 혼합된 빈도분석모델의 필요성이 대두되고 있다.

혼합분포형과 관련된 국내연구 사례에서는, 강우자료를 집단별로 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). 본 연구에서는 강우자료를 특정사건 또는 기간을 조사하여 분류하지 않았으며, Bayesian MCMC(Markov Chain Monte Carlo) 기법을 적용하여 매개변수 추정 및 강우빈도분석을 수행하였다. 제안한 모델의 적합성을 판단하기 위하여 모의실험을 실시하였고, 매개변수 추정의 정확도를 비교하였다. 최종적으로 본 연구는 연최대일강수량 자료의 핵밀도함수가 다중첨두의 특성을 가지는 지점에 대해 비정상성 혼합모델을 적용할 경우 정상성 모델에 비해 어떠한 점이 개선되는지에 대해 중점을 두고 있다. 이를 위해 정상성 및 비정상성 혼합모델의 통계치를 비교하였다.

## 2. 대상자료

본 연구에서는 기상청 산하 관측소에서 강우자료 기간이 40년 이상의 지점의 일 단위 시계열 자료로부터 극치 자료를 추출하였고, 극치 자료의 핵밀도함수가 2개 이상의 첨두를 가지는 남해, 부산, 영덕 지점을 선정하였다. 채택한 지점의 극치 강우 시계열의 1차 선형 회귀함수, 핵밀도함수는 Fig.1과 같다. 남해지점의 1차 선형회귀계수는  $+0.99 \text{ mm/yr}$ , 부산과 영덕 지점에서는 기온기가 통계적으로 유의하지 않은 것으로 판단되었다. 강우가 발생하는 메커니즘은 국지성, 장마, 태풍 등 여러 경우가 있고, 이러한 강우자료를 하나의 집합으로 분석하는 점에 대해 문제가 있다고 판단하였다. 본 연구에서는 극치 자료 안의 비슷한 성질을 가지는 자료들을 각 집단으로 분류하고 집단별 경향성을 분석하는 모델을 제시하였다.

## 3. 비정상성 혼합분포모델 매개변수 추정기법

본 절에서는 이전 연구에서 다루었던 Bayesian 기법 기반의 정상성 혼합분포모델의 매개변수 추정기법(Choi *et al.*, 2018)에 이은 비정상성 기법에 관해 서술하였고, 제안한 방법론의 적합성을 판단하기 위해 모의실험을 하였다. Bayesian 기법을 통한 매개변수 추정기법에서는 매개변수를 하나의 확률변수로 취급한다. 즉, 매개변수가 단일 값이 아닌 확률 분포의 형태로 부여되며 최종적으로 매개변수의 사후 분포(posterior distribution)를 추정하는 데 목적을 두며, Bayes' 정리(Eq. (1))를 기반으로 한다.

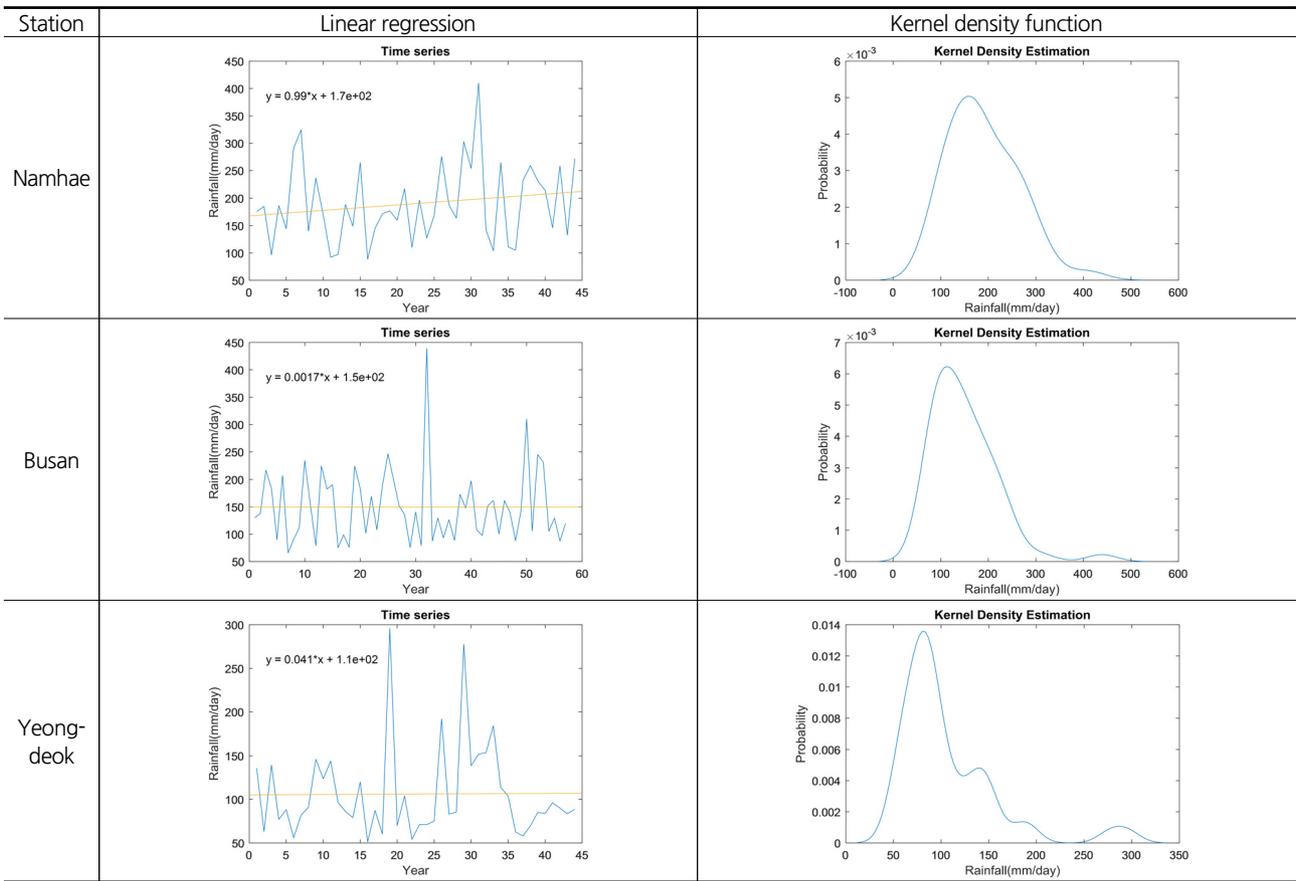


Fig. 1. A graphical representation of annual daily maximum rainfall data at Namhae, Busan and Yeongdeok stations. Extreme time series data with linear regression are illustrated in the left panel while kernel density functions are displayed in the right panel

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

여기서,  $p(\boldsymbol{\theta} \mid \mathbf{x})$ 은 사후 분포로, 앞서 제시한 경우자료가 Gumbel 혼합분포를 따르며, 각 분포의 위치매개변수(location parameter)가 선형 경향성을 가진다고 가정하며,  $\boldsymbol{\theta} = [\mu_1(t), \sigma_1, \mu_2(t), \sigma_2, \rho]$ 는 Gumbel 분포에 대한 매개변수들의 집합을 나타낸다.  $p(\mathbf{x})$ 는 관측자료  $\mathbf{x}$ 의 주변분포(marginal distribution),  $p(\boldsymbol{\theta})$ 는 매개변수들의 사전분포를,  $p(\mathbf{x} \mid \boldsymbol{\theta})$ 는 극치 자료  $\mathbf{x}$ 의 우도함수를 의미하며 Eq. (2)로 표현된다.  $\rho$ 는 혼합분포의 혼합비(mixing ratio)를 나타낸다.

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

여기서,  $N$ 은 강수시계열의 자료연수,  $\mu_1(t) = a + b \cdot t$ ,  $\mu_2(t) = c + d \cdot t$ 를 나타낸다.  $\mu_1(t)$ 과  $\mu_2(t)$ 를 통해 시간에 따른 자료의 선형성을 고려할 수 있다. 매개변수들의 사전분포

( $p(\boldsymbol{\theta})$ )는 사전정보를 기반으로 하는 Informative 사전분포와 사전정보에 크게 의존하지 않는 Noninformative 사전분포가 있으며, Noninformative 사전분포로는 정규분포, 균일분포, 지수분포가 대표적으로 활용된다(Gelman *et al.*, 2004). 사전 분포를 결정하는데 공액분포(conjugate distribution), 즉 사전 분포와 사후분포가 동일하게 결정되는 사전분포를 활용하는 것이 매개변수 추정에 안정성을 도모할 수 있으나, 공액분포 추정이 어려운 경우 비공액분포를 사용하는 것이 일반적이다(Gelman *et al.*, 2004). 본 연구에서는 각 매개변수들의 사전분포로서 정규분포와 Gamma 분포, Dirichlet 분포를 사용하였다. 다시 말해, 위치매개변수에 대한 사전분포로서 분산이 큰 정규분포를 활용하였으며, 규모매개변수에 대해서는 음의 값을 방지하기 위하여 Gamma 분포를 활용하였다. 또한, 각 모집단별 혼합비를 의미하는 매개변수는 총합이 '1'인 것을 이용하여 Dirichlet 분포를 채택하였다. 본 연구에서 추정이 필요한 7개의 매개변수의 수에 비해 채택된 기상청 3개 지점의 자료 수가 충분히 큰 점을 고려할 때, Noninformative 사전분포를 통한 추정 시 큰 무리가 없을 것으로 판단하였다.

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

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

$$p(\sigma_{1,2}) \sim \text{Gamma}(k_{\sigma_{1,2}}, s_{\sigma_{1,2}}) \quad (3c)$$

$$p(\rho) \sim \text{Dirichlet}(K, \alpha_{i=1:K}) \quad (4)$$

Eqs. (3a) and (3b)의  $a, b, c, d$  중  $a, c$ 는 위치매개변수의 절편을 의미하며,  $b, d$ 는 위치매개변수의 기울기를 의미한다. 매개변수  $a, b, c, d$ 는 각각  $\mu_{a,b,c,d}, \sigma_{a,b,c,d}^2$ 를 매개변수로 하는 정규분포를 따르고, Eq. (3c)에서 규모매개변수  $\sigma_1, \sigma_2$ 는 각각  $k_{\sigma_1}, s_{\sigma_1}$ 와  $k_{\sigma_2}, s_{\sigma_2}$ 를 갖는 Gamma 분포를 따르며, 혼합비  $\rho$ 는 Dirichlet 분포를 따른다. 혼합분포모형을 구성하는 2개의 Gumbel 분포형의 위치매개변수가 비정상성을 보인다는 가정하에 확률론적 추론(statistical inference)을 수행했으며, 7개의 매개변수에 대한 결합 확률밀도함수(joint distribution)는 Eq. (5)와 같다.

$$p(a, c, \sigma_1, b, d, \sigma_2, \rho) \propto 1 \quad (5)$$

Eqs. (3) and (4)에서 정의되는 매개변수들의 사전분포들을 Eq. (2)에 대입시킴으로써 매개변수들의 사후분포를 Eq. (6)과 같이 추정할 수 있다.

$$p(\Theta | x) \propto \left[ \rho \prod_{n=1}^N \text{Gumbel}(x_n | \mu_1(t), \sigma_1) \cdot N(\mu_a, \sigma_a^2) \cdot N(\mu_b, \sigma_b^2) \cdot \Gamma(k_{\sigma_1}, s_{\sigma_1}) \right. \\ \left. + (1-\rho) \prod_{n=1}^N \text{Gumbel}(x_n | \mu_2(t), \sigma_2) \cdot N(\mu_c, \sigma_c^2) \cdot N(\mu_d, \sigma_d^2) \cdot \Gamma(k_{\sigma_2}, s_{\sigma_2}) \right] \quad (6)$$

Eq. (6)에서 모든 매개변수에 대한 적분을 통해 직접 추정하는 것은 불가능하다. 따라서, 깃스표본법을 이용한 Bayesian MCMC 기법을 도입하여 매개변수들의 사후분포를 추정함으로써 문제점을 해결하였다. 깃스표본법은 원하는 다변량 확률분포에서 IID (independent identically distributed) 표본을 추출하는 것이 복잡할 때 사용하는 방법이다. 현재 시점의 매개변수들의 값은 정확하게 바로 직전의 추출된 값들이 사용되게 되며, 조건부 분포에서 추출된 난수들이 안정 상태에 도달하는 것이 주어진 다변량 확률분포를 정확히 따르는 난수가 되는 척도가 되며 깃스표본법을 구현하는데 가장 중요한 부분

이 된다(Kwon *et al.*, 2012). 일반적인 Monte Carlo 기법은 확률변수들 간의 독립성을 가정으로 이루어지는 샘플링 방법이라면, MCMC 기법은 다변량에 대해서 종속성을 기준으로 조건부 샘플링이 가능한 방법이라 할 수 있다. Bayesian MCMC 기법은 사후분포 추정 시 다변량에 대한 복잡한 적분을 위해서 적용되는 수치해석 기법으로 매개변수를 추정하는 수단으로 활용되고 있다(Kwon *et al.*, 2008a).

### 3.1 모의실험

비정상성 혼합 Gumbel 분포 빈도모형을 관측자료에 적용하기 전에 모의실험으로 검증하는 과정이 필요하다. 본 연구에서는 강우자료의 발생 원인을 두 가지로 판단하여 두 모집단의 매개변수를 임의로 설정하였다. 혼합분포형을 표현하기 위해서는 위치매개변수( $\mu_1, \mu_2$ )의 경향성을 고려하기 위하여  $a, b, c, d$  등 4개의 회귀계수들과 2개의 규모매개변수  $\sigma_1, \sigma_2$ , 1개의 혼합비(mixing ratio)  $\rho$  등 총 7개의 매개변수가 필요하다. 모의실험 순서는 다음과 같다.

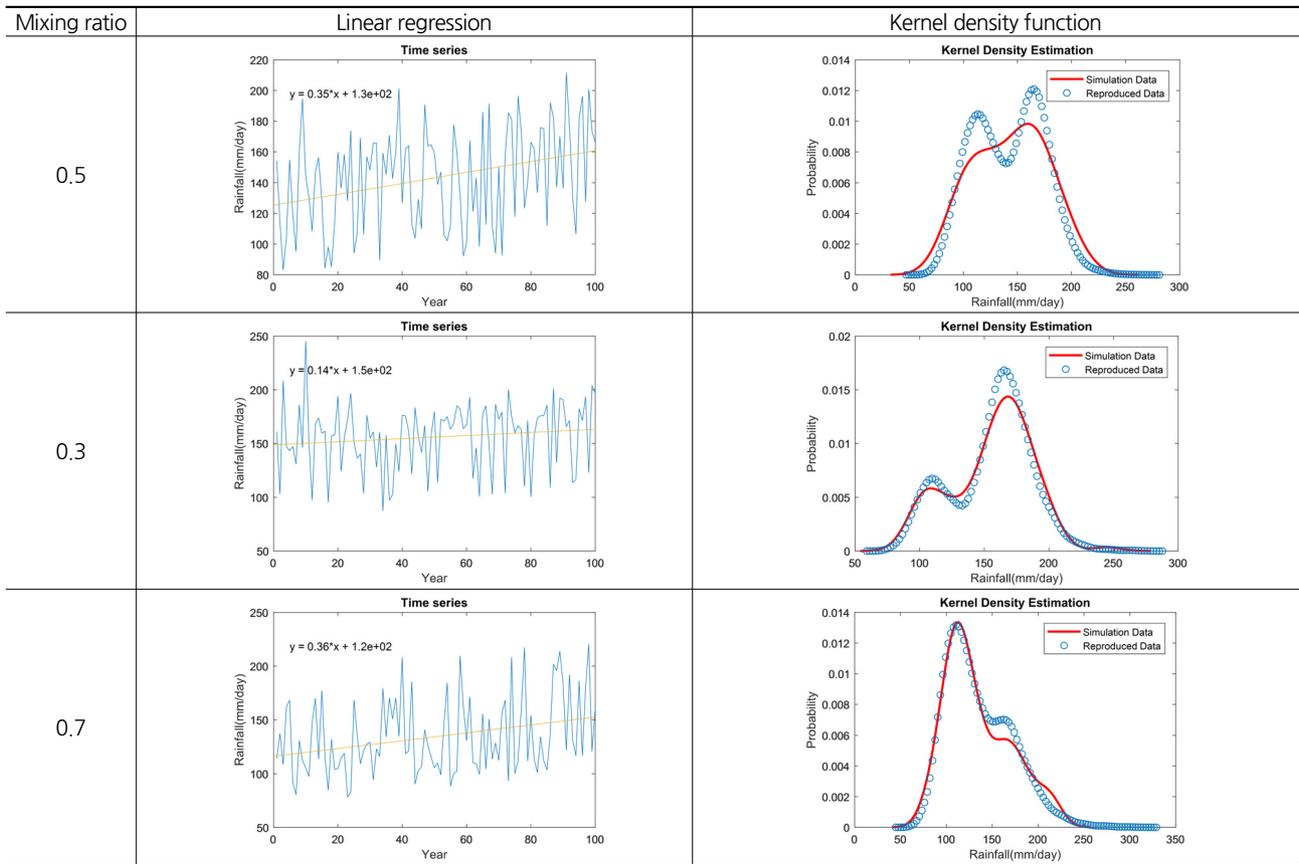
- (1) 매개변수( $a, b, c, d, \sigma_1, \sigma_2, \rho$ )를 기지의 값으로 가정하고 모의발생시켜 하나의 강우자료로 합성한다.
- (2) 합성한 강우자료를 앞서 제안한 방법론을 통해 두 개의 모집단으로 분류하고 매개변수를 추정한다.
- (3) 추정된 매개변수로 모의발생시 기존 모의발생 자료와의 상관성 및 경향성을 통하여 방법론을 검증한다.

두 개 모집단의 구성 비율을 나타내는 매개변수  $\rho$ 는 전체자료에서 한 모집단의 비율을 의미하며 나머지 모집단의 비율은  $(1-\rho)$  값으로 정의된다. 혼합비를 크게 세 가지 경우로 가정한 모의실험 결과는 다음과 같다. Case I ( $\rho=0.5$ ), Case II ( $\rho=0.3$ ), Case III ( $\rho=0.7$ )이다.

위의 순서에 따른 Case별 모의실험결과는 Table 1과 Fig. 2에 제시하였다. Table 1 및 Fig. 2에서 제시된 바와 같이 추정된 매개변수의 불확실성 구간 안에 가정된 두 개의 모집단의 매개변수가 포함되는 것을 확인할 수 있으며, 중앙값(median)을 기준으로 평가해보면 기지의 매개변수와 유사한 결과를 얻을 수 있었다. 실제 기지의 매개변수와 다소 오차가 발생하는 요인은 한정된 샘플(100개)로 인한 샘플링 오차에 기인하는 것으로 판단되며, 전반적으로 제안된 모형의 적합성을 확인할 수 있었다. 두 개 모집단에 가정된 경향성과 적용된 혼합비에 따라 전체 시계열이 가지는 경향성이 다르게 된다. 모의실험 결과를 토대로 실제 관측자료에 대한 적용성을 추가로 검토하였다.

**Table 1.** Estimated parameters and their credible intervals of synthetic ADMRs for three different mixing ratios

Known parameter	Case I (p = 0.5)			Case II (p = 0.3)			Case III (p = 0.7)				
	2.5%	median (mean)	97.5%	2.5%	median (mean)	97.5%	2.5%	median (mean)	97.5%		
$\mu_1$	$a$	100	93.0	<b>102.5 (102.7)</b>	113.0	86.3	<b>97.5 (98.4)</b>	114.7	79.7	<b>97.5 (96.5)</b>	108.3
	$b$	0.15	0.05	<b>0.17 (0.17)</b>	0.29	0.03	<b>0.19 (0.19)</b>	0.35	0.08	<b>0.24 (0.24)</b>	0.41
$\mu_2$	$c$	150	141.2	<b>151.7 (151.4)</b>	159.6	142.3	<b>150.3 (150.4)</b>	158.2	112.7	<b>157.5 (152.2)</b>	176.1
	$d$	0.30	0.13	<b>0.25 (0.25)</b>	0.39	0.15	<b>0.27 (0.27)</b>	0.39	0.07	<b>0.29 (0.30)</b>	0.54
$\sigma_1$	15	13.4	<b>18.9 (19.2)</b>	26.4	7.9	<b>13.2 (14.6)</b>	28.0	9.2	<b>18.0 (17.7)</b>	25.1	
$\sigma_2$	15	6.5	<b>10.5 (10.6)</b>	15.3	9.3	<b>13.4 (13.3)</b>	16.9	8.0	<b>17.5 (18.0)</b>	30.9	
$\rho$		0.42	<b>0.57 (0.57)</b>	0.73	0.18	<b>0.28 (0.30)</b>	0.47	0.30	<b>0.76 (0.71)</b>	0.92	



**Fig. 2.** A graphical representation of synthetic annual daily maximum series for different mixing ratios

#### 4. 정상성 및 비정상성 혼합모델의 매개변수 추정결과

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 인지하

기는 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 핵밀도함수가 일치하는 정도에 따라 간접적으로 추정의 정확도를 판단하였다. 남해, 부산, 영덕지점의 연최대강우량에 대한 확률밀도함수의 매개변수 추정 결과는 Tables 2, 3 and 4와 Fig. 3에 나타내었다.

**Table 2.** Comparison of the estimated parameters between stationary and nonstationary mixture models at Namhae station

Namhae station												
Percentile (%)/ Mean	Stationary mixture model					Nonstationary mixture model						
	Gumbel-1		Gumbel-2		$\rho$	Gumbel-1			Gumbel-2		$\rho$	
	$\mu_1$	$\sigma_1$	$\mu_2$	$\sigma_2$		$\mu_1$		$\sigma_1$	$\mu_2$			$\sigma_2$
						$a$	$b$		$c$	$d$		
2.5	74.6	14.4	149.6	16.0	0.02	67.6	0.05	7.8	122.5	0.43	19.1	0.08
<b>50</b>	<b>136.3</b>	<b>44.1</b>	<b>206.6</b>	<b>48.3</b>	<b>0.64</b>	<b>99.5</b>	<b>0.66</b>	<b>26.2</b>	<b>162.4</b>	<b>1.48</b>	<b>44.1</b>	<b>0.37</b>
97.5	164.6	68.5	347.9	76.4	0.98	145.6	1.58	61.6	234.5	2.84	66.0	0.94
mean	<b>131.7</b>	<b>43.1</b>	<b>217.7</b>	<b>46.9</b>	<b>0.58</b>	<b>102.3</b>	<b>0.68</b>	<b>29.3</b>	<b>165.7</b>	<b>1.52</b>	<b>44.1</b>	<b>0.42</b>

**Table 3.** Comparison of the estimated parameters between stationary and nonstationary mixture models at Busan station

Busan station												
Percentile (%)/ Mean	Stationary mixture model					Nonstationary mixture model						
	Gumbel-1		Gumbel-2		$\rho$	Gumbel-1			Gumbel-2		$\rho$	
	$\mu_1$	$\sigma_1$	$\mu_2$	$\sigma_2$		$\mu_1$		$\sigma_1$	$\mu_2$			$\sigma_2$
						$a$	$b$		$c$	$d$		
2.5	77.3	11.9	117.5	21.4	0.06	55.7	0.06	7.4	103.5	-1.37	23.9	0.09
<b>50</b>	<b>100.6</b>	<b>29.9</b>	<b>162.3</b>	<b>47.2</b>	<b>0.56</b>	<b>76.6</b>	<b>0.47</b>	<b>16.6</b>	<b>163.3</b>	<b>-0.40</b>	<b>44.5</b>	<b>0.39</b>
97.5	125.8	54.3	426.3	77.7	0.99	148.1	1.21	52.6	208.0	-0.02	61.6	0.82
mean	<b>101.1</b>	<b>31.0</b>	<b>179.4</b>	<b>47.8</b>	<b>0.57</b>	<b>81.0</b>	<b>0.51</b>	<b>20.1</b>	<b>161.8</b>	<b>-0.48</b>	<b>44.5</b>	<b>0.38</b>

**Table 4.** Comparison of the estimated parameters between stationary and nonstationary mixture models at Yeongdeok station

Yeongdeok station												
Percentile (%)/ Mean	Stationary mixture model					Nonstationary mixture model						
	Gumbel-1		Gumbel-2		$\rho$	Gumbel-1			Gumbel-2		$\rho$	
	$\mu_1$	$\sigma_1$	$\mu_2$	$\sigma_2$		$\mu_1$		$\sigma_1$	$\mu_2$			$\sigma_2$
						$a$	$b$		$c$	$d$		
2.5	64.4	13.2	84.4	14.5	0.20	54.4	0.01	12.0	74.8	0.17	16.1	0.38
<b>50</b>	<b>78.2</b>	<b>23.3</b>	<b>153.2</b>	<b>43.2</b>	<b>0.83</b>	<b>69.6</b>	<b>0.25</b>	<b>19.9</b>	<b>121.7</b>	<b>0.93</b>	<b>40.6</b>	<b>0.76</b>
97.5	89.7	42.6	311.1	82.0	0.98	83.5	0.74	32.4	260.3	2.48	72.7	0.97
mean	<b>77.8</b>	<b>24.0</b>	<b>175.4</b>	<b>44.4</b>	<b>0.76</b>	<b>69.5</b>	<b>0.28</b>	<b>20.5</b>	<b>131.6</b>	<b>1.04</b>	<b>41.7</b>	<b>0.75</b>

최적분포형을 선정하는 기준에는 AIC (akaike information criterion), BIC (bayesian information criterion), DIC (deviance information criterion) 등이 있다. 자료의 개수( $n$ ), 매개변수의 개수( $k$ ), 우도( $\hat{L}$ )를 고려하여 우선순위를 선정하는 BIC, 즉 Eq. (7)을 사용하였다.  $\hat{L}$ 은 모델의 최우도값(maximum likelihood estimator)이며, Eq. (8)로 나타내며,  $\hat{\theta}$ 는 최우도함수의 매개변수이다.

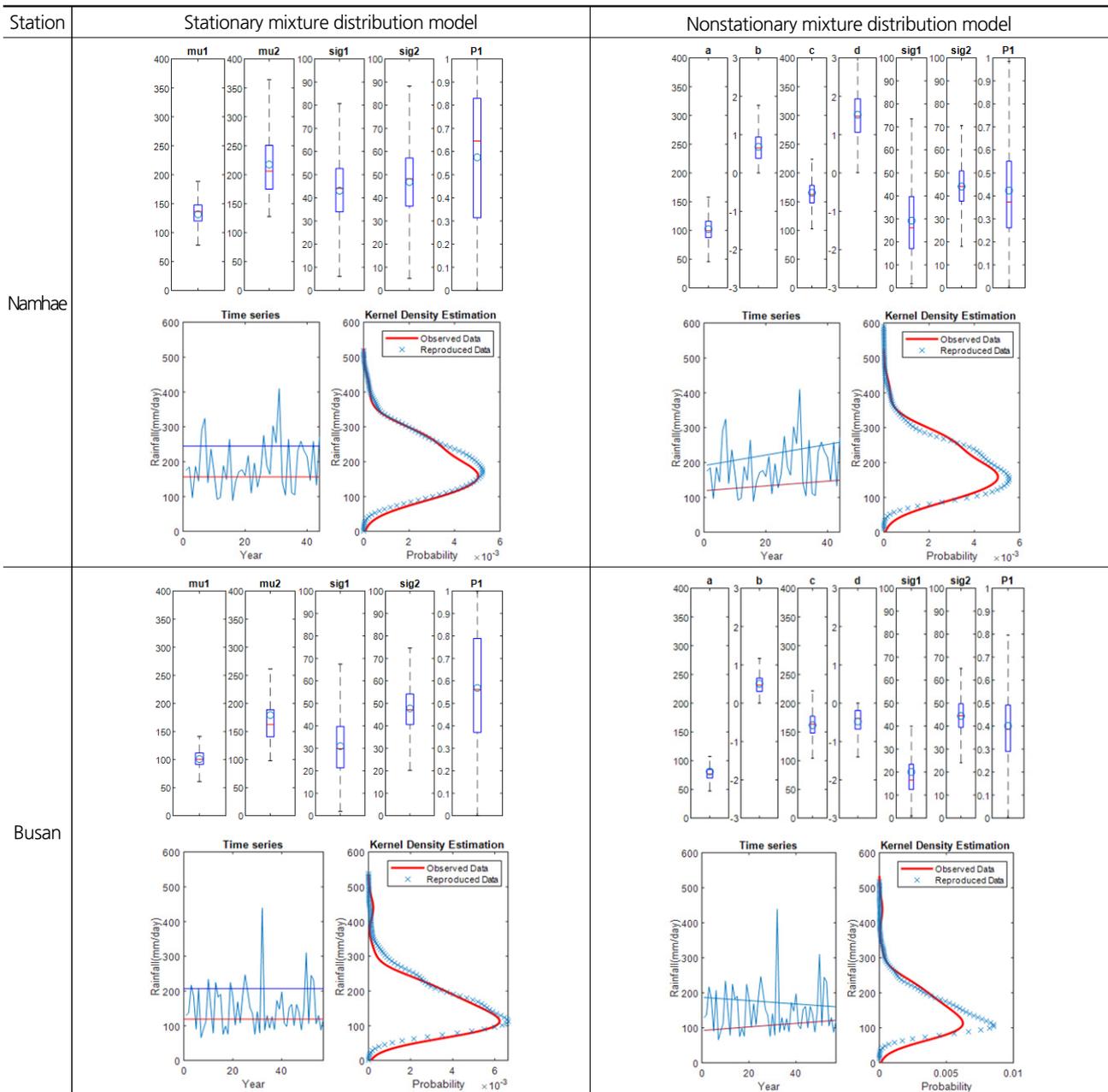
$$BIC = \ln(n)k - 2\ln(\hat{L}) \tag{7}$$

$$\hat{L} = p(x|\hat{\theta}) \tag{8}$$

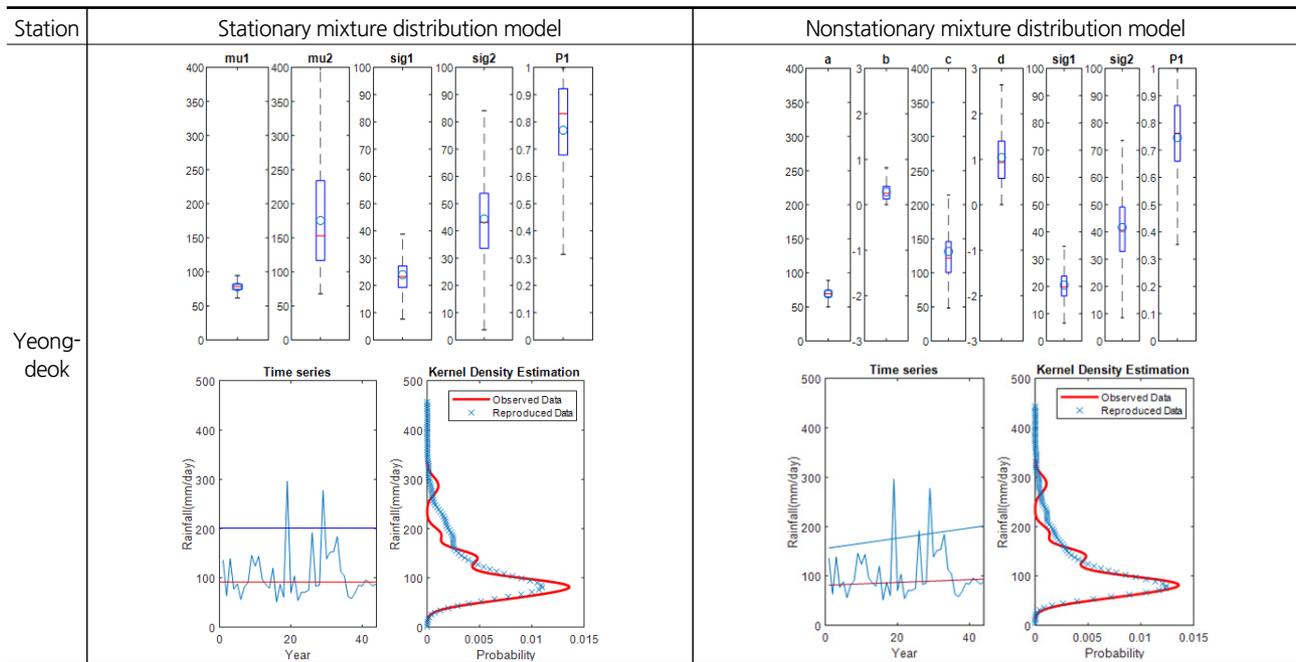
적은 개수의 매개변수와 큰 우도값을 토대로 BIC값이 작게 산정될 때 통계적으로 우수한 분포형이라 판단할 수 있다. 과적합(over-fitting) 문제에 있어서 매개변수의 개수에 대한 벌점(penalty)이 크게 작용하므로 정상성 모델보다 비정상성 모델이 전반적으로 불리하게 작용한다. Table 5에 정리한 것처럼 남해와 부산 지점의 비정상성 모델의 우도(likelihood)가 개선되는 효과가 매개변수 개수 증가의 영향보다 크므로 통계론적 측면에서 비정상성 모델이 정상성 모델보다 더 적합하다

**Table 5.** Comparison of the estimated likelihoods and BICs between stationary and nonstationary mixture models at Namhae, Busan, Yeongdeok stations

Statistical value	Median (Mean)	Namhae		Busan		Yeongdeok	
		Stationary model	Nonstationary model	Stationary model	Nonstationary model	Stationary model	Nonstationary model
Likelihood	median (mean)	-241.1 (-240.0)	<b>-230.7 (-232.0)</b>	-299.7 (-298.5)	<b>-287.6 (-289.5)</b>	-213.7 (-214.1)	<b>-211.3 (-211.0)</b>
BIC	median (mean)	501.2 (499.0)	<b>487.9 (490.6)</b>	619.5 (617.2)	<b>603.6 (607.2)</b>	<b>446.4 (447.2)</b>	449.0 (448.5)



**Fig. 3.** Comparison of plots between stationary and nonstationary mixture models at Namhae, Busan, Yeongdeok stations. All the parameters for the models were estimated within a Bayesian framework



**Fig. 3.** Comparison of plots between stationary and nonstationary mixture models at Namhae, Busan, Yeongdeok stations. All the parameters for the models were estimated within a Bayesian framework (continue)

고 판단된다. 영덕 지점의 경우 비정상성 모델에 의해 개선되는 효과가 매개변수 개수의 페널티를 감수하지 못하여 정상성 모델이 BIC 측면에서 적합하다.

각 지점마다 강우자료의 특성이 다르지만 공통적으로 나타나는 부분은 다음과 같다. Fig. 3의 정상성 및 비정상성 모델에서 첫 번째 분포형의 매개변수 불확실성이 두 번째 분포형의 경우에 비해 작다. 이는 관측자료의 분포를 통해 설명할 수 있다. 대부분 자료가 첫 번째 분포형 주변에 분포하는 것을 Fig. 1의 핵밀도함수를 통해 식별할 수 있다. 샘플의 수가 많아질수록 자료의 설득력이 커지기 때문에 매개변수의 불확실성이 줄어들게 된다. 반대로 두 번째 분포형쪽으로 갈수록 관측 자료 수가 줄어들어 따라 불확실성이 커지게 된다. 정상성 모델에 비해 비정상성 모델의 경우, 혼합비와 규모매개변수의 불확실성이 줄어든다. 특히, 첫 번째 분포형의 규모매개변수의 신뢰구간은 현저히 좁아지는 것을 식별할 수 있다. 앞서 설명했던 자료의 설명력이 커짐에 따라 분산이 줄어들게 되는 것을 확인할 수 있다. Fig. 3에서 지점별로 차이를 보이는 부분은 다음과 같으며, 편의상 작은 극치값을 대표하는 집단을 I 집단, 큰 극치값을 대표하는 집단을 II 집단으로 표기하였다. 남해 지점의 비정상성 혼합분포모델의 I, II 집단 모두 기댓값이 증가하는 특성을 보이고, 부산 지점은 I 집단은 완만하게 증가하고, II 집단은 완만하게 감소하는 형태를 보인다. 영덕 지점은 I 집단은 거의 변동이 없고, II 집단의 경우에 급격

**Table 6.** Estimations of design rainfalls of stationary and nonstationary models and their comparison corresponding to different return periods

Return Periods (year)	Percentile (%) / Mean	Namhae		Busan		Yeongdeok	
		Stationary (mm)	Non-stationary (mm)	Stationary (mm)	Non-stationary (mm)	Stationary (mm)	Non-stationary (mm)
10	2.5	253	267	204	186	138	146
	<b>median</b>	<b>286</b>	<b>305</b>	<b>233</b>	<b>220</b>	<b>169</b>	<b>185</b>
	97.5	331	357	274	264	258	260
	<b>mean</b>	<b>288</b>	<b>307</b>	<b>234</b>	<b>221</b>	<b>175</b>	<b>190</b>
50	2.5	320	333	270	249	196	208
	<b>median</b>	<b>372</b>	<b>387</b>	<b>315</b>	<b>297</b>	<b>261</b>	<b>272</b>
	97.5	452	469	419	361	370	368
	<b>mean</b>	<b>376</b>	<b>391</b>	<b>321</b>	<b>299</b>	<b>267</b>	<b>276</b>
100	2.5	345	360	295	275	220	231
	<b>median</b>	<b>409</b>	<b>423</b>	<b>349</b>	<b>329</b>	<b>295</b>	<b>303</b>
	97.5	506	516	467	404	419	415
	<b>mean</b>	<b>413</b>	<b>427</b>	<b>357</b>	<b>332</b>	<b>300</b>	<b>308</b>
500	2.5	405	420	354	334	272	279
	<b>median</b>	<b>495</b>	<b>505</b>	<b>429</b>	<b>404</b>	<b>363</b>	<b>371</b>
	97.5	627	628	573	504	543	528
	<b>mean</b>	<b>500</b>	<b>510</b>	<b>438</b>	<b>408</b>	<b>375</b>	<b>379</b>

하게 증가하는 형태이다. Fig. 1의 1차 선형 회귀함수를 통해 살펴본 지점별 모멘트의 동향을 설명할 수 있다. 남해 지점은

I, II 집단이 모두 증가함으로써 자료의 전체 모멘트가 증가하는 쪽, 즉 기온기가 양수(+)인 값을 가지고, 부산과 영덕 지점은 두 집단 간 상쇄되는 효과로 회귀계수가 '0'에 근접하는 값을 갖게 된다. 영덕 지점의 경우 II 집단의 동향에 의해 극치 자료의 기댓값이 증가하는 것으로 판단할 수 있으나, 혼합비에서 알 수 있듯이 II 집단의 자료 개수는 I 집단보다 현저히 적기 때문에 전체값에 대한 작은 영향력을 가진다. 본 연구 결과에 따라 남해와 부산 지점의 극치 자료는 비정상성을, 영덕 지점은 정상성으로 판단할 수 있다.

## 5. 결론

우리나라에서는 빈도해석 시 일반적으로 확률가중모멘트법을 적용한 Gumbel 분포형을 사용하고 있다. 서로 다른 발생요인으로 이루어진 자료의 모집단 추정에는 혼합분포형이 적합하다는 많은 연구사례가 있다. 이전 연구에서는 혼합분포에 기반한 정상성 빈도해석 기법을 제안했고, 단일분포모형을 적용한 경우와 통계치 비교를 통해 혼합분포모형이 불확실성에 대한 신뢰성을 확보하는데 우수한 성능을 발휘하는 것을 확인하였다(Choi et al., 2018). 본 연구에서는 남해, 부산, 영덕 지점의 강우자료에 대해서 비정상성 혼합모형의 적합성을 평가하였으며, 우도와 BIC 값을 통해 정량적으로 기존 정상성 혼합모형과 비교하였고, 도출된 결론은 다음과 같다.

- 1) 일반적인 강우빈도해석법은 연최대치강우가 단일 모집단 및 정상성 가정에 실시된다. 채택된 지점의 극치 자료를 평가해 보면 통계적으로 하나의 분포형으로 발생하는 모집단이라고 보기 어렵다. 즉, 발생 비율면에서 차이가 있지만, 최소 2개 이상의 특성을 가진 모집단으로 구분되고 있다. 또한, 전지구적인 기온 상승에 의한 영향으로 강우자료의 모멘트가 비정상성을 보인다. 이러한 자료를 정상성 가정에 한하여 적용시키는데 통계적인 한계가 있으며, 제안한 비정상성 혼합분포모형은 신뢰할 만한 추정치를 제공한다.
- 2) 매개변수가 7개인 비정상성 혼합모형이 매개변수가 5개인 정상성 혼합모형보다 매개변수가 많아 BIC 값 비교 시 불리하게 작용한다. 비정상성 모형 적용 시 모든 지점에서 우도값이 개선되었다(Table 5). 남해, 부산 지점에서는 우도값이 매개변수 증가로 인한 별점보다 상대적인 우위를 보였고 비정상성 모형이 우수한 것으로 평가되었다. 영덕 지점의 경우 매개변수가 증가한 만큼 우도값이 개선되는 효과가 없어 정상성 모형이 효율적인 것으로 평가되었다. 전반

적으로 비정상성 모형이 극치강우량 자료의 분포 특성을 보다 실질적으로 묘사하고 있는 것으로 판단할 수 있다.

- 3) 설계강우량 산정시 정상성 모형의 경우보다 정밀하게 추정되는 효과가 있다(Table 6). 비정상성 모형은 평균 기온 상승에 따른 우리나라 여름철 강우 지속시간 및 강우강도의 변화에 따른 극치 자료의 이중첨두 특성과 모멘트 증감 동향을 반영했고, 결과적으로 설계강수량의 불확실성이 줄어들 것으로 기대된다.

## 감사의 글

본 연구는 국토교통부/국토교통과학기술진흥원의 지원으로 수행되었음(과제번호 19AWMP-C140010-02).

## References

- Choi, H. G., Uranchimeg, S., Kim, Y. T., and Kwon, H. H. (2018). "A Bayesian approach to Gumbel mixture distribution for the estimation of parameter and its use to the rainfall frequency analysis." *Journal of the Korean Society of Civil Engineers*, KSCE, Vol. 38, No. 2, pp. 249-259.
- El Adlouni, S., Bobée, B., and Ouarda, T. B. M. J. (2008). "On the tails of extreme event distributions in hydrology." *Journal of Hydrology*, Vol. 355, No. 1-4, pp. 16-33.
- Gelman, A., Chew, G. L., and Shnaidman, M. (2004). "Bayesian analysis of serial dilution assays." *Biometrics*, Vol. 60, No. 2, pp. 407-417.
- 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*, KWRA, Vol. 32, No. 3, pp. 247-254.
- 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: A Journal of the Royal Meteorological Society*, 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. 223-237.
- 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 1." *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*, KWRA, Vol. 45, No. 5, pp. 505-516.
- 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 1." *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*, KWRA, Vol. 33, No. S1, pp. 124-129.
- Lee, S. H., and Kwon, W. T. (2004). "A variation of summer rainfall in Korea." *Journal of the Korean Geographical Society*, Vol. 39, No. 6, pp. 819-832.
- Ministry of Construction & Transportation (2000). *Korea probabilistic rainfalls map : development of water resources management techniques*, - research report, pp. 2-3.
- 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.
- 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 the Korean Society of Hazard Mitigation*, Vol. 14, No. 4, pp. 93-100.
- Smith, J. A. (1987). "Estimating the upper tail of flood frequency distributions." *Water Resources Research*, Vol. 23, No. 8, pp. 1657-1666.
- 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*, KWRA, Vol. 45, No. 3, pp. 263-274.