

# 최심신적설량 빈도분석을 위한 임계값을 가지는 일반화된 혼합분포모형 개발

김호준<sup>1</sup> · 김장경<sup>2</sup> · 권현한<sup>3\*</sup>

<sup>1</sup>세종대학교 건설환경공학과 박사과정, <sup>2</sup>베이지안웍스 공학박사, <sup>3</sup>세종대학교 건설환경공학과 교수

## Development of Snow Depth Frequency Analysis Model Based on A Generalized Mixture Distribution with Threshold

Ho Jun Kim<sup>1</sup>, Jang-Gyeong Kim<sup>2</sup>, and Hyun-Han Kwon<sup>3\*</sup>

<sup>1</sup>Ph.D Course, Dept. of Civil and Environmental Engineering, Sejong University

<sup>2</sup>Ph.D, Bayesian Works Research Institute

<sup>3</sup>Professor, Dept. of Civil and Environmental Engineering, Sejong University

### 요약

기후변화로 인해 다양한 자연재해의 발생빈도 및 강도가 증가하고 있으며, 이를 대비하기 위하여 행정안전부에서 가뭄과 대설까지 포함한 자연재해저감 종합계획을 발표하였다. 강설량은 기온과 지형적 요인의 영향을 크게 받는다. 산악지형이 많은 강원도는 강설량이 많아 큰 적설량이 관측되지만, 겨울철 평균 온도가 상대적으로 높은 남부지방은 적설량이 작다. 무강설과 결측으로 인해 관측값에 0이 포함된 경우가 존재한다. 자료에 포함된 0은 통계적으로 민감하게 작용하며, 최적 확률분포 선정과 매개변수 추정이 어려워지는 문제점이 발생한다. 본 연구에서는 창원, 통영, 진주 관측소의 최심신적설에 대해 혼합분포를 적용하여 0을 구분하였고, 0에 근사한 값을 나누는 기준인 임계값을 매개변수  $\delta$ 로 가정함으로써 무적설 기준을 자동으로 모형에서 추정하도록 하였다. Bayesian 기법 활용하여 혼합분포모형의 매개변수를 추정하였고, 산정된 빈도별 확률적설심의 불확실성을 정량화하였다. 대관령 지점과 비교한 결과, 본 연구의 혼합분포모형은 적설량이 적은 지점에 대해 적용성이 우수한 것으로 평가되었다.

**핵심용어:** 혼합분포함수, 최심신적설, Bayesian 모형, 빈도분석

### ABSTRACT

An increasing frequency and intensity of natural disasters have been observed due to climate change. To better prepare for these, the MOIS (ministry of the interior and safety) announced a comprehensive plan for minimizing damages associated with natural disasters, including drought and heavy snowfall. The spatial-temporal pattern of snowfall is greatly influenced by temperature and geographical features. Heavy snowfalls are often observed in Gangwon-do, surrounded by mountains, whereas less snowfall is dominant in the southern part of the country due to relatively high temperatures. Thus, snow depth data often contains zeros that can lead to difficulties in the selection of probability distribution and estimation of the parameters. A generalized mixture distribution approach to a maximum snow depth series over the southern part of Korea (i.e., Changwon, Tongyeong, Jinju weather stations) are located is proposed to better estimate a threshold ( $\delta$ ) classifying discrete and continuous distribution parts. The model parameters, including the threshold in the mixture model, are effectively estimated within a Bayesian modeling framework, and the uncertainty associated with the parameters is also provided. Comparing to the Daegwallyeong weather station, It was found that the proposed model is more effective for the regions in which less snow depth is observed.

**Keywords:** Mixture distribution function, Maximum snow depth, Bayesian model, Frequency analysis

\*Corresponding author: Hyun-Han Kwon, hkwon@sejong.ac.kr ORCID: <https://orcid.org/0000-0003-4465-2708>

Received: 18 September 2020, Revised: 16 October 2020, Accepted: 10 November 2020



## 1. 서론

기상재해는 이상기후 현상으로 인해 발생하며, 발생하는 기상현상에 따라 계절적으로 구분된다. 우리나라에 발생하는 강수량은 여름철에 집중되어 홍수와 집중호우를 발생시키며, 편중된 강수현상은 봄철 가뭄의 원인이 된다. 겨울철 강수는 연강수량의 10% 이하로 홍수기에 비해 작지만, 눈의 형태인 강설은 여러 피해를 발생시킨다. 특히, 겨울철 고위도 대륙의 시베리아 고기압이 따뜻한 해양의 수증기를 만나 눈구름을 형성하고 산악지형에 의해 상승하면서 대설을 자주 발생시킨다(Lee, 1979). 기후변화로 인한 이상기후 빈도 증가는 자연재해로 이어지고 있으며, 범국가적 차원에서 자연재해저감에 대한 관심이 높아지고 있다. 따라서, 최근 행정안전부는 개정된 자연재해대책법에 근거하여 풍수해 중심의 “풍수해저감 종합계획”을 가뭄과 대설을 포함한 “자연재해저감 종합계획”으로 변경하였다(MOIS, 2018). 이는 가뭄과 함께 겨울철 대설을 대비하기 위하여 국가적 차원에서 사전예방 및 대책을 수립하기 위함이다. 행정안전부의 “자연재해저감 종합계획 세부수립기준”에서 대설피해가 큰 도로 및 농업시설의 설해위험요인지수와 설해취약지수를 선정하여 대설재해 위험도를 산정하였고, 최심신적설 자료를 지표로 채택하였다(MOIS, 2019). 그러나 적설 빈도분석은 “홍수량 산정 표준지침”과 같은 산정기준 및 가이드라인이 존재하지 않아 국내 연구에서는 기존 수문자료와 동일한 방법으로 적설 빈도분석 연구를 수행하고 있다. Yu et al.(2014)은 확률가중모멘트법에 의한 Weibull 분포가 가장 적합하다고 판단하여 연최대 최심적설량에 대한 빈도별 확률적설심 산정과 신적설심 지도를 작성하였다. Kim et al.(2012)은 지리적으로 구분된 대설지역의 최심신적설에 대한 시공간적 경향성 분석과 빈도분석을 수행한 결과, 감소경향을 보였으며 국부적으로 폭설이 발생한다고 판단하였다. 기후변화로 인해 전지구적 기온 상승은 겨울철 강수 형태에 영향을 미치며, 많은 연구에서는 기후변화 시나리오를 활용한 미래 확률적설심 산정 방법들이 연구되고 있다. Park et al.(2014)과 Kim et al.(2014)은 기후변화 시나리오를 통한 미래 강수량에 대한 적설량의 비율을 산정하여 미래 확률적설심 증감률이 기후변화로 인해 지속적으로 감소한다고 평가하였다.

강수는 대기온도에 따라 형태변화가 이루어지고 지형적 영향을 받기에 산악지형이 많은 강원도에서는 많은 강설량이 관측된다. 그러나 기후변화로 인하여 겨울철 평균 기온이 상승하여 겨울철 강수 형태가 강설에서 강우로 점점 변하고 있다. 특히, 남부지방에서는 강설이 발생하지 않는 해가 증가하고 있으며 이로 인해 관측값에 ‘0’이 포함된 경우가 존재한다. 무강설 또는 결측된 자료, 즉 ‘0’은 통계적으로 민감하게 반응하며, 특히 연최대치 계열의 자료를 분석할 때, 최적 확률분포형 선정과 매개변수 추정이 어려워지는 문제점이 발생한다. 이러한 문제점을 해결하기 위하여 Park and Chung(2019)은 연최대치 계열 자료를 특정 임계값(threshold)으로 표현되는 절단자료(censored data) 가정과 조건부 결합확률분포를 적용하여 빈도분석을 수행하였고 대부분 고빈도에서 확률적설심이 작아지는 것을 확인하였다. Park et al.(2019)은 혼합분포를 활용하여 ‘0’이 포함된 연최대 최심신적설을 이산확률분포와 연속확률분포로 나누어 분석하였으며, 혼합분포 기반의 확률적설심이 단일분포를 적용한 결과보다 작게 산정되었다. 수문 분야에서의 혼합분포는 하나 이상의 원인에 의해 발생하는 사상을 분석하는데 활용된다. 예를 들어 태풍과 장마와 같이 다른 형태의 강우에 대한 빈도해석에 활용되고 있다(Haan, 1977). Yue et al.(1999)은 홍수위험도 평가 시 첨두홍수량(flood peaks)과 홍수지속시간(duration)을 대상으로 혼합검별분포모형(gumbel mixed model)을 활용해 빈도분석을 수행하였다. Yoon et al.(2012)은 극한 호우사상을 태풍과 집중호우로 구분한 혼합검별분포모형을 제시하여 확률강우량을 산정하였고, 혼합분포모형을 이용한 확률강우량이 감소하는 것을 확인하였다. Choi et al.(2019)은 2개 이상의 첨두 형태의 확률밀도함수로 표현된 극치 강우사상에 대해서 혼합분포함수를 활용하여 Bayesian 비정상성 빈도분석을 수행하였다.

본 연구에서는 두 가지 이상의 확률분포로 표현되는 혼합분포함수를 적용하였다. 기존 연구와는 다르게 임계값을 ‘0’으로 가정하지 않고 임계값에 해당하는  $\delta$ 를 추정해야 할 매개변수로 처리하여 혼합분포를 구성할 수 있는 빈도해석 기법을 개발하였다. 이를 통하여 확률분포의 적합성을 높이는데 주안점을 두었으며, 매개변수 추정 시 Bayesian 방법을 도입하여 매개변수 추정과 함께 불확실성을 정량화하여 제시하였다. 모형의 적합성을 평가하기 위하여, 대표적인 통계적 지표인 BIC

(Bayesian information criterion)를 사용해 최적 확률분포형을 선정하였고 최종적으로 재현기간별 확률적설심을 제시하였다.

## 2. 분석자료 및 연구방법

### 2.1 분석자료

적설은 고체형태 강우(눈, 싸락눈 등)를 제거하지 않고 누적된 눈의 깊이를 뜻하지만, 신적설은 특정 기간을 전제로 새롭 게 쌓인 눈의 깊이를 의미한다. 본 연구에서 활용한 일 최심신적설은 24시간 동안의 가장 큰 적설깊이를 의미하고, 관측 후 제 거하는 과정을 거친다. 기상청 산하 중관기상관측소(ASOS, automated synoptic observing system) 중 높은 겨울철 평균 기 온과 적은 강설량으로 인해 관측값에 ‘0’이 포함된 남부지방의 창원(155), 통영(162), 진주(192)를 대상지점으로 선정하였 다. 최심신적설량에 따른 혼합분포모형을 비교하기 위하여 상대적으로 적설량이 많은 대관령(100)을 비교지점으로 선정하 였다. 관측연도가 지점마다 상이하고 통영과 진주는 적설에 대한 목측 관측이 중단되어 무인적설(CCTV, 레이저식 적설계) 을 활용하고 있다. 목측 관측자료를 바탕으로 분석하기 위해 창원 1987~2019년(33년), 통영과 진주 1971~2015년(45년), 대 관령 1972~2015년(44년) 자료를 분석하였다. 목측 관측자료는 무인 관측자료와 달리 시스템의 오류가 없기에 결측치가 적 어 자료의 품질이 좋으며, 관측 연수가 훨씬 길어 통계적 분석에 적합하다. 해당 지점의 전년도 10월부터 당해 연도 4월까지 의 일단위 최심신적설을 연최대치 최심신적설로 산정하였으며, 이는 연도별로 산정하는 것이 아닌 홍수기와 같이 연속적 사 상발생기간으로 적용하기 위함이다. 예를 들어, 2000년은 1999년 10월부터 2000년 4월까지의 자료를 의미한다. Fig. 1은 연 최대 최심신적설 시계열 자료를 나타내며, Fig. 2은 창원(155), 통영(162), 진주(192) 기상관측소의 위치를 나타낸다.

확률지는 관측자료가 이론적 확률분포형에 적합하기를 시각적으로 평가하는 방법이다. Lognormal 분포는 입력값으로 ‘0’을 가질 수 없기에 연 최심신적설의 양수값을 Lognormal 분포 확률지인 Fig. 3에 제시하였으며, 대부분의 관측값이 이론 적인 분포를 나타내는 일직선상에 분포되는 것을 확인할 수 있다. 따라서 자료의 Lognormality를 가정할 수 있으며, 혼합분 포함수 구성 시 Lognormal 분포를 기본적으로 활용하였다.

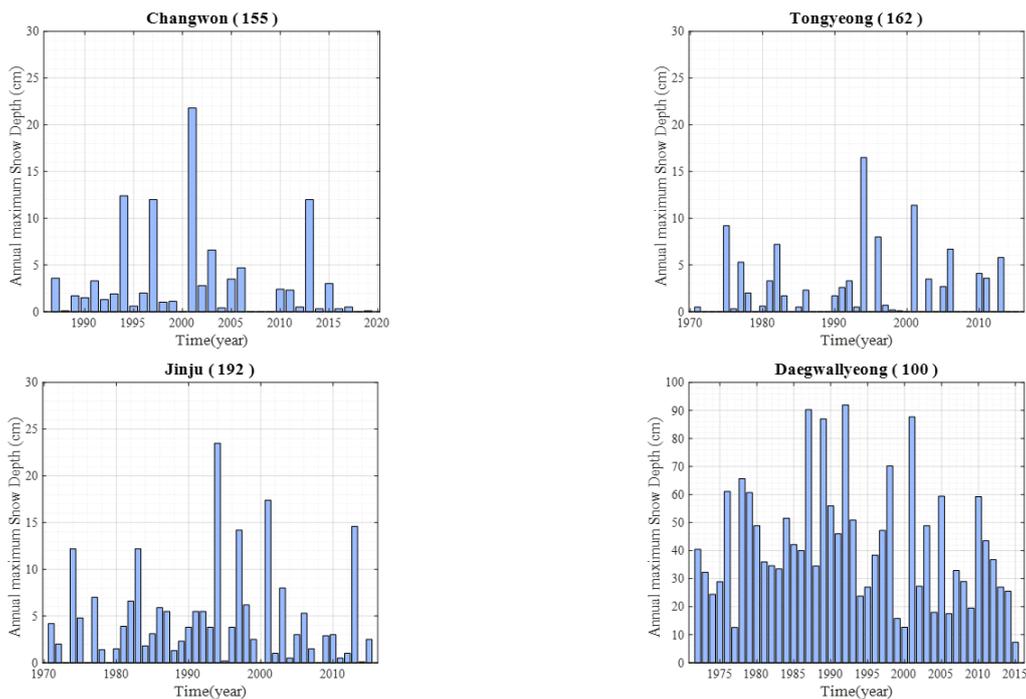


Fig. 1. Annual maximum snow depth series for 4 stations

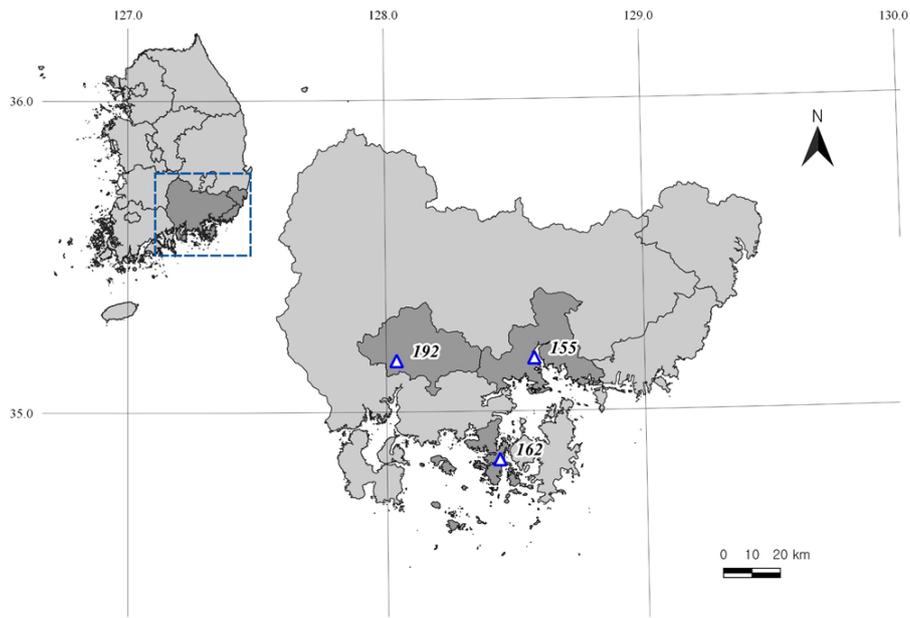


Fig. 2. Locations of weather stations used in this study

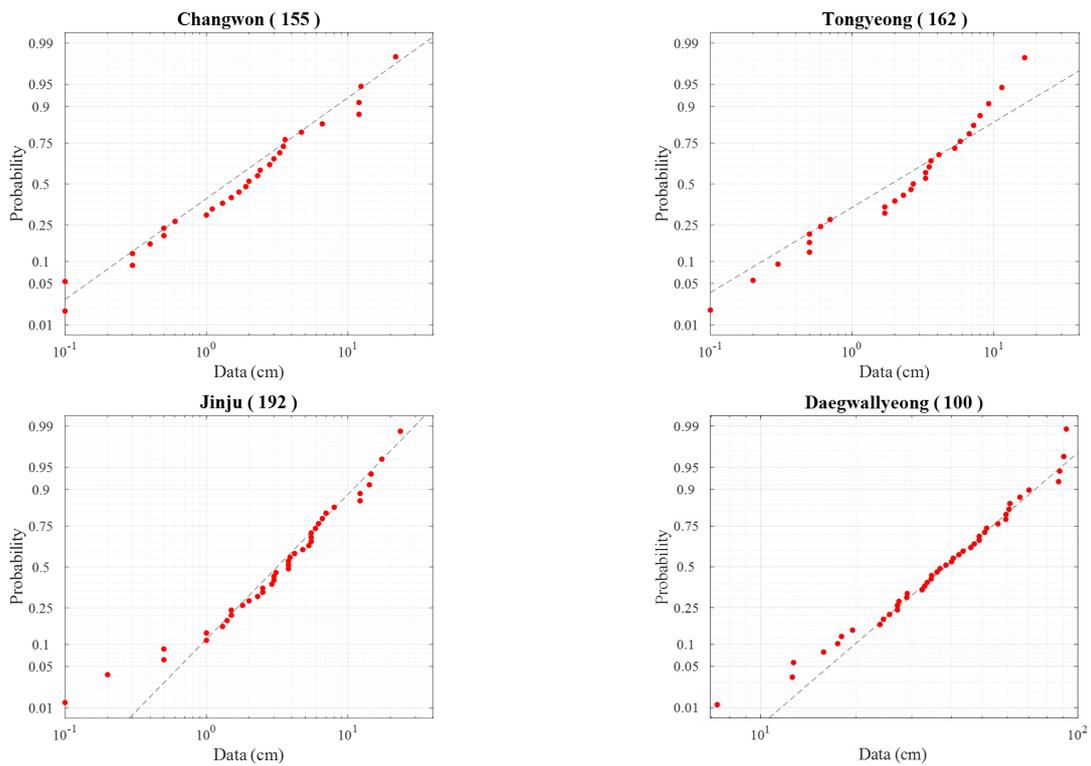


Fig. 3. Log-normal probability plots of annual maximum snow depth series for 4 stations

## 2.2 혼합 분포 함수

확률분포는 확률변수의 종류에 따라 이산확률분포, 연속확률분포 또는 이산확률분포와 연속확률분포를 혼합하여 활용할 수 있다(Kedem et al., 1990).  $f_i(X)$ ,  $i = 1, 2, 3, \dots, m$ 은 확률밀도함수이고,  $\pi_i$ ,  $i = 1, 2, 3, \dots, m$ 는 각 확률밀도함수의 비

율을 의미하며,  $\pi_i \geq 0$  과  $\sum_{i=1}^m \pi_i = 1$  을 만족한다면, 혼합분포의 확률밀도함수는 Eq. (1)과 같이 나타난다.

$$f_X(X) = \sum_{i=1}^m \pi_i f_i(X) \quad (1)$$

$f_i(X)$ 는  $i$  번째 분포에 따라 주어지는  $X$ 에 대한 확률밀도를 의미하며, 매개변수  $\pi_i$ 는 확률분포인  $f_i(X)$ 에 따라 달라지며, 이는 각 확률밀도함수의 혼합비(mixing ratio)를 나타낸다(Haan, 1977).  $X$ 에 대한 누적확률밀도함수는 Eq. (2)과 같이 표현된다.

$$F_X(X) = \int_{-\infty}^X \sum_{i=1}^m \pi_i f_i(t) dt = \sum_{i=1}^m \pi_i \int_{-\infty}^X f_i(t) dt \quad (2)$$

일반적으로 '0'을 임계값으로 혼합분포함수를 적용한다. 그러나 '0'뿐만 아니라 '0'에 근접한 값을 임계값  $\delta$ 로 가정하여 연최대 최심신적설인  $x$ 를 구분하는 경우 임계값 이상의 자료에 대해서 확률분포 적합 시 모형의 우도(likelihood)의 개선을 기대할 수 있다. 이는  $\delta$ 보다 작은 값, 즉 '0'과 근접한 값을 '0'으로 취급하여 통계적으로 혼합분포함수에 적용하기 위함이다. 여기서, 임계값  $\delta$ 는 특정값으로 고정된 값이 아니라, 모형 전체의 우도를 가장 크게 할 수 있는 방향으로 추정되는 매개변수로 고려하였다. Eq. (3)은 이산분포와 연속분포로 구성된 혼합분포함수의 기본식을 나타낸다.  $F(x)$ 은  $x$ 의 연속분포함수이고,  $H(x)$ 은 계단 함수(step function)로 정의된다.

$$G(x) = (1 - \pi)F(x) + \pi H(x) \quad (3)$$

$$\pi = \frac{\sum P(x)}{N}, P(x) = \begin{cases} 0 & \text{if } x \geq \delta \\ 1 & \text{if } x < \delta \end{cases} \quad (4)$$

Eq. (3)에서 가중치를 의미하는  $\pi$ 는 무적설의 비율에 따라 산정된다.  $P(x)$ 은 Eq. (4)과 같이  $x$ 가  $\delta$  미만일 경우 1,  $\delta$  이상일 경우 0을 가지며, 이를 합산하여 자료 전체 기간인  $N$ 으로 나눈다.

$$F(x) = \begin{cases} F(x) & \text{if } x \geq \delta \\ 0 & \text{if } x < \delta \end{cases} \quad (5)$$

$$H(x) = \begin{cases} 0 & \text{if } x \geq \delta \\ 1 & \text{if } x < \delta \end{cases} \quad (6)$$

연속분포함수  $F(x)$ 와 이산분포함수  $H(x)$ 는  $x$ 의 크기에 따라 결정된다.  $F(x)$ 은 Eq. (5)과 같이  $x$ 가  $\delta$  미만일 경우 0,  $\delta$  이상일 경우  $F(x)$  그대로 표현되며, 이산분포를 의미하는 계단함수인  $H(x)$ 은 Eq. (6)과 같이  $x$ 가  $\delta$  미만일 경우 1,  $\delta$  이상일 경우 0을 가진다. 최종적으로 혼합비율  $\pi$ 를 포함한 혼합분포함수  $G(x)$ 가 Eq. (7)과 같이 표현된다.

$$G(x) = \begin{cases} (1 - \pi)F(x) & \text{if } x \geq \delta \\ \pi & \text{if } x < \delta \end{cases} \quad (7)$$

### 2.3 혼합분포 매개변수 추정을 위한 Bayesian 모형

Bayesian 추론기법은 관측된 자료와 모수에 대한 확률모형과 사전분포(prior distribution)을 기준으로 자료 증가로 정보가 갱신되어 최종적으로 사후분포(posterior distribution)을 추정하는 기법을 말한다(Gelman et al., 2004). 베이지 이론(Bayes's rule)을 기반의 Bayesian 추론기법은 매개변수 최적화 및 불확실성 정량화를 목적으로 두며, 기존 방법(모멘트법, 최우도법, 확률가중모멘트법)과 다르게 매개변수를 단일 값이 아닌 확률분포 형태로 산정된다. 베이지 이론에 따르면, 독립된 두 사상 A와 B가 존재할 때, 결합확률  $P(A, B)$ 은 주변확률  $P(B)$ 과 조건부확률  $P(A|B)$ 을 통해 구할 수 있다. 주어진 확률변수  $y$ 와 매개변수들의 집합  $\theta$ 가 있을 때, 두 확률변수의 결합확률분포는 Eq. (8)과 같이 사전분포  $p(\theta)$ 와 우도  $p(y|\theta)$ 의 곱으로 표현할 수 있다. Bayes 정리에서 매개변수  $\theta$ 와 변량  $y$ 의 조건부확률과 주변확률의 관계는 Eq. (9)과 같으며, Eq. (10)에서  $p(\theta|y)$ 는 사후분포를 나타내며 사전분포  $p(\theta)$ 와 우도  $p(y|\theta)$ 의 곱으로 추정될 수 있다(Gelman et al., 2004).

$$p(\theta, y) = p(\theta)p(y|\theta) \quad (8)$$

$$p(\theta|y) = \frac{p(\theta, y)}{p(y)} = \frac{p(\theta)p(y|\theta)}{p(y)} \quad (9)$$

$$p(\theta|y) \propto p(y|\theta)p(\theta) \quad (10)$$

모형 및 분포 선정시 대표적으로 AIC, BIC 등과 같은 통계적 지표를 활용한다. AIC(Akaike information criterion)는 확률분포 간의 차이를 의미하는 Kullback-Leibler 거리에 의해 정의되며 모형의 대수우도함수와 모형의 복잡한 정도를 나타내는 매개변수의 개수로 표현된다. 본 연구에서는 최적 확률분포 선정에 위해 우도와 매개변수의 개수로 계산되는 통계적 지표인 BIC를 활용하였다. BIC는 Schwarz(1978)가 제안하였으며, Bayesian 사후확률의 형식으로 정의되는 정보 기준이며, AIC와 달리 표본의 크기에 따라 별점이 정해지며, 복잡한 모형의 경우 AIC보다 별점이 크게 산정된다. AIC와 동일하게 BIC 값이 작을수록 모형의 적합도가 높음을 의미한다. BIC의 일반적인 식은 Eq. (11)과 같으며,  $L$ 은 우도를 의미하며,  $k$ 는 매개변수의 개수,  $n$ 은 자료의 개수를 나타낸다.

$$BIC = -2\ln(L) + k\ln(n) \quad (11)$$

본 연구에서는 Eq. (7)과 같이 이산분포와 연속분포로 결합된 혼합분포함수의 매개변수를 추정하기 위하여 Bayesian 기법을 활용하였다. 혼합분포함수  $G(x)$ 의 연속확률분포를 의미하는  $F(x)$ 에 적합한 확률분포 선정하기 위하여 BIC를 추정하여 검토하였으며, 여러 확률분포 중 Lognormal 분포를 최적분포로 선정하였다. 연속된 곱으로 표현되는 우도함수는 로그함수를 통해 합으로 변환되며, 이를 대수우도함수(log likelihood)라 한다.  $\delta$  이상의  $x$ 에 대한 Lognormal 분포의 우도함수를 Eq. (12)에 제시하였으며,  $\delta$  미만의  $x$ 는 Eq. (4)에서 계산된  $\pi$ 로 산정된다.  $x$ 은 연최대 최심신적설이며,  $n$ 은  $\delta$ 보다 큰 자료의 개수를 의미한다.

$$L = \begin{cases} \pi & \text{if } x < \delta \\ (1 - \pi) \times \frac{1}{\prod_{i=1}^n x_i} \left( \frac{1}{\sigma \sqrt{2\pi}} \right)^n \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (\log(x_i) - \mu)^2 \right\} & \text{if } x \geq \delta \end{cases} \quad (12)$$

사후분포를 추정하기 위해서는 사전분포를 정의해야 하며, Eq. (13)과 같이 평균  $\mu$ 와 매개변수  $\delta$ 은 정규분포(normal distribution), 표준편차  $\sigma$ 은 감마분포(gamma distribution)를 부여하였다. 매개변수  $\delta$ 은 0 이상이므로 절단정규분포(truncated normal distribution)를 부여하였다. Eq. (10)과 같이 사후분포는 우도함수와 사전분포의 곱으로 산정되며,  $\theta = [\mu, \sigma, \delta]$ 는 추정되는 매개변수이며,  $p(\theta)$ 는 매개변수들의 사전분포이다. 우도함수  $p(y|\theta)$ 는 Eq. (14)과 같이 표현되며, 최종적으로 사후분포  $p(\theta|y)$ 는 사전분포와의 곱을 통해 Eq. (15)과 같이 추정할 수 있다.

$$\mu \sim N(1, 10) \quad (13a)$$

$$\sigma \sim \Gamma(1, 1) \quad (13b)$$

$$\delta \sim N(0, 1)|(0, \sim) \quad (13c)$$

$$p(y|\theta) = \prod_{i=1}^I L(x_i|\mu, \sigma, \delta) \quad (14)$$

$$p(\theta|y) \propto \prod_{i=1}^I L(x_i|\mu, \sigma, \delta) \cdot N(\mu|1, 10) \cdot \Gamma(\sigma|1, 1) \cdot N(\delta|0, 1)|(0, \sim) \quad (15)$$

매개변수를 추정하기 위하여 Bayesian Markov Chain Monte Carlo(MCMC)기법 중 깃스샘플러(gibbs sampler) 방법을 이용하였다. 깃스샘플러 방법은 두 개 이상 확률변수의 결합확률분포로부터 일련의 표본을 생성하는 확률적 알고리즘이다. 이 방법을 통해 추정된 매개변수의 샘플을 활용하여 사후분포를 산정하였다. Bayesian MCMC 기법에 대한 자세한 내용은 기존 연구문헌을 참조할 수 있다(Gilks et al., 1995; Kwon et al., 2008).

일반적으로 누적분포함수(CDF, cumulative distribution function)는 어떤 변량  $X$ 가 특정변수  $x$ 보다 크지 않을 확률인 비초과확률을 의미하며, 재현기간  $T$ 와 관계는 Eq. (16)과 같이 표현된다.

$$P(X \leq x) = 1 - \frac{1}{T} \quad (16)$$

혼합분포  $G(x)$ 는 재현기간  $T$ 와 Eq. (17)과 같은 관계를 갖는다. 재현기간이 커짐에 따라 누가발생확률은 1로 수렴하게 된다. 혼합분포  $G(x)$ 는 Eq. (18)과 같이 나타나며, Eq. (17), (18)의 과정을 거쳐 Eq. (19)과 같이 나타난다.

$$G(x) = 1 - \frac{1}{T} \quad (17a)$$

$$(1 - \pi)F(x) + \pi = 1 - \frac{1}{T} \quad (17b)$$

$$F(x) = \frac{1}{(1 - \pi)} \left(1 - \frac{1}{T} - \pi\right) \quad (18)$$

$$x = F^{-1} \left( \frac{1}{(1 - p)} \left(1 - \frac{1}{T} - p\right) \right) \quad (19)$$

### 3. 연구결과

남부지방의 관측소에서는 높은 겨울철 평균 기온으로 인해 강설이 아닌 강우의 형태로 강수가 발생하여 연최대 최심신적 설 자료에 0이 포함된다. 따라서, 혼합분포를 활용하여 무적설을 구분하였으며, 확률적설심을 산정하기 위해 빈도분석 모형을 개발하였다. 기존 연구 중 혼합분포함수를 활용한 연구(Park et al., 2019)가 존재하지만, 본 연구에서는 임계값을 매개변수  $\delta$ 로 새롭게 추정함으로써 무적설 기준을 정량적으로 정의할 수 있으며, 빈도해석 모형의 매개변수 추정방법 시 불확실성을 정량화하였다.

#### 3.1 BIC를 활용한 확률분포형 선정 및 혼합분포 비교

BIC를 활용하여 Eq. (7)의  $F(x)$ 에 해당하는 최적 확률분포형을 선정하였다.  $\delta$ 보다 큰 값에 해당하는 Eq. (7)에 나타난  $F(x)$ 의 최적 확률분포형을 선정하기 위해 BIC를 사용했다. 대관령을 제외한 대상 지점들은 분석한 5개의 확률분포(Lognormal, Gamma, Exponential, Gumbel, Weibull) 중 Lognormal의 BIC가 제일 작게 산정되어 이를 채택하였으며, Table 1에 지점별 BIC 결과를 나타내었다. 대관령의 경우도 Lognormal과 Gamma 분포의 BIC 값의 차이가 미소하게 나타나고 있으며 본 연구에서는 Lognormal 분포를 이용하여 분석을 실시하였다.

**Table 1.** BIC values for different distributions across weather stations

Station No.	Distribution				
	Lognormal	Gamma	Exponential	Gumbel	Weibull
155	<b>262.047</b>	263.558	266.371	267.056	283.562
162	<b>270.851</b>	276.128	278.558	272.592	274.233
192	<b>427.194</b>	430.360	429.670	439.169	433.398
100	605.819	<b>603.582</b>	638.326	660.635	615.564

본 연구에서 제시한 임계값 기준의 혼합분포함수(모형 A)와 0을 기준으로 한 혼합분포함수(모형 B)를 비교하고자 한다. 혼합분포 A와 B의 혼합비  $\pi$ 와 BIC를 Table 2에 나타내었으며,  $\pi$ 는 기준(0 또는  $\delta$ )보다 작은  $x$ 의 비율을 의미하고(Eq. (4)) BIC는 모형의 우수성을 상대적으로 비교할 수 있는 지표이다. 먼저, 대상지점 3개 지점(155, 162, 192)은 임계값  $\delta$ 가 클수록 혼합비  $\pi$ 가 크게 산정되었다. 혼합분포 A를 적용한 결과 창원지점의  $\delta$ 는 5.067 mm이고 혼합비  $\pi$ 는 0.36으로 분석되었다. 혼합분포 B를 통해 계산된 혼합비  $\pi$  값 보다 0.194 크게 추정되었다. 이는 0 mm부터 5.067 mm까지 무적설로 취급하여 혼합비  $\pi$ 가 커진 것을 의미한다. 또한, 0에 근사한 값을 제외함으로써 적설자료에 대한 연속분포 적합성을 개선할 수 있었다. 이와 반대로, 강설량이 많은 대관령(100)의 혼합분포 A와 B를 통해 계산된 혼합비  $\pi$ 는 0으로 나타났다. 즉, 대관령 지점은 연

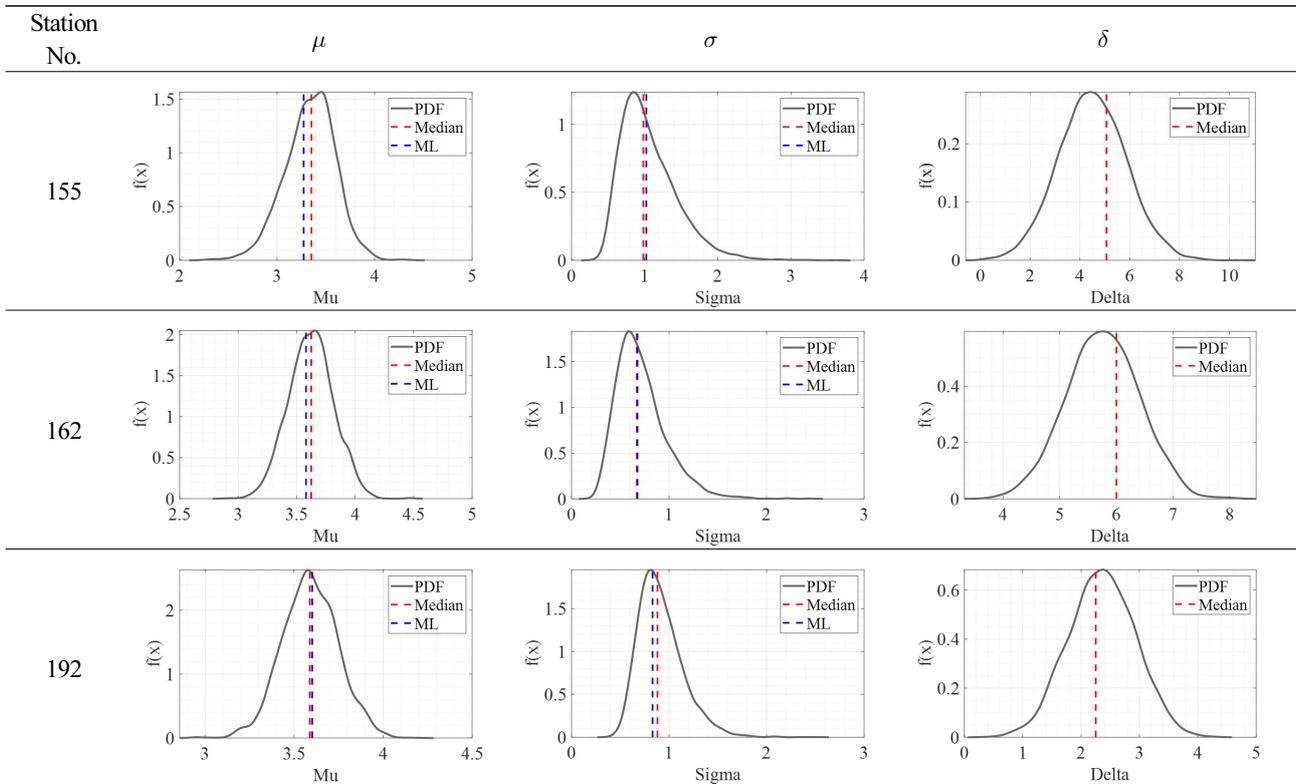
최대 최심신적설( $x$ )이 기본적으로 남부지방에 비해서 크며 임계값  $\delta$ 가 혼합분포모형에서 역할은 제한적이다 할 수 있다. 다시 말해, 대관령 지방의 경우 기존 매개변수에  $\delta$ 까지 추가하는 것은 통계적으로 불필요함을 의미한다. 즉, 혼합분포 A에서 추가한 매개변수인  $\delta$ 로 인하여 BIC값이 크게 산정되었다.

**Table 2.** Comparison of mixture distribution model A and B across weather stations

Station No.	Mixture distribution model A			Mixture distribution model B	
	BIC	$\pi$	$\delta$ (unit : mm)	BIC	$\pi$
155	262.0	0.364	5.067	291.4	0.152
162	270.9	0.556	6.002	324.1	0.400
192	427.2	0.156	2.251	440.1	0.111
100	603.6	0	0.786	599.9	0

### 3.2 분석결과

본 연구에서는 Bayesian기법을 활용하여 매개변수의 사후분포  $p(\theta|y)$ 를 추정하고 Fig. 4에 제시하였다. 사후분포의 중앙값(median)을 빨간 점선, 최우도법(likelihood method)을 통해 추정된 값을 파란 점선으로 나타내어 비교하였다. 임계값  $\delta$ 에 따라서 적설빈도해석 모형에 적용되는 적설자료가 다르기 때문에 추정되는 매개변수 또한 기존 방법과 일부 상이하게 나타내게 된다. 사후분포는 가정한 사전분포 형태로 수렴되어 추정되는 것을 그림에서 확인할 수 있다.



**Fig. 4.** Posterior distributions of model parameters at 3 stations (Changwon, Tongyeong, Jinju)

추정된 매개변수와 Eq. (19)를 통해 빈도분석을 수행하였으며, 빈도별 확률적설심을 Table 3에 제시하였다. 다른 지점들과 비교하였을 때, 통영 지점의 확률적설심이 제일 작은 것으로 평가되었다. 이는 통영지점의 혼합비  $\pi$ 가 0.55로 가장 컸으며, 대부분의 값이 무적설로 취급됨을 알 수 있다. 또한, 통영지점의 연최대 최심신적설의 평균이 제일 작았기 때문에 확률적설심에 동일하게 반영된 것으로 판단된다. Bayesian 추론기법은 매개변수 최적화와 더불어 매개변수의 불확실성을 정량적으로 제시할 수 있는 장점이 있다. 확률적설심과 함께 불확실성 구간을 산정하였으며, Fig. 5와 Table 3에 나타내었다. Fig. 5의 파란 실선은 중앙값으로 확률적설심을 의미하고, 검정 및 빨간 점선은 각각 2.5%, 97.5% 불확실성 구간을 뜻한다. 그래프의 표식은 Table 3의 재현기간별 확률적설심을 의미한다. 본 연구에서 제시한 불확실성은 정규분포로 가정하여 좌우대칭의 형태로 표현되는 기존 confidence interval(CI)와 달리 Bayesian 모형으로부터 자료의 크기가 커지면 분산이 커지는 이분산성(heteroscedasticity)의 특징을 가진다.

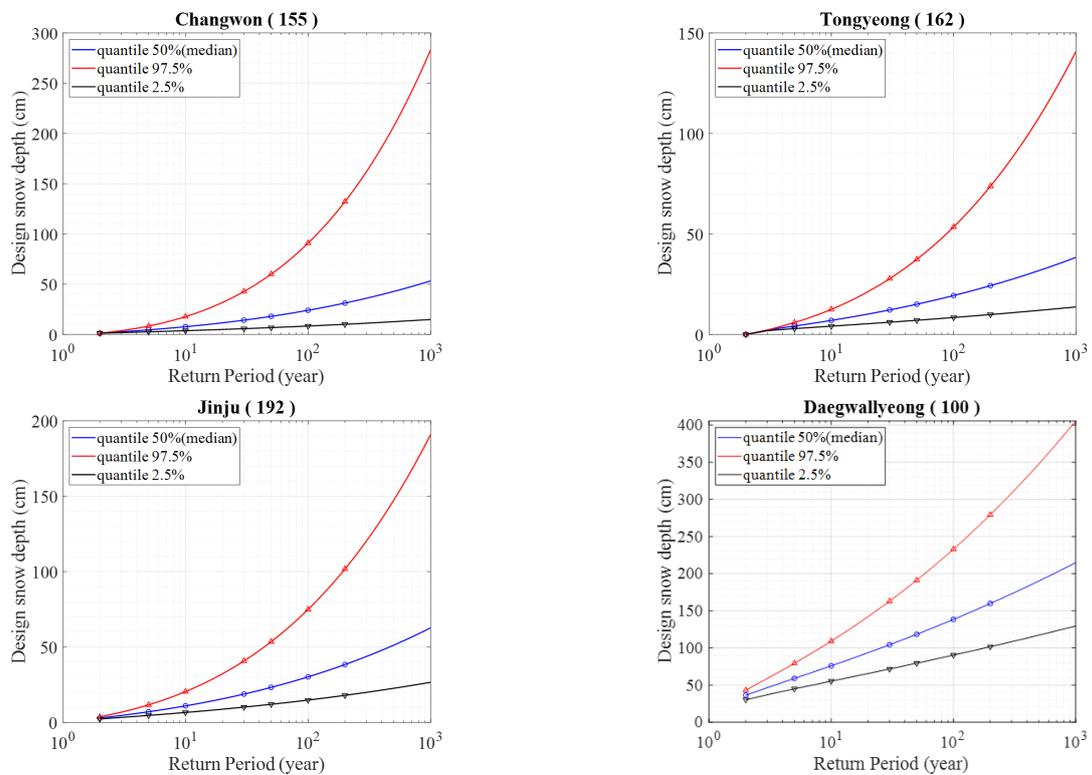


Fig. 5. Estimation of design snow depth corresponding to different return periods (unit : cm)

Table 3. Estimation of design snow depth and their comparisons corresponding to different return periods (unit : cm)

Return Period	Changwon (155)			Tongyeong (162)			Jinju (192)			Daegwallyeong (100)		
	2.5%	50%	97.5%	2.5%	50%	97.5%	2.5%	50%	97.5%	2.5%	50%	97.5%
5	2.7	4.6	8.4	2.9	4.2	5.9	4.7	7.1	11.7	30.2	36.3	43.1
10	3.8	7.7	17.7	4.1	7.0	12.5	6.6	11.0	20.4	44.9	58.9	79.3
30	5.8	14.3	42.8	6.1	12.2	27.7	10.2	18.8	40.8	55.2	75.9	109.2
50	6.8	18.0	60.0	7.0	15.0	37.3	12.0	23.3	53.5	71.6	104.3	163.0
100	8.4	24.1	90.9	8.4	19.3	53.5	14.9	30.2	74.9	79.4	118.3	191.2
200	10.1	31.3	132.1	9.9	24.3	73.7	18.0	38.4	101.8	90.3	138.4	233.0

## 4. 결론

우리나라의 적설량에 대한 분석지침은 ‘홍수량 산정 표준지침’과 다르게 명확한 기준이 없다. 강우는 빠른 유출로 인해 홍수를 발생시키는 반면, 강설은 도로와 농업시설 등을 마비시키는 다른 유형의 피해로 나타난다. 또한, 강설량은 강우와 다르게 기온의 영향을 크게 받아 관측소 위치에 따라 분명한 차이가 존재한다. 따라서, 적설 특성을 파악하여 강우자료와 차이점을 인지하고 적합한 평가방법을 사용하는 것이 필요하다. 본 연구에서는 겨울철 높은 평균 기온으로 인해 강설량이 적은 지역을 대상지점으로 선정한 후 확률적설심을 평가하였다. 이산확률분포와 연속확률분포로 구성된 혼합분포를 적용하여 0이 포함된 연 최대 최심신적설의 빈도분석을 수행하였다. 도출된 결론은 다음과 같다.

첫째, 본 연구에서는 ‘0’값을 포함하는 적설자료에 대한 분석을 위해서 일반화된 혼합분포의 형태를 제시하였다. 혼합분포를 구성하는 연속확률분포의 선정을 포함하여 매개변수의 최적화 등을 동시에 수행할 수 있도록 BIC값 기준의 모형 선택 및 매개변수 추정 방안을 제시하였다.

둘째, 기존 적설심 추정을 위한 혼합분포함수에서는 0을 임계값을 가정하여 혼합분포를 구성한 반면 본 연구에서는 임계값을 매개변수  $\delta$ 로 가정하여 모형 자체적으로 가장 최적의 임계값을 선정할 수 있도록 하였다. 이를 통해 구축된 혼합분포모형이 BIC값 기준 기존 해석 모형에 비해 개선된 결과를 나타내었다.

셋째, 대관령 지점과 같이 적설량이 많은 지역에 대해서 임계값  $\delta$ 를 포함한 일반화된 혼합분포를 적용한 결과 임계값  $\delta$ 가 0으로 수렴하게 되는 것을 확인할 수 있었다. 즉, 혼합비가 1로 수렴하게 되며 하나의 분포로 모형을 구성하는 것이 통계적으로 우수하다는 것을 입증하는 것으로서 BIC 값을 통해서도 확인할 수 있었다.

넷째, Bayesian 기법을 적용하여 혼합분포모형의 매개변수를 확률분포형태로 제시하였다. 기존 최우도법을 통해 추정된 매개변수와 비교한 결과, 매개변수의 평균값은 거의 동일하게 추정되고 있는 것을 확인할 수 있다. 그러나, 기존 최우도법과는 다르게 재현기간이 커질수록 분산이 커지는 이분산성의 특징을 효과적으로 평가할 수 있었으며 불확실성 구간 또는 좌우대칭의 정규분포가 아닌 실제 자료의 확률분포를 따르는 특징 또한 확인할 수 있었다.

강설은 강우에 비해 지역적 편차가 훨씬 크므로 기존의 방법으로 확률적설심을 산정하는 것은 과대추정의 오류를 범할 수 있으므로 주의가 필요하다. 본 연구에서는 강설량이 적은 지점에 대해서는 혼합분포모형을 적용하는 것이 적절하다고 판단하였으며, Bayesian 기법을 활용한 최심신적설의 빈도분석방법을 제시함으로써 불확실성 구간을 산정하였다. 향후 연구로 기후변화 시나리오를 활용하여 강설량 변화에 파악하고, 온도 상승에 따라 강설량이 줄어들어 혼합분포모형이 적용이 확대 될 것으로 기대된다.

## Acknowledgment

본 결과물은 환경부의 재원으로 한국환경산업기술원의 지능형 도시수자원 관리사업의 지원을 받아 연구되었습니다 (2019002950001).

## References

- Choi, H. G., Kwon, H. H., and Park, M. H. (2019). A Development of Nonstationary Rainfall Frequency Analysis Model based on Mixture Distribution. *Journal of Korea Water Resources Association*. 52(11): 895-904.
- Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004). *Bayesian Data Analysis* (2nd ed.). Boca Raton: Chapman and Hall/CRC. CRC press. USA.
- Gilks, W. R. Best, N. G., and Tan, K. K. C. (1995). Adaptive Rejection Metropolis Sampling within Gibbs Sampling, *J. R. Stat. Soc., Ser. C*. 44: 455-472.

- Haan, C. T. (1977). *Statistical Methods in Hydrology*. The Iowa State University Press. Ames. Iowa. USA.
- Kedem, B., Chiu, L. S., and Karni, Z. (1990). An Analysis of the Threshold Method for Measuring Area-average Rainfall. *Journal of Applied Meteorology*. 29(1): 3-20.
- Kim, S. B., Shin, H. J., Ha, R., and Kim, S. J. (2012). Spatio-temporal Analysis of Snowfall for 5 Heavy Snowfall Areas in South Korea. *Journal of the Korean Society of Civil Engineers*. 32(2B): 103-111.
- Kim, Y., Kim, S., Kang, N., Kim, T., and Kim, H. (2014). Estimation of Frequency based Snowfall Depth Considering Climate Change using Neural Network. *Journal of Korean Society of Hazard Mitigation*. 14(1): 93-107.
- Kwon, H. H., Brown, C., and Lall, U. (2008). Climate Informed Flood Frequency Analysis and Prediction in Montana using Hierarchical Bayesian Modeling. *Geophysical Research Letters*. 35(5).
- Lee, B. S. (1979). The Distribution of the Fresh Snowfall in South Korea. *Geography Education, Department of Geography*. 9: 224-233.
- Ministry of the Interior and Safety (MOIS) (2018). *Enforcement Decree of The Countermeasures Against Natural Disasters Act*. Sejong: MOIS.
- Ministry of the Interior and Safety (MOIS) (2019). *Comprehensive Plans to Mitigate Natural Disasters*. Sejong: MOIS.
- Park, H. and Chung, G. (2019). Frequency Analysis for Annual Maximum of Daily Snow Accumulations using Conditional Joint Probability Distribution. *Journal of Korea Water Resources Association*. 52(9): 627-635.
- Park, H. S., Jeong, S., and Chung, G. (2014). Frequency Analysis of Future Fresh Snow Days and Maximum Fresh Snow Depth using Artificial Neural Network under Climate Change Scenarios. *Journal of Korean Society of Hazard Mitigation*. 14(6).
- Park, K. W., Kim, D., Shin, J. Y., and Kim, T. W. (2019). Statistical Frequency Analysis of Snow Depth using Mixed Distributions. *Journal of Korea Water Resources Association*. 52(12): 1001-1009.
- Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*. 6(2): 461-464.
- Yoon, P., Kim, T., Yang, J., and Lee, S. (2012). Estimating Quantiles of Extreme Rainfall using a Mixed Gumbel Distribution Model. *Journal of Korea Water Resources Association*. 45(3): 263-274.
- Yu, I., Kim, H., Chung, G., and Jeong, S. (2014). Estimation of Snowfall Frequency and Selection of Appropriate Probability Distribution in Korea. *Journal of the Korean Society of Hazard Mitigation*. 14(4): 101-110.
- Yue, S., Ouarda, T. B. M. J., Bobée, B., Legendre, P., and Bruneau, P. (1999). The Gumbel Mixed Model for Flood Frequency Analysis. *Journal of Hydrology*. 226(1-2): 88-100.

#### **Korean References Translated from the English**

- 김셋별, 신형진, 하림, 김성준 (2012). 우리나라 5 대 대설지역의 적설량 변화 분석. *대한토목학회논문집 B*. 32(2B): 103-111.
- 김연수, 김수전, 강나래, 김태균, 김형수 (2014). 신경망 기법을 이용한 기후변화에 따른 미래 확률적설량 산정. *한국방재학회지*. 14(1): 93-107.
- 박경운, 김동욱, 신지예, 김태웅 (2019). 혼합분포함수를 적용한 최심신적설량에 대한 수문통계학적 빈도분석. *한국수자원학회 논문집*. 52(12): 1001-1009.
- 박희성, 정건희 (2019). 적설 자료의 빈도해석을 위한 확률밀도함수 개선 연구. *한국수자원학회논문집*. 52(9): 627-635.
- 박희성, 정상만, 정건희 (2014). 기후변화 시나리오와 신경망 모형을 이용한 미래 신적설일수와 최심신적설 빈도 분석. *한국방재학회지*. 14(6): 365-377.
- 유인상, 김하룡, 정건희, 정상만 (2014). 우리나라 빈도별 신적설심 산정 및 적정 확률분포형 선정. *한국방재학회지*. 14(4): 101-109.
- 윤필용, 김태웅, 양정석, 이승오 (2012). 혼합 검분포모형을 이용한 확률강우량의 산정. *한국수자원학회 논문집*. 45(3): 263-274.
- 이병설 (1979). 남한의 강설분포에 관한 연구. *지리교육논집*. 9: 224-235.
- 자연재해대책법 개정(시행령, '18.10.23) (2018). 세종: 행정안전부.
- 최홍근, 권현한, 박문형 (2019). 혼합분포 기반 비정상성 강우 빈도해석 기법 개발. *한국수자원학회논문집*. 52(11): 895-904.
- 행정안전부 (2019). *자연재해저감 종합계획 세부수립기준*. 세종: 행정안전부.