

극치강우사상을 포함한 강우빈도분석의 불확실성 분석

Analysis of Uncertainty of Rainfall Frequency Analysis Including Extreme Rainfall Events

김 상 욱* / 이 길 성** / 박 영 진***

Kim, Sang Ug / Lee, Kil Seong / Park, Young-Jin

Abstract

There is a growing dissatisfaction with use of conventional statistical methods for the prediction of extreme events. Conventional methodology for modeling extreme event consists of adopting an asymptotic model to describe stochastic variation. However asymptotically motivated models remain the centerpiece of our modeling strategy, since without such an asymptotic basis, models have no rational for extrapolation beyond the level of observed data. Also, this asymptotic models ignored or overestimate the uncertainty and finally decrease the reliability of uncertainty. Therefore this article provide the research example of the extreme rainfall event and the methodology to reduce the uncertainty. In this study, the Bayesian MCMC (Bayesian Markov Chain Monte Carlo) and the MLE (Maximum Likelihood Estimation) methods using a quadratic approximation are applied to perform the at-site rainfall frequency analysis. Especially, the GEV distribution and Gumbel distribution which frequently used distribution in the fields of rainfall frequency distribution are used and compared. Also, the results of two distribution are analyzed and compared in the aspect of uncertainty.

Keywords : uncertainty, extreme rainfall event, bayesian MCMC, MLE

요 지

극치사상을 예측하기 위한 기존의 빈도분석 결과의 이용에 대한 많은 문제점들이 부각되고 있다. 특히, 통계적 모형을 이용하기 위해서 흔히 사용되는 점근적 모형 (asymptotic model)의 합리적인 검토 없는 외삽 (extrapolation)은 산정된 확률 값을 과대 또는 과소평가하는 문제를 일으켜, 예측결과에 대한 불확실성을 과다하게 산정함으로써 불확실성에 대한 신뢰도를 감소시키는 문제가 있다. 그러므로 본 연구에서는 국내에서 극치강우사상을 포함한 강우자료의 빈도분석에 대한 연구사례를 제공하고 점근적 모형을 사용하는 경우 발생하는 불확실성을 감소시키기 위한 방법론을 제시하였다. 이를 위하여 본 연구에서는 극치강우사상의 빈도분석을 수행하는 데 있어서 최근 들어 여러 분야에서 다양하게 적용되고 있는 Bayesian MCMC (Markov Chain Monte Carlo) 방법을 사용하였으며, 그 결과를 최우추정방법 (Maximum likelihood estimation method)과 비교하였다. 특히 강우사상의 점 빈도분석에 흔히 이용되는 확률밀도함수로 GEV (Generalized Extreme Value) 분포와 Gumbel 분포를 모두 고려하여 두 분포의 결과를 비교하였으며, 이 과정에서 각각의 산정결과 및 불확실성은 근사식을 이용한 최우추정방법과 Bayesian 방법을 이용하여 각각 비교 및 분석되었다.

핵심용어 : 불확실성, 극치강우사상, bayesian MCMC, 최우추정방법

* 교신저자, 국회입법조사처 경제산업조사실 국토해양팀 입법조사관

Legislative Researcher, National Assembly Research Service, Seoul 150-703, Korea (e-mail: sukim@assembly.go.kr)

** 서울대학교 공과대학 건설·환경공학부 교수

Professor, Department of Civil and Environmental Engineering, Seoul National University (e-mail: kilselee@snu.ac.kr)

*** 서일대학교 토목과 부교수

Associate Professor, Department of Civil Engineering, Seoil College (e-mail: profpark@seoil.ac.kr)

1. 서론

강우자료를 이용한 빈도분석과 관련이 있는 많은 국내 외의 기존 연구(Baghirathan and Shaw, 1978; Pitman, 1980; 유철상 등, 2007; 성기원, 2008; 김기욱 등, 2007; 이정규와 추현재, 2006)가 진행되어 있는 반면에 극치강우사상을 포함한 강우자료를 이용한 빈도분석에 대한 연구는 외국의 경우 진행된 사례(Coles and Powell, 1996; Parent and Bernier, 2003; Coles et al., 2003; Sisson et al., 2006; Coles and Tawn, 1996)가 비교적 풍부한 반면 국내에서는 김병식 등(2008)과 김보경 등(2008)의 연구 등을 제외하면 찾아보기가 어렵다.

극치사상을 통계적 모형을 이용하여 평가하기 위해서는 과거자료의 추계학적인 변동(stochastic variation)을 다루기 위한 가정조건을 이용한 점근적 모형(asymptotic model)이 흔히 이용되고 있다. 그러나 이와 같은 기존의 통계적 방법은 측정된 과거자료의 외측 구간에서의 외삽(extrapolation)에 대한 합리적인 검토없이 사용됨으로써 최종적으로 산정된 확률 값이 과대 또는 과소평가되는 문제가 있고, 예측결과에 대한 불확실성을 과다하게 산정함으로써 불확실성에 대한 신뢰도를 감소시키는 문제가 있다.

특히 최근에는 극치사상을 예측하기 위한 기존의 표준적인 통계분석 절차에 대한 문제점들이 통계학 분야(Malakoff, 1999)를 비롯하여 환경 및 생물학 분야(Ludwig, 1996; Smith, 1989), 수자원공학 분야(Lee and Kim, 2008) 등에서 다양하게 제시되고 있으며, 제시되는 문제점들은 극치사상을 나타내는 확률 값을 과소평가함으로써 일어날 수 있는 다양한 피해들과 관련된 연구들이라 할 수 있다. Pinter et al. (2001)은 특히 각각 다른 유역의 토지이용도의 변환에 따른 하천의 동적구조가 가지고 있는 특징에 초점을 맞추어 홍수피해의 통계적 평가절차를 새롭게 수립해야 할 필요성이 있음을 제시한바 있다. 또한 Coles et al. (2003)은 점근적 모형을 사용한 극치사상의 통계적 평가절차에 대한 문제를 제시하면서 특히 점근적인 모형을 사용하여 외삽구간에 대한 평가를 수행할 경우에는 반드시 모형에 의한 예측결과에 대한 불확실성을 적절히 산정하여 포함시킬 것과 분석대상 자료의 정상성(stationarity)에 대한 명확한 이해를 제공해야함을 주장한 바 있다.

그러므로 본 연구에서는 국내에서 극치강우사상을 포함한 강우자료의 빈도분석에 대한 연구사례를 제공하고 점근적 모형을 사용하는 경우 발생하는 불확실성을 감소시키기 위한 방법론을 제공하는 것을 목적으로 하였다. 이를 위하여 본 연구에서는 극치강우사상의 빈도분석을 수행하는 데 있어서 최근 들어 여러 분야에서 다양하게 적용되고 있는 Bayesian 방법을 사용하였으며, 그 결과를

최우추정방법(maximum likelihood estimation method)과 비교하였다. 특히 강우사상의 점 빈도분석에 흔히 이용되는 확률밀도함수로 GEV(Generalized Extreme Value) 분포와 Gumbel 분포를 모두 고려하여 두 분포의 결과를 비교하였으며, 이 과정에서 각각의 산정결과 및 불확실성을 근사식을 이용한 최우추정방법과 Bayesian 방법으로 나누어 두 분포에 따른 결과를 비교 및 분석하였다.

근사식을 사용한 신뢰구간 산정방법은 확률밀도함수의 모수를 산정함에 있어서 정규분포성(normality), 선형성(linearity) 등의 가정이 필요하므로, 불확실성을 산정함에 있어서 비현실적인 값을 산정하거나 과대 추정되는 경우가 있는 것으로 알려져 있다(Reis Jr. and Stedinger, 2005). 그러나 Bayesian 접근방법을 사용한 모수 및 불확실성의 추정은 근사식을 사용하기 위한 가정 조건이 필요하지 않기 때문에 특히 불확실성을 표현하는 데 있어서 근사식을 사용한 방법보다 우월한 것으로 알려져 있으며, 최근 들어 Bayesian 방법론을 이용한 수공학 분야를 포함한 과학 분야에서의 불확실성을 산정하기 위한 많은 연구들이 진행된 바 있다(Coles and Powell, 1996; Kuczera and Parent, 1998; Kuczera, 1999; O'Connell et al., 2002; Reis Jr. et al., 2005; Reis Jr. and Stedinger, 2005; Kavetski et al., 2006; Seidou et al., 2006; Lee and Kim, 2008; 김상욱과 이길성, 2008a; 김상욱과 이길성, 2008b; 김상욱과 이길성, 2008c; 김상욱과 이길성, 2008d)

특히 김상욱과 이길성(2008a, 2008b, 2008d)은 Bayesian 방법 중 Bayesian MCMC방법에 의한 불확실성의 산정이 기존 방법보다 불확실성을 효과적으로 감소시킬 수 있다는 연구결과를 제시한 바 있으므로 본 연구에서도 이와 같은 이론을 사용하여 극치값을 포함한 강우빈도분석에서도 유사한 결론을 얻을 수 있을지를 살펴보았다. 단, 김상욱과 이길성(2008c)의 연구는 Bayesian회귀분석에 대한 연구로 최우추정방법과의 비교 등을 참고하였다.

즉, 본 연구에서는 기존에 국내외에서 진행된 일반적인 강우빈도분석에 대한 연구, 극치값을 포함한 강우빈도분석에 대한 연구 및 불확실성에 대한 연구를 기반으로 하여 극치값을 포함한 강우빈도분석에서 나타날 수 있는 불확실성을 Bayesian 방법론 중 하나인 Bayesian MCMC 방법을 이용하여 산정함으로써 극치값이 포함된 자료의 강우빈도분석에 있어서의 불확실성을 보다 현실적으로 나타내고자 하였다.

2. 확률적 추정과 불확실성의 정량화: Bayesian MCMC와 2차 근사를 이용한 MLE

어떤 확률밀도함수의 모수를 추정하기 위해서는 다양

한 방법이 적용될 수 있지만 본 연구에서는 최우추정방법과 확률적 추정방법인 Bayesian 방법을 사용하였다. Bayesian 방법을 사용하는 경우에는 불확실성의 산정이 적용 과정 중에 얻어지는 반면에 최우추정방법을 사용하는 경우에는 불확실성을 산정하기 위해서는 선형성 등을 가정조건으로 하는 2차 근사식을 이용한 불확실성 산정방법을 별도로 적용해야 한다.

베이즈의 정리는 A가 먼저 발생하고 그 후에 B가 발생하는 두 개의 사건 A, B가 서로 종속적일 경우 A의 사건에 의해 B 사건의 확률이 달라진다는 것이다. 베이즈의 정리를 연속 확률밀도함수(probability density function)로 나타내면 베이즈의 정리는 Eq. (1)과 같이 표현될 수 있다.

$$\pi(\theta|x_1, x_2, \dots, x_n) = \frac{f(x_1|\theta) \cdots f(x_n|\theta)\pi(\theta)}{\int_{\theta} f(x_1|\theta) \cdots f(x_n|\theta)\pi(\theta)d\theta} \quad (1)$$

Eq. (1)에서 좌변의 $\pi(\theta|x_1, x_2, \dots, x_n)$ 는 사후분포(posterior distribution), 우변 분자의 $\pi(\theta)$ 는 사전분포(prior distribution)라 명명되며, 우변의 분모는 상수로서 주변분포(marginal distribution)이고, 우변의 분자의 $f(x_1|\theta) \cdots f(x_n|\theta)$ 는 발생할 수 있는 모든 가능성을 고려한 우도함수(likelihood function)이다.

Bayesian 방법을 이용한 모수의 추정은 자료로부터 얻은 모수에 대한 정보와 모수에 대한 과거의 경험 또는 주관을 사전분포로 표현함으로써 보다 정확한 모수의 불확실성에 대한 탐색에 그 목적이 있다고 할 수 있다. 그러나 Bayesian 방법을 이용하여 사후분포를 계산하는 것은 쉬운 작업이 아니다. 특히 Eq. (1)의 우변의 분모의 적분은 결정된 확률밀도함수에 따라 적분이 가능할 수도 있으나, 대부분의 확률밀도함수는 수학적으로 적분하기 어려운 경우가 많다. 또한 확률밀도함수로 표현되는 우도함수와 사전분포를 이용하여 모수의 추출이 가능한 형태의 사후분포를 산정하는 것은 수학적으로 어려운 경우가 많다. 그러나 최근 계산능력의 하드웨어부분의 발전과 주변 확률밀도함수의 적분이 필요 없는 Bayesian 계산방법의 발전으로 인하여 서론에 언급한 연구사례와 같이 최근 들어 수자원공학 분야에서도 다시 활발히 적용되고 있는 실정이다.

최근에는 Bayesian 방법에 의해 유도된 사후분포로부터 모수를 추정하기 위하여 Bayesian MCMC (Markov Chain Monte Carlo)방법이 많이 사용되고 있다. Bayesian MCMC방법이란 마코프 연쇄(Markov chain)를 이용하여 모수간의 관계를 구성하고 이를 상당히 큰 수만큼 반복하는 몬테카를로 적분기법(Monte carlo integration)을 이용하여 최종적으로 모수의 통계적 특성을 산정하는 방

법이다. 여러 가지 Bayesian 계산 방법 중에서 가장 활발히 사용되고 있는 알고리즘은 Metropolis-Hastings 알고리즘으로 기본적인 개념은 Metropolis et al. (1953)에 의하여 만들어 졌으나 최근 들어 개선되어져 활발히 이용되고 있다. 마코프 연쇄는 불변(invariant)인 임의의 함수, f 와 임의의 제안분포(proposal distribution), q 를 이용하여 다음과 같이 정의될 수 있다. 또한 Eq. (2)에서 f 를 구하고자 하는 사후분포, $\pi(\theta|D)$ 가 불변하다는 가정 하에 대체하면 좌변 또는 우변이 큰 부등식 Eq. (3)을 얻을 수 있다.

$$q(\theta_{j+1}|\theta_j)f(\theta_j|D) = q(\theta_j|\theta_{j+1})f(\theta_{j+1}|D) \quad (2)$$

$$q(\theta_{j+1}|\theta_j)\pi(\theta_j|D) > q(\theta_j|\theta_{j+1})\pi(\theta_{j+1}|D) \quad (3)$$

Eq. (3)의 부등식을 등식으로 바꿀 수 있는 값은 반드시 존재하는데, 이를 ρ 로 표현하여 고려하면 다음과 같은 등 Eq. (4)를 얻을 수 있다.

$$\rho(\theta_j, \theta_{j+1})q(\theta_{j+1}|\theta_j)\pi(\theta_j|D) = q(\theta_j|\theta_{j+1})\pi(\theta_{j+1}|D) \quad (4)$$

Eq. (4)를 ρ 에 대하여 정리하면 최종적인 Eq. (5)를 얻을 수 있고, ρ 를 채택확률(acceptance probability)이라 한다.

$$\rho(\theta_j, \theta_{j+1}) = \min \left[\frac{q(\theta_j|\theta_{j+1})\pi(\theta_{j+1}|D)}{q(\theta_{j+1}|\theta_j)\pi(\theta_j|D)}, 1 \right] \quad (5)$$

얻어진 채택확률을 이용하여 추정하고자 하는 모수를 구하는 Metropolis-Hastings 알고리즘은 다음과 같은 과정으로 표현할 수 있다.

- Step 1) $j=0$ 에서의 임의의 θ_0 를 선정한다.
- Step 2) 제안분포로부터 제안 모수 θ_* 를 무작위적으로 생성한다.
- Step 3) Eq. (5)로부터 채택확률을 계산한다.
- Step 4) 0과 1사이의 균일분포로부터 무작위수 u 를 생성한다.
- Step 5) 만약 $u < \rho$ 이면, θ_* 를 θ_{j+1} 로 교체하고 반대의 경우에는 θ_j 를 θ_{j+1} 로 교체한다.
(즉 $u > \rho$ 인 경우에는 제안분포로부터 생성된 θ_* 를 사용하지 않는다.)
- Step 6) j 를 1 증가시키고 Step 2)로 돌아간다.
- Step 7) 충분히 큰 수만큼 위 과정을 반복한다.
- Step 8) 발생된 모수를 이용하여 평균값을 계산한다.

위의 알고리즘을 이용하여 추정된 모수들은 최종적으

로 몬테카를로 적분을 통하여 추정된 모수들의 평균과 같은 통계적 특성치를 근사적으로 얻을 수 있으며, 기존 통계학에서 사용되는 신뢰구간 (confidence interval)에 해당되는 신용구간 (credible interval)도 얻을 수 있고 이를 이용하여 추정된 모수의 불확실성을 표현할 수 있다. 반복하여 추출된 모수의 개수, n 과 유의수준 (significance level), α 에서의 Bayesian MCMC 방법에 의한 신용구간, $100(1-\alpha)\%$ 는 다음 Eq. (6)과 같다.

$$(\theta_{[(\alpha/2)n]}, \theta_{[(1-\alpha/2)n]}) \quad (6)$$

최우추정 (maximum likelihood estimation) 방법은 특정 공간에서의 모집단을 임의의 확률밀도함수로 가정 한 후 자료가 추출될 수 있는 확률을 나타내는 우도함수 (likelihood function)를 최대가 되도록 하는 확률밀도함수의 모수를 결정하는 일종의 최적화 방법이라 할 수 있다. 즉 n 개의 자료와 이를 나타낼 수 있는 모수가 포함되어 있는 임의의 확률밀도함수, $f_X(x)$ 가 존재하면 우도함수, L 은 다음과 같이 표현될 수 있다.

$$L(\theta) = \prod_{i=1}^n f_X(x_i | \theta) \quad (7)$$

최우추정 방법은 우도함수를 최대화시키도록 하는 θ 를 최적화시키는 방법이므로, 우도함수의 형태에 따른 여러 가지 최적화방법을 고려할 수 있다. 우도함수에 따른 최적화방법을 적용하여 추정하고자 하는 모수를 구한 이후에는 모수의 불확실성을 나타내기 위하여 다음 식과 같이 2차 근사식을 이용한 우도비함수 (likelihood function ratio)를 이용할 수 있으며, 2차 근사식은 $\theta = \hat{\theta}$ 에서 최대 값을 가져야 하므로 다시 임의의 상수 I 를 사용하여 2차 식으로 표현될 수 있다.

$$\log RL(D | \theta) = \log \frac{L(D | \theta)}{L(D | \hat{\theta})} \cong a + b\theta + c\theta^2 \cong -\frac{1}{2}I(\theta - \hat{\theta})^2 \quad (8)$$

여기서, RL은 상대우도 (relative likelihood)를 나타내며 D 는 주어진 자료, $\hat{\theta}$ 는 θ 의 추정치이고 a, b, c 는 상수이다. Meeker and Escobar (1998)는 Eq. (8)에서 θ 의 불확실성을 추정하기 위하여 정상성 (Normality)을 가정으로 하는 다음의 근사식을 제안한 바 있다.

$$\frac{\partial \log L}{\partial \theta} \cong -I(\theta - \hat{\theta}) \sim N(0, I) \quad (9)$$

Eq. (9)으로부터 $\hat{\theta} \sim N(0, I^{-1})$ 임을 알 수 있고 유의수준 α 에서 다음과 같은 신뢰구간을 유도할 수 있으며, 이를 2차 근사식에 의한 정량화된 불확실성으로 간주할 수 있다.

$$\Pr\left[\hat{\theta} - \frac{z_{(1-\alpha/2)}}{\sqrt{I}} < \theta \leq \hat{\theta} + \frac{z_{(1-\alpha/2)}}{\sqrt{I}}\right] \approx 1 - \alpha \quad (10)$$

최근에는 이와 같은 최우추정방법보다 확률가중모멘트 방법 (probability weighted moment method)이 정확도가 우수한 결과를 나타냄에 따라 이 방법이 많이 사용되고 있으나, 본 연구에서는 정확도 높은 결과의 추정보다는 Bayesian방법의 효율적인 불확실성에 대한 비교평가가 목적이므로 비교적 구성이 쉽고 비교가 간단한 최우추정 방법만을 비교대상으로 선정하였다. 그러나 향후 확률가중모멘트방법을 이용하여 그 정확도와 근사법에 의한 불확실성을 산정하여 Bayesian방법의 결과와 비교하는 것도 의미있는 연구가 될 수 있으리라 판단된다.

3. 확률밀도함수의 선정 및 사전분포의 결정방법

강우자료의 길이가 충분하다고 판단되는 경우 해당 지점에서의 점 빈도분석 (at-site frequency analysis)을 수행할 수 있으며, 이를 위해서는 사용하려는 자료의 기본적인 통계적 특성의 계산과 더불어 빈도해석의 기본적 조건인 자료의 무작위성에 대한 예비해석을 실시해야 한다. 이와 같은 일련의 과정은 본 논문의 적용부분에 그 결과를 나타내기로 한다. 예비해석이 완료되면, 대상자료의 빈도해석에 사용할 확률밀도함수를 선정해야 한다. 건설교통부 (2000)는 강우 빈도분석에 많이 사용되는 13가지의 확률밀도함수를 우리나라의 68개 대상지점에 적용하면서 χ^2 -검정, Kolmogorov-Smirnov검정, Cramer Von Mises 검정, Probability Plot Correlation Coefficient (PPCC)검정을 이용하여 대부분 지점에서 Gumbel 분포가 타당함을 제시한 바 있다. 그러나 Coles et al. (2003)은 극치강우사상이 포함된 강우자료를 이용한 빈도분석을 수행하면서, Gumbel 분포와 GEV (Generalized Extreme Value) 분포를 비교하여 극치강우자료의 경우 GEV 분포의 유용성을 제시한 바 있다.

그러므로 본 연구에서는 극치강우사상을 포함한 강우자료의 빈도분석이 연구의 목적이므로 국내에서는 GEV 분포를 이용한 극치강우사상의 분석 사례가 부족하므로 별도의 검정절차 없이 Gumbel 분포와 GEV 분포 두 개를 선정하여 빈도분석을 진행하였으며, 최종적으로는 두 분포에 따른 불확실성을 정량화하여 나타내었다. 본 연구에서 사용된 GEV 분포의 확률밀도함수 (probability density function), 확률분포함수 (cumulative distribution function) 및 우도함수 (likelihood function)을 나타내면 각각 Eq. (11), Eq. (12), Eq. (13)과 같다. GEV 분포에서 κ 가 0일 때를 GEV-1 또는 Gumbel 분포라 명명하여 사

용하고 있으며, Gumbel 분포의 확률밀도함수, 확률분포 함수 및 우도함수는 Eq. (14), Eq. (15), Eq. (16)과 같다.

$$f(x) = \frac{1}{\alpha} \left[1 - \kappa \left(\frac{x - \xi}{\alpha} \right) \right]^{1/\kappa - 1} \exp \left[- \left(1 - \kappa \left(\frac{x - \xi}{\alpha} \right) \right)^{1/\kappa} \right] \quad (11)$$

$$F(x) = \exp \left[- \left(1 - \kappa \left(\frac{x - \xi}{\alpha} \right) \right)^{1/\kappa} \right] \quad (12)$$

$$L(\mathbf{x} | \alpha, \kappa, \xi) = \prod_{i=1}^N \left\{ \frac{1}{\alpha} \left[1 - \kappa \left(\frac{x_i - \xi}{\alpha} \right) \right]^{1/\kappa - 1} \exp \left[- \left(1 - \kappa \left(\frac{x_i - \xi}{\alpha} \right) \right)^{1/\kappa} \right] \right\} \quad (13)$$

$$f(x) = \frac{1}{\alpha} \exp \left[- \left(\frac{x - \xi}{\alpha} \right) \right] - \exp \left[- \left(\frac{x - \xi}{\alpha} \right) \right] \quad (14)$$

$$F(x) = \exp \left[- \exp \left[- \left(\frac{x - \xi}{\alpha} \right) \right] \right] \quad (15)$$

$$L(\mathbf{x} | \alpha, \xi) = \prod_{i=1}^N \left\{ \frac{1}{\alpha} \exp \left[- \left(\frac{x_i - \xi}{\alpha} \right) \right] - \exp \left[- \left(\frac{x_i - \xi}{\alpha} \right) \right] \right\} \quad (16)$$

위의 식에서 α 는 척도모수 (scale parameter), κ 는 형상모수 (shape parameter)이며 ξ 는 위치모수 (location parameter)이다. 확률밀도함수의 선정에 따른 우도함수가 결정되었으므로 Eq. (1)을 GEV 분포와 Gumbel 분포함수를 이용하여 표현하면 각각 다음의 식과 같다.

$$\pi(\alpha, \kappa, \xi | \mathbf{x}) = \frac{L(\mathbf{x} | \alpha, \kappa, \xi) \pi(\alpha, \kappa, \xi)}{\int_0^\infty \int_0^\infty L(\mathbf{x} | \alpha, \kappa, \xi) \pi(\alpha, \kappa, \xi) d\alpha d\kappa d\xi} : \text{GEV} \quad (17)$$

$$\pi(\alpha, \xi | \mathbf{x}) = \frac{L(\mathbf{x} | \alpha, \xi) \pi(\alpha, \xi)}{\int_0^\infty \int_0^\infty L(\mathbf{x} | \alpha, \xi) \pi(\alpha, \xi) d\alpha d\xi} : \text{Gumbel} \quad (18)$$

여기서, $\pi(\alpha, \kappa, \xi)$ 및 $\pi(\alpha, \xi)$ 는 형상모수, 척도모수, 위치모수의 결합확률밀도함수로써 모수의 특징을 나타내는 사전분포이다. Bayesian 방법을 적용하기 위해서는 적절한 사전분포를 선정하는 것이 가장 중요하고도 어려운 작업이며, 사전분포는 크게 무정보적 사전분포 (noninformative prior distribution)와 정보적 사전분포 (informative prior distribution)로 구분할 수 있다. 과거 컴퓨터의 계산능력이 발전하지 않았을 때에는 Eq. (17) 또는 Eq. (18)의 사후분포를 계산하기 위해서 무정보적 사전분포나 공액사전분포 등을 이용하여 계산을 쉽게 할 수 있도록 사후분포를 변경하여 모수를 추정하였으나, 최근에는 하드웨어 및 Metropolis-Hastings 알고리즘과 같은 소프트웨어의 발전에 힘입어 다양한 사전분포의 적용사례가 연구되고 있다.

특히 빈도분석과 관련된 사전분포의 구축은 크게 년최대시계열자료 이외에 일정 문지방값 (threshold value)을 이용한 부분시계열자료를 이용하는 방법, 인근 지역의 정보 (regional information)를 이용하여 사전분포화 하는 방

법, 계측기를 이용한 계측이전시기의 자료 (paleohydrologic data)를 이용하는 방법, 전문가의 지식을 주관적 관점에서 사전분포화하는 방법 등이 사용될 수 있다. 그러나 위의 방법들 중에서 홍수의 흔적 등을 이용하는 계측기 이전시기의 자료를 이용하는 방법이나 전문가의 지식을 이용하는 사전분포는 객관적인 결과에 불확실성을 추가할 수 있으므로 바람직하지 않다고 볼 수 있다. Lee and Kim (2008)은 저수량의 빈도분석을 Bayesian 방법을 이용하여 수행하면서 에르고딕 (ergodic)가정을 이용한 주변 지역자료를 이용한 정보적 사전분포를 구축하고 이를 무정보적 사전분포와 비교함으로써 지역정보를 이용한 정보적 사전분포의 우수성을 입증한 바 있다. 본 연구에서도 이와 같은 방법을 이용함으로써 정보적 사전분포를 구축하여 Eqs. (17) and (18)의 사후분포로부터 필요한 개수의 모수를 추정하였다.

4. 자료의 선정 및 자료의 예비해석

2002년 우리나라에 큰 재해를 유발한 제15호 태풍 루사 (RUSA)는 종전 일최대강수량을 기록했던 장릉관측소의 547.4mm를 크게 초과하여 가장 피해가 극심했던 강원도 지방에 위치한 강릉관측소에서는 870.5mm라는 기록적인 일최대강수량을 기록한 태풍이었다. 이때 강릉관측소에서 관측된 시간최대강수량 역시 종전의 기록을 크게 초과한 80.0mm로 기록된바 있다. 그러므로 본 연구에서는 극치강수량사상에 대한 빈도분석을 목적으로는 2002년 루사의 영향 하에 역대 최대 일최대강수량을 기록한 강릉관측소를 선정하였으며, 이 당시 태풍의 영향을 크게 받지 않은 서해안에 위치한 인천관측소를 추가로 선정하여 두 관측소에서의 강우자료를 이용하여 빈도분석을 수행하고 그 결과를 비교하였다.

확률강우량을 산정하기 위한 빈도해석을 실시하기 전에 대상 자료를 도시하여 그 변동성 또는 경향성을 파악하는 것은 자료의 개략적인 성질을 예측하는데 많은 도움을 줄 수 있다. 그러므로 먼저 인천과 강릉관측소의 1961년부터 2007년까지 47년간 동안 측정된 17,155개의 일강우자료를 수집한 후, 이로부터 년최대강수량 (annual maximum precipitation)을 산정하여 각각 Figs. 1 and 2에 나타내었다. 태풍 루사의 내습이 있었던 2002년 제외하면 두 관측소는 대략적으로는 년최대강수량의 분포가 비슷한 것을 알 수 있었다. 단, 2002년 태풍 루사로 인하여 강릉관측소의 2002년 년최대강수량은 870.5mm로 기록되어 이전의 강우기록에 비해 2배를 초과하므로 충분히 극치강우사상 (extreme rainfall event)으로 간주할 수 있음을 알 수 있다. 두 관측지점에서 관측된 자료의 기본적인 통계량은

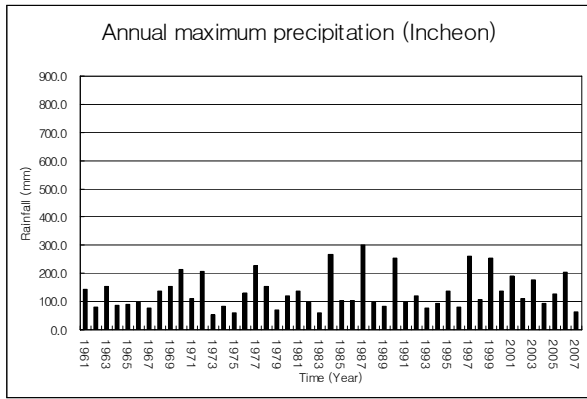


Fig. 1. Annual Maximum Precipitation at Incheon Rainfall Station

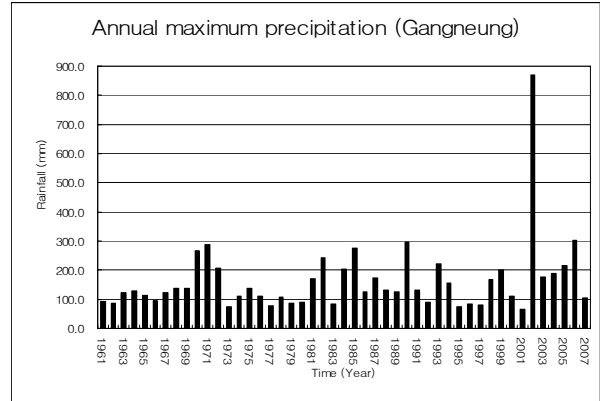


Fig. 2. Annual Maximum Precipitation at Gangneung Rainfall Station

Table 1. Statistical Characteristics in Two Rainfall Station

| Station | Mean (mm) | Standard Deviation (mm) | Coefficient of Variation | Coefficient of Skewness | Coefficient of Kurtosis |
|-----------|-----------|-------------------------|--------------------------|-------------------------|-------------------------|
| Incheon | 133.6 | 63.2 | 0.473 | 1.066 | 3.395 |
| Gangneung | 163.5 | 124.0 | 0.758 | 4.237 | 25.512 |

Table 2. Results of Preliminary Statistical Tests

| Tests for Randomness | Station | Test Statistic | Table (95%) | Result |
|---------------------------------|-----------|----------------|-------------|----------|
| Anderson corr. test | Incheon | 0.000 | 0.55 | accepted |
| | Gangneung | 0.000 | | accepted |
| Run Test | Incheon | 1.337 | 1.960 | accepted |
| | Gangneung | 1.825 | | accepted |
| Spearman Rank corr. coeff. test | Incheon | 0.866 | 2.016 | accepted |
| | Gangneung | 1.283 | | accepted |
| Turning Point Test | Incheon | 1.411 | 1.960 | accepted |
| | Gangneung | 0.706 | | accepted |

Table 1에 나타내었는데 강릉관측소의 자료는 인천관측소의 강우자료의 통계적 특성에 비해 왜곡도 (Coefficient of skewness)와 첨예도 (Coefficient of kurtosis)가 굉장히 크게 산정된 것을 알 수 있어, 강릉관측소의 분포는 인천관측소의 분포보다 상당히 왼쪽으로 치우친 분포이며 분포의 중앙부분에 관측치가 몰려있는 상당히 뾰족한 분포임을 대략적으로 알 수 있다.

자료의 도서를 통한 개략적인 통계적 성질의 예측과 함께 빈도해석을 수행하기 전에 반드시 거쳐야 할 과정이 수집한 자료가 통계적으로 무작위성 (randomness)를 가지고 있는지를 검토하는 것이다. 이와 같은 무작위성 검증은 건교부 (2000)가 수행한 것과 동일한 Anderson correlation test, Run test, Spearman rank correlation coefficient test 및 Turning point test의 4가지 검정을 이용하였으며, 그 결과를 Table 2에 나타내었다. Table 2에

서 나타낸 바와 같이 인천 및 강릉관측소에서 산정된 연 최대강수량자료는 빈도해석을 수행하기에 충분한 무작위성을 가지고 있는 것으로 나타났다.

5. 2차 근사를 이용한 최우추정법과 Bayesian MCMC 방법의 결과 비교

5.1 2차 근사를 이용한 최우추정법의 수행

최우추정 (Maximum likelihood estimation, MLE)방법은 우도함수를 최대화하는 모수를 결정하는 최적화 문제이다. 본 연구에서는 우도함수를 선형화하기 위하여 양변에 로그를 취한 후, 형상모수와 척도모수에 대한 편미분 방정식의 값이 0이 된다는 편미분 방정식을 구성한 후, 이를 반복법을 이용하여 모수의 최적값을 추정하였다. 그러나 이러한 도해적인 방법을 통해 최적화 알고리즘을 경사

범만으로 결정하여 사용하는 것은 이후 추정결과에서 모수들의 변화에 따라 불연속적인 모수들로 추정될 수 있는 우려가 있으므로, 최우추정방법을 이용한 모수 추정 시 지역최적화 알고리즘을 사용함으로써 보다 나은 최우추정결과를 얻을 수 있을 것으로 판단된다. 본 연구의 주된 관심사는 추정된 모수뿐만 아니라, 모수의 추정과정에서 발생하는 불확실성을 산정하는 것이므로, 이를 나타내기 위하여 앞에서 제시한 2차 근사식을 이용한 신뢰구간의 범위에 대한 근사식을 이용하였다.

5.2 사전분포의 구성 및 Metropolis-Hastings 알고리즘의 수행

앞에서 기술한 바와 같이 사전분포의 선정은 Bayesian 방법을 수행하는 데 있어서 가장 중요한 과정이며, 본 연구에서는 지역정보를 이용하여 얻어진 사전분포를 결정하였다. 먼저 지역정보를 얻기 위하여 건교부(2000)가 사용하였던 강우관측소 중에서 인천관측소 및 강릉관측소의 인근에 위치한 10개의 강우관측소(강화, 서울, 수원, 이천, 양평, 대관령, 속초, 태백, 춘양, 울진)에서 수집할 수 있는 최대기록년도의 강우자료를 수집하였다. 수집된 10개 관측소의 강우자료를 Gumbel 및 GEV를 이용하여 각 분포의 모수를 추정하였다(Table 3). 단, 사전분포를 합리적으로 구축하기 위해서는 수집된 자료의 동질성 분석을 수행하고 이에 따라 분할된 그룹마다 각각의 사전분포를 다르게 구축하여 최종결과를 산정하는 것이 가장 합리적인 방법이

지만, 연구의 목적이 합리적인 사전분포의 구축이기보다는 Bayesian 방법론을 이용한 불확실성의 정량화에 대한 적정성 검토이므로 수집된 자료가 동질하다는 가정 하에 별도의 그룹화 작업을 진행하지 않고 연구를 진행하였다.

Table 3에서 얻어진 10개 관측소에 대한 모수추정결과를 위치모수(location parameter), 척도모수(scale parameter), 형태모수(shape parameter)별로 구분한 뒤, 각각의 모수를 나타내는데 있어 객관적으로 일정정도 적합하며 이후의 계산에 있어 간편성이 있다고 판단되는 확률분포를 사전분포로 선정하였다. 이러한 사전분포의 선정은 Bayesian방법에 있어서 사용자의 주관적인 입장으로 포함시키는 과정으로 본 연구에서는 무조건적인 주관적 요소에 의한 사전분포의 선정에 있어 나타나는 단점을 극복하기 위해 각 분포를 확률도시함으로써 일정정도의 적절성을 확보하였다고 할 수 있다.

즉 Gumbel 및 GEV 분포의 위치모수 및 척도모수는 모두 2변수 Weibull 분포에 일정정도 적합되는 것으로 나타났다. GEV 분포의 형태모수는 지수분포에 일정정도 적합한 것으로 나타났다. 단, 이와 같은 적합성에 대한 표시는 사전분포의 적합성에 대한 일정정도의 객관적 검토를 위한 것이기 때문에 정량적으로 수행하지 않고 확률도시를 통해 정성적으로만 수행하였으며, 각 분포선정의 적절성을 나타내는 확률도시 결과는 Fig. 3에 나타내었다. 이 그림 중 Fig. 3의 (c)와 (d)는 다른 확률도시 결과에 비해 상대적으로 잘 맞는다고 판단하기 어려우나, 이를 보다 정량

Table 3. Results of Estimates at 10 Stations

| Distribution | Station | Parameters | | |
|--------------|---------------|------------|--------|--------|
| | | Location | Scale | Shape |
| Gumbel | Ganghwa | 97.589 | 31.219 | 0 |
| | Seoul | 125.252 | 47.335 | 0 |
| | Suwon | 115.254 | 42.968 | 0 |
| | Icheon | 95.197 | 33.744 | 0 |
| | Yangpyeong | 104.437 | 37.356 | 0 |
| | Daegwallyeong | 82.543 | 33.173 | 0 |
| | Sokcho | 95.058 | 37.326 | 0 |
| | Taebaek | 39.313 | 28.982 | 0 |
| | Chunyang | 89.744 | 31.317 | 0 |
| | Uljin | 120.814 | 46.882 | 0 |
| GEV | Ganghwa | 95.643 | 30.119 | -0.099 |
| | Seoul | 122.274 | 45.026 | -0.120 |
| | Suwon | 112.156 | 40.486 | -0.136 |
| | Icheon | 94.884 | 33.555 | -0.017 |
| | Yangpyeong | 102.491 | 36.369 | -0.091 |
| | Daegwallyeong | 82.015 | 32.873 | -0.03 |
| | Sokcho | 91.019 | 33.254 | -0.183 |
| | Taebaek | 30.83 | 10.994 | -0.901 |
| | Chunyang | 87.767 | 29.805 | -0.117 |
| | Uljin | 118.447 | 45.450 | -0.090 |

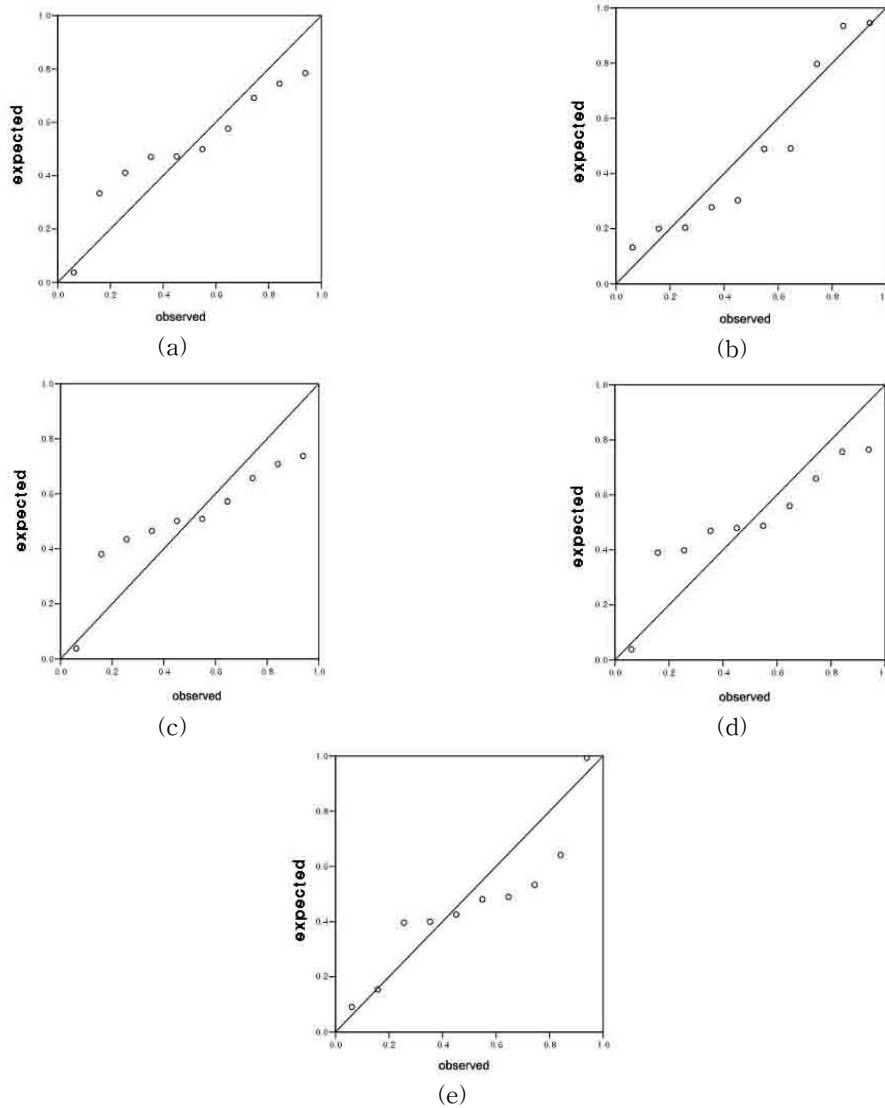


Fig. 3. Probability Plots of the Selected Distribution

(a) Gumbel: location, (b) Gumbel: scale, (c) GEV: location, (d) GEV: shape, (e) GEV: shape

화하여 결정하는 과정을 거치지 않고 일정정도 잘 맞는다
는 주관적인 판단 하에 사전분포를 결정하였다. 또한 항
후에는 확률도시의 정확도에 따라 사전분포를 여러 가지
로 선정하여 이를 이용하여 결과를 분석함으로써 최종적
인 추정치가 어떻게 변화하는지도 연구해 볼 필요가 있을
것으로 판단된다.

Gumbel 분포의 사후분포식 Eq. (17)에서 사전분포인
 $\pi(\alpha, \xi)$ 는 각각의 모수가 통계적으로 독립적이라는 가정
하에 다음과 같은 식으로 나타낼 수 있다.

$$\pi(\alpha) = \frac{b_1}{a_1} \left(\frac{\alpha}{a_1} \right)^{b_1-1} \exp \left[- \left(\frac{\alpha}{a_1} \right)^{b_1} \right] \quad (19)$$

$$\pi(\xi) = \frac{d_1}{c_1} \left(\frac{\xi}{c_1} \right)^{d_1-1} \exp \left[- \left(\frac{\xi}{c_1} \right)^{d_1} \right] \quad (20)$$

$$\therefore \pi(\alpha, \xi) = \frac{b_1 d_1}{a_1 c_1} \left(\frac{\alpha}{a_1} \right)^{b_1-1} \left(\frac{\xi}{c_1} \right)^{d_1-1} \exp \left[- \left(\left(\frac{\alpha}{a_1} \right)^{b_1} + \left(\frac{\xi}{c_1} \right)^{d_1} \right) \right] \quad (21)$$

또한 GEV 분포의 경우 사전분포, $\pi(\alpha, \kappa, \xi)$ 는 Eq. (19)
와 Eq. (20) 외에 형태모수에 대한 Eq. (22)를 고려함으로
써 최종적인 사후분포인 Eq. (23)으로 나타내질 수 있다.

$$\pi(\kappa) = \exp(-\lambda \kappa) \quad (22)$$

$$\therefore \pi(\alpha, \xi) = \frac{b_2 d_2}{a_2 c_2} \left(\frac{\alpha}{a_2} \right)^{b_2-1} \left(\frac{\xi}{c_2} \right)^{d_2-1} \exp \left[- \left(\left(\frac{\alpha}{a_2} \right)^{b_2} + \left(\frac{\xi}{c_2} \right)^{d_2} + \lambda \kappa \right) \right] \quad (23)$$

위의 Eqs. (21) and (23)에서 $a_1, b_1, c_1, d_1, a_2, b_2, c_2, d_2$
및 λ 는 위의 Table 4에서 추정된 값으로 초모수(hyper
parameter)이며, 이 값들을 최우추정방법을 사용하여 추
정한 결과는 다음과 같다.

Metropolis-Hastings 알고리즘을 수행하기 위해서는
적절한 제안분포(Proposal distribution)가 필요하며, 최종

Table 4. Results of Estimates for Derivation of Prior Distribution

| Distribution | Parameter | Selected Distribution | Estimates of Selected Distribution | |
|--------------|--------------------|-----------------------|------------------------------------|-----------|
| Gumbel | Location (ξ) | 2-parameter Weibull | Scale (d_1) | 109.54183 |
| | | | Shape (c_1) | 3.18719 |
| | Scale (α) | 2-parameter Weibull | Scale (a_1) | 39.81675 |
| | | | Shape (b_1) | 6.15267 |
| GEV | Location (ξ) | 2-parameter Weibull | Scale (d_2) | 109.22226 |
| | | | Shape (c_2) | 2.57093 |
| | Scale (α) | 2-parameter Weibull | Scale (a_2) | 39.31696 |
| | | | Shape (b_2) | 2.54266 |
| | Shape (κ) | Exponential | Scale (λ) | 5.60538 |

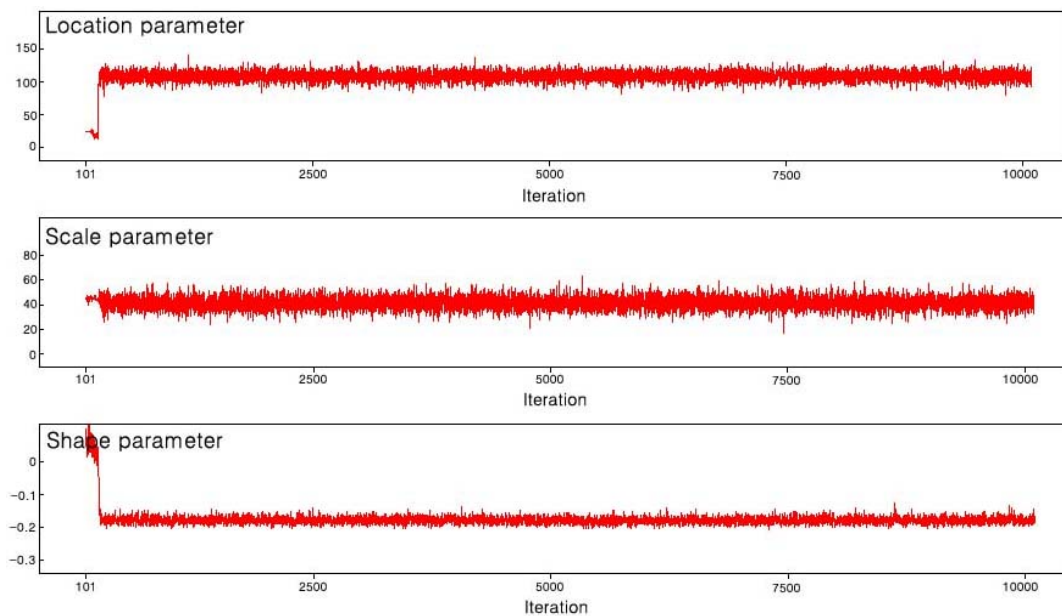


Fig. 4. Trace of Metropolis-Hastings Algorithm at Incheon Station: GEV

적으로 채택률 (Acceptance rate)을 계산하여 적절한 제안 분포인지를 검증해 보아야 한다. Chip and Greenberg (1995)는 Metropolis-Hastings 알고리즘을 수행하기 위한 5가지의 적절한 사전분포를 제안한 바 있는데, 본 연구는 5개의 제안분포 중 Eq. (24)와 같이 이변량 정규분포 (Bivariate normal distribution)를 사용한 확률보행 (Random walk) 형태의 제안분포를 사용하였다.

$$q(\theta_{j+1}|\theta_j) = q_1(|\theta_{j+1} - \theta_j|) \quad (24)$$

여기서, q_1 은 이변량 정규분포이며, 알고리즘 내에서 추정하고자 하는 모수가 적절한 값을 가지도록 정규분포의 분산은 수동적인 방법을 통하여 산정하였다. 이변량 정규분포를 사용하면 채택확률을 산정하는 식에서 $q(\theta_j|\theta_{j+1}) = q(\theta_{j+1}|\theta_j)$ 가 되어 채택확률의 산정이 사후분포에만 관련되므로 알고리즘의 효율성이 좋은 것으로 Chip

and Greenberg (1995)가 제안한 바 있으며 본 연구에서도 이들의 결론에 의해 이변량 정규분포를 사용하였고, 채택률을 산정하여 제안분포의 적정성을 확인하였다.

우도함수, 2가지 종류의 사전분포, 제안분포가 모두 구축되었으므로 Metropolis-Hasting 알고리즘을 사용하여 Bayesian MCMC 방법에 의해 추정하고자 하는 모수를 추정하였다. 본 연구에서는 알고리즘의 최대 반복횟수를 10,000번으로 설정하였으며, 최초 샘플링된 100개의 추정치는 모형의 안정성을 위하여 평균계산 시 제외하였으며 (Burn-in=100), Burn-in 값은 모의된 모수가 변동하지 않는 구간을 주관적으로 선정하여 산정하였다. 사용된 Metropolis-Hastings 알고리즘은 Lee and Kim (2008)이 사용했던 Matlab 코드를 본 연구에 맞도록 일부 수정하여 수행되도록 하였다.

다음 Fig. 4는 인천 관측소지점에 대한 GEV 분포의

Metropolis-Hastings 알고리즘 수행절차를 나타낸 것이며, Fig. 5는 인천 관측소지점에 대한 Gumbel 분포의 Metropolis-Hastings 알고리즘의 수행절차를 나타낸 것이다. Gumbel 분포의 경우 Burn-in값을 100으로 했을 때 크게 진동하는 초기값들이 나타나지 않아 안정적인 결과를 얻을 수 있었던 반면, 3개의 모수를 가지는 GEV 분포의 경우에는 Burn-in 값을 100으로 하였을 때에도 불안정한 값이 나타나는 것을 알 수 있었다. 이로부터 분포에 따른 Burn-in 값을 별도로 설정해야 보다 합리적인 Bayesian 평균값을 산정할 수 있다는 사실을 확인하였으나, 본 연구에서는 큰 영향을 미치지 않아 Gumbel 분포와 GEV 분포에 모두 같은 Burn-in 값인 100을 이용하여 Bayesian 평균값을 최종적으로 산정하였다.

알고리즘의 안정성을 검증하기 위해서는 위에서 언급한 바와 같이 채택률을 계산할 필요가 있다. 채택률이란 알고리즘의 반복횟수에 대한 추정된 모수의 채택된 비율을 의미하며, 채택률이 낮은 경우에는 모수의 새로운 값이 추정

되지 않으므로 제안분포 등을 새로 고려해야 하고 반대로 채택률이 너무 높은 경우에는 모수를 탐색하는데 있어서 과도한 시간이 필요하므로 반복횟수를 증가시키게 되어 계산시간이 너무 길어지는 비효용성을 가지게 된다. 적절한 채택률은 알고리즘의 효율성과 크게 관계있으며 Roberts et al. (1994)은 약 0.45를 Gamerman (1997)은 0.2에서 0.5 사이를 적절한 채택률로 제시한 바 있는데 본 연구에서 산정된 채택률은 0.41로 기존 연구자들이 제시한 채택률의 범위 내에 있어 알고리즘이 합리적이라는 것을 알 수 있었다.

5.3 2차 근사를 이용한 최우추정법과 Bayesian MCMC의 결과 비교

앞의 Eq. (10)에 나타낸 바 있는 2차 근사를 이용하여 최우추정법에 의한 모수추정결과 중 하한값 (2.5%), 최빈값 (mode), 상한값 (97.5%)과 Bayesian MCMC방법에 의한 하한값 (2.5%), 평균값 (mean), 상한값 (97.5%)을 Table 5에 나타내었으며, 이에 대한 간단한 분석결과를

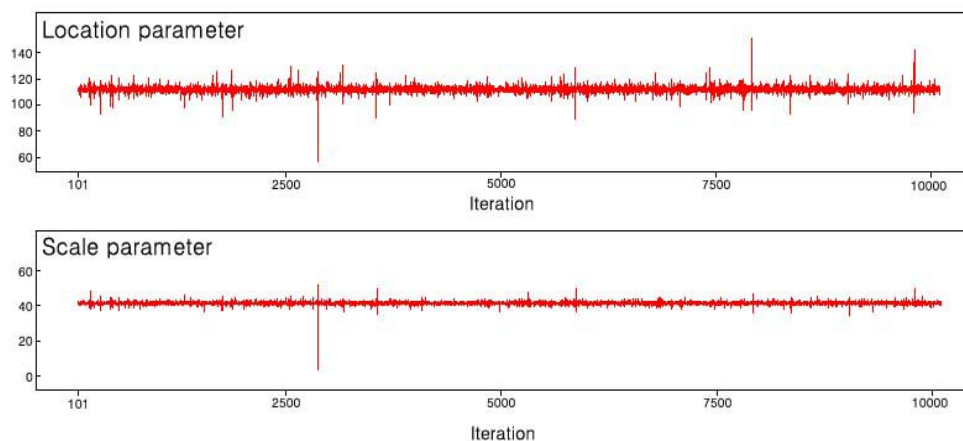


Fig. 5. Trace of Metropolis-Hastings Algorithm at Incheon Station: Gumbel

Table 5. Estimates of Parameters Using MLE and Bayesian MCMC

| Station (Distribution) | Parameter | MLE (2.5%) | MLE (Mode) | MLE (97.5%) | MLE (Uncertainty) | MCMC (2.5%) | MCMC (Mean) | MCMC (97.5%) | MCMC (Uncertainty) |
|------------------------|-----------|------------|------------|-------------|-------------------|-------------|-------------|--------------|--------------------|
| Incheon (GEV) | Location | 61.458 | 101.565 | 141.673 | 80.215 | 94.546 | 104.573 | 114.6 | 20.054 |
| | Scale | 39.734 | 39.911 | 40.099 | 0.365 | 39.879 | 39.925 | 39.971 | 0.092 |
| | Shape | -0.003 | -0.197 | -0.341 | -0.338 | -0.108 | -0.197 | -0.265 | -0.157 |
| Incheon (Gumbel) | Location | 66.005 | 106.004 | 146.004 | 79.999 | 99.007 | 109.004 | 119.004 | 19.997 |
| | Scale | 33.427 | 43.887 | 53.887 | 20.46 | 38.878 | 43.887 | 48.887 | 10.009 |
| | Shape | - | - | - | - | - | - | - | - |
| Gang-neung (GEV) | Location | 72.169 | 112.309 | 152.451 | 80.282 | 105.284 | 115.319 | 125.355 | 20.071 |
| | Scale | 42.855 | 43.078 | 43.317 | 0.462 | 43.038 | 43.096 | 43.154 | 0.116 |
| | Shape | -0.289 | -0.402 | -0.546 | -0.257 | -0.344 | -0.402 | -0.45 | -0.106 |
| Gang-neung (Gumbel) | Location | 83.734 | 123.743 | 163.743 | 80.009 | 116.734 | 126.743 | 136.743 | 20.009 |
| | Scale | 37.928 | 57.927 | 77.927 | 39.999 | 47.972 | 57.927 | 62.927 | 14.955 |
| | Shape | - | - | - | - | - | - | - | - |

나타내면 다음과 같다.

- 1) 전체적인 불확실성 산정량은 Bayesian MCMC 방법에 의한 불확실성이 2차근사를 이용한 불확실성의 산정결과보다 크게 감소되어 나타나짐을 알 수 있었다.
- 2) 최우추정법에 의해 산정된 최빈값(mode)과 Bayesian MCMC의 평균값은 크게 다르지 않고 유사하게 나타남을 알 수 있었다.
- 3) 지역적인 차별성으로써 GEV 분포 및 Gumbel 분포의 위치모수, 척도모수, 형태모수는 인천 관측소보다 강릉 관측소의 자료를 이용한 경우가 큰 값을 보임을 알 수 있다(단, GEV 분포의 형태모수는 절대값을 사용했을 때임).
- 4) 분포적 차별성으로써 인천관측소 및 강릉관측소에서 GEV 분포의 척도모수의 불확실성은 상당히 작은 값으로 산정된데 반하여, Gumbel 분포의 척도모

수의 불확실성은 큰 값으로 산정된 것을 알 수 있었다. 이러한 현상은 GEV 분포의 경우 형태모수가 추가됨으로써 척도모수의 불확실성이 감소되는 효과를 일으킨 것으로 판단된다.

추정된 모수들을 이용하여 각각의 모수에 해당되는 확률강우량을 산정하였다. 산정된 확률강우량 중에서 각 중수공구조물의 설계에 이용되는 빈도가 높은 10년, 50년, 100년, 200년, 300년, 400년 및 500년 빈도에 대한 인천 관측소 및 강릉 관측소에서의 확률강우량을 Table 6에 나타내었으며, 일반적으로 사용하는 확률강우량과 달리 2차근사에 의한 최우추정법과 Bayesian MCMC 방법에 의한 불확실성까지 함께 나타내었다. 산정된 확률강우량에 대한 비교를 나타내면 다음과 같다.

- 1) 불확실성측면에서는 모수추정에서 나타난 것과 동일한 결과를 얻을 수 있었다. 즉, Bayesian MCMC

Table 6. Results of Quantile Estimation Including Uncertainty

| Station (Distribution) | Return Period | MLE (2.5%) | MLE (Mode) | MLE (97.5%) | MLE (Uncertainty) | MCMC (2.5%) | MCMC (Mean) | MCMC (97.5%) | MCMC (Uncertainty) |
|------------------------|---------------|------------|------------|-------------|-------------------|-------------|-------------|--------------|--------------------|
| Incheon (GEV) | 10 | 151.3 | 214.6 | 277.4 | 126.1 | 196.1 | 217.6 | 237.6 | 41.5 |
| | 50 | 217.5 | 336.0 | 469.0 | 251.5 | 288.1 | 339.0 | 388.0 | 99.9 |
| | 100 | 245.6 | 400.4 | 588.5 | 342.9 | 332.2 | 403.5 | 474.2 | 142.0 |
| | 200 | 273.7 | 474.0 | 739.7 | 466.0 | 379.5 | 477.2 | 577.5 | 198.0 |
| | 300 | 290.1 | 522.0 | 846.0 | 555.9 | 408.8 | 525.1 | 647.3 | 238.4 |
| | 400 | 301.7 | 558.3 | 930.8 | 629.1 | 430.5 | 561.5 | 701.5 | 271.1 |
| | 500 | 310.8 | 588.0 | 1002.6 | 691.9 | 447.7 | 591.2 | 746.5 | 298.8 |
| Incheon (Gumbel) | 10 | 142.3 | 204.8 | 267.3 | 125.0 | 186.5 | 207.8 | 229.0 | 42.5 |
| | 50 | 198.2 | 277.2 | 356.3 | 158.0 | 250.7 | 280.2 | 309.8 | 59.0 |
| | 100 | 221.9 | 307.9 | 393.9 | 172.0 | 277.9 | 310.9 | 343.9 | 66.0 |
| | 200 | 245.5 | 338.4 | 431.4 | 185.9 | 304.9 | 341.4 | 377.9 | 73.0 |
| | 300 | 259.2 | 356.3 | 453.3 | 194.0 | 320.7 | 359.3 | 397.8 | 77.0 |
| | 400 | 269.0 | 368.9 | 468.8 | 199.8 | 331.9 | 371.9 | 411.8 | 79.9 |
| | 500 | 276.6 | 378.7 | 480.8 | 204.3 | 340.6 | 381.7 | 422.8 | 82.1 |
| Gang-neung (GEV) | 10 | 208.0 | 269.9 | 344.2 | 136.2 | 251.5 | 273.0 | 293.5 | 42.0 |
| | 50 | 381.9 | 519.5 | 741.0 | 359.2 | 459.1 | 522.7 | 584.6 | 125.5 |
| | 100 | 484.3 | 686.2 | 1051.0 | 566.7 | 589.1 | 689.4 | 789.5 | 200.4 |
| | 200 | 609.0 | 905.9 | 1502.8 | 893.7 | 753.7 | 909.2 | 1068.9 | 315.2 |
| | 300 | 694.4 | 1065.7 | 1857.9 | 1163.5 | 869.7 | 1069.1 | 1277.4 | 407.6 |
| | 400 | 761.3 | 1196.0 | 2161.9 | 1400.6 | 962.4 | 1199.4 | 1450.1 | 487.7 |
| | 500 | 817.1 | 1307.8 | 2432.9 | 1615.7 | 1040.9 | 1311.3 | 1600.4 | 559.5 |
| Gang-neung (Gumbel) | 10 | 169.1 | 254.1 | 339.1 | 170.0 | 224.6 | 257.1 | 278.4 | 53.8 |
| | 50 | 231.7 | 349.8 | 467.8 | 236.1 | 303.8 | 352.8 | 382.3 | 78.5 |
| | 100 | 258.2 | 390.2 | 522.2 | 264.0 | 337.2 | 393.2 | 426.2 | 89.0 |
| | 200 | 284.6 | 430.5 | 576.4 | 291.8 | 370.6 | 433.5 | 470.0 | 99.4 |
| | 300 | 300.0 | 454.0 | 608.1 | 308.1 | 390.0 | 457.0 | 495.6 | 105.5 |
| | 400 | 310.9 | 470.7 | 630.5 | 319.6 | 403.8 | 473.7 | 513.7 | 109.9 |
| | 500 | 319.4 | 483.7 | 648.0 | 328.5 | 414.5 | 486.7 | 527.7 | 113.2 |

방법에 의한 확률강우량이 2차근사를 이용한 최우 추정방법의 결과보다 작은 불확실성을 가짐을 알 수 있었으나 최우추정법의 최빈값에서의 확률강우량과 Bayesian MCMC 방법의 평균에서의 확률강우량은 거의 동일하게 산정됨을 알 수 있었다.

- 2) 인천 관측소지점에 비하여 강릉 관측소지점의 확률 강우량이 큰 값으로 산정됨을 알 수 있다.
- 3) 각각의 관측소에서 GEV 분포에 의한 확률강우량이 Gumbel 분포에 의한 확률강우량보다 큰 값으로 산정되는 것을 알 수 있다. 이에 대한 분석은 다음 절의 빈도곡선 (Fig. 6)을 작성하여 보다 상세히 논하였다.

6. 지점별 및 분포별 강우 빈도곡선의 작성

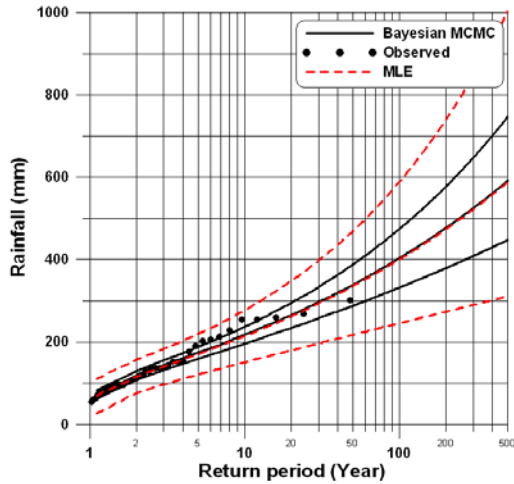
Table 6에서 추정된 빈도 강우량을 이용하여 빈도곡선을 작성하기 위해서는 47개의 실측 강우량 자료의 위치를

결정하기 위한 확률도시식이 필요하다. 본 연구에서는 강우량의 확률도시에 주로 사용되는 다음의 Weibull 식을 이용하여 실측 유량자료의 위치를 산정하였다.

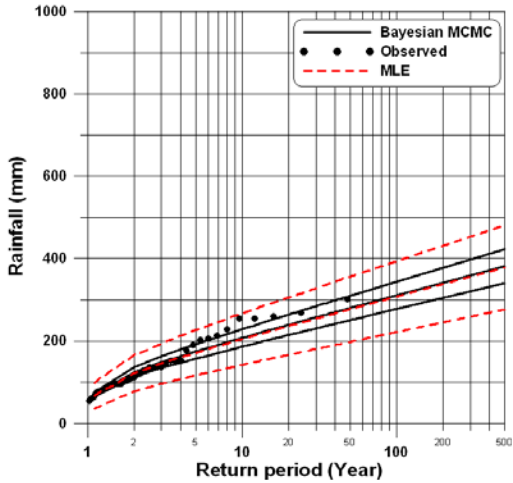
$$T = \frac{m+1}{n} \quad (25)$$

산정된 47개의 실측 강우량과 확률강우량을 이용하여 빈도곡선을 각 지점별과 사용된 확률분포식별로 구분하여 작성하였으며, 이를 Fig. 6에 나타내었다.

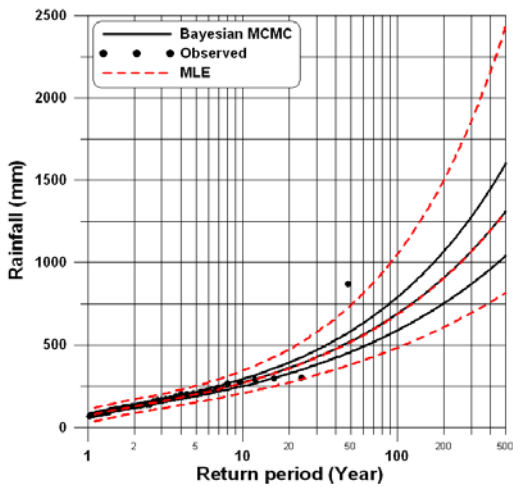
Fig. 6은 위의 5.3절에서 기술한 두 가지 추정방법에 의한 비교결과를 보다 명확히 보여준다. 즉 불확실성 측면에서는 모든 경우에 대하여 Bayesian MCMC 방법에 의한 분석결과가 2차 근사식을 이용한 최우추정방법의 결과보다 감소시켜 나타내기 때문에 불확실성을 산정하고자 하는 경우에는 Bayesian MCMC 방법을 사용하는 것이 훨씬 유리할 것으로 판단되었다. 그러나 각 방법의 최빈



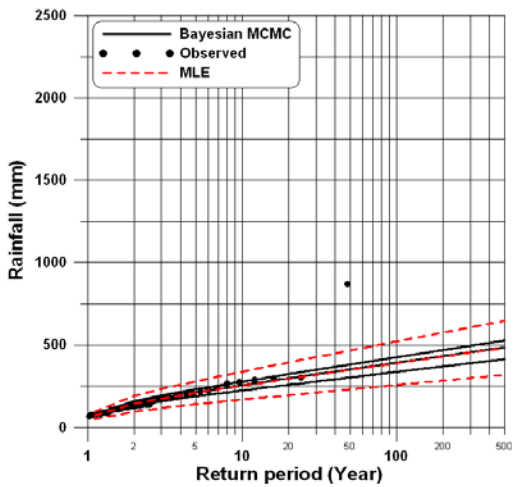
(a) GEV - Incheon



(b) Gumbel - Incheon



(c) GEV - Gangneung



(d) Gumbel - Gangneung

Fig. 6. Rainfall Frequency Curves Including Uncertainty at 2 Stations

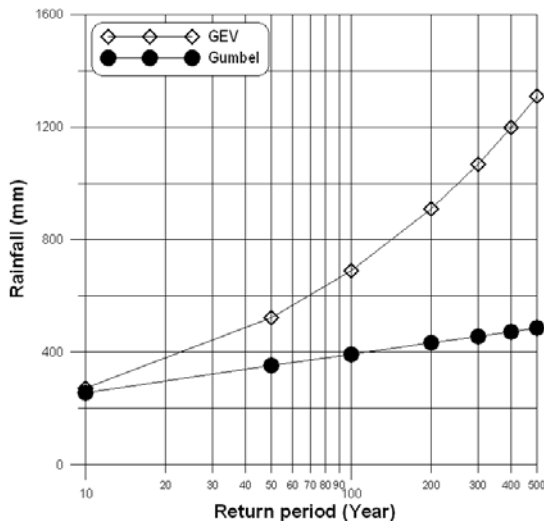


Fig. 7. Comparison with GEV and Gumbel Distribution

값과 평균값은 비슷한 값을 나타내기 때문에 불확실성의 산정이 의미가 없는 경우라면 복잡한 절차를 필요로 하는 Bayesian MCMC 방법보다는 간단히 확률강우량을 산정할 수 있는 최우추정법을 사용하는 것이 효율적일 수 있다.

지역적인 측면에서는 인천 관측소의 경우에는 GEV 분포와 Gumbel 분포 두 가지가 모두 큰 차이없이 사용될 수 있는 것으로 판단되었다. 그러나 이와 달리 극치강우사상을 포함하고 있는 강릉 관측소의 경우에는 Gumbel 분포는 형태모수가 없어 그 기능을 할 수 없으므로 GEV 분포에 비해 확률강우량을 크게 감소시켜 추정하는 오류를 범하게 됨을 알 수 있었다.

극치강우사상에 대한 분석을 보다 명확하게 하기 위하여 강릉 관측소에서의 각 빈도별 확률강우량에 대한 다음의 그림을 다시 나타내었다. 단, 두 방법에 의한 최빈값 및 평균값에서의 확률강우량은 거의 동일하므로 Bayesian MCMC 방법의 평균값만 이용하여 그래프를 작성하였다. Fig. 7을 보면 GEV 분포를 사용하는 경우 200년 빈도 확률강우량이 Gumbel 분포의 2배 가까이 증가되어 나타나는 것을 알 수 있으며, 500년 빈도에서는 대략적으로 3배 가량의 차이가 나는 것을 확인할 수 있다. 그러므로 강릉 지점의 47년간 최대강우일강우량이 870.5 mm인 것을 감안하면 Gumbel 분포를 이용해서는 이와 같은 값을 추정할 수 없을 것으로 판단된다.

7. 결론 및 향후과제

본 연구는 Bayesian 방법을 이용하여 극치강우사상을 포함한 강우빈도분석을 수행하기 위한 연구로써 Bayesian MCMC 방법과 2차 근사식을 이용한 최우추정방법을 이

용하기 위해 구축된 자료에 기반한 사전분포, 제안분포, 신뢰구간의 산정식 등을 이용하여 인천과 강릉관측소에 대한 점 빈도분석을 수행하고 그 결과로써 불확실성을 포함한 확률강우량의 빈도곡선을 작성하였다.

두 가지 추정방법의 분석을 위하여 먼저 자료의 길이가 47개년인 인천관측소와 강릉관측소를 선정하였다. 극치강우사상을 포함한 경우 결과를 비교하기 위하여 인천관측소와는 달리 2002년 태풍 RUSA로 인한 극치강우사상을 포함하고 있는 강릉관측소를 선정하였다. 또한 확률분포는 국내 기존 강우빈도의 분석에 많이 사용된 바 있는 GEV 분포와 Gumbel 분포를 선정하였다. 선정된 자료를 이용하여 먼저 2차 근사를 이용한 최우추정법을 수행하여 그 결과를 하한(2.5%), 최빈값, 상한(97.5%)에 대해 정리하였다. 또한 Bayesian 방법을 적용하기 위한 과정 중 사전분포는 인천 및 강릉관측소를 제외한 전국의 10개 관측소를 대상으로 하여 분석한 GEV 분포와 Gumbel 분포를 이용하여 구축하였다.

이와 같이 지역적정보로부터 얻어진 각각의 분포별 사전분포와 각각의 우도함수를 구축하여 이로부터 사후분포를 구성하였으며, 최종적으로 구축된 각각의 사후분포로부터 Metropolis-Hastings 알고리즘을 이용하여 10,000개의 모수를 추정하고 이로부터 최종적인 하한(2.5%), 평균값, 상한(97.5%)을 추출하였다.

추정된 모수에 대한 분석결과를 살펴보면 Bayesian MCMC 방법에 의한 불확실성이 2차근사를 이용한 불확실성의 산정결과보다 크게 감소되어 나타나짐을 알 수 있었으나, 최우추정법에 의해 산정된 최빈값(mode)과 Bayesian MCMC의 평균값은 크게 다르지 않고 유사하게 나타남을 알 수 있었다. 또한 지역적인 차별성으로써 GEV 분포 및 Gumbel 분포의 위치모수, 척도모수, 형태모수는 인천 관측소보다 강릉 관측소의 자료를 이용한 경우가 큰 값을 보임을 알 수 있었고, 분포적 차별성으로써 인천관측소 및 강릉관측소에서 GEV 분포의 척도모수의 불확실성은 상당히 작은 값으로 산정된데 반하여, Gumbel 분포의 척도모수의 불확실성은 큰 값으로 산정된 것을 알 수 있었다.

추정된 모수를 이용하여 인천관측소와 강릉관측소에서의 확률강우량을 나타내는 빈도곡선을 작성할 수 있었다. 그러나 본 연구에서는 기존의 빈도곡선과는 달리 불확실성까지를 포함하여 빈도곡선을 나타냄으로써 불확실성을 정량화함으로써 불확실성 결과를 이용하여 각종 수공구조물 등의 최종적인 설계안을 결정할 수 있도록 하였다. 최종적으로 구축된 빈도곡선을 살펴보면, 인천 관측소의 경우에는 GEV 분포와 Gumbel 분포 두 가지가 모두 큰 차이 없이 사용될 수 있는 것으로 판단되었으나 이와 달

리 극치강우사상을 포함하고 있는 강릉 관측소의 경우에는 Gumbel 분포는 형태모수가 없어 그 기능을 할 수 없으므로 GEV 분포에 비해 확률강수량을 크게 감소시켜 추정하는 오류를 범하게 됨을 알 수 있었다.

이와 같은 결론을 바탕으로 하여 강우량을 이용하여 확률강수량을 산정하는 경우나 또는 빈도곡선을 산정하는 경우 해당 분석지점의 기존 자료가 극치강우사상을 포함하고 있다면 명확한 불확실성의 산정을 위하여 과정이 다소 복잡하더라도 Bayesian 방법을 이용하여 감소된 형태의 불확실성의 범위도 함께 제시하는 것이 확률강수량 산정 시 왜곡된 외삽문제로 인한 설계시의 오류를 감소시킬 수 있으리라 판단된다. 또한 국내의 경우 강우량의 빈도 분석에 가장 흔히 사용되고 있는 Gumbel 분포는 극치강우사상을 포함하고 있는 경우에는 확률값을 크게 과소평가하게 되므로, 분석하고자 하는 지점이 극치강우사상을 포함하고 있다면 확률분포의 선정 시 필히 GEV 분포와의 비교분석을 통하여 가장 합리적인 확률분포를 선정해야함을 알 수 있었다.

감사의 글

본 연구는 21세기 프런티어 연구개발 사업인 수자원의 지속적 확보기술개발 사업단(과제번호 1-7-3)의 서울대학교 공학연구소를 통한 연구비 지원(30%)과 서울대학교 BK21 안전하고 지속가능한 사회기반건설사업단의 연구비 지원(70%)에 의해 수행되었습니다. 연구비 지원에 심심한 감사의 뜻을 표합니다.

참고 문헌

건설교통부(2000). 1999년도 수자원관리기법개발연구조사 보고서: 제1권 한국확률강수량도작성.

김기욱, 유철상, 박민규, 김현준(2007). “강우빈도해석에서의 측우기자료의 유용성 평가.” **한국수자원학회 논문집**, 제40권, 제11호, pp. 851-859.

김병식, 김보경, 경민수, 김형수(2008). “기후변화가 극한 강우와 I-D-F 분석에 미치는 영향평가.” **한국수자원학회 논문집**, 제41권, 제4호, pp. 379-394.

김보경, 김병식, 김형수(2008). “극한지수를 이용한 극한 기상사상의 변화 분석.” **대한토목학회 논문집**, 제28권, 제1B호, pp. 41-53.

김상욱, 이길성(2008a). “Bayesian MCMC를 이용한 저수량 점 빈도분석: I. 이론적 배경과 사전분포의 구축.” **한국수자원학회 논문집**, 제41권, 제1호, pp. 35-47.

김상욱, 이길성(2008b). “Bayesian MCMC를 이용한 저수량 점 빈도분석: II. 적용과 비교분석.” **한국수자원**

학회 논문집, 제41권 제1호, pp. 49-63.

김상욱, 이길성(2008c). “Bayesian 다중회귀분석을 이용한 저수량(Low flow) 지역 빈도분석.” **한국수자원학회 논문집**, 제41권 제3호, pp. 324-340.

김상욱, 이길성(2008d). “베이지안 회귀분석을 이용한 수위-유량 관계곡선의 불확실성 분석.” **한국수자원학회 논문집**, 제41권 제9호, pp. 943-958.

성기원(2008). “제한적 강우-지속기간 자료를 이용한 실용적 IDF 관계식의 유도.” **한국수자원학회 논문집**, 제41권, 제6호, pp. 57-595.

유철상, 박창열, 김경준, 전경수(2007). “모포마분포를 적용한 분단위 강우강도-지속시간-재현기간 관계의 유도.” **한국수자원학회 논문집**, 제40권, 제8호, pp. 643-654.

이정규, 추현재(2006). “Huff의 4분위법을 이용한 지속기간별 연 최대치 강우의 시간분포 특성연구.” **대한토목학회 논문집**, 제26권, 제5B호, pp. 519-528.

Baghirathan, V.R., and Shaw, E.M.(1978). “Rainfall depth-duration-frequency studies for Sri Lanka.” *Journal of Hydrology*, Vol. 37, No. 3, pp. 223-239.

Chip, S., and Greenberg, E.(1995). “Understanding the Metropolis-Hastings algorithm.” *Journal of the American Statistical Association*, Vol. 49, No. 4, pp. 327-335.

Coles, S., Pericchi, L.R., and Sisson, S.(2003). “A fully probabilistic approach to extreme rainfall modeling.” *Journal of Hydrology*, Vol. 273, pp. 35-50.

Coles, S., and Powell, E.A.(1996). “Bayesian methods in extreme value modeling: A review and new developments.” *International Statistical Review*, Vol. 64, No. 1, pp. 119-136.

Coles, S., and Tawn, J.A.(1996). “A Bayesian analysis of extreme rainfall data.” *Applied Statistics*, Vol. 45, No. 4, pp. 463-478.

Gamerman, D.(1997). *Markov Chain Monte Carlo-Stochastic Simulation for Bayesian Inference*. Chapman & Hall, London, UK.

Kavetski, D., Kuczera, G., and Fanks, S.W.(2006). “Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory.” *Water Resources Research*, Vol. 42, W03407.

Kuczera, G.(1999). “Comprehensive at-site flood frequency analysis using Monte Carlo Bayesian inference.” *Water Resources Research*, Vol. 35, No. 5, pp. 1551-1557.

Kuczera, G., and Parent E.(1998). “Monte Carlo

- assessment of parameter uncertainty in conceptual catchment models: The Metropolis algorithm." *Journal of Hydrology*, Vol. 211, pp. 69–85.
- Lee, K.S., and Kim, S.U. (2008). "Identification of uncertainty in low flow frequency analysis using Bayesian MCMC method." *Hydrological Processes*, Vol. 22, pp. 1949–1964.
- Ludwig, D. (1996). "Uncertainty and the assessment of extinction probabilities." *Ecological Applications*, Vol. 6, No. 4, pp. 1067–10076.
- Malakoff, D. (1999). "Bayes offers a new way to make sense of numbers." *Science*, Vol. 286, No. 5444, pp. 1460–1464.
- Meeker, W.Q., and Escobar, L.A. (1998). Statistical methods for reliability data. *John Wiley and Sons*, New York, NY.
- Metropolis, N., Rosenbluth, A.W., Teller A.H., and Teller E. (1953). "Equations of state calculations by fast computing machines." *Journal of Chemical Physics*, Vol. 21, pp. 1087–1092.
- O'Connell, D.R.H., Ostenaar, D.A., Levish, D.R., Klinger, and R.E. (2002). "Bayesian flood frequency analysis with paleohydrologic bound data." *Water Resources Research*, Vol. 38, No. 5, pp. 1–14.
- Parent, E., and Bernier, J. (2003). "Encoding prior experts judgements to improve risk analysis of extreme hydrological events via POT modeling." *Journal of Hydrology*, Vol. 283, pp. 11–18.
- Pinter, N., Thomas, R., and Wlosinski, H. (2001). "Assessing flood hazards on dynamic rivers." *EOS Transactions*, Vol. 82, No. 31, pp. 339–340.
- Pitman, W.V. (1980). "A depth-duration-frequency diagram for point rainfall in SWA–Namibia." *Water SA*, Vol. 6, No. 4, pp. 157–162.
- Reis Jr., D.S., and Stedinger, J.R. (2005). "Bayesian MCMC flood frequency analysis with historical information." *Journal of Hydrology*, Vol. 313, pp. 97–116.
- Reis Jr., D.S., Stedinger, J.R., and Martins, E.S. (2005). "Bayesian generalized least squares regression with application to log pearson type III regional skew estimation." *Water Resources Research*, Vol. 41, W10419.
- Roberts, G.O., Gelman, A., and Gilks, W.R. (1994). *Weak convergence and optimal scaling of random walk metropolis-hastings algorithms*. Technical Report, University of Cambridge.
- Seidou, O., Ouarda, T.B.M.J., Barbet, M., Bruneau, P., and Bobee, B. (2006). "A parametric Bayesian combination of local and regional information in flood frequency analysis." *Water Resources Research*, Vol. 42, W11408.
- Sisson, S.A., Pericchi, L.R., and Coles, S.G. (2006). "A case for a reassessment of the risks of extreme hydrological hazards in the Caribbean." *Stochastic Environmental Research and Risk Assessment*, Vol. 20, pp. 296–306.
- Smith, R.L. (1989). "Extreme value analysis of environmental times series: An example based on ozone data." *Statistical Science*, Vol. 4, pp. 367–393.

| | |
|-----------------------------|------------------|
| 논문번호: 09-053 | 접수: 2009.05.13 |
| 수정일자: 2009.12.28/2010.02.22 | 심사완료: 2010.02.22 |