

## Bayesian MBLRP 모형을 이용한 시간강수량 모의 기법 개발

김장경\* · 권현한\*\* · 김동균\*\*\*

Kim, Jang Gyeong\*, Kwon, Hyun Han\*\*, Kim, Dong Kyun\*\*\*

### A Development of Hourly Rainfall Simulation Technique Based on Bayesian MBLRP Model

#### ABSTRACT

Stochastic rainfall generators or stochastic simulation have been widely employed to generate synthetic rainfall sequences which can be used in hydrologic models as inputs. The calibration of Poisson cluster stochastic rainfall generator (e.g. Modified Bartlett-Lewis Rectangular Pulse, MBLRP) is seriously affected by local minima that is usually estimated from the local optimization algorithm. In this regard, global optimization techniques such as particle swarm optimization and shuffled complex evolution algorithm have been proposed to better estimate the parameters. Although the global search algorithm is designed to avoid the local minima, reliable parameter estimation of MBLRP model is not always feasible especially in a limited parameter space. In addition, uncertainty associated with parameters in the MBLRP rainfall generator has not been properly addressed yet. In this sense, this study aims to develop and test a Bayesian model based parameter estimation method for the MBLRP rainfall generator that allow us to derive the posterior distribution of the model parameters. It was found that the HBM based MBLRP model showed better performance in terms of reproducing rainfall statistic and underlying distribution of hourly rainfall series.

**Key words** : Bartlett-Lewis, Bayesian, Uncertainty, MBLRP

#### 초 록

추계학적 강수발생 및 모의기법은 수문학적 모형의 입력 자료로써 널리 이용되고 있다. 그러나 Modified Bartlett-Lewis Rectangular Pulse (MBLRP)와 같은 추계학적 포아송 클러스터 강수생성 모형에 대해서 국부최적화 방법을 통한 매개변수 추정 방법은 매개변수의 신뢰성에 상당한 영향을 주는 것으로 알려져 있다. 최근에는 MBLRP 모형의 국부해추정 문제를 해소하기 위하여 Particle Swarm Optimization (PSO) 또는 Shuffled Complex Evolution developed at The University of Arizona (SCE-UA) 등 매개변수 추정 성능이 우수한 전역최적화기법이 도입되고 있지만, 제한된 매개변수 공간에서 항상 신뢰성 있는 매개변수 추정이 가능한 것은 아니다. 뿐만 아니라, 모형의 매개변수들이 갖고 있는 불확실성에 관한 연구는 아직 충분히 논의되지 않았다. 이러한 관점에서 본 연구는 Bayesian 기법과 연계한 MBLRP 모형을 개발하였으며 각 매개변수들의 사후분포(Posterior Distribution)를 유도하여 매개변수가 내포하는 불확실성을 정량적으로 평가하였다. 그 결과 관측값에 대한 시간단위 이하 강수발생 통계치를 효과적으로 복원하고 있음을 확인할 수 있었다.

**검색어** : Bartlett-Lewis, 베이지안, 불확실성, MBLRP

\* 정희원 · 전북대학교 토목공학과 박사과정 (kjk2388@jbnu.ac.kr)

\*\* 정희원 · 교신저자 · 전북대학교 토목공학과 부교수 (Corresponding Author · Chonbuk National University · hkwon@jbnu.ac.kr)

\*\*\* 정희원 · 홍익대학교 건설도시공학부 조교수 (kim.dongkyun@hongik.ac.kr)

Received January 9, 2014/ revised February 28, 2014/ accepted March 4, 2014

## 1. 서론

강수자료는 수공학분야에서 기본적인 입력 자료로서 목적에 따라 다양한 시계열의 형태로 이용되나 댐과 같은 대규모 수공구조물 설계에서는 상대적으로 자료 연한이 짧은 문제로 인해 신뢰성 있는 수자원계획 수립이 어렵다(Wheater et al., 2006). 이러한 문제를 개선하기 위한 방법 중 가장 대표적인 방법은 강수모의기법을 이용하는 것이다. 강수모의기법의 가장 핵심적인 부분은 과거자료의 통계특성을 효과적으로 복원할 수 있어야 하며 다양한 기상학적 특성들 또한 적절하게 모의될 수 있어야 한다(Boughton and Droop, 2003).

강수모의 시 일강수량 이하의 시간 규모에서는 구형펄스모형(Rectangular Pulses Model) 기반의 강수모형이 일반적으로 활용되고 있으며, 포이송 클러스터 프로세스 기반의 지점 강수모형인 Bartlett-Lewis Rectangular Pulse (BLRP) 모형이 대표적으로 이용되고 있다. BLRP 모형은 공간상의 한 점에 대한 연속시간 강수모형으로 일단위 및 일단위 이하 시간 규모의 통계특성을 재현하는데 유용하며, 다양한 시간 규모에 대하여 대부분의 통계특성을 비교적 정확하게 재현하는 것으로 알려져 있다(Rodriguez-Iturbe et al., 1987). 그러나 무강수기간(zero depth period)의 비율을 정확하게 재현하지 못하는 문제가 지적된 바 있으며, 이러한 문제를 해소하기 위해서 Rodriguez-Iturbe et al. (1988)은 기존 모형에 매개변수를 1개 더 추가한 Modified Bartlett-Lewis Rectangular Pulse (MBLRP) 모형을 제시하였다.

MBLRP 모형은 복잡한 강수의 물리적 발생특성을 계층적 구조로 간략화 함으로서 비교적 적은 수의 매개변수로 다양한 지역 및 계절의 강수 특성을 효과적으로 복원할 수 있는 장점을 갖고 있다(Onof et al., 2000). 특히 우리나라는 강수관측 초기 일단위 이하의 자료가 계속되지 않아 시간단위 강수에 대한 분석 시 신뢰성이 결여되는 단점이 있다. 이러한 점에서 MBLRP 강수모의기법은 일단위 이하의 강수자료에 대해서 신뢰성 있는 기초자료를 제공함으로써 일단위 이하 관측 자료가 갖는 사공간적 한계를 효과적으로 보완해 줄 수 있을 것으로 판단된다. 그러나 MBLRP 모형을 구축하는데 있어서 가장 어려운 점은 매개변수 추정이 쉽지 않다는 점이다. 특히, 기존 최적화 방법의 경우 최적해가 수렴되지 않아 모형의 매개변수 추정에 실패하는 사례가 많았으며, 모형이 매개변수들에 민감하게 반응하는 등의 문제가 있었다(Verhoest et al., 1997).

최근에는 이러한 문제점을 개선하기 위한 노력으로 Downhill Simplex Method (DSM), Simplex-Simulated Annealing (SSA), Particle Swarm Optimization (PSO), Shuffled Complex Evolution developed at The University of Arizona (SCE-UA) 등 다양한 전역 최적화기법들을 도입하여 만족할 만한 재현 능력을 입증한

바 있다(Vanhaute et al., 2012). 그러나 강수량의 1, 2차 모멘트 이상의 통계적 특성을 효과적으로 재현하지 못하고 극치강수량이 관측값에 비해 현저하게 과소 추정되는 문제점은 여전히 남아있다. 뿐만 아니라 모형의 매개변수들이 갖고 있는 불확실성에 관한 연구는 아직 충분히 논의되지 않았다.

국내의 포이송 클러스터 기반의 강수모의기법과 관련된 연구는 다음과 같다. Kim and Yoo (2008)은 관측강수의 시간적 군집특성을 적절하게 재현해 주는 강수모형을 선정하기 위하여 포이송 과정에 근거한 RPPM과 NS-RPPM 프로세스 모형을 비교한 바 있다. Kyoung et al. (2008)은 카오스 이론을 적용해 일강수 자료를 시간단위 자료로 분해하였으나 극치값의 경우 과소 추정되는 문제점을 지적하였다. Kim et al. (2013)은 홍수량 산정 모형의 입력 강수자료를 생성하기 위해서 MBLRP 모형을 활용한 바 있으나, 극치강수량이 과소 추정되는 문제로 인해 신뢰성 있는 홍수량 추정이 어려움을 지적한 바 있다.

이러한 점에서 본 연구에서는 일 단위 이하 강수량의 평균적 특성뿐만 아니라 극치강수량의 통계적 특성을 동시에 구현하고 MBLRP 모형의 매개변수들이 갖고 있는 불확실성을 정량적으로 평가하기 위한 방법으로 Bayesian 기법을 연계한 Bayesian MBLRP 모형을 제안하고자 한다. 본 연구에서 제시하는 Bayesian MBLRP 모형의 주된 목적은 6개 매개변수의 사후분포(posterior distribution)를 제공함으로써 매개변수의 불확실성을 정량적으로 분석할 수 있으며, 추정된 사후분포로부터 Monte Carlo 모의를 통해 다양한 시간 규모에서 극치강수량에 대한 재현효과를 극대화 하는데 있다.

본 논문의 구성은 다음과 같다. 2장에서는 Bayesian 기법과 연계한 MBLRP모형의 이론적 배경을 제시하였으며 3장에서는 전지구적 시간 강수량 자료를 활용하여 제시된 방법론의 적합성을 평가하였다. 마지막으로 4장에서는 연구 결론 및 향후연구방향을 제시하였다.

## 2. 연구방법

### 2.1 MBLRP Model

추계학적 관점에서 가상 강수시계열과 관측 강수시계열의 유사성은 통계값의 유사성에 의해 결정된다. Rodriguez-Iturbe et al. (1987)은 이러한 점을 고려하여 강수와 셀 특성을 5개의 매개변수( $\lambda, k, \mu_x, \phi, \eta$ )로 표현한 Bartlett-Lewis Rectangular Pulse (BLRP) 모형을 개발하였다. 모형의 매개변수  $\lambda$ 는 포이송 과정을 따르는 강수의 도달시간을 결정한다.  $\beta$ 는 포이송 과정을 따르는 셀의 도달시간을 결정한다. 셀의 깊이는  $\mu_x$ 를 매개변수로 갖는 지수분포에 의해 결정되며, 강수 전체의 길이는 지수분포를 갖는

매개변수  $\gamma$ 에 의해 결정된다. 각각의 셀 길이들은  $\eta$ 를 매개변수로 갖는 지수분포로 표현되며, 수학적 편의를 위해, 매개변수  $\gamma$ 와  $\beta$ 는  $k = \beta/\eta$ ,  $\phi = \gamma/\eta$ 로 무차원화 한다. 그러나 BLRP 모형은 다양한 시간스케일에 대하여 통계치 재현능력이 우수한 것에 비해 무강수기간의 비율을 정확하게 재현하지 못하는 문제가 있었다. 이러한 문제를 해소하기 위해 Rodriguez-Iturbe et al. (1988)은 기존 모형에 매개변수를 하나 더 추가한 Modified Bartlet-Lewis Rectangular Pulse (MBLRP) 모형을 제안하였다.

MBLRP 모형에서 셀 길이는 더 이상  $\eta$ 를 매개변수로 갖는 지수분포로부터 취하지 않고 형상인자(shape factor)  $\alpha$ 와 규모인자(scale factor)  $1/\nu$ 를 갖는 감마분포로 결정된다. 따라서 매개변수  $\eta$ 는 매개변수  $\alpha$ 와  $1/\nu$ 로 대체되어 최종적으로는 6개 매개변수 ( $\lambda, k, \mu_x, \phi, \alpha, 1/\nu$ )에 의해 모형이 결정된다. 모형의 통계치는 집성시간별(agggregated time) 평균, 분산, 지체시간  $s$ 인 자기공분산, 무강수기간의 비율로 다음 Eqs. (1), (2), (3) and (4a)~(4d)와 같이 표현된다. 여기서,  $\mu_c$ 는 강수 당 셀의 수 평균값으로  $\mu_c = 1 + k/\phi$ 로 표현된다.  $c$ 는 강수 당 셀의 수를 의미한다.  $T$ 는 대상 집성시간이고  $s$ 는 지체시간으로 일반적으로 1이 사용된다. 편의상  $k_1$ 과  $k_2$ 는 Eqs. (5a)~(5b)와 같이 매개변수화 하였다.

$$Mean = \lambda \mu_x \mu_c \frac{\nu}{\alpha - 1} T \tag{1}$$

$$Variance \tag{2}$$

$$= \frac{2\nu^{2-\alpha} T}{\alpha - 2} \left( k_1 - \frac{k_2}{\phi} \right) - \frac{2\nu^{3-\alpha}}{(\alpha - 2)(\alpha - 3)} \left( k_1 - \frac{k_2}{\phi^2} \right)$$

$$+ \frac{2}{(\alpha - 2)(\alpha - 3)} \left( k_1 (T + \nu)^{3-\alpha} - \frac{k_2}{\phi^2} (\phi T + \nu)^{3-\alpha} \right)$$

$$Autocovariance (lag - s) \tag{3}$$

$$= \frac{k_1}{(\alpha - 2)(\alpha - 3)}$$

$$\{ (T(s - 1) + \nu)^{3-\alpha} + (T(s + 1) + \nu)^{3-\alpha} - 2(Ts + \nu)^{3-\alpha} \}$$

$$+ \frac{k_2}{\phi^2 (\alpha - 2)(\alpha - 3)}$$

$$\{ 2(\phi Ts + \nu)^{3-\alpha} - (\phi T(s - 1) + \nu)^{3-\alpha} - (\phi T(s + 1) + \nu)^{3-\alpha} \}$$

$$P(\text{zero rain}) = e^{(-\lambda T - f_1 + f_2 + f_3)} \tag{4a}$$

$$f_1 = \frac{\lambda \nu}{\phi(\alpha - 1)} \left( 1 + \phi \left( k + \frac{\phi}{2} \right) - \frac{1}{4} \phi (k + \phi)(k + 4\phi) \right) \tag{4b}$$

$$\left( + \frac{\phi(k + \phi)(4k^2 + 27k\phi + 72\phi^2)}{72} \right)$$

$$f_2 = \frac{\lambda \nu}{(k + \phi)(\alpha - 1)} \left( 1 - k - \phi + \frac{3}{2} \phi k + \phi^2 + \frac{k^2}{2} \right) \tag{4c}$$

$$f_3 = \frac{\lambda \nu}{(k + \phi)(\alpha - 1)} \left( \frac{\nu}{\nu + T(k + \phi)} \right)^{\alpha - 1}$$

$$\frac{k}{\phi} \left( 1 - k - \phi + \frac{3}{2} \phi k + \phi^2 + \frac{k^2}{2} \right) \tag{4d}$$

$$k_1 = \left( 2\lambda \mu_x \mu_c^2 + \frac{\lambda \mu_x k \phi \mu_x^2}{\phi^2 - 1} \right) \left( \frac{\nu^\alpha}{\alpha - 1} \right) \tag{5a}$$

$$k_2 = \left( \frac{\lambda \mu_x k \mu_x^2}{\phi^2 - 1} \right) \left( \frac{\nu^\alpha}{\alpha - 1} \right) \tag{5b}$$

한편 본 연구에서는 무강수기간의 비율에 대한 식을 Rodriguez-Iturbe et al. (1988)의 것을 그대로 인용하지 않고  $f_1$ 에 표현된 세 개의  $\phi$ 항을 수정하여 셀 길이를 무차원화하였다. 추가되는 매개변수의 수정으로 강수 셀 길이의 값은 더 이상 순수한 지수분포로 제한되지 않으며, 지수분포의 한 형태인 감마분포로부터 생성되어 강수의 표현에 있어 모형을 유연하게 하였다.

### 2.2 Bayesian Modeling

Bayesian 확률은 사전확률(prior probability)을 기준으로 자료의 증가에 의해서 정보가 갱신되며 최종적으로 조건부로 바뀌는 사후확률(posterior probability)을 추정하는 과정이다. Bayes 정리(Bayes's rule)는 Thomas Bayes(영국, 1702~1761)에 의해 개념이 정립되었으며, 두 확률 변수의 조건부확률(conditional probability)과 주변확률(marginal probability) 사이의 관계를 나타내는 정리로, 확률론의 Bayesian 해석에 따르면 새로운 증거에 기반하여 과거의 정보를 향상시키거나 개선한다고 할 수 있다 (Gelman et al., 2003).

주어진 확률변수  $\theta$ 와  $y$ 가 있을 때, 두 확률변수들의 결합확률분포(joint probability distribution)는 Eq. (6)과 같이 사전분포(prior distribution)  $p(\theta)$ 와 우도(likelihood)  $p(y | \theta)$ 의 곱으로 표현할 수 있다. Bayes 정리에서 확률사건  $\theta$ 와  $y$ 의 조건부확률과 주변확률의 관계는 Eq. (7)와 같으며, Eq. (8)은 좌변  $p(\theta | y)$ 는 사후분포를 나타내며 사전분포와 우도  $l(\theta | y)$ 의 곱으로 추정될 수 있다.

$$p(\theta, y) = p(\theta)p(y | \theta) \tag{6}$$

$$p(\theta | y) = \frac{p(\theta, y)}{p(y)} = \frac{p(\theta)p(y | \theta)}{p(y)} \tag{7}$$

$$p(\theta | y) \propto l(\theta | y)p(\theta) \tag{8}$$

여기서  $p(\theta)$ 는  $\theta$ 의 사전확률 또는 경계 확률이며, ‘사전’이라는 의미는 사건  $y$ 에 관한 기본적인 정보를 의미한다.  $p(\theta | y)$ 는  $y$ 가 주어졌을 때  $\theta$ 의 조건부 확률을 말하며,  $y$ 의 특정 값에 의해 결정되기 때문에 사후확률이라고도 불린다.  $p(y|\theta)$ 는  $\theta$ 가 주어졌을 때  $y$ 의 조건부 확률을 말하며,  $p(y)$ 는  $y$ 의 주변확률로써 정규화 상수의 역할을 한다.

### 2.3 Markov Chain Monte Carlo 모의

기지의 확률분포(probability density function, PDF)로부터 추출된 난수를 바탕으로 적분값이나 신뢰구간 등을 추정하는 방법으로 Monte Carlo Simulation (MCS) 기법이 있다. MCS 기법은 주어진 확률분포로부터 동일한 분포를 따르며 상호독립인(Independent and Identical) 난수를 얻기 위해 활용된다. 그러나 다변량확률분포 (Multivariate Probability Density Function, MPDF)의 경우에는 기존 MCS를 통해서 변량을 모의할 수 없으며, 이러한 특수한 경우에 대해서 적용 가능한 방법으로 Markov Chain Monte Carlo (MCMC) 기법이 이용되고 있다.

MCMC 기법은 MPDF가 복잡하여 결합확률을 정확히 고려하여 모의하기 어려운 경우에 적용하는 방법으로 서로 독립인 형태의 변량 대신에 Markov Chain을 따르는 변량을 연속적으로 추출하여 사용한다. Markov Chain을 통해서 초기에 추출된 변량들은 정확하게 주어진 확률분포를 따르지는 않지만 연속적인 모의를 통해서 충분한 시간이 흐르게 되면 가정된 확률분포에 수렴하게 된다. MCMC 기법은 복잡한 MPDF를 갖는 경우 매개변수 추정 및 사후분포(posterior distribution)를 추론(inference)하기 위한 방법으로 활용된다. 특히 사후분포가 확률분포로서 적합성을 가지기 위해서는 전체 적분값이 1을 가져야 하는데, 이러한 확률분포의 적합성을 갖기 위해서는 사후분포의 분모 즉 상수항의 적분이 필수적으로 필요하게 된다. 그러나 다변량 함수의 적분을 수학적으로 추정하거나 추론을 하는 것은 불가능하므로 이러한 경우에 가장 적절한 방법으로써 MCMC 기법이 활용된다.

본 연구에서는 MBLRP 모형의 6개 매개변수의 다변량 분포가 복잡하여 수리적인 추정치의 계산이나 각 매개변수들에 대한 주변 분포(marginal distribution)의 계산이 난해한 이유로 MCMC기법을 활용하였다. MCMC 기법의 대표적인 방법에는 메트로폴리스-헤스팅 알고리즘(Metropolis-Hastings algorithm)과 깁스표본법 (Gibbs Sampling) 등이 있으며, 본 연구에서는 깁스표본법을 이용하였다. 깁스표본법은 원하는 다변량 확률분포에서 iid (independent identically distributed) 표본을 추출하는 것이 복잡하거나 난해한 경우 이용 가능한 방법으로서 3개의 변수를 갖는 다변량 확률분포를 이용하여 설명하면 다음과 같다.

첫째, 3개의 변수를 갖는 다변량 확률밀도함수를  $f(\theta_1, \theta_2, \theta_3)$ 라

고 하면, 깁스표본법은 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. 일반적으로 깁스표본법을 적용하기 위해서는 Eq. (9)와 같은 3개의 조건부 분포를 알고 있어야 한다.

$$f(\theta_1 | \theta_2, \theta_3) \tag{9}$$

$$f(\theta_2 | \theta_1, \theta_3)$$

$$f(\theta_3 | \theta_1, \theta_2)$$

둘째, 위의 조건부 확률분포로부터 iid 난수의 추출이 가능해야 깁스표본법을 통해서 매개변수의 사후분포 추정이 가능하다. 알고리즘은 다음과 같다.

- (1) 세 변수에 대한 초기 값( $\theta_1^{(0)}, \theta_2^{(0)}, \theta_3^{(0)}$ )을 부여한다.
- (2)  $t$ 번째 난수 벡터 ( $\theta_1^{(t)}, \theta_2^{(t)}, \theta_3^{(t)}$ )가 주어졌을 때  $t+1$ 번째 난수 벡터를 다음과 같은 조건부 분포에서 추출한다.
  - ①  $\theta_1^{t+1} \sim f(\theta_1 | \theta_2^{(t)}, \theta_3^{(t)})$
  - ②  $\theta_2^{t+1} \sim f(\theta_2 | \theta_1^{(t+1)}, \theta_3^{(t)})$
  - ③  $\theta_3^{t+1} \sim f(\theta_3 | \theta_1^{(t+1)}, \theta_2^{(t+1)})$
- (3) 위의 과정을 충분히 반복한 후 초기의 일정부분 난수를 제거한 이후의 난수들을 이용한다. 이러한 제거과정을 Burning이라고 하며 Bayesian 해석에서 중요한 구성단계이다(George and McCulloch, 1993).

상기 알고리즘과 같이 깁스표본법은 조건부 분포들에서 조건으로 주어지는 변수들의 값은 정확하게 바로 직전의 단계에서 주어진 값들이 사용된다. 따라서 조건부 분포에서 추출된 난수들이 안정

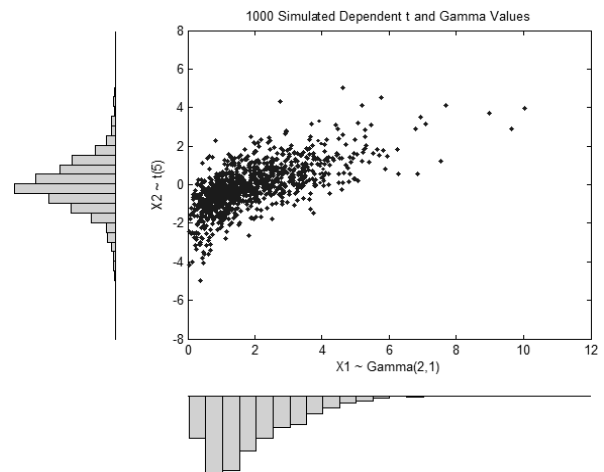


Fig. 1. Concept of Correlated Random Number Generation Based on t and Gamma Distribution

상태에 도달하는 것이 주어진 다변량 확률분포를 정확히 따르는 난수가 되는 척도가 되며 깃스프분법을 구현하는 데 가장 중요한 부분이 된다. Fig. 1은 t확률분포와 Gamma확률분포를 따르는 2변량에 대해서 깃스프분법을 적용한 결과로서 2개의 난수가 서로 상관성을 가지고 모의되고 있는 것을 확인할 수 있다.

**2.4 Bayesian MBLRP 모형의 매개변수 추정방법**

매개변수의 사후분포를 추정하기 위해서는 매개변수의 사전분포를 가정해야 한다. 본 연구에서는 각 매개변수간 물리적 범위를 고려할 수 있는 Truncated 정규분포를 매개변수의 사전분포로 가정하였다. 매개변수에 대한 결합확률밀도함수는 Eq. (10)과 같으며, Bayes 정리에 의한 매개변수의 사후분포는 Eq. (11)과 같이 표현된다.

$$p(\lambda, k, \mu_x, \phi, \alpha, 1/\nu) \propto 1 \tag{10}$$

$$p(\theta | M) = \frac{p(\theta, M)}{p(M)} \propto p(M | \theta) \cdot p(\theta) \tag{11}$$

여기서  $\theta = [\lambda, k, \mu_x, \phi, \alpha, 1/\nu]$  는 전체매개변수들의 집합을 나타내며,  $p(\theta)$  는 매개변수들의 사전분포를 의미한다.  $M$ 은 시간강수량자료( $x$ )의 집성시간별 통계치 즉, 평균, 분산, 1차 자기상관계수, 무강수비율을 나타내며 우도함수  $p(M | \theta)$  는 Eq. (12)와 같다.  $t$ 는 시간강수량의 집성시간을 나타낸다. 매개변수의 사후분포는 Eq. (13)과 같이 나타낼 수 있다.

$$p(M | \theta) = \prod_{t=1,3,12,24} N(M_t | \lambda, k, \mu_x, \phi, \alpha, \nu^{-1}) \tag{12}$$

$$p(M | \theta) \propto \prod_{t=1,3,12,24} N(M_t | \lambda, k, \mu_x, \phi, \alpha, \nu^{-1}) \cdot N(\mu_\lambda, \sigma_\lambda^2) \cdot N(\mu_k, \sigma_k^2) \cdot N(\mu_{\mu_x}, \sigma_{\mu_x}^2) \cdot N(\mu_\phi, \sigma_\phi^2) \cdot N(\mu_\alpha, \sigma_\alpha^2) \cdot N(\mu_{\nu^{-1}}, \sigma_{\nu^{-1}}^2) \tag{13}$$

Eq. (13)은 모든 매개변수에 대한 적분을 통해 직접적으로 추정하는 것은 불가능하며 본 연구에서는 앞서 언급한 MCMC방법을 도입하여 매개변수들의 사후분포를 추정하게 된다. 본 연구에서는 모형의 Convergence를 확증하기 위해서 3개의 Chain을 독립적으로 시행하여 Sampling이 효과적으로 혼합(mixing)되도록 하였다. 한편 매개변수 추정을 위한 목적함수는 Eq. (14)과 같으며 (Cowpertwait et al., 2007), 결과적으로 Eq. (14)의 값이 최소화될 때 최적 매개변수가 된다.

$$f(x) = \sum_{i=1}^k \left[ \left( \frac{M_i(x)}{M_i'} - 1 \right)^2 + \left( \frac{M_i'}{M_i(x)} - 1 \right)^2 \right] \tag{14}$$

여기서  $M_i'$ 는 관측값으로부터 추정된 통계치를 나타내며  $M_i(x)$ 는 모의된 강수량으로부터 추정된 통계치를  $x$ 는  $k=6$ 개의 매개변수를 나타낸다.

**3. 연구결과**

**3.1 Bayesian MBLRP 모형의 매개변수 추정결과**

본 연구에서는 전주지역을 연구대상으로 하였으며 Bayesian MCMC 기법을 통해 MBLRP 모형의 매개변수( $\lambda, 1/\nu, \alpha, \mu_x, \phi, k$ )의 사후분포를 추정하였다. 매개변수의 최적화 및 불확실성 구간을 추정하기 위한 Bayesian 모형 구축은 MBLRP 모형의

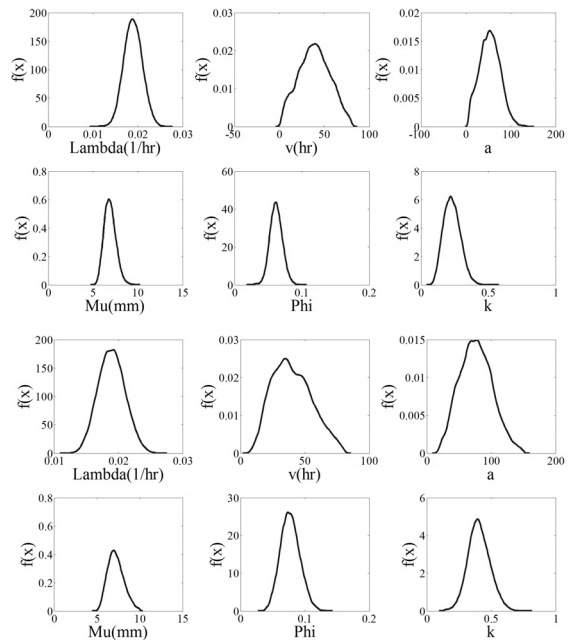


Fig. 2. Posterior Distribution of MBLRP Parameters

Table 1. The Estimated Optimal Parameter Using SCE-UA Algorithm in Jeonju Station

Parameters	July	August
$\lambda$	0.0178	0.0194
$\nu$	39.0616	31.9035
$\alpha$	39.0261	70.6537
$\mu_x$	6.3874	7.1972
$\phi$	0.0614	0.0675
$k$	0.1660	0.4170

통계치의 우도함수를 나타내는 Eq. (13)과 목적함수를 나타내는 Eq. (14)를 이용하였다. 매개변수 사후분포 추정을 위한 초기값은 SCE-UA 알고리즘으로부터 분석된 최적해를 이용하였으며(Table 1), 경계조건은 0부터 SCE-UA 알고리즘으로 분석된 최적해의 최대 2배까지 설정하였다. 샘플링은 20,000번 시행하였으며, 매개변수 초기값의 불확실성을 고려하여 초기 5,000번의 샘플은 사후분포 추정에 이용하지 않았다. 추정된 매개변수 사후분포의 결과는 정규분포형을 따르고 있음을 확인할 수 있으며, 각 매개변수간의 결합확률밀도함수를 효과적으로 계산되었음을 알 수 있다(Fig. 2 and Table 2).

### 3.2 Bayesian MBLRP 모형의 재현능력 평가

추계학적 관점에서 강수재현능력은 관측값의 통계치와 모의값의 통계치의 유사성으로 평가할 수 있다. 이를 위해 Bayesian MBLRP 모형을 통해 추정된 매개변수의 사후분포로부터 Monte Carlo 모의를 통해서 다수의 강수계열을 생성하여 통계치를 추정하였다. 검토에 사용한 자료는 기상청 산하 전주 기상관측소의 40년간 (1973~2012년) 시간강수량 자료를 이용하였으며, 비교적 강우강도가 큰 7, 8월을 대상으로 검토하였다. 재현능력 평가를 위한 통계치는 강수의 집성시간별 평균강수량, 분산, 1차 자기상관계수, 무강수비율이며 집성시간은 1, 3, 12, 24시간으로 하였다.

Table 2. Statistics of Posterior Distribution Associated with MBLRP Model Parameters

Month	Parameters	Mean	Standard Deviation	Probability (2.50%)	Median	Probability (97.50%)
July	$\lambda$	0.019	0.002	0.015	0.019	0.023
	$\nu$	39.100	17.250	6.927	38.950	72.490
	$\alpha$	53.610	22.620	11.560	53.310	99.070
	$\mu_x$	6.888	0.664	5.731	6.843	8.309
	$\phi$	0.062	0.009	0.044	0.062	0.081
	$k$	0.231	0.062	0.122	0.227	0.360
August	$\lambda$	0.019	0.002	0.015	0.019	0.023
	$\nu$	40.310	15.180	15.250	38.890	72.250
	$\alpha$	76.310	25.170	31.030	75.610	128.100
	$\mu_x$	7.189	0.931	5.601	7.110	9.203
	$\phi$	0.077	0.015	0.050	0.077	0.108
	$k$	0.405	0.086	0.243	0.401	0.582

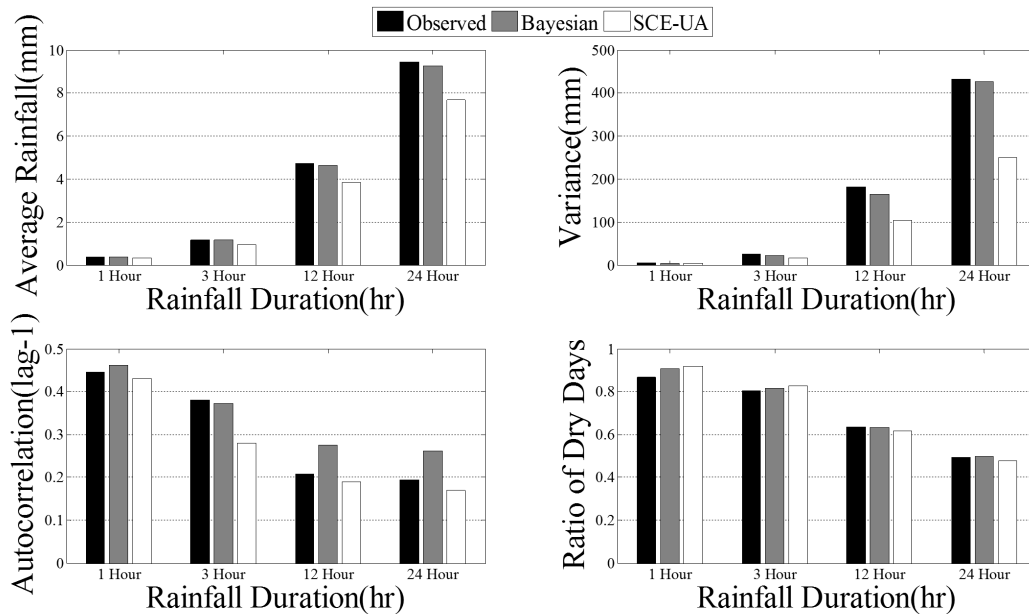


Fig. 3. Comparison of Performance in Terms of Reproducing Rainfall Statistic on July

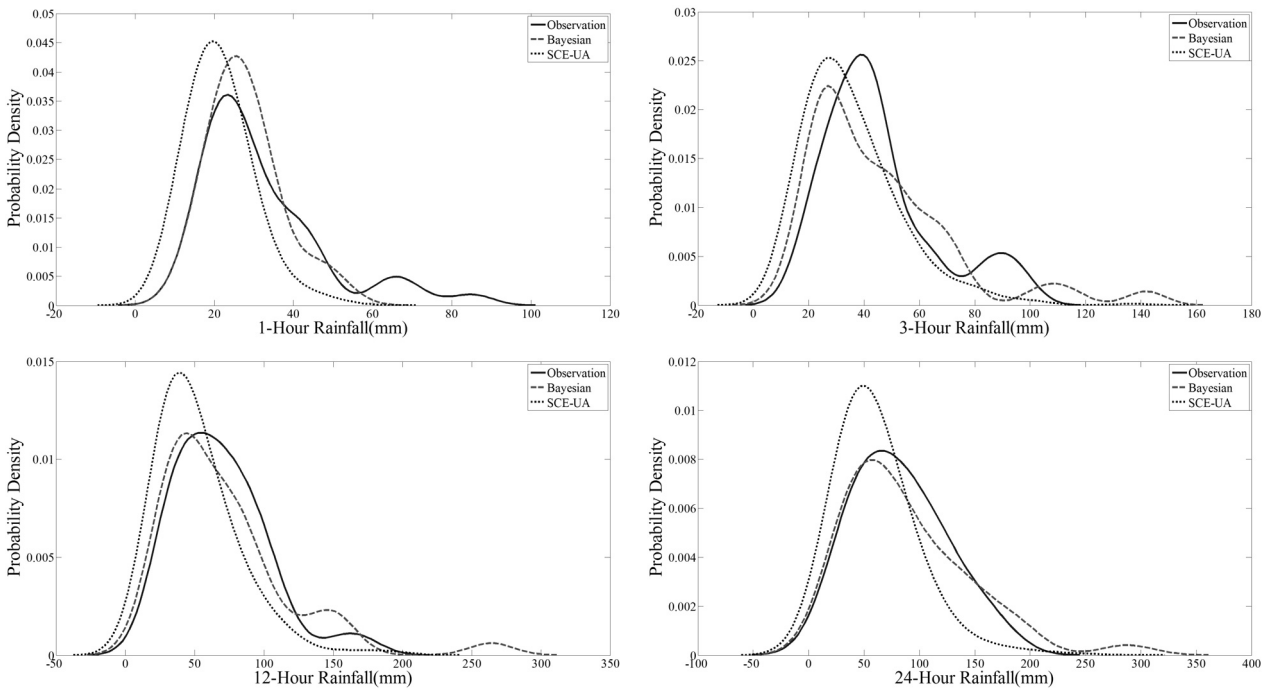


Fig. 4. Probability Density Function of Annual Maximum Rainfall Estimated from Bayesian MBLRP Model on July

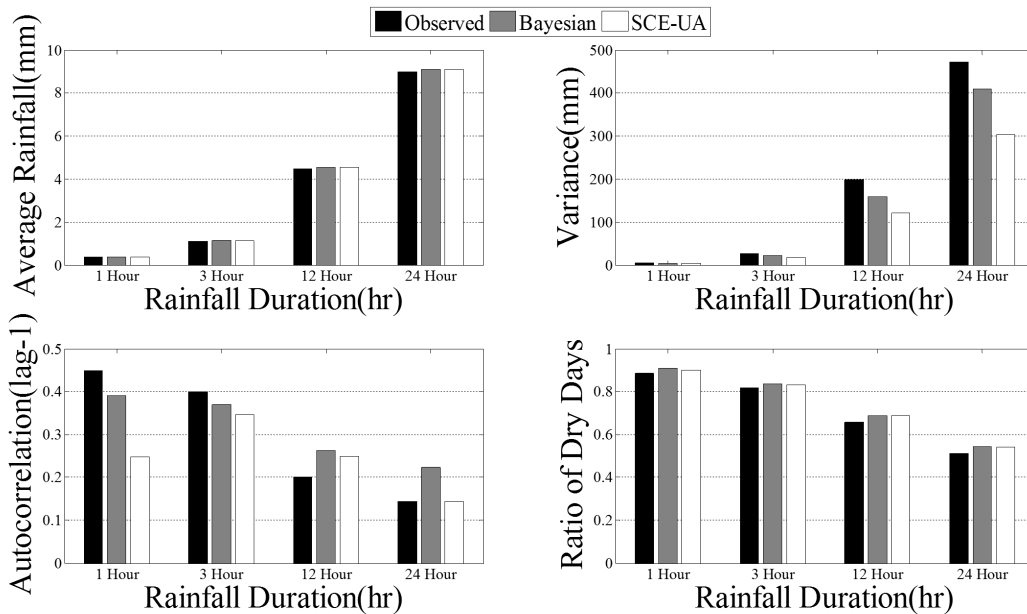


Fig. 5. Comparison of Performance in Terms of Reproducing Rainfall Statistic on August

본 연구에서는 제안한 모형의 성능을 평가하기 위해 우수한 매개변수 추정 성능을 보이는 것으로 알려진 SCE-UA 전역최적화 기법(Duan et al., 1994)을 이용한 모의결과와 비교 분석하였다. 그 결과 Bayesian 모의값이 SCE-UA 모의값에 비해 관측값의

통계치 적합이 우수한 것으로 나타났다(Figs. 3~6 and Table 3). 확률밀도함수(Figs. 4 and 6)를 비교해보면 Bayesian 모형이 관측값의 확률분포와 보다 유사하게 모의되고 있음을 확인할 수 있었다. 특히, 확률분포의 Tail부분에 거동이 상당히 유사하게 모의

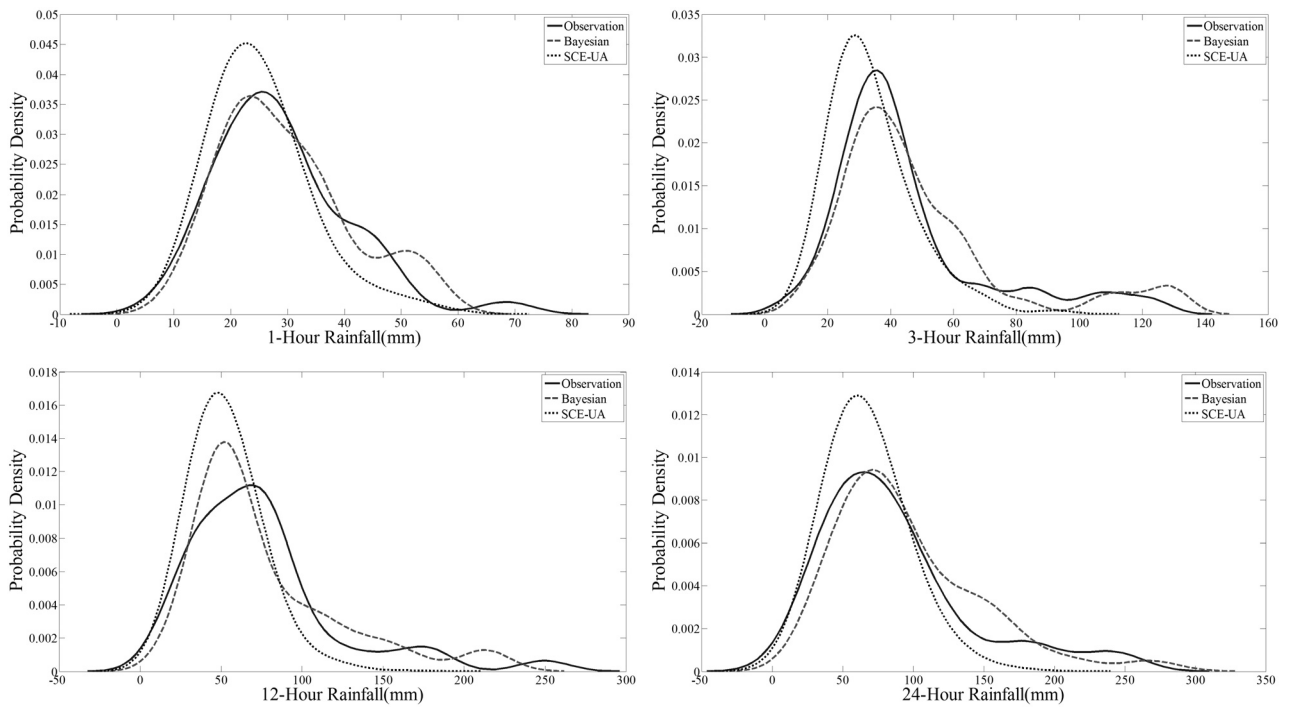


Fig. 6. Probability Density Function of Annual Maximum Rainfall Estimated from Bayesian MBLRP Model on August

Table 3. Comparison of Rainfall Statistics Between Observation, Bayesian Model and SCE-UA Algorithm Over Different Aggregation Time in July and August

Month	Statistics	Aggregation Time	Observation	Bayesian Simulation	SCE-UA Simulation
July	Mean	1hour	0.39	0.39	0.32
		3hour	1.18	1.16	0.96
		12hour	4.71	4.62	3.84
		24hour	9.42	9.24	7.68
	Variance	1hour	4.80	4.15	3.17
		3hour	25.54	22.67	16.52
		12hour	181.83	164.92	104.51
		24hour	432.85	426.85	250.23
	Autocorrelation	1hour	0.45	0.46	0.43
		3hour	0.38	0.37	0.28
		12hour	0.21	0.28	0.19
		24hour	0.19	0.26	0.17
Zero Depth Probability	1hour	0.87	0.91	0.92	
	3hour	0.80	0.82	0.83	
	12hour	0.64	0.63	0.62	
	24hour	0.49	0.50	0.48	
August	Mean	1hour	0.37	0.38	0.38
		3hour	1.12	1.14	1.14
		12hour	4.49	4.56	4.55
		24hour	8.98	9.11	9.10



Table 3. Comparison of Rainfall Statistics between Observation, Bayesian Model and SCE-UA Algorithm Over Different Aggregation Time in July and August (continue)

August	Variance	1hour	4.85	4.24	3.88
		3hour	26.50	21.84	16.97
		12hour	198.63	159.31	121.04
		24hour	471.99	409.54	303.74
	Autocorrelation	1hour	0.45	0.39	0.25
		3hour	0.40	0.37	0.35
		12hour	0.20	0.26	0.25
		24hour	0.14	0.22	0.14
	Zero Depth Probability	1hour	0.89	0.91	0.90
		3hour	0.82	0.84	0.83
		12hour	0.66	0.69	0.69
		24hour	0.51	0.55	0.54

Table 4. Comparison of Design Rainfall between Observation, Bayesian Model and SCE-UA Algorithm Over Different Aggregation Time in July and August

Month	Aggregation Time	Method	Design Rainfall according to Return Period(mm)							
			2yr	5yr	10yr	20yr	30yr	50yr	100yr	200yr
July	1hour	Observation	29.1	40.6	48.3	55.6	59.8	65.1	72.2	79.3
		Bayesian	26.3	34.3	39.6	44.7	47.6	51.3	56.2	61.2
		SCE-UA	20.0	27.0	31.6	36.0	38.6	41.8	46.1	50.3
	3hour	Observation	41.6	58.0	68.8	79.1	85.1	92.5	102.6	112.6
		Bayesian	40.9	60.5	73.4	85.9	93.1	102.0	114.1	126.1
		SCE-UA	32.4	46.7	56.2	65.3	70.5	77.1	85.9	94.7
	12hour	Observation	62.9	90.9	109.4	127.1	137.3	150.1	167.3	184.5
		Bayesian	63.1	96.6	118.8	140.1	152.4	167.7	188.3	208.9
		SCE-UA	46.4	68.7	83.5	97.6	105.8	116.0	129.7	143.4
	24hour	Observation	75.3	111.6	135.7	158.8	172.0	188.6	211.0	233.3
		Bayesian	79.1	123.0	152.1	180.0	196.0	216.1	243.2	270.2
		SCE-UA	54.7	80.5	97.5	113.9	123.3	135.0	150.9	166.7
August	1hour	Observation	27.0	37.4	44.3	50.9	54.7	59.5	65.9	72.3
		Bayesian	27.7	37.7	44.2	50.5	54.1	58.7	64.8	70.8
		SCE-UA	23.7	31.4	36.5	41.3	44.1	47.6	52.4	57.1
	3hour	Observation	40.6	59.0	71.1	82.8	89.5	97.9	109.2	120.5
		Bayesian	44.1	64.1	77.4	90.1	97.4	106.6	118.9	131.2
		SCE-UA	31.5	42.5	49.7	56.7	60.7	65.7	72.5	79.2
	12hour	Observation	66.9	102.4	125.9	148.5	161.4	177.7	199.5	221.3
		Bayesian	68.3	101.4	123.3	144.3	156.4	171.5	191.9	212.2
		SCE-UA	49.3	66.9	78.5	89.7	96.1	104.2	115.0	125.8
	24hour	Observation	77.7	119.2	146.7	173.1	188.3	207.3	232.9	258.5
		Bayesian	86.6	127.0	153.8	179.5	194.2	212.7	237.6	262.4
		SCE-UA	62.3	85.3	100.5	115.0	123.4	133.9	148.1	162.1

되고 있으며 최종적으로 극치강수량 재현효과가 상당 부분 개선되었음을 알 수 있다. 그러나 다른 통계치들과 다르게 자기상관계수가 집성시간이 커질수록 과대 추정되는 문제점을 확인할 수 있었다. 이는 강수의 지속성 또는 강수량의 크기가 관측치보다 일부 크게 생성되는 문제로 기인한다고 할 수 있다. 그러나 12시간이상에서는 상대적으로 자기상관계수가 크지 않으며, 통계적인 유의성이 없는 점을 고려한다면 강수모의기법으로의 활용성에는 큰 지장이 없을 것으로 판단된다.

### 3.3 Bayesian MBLRP 모형의 극치강수량 재현 효과

앞선 연구를 통해 Bayesian MBLRP 모형이 일단위 이하 강수량을 효과적으로 모의하고 있음을 확인하였다. 본 절에서는 극치강수량 재현효과를 검토하기 위해 관측강수량과 모의강수량의 빈도해석을 수행하였다. 빈도해석에 사용한 확률분포형은 우리나라에서 극대치강수량자료의 확률분포로 주로 사용되는 Gumbel 확률분포를 이용하였다. Gumbel 분포의 확률밀도함수(PDF)와 누가확률밀도함수(CDF)는 Eqs. (15) and (16)과 같다.

$$f(x) = \frac{1}{\sigma} \exp\left[-\frac{(x-\mu)}{\sigma} - \exp\left[-\frac{(x-\mu)}{\sigma}\right]\right] \quad (15)$$

$$F(x) = \exp\left\{\exp\left[-\frac{(x-\mu)}{\sigma}\right]\right\} \quad (16)$$

여기서  $f(x)$ 는 확률밀도함수를 나타내며,  $F(x)$ 는 누가확률밀도함수이다. 또한,  $\mu$ 는 위치매개변수(location parameter)이고,  $\sigma$ 는 규모매개변수(scale parameter)이다.

빈도해석에 사용한 재현기간은 2년, 5년, 10년, 20년, 30년, 50년, 100년, 200년을 기준으로, 집성시간 1시간, 3시간, 12시간, 24시간의 연속강수사상에 대하여 관측값 및 Bayesian 모의값, SCE-UA 모의값을 비교 분석하였다(Table 4). 비교 분석결과 Bayesian MBLRP의 극치강수량 재현 효과는 SCE-UA 모의결과에 비해 RMSE (Root Means Square Error) 기준 최소 10%~최대 92%까지 전반적인 극치강수량 재현성능이 향상되었다.

## 4. 결론

강수자료는 수자원 계획 수립 시 다양한 수문 모형의 입력 자료로써 이용되고 있지만, 시공간의 제약으로 신뢰성 있는 강수 자료 취득은 쉽지 않다. 특히, 과거에는 일단위 이하의 강수자료 취득이 어려웠기 때문에 이를 활용한 수자원 계획 수립 시 어려움이 따른다. 이러한 이유로 강수자료를 모의할 수 있는 다양한 강수모형이 개발되고 활용되어지고 있다. 국내외에서 가장 대표적인 강수모의

기법인 MBLRP모형을 포함해서 대부분의 추계학적 강수모의기법은 2차모멘트 이상의 통계치에서 과소 추정되는 문제점이 나타나고 있다. 이러한 점에서 본 연구에서는 기존 MBLRP 모형의 대표적인 문제점인 극치강수량 과소 추정 문제점을 개선하기 위한 방법으로 계층적 Bayesian 모형과 MBLRP 기법을 연계한 Bayesian MBLRP 모형을 개발하였다.

Bayesian MBLRP 모형은 각 매개변수간 결합확률분포를 계산하여 매개변수의 사후분포를 추정하므로 매개변수의 불확실성을 정량적으로 분석할 수 있으며, 추정된 사후분포로부터 Monte Carlo 모의를 통해 다양한 시간 규모에서 극치강수량에 대한 재현을 효과적으로 복원할 수 있는 장점을 확인할 수 있었다.

정량적인 재현능력 평가를 위해 전역 최적화기법인 SCE-UA 알고리즘을 적용한 모의 통계치와 비교 분석하였다. 그 결과 1차 모멘트인 평균강수량과 2차 모멘트인 분산을 기준으로 SCE-UA 모의 값에 비해 전주 기상관측소의 7월 평균강수량은 약 16%, 분산은 약 20~40%, 8월 분산은 약 7~22%의 오차개선효과가 있었다. 반면 8월 평균강수량 및 1차 자기상관계수, 무강수기간 비율 등은 SCE-UA 모형과 비슷한 수준을 나타냈다. 한편 극치강수량 재현효과는 약 4~39% , 평균적으로 약 20% 오차개선효과가 있었다.

현재 개발된 모형은 단일지점 적용을 위한 모형으로서 홍수량 산정시 유역전반의 강수특성을 대표하기에는 무리가 있다. 이러한 점에서 다지점에 적용할 수 있는 다변량 Bayesian MBLRP 모형 개발이 필요할 것으로 판단된다.

## 감사의글

본 연구는 국토교통부 물관리연구사업의 연구비지원(13AWMP B066744-01)에 의해 수행되었습니다. 저자들은 전북대학교 방재연구센터에 소속되어 연구를 수행하였습니다.

## References

Boughton, W. and Droop, O. (2003). "Continuous simulation for design flood estimation - A review." *Environmental Modelling and Software*, Vol. 18, No. 4, pp. 309-318.

Cowpertwait, P., Isham, V. and Onof, C. (2007). "Point process models of rainfall: Developments for Fine-scale Structure." *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, Vol. 463, No. 2086, pp. 2569-2587.

Duan, Q., Sorooshian, S. and Gupta, V. K. (1994). "Optimal use of the SCE-UA global optimization method for calibrating watershed models." *Journal of Hydrology*, Vol. 158, No. 3, pp. 265-284.

Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (2003).

- Bayesian data analysis*. CRC press, United States of America.
- 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.
- Kim, K. W. and Yoo, C. S. (2008). "A selection of the point rainfall process model considered on temporal clustering characteristics." *Journal of Korea Water Resources Association*, Vol. 41, No. 11, pp. 747-759 (in Korean).
- Kyoung, M. S., Sivakumar, B., Kim, H. S. and Kim, B. S. (2008). "Chaotic disaggregation of daily rainfall time series." *Journal of Korea Water Resources Association*, Vol. 41, No. 9, pp. 959-967 (in Korean).
- Kim, D. K., Shin, J. Y., Lee, S. O. and Kim, T. W. (2013). "The application of the poisson cluster rainfall generation model to the flood analysis." *Journal of Korea Water Resources Association*, Vol. 46, No. 5, pp. 439-447 (in Korean).
- Onof, C., Chandler, R., Kakou, A., Northrop, P., Wheatler, H. and Isham, V. (2000). "Rainfall modelling using Poisson-cluster processes: A review of developments." *Stochastic Environmental Research and Risk Assessment*, Vol. 14, No. 6, pp. 384-411.
- Rodriguez-Iturbe, I., Cox, D. and Isham, V. (1987). "Some models for rainfall based on stochastic point processes." *Proceedings of the Royal Society of London. A, Mathematical and Physical Sciences*, Vol. 410, No. 1839, pp. 269-288.
- Rodriguez-Iturbe, I., Cox, D. and Isham, V. (1988). "A point process model for rainfall: Further developments." *Proceedings of the Royal Society of London. A, Mathematical and Physical Sciences*, Vol. 417, No. 1853, pp. 283-298.
- Vanhaute, W., Vandenberghe, S., Scheerlinck, K., De Baets, B. and Verhoest, N. (2012). "Calibration of the modified Bartlett-Lewis model using global optimization techniques and alternative objective functions." *Hydrology and Earth System Sciences*, Vol. 16, No. 3, pp. 873-891.
- Verhoest, N., Troch, P. A. and De Troch, F. P. (1997). "On the applicability of Bartlett-Lewis rectangular pulses models in the modeling of design storms at a point." *Journal of Hydrology*, Vol. 202, No. 1, pp. 108-120.
- Wheatler, H., Isham, V., Chandler, R., Onof, C., Stewart, E., Bellone, E., Yang, C., Lekkas, D., Lourmas, G. and Segond M. (2006). *Joint DEFRA/EA flood and coastal erosion risk management R&D programme: Improved Methods for National Spatial-temporal Rainfall and Evaporation Modelling for BSM*, R&D Technical Report.