기후 및 지형정보가 고려된 극한 일 강수량 산정을 위한 계층적 베이지안 모형

Hierarchical Bayesian Model for Assessing Extreme Daily Rainfall Through Combining Climatic and Topographic Information

이정훈* · 원정은** · 최정현*** · 김상단****

Lee, Jeonghoon*, Won, Jeongeun**, Choi, Jeonghyeon***, and Kim, Sangdan****

Abstract

Frequency analysis of the annual maximum rainfall time series is essential for designing infrastructures to provide protection against local floods and related events. However, the results of the frequency analysis obtained are ambiguous. In this study, we aimed to develop a spatial hierarchical Bayesian model framework through combining the climatic and topographic information. To confirm the applicability of the proposed method, the results of at-site frequency analysis and regional frequency analysis using the index flood method were compared in the Busan-Ulsan-Gyeongnam region. Furthermore, a hierarchical Bayesian model was developed, in which the parameters of the generalized logistic distribution comprised relatively simple covariate relationships upon considering the possibility of expansion into various probability distributions and more complex covariate structures. The uncertainty of this model was analyzed using the coefficient of variation of rainfall quantile ensemble. The results confirmed that the regional frequency analysis using the hierarchical Bayesian model combined with the climatic and topographic information could provide an accurate estimate of extreme daily rainfall with relatively good agreement with the estimate at a specific site, but is a more reliable approach.

Key words : Generalized Logistic Distribution, Hierarchical Bayesian Model, Rainfall Extremes, Regional Frequency Analysis, Uncertainty

요 지

연최대강우시계열의 빈도해석은 지역 홍수방어와 인프라 시설 설계를 위한 필수적인 절차이다. 하지만 빈도해석의 결과에는 많은 불확실성이 내재되어 있다. 본 연구의 목적은 공간 계층적 베이지안 모형의 프레임 워크를 제안하는 것이다. 제안된 방법론의 적용 가능성을 살펴보기 위해 부산-울산-경남지역을 대상으로 기존의 지점빈도해석과 지수홍수법의 의한 지역빈도해석의 결과를 비교하였다. 본 연구에서는 비교적 단순한 공변량 관계를 맺는 일반화된 로지스틱 분포의 매개변수로 계층적 베이지안 모형을 구성하였으며, 추후 다양한 확률분포 및 더 복잡한 공변량 구조로 확장될 가능성을 염두에 두었다. 추가로 확률강수량 양상불의 변동 계수를 이용하여 모형의 불확실성을 살펴보았다. 기후 및 지형정보가 결합된 계층적 베이지안 모형을 이용한 지역빈도해석은 지점에서의 추정치와 비교적 잘 일치하면서 더 신뢰성 있는 극한 일 강수량 추정치를 제공할 수 있다는 것을 확인하였다.

핵심용어 : 일반화된 로지스틱 분포, 계층적 베이지안 모형, 극한강우, 지역빈도해석, 불확실성

*정회원, 부경대학교 환경공학과 전임연구원(E-mail: bravo281@hanmail.net)

Member, Research Fellow, Department of Environmental Engineering, Pukyong National University

**정회원, 부경대학교 지구환경시스템과학부 환경공학전공 박사과정(E-mail: won6922@pukyong.ac.kr)

Member, Ph.D. Student, Division of Earth Environmental System Science (Major of Environmental Engineering), Pukyong National University

***정회원, 부경대학교 지구환경시스템과학부 환경공학전공 박사과정(E-mail: jeonghyeon20@pukyong.ac.kr)

Member, Ph.D. Student, Division of Earth Environmental System Science (Major of Environmental Engineering), Pukyong National University

****교신저자, 정회원, 부경대학교 환경공학과 교수(Tel: +82-51-629-6529, Fax: +82-51-629-6523, E-mail: skim@pknu.ac.kr)

Corresponding Author, Member, Professor, Department of Environmental Engineering, Pukyong National University

1. 서론

극한강우사상의 크기와 빈도를 추정하는 것은 홍수에측 및 사회기반시설의 설계를 위해 필수적이다(Gumbel, 1958; Stedinger and Gris, 2008; Luke et al., 2017). 극한사상은 지역의 기반시설과 시민들에게 직접적인 영향을 미치며 높은 경제적 파급효과를 가진다(Fowler and Kilsby, 2000; Cheng and AghaKouchak, 2014). 하지만 관측자료가 충분하지 않은 일부 지역에서는 국지적인 극한사상을 예측하기 어려울 수 있으며, 상대적으로 관측망이 잘 갖춰진 지역일지라도 빈도해석에 기후 및 공간적인 정보를 통합하는 것은 쉬운 일이 아니다(Cooley et al., 2007).

Fisher and Tippett (1928)의 연구를 확장한 Gumbel (1941)의 초기 연구 이후, 지점빈도해석은 다양한 재현기간에 해당하는 강우량을 추정하는 데 널리 사용되어왔다(Davison and Smith, 1990; Bonnin et al., 2006; England et al., 2018). 일반적으로 지점빈도해석은 일반화된 극치 분포 또는 일반화된 로지스틱 분포와 같은 확률분포형에 극한값을 적합시키는 과정이 포함된다(Jenkinson, 1955; Hosking and Wallis, 1993). 그런 다음 적합한 분포형을 사용하여 다양한 크기의 강우량에 대한 발생 확률을 추정한다(Coles, 2001). 하지만 재현기간이 클수록 극한강우량 추정치의 불확실성이 증가하게 된다(Davison et al., 2012; Papalexiou et al., 2018).

극한강우 등의 자연재해는 물리적 과정으로 인해 발생하며 공간적 범위가 넓으므로 공간적 극한 현상에 대한 더 나은 이해와 평가가 요구된다(Davison et al., 2012). 하지만 극한 현상은 드물게 관측되므로 이에 해당하는 관측자료는 많지 않다(Hosking and Wallis, 1997).

공간적인 극한강우를 다루는 이론에는 copulas, max-stable 이론, 지수홍수법으로 대표되는 지역빈도해석, 계층적 베이지안 모형(hierarchical Bayesian model, HBM) 등이 있다(Cooley, 2009). 각 관측소의 자료를 개별적으로 모델링하는 대신에 공간이론을 적용하면서 가지게 되는 가장 중요한 장점은 유사한 관측소들의 정보를 공유하여 확률분포 매개변수 추정의 신뢰도를 향상하는 것에 있다(Casson and Coles, 1999). 지수홍수법을 이용한 지역빈도해석의 경우에는 지점별 연 최대 강수량 자료의 표준화를 통하여 균질한 지역 내의 자료폴링과정을 필수적으로 수행하고 있다(Dalrymple, 1960; Kim et al., 2008; Heo and Kim, 2019). 또한 Renard (2011)가 언급한 바와 같이 공간이론은 측정되지 않거나 불완전하게 측정된 관측소에서 확률분포 매개변수 추정이 가능하다는 장점이 있다. 그뿐만 아니라 베이지안 통계를 사용하면 극한강우를 모델링할 때 발생하는 불확실성을 적절하게 설명할 수 있다(Epstein, 1985).

본 연구에서는 HBM을 이용하여 지역의 극한강우를 설명하고자 한다. HBM은 지역의 극한강우에 영향을 미치는 지형적인 또는 기후적인 요인과 관련된 공변량 정보를 사용

하여 공간에서 모형 매개변수를 추정할 수 있다(Wikle et al., 1998; Ribatet et al., 2012; Love et al., 2020). HBM을 이용한 공간적인 극한강우분석의 장점은 아래와 같이 요약될 수 있다. (1) HBM은 연구영역을 균질한 영역으로 분해할 필요가 없다(Cooley et al., 2007); (b) 불확실성 분석이 용이하다(Banerjee et al., 2014); (c) 비정상성분석을 위한 확장이 가능하다(Cheng et al., 2014; Economou et al., 2014); (d) 지역의 지형적인 특징(예: 고도, 위도, 경도 등)과 기후적인 변수(예: 평균 강수량, 지표면 기온, 이슬점 등)를 포함할 수 있다(Ahn et al., 2017). 그러나 공간 HBM을 적용할 때 적절한 공변량을 탐색하고 공변량들과 모형 매개변수들 사이의 구조를 개발하는 것이 잠재적인 어려움으로 작용한다(Najafi and Moradkhani, 2013).

HBM은 세계 여러 지역에서 극한강우를 연구하는 데 사용되어왔다. Cooley et al. (2007)은 미국 콜로라도, Gaetan and Grigoletto (2007)은 이탈리아 Triveneto, Aryal et al. (2009)은 호주 남서부, Renard (2011)은 프랑스 남부, Sun et al. (2014)은 호주 퀸즐랜드 남동부, Dyrddal et al. (2015)은 노르웨이를 대상으로 HBM을 적용한 바 있다. Ragulina and Reitan (2017)은 전 지구영역을 대상으로 적용성을 평가하였다. Armal et al. (2018)은 미국의 극한 강우사상의 연간 빈도의 추세 및 기후 변화 및 기후변동이 기여도를 살펴보기 위해 HBM을 사용한 바 있다. Bracken et al. (2018)은 여러 수문 변수의 비정상성 빈도해석을 수행하기 위한 HBM을 제안한 바 있다. Love et al. (2020)은 HBM 구조에 지리 및 기후정보가 공변량으로 사용되는 경우 극한 강우의 공간 분석 결과가 개선된다는 것을 확인하였다. 국내에서는 Kim et al. (2014)이 표준화된 연 최대 강수량 시계열을 지형특성과 결합된 HBM에 적용하여 한강유역에서 지역빈도해석을 수행한 바 있으며, Jo et al. (2015)은 Peak-over-threshold 시계열을 이용하여 서울 지점의 비정상성 빈도해석에 HBM을 적용한 바 있다. 또한, Kim et al. (2017)은 강우 극치자료의 Scaling 특성을 고려하여 지역빈도해석을 수행하는 HBM을 제안한 바 있으며, Kim et al. (2018)은 기상변동성을 다양한 시간규모에서 고려하기 위해 HBM을 활용한 바 있다.

이러한 선행연구들을 기반으로 본 연구에서는 HBM을 이용한 지역빈도해석 모형을 개발하고 이를 한반도 동남부 부산-울산-경남지역에 적용해보고자 하였다. 제안된 모형은 다양한 공변량 관계를 수용할 수 있는 유연성을 가지고 있으므로 기후정보와 지형정보를 연계하여 공변량을 선택한 후, 상대적으로 간단한 공변량 구조를 가진 HBM을 구성하였다. 또한 제안된 방법은 연 최대 강수량 시계열에 별도의 정규화 과정을 요구하지 않는다. 이를 통하여 미계측 지역의 추정 시 정규화 과정에 따른 불확실성을 가능한 방지하고자 하였다. 원칙적으로 제안된 방법은 균질한 영역을 별도로 정의할 필요는 없다. 하지만, 지역빈도해석을 수행하는 적용영역이 너무 넓은 경우에는 공변량의 종류 및 구성 방법이

더 복잡하게 될 수밖에 없으므로 부산-울산-경남지역으로 적용영역을 한정시켰다.

2. 연구방법

2.1 자료

한국 기상청에서 운영 중인 종관기상관측소(Automated Synoptic Observing System, ASOS)들 중 부산-울산-경남지역의 10개 지점(울산, 부산, 통영, 진주, 거창, 합천, 밀양, 산청, 거제, 남해)의 1973년부터 2015년까지 43년 동안의 연 최대 일 강우량 자료(Annual Maximum Rainfall, AMR)가 사용되었다. Table 1에 사용된 관측소의 기본 정보를 수록하였다.

2.2 베이지안 계층 모형

본 연구에서는 연 최대 일 강우량 시계열의 빈도해석을 위한 확률분포형으로 일반화된 로지스틱(Generalized logistic, GLO) 분포를 적용하였다. GLO 분포의 확률밀도함수 $f(x)$ 는 아래 Eq. (1)과 같다.

$$f(x) = \frac{1}{\alpha} \left\{ 1 - \frac{k(x-\xi)}{\alpha} \right\}^{1/k-1} \left[1 + \left\{ 1 - \frac{k(x-\xi)}{\alpha} \right\}^{1/k} \right]^{-2} \quad (1)$$

여기서 α 은 축척 매개변수, k 은 형상 매개변수, ξ 은 위치 매개변수이다. 연 최대 시계열 x 의 범위는 형상 매개변수 k 의 부호에 따라 결정된다. 형상 매개변수가 $k < 0$ 이면 x 은 $\xi + \alpha/k < x < \infty$ 의 범위를 가져야 하므로, 연 최대 시계열의 최솟값이 $\xi + \alpha/k$ 보다 커야 GLO 분포의 적용이 가능해진다. 만일 $k > 0$ 이면 x 은 $-\infty < x < \xi + \alpha/k$ 의 범위를 가지게 되므로, $\xi + \alpha/k$ 이 연 최대 시계열의 최댓값보다 훨씬 큰 값이 되어야 GLO 분포를 이용한 빈도해석이 의미를 갖게 된다.

사용된 10개 지점에서 L-모멘트법으로 매개변수를 추정하여 살펴본 결과, 추정된 형상 매개변수는 음의 값이었으며, 각 지점별 연 최대 시계열의 최솟값이 L-모멘트로 추정된 $\xi + \alpha/k$ 보다 큰 값을 가지고 있음을 확인하였다. 또한, 적합도 검정 결과에서도 모두 GLO 분포 적용이 가능한 것으로 나타났다. 따라서 부산-울산-경남지역에 GLO 분포를 적용하는 것은 타당한 것으로 판단된다.

HBM의 첫 번째 단계에서 우도함수는 GLO 분포의 시간과 공간에 대한 곱으로 아래와 같이 표현된다.

$$L_1 = \prod_{j=1}^S \prod_{i=1}^N f(x_{ij} | \alpha_{ij}, \beta_{ij}, x_{oij}) \quad (2)$$

여기서 i 는 시간(year)을 의미하며, j 는 공간(site)을 의미한다.

HBM의 두 번째 단계에서 GLO 분포의 매개변수는 아래와 같은 다변량 정규분포를 갖는 것으로 정의된다.

$$\phi \sim N(\mu_\phi, \Sigma_\phi) \quad (3)$$

$$\psi \sim N(\mu_\psi, \Sigma_\psi) \quad (4)$$

$$\zeta \sim N(\mu_\zeta, \Sigma_\zeta) \quad (5)$$

여기서 $\phi_j = \ln[\alpha_j]$, $\psi_j = \ln[-k_j]$, $\zeta_j = \ln[\min(x_j) - \alpha_j/k_j - \xi_j]$ 이다. 따라서 μ_ϕ , μ_ψ , μ_ζ 는 $1 \times S$ 의 차원인 평균 벡터가 되며, Σ_ϕ , Σ_ψ , Σ_ζ 는 $S \times S$ 의 차원인 공분산 행렬이 된다.

GLO 분포 매개변수의 평균 벡터 μ_ϕ , μ_ψ , μ_ζ 는 아래와 같은 공변량의 함수로 표현된다.

$$\mu_{\phi_j} = a_{\phi,1} + a_{\phi,2}(\text{cov}_{\phi,1})_j + a_{\phi,3}(\text{cov}_{\phi,2})_j \quad (6)$$

$$\mu_{\psi_j} = a_{\psi,1} + a_{\psi,2}(\text{cov}_{\psi,1})_j + a_{\psi,3}(\text{cov}_{\psi,2})_j \quad (7)$$

$$\mu_{\zeta_j} = a_{\zeta,1} + a_{\zeta,2}(\text{cov}_{\zeta,1})_j + a_{\zeta,3}(\text{cov}_{\zeta,2})_j \quad (8)$$

여기서 a 는 평균 초 매개변수(hyper-parameter)이며, cov

Table 1. Site Information

Site	Latitude (°)	Longitude (°)	Elevation (m)	Annual precipitation (mm)	Minimum of AMR (mm/day)	Maximum of AMR (mm/day)
Ulsan	35.5601	129.3203	34.57	1,257.7	56.2	417.8
Busan	35.1047	129.0320	69.56	1,504.4	75.2	439.0
Tongyeong	34.8455	128.4356	31.24	1,433.4	68.0	340.5
Jinju	35.2084	128.1191	21.32	1,515.2	73.1	318.0
Geochang	35.6712	127.9110	221.40	1,292.5	66.0	253.0
Hapcheon	35.5650	128.1699	26.72	1,277.2	58.5	288.5
Miryang	35.4915	128.7441	8.31	1,230.6	33.7	271.4
Sancheong	35.4130	127.8791	138.22	1,525.6	60.1	332.5
Geoge	34.8882	128.6045	44.83	1,815.7	88.0	438.3
Namha	34.8166	127.9264	45.71	1,825.0	88.5	410.0

는 공변량이다. 참고로 공변량은 연평균 강수량과 같은 기후 자료 및 위도, 경도, 고도와 같은 지형자료가 이용될 수 있다.

GLO 분포 매개변수의 공분산 행렬 Σ_ϕ , Σ_ψ , Σ_ζ 는 아래와 같은 지수 covariogram의 형태로 표현된다.

$$\Sigma_\phi = b_{\phi,1} e^{-b_{\phi,2}d} \quad (9)$$

$$\Sigma_\psi = b_{\psi,1} e^{-b_{\psi,2}d} \quad (10)$$

$$\Sigma_\zeta = b_{\zeta,1} e^{-b_{\zeta,2}d} \quad (11)$$

여기서 b 는 공분산 초 매개변수이며, d 는 지점들 사이의 거리이다.

HBM의 세 번째 단계에서 초 매개변수 a 와 b 의 사전분포는 균등분포로서 정의된다. 정리하자면 본 연구에서 구성된 HBM의 평균 초 매개변수 a 의 개수는 9개이며, 분산 초 매개변수 b 의 개수는 6개이다.

2.3 지역 빈도해석을 위한 마코프 체인 몬테카를로 알고리즘

일반적인 베이지안 추론을 위한 마코프 체인 몬테카를로 (Markov Chain Monte-Carlo, MCMC) 방법의 수치적인 구현은 메트로폴리스-헤스팅스(Metropolis Hastings, MH) 알고리즘 등이 정형화되어 있으므로 어렵지 않으나, HBM의 경우에는 적절한 절차가 수치적인 구현을 위해 설정되지 않으면 사후분포를 올바르게 추출하는 것이 어려울 수 있다 (Berger et al., 2001). 따라서 본 연구에서는 MH 알고리즘에 기초하여 아래와 같은 절차를 수립한 후, 매개변수 및 초 매개변수의 사후분포를 추출하였다. MCMC의 초기화 과정은 다음과 같다.

- ① $a_{\phi,k}$ 의 사전분포에서 $a_{\phi,k}^1$ 을 무작위로 추출
- ② $b_{\phi,k}$ 의 사전분포에서 $b_{\phi,k}^1$ 을 무작위로 추출
- ③ Eqs. (6)과 (9)를 이용하여 μ_ϕ^1 과 Σ_ϕ^1 을 계산
- ④ Eq. (3)으로부터 ϕ_j^1 (여기서 $j=1,2,\dots,S$)을 무작위로 추출
- ⑤ 동일한 방법으로 a_ψ^1 과 b_ψ^1 을 무작위로 추출하여 μ_ψ^1 과 Σ_ψ^1 을 계산한 후, ψ_j^1 을 무작위로 추출
- ⑥ 동일한 방법으로 a_ζ^1 과 b_ζ^1 을 무작위로 추출하여 μ_ζ^1 과 Σ_ζ^1 을 계산한 후, ζ_j^1 을 무작위로 추출

초기화된 MCMC를 이용한 반복 추출방법은 다음과 같으며, 우선 매개변수 ϕ_j 의 갱신은 :

- ① 제안분포 $N(\phi_j^{m-1}, v\Sigma_\phi^{m-1})$ 로부터 ϕ_j^* 을 무작위로 추출하며, 여기서 v 는 추출 교란을 위한 계수를 의미
- ② 매개변수 ϕ_j 의 갱신을 위한 r_ϕ 계산

$$r_\phi = \sum_{j=1}^S \sum_{i=1}^T \ln[f(x|\phi_j^*, \psi_j^{m-1}, \zeta_j^{m-1})] \quad (12)$$

$$- \sum_{j=1}^S \sum_{i=1}^T \ln[f(x|\phi_j^{m-1}, \psi_j^{m-1}, \zeta_j^{m-1})]$$

$$+ \sum_{j=1}^S \ln[N(\phi_j^*|\mu_\phi^{m-1}, \Sigma_\phi^{m-1})]$$

$$- \sum_{j=1}^S \ln[N(\phi_j^{m-1}|\mu_\phi^{m-1}, \Sigma_\phi^{m-1})]$$

- ③ r_ϕ 와 0 중에서 작은 값이 0에서 1 사이에서 무작위로 추출한 균등난수에 자연로그를 취한 값보다 크면 $\phi_j^m = \phi_j^*$ 가 되며, 아니면 $\phi_j^m = \phi_j^{m-1}$ 이 됨.

다음 매개변수 $a_{\phi,k}$ 의 갱신의 갱신방법은 :

- ① 제안분포 $N(a_{\phi,k}^{m-1}, V_{a_{\phi,k}})$ 로부터 $a_{\phi,k}^*$ 을 무작위로 추출하며, 여기서 $V_{a_{\phi,k}}$ 는 추출 교란을 위한 분산으로 초 매개변수 $a_{\phi,k}$ 의 사전분포의 분산보다 작은 값으로 설정
- ② Eq. (6)을 이용하여 μ_ϕ^* 을 계산
- ③ 초 매개변수 $a_{\phi,k}^*$ 의 갱신을 위한 r_{a_ϕ} 계산

$$r_{a_\phi} = \sum_{j=1}^S \ln[N(\phi_j^m|\mu_\phi^*, \Sigma_\phi^{m-1})] - \sum_{j=1}^S \ln[N(\phi_j^m|\mu_\phi^{m-1}, \Sigma_\phi^{m-1})] \quad (13)$$

- ④ r_{a_ϕ} 와 0 중에서 작은 값이 0에서 1 사이에서 무작위로 추출한 균등난수에 자연로그를 취한 값보다 크면 $a_{\phi,k}^m = a_{\phi,k}^*$ 및 $\mu_\phi^m = \mu_\phi^*$ 가 되며, 아니면 $a_{\phi,k}^m = a_{\phi,k}^{m-1}$ 및 $\mu_\phi^m = \mu_\phi^{m-1}$ 이 됨.

다음 매개변수 $b_{\phi,k}$ 의 갱신의 갱신방법은 :

- ① 제안분포 $N(b_{\phi,k}^{m-1}, V_{b_{\phi,k}})$ 로부터 $b_{\phi,k}^*$ 을 무작위로 추출하며, 여기서 $V_{b_{\phi,k}}$ 는 추출 교란을 위한 분산으로 초 매개변수 $b_{\phi,k}$ 의 사전분포의 분산보다 작은 값으로 설정
- ② Eq. (9)를 이용하여 Σ_ϕ^* 을 계산
- ③ 초 매개변수 $b_{\phi,k}^*$ 의 갱신을 위한 r_{b_ϕ} 계산

$$r_{b_\phi} = \sum_{j=1}^S \ln[N(\phi_j^m|\mu_\phi^m, \Sigma_\phi^*)] - \sum_{j=1}^S \ln[N(\phi_j^m|\mu_\phi^m, \Sigma_\phi^{m-1})] \quad (14)$$

- ④ r_{b_ϕ} 와 0 중에서 작은 값이 0에서 1 사이에서 무작위로 추출한 균등난수에 자연로그를 취한 값보다 크면 $b_{\phi,k}^m = b_{\phi,k}^*$ 및 $\Sigma_\phi^m = \Sigma_\phi^*$ 가 되며, 아니면 $b_{\phi,k}^m = b_{\phi,k}^{m-1}$ 및 $\Sigma_\phi^m = \Sigma_\phi^{m-1}$ 이 됨.

나머지 매개변수 ψ_j , $a_{\psi,k}$, $b_{\psi,k}$ 및 ζ_j , $a_{\zeta,k}$, $b_{\zeta,k}$ 도 유사한 절차를 이용하여 갱신할 수 있다. 이때, 매개변수의 수렴 (convergence)을 위해서 서로 다른 초기 값을 가진 3개의 Chain을 구성하였으며 Brooks and Gelman (1998)이 제시한 potential scale reduction factor가 1에 가까이 수렴하는 것을

확인하였다.

3. 연구결과 및 토론

3.1 공변량 선택결과

HBM을 구성하기 위해서는 Eqs. (6)~(8)에 명시된 공변량의 선택이 필요하다. GLO 분포의 매개변수의 평균 μ_ϕ , μ_ψ , μ_ζ 을 위한 공변량을 찾기 위하여 지점별로 L-모멘트법을 이용하여 추정된 매개변수 ϕ , ψ , ζ 와 후보 공변량들 사이의 관계를 탐색하였다(Fig. 1 참조). 축척 매개변수의 경우에는 연평균 강수량과 위도를 사용하는 것이 가장 좋은 상관성이 보여주었으며, 형상 매개변수는 경도와 위도, 위치 매개변수는 경도와 고도가 가장 높은 상관성을 나타내었다. 따라서 각 매개변수별로 두 개의 공변량을 사용하여 Eqs. (9)~(11)와 같이 구성하였다. 참고로 Fig. 1에서 연평균 강수량, 위도, 경도, 고도의 값은 모두 표준화 과정을 거친 Z-score 값들(즉, $Z = (X - \mu_X) / \sigma_X$)을 적용하였다.

3.2 매개변수 사후분포 결과

Fig. 1의 정보에 기초하여 사전 설정된 초 매개변수들의 사전분포와 MCMC를 이용하여 추정된 초 매개변수들의 사후분포를 살펴보면 Fig. 2와 같다. 그림에서 빨간색 점선이 초 매개변수들의 사전분포를 의미한다. 평균 초 매개변수 a 는 설정된 사전분포로부터 사후분포가 잘 수렴되었으나, 공분산 초 매개변수 b 는 사전분포와 크게 다르지 않은 사후분포가 추정되었다.

특히 공분산 초 매개변수의 감쇄 매개변수 $b_{\phi,2}$, $b_{\psi,2}$, $b_{\zeta,2}$ 는 사전분포와 사후분포가 거의 같은 형태를 보여주고 있다. 감쇄 매개변수의 사전분포의 범위를 크게 넓히거나 좁혀도 유사한 결과가 획득되었다. 그러나 모형의 최종적인 결과인 극한강우량의 앙상블의 사후분포는 감쇄 매개변수의 사전분포의 범위에 그다지 민감도하게 반응하지 않는다는 것을 확인할 수 있었다. 참고로 본 연구와 동일하게 지수 variogram을 적용한 Cooley et al. (2007)의 연구의 경우에는 감쇄 매개변수의 사전분포의 범위를 0.075에서 0.6으로 매우 좁게

지정하기도 하였다. Eqs. (6)~(11)의 구조를 살펴보면, 평균 초 매개변수 a 에 의해 GLO 분포 매개변수의 공간 상관성이 명시적으로 설명될 수 있으며, 공분산 초 매개변수 b 는 평균 초 매개변수 a 에 의해 설명되지 않는 공간 상관성을 부차적으로 설명하고 있음을 살펴볼 수 있다. Fig. 2의 결과와 감쇄 매개변수의 민감도 분석으로부터 제안된 모형 구조에서 GLO 분포 매개변수의 공간 상관성의 대부분은 평균 초 매개변수 a 에 의해 설명되고 있음을 파악할 수 있다.

Fig. 3은 부산 지점만을 이용하여 MH 알고리즘으로 추출한 GLO 분포 매개변수의 사후분포(즉, 지점빈도해석의 결과)와 지역의 모든 자료를 이용하는 HBM 방법으로 추출한 부산 지점 GLO 분포 매개변수의 사후분포를 보여주고 있다. 위치 매개변수는 두 방법 모두 유사한 사후분포의 분산도를 보여주고 있지만, 축척 매개변수와 형상 매개변수의 경우에는 HBM 방법으로 추출한 사후분포의 분산도가 더 작은 것을 확인할 수 있다. 이러한 결과는 HBM 방법이 지역의 모든 이용 가능한 자료를 풀링함으로써 불확실성이 더 작은 매개변수를 추정할 수 있다는 사실을 말해주고 있다.

Fig. 4는 두 개 방법별로 부산-울산-경남지역의 모든 지점들에서 추출한 매개변수 앙상블의 box-plot를 보여주고 있다. 부산 지점과 유사하게 축척 매개변수와 형상 매개변수는 HBM 방법에 의한 사후분포의 분산도가 더 작게 나타나고 있음을 발견할 수 있다.

3.3 빈도해석 결과

Fig. 5는 지점빈도해석 및 지수홍수법을 이용한 지역빈도해석, 본 연구에서 제안된 HBM 방법을 이용한 지역빈도해석으로 추정된 재현기간별 확률강우량을 보여주고 있다. 지점에 따라 다소 차이가 있지만, 세 가지 방법 모두 크게 다르지 않은 확률강우량을 나타내고 있음을 살펴볼 수 있다. 전반적으로 지점빈도해석이 가장 작은 확률강우량을 나타내고 있으며, 지수홍수법이 가장 큰 확률강우량을 제시하고 있다. 지역 평균적으로 살펴볼 때, HBM과 지수홍수법에 의한 확률강우량은 서로 크고 작음이 교차하여 1% 이하의 차이를 보였으며, HBM과 지점빈도해석에 의한 확률강우량

Fig. 1. Relation between GLO Parameters and Covariates

(a) Hyper-parameter a

(b) Hyper-parameter b

Fig. 2. Posterior Distribution of Hyper-Parameter a and b

Fig. 3. Posterior Distribution of Parameters of GLO Distribution for At-Site and BHM-Based Regional Approaches at Busan Site

Fig. 4. Box-Plot of GLO Parameters for At-Site and HBM-Based Regional Frequency Analysis

Fig. 5. Daily Rainfall Quantile for Various Return Levels

의 차이는 5% 정도인 것으로 나타났다(Fig. 5(k) 참조).

적용된 10개 지점에서 확률강우량의 공간적 변동성은 HBM 방법을 이용한 지역빈도해석이 가장 작게 나타났으며, 지점빈도해석이 가장 큰 공간적 변동성을 보였다. 전체적으로 살펴볼 때, HBM 방법에 의한 확률강우량의 크기는 지수홍수법에 의한 지역빈도해석과 비슷하였으나 공간적인 변동성은 지점빈도해석과 상대적으로 더 유사함을 발견할 수 있었다.

Fig. 6은 지점빈도해석, 지수홍수법, HBM 방법으로 각각 추정된 합천 지점의 재현기간 50년 확률강우량의 사후분포를 보여주고 있다. 합천 지점의 재현기간 50년 확률강우량의 경우, 앙상블 평균의 측면에서는 지점빈도해석과 HBM 방법이 비슷한 값을 주고 있다. 추정된 확률강우량(즉, 앙상블 평균)의 불확실성 측면에서 살펴보면, 지점빈도해석이 가장 큰 불확실성을 보이고 있으며 지수홍수법과 HBM 방법은 비슷한 불확실성을 나타내고 있음을 발견할 수 있다.

이러한 개념을 확장하여, Fig. 7에 적용된 방법별 재현기간별 확률강우량 앙상블의 변동계수를 지점별로 나타내었다.

추정된 확률강우량의 불확실성이 클수록 앙상블의 변동계수가 크다고 볼 수 있다. 세 방법 모두 재현기간이 증가할수록 변동계수가 증가하고 있음을 파악할 수 있다. 재현기간 10년

Fig. 6. Posterior Distribution of Quantile for Return Level of 50-Year at Hapcheon Site

Fig. 7. Coefficient of Variation of Quantile Ensemble for Various Return Levels. Black-◇, Red-○, and Blue-□ Lines are Derived from At-Site, IFM-Based Regional, and HBM-Based Regional Frequency Analyses, Respectively

이하의 확률강우량에서는 방법별 변동계수의 차이가 크지 않았으며, 재현기간이 증가할수록 지점빈도해석으로 추정된 확률강우량 앙상블의 변동계수가 상대적으로 더 많이 증가하고 있으며, 이는 지점빈도해석의 불확실성이 가장 크다는 것을 의미한다. 지점별로 살펴보았을 경우에는 울산지점, 부산지점, 밀양지점에서는 홍수지수법의 불확실성이 가장 작았으며, 나머지 7개 지점에서는 HBM의 불확실성이 가장 작은 것으로 나타났다.

HBM을 이용한 지역빈도해석의 장점들 중 하나는 미계측 지역에서의 확률강우량 추정이 가능하다는 것이다. 부산-울산-경남지역에 적용된 모델의 경우 연평균 강수량, 고도, 위도, 경도에 관한 정보가 주어지면 부산-울산-경남 내 임의

의 지역에서 확률강우량을 추정할 수 있다. 고도, 위도, 경도의 경우 GIS 정보로부터 취득이 가능하나, 연 평균 강수량은 지점에서의 정보만 존재한다. 따라서 10개 지점의 연평균 강수량을 크리깅 방법을 이용하여 부산-울산-경남 지역의 격자별(공간 해상도 1 km × 1 km)로 분포시켜 연 평균 강수량을 계산하였다(Fig. 8(a) 참고). 이들 정보를 이용하여 추정한 부산-울산-경남지역의 재현기간 100년 확률강우량을 Fig. 8(d)에 나타내었다. 비교를 위하여 지점빈도해석과 지수홍수법으로 추정된 지점별 확률강우량의 크리깅 결과가 Figs. 8(b)와 8(c)에 도시되었다.

먼저 빈도해석방법에 따른 부산-울산-경남지역의 공간적인 분포를 살펴보면 지점빈도해석, 지수홍수법, HBM을 이

(a) Study area and annual maximum daily rainfall

(b) At-site

(c) Regional

(d) HBM

Fig. 8. Spatially Distributed Daily Rainfall Quantile for Return Level of 100-Year

용한 지역빈도해석순으로 변동성이 크며 지역적인 편차가 있지만 확률강우량 또한 높게 산정되는 것을 시각적으로 확인할 수 있다. 지점빈도해석 결과의 평균값은 405.8 mm, 최소 및 최대값은 각각 294.0 mm, 534.5 mm였으며, 표준편차는 42.4 mm로 나타났다. 지수홍수법에 의한 평균 확률강우량은 449.8 mm, 최소 및 최대값은 각각 366.4 mm, 588.2 mm였으며, 표준편차는 53.4 mm로 나타났다. 지수홍수법으로 추정된 확률강우량은 남서방향에서 최대치를 나타내며 북상할수록 낮아지는 공간적 분포를 보여주고 있다. 이는 각 지점들의 평균 연최대 강수량(Fig. 8(a) 참고)의 공간적 분포와 흡사한 형태로 지수홍수법은 평균 연최대 강수량의 수치에 많은 영향을 받는다는 것을 알 수 있다. 마지막으로 HBM을 이용한 지역빈도해석 결과의 평균값은 516.8 mm, 최소 및 최대값은 각각 380.9 mm, 752.7 mm였으며, 표준편차는 70.0 mm로 나타났다. 확률강우량은 서쪽에서 동쪽으로 갈수록 높아지는 경향을 확인할 수 있으며, 평균 확률강우량 및 최대값이 다른 빈도해석결과들보다 상대적으로 높은 수치를 나타내며 변동성도 큰 것으로 나타났다. 이는 단순히 연최대 일강우 시계열만을 사용하는 것이 아니라 여러 가지 지형 및 기후정보(본 연구에서는 연평균 강수량, 고도, 위도, 경도)를 반영하였기 때문이다. 하지만 그런 점을 감안하더라도 추정된 확률강우량이 다른 빈도해석방법에 비해 높는데, 이에 대한 원인은 우리나라 강우지점들이 대부분 고도가 낮은 지역에 위치함에 따라 실제 매개변수 추정시 사용되는 DEM의 고도 범위가 다르다는 이유도 일부 존재하지만, HBM의 미계측지역의 매개변수 추정과정에서 사용되는 공변량을 이용한 회귀분석식에 내재된 불확실성이 가장 큰 것으로 판단된다. 본 연구에서는 공변량을 선택하기 위해 가장 좋은 상관성을 나타내는 기후 및 지형정보를 이용하였지만 Fig. 1(a)을 참고하면 ϕ 와 ζ 추정에 사용된 회귀분석식의 결정계수가 낮다는 것을 확인할 수 있다. 이러한 한계점은 추후 연구에서 다양한 공변량과 매개변수와의 상관관계를 고려하여 개선될 필요가 있다. 하지만 지점별로 크리깅 하는 경우는 특정지점을 중심으로 확률강우량이 분포되지만, HBM은 확률강우량을 격자별 정보를 바탕으로 추정하기 때문에 확률강우량의 공간적인 거동을 더 우수하게 표현할 수 있음에 따라 기존의 지수홍수법에 의한 빈도해석의 좋은 대안이 될 수 있을 것이라 기대된다.

4. 결 론

기존의 강수량 지점빈도해석 또는 지수홍수법에 의한 지역빈도해석의 한계는 공간 전반에 걸쳐 극한강우에 대한 정보를 통합하는 것이 간단하지 않다는 것이었다. 본 연구에서는 일 강수량의 연 최대 시계열의 공간분석을 위한 HBM 프레임 워크를 적용하여 공간정보 통합의 문제를 풀어보고자 하였다. 본 연구에서 적용된 공간 HBM 프레임 워크에는

사용된 공변량과 관련된 모형 불확실성에 대한 설명이 포함된다.

기존의 선행 국내연구들과 본 연구의 가장 큰 차이는 기후정보와 지형정보가 결합된 베이지안 계층 모형이 적용되었다는 것에 있다. 외국의 경우에는 기후정보와 지형정보를 결합한 사례가 있었으나, 국내의 경우에는 지형정보를 결합한 연구사례는 찾아볼 수 있으나, 기후정보를 매개변수 추정에 결합한 사례는 찾아보기 어렵다. 기후정보(본 연구에서는 연 평균 강수량)와 지형정보(본 연구에서는 고도, 위도, 경도)가 결합된 공간분석을 수행함으로써 기존의 지수홍수법에 의한 지역빈도해석보다 확률강우량의 공간적인 분포를 더 잘 설명할 수 있었다.

지수홍수법을 이용한 지역빈도해석과 비교해볼 때 HBM 방법을 이용한 지역빈도해석의 가장 큰 차이는 지점별 연 최대 강수량 시계열의 정규화가 불필요하다는 것이다. 즉, 극한강우를 공간적으로 해석하면서도 적용된 확률분포형(본 연구의 경우 GLO 분포)의 매개변수가 지점별로 추정될 수 있다. 이러한 차이로 인하여 지점에서의 자료를 최대한으로 반영하면서도 불확실성을 감소시킬 수 있었다. 제안된 방법론을 이용하여 부산-울산-경남지역에 공간적으로 분포된 확률강우량 지도를 제공할 수 있었으며, 다양한 지속기간에 대한 확률강우량 지도를 생성하는데 사용될 수 있을 것이다. 또한 큰 어려움 없이 부산-울산-경남 이외의 지역에 적용하는 것도 가능할 것이다. 더 나아가 연 평균 강수량과 같은 기후정보가 GLO 분포 매개변수 추정을 위한 공변량으로 적용될 수 있기 때문에, 공변량 기반의 미래 확률강우량 전망에도 적용될 수 있을 것으로 기대된다.

본 연구에서는 GLO 분포의 매개변수를 비교적 간단한 방식(두 개 공변량의 선형 방정식)으로 처리하여 지역빈도해석을 수행하였다. 다른 지역 또는 더 넓은 지역에 적용하려면 지역 구분에 대한 명시적인 절차가 필요할 것이며, 확률분포를 비롯하여 공변량의 개수와 종류, 표현 방정식도 지역에 맞게 다시 구축하는 것이 바람직할 것이다. 또한, 미계측지역 결과에 대한 불확실성을 감안하여 자동기상관측장비(Automatic Weather System, AWS) 자료와의 비교도 방법론의 검증에 도움이 되리라 판단된다.

감사의 글

본 결과물은 환경부의 재원으로 한국환경산업기술원의 지능형 도시 수자원 관리사업의 지원을 받아 연구되었습니다(2019002950004).

References

- Ahn, K., Palmer, R., and Steinschneider, S. (2017). A Hierarchical bayesian model for regionalized seasonal

- forecasts: application to low flows in the northeastern united states. *Water Resources Research*, Vol. 53, pp. 503-521.
- Armal, S., Devineni, N., and Khanbilvardi, R. (2018). Trends in extreme rainfall frequency in the contiguous United States: Attribution to climate change and climate variability modes. *Journal of Climate*, Vol. 31, pp. 369-385.
- Aryal, S., Bates, B., Campbell, E., Li, Y., Palmer, M., and Viney, N. (2009). Characterizing and modeling temporal and spatial trends in rainfall extremes. *Journal of Hydrometeorology*, Vol. 10, pp. 241-253.
- Banerjee, S., Carlin, B., and Gelfand, A. (2014). *Hierarchical modeling and analysis for spatial data*. New York, NY, USA: Chapman and Hall/CRC.
- Berger, J., DeOliveira, V., and Sanso, B. (2001). Objective Bayesian analysis of spatially correlated data. *Journal of the American Statistical Association*, Vol. 96, pp. 1361-1374.
- Bonnin, G., Martin, D., Lin, B., Parzybok, T., Yekta, M., and Riley, D. (2006). *Precipitation-frequency atlas of the United States*. NOAA Atlas 14, Volume 1, Version 4; NOAA: Silver Spring, MD, USA.
- Bracken, C., Holman, K.D., Rajagopalan, B., and Moradkhani, H. (2018). A Bayesian hierarchical approach to multi-variate nonstationary hydrologic frequency analysis. *Water Resources Research*, Vol. 54, pp. 243-255.
- Brooks, S.P., and Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of computational and graphical statistics*, Vol. 7, pp. 434-455.
- Casson, E., and Coles, S. (1999). Spatial regression models for extremes. *Extremes*, Vol. 1, pp. 449-468.
- Cheng, L., and AghaKouchak, A. (2014). *Nonstationary precipitation intensity-duration-frequency curves for infrastructure design in a changing climate*. Scientific Report 4, Article No. 7093.
- Cheng, L., AghaKouchak, A., Gilleland, E., and Katz, R. (2014). Non-stationary extreme value analysis in a changing climate. *Climate Change*, Vol. 127, pp. 353-369.
- Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Statistics, London, UK: Springer.
- Cooley, D. (2009). *Extreme value analysis and the study of climate change*. *Climate Change*, 97, Article No. 77.
- Cooley, D., Nychka, D., and Naveau, P. (2007). Bayesian spatial modeling of extreme precipitation return levels. *Journal of American Statistical Association*, Vol. 102, pp. 824-840.
- Dalrymple, T. (1960). *Flood-frequency analyses*. U.S. Geological Survey Water-Supply paper, 1543-A.
- Davison, A., and Smith, R. (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society: Series B (Methodological)*, Vol. 52, pp. 393-425.
- Davison, A., Padoan, S., and Ribatet, M. (2012). Statistical modeling of spatial extremes. *Statistical Science*, Vol. 27, pp. 161-186.
- Dyrddal, A., Lenkoski, A., Thorarinsdottir, T., and Stordal, F. (2015). Bayesian hierarchical modeling of extreme hourly precipitation in Norway. *Environmetrics*, Vol. 26, pp. 89-106.
- Economou, T., Stephenson, D., and Ferro, C. (2014). Spatio-temporal modelling of extreme storms. *Annals of Applied Statistics*, Vol. 8, pp. 2223-2246.
- England, J., Cohn, T., Faber, B., Stedinger, J., Thomas, W., Veilleux, A., et al. (2018). *Guidelines for determining flood flow frequency—Bulletin 17C*. In *Techniques and Methods*; U.S. Geological Survey: Reston, VA, USA, p. 168.
- Epstein, E. (1985). *Statistical inference and prediction in climatology: A bayesian approach*. Boston, MA: American Meteorological Society, Meteorological Monographs 20.
- Fisher, R., and Tippett, L. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, Vol. 24, pp. 180-190.
- Fowler, H., and Kilsby, C. (2000). A regional frequency analysis of United Kingdom extreme rainfall from 1961 to 2000. *International Journal of Climatology*, Vol. 23, pp. 1313-1334.
- Gaetan, C., and Grigoletto, M. (2007). A hierarchical model for the analysis of spatial rainfall extremes. *Journal of Agricultural, Biological and Environmental Statistics*, Vol. 12, pp. 434-449.
- Gumbel, E. (1941). Probability-interpretation of the observed return-periods of floods. *Eos, Transactions American Geophysical Union*, Vol. 22, pp. 836-850.
- Gumbel, E. (1958). *Statistics of Extremes*. New York, NY, USA: Columbia University Press.
- Heo, J., and Kim, H. (2019). Regional frequency analysis for stationary and nonstationary hydrological data. *Journal of Korean Water Resources Association*, Vol. 52, pp. 657-669.
- Hosking, J., and Wallis, J. (1993). *Some statistics useful*

- in regional frequency analysis. *Water Resources Research*, Vol. 29, pp. 271-281.
- Hosking, J., and Wallis, J. (1997). *Regional frequency analysis: An approach based on l-moments*. New York: Cambridge University Press.
- Jenkinson, A. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society*, Vol. 81, pp. 158-171.
- Jo, A., Cho, H., and Kim, G. (2015). Nonstationary frequency analysis using a hierarchical Bayesian model. *Journal of Korean Society of Hazard Mitigation*, Vol. 15, pp. 19-24.
- Kim, J.Y., Kwon, H.H., and Lee, B.S. (2017). A bayesian glm model based regional frequency analysis using scaling properties of extreme rainfalls. *Journal of The Korean Society of Civil Engineers*, Vol. 37, pp. 29-41.
- Kim, J., Kwon, H., and Lim, J. (2014). Development of hierarchical spatial regional frequency analysis model considering geographical characteristics. *Journal of Korean Water Resources Association*, Vol. 47, pp. 469-482.
- Kim, J., Nam, W., Shin, J., and Heo, J. (2008). Regional frequency analysis of south Korean rainfall data using Forgem method. *Journal of Korean Water Resources Association*, Vol. 41, pp. 405-412.
- Kim, Y.T., Lee, M.S., Chae, B.S., and Kwon, H.H. (2018). A Development of Summer Seasonal Rainfall and Extreme Rainfall Outlook Using Bayesian Beta Model and Climate Information. *Journal of The Korean Society of Civil Engineers*, Vol. 38, 655-669.
- Love, C.A., Skahill, B.E., England, J.F., Karlovits, G., Duren, A., and AghaKouchak, A. (2020). Integrating Climatic and Physical Information in a Bayesian Hierarchical Model of Extreme Daily Precipitation. *Water*, Vol. 12, pp. 2211.
- Luke, A., Vrugt, J., AghaKouchak, A., Matthew, R., and Sanders, B. (2017). Predicting nonstationary flood frequencies: Evidence supports an updated stationarity thesis in the United States. *Water Resources Research*, Vol. 53, pp. 5469-5694.
- Najafi, M., and Moradkhani, H. (2013). Analysis of runoff extremes using spatial hierarchical Bayesian modeling. *Water Resources Research*, Vol. 49, pp. 6656-6670.
- Papalexiou, S., AghaKouchak, A., and Foufoula-Georgiou, E. (2018). A diagnostic framework for understanding climatology of tails of hourly precipitation extremes in the United States. *Water Resources Research*, Vol. 54, pp. 6725-6738.
- Ragulina, G., and Reitan, T. (2017). Generalized extreme value shape parameter and its nature for extreme precipitation using long time series and the Bayesian approach. *Hydrological Sciences Journal*, Vol. 62, pp. 863-879.
- Renard, B. (2011). A bayesian hierarchical approach to regional frequency analysis. *Water Resources Research*, Vol. 47, pp. W11513.
- Ribatet, M., Cooley, D., and Davison, A. (2012). Bayesian inference from composite likelihoods, with an application to spatial extremes. *Statistica Sinica*, Vol. 22, pp. 813-845.
- Stedinger, J., and Gris, V. (2008). Flood frequency analysis in the United States: Time to update. *ASCE Journal of Hydrologic Engineering*, Vol. 13, pp. 199-204.
- Sun, X., Thyer, M., Renard, B., and Lang, M. (2014). A general regional frequency analysis framework for quantifying local-scale climate effects: A case study of ENSO effects on Southeast Queensland rainfall. *Journal of Hydrology*, Vol. 512, pp. 53-68.
- Wikle, C., Berliner, L., and Cressie, N. (1998). Hierarchical bayesian space-time models. *Environmental and Ecological Statistics*, Vol. 5, pp. 117-154.

Received	May 11, 2021
Revised	May 12, 2021
Accepted	July 28, 2021