

A Hierarchical Bayesian Modeling of Temporal Trends in Return Levels for Extreme Precipitations

Yongku Kim^{a,1}

^aDepartment of Statistics, Kyungpook National University

(Received February 10, 2015; Revised March 12, 2015; Accepted March 12, 2015)

Abstract

Flood planning needs to recognize trends for extreme precipitation events. Especially, the r -year return level is a common measure for extreme events. In this paper, we present a nonstationary temporal model for precipitation return levels using a hierarchical Bayesian modeling. For intensity, we model annual maximum daily precipitation measured in Korea with a generalized extreme value (GEV). The temporal dependence among the return levels is incorporated to the model for GEV model parameters and a linear model with autoregressive error terms. We apply the proposed model to precipitation data collected from various stations in Korea from 1973 to 2011.

Keywords: Bayesian analysis, daily precipitation, extremes, generalized extreme value distribution, return level, temporal trend.

1. 서론

요즘 집중호우와 같은 이상기후는 우리에게 미치는 영향이나 피해가 일반적인 사건에 비해서 크기 때문에 이러한 사건에 대한 패턴 분석이나 예측에 관한 연구가 최근 활발하게 이루어지고 있다. 일반적으로 이러한 사건은 아주 큰 값을 가지지만 그 발생 빈도는 상대적으로 아주 작은 경향을 가지는데, 특히 통계학 분야에서는 이러한 사건에 대한 분포와 그 분포에 관련된 추론에 관한 연구가 활발하게 진행되고 있다. 예를 들어, 대기 중 이산화탄소(CO₂) 농도 증가와 같은 기후조건 변화에 따른 대기 순환 모의 모형 결과는 온실 효과와 함께 강수 강도의 증가를 예측하고 있는데, 이는 지구온난화에 따른 강한 대류성 강수의 증가와 대규모 비대류성 강수의 감소에 기인한다고 알려져 있다. 우리나라의 경우, 집중호우에 의한 홍수 피해가 증가함에 따라 국내 강수자료의 체계적인 분석을 통한 강수량의 비정상성에 대한 해석과 강수의 비정상성이 반영된 설계 강수량 산정 방법의 개발이 시급한 실정이다. 실제 국내에서도 이미 비정상 강수 빈도 해석에 관한 활발한 연구가 진행되고 있으며, Jang 등 (2011)은 연 최대 강수량의 회귀직선에 대한 잔차의 수문학적 빈도 해석을 바탕으로, 가까운 미래로 설정된 목표연도의 확률 강수량을 산정하는 방법을 제안하였으며, Lee (2010)는 우리나라(대한민국) 강수량 계열의 특성 평가를 통한 경

This work was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry Science, ICT & Future Planning(No. 2014R1A1A1007075).

¹Department of Statistics, Kyungpook National University, 80 Daehak-Ro, Daegu 702-701, Korea.

E-mail: kim.1252@knu.ac.kr

향성을 고려한 비정상성 빈도해석을 위한 모형을 제시하는 등 강수계열의 비정상성 평가 및 비정상성 빈도 해석 기법을 제시 하였다. Strupczewski 등 (2001)은 Akaike Information Criterion(AIC)의 관점에서 56개의 모델 중 비정상 홍수 빈도에 관한 최적화 모형을 비교 분석하였다. Cunderlik 등 (2008)은 지점에 대한 홍수 빈도 모형에 대한 비정상 접근법을 제안하였다. 통계학 분야에서는 극단값 분포에 관한 연구가 주로 이루어져 왔는데, 통계적인 관점에서의 여러가지 응용에 대한 일반적인 이론과 분석 방법은 Embrechts 등 (1997)과 Coles (2001)에서 참고할 수 있다.

최근에 들어 강수 모형의 비정상성과 함께 집중호우와 같은 극단적인 사건이 언제 일어날 것인지 그리고 얼마나 큰 사건이 일어날 지에 대한 관심이 증가하고 있는데, 특히 두 번째 물음에 관한 답을 반환 기간을 통해서 알아볼 수 있다. 반환 기간은 지진이나 홍수와 같은 익스트림 이벤트의 발생 가능성에 대한 추정치이며, 일반적으로 장기간에 걸친 평균 반복 간격을 나타내는 기록 데이터에 기초하여 작성된 통계치이다. 따라서 특정 위험 구역에서 프로젝트를 진행할 때, 특정 반환 기간과 이벤트를 견딜 수 있도록 구조를 설계한다. 여기에서 발생하는 이벤트의 확률은 시간에 걸쳐 변화하지 않는 것으로 가정하고 과거의 이벤트와는 독립이라고 가정한다. 이론적인 반환 기간은 이벤트가 특정 수준을 초과할 확률의 역수로 표현된다. 지금까지는 주로 단변량 극단값에 관한 여러가지 통계모형 및 분석방법이 주로 연구되어져 왔다면 최근에는 이러한 모형이나 방법을 다변량 극단값 또는 공간구조를 가지는 극단값으로 확장하는 연구가 활발하게 진행되고 있다. 다변량 극단값 또는 공간구조를 가지는 극단값에 관련된 연구는 주로 이들 극단값들 사이의 종속성을 어떻게 설명하는지에 중점을 두고 진행되어 왔는데 이러한 연구 결과는 집중호우와 같은 극단기후사건들의 모형에 적용될 수 있다. 다변량 극단값 분포에 관한 연구는 de Haan (1985), Coles 등 (1999), Schlather와 Tawn (2003), Heffernan과 Tawn (2004)에서 참고할 수 있다. 최근에는 베이지안 모형에 관한 연구가 활발해지면서 베이지안 다층구조를 이용하여 다변량 극단 분포를 구성하는 방법에 관한 연구가 진행되었는데, Cooley 등 (2006)은 베이지안 다층구조를 가지는 Generalized Extreme Value 모형을 소개하였고, Casson과 Coles (1999)는 어떤 지점에서의 특정값 이상의 극단값에 대한 공간모형을 연구하였다. 최근 Cooley 등 (2007)은 극단값 분포의 모수에 공간구조를 적용하고 이를 통한 공간구조를 가지는 극단값의 분포를 구성하고 이를 통해 집중호우에 관한 반환주기에 대한 공간 구조를 구현하였다. 특히 반환주기는 극단값 분포의 모수들의 함수 형태로 표시되는데 베이지안 모형을 이용하여 이 반환주기의 분포를 구할 수 있다. 하지만 비정상적인 강수자료에 대한 극단값 분포에 대한 연구나 반환주기에 대한 연구는 제한적이라고 할 수 있다.

본 논문에서는 비정상 극치 강수 자료에 대해 계층적 베이지안 모형을 적용하여 시간에 따른 모수의 변화를 추정하며, 미래 확률 강수량에 대한 극단값 분포를 예측하고 더 나아가 반환기간에 대한 경향과 예측 값을 얻고자 한다. 이전의 고전적 통계 방법을 통한 강수 자료의 모수 추정연구의 경우, 자료의 정상성 가정 하에 고정된 모수를 추정하는 방법으로, 최근 나타난 비정상 강수 사상과 같이 강수량이 가지는 분포의 모수적 변화가 예상되는 경우 해석상 문제가 발생한다. 이러한 문제점을 해결하기 위해 모형의 관심모수에 시간에 따른 자기 상관 선형 회귀 함수를 적합한 계층적 베이지안 모형을 고려한다. 결과적으로 복잡한 다차원의 모수 추정을 위해 마르코프 연쇄-몬테 칼로(Markov Chain Monte Carlo; MCMC) 방법을 이용한 베이지안 추정을 실시한다. 또한 모형에기반하여 산정된 예측 모수를 통해 목표 연도별 재현기간별 반환기간을 추출하여 강수관련 사회 제반시설을 위한 척도를 제시하기를 기대한다. 제안된 모형의 효율성을 확인하기 위해서 1973년부터 2011년까지 39년 동안의 우리나라 여러 지역의 기상 관측소에서 관측된 일일 강우량 자료를 사용하여 대표적인 극단값 분포인 Generalized Extreme Value(GEV) 분포에 적합 시키고, 계층적 베이지안 모형을 이용하여 이들 분포의 모수들에 자기상관 시간모형을 소개한 후 우리나라 여러 지역에 대한 반환기간에 대한 시간에 따른 경향을 확인하였다.

2. 극단값 분포

극대값의 분포이론은 주로 확률분포의 꼬리부분에 대한 통계모형을 제공하는데 일반적인 단변량 극대값에 대한 분포에 대한 이론은 잘 알려져 있고 이는 근사적인 분포인 Generalized Extreme Value(GEV) 분포를 이용해서 설명된다. 이는 독립적이고 동일한 분포를 가진 극단치를 무작위로 샘플링 했을 때, 그 분포가 Gumbel, Frechet 그리고 Weibull 분포 중에 하나를 따른다는 극단치이론과 그 맥을 같이 한다. 일반적으로 극단치이론은 확률분포의 꼬리의 형태에 대한 통계적 이론으로 극단적인 사건에 적합한 보다 나은 분포를 제공하기 때문에 위험분석에 있어서는 좀 더 유연성을 가질 수 있다. 여러 분야에서 응용되는 극단치 이론모형은 기본적으로 두 개의 모형으로 나눌 수 있다. 우선 동등한 크기로 전체 표본을 나누어 각 블록의 최대값(block maxima) 혹은 최소값으로 구성된 관측치들의 분포를 추정하는 모형과 높은 분계점(threshold)을 초과한 모든 관측값에 대한 모형이다. 일반적으로 전자의 관측값에 대해서는 GEV 분포를 가정하고 후자의 관측값에 대해서는 Generalized Pareto 분포를 가정한다. GEV 분포는 Gumbel, Frechet 그리고 Weibull 분포를 통합적인 모형으로 표현한 것으로 분포함수는 다음과 같이 표현할 수 있다.

$$F(z) = P(Z < z) = \begin{cases} \exp \left\{ - \left[1 + \xi \left(\frac{z - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right\}, & \xi > 0, \\ \exp \left\{ - \exp \left(\frac{z - \mu}{\sigma} \right) \right\}, & \xi = 0, \end{cases} \quad (2.1)$$

여기에서 μ 는 위치모수이고, σ 는 척도모수이며, ξ 는 형상모수이다. 특히 GEV 분포는 $\xi > 0$ 일 때 Frechet 분포가 되고, $\xi < 0$ 일 때 Weibull 분포가 되며, $\xi = 0$ 일 때 Gumbel 분포가 된다. 참고로 Figure 2.1에서는 기상청(KMA)으로부터 얻은 1973년 1월 1일부터 2011년 12월 31일까지로 39년동안 우리나라 대표적 강우지역 (울릉도, 남해안 지역, 대구)에 대한 측정된 일별 강우량 관측자료를 사용하여 연간 블록 최대값 자료를 표시한 것이다.

앞에서 소개한 극단값 분포에 대한 반환주기는 각 극단값 분포의 모수의 함수형태로 표현되는데 r 년 반환주기는 이 특정 수준을 초과하는 이벤트가 일어날 확률이 $1/r$ 인 경우이다. 일반적으로 GEV 분포에 대해서는 r 년 반환주기는 다음과 같이 표현된다.

$$z_r = \mu - \frac{\sigma}{\xi} \left\{ 1 - \left[-\log \left(1 - \frac{1}{r} \right) \right]^{-\xi} \right\}, \quad (2.2)$$

여기에서 μ , σ 와 ξ 는 GEV 분포의 모수이다. 따라서 극단값 분포에 대한 반환주기는 각 블록의 최대값(block maxima)이나 높은 분계점(threshold)을 초과한 모든 관측값의 분포를 추정하여 추정된 각 극단값 분포의 모수를 이용하여 구할 수 있다. 단, 어떤 특정한 높은 분계점(threshold)을 초과한 모든 관측값의 분포인 Generalized Pareto 분포의 경우에도 유사한 방법으로 반환기간을 구할 수 있다 (Cooley 등, 2007).

3. 반환기간을 위한 계층적 베이지안 모형

2장에서 소개된 기존의 반환주기에 시간에 따른 비정상성을 적용하기 위해서 베이지안 모형에서 많이 사용되는 다층구조를 이용하였다. 반환주기의 분포에 대한 모수적인 정보가 부족하므로 반환주기에 직접 시간모형을 적용하는 대신에 각 극단값 분포의 모수에 시간구조를 가지는 랜덤효과 모형을 가정함으로써 간접적으로 반환주기에 시간적 연관성을 소개하였다. 즉, 단위시간 i 의 반환주기는 각 단위시간의

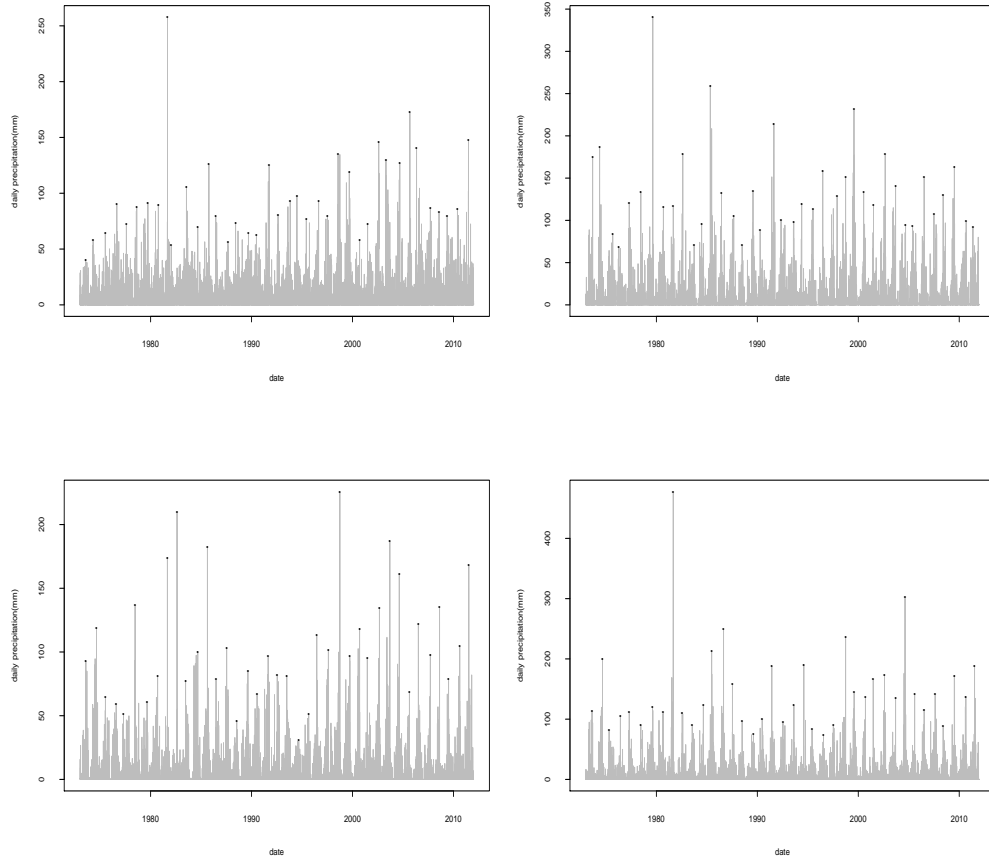


Figure 2.1. Yearly maximum precipitations (block maxima) of Ulleung-do, Geoje, Haenam and Daegu (clockwise from the left).

관측 극단값들에 대한 극단값 분포의 모수들인 위치모수 μ_i , 척도모수 σ_i 그리고 형상모수 ξ_i 의 함수로 표현된다. 따라서 이들 극단값 분포의 모수들에 시간구조를 가지는 사전분포를 가정하고자 한다.

계층적 베이저안 랜덤효과 모형은 시간에 따른 개체내의 상관을 각 개체에 랜덤 효과를 사용하여 모형화할 뿐만 아니라 사전 분포의 초모수에 적절한 분포를 부여함으로써 다른 개체들로부터 정보를 빌려오는 것이 가능하게 한다. 즉 경시적 데이터에 계층적 베이저안 모형을 구축하여 관심 있는 개체에 대해 시간으로부터 정보를 빌려오는 것이 가능하다.

먼저 z_{ij} , $i = 1, \dots, T$, $j = 1, \dots, n_i$ 를 i 번째 단위 시간에서의 j 번째 블록 최대 강수량이라고 하고 아래와 같은 분포를 따른다고 가정하자.

$$z_{ij} \stackrel{ind}{\sim} \text{GEV}(\mu_i, \sigma_i, \xi_i), \quad i = 1, \dots, T, \quad j = 1, \dots, n_i. \quad (3.1)$$

즉, 여기에서 μ_i , σ_i 그리고 ξ_i 는 i 번째 단위 시간에서의 GEV 분포의 위치모수, 척도모수 그리고 형상모수이다. 고전적 통계 추정을 통해 산정한 사전적 연구결과에서 이들 각 모수에서의 시간의 흐름에 따른 경향성이 나타나므로 이를 사전 정보로 간주하여 아래와 같은 랜덤효과 모형을 정의한다. 각각

$i = 1, \dots, T$ 에 대해서,

$$\mu_i = \alpha_0 + \alpha_1 t_i + u_i, \quad u_i = \phi_1 u_{i-1} + e_{1i}, \quad (3.2)$$

$$\log(\sigma_i) = \beta_0 + \beta_1 t_i + v_i, \quad v_i = \phi_1 v_{i-1} + e_{2i}, \quad (3.3)$$

$$\xi_i = \gamma_0 + \gamma_1 t_i + w_i, \quad w_i = \phi_1 w_{i-1} + e_{3i}, \quad (3.4)$$

여기에서 t_i 는 단위 시간에 대한 시간지표 변수이고 $e_{1i} \stackrel{ind}{\sim} N(0, \tau_1^2)$, $e_{2i} \stackrel{ind}{\sim} N(0, \tau_2^2)$ 그리고 $e_{3i} \stackrel{ind}{\sim} N(0, \tau_3^2)$ 이다. 참고로 시간에 따른 경향성이 보이지 않는 모수에 대해서는 이러한 랜덤효과 모형을 생략할 수 있고 전체자료를 통해 추정된 최우추정량(MLE)으로 고정하기도 한다. 형상모수의 경우에는 척도모수처럼 로그변환 모형을 고려할 수 있다. 따라서 본 연구에서 적용되는 모형은 다음과 같은 계층적 베이지안 모형으로 정리할 수 있다.

- Data model : $[z_{ij} | \mu_i, \sigma_i, \xi_i]$ for $i = 1, \dots, T$ and $j = 1, \dots, n_i$

$$z_{ij} | \mu_i, \sigma_i, \xi_i \stackrel{ind}{\sim} \text{GEV}(\mu_i, \sigma_i, \xi_i).$$

- Process model

- $[\mu_i | \mu_{i-1}, \alpha_0, \alpha_1, \phi_1, \tau_1^2] : \mu_i \sim N(\alpha_0 + \alpha_1 t_i + \phi_1 [\mu_{i-1} - \alpha_0 - \alpha_1 t_{i-1}], \tau_1^2).$
- $[\sigma_i | \sigma_{i-1}, \beta_0, \beta_1, \phi_2, \tau_2^2] : \log(\sigma_i) \sim N(\beta_0 + \beta_1 t_i + \phi_2 [\log(\sigma_{i-1}) - \beta_0 - \beta_1 t_{i-1}], \tau_2^2).$
- $[\xi_i | \xi_{i-1}, \gamma_0, \gamma_1, \phi_3, \tau_3^2] : \xi_i \sim N(\gamma_0 + \gamma_1 t_i + \phi_3 [\xi_{i-1} - \gamma_0 - \gamma_1 t_{i-1}], \tau_3^2).$

- Prior model

$$[\alpha_0], [\alpha_1], [\phi_1], [\tau_1^2], [\beta_0], [\beta_1], [\phi_2], [\tau_2^2], [\gamma_0], [\gamma_1], [\phi_3], [\tau_3^2].$$

일반적으로 선형모형 회귀계수인 $(\alpha_0, \alpha_1, \beta_0, \beta_1, \gamma_0, \gamma_1)$ 에 대해서는 서로 독립인 무정보 사전분포를 사용될 수 있고, 자기상관계수인 (ϕ_1, ϕ_2, ϕ_3) 에 대해서는 균일분포를 사용할 수 있다. 그리고 모형의 분산인 $(\tau_1^2, \tau_2^2, \tau_3^2)$ 에 대해서는 각각 독립인 제프리 사전분포를 고려할 수 있다.

4. 국내 강우자료에 기초한 모형 적합

강수 예측에 있어 주된 관심사는 정상적인 상황에서의 강수량이 아니라 극단적인 상황 혹은 과거에 발생한 적 없는 큰 강수량을 나타내는 경우이다. 강수 극치 사상에 있어 미래 강수 예측에 과소 추정이 발생한다면 이례적인 자연재해를 통한 큰 사회적 손실을 가져 오는 반면, 과대 추정이 발생할 경우 예산의 과다로 인한 사회 효율성의 감소를 가져온다. 우리나라의 연평균 총강수량은 500~1500mm 정도이고 대부분의 지역이 800~1000mm가 된다. 이것은 같은 온대에 있는 일본에 비하면 1/2이고 중국 북부에 비해서는 약 2배에 달한다. 대체로 남에서 북으로 갈수록 강수량은 감소하나 풍향, 해안과의 거리, 지형의 영향을 크게 받아서 지역에 따른 강수량의 차이는 많다. 연강수량이 가장 많은 지역은 남해안 지방으로 1500mm에 달하는데, 특히 제주도 남동해안이 1800mm 내외로 가장 많고 울릉도는 1400~1500mm의 강수량을 갖는다. 같은 제주도에서도 지형에 따라 큰 차이 있고 육지에서도 비슷한 경향을 보인다. 본 논문에서는 Korea Meteorological Administration(KMA)로부터 얻은 1973년 1월 1일부터 2011년 12월 31일까지로 39년동안 우리나라 여러 지역에서 측정된 일별 강수량 관측자료를 사용하여 제안된 모형을 적합하였다. 특히 앞에서 언급한 것처럼 우리나라 대표적 강우지역인 울릉도, 전

남 해남, 경남 거제 그리고 대구 지역에 대한 분석에 초점을 맞추었다. 여기서는 복잡한 데이터구조와 계산을 고려하여 1년을 365일로 적용하였고 윤달의 경우 2월 29일을 제외하였다.

KMA로부터 얻은 39년 동안의 각 블록의 최대값(block maxima)인 연 최고강수량을 이용하여 극단값 분포인 GEV 분포를 추정할 수 있는데 이 때 주어진 39개의 자료는 주어진 단위기간동안의 모수값의 변화를 추정하기에는 충분하지 않다. 각 단위기간 동안의 GEV 분포의 모수를 효과적으로 추정하기 위해서는 적어도 20개 정도의 극단값 자료가 필요하고 또한 시간적 변화를 보기 위해서는 여러 개의 단위기간이 필요하다. 따라서 본 연구에서는 이러한 문제점을 극복하기 위해서 다음과 같은 이동창문 개념을 이용한 pseudo 자료를 생성한다. 먼저 $y_i, i = 1, \dots, N$ 를 i 년도 최대 강수량이라고 하자. 이 때 단위기간의 크기를 K 라고 하면,

$$\begin{aligned}(z_{11}, \dots, z_{1K}) &= (y_1, \dots, y_K) \\ (z_{21}, \dots, z_{2K}) &= (y_{1+c}, \dots, y_{K+c}) \\ &\vdots \\ (z_{M1}, \dots, z_{MK}) &= (y_{1+N-K}, \dots, y_N),\end{aligned}$$

여기에서 $M = (N - K)/c - 1$ 이고 c 는 자료의 겹침의 정도를 결정하는 lag에 해당하는데 전체의 수를 고려하여 적당한 값을 정하여 사용한다. 즉, 잔체자료의 수가 많은 경우에는 큰 c 를 사용하여 겹치는 부분을 작게하면 되고 전체자료의 수가 작을 경우에는 상대적으로 작은 c 를 사용하여 겹치는 부분을 크게하면 된다. 단, 이렇게 자료를 중복하여 사용하는 경우 시간에 따른 모수값을 smoothing하는 효과가 있는 반면 이 값들이 서로 독립이 아니므로 모수에 대하여 시간에 대한 경향성 이외에 자기상관관계를 모형에 고려하여야 한다. 본 논문에서는 단순 자기상관 모형인 AR(1)을 고려하였지만 자료에 따라서 AR(p)와 같은 일반적인 모형을 고려할 수 있다.

복잡한 다층구조에서의 베이지안 모수추정은 관측값 \mathbf{z} 가 주어졌을 때, 관련 모수 (θ_1, θ_2) 에 대한 사후분포는 다음과 같이 구성된다.

$$\pi(\theta_1, \theta_2 | \mathbf{z}) \propto p(\mathbf{z} | \theta_1) \pi(\theta_1 | \theta_2) \pi(\theta_2),$$

여기에서 $p(\mathbf{z} | \theta_1)$ 는 data model로 부터의 극단 관측값에 대한 확률분포함수이고 $\pi(\theta_1 | \theta_2)$ 는 process model로 부터의 모수에 관한 모형이며 마지막으로 $\pi(\theta_2)$ 는 prior model로 부터의 그 밖의 모수에 대한 사전분포이다. 즉, 극단값 분포인 GEV 분포함수는 $p(\mathbf{z} | \theta_1)$ 에 해당하며 θ_1 는 극단값 분포의 위치모수 μ , 척도모수 σ 그리고 형상모수 ξ 가 된다. (4)–(6)과 같은 극단값 분포의 모수들에 대한 시간모형은 $\pi(\theta_1 | \theta_2)$ 단계에 해당한다. 마지막으로 $\pi(\theta_2)$ 는 시간모형에 사용된 모수에 대한 사전분포가 된다. 본 논문에서 제안한 모형의 관심모수에 대한 결합사후확률분포를 구하는 것은 어려우므로 각 변수의 조건부확률분포로부터 랜덤표본을 반복적으로 생성하면 적절한 조건 하에서 이들의 극한분포가 결합 확률밀도함수가 된다는 사실에 근거하여 난수 추출이 용이한 깃스 표집기를 형성한다. 깃스 표집기를 가동하기 위한 각 확률 변수의 조건부 밀도함수를 구할 수 있는데 깃스 표집기를 위한 대부분의 조건부 분포가 잘 알려진 분포족에 속하지않는 경우가 많아서 정규화 상수의 계산에도 무리가 따른다. 이 경우 깃스 표집기 내부에 정규화 상수 없이도 목표함수를 따르는 난수 추출이 가능한 메트로폴리스-헤스팅스 알고리즘(Metropolis-Hastings algorithm)을 삽입하여 모수의 추정을 시도한다. 이러한 방법에 기초하여 여러 관심모수에 대한 사후분포를 추정할 수 있다. 난수 추출을 위한 깃스 표집기의 효율성을 확인하기 위해 병렬적 구조를 갖춘 시뮬레이션 체인을 형성 하였으며, 겔만-루빈통계량(Gelman-Rubin statistics, G-R statistics) 및 메트로폴리스-헤스팅스 알고리즘의 기각률을 확인하였다. 관심모수는 매 시간단위

에서 독립이며, 각 단위시간별 관심모수를 1차원적 모수로 간주하여 메트로폴리스-해스팅스 알고리즘을 통한 추출이 가능하다. 본 논문의 메트로 폴리스-해스팅스 알고리즘이 가지는 기각률은 전체적으로 30-50%로 나타나는 것으로 확인할 수 있다.

각 지역의 강우량 자료를 위한 계층적 베이زي안 모형에 대한 결합사후분포는 다음과 같다.

$$\begin{aligned}
& \pi(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\xi} | \{z_{ij}\}_{i=1, \dots, M; j=1, \dots, K}) \\
& \propto \prod_{i=1}^M \prod_{j=1}^K \left[1 + \xi_i \left(\frac{z_{ij} - \mu_i}{\sigma_i} \right) \right]^{-\frac{1}{\xi_i}} \\
& \times \tau_1^{-M} \exp \left\{ -\frac{1}{2\tau_1^2} \sum_{i=1}^M (\mu_i - (\alpha_0 + \alpha_1 t_i + \phi_1 [\mu_{i-1} - \alpha_0 - \alpha_1 t_{i-1}]))^2 \right\} \\
& \times \prod_{i=1}^M (\log(\sigma_i) \tau_2^2)^{-1} \exp \left\{ -\frac{1}{2\tau_2^2} (\log(\sigma_i) - \beta_0 + \beta_1 t_i + \phi_2 [\log(\sigma_{i-1}) - \beta_0 - \beta_1 t_{i-1}])^2 \right\} \\
& \times \tau_3^{-M} \exp \left\{ -\frac{1}{2\tau_3^2} \sum_{i=1}^M (\xi_i - (\gamma_0 + \gamma_1 t_i + \phi_3 [\xi_{i-1} - \gamma_0 - \gamma_1 t_{i-1}]))^2 \right\}.
\end{aligned} \tag{4.1}$$

위의 결합사후분포에 기초한 각 모수의 조건부사후분포는 다음과 같이 구해진다.

- μ_i 의 조건부사후분포 ($i = 1, \dots, M$)

$$\begin{aligned}
\mu_i | \text{rest} & \propto \prod_{j=1}^K \left[1 + \xi_i \left(\frac{z_{ij} - \mu_i}{\sigma_i} \right) \right]^{-\frac{1}{\xi_i}} \\
& \times \exp \left\{ -\frac{1}{2\tau_1^2} (\mu_i - (\alpha_0 + \alpha_1 t_i + \phi_1 [\mu_{i-1} - \alpha_0 - \alpha_1 t_{i-1}]))^2 \right\} \\
& \times \exp \left\{ -\frac{1}{2\tau_1^2} (\mu_{i+1} - (\alpha_0 + \alpha_1 t_{i+1} + \phi_1 [\mu_i - \alpha_0 - \alpha_1 t_i]))^2 \right\}.
\end{aligned}$$

- σ_i 의 조건부사후분포 ($i = 1, \dots, M$)

$$\begin{aligned}
\sigma_i | \text{rest} & \propto \prod_{j=1}^K \left[1 + \xi_i \left(\frac{z_{ij} - \mu_i}{\sigma_i} \right) \right]^{-\frac{1}{\xi_i}} \\
& \times (\log(\sigma_i))^{-1} \exp \left\{ -\frac{1}{2\tau_2^2} (\log(\sigma_i) - \beta_0 + \beta_1 t_i + \phi_2 [\log(\sigma_{i-1}) - \beta_0 - \beta_1 t_{i-1}])^2 \right\} \\
& \times \exp \left\{ -\frac{1}{2\tau_2^2} (\log(\sigma_{i+1}) - \beta_0 + \beta_1 t_{i+1} + \phi_2 [\log(\sigma_i) - \beta_0 - \beta_1 t_i])^2 \right\}.
\end{aligned}$$

- ξ_i 의 조건부사후분포 ($i = 1, \dots, M$)

$$\begin{aligned}
\mu_i | \text{rest} & \propto \prod_{j=1}^K \left[1 + \xi_i \left(\frac{z_{ij} - \mu_i}{\sigma_i} \right) \right]^{-\frac{1}{\xi_i}} \\
& \times \exp \left\{ -\frac{1}{2\tau_3^2} \sum_{i=1}^M (\xi_i - (\gamma_0 + \gamma_1 t_i + \phi_3 [\xi_{i-1} - \gamma_0 - \gamma_1 t_{i-1}]))^2 \right\} \\
& \times \exp \left\{ -\frac{1}{2\tau_3^2} \sum_{i=1}^M (\xi_{i+1} - (\gamma_0 + \gamma_1 t_{i+1} + \phi_3 [\xi_i - \gamma_0 - \gamma_1 t_i]))^2 \right\}.
\end{aligned}$$

- (α_0, α_1) 의 조건부사후분포

$$\alpha_0, \alpha_1 | \text{rest} \propto \exp \left\{ -\frac{1}{2\tau_1^2} \sum_{i=1}^M (\mu_i - (\alpha_0 + \alpha_1 t_i + \phi_1 [\mu_{i-1} - \alpha_0 - \alpha_1 t_{i-1}]))^2 \right\}.$$

- (β_0, β_1) 의 조건부사후분포

$$\beta_0, \beta_1 | \text{rest} \propto \exp \left\{ -\frac{1}{2\tau_2^2} \sum_{i=1}^M (\log(\sigma_i) - \beta_0 + \beta_1 t_i + \phi_2 [\log(\sigma_{i-1}) - \beta_0 - \beta_1 t_{i-1}]))^2 \right\}.$$

- (γ_0, γ_1) 의 조건부사후분포

$$\gamma_0, \gamma_1 | \text{rest} \propto \exp \left\{ -\frac{1}{2\tau_3^2} \sum_{i=1}^M (\xi_i - (\gamma_0 + \gamma_1 t_i + \phi_3 [\xi_{i-1} - \gamma_0 - \gamma_1 t_{i-1}]))^2 \right\}.$$

- τ_1 의 조건부사후분포

$$\tau_1 | \text{rest} \propto \tau_1^{-M-1} \exp \left\{ -\frac{1}{2\tau_1^2} \sum_{i=1}^M (\mu_i - (\alpha_0 + \alpha_1 t_i + \phi_1 [\mu_{i-1} - \alpha_0 - \alpha_1 t_{i-1}]))^2 \right\}.$$

- τ_2 의 조건부사후분포

$$\tau_2 | \text{rest} \propto \tau_2^{-M-1} \exp \left\{ -\frac{1}{2\tau_2^2} \sum_{i=1}^M (\log(\sigma_i) - \beta_0 + \beta_1 t_i + \phi_2 [\log(\sigma_{i-1}) - \beta_0 - \beta_1 t_{i-1}]))^2 \right\}.$$

- τ_3 의 조건부사후분포

$$\tau_3 | \text{rest} \propto \tau_3^{-M-1} \exp \left\{ -\frac{1}{2\tau_3^2} \sum_{i=1}^M (\xi_i - (\gamma_0 + \gamma_1 t_i + \phi_3 [\xi_{i-1} - \gamma_0 - \gamma_1 t_{i-1}]))^2 \right\}.$$

- ϕ_1 의 조건부사후분포

$$\phi_1 | \text{rest} \propto \exp \left\{ -\frac{1}{2\tau_1^2} \sum_{i=1}^M (\mu_i - (\alpha_0 + \alpha_1 t_i + \phi_1 [\mu_{i-1} - \alpha_0 - \alpha_1 t_{i-1}]))^2 \right\}.$$

- ϕ_2 의 조건부사후분포

$$\phi_2 | \text{rest} \propto \exp \left\{ -\frac{1}{2\tau_2^2} \sum_{i=1}^M (\log(\sigma_i) - \beta_0 + \beta_1 t_i + \phi_2 [\log(\sigma_{i-1}) - \beta_0 - \beta_1 t_{i-1}]))^2 \right\}.$$

- ϕ_3 의 조건부사후분포

$$\phi_3 | \text{rest} \propto \exp \left\{ -\frac{1}{2\tau_3^2} \sum_{i=1}^M (\xi_i - (\gamma_0 + \gamma_1 t_i + \phi_3 [\xi_{i-1} - \gamma_0 - \gamma_1 t_{i-1}]))^2 \right\}.$$

여기에서 모든 관심모수의 조건부사후분포가 일반적으로 잘 알려진 분포족에 속하지 않음을 알 수 있다. 메트로폴리스-헤스팅스 알고리즘을 위한 일반적인 후보 생성 밀도 함수의 선택은 $q(y|x) = q_1(|y-x|)$ 로 주어지는 확률 보행(random walk) 연쇄이며, 본 연구의 경우 목표함수가 가지는 경험적 밀도함수

Table 4.1. Posterior means and standard errors of all model parameters at Ulleung-do, Geoje, Haenam and Daegu.

Coefficient	울릉도		거제		해남		대구	
	Est.	S.E	Est.	S.E	Est.	S.E	Est.	S.E
α_0	66.586	1.389	100.744	1.803	97.692	3.025	69.567	1.666
α_1	2.187	0.224	1.642	0.291	2.045	0.488	1.965	0.269
β_0	2.887	0.062	3.662	0.069	3.300	0.063	3.351	0.054
β_1	0.033	0.010	-0.039	0.011	0.045	0.010	0.032	0.008
γ_0	0.386	0.066	0.333	0.065	0.724	0.089	0.244	0.021
γ_1	-0.040	0.011	-0.022	0.010	-0.066	0.014	-0.034	0.003
ϕ_1	0.409	0.339	-0.340	0.329	0.089	0.351	-0.367	0.266
ϕ_2	0.078	0.307	0.604	0.320	0.345	0.341	0.076	0.304
ϕ_3	0.184	0.316	10.534	0.389	0.269	0.326	-0.131	0.286
τ_1	1.865	0.869	2.422	0.978	4.241	1.893	1.812	0.849
τ_2	0.077	0.003	0.085	0.004	0.087	0.004	0.065	0.003
τ_3	0.086	0.004	0.085	0.004	0.120	0.007	0.025	0.001

와 유사한 형태의 절단정규분포를 후보생성 밀도함수로 설정한다.

$$q_1(y | x) = \frac{\phi(y|x, \sigma^2)}{1 - \Phi(0|x, \sigma^2)}, \tag{4.2}$$

여기서 $\phi(y|x, \sigma^2) = (1/\sqrt{2\pi\sigma^2}) \exp\{-(1/2\sigma^2)(y - x)^2\}$ 이며 $\Phi(0|x, \sigma^2) = \int_0^\infty \phi(y|x, \sigma^2) dy$ 이다. 목표 함수가 가지는 확률변수의 범위는 0보다 크거나 같은 구간임을 감안하여 절단 정규 분포의 하한 값을 0으로 결정한다. 이때 분산 σ^2 을 후보 생성 밀도 함수의 조율 모수(tuning parameter)라 정의 하며, 조율 모수의 조절을 통해 적절한 채택-기각률을 가지는 효율적인 메트로폴리스-헤스팅스 알고리즘을 형성한다.

표본 추출을 위한 깃스 표집기의 효율성과 체인의 수렴도를 확인하기 위해 병렬적 구조를 갖춘 시뮬레이션 체인을 형성 하였으며, 메트로폴리스-헤스팅스 알고리즘의 기각률과 겔만-루빈통계량(Gelman-Rubin statistics 또는 G-R statistics)을 각각 확인하였다. 관심모수는 매 시간단위에서 독립이며, 각 단위시간별 관심모수를 1차원적 모수로 간주하여 메트로폴리스-헤스팅스 알고리즘을 통한 추출이 가능하다. 본 논문의 메트로폴리스-헤스팅스 알고리즘이 가지는 기각률은 전체적으로 30-50%로 나타나고 겔만-루빈통계량은 1.03-1.37의 값을 가지는 것으로 확인할 수 있었다. 본 연구에서는 3개의 병렬 체인을 이용하여 체인의 수렴 여부를 확인하였고 30,000개의 표본을 생성한 후, 초기값의 영향을 제거하기 위하여 처음의 15,000개의 표본을 burn-in 과정으로 제거하였다.

Table 4.1에서는 본 논문에서 제안한 모형을 우리나라의 대표 강우지역(울릉도, 남해안 거제, 남해안 해남, 대구)의 강우자료에 적합한 후, 얻은 관련 모수의 사후분포 평균과 표준오차를 보여준다. 위치모수의 경우 모든 지역에서 시간에 따라 증가하는 경향을 보여주었고 척도모수의 경우에도 거제를 제외한 모든 지역에서 시간에 따라 증가하는 경향을 보여주었다. 하지만 형상모수의 경우에는 모든 지역에서 시간에 따라 감소하는 경향을 보여주었다. 이러한 경향을 네 지역에 대한 각 단위시간의 GEV 분포의 위치모수(왼쪽), 척도모수(중앙) 그리고 형상모수(오른쪽)의 추정값을 통해서 확인할 수 있다 (Figure 4.1).

이제 추정된 극단값 분포의 모수들에 대한 사후분포를 이용하여 반환기간에 대한 사후분포를 알아보고자 한다. 정확한 관심모수의 사후분포가 알려지지 않았으므로 반환기간의 사후분포도 몬테카를로 방법을 이용하여 구할 수 있다. 즉, 관심모수의 사후분포로부터 추출된 깃스 표본을 이용하여 반환기간을

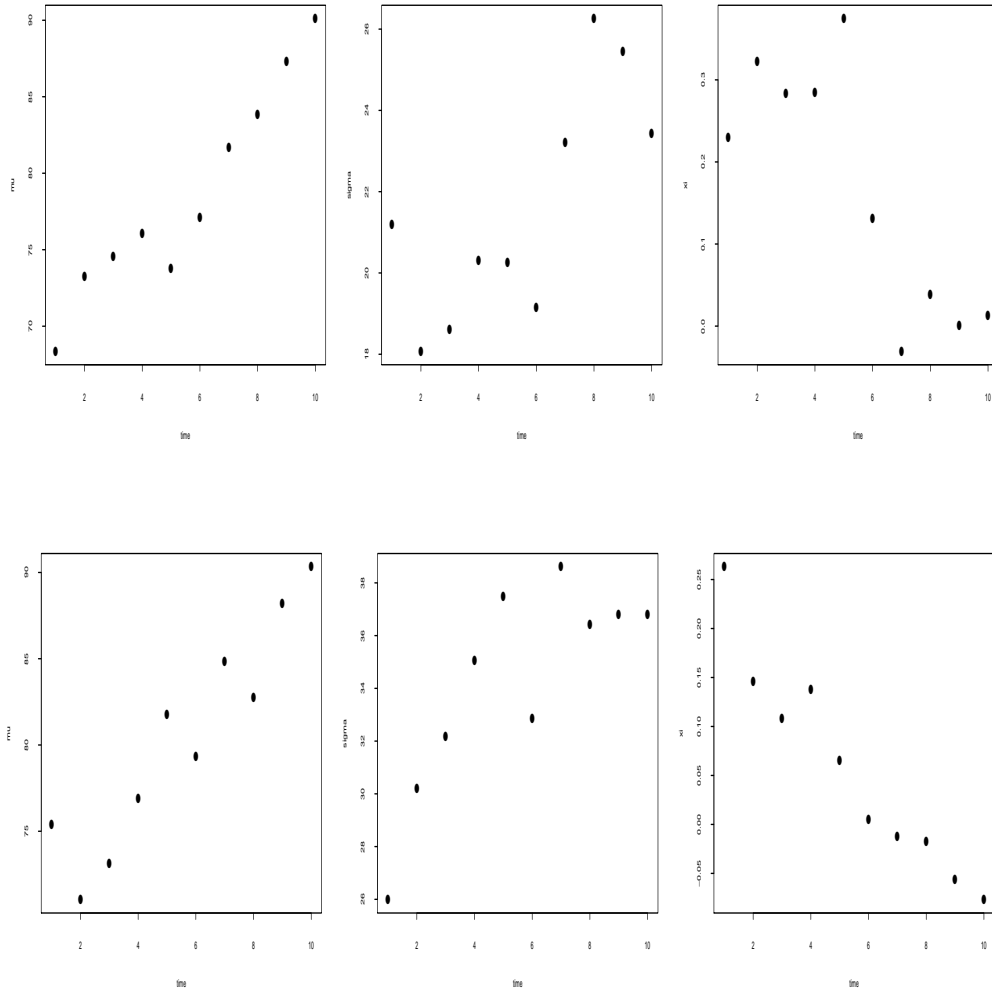


Figure 4.1. Estimated location (left), scale (middle) and shape (right) parameters of GEV distributions for each time interval at Ulleung-do (left) and Daegu (right).

구성하고 이렇게 얻어진 반환기간값들을 이용하여 반환기간에 대한 사후분포를 얻을 수 있다. Figure 4.2에서는 이렇게 얻어진 GEV 분포에 기초한 반환기간의 사후분포의 평균의 시간적 경향성을 확인할 수 있다. 지역에 따라 정도의 차이는 있지만 전체적으로 반환기간의 값이 감소하는 경향을 확인할 수 있었다. 이렇게 얻어진 시간정보를 이용하여 미래의 반환기간에 대한 시간적 예측도 가능하다.

5. 결론

본 논문에서 대표적 극단값 분포인 GEV 분포와 이를 이용한 시간적구조를 가지는 반환기간에 대한 계층적 베이저안 모형을 소개하였고 제안된 모형을 우리나라 대표적인 강우지역인 4개의 지역에서 39년간 측정된 일별 강우량 관측자료에 적용하였다. 효과적으로 관심모수값인 반환기간에 공간모형을 구성

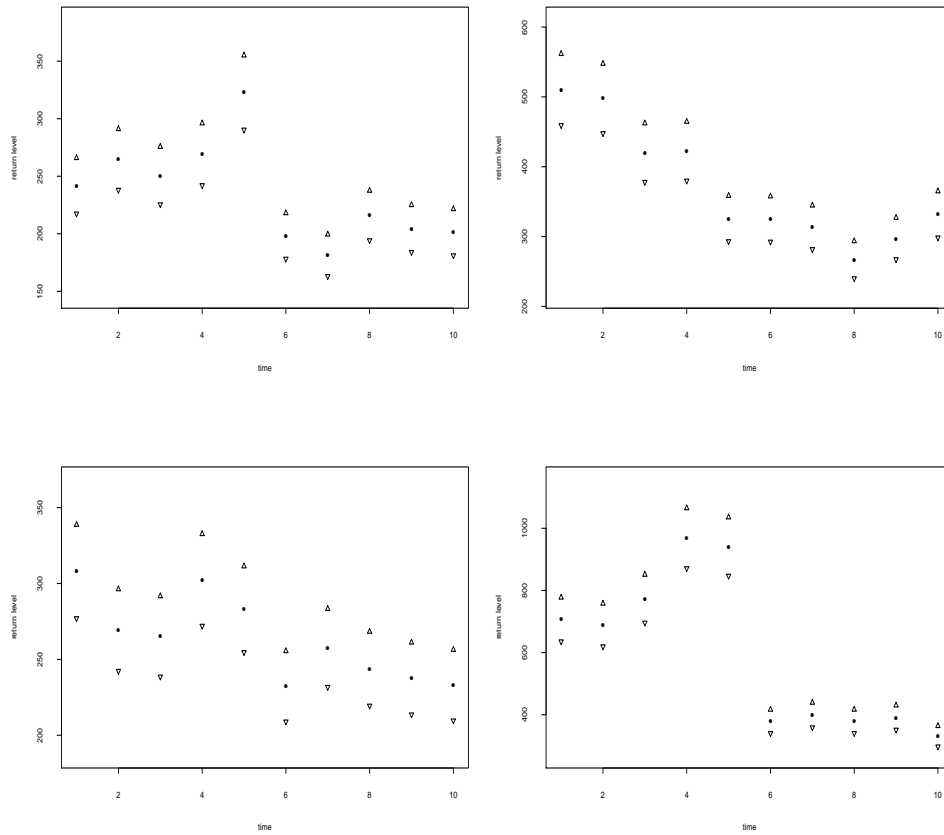


Figure 4.2. Posterior means of the return levels at Ulleung-do, Geoje, Haenam and Daegu (clockwise from the left).

하기 위해서 계층적 구조를 이용하여 각 극단값 분포의 모수들에 시간구조를 가지는 모형을 가정하였다. 이를 통해서 간접적으로 반환기간에 시간적 관련성을 설명하였다. 모수 추정 결과 시간에 따라 추정된 관심모수가 선형적 경향을 가지고 또한 지역에 따라 다른 경향을 보일 수 있음을 알 수 있었다. 사후평균의 측면에서는 모형을 적용한 네 지역에서는 반환기간이 줄어드는 경향을 확인하였다. 하지만 전국적인 경향성을 보기 위해서는 다른 지역에 대한 추가적인 연구는 필요한 것으로 보인다. 일반적으로 실제 모수의 변화에 따른 극치 강수량의 경우 형상모수보다 척도모수의 영향을 많이 받지만 추론의 측면에서는 형상모수에 더 민감하게 반응하는 것을 알 수 있다. 추가적으로 제안된 모형에 기초한 반환기간에 대한 효과적인 미래예측에 관한 연구도 가능할 것으로 기대된다.

References

Casson, E. and Coles, S. (1999). Spatial regression models for extremes, *Extremes*, 1, 449–468.
 Coles, S. G. (2001). *An Introduction to Statistical Modeling of Extreme Values*, Springer-Verlag, London.
 Coles, S., Heffernan, J. and Tawn, J. (1999). Dependence measures for extreme value analysis, *Extremes*, 2, 339–365.

- Cooley, D., Naveau, P., Jomelli, V., Rabatel, A. and Grancher, D. (2006). A Bayesian hierarchical extreme value model for lichenometry, *Environmetrics*, **17**, 555–574.
- Cooley, D., Nychka, D. and Naveau, P. (2007). Bayesian spatial modeling of extreme precipitation return levels, *Journal of the American Statistical Association*, **102**, 824–840.
- Cunderlik, J. M., Burn, D. H., Posbberg, D., Robinson, B. A. and Zyzolowski, G. A. (2008). Generalized likelihood uncertainty estimation (GLUE) using adaptive Markov Chain, *Journal of Hydrology*, **276**, 210–223.
- de Haan, L. (1985). *Extremes in Higher Dimensions: The Model and Some Statistics*, In Proceedings of the 45th Session of the International Statistical Institute.
- Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*, Springer-Verlag, Berlin.
- Heffernan, J. E. and Tawn, J. A. (2004). A conditional approach for multivariate extreme values, *Journal of the Royal Statistical Society, Series B*, **66**, 497–546.
- Jang, S. W., Seo, L., Kim, T. W. and Ahn, J. H. (2011). Non-stationary rainfall frequency analysis based on residual analysis, *Journal of the Korean Society of Civil Engineers*, **31**, 449–457.
- Lee, J. J. (2010). *Assessment of Non-Stationarity in Precipitation and Development of Nonstationary Frequency Analysis*, Ph.D Thesis, Chonbuk National University.
- REVSTAT - Statistical Journal*, **4**, 81–109.
- Schlather, M. and Tawn, J. (2003). A dependence measure for multivariate and spatial extreme values: Properties and inference, *Biometrika*, **90**, 139–156.
- Strupczewski, W. G., Singh, V. P. and Mitosek, H. T. (2001). Non-stationary approach to at-site flood frequency modeling III: Flood analysis of Polish rivers, *Journal of Hydrology*, **248**, 152–167.

한국지역 집중호우에 대한 반환주기의 베이지안 모형 분석

김용구^{a,1}

^a경북대학교 통계학과

(2015년 2월 10일 접수, 2015년 3월 12일 수정, 2015년 3월 12일 채택)

요약

본 논문에서는 비정상 극치 강수 자료에 대해 계층적 베이지안 모형을 적용하여 시간에 따른 모수의 변화를 추정하며, 미래 확률 강수량에 대한 극단값 분포를 예측하고 더 나아가 반환기간에 대한 경향과 예측 값을 얻고자 한다. 이전의 고전적 통계 방법을 통한 강수 자료의 모수 추정연구의 경우, 자료의 정상성 가정 하에 고정된 모수를 추정하는 방법으로, 최근 나타난 비정상 강수 사상과 같이 강수량이 가지는 분포의 모수적 변화가 예상되는 경우 해석상 문제가 발생한다. 이러한 문제점을 해결하기 위해 모형의 관심모수에 시간에 따른 자기 상관 선형 회귀 함수를 적합한 계층적 베이지안 모형을 고려한다. 제안된 모형의 효율성을 확인하기 위해서 1973년부터 2011년까지 39년 동안의 우리나라 여러지역의 기상 관측소에서 관측된 일일 강수량 자료가 사용하여 대표적인 극단값 분포인 Generalized Extreme Value(GEV) 분포에 적합시키고, 계층적 베이지안 모형을 이용하여 이들 분포의 모수들에 자기상관 시간 모형을 소개한 후 우리나라 여러지역에 대한 반환기간에 대한 시간에 따른 경향을 확인하였다.

주요용어: 베이지안 분석, 강우모형, 극단값, generalized extreme value 분포, 반환주기, 시간모형.

이 논문은 2014년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (No. 2014R1A1A1007075).

¹(702-701) 대구광역시 북구 대학로 80, 경북대학교 통계학과. E-mail: kim.1252@knu.ac.kr