

극치수문자료의 경향성 분석 개념 및 비정상성 빈도해석

Concept of Trend Analysis of Hydrologic Extreme Variables and Nonstationary Frequency Analysis

이정주*, 권현한**, 김태웅***

Jeong-Ju Lee, Hyun-Han Kwon, Tae-Woong Kim

요 지

최근 기상변동성 증가 및 기후변화 영향으로 수문순환과정이 과거와는 다른 양상으로 전개되고 있으며 전반적으로 극치사상의 빈도 및 강도의 증가현상이 지배적이다. 이러한 영향을 정량적으로 검토하기 위해서 경향성분석 방법 등이 도입되어 극치수문사상의 변동경향을 평가하는데 이용되고 있다. 대표적인 방법으로 선형회귀분석, Mann-Kendall 경향성 분석 등이 있으나 기본적인 가정(assumption)의 제약으로 극치수문자료 계열의 특성을 효과적으로 분석하는데 무리가 있다. 대표적이고 일반적으로 적용되는 선형회귀분석의 경우 자료가 정규분포(normal distribution)의 특성을 가질 때 유효한 방법으로서 극치수문자료와 같이 Heavy Tail를 가지는 분포특성을 표현하는데는 무리가 따른다. 이밖에도 기존 선형회귀분석을 극치수문자료에 적용할 경우 추정된 결과를 수자원설계의 관심사항인 빈도해석 등에 직접적으로 연계시켜 해석할 수 없는 단점이 있다. 이는 자료계열의 분포특성을 정규분포로 가정하기 때문에 발생하는 문제로서 극치수문자료계열의 분포특성을 반영할 수 있는 방법론의 개발이 필요하다. 본 연구에서는 이러한 점을 개선하기 위해서 극치분포(extreme distribution)를 선형회귀분석에 적용하는 비정상성빈도해석(nonstationary frequency analysis) 방법론의 개념을 제시하고자 한다. 비정상성빈도해석을 위해서 Bayesian 기법이 도입되며 Bayesian 기법의 특성상 관련변수들이 사후분포(posterior distribution)로 귀결되기 때문에 경향성에 대한 정량적이고 확률적인 분석이 가능한 장점이 있다. 본 연구를 통해 개발된 방법론은 국내외 주요 강수지점에 대해서 적용되며 경향성, 분포특성, 빈도별 강수량에 대한 체계적인 분석이 이루어진다.

핵심용어 : 극치수문량, 경향성, 비정상성, Bayesian 해석

1. 서 론

일반적으로 경향성을 파악하는데 많이 이용되는 선형회귀분석에서 변수의 선택문제는 최적의 모형을 찾는 데 있어서 중요한 부분을 차지한다. 변수 x 와 y 간에 직선회귀모형(straight line regression model)을 적합시킬 경우에는 일반적으로 몇 가지 가정이 전제조건을 이룬다(박성현, 1997). 그중에서 극치수문자료의 경향성(trend) 해석과 관련하여 정규분포 가정과 등분산

* 정회원 · 전북대학교 토목공학과 박사수료 · E-mail : julee@jbnu.ac.kr
** 정회원 · 전북대학교 토목공학과 조교수 · E-mail : hkwon@jbnu.ac.kr
*** 정회원 · 한양대학교 건설환경공학과 조교수 · E-mail : twkim72@hanyang.ac.kr

(homoscedastic)의 가정을 유념할 필요가 있다.

Katz 등(2002)은 최우도추정(Maximum likelihood)에 의한 GEV 분포형 적합 과정에서, 수문학적 변수들의 분포는 Heavy Tail을 가지고 있는 것에 대한 여러 사례들을 언급하였다. 유출량의 경우에 있어서 나타나는 Heavy Tail 특성과(Anderson과 Meerschaert, 1998; Farquharson 등, 1992; Madsen과 Rosbjerg, 1997) 더불어 극치강수량(시단위, 일단위)의 경우도 그 분포형이 Heavy Tail을 갖는 것으로 알려져 있다(Egozcue와 Ramis, 2001). x, y 평면에서의 최소제공법 개념으로 단순회귀분석을 통해 극치강수량의 시간변화에 따른 확률분포의 경향성을 분석할 경우 분석자료가 극치분포를 따르게 되기 때문에 정규분포 가정과 등분산의 가정을 충족하지 못하며, 수문학적 빈도해석을 위한 극치분포특성을 충족시키지 못하는 문제가 있다.

본 논문에서는 극치수문자료의 경향성 평가를 위해서 극치분포를 활용한 경향성 평가 방법을 제시하고자 하며 이를 해석하기 위한 방법론으로 Bayesian 모형을 활용하였다.

2. 대상자료 및 연구 방법

2.1 대상자료

본 연구에서는 분석을 위한 극치수문사상을 추출하기 위해 서울지점의 시단위 강우자료를 이용하였다. 24시간 지속 이동시간 최대강우량을 산출하여 연최대강우량시계열(Annual Maxima Rainfall)을 산정하였다. 자료계열의 기간은 1961년부터 2008년까지 48년간이다. 우선 서울지방 극치강수량 시계열의 분포형 적합도를 도해적으로 검토하였다. 정규분포의 경우 이론적인 Quantile 값을 크게 벗어나는 반면, Gumbel 분포의 경우 Upper 및 Lower Tail부분의 거동을 잘 묘사하고 있는 것을 볼 수 있다(Figure 1).

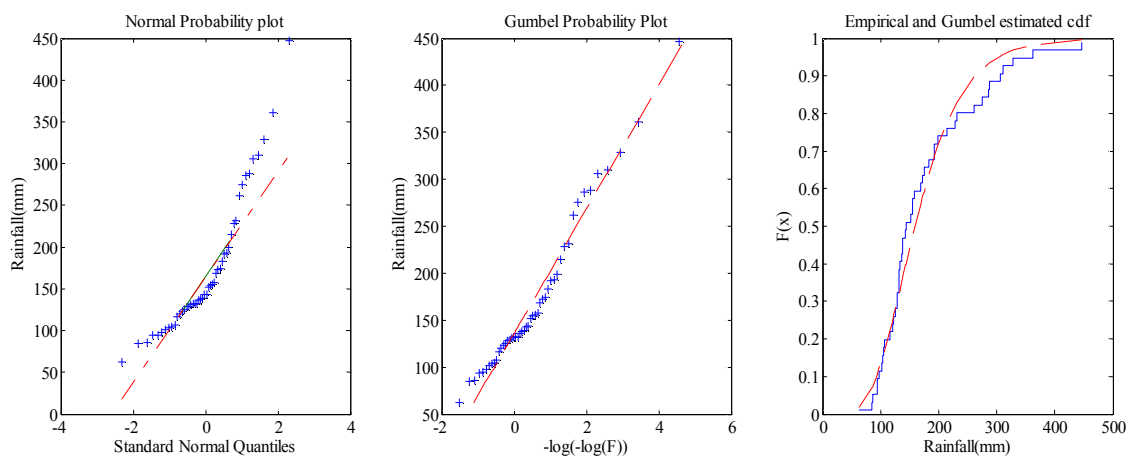


Figure 1. Visual inspection of tail behavior through probability plot. (a) Normal probability plot for Annual Maxima, (b) Gumbel probability plot for Annual Maxima, (c) Empirical CDF versus Theoretical Gumbel CDF

2.2 경향성을 고려한 비정상성 빈도해석 방법

비정상성 빈도해석을 위한 방법으로서 Bayesian모형을 사용하였으며, Markov-Chain Monte

Carlo 모의를 통해 사후분포의 매개변수를 추정하였다. 극치사상에 나타나는 경향성을 처리하기 위한 방법론으로 극치계열 자료 Y 가 앞서 제시한 Gumbel 분포를 따른다고 가정할 때, Gumbel 분포형의 위치매개변수와 규모매개변수는 다음과 같이 시간에 따라 다른 값을 갖도록 구성할 수 있다.

$$Y \sim \text{Gumbel}(\mu(t), \sigma(t)) \quad (1)$$

Bayesian 모형 하에서, 식(1)의 모든 매개변수들은 확률분포를 가지게 된다. 즉, $\mu(t)$ 와 $\sigma(t)$ 가 사전분포를 가지게 되며 각 매개변수는 시간에 함수로서 다음의 식(2)와 식(3)과 같이 선형관계를 갖는다고 가정한다. 여기서, t 는 시간을 의미하며 α 와 β 는 시간과 매개변수 μ 와 σ 의 사전분포와 선형적으로 연계시키는 회귀분석의 매개변수를 나타낸다. 본 연구에서는 시간에 따라 2개의 매개변수가 동시에 변화한다고 가정하여 확률적 추론(statistical inference)를 실시하였다.

$$p(\mu(t)) \sim N(\alpha_0 + \alpha_1 \times t, \sigma_\alpha^2) \quad (2)$$

$$p(\sigma(t)) \sim N(\beta_0 + \beta_1 \times t, \sigma_\beta^2) \quad (3)$$

Bayes 정리에 의해서 매개변수의 사후분포는 다음과 같이 전개될 수 있다.

$$p(\Theta | y) = \frac{p(\Theta, y)}{p(y)} \propto p(y | \Theta) \cdot p(\Theta) \quad (10)$$

여기서 $\Theta = [\mu(t), \sigma(t), \alpha_0, \alpha_1, \beta_0, \beta_1, \sigma_\alpha, \sigma_\beta]$ 는 전체매개변수들의 집합을 나타내며, $p(\Theta)$ 는 매개변수들의 사전분포를 $p(y | \Theta)$ 는 극치자료 y 의 우도함수(likelihood function)를 의미한다.

$$p(y | \Theta) \propto \prod_{t=1}^T \text{Gumbel}(y_t | \mu(t), \sigma(t)) \cdot N(u(t) | \alpha_0 + \alpha_1 t, \sigma_\alpha^2) \cdot N(\sigma(t) | \beta_0 + \beta_1 t, \sigma_\beta^2) \quad (12)$$

식(12)는 모든 매개변수에 대한 적분을 통해 직접적으로 추정하는 것은 불가능하며 본 연구에서는 앞서 언급한 Markov Chain Monte Carlo (MCMC)방법을 도입하여 매개변수들의 사후분포를 추정하게 된다. 본 연구에서는 MCMC 방법 중 깃스표본법을 이용하여 회귀매개변수들을 추정하였으며 모형의 Convergence를 확증하기 위해서 3개의 Chain을 독립적으로 시행하여 Sampling이 효과적으로 혼합(mixing)되도록 하였다. 그림 3은 본 연구에서 제안하는 Bayesian Network 모형에 대한 개념도를 나타낸다.

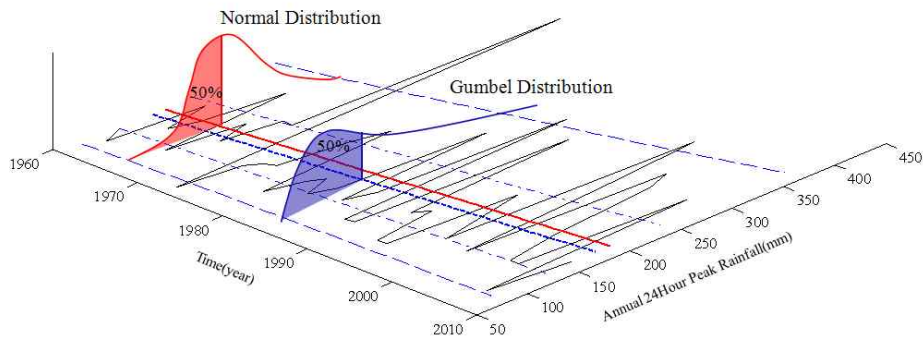


Figure 2. Concept on Bayesian Regression Analysis with Gumbel Distribution

3. 적용 및 결과

자료계열을 정규분포로 가정할 경우와 Gumbel 분포로 가정할 경우의 경향성을 비교한 Figure 3을 통해 중앙값을 중심으로 Upper Tail이 두꺼운 좌우비대칭의 불확실성 구간을 확인할 수 있다.

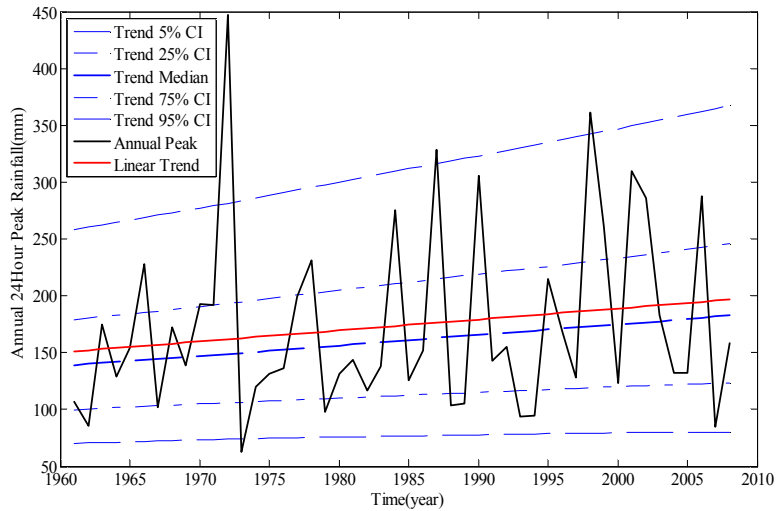


Figure 3. Comparison of trend analysis of extreme rainfall time series between Normal distribution and Gumbel distribution



Figure 4. Model parameters of trend model in nonstationary frequency analysis and associated uncertainties. Here, alpha(0) and beta(0) indicate intercepts while alpha(1) and beta(1) represent slopes in location and scale parameter of Gumbel distribution

Figure 5는 경향성을 고려하여 추정된 100년 빈도 강수량과 기존 정상성을 가정으로 추정된 100년 빈도 강수량을 비교한 그림이며 비정상성 빈도해석으로부터 추정된 100년 빈도 강수량의 불확

실성 구간을 사후분포로부터 추정하여 도시하였다. 경향성이 존재한다고 가정한다면 1981년을 전후로 현재의 100년 빈도 강수량의 빈도가 100년 빈도 이하로 낮아지고 있음을 확인할 수 있다.

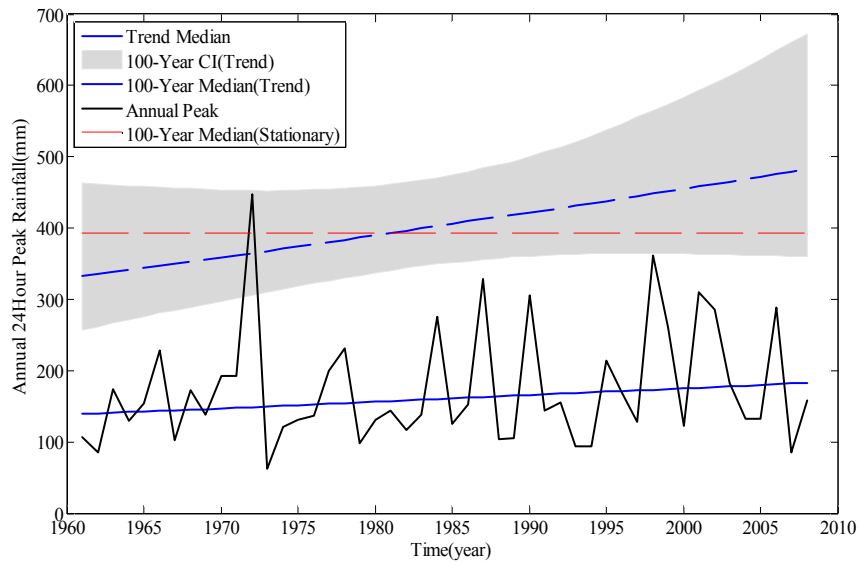


Figure 5. Comparison of trend analysis of extreme rainfall time series between Normal distribution and Gumbel distribution

4. 결 론

본 논문에서는 극치수문자료의 경향성 분석 개념을 소개하고 이를 빈도해석과 연계시켜 해석하는 방법론을 제시하고자 Gumbel 극치분포를 기반으로, 시간변화에 의한 수문빈도 특성 변화를 모의할 수 있는 Bayesian 모형을 구성하였다. 사후분포의 매개변수는 깃스표본법에 의한 Markov Chain Monte Carlo Simulation을 통해 추정하였으며, 이를 통해 경향성을 고려한 확률강우량과 불확실성 구간을 추정하였다.

참 고 문 헌

1. 박성현(1997). 회귀분석, 민영사.
2. Anderson, P. L., and Meerschaert, M. M.(1998). Modeling river flows with heavy tails, *Water Resour Res*, 34(9), 2271-2280.
3. Egozcue, J. J. and Ramis, C.(2001). Bayesian hazard analysis of heavy precipitation in eastern Spain, *International Journal of Climatology*, Vol. 21, No. 10, pp. 1263-1279.
4. Farquharson F. A. K., Meigh J. R., Sutcliffe J. V.(1992). Regional flood frequency analysis in arid and semi-arid areas, *Journal of Hydrology*, Vol. 138, pp. 487 - 501.
5. Katz, R. W., Parlange, M. B., Naveau, P.(2002). Statistics of extremes in hydrology, *Advances in Water Resources*, Vol. 25, No. 8-12, pp. 1287-1304.
6. Madsen H., Rasmussen P. F., Rosbjerg D.(1997). Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events. 1. at-site modeling. *Water Resour Res.*, Vol. 33, pp. 747 - .57.