

극치수문자료의 계절성 분석 개념 및 비정상성 빈도해석을 이용한 확률강수량 해석

Concept of Seasonality Analysis of Hydrologic Extreme Variables and Design Rainfall Estimation Using Nonstationary Frequency Analysis

이 정 주* / 권 현 한** / 황 규 남***

Lee, Jeong-Ju / Kwon, Hyun-Han / Hwang, Kyu-Nam

Abstract

Seasonality of hydrologic extreme variable is a significant element from a water resources management point of view. It is closely related with various fields such as dam operation, flood control, irrigation water management, and so on. Hydrological frequency analysis conjunction with partial duration series rather than block maxima, offers benefits that include data expansion, analysis of seasonality and occurrence. In this study, nonstationary frequency analysis based on the Bayesian model has been suggested which effectively linked with advantage of POT (peaks over threshold) analysis that contains seasonality information. A selected threshold that the value of upper 98% among the 24 hours duration rainfall was applied to extract POT series at Seoul station, and goodness-fit-test of selected GEV distribution has been examined through graphical representation. Seasonal variation of location and scale parameter (μ and σ) of GEV distribution were represented by Fourier series, and the posterior distributions were estimated by Bayesian Markov Chain Monte Carlo simulation. The design rainfall estimated by GEV quantile function and derived posterior distribution for the Fourier coefficients, were illustrated with a wide range of return periods. The nonstationary frequency analysis considering seasonality can reasonably reproduce underlying extreme distribution and simultaneously provide a full annual cycle of the design rainfall as well.

Keywords : hydrologic extreme variable, peaks over threshold, seasonality, nonstationary, bayesian model, fourier analysis

요 지

수문자료의 계절성은 수자원관리의 관점에서 매우 중요한 요소로서 계절성의 변동은 댐의 운영, 홍수조절, 관개용수 관리 등 다양한 분야와 밀접한 관계를 가지고 있다. 수문빈도해석을 위해 POT 자료와 같은 부분기간치계열을 사용함으로써 자료의 확충, 계절성 확보, 발생빈도모형의 구축 등이 가능하다. 본 연구에서는 POT 자료의 장점을 효과적으로 빈도해석에 연계시키는 방법론으로서 POT 자료로부터 계절성을 추출하고 이를 빈도해석과 연계시켜 Bayesian 기법을 기반으로 하는 비정상성 빈도해석 모형을 구축하였다. 서울지점의 관측 자료로부터 98% Threshold를 적용하여 POT 자료

* 전북대학교 공과대학 토목공학과 박사수료

Ph.D. Candidate, Dept. of Civil Engrg., Chonbuk National Univ., Jeonju 561-756, Korea (e-mail: julee@jbnu.ac.kr)

** 교신저자, 전북대학교 공과대학 토목공학과 조교수

Corresponding Author, Assistant Professor, Dept. of Civil Engrg., Chonbuk National Univ., Jeonju 561-756, Korea (e-mail: hkwon@jbnu.ac.kr)

*** 전북대학교 공과대학 토목공학과 교수

Professor, Dept. of Civil Engrg., Chonbuk National Univ., Jeonju 561-756, Korea (e-mail: khwang@jbnu.ac.kr)

를 추출하였으며, GEV 분포에 대한 적합성을 검토하였다. 위치 및 규모매개변수의 계절적 변동성을 Fourier 급수로 표현하고, Bayesian Markov Chain Monte Carlo 모의를 통해 매개변수들의 사후분포를 추정하였으며, 사후분포와 Quantile 함수를 이용하여 재현기간에 따른 확률강수량을 추정하였다. 계절성을 고려한 비정상성빈도분석 결과 7~8월의 비정상성 확률강수량과 기존 정상성빈도분석의 결과가 유사한 값을 나타내고 있으며 동시에 계절성을 반영한 확률강수량의 거동을 효과적으로 모의가 가능하였다.

핵심용어 : 극치수문량, 부분기간치계열, 계절성, 비정상성, Bayesian 해석, Fourier 해석

1. 서 론

수문자료의 계절성은 수자원관리의 관점에서 매우 중요한 요소로서 계절성의 변동은 댐의 운영, 홍수조절, 관개용수 관리 등 다양한 분야와 밀접한 관계를 가지고 있다. 그러나 지금까지의 수문자료의 계절성 평가는 주로 이수 관점에서 이루어지고 있으며 치수 관점에서 극치수문량의 계절성을 평가하는 연구는 미진한 실정이다. 이는 극치수문량을 해석하는 방법론으로서 연최대치계열 (annual maxima) 즉, Block Maxima가 이용됨에 따라 나타나는 문제점이다. 그러나 부분기간치계열 (partial duration series)을 활용하게 되면 자료의 확충뿐만 아니라 자연적으로 극치수문량의 계절성에 대한 평가 또한 가능하다 (Clarke et al., 2009). 이러한 분석과정을 POT 분석이라 하며 일정 Threshold 이상의 자료를 모두 취하여 빈도분석에 이용함으로써 극치분포 Upper Tail의 추가정보를 제공하게 되고, 극치분포의 매개변수 (parameter)와 Quantile을 보다 적절하게 추정할 수 있는 방법이다 (Madsen et al., 1997a; 1997b). 연최대치계열을 이용하는 일반적인 극치빈도분석의 경우는 연최대값이 보통 7월과 8월에만 존재하게 되지만, POT 분석의 경우 여러 달에 걸쳐 빈도분석을 위한 자료가 구성되게 된다. 이를 빈도분석으로 연계시키기 위해서는 계절성을 비정상성으로 고려하여 모형화 할 수 있는 방법론의 개발이 필요하다. 본 연구에서는 이러한 목적을 위해서 계절성을 고려할 수 있는 비정상성 빈도분석 기법의 개념을 제시하고 모형으로 개발하고자 한다.

계절성 및 POT 분석과 관련하여 기존에 수행된 국내의 연구들을 살펴보면, 노재식과 이길춘 (1992, 1993)이 한강유역의 6개 관측점에서의 14년간의 홍수기록 자료를 이용한 지점 빈도분석을 수행하였으며, 부족한 자료보유기간을 보완하기 위해 POT 모형을 적용한 사례가 있으며, POT모형을 이용하여 전국 주요 하천에서의 홍수량 빈도분석을 수행하고, 유역특성인자들과의 상관성 분석을 통하여 유역면적, 유역형상계수, 하천경사, 유역표고차 등에 의한 통계학적 지역화 회귀모형을 개발한 사례가 있다.

유철상 (2002)은 금강유역의 30년 이상 일강우자료를 가진 28개 지점에 대해서 월별 및 연평균 강수량의 계절성과 공간상관성에 따른 면적평균강수량의 추정오차 변화를 평가하였다. 유철상 등 (2002)은 또한 기상학적 가뭄의 빈도분석을 위해 조선시대 측우기 자료를 포함한 약 200년간의 월강수량 자료를 이용하여 Threshold 이하의 강수일의 재현 및 지속특성을 분석하였다. 각 월의 강수량이 정규분포를 따른다고 가정하여 정규화 하였으며, 절단수준 (truncation level)을 -0.25, -0.50, -0.75, -1.00으로 정하여 적용하였고, 정규화된 자료를 기간 (월)별로 이동평균하여 그 거동을 분석하였다. 진영훈 등 (2005)은 1904~2000년까지의 목포지점 강수량 자료에 대해서 이산형 Wavelet 변환을 통해 강수의 장·단기 성분을 추출하였다. Wavelet 변환을 통해 계절성 및 연주기성과 장시간 주기성을 파악하였다. 엄명진 등 (2008)은 극치수문자료의 GPD (generalized Pareto distribution)분포 모형에 적용할 Threshold를 산정함에 있어서 최적화된 값을 찾기 위해 고산, 제주, 서귀포, 성산포지점의 시단위자료를 이용하여 최소무강우기간을 산정하여 강우사상을 분류하고, 지속시간별로 모멘트법, 최우도법, 확률가중모멘트법을 이용하여 매개변수를 추정하였으며, Hall and Welsh (1985), Hall (1990), Drees and Kaufmann (1998) 및 Gomes and Pestan (2007)이 제안한 임계값 결정방법들을 비교하여 적합성을 분석하였다.

국외의 경우 최근 들어 극치수문자료의 경향성 및 변동성과 관련한 POT 분석이 이루어지고 있다. An and Pandey (2005)는 풍속자료의 극치분포를 Standard Gumbel, Modified Gumbel, POT, MIS (method of independent storms)의 네 가지 방법에 의해 분석하고 Quantile 추정값을 도출하였으며, Pujol et al. (2007)은 프랑스 중부지방의 92개 관측소 일강우자료를 이용하여 전체를 7개의 기상학적 동질지역으로 분류하고, 각각에 대해서 POT사상을 추출하였다. 극치사상의 비정상성을 판별함에 있어서 POT 모형의 규모매개변수 (scale parameter)와 Poisson 분포의 λ 가 일정하지 않으며, 시간에 따라 선형 변화한다고 가정함으로써 지역적인 규모에서 극치사상의 발생

빈도와 강도의 정상성에 대한 판별을 가능케 하였으며, Poisson분포와 GPD분포에 의해 POT 자료를 분석할 결과 통계적으로 유의한 증가경향의 변동성을 갖는 지역을 분류하였다. Cooley et al. (2007)은 극치강우사상에 대하여 강우강도와 빈도로 분리된 계층모델 (separate hierarchical models)을 구성하였으며, 강우강도에 대해서는 높은 Threshold를 넘는 일강우량을 GPD를 이용하여 모델링하였고, 빈도에 대해서는 관측소들의 POT 발생횟수를 변수로 모델링하였으며, Markov Chain Monte Carlo (MCMC) 알고리즘을 이용한 불확실성 추정과 공간보간법을 이용하여 콜로라도 지역의 지역빈도를 해석하였다. Mendez et al. (2008)은 태풍의 지속시간과 계절성을 고려한 유의파고 (significant wave height)의 장주기 극치분포의 통계학적 모델을 구성하였다. 시간 종속성을 고려한 POT 자료와 Harmonic Function에 의해서 연주기 (annual cycle)와 반년주기 (semiannual cycle)를 고려한 모형을 구성하고, 계절적인 변동성을 고려함으로써 모델의 추정 오차를 줄였다. 이 외에도 기상학적 인자들로서 SOI (Southern Oscillation index), AOI (Antarctic Oscillation index), 그리고 SI (solar irradiance) 등을 고려한 계층적 공간모형을 이용해 POT 분석을 수행한 Aryal et al. (2009)의 연구도 극치강수량의 시공간적 경향성에 대한 해석방법을 다루고 있다.

우리나라 강우의 분포 (annual distribution)는 뚜렷한 연주기를 가지며 이와 함께 극치강수량의 부분간치 자료의 경우도 연주기를 가지는 것이 일반적이다. 이러한 경우에 전 자료를 하나의 모형으로 적합 시키는 경우 극치자료의 빈도해석 시 자료의 시간적 분포특성을 고려할 수 없는 단점이 있다. 이러한 점에서 본 연구에서는 서울지방의 24시간 임의지속시간 강수량을 대상으로 POT 자료를 구축하고 이들 자료에서 나타나는 계절성을 고려할 수 있는 비정상성빈도해석 기법을 제안하고자 한다. 이를 위해서 본 연구에서는 매개변수 추정, 불확실성분석, 빈도강수량 추정 등 일련의 빈도해석을 위해서 Bayesian 통계 기법을 활용하였다.

2. 대상자료 및 연구방법

2.1 대상자료

POT분석을 통한 극치수문자료의 해석이 제공하는 확장성과 보완성은 서론에서도 언급하였다 (Clarke et al., 2009; Madsen et al., 1997a, b). 높은 Threshold이상의 자료계열에 대해서 Pareto분포 또는 GPD분포를 이용한 모델링은 극치자료의 통계를 위해서 가장 많이 쓰이는 접근법이다. GPD는 분포의 높은 Threshold를 통해 얻어진

POT 자료계열에만 통계학적 유의성을 가지는 분포형으로서 낮은 Threshold를 이용한 POT 자료는 GPD 분포에 적합시킬 수 없다. 이런 이유로 heavy tail 분포를 가지고 있는 경우에, 낮은 Threshold를 초과하는 분포에 대해서도 적용성이 높도록 확률분포모델을 개선한 연구가 최근 발표된 바 있다 (Beirlant et al., 2009).

극치값의 분석에 있어서 국부최소값 (local minima)에 의한 오류를 방지하고, 부분적인 결측 구간이 있는 자료의 보안을 위해서 POT계열이 이용되기도 하지만, Threshold 값의 선택에 있어서는 신중할 필요가 있다고 알려져 있다 (Clarke et al., 2009). 국내의 연구사례를 통해서도 엄명진 등 (2008)이 네 개의 관측소 자료에 대해 산정방법별로 그 적정성을 비교한 연구가 있으나, 관측소마다 상이한 강우의 분포형태로 인해, 일반적으로 적정한 Threshold 값의 크기가 얼마인지는 여전히 정하기 어려운 문제이다. GPD 분포의 경우 Threshold에 따른 분포적합성이 크게 달라진다.

본 연구에서는 계절성의 평가를 위해 극히 높은 값의 Threshold 보다는 다양한 발생 시기 정보를 포함할 수 있는 POT 자료를 추출하기 위해서, 서울지점의 시단위 강우자료를 대상으로 전기간 강우량계열의 상위 98%에 해당하는 값으로 Threshold를 정하였으며, GEV분포형을 가정하고 적합성을 검토한 후 연구에 이용하였다. 서울지점 시단위 강우자료계열의 보유기간은 1961년부터 2008년까지 48년간이며, 98%에 해당하는 Threshold 값은 46.7 mm이다. Fig. 1은 서울지점의 시단위 강수량자료를 이용한 24시간 지속 일강우량 시계열이며, Fig. 2는 추출된 POT series의 연중 월별 분포를 나타낸 Boxplot이다.

전체 48년의 관측기간동안 연도별로 추출된 POT의 선형 경향성을 Fig. 3에 도시하였다. POT 계열의 양적 평균은 84.9 mm, 증가 기울기는 0.0407로 나타났다. 또한, 연도별로 추출된 POT의 발생빈도를 Fig. 4에 도시하였다. 연도별 POT 발생횟수의 평균은 7.3회이며 발생횟수의 증가 기울기는 0.0243이다.

본 연구에서 사용할 확률분포형을 결정하기 위해 추출된 POT를 대상으로 모형의 적합성을 평가하였다. 앞서 언급했듯이 자료의 계절성에 대한 비정상성 빈도해석을 위해서 GEV 분포형을 가정하였으며, GEV 분포가 서울지역의 POT 사상에 적합한지를 평가하기 위해서 시각적인 검토를 실시하였다. Fig. 5의 첫 번째 그림은 정규분포 확률지에 서울강수량 POT를 도시한 결과이며, 정규분포의 경우 자료계열이 이론적인 정규분포의 Quantile (직선) 값을 크게 벗어나고 있는 것을 확인할 수 있다. 반면 GEV 분포의 경우 경험적인 확률밀도함수와 누가확률밀도함수 모두에서 유사한 거동을 보임을 확인할 수 있다. 따라서

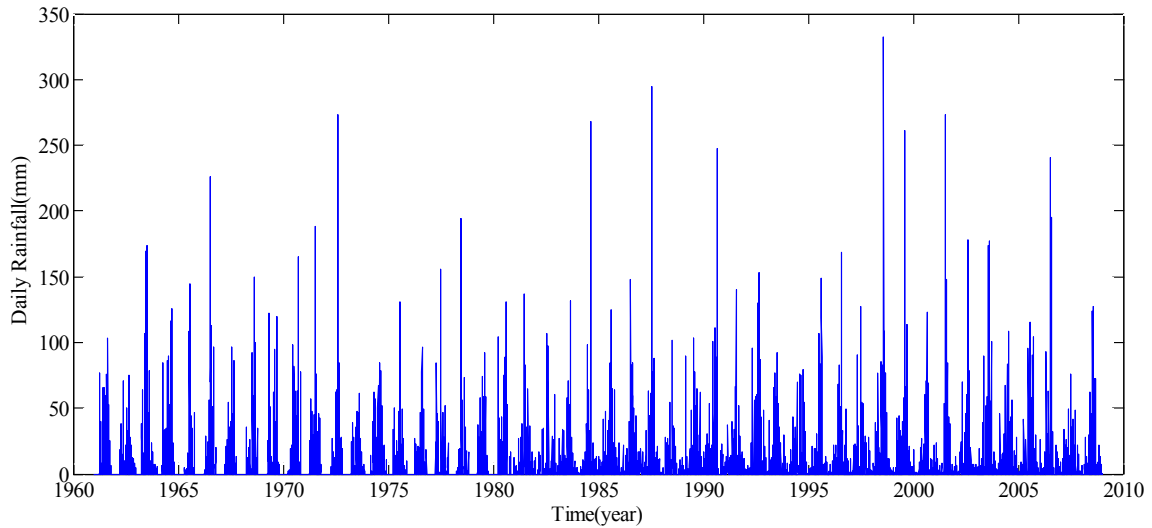


Fig. 1. Daily Rainfall Time Series at Seoul Station

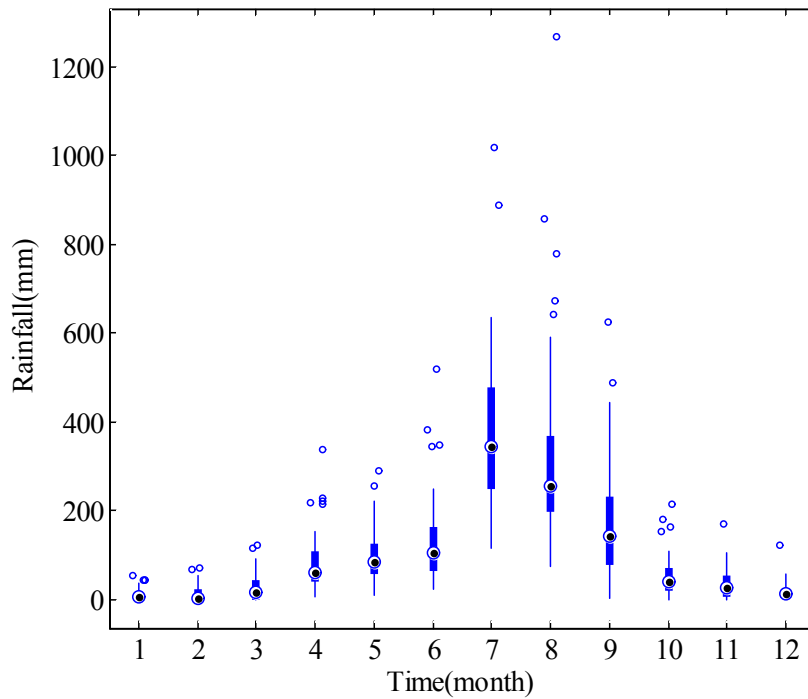


Fig. 2. Annual Distribution of Precipitation at Seoul Station. A Strong Seasonal Cycle is Identified. The Void Circle Indicates Outlier while the Filled Circle Indicates Median Value

GEV 분포형을 이용한 POT 계열의 계절성 평가에 무리가 없을 것으로 판단된다.

2.2 Markov Chain Monte Carlo 모의

수리적으로 계산이 불가능하거나 복잡한 적분, 추정 등의 문제에 사용되는 Monte Carlo 기법은 최근에 수리 수문학 분야에서 위험도 및 불확실성을 평가하는 수단으로 널리 이용되고 있다 (Kwon and Moon, 2006). Monte Carlo

기법은 관심이 있는 값을 확률변수의 기대값으로 표현하고 이를 모의를 통하여 추출된 동일한 분포를 따르며 서로 독립 (independent and identically distributed: iid)인 표본들의 표본평균을 이용하여 추정하는 방법이라고 할 수 있다.

이에 반해 MCMC 기법은 주어진 다변량 확률분포가 복잡하여 이를 따르는 iid 난수를 얻을 수 없는 경우에 사용가능한 기법으로서 iid 난수 대신 Markov Chain 난수

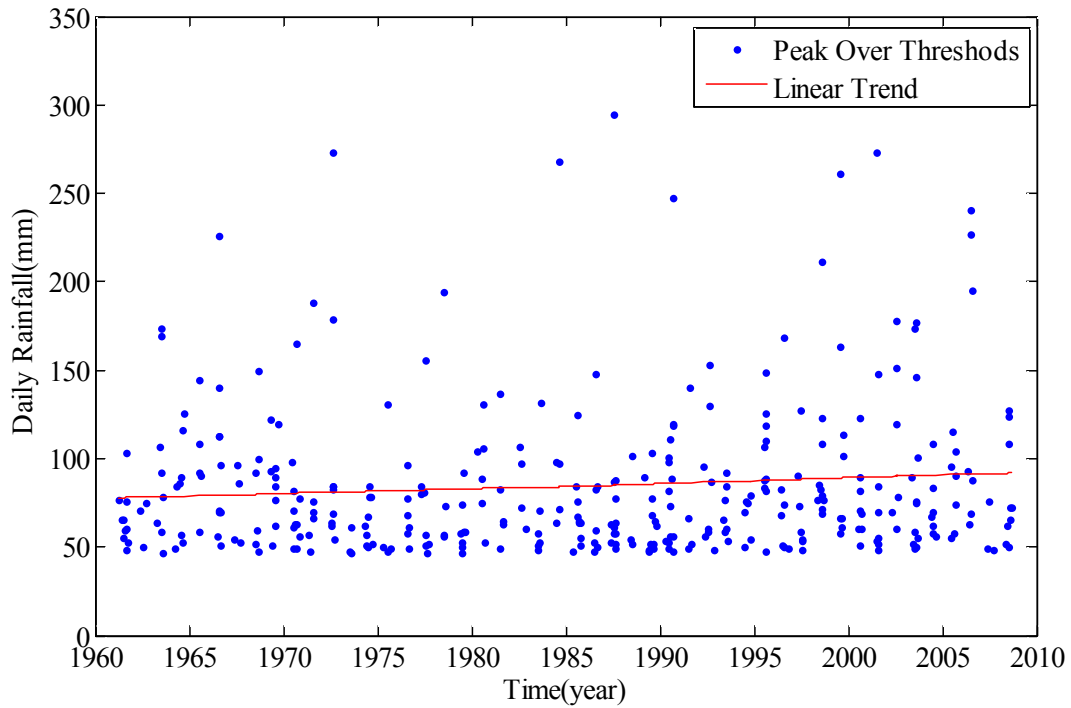


Fig. 3. Peaks Over Threshold Time Series at Seoul Station.
The Dots Show POTs and the Solid Line is a Trend Line Showing a Slight Increasing Trend

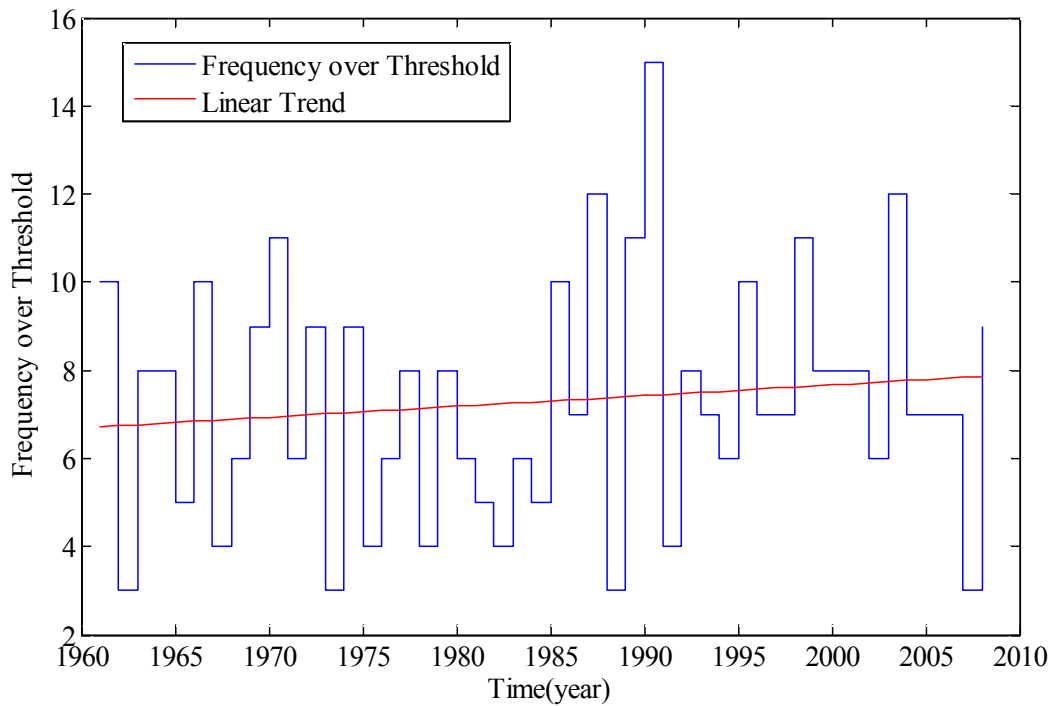


Fig. 4. Frequency of Peaks Over Threshold with Trend Line.
The Frequency is Aggregated Over a Year, and Increased Trend is Identified.

를 추출하여 사용한다. Markov Chain을 통해 난수를 발생시킨다고 해서 정확하게 관심이 되는 확률분포를 따르지 않지만 이를 일정 시간동안 반복 후에 얻어지는 난수들은 추출을 원하는 분포에 수렴하게 된다. 따라서

MCMC 기법은 복잡한 다변량 확률분포 및 매개변수의 추정을 요하는 문제에서 주로 사용되며 또한 Bayesian 통계 기법에서 사후분포의 추론에 이용될 수 있다. 본 연구에서는 2가지 관점에서 MCMC 기법을 이용하게 된다. 즉

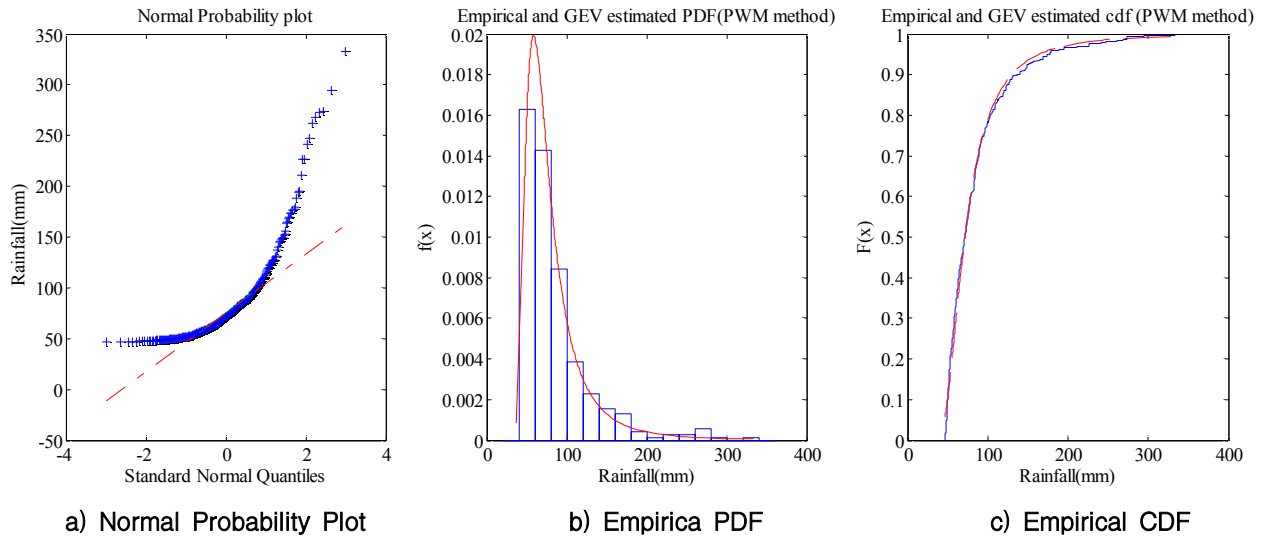


Fig. 5. Graphical Representation of POT Series Using

종속변수 Y 에 대해서 조건부 분포를 갖는 각 독립변수의 사후분포를 추정함과 동시에 불확실성의 정량적인 해석이 가능한 이유로 MCMC 기법을 사용하였다.

MCMC 기법의 대표적인 방법으로 메트로폴리스-헤스팅 알고리즘 (Metropolis-Hastings algorithm)과 깁스 표본법 (Gibbs sampling) 등이 있으며, 본 연구에서는 깁스 표본법을 이용하였다. 깁스 표본법은 원하는 다변량 확률 분포에서 iid 표본을 추출하는 것이 복잡하거나 난해한 경우 이용 가능한 방법으로서 2개의 변수를 갖는 다변량 확률 분포를 이용하여 설명하면 다음과 같다. 2개의 변수를 갖는 다변량 확률밀도함수를 $f(\theta_1, \theta_2)$ 라고 하자. 깁스 표본법은 확률밀도함수로부터 직접 표본을 추출할 수는 없으나 각각의 변수들에 대해서 다른 두 변수들이 주어졌을 때의 조건부 분포가 알려져 있고 이로부터의 표본추출이 가능한 경우에 사용할 수 있다. 알고리즘을 간단히 정리하면 다음과 같다.

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

위의 알고리즘에서 보듯이 깁스 표본법은 조건부 분포들에서 조건으로 주어지는 변수들의 값은 정확하게 바로 직전의 단계에서 주어진 값들이 사용되게 되며 따라서 조건부 분포에서 추출된 난수들이 안정 상태에 도달하는 것

이 주어진 다변량 확률분포를 정확히 따르는 난수가 되는 척도가 되며 깁스 표본법을 구현하는데 가장 중요한 부분이 된다.

2.3 계절성을 고려한 비정상성 빈도해석

본 연구에서 적용할 확률분포형을 결정하기 위하여 GEV (generalized extreme value distribution) 확률분포를 이용하였다. GEV 분포형의 확률밀도함수와 누적확률밀도함수는 다음 Eqs. (1)~(2)와 같다. 여기서 μ 는 위치 매개변수 (location parameter), σ 는 규모매개변수 (scale parameter), ξ 는 형상매개변수 (shape parameter)이다.

$$f(x; \mu, \sigma, \xi) = \frac{1}{\sigma} \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]^{(-1/\xi) - 1} \quad (1)$$

$$F(x; \mu, \sigma, \xi) = \exp \left[- \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right)^{-1/\xi} \right] \right] \quad (2)$$

POT 계열의 분포가 GEV 분포를 따른다고 가정할 때, 계절적인 변동성을 고려한 POT 자료 Y 의 위치 및 규모 매개변수에 대하여 아래와 같이 시간에 따른 변동성을 갖는 식으로 표현할 수 있다. 본 논문에서는 형상매개변수는 시간에 따라 일정하다고 가정하였다.

$$Y \sim GEV(\mu(t), \sigma(t), \xi) \quad (3)$$

Bayesian 모형 하에서, Eq. (3)의 모든 매개변수들은 확률분포를 가지게 된다. 즉, $\mu(t)$, $\sigma(t)$, ξ 가 사전분포를 가지게 되며 $\mu(t)$, $\sigma(t)$ 는 시간에 따른 주기성을 갖는 함수로서 다음의 Eqs. (4) and (5)와 같이 Fourier 급수의 형태로 표현할 수 있다. 여기서, t 는 시간을 의미하며 α 와 β 는 시간과 매개변수 μ 와 σ 의 사전분포를 연계시키는 Fourier 계수를 나타낸다. 본 연구에서는 시간에 따라 2개의 매개변수

가 동시에 변화한다고 가정하여 통계적 추론 (statistical inference)을 실시하였다. 주기함수는 시간의 변화에 따른 POT의 발생일자 $td(t)$ 를 계절성의 범위 T 에 대한 상대주기함수로 표현하였다. 여기서, $td = 1, \dots, 365$ 의 범위를 가지며 $T = 365$ 의 값을 갖는다.

$$p(\mu(t)) \sim N\left(\alpha_0 + \alpha_1 \times \sin\left(2\pi \times \frac{td(t)}{T}\right) + \alpha_2 \times \cos\left(2\pi \times \frac{td(t)}{T}\right), \sigma_\alpha^2\right) \quad (4)$$

$$p(\sigma(t)) \sim N\left(\beta_0 + \beta_1 \times \sin\left(2\pi \times \frac{td(t)}{T}\right) + \beta_2 \times \cos\left(2\pi \times \frac{td(t)}{T}\right), \sigma_\beta^2\right) \quad (5)$$

$$\xi \sim G(k_\xi, s_\xi) \quad (6)$$

시간에 따른 계절성을 고려하는데 있어서 1차 Fourier 급수만을 고려하였다. 이는 우리나라의 경우 연주기가 뚜렷하고 모형의 매개변수 또한 줄일 수 있기 때문에 1차 급수만으로 모형을 구성하였다. 본 연구에서는 이러한 주기성을 비정상성으로 고려하는 방법론의 전개 및 해석 방법에 대한 개념을 제시하는데 주안점을 두었다. 그러나 자료의 주기특성이 복잡한 경우에는 Spline기법과 같은 Smoothing 기법을 적용하여 해석이 가능하다. 본 연구에서는 48년간의 자료를 대상으로 하며 총 8개의 매개변수의 사후분포를 추정하는 것이 최종 목적이다. 6개의 회귀매개변수는 정규분포로 가정하였다. α 와 β 의 분산 σ_α 와 σ_β 는 Gamma 분포로 가정하였으며 이를 Eqs. (7)~(10)과 같이 나타낼 수 있다. 결과적으로 9개의 매개변수에 대한 결합 확률밀도함수 (joint distribution)는 Eq. (11)과 같다.

$$\alpha_n \sim N(\mu_{\alpha_n}, \sigma_{\alpha_n}) \quad n = 0, \dots, 2 \quad (7)$$

$$\beta_n \sim N(\mu_{\beta_n}, \sigma_{\beta_n}) \quad n = 0, \dots, 2 \quad (8)$$

$$\sigma_\alpha \sim G(k_\alpha, s_\alpha) \quad (9)$$

$$\sigma_\beta \sim G(k_\beta, s_\beta) \quad (10)$$

$$p(\alpha_0, \alpha_1, \alpha_2, \beta_0, \beta_1, \beta_2, \sigma_\alpha, \sigma_\beta, \xi) \propto 1 \quad (11)$$

Bayes 정리에 의해서 매개변수의 사후분포는 다음과 같이 전개될 수 있다.

$$p(\Theta | y) = \frac{p(\Theta, y)}{p(y)} \propto p(y | \Theta) \cdot p(\Theta) \quad (12)$$

여기서 $\Theta = [\mu(t), \sigma(t), \xi, \alpha_0, \alpha_1, \alpha_2, \beta_0, \beta_1, \beta_2, \sigma_\alpha, \sigma_\beta]$ 는 전체매개변수들의 집합을 나타내며, $p(\Theta)$ 는 매개변수들의 사전분포를 $p(y | \Theta)$ 는 POT자료 y 의 우도함수 (likelihood function)를 의미한다.

$$p(y | \Theta) = \prod_{t=1}^T GEV(y_t | \mu(t), \sigma(t), \xi) \quad (13)$$

여기서, T 는 POT계열의 자료수를 나타낸다. Eqs. (4)~(6)과 Eqs. (11) and (13)에서 정의되는 매개변수들의 사전분포들을 Eq. (14)에 대입시킴으로서 매개변수들의 사후분포를 추정할 수 있다. Eq. (14)에서는 GEV 분포의 위치 및 규모 매개변수와 회귀매개변수들을 위한 비공액사전분포 (non-conjugate prior)를 이용하여 여러 매개변수들을 추정한다. 공액사전분포 (conjugate prior)는 사전분포와 우도가 공액을 이룬다는 의미로서 사전분포와 사후분포가 동일한 분포를 나타내도록 사전분포를 선택하는

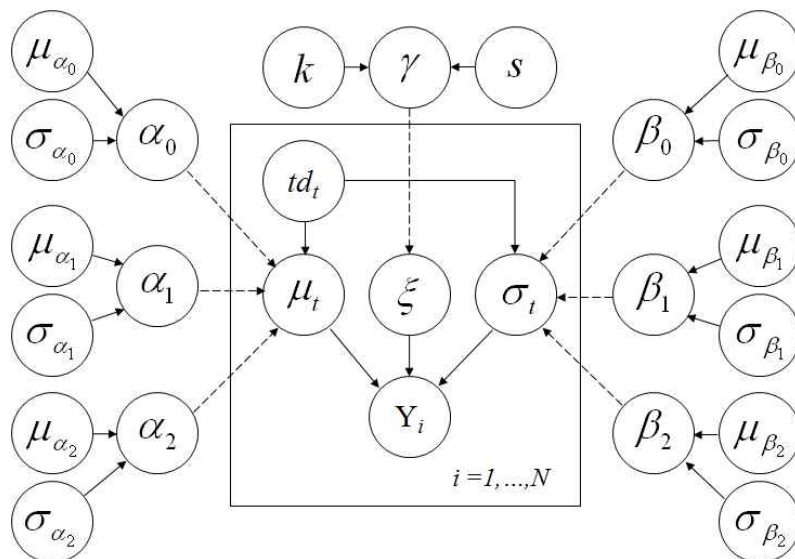


Fig. 6. A Conceptual Bayesian Network Model for Nonstationary Frequency Analysis Considering Seasonality

방법으로서 편의성을 기준으로 선택되는 간단한 방법이다. 즉, 우도함수가 이항분포를 따르는 경우 Beta 분포를 사전분포로 사용하게 되면 공액분포가 된다. 본 연구에서는 비공액사전분포를 이용하여 사후분포를 추정하였다.

$$p(y | \theta) \propto \prod_{t=1}^T GEV(y_t | \mu(t), \sigma(t), \xi) \cdot N\left(\alpha_0 + \alpha_1 \times \sin\left(2\pi \times \frac{td(t)}{T}\right) + \alpha_2 \times \cos\left(2\pi \times \frac{td(t)}{T}\right), \sigma_\alpha^2\right) \cdot N\left(\beta_0 + \beta_1 \times \sin\left(2\pi \times \frac{td(t)}{T}\right) + \beta_2 \times \cos\left(2\pi \times \frac{td(t)}{T}\right), \sigma_\beta^2\right) \cdot G(k_\xi, s_\xi) \quad (14)$$

Eq. (14)는 모든 매개변수에 대한 적분을 통해 직접적으로 추정하는 것은 불가능하며 본 연구에서는 앞서 언급한 MCMC 기법을 도입하여 매개변수들의 사후분포를 추정하게 된다. 본 연구에서는 MCMC 기법 중 깃스프분법을 이용하여 회귀매개변수들을 추정하였으며 모형의 Convergence를 확증하기 위해서 3개의 Chain을 독립적으로 시행하여 Sampling이 효과적으로 혼합(mixing)되도록 하였다. 형상매개변수의 경우는 Gamma 분포로 적용하였다. Fig. 6은 본 연구에서 구성한 Bayesian Network 모형에 대한 개념도를 나타낸다.

3. 적용

본 연구에서는 서울지점의 시단위 강수량 자료를 이용

해 추출한 POT 자료계열을 대상으로 모형의 적합성을 평가하였다. Fig. 7은 상위 98% 이상의 범위에서 Threshold 변화에 대한 Mean Excess Plot이다. Threshold를 초과하는 값들의 평균이 Threshold 변화에 따라서 선형에 가까우면 효과적으로 Threshold가 선택되었다고 할 수 있으며 이중 가장 작은 Threshold를 선택하도록 권장된다. 본 연구에서는 상위 98%에 해당하는 Threshold를 이용하여 추출한 POT가 극치분포의 경향을 잘 반영하고 있는지 검토하기 위해 Mean Excess Plot을 이용하였다. 채택한 Threshold에 가까운 평균의 분포가 선형에 가까움을 볼 수 있다.

Bayesian MCMC 모의를 통해 추정된 모형 매개변수들의 사후분포는 Fig. 8과 같다. Table 1은 추정된 사후분포의 매개변수 값을 보여주고 있다. 계절적 변동성에 대하여 매개변수들의 사후분포로부터 판단할 수 있는 사실은 서울지점 극치강수량의 경우 Fourier 급수의 매개변수가 Median값을 기준으로 모두 음의 값을 가지며 이는 단일 Peak를 가지는 주기함수의 형태를 의미한다. 사후분포로부터 위치매개변수를 구성하는 $\alpha_{0,1,2}$ 의 변동성이 규모 매개변수를 구성하는 $\beta_{0,1,2}$ 보다 크음을 쉽게 인지할 수 있다. Fourier 급수에서 Sine을 구성하는 매개변수보다는 Cosine을 구성하는 매개변수의 불확실성이 크게 나타내고 있음을 Fig. 8과 Table 1에서 확인할 수 있다. 이러한 특성은 p -value를 통해서 효율적으로 판단할 수 있다.

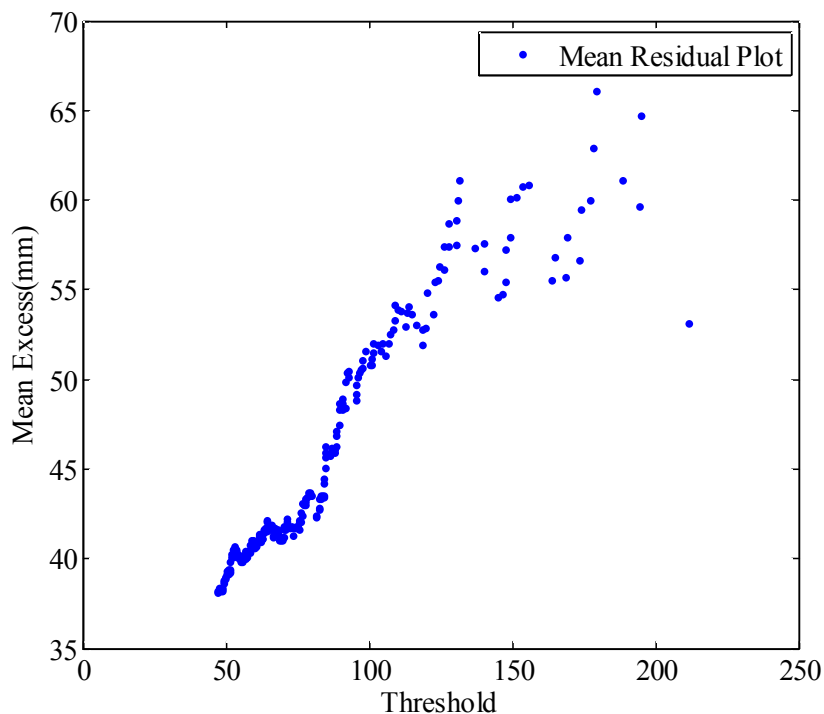


Fig. 7. Mean Excess Plot given different Thresholds.
Note that Linear Line Over threshold Indicates Generally Appropriate Selection for Threshold

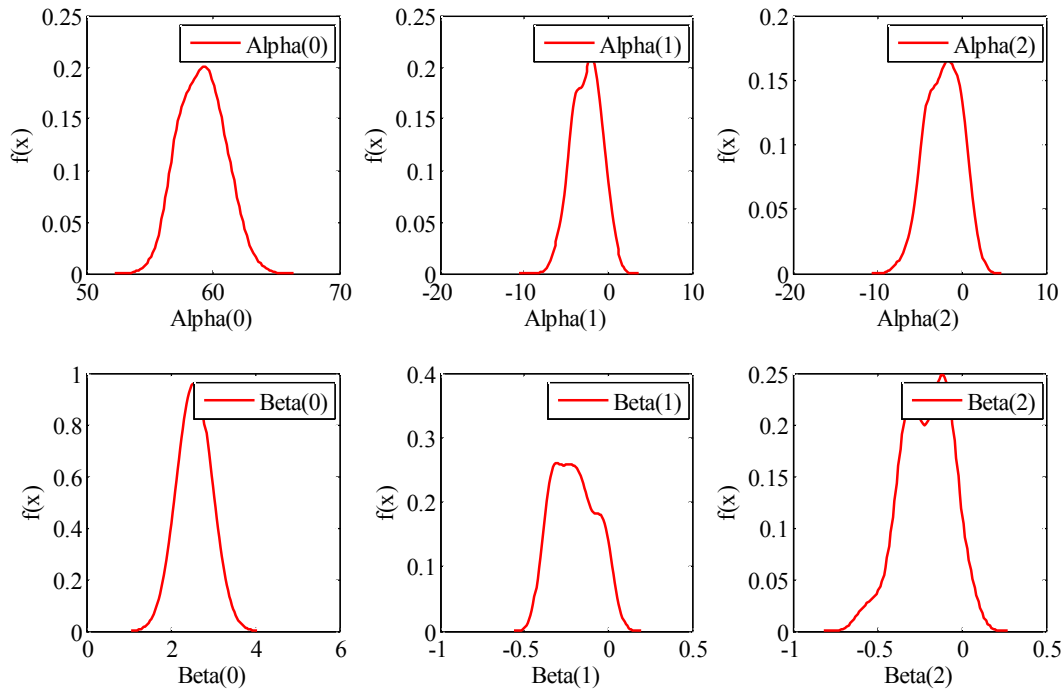


Fig. 8. Model Parameters of Fourier Series in Nonstationary Frequency Analysis Considering Seasonality

Table 1. Posterior Density of Parameters and Associated Uncertainty Bounds

Parameters	Mean	SD	2.5%	Median	97.5%
α_0	59.12	1.72	56.12	59.18	62.41
α_1	-2.57	1.68	-6.07	-2.44	0.45
α_2	-2.16	2.05	-6.28	-2.10	1.50
β_0	2.56	0.11	2.35	2.56	2.78
β_1	-0.20	0.12	-0.39	-0.21	0.03
β_2	-0.21	0.14	-0.54	-0.20	0.05
μ	61.68	1.65	58.63	61.72	65.11
σ	16.37	1.79	13.19	16.43	19.82
ξ	0.61	0.08	0.46	0.61	0.77

Bayesian 방법의 경우 사후분포로부터 매개변수들의 p -value를 추정할 수 있으며 이를 통해 매개변수들의 유의성을 효과적으로 검토할 수 있다. 매개변수들의 p -value를 추정해보면 $\alpha_0=0$, $\alpha_1=0.06$, $\alpha_2=0.16$, $\beta_0=0$, $\beta_1=0.05$, $\beta_2=0.06$ 의 값을 갖는다. 따라서 α_2 를 제외하고는 유의수준 0.1에서 매개변수들은 통계적으로 유의하다. 이 밖에도 형상매개변수 ξ 는 0.61로서 Heavy Tail의 특성을 단적으로 보여주고 있으며 불확실성이 매우 작은 사후분포의 형태임을 Table 1로부터 알 수 있다. 이는 정상성가정으로 추정된 ξ 의 값이 0.63으로 비정상성을 고려한 경우와 매우 유사한 값을 갖는 것을 토대로 분포의 특성이 계절성을

고려한 비정상성분석을 통해서도 유지되고 있음이 확인되었다.

본 연구에서 수립된 모형은 POT의 발생시기 및 시간 변화에 따른 분포의 변화를 반영하고 있으며, 이에 대한 비정상성 빈도분석을 통해 추정된 사후분포의 매개변수들을 이용하여 원하는 확률수문량을 산정할 수 있다. Eq. (15)는 추정된 사후분포의 매개변수를 이용한 GEV 분포의 Quantile 함수이며, 이를 통해 재현기간 R_T 에 해당하는 확률수문량을 추정할 수 있다. Fig. 9는 추출된 POT의 발생시기별 분포와 그 평균을 도시한 그림이며, 점선으로 표시된 50년 빈도의 설계강우량이 Quantile 함수로부터

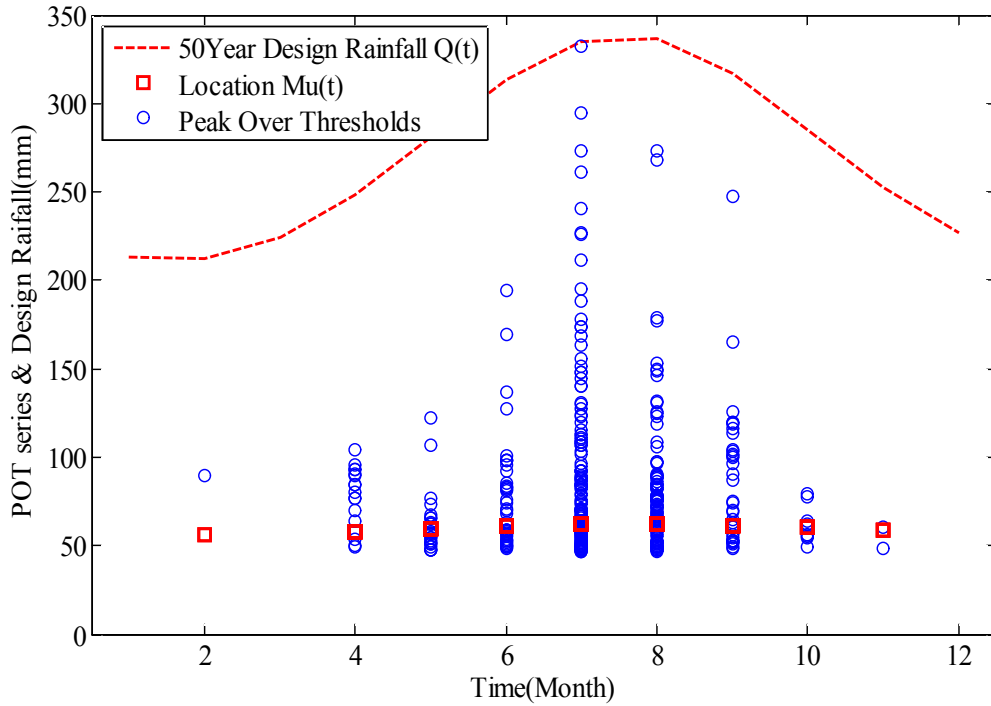


Fig. 9. POT Distribution and the Estimated 50-year Design Rainfall Cycle. The Design Rainfall Cycle is Interpolated with Estimated Posterior Distributions of Fourier Model Parameters

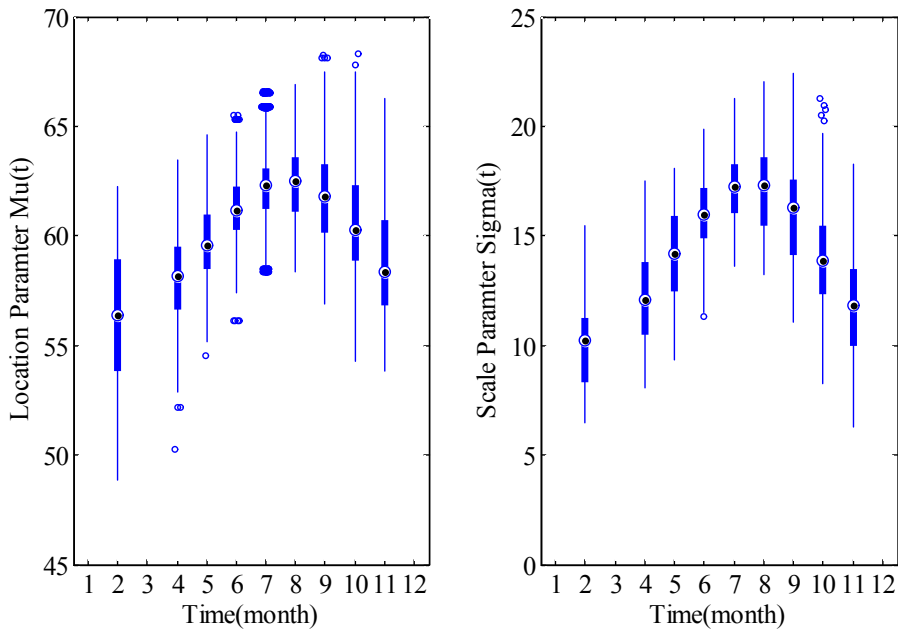


Fig. 10. Annual Distribution of Location and Scale Parameters. The Circle Indicates Outlier While the Filled Circle Shows Median Value

추정된 값이다. 그림과 같이 POT 자료계열은 7월을 중심으로 종의 형태를 가지는 전형적인 연주기 특성을 보여주고 있으며 추정된 50년 빈도 확률강수량의 분포 또한 연주기의 형태로 추정된 것을 확인할 수 있다.

$$X(t) = \mu(t) + \frac{\sigma(t)}{\xi} \cdot \left[-\log\left(1 - \frac{1}{R_T}\right)^{-\xi} - 1 \right] \quad (15)$$

사후분포로부터 추정된 위치매개변수 μ 와 규모매개변수 σ 의 분포를 월별로 추출하여 Fig. 10에 도시하였다. 규모매개변수의 경우는 계절적 분포에 따른 분산의 차이가

크지 않은 반면, 위치매개변수의 경우 변동성이 계절에 따라 크게 나타나고 있다. 이는 앞서 Table 1과 Fig. 8을 통해서 설명했듯이 위치매개변수를 구성하는 $\alpha_{0,1,2}$ 가 상대적으로 큰 불확실성으로부터 기인하는 결과이며 통계적으로 극치계열에서 분산보다는 평균이 계절적으로 변동성이 큼을 의미한다 하겠다. 극치강우가 집중되는 시기에서 위치 및 규모 매개변수 모두에서 불확실성이 적게

되는 것은 상대적으로 자료의 개수가 많이 존재하기 때문이며 역으로 건기에는 적은 극치사상들로 인해 불확실성이 증가하고 있음을 확인할 수 있다.

마지막으로 계절성을 고려한 비정상성 빈도해석 모형을 구축하고, POT 자료에 대한 확률분포 모의를 통해 추정된 사후분포의 매개변수들을 이용하여 재현기간별 확률강우량을 Table 2와 Fig. 11에 도시하였다. 재현기간은

Table 2. Design Rainfall Estimates by Nonstationary Frequency Analysis Considering Seasonality

Return Period		2 yr	4 yr	5 yr	10 yr	20 yr	30 yr	50 yr	100 yr
Stationary		68.10	91.66	101.24	140.23	199.31	247.42	327.94	487.12
Non-stationary Quantile	Jan.	60.15	74.35	80.08	103.26	138.03	166.15	212.92	304.60
	Feb.	60.00	74.13	79.84	102.90	137.50	165.48	212.01	303.24
	Mar.	61.02	76.18	82.30	107.05	144.19	174.21	224.15	322.06
	Apr.	62.98	80.20	87.15	115.26	157.44	191.55	248.27	359.48
	May	65.42	85.43	93.51	126.17	175.17	214.79	280.69	409.88
	Jun.	67.73	90.56	99.78	137.06	193.00	238.23	313.44	460.92
	Jul.	69.20	93.92	103.90	144.25	204.81	253.76	335.19	494.83
	Aug.	69.36	94.21	104.23	144.79	205.64	254.85	336.68	497.12
	Sep.	68.16	91.31	100.66	138.44	195.15	240.99	317.24	466.74
	Oct.	66.00	86.38	94.61	127.88	177.80	218.17	285.30	416.92
	Nov.	63.54	81.08	88.16	116.80	159.77	194.52	252.30	365.60
	Dec.	61.42	76.79	83.00	108.08	145.73	176.17	226.79	326.04

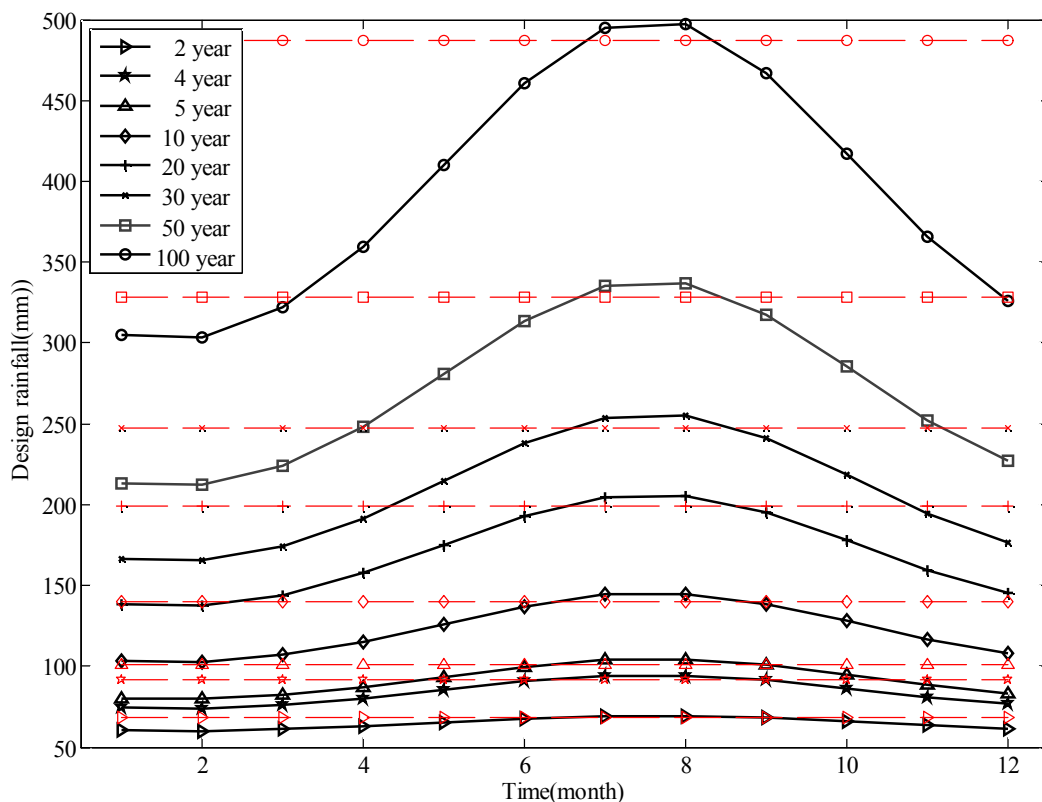


Fig. 11. Comparison of Design Rainfall Between Stationary and Nonstationary Frequency Analysis.

2, 4, 5, 10, 20, 30, 50, 100년에 대해서 Quantile 값을 추정하였고, 동일 재현기간에 대하여 기존의 정상성 빈도 해석을 통해서 추정된 확률강수량을 점선으로 도시하였다. 모든 경우에서 기존의 확률강수량은 계절성을 고려한 비정상성빈도해석의 7~8월 기준 값들과 거의 유사함을 확인할 수 있다. 이는 계절성을 고려하더라도 효과적으로 극치강수량의 분포특성을 고려할 수 있음을 의미한다. 이와 더불어 확률강수량의 시간적 분포에 대한 정보를 동시에 제공할 수 있으며 추정된 Fourier 급수의 사후분포를 통해 극치강수량이 존재하지 않는 월의 극치분포 또한 고려할 수 있는 장점이 있다.

4. 결론

최근 수문빈도해석을 위해 Block Maxima와 같은 연최대치계열보다는 POT 자료와 같은 부분기간치계열을 활용하고자 하는 연구가 많이 진행되고 있다. 이는 POT 자료를 이용함으로써 얻을 수 있는 장점들을 활용하는데 목적이 있다. 즉, 자료의 확충뿐만 아니라 계절성 확보, 발생빈도(occurrence) 모형 구축 가능 등의 다양한 이유로 POT 자료의 사용이 장려되고 있다 하겠다. 그러나 우리나라에서는 이러한 연구가 활발히 진행되고 있지 않으며 POT 자료의 장점을 효과적으로 빈도해석에 연계시켜 해석할 수 있는 방법론이 개발되지 않고 있다. 이러한 점에서 본 연구를 통해 POT 자료로부터 계절성을 추출하고 이를 빈도해석과 연계시켜 체계적으로 확률강수량을 추정할 수 있는 Bayesian 기법을 기반으로 하는 비정상성 빈도해석 개념을 소개하고 서울지방 24시간 임의지속시간 강수량에 대해서 적용하여 방법론의 적합성을 평가하였다. 본 연구의 진행과정과 결과를 요약하면 다음과 같다.

- 1) 임의지속 서울강수량에 대해서 98% Threshold를 적용하여 POT 자료를 구성하였으며 Mean Excess Plot를 활용하여 선택된 POT 자료의 적합성을 평가하였다.
- 2) POT 방법으로 추출된 극치강수량이 월강수량과 같이 계절성이 존재함을 확인할 수 있었으며 이를 매개변수적으로 모의하기 위해서 1차 Fourier 급수를 도입하였다.
- 3) POT 자료에 대해서 확률강수량을 추정하기 위해서 GEV분포형을 선택하였으며 적합도 검정 및 도시적 해석을 통해 분포형의 적합성을 평가하였다. GEV 분포형의 3개의 매개변수 중 위치 및 규모 매개변수가 계절적으로 변한다고 가정하였으며 형상매개변수는 계절적으로 동일하다고 가정하였다.
- 4) 위치 및 규모매개변수의 계절적 변동성을 1차

Fourier와 연계시켰으며 연관된 매개변수들의 사후분포를 Bayesian Markov Chain Monte Carlo 모의통계를 통해서 추정하였으며 최종적으로 시간에 따른 확률강수량을 추정하였다.

- 5) 계절성을 고려한 비정상성빈도해석 결과 계절성을 효과적으로 모의하면서 확률강수량의 추정이 가능하였으며 비정상성빈도해석로 추정된 7~8월의 확률강수량과 기존 정상성빈도해석의 결과가 유사한 값을 나타내었다. 이는 본 연구를 통해서 제시된 방법론이 극치강수량의 계절성과 양적특성을 동시에 고려할 수 있음을 의미한다 하겠다.

본 연구를 통해 도출되는 확률강수량의 계절적 분포는 다양한 계절적 특성치를 요구하는 수공구조물 설계 시에 활용될 수 있을 것으로 사료된다. 본 연구결과를 바탕으로 추후에는 계절성과 증가경향성을 동시에 모의할 수 있는 방법론이 필요하며, 이와 더불어 지구온난화와 같은 대규모 기상변화 특성을 고려할 수 있는 기상인자와 연계한 해석방법이 필요할 것으로 판단된다.

참고문헌

노재식, 이길춘 (1992). “하천 홍수량의 크기 및 빈도 결정.” **대한토목학회논문집**, 대한토목학회, 제12권, 제2호, pp. 141-150.

노재식, 이길춘 (1993). “하천홍수량의 지역화 회귀모형 개발을 위한 지역빈도해석.” **대한토목학회논문집**, 대한토목학회, 제13권, 제3호, pp. 139-154.

엄명진, 조원철, 허준행 (2008). “GPD 모형 산정을 위한 부분시계열 자료의 임계값 산정방법 비교.” **한국수자원학회논문집**, 한국수자원학회, 제41권, 제5호, pp. 527-544.

유철상 (2002). “강수의 계절성과 면적평균강수량의 추정 오차.” **한국수자원학회논문집**, 한국수자원학회, 제35권, 제5호, pp. 575-581.

유철상, 류소라, 김정환 (2002). “서울지점 월강수량 자료에 나타난 가뭄의 장기 재현특성.” **대한토목학회논문집**, 대한토목학회, 제22권, 제3B호, pp. 281-289.

진영훈, 박성천, 이연길 (2005). “수문시계열의 장·단기 성분 추출을 위한 웨이블릿 변환의 적용.” **대한토목학회논문집**, 대한토목학회, 제25권, 제6B호, pp. 493-499.

An, Y., and Pandey, M.D. (2005). “A comparison of methods of extreme wind speed estimation.” *Journal of Wind Engineering and Industrial Aerodynamics*, Vol. 93, No. 7, pp. 535-545.

Aryal, S.K., Bates, B.C., Campbell, E.P., Li, Y., Palmer,

- M.J., and Viney, N.R. (2009). "Characterizing and modeling temporal and spatial trends in rainfall extremes." *Journal of Hydrometeorology*, Vol. 10, No. 1, pp. 241-253.
- Beirlant, J., Joossens, E., and Segers, J. (2009). "Second-order refined peaks-over-threshold modelling for heavy-tailed distributions." *Journal of Statistical Planning and Inference*, Vol. 139, No. 8, pp. 2800-2815.
- Clarke, R.T., de Paiva, R.D., and Uvo, C.B. (2009). "Comparison of methods for analysis of extremes when records are fragmented: A case study using Amazon basin rainfall data." *Journal of Hydrology*, Vol. 368, No. 1-4, pp. 26-29.
- Cooley, D., Nychka, D., and Naveau, P. (2007). "Bayesian spatial modeling of extreme precipitation return levels." *Journal of American Statistical Association*, Vol. 102, No. 479, pp. 824-840.
- Drees, H., and Kaufmann, E. (1998). "Selecting the optimal sample fraction in univariate extreme value estimation." *Stochastic Processes and Their Applications*, Vol. 75, pp. 149-172.
- George, E.I., and McCulloch, R.E. (1993). "Variable Selection Via Gibbs Sampling." *Journal of American Statistical Association*, Vol. 88(423), pp. 881-889.
- Gomes, M.I., and Pestana, D. (2007). "A sturdy reduced-bias extreme quantile (VaR) estimator." *Journal of American Statistical Association*, Vol. 102, No. 477, pp. 280-292.
- Hall, P. (1990). "Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems." *Journal of Multivariate Analysis*, Vol. 32, pp. 177-203.
- Hall, P., and Welsh, A.H. (1985). "Adaptive estimates of parameters of regular variation." *Annals of Statistics*, Vol. 13, pp. 331-341.
- Kwon, H.H., and Moon, Y.I. (2006). "Improvement of overtopping risk evaluations using probabilistic concepts for existing dams." *Stochastic Environmental Research And Risk Assessment*, Vol. 20, No. 4, pp. 223-237.
- Madsen, H., Pearson, C.P., and Rosbjerg, D. (1997a). "Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events. 2. regional modeling." *Water Resources Research*, Vol. 33, pp. 759-769.
- Madsen, H., Rasmussen, P.F., and Rosbjerg, D. (1997b). "Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events. 1. at-site modeling." *Water Resources Research*, Vol. 33, pp. 747-757.
- Mendez, F.J., Menendez, M., Luceno, A., Medina, R., and Graham, N.E. (2008). "Seasonality and duration in extreme value distributions of significant wave height." *Ocean Engineering*, Vol. 35, No. 1, pp. 131-138.
- Pujol, N., Neppel, L., and Sabatier, R. (2007). "Regional approach for trend detection in precipitation series of the French Mediterranean region." *Comptes Rendus Geoscience*, Vol. 339, No. 10, pp. 651-658.

논문번호: 10-019	접수: 2010.03.07
수정일자: 2010.07.09/08.05	심사완료: 2010.08.05