

극치수문자료의 계절성 분석 개념 및 비정상성 빈도해석을 이용한 유효확률강수량 해석

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

권현한*, 이정주**, 이동률***

Hyun-Han Kwon, Jeong Ju Lee, Dong-Ryul Lee

요 지

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 이수과점에서 이루어지고 있으며 치수관점에서 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. 이러한 분석과정을 POT(peak over threshold)분석이라 하며 일정 기준값(threshold) 이상의 자료를 모두 취하여 빈도해석에 이용하는 방법으로서 기존 방법의 경우 연최대값이 일반적으로 7월과 8월에만 존재하게 되지만 POT 분석의 경우 여러 달에 걸쳐 빈도해석을 위한 자료가 구성되게 된다. 이를 빈도해석으로 연계시키기 위해서는 계절성을 비정상성으로 고려하여 모형화 할 수 있는 방법론의 개발이 필요하다. 본 연구에서는 이러한 목적을 위해서 계절성을 고려할 수 있는 비정상성빈도해석 기법의 개념을 제시하고 모형으로 개발하고자 한다. GEV 또는 Gumbel 분포의 매개변수와 계절성을 연계시키기 위해서 Fourier 급수가 활용되며 매개변수는 Bayesian 기법을 통해 최적화 된다. 이를 통하여 설계강수량의 계절적 분포를 정량적으로 해석할 수 있으며 미래의 극치강수량에 대한 분포특성 또한 확률적으로 해석이 가능하다. 본 연구에서 제안된 방법은 국내외 시간강수량자료에 적용되어 적합성과 적용성이 평가된다.

핵심용어 : 극치수문량, 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 이수과점에서 이루어지고 있으며 치수관점에서 극치수문량의 계절

* 정회원 · 전북대학교 토목공학과 조교수 · E-mail : hkwon@jbnu.ac.kr
** 정회원 · 전북대학교 토목공학과 박사수료 · E-mail : julee@jbnu.ac.kr
*** 정회원 · 한국건설기술연구원 수자원연구실 책임연구원 · E-mail : dryi@kict.re.kr

성을 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 등, 2009). 이러한 분석과정을 POT(peak over threshold) 분석이라 하며 일정 기준값(threshold) 이상의 자료를 모두 취하여 빈도해석에 이용함으로써 극치분포 Upper Tail의 추가정보를 제공하게 되고, 극치분포의 매개변수(parameter)와 Quantile을 보다 적절하게 추정할 수 있는 방법이다(Madsen 등, 1997a; 1997b). 연최대치계열을 이용하는 일반적인 극치빈도해석의 경우는 연최대값이 보통 7월과 8월에만 존재하게 되지만, POT 분석의 경우 여러 달에 걸쳐 빈도해석을 위한 자료가 구성되게 된다. 이를 빈도해석으로 연계시키기 위해서는 계절성을 비정상성으로 고려하여 모형화 할 수 있는 방법론의 개발이 필요하다. 본 연구에서는 이러한 목적을 위해서 계절성을 고려할 수 있는 비정상성빈도해석 기법의 개념을 제시하고 모형으로 개발하고자 한다.

2. 대상자료 및 분석방법

2.1 대상자료

본 연구에서는 방법론에 치중하기 위해 우선 서울지역 한 지점의 시단위 강우자료를 이용하여 분석을 위한 POT자료를 추출하였다. Threshold는 상위 98%에 해당하는 값으로 정하였으며, 분포형 적합성을 검토한 후 GEV분포형을 연구에 적용하였다.

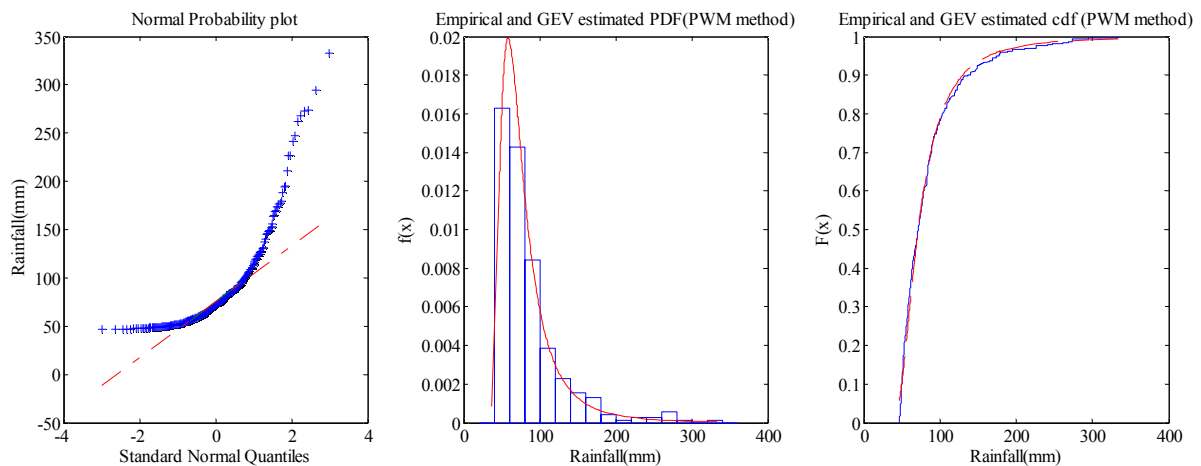


Figure 1. Graphical representation of POT series using a) Normal probability plot, b) Empirical PDF and c) Empirical CDF

2.2 계절성을 고려한 비정상성 빈도해석

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

$$Y \sim GEV(\mu(t), \sigma(t), \xi) \quad (1)$$

Bayesian 모형 하에서, 식(1)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 식(2), 식(3)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 매개변수가 동시에 변화한다고 가정하여 확률적 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.

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

$$p(\sigma(t)) \sim N\left(\beta_0 + \beta_1 \times \sin\left(2\pi \times \frac{td(t)}{T}\right) + \beta_2 \times \cos\left(2\pi \times \frac{td(t)}{T}\right), \sigma_\beta^2\right) \quad (3)$$

$$\xi \sim G(k_\xi, s_\xi) \quad (4)$$

우리나라의 경우 강우의 연주기가 뚜렷하기 때문에 1차 Fourier 급수만을 고려하였다. 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 식(5)~식(8)과 같이 나타낼 수 있다. 결과적으로 9개의 매개변수에 대한 결합확률밀도함수(joint distribution)는 식(9)와 같다.

$$\alpha_n \sim N(\mu_{\alpha_n}, \sigma_{\alpha_n}) \quad n = 0, \dots, 2 \quad (5)$$

$$\beta_n \sim N(\mu_{\beta_n}, \sigma_{\beta_n}) \quad n = 0, \dots, 2 \quad (6)$$

$$\sigma_\alpha \sim G(k_\alpha, s_\alpha) \quad (7)$$

$$\sigma_\beta \sim G(k_\beta, s_\beta) \quad (8)$$

$$p(\alpha_0, \alpha_1, \alpha_2, \beta_0, \beta_1, \beta_2, \sigma_\alpha, \sigma_\beta, \xi) \propto 1 \quad (9)$$

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

$$p(y | \Theta) = \prod_{t=1}^T GEV(y_t | \mu(t), \sigma(t), \xi) \quad (11)$$

여기서, T 는 POT계열의 자료수를 나타낸다.

3. 적용 및 결과

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.

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 (mean)	61.68	1.65	58.63	61.72	65.11
σ (mean)	16.37	1.79	13.19	16.43	19.82
ξ	0.61	0.08	0.46	0.61	0.77

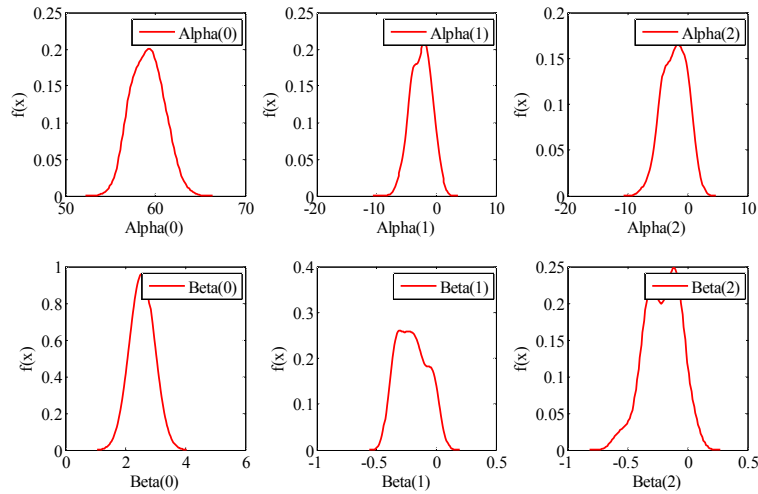


Figure 2. Model Parameters of Fourier Series in Nonstationary Frequency Analysis Considering seasonality

본 연구에서 수립된 모형은 POT의 발생시기 및 시간변화에 따른 분포의 변화를 반영하고 있으며, 이에 대한 비정상성 빈도분석을 통해 추정된 사후분포의 매개변수들을 이용하여 원하는 확률수문량을 산정할 수 있다. 식(12)는 추정된 사후분포의 매개변수를 이용한 GEV 분포의 Quantile 함수이며, 이를 통해 재현기간 R_T 에 해당하는 확률수문량을 추정할 수 있다. Figure 3은 추출된 POT의 발생시기별 분포와 그 평균을 도시한 그림이며, 점선으로 표시된 50년 빈도의 설계강우량이 Quantile 함수로부터 추정된 값이다. 그림과 같이 POT자료계열은 7월을 중심으로 종의 형태를 가지는 전형적인 연주기 특성을 보여주고 있으며 추정된 50년 빈도 확률강수량의 분포 또한 연주기의 형태로 추정된 것을 확인할 수 있다.

$$X(t) = \mu(t) + \frac{\sigma(t)}{\xi} \cdot \left[-\log\left(1 - \frac{1}{R_T}\right)^{-\xi} - 1 \right] \quad (12)$$

계절성을 고려한 비정상성 빈도분석 모형을 구축하고, POT 자료에 대한 확률분포 모의를 통해 추정된 사후분포의 매개변수들을 이용하여 재현기간별 Quantile 값을 이용하여 비정상성을 고

려한 확률강우량을 산정하여 Figure 4에 도시하였다.

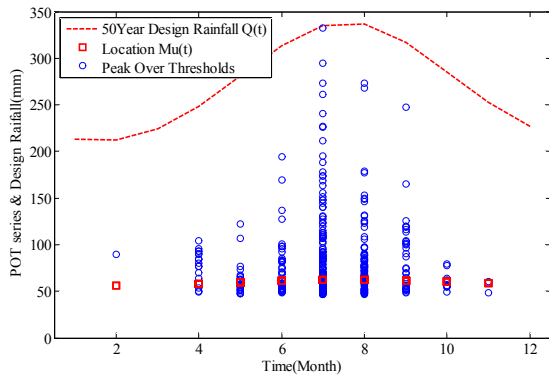


Figure 3. POT distribution and the estimated 50-year Design Rainfall cycle. The design rainfall cycle is interpolated with estimated posterior distributions of Fourier model parameters

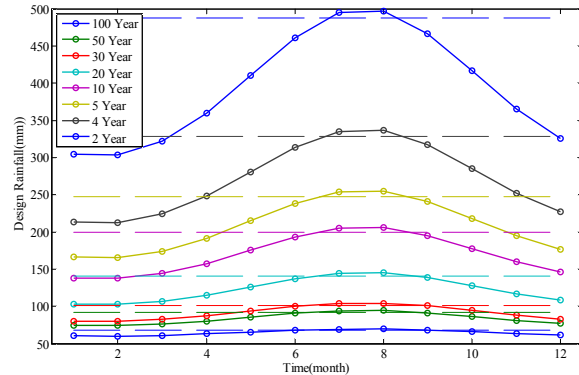


Figure 4. Comparison of design rainfall between stationary and nonstationary frequency analysis

모든 경우에서 기존의 확률강우량은 계절성을 고려한 비정상성빈도해석의 7월-8월 기준 값들과 거의 유사함을 확인할 수 있다. 이는 계절성을 고려하더라도 효과적으로 극치강우량의 분포특성을 고려할 수 있음을 의미한다. 이와 더불어 확률강우량의 시간적 분포에 대한 정보를 동시에 제공할 수 있으며 추정된 Fourier 급수의 사후분포를 통해 극치강우량이 존재하지 않는 월의 극치분포 또한 고려할 수 있는 장점이 있다.

4. 결론

본 연구에서는 임의지속 서울강우량에 대해 98% Threshold의 POT자료를 구성하였으며, GEV 분포형의 매개변수 중 위치 및 규모매개변수에 대해 계절적 변동성을 가정한 베이지안 모형을 구성하고, 사후분포의 매개변수를 MCMC모의를 통해 추정하였다. 이를 통해 계절성을 효과적으로 모의한 확률강우량을 추정할 수 있었으며, 이러한 확률강우량의 계절적 분포는 다양한 계절적 특성치를 요구하는 수공구조물의 설계에 이용할 수 있을 것으로 사료된다.

참고 문헌

1. Clarke, R. T., R. D. de Paiva, and C. B. Uvo(2009). Comparison of methods for analysis of extremes when records are fragmented: A case study using Amazon basin rainfall data, *Journal of Hydrology*, 368(1-4), 26-29.
2. Madsen, H., Pearson C. P., Rosbjerg D.(1997a). Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events. 2. regional modeling. *Water Resour. Res.*, Vol. 33, pp. 759 - 769.
3. Madsen, H., Rasmussen P. F., Rosbjerg D.(1997b). 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 - 757.