

Freund 이변량 지수분포의 매개변수 추정과정 검토

Review of Parameter Estimation Procedure of Freund Bivariate Exponential Distribution

박 철 순* / 유 철 상**

Park, Cheolsoon / Yoo, Chulsang

Abstract

This study reviewed the parameter estimation procedure of the Freund bivariate exponential distribution for the decision of the annual maximum rainfall event. The method of moments was reviewed first, whose results were compared with those from the method of maximum likelihood. Both methods were applied to the hourly rainfall data of the Seoul rain gauge station measured from 1961 to 2010 to select the annual maximum rainfall events, which were also compared each other. The results derived are as follows. First, when applying the method of moments for the parameter estimation, it was found necessary to consider the correlation coefficient between the two variables as well as the mean and variance. Second, the method of maximum likelihood was better to reproduce the mean, but the method of moments was better to reproduce the annual variation of the variance. Third, The annual maximum rainfall events derived were very similar in both cases. Among differently selected annual maximum rainfall events, those with the higher rainfall amount were selected by the method of maximum likelihood, but those with the higher rainfall intensity by the method of moments.

Keywords : freund bivariate exponential distribution, parameter estimation, method of maximum likelihood, method of moments, annual maximum rainfall event

요 지

본 연구에서는 연최대치 독립 호우사상의 결정에 사용되는 Freund 이변량 지수분포의 매개변수 추정과정을 구체적으로 검토하였다. 먼저, 모멘트법을 이용하는 경우를 구체적으로 검토하고, 그 결과를 최우도법을 적용한 결과와 비교하였다. 두 방법을 1961~2010년 서울지점의 시강우 자료에 적용하여 연최대치 독립 호우사상을 선정하고, 그 결과를 비교·검토하였다. 이러한 과정을 통해 얻은 결과는 다음과 같다. 첫째, 매개변수 추정방법으로 모멘트법을 적용하는 경우에는 두 변량의 평균과 분산뿐만 아니라 상관계수도 고려해 주어야 하는 것으로 나타났다. 둘째, 최우도법은 두 변량의 평균에 대한 재현성이 우수하고, 모멘트법은 분산의 경년변동을 잘 나타내는 것으로 나타났다. 셋째, 모멘트법과 최우도법을 통해 선정된 연최대치 독립 호우사상들은 대체로 유사한 것으로 나타났다. 다르게 선정된 호우사상은 최우도법의 경우에는 총 강우량이 큰 것, 모멘트법의 경우에는 강우강도가 큰 것으로 나타났다.

핵심용어 : Freund 이변량 지수분포, 매개변수 추정, 최우도법, 모멘트법, 연최대치 호우사상

* 고려대학교 공과대학 건축사회환경공학부 석사과정 (e-mail: impromn@empal.com)

** 교신저자, 고려대학교 공과대학 건축사회환경공학부 교수·공학박사 (e-mail: envchul@korea.ac.kr)

Corresponding Author, Prof., School of Civil, Environmental and Architectural Engineering, College of Engineering, Korea University, Seoul 136-713, Korea

1. 서론

전통적인 수문학적 강우분석은 실제 발생된 독립 호우 사상 자체의 특성을 고려한 분석이 아니라 주어진 지속기간 동안 기록된 강우량만을 분석하는 방법이다(건설교통부, 2010). 따라서 이와 같은 방법을 통해 실제 발생된 여러 호우사상의 재현기간을 명확히 설명하기란 쉽지 않다. 예를 들어, 20 시간동안 200 mm의 강우량을 발생시킨 호우사상과 30 시간동안 250 mm의 강우량을 발생시킨 호우사상 중 어느 경우가 더욱 큰 호우사상인지가 명확하지 않은 것이다. 당연히 주어진 지속기간에 대한 연최대치 강우 계열이 실제로 연최대치 독립 호우사상 계열과 어떤 관계에 있는지도 명확하지 않다.

이상과 같은 기존 강우분석 방법의 한계를 극복하기 위한 방법이 실제 발생된 독립 호우사상 자체의 특성을 분석하는 것이다. 즉, 독립 호우사상의 구성요소인 지속기간(시점과 종점)과 총강우량, 그리고 평균 강우강도 등을 고려하여 해당 호우사상의 재현기간을 결정하는 방법이다. 이와 같은 관점에서 도입된 것이 바로 이변량 빈도해석이다. 즉, 기존과 같이 총강우량만을 고려하는 단변량 빈도해석이 아닌 독립 호우사상 자체의 특성을 고려하고자 하는 것이다. 이미 국외에서는 이에 대한 많은 연구가 진행되어 왔으며(Yue, 2000; Yue and Rasmussen 2002; Kao and Govindaraju, 2007), 국내에서도 이와 관련된 여러 연구사례를 찾아 볼 수 있다(권영문과 김태웅, 2009; 박민규 등, 2010; 유철상과 박민규, 2011; Lee et al., 2010).

그러나 이변량 빈도해석의 대상이 되는 호우사상의 결정은 쉽지 않은 문제이다. 이는 호우사상의 크기를 판별할 수 있는 기준이 명확하지 않기 때문이다. 호우사상은 강우 지속기간, 총강우량, 평균 강우강도 및 침투 강우강도 등의 여러 가지 구성요소로 이루어져 있다. 따라서 어떤 하나의 요소만을 고려해 연최대치 독립 호우사상을 판단하는 것은 명확하지 않다. 예를 들어, 다른 요소들의 크기를 배제하고 호우사상의 총강우량이 큰 것을 연최대치 독립 호우사상으로 결정하는 경우, 지속기간은 매우 길고, 강우강도는 매우 작은 호우사상이 연최대치 호우사상으로 선정될 가능성이 있다. 평균 강우강도를 기준으로 할 경우, 지속기간이 매우 짧은 호우사상이 연최대치 호우사상이 될 가능성이 있다. 당연히 연최대치 호우사상의 결정에도 개개의 독립 호우사상을 비교할 필요성이 있다. 가장 보편적인 방법은 각 호우사상의 재현기간 평가하는 것이며, 이를 통해 특정 연도에 대한 연최대치 독립 호우사상을 결정할 수 있다. 연최대치 독립 호우사상계열 구

축에 이변량 빈도해석을 적용한 연구사례로 박민규와 유철상(2011)이 있다. 이 연구에서는 Freund (1961)가 제시한 이변량 지수분포를 이용하여 매 연도별로 발생된 독립 호우사상의 재현기간을 각각 산정하고, 그 중 가장 큰 재현기간을 갖는 호우사상을 해당 연도의 연최대치 독립 호우사상으로 선정한 바 있다.

본 연구에서는 연최대치 독립 호우사상의 결정에 사용되는 Freund 이변량 지수분포의 매개변수 추정과정을 구체적으로 검토해보고자 한다. 이를 위해 1961-2010년 서울지점의 시강우 자료를 대상으로 독립 호우사상을 선정하고, 연최대치 호우사상의 결정에 사용되는 Freund 이변량 지수분포의 매개변수 특성을 살펴보았다. 모멘트법을 이용한 매개변수 추정과정을 여러 경우에 대해 구체적으로 검토하였으며, 그 결과를 최우도법을 적용한 결과와 비교하였다. 마지막으로 모멘트법 및 최우도법을 적용하여 유도된 연최대치 독립 호우사상을 비교·검토하였다.

2. 이변량 지수분포를 이용한 독립 호우사상의 재현기간 산정

2.1 Freund 이변량 지수분포

Freund 이변량 지수분포를 따르는 두 변량 X_1 및 X_2 의 확률밀도함수는 다음과 같이 표현된다.

$$P_{X_i}(x_i) = \alpha_i e^{-\alpha_i x_i}, \quad x_i > 0, \quad \alpha_i > 0 (i=1,2) \quad (1)$$

여기서, $\alpha_i (i=1,2)$ 는 각 변량간의 관계를 독립으로 간주하여 얻은 위 지수분포의 매개변수이며, 이는 Eq. (1)로 표현되는 확률밀도함수의 형태에 영향을 미치게 된다. 즉, α_i 가 비교적 작은 경우, 독립변수의 증가에 따라 위 확률밀도함수는 완만하게 수렴하는 형태를 갖는 반면, α_i 가 커짐에 따라서 급격히 수렴하는 형태를 나타내게 된다. 그러나 호우사상을 구성하는 두 변량 X_1 및 X_2 의 경우, 둘은 서로 상관성을 가지고 영향을 미치게 되므로 두 변량은 서로 종속이 된다. 이 경우에 두 변량 X_1 과 X_2 의 결합밀도함수(joint density function)는 다음과 같이 정의된다(Freund, 1961).

$$\begin{aligned} P_{X_1, X_2}(x_1, x_2) &= \alpha_1 \alpha_2' e^{-\alpha_2' x_2 - x_1(\alpha_1 + \alpha_2 - \alpha_2')}, \quad 0 \leq x_1 < x_2 \\ &= \alpha_1' \alpha_2 e^{-\alpha_1' x_1 - (x_2 + \alpha_2 - \alpha_1')x_1}, \quad 0 \leq x_2 < x_1 \end{aligned} \quad (2)$$

위 식에서 $\alpha_i' (i=1,2)$ 는 두 변량이 종속인 상태에서 위 결합밀도함수의 형태를 결정하는 매개변수이다. 이러한 매개변수의 특성은 Leurgans et al. (1982)이 제시한 Fig.

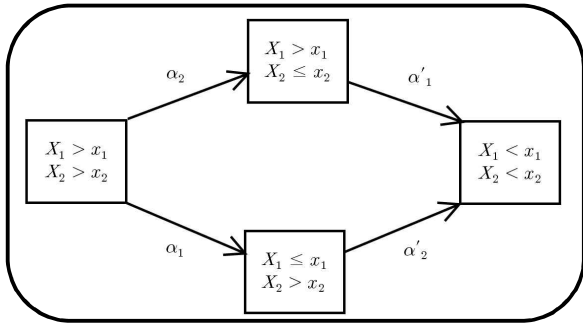


Fig. 1. Flow chart of parameters influencing on the probability density function (Leurgans et al., 1982)

1을 통해 쉽게 설명이 가능하다. 즉, 두 변량이 독립인 경우, $X_i \leq x_i (i=1,2)$ 가 되는 확률밀도함수에 영향을 미치는 매개변수는 $\alpha_i (i=1,2)$ 이다. 그러나 두 변량이 종속인 경우에는 $X_i < x_i (i=1,2)$ 가 되는 확률밀도함수에 영향을 미치는 매개변수가 α_i 에서 α'_i 로 변화하게 되는 것이다. 이는 위 확률밀도함수를 유도하기 위해 Freund (1961)가 정의한 α'_i 의 의미이기도 하다. 여기서, α'_1, α'_2 는 각각 $0 \leq x_2 < x_1$ 인 경우 및 $0 \leq x_1 < x_2$ 인 경우에 해당하는 확률밀도함수에 영향을 미치는 매개변수이다.

Freund 이변량 지수분포의 결합생존함수(joint survival function)는 다음과 같이 표현된다.

$$\begin{aligned} \bar{F}(x_1, x_2) &= \Pr[X_1 > x_1, X_2 > x_2] \quad (3) \\ &= \frac{1}{\gamma_2} \left\{ \alpha_1 e^{-\gamma_2 x_1 - \alpha_2' x_2} + (\alpha_2 - \alpha_2') e^{-(\alpha_1 + \alpha_2) x_2} \right\}, \quad 0 \leq x_1 \leq x_2 \\ &= \frac{1}{\gamma_1} \left\{ \alpha_2 e^{-\gamma_1 x_2 - \alpha_1' x_1} + (\alpha_1 - \alpha_1') e^{-(\alpha_1 + \alpha_2) x_1} \right\}, \quad 0 \leq x_2 \leq x_1 \end{aligned}$$

위 식에서 $\gamma_i = \alpha_1 + \alpha_2 - \alpha_i' (i=1,2)$ 이다. 아울러 $\alpha_1 + \alpha_2 \neq \alpha_i' (i=1,2)$ 인 경우, 변량 X_i 의 주변밀도함수(marginal density function)는 다음과 같이 표현된다.

$$P_{X_i}(x_i) = \frac{1}{\alpha_1 + \alpha_2 - \alpha_i'} \left\{ (\alpha_i - \alpha_i') (\alpha_1 + \alpha_2) e^{-(\alpha_1 + \alpha_2) x_i} + \alpha_i' \alpha_{3-i} e^{-\alpha_i' x_i} \right\}, \quad x_i \geq 0 \quad (4)$$

이상과 같이 정의된 결합밀도함수에 대한 결합적률생성함수(joint moment generating function)는 다음과 같이 표현되고(Kotz at al., 2000),

$$E[e^{t_1 X_1 + t_2 X_2}] = \frac{1}{\alpha_1 + \alpha_2 - t_1 - t_2} \left\{ \alpha_2 \left(1 - \frac{t_1}{\alpha_1'} \right)^{-1} + \alpha_1 \left(1 - \frac{t_2}{\alpha_2'} \right)^{-1} \right\} \quad (5)$$

위 식을 이용하여 나타낸 두 변량의 평균과 분산, 그리고 상관계수는 다음과 같이 유도된다(Kotz at al., 2000).

$$E[X_1] = \frac{\alpha_1' + \alpha_2}{\alpha_1' (\alpha_1 + \alpha_2)} \quad (6)$$

$$E[X_2] = \frac{\alpha_2' + \alpha_1}{\alpha_2' (\alpha_1 + \alpha_2)} \quad (7)$$

$$var[X_1] = \frac{\alpha_1'^2 + 2\alpha_1\alpha_2 + \alpha_2^2}{\{\alpha_1' (\alpha_1 + \alpha_2)\}^2} \quad (8)$$

$$var[X_2] = \frac{\alpha_2'^2 + 2\alpha_1\alpha_2 + \alpha_1^2}{\{\alpha_2' (\alpha_1 + \alpha_2)\}^2} \quad (9)$$

$$corr(X_1, X_2) = \frac{\alpha_1' \alpha_2' - \alpha_1 \alpha_2}{\sqrt{(\alpha_1'^2 + 2\alpha_1\alpha_2 + \alpha_2^2)(\alpha_2'^2 + 2\alpha_1\alpha_2 + \alpha_1^2)}} \quad (10)$$

2.2 매개변수 추정방법

이변량 지수분포의 매개변수 추정방법으로 모멘트법(method of moments)과 최우도법(method of maximum likelihood)을 이용할 수 있다. 먼저, 모멘트법은 표본자료의 통계치를 모집단의 통계치와 일치시킴으로서 모집단의 매개변수를 추정하는 방법이다. 이변량 지수분포의 경우, 총 4가지의 매개변수를 포함하고 있으며, 따라서 최소 4가지 이상의 통계치를 고려하여야만 적절한 매개변수 추정이 가능하다. 참고로, 고려하는 두 변량의 값 차이가 크게 되면, 큰 값을 갖는 변량에 매개변수의 결정이 의존하게 되므로, 두 변량의 크기를 유사하게 맞출 필요가 있다. 통상 각 변량의 평균으로 나누어 표준화된 값을 사용하게 된다.

최우도법을 이용한 매개변수의 추정은 다음과 같은 우도함수(likelihood function)를 이용한다(Freund, 1961).

$$\begin{aligned} L &= (\alpha_1 \alpha_2')^{R_1} (\alpha_1' \alpha_2)^{R_2} \exp \left[-\alpha_1' \sum_{X_i > Y_i} X_i - \alpha_2' \sum_{X_i < Y_i} Y_i \right. \\ &\quad \left. - (\alpha_1 + \alpha_2 - \alpha_1') \sum_{X_i > Y_i} Y_i - (\alpha_1 + \alpha_2 - \alpha_2') \sum_{X_i < Y_i} X_i \right] \quad (11) \end{aligned}$$

위 식에서 X_i 및 Y_i 는 각각 i 번째 호우사상의 변량 X 및 Y 이고, $R_i (i=1,2)$ 는 호우사상을 구성하는 특정 변량의 표준화된 크기가 다른 변량의 표준화된 크기보다 작은 경우의 호우사상 수이다. 이 경우에 $R_1 + R_2 = n$ 이 성립하며, 각 매개변수에 대한 최우추정량은 다음과 같이 유도된다(Kotz at al., 2000).

$$\hat{a}_1 = \frac{R_1}{\sum_{X_i < Y_i} X_i + \sum_{X_i > Y_i} Y_i} \quad (12)$$

$$\hat{a}_2 = \frac{R_2}{\sum_{X_i < Y_i} X_i + \sum_{X_i > Y_i} Y_i} \quad (12)$$

$$\hat{a}_1' = \frac{R_2}{\sum_{X_i > Y_i} (X_i - Y_i)} \quad (13)$$

$$\hat{a}_2' = \frac{R_1}{\sum_{X_i < Y_i} (Y_i - X_i)} \quad (14)$$

2.3 재현기간 산정방법

이변량 빈도해석 결과로 얻어지는 재현기간을 결합재현기간(joint return periods)이라고 하며, 정의방법에 따라서 두 가지의 형태가 가장 잘 알려져 있다(Yue and Rasmussen, 2002; Michele et al., 2005). 첫 번째 형태는, 두 변량 중 하나의 경우라도 임계값 이상이거나 두 변량 모두가 임계값 이상이 되는 호우사상이 발생할 재현기간(T_{or})이다. 두 번째 형태는, 두 변량 모두 임계값을 초과하는 호우사상이 발생할 재현기간(T_{and})으로 정의된다. 이와 같은 두 가지 형태의 재현기간은 다음과 같이 표현된다.

$$T_{or} = \frac{1}{\mu \times (1 - F_{X_1, X_2}(x_1, x_2))} \quad (15)$$

$$T_{and} = \frac{1}{\mu \times (1 + F_{X_1, X_2}(x_1, x_2) - F_{X_1}(x_1) - F_{X_2}(x_2))} \quad (16)$$

여기서 $F_{X_i}(x)$ 는 변량 X_i 에 대한 주변누가확률분포, $F_{X_1, X_2}(x_1, x_2)$ 는 $\Pr[X_1 \leq x_1, X_2 \leq x_2]$ 로 정의되며, μ 는 전기간에 대한 연평균 호우발생횟수를 의미한다.

기존의 단변량 빈도해석 결과를 통해 추정된 재현기간을 T 라고 하면, 위 제시된 두 가지 형태의 결합재현기간과 T 의 크기 관계는 일반적으로 $T_{or} < T < T_{and}$ 인 것으로 알려져 있다(박민규, 2009). 즉, T_{or} 에 근거하여 추정된 설계수문량은 기존의 단변량 빈도해석을 통해 추정된 설계수문량에 비해 과소 추정될 가능성이 크다. 이는 호우

사상을 구성하는 두 변량 중에서 하나의 변량이라도 특정 재현기간을 만족시키면, 그 때의 호우사상을 설계호우사상으로 선정하고, 이를 통해 설계수문량을 추정하기 때문이다. 그러나 T_{and} 에 근거하여 선정한 설계호우사상은 두 변량 모두가 특정 재현기간을 초과해야 하므로, 이 설계호우사상을 통해 추정된 설계수문량은 기존의 단변량 빈도해석 결과에 비해 크게 추정되게 된다.

따라서 T_{or} 에 근거하여 연최대치 독립 호우사상계열을 구축하는 것은 위험부담이 있다. 즉, 작은 강우강도로 오랜 기간 동안 내린 호우사상 또는 매우 짧은 기간 동안 큰 강우강도를 보인 호우사상의 재현기간이 크게 평가될 경우가 많기 때문이다. 따라서 본 연구에서는 연최대치 독립 호우사상을 결정하기 위한 결합재현기간으로 T_{and} 를 적용하였다.

3. 대상 자료

3.1 원자료의 특성

본 연구에서는 1961년부터 2010년까지의 서울지점 시강우 자료를 대상으로 하였다(Fig. 2). 서울지점의 경우 다른 지점들에 비해서 비교적 장기간의 강우자료가 축적되어 있다. 이 자료에 대한 기본 통계특성은 Table 1에 정리하였다.

3.2 독립 호우사상의 결정

독립 호우사상을 구분하기 위해서는 먼저 적절한 IETD (InterEventTime Definition) 및 절단값(threshold)을 결정

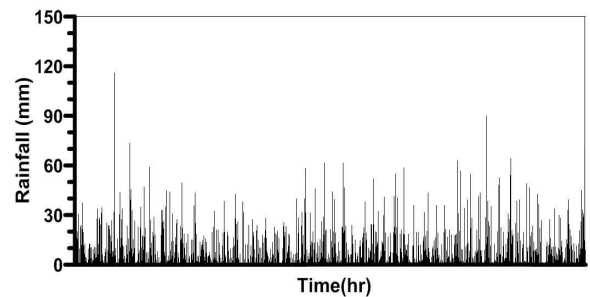


Fig. 2. Hourly rainfall time series in Seoul from 1961 to 2010 (rainy season)

Table 1. Basic Statistics of Hourly Rainfall in Seoul from 1961 to 2010 (rainy season)

Average (mm/hr)	Standard deviation (mm/hr)	Coefficient of skewness	Coefficient of kurtosis	Maximum value (mm/hr)
0.3	2.1	13.4	283.6	116

하여야 한다. IETD의 결정을 위해 Restrepo-Posada and Eagleson (1982)이 제안한 방법을 적용할 경우, 국내의 경우에는 10시간 또는 12시간의 IETD가 적절한 것으로 알려져 있기는 하다(이동률과 정상만, 1992; 김남원, 1998; 권재호 등, 2004). 그러나 본 연구에서는 IETD 및 절단값의 증가에 따른 독립 호우사상의 빈도수 변화를 살펴보고(Fig. 3), 이를 통해 적절한 IETD 및 절단값을 결정해보았다.

Fig. 3(a)를 살펴보면, IETD를 짧게 적용한 경우에 독립 호우사상의 빈도수가 매우 크나, IETD의 증가에 따라 그 빈도수가 급격히 감소하는 것으로 나타났다. 위 Fig. 3(a)에서 변곡점의 위치를 대략적으로 추정해보면, 대략 8~12시간 정도로 파악된다. 이러한 결과는 IETD의 증가에 따른 독립 호우사상의 빈도수 감소량을 총 빈도수 감소량에 대한 비율로 나타낸 Fig. 3(b)를 통해 보다 명확히 확인된다. 이러한 결과는 국내의 경우에 대해 적절한 IETD를 결정하고자 했던 연구들과도 유사한 값이기도 하다.

본 연구에서는 독립 호우사상의 선정에 IETD 뿐만 아니라 절단값을 함께 적용하였다. 큰 절단값을 적용하는 경우에는 작은 IETD를 적용하는 경우와 마찬가지로 단일 호우사상으로 판단되었던 호우사상들이 두개 이상의 호우사상으로 분리될 가능성이 커지게 된다. 그러나 절단값과 IETD 사이에는 일종의 상쇄작용(trade-off)이 있어 그 결정에 신중할 필요가 있다. 본 연구에서 고려한 절단값은 0~2mm이다. 이 절단값의 범위에 IETD 8~12시간을 함께 고려하여 최종적으로 IETD로는 12시간을, 절단값으로는 1mm를 결정하였다. 이상과 같이 결정한 IETD

(12시간) 및 절단값(1 mm)을 적용하여 선정된 전체 독립 호우사상의 구성요소별(지속기간, 총강우량, 평균 강우강도) 통계 특성은 Table 2와 같다.

3.3 대상 변량의 선정

이변량 지수분포를 적용하기 위해서는 호우사상을 구성하는 여러 요소들(지속기간, 총강우량, 평균 강우강도) 중에서 두 가지를 대상 변량으로 선정하여야 한다. 본 연구와 같이 이변량 지수분포를 적용할 경우 총강우량, 지속기간 및 강우강도 중 어떤 구성요소를 대상변량으로 선정하여도 크게 문제가 되지 않는다고(Zhang and Singh, 2007). 조건에 맞는 이변량 지수분포를 사용하면 되기 때문이다. 그러나 강우 지속기간의 경우 그 분포형태가 두개의 최빈값(bi-modality)을 갖는 경향이 있어 잘 고려하지 않는 편이다(박민규와 유철상, 2011). 아울러 강우 지속기간과 평균 강우강도와 같이 음의 상관성(Table 3)을 갖는 요소를 대상변량으로 하는 경우에는 양의 상관성을 갖거나 또는 두 변량을 독립으로 가정한 경우에 비해서 추정된 확률수문량이 과소하게 된다(Cordova and Rodriguez-Iturbe, 1985; Goel et al., 2000; 권영문과 김태웅, 2009). 따라서 본 연구에서는 강우 지속기간을 배제하고 총강우량과 평균 강우강도를 대상 변량으로 선정하였다.

4. 이변량 지수분포의 적용 결과

Freund 이변량 지수분포의 매개변수 추정에는 본 연구

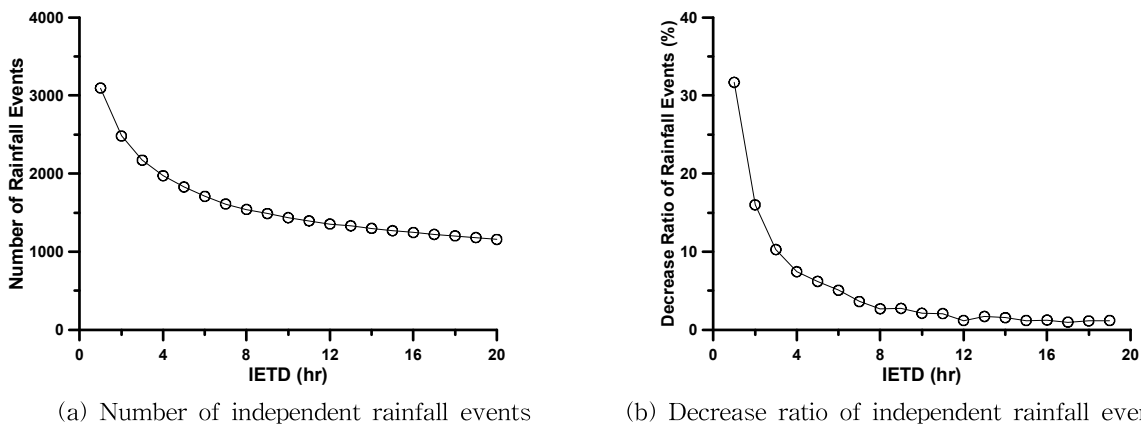


Fig. 3. Change in Frequency of Independent Rainfall Events with Respect to IETD

Table 2. Statistics for Independent Rainfall Events (IETD: 12 hour, threshold: 1 mm)

	Duration (hr)	Total rainfall (mm)	Rainfall intensity (mm/hr)
Average	10.8	36.2	3.8
Standard deviation	12.7	54.7	3.9

에서 선정한 두 변량(총강우량, 평균 강우강도)의 크기를 평균으로 나누어 표준화시킨 자료를 사용하였다. 아울러 매개변수의 추정에는 매 연도별로 수행하였다. 참고로 전 기간에 대해서 또는 매 연도별의 매개변수 추정이 가능하나, 강우과정의 변동성을 고려하는 경우라면 후자의 방법이 보다 합리적일 것으로 확인되었다(박민규와 유철상, 2011). 본 연구에서는 먼저 한 개 년도에 대한 매개변수 추정과정을 구체적으로 살펴보고, 그 결과를 나머지 년도에도 일괄 적용하였다.

4.1 1961년 발생한 호우사상의 통계적 특성

본 연구에서는 가용한 총 50개(1961~2010년) 년도의 자료 중 먼저 1961년도에 발생한 호우사상을 분석하였다. 1961년에 발생한 전체 독립 호우사상의 구성요소별 기본 통계치를 정리하면 Table 4와 같다. 강우강도는 3.1 mm/hr, 총강우량은 28.9 mm으로 전체 자료기간에 해당하는 평균 호우사상 보다는 약간 작은 것으로 파악된다. 1961년에 발생한 독립 호우사상의 강우강도와 지속기간을 박스(box)형태로 나타내면 Fig. 4와 같다.

4.2 모멘트법을 통한 매개변수 추정

본 연구에서는 먼저 모멘트법을 이용하여 이변량 지수

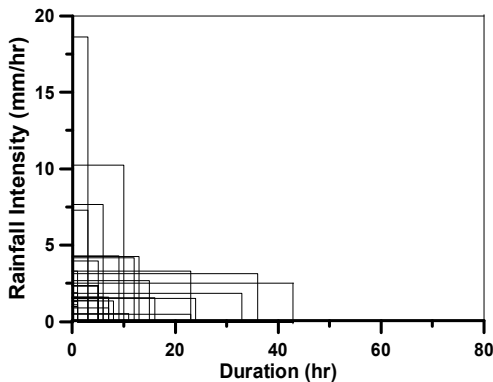


Fig. 4. Independent rainfall events occurred in 1961 (33 events)

Table 3. Correlation Coefficients between Components of Independent Rainfall Events

Duration and total rainfall	Duration and rainfall intensity	Total rainfall and rainfall intensity
0.74	-0.10	0.32

Table 4. Basic Statistics of Independent Rainfall Events in 1961

	Duration (hr)	Total rainfall (mm)	Rainfall intensity (mm/hr)
Average	10.4	28.9	3.1
Standard deviation	11.0	33.0	3.5

분포의 매개변수를 추정하였다. 모멘트법을 이용하는 경우, Eqs. (6)~(10)과 같이 제시된 5가지의 식을 이용하여 매개변수를 추정할 수 있다. 매개변수의 수가 4개 이므로 이 중 4개의 식을 선택하여 사용할 수도 있고 또는 5개 식을 모두 고려할 수도 있다. 그러나 주어진 식 자체가 비선형이므로 그 해는 일종의 최적화 기법에 의존할 수밖에 없다. 본 연구에서는 Levenberg-Marquardt 근사법을 이용하여 매개변수를 결정하였다(Madsen et al., 2004).

앞서 언급한 바와 같이 가용한 이론적 통계치는 5개이고 추정이 필요한 매개변수의 수는 4개이다. 따라서 매개변수의 추정에 다양한 선택의 여지가 발생하게 된다. 가장 간단한 것은 총 4개의 식을 이용하는 경우로서 통상 Eqs. (6)~(9)를 이용하게 된다. 두 변량간의 상관관계를 고려하지 않는 경우이다. 또 하나는 식 5개를 모두 고려하는 경우이고, 추가로 각 통계치의 규모차이를 고려하여 (특히 상관계수) 적용하는 방법이 있을 수 있다. 본 연구에서는 이런 경우들에 따라 매개변수의 추정결과가 어떻게 달라지는지를 평가하였다. 각각의 경우에 대한 목적함수를 다음과 같이 표현할 수 있다.

먼저, 4개의 통계치를 고려하는 경우(Case 1)에는 그 목적함수가 다음과 같이 나타난다.

$$\min \left\{ \frac{E[X_1] - ave_1}{ave_1} \right\}^2 + \left\{ \frac{E[X_2] - ave_2}{ave_2} \right\}^2 + \left\{ \frac{var[X_1] - var_1}{var_1} \right\}^2 + \left\{ \frac{var[X_2] - var_2}{var_2} \right\}^2 \quad (17)$$

위 식은 물론 각 변량의 통계치로 표준화 한 경우이나 실제 적용 전에 자료를 평균으로 나누어 이미 표준화하였기 때문에 큰 의미는 없다. 두 번째의 경우(Case 2)는

$$\min \{E[X_1] - ave_1\}^2 + \{E[X_2] - ave_2\}^2 + \{var[X_1] - var_1\}^2 + \{var[X_2] - var_2\}^2 + \{corr[X_1, X_2] - corr\}^2 \quad (18)$$

마지막으로 세 번째의 경우(Case 3)는 다음과 같다.

$$\min \left\{ \frac{E[X_1] - ave_1}{ave_1} \right\}^2 + \left\{ \frac{E[X_2] - ave_2}{ave_2} \right\}^2 + \left\{ \frac{var[X_1] - var_1}{var_1} \right\}^2 + \left\{ \frac{var[X_2] - var_2}{var_2} \right\}^2 + \left\{ \frac{corr[X_1, X_2] - corr}{corr} \right\}^2 \quad (19)$$

위 세 식에서 $ave_1, ave_2, var_1, var_2, corr$ 은 각각 관측자료로부터 산정한 변량 X_1 및 X_2 의 평균과 분산, 그리고 상관계수이다. 본 연구에서는 이상과 같은 여러 경우에 대해서 매개변수를 각각 추정하고, 그 결과를 비교·검토하였다. 그 결과는 Table 5와 같다.

위 결과를 보면 매개변수 추정에 고려되는 통계치에 따라 또한 표준화의 유무에 따라 그 결과가 상이하게 나타난다는 것을 알 수 있다. 특히, 두 변량 사이의 상관관계를 고려한 경우인 Case 2 및 Case 3과 이를 배제한 경우인 Case 1의 결과가 매우 다르게 나타났다. Case 2와 Case 3은 표준화의 유무 차이를 나타내는 경우이나, 실제로는 이미 표준화된 자료를 사용하여 평균과 분산을 단위 크기로 맞춘 상태이므로, 오직 상관계수의 크기를 절대적으로 고려하느냐 또는 상대적으로 고려하느냐의 차이를 나타낼 뿐이다. 추정된 매개변수의 차이는 이 차이에서 기인한 것이고, 실제 그 영향이 아주 크지는 않은 것으로 나타난다. 그러나 위 결과만을 가지고는 어느 경우가 가장 적절한 경우인지를 판단하기는 쉽지 않다. 즉, Case 1은 두 변량의 평균과 분산을 적절히 반영하는 매개변수일 가능성이 크나, 상관계수를 또한 적절히 반영하게 될지는 명확하지 않다. 반면에 Case 2와 Case 3의 경우에는 Case 1과 다르게 상관계수까지를 반영하는 매개변수일 가능성이 크나 두 변량의 평균과 분산을 또한 적절히 반영하게 될지는 분명하지 않다. 이를 확인하기 위해 관측자료의 통계치(두 변

량의 평균과 분산, 그리고 상관계수)와 각 경우에서 추정된 매개변수를 통해 산정한 통계치를 비교하였다(Table 6).

Table 6을 살펴보면, 모든 경우에서의 매개변수가 두 변량의 평균과 분산은 적절히 반영하고 있는 것으로 나타났다. Case 1은 물론 평균과 분산만으로 매개변수를 추정한 경우이므로 이 통계치들에 대한 재현 정도가 매우 높으며, Case 2 및 Case 3은 상대적으로 조금 떨어지는 것으로 확인된다. 그러나 상관계수를 배제하고 매개변수를 추정한 경우인 Case 1은 관측자료의 상관성을 전혀 반영하지 못하는 것으로 나타났다. 이에 반해, Case 2와 Case 3은 상관계수에 대한 재현성이 Case 1보다 훨씬 우수하나, 그렇다고 만족할 만한 수준인 것으로 보이지는 않는다. Case 3은 Case 2에 비해 상관계수에 대한 고려비중이 커 역시 상관계수의 재현성도 약간 나은 것으로 나타나고 있다.

이상의 결과들을 종합해 보면, Freund 이변량 지수분포의 매개변수 추정방법으로 모멘트법을 적용하는 경우에는 두 변량의 평균과 분산뿐만 아니라 상관계수를 고려하여야 함을 알 수 있다. 아울러 Eq. (19)와 같이 각 항을 표준화하여 사용하는 것이 매개변수의 추정에 보다 유리함을 확인할 수 있었다. 최종적으로 본 연구에서는 모멘트법을 대표하는 매개변수 추정방법으로 Case 3을 선택하였다.

4.3 최우도법을 통한 매개변수 추정

본 절에서는 최우도법을 통해 추정된 매개변수와 앞서 선택한 모멘트법을 통해 추정된 매개변수를 비교·검토하였다. 이러한 검토는 전체 자료기간에 대해 매해 수행하였다. 먼저, 추정된 매개변수를 비교하면 다음 Fig. 5와 같다.

Fig. 5에서 살펴볼 수 있는 것처럼, 모멘트법과 최우도법을 통해 추정된 매개변수는 큰 차이가 있다. 가장 중요한 차이점은 모멘트법의 경우 매개변수의 경년변화가 매

Table 5. Comparison of Parameter Estimation Results (1961 year)

	α_1	α_2	α'_1	α'_2
Case 1	1.2782	1.2682	0.8199	0.8264
Case 2	0.8008	0.7901	0.9879	1.0004
Case 3	0.7941	0.7845	1.0141	1.0259

Table 6. Comparison of Statistics of Observed and Calculated by Estimated Parameters (1961 year)

	$E[X_1]$	$E[X_2]$	$var[X_1]$	$var[X_2]$	$corr[X_1, X_2]$
Observed	1.0000	1.0000	1.2669	1.2550	0.4460
Case 1	1.0001	1.0001	1.2669	1.2552	-0.1703
Case 2	1.1312	1.1317	1.1601	1.1478	0.1231
Case 3	1.1235	1.1238	1.1276	1.1167	0.1434

우 큰 반면 최우도법의 경우에는 그렇지 않다는 점이다. 이는 모든 매개변수에서 발견되는 공통점이다. 따라서 모멘트법을 적용하는 경우 당해년의 호우사상 특성에 보다 민감한 추정치를 제공하고, 반대로 최우도법의 경우에는

상대적으로 안정적인 매개변수 추정 결과를 준다고 생각할 수 있다. 그러나 이 두 경우 중 무엇이 보다 적절한지를 이 결과로 판단하기는 힘들다.

다음 Fig. 6은 모멘트법 및 최우도법을 적용하여 추정

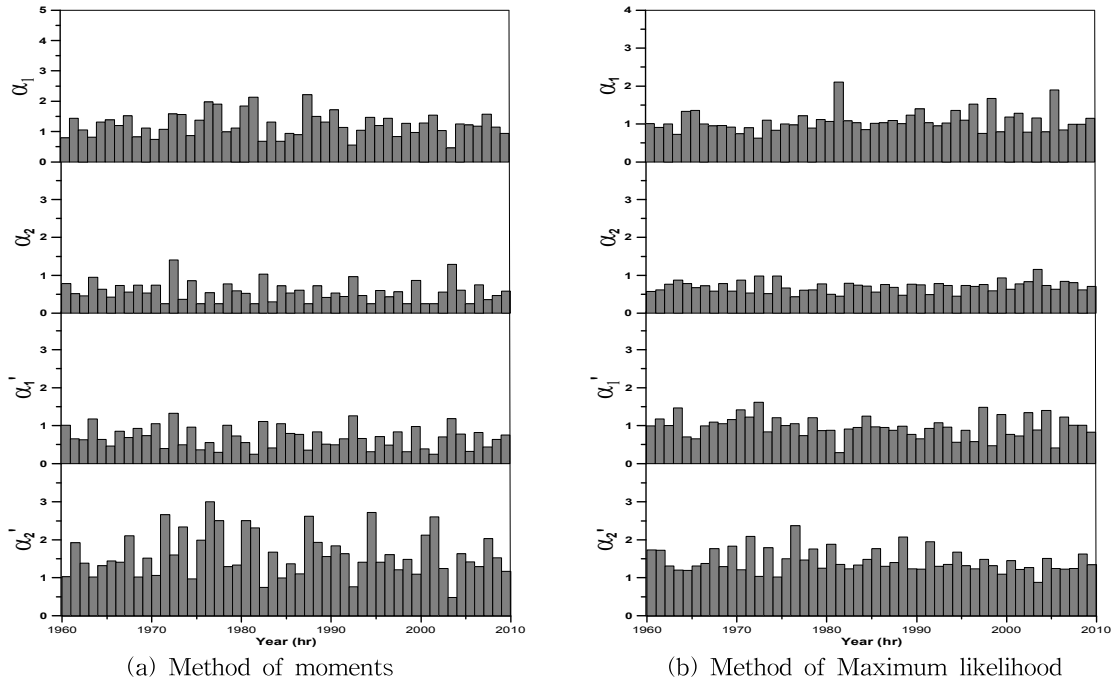


Fig. 5. Comparison of Parameters Estimated by the Method of Moments and the Method of Maximum Likelihood

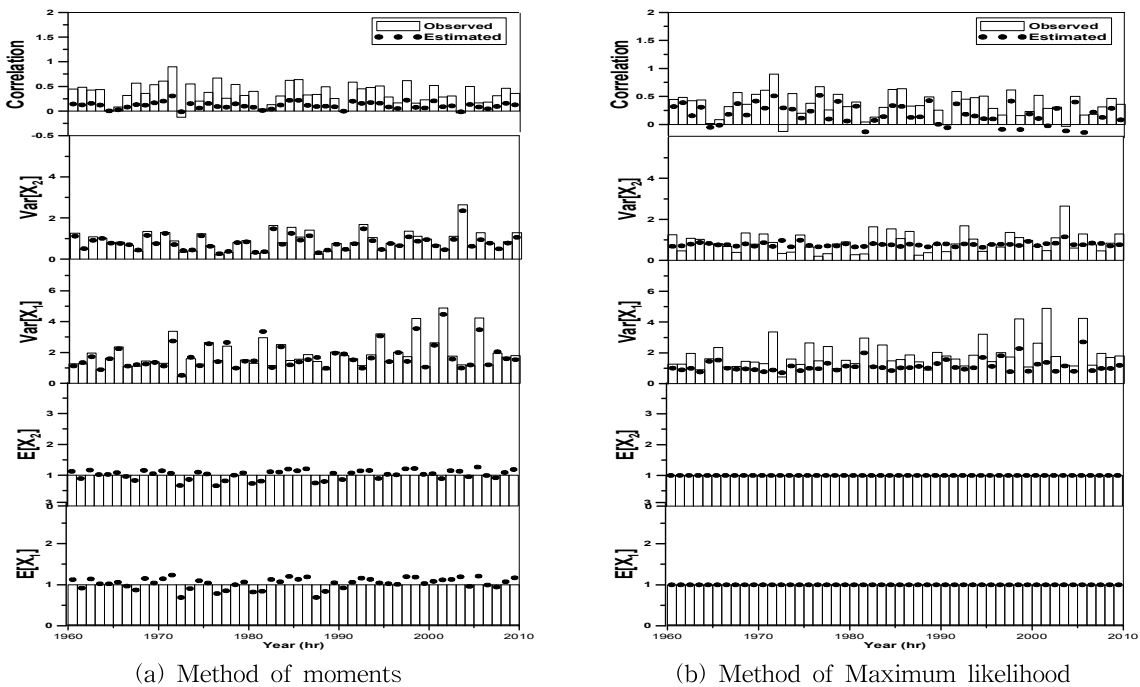


Fig. 6. Comparison of Statistics of Observed and Calculated by Estimated Parameters

한 매개변수를 가지고 구한 5가지의 통계치(두 변량의 평균과 분산, 그리고 상관계수)를 관측자료를 이용하여 구한 통계치와 비교한 것이다.

이 Fig.에서는 모멘트법을 통해 추정된 매개변수와 최우도법을 통해 추정된 매개변수의 특성이 뚜렷하게 차이남을 확인할 수 있다. 평균과 상관계수의 경우에는 최우도법에 의한 추정치가 우수한 것으로 나타난다. 그러나 분산의 경우에는 모멘트법을 이용하는 경우가 더 우수한 것으로 보인다. 보다 구체적으로 살펴보면, 먼저, 최우도법은 평균에 대한 재현성은 완벽하나, 분산의 경우에는 경년변동을 잘 파악하지 못하는 단점을 보인다. 상관계수의 경우는 그 경년변동을 비교적 잘 재현하는 것으로 나타난다. 이에 반해 모멘트법의 경우, 평균의 경우는 대체로 잘 재현하고 있고, 분산의 경우에도 최우도법에 비해 확실히 우수한 경년변동의 재현능력을 보인다. 그러나 상관계수의 경우에는 최우도법의 그것에 비해 떨어지는 결과를 보이고 있다. 따라서 전체적으로는 최우도법이 더 낫아 보이나, 특히 호우사상의 변동성을 반영한다는 측면에서는 모멘트법에도 장점이 있어 보인다.

4.4 선정된 연최대치 독립 호우사상의 비교

본 연구에서는 앞서 결정된 매개변수와 T_{and} 개념에 기초한 재현기간을 적용하여 연도별로 발생된 모든 독립 호우사상의 재현기간을 산정하였다. 이와 같이 결정된 재현기간을 기준으로 매년 최대의 재현기간을 갖는 독립 호우사상을 해당 연도의 연최대치 독립 호우사상으로 결정하였다. 이러한 과정을 통해 구축된 연최대치 독립 호우사상계열은 Fig. 7과 같다. Fig. 7에는 모멘트법과 최우도법을 이용한 결과를 함께 도시하여 두 경우의 비교가 가능하도록 하였다. 아울러, 구축된 연최대치 독립 호우사상을 강우강도 및 지속기간의 박스 형태로 중첩하여 나타내면 Fig. 8과 같다.

Figs. 7 and 8에서 살펴볼 수 있는 것처럼, 모멘트법과 최우도법을 통해 선정된 연최대치 독립 호우사상들은 매우 유사하다. 실제로 전체 년도(50년) 중에서 47개 년도에 대한 연최대치 독립 호우사상이 동일하게 선택되었다. 이는 전체의 95%로 연최대치 독립호우사상을 결정하는 두 방법에 큰 차이가 없음을 나타내는 결과로 이해할 수 있기도 하다. 이렇게 구축된 연최대치 독립 호우사상 계열의 구성요소별 평균을 비교하면 Table 7과 같다. 이 표에

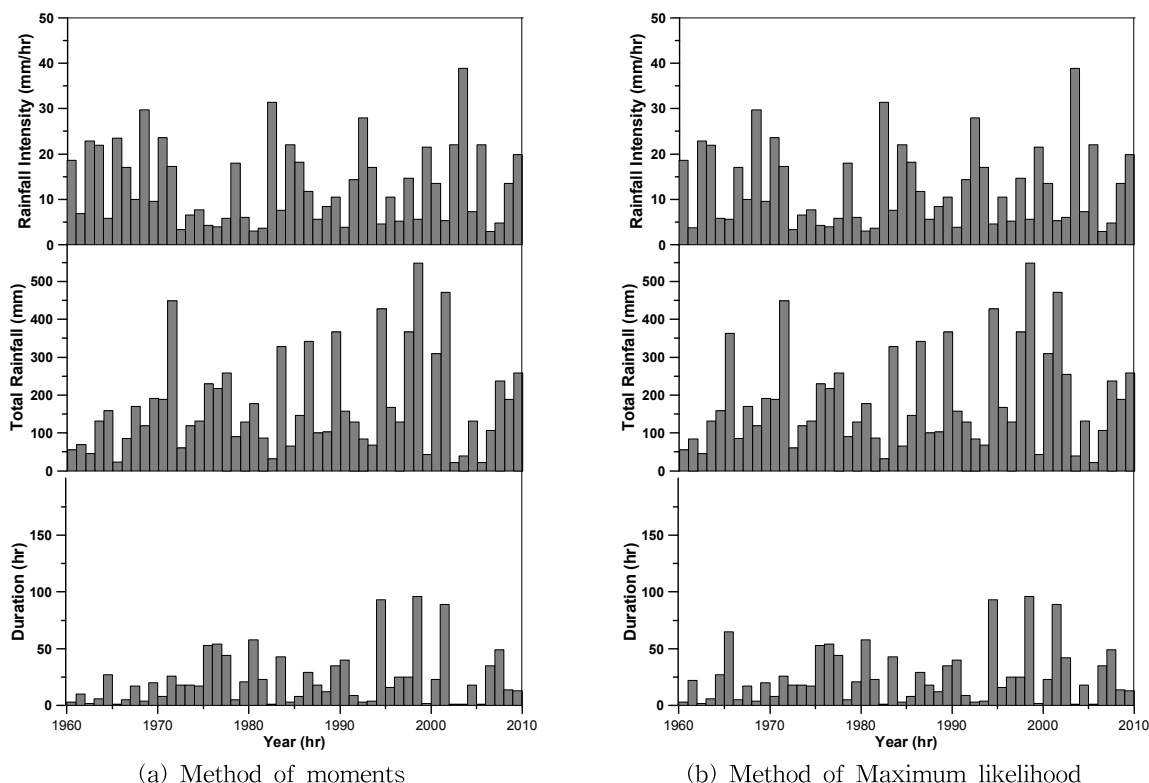


Fig. 7. Comparison of Characteristics for the Annual Maximum Independent Rainfall Events Selected by Applying the Method of Moments and the Method of Maximum Likelihood

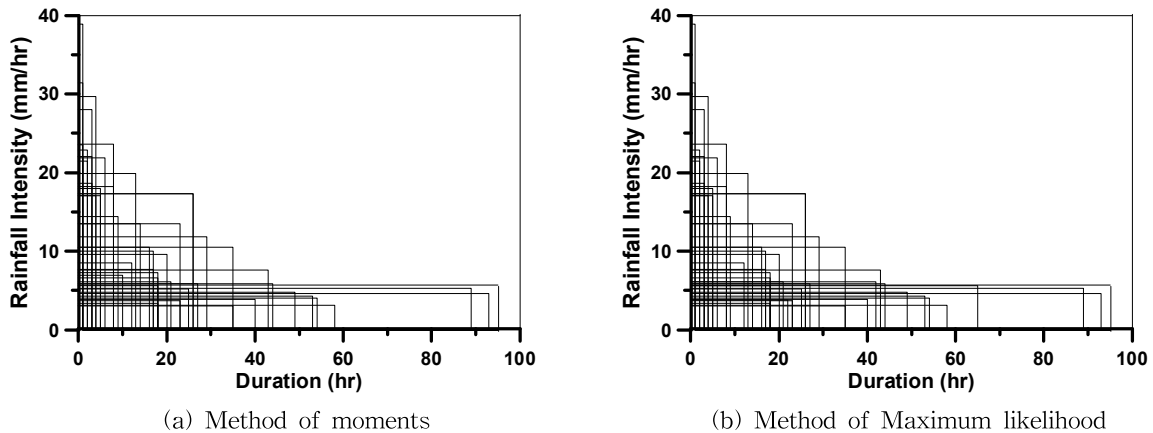


Fig. 8. Characteristics of Average Rainfall Intensity and Duration for Annual Maximum Independent Rainfall Events Selected by Applying the Method of Moments and the Method of Maximum Likelihood

Table 7. Comparison of Annual Maximum Independent Rainfall Events Selected by Applying the Method of Moments and the Method of Maximum Likelihood

	Duration (hr)	Total rainfall (mm)	Rainfall intensity (mm/hr)
Method of moments	22.9	171.6	13.2
Method of Maximum likelihood	25.3	183.4	12.5

나타난 차이는 결국 3개의 다른 호우사상으로 부터 발생한 것이다. 최우도법을 적용한 경우에는 총 강우량이 큰 것이, 모멘트법을 적용한 경우에는 강우강도가 큰 것이 선택된 것으로 파악되었다.

이상과 같은 두 가지 방법 중에서 모멘트법의 적용시에는 다양한 모멘트의 조합을 고려하여 의도한 결과가 유도되는지를 확인해야 하며, 이러한 과정이 수행되기 어려운 경우에는 최우도법을 적용하는 것이 안전한 선택이 될 것으로 보인다.

5. 결론

본 연구에서는 연최대치 독립 호우사상의 결정에 사용되는 Freund 이변량 지수분포의 매개변수 추정과정을 구체적으로 검토하였다. 먼저, 이를 위해 1961~2010년 서울 지점의 시강우 자료를 대상으로 독립 호우사상을 선정하고, 연최대치 호우사상의 결정에 사용되는 Freund 이변량 지수분포의 매개변수 특성을 살펴보았다. 모멘트법을 이용한 매개변수 추정과정을 여러 경우에 대해 구체적으로 검토하였으며, 그 결과를 최우도법을 적용한 결과와 비교하였다. 마지막으로 모멘트법 및 최우도법을 적용하여 유도된 연최대치 독립 호우사상을 비교·검토하였다. 그

결과를 정리하면 다음과 같다.

첫째, 매개변수 추정방법으로 모멘트법을 적용하는 경우에는 두 변량의 평균과 분산뿐만 아니라 상관계수 또한 고려해야 하는 것이 필요한 것으로 나타났다. 아울러 각각의 통계치를 동일한 비율로 고려해 주는 것이 그렇지 않은 경우보다 우수한 것으로 확인되었다.

둘째, 모멘트법을 통해 추정된 매개변수와 최우도법을 통해 추정된 매개변수의 특성이 뚜렷하게 차이남을 확인할 수 있었다. 먼저, 최우도법은 평균에 대한 재현성은 완벽하나, 분산의 경우 경년변동을 잘 파악하지 못하는 단점을 보인다. 상관계수의 경우는 그 경년변동을 비교적 잘 재현하는 것으로 나타난다. 이에 반해 모멘트법의 경우, 평균의 경우는 대체로 잘 재현하고 있고, 분산의 경우에도 최우도법에 비해 확실히 우수한 경년변동의 재현능력을 보인다.

셋째, 모멘트법과 최우도법을 통해 선정한 연최대치 독립 호우사상들은 매우 유사하게 나타났다. 실제로 전체 50개년 중에서 47개 년도에 대한 연최대치 독립 호우사상이 동일하게 선택되었다. 다르게 선택된 3개의 호우사상은 최우도법의 경우에는 총 강우량이 큰 것이었고, 모멘트법의 경우에는 강우강도가 큰 것으로 확인되었다.

감사의 글

본 연구는 2008년도 정부(교육과학기술부)의 재원으로 한국연구재단(No. KRF-2008-313-D01083) 및 2010년도 한국연구재단(No. 2010-0014566)의 지원을 받아 수행된 연구입니다.

참고문헌

- 건설교통부(2010). 설계홍수량 산정 선진화 기획 연구보고서.
- 권영문, 김태웅(2009). “이변량 강우 빈도해석을 이용한 서울 지역 I-D-F 곡선 유도.” **대한토목학회논문집**, 대한토목학회, 제29권, 제2B호, pp. 155-162.
- 권재호, 박무중, 김중훈(2004). “비점오염원 산정을 위한 강우 분석.” **한국수자원학회 2004 학술발표회**, 한국수자원학회, pp. 666-670.
- 김남원(1998). 강우의 시·공간 분포특성: 점 강우모형 매개변수 추정. 한국건설기술연구원.
- 박민규(2009). 극치 호우사상의 확률적 발생특성 및 기후변화에 따른 변동성. 박사학위논문, 고려대학교.
- 박민규, 유철상(2011). “독립호우사상의 확률론적 해석: 1. 연최대 호우사상 계열의 작성.” **한국방재학회지**, 한국방재학회, 제11권, 제2호, pp. 127-136.
- 박민규, 유철상, 김현준(2010). “연최대 호우사상 계열을 이용한 측우기자료 및 현대자료의 비교.” **대한토목학회논문집**, 대한토목학회, 제30권, 제2B호, pp. 137-147.
- 유철상, 박민규(2011). “독립호우사상의 확률론적 해석: 2. 호우사상의 재현기간.” **한국방재학회지**, 한국방재학회, 제11권, 제2호, pp. 137-146.
- 이동률, 정상만(1992). “한강유역 강우의 시·공간적 특성.” **한국수자원학회지**, 한국수자원학회, 제25권, 제4호, pp. 75-85.
- Cordova, J.R., and Rodriguez-Iturbe, I. (1985). “On probabilistic structure of storm surface runoff.” *Water Resources Research*, Vol. 21, No. 5, pp. 755-763.
- Goel, N.K., Kurothe, R.S., Mathur, B.S., and Vogel, R.M. (2000). “A derived flood frequency distribution for correlated rainfall intensity and duration.” *Journal of Hydrology*, Vol. 228, pp. 56-67.
- Freund, J.E. (1961). “A bivariate extension of the exponential distribution.” *American Statistical Association Journal*, Vol. 56, No. 296, pp. 971-977.
- Kao, S-C., and Govindaraju, R.S. (2007). “A bivariate frequency analysis of extreme rainfall with implications for design.” *Journal of Geophysical Research Atmospheres*, Vol. 112, D13119.
- Kotz, S., Balakrishnan, N., and Johnson, N.L. (2000). *Continuous Multivariate Distributions Volume 1: Models and Applications*, John Wiley & Sons, INC.
- Lee, C.H., Kim, T-W., Chung, G., Choi, M., and Yoo, C. (2010). “Application of bivariate frequency analysis to the derivation of rainfall-frequency curves.” *Stochastic Environmental Research and Risk Assessment*, Vol. 24, pp. 389-397.
- Leurgans, S., Tsai, T.W.Y., and Crowley, J. (1982). “Multivariate distributions for the life lengths of components of a system sharing a common environment.” *Journal of Applied Probability*, Vol. 23, pp. 418-431.
- Madsen, K., Nielsen, H.B., and Tingleff, O. (2004). *Methods for non-linear least squares problems. Informatics and mathematical modelling*. Technical University of Denmark.
- Michele, C.D., Salvadori, G., Canossi, M., Petaccia, A., and Rosso, R. (2005). “Bivariate Statistical Approach to Check Adequacy of Dam Spillway.” *Journal of Hydrologic Engineering*, Vol. 10, No. 1, pp. 50-57.
- Restrepo-Posada, P.J., and Eagleson, P.S. (1982). “Identification of Independent Rainstorms.” *Journal of Hydrology*, Vol. 55, pp. 303-319.
- Yue, S. (2000). “The Gumbel mixed model applied to storm frequency analysis.” *Water Resources Management*, Vol. 14, pp. 377-389.
- Yue, S., and Rasmussen, P. (2002). “Bivariate frequency analysis: discussion of some useful concepts in hydrological application.” *Hydrological Processes*, Vol. 16, pp. 2881-2898.
- Zhang, L., and Singh, V.P. (2007). “Bivariate rainfall frequency distributions using Archimedean copulas.” *Journal of Hydrology*, Vol. 332, pp. 93-109.

논문번호: 11-104	접수: 2011.09.07
수정일자: 2011.11.17/12.21	심사완료: 2011.12.21