이변량 빈도해석을 위한 Archimedean Copula
Archimedean Copula for bivariate Frequency Analysis

성장현*, 백희정, 권원태***, 김영오******

Jang Hyun Sung, Hee-Jeong Baek, Won tae Kwon, Young-Oh Kim

요 지

수문설계 인자인 확률홍수량 산정시 짧은 홍수량 자료 길이로 인해 홍수량을 직접 이용하기 보다는 강우자료와 강우-유출모형에 의존하고 있는 현시점에서 무엇보다 중요한 것은 신뢰할 만한 확률강우량이 산정되어야 한다는 것이다. 하지만 지금까지의 강우빈도해석(rainfall frequency analysis)은 강도(intensity), 지속기간(duration), 깊이(depth) 사이의 연관성은 고려하지 않은 단편적인 방법론에 그치고 있다. 즉, 강우를 표현하는 인자들 간 독립(independency)이라는 가정을 거친 후, 간단한 단변량(univariate) 강우빈도분포(rainfall frequency distribution)로 확률강우량을 산정하고 있다는 것이다. 간단한 모형에 따른 이점은 있으나 최근의 강우 형태는 매우 복잡한 양상을 띠고 있어, 단변량 강우빈도분포로 이를 대표하기에는 무리가 따른다. 따라서 본 연구에서는 강우빈도해석의 인자가 독립적이며 정규분포(normal distribution)라 가정하지 않고, 세 개의 주변분포(marginal distribution)의 형태가 같지 않다는 점, 또한 가정하지 않는 방법론 중, 그 가치를 널리 인정받고 있는 Archimedean Copula (AC)에 대한 연구를 수행하였다. AC를 이용하여 강도, 지속기간, 깊이 사이의 종속성 중, 두 가지 변량을 고려한 이변량(bivariate) 강우빈도해석을 수행하였고 그 효용성을 검토해 보았다.

핵심용어: 강도, 지속기간, 깊이, Archimedean Copula, 이변량 강우빈도해석

1. 서 론

그 동안의 강우빈도해석에서는 단편적으로 지속시간별 총 강우량을 대상으로 하는 단변량 빈도해석 방법이 주로 사용되어져 왔으나 이러한 방법으로는 최근의 복잡한 강우를 현실적으로 묘사하기가 힘들다. 또한 확률홍수량 산정에 있어 전적으로 강우자료에 의존하는 우리나라의 현실에서는 무엇보다도 가능한 노력을 집중하여 타당한 강우빈도곡선을 구하는 것이 최우선이라고 할 수 있겠다. 빈도해석 선진국 격인 미국이나 영국 등지의 유럽에서는 주로 홍수빈도해석에 많은 관심을 집중하고 있으며 더불어 다양한 강우빈도해석 기법에 대한 최신 이론을 개발하고 있는데 2000년대 후반부터는 강우의 경우 홍수와는 달리 자료 확장을 위한 연구보다는 다변량 해석을 중심으로 한 빈도곡선의 정밀묘사에 많은 노력을 기울이고 있다. 따라서, 본 연구에서는 다변량 빈도해석 기법 중, 그 장점을 인정받고 있는 AC를 적용하여 빈도곡선을 구하고 다양한 시선으로 검토해 보았다.

* 정회원, 국립기상연구소 기후연구과 연구원 · E-mail: jhsung1@korea.kr
** 국립기상연구소 기후연구과 연구관 · E-mail: heejbaek@korea.kr
*** 국립기상연구소 기후연구과 과장 · E-mail: wontk@korea.kr
**** 서울대학교 건설환경공학부 부교수 · E-mail: yokim05@snu.ac.kr

2. 핵심이론

2.1 Copula의 특징

Cordova and Rodriguez-Iturbe (1985)는 강도와 지속기간의 상관성이 지표유출에 유효한 영향을 준다는 것을 확인하였으며 Singh과 Singh (1991)은 강도와 깊이 사이를 결합분포를 묘사할 수 있는 지수분포를 바탕으로 이변량 확률밀도함수(probability density function)를 제안한 바 있다. 본 연구에서 적용하게 될 여러 coupla 중, Gumbel-Hougaard Copula (G-HC)를 먼저 설명하면 아래와 같다.

변수 X_1, X_2와 X_3의 주변분포를 $u_1 = F_{X_1}(x_1)$, $u_2 = F_{X_2}(x_2)$, $u_3 = F_{X_3}(x_3)$라 생각하면 u_1, u_2와 u_3는 0과 1 사이의 일양분포(uniform distribution)를 따른다고 볼 수 있다. 이때 G-HC의 C_θ^3의 누가분포함수(cumulative distribution function)는 식 1과 같다.

$$C_\theta^3 = (u_1, u_2, u_3) = \exp(-[(-\ln u_1)^\theta + (-\ln u_2)^\theta + (-\ln u_3)^\theta]^{1/\theta})] \qquad \theta \geq 1 \qquad (1)$$

위의 식을 미분하여 얻은 확률밀도함수는 식 2와 같다.

$$\begin{aligned} C_\theta^3 &= \frac{\partial^3 C_\theta^3}{\partial u_1 \partial u_2 \partial u_3} \\ &= \frac{(-\ln u_1 \ln u_2 \ln u_3)^{\theta-1}}{u_1 u_2 u_3} \bullet \\ &\exp(-w^{1/\theta})[w^{3/\theta-3} + (3/\theta-3)^{2/\theta-3} + (\theta-1)(2/\theta-1)w^{1/\theta-3}] \end{aligned} \qquad (2)$$

여기서, $w = (-\ln u_1)^\theta + (-\ln u_2)^\theta + (-\ln u_3)^\theta$이다. 위 식 2와 같은 누가분포함수는 생성함수(generating function), $\varphi(t) = (-\ln t)^\theta$, $\varphi(0) = \infty$, $\varphi(1) = 0$으로 생성된다. 이변량 copula는 $[0,1] \rightarrow [1,0]^2$안에서 아래와 같은 특성들을 가지는 함수이다.

- $C(1,u) = v$, $C(u,0) = C(0,v) = 0$
- $C(u_1, u_2) + C(v_1, v_2) - C(u_1, v_2) - C(v_1, u_2) = 0$

  $u_1 > v_1$, $u_2 > v_2$, 이고 $u_1, u_2, v_1, v_2 \in [0,1]$

G-HC는 대표적인 AC이고, 본 연구에서는 G-HC 이외에 Clayton 및 Frank Copula (CC, FC)를 적용하였다. CC 및 FC를 간단히 실펴보면 아래와 같다.

$\theta \geq -1$, $\theta \neq 0$를 만족하는 θ에 대하여 CC는 식 3과 같다.

$$C_{clay}(x,y) = \max([x^{-\theta} + y^{-\theta} - 1]^{-1/\theta}, 0) \qquad (3)$$

CC의 대표적인 특징 중 하나는 왼쪽 꼬리의 종속성을 잘 설명한다는 점이다.

0이 아닌 실수 θ에 대하여 FC는 식 4와 같다.

$$C_{frank}(x,y) = -\frac{1}{\theta} ln(1 + \frac{(e^{-\theta x}-1)(e^{-\theta y}-1)}{(e^{-\theta}-1)}) \tag{4}$$

FC는 왼쪽과 오른쪽 꼬리의 종속성을 똑같이 증대시키는 장점을 가지고 있다.

2.2 매개변수 추정법

Copula함수의 매개변수를 추정하는 방법에는 Maximum Likelihood Method (MLM), Inference Functions for Margins (IFM) 그리고 Canonical Maximum Likelihood Method (CMLM)이 있다. 각 방법론은 아래에 서술하였다.

- MLM

결합밀도함수(f), 한계밀도함수(f_i) 그리고 copula 밀도함수(c) 사이의 관계는 다음과 같은 canonical representation으로 나타낼 수 있다(식 5).

$$f(x_1, x_2, ..., x_n) = c(F_1(x_1), F_2(x_2), ..., F_n(x_n)) \bullet \prod_{i=1}^{n} f_i(x_i) \tag{5}$$

$\aleph = \{x_{1t}, x_{2t}, ..., x_{nt}\}_{t=1}^{T}$를 표본 데이터 행렬이라고 하면 $\Theta = (\varphi_1, \varphi_2, ..., \varphi_n, \theta)$를 추정할 모수 벡터, φ_t를 한계분포함수 F_i의 모수 벡터, 그리고 θ를 copula 함수의 모수 벡터라 할 때 로그우도함수(log-likelihood function)는 다음과 같이 표현할 수 있다(식 6).

$$l(\Theta) = \sum_{t=1}^{T} \ln c(F_1(x_1^t;\varphi_1, ..., F_n(x_n^t;\varphi_n);\theta) + \sum_{t=1}^{T}\sum_{i=1}^{n} \ln f_i(x_i^t;\varphi_i) \tag{6}$$

최우추정량 $\hat{\Theta}$는 식 7과 같이 추정한다.

$$\hat{\Theta} = \text{argmax} l(\Theta) \tag{7}$$

- IFM

로그우도 함수에서 한계분포의 모수 종속구조 모수인 ϕ_i와 종속구소 모수인 copula 모수 θ를 분리할 수 있다는 사실에 착안하여 다음과 같이 2단계로 추정한다.

1단계 : 한계분포 F_i의 파라미터 ϕ_i를 최우추정한다(식 8).

$$\hat{\varphi}_i = \text{argmax} l^i(\varphi_i) = \text{argmax} \sum_{i=1}^{T} \ln f_i(x_i^t;\varphi_i) \tag{8}$$

2단계 : 주어진 $\hat{\phi}$하에서 copula 매개변수 θ를 최우추정한다(식 9).

$$\hat{\theta} = \text{argmax} l^c(\theta) = \text{argmax} \sum_{i=1}^{T} \ln C(F_i(x_i^t;\hat{\varphi}_i), ..., F_n(x_n^t;\hat{\varphi}_n);\theta) \tag{9}$$

- CMLM

MLM과 IFM은 단변량 한계분포에 대해 일정한 분포를 가정한다. 그러나 CMLM은 한계분포

의 분포형태를 미리 가정하지 않고 실증적 한계분포변환(empirical marginal transformation)의 개념에 의존한다. 이러한 변환은 미지의 모수적 한계분포 $F_i(\bullet)$을 다음과 같은 실증분포함수(empirical distribution function) $\hat{F}_i(\bullet)$로 근사시킨다. CMLM은 다음의 두 단계로 실행된다.

1단계 : 실증한계분포(empirical marginal distribution)를 이용하여 초기 자료집합 $X=(X_{1t},X_{2t},....,X_{nt})|_{t=1}^{T}$를 일양변량으로 변환시킨다(식 10).

$$\hat{F}_i(\bullet)=\frac{1}{T}\sum_{t=1}^{T}1_{X_{it}\leq,}\ \text{for}\ \ i=1,....,n \tag{10}$$

여기서, $1_{X_{it}\leq}$는 지시함수(indicator function)를 나타낸다.

2단계 : copula의 파라미터 θ는 식 11과 같이 추정한다.

$$\hat{\theta}=\mathrm{argmax}\sum_{t=1}^{T}\ln c(\hat{u}_1^t,\hat{u}_2^t,...,\hat{u}_n^t;\theta) \tag{11}$$

3. 핵심기법의 적용

서울지점의 시강우 자료의 깊이와 총량을 고려해서 아래와 같은 개념(식 12)으로 강우자료를 구성하였다.

$$V=\sum_{S=1}^{D_M}R_{MS} \tag{12a}$$

$$R=\frac{V}{D_M} \tag{12b}$$

여기서, D_M은 강우의 지속시간, R_{MS}는 S번째 시간의 강우강도, R은 지속시간 동안의 평균강우강도이다. 구성된 각 자료의 적합도 검정을 수행한 결과, 강도는 대수정규(log normal)분포, 깊이와 지속시간은 지수(exponential)분포에 적합하였다. 강우 변량간 상관정도는 아래 표 1과 같다.

표 1. 상관계수로 본 강우 변량간 상관정도

강도와 깊이	강도와 지속시간	깊이와 지속시간
0.45	- 0.40	0.70

위의 “2. 핵심이론”에서 언급된 방법론을 고려하여 세 가지 방법론에 MLM을 적용하여 AC 결과를 얻었으며 최종적으로 구해진 이변량 강우빈도곡선은 그림 1과 같다.

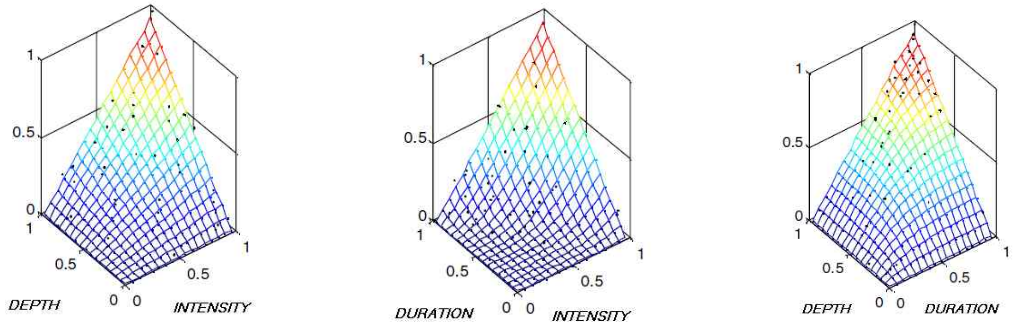

그림 1. 강우자료의 결합 누가분포함수

4. 결론

본 연구에서는 강우 변량간 종속성을 고려할 수 있는 AC를 적용하였다. AC에서도 우수한 결과를 인정받고 있던 G-HC, CC, FC을 통해 이변량 강우빈도곡선을 제시하였다. 본 연구에서는 이변량 강우빈도곡선 만을 제시하였지만 향후에는 본 결과와 단변량 강우빈도곡선을 자세하게 검토할 것이며 아울러, 다양한 극치사상(extreme event)의 결합(강우, 해수면 온도 등)도 시도해 볼 예정이다.

감사의 글

이 연구는 “NIMR-2010-B-2 (기후변화 예측기술 지원 및 활용연구)”의 지원으로 수행되었습니다.

참고문헌

1. Cordova, J. R., and Rodriguez-Iturbe, I. (1985). “On probabilistic structure of storm surface runoff.” Water Resour. Res., 215, pp. 755-763.
2. Singh, K., and Singh, V. P. (1991). “Derivation of bivariate probability density functions with exponential marginals.” Stchastic Hydrol. Hydraul., 5(1), pp. 55-68.