Copula 방법을 통한 강우 빈도 해석

Rainfall Frequency Analysis Based on the Copula Method

주경원* · 신주영 · 김수영*** · 허준행******

Kyung-Won Joo, Ju-Young Shin, Sooyoung Kim, Jun-Haeng Heo

요 지

강우사상은 강우량, 지속기간, 강우강도 등의 특성으로 표현될 수 있으며 이런 인자들을 같이 고려할수록 그 현상을 보다 종합적으로 표현할 수 있다. 하지만 현재 일반적으로 이루어지는 일변량 빈도해석절차에서는 지속기간을 고정시켜놓고 각 지속시간에 따른 결과만을 도출해 낼 수 있기 때문에 지속기간에 대해 제약적이고 입력자료에 존재하지 않는 지속기간에 대한 결과를 얻기가 어렵다. Copula모델은 두 일변량 분포형을 다변량 분포형으로 연결하여 주는 모델이다. 따라서 강우량과 지속기간을 변수로 사용하면 Copula모델을 통한 이변량 강우빈도해석은 보편적으로 이루어지고 있는 일변량 지점빈도해석보다 지속기간에 대해 유연한 결과를 나타낼 수 있다. 즉, 강우와 지속기간이 동시에 변수로 사용되기 때문에 임의의 지속기간이나 강우에 대해서 확률강우량 및 확률지속기간을 얻을 수 있다. 본 연구에서는 서울지점을 대상으로 1961~2009년 동안 발생한 강우사상 중 각 년도에서 최대강우량이 발생한 사상을 추출하여 입력자료로 사용하였다. Copula 모형은 Gumbel-Hougaard, Frank, Joe, Clayton, Galambos등 총 5개의 모델을 적용하였고 각 Copula의 매개변수는 준모수방법인 maximum pseudolikelihood estimator를 이용하여 추정하였다.

핵심용어 : Copula 모델, 빈도해석

1. 서론

확률강우량의 산정은 수공구조물의 설계기준에 있어 중요한 기준이 된다. 확률강우량은 통계학적인 빈도개념을 가지는 수문량으로, 취득가능한 모든 자료를 이용하여 모집단의 특성을 대변하는 분포형을 가정하고, 선정된 분포형을 이용하여 산정하게 된다(고연우 등, 2000). 하지만 일변량 빈도해석절차를 수행하는 경우 강우량만을 변량으로 사용하기 때문에 강우사상의 지속시간이나 강우강도등을 고려할 수 없다는 단점이 있다. 실제 강우의 물리적 특성을 더 잘 반영하려면 한 가지 변량으로는 그 특성을 파악하기가 어려우며 이에 따라 이변량, 다변량 빈도해석을 통한 확률강우량의 추정은 필수적이라 할 수 있었다. 일변량 빈도해석절차에서의 이러한 한계들을 뛰어넘기 위해 이변량 빈도해석을 적용한 여러 모델들이 개발되고 여러 연구들이 이루어지고 있다(권영문과 김태웅, 2009). 일반적으로 이루어지고 있는 이변량 빈도해석 모델로 bivariate normal, log-normal, gamma, 그리고 extreme-value distribution이 있다. 하지만 이런 모델들의 가장 큰 제약은 각 변량의 분포형이 같아야 한다는 점이다(Genest C. and Favre A-C., 2007). Copula 모델의 가장 큰 장점은 두 변량의 분포형이 같아야 한다는 제한을 받지 않으며 변량 각각의 최적

* 정회원 · 연세대학교 대학원 토목공학과 석사과정 · E-mail : kwjy1@yonsei.ac.kr
** 정회원 · 연세대학교 산업공학연구소 연구원 · E-mail : hyjyshin@gmail.com
*** 정회원 · 연세대학교 대학원 토목공학과 박사과정 · E-mail : sykim79@yonsei.ac.kr
**** 정회원 · 연세대학교 사회환경시스템공학부 토목환경공학과 교수 · E-mail : jhheo@yonsei.ac.kr

분포형을 적용하여 결합할 수 있다는 것에 있다. 본 연구에서는 서울의 기상청 유인관측소의 자료를 적용하였으며, 5가지의 Copula 모형을 적용하여 매개변수 추정 및 확률강우량 산정절차를 진행하였다

2. Copula model

본 연구에서는 5개의 Copula 모델(Gumbel-Hougaard(GUH), Frank(FRK), Joe(JOE), Clayton(CLN), Galambos(GLB))을 사용하였으며 매개변수 추정방법으로는 maximum pseudolikelihood estimator를 이용하였다.

2.1. Copula model

Copula는 두 집합 U_2 (U_i는 0부터 1까지의 값을 가지는 집합)의 결합분포형으로 식 (1)~(2)와 같이 정의된다.

$$H(x_1, x_2) = C[F_1(x_1), F_2(x_2)] \tag{1}$$

$$U_k = F_k(X_k),\ (k = 1, 2) \tag{2}$$

본 연구에서 사용된 Copula 모델의 형태와 매개변수 θ의 범위는 표 1과 같다.(Nelson RB., 2006; Frank MJ., 1979; Joe H., 1993; Galambos J., 1975)

표 1. Copula 모델

Copula 모델	수식	θ
Gumbel-Hougaard	$C_\theta(u,v) = \exp(-[(-\ln u)^\theta + (-\ln v)^\theta]^{1/\theta})$	$\theta \geq 1$
Frank	$C_\theta(u,v) = \dfrac{1}{\ln\theta} \ln\left[\dfrac{(\theta^u - 1)(\theta^v - 1)}{\theta - 1}\right]$	$\theta \geq 0$
Joe	$C_\theta(u,v) = 1 - [(1-u)^\theta + (1-v)^\theta - (1-u)^\theta(1-v)^\theta]^{1/\theta}$	$\theta \geq 1$
Clayton	$C_\theta(u,v) = (\max(u^{-\theta} + v^{-\theta} - 1), 0)^{-1/\theta}$	$\theta \geq -1$
Galambos	$C_\theta = uv\exp([(-\ln u)^{-\theta} + (-\ln v)^{-\theta}]^{-1/\theta})$	$\theta \geq 0$

2.2. 매개변수의 추정

본 연구에서는 maximum pseudolikelihood estimator를 이용하여 매개변수를 추정하였으며 식은 (3)과 같다.

$$l(\theta) = \sum_{i=1}^{n} \log\ c_\theta\left(\frac{R_i}{n+1}, \frac{S_i}{n+1}\right)\} \tag{3}$$

여기서, c_θ는 C_θ의 θ에 대해 미분된 식이고, R_i, S_i는 각 변량의 순위(rank)이다. $l(\theta)$를 최대로 하는 θ를 찾음으로서 매개변수를 추정할 수 있다. 본 연구에서는 최대로 하는 θ를 찾기 위해 Newton-Rhapson 법을 이용하였다.

3. 적용

3.1. 대상지점

입력자료로는 기상청의 서울지점 시강우 자료를 사용하였으며 자료의 개요는 표 2와 같다.

표 2. 입력 강우자료 개요

지점번호	지점명	시작년도	끝년도	자료년도수
000108	서울(Seoul)	1961	2009	49

3.2. 강우자료의 추출

Copula 모델을 이용한 이변량 빈도해석의 목적상 강우사상의 총강우량, 지속기간, 강우강도를 모두 변량으로 사용하는 방향으로 연구를 진행하였으며, 모든 강우사상을 추출하여 각 항목별로 연 최대치를 갖는 사상을 뽑아서 사용하였다.

3.3. 각 변량의 확률분포형 결정 및 매개변수 추정

연최대강우량 사상, 연최대강우강도 사상 자료에 대해서 각 변량의 적정확률분포형과 매개변수를 추정하였으며 결과는 표 3과 같다.

표 3. 각 변량의 확률분포형 및 매개변수 추정결과

기준최대치	변량	분포형	XLO	XSC	XSH
강우량	강우량	GLO	198.86	57.313	-0.289
	지속기간	GEV	43.731	25.547	0.091
강우강도	강우강도	GEV	9.682	5.559	-0.235
	지속기간	GEV	1.702	2.076	-0.687

강우량을 기준으로 추출한 연최대사상에서 강우량은 GLO분포형, GEV분포형을 가장 잘 따르는 것으로 나타났으며 강우강도를 기준으로 추출한 연최대사상에서는 강우강도, 지속기간 모두 GEV 분포를 가장 잘 따르는 것으로 나타났다.

3.4. Copula 모델의 매개변수 추정

다음 과정으로 5개의 Copula 모델에서 각각의 매개변수 를 찾기 위해 2.2절에서 소개한 maximum pseudolikelihood estimator를 이용하였으며 Newton Rhapson 방법을 이용하여 추정하였다. 강우량을 기준으로 연최대사상을 추출하여 강우량, 지속기간을 입력자료로 사용하였고 결과는 표 4와 같다.

표 4. Copula 모델의 매개변수 θ

Copula	Gumbel-Hougaard	Frank	Joe	Clayton	Galambos
θ	1.67630	4.28616	1.91942	1.03182	1.87197

3.5. 확률강우량

각 Copula 모델에 GEV(강우량), GLO(지속기간)분포형을 결합시켜 Copula 모델의 확률분포형을 구할 수 있다. 이변량 빈도해석의 특성 상 위와 같은 결과를 얻으면 임의의 지속기간이나 임의의 강우량에 대한 marginal distribution을 얻을 수 있으며 대표로 6시간, 48시간의 지속기간에 대한 확률강우량은 다음과 같다.

표 5. 6시간과 24시간에 대한 Copula 모델별 확률강우량

재현	6시간(mm)					48시간(mm)				
기간	GUH	FRK	JOE	CLN	GLB	GUH	FRK	JOE	CLN	GLB
2년	124.55	125.35	157.11	80.20	93.86	147.80	139.22	161.25	138.63	128.90
5년	178.63	172.02	215.87	114.08	129.52	203.91	189.09	220.81	198.15	166.51
10년	214.61	204.55	254.27	142.19	151.47	241.66	224.79	259.63	247.74	189.41
30년	272.62	263.84	315.09	197.25	183.58	303.53	292.33	321.14	344.80	223.31
50년	301.64	298.16	344.80	228.73	198.17	334.67	331.93	351.22	400.88	238.84
100년	343.80	354.43	387.24	279.86	218.11	380.14	397.20	394.18	490.65	260.48
200년	389.50	424.69	433.84	341.37	238.29	431.04	478.26	441.46	599.95	282.37
300년	459.65	542.25	500.74	441.04	266.29	506.39	612.88	509.31	782.13	313.52
500년	517.96	650.32	553.45	530.60	287.46	568.47	736.02	562.66	955.63	336.93

연최대강우량을 갖는 사상을 추출하다 보니 지속기간이 긴 사상들을 위주로만 입력자료로 사용하게 되어 짧은 지속기간보다는 긴 지속기간에서 더 신뢰할만한 확률강우량을 추정하였다. 분포형별로는 재현기간에 증가에 따라 Galambos 모델은 과소추정되는 경향이 있으며 Clayton 모델은 과대추정되는 경향을 보인다.

4. 결론

이변량 빈도해석을 하기 위해 기상청 서울 유인관측소 지점자료를 대상으로 Copula 모델을 이용하여 확률강우량을 산정하였다. 기존 일변량 빈도해석에서 이루어지던 지속기간별 최대강우량을 추출하는 방법 대신 강우를 사상별로 구분하여 입력자료로 사용하였다. Copula 모형은 5개의 모형을 적용하였으며 매개변수 추정은 maximum pseudolikelihood estimator를 이용하였다. 입력자료의 특성상 짧은 지속기간에서는 확률강우량을 잘 모의하지 못하는 것으로 나타나며 짧은 지속기간에 대한 연관성을 높일 수 있는 방법의 개발이 필요하다. 또한 한국 강우사상에 적합한 Copula 모델의 정립과 추후에 적합도 검정 등을 통하여 매개변수와 확률강우량의 신뢰성을 검정하고 확률강우량이 도출되기까지의 과정에 대한 검증 및 보완이 필요하다.

감 사 의 글

본 연구는 국토해양부 한국건설교통기술평가원의 2009 건설기술혁신사업인 ‘기후변화에 의한 수문 영향분석과 전망’ 과제에 의해 지원되었습니다.

참 고 문 헌

1. 고연우, 허준행, 김규호 (2000), 확률강우량 산정시 GEV분포형 및 Gumbel 분포형의 역전현상

2. 권영문과 김태웅 (2009), 이변량 강우 빈도해석을 이용한 서울지역 I-D-F 곡선 유도
3. Frank MJ., (1979), "On the simultaneous associativiry of (x,y) and $x+y-f(x,y)$.", Aequationes Math, No. 19, pp. 194-226
4. Galambos J., (1975), "Order statistics of samples from multivariate distributions" Journal of the American Statistical Association, No. 70, pp. 674-680
5. Genest C. and Favre A-C., (2007), "Everything you always wanted to know about copula modeling but were afraid to ask." Journal of hydrology, no. 12, issue. 4, pp. 347-368
6. Joe H., (1993), "Parametric families of multivariate distributions with given margins." Journal of Multivarite Analysis, No. 46, pp. 262-282
7. Nelson RB., (2006), An Introduction of Copulas, Springer-Verlag, New York