

빈도해석 방법에 의한 확률강우량 산정: 매개변수적 방법과 비매개변수적 방법

문영일¹⁾, ○차영일²⁾, 전시영³⁾

1. 서론

치수 또는 이수의 목적을 위한 수공구조물을 설계할 때 홍수량, 강우량 등의 수문자료에 대한 적당한 확률분포형을 선정하는 것은 경제적인 수공구조물의 계획과 설계에 있어서 매우 중요한 요소 중의 하나이다. 우리나라의 경우 일반적으로 강우량자료는 홍수량자료에 비해 관측이 용이하기 때문에 비교적 풍부한 자료의 축적이 이루어져 있다. 따라서 미계측 지역의 홍수량산정을 위하여 강우자료에 대한 빈도분석을 통하여 빈도강우량을 구한 후 이를 강우-유출모형의 입력자료로 이용하여 홍수량을 구하는 경향이 많다. 이러한 강우량 자료의 빈도해석방법에는 크게 매개변수적 빈도해석방법(parametric frequency analysis)과 비매개변수적 빈도해석방법(nonparametric frequency analysis)이 있다.

우리나라에서 사용되는 빈도해석방법은 대부분 매개변수적 빈도해석방법을 사용하고 있다. 주로 사용되는 분포들은 정규분포, 2변수 Log-Normal 분포, 3변수 Log-Normal 분포, 3변수 Gamma 분포, GEV 분포, Wakely분포, Gumbel분포, Pearson Type III분포, Log-Pearson Type III, Type-I 극치분포, Type-III 극치분포 등이다. 고전적으로 알려진 매개변수적 빈도해석 방법의 단점인 짧은 기록의 자료 및 왜곡된 자료에 대한 매개변수의 신뢰도는 확률가중모멘트법(Greenwood 등, 1979; Hosking, 1989), L-모멘트법(Hosking, 1990) 등 여러 가지 매개변수 결정 방법의 발달로 보완된 상태이다. 그러나, 매개변수적 빈도해석 방법을 사용할 때, 한 지점에 적합도 검정에 통과한 다수의 분포형이 선정되는 경우에 어떤 분포형을 사용하느냐 하는 문제점이 제기될 수 있다. 또한 수문자료가 여러 가지 원인으로 인하여 복합분포(mixed distribution) 형태를 가질 때, 매개변수적 해석방법은 bimodal을 갖는 확률밀도함수를 해석하는 데는 여러 가지 어려움이 따른다. 빈도해석에서 선정된 확률밀도함수는 사용된 관측자료의 분포특성을 정확히 설명하여야 신뢰있는 빈도해석 결과를 기대할 수 있다. 매개변수적 빈도해석의 가장 어려움 중에 하나가 특정 확률분포형을 자료의 모분포로 가정하여 임의적으로 빈도해석을 하는데 있다고 볼 수 있다. 한 지역에서 여러 개의 확률분포형이 적합도 판정(χ^2 검정, Kolmogorov-Smirnov 검정, Cramer von Mises검정, 편차검정 등)을 받고, 채택지 점수에 의한 한 지점의 최적 분포형을 선정되어도 몇 년 후에 관측기간이 증가되었을 때 같은 분포형이 선택된다는 보장이 없기 때문에 이러한 가정은 항상 타당한 것은 아니다.

비매개변수적 빈도해석 방법은 자료에 대하여 어떤 분포형의 가정이 필요 없이 본래 관측자료가 나타내는 분포형을 자료의 확률밀도함수로 간주한다. 따라서 분포형 선정의 어려움이 해소되고, bimodal을 갖는 분포의 경우도 자료의 성질이 잘 반영된다. 이러한 비매개변수적 해석 방법의 대표적인 것이 변동 핵밀도함수(Variable Kernel Density Function)방법이다. 지금까지 외국에서는 변동 핵밀도함수를 이용한 빈도해석 기법이 홍수량, 갈수량 및 강우량의 산정에 많이 적용(Guo, 1991; Lall 등, 1993; Moon 등, 1993; Moon과 Lall, 1994; Moon, 1996; Adamowski, 1996; Rajagopalan 등, 1996)되고 있다.

본 연구에서는 전국 기상청 산하 71개 관측소 중에서 각 유역을 대표 할 수 있도록 총 14개 관측소지점의 강우량 자료에 대하여 매개변수에 의한 확률강우량과 비매개변수에 의한 확률강

1) 서울시립대 토목공학과 조교수
2) 서울시립대 토목공학과 석사과정
3) 원광대학교 토목환경공학과 교수

우량을 추정하여 비교 분석하였으나 지면관계상 한강유역에 대한 결과만 제시하였다.

2. 빈도해석 방법

2.1. 매개변수적 빈도해석

매개변수적 빈도해석 방법은 아래 그림 1과 같이 수문자료에 기존의 여러 개 확률분포형을 적용한 후 적합도 검정을 통하여 채택된 분포형을 선택해 자료의 모분포로 가정하여 빈도분석하는 것이다.

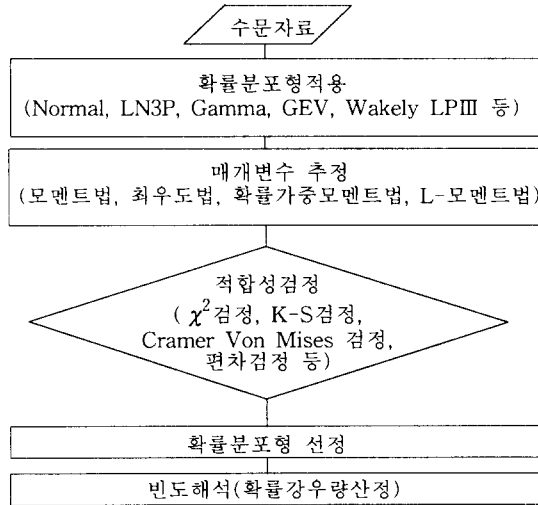


그림1. 매개변수적 빈도해석 방법.

매개변수적 빈도해석 방법은 가장 전통적인 분포형인 정규분포로부터 근래에 개발된 Wakeby, Logistic 분포에 이르기까지 매우 많은 종류의 확률분포형이 사용되고 있다. 각각의 확률분포형은 매개변수의 개수와 그 모양에 따라 서로 다른 특징을 가지기 때문에 어떤 확률분포가 가장 적절한 분포인지에 대한 의견은 다양하다. 또한 같은 분포형이라도 어떤 방법으로 매개변수를 추정하느냐에 따라 많은 차이가 있다.

본 연구에서는 매개변수적 빈도해석 방법으로 많이 이용되는 정규분포, 2변수 Log-Normal 분포, 3변수 Log-Normal 분포, 3변수 Gamma 분포, Log-Pearson typeIII, GEV1 분포를 사용하였다. 이들 6개 분포형의 매개변수 결정은 모집단의 모멘트와 표본자료의 모멘트를 같다고 가정한 모멘트법 및 확률가중모멘트법의 선형조합으로 표시한 L-모멘트법을 사용하였다. 적합도 검정으로는 확률밀도함수에 대하여서는 χ^2 검정과 누가분포함수에 대하여서는 Kolmogorov-Smirnov 검정 등을 유의수준 5% 내에서 검토하였다.

2.2. 비매개변수적 빈도해석

비매개변수적 확률밀도함수(probability density function) 구성 방법으로는 핵밀도함수(kernel density function)방법, orthogonal series, k^{th} nearest neighbor, maximum likelihood 와 히스토그램 등이 있으나, 빈도해석 방법으로 많이 사용되는 것은 이론적으로 가장 잘 발달된 핵밀도함수(kernel density function)에 의한 방법이다. 핵밀도함수 추정법의 기본 개념은 히스토그램에서 출발했다. 아마도 히스토그램이 가장 오래되고 쉽게 사용되는 확률밀도함수 추정법

일 것이다. 그러나 히스토그램의 단점은 계급구간이 변화는 점에서 불연속적이고 구간간격과 시작점의 선택에 따라 확률밀도함수의 모양이 달라지는 것이다. 이와 같은 히스토그램의 단점을 보완하여 Rosenblatt(1956)는 핵밀도함수 추정식을 모든 실수 x 에 대하여 아래와 같이 정의하였다.

$$f(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} K\left(\frac{x - X_i}{h}\right) \quad (1)$$

여기에서 X_1, X_2, \dots, X_n 은 독립적으로 동일하게 분포된 실관측치를 나타내며, $K(\cdot)$ 는 핵함수이고 h 는 n 이 무한대로 갈 때 0으로 접근하는 값을 갖는 양의 bandwidth 이다. 자료가 발생된 위치에 히스토그램의 box 대신 핵함수를 자료의 개수만큼 위치하도록 하여 구간을 이동시킬 수 있는 이동 히스토그램이 핵밀도함수 추정식 개념이다. 즉, 각 관측치에 폭 $2h$ 인 핵함수 $K(\cdot)$ 를 box대신 각 관측치에 놓고, 이 들을 합산하므로 식(1)이 만들어진다. 일반적으로 핵함수(kernel function)는 다음조건을 만족하여야 한다.

$$\int K(t)dt = 1, \text{ 여기서 } t = \frac{x - X_i}{h} \quad (2)$$

핵밀도함수(kernel density function)에 의한 빈도해석 방법에는 고정 핵밀도함수(fixed kernel density function) 방법과 변동 핵밀도함수(variable kernel density function) 방법이 있다. 식(1)에서 h 가 변화 없이 일정하면 고정 핵밀도함수(fixed kernel density function)라 하고, h 가 자료의 상황에 따라 변화하면 이를 변동 핵밀도함수(variable kernel density function)라 한다. 고정 핵밀도 추정법은 균대균대 길게 늘어진 자료분포에 적용될 경우 결점을 갖는다. 왜냐하면, 일정한 bandwidth, h 를 전체자료에 적용될 경우 추정치의 꼬리(tail) 부분에서 자료가 드물게 분포할 경우에 추정치의 값이 불연속 또는 거친 밀도함수의 모양을 갖기 때문이다. 따라서 고정 핵밀도함수에 의한 빈도해석보다 변동 핵밀도함수에 의한 빈도해석의 결과가 우수하다(Lall 등, 1993; Moon 등, 1993; Moon과 Lall, 1994). Breiman 등(1977)은 고정 핵밀도함수 추정법의 특성에다 자료의 지역적인 밀도를 고려하는 k^{th} nearest neighbor방법을 결합한 변동 핵밀도함수 추정법을 제안하여 자료가 발생된 위치에 놓여지는 핵함수의 폭이 자료의 밀도에 따라 변하게 다음과 같이 정의하였다.

$$f(x) = \frac{1}{nh} \int \frac{1}{d_{j,k}} K\left(\frac{x - X}{hd_{j,k}}\right) dx \quad (3)$$

여기서 k 를 양의 정수로 놓고, $d_{j,k}$ 를 한 개의 자료 x_j 에서 그 나머지 자료 $(n-1)$ 개 중에서 k 번째로 가까운 지점에 있는 자료까지의 거리다. 따라서 자료의 분포가 적은 낮은 밀도지역에서 $d_{j,k}$ 는 커지고 변동 핵함수의 모양은 넓게 퍼지게 된다. 자료의 분포가 많은 높은 밀도지역에서는 그 반대현상이 일어나, 변동 핵함수는 좁게 밀착된 형태를 가질 것이다. Bowman(1985), Moon과 Lall(1994)은 변동 핵밀도함수 추정법이 밀도함수의 꼬리 부분에서의 극치값 추정이나 또는 자료가 비대칭 분포일 때 장점이 많다는 것을 보여 주었다.

본 연구에서는 자료의 분포형태에 따라 잘 대응하는 변동 핵밀도함수를 사용하여 빈도해석을 하였고, Gaussian, Epanechnikov, Rectangular, Cauchy 등의 핵함수와 cross validated IMSE, maximum likelihood (Lall 등, 1993; Moon 등, 1993), Adamowski Criterion(Adamowski, 1981), Plug-in(Sheather와 Jones, 1991) 등의 방법을 사용하여 최적의 h 를 구하였다.

3. 적용 및 결과

본 연구에서는 전국 기상청 산하 71개 관측소 중에서 각 유역을 대표 할 수 있도록 총 14개 관측소지점(표1)의 강우량 자료에 대하여 일 최대치 계열, 월 최대치 계열 및 년 계열을 작성한 후, 각 계열에 대하여 총 6개의 매개변수에 의한 분포함수(normal 분포, 2변수 log-normal 분포, 3변수 log-normal 분포, 3변수 gamma 분포, log-Pearson typeIII, GEV1 분포)에 의한 확률강우량과 변동 핵밀도함수를 이용한 비매개변수에 의한 확률강우량을 추정하여 비교 분석하였다. 매개변수적 분포의 적합도 검정은 유의수준 0.05인 K-S검정과 χ^2 검정을 적용하여 이루어졌으며, 한강유역의 지점별 적합도 검정 결과 거의 모든 지점에서 적합한 것으로 판정되었으나,

서울지점의 일 최대강우량자료에서 normal 분포가 기각되는 것으로 판정되었다.

표1. 유역별 전국 14개 관측소지점의 강우자료.

유역	지점명	자료관측기간	자료크기	결측기간
한 강	춘천	1966~1996 (31년)	31	
	서울	1907~1996 (90년)	85	1907.1~1907.9, 1950.9~1953.11
	충주	1973~1996 (24년)	24	
낙동강	대구	1909~1996 (88년)	88	
	영주	1973~1996 (24년)	24	
	밀양	1973~1996 (24년)	24	
	산청	1973~1996 (24년)	24	
금강	대전	1969~1996 (28년)	25	
	추풍령	1949~1996 (48년)	45	1950.6~1950.12, 1951.7~1952.9
	부여	1973~1996 (24년)	24	
영산강	광주	1939~1996 (58년)	57	1939.1~1939.4
	목포	1906~1996 (91년)	91	
섬진강	임실	1973~1996 (24년)	24	
	승주	1973~1996 (24년)	24	

그림2~4에서는 한강유역의 3개 지점인 서울, 충주, 춘천 관측소 자료에 대한 일 최대강우량, 월 최대강우량, 년 강우량의 확률밀도함수의 형태를 각각 나타내고 있다. 서울관측소와 충주 관측소 자료는 bimodal 형태를 보여주고 있다. 분석된 14개 지점중 한강유역의 서울관측소의 강우량 자료를 사용하여 비매개변수적 변동 핵밀도함수(VK)에 의한 방법과 6개의 매개변수적 분포함수를 누가분포함수(cumulative distribution function, CDF)에 대하여 비교하였다. 그림5~7에서 보면, 변동 핵밀도함수(VK)의 CDF와 실측자료의 empirical CDF는 상당히 근접한 모습을 보여 주고 있다. 그러나, 대부분 매개변수적 분포함수의 CDF는 다른 거동의 모습을 보여 준다. 실측자료를 살펴보면 서울관측소의 강우량자료가 그림 2~4에서 bimodal의 모양을 갖는 것을 알 수 있다. 그림5와 6에서의 일 최대강우량 및 월 최대강우량 자료에 대한 매개변수적 분포함수의 결과는 관측자료보다 크게 추정되었고, 그림7의 년 강우량 자료에 대한 매개변수적 분포함수의 결과는 분포형에 따라 크기가 변하여 Log-Pearson TypeIII 분포는 관측자료보다 크게 추정되었고 3변수 Gamma 및 정규분포는 작게 나타났다. 매개변수적 분포함수에 비하여 변동 핵밀도함수 방법은 자료 자체의 모습을 표현하여 주므로, 자료의 밀도함수의 모양과 상관이 없이 일관성 있게 결과가 나타났다. 아래 표2에서는 비매개변수적 변동 핵밀도함수를 사용하여 서울관측소의 과거 관측강우량에 대한 확률강우량의 재현기간을 일 최대강우량, 월 최대강우량, 년 강우량별 보여주고 있다.

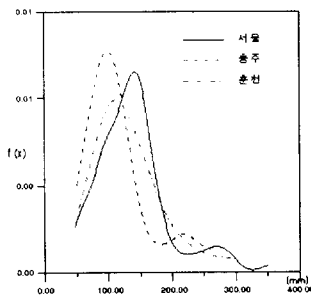


그림2. 일 최대강우량의 확률밀도함수.

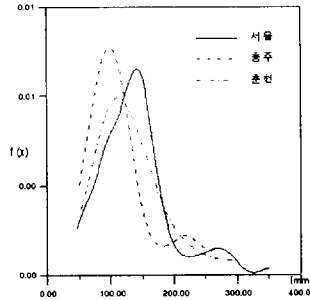


그림3. 월 최대강우량의 확률밀도함수.

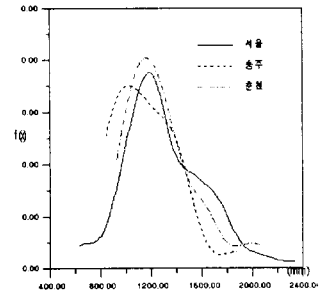


그림4. 년 강우량의 확률밀도함수.

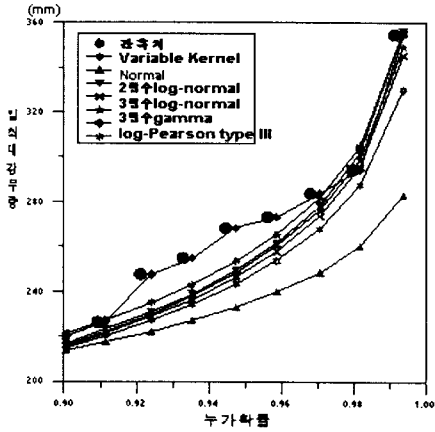


그림5. 일 최대강우량에 대한
누가분포함수(서울관측소).

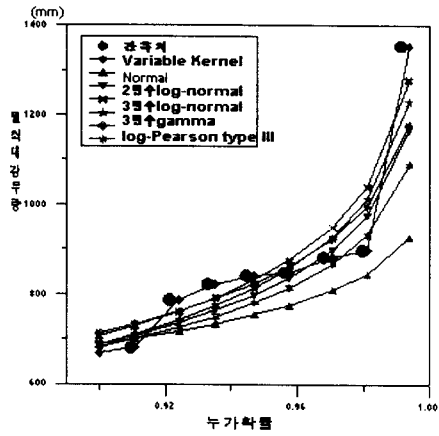


그림6. 월 최대강우량에 대한
누가분포함수(서울관측소).

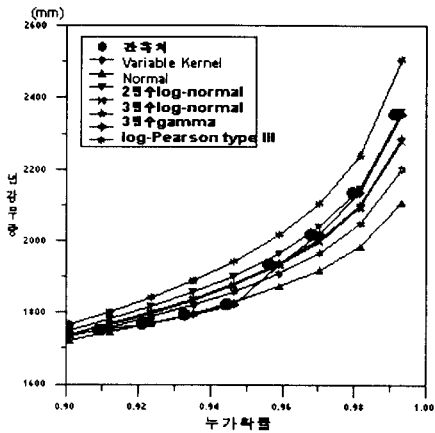


그림7. 년 강우량에 대한
누가분포함수(서울관측소).

표2. 서울관측소 과거 관측강우량에 대한 확률강우량의 재현기간. 단위(mm)

일 최대강우량			월 최대강우량			년 강우량		
년월일	강우량	재현기간	년월	강우량	재현기간	연도	강우량	재현기간
1920.8. 2	354.6	155	1940.7	1354.1	158	1990	2355.4	144
1987.7.28	294.5	54	1966.7	897.6	52	1940	2135.0	54
1940.7.10	283.8	33	1972.8	881.7	34	1960	2018.8	33
1972.8.19	273.1	24	1926.7	847.5	23	1925	1934.0	24
1984.9. 1	268.1	19	1925.7	841.0	18	1926	1822.4	18
1915.8.22	254.6	15	1930.7	821.2	15	1964	1793.9	15
1990.9.11	247.4	13	1995.8	786.5	13	1972	1769.5	13
1966.7.15	226.2	11	1920.8	680.4	11	1987	1751.3	11
1956.6.22	219.8	10	1936.8	668.7	10	1969	1736.7	10
1978.6.25	194.5	8	1987.7	651.1	8	1916	1710.2	8

4. 결론

본 연구에서는 전국 14개 지점의 일 최대 강우계열, 월 최대 강우계열 및 년 강우계열에 대하여 비매개변수적 변동 핵함수에 의한 빈도해석 및 매개변수적 빈도해석의 6개 분포형(정규분포, 2변수 Log-Normal분포, 3변수 Log-Normal 분포, 3변수 Gamma 분포, Log-Pearson type III, GEV1 분포)을 적용하여 각 강우계열의 분포함수와 재현기간을 비교하였다. 일반적으로 동일 지점의 자료에 대하여 여러 개의 매개변수적 분포함수가 적합도 판정을 받았으며, 이에 따른 관측자료의 재현기간은 분포형에 따라 다르게 나타났다. 그러나, 비매개변수적 변동 핵밀도함수에 의한 빈도해석은 자료의 밀도함수의 모양과 상관없이 일관성 있게 결과가 나타났다. 특히 관측자료가 bimodal 형태를 가지고 있을 시에는 매개변수적 빈도해석의 결과는 실측자료와 상당히 다른 거동의 모습을 보여주는 반면, 변동 핵밀도함수 추정의 결과 실측자료의 거동과 비슷하게 나타났다. 따라서 기존의 매개변수적 빈도해석은 자료의 형태에 따라 특정 확률분포형 선정에 어려움이 따를 수 있으나 비매개변수적 변동 핵밀도함수에 의한 빈도해석방법은 일관성 있는 결과를 산정함으로써 수공구조물의 적정한 크기를 정하는데 도움을 줄 수 있을 것이다.

5. 참고문헌

- Adamowski, K. 1981. Plotting formula for flood frequency. *Water Resources Bulletin* 17(2):197-202.
- Adamowski, K. 1996. Nonparametric Estimation of Low-Flow Frequencies. *Journal of Hydraulic Engineering* 122(1):46-49.
- Bowman, A. W. 1985. A comparative study of some kernel-based nonparametric density estimators. *Journal of Statistical Computing and Simulation* 21:313-327.
- Breiman, L., W. Meisel, and E. Purcell. 1977. Variable kernel estimates of multivariate densities. *Technometrics* 19(2):135-144.
- Greenwood, J.A., Landwehr, J.M., Matalas, N.C., and Wallis, J.R. 1979. "Probability weighted moment : definition and relation to parameters of several distributions expressible in inverse form." *Water Resources Research* 15(5):1049-1054.
- Guo, S. L. 1991. Nonparametric variable kernel estimation with historical floods and paleoflood information. *Water Res. Research* 27(1):91-98.
- Hosking, J.R.M. 1989. The theory of probability weighed moments. Research Report, RC 12210, IBM Research Division, T.J. Watson Research Center. New York.
- Hosking, J. R. 1990. "L-moments Analysis and estimation of distribution using linear combinations of order statistics.", *Journal of Royal Statistical Society* 52(1):105-124.
- Lall, U. , Young-II Moon, and K. Bosworth. 1993. Kernel flood frequency estimators :bandwidth selection and kernel choice. *Water Resources Research* 29(4):1003-1015.
- Moon, Young-II, Lall, U., and Bosworth, K. 1993. A comparison of tail probability estimators. *Journal of Hydrology* 151:343-363.
- Moon, Young-II and Lall, U. 1994. Kernel Quantile Function Estimator for Flood Frequency Analysis. *Water Resources Research* 30(11):3095-3103.
- Moon, Young-II. 1996. Nonparametric flood frequency analysis. *Journal of the Institute of Metropolitan Studies* 22(1):231-248.
- Rajagopalan, B., Lall, U. and Tarboton, D. 1996. Nonhomogeneous markov model for daily precipitation. *Journal of Hydrologic Engineering* 1(1):33-40.
- Rosenblatt, M. 1956. Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* 27:832-837.
- Sheather, S.F., and Jones, M.C. 1991, A reliable data-based band-width selection method for kernel density estimation. *J. Roy. Statistical Soc.:*B. 53, 683-690.