

비매개변수적 핵밀도함수와 대수핵밀도함수에 의한 확률강수량

○차영일¹⁾, 조성진²⁾, 권현한³⁾, 문영일⁴⁾

1. 서론

수공구조물의 설계 시 홍수량, 강우량 등의 수문자료에 대한 적당한 확률분포형을 선정하는 것은 안전하고 경제적인 수공구조물의 계획과 설계에 있어서 매우 중요한 요소 중의 하나이다. 우리나라 수공구조물에 대한 설계기준이 되는 빈도해석방법은 매개변수적 빈도해석방법이다. 그러나 매개변수적 방법은 한 지점에 적합도 검정에 통과한 다수의 분포형이 선정되는 경우에 어떤 분포형을 사용하느냐 하는 객관성의 문제점이 제기될 수 있다. 또한 수문자료가 여러 가지 원인으로 인하여 확률밀도함수의 첨두가 여러 개인 복합분포(mixed distribution) 형태를 가질 때, 매개변수적 해석방법으로는 이러한 확률밀도함수를 해석하는데 여러 가지 어려움이 따른다. 매개변수적 빈도해석에 대한 대안으로 제시되는 비매개변수적 빈도해석 방법은 자료에 대하여 어떤 분포형의 가정이 필요 없이 본래 관측자료가 나타내는 분포형을 자료의 확률밀도함수로 간주한다. 그러므로 분포형 선정의 어려움이 해소되고, bimodal을 갖는 분포의 경우도 자료의 특성이 잘 반영된다. 이러한 비매개변수적 해석방법의 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; Guo 등, 1996; Rajagopalan 등, 1996; Moon, 2000)되고 있다. 그러나 변동 핵밀도함수를 이용한 빈도해석도 관측자료의 구조에 따라 자료크기보다 매우 큰 재현기간에서는 추정값이 작은 경향을 보이는 문제가 있다(차영일, 2000). 이에 대한 방법으로 꼬리 부분이 두꺼운 핵함수의 개발과 최적의 광역폭의 선택, 자료의 변환, 또는 준매개변수적 방법(Semi-Parametric Method)(문영일, 2000) 등 여러가지 방법이 제안되고 있다. 본 연구에서는 그 중에서 자료의 대수변환을 이용하여 변동 핵밀도함수의 문제점을 해결하였다. 이 방법은 자료에 대수를 취해 변환 후 변동 핵밀도함수를 이용하여 빈도해석을 하는 방법으로 대수 변동 핵밀도함수라 하며 변동 핵밀도함수에 의한 빈도해석 시 큰 재현기간의 작은 추정값을 보완 할 수 있다.

본 연구에서는 기상청 산하 한강유역 관측소지점의 강우량 자료에 대하여 매개변수에 의한 확률강수량과 비매개변수에 의한 방법 중 변동 핵밀도함수와 대수 변동 핵밀도함수의 확률강수량을 추정하여 비교 분석하였다.

매개변수에 의한 방법은 모멘트법, 최우도법, 확률가중모멘트법, L-moments법을 사용하였고, 분포형은 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 등을 적용하였다.

-
- 1) 서울시립대학교 토목공학과 박사과정
 - 2) 서울시립대학교 토목공학과 석사과정
 - 3) 서울시립대학교 토목공학과 석사과정
 - 4) 서울시립대학교 토목공학과 조교수

2. 비매개변수적 밀도함수에 의한 빈도해석

2.1. 변동 핵밀도함수 추정법

핵밀도함수추정법은 다음 식 (1)과 같다(Rosenblatt, 1956).

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

여기서, h 는 광역폭(bandwidth)이고, n 의 자료의 개수, x 는 임의의 실수, X_i 는 독립적으로 분포된 실 관측치, $K(\cdot)$ 는 핵함수이다. 식 (1)은 그림 1과 같이 관측된 자료위치에 핵함수를 각각 위치하여 합산한 후 확률밀도함수로 표현하였다. 그림 1에서 관측자료의 개수 10, 광역폭 h 가 1인 핵함수를 사용한 핵밀도함수를 보이고 있다.

식 (1)의 고정 핵밀도 추정법은 꼬리(tail)에서 길게 늘어진 자료분포에 적용될 경우, 일정한 광역폭을 전체자료에 적용하기 때문에 추정치의 값이 불연속 또는 거친 밀도함수의 모양을 가질 수 있다. 이런 고정 핵밀도 추정법의 단점을 보완한 방법이 변동 핵밀도함수 추정법이다. Breiman은 고정 핵밀도함수 추정법의 특성에다 자료의 지역적인 밀도를 고려하는 다음과 같은 식 (2)의 k th nearest neighbor방법을 사용하여 자료의 밀도에 따라 광역폭 $hd_{j,k}$ 가 변하는 변동 핵밀도함수 추정법을 제안했다(Breiman 등, 1977).

$$f(x) = \frac{1}{nh} \int \frac{1}{d_{j,k}} K\left(\frac{x - X_j}{hd_{j,k}}\right) dx \quad (2)$$

여기서 $hd_{j,k}$ 는 한 개의 자료 x_j 에서 그 나머지 자료 $(n-1)$ 개 중에서 k 번째로 가까운 지점에 있는 자료까지의 거리로 변동광역폭이다. 따라서 꼬리부분같이 자료의 분포가 적은 낮은 밀도지역에서 $hd_{j,k}$ 는 커지고 핵함수의 모양은 넓게 퍼지게 된다. 자료의 분포가 많은 높은 밀도지역에서는 핵함수는 좁게 밀착된 형태를 가진다. Moon 과 Lall(1994)은 변동 핵밀도함수 추정법이 밀도함수의 꼬리 부분인 최빈값 추정이나 또는 자료가 비대칭 분포일 때 장점이 많다는 것을 보여 주었다. 그러나 변동 핵밀도함수 추정법도 관측자료크기보다 매우 큰 재현기간에 대한 추정값은 상대적으로 작은 경향을 보이는 문제가 있다. 따라서 본 연구에서는 대수 변동 핵밀도함수를 이용하여 문제 해결을 제시하였다.

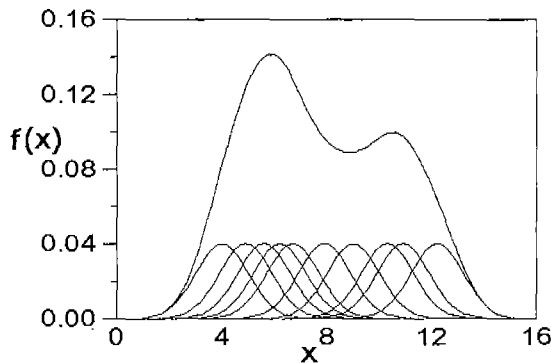


그림 1. 핵밀도함수

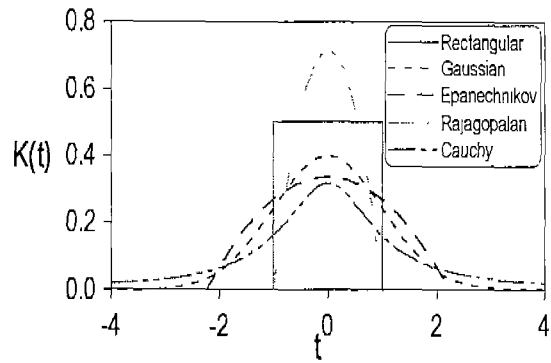


그림 2 여러 가지 핵함수

표 1. 여러 핵함수의 종류와 식

Kernel	K(t)
Rectangular	1/2 for $ t < 1$, 0 if otherwise
Gaussian	$\{1/\sqrt{2\pi}\} \times \exp(-t^2/2)$
Epanechnikov	$3/4(1-1/5t^2)/\sqrt{5}$ for $ t < \sqrt{5}$
Rajagopalan	$[3h/(1-4h^2)] \times (1-t^2)$, 여기서 $ t \leq 1$
Cauchy	$1/\{\pi(1+t^2)\}$

2.2. 대수 변동 핵밀도함수(Log-Variable Kernel Density Function) 추정법

대수 변동 핵밀도함수법은 관측자료에 대수를 취하여($y=\ln x$) 변환 후 변동 핵밀도함수를 이용하여 빈도해석을 하는 방법으로 변동 핵밀도함수법에 의한 빈도해석 시 외삽지점의 작은 추정값을 보완 할 수 있고 식 (3)과 같이 표현할 수 있다.

$$f(y) = \frac{1}{nh} \int \frac{1}{d_{j,k}} K\left(\frac{y - Y_j}{hd_{j,k}}\right) dy \quad (3)$$

여기서 $hd_{j,k}$ 는 한 개의 자료 $Y_j(Y_j=\ln X_j)$ 에서 그 나머지 자료 $(n-1)$ 개 중에서 k 번째로 가까운 지점에 있는 자료까지의 거리이다. 핵밀도함수 추정 방법을 실제로 적용할 때 핵함수 $K(t)$ 와 광역폭(bandwidth) h 를 선택하는 것이 필요하다. 많이 쓰이고 있는 핵함수들이 표 1에 주어졌고, 자료의 특성에 따라 각각 다른 핵함수를 적용할 수 있다(Lall 등, 1993).

광역폭(bandwidth) h 값은 핵밀도함수 추정법에 있어서 매우 중요하며 너무 큰 h 는 큰 편차(bias)로 인해 너무 완만(oversmooth)한 밀도함수의 추정과 정보의 손실을 가져오고, 너무 작은 h 는 큰 분산(variance)과 거친(rough) 추정치를 나타낸다. 주어진 자료에서 최적의 광역폭 h 를 구하는 여러 방법이 제안되고 있으며, 그 중에서 대표적인 것이 MISE(mean integrated squared error), ML(maximum likelihood), Breiman method, 그리고 Adamowski Criterion(Moon 등, 1994) 등으로 이를 값으로부터 최적의 h 를 구할 수 있다.

3. 적용 및 결과

본 연구에서는 기상청 산하 한강유역 관측소의 강수량 자료에 대하여 년 계열을 작성한 후, 각 계열에 대하여 이용 가능한 매개변수에 따른 분포함수에 의한 확률강수량과 변동 핵밀도함수와 대수 변동 핵밀도함수를 이용한 비매개변수에 의한 확률강수량을 추정하여 비교하였다. 매개변수적 분포의 적합도 검정은 유의수준 5%인 Kolmogorov-Smirnov검정과 χ^2 검정을 적용하여 이루어졌다.

지면상 한강유역(표 2)중 자료의 크기가 비교적 큰 강릉, 서울, 수원, 인천지점인 4개 관측소에 대하여 결과를 비교 분석하여 제시하였다. 그림 5~8에서는 한강유역의 4개 지점에 대한 년 최대치 계열의 누가분포함수를 보여주고 있으며, 표 3은 각 지점에 대한 재현기간별 확률강수량을 보여주고 있다. 그림 5~8에서와 같이 대부분 매개변수적 분포함수의 CDF는 실측자료의 empirical CDF와 다른 거동의 모습을 보이는 반면에, 변동 핵밀도함수(VK)와 대수 변동 핵밀도함수의 CDF는 상당히 근접한 모습을 보여 주고 있다. 그러나, 변동 핵밀도함수는 자료보다 상대적으로 큰

표 2. 한강유역 지점별 강우자료(기상청 강수 D/B, □ : 자료의 크기가 큰 지점)

유역	지점명	자료관측기간	자료크기	결측기간
한 강	강릉	1912~1999 (88년)	86	1950.9~10, 1951.1~ 8
	강화	1973~1999 (27년)	27	
	대관령	1972~1999 (28년)	28	
	서울	1907~1999 (93년)	88	1907.1~1907.9, 1950.9~1953.11
	속초	1968~1999 (32년)	32	
	수원	1964~1999 (36년)	36	
	양평	1973~1999 (27년)	27	
	이천	1973~1999 (27년)	27	
	인제	1973~1999 (27년)	27	
	인천	1949~1999 (51년)	49	1950.6~1951.9
	제천	1973~1999 (27년)	27	
	춘천	1966~1999 (34년)	34	
	충주	1973~1999 (27년)	27	
	홍천	1973~1999 (27년)	27	

재현시간에 대해서는 작은 추정값을 보이는 문제가 있으나, 대수 핵밀도함수는 변동 핵밀도함수의 그런 문제점을 잘 보완해 주고 있다.

표 3은 한강유역의 4개 지점에 대한 년 최대치 계열의 각 추정방법에 대한 100년 및 200년 빈도의 확률강수량을 보여주고 있다. 적합도검정을 2개의 분포형만이 통과한 인천지점을 제외한 모든 지점인 서울, 강릉, 수원이 변동 핵밀도함수법에 의한 추정법은 모든 확률강수량이 이 최대치와 최소치사이에 존재한다. 그러나 변동 핵밀도함수에 의한 확률강수량은 100년과 200년의 확률강수량의 차이가 거의 보이지 않는 단점이 있다. 그러나 대수 변동 핵밀도함수로 추정된 확률강수량은 매개변수적 분포함수와 비슷하게 지속적인 증가를 보이고 있어 변동 핵밀도함수에 의한 빈도해석의 단점을 보완한다 할 수 있다.

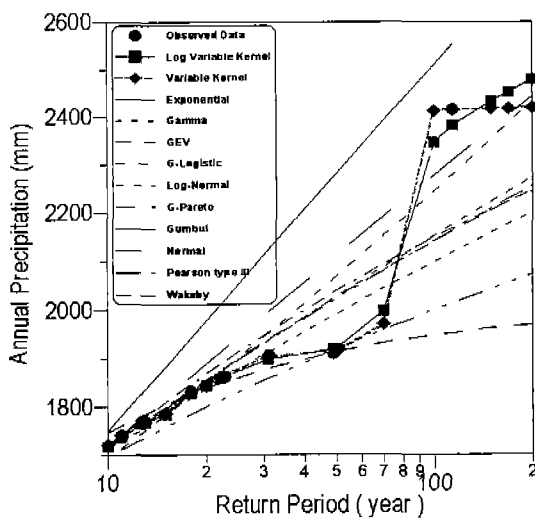


그림 5. 강릉지점의 누가분포함수(년강수량)

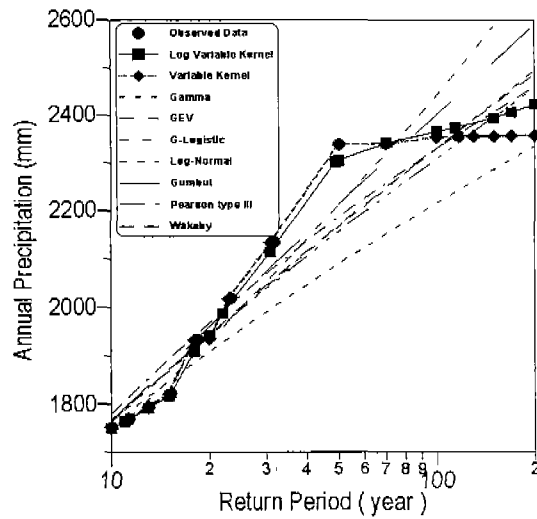


그림 6. 서울지점의 누가분포함수(년강수량)

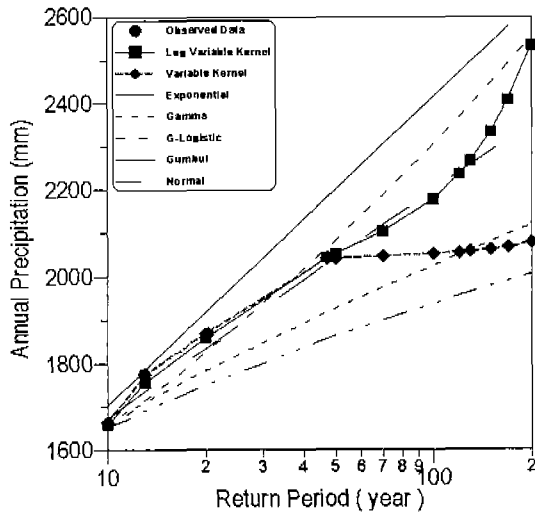


그림 7. 수원지점의 누가분포함수(년강수량)

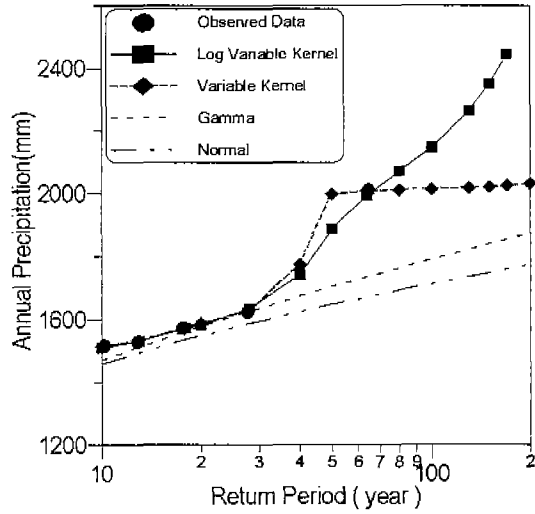


그림 8. 인천지점의 누가분포함수(년강수량)

표 3. 각 지점의 재현기간별 확률강수량(단위 : mm, □ : 최대치, 최소치)

지점	서울		강릉		수원		인천	
	100	200	100	200	100	200	100	200
Log-Variable Kernel	2365	2423	2347	2479	2178	2535	2147	2610
Variable Kernel	2353	2356	2414	2420	2047	2079	2015	2032
Exponential			2510	2739	2415	2629		
Gamma	2215	2335	2099	2200	2026	2119	1792	1875
GEV	2330	2486	2141	2248				
G-Logistic	2442	2696	2245	2433	2305	2555		
Log-Normal	2330	2493	2152	2272				
G-Parato			1949	1971				
Gumbel	2401	2590	2278	2444	2198	2353		
Normal			2000	2074	1939	2007	1714	1775
Pearson type III	2309	2456	2142	2256				
Wakeby	2330	2458	2149	2245				

4. 결론

본 연구에서는 기상청 산하 한강유역 관측소의 년 강우계열에 대하여 비매개변수적 방법인 변동 핵함수와 대수 변동 핵함수에 의한 빈도해석 및 매개변수적 빈도해석으로 추정된 년 강수량의 분포함수와 재현기간을 비교하였다. 일반적으로 매개변수적 빈도해석은 동일지점의 자료에 대하여 여러 개의 분포함수가 적합도 판정을 받았으며, 이에 따른 관측자료의 재현기간은 분포형에 따라 다르게 나타났다. 그러나, 비매개변수적 변동 핵밀도함수에 의한 빈도해석은 그 특성상 자료의 밀도함수의 모양과 상관없이 일관성 있게 결과가 나타났다. 특히 관측자료의 확률밀도함수가 첨두가 두 개인 bimodal 형태를 가지고 있을 시에는 매개변수적 빈도해석의 결과보다는, 대수 변동 핵밀도함수와 변동 핵밀도함수 추정의 결과가 실측자료의 거동과 비슷하게 나타났다. 하지만 변동 핵밀도함수법은 관측자료 크기보다 큰 재현기간에서의 추정치가 작게 나오는 문제가 있다. 이에 비하여 대수 변동 핵밀도함수에 의한 추정법은 이런 문제점을 잘 보완해주고 있다. 외삽의 문제는 대수 변동 핵밀도함수법에 의한 방법으로도 해소 될 수 있으나, 앞으로 응용분야에 적당한

핵함수의 개발, 자료의 특성에 맞는 최적의 광역폭의 선택방법, 그리고 준매개변수적 방법 등 지속적인 연구가 필요 할 것이다.

참고문헌

- 차영일, 2000. "매개변수와 비매개변수적 방법에 의한 우리나라의 지역별 확률강우량 추정." 서울 시립대학교, 공학석사논문.
- Adamowski, K., 1996. "Nonparametric Estimation of Low-Flow Frequencies." *Journal of Hydraulic Engineering*, Vol. 122, No. 1, pp. 46~49.
- Breinan, L., W. Meisel, and E. Purcell, 1977. "Variable kernel estimates of multivariate densities." *Technometrics*, Vol. 19, No. 2, pp. 135-144.
- Guo, S. L. 1991. "Nonparametric variable kernel estimation with historical floods and paleoflood information." *Water Res. Research*, Vol. 27, No. 1, pp. 91-98.
- Guo, S. L., R. K. Kachroo, and R. J. Mngodo, 1996. "Nonparametric kernel estimation of low flow quantiles." *J. of Hydrology*, Vol. 185, pp. 335-348.
- Lall, U., Young-II Moon, and K. Bosworth, 1993. "Kernel flood frequency estimators :bandwidth selection and kernel choice." *Water Resources Research*, Vol. 29, No. 4, pp. 1003-1015.
- Moon, Young-II, and U. Lall, 1994. "Kernel Quantile Function Estimator for Flood Frequency Analysis." *Water Resources Research*, Vol. 30, No. 11, pp. 3095-3103.
- Moon, Young-II, U. Lall, and K. Bosworth, 1993. "A comparison of tail probability estimators." *Journal of Hydrology* 151, pp. 343-363.
- Moon, Young-II, 1996. "Nonparametric flood frequency analysis." *Journal of the Institute of Metropolitan Studies*, Vol. 22, No. 1, pp. 231-248.
- Moon, Young-II, 2000. "The Study of Parametric and Nonparametric Mixture Density Estimator for Flood Frequency Analysis." *Water Engineering Research*, Vol. 1, No. 1, pp. 61-73.
- Rajagopalan, B., U. Lall, and D. Tarboton, 1996. "Nonhomogeneous markov model for daily precipitation." *Journal of Hydrologic Engineering* Vol. 1, No. 1, pp. 33-40.
- Rosenblatt, M., 1956. "Remarks on some nonparametric estimates of a density function." *Ann. Math. Statist.* 27, pp. 832-837.