# 경계 핵함수를 이용한 가뭄 빈도해석

○차영일[1], 문영일[2], 윤재호[1], 김동권[1]

## 1. 서 론

국내 및 전 세계적으로 근래의 이상기후 및 도시화로 인한 인구 집중으로 홍수에 의한 침수와 물 수요의 증가로 인한 가뭄의 피해는 더욱 커져 가고 있다. 또한 최근 연구에 의하면 년 강수량의 변동폭이 커지고 있어 가뭄과 홍수의 크기가 더욱 커지고 있다는 사실을 증명해주고 있다. 가뭄의 연구와 그 크기를 정량화 시키는 일은 앞으로의 수자원확충과 가뭄의 피해를 최소화하는데 매우 중요한 연구라 할 수 있을 것이다. 지금까지의 가뭄 빈도해석은 임의의 관측지점에서 갈수량 및 강수량에 대하여 대부분의 연구는 관측자료를 특정 확률밀도함수(PDF)를 가진 모집단으로부터 나왔다는 가정에 근거한 통계학적인 관점에 집중해왔다. 그러나 항상 어떤 단일 확률밀도함수가 관측된 자료에 대한 최선의 선택이 될 수는 없다. 관측 지점에서의 수문자료는 여러 가지 원인을 나타내는데, 이는 통계학적으로 이질적인 모집단 또는 복합 분포형 등을 초래하게 된다. 일반적으로 n=20~70 정도의 짧은 기록들로부터 임의 또는 미지의 모집단의 한정된 혼합을 동일하게 보는 것은 그리 논리적이라고 볼 수 없다. 일반적으로 가뭄빈도해석에 이용되는 자료 계열은 저유량에 대한 3, 5, 15일 등의 평균유량(CMS)과 또는 강수량의 경우 지속기간 1,2, …,24 개월의 이동누가우량에 대한 부분기간치계열을 이용하는 것이 일반적이나, 본 연구에서는 년 강수량계열을 이용하여 가뭄 빈도해석을 실시하였다.

따라서 본 연구에서는 관측자료($y_i$, i=1, 2,…, n)와 임의의 도시위치공식(plotting position formula)($p_i$, i=1, 2,…, n)을 사용하여 단일 자료 분포형뿐만 아니라 혼합된 분포형에서도 일관성 있게 적용할 수 있는 경계 핵 밀도함수를 이용한 빈도해석 방법을 제시하였다. 이 해석방법은 첫째로 기본적인 확률밀도함수에 대한 가정이 없다는 점에서 둘째로 꼬리의 거동이 뚜렷하게 만들어지고 추정량들이 데이터의 경험적 빈도값을 핵함수로 완화(smoothing)시킨다는 점에서는 완전하게 비매개변수적 기법이다. 지금까지 많은 비매개변수적 해석방법이 갈수량의 산정에 적용(Adamowski, 1996; Guo, 1996)되었다. 이러한 방법은 임의의 분위값(quantiles)에 대한 초과확률을 구하는 방식을 취한 핵밀도함수(Kernel Density Function)방법이나 본 연구에서는 반대로 임의의 초과확률에 대한 분위값(quantiles)을 구하는 방식을 취하여 가뭄빈도분석을 수

1) 서울시립대학교 토목공학과 박사과정
2) 서울시립대학교 토목공학과 조교수

행하였다.

## 2. 매개변수적 가뭄빈도해석

일반적인 수문자료의 빈도해석은 그림 1과 같은 방법으로 실시된다. 빈도해석에 주로 사용되는 확률분포형은 Log-normal분포, Gamma분포, Generalized Extreme Value분포, Wakeby분포, Gumbel분포, Log-Pearson TypeⅢ분포 등이 있고, 각각의 확률분포형은 매개변수의 개수와 그 형태가 서로 다른 특징을 가지고 있기 때문에 어떤 확률분포가 가장 적정한 분포인지에 대한 의견은 다양하며, 같은 분포형이라도 어떤 방법으로 매개변수를 추정하느냐에 따라 많은 차이가 있다.

수문자료의 분포특성을 잘 표현할 수 있는 매개변수의 추정은 분포함수를 이용한 매개변수적 빈도해석에 있어서 매우 중요하며 어려운 문제이다. 확률분포의 매개변수(parameters)를 추정하기 위해 사용하는 일반적인 방법은 여러 가지가 있으며 많이 쓰이는 방법으로는 모멘트법(method of moments), 최우도법(method of maximum likelihood), 확률가중모멘트법(method of probability weighted moments)과 L-모멘트법(L-moments)등이 있다. 또한 최적의 분포형을 선정하는 방법인 적합도 검정은 $\chi^2$ 검정, K-S 검정(Kolmogorov-Smirnov), PPCC 검정(Probability Plot Correlation Coefficient) 등이 있다.

## 3. 경계 핵 함수에 의한 추정식

### 3.1. 내부핵(Interior Kernel)함수에 의한 빈도해석

경계 핵 함수에 의한 빈도해석 추정식의 구성은 관측자료의 경험적 발생확률값과 핵함수의 핵완화와 초과확률에 대한 경계 핵함수에 근거를 두고 있다. 경험적 발생확률은 임의의 표준도시공식을 통하여 구할 수 있고, 표준도시공식으로 Adamowski (1981) 공식을 사용하면 다음 식(1)과 같다.

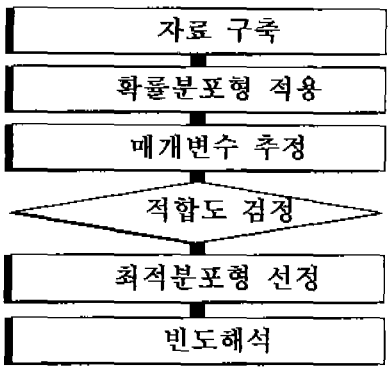


그림 1 매개변수적 가뭄빈도해석

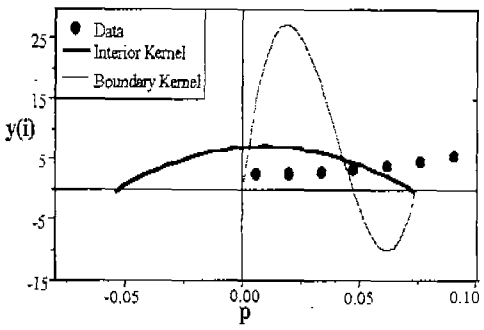


그림 2. x(0.01)의 핵함수 회귀추정(h=0.064)

$$p_i = \frac{i-0.25}{n+0.5} \tag{1}$$

여기서, i(i=1, 2,⋯, n)는 내림차순의 배열이고 $p_i$는 표준도시공식을 이용하여 경험상으로 추정되는 확률 값이다.

빈도해석 함수의 분위값(quantile) $\hat{x}(p)$는 Gasser-Muller (1984)가 식(2)와 같이 제시한 핵함수를 이용한 회귀추정식으로부터 추정할 수 있다.

$$\begin{aligned}\hat{x}(p) &= \sum_{i=1}^{n}\frac{1}{h}\int_{s_{i-1}}^{s_i} y_i K\left(\frac{p-u}{h}\right)du \\ &= \sum_{i=1}^{n}\frac{1}{h} y_i \int_{s_{i-1}}^{s_i} K\left(\frac{p-u}{h}\right)du \\ &= \sum_{i=1}^{n} y_i w_i\end{aligned} \tag{2}$$

여기서 h는 점 p와 관련된 대역폭(bandwidth), K( · )는 핵함수(kernel function)이며, $s_i = (p_i + p_{i+1})/2$, ( $i = 1, \cdots, n-1$), $s_0 = 0, s_n = 1$이다. 이 때 p는 초과확률을 나타내는 구간 [0,1]에서의 임의의 값이다.

### 3.2. 핵함수(Kernel Function)

일반적으로 사용되는 핵함수 K(t)는 t=0에서 최대치를 가지고 연속적이며, 대칭인 방정식의 형태를 가지고, Epanechnikov 핵함수는 다음 식 (3)과 같다.

$$K(t) = 0.75(1-t^2) \tag{3}$$

여기서 $|t| < 1$.

### 3.3. 대역폭(Bandwidth)

대역폭 h는 경계 핵함수법에 의한 빈도해석시 중요한 인자로서 추정되는 회귀식함수의 완만함과 거칠기를 결정한다. 작은 대역폭은 임의의 점에서 $\hat{x}(p)$ 산정에 적은 수의 관측자료가 고려되고, 보다 큰 대역폭에서는 상대적으로 많은 수의 관측자료가 고려되어 유연한 회귀식을 만든다. 따라서 대역폭이 증가함에 따라 편의(bias)는 증가하고 분산은 감소한다. Muller(1991)는 최적의 광역폭 선택을 위한 평균제곱오차(Mean Squared Error)를 다음 식 (4)와 같이 제시하였다.

$$\begin{aligned} MSE(\hat{x}(p)) &= E[\hat{x}(p)-x(p)]^2 \\ &\sim \frac{\sigma^2}{nh}\int_{-1}^{q}(K_x(q,t))^2dt+\frac{1}{4}h^4\{x''(p)\}^2\left\{\int_{-1}^{q}K_x(q,t)t^2dt\right\}^2 \end{aligned} \tag{4}$$

편의와 분산을 고려한 최적화된 대역폭을 찾는 방법에는 Gasser 등(1991) 에 의해 제안되었다. Gasser 등(1991)에 의해 $0<p<1$에 걸친 범위에 대해 식(4)에서 평균제곱오차(MSE)를 최소화시키는 최적의 대역폭은 다음 식 (5)와 같이 제시하였다.

$$h=\left\{\frac{1.5}{n}\frac{c_1}{c_2}\frac{\sigma^2}{\int_0^1\{x''(p)\}dt}\right\}^{0.2} \tag{5}$$

여기서, $c_1=2\int_{-1}^{1}K_x(q,t)^2dt$, $c_2=4\int_{-1}^{1}K_x(q,t)t^2dt$.

### 3.4. 경계 핵(Boundary Kernel) 함수에 의한 빈도해석

일반적으로 가뭄빈도해석 수행시 하위 비초과확률($0<p<0.1$)에 관심을 가지게 된다. 그러나 재현기간별 빈도해석시 주어지는 전형적인 관측자료의 크기는 20~90개의 범위를 가지게 된다. 결과적으로 빈도해석시 홍수해석에는 $p>p_n$에 대한 주어진 자료의 외삽이 필요하게 되고(문영일, 2000), 반대로 가뭄빈도해석에는 $p>p_1$의 외삽이 필요하다. 그러나 지금까지 사용한 Epanechnikov 핵함수와 같이 내부 핵(Interior Kernel) 함수를 사용하면 그림 2와 같이 한정된 영역($0<p$)을 벗어나기 때문에 경계 핵(Boundary Kernel) 함수가 필요하다. 경계 핵함수는 구간 [0, 0+h] 내에서 가중된 포선형의 편의를 제거하기 위해 필요하다.

Muller(1991)는 내부 핵함수에 대응하는 경계 핵함수를 식(4)의 평균제곱오차(MSE) 의 최적화에 따라 제시하였다. 여기에서 Epanechnikov 핵함수를 사용하여 내부 핵함수와 경계 핵함수를 살펴보면 각각 다음 식(6), 식(7)과 같다.

내부 핵함수 :

$$K(t)=0.75(1-t^2),\ \text{여기서}\ |t|<1 \tag{6}$$

경계 핵함수:

$$K(q,t)=6(1+t)(q-t)\frac{1}{(1+q)^3}\times\left\{1+5\left(\frac{1-q}{1+q}\right)^2+10\frac{1-q}{(1+q)^2}t\right\} \tag{7}$$

여기서 $t=(p-p_i)/h$이며, 가뭄빈도해석에 적용되는 $K(q,t)$는 왼쪽 경계면 구간 [0, h]에서 $q=p/h$ 이다. 이때 q=1이면 경계 핵함수 식(7)은 내부 핵함수 Epanechnikov 핵함수와 동일하다.

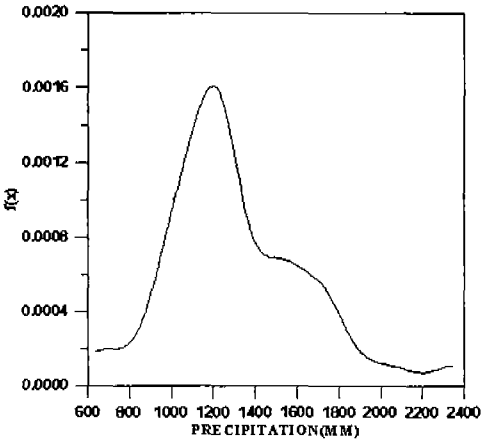


그림 3. 서울지점의 확률밀도함수

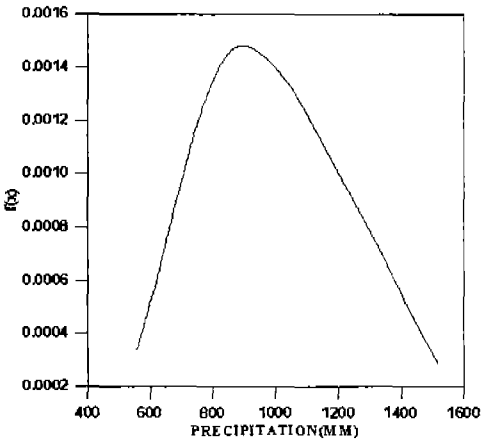


그림 4. 대구지점의 확률밀도함수

# 4. 적용 및 분석결과

## 4.1. 확률 갈수량 산정

그림 3와 그림 4는 서울과 대구지점의 년 강수량자료의 확률밀도함수이다. 그림 3에서 서울지점은 bimodal의 경향을 보이고 있고, 그림 4에서 대구지점은 비교적 양호한 단일 확률밀도함수를 보이고 있다. 그림 5과 그림 6은 서울과 대구지점의 년강수량자료에 대한 가뭄 누가분포함수의 그래프들이다. 관측치는 Adamowski 도시공식에 의해 도시되었고, 경계 핵 함수에 의한 추정치와 검정을 통과한 매개변수적 빈도해석에 의해 추정된 가뭄 누가분포함수(CDF)를 보여주고 있다. 그래프에서 x축은 가뭄재현기간이고, y축은 강수량이며 단위는 각각 년과 mm이다. 그림 5과 그림 6의 그래프에서 서울은 100년 빈도 가뭄강수량이 지수분포는 931mm, 정규분포는 515mm로 약 45%의 편차를 보이고 있고, 또한 대구지점은 Generalized Pareto분포는 648mm, 정규분포는 451mm로 약 30%의 편차로 역시 큰 편차를 보이고 있다. 그러나 경계핵 밀도함수에 의한 추정치는 관측치는 서울의 경우 601mm, 대구지점은 566mm로 두 지점

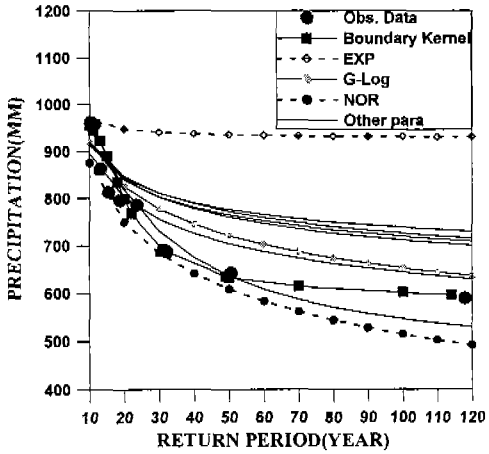


그림 5 서울지점의 재현기간별 가뭄강수량

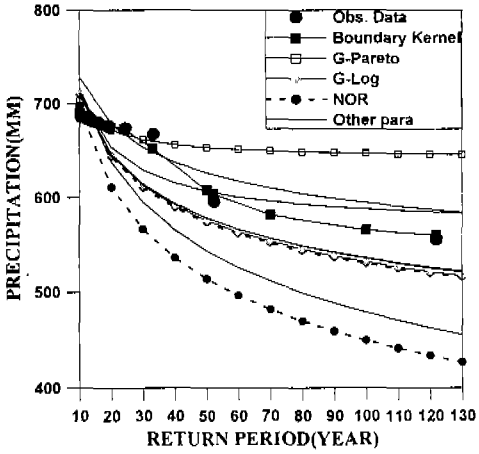


그림 6 대구지점의 재현기간별 가뭄강수량

모두 매개변수에 의한 추정치의 범위 안에 들어가는 값을 제시하고 있다. 그러나 매개변수적 빈도해석은 선택의 범위가 너무 커서 객관적인 선택을 하는데 있어서 어려움이 있다.

## 5. 결 론

년 강수량 계열에 대한 가뭄 빈도해석 시 매개변수에 의한 확률 년 최소강수량추정은 한 지점에서 여러 개의 확률분포형이 적합도 판정을 받는 경우가 일반적이고 또한 매개변수 산정 방법에 따라 추정값의 차이가 있었다. 서울지점은 100년 빈도 가뭄강수량이 기존방법은 확률분포형에 따라 [515mm, 931mm] 사이의 값을 제시한 반면 경계핵 밀도함수는 601mm, 대구지점은 확률분포형에 따라 [451mm, 648mm] 경계핵 밀도함수는 566mm을 각각 산정하였다. 따라서 매개변수적 방법은 분포형에 따라 100년 빈도 가뭄강수량의 경우 서울은 45% 이상, 대구는 30%이상의 차이가 있어 재현기간별 적절한 추정값을 선택하는데 어려움이 있었다. 반면에 경계핵 밀도함수 추정법은 어느 특정 분포형의 가정이 필요 없어 분포형 선택의 어려움이 해소되어 일관된 하나의 값을 제시하였다. 또한 일반적인 비매개변수적 빈도해석 방법에서 자료에 따라 나타나는 외삽의 문제를 경계 핵함수를 사용하여 해결할 수 있었다.

### 참고문헌


문영일. 2000. Boundary Kernel 함수를 이용한 빈도해석, 한국수자원학회, 2000년도 학술발표회 논문집, pp. 71-76.

문영일, 차영일, 전시영. 2000. 매개변수적 빈도해석시 매개변수 추정방법과 검정방법에 의한 적정분포형 선택에 관한 연구, 대한토목학회, 2000년도 학술발표회 논문집 (II), pp. 107-110.

Adamowski, K. 1981. Plotting formula for flood frequency. Water Resources Bulletin 17(2), pp. 197-202.

Adamowski, K. 1996. Nonparametric Estimation of Low-Flow Frequencies. Journal of Hydraulic Engineering, Vol. 122, No. 1, pp. 46~49.

Gasser, T., and Muller, H. G. 1984. Estimating regression functions and their derivatives by the kernel method. Scandinavian Journal of Statistics 11, pp. 171-185.

Gasser, T., Kneip, A., and Kohler, W. 1991. A flexible and fast method for automatic smoothing. Journal of the American Statistical Association 86(514), pp. 643-652.

Guo, S. L. 1996. Nonparametric kernel estimation of low flow quantiles. J. of Hydrology, Vol. 185, pp. 335-348.

Muller, U. G. 1991. Smooth optimum kernel estimators bear endpoint, Biometrika, 78(3), pp. 521-530.