

극치수문자료의 확률량 산정에 따른 안전도 해석

Safety Analysis of estimating quantiles of extreme hydrologic data

○김경덕*, 허준행**, 조원철***

1. 서론

확률홍수량을 산정하기 위해서는 적정 확률분포형을 선정하게 되는데, 실측자료의 분포특성을 정확히 나타내는 분포형을 결정한다는 것은 매우 어렵다. 따라서 확률분포형을 선정할 때는 여러 가지 확률분포형을 실측자료에 적용하여 합리적인 매개변수 추정과 매개변수 적합성 확인, 적합도 검정 등을 통하여 가장 적합한 분포형을 선정해야 한다.

그러나 자료의 구축에 있어서 발생하는 자료오차와 모집단에 대한 분포형을 가정할 때 생기는 모형오차에 의하여 산정된 확률홍수량은 불확실성이 내포되어 있다. 따라서 본 연구에서는 합리적인 빈도해석을 통하여 확률홍수량을 산정하고, 산정된 확률홍수량에 대하여 불확실성/안전도 해석을 실시하려 한다.

홍수량자료나 저유량 자료 등 극치자료에 대한 매개변수적 빈도해석은 1910년대 이후 활발하게 진행되어 왔다. 최근까지 이러한 매개변수적 빈도해석은 매개변수 추정기법, 적합도 검정방법 등에 있어서 많은 발전을 보여왔다. 국내에서는 1960년대 이후 빈도해석에 대한 연구가 활발히 진행되고 있는데, 최근까지 확률분포형, 매개변수 추정방법, 적합도 검정방법 등에 있어서 미진한 면이 많았다. 따라서 본 연구에서는 최근의 세계적 추세에 발맞추어 매개변수적 빈도해석을 실시하고자 한다. 확률분포형은 외국에서 추천하고 있는 GEV, Wakeby 분포 등을 추가하였으며, 모멘트법, 최우도법, 확률가중 모멘트법 등으로 매개변수를 추정하였다(김경덕 등, 1995, 1996). 추정된 매개변수는 적합성 조건을 조사하였으며, 적합성 조건을 만족하는 확률분포형에 대하여 χ^2 검정, Kolmogorov - Smirnov 검정, Cramer von Mises 검정 등의 적합도 검정을 실시하였다. 최근에 간편하면서도 강력한 기법으로 알려진 probability plot correlation coefficient(PPCC) 검정(Filliben, 1975; Vogel, 1986; Chowdhury 등, 1991)을 추가하여 적정확률분포형 선정에 다각적인 접근을 시도하였다. 본 연구에서는 최근의 빈도해석 기법을 통하여 확률홍수량을 산정하려 하며, 모의발생을 통해 산정된 확률홍수량에 대하여 안전계수를 제시하고자 한다.

* 연세대학교 대학원 토목공학과 박사과정

** 연세대학교 공과대학 토목공학과 부교수

*** 연세대학교 공과대학 토목공학과 교수

2. 본 론

대상자료는 1954년부터 1995년까지의 고안지점 홍수량자료이며, 이를 그림으로 나타낸 것이 그림 1이다. 그리고 이 자료에 대하여 무작위성(randomness)을 조사하였다.

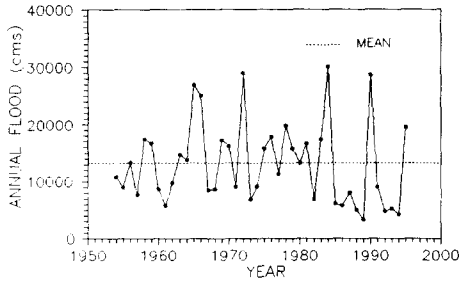


그림 1 고안지점의 경년별 홍수량

무작위성 검토는 Run test, Spearman's rank correlation coefficient test, Turning point test 등을 실시하였으며, 그 결과를 나타내면 표 1과 같다.

표 1 무작위성 검토결과

유의 수준	Run test			Spearman's rank correlation coeff. test			Turning point test		
	Uc	$u_{1-a/2}$	결과	Tc	$t_{1-a/2}(N-2)$	결과	Uc	$u_{1-a/2}$	결과
0.01	0.611	2.240	○	1.327	2.704	○	0.125	2.240	○
0.05	0.611	1.96	○	1.327	2.021	○	0.125	1.960	○

표 1에서 보면 알 수 있듯이 3가지 검정에 있어서 독립성이 검증되었다. 이 자료에 대하여 빈도해석을 다음과 같이 실시하였다.

2.1 확률분포형의 적용

확률가중 모멘트법에 의한 매개변수 추정치와 각 확률분포형이 요구하는 매개변수 적합성 조건을 검토한 결과는 다음 표 2와 같다.

표 2 추정된 매개변수와 적합성 결과

확률분포형	자료 크기	위치 매개변수	최소값	최대값	규모 매개변수	형상 매개변수	적합성 결과
gamma-2	39	0.0	3,330	30,130	4313.673	3.112	O-K
gamma-3	39	788.882	3,330	30,130	4532.368	2.716	O-K
GEV	39	9624.970	3,330	30,130	5548.041	-0.047	O-K
Gumbel	39	9747.445	3,330	30,130	5808.059	-	O-K
log-Gumbel 2	39	-	3,330	30,130	8977.056	2.587	O-K
lognormal-2	39	0.0	3,330	30,130	9.364	0.570	O-K
lognormal-3	39	-4396.402	3,330	30,130	9.684	0.414	O-K
log-Pearson type III	39	16.025	3,330	30,130	-0.051	132.550	O-K
Weibull-2	39	0.0	3,330	30,130	14759.530	1.888	O-K
Weibull-3	39	2415.906	3,330	30,130	11800.300	1.466	O-K
		위치 매개변수	A	B	C	D	
Wakeby-4	39	0.0	3886.609	144.648	-39021.070	-0.310	O-K
Wakeby-5	39	3516.02	772.714	3.945	-40283.540	-0.286	O-K

2.2 도시적 해석과 적합도 검정

추정된 매개변수가 적합성 조건을 만족하는 확률분포형에 대하여 도시적 해석과 적합도 검정을 실시하였다. 도시적인 해석은 군집화된 자료에 대하여 빈도해석을 실시하므로써, 경험 확률밀도함수와 누가분포함수를 산정하고 추정된 매개변수를 이용하여 적합된 확률밀도함수와 누가분포함수를 도시하여 비교하는 기법이다. 그림 2는 확률밀도함수, 그림 3은 누가분포함수에 대한 도시적인 해석을 나타내고 있다.

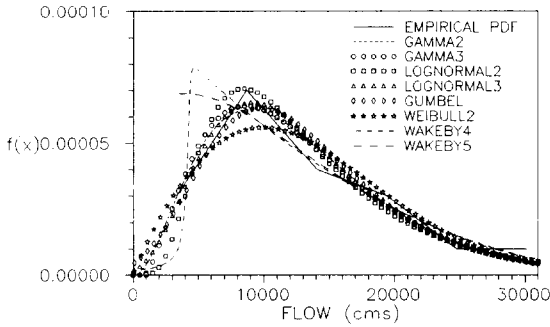


그림 2 경험 확률밀도함수와 적합된 확률밀도함수

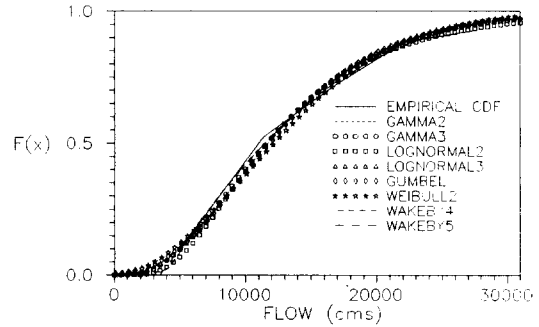


그림 3 경험 누가분포함수와 적합된 누가분포함수

적합도 검정으로는 χ^2 -검정, Kolmogorov-Smirnov 검정, Cramer von Mises 검정을 실시하였다. 유의수준 5%에 대하여 적합도 검정결과는 다음 표 3과 같다.

표 3 적합도 검정결과(유의수준 5%)

확률분포형	χ^2 -검정			Kolmogorov-Smirnov 검정			Cramer von Mises 검정		
	계산값	표값	판정	계산값	표값	판정	계산값	표값	판정
gamma-2	4.90	7.81	○	0.11	0.18	○	0.06	0.46	○
gamma-3	4.90	5.99	○	0.10	0.18	○	0.06	0.46	○
GEV	6.29	5.99	×	0.11	0.18	○	0.07	0.46	○
Gumbel	7.43	7.81	○	0.12	0.18	○	0.07	0.46	○
log-Gumbel 2	9.71	7.81	△	0.15	0.18	○	0.26	0.46	○
lognormal-2	6.86	7.81	○	0.12	0.18	○	0.07	0.46	○
lognormal-3	3.14	5.99	○	0.11	0.18	○	0.07	0.46	○
log-Pearson type III	8.24	5.99	×	0.10	0.18	○	0.06	0.46	○
Weibull-2	7.71	7.81	○	0.12	0.18	○	0.07	0.46	○
Weibull-3	4.29	5.99	○	0.09	0.18	○	0.05	0.46	○
Wakeby-4	1.82	5.99	○	0.07	0.18	○	0.04	0.46	○
Wakeby-5	3.63	3.84	○	0.07	0.18	○	0.05	0.46	○

다음 표 4는 고안지점의 홍수자료에 대하여 적합도 검정과 도시적인 해석에서 채택된 분포형중에서 gamma-2, Weibull-2, Gumbel, lognormal-2 분포형의 PPCC 검정을 실시한 결과이다.

표 4 PPCC 검정결과(유의수준 5%)

	gamma-2	Gumbel	lognormal-2	Weibull-2
계산값	0.98365	0.98344	0.96658	0.98459
표값	0.97761	0.96148	0.97280	0.96198
판정	○	○	×	○

표 4에서 보면, lognormal-2 분포형은 7 됨을 알 수 있으며, gamma-2, Weibull-2, Gumbel 분포형은 채택되었다.

2.3 확률홍수량 산정

이상의 도시적인 해석, 적합도 검정, PPCC 검정 등을 종합적으로 고려하였을 때, gamma-2, Gumbel, Weibull-2, 분포형이 적합한 확률분포형으로 선정되었다. 다음 표 5는 채택된 3가지 확률 분포형에 따른 확률홍수량을 나타내고 있다.

표 5 고안지점의 확률홍수량

재현기간 (year)	확률분포형별 확률홍수량 (cms)		
	gamma-2	Gumbel	Weibull-2
2	11726.5	11876.1	12155.2
5	18603.6	18459.1	18990.6
10	23057.5	22817.7	22957.4
20	27205.8	26998.5	26391.1
50	32400.5	32410.1	30397.9
100	36180.5	36465.3	33141.1
200	39866.8	40505.8	35696.0
300	41987.0	42865.6	37117.8
400	43478.4	44538.9	38097.9
500	44627.3	45836.5	38843.0
1000	48161.6	49865.2	41080.6

2.4 불확실성 해석

자료오차와 모형오차를 고려하기 위하여 산정된 확률홍수량에 대하여 안전계수를 결정하려 한다. 이를 위해서는 적정 확률분포형을 모집단의 확률분포형으로 가정하고, 모의실험을 통하여 자료를 발생시켜서 안전계수를 얻게 된다.

- 대상분포형 : Gumbel, gamma-2, Weibull-2 분포형
- 모의발생 시행횟수 : 1,000번
- 자료크기 종류 : 10개(20, 30, 40, 42, 50, 60, 70, 80, 90, 100)
- 재현기간 종류 : 12개(2, 3, 5, 10, 20, 30, 50, 70, 100, 200, 300, 500)
- 각 분포형별로 모집단의 분포형으로 가정하고, bias와 root mean square error 산정
(Cunnane, 1989)
- 모의발생된 1000개의 자료집단으로부터 확률홍수량 산정
- 산정된 1000개의 확률홍수량에 대하여 nonparametric estimation을 통하여 relative bias를 산정하여 안전계수 결정(Adamowski, 1989, 1996; Mkhandi, 등, 1996)

Gumbel, gamma-2, Weibull-2 분포형을 모집단의 확률분포형으로 가정하고 모의실험을 통하여 얻어진 root mean square error는 다음 표 6과 같다.

표 6 모의실험에 의한 root mean square error 결과(재현기간 100년)

자료 크기	가정분포형								
	Gumbel			gamma-2			Weibull-2		
	적용분포형								
	GUM	GAM2	WBU2	GUM	GAM2	WBU2	GUM	GAM2	WBU2
20	17.054	18.369	20.139	16.262	17.893	19.387	16.237	19.621	15.430
30	14.283	15.005	17.469	14.405	15.537	17.000	15.892	18.389	14.277
40	12.441	13.039	15.314	12.433	13.360	15.498	14.644	16.378	12.539
42	12.091	12.687	15.270	12.914	13.609	15.953	14.627	16.488	12.493
50	11.099	11.656	14.824	11.782	12.407	14.599	14.350	15.889	11.794
60	10.417	10.834	13.830	11.532	12.098	14.093	13.931	15.359	11.198
70	9.400	9.951	13.307	10.941	11.309	13.700	13.193	14.503	10.918
80	8.913	9.077	12.334	10.188	10.546	13.156	12.984	14.263	10.355
90	8.328	8.495	12.144	9.693	10.082	12.713	12.815	13.986	10.142
100	7.644	7.843	11.592	9.321	9.597	12.271	12.469	13.633	9.689

Kernel 밀도함수는 식 (1)로 주어지며, Epanechnikov kernel 함수는 식 (2)와 같다. bandwidth h에 대하여 매우 민감하게 반응하는데, cross validation 기법중 많이 사용되고 있는 Maximum likelihood cross validation(MLCV)과 Least square cross validation(LSCV)은 식 (3), (4)로 나타낼 수 있다.

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

$$K(t) = \frac{3}{4}(1 - t^2), \quad -1 \leq t \leq 1 \quad (2)$$

$$CV_{KL}(h) = n^{-1} \sum_{i=1}^n \log \left[\sum_{j \neq i} K\left(\frac{X_i - X_j}{h}\right) \right] - \log [(n-1)h] \quad (3)$$

$$CV(h) = \int \hat{f}_h^2(x) dx - \frac{2}{n} \sum_{i=1}^n \hat{f}_{h,i}(X_i) \quad (4)$$

다음 표 7은 모의발생된 자료집단으로부터 산정된 확률홍수량에 대하여 MLCV, LSCV 기법에 의해 Epanechnikov kernel 함수의 bandwidth h를 구하고 다음 식 (5), (6)에 의해서 얻어진 안전계수를 나타내고 있다.

$$SF = 1.0 + RB(\%)_{P(\hat{Q}_T)=0.05} \quad (5)$$

$$RB(\%)_{P(\hat{Q}_T)} = \{[\hat{Q}_T - E(\hat{Q}_T)]/E(\hat{Q}_T)\} \times 100, \quad P(\hat{Q}_T) = 0.95 \quad (6)$$

표 7 Gumbel 분포형을 모집단으로 가정할 경우의 안전계수(재현기간 100년, 자료기간 42년)

분포형	MLCV	LSCV
Gumbel	1.0124	1.0025
Gamma2	1.0063	1.0092
Weibull2	1.0143	1.0028

3. 결 론

고안지점의 홍수자료를 이용한 매개변수적 방법을 통해 최적확률분포형을 선정하고, 이를 이용하여 확률홍수량을 추정하였다. 또한 모형오차와 자료오차에 의한 불확실성에 대하여 모의발생을 통한 안전계수를 산정함으로써 수공구조물의 설계 혹은 수자원 계획 등에 있어서 보다 안정된 설계량을 제시하였다.

참고문헌

- 건설부(1993~1995). 한국 수문조사서.
- 김경덕, 허준행, 조원철(1996). "연최대치 강우자료의 적정 확률분포형 선정에 관한 연구.", 대한토목학회 논문집, 제 16권, 제 II-4호, pp. 335~344.
- 김경덕, 서규우, 허준행, 조원철(1995). "한강 인도교 지점에서의 홍수빈도해석(II) -확률홍수량 산정-.", 대한토목학회 논문집, 제 16권, 제 II-1호, pp. 23~31.
- 이승재, 서규우, 허준행, 조원철(1995). "한강 고안지점의 홍수위 환산과 홍수빈도해석.", 한국수자원학회 논문집, 제 28권, 제 5호, pp. 191~204.
- 허준행(1996). "수문통계학의 기초(III).", 한국수자원학회지 기술강좌, 제29권, 제 5호, pp. 67~79.
- Adamowski, K.(1989). "A Monte Carlo comparison of parametric and nonparametric estimation of flood frequencies.", *Journal of Hydrology*, 108, pp. 295-308.
- Adamowski, K.(1996). "Nonparametric estimation of low flow frequencies.", *Journal of Hydraulic Engineering, ASCE*, Vol. 122, No. 1, pp. 46-49.
- Chowdhury, J. D.(1991). "Goodness-of-fit tests for regional generalized extreme value flood distributions.", *Water Resources Research*, Vol. 27, No. 7, pp. 1765-1776.
- Cunnane, C.(1989). "Statistical distributions for flood frequency analysis.", WMO Operational Hydrology Report, No. 33, Geneva, Switzerland.
- Filliben, J. J.(1975). "The probability plot correlation coefficient test for normality.", *Technometrics*, Vol. 17, No. 1, pp. 111-117.
- Mkhandi, S. H., Kachroo, R. K., and Guo, S. L.(1996). "Uncertainty analysis of flood quantile estimates with reference to Tanzania.", *Journal of Hydrology*, 185, pp. 317-333.
- Vogel, R. M.(1986). "The probability plot correlation coefficient test for the normal, lognormal, and Gumbel distributional hypotheses.", *Water Resources Research*, Vol. 22, No. 4, pp. 587-590.
- Vogel, R. M., and McMartin, D. E.(1991). "Probability plot goodness-of-fit and skewness estimation procedures for the Pearson type 3 distribution.", *Water Resources Research*, Vol. 27, No. 12, pp. 3149-3158.