

## 수문자료의 확률해석 - 기초와 응용

### Frequency Analysis of Hydrologic Data - Principles and Applications

박 승 우\*  
Park, Seung-Woo

#### 1. 머리말

수공 구조물의 계획과 설계는 미래의 예측할 수 없는 특정 크기의 수문사상에 대하여 제대로 기능을 할 수 있도록 해야한다. 여기서, 수문사상은 강우량이나 홍수량, 혹은 갈수량 등과 같은 특이 사상을 대상으로 한다. 이와 같은 수문량이 어떤 특정치를 초과(超過)하거나, 하회(下廻)할 확률을 추정함으로써 수공구조물의 경제적, 사회적 평가에 이용하는 것이다.

수공 구조물이 앞으로 발생할 수 있는 모든 크기의 홍수량이나 갈수량을 완전히 통제할 수 있는 규모나 크기로 계획하는 일은 현실적으로 불가능하므로, 특정한 생기 확률(Probability of occurrence)을 가지는 수문량에 대하여 계획하고 있다.

수문자료의 확률해석이란 확률론적 방법을 이용하여, 수문량의 크기에 대한 생기 빈도를 해석하는 것으로 빈도해석이라고도 한다. 확률론적 방법이란 대상이 되는 수문사상의 발생이 순수한 확률적인 특징을 가지는 경우에 해당한다. 만약, 계절별 강수량과 같이 일정한 주기성이나 경향성 혹은 특이성을 갖는 경우는 확률적이지 아니므로, 그 분석 대상이 되지 못한다.

본 소강좌에서는 확률론적 수문분석기법의 기초적인 내용을 살펴보고, 실무에서 흔히 접하는 문제를 중심으로 응용방법을 논의하도록 한다.

#### 2. 확률과 재현기간

확률은 서로 독립적으로 반복되는 사상이 여러 차례의 시도 끝에 그 사상이 생기할 도수하는 평균적인 값으로 정의한다.

$$P(E_i) = \frac{n_i}{N}$$

여기서,  $P(E_i)$ 은  $E_i$ 사상의 확률,  $n_i$ 은  $E_i$ 의 사상이 생기는 도수,  $N$ 은 총시도회수이다.  $N$ 이 충분히 큰값일 때, 상대도수  $\frac{n_i}{N}$ 을 확률이라 한다.  $P(E_i)$ 이 0.10이면, 10회중 1회는  $E_i$ 의 사상이 나타나는 것을 뜻한다.

확률을 표시하는 다른 방법은 누가확률분포함수이다. 누가확률분포함수  $F(x)$ 는 변량  $X$ 가 특정값  $x$ 보다 크지 않을 확률 (비초과확률)을 말하며, 다음과 같다.

$$F(x) = P(X \leq x)$$

어떤 사상의 확률이 주어지면, 그 사상이 나타난 후 다시 나타날 때까지의 횟수나 기간을 정할 수 있다.  $P(E_i)=0.10$ 의 확률을 가질 때,  $E_i$ 사상이 평균적으로 10회 (또는 10 기간)후에 다시 나타나는 것을 뜻하므로, 이를 수식으로 나타내면 다음과 같다.

$$T = \frac{1}{P(E_i)}$$

\*서울대학교 농업생명과학대학 (swpark@snu.ac.kr)

여기서, T= 재현기간 (Return period) 혹은 재현간격(Recurrence interval)이다. 재현기간은 연 최대 홍수량과 같이 연도별 자료에 대한 빈도해석에서 쓰고, 재현 간격은 주사위 문제와 같이 기간으로 나타나지 않는 경우에 적용한다.

### 3. 확률분포함수

수문변량의 빈도해석에서는 수문자료의 빈도분포특성을 대표하는 이론적 확률분포함수를 사용한다. 수문통계에서 사용되는 이론확률분포함수에는 1) 정규분포함수, 2) 대수정규분포함수, 3) 3변수 대수정규분포함수, 4) Gamma 분포함수 (여기서 3변수 분포는 Pearson Type III 분포라고 함), 5) Log Pearson Type III 분포함수, 6) Type I 극치분포함수 (Gumbel 분포함수), 7) Weibull분포함수 (혹은 Type III 극치분포함수)등이 있다.

확률해석에서는 이상의 다양한 확률분포함수의 매개변수를 추정하고, 적합성 분석을 통해 이론확률분포함수를 정하고, 이를 기초로 재현기간별 수문변량의 값을 정하게 된다. 그런데, 수문자료의 통계적 특성을 나타내는 이론확률분포함수가 2개 이상이 되는 경우가 많다. 이 경우, 각각의 분포함수로부터 얻은 재현기간별 수문변량의 값이 다를 수 있다. 따라서, 수문자료의 빈도해석에서 여러 가지 함수를 적용하는데서 오는 문제를 극복하기 위해 보통 통일된 방식을 적용한다.

미국 연방정부(1967)에서는 Log Pearson Type III분포함수를 홍수량 해석에 채택하였는데, 이를 Bulletin 15라고 한다. 1976년에는 Bulletin 17를 발표하고 1981년 수정하였는데, 여기서는 Log-Pearson Type III분포함수의 적용에서 지역적 왜곡계수의 추정치를 보정하도록 하였으며, 지역빈도해석법을 정립하였다. Log-Pearson Type III분포함수는 다음과 같다.

$$f(x) = \frac{1}{\alpha \Gamma(\beta)} \frac{1}{x} \left( \frac{\ln x - y_0}{\alpha} \right)^{\beta-1} \left( -\frac{\ln x - y_0}{\alpha} \right)$$

여기서,  $\alpha, \beta =$  매개변수,  $y_0 =$  위치변수이다.

영국에서는 Gumbell 분포함수의 일종인 일반화 극치(generalized extreme value, GEV)분포함수를 채택하고 있다.

건설부 (1988)은 전국의 확률강우량 추정에서 적정확률분포형을 연최대치 계열에 대해 극치I 분포함수로 제시하였다. 극치I 분포함수 (Extreme Value Type I Distribution Function)의 일반형은 다음과 같다.

$$f(x) = \frac{1}{\alpha} \exp \left( -\frac{x-x_0}{\alpha} - \exp \left( -\frac{x-x_0}{\alpha} \right) \right)$$

여기서,  $\alpha$ 는 축척변수,  $x_0$ 는 최대확률의 발생 위치를 나타낸다.

한편, 특별히 계획 대상 구역의 지점별 확률분포형이 서로 다를 경우로서 지점간 확률강우량의 차이가 클 경우에는 지점별 확률분포형에 따르고, 그 차이가 크지 않을 경우는 많이 나타나는 확률분포형을 채택할 수 있다고 하였다 (건설부, 1993).

### 4. 매개변수의 추정

확률분포함수의 매개변수 추정법에는 L모멘트(moment)법, 최우법 (maximum likelihood method), 확률가중모멘트법 등이 있다. L 모멘트법은 통계적 모멘트와 확률분포함수의 매개변수와의 관계를 이용하는 방법이며, 최우법은 통계적인 최적화방법을 적용하는 것이다. 보통 L 모멘트법이 확률해석에서 널리 쓰고 있으나, 빈도해석 프로그램에서는 다양한 매개변수 추정법을 적용하도록 구성하는 경우가 많다.

### 5. 확률분포함수의 적합성 검정

앞에서 언급한 확률분포함수의 적용에 앞서,

수문자료가 특정 분포함수에 적합한가를 판정해야 하는데, 이를 적합성 검정이라 한다. 적합성 검정법에는 1) 확률지법, 2) Chi square ( $X^2$ )법, 3) Kolmogorov-Smirnov 검정법 등이 있다.

확률지법은 변량  $x$ 와 그 누가확률  $F(x)$ 가 직선적 관계로 표시되도록  $F(x)$ 축의 눈금을 정한 특수 방안이다. 따라서, 수문자료의 순위로부터 도시위치 (plotting position)를 정하고, 이를 확률지에 도시하여 그 점들의 관계가 직선적으로 나타나는지를 판정하면 된다. 여기서, 확률도시 결과의 판정에 적용되는 통계적 기법에는 상관계수법과 Kolmogorov-Smirnov법(이하 K-S법) 등이 있다. K-S법은 각 누가분포함수 도시점에 대한 이론값과 실제값 중의 절대값의 차이가 허용오차 범위 이내에 있는지를 검정한다.

확률도시위치는  $N$ 개의 자료를 큰 값부터 크기 순으로 배열할 때,  $i$ 번째의 수문변량과 같거나 큰 값의 발생확률  $E(X \geq X_i)$ 은

$$E(X \geq X_i) = \frac{i}{N+1}$$

로 정의하거나 (Weibull 도시법), USDA의 NRCS에서는  $\frac{2i}{2N-1}$ 로 하고 있는 등, 몇가지 이론식이 제안되어있다.

L-모멘트법으로부터 특정한 확률분포함수에 해당하는 재현기간별 이론 수문변량을 얻을 수 있는데, 그 일반식은 다음과 같다.

$$X_T = \bar{X} + X_{T\sigma}$$

여기서,  $X_T$ 는 재현기간  $T$ 년의 수문변량,  $X_T$ 는 빈도계수,  $\bar{X}$ 와  $\sigma$ 는 수문변량의 평균과 표준편차이다. 그런데, 빈도계수나 그 밖의 매개변수의 값은 적용하는 확률분포함수에 해당하는 값을 써야하며, 주요 확률분포함수의 값은 참고문헌 (윤용남, 1998)에서와 같다.

한편, Chi-square법과 K-S법에서는 이론

값과 실제값이 통계적으로 동일한 모집단인지를 검정하는 방법이다. 보다 상세한 내용은 수문통계학 또는 통계학 등의 참고문헌에 따르면 한다.

## 6. 홍수량자료의 빈도해석의 기초

수문자료의 빈도해석은 경향성, 주기성, 특이성 등이 나타나지 않는 무작위 변량의 경우로 제한된다. 그리고, 확률분포함수를 적용하므로, 일정이상의 자료기간을 보유하는 것 등 여러 가지 조건을 충족해야 한다. 가장 일반적인 빈도해석의 자료계열로는 연홍수량계열 (연최대치 혹은 연초과치계열),

연강수량계열, 지속시간별 연 최대강우강도계열, 지속시간별 하천의 연 최소 유량 등이다. 그밖에도 부분기간자료계열과 합성자료계열 등이 있다. 이들 자료의 빈도해석에 앞서, 면밀한 검토가 필요한데, 그 내용은 다음과 같다.

### 가. 자료 기간

모든 빈도해석에서는 자료 기간은 충분히 길어야 한다. 최소자료기간은(SCS, 1971)

$$Y_m = (4.30 t_{10} \log R)^2 + 6$$

여기서,  $Y_m$ 은 최소 자료기간 (연),  $t_{10}$ 은 Student t 분포의 10% 유의수준 값, 그리고  $R$ 은 확률도시지에서의 2년과 100년 사상의 값의 비, 혹은

$$R = \frac{100년 사상의 크기}{2년 사상의 크기}$$

이다.

### 나. 연최고치과 연초과치계열

연최고치계열은 연 최대값 만으로 이루어진

자료를 말하며, 연초과치계열은 특정 값 이상의 자료로부터 자료년수 만큼 구성한 자료를 말한다. 연최고치계열은 자료 년수 이상의 재현기간의 홍수량을 정할 때, 그리고 연초과치계열은 그 이하의 재현기간의 값을 추정할 때 쓴다.

홍수량의 빈도해석에서 연초과치 계열과 연최고치계열의 평균재현기간 사이는 다음의 관계가 있다(SCS, 1971).

$$T_E = \frac{1}{\log_e T_M - \log_e (T_M - 1)}$$

여기서,  $T_E$ ,  $T_M$ 은 각각 연초과치계열과 연최대치계열의 평균재현기간이다.

다. 전기간계열과 부분기간계열

전기간계열은 자료년수 만큼을 고려하는 시계열이며, 부분기간계열은 자료년수보다 작은 시계열이다. 전기간계열은 연최고치, 연초과치계열을 적용한다.

그런데, 특정 값 이상의 사상만을 고려할 때 자료 연수보다 작은 경우가 발생한다. 예를 들어, 과거 50년 동안 하천 제방을 월류하는 홍수사상이 12회 만큼 발생할 경우, 부분기간계열로 처리하여 1~12 순위까지 만을 고려하고, 자료기간의 50년으로 하는 방법이 그것이다. 또 다른 예는 홍수로 인해 고수부지의 범람은 매년 발생하지 않을 수 있는데, 이 경우도 연초과치계열을 적용하고, 자료연수 이하의 부분기간계열을 분석할 수도 있다. 이때, 확률도식적의 N값은 전기간으로하여 분석해야 한다.

7. 강우량자료의 빈도해석

강우량자료의 빈도해석은 연폭우사상(연간 최대홍수량을 일으키는 폭우사상으로 반드시 최대 강우량을 갖는 경우와 일치하지 않는다)의 강우량 자료계열, 연 최대 면적강우량자료

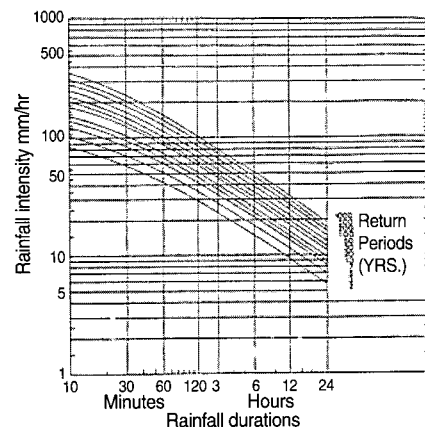
계열, 지속시간별 연 최대강우강도 자료계열 등이 대상이다. 가뭄의 해석에서 연도별 생육기간중 과우일수 자료계열 등의 빈도해석을 적용하는 예도 있다.

점 강우량자료계열의 빈도해석은 홍수량의 경우와 대동소이하다. 다만, 강우자료의 특성이 홍수량과 다르므로, 적정 확률분포함수가 다르기도 한다. 건설부 (1988)의 Type I 극치분포함수가 최적 확률분포함수로 제시한 경우도 그 까닭이다.

건설부 (1993)은 지속시간별 최대강우강도의 빈도해석결과에서 자료계열에 따른 환산계수를 제시하고 있다. 설계빈도가 10년 이하의 경우에 적용되는 것으로, 연초과치계열의 자료 추출이 어려울 때, 연최고치계열로부터 빈도해석을 실시하고, <표-1>의 환산계수를 적용한다.

<표-1> 연초과치 계열 대 연최대치 계열의 환산계수

지속시간	재현기간(년)			비고
	2	5	10	
60분	1.157 (1.136)	1.107 (1.136)	1.033 (1.042)	( ) 안에 나타낸 수치는 미국에서 이용되고 있는 환산계수
24시간	1.029 (1.042)	1.008 (1.010)	1.012 (1.010)	



<그림-1> 서울지점의 강우강도-지속시간-빈도 곡선(건설부, 1988a)

한국건설기술연구원 (건설부, 1988)은 강우 강도-지속시간-생기빈도곡선을 제시하였다. 대상지점의 장기간 자기우량기록지로부터 지속시간별 연최대강우강도 자료계열로부터 단시간의 지속시간별로 연최대치계열을 작성하여 빈도해석하고, 그 결과를 도식적으로 나타낸 것이다(<그림-1> 참조).

## 8. 갈수량의 빈도해석

홍수량과 달리 갈수량의 빈도해석은 최소치계열을 분석대상으로 하고, 확률분포함수의 우측이 아닌 좌측의 값을 적용한다. 갈수량자료계열의 적정확률분포함수도 홍수량과는 달리 대수정규분포함수나 Gumbel 분포함수 등이 더 적합한 것으로 나타난다. 그러나, 특정한 분포함수가 추천된 것이 없는 데, 갈수량 자료의 통계적 특성에 기인하기 때문으로, 각 측정점의 자료의 특성에 적합한 함수를 택하여 적용하는 방식을 취한다.

## 9. 지역빈도해석법

지금까지의 수문자료는 한 측정점의 관측치를 의미한 경우로서 통상 점 빈도해석이라 한다. 그런데, 유역내의 여러 측정점에서의 수문자료 빈도해석의 결과로부터 그 유역의 특성값을 얻어야 하는 경우가 있는 데, 이와 같은 응용방법을 지역빈도해석(Regional frequency analysis)라고 한다.

지역빈도해석의 대표적인 것은 강우자료에서 적용되는 평균우량깊이-유역면적-지속시간(Depth-area-duration)의 자료에 대한 빈도해석이다. 그리고, 유역의 홍수량 지역빈도해석에서는 지표홍수량법(Index-flood method)을 이용하는 예가 있으나, 아직 우리나라에서는 아직 시도되지 않고 있다.

## 10. 기타 빈도해석의 문제

이상으로 수문자료의 빈도해석에 관한 이론과 적용 방법에 대하여 살펴보았다. 빈도해석법은 많은 문헌에 상세히 소개되고 있으므로, 개략적인 내용을 위주로 해설하였다.

그런데, 실무에서 빈도해석과 관련하여 잘못 적용할 수 있는 경우가 많은 데 그 내용을 살펴보면 다음과 같다.

### 가. 설계 강수량로부터 계획홍수량 추정 문제

계획 홍수량의 추정의 가장 적합한 방식은 관측치를 빈도해석해서 얻는 것이다. 그런데, 수공구조물의 계획에서는 홍수기록이 거의 없으므로, 설계 강우량을 얻고 강우-유출관계로부터 추정하는 방식을 적용하는 경우가 많다. 이 때, 강우량자료계열의 빈도해석결과와 홍수량자료의 결과와는 동일 연도의 자료가 상이한 재현기간을 나타낸다. 즉, 강우량과 홍수량의 동일한 재현기간을 갖지 못하게 된다. 합리식에서의 유출계수 C값에서는 재현기간별로 상이한 값을 적용하여 보정한다. 그 결과, 재현기간이 긴 홍수량 추정시 적용하는 유출계수는 1.0보다 큰 값을 나타내는 경우도 있다. 마찬가지로, 강우량 자료계열을 활용하여 홍수량을 추정할 경우는 이에 상응하는 재현기간의 변화를 감안하는 일이 필요하다.

### 나. 설계 저수량의 빈도해석 문제

농업용 저수지의 내용적 설계와 같이 필요 저수량의 결정에서 연도별 모의된 필요 저수량을 빈도해석 할 경우 연도별 저수량 변화 경향에 대한 자기 상관성도 고려하여야 한다. 저수량 변화는 계절적 요인 및 물 관리의 관행과 함께 연도별로 모의된 필요 저수량을 빈도 해석할 경우 연도별 저수량 변화 경향에 대한 자기 상관성도 고려하여야 한다. 저수량 변화는 계

절적 요인 및 물 관리의 관행과 함께 유입량과 유출량의 차이로부터 나타나므로, 초기 저류량은 다음 해의 저류량에 영향을 미치는 경우가 있다. 특히, 대규모 저류시설의 경우에는 자기 상관성을 나타낼 수 있으므로, 이를 소거하지 않으면 자료계열은 빈도해석의 대상이 되지 못할 수 있음에 유의해야 한다.

#### 다. 비초과 연 계열치 및 확률 2.23년 혹은 평년 빈도

하천정비계획 등에서는 연2회 정도의 생기 빈도를 갖는 홍수량을 계획 홍수량으로 정한다. 이때, 재현기간 0.5년 홍수량이라고 하는 것은 확률이 2인 경우를 말하므로, 잘 못된 정의이다. 이 경우, 연2회 생기 빈도 홍수량이라고 정의해야 한다. 이 경우는 자료년수보다 많은 수의 자료를 빈도해석의 대상으로 해야하므로 비연초과치 계열문제라고 한다.

#### 참고문헌

1. 건설부, 1988. 한국확률강우량도의 작성, 부록: 한국 확률강우량도. 한국건설기술연구원, 서울.
2. 건설부, 1993. 하천시설기준. 서울. p.981.
3. 건설부, 1993. 댐시설기준, 서울.
4. 윤용남, 1998. 공업수문학, 청문각. p.656.
5. USDA NRCS, 1971. Hydrology. SCS National Engineering Handbook Section 4, Chapter 18.
6. Stedinger, J. R., R. M. Vogel, and E. Foufoula-Georgiou, 1993. Frequency Analysis of Extreme Events. D. R. Maidment (ed.), Handbook of Hydrology, Chapter 18, McGraw-Hill Inc.
7. Interagency Advisory Committee on Water Data, 1982. Guidelines for Determining Flood Flow Frequency. Bulletin 17B. USDI, USGS, Office of Water Data Coordination, Reston, Va.