

極值頻度分析에 있어서 Outliers의 取扱에 관하여

李 淳 赫

(忠北大學校 農科大學)

1. 緒 言

自然의 水文現象에서 얻어지는 各種 水文資料를 기본으로 水利構造物의 設計水文量을 결정하기 위해서는 우선적으로 관측된 資料의 特性에 근사한 確率分布型을 찾아냄으로써 기대하는 再現期間년 設計水文量을 推定하게 된다. 이를 위해서는 水文資料 系列 중에서 目的하는 時系列을 選定하고 適正確率分布型을 구한후 이에 관한 統計學的 特性들의 誘導와 함께 궁극적인 設計水文量의 推定이 가능하게 되는 것이다. 그러나 일련의 水文資料系列에서 資料의 일반적인 均衡分布의 상태로부터 훨씬 위 또는 아래로 격리되는 Data point를 나타내는 Outliers의 유발로 인해 不適切한 統計學的 媒介變數值들이 유도될 수 있으며 이로써 適正한 設計水文量의 提示에 不確實性을 종종 나타나게 한다. 이는看過할 수 없는 設計水文上의 重大事로서 本稿에서는 上記한 Outliers의 高低偏倚시 特別히 문제가 되는 高位 Outliers의 處理分析方法에 관해서 Log-Pearson Type III의 分布型을 例로하여 重點의으로 다루고자 한다.

2. Outliers의 取扱

가. Station Skew 와 Generalized Skew Coefficient

資料의 Outliers에 대한 除去, 혹은 修正은 各種 水文資料系列로 부터 계산된 統計學的 媒介變數들에 커다란 影響을 미치게 한다. 그러므로 이들 Outliers를 取扱하는 節次는 數學的이며 水文統計學的 思考를 겸하는 최종적인 分析이 要求된다. 特別히 一般化 歪曲度係數의 이용은 Outliers의 影響을 減少시킬 수 있다. 여기에서 一般化 歪曲度係數(Generalized Skew Coefficient)라 함은 對象流域의 觀測所周圍에 散在한 적어도 25年 이상의 記錄을 갖고있는 觀測所群의 관측치로 부터 추정되는 係數를 의미한다. 觀測所의 Skew Coefficient가 正이면 資料 프루팅의 均衡化 傾向에서 편기된 Data Point는 우선적으로 高位의 Outliers로 看做되며 負이면 低位의 Outliers로서 看做된다. 만약에 Skew Coefficient의 값이 거의 零에 근접되면 高位 및 低位의 Outliers로 동시에 생각할 수 있다. 어느 觀測所의 觀測資料에 의한 Skew Coefficient는 極值水文事象인 경우 特別히 민감한 反應을 나타내기 때문에 소수의 水文資料로 부터 精確한 추정치를 이끌어낸 매우 어렵게 되어 일반적으로 단기간 기록에 대한 歪曲度係數를 추정할 시에는 一般化歪曲度係數를 사용할 것을 권장하고 있다.

그러나 觀測所의 資料觀測年限이 길면 길어질수록 實測資料로 부터 계산된 歪曲度는 매우 신빙성이 커지게 되므로 100年 이상의 水文資料가 가능하다면 一般化歪曲度係數는

필요없고 관측소 歪曲度係數를 사용함이 바람직하다.

만약에 資料의 관측년수가 25年에서 100年 사이의 기록치라면 관측소왜곡도 加重值, $\frac{N-25}{75}$ 를 구한후 이를 이용하여 一般化歪曲度係數의 加重值, $1 - \frac{(N-25)}{75}$ 를 구한후 Log化된 왜곡도계수와 Station Skew의 加重值 및 Generalized Skew와 Generalized Skew의 加重值와의 積의 合으로써 最終的인 加重된 Skew Coefficient의 求值가 가능하게 된다. 上記式에서 제시된 N은 記錄年數를 의미한다. 前述한 Log-Pearson Type III 分布의 基本統計值인 平均値를 除外한 標準偏差와 觀測所歪曲度係數는 다음의 식(1) 및 (2)와 같다.

$$S = \left[\frac{\sum(X - \bar{X})^2}{N - 1} \right]^{0.5} \dots\dots\dots (1)$$

$$G = \frac{N \sum(X - \bar{X})^3}{(N - 1)(N - 2)S^3} \dots\dots\dots (2)$$

- 여기서 N : 水文資料系列의 數
- \bar{X} : 對數平均
- S : 對數化된 標準偏差
- G : 對數化된 歪曲度係數

나. Historic Data의 適用

連續記錄 이전에 간헐적으로 記錄된 最大洪水量의 觀測值가 있을시에는 이들은 近年까지 연속하여 記錄된 觀測值들과 함께 設計水文量의 추구에 보다 효율적으로 이용될 수 있다. 上記한 간헐적인 과거 最高기록치를 고려한 加重值와 함께 例로 제시된 Log-Pearson Type III 頻度曲線에 적용할 수 있는 解析방법은 다음과 같다.

1) 연속적인 관측기록 이전의 간헐적인 최대수문사상들을 갖는 가장 오래된 해를 始發年으로하여 近年까지의 연속관측된 年數를 H라하고 이중 연속 및 불연속적인 最大水文事象들의 數를 Z라 할때 이들에 관한 加重值를 1로하여 남은 連續記錄이 되는 N개의 水文

事象들은 이들에 관한 分布가 (H-Z) 期間을 代表한다는 가정에서 (H-Z)/N의 加重值가 주어진다.

2) Historic Data를 이용한 基本統計值의 修正에 관한 계산은 다음의 式(3), (4), (5) 및 (6)을 사용하여 各年の 水文資料에 대한 加重統計值를 구함으로써 高位 Outlier의 修正이 가능하게 된다.

$$W = \frac{H-Z}{H} \dots\dots\dots (3)$$

$$\bar{M} = \frac{W \sum X + \sum Xz}{H} \dots\dots\dots (4)$$

$$\bar{S}^2 = \frac{W \sum (X - \bar{M})^2 + \sum (Xz - \bar{M})^2}{H - 1} \dots\dots (5)$$

$$\bar{G} = \frac{H}{(H - 1)(H - 2)} \frac{[W \sum (X - \bar{M})^3 + \sum (Xz - \bar{M})^3]}{\bar{S}^3} \dots\dots(6)$$

- 여기서
- W : N개의 水文事象에 적용되는 加重係數
- H : 總記錄年數
- Z : 連續記錄年 이전의 最大水文事象記錄年數
- N : 連續記錄年數
- Xz : 連續 및 不連續記錄年중 最大水文事象值의 對數

- \bar{M} : 總記錄年數에 대한 加重平均値
- \bar{S}^2 : 總記錄年數에 대한 加重分散
- \bar{G} : 總記錄年數에 대한 加重歪曲度係數

3) 基本統計值가 이미 近年까지 계속된 連續記錄值들 만으로 계산되었다면 이 統計值들은 前述한 式(3)과 다음의 式(4'), (5') 및 (6')를 사용하여 不連續的인 記錄과 近年까지의 連續記錄에 대한 각각의 加重值를 주어 總記錄年에 대한 加重平均値, 加重平均偏差, 加重歪曲度係數를 각각 구해야 한다.

$$M' = \frac{W \cdot N \cdot \bar{M} + \sum Xz}{H} \dots\dots\dots (4')$$

$$\bar{S}'^2 = \frac{W(N-1)S^2 + WN(M - \bar{M})^2 + \sum (Xz - \bar{M})^2}{H - 1} \dots (5')$$

$$\bar{G}' = \frac{H}{(H-1)(H-2)S^3}$$

$$\frac{[W(N-1)(N-2)S^3G + 3W(N-1)(M-\bar{M})S^2 + WN(M-\bar{M})^3 + \sum(Xz-\bar{M})^3]}{\dots\dots\dots(6')}$$

4) 加重된 頻度曲線의 作圖는 式(7)에 의해서 求解된 再現期間별 確率洪水量을 對數 確率紙상에 프룻팅한다.

個個 水文事象에 대한 加重 프룻팅포지션 들은 式(8)에 의해 계산된다.

$$\text{Log } Q = \bar{M} + K\bar{S} \dots\dots\dots(7)$$

$$\overline{PP} = \frac{\bar{m} - a}{H + 1 - 2a} \times 100 \dots\dots\dots(8)$$

여기서 K : Skew Coefficient에 따른 頻度 係數

a : 프룻팅포지션 공식에 따른 常數로서 Weibull公式에서 a=0, Beard公式에서 a=0.3 Hazen公式에서 a=0.5이다.

또한 각각의 水文事象의 加重順番인 m는 다음의 式(9) 및 (10)에 의해서 구한다.

$$1 \leq E \leq Z \text{ 일때 } \bar{m} = E \dots\dots\dots(9)$$

$$(Z+1) \leq E \leq (Z+N) \text{ 일때 } m = WE - (W-1)(Z+0.5) \dots\dots\dots(10)$$

여기서 m̄ : 프룻팅포지션을 구할시 사용되 는 각 水文事象의 加重順番

E : 水文事象을 크기순으로 配列시 番號로서 1로부터 (Z+N) 까 지이다.

다. 高位 Outliers의 調整例

1) 標本流域의 極值系列

極值系列중 年最大值系列(Annual Maximum Series)을 택하고 記錄年數 및 年最大流量值 는 다음의 Table-1로 가정한다.

2) 基本統計値와 高位 Outlier의 검증

上記 Table-1의 39년간의 記錄年數에 대

Table-1. Annual Maximum Series in Example

Year	Annual Max. Discharge	Year	Annual Max. Discharge
1935	1460	1946	1400
1936	4050	1947	3240
1937	3570	1948	2710
1938	2060	1949	4520
1939	1300	1950	4840
1940	1390	1951	8320
1941	1720	1952	13900
1942	6280	1953	71500
1943	1360	1954	6250
1944	7440	1955	2260
1945	5320	1973	5660

한 對數平均値, 對數標準偏差, 對數歪曲度係數는 各各 3.555, 0.464 및 0.357로 구하여 졌다. 그리고 標本流域의 一般歪曲度係數를 -0.2로 假定한다면

$$\begin{aligned} \text{觀測所歪曲度の 加重値} &= 0.187 \\ \text{一般化歪曲度の 加重値} &= 1 - 0.187 = 0.813 \\ \text{加重歪曲度係數} &= 0.187(0.357) + 0.813(-0.2) \\ &= -0.096 \approx -0.1 \end{aligned}$$

따라서 計算된 對數歪曲度係數가 正이므로 Historic Data를 적용시에는 高位의 Outliers가 있을 수 있음을 示唆하고 있다. 本例의 水文資料들중 1953년의 年最大洪水量, 71,500 CMS는 Fig. 1에서 보는바와 같이 Weighted Skew와 Station Skew에 의한 洪水頻度曲線의 兩曲線 보다도 실제상으로 他 프룻팅포지션에 비해 훨씬 상부에 위치하고 있으므로 上記한 水文事象은 高位의 Outlier로서 확인검 증될 수 있다. 여기에서 實測水文事象들의 프룻팅포지션은 Weibull법에 의해서 계산된 洪水頻度曲線에 연해 프룻트되었다.

上記와 같은 계산과 함께 관측치 및 加重 Skew와 觀測所 Skew에 따른 洪水頻度曲線을 나타낸 결과는 Fig. 1과 같다.

3) 高位 Outliers의 調整

가) 基本統計値의 再計算

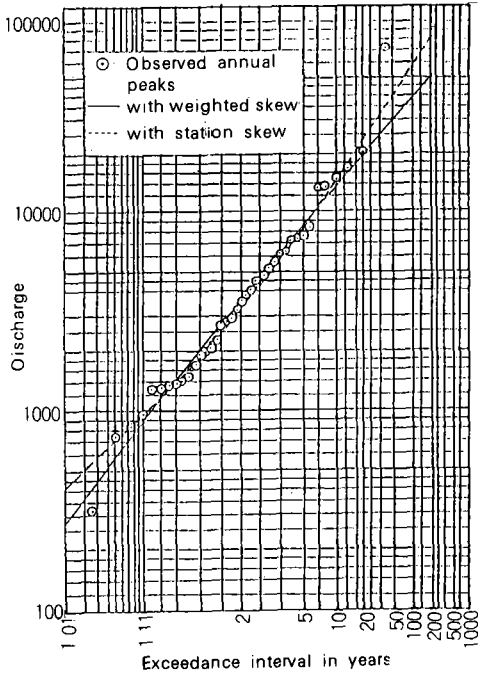


Fig. 1. Flood Frequency Curves

高位 Outlier로 판정된 1953년의 71,500 CMS의 값은 除外하고 基本統計值를 다시計算하여야 하며 여기서 구해진 統計值들은 다음과 같다.

- 對數平均値 : 3.521
- 對數標準偏差 : 0.418
- 對數歪曲度係數 : -0.095
- 年數 : 38

나) 統計值 修正을 위한 Historic Data의 이용
 本例에서의 Historic Period (H)는 1892~1973년의 82년으로 가정한다. 그리고 連續記録年數, N은 1953년을 제외한 1935~1973년 간의 38년이 된다. 또한 상술한 82년간에 큰 水文事象數, Z는 1953년의 水文事象인 1로 한다. 따라서 加重係數, W와 總記録年에 대한 加重平均値, M'의 값은 다음과 같다.

$$W = \frac{82-1}{38} = 2.1316$$

$$W \cdot N \cdot M = 285.204$$

$$\sum Xz = 4.854$$

$$M' = 290.058/82 = 3.537$$

그리고 전술한 式 (5'), (6')에 의거하여 구한 加重標準偏差, \bar{S}' 와 加重歪曲度係數, \bar{G}' 는 각각 다음과 같다.

$$\bar{S}' = 0.438, \bar{G}' = 0.180$$

따라서 最終的인 一般化加重歪曲度係數는 Station Skew의 加重値와 加重歪曲度係數의 積 그리고 一般化歪曲度係數의 加重値와 標本流域의 一般化歪曲度係數의 積과의 總和로서 구해질 수 있으며 그 結果値는 다음과 같다.

$$0.187(0.180) + 0.813(-0.2) = -0.129$$

$$\approx -0.1$$

다) 加重프롯팅포지션의 계산

水文事象중 最大値의 順位 $m=1$ 이고 그 다음 크기들에 관한 事象들의 加重順位는 式 (10)에 의거하여 구할 수 있으며 本例에서 사용한 Weibull法에서 $a=0$ 이므로 이에의한 프롯팅포지션은 다음의 式 (11)이 된다.

$$\overline{PP} = \frac{\bar{m}}{H+1} \times 100 \dots\dots\dots (11)$$

따라서 크기순 10년에 걸친 Weibull 프롯팅포지션의 계산결과와 예는 다음의 Table-2와 같다.

Table-2. Computation of weighted plotting positions

Year	Annual Max. Discharge	Weighted factor (W)	Order (\bar{m})	Weighted order (m')	Weibull plotting position (\overline{PP})
1953	71500	1.0000	1	1.00	1.20
1962	20600	2.1316	2	2.57	3.09
1969	17300	2.1316	3	4.70	5.66
1960	15100	2.1316	4	6.83	8.23
1952	13900	2.1316	5	8.96	10.80
1971	13400	2.1316	6	11.09	13.36
1951	8320	2.1316	7	13.22	15.93
1965	7500	2.1316	8	15.36	18.50
1944	7440	2.1316	9	17.49	21.07
1966	7170	2.1316	10	19.62	23.64

라) 修正된 洪水頻度曲線의 誘導

一般化加重歪曲係數 -0.1 을 이용한 超過確率별 頻度係數에 의해서 구한 再現期間별 超過確率洪水量값은 다음의 Table-3과 같으며 이에 관한 최종적인 修正된 洪水頻度曲線은 Fig. 2와 같다.

Table-3. Computation of Frequency Curve Coordinates

Exceedance Probability (P)	Frequency Factor (K)	Discharge (Q)
0.99	-2.39961	306
0.90	-1.29178	936
0.50	0.01662	3500
0.10	1.27037	12400
0.05	1.61594	17600
0.02	1.99973	25900
0.01	2.25258	33400
0.005	2.48187	42100

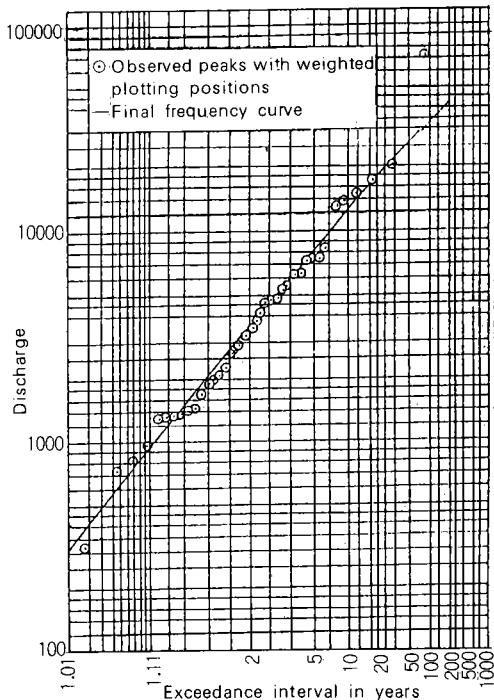


Fig. 2. Final Flood Frequency Curve

參 考 文 獻

1. Beard, L. R., Flood Flow Frequency Techniques, Center for Research in Water Resources, The Univ. of Texas at Austin, 1974.
2. Benson, M. A., Evaluation of Methods for evaluating the Occurrence of Floods, USGS Water Supply Paper, 1580-A, 1962
3. _____, Factors Influencing the Occurrence of Floods in a Humid Region of Diverse Terrain, USGS Water Supply Paper 1580-B, 1962.
4. Carrigan, P. H., Jr., and C. S. Huzzen, Serial Correlation of Annual Floods, International Hydrology Symposium, Fort Collins, 1967.
5. Corps of Engineers, U. S. Army, Storm Rainfall in the United States, Washington, 1945.
6. Chow, V. T., Handbook of Applied Hydrology, McGraw-Hill Book Co., New York, 1964, pp.8~28, 8~29.
7. Dalrymple, T., Flood Frequency Analyses, Part 3, Flood Flow Techniques, USGS Water Supply Paper 1543-A, 1960.
8. Gann, E. E., Generalized Flood-Frequency Estimates for urban Areas in Missouri, USGS Open-File Report, 1971.
9. Gumbel, E. J., The Calculated Risk in Flood Control, Applied Science Research, Section A, Vol.5, 1955.
10. Haan, C. T., Statistical Method in Hydrology, I.S.U. Press, Ames, 1977.
11. Hardison, C. H., Generalized Skew Coefficients of Annual Floods in the United States and Their Application, W.R.R., Vol.10, No. 5, pp.745~752.
12. Harter, H. L., Some Optimization Problems in Parameter Estimation, edited by Jagdish S. Rustagi, Optimizing Methods in Statistics, Academic Press, New York, 1971.
13. 角屋 陸, 水流量のPlotting Position について, 京都大學防災研究所年報 第3號,

- 1959.
14. Kadoya, Mutsumi, Application of Extreme value distribution in Hydrologic Frequency Analysis, Kyoto Univ., Bulletin No.66, 1964.
 15. 李淳赫, 韓重錫, 極值流量的最適分布型과 極值確率流量에 관한 研究, 韓國農工學會誌, 22(1), 1980
 16. 李淳赫, 朴明根, 年超過值系列의 洪水頻度分析에 관한 研究, 韓國農工學會誌, 24(1), 1982.
 17. Matalas, N. C., Autocorrelation of Rainfall and Streamflow Minimums, U.S. Geological Survey Prof. Paper 434B, 1963.
 18. U. S. Water Resources Council, Guide Lines for Determining Flood Flow Frequencies, 1967.
 19. Yevjevich, Vujica, Probability and Statistics in Hydrology, Water Resources Publications,