

加重回歸分析에 의한 地域化歪曲係數의 推定

Estimation of Regional Skew Coefficient with Weighted Least Squares Regression

趙 國 光* · 權 純 國**
Cho, Kuk Kwang · Kwun, Soon Kuk

Summary

The application of the Log-Pearson Type III distribution recommended by Water Resources Council, U. S. A. for flood frequency analysis requires the estimation of the regionalized skew coefficient. In this study, regionalized skew coefficients are estimated using a weighted regression model which relates at-site skews based on logarithms of observed annual flood peak series to both basin characteristics and precipitation data in the Han river and the Nakdong river basin. The model is developed with weighted least squares method in which the weights are determined by separating residual variance into that due to model error and due to sampling error. As the result of analysis, regionalized skews are estimated as -0.732 and -0.575 in the Han river and the Nakdong river basin, respectively.

I. 緒 論

홍수터 관리를 위한 潛在洪水被害의 推定이나 水資源開發事業의 計劃 및 設計에 있어서 하나의 중요한 단계는 洪水頻度を 算定하는 것이며 이는 주로 美國聯邦水資源委員會(Water Resources Council : WRC)가 추천한 標準洪水頻度分析方法인 年最大洪水量系列의 對數值에 Pearson Type III 分布를 적용시키므로써 算定된다. 그러나 상대적으로 짧은 記錄值를 사용하여 歪曲係數를 推定하게 될 경우에는 標本誤差의 影響을 크게 받게 된다.

歪曲係數 推定量을 개선하기 의하여 美國의

WRC¹⁶⁾는 洪水頻度 算定을 위한 指針書에서 洪水資料의 對數值를 이용한 地域 및 地點歪曲係數의 加重推定量을 사용하도록 추천하였다. Tasker¹⁷⁾는 보다 개선된 加重方法을 제안하였으며, Tung and Mays^{14, 15)}는 여러가지 加重方法들을 비교 평가하였다. 그 후 폐쇄된 WRC의 업무를 이어받은 Interagency Advisory Committee on Water Data⁶⁾ 内の Hydrology Subcommittee는 지침서를 개정하면서 인근의 觀測點들로부터 蒐集된 情報를 基礎로한 地域化 혹은 一般化歪曲係數를 使用하도록 추천하였다. Tasker and Stedinger¹³⁾는 Stedinger and Tasker⁸⁾와 Tasker¹²⁾에 의해서 개발된 加重最小自乘 回歸推定量들

* 水原專門大學 土木科

**서울大學校 農科大學

키워드 : 水文統計, 洪水頻度分析, 年洪水尖頭值系列, Log Pearson III型 分布, 加重回歸分析, 地域化回歸模型, 加重回歸模型, 地域化歪曲係數.

에 대한 연장으로서 觀測된 年最大洪水量 系列의 對數值들로부터 算定된 標本歪曲係數들을 기초로한 歪曲係數의 地域化回歸推定量을 유도하는 방법을 제시하였다. 이 방법은 標本の 크기가 地點마다 변할때 回歸模型의 媒介變數들에 대한 보다 더 精確한 推定量과 보다 더 현실적이고 精確한 推定의 分散에 대한 推定量을 제공해준다.

本 研究에서는 Tasker and Stedinger¹³⁾의 방법을 利用하여 한강 및 낙동강 流域의 流量觀測點에서 蒐集된 年最大洪水量 系列의 對數值들로부터 算定된 標本歪曲係數들을 流域特性 및 降雨量資料와 연관시키는 地域化回歸模型을 加重回歸分析法에 의하여 개발하므로써 各 流域에서의 地域化歪曲係數 推定量을 算定하였다.

II. 基本理論

$g = (g_1, g_2, \dots, g_n)^T$ 는 어느 地域內의 n 개 觀測點에서 歪曲係數의 참값을 나타내는 $(n \times 1)$ 벡터이며, \hat{g} 은 g 의 不偏벡터推定量을 나타낸다고 하자. Hydrology Subcommittee의 지침서에서는 g 의 各 성분 g_i 를 식(1)에 의하여 推定하도록 추천하였다.

$$G_i = \frac{N_i \sum_{t=1}^{N_i} (Z_{it} - \bar{Z}_i)^3}{(N_i - 1)(N_i - 2)S_i^3} \dots \dots \dots (1)$$

여기서 N_i 는 觀測點 i 에서 觀測된 年最大洪水量의 數이고, Z_{it} 는 t 년에 觀測點 i 에서의 年最大洪水量의 對數值이며, \bar{Z}_i 는 觀測點 i 에서의 年最大洪水量 對數值의 平均值이고, S_i 는 觀測點 i 에서의 年最大洪水量 對數值의 標準偏差이다.

G_i 는 일반적으로 g_i 에 대한 偏倚推定量이므로 $|g_i| < 1.0$ 이고 $N_i \geq 20$ 일때 g_i 에 대한 不偏推定量 \hat{g}_i 은 補正因子를 利用하여 식(2)와 같이 주어질 수 있다.

$$\hat{g}_i = \left(1 + \left(\frac{6}{N_i} \right) \right) \cdot G_i \dots \dots \dots (2)$$

\hat{g}_i 이 g 의 不偏推定量이 되기 위해서는 식(3) 및 (4)의 관계가 성립된다.

$$E(\hat{g}) = g \dots \dots \dots (3)$$

$$E[(\hat{g} - g)(\hat{g} - g)^T] = \sum(\hat{g}) \dots \dots \dots (4)$$

여기서 $\sum(\hat{g})$ 은 대각선上的 g_i 에 관한 \hat{g}_i 의 標本分散 $\text{Var}(\hat{g}_i)$ 을 나타내는 對角線行列로서 加重回歸分析法에서는 g 에 대한 벡터推定量 \hat{g} 의 標本공분산 行列의 합리적인 근사치를 의미하는 것으로 생각된다.

無作爲로 선정된 觀測點 i 에서의 推定量 \hat{g}_i 이 流域特性因子들과 線型的인 관계가 있다고 가정하면 식(5)와 같은 回歸模型을 생각할 수 있다.

$$\hat{g} = X\beta + \epsilon \dots \dots \dots (5)$$

여기서 X 는 $(k-1)$ 개의 流域特性因子로 구성되며 單位列벡터가 첨가된 $(n \times k)$ 行列이고, β 는 推定될 예정인 미지의 係數들로 구성된 $(k \times 1)$ 벡터이며, ϵ 은 $(n \times 1)$ 인 잔차벡터로서 식(6) 및 (7)과 같은 平均과 分散을 갖는 것으로 期待된다.

$$E(\epsilon) = 0 \dots \dots \dots (6)$$

$$E(\epsilon\epsilon^T) = \gamma^2 I_n + \sum(\hat{g}) = \Lambda(\gamma^2) \dots \dots \dots (7)$$

여기서 γ^2 는 模型誤差分散이고, I_n 은 $(n \times n)$ 單位行列이며, $\sum(\hat{g})$ 은 觀測點 i 에서 平均에 대한 \hat{g}_i 의 標本分散인 $\text{Var}(\hat{g}_i)$ 을 要素로하는 對角線行列이다.

$\Lambda(\gamma^2)$ 를 알 수 있다면, β 의 最小分散不偏推定量은 加重平方和를 最小로 해주므로써 식(8)에 의해서 求解될 수 있다.

$$\hat{\beta} = (X^T \Lambda^{-1} X)^{-1} X^T \Lambda^{-1} \hat{g} \dots \dots \dots (8)$$

年最大洪水量의 對數值가 正規分布를 한다고 할 경우, $\text{Var}(\hat{g}_i)$ 은 식(9)에 의하여 推定될 수 있다.

$$\text{Var}(\hat{g}_i) = \frac{6N_i(N_i-1) \left[1 + \left(\frac{6}{N_i}\right)^2\right]}{(N_i-2)(N_i+1)(N_i+3)} \dots\dots(9)$$

한편, 年最大洪水量의 對數值가 Pearson Type III分布를 할 경우 $\text{Var}(\hat{g}_i)$ 은 식(10)에 의하여 推定될 수 있다.

$$\text{Var}(\hat{g}_i) = \left[1 + \left(\frac{6}{N_i}\right)\right]^2 \cdot 10^{(a-b \cdot \log(N_i/10))} \dots(10a)$$

$$a = \begin{cases} -0.33 + 0.08 |\bar{G}|, & |\bar{G}| < 0.90 \\ -0.52 + 0.30 |\bar{G}|, & |\bar{G}| > 0.90 \end{cases} \dots(10b)$$

$$b = \begin{cases} 0.94 - 0.26 |\bar{G}|, & |\bar{G}| < 1.50 \\ 0.55, & |\bar{G}| > 1.50 \end{cases} \dots(10c)$$

식(10)에서 \bar{G} 는 標本地點歪曲係數 \hat{g}_i 가 아니고 地域平均歪曲係數이다.

模型誤差, γ^2 는 다음의 式(11)을 이용하여 반복계산 하므로써 얻어질 수 있다.

$$\begin{aligned} & (\hat{g} - X\hat{\beta})^T \left[\hat{\gamma}_{wls}^2 I_n + \sum (\hat{g})^{-1} (\hat{g} - X\hat{\beta}) \right] \\ & = n - k \dots\dots\dots(11) \end{aligned}$$

여기서 γ^2 의 推定量, $\hat{\gamma}^2$ 은 普通最小自乘法에 의하여 계산된 殘差平均平方誤差와 유사한 一般化殘差平均平方誤差이며, $\hat{\beta}$ 은 式(8)에 의해서 주어진다.

$X_0 = (1, X_{01}, X_{02}, \dots, X_{0K})$ 를 어느 特定 地點의 流域特性을 나타내는 벡터라하면 回歸模型에 의해서 推定된 그 地點에서의 一般化歪曲係數 값은 $X_0\hat{\beta}$ 이 된다. 期待值 $X_0\hat{\beta}$ 대한 推定量의 標本分散은 式(12)와 같다.

$$\text{Var}(X_0\hat{\beta}) = X_0(X^T\hat{\Lambda}^{-1}X)^{-1}X_0^T \dots\dots\dots(12)$$

따라서 미지의 歪曲係數, g_0 에 대한 $X_0\hat{\beta}$ 의 分散은 式(13)과 같이 되므로

$$\gamma^2 + \text{Var}(X_0\hat{\beta}) \dots\dots\dots(13)$$

推定の 分散에 대한 加重回歸推定量은 式(14)와 같이 주어진다.

$$\gamma^2 + X_0(X^T\hat{\Lambda}^{-1}X)^{-1}X_0^T \dots\dots\dots(14)$$

III. 使用資料

1. 流量資料

과거에 閉鎖되었거나 1986年 현재까지 建設部에서 運營해오고 있는 한강 및 낙동강 流域의 流量觀測點 중에서 저수지 水門操作의 영향을 받지 않는 觀測點으로서 5年 以上の 流量資料를 구할 수 있는 地點만을 選定하여 漢江流域은 16개 地點을, 낙동강 流域은 28개 地點을 分析對象으로 하였으며 이들로부터 年最大洪水量資料를 추출하여 自然對數를 취한 후 地點歪曲係數를 구하여 回歸模型의 從屬變數로 하였다. 각 流域의 分析對象 觀測所名 및 記錄值年數는 Table-1과 같으며 觀測所들의 위치는 Fig. 1 및 2와 같다.

2. 流域特性

對象流域의 流域特性을 나타내는 파라메타는 集水面積, 河川頻度, 主河川의 河床傾斜, 主河川長 및 貯留指數 등이며 이들의 값은 1:25,000 및 1:250,000 地形圖를 이용하여 구하였다.

가. 集水面積 (AREA)

各 觀測點別 集水面積은 主要 流域報告書와 流量測定調査 報告書등을 이용하여 구하였으며, 面積이 불확실하거나 報告書에 나타나지않는 觀測點들에 대해서는 1:250,000 地形圖와 求積器로 面積을 구하였다.

나. 河川頻度 (STMFRQ)

河川頻度는 1:25,000 地形圖上에서 識別이 가능한 모든 河道마다 河川次數를 記入한 후 各 觀測點의 集水區域內에 있는 河川次數別 河道의

Table-1. Streamflow-gaging stations in the Han and Nakdong river basin.

River basin	Station	Name
Han	1. Gwangjang(24)	9. Geoun(23)
	2. Goan(46)	10. Moggye(62)
	3. Yeosu(58)	11. Ganhyeon(21)
	4. Danyang(56)	12. Chungju1(48)
	5. Hongcheon(23)	13. Chungju2(15)
	6. Hoengseong(24)	14. Yeongweo11(54)
	7. Hupo(22)	15. Jeongseon1(37)
	8. Pyeongchang(28)	16. Jucheon(24)
Nakdong	1. Yerimgyo(19)	15. Imhal(24)
	2. Jeongam(31)	16. Imha2(22)
	3. Changri(22)	17. Bonghwa(23)
	4. Gaejin(18)	18. Hyunpung(54)
	5. Goryeonggyo(16)	19. Songriweon2(16)
	6. Dongchon(54)	20. Goryeong(24)
	7. Waegwan(52)	21. Keumho(24)
	8. Seonsan(20)	22. Gimcheon(36)
	9. Yonggok(20)	23. Nakdong(56)
	10. Ian(18)	24. Andong1(46)
	11. Jeomchon(21)	25. Jindong(52)
	12. Dalji(23)	26. Susan(53)
	13. Sanyang(18)	27. Masuweon(31)
	14. Weolpo(18)	28. Habcheon(24)

Note : numbers in the () are years of record length.



Fig. 1. Location of streamflow-gaging stations in the Han river basin.

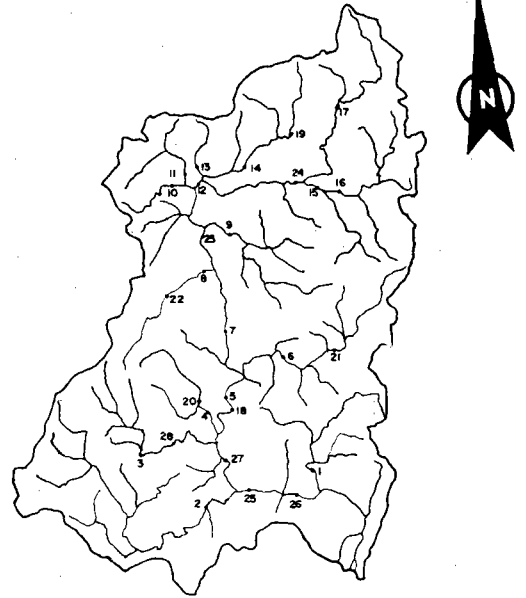


Fig. 2. Location of streamflow-gaging stations in the Nakdong river basin.

數를 모두 더한 값을 해당 觀測點의 集水面積으로 나누어 구하였다.

다. 河床傾斜 (SLOPE) 및 主河川長 (MSL)
主河川の 河床傾斜로는 極高低 기울기를 배제하기 위하여 觀測點으로부터 해당 集水區域의 最遠點까지의 主河川長의 10% 및 85% 되는 두點사이의 傾斜인 S1085¹⁾를 사용하였다.

라. 貯留指數 (LAKE1, LAKE2)
流域內에 있는 貯留지가 洪水量에 미치는 영향을 나타내는 지수로서 貯留지의 滿水面積이 排水面積의 1% 이상인 경우에만 適用한다.³⁾ LAKE1은 觀測點의 集水區域內에 있는 貯留지의 排水面積의 合을 集水面積으로 나눈 값이고, LAKE2는 觀測點의 集水區域內에 있는 貯留지의 滿水面積의 合을 集水面積으로 나눈 값이며, 分析對象 貯留지로는 韓國水資源公社의 韓國電力公社에서 管轄하는 大陂과 市·郡 및 農地改良組合에서 管轄하는 中·소규모 貯留지를 택하였다.

3. 降雨量資料

降雨量資料로는 年平均等雨量線圖³⁾를 이용하여 각 觀測點의 集水區域內에서 等雨線 사이의 면적을 加重平均하여 구한 年平均降雨量(AAP)을 사용하였다.

IV. 加重回歸模型의 開發

加重回歸模型은 한강 및 낙동강유역을 통합하여 하나의 유역으로 간주하여 구해지며 流域特性 및 降雨量資料 이외에 두 유역을 구분하기 위한 變數 Z를 獨立變數로 추가하여 한강유역은 Z=0, 낙동강유역은 Z=1로 하였다. 加重回歸의 從屬變數는 年最大洪水量系列의 대수치의 地點歪曲係數에 偏倚 補正因子 $[1+(6/N_i)]$ 를 곱해준 값으로 하였으며 여기서 N_i 는 i 관측점의 記錄值年數이다. 回歸模型은 다음의 식(15)와 같이 구성되며 段階的回歸에 의하여 最良回歸方程式을 구할 수 있다.

$$\begin{aligned} \text{SKEW} = & a_0 + a_1(Z) + a_2 \ln(\text{AREA}) \\ & + a_3 \ln(\text{STMFRQ}) + a_4 \ln(\text{SLOPE}) \\ & + a_5 \ln(\text{MSL}) + a_6 \ln(\text{AAP}) \\ & + a_7 \ln(\text{LAKE1}) + a_8 \ln(\text{LAKE2}) \dots (15) \end{aligned}$$

여기서

- Z : 두 유역 사이의 歪曲係數의 연속성을 고려하기 위한 假變數
- AREA : 각 觀測點의 集水面積(km²),
- STMFRQ : 각 관측점의 集水區域의 河川頻度(km⁻²),
- SLOPE : 각 集수구역내의 主河川의 平均河床傾斜(m/km),
- MSL : 각 集수구역내의 主河川長(km),
- AAP : 각 集수구역내의 年平均降雨量(mm),
- LAKE1 및 LAKE2 : 貯留指數.

V. 結果 및 考察

標本歪曲係數의 추정의 分散이 最小가 되도록 獨立變數들을 조합하여 加重回歸分析을 한 결과 7개의 模型이 구성될 수 있으며 각 모형에 대한 回歸係數, 및 오차분산 값은 Table-2에 요약되어 있다. 7개의 模型 중에서 模型誤差分散이 가장 작으면서 推정의 分散이 작은 것을 선정한 결과 模型 6이 最良回歸模型이 되며 방정식은 식(16)과 같다.

Table-2. Result of weighted least squares regression for regionalized skew.

Step No.	Constant	Coefficient of								Error variances		
		Z	AREA	STMFRQ	SLOPE	MSL	AAP	LAKE1	LAKE2	AME	ASE	AVP
1	-3.98338	0.49794	0.06400	-0.82581	-0.46596	-0.99134	1.17011	-0.64046	-1.24612	0.3999	0.1418	0.5417
2	-4.03617	0.50133	-	-0.82166	-0.47561	-0.88203	1.17342	-0.70467	2.76939	0.3821	0.1215	0.5036
3	-4.04385	0.49751	-	-0.82208	-0.48118	-0.89163	1.18184	-0.63574	-	0.3648	0.1032	0.4680
4	3.91738	0.41902	-	-0.88917	-0.36193	-0.80942	-	-0.66949	-	0.3555	0.0868	0.4423
5	2.19733	0.40270	-	-0.75087	-	-0.54633	-	-0.55624	-	0.3481	0.0696	0.4177
*6	2.41535	0.36488	-	-0.75374	-	-0.60510	-	-	-	0.3410	0.0557	0.3968
7	2.34986	0.00285	-	-0.61758	-	-	-	-	-	0.3592	0.0436	0.4028

(Note) AME : Average Model Error Variance

ASE : Average Sampling Error Variance

AVP : Average Variance of Prediction (AME+ASE)

* : Selected model

$$\begin{aligned} \text{SKEW} &= 2.41535 + 0.36488(Z) - 0.75374 \\ &\ln(\text{STMFRQ}) - 0.60510 \ln(\text{MSL}) \\ &\dots\dots\dots (16) \end{aligned}$$

回歸分析에서 獨立變數로 도입된 Z는 두 유역 사이의 歪曲係數의 연속성을 고려하기 위한 假變數이며, 유역 사이에 歪曲係數 값들이 線型性을 가질 경우에는 다른 假變數(예컨데, 緯度 및 經度)를 獨立變數에 포함시킬 수도 있을 것이다. 또한 두 유역에 인접된 다른 유역을 분석에 포함시키므로써 유역의 경계를 고려한 보다 더 합리적인 地域化歪曲係數를 산정할 수 있을 것으로 사료된다.

VI. 結 論

한강 및 낙동강 유역에서의 地域化歪曲係數를 산정하기 위하여 加重回歸分析法을 이용하여 각 유역의 유량관측점에서 수집된 年最大洪水量 자료의 對數値로부터 산정된 地點歪曲係數와 對數變換된 유역특성 및 강우량 자료를 연관시키는 加重回歸模型을 개발하였으며 回歸方程式은 다음과 같다.

$$\begin{aligned} \text{SKEW} &= 2.41535 + 0.36488(Z) - 0.75374 \\ &\ln(\text{STMFRQ}) - 0.60510 \ln(\text{MSL}) \end{aligned}$$

回歸模型을 이용하여 각 유역의 地域化歪曲係數를 산정한 결과 한강유역의 경우는 -0.732, 낙동강유역은 -0.575였다.

本 研究는 韓國科學財團의 研究費支援에 의하여 遂行되었음

參 考 文 獻

1. Benson, M. A. 1959. Channel slope factor in flood frequency analysis. J. Hydraul. Div., ASCE, 85(HY4) : 1-19.

2. Benson, M. A. and N. C. Matalas, 1967. Synthetic hydrology based on regional statistical parameters, Water Res. Res., 3(4) : 391-935.
3. Binnie and Partners, 1978. Hydrological services rural infrastructure project, Final report Vol. 1-6
4. Bobee, B. and R. Robitaille, 1975. Correction of bias in the estimation of the coefficient of skewness, Water Res. Res., 11(6) : 851-854
5. Draper, N. R. and H. Smith, 1981. Applied Regression Analysis, 2nd ed., John Wiley & Sons, Inc., New York
6. Interagency Advisory Committee on Water Data, 1982. Guidelines for determining flood flow frequency, Bull. # 17B of the Hydrology Subcommittee, OWDC, U. S. Geol. Surv., Reston, VA.
7. Natural Environment Research Council, 1975. Flood studies report, Vol. I : Hydrological studies, London.
8. Stedinger, J. R. and G. D. Tasker, 1985. Regional hydrologic analysis, 1, Ordinary, weighted, and generalized least squares compared, Water Res. Res., 21(9) : 1421-1432.
9. Stedinger, J. R. and G. D. Tasker, 1986a. Correction to "Regional hydrologic analysis, 1, Ordinary, weighted, and generalized least squares compared", Water Res. Res., 22(5) : 844.
10. Stedinger, J. R. and G. D. Tasker, 1986b. Regional hydrologic analysis, 2, Model-error estimators, estimation of sigma and Log-Pearson Type III distribution, Water Res. Res., 22(10) : 1487-1499.
11. Tasker, G. D., 1978. Flood frequency analysis with a generalized skew coefficient, Wa-

- ter Res. Res., 14(2) : 373-376.
12. Tasker, G. D., 1980. Hydrologic regression and weighted least squares, Water Res. Res., 16(6) : 1107-1113.
 13. Tasker, G. D. and J. R. Stedinger, 1986. Regional skew with weighted least squares regression, J. Water Resour. Pla. Manage., ASCE, 112(2) : 225-237.
 14. Tung, Y. K. and L. W. Mays, 1981a. Reducing hydrologic parameter uncertainty, J. Water Resour. Pla. Manage., ASCE, 107 (WR1) : 245-262.
 15. Tung, Y. K. and L. W. Mays, 1981b. Generalized skew coefficients for flood frequency analysis, Water Res. Bull., 17(2) : 262-269.
 16. Water Resources Council, 1977. Guidelines for determining flood flow frequency, Bull. # 17A, U. S. Water Resources Council, Washington, D. C..
 17. 건설부, 1974. 한국하천조사서.
 18. 건설부, 1974. 낙동강유역 조사보고서.
 19. 건설부, 1977. 한강하천정비 기본계획.
 20. 건설부, 1979. 한국수문조사서 (우량, 수위 ; 1962-1978).
 21. 건설부, 1981. 하천편람.
 22. 건설부, 1982. 한국하천일람
 23. 건설부, 1983. 홍수량측정 조사보고서.
 24. 건설부, 1985. 홍수량측정 조사보고서.
 25. 건설부, 1986. 홍수량측정 조사보고서.
 26. 건설부, 1987. 홍수량측정 조사보고서.
 27. 건설부 한강홍수통제소, 1985. 충주다목적댐 건설에 따른 한강홍수예경보 프로그램 개선 보고서.
 28. 건설부 한강홍수통제소, 1986. 낙동강 홍수예경보 프로그램 개발 최종보고서.
 29. 건설부 한강홍수통제소, 1986. 낙동강 홍수량측정조사 보고서.
 30. 건설부 한강홍수통제소, 1986. 한강홍수예경보.
 31. 건설부, 한국수문조사년도 (1964-1986).
 32. 내무부 토목국, 1985. 한국하천편람, 제2집.
 33. 농림수산부, 농업진흥공사, 1987. 수리시설물수원공일람표, 5-1권-5-5권.
 34. 박성현, 1981. 회귀분석, 대영사.
 35. 선우 중호, 1987. 수문학, 동명사.
 36. 조국광, 1989. 하천홍수량의 지역화회귀모형에 의한 유량관측망의 계획, 서울대학교 대학원 박사학위논문.