

〈論 文〉

## 연유출량 추정모형의 개선방안 A Study on the Improvement of Annual Runoff Estimation Model

李 相 勳\*  
LEE Sang Hoon

**Abstract** □ The most significant factor in estimating annual runoff must be the precipitation. But in the previous study, the watershed area instead of precipitation was included as an independent variable in regression model in the process of checking accurate data. The criterion of accurate data was the runoff ratio in the range of 20% to 100%. In this study the valid range of evapotranspiration was adopted as a criterion of accurate data and the same data were reexamined. It came up with following model which has a high coefficient of determination and conforms to hydrologic theory.

$$R = -518.25 + 0.8834P \text{ where, } R : \text{runoff depth(mm)} \text{ } P : \text{precipitation(mm)}$$

This regression model was found to be stable by cross-validation and is proposed as annual runoff estimation model applicable to ungauged small and medium watersheds in Korea.

**요 지 :** 연유출량에 영향을 미치는 가장 직접적인 인자는 강수량인데 회귀분석을 이용한 이전의 연구에서는 유출률이 20% 미만 또는 100% 이상인 경우에는 강수와 유량자료는 이상점(outlier)으로서 분석에서 제외시킨 결과 강수량은 독립변수로서 의의가 없고 대신 유역면적을 중요한 독립변수로 포함시켰다. 본 연구에서는 유출률대신 (연강수량-연유출량)을 연증발산량의 좋은 추정치로 간주하고 우리나라에서 가능한 연증발산의 범위를 벗어나는 자료를 제외시키고 회귀분석을 한 결과 수문학적인 이론에 부합되며 결정계수가 높은 다음과 같은 회귀분석식을 얻었다.

$$R = -518.25 + 0.8834P \text{ 단, } R : \text{유출고(mm)} \text{ } P : \text{연강수량(mm)}$$

이 회귀분석식은 cross-validation을 거친 결과 계수가 매우 안정되어 있어서 우리나라의 미계측 중소수계에서 사용할 수 있는 좋은 연유출량 추정모델로서 제안한다.

### 1. 서 론

유역단위의 용수수급계획을 세울 때 기본이 되는 자료는 연유출량으로서 1년 동안에 한 유역에서 이용할 수 있는 수자원의 총량을 추정하는 일은 매우 중요하다. 가용수자원을 파악하는 가장 좋은 방법은 대상유역에서 유출량을 측정하는 것이겠으나 장기간의 유출량측정은 매우 어렵기 때문에 연유출량추정모형을 개발하여 간접적으로 추정하는 방법을 모색하게 된다. 우리나라에서는 미계측지점에 대한 연유출량추정에 대한 공식적인 방법이 제안되어 있지 않고 월유출량 추정모형을 이용하거나 실무에서 많이

이용하고 있는 면적비법이나 평균유출고법을 연단위로 확대 적용하고 있다.

수문학적 관점에서 볼 때에 연유출과 월유출의 유출 특성은 다르다. 월유출에서는 토양수분의 함량변화가 중요한 변수가 되며 증발산량이 월에 따라 크게 변화하게 된다. 그러나 연유출에서는 토양수분량 변화량이 크게 문제가 안되며 유출량은 유역의 특성과 강수량에 크게 의존하게 되므로 추정모형은 훨씬 간단해 질 수 있게 된다. 이러한 점에 착안하여 건설기술연구원에서는 우리나라의 미계측 중소유역에 적용할 수 있는 연유출량추정모형을 다음과 같은 형태로 제안하였다(건설기술연구원, 1989).

$$Q = a + bA \tag{1}$$

\* 정회원, 수원대학교 환경공학과 조교수

단, Q : 연유출량(톤/년)

A : 유역면적(km<sup>2</sup>)

a,b : 계수

또한 국제수문개발계획(IHP)의 일환으로 추진되는 대표유역연구의 조사보고서에서는 한강의 평창강, 금강의 보청천, 낙동강의 위천유역에 대해서 7년간의 유량측정자료를 정리하여 각각 다음과 같은 형태의 연유출량추정모형을 제시하였다(건설부, 1989).

$$R = a + bP \quad (2)$$

$$R = a + bP + cET \quad (3)$$

단, R : 연유출량(mm)

P : 강수량(mm)

ET : 연증발산량(mm)

a,b,c : 계수

이들 두 보고서는 똑같이 통계기법인 회귀분석을 이용하였으며 모형의 정확성을 상관계수의 크기를 기준으로 삼았다. 앞으로도 회귀분석을 이용한 연유출량추정모형 연구는 계속될 것으로 예상되는데 회귀분석을 적용하여 수문모형을 연구할 때에는 회귀분석의 특성과 한계를 인식하는 일이 중요하다.

본 연구에서는 수문모형연구에서 자주 이용되는 회귀분석에 대해서 주의해야 할 사항을 검토한 후, 발표된 연유출량추정모형에서 사용한 자료를 재해석하여 수문학적으로 타당하고 통계적으로 안정된 모형을 제시하며, 앞으로의 연유출량추정연구가 지향해야 할 방향을 지적하고자 한다.

## 2. 연유출량추정모형의 해석

### 2.1 강수량과 증발산량을 포함한 모형

1989년에 발표된 IHP대표유역연구조사보고서에서는 1983년부터 1989년까지 7년간 측정된 수문자료를 정리하여 장기물수지모델로서 3유역에 대해서 다음 표 1과 같이 연유출량추정모형을 발표하였다. 여기서 P는 측정된 유역평균강수량(mm)이고 ET는 유역증발산량으로서 기상자료에 의한 Penman방법

에 의하여 월잠재증발산량을 구하고 여기에 월별증발산계수를 곱하여 얻은 월증발산량을 합산하여 구해진다.

표 1. IHP대표유역의 장기물수지모델

(Long-Term Water Balance Model at IHP Representative Watersheds)

유역	물수지모델	상관계수	F치
평창강	$R = 155.698 + 0.388 \times P$	0.877	16.81
	$R = 489.830 + 0.341 \times P - 0.417 \times ET$	0.936	14.25
보청천	$R = 18.109 + 0.467 \times P$	0.716	3.06
	$R = -166.019 + 0.476 \times P + 0.239 \times ET$	0.819	1.24
위 천	$R = 97.166 + 0.255 \times P$	0.760	13.68
	$R = 500.146 + 0.235 \times P - 0.521 \times ET$	0.885	7.24

(참고) 각변수의 범위 R: 유출량(mm) P: 강수량(mm) ET: 증발산량(mm)

여기서 상관계수는 결정계수의 평방근을 의미한다.

이들 두가지 형태의 모형중에서 상관계수가 높은 두번째 모형을 각 유역의 물수지모델로 채택하고 있다. 이 모델의 해석에서 두가지 문제점이 대두된다. 첫째, 물수지모델을 수문학적인 관점에서 물리적으로 해석하면, 같은 양의 강우가 내렸을 때 증발산량이 크면 유출량은 적어진다는 이론적 결론과 반하여 보청천의 모델에서는 ET항의 계수가 +로서 증발산량이 많을수록 유출량도 증가한다는 해석이 내포되어 있다는 점이다. 둘째, 좋은 모델의 준거로서 “높은 상관계수”를 택했지만 일반적으로 회귀분석모형의 유의성은 결정계수(coefficient of determination)로 비교하는데 결정계수란 “종속변수(Y)의 변동(variation)중에서 회귀분석식에 의해 설명되는 변동의 비율”로 해석한다. 결정계수는 단순회귀의 경우에는 두 변수의 상관계수 r의 제곱과 동일하나 독립변수가 2개 이상이면 상관계수 r은 다중상관계수로서 그 의미가 달라지며 결정계수와 직접 관련시킬 수가 없다. F치는 회귀분석식에 의해 설명되는 변량(variance)을 잔차(residual)변량으로 나눈 비율로서 F값이 클수록 좋지만 같은 크기의 결정계수일 경우 사례수 n이 증가하면 항상 커지므로 모형의 유의성을 표현하는 방법으로는 적당치 않다

(Wonnacott, 1977).

2.2 강수량과 유역면적을 포함한 모형

1989년에 발표된 건설기술연구원의 보고서(이하 건기원 연구)에서는 우리나라 전역을 대상으로 수위 유량곡선이 작성되어 있는 총 95개 지점을 조사하였다. 이 중에서 수위자료가 양호하고 상류지역에서 유출에 심각한 인공조작을 받지 않는 지점을 46개 선정하였다. 연유출량은 일별유출량을 일년동안 합산하고 유역평균강수량은 유역내 강우측정지점의 강수량을 산술평균으로 구하였다. 연강수량과 연유출량을 비교하여 유출률이 100%이상으로 크게 나타나거나 20%이하로 작게 나타나는 연도는 분석에서 제외하였는데 이 과정에서 9개 지점이 제외되었으며 최종분석에 이용된 것은 37개 지점의 355개 자료이다. 이들 355개 자료로써 각 지점별로 유역평균강수량과 유역평균유출량을 계산한 후, 각 유역의 유역 특성인자로서 하천연장, 하천경사, 형상계수, 유역면적을 조사하였다. 각 인자들간의 상관분석결과 모형의 입력변수로 고려할 수 있는 것은 유역면적뿐이었으며 연평균강수량은 연평균유출량에 무상관으로 나타났다(상관계수 0.10). 연평균강수량은 이론적으로는 연평균유출량에 상관이 있는 변수이므로 유역면적과 함께 입력변수로 택하여 분석을 실시한 결과 표 2와 같은 결과를 얻었다.

표 2. 연유출량추정을 위한 회귀분석결과와 통계치  
(Statistics and Result of Regression Analysis for Annual Runoff Estimation)

구분	회귀모형	회귀분석결과	상관계수	표준개산오차
1)	$Q=a+bA$	$Q=-39681300+734859A$	0.920	201000
2)	$Q=aA^b$	$Q=570200A^{1.06}$	0.938	206000
3)	$Q=a+bP+cA$	$Q=-18880000+117400P+737136A$	0.915	200000
4)	$Q=aP^bA^c$	$Q=65500P^{.289}A^{1.027}$	0.935	206100

(참고) 각변수의 단위 Q:유출량(톤/년) A:유역면적(km<sup>2</sup>)  
P:강수량(mm)

이들 4개의 식을 비교해 볼 때 선형모형인 1), 3)식이 비선형모형인 2), 4)식보다 표준개산오차가 작으

므로 최종 모형으로 선정되었다. 그러나 그후 수문학회지에 발표한 논문에서는 상관계수가 가장 높은 2)식을 최종모형으로 발표하였다(김양수의 3인, 1991). 1)식과 3)식을 비교해 보면 1)식이 상관계수가 높으며 간단하므로 가장 좋은 모형이라고 해석할 수 있다. 이 연구결과와 해석에 있어서 다음과 같은 문제가 제기될 수 있다.

첫째, 유출량에 영향을 미치는 가장 중요한 인자는 강수임에도 불구하고 1)식에서는 강수량대신 유역면적이 포함되어 있으며, 강수량이 포함된 3)식의 결정계수가 유역면적만을 포함하는 1)식의 결정계수보다 오히려 작아진다는 점이다. 이에 대해 동 연구에서는 “이것에 대한 원인은 명확히 판단하기 어려우나 이용자료에 문제가 있는 것으로 추측할 수 있다”고 언급하고 있는데 본 연구에서는 너무 큰 수치에 따르는 계산상의 문제외에 이용자료에 어떠한 문제가 있는지를 밝히고자 한다.

둘째, 1)식의 형태로 볼때에 유역면적이 54.0km<sup>2</sup> 이하가 되면 Q값이 음수가 된다. 모형개발에 사용된 유역면적은 최소 36km<sup>2</sup> 최대 2949km<sup>2</sup>로서 200km<sup>2</sup> 이하의 지점은 3개에 불과했다. 그러므로 1)식을 소유역에 적용할 때에는 주의를 요한다.

셋째, 종속변수 Q의 scale이 독립변수와는 너무 크게 차이가 난다는 점이다. 유출고를 R(단위: mm)이라고 하면  $Q=1000AR$ 이 된다. Q와 R은 쉽게 계산하여 환원할 수 있는데 Q는 R을 1000A배 만큼 부풀린 값이다. 물리적인 의미로 보면 R을 종속변수로 보나 Q를 종속변수로 보나 똑같다. IHP연구에서는 R을 종속변수로 채택하였으며 모형에 사용되는 세 변수 R과 P, A의 범위는 100에서 3000정도로서 비슷하게 된다. 그러나 Q의 범위는 1000만에서 10억으로서 독립변수와는 너무 scale이 달라지게 된다. 종속변수를 R로 채택할 때와 Q로 채택할 때 각 지점의 A가 달라지므로 회귀분석식의 기울기는 달라지며 결정계수가 커질지 작아질지는 예측할 수 없다. 똑같은 37개의 자료를 이용하여 종속변수를 R로 채택하면 결정계수는 0.8464에서 0.0115로 적어져 회귀분석은 의의가 없게 된다. 건기원 연구

에서 Q를 종속변수로 채택한 이유는 높은 결정계수를 얻기 위한 것으로 추측된다.

이러한 문제점에도 불구하고 건기원모형은 강우자료가 없는 지역의 연유출량추정에 이용되는 면적비법의 근거를 제시했다는 점에서 의의가 있다고 볼 수 있다.

### 3. 회귀분석적용시 유의점

#### 3.1 회귀분석결과의 제시형태

회귀분석은 물리적 현상에 적용할 때 어느 의미에서 보면 암흑상자(black box)로서 상자안에서 일어나는 일은 알 수 없으며 최종적으로 주어지는 숫자의 해석에는 매우 조심해야 한다. 회귀분석은 다음과 같이 설명된다.

어느 종속변수 Y의 측정치  $Y_i$ 는 독립변수 X에 의해 결정되어지는 부분과 미지의 오차  $e_i$ 의 합으로 나타낼 수 있다.

$$Y_i = f(X) + e_i \quad (4)$$

만일  $f(X)$ 가 X의 일차함수라고 가정한다면

$$Y_i = a + bX_i + e_i \quad \text{여기서 } a, b \text{는 미지의 계수} \quad (5)$$

와 같이 나타낼 수 있다. 여러개의 측정치  $(X_i, Y_i)$ 를 구했을 때 오차 e의 제곱합을 최소로 하는 계수 a와 b를 구하여 일차식으로 나타내면

$$\hat{Y} = a + bX_i \quad (6)$$

가 되며  $\hat{Y}$ 는 변수가 하나일 때에는 직선식이 되며 이 식이 회귀분석식이다. 회귀분석식의 계수 a와 b는 다음과 같이 원자료에서 계산된다.

$$b = (\sum x_i Y_i / \sum x_i^2) \quad (7)$$

$$a = \bar{Y} - b\bar{X} \quad (8)$$

단,  $x_i = X_i - \bar{X}$ ,  $\bar{X} = X$ 의 평균,  $\bar{Y} = Y$ 의 평균 회귀분석식의 사용이 의미가 있는가를 판단하기 위해서 분산분석을 하게 된다.

관측값  $Y_i$ 와 Y의 평균  $\bar{Y}$ 의 차이를 총편차(total deviation)라고 하며 다음과 같이 두개의 편차의 합으로 나타낼 수 있다.

$$Y_i - \bar{Y} = (Y_i - \hat{Y}) + (\hat{Y} - \bar{Y}) \quad (9)$$

$$\text{또는 } D_i = e_i + d_i \quad (10)$$

위 관계를 그림으로 나타내면 그림 1과 같다. 식(9)의 양변을 제곱하고  $i = 1, 2, \dots, n$ 에 대하여 합하고 단순회귀의 정규방정식의 결과를 이용하면 다음과 같이 된다(김우철외 7인, 현대통계학, 1989).

$$\sum (Y_i - \bar{Y})^2 = \sum (Y_i - \hat{Y})^2 + \sum (\hat{Y} - \bar{Y})^2 \quad (11)$$

$$SST = SSE + SSR$$

단, SST : total sum of squares, 총변동

SSE : error sum of squares, 설명안되는 변동

SSR : regression sum of squares, 회귀 분석에 의해 설명되는 변동

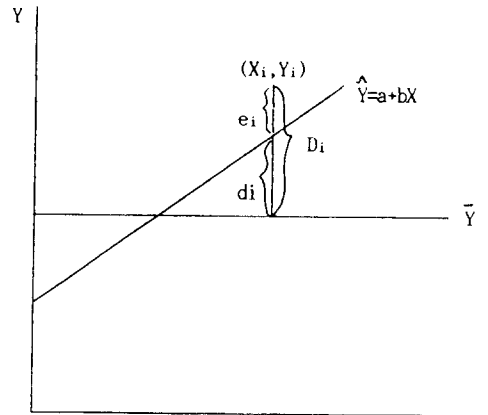


그림 1. 총편차 Di의 분해

(Analysis of Total Deviation Di)

단순회귀모형에서 회귀직선에 의해 종속변수가 설명되어지는 정도를 나타내는 결정계수  $R^2$ 는 다음과 같이 계산된다.

$$R^2 = SSR / SST = \sum d_i^2 / \sum D_i^2 \quad (12)$$

결정계수  $R^2$ 의 범위는  $0 \leq R^2 \leq 1$ 로서 측정값들이 모두 회귀직선상에 위치한다면  $SSE = 0$ 이 되고  $R^2 = 1$ 이 된다. 반대로 Y와 X사이에 회귀관계가 전혀 없어서 추정된 회귀직선의 기울기가 0이라면  $\hat{Y} = \bar{Y}$ 가 될 것이며  $R^2 = 0$ 이 된다.

회귀직선의 유의성은 회귀직선식의 기울기  $b = 0$

인가를 통계적으로 검증하는데 검정통계치는 F값으로 다음과 같이 계산한다.

$$MSR = SSR/p \quad (13)$$

$$MSE = SSE/(n-1-p) \quad (14)$$

$$F = MSR/MSE \quad (15)$$

단, MSR=회귀분석식의 분산

MSE=잔차(residual)의 분산

n=사례수

p=독립변수의 수

MSE는 분모가(n-1-p)이므로 사례수가 많아지면 당연히 작아지며 따라서 F값은 커지게 된다. F값추정치인 모집단(population)은 F분포를 가지며 영가설 Ho:b=0을 받아들일 확률은 F값이 클수록 작아진다. 즉 F값이 클수록 b=0이라고 주장할 수 없고 따라서 회귀직선은 유의성이 있게 된다. 여기서 회귀직선이 유의성이 있다는 주장은 b=0이 아니라는 의미일 뿐 계산된 b의 크기가 의미있는지의 여부는 말할 수 없다. 예를 들면 계산되어 나온 b의 추정치  $\hat{b}=0.001$ 이나  $\hat{b}=20.0$ 이나 F값이 크다면 회귀직선의 기울기는 유의하다는 해석이 가능하다. 그러나 X변수의 단위가 10단위인데  $\hat{b}=0.001$ 이라면 X항의 기여도는 산술적으로 미미할 것이며 의미가 없을 것이다. 그러므로  $\hat{b}$ 의 크기가 의미가 있는지를 검정할 필요가 있으며 이것은  $\hat{b}$ 의 신뢰구간을 구하거나 또는 t값을 계산하여 확인할 수 있다.  $\hat{b}$ 의 t값은 다음과 같이 구해진다.

$$t = \hat{b}/s_b \quad (16)$$

$$\text{단, } s_b = s/(\sum x_i^2) \quad s = (\sum(Y_i - \hat{Y})^2)^{1/2}$$

$s_b$ 를 표준오차라고 하며(표 2에서는 표준개산오차라고 표현했음) b의 표준편차라고 해석할 수 있다. 이와 같이 계산된 t값추정치인 모집단은 t분포를 따르며 t분포의 확률값은 통계책에 표의 형태로 제시되며 통계 package에서는 Prob>|T|값으로서 회귀분석 결과의 일부로서 제시된다. Prob값의 해석은 t값이 크면 Prob값이 작아지며 Prob값이 0.05 또는 0.01이하이면  $\hat{b}$ 의 값은 유의있는 것으로 해석하면 된다. 독립변수가 2개 이상이 되면 다중

회귀분석이 되고 계수는 각각의 독립변수에 대해서 계산된다. 각각의 계수의 Prob의 해석은 동일하다.

이상을 종합하면 회귀분석의 의의는 상관계수와 F값보다는 결정계수 R<sup>2</sup>값과 기울기 b의 Prob값으로 판단하는 것이 바람직하며 회귀분석의 결과를 제시할 때에도 R<sup>2</sup>값과 b의 Prob값을 제시하여야 할 것이다.

### 3.2 이상점(outlier)의 영향

회귀분석을 이용할 때에 주의하여야 할 점은 종속변수 Y가 나머지 측정치 경향과 비교했을 때 매우 큰 값 또는 매우 작은 값으로 관측되었을 때 이것을 분석에 포함시켜야 할 것인지를 신중히 검토해야 한다는 것이다. 일반적으로 이상점은 X와 Y변수에 대한 散布圖를 작성함으로써 시각적으로 어느 정도 찾아낼 수 있으나 정확한 방법은 회귀식의 결과로 산출된 표준잔차(standard residual)가 +2보다 크거나 -2보다 작을 경우 이상점으로 간주하여 측정과정이나 기록상의 오류가 있는지를 검토해야 한다(이희연, 1989). 하나의 이상점을 분석에 포함시켰을 때 이상점이 미치는 영향은 이상점이 회귀분석선에서 벗어나는 정도보다는 이상점이 나타나는 위치에 따라 다르다.

그림 2는 평창강에서 얻어진 7개의 자료를 가지고 회귀분석식을 구한 결과를 그래프로 보여준다. 여기에서 A와 B는 모두 회귀분석식에서 200mm의 편차를 가진 가상의 이상점들이다. 7개의 원자료와 각각의 이상점을 포함시켰을 때의 회귀분석식의 변화는 다음과 같다.

최초 7개의 자료를 이용한 회귀식

$$R1 = 155.698 + 0.388P \quad (17)$$

7개+이상점 A를 이용한 회귀식

$$R2 = 414.340 + 0.208P \quad (18)$$

7개+이상점 B를 이용한 회귀식

$$R3 = -36.539 + 0.546P \quad (19)$$

똑같이 200mm의 편차를 가진 이상점이지만 이상점 A는 절편을 258mm나 증가시키며 기울기를 작게 하

는데, 이상점 B는 절편을 191mm 감소시키며 기울기를 크게 한다. 그러므로 회귀분석을 이용하여 유출량 추정모형을 만들 때에는 이상점을 잘 검토하여 측정오차가 크다고 판단되면 분석에 포함시키지 말아야 할 것이다. 만일 하나의 이상점을 무분별하게 포함시키면 그전까지 축적된 좋은 자료의 효과가 무산되고 결정계수를 아주 작아지게 하는 결과가 나타날 수 있다.

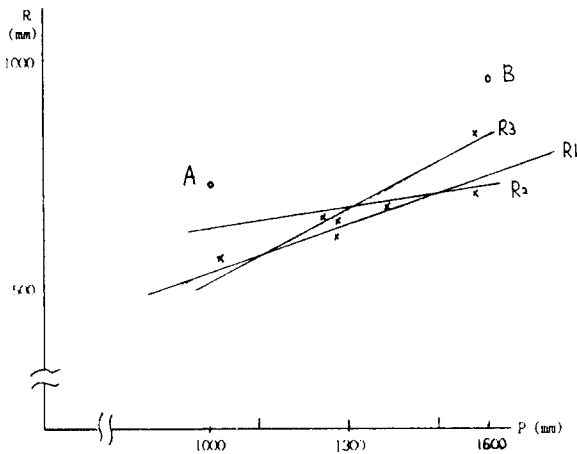


그림 2. 연유출량추정 회귀분석식과 가상의 이상점들 (Regression Equation Estimating Annual Runoff and Hypothetical Outliers)

3.3 물리적 의미

회귀분석선의 기울기 b의 물리적의미는 독립변수 X가 1단위 증가할 때에 예측되는 종속변수 Y의 증가량으로 해석된다. 이의 증명은 다음과 같다.

$$Y = a + bX$$

$$Y_{변화된} = a + b(X + 1) = a + bX + b$$

$$Y \text{의 증가량} = b$$

그러므로 표 2의 1)식의 기울기 734859의 해석은 유역면적이 1km<sup>2</sup>증가할 때 증가되는 연유출량(톤)을 의미한다. 마찬가지로 표 1에서 평창강의 연유출량 추정모델의 처음식은 연강수량이 1mm증가할 때에 유

역의 유출고(mm)는 0.388mm 증가함을 의미한다. 보청천의 경우 회귀분석식에 유역증발산 ET가 포함되면 다중회귀분석이 되며 ET의 계수 0.239의 해석도 동일하다. ET의 계수 0.239는 강수량이 똑같은 경우 ET가 1mm 증가하면 유출고는 0.239mm 증가한다는 뜻인데 이것은 명백히 물리적 현상으로는 받아들이기 어렵다. 이러한 통계적 현상이 나타나는 이유는 P와 ET사이에 상관계수가 높은 이른바 공선성(multicollinearity) 때문인 것으로 해석할 수 있다. 표 1의 2)식의 해석에도 약간의 문제가 있다. Q = aA<sup>b</sup>에서 b=1.021로서 면적 A가 1% 증가하면 Q의 증가가 1%보다 많이 증가한다는 뜻이다. 이 결과는 유역면적이 커질수록 유출량의 총량은 증가하겠으나 그 증가율은 A의 증가율보다는 작아진다는, 즉 b < 1이라는 일반적인 결과와 잘 맞지 않는다(Chow, 1965). 그러므로 회귀분석결과가 통계적으로는 의의 있을지라도 물리적으로 해석할 때에 어려움이 있다면 보편적인 모델로서 받아들이는 데에는 신중해야 할 것이다.

4. 증발산량을 이용한 수문자료의 정확성검증

4.1 물수지방정식

유역을 단위로 할 때 유출에 영향을 미치는 인자는 크게 기후적인자와 地相의인자로 구별되는데 기후적인자에는 강우, 차단, 증발산, 지표면저류, 침투 등이 포함되며 지상적인자에는 유역특성, 유역의 크기, 유역의 형상, 경사, 하천밀도, 토양의 성질등이 포함된다(권순국의 5인, 1984). 그러나 유출량을 양적으로 따져보면 다음과 같이 물수지방정식으로 나타낼 수 있다.

$$P = R + ET + \Delta S \tag{20}$$

여기수 ΔS는 지하수를 포함한 유역의 저류량변화를 나타낸다. ΔS는 1회의 강우사상이나 주, 월, 또는 계절단위에서는 무시할 수 없다. 그러나 1년 동안의 경우, 저수지나 댐이 있고 유역이 크다면 저류량변화를 무시하기 어렵지만 큰 저수지가 없는 중소유역

의 경우에는 1년동안의 저류량의 변화는 무시할 수 있게 되며 물수지방정식은 간단하게 된다. 즉 1년 동안 그 유역에 내린 강수의 총량은 유출의 총량과 증발산의 총량으로 나누어진다. 강수량은 비교적 정확히 측정되는데 유출량은 수자원이용에서 매우 중요하나 측정이 매우 어렵기 때문에 어떻게 하면 유출량을 예측할 수 있는 모형을 얻을 수 있는가 하는 것이 쫓점이 된다.

연단위의 물수지방정식에서 보면  $R=P-ET$ 가 되므로  $ET$ 만 정확히 측정될 수 있다면 유출은  $(P-ET)$ 로서 쉽게 계산될 수 있다. 그러나 증발산의 측정은 유출의 측정보다 더 어려우며 증발산의 추정 자체가 매우 어려운 과제이다. 증발산의 추정은 기상 관측소에서 이루어지는 증발의 측정에서부터 시작하여 2단계로 이루어진다.

1단계 계기증발량  $\times$  접시계수 = 잠재증발산

2단계 잠재증발산  $\times$  유역증발산계수 = 추정증발산  
증발산의 추정은 접시계수와 증발산계수의 두가지 계수가 필요하므로 보다 더 복잡함을 알 수 있다. 강수는 불규칙한 기상현상으로서 변화가 심하여 홍수년과 갈수년의 강수량의 차이가 크며 강수의 변동은 유출의 변동에 직접적으로 영향을 준다. 그러나 증발산은 비교적 안정된 요소로서 그 변화가 심하지 않다. 왜냐하면 비가 많이 오면 증발산이 많아지는

가장 많으며 평균 1231mm이다(한국기후표 제1권, 1982). 그러나 계기증발량을 유역증발산으로 환산하기 위해서는 접시계수와 증발산계수를 곱하여야 하나 이에 관해서는 뚜렷한 연구결과가 없으며 일반적으로 접시계수를 0.7로 보므로 계기증발량의 70% 이하가 유역증발산이라고 추정해 볼 수 있다.

한 유역에서 강수량과 유출량만 정확히 측정되었다면 연증발산량의 가장 정확한 추정치는  $(P-R)$  값이 될 것이다. 북미대륙의 태평양 연안 유역에서 측정된 자료를 보면  $ET$ 의 평균은 600~900mm이고 표준편차는 50~100mm로 보고되었다(Leopold, 1978). 우리나라와 위도가 비슷한 미국의 James (37.2N)와 North(41.5N), 오스트렐리아의 Shoalhaven(35.3S), 뉴질랜드의 Waikato(37.6S) 등의 4개 지점에서 측정한  $(P-R)$ 값의 범위는 최소 653mm 최대 788mm로 보고되었다(Morton, 1983). 조희구와 이태영(1975)은 낙동강유역의 증발산을 516mm로 추정하였다.

우리나라에서 강수량과 유출량을 정확히 측정한 지역은 IHP대표시험유역을 들 수 있다. 대표시험유역에서는 증발산의 추정연구도 실시되었는데 유역의 실제증발산량의 추정은 Penman법에 의해 기상자료를 이용하여 잠재증발산량을 구하고 여기에 월별유역증발산계수를 곱하여 구하였다. 이값은 결국 증발

표 3. IHP대표유역의 강수량, 유출량, 증발산량추정치 자료

(Precipitation, Runoff, and Evapotranspiration Estimation Data at IHP Representative Watersheds)

지역	항목	1983	1984	1985	1986	1987	1988	1989	평균
평창강	P	1277.3	1570.9	1288.0	1245.1	1578.7	1021.8	1387.1	1338.4
	R	606.1	836.9	646.6	660.0	722.5	571.9	679.8	674.8
	ET'	698.5	584.1	770.0	591.4	638.9	639.5	661.7	655.3
	P-R	671.2	734.0	641.4	585.1	856.2	449.9	707.3	663.6
보청천	P	1357.5	910.4	1535.8	1226.6	1684.3	920.1	1293.8	1275.5
	R	559.1	355.3	791.2	382.2	943.9	741.7	520.0	613.1
	ET'	808.6	696.1	674.1	673.6	681.0	751.7	733.6	717.0
	P-R	798.4	555.1	744.6	844.4	740.4	178.4	773.8	662.4
위천	P	1042.6	1125.9	1295.4	972.8	1263.4	740.9	1030.9	1067.5
	R	268.9	353.9	437.2	353.5	490.6	290.2	456.9	378.8
	ET'	737.4	727.2	781.7	704.0	700.9	747.0	808.6	743.9
	P-R	773.7	772.0	858.2	619.3	772.8	450.7	574.0	688.7

립할 수 있다. 강수량과 유출량자료를 검증할 때에 (P-R)값이 예상할 수 있는 증발산의 범위를 벗어나면 그 해의 자료는 부정확한 것으로 분석에서 제외할 수 있다. 이러한 부정확한 개별 자료는 회귀분석식 이상점으로서 작용하며 한 두개만 있어도 회귀분석선의 기울기를 크게 변동시키고 R<sup>2</sup>값을 작게 하기 때문에 매우 주의해야 한다. 건기원연구에서는 이러한 점에 착안하여 개별연도의 강수량과 유출량 자료를 검증할 때에 유출률의 범위를 20~100%로 정하고 이 범위를 벗어나면 분석에서 제외시켰다. 사실 유출률 100%는 가능하지 않으며 범위를 너무 넓게 잡은 흠이 있다. 1988년 보청천의 유출률은 81%로서 유출률로서는 문제점이 발견되지 않으나 (P-R)값으로 보면 의심스러운 이상점이 된다.

(P-R)값을 기준으로 수문자료를 검증할 수 있는 다른 예를 제시한다. 중규모 다목적댐예비타당성조사최종보고서(건설부, 1986)에서 제시하는 한강유역의 평균유출계수는 0.71로서 금강의 0.52와 낙동강의 0.49에 비해 높다. 한강의 유출계수는 1968년부터 1983년까지 16년자료를 이용하여 계산했는데 이중에는 의심스러운 3개 자료가 포함되어 있다.

$$1975년 (P-R) = (1132.7 - 981.6) = 151.2$$

$$1977년 (P-R) = (865.5 - 725.2) = 140.3$$

$$1980년 (P-R) = (1159.6 - 1023.6) = 136.0$$

1년에 150mm의 유역증발산은 대유역의 저류효과를 인정하여도 다른 연구결과와 비교해 볼 때 너무 작다고 생각된다(Binnie & Parters and Hyundai Engineering Co., 1978). 이들 3개년의 부정확한 자료가 포함되어 있기 때문에 한강의 유출계수가 높은 것으로 계산되었다고 말할 수 있다. 1959~1988까지의 최근의 연구결과에 의하면 한강의 유출계수는 0.57, 금강의 유출계수는 0.49, 낙동강의 유출계수는 0.49로서 한강의 유출계수는 전국평균인 0.55보다 약간 높을 뿐이다(한국수자원공사, 1990). 이처럼 (P-R)값의 크기를 이용하여 자료의 신빙성을 검증할 수 있는데 증발산의 가능한 범위를 정하기 위해서는 유출량을 정확히 측정된 자료의 획득이 중요하게 된다.

본 연구에서는 가장 정확히 유출량을 측정한 자료로서 IHP대표유역의 자료를 이용하되, 1988년 보청천의 자료는 제외한 40개의 자료를 사용하였다. 표 4에서는 ET추정치의 통계적 특성을 보여준다.

만일 모수(parameter)로서의 ET가 정상분포를 이룬다면  $\pm 2$ 표준편차 사이에 98%의 사례가 다 포함될 수 있을 것이므로 두개의 추정치들을 공통으로 합쳐 계산한 평균값 700mm의 약 2 표준편차의 범위



인 500~900mm를 ET의 가능한 범위로 잡았다. 즉 어느 연도의 (P-R)값이 500mm미만이거나 900mm 이상이면 그 자료는 의심스러운 이상점으로 간주하여 회귀분석에서 제외한다.

표 4. IHP대표유역에서의 ET추정치의 통계적 특성  
(Statistical Characteristics of ET Estimates at IHP Representative Watersheds)

추정치	사례수	최소값	최대값	평균값	표준편차
ET*	20	584.1	808.6	702.9	63.4
(P-R)	20	449.9	858.2	696.1	123.9
공통	40	449.9	858.2	699.5	97.2

5. 연유출량추정모형의 비교

5.1 새로운 연유출량추정 모형

자료의 검증을 유출률과 (P-R)을 기준으로 하여 건기원연구에서 이용한 똑같은 자료를 분석하여 얻어지는 새로운 모형을 이하 편의상 E모형이라고 칭하고 그 효용성을 비교하여 보았다. 기본자료는 모두 37개 유역의 총 355년(1961~1988사이)의 자료이나 (P-R)값이 500~900mm의 범위를 벗어나면 제외시킨 결과 30개 유역 160년의 자료만이 분석에 이용되었다. 제외된 195개 자료는 ET가 900mm 이상이어서 제외된 것이 33개, ET가 500mm 미만이어서 제외된 것이 162개 이었는데 특히 ET가 150mm 이하인 자료가 26개나 되었다. 분석에 포함된 자료의 조건은 1) 유출률(R/P)이 0.2이상 1.0 미만이고 2) (P-R)이 500mm이상 900mm미만인데

표 5. E모형 개발에 사용된 자료의 통계적 특성  
(Statistical Characteristics of the Data Used for the Development of E-Model)

특성	P	R	(P-R)	R/P
평균	1224.3	563.3	661.0	0.444
표준편차	239.9	236.7	109.0	0.120
변동계수(%)	19.6	42.0	16.5	27.1
최대	1969.0	1213.2	899.2	0.680
최소	682.5	132.1	500.9	0.163

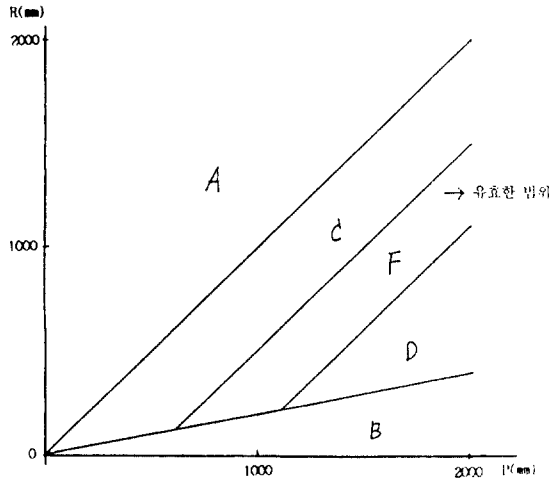


그림 3. E모형개발에 사용된 자료의 유효한 범위  
(Region of Valid Data Used for Model Development)

160개 자료가 남았다. 그림 3에서 A지역은 유출률이 100%이상이 되어 제외되고 B지역은 유출률이 20%미만이 되어 제외된다. C지역은 연증발산이 500mm미만되어 제외되고 D지역은 연증발산이 900mm이상이 되어 제외된다. 유효한 자료의 범위인 F지역에 포함된 160개 사례의 통계적 특성은 표 5와 같다.

표 5를 보면 평균 유출률은 0.444로서 강수량의 44%만이 유출로 나타나는데 이 값은 지금까지 보고된 여러 유역의 유출률보다는 적은 값이나 IHP대표 유역의 평균 유출률인 0.453에 근사한 값으로서 지금까지의 유출량측정이 과다한 값으로 나타났을 가능성을 시사하고 있다. 측정치의 변동은(표준편차/평균)으로 계산되는 변동계수로 나타내는데 강우에 비해 유출의 변동이 심함을 알 수 있으며 (P-R)의 변동이 가장 적다. 이들 160개 자료를 이용하여 세가지 형태의 회귀분석을 실시한 결과는 표 6과 같다.

세가지 모형을 비교해 보면 가장 간단한 형태인

표 6. 연유출량추정을 위한 E모형 개발

(Development of E-Model for Annual Runoff Estimation)

구분	회귀모형	회귀분석식	R <sup>2</sup>	t	Pr >  T
E1	R=a+bP	R=-518.25+0.8834P	0.8018	25.29(H <sub>0</sub> :P=0)	0.0001
E2	LR=a+bLP	R=0.0002558P <sup>2</sup> 045322	0.7589	22.30(H <sub>0</sub> :LP=0)	0.0001
E3	R=a+bP+cA	R=-519.025+0.8830P+0.00126A	0.8018	25.03(H <sub>0</sub> :P=0) 0.097(H <sub>0</sub> :A=0)	0.0001 0.9225

(참고) LR과 LP는 R과 P의 대수값을 나타낸다.

표 7. E모형의 안정성 비교

(Comparison of Stability of E-Model)

회귀분석식	사례수	선정기준	R <sup>2</sup>	t	Pr >  T
R=-518.25+0.8834P	160	전체	0.8018	25.29	0.0001
R=-517.94+0.8862P	80	홀수번째	0.8439	20.53	0.0001
R=-515.98+0.8785P	80	짝수번째	0.7508	15.33	0.0001
R=-550.84+0.9135P	53	3배수번째	0.7553	12.55	0.0001
R=-507.70+0.8465P	40	4배수번째	0.7436	10.50	0.0001
R=-458.09+0.8349P	30	유역평균치	0.7841	10.09	0.0001

E1모형이 결정계수도 가장 높고 P도 유의있는 변수로 나타났다. E2의 모형은 logR과 logP의 회귀분석인데 변수의 물리적해석에 어려움이 있으므로 비록 설명계수가 높더라도 바람직한 모형은 아닌 것으로 생각된다. E3모형은 유역면적을 추가했는데 설명계수는 높아도 변수 A의 t값이 작고 A=0일 확률이 92%나 되므로 변수로서 의의가 없다. 따라서 연유출량추정모형으로서의 간단하며 수문학적 이론에도 부합하는 E1모형이 가장 적합하며 증발산검증을 거친 모형임을 나타내기 위하여 E모형이라고 칭하였다. E모형이 얼마나 안정적인지를 알아보기 위하여 사례수를 줄여가며 결정계수와 기울기를 알아보는 cross-validation을 시도하였다. 표 7의 결과를 보면 사례수를 무작위로 줄여가더라도 결정계수와 기울기는 크게 변하지 않으며 상수항의 크기도 변화가 적다.

5.2 모형의 비교

유출량추정모형의 개발 목적이 미계측하천의 연유출량을 예측하기 위한 것이므로 건기원모형에서 모

형검증을 위하여 사용한 10개 지점의 자료를 이용하여 E모형의 효용성을 비교하였다. 건기원모형은 중속변수가 Q이지만 Q를 유역면적으로 나누어 유출고를 비교하였다. 모형의 비교에 많이 쓰이는 기준은 RMSE이며 다음과 같이 정의된다.

$$RMSE = (\sum(Y_i - \hat{Y})^2/n)^{1/2} \tag{21}$$

10개 자료의 추정치와 측정치의 RMSE를 비교해 본 결과 건기원모형이 194mm, E모형이 180mm로서 E모형이 약간 나은 것으로 나타났다. 그러나 보다 중요한 차이점은 E모형은 유출은 강우에 의하여 결정된다는 수문학적이론에 잘 부합한다는 점이다. 건기원모형의 개발과정에서는 강우만을 독립변수로 하면 결정계수가 너무 작아 변수로서 의의가 없었는데, 거꾸로 E모형에서는 강우가 독립변수로서 가장 의의가 있었으며 오히려 유역면적은 독립변수로서의 의의가 없는 것으로 나타났다.

E모형과 IHP모형의 비교는 IHP모형의 사례수가 9개에 불과하므로 무리이다. 사례수가 적으므로 평창강, 보청천, 위천에서 각각 개발된 모형의 절편과 기울기가 안정되어 있지 않음을 알 수 있다. 그러나 E모형은 사례수를 19%까지 크게 줄여도 기울기와

절편이 안정되어 있어서 회귀분석모형으로서 장점을 보여주고 있다.

### 6. 결 론

연유출량추정모형의 개발에는 정확한 강수량과 유출량의 측정자료를 요구하는데 지금까지의 연구에서는 유량측정자료를 개별적으로 확인하여 의심스러운 유량자료는 1차적으로 제외시키고 2차적으로 유출률을 기준으로 하여 유출률이 과다하거나 과소한 사례는 제외시키는 방법을 채택했다. 건설기술연구원에서는 전국의 미계측하천에 적용할 수 있는 연유출량추정모형을 개발하기 위하여 전국에서 가장 유량 측정자료가 양호한 37개 지점 355년의 자료를 수집한 후 유출률 20%미만 또는 100%이상의 자료를 이상점으로 간주하였다. 37개의 유역평균치를 이용하여 회귀분석한 결과 강수량을 독립변수로 할 때에는 모형의 결정계수가 너무 적어서 강수량 대신 유역면적을 독립변수로 포함시킨 연유출량 추정모형을 발표했다.

연유출량추정을 위한 회귀분석의 응용은 몇가지 주의사항을 인식하고서 실행되어야 한다. 회귀분석은 적용하려는 분야의 이론적인 범위안에서 그 결과를 해석해야 하며 단순히 통계적 결정계수의 크기로 회귀분석의 양호도를 말할 수는 없다. 연유출량을 종속변수로 할 때에 강수량이 가장 중요한 독립변수로 간주된다. 그밖의 다른 변수, 유역면적이나 증발산등은 2차적으로 독립변수로 넣을 수 있으나 그 결과가 수문학적 이론에 부합되지 않을 때에는 결정계수를 높더라도 모형에 포함시키지 않는 것이 좋다. 회귀분석의 결정계수는 사례수가 많다고 반드시 높아지는 것이 아니고 측정오차가 큰 것으로 간주되는 이상점에 의해 큰 영향을 받으므로 이상점의 적절한 제거가 회귀분석의 성과를 좌우하는 가장 중요한 문제로 대두된다.

중소유역에서 1년을 단위로 한 물수지분석에서는 양적으로 볼 때에  $P=R+ET$ 의 관계식이 성립한다고 가정할 수 있다. 증발산은 강수나 유출보다 변동

이 크지 않으므로 우리나라에서 관측될 수 있는 증발산의 범위를 정하고 (P-R)로 계산한 ET가 가능한 범위를 벗어나면 그 해의 자료는 이상점으로서 분석에서 제외시키는 방법을 제안하였다.

ET의 가능한 범위는 우리나라에서 가장 정확한 수문자료를 측정한 IHP대표시험유역의 자료를 이용하여 구하였다. IHP대표시험유역인 한강의 평창강, 금강의 보청천, 낙동강의 위천에서 9년간 측정된 (P-R)을 정확한 자료로 간주하고, 또한 세 유역에서 기상자료에서부터 구한 ET추정치를 정확한 자료로 간주하여 이들 두 종류의 자료 40개의 평균과 표준편차를 구하였다. 그 결과 ET의 평균은 700mm, 표준편차는 100mm로 밝혀졌는데 평균을 기준으로  $\pm 2$ 표준편차의 범위인 500mm~900mm를 ET의 가능한 범위로 잡았다.

전국적으로 가장 정확한 유출측정자료라고 간주되는 355개의 자료중에서 (P-R)값이 500mm 미만이거나 900mm 이상인 경우는 이상점으로서 제외시키고 나머지 160개의 자료를 가지고 유출고(R)를 종속변수로, 강수량(P)을 독립변수로 하여 회귀분석한 결과 다음과 같은 회귀분석식을 얻었다.

$$R = -518.25 + 0.8834P (n=160, R^2=0.8018, t=25.29 (H_0: P=0), Pr > |T| = 0.0001)$$

이 식은 결정계수가 0.80으로서 높고 유출량이 강수량에 의해서 결정된다는 수문학적 이론에도 부합된다. 이 모형의 장점은 사례수가 줄어도 계수가 크게 변하지 않는 안정적인 회귀분석식이라는 점이다. 어느 하천의 유출량을 장기간 측정된 자료가 없고 강수량 자료만 이용하여 연유출량을 추정하고자 할 때에는 당분간 적용할 수 있는 연유출량추정모형으로서 제안한다.

앞으로 이 모형을 개량하기 위해서는 유역의 실제 증발산에 대한 연구가 이루어져서 본 연구에서 채택한 ET의 가능한 범위인 500~900mm가 타당한지를 검토해야 할 것이다. 앞으로 정확한 유출모형을 개발하기 위해서는 유출량의 정확한 측정은 물론 유역 증발산의 연구에도 많은 관심을 가져야 할 것이다.

### 감사의 글

본 연구는 한국과학재단의 연구비 지원에 의하여 수행되었으며 지원해 주신 한국과학재단에 심심한 사의를 표합니다. 또한 자료이용에 협조하여 주신 한국건설기술연구원에도 감사를 표합니다.

### 참 고 문 헌

1. 건설부, 1989, 국제수문개발계획(IHP) 대표 유역연구조사보고서
2. 건설부, 1986, 중규모다목적댐에비타당성조사최종보고서
3. 국토개발연구원, 1987, 중소수계의 하천유출량추정모델개발
4. 권순국의 5인, 1984, 응용수문학
5. 김양수의 5인, 1991, 년유출량추정모형개발, 한국수문학회지 제 24 권 제 3 호, pp.95-104
6. 김우철외 7인, 1989, 현대통계학
7. 이희연, 1989, 지리통계학
8. 윤용남, 1987, 공업수문학
9. 조희구, 1973, "기후요소에 의한 증발량 연구", 한국수문

학회지 제 6 권 제 1 호, pp.5-12

10. 조희구, 이태영, 1975, "낙동강유역의 증발산량과 물수지", 한국수문학회지 제 8 권 제 2 호, pp.81-92
11. 중앙기상대, 1982, 한국기후표
12. 한국건설기술연구원, 1989, 장기수자원계획을 위한 유역물수지 추정방법개발
13. 한국수자원공사, 1990, 수자원장기종합계획('90-2001)보고서
14. Binnie & Partners and Hyundai Engineering Co., 1978, Hydrological Services Rural Infrastructure Project Final Technical Report.
15. Chow, Ven Te, 1964, Handbook of Applied Hydrology, McGraw-Hill Book Company.
16. Dunne, T. and L. B. Leopold, 1978, Water in Environmental Planning, W. H. Freeman and Company.
17. Morton, F. I., 1983, "Operational Estimates of Area Evapotranspiration and Their Significance to the Science and Practice of Hydrology," Journal of Hydrology, 66.
18. Wonnacott, T. H. and R. J. Wonnacott, 1977, Introductory Statistics, John Wiley & Sons.

〈접수 : 1992년 6월 13일〉