

수위-유량관계식에 새로운 양방향 회귀모형의 적용

An Application of a New Two-Way Regression Model for Rating Curves

이 창 해*

Lee, Chang Hae

Abstract

Whether rating curves are used in practice or new ones are derived, the characteristics of regression analysis are often neglected. For example, a discharge rating curve, which is established from a regression of observed water levels (H) on observed flowrates(Q), is sometimes used for estimating a design water level corresponding to a simulated design flood runoff. However, if independent and dependent variables are changed with each other, the regression equation is changed in existing regression analysis, which is derived from vertical errors between observed data and regression line. Thus, regression equations should not be applied inversely. To avoid this problem, A new two-way variable least-squares regression analysis is proposed. The new method was applied to the rating curves of five water level stations on main stream of Nakdong River. The three kinds of regression models, which are respectively regression of Q versus H (model 1), H versus Q (model 2) and two-way (model 3), showed that the new method can reduce inadvertent mistakes when applied in practice.

keywords : Regression model, rating curve, Excel VBA

요 지

수위-유량관계식의 유도와 실무적용에 있어 통상적으로 회귀분석의 특성을 간과하고 사용하는 경우가 종종 발생한다. 예를 들어 실무에서는 관측수위로부터 관측유량으로 회귀분석되어 만들어진 수위-유량관계식을 홍수모형으로부터 모의된 설계홍수유출량으로부터 설계홍수위를 환산하는데 사용되기도 한다. 그러나 독립과 종속변수가 서로 바뀌면, 관측치와 회귀식간 연직거리의 잔차들로부터 유도된 기존의 회귀분석에 의하여, 회귀식이 서로 달라지기 때문에 역으로 적용하여서는 안 된다. 본 연구에서는 이런 문제점을 해결하기위해 회귀식의 변수들을 상호 교환할 수 있는 최소자승 회귀분석의 새로운 알고리즘을 제안하였다. 새로운 방법을 낙동강유역의 분류 5개 수위표지점의 수위-유량관계식에 대하여 적용하였다. 3가지 회귀식이 유도되었는데, 이들은 각각 수위로부터 유량으로(model 1), 유량으로부터 수위로(model 2) 그리고 양방향(model 3)으로 유도된 수위-유량관계식을 비교하여 실무에서 잘못 적용되는 실수를 줄일 수 있는 새로운 방법을 제시하였다.

핵심용어 : 회귀모형, 수위-유량관계식, Excel VBA

* 대전대학교 공과대학 환경공학과 부교수

Associate Professor, Dept. of Environmental Engineering, Daejin University, Gyeonggi-do, 487-711, Korea
(e-mail: chlee@daejin.ac.kr)

1. 서 론

물을 다루는데 있어서 유량에 대한 정확한 정보는 생명과 재산을 보호하기 위하여 매우 중요할 뿐만 아니라, 수자원 분석의 기본조건이며 댐운영, 수리 및 수문 모형의 개발, 수리권과 물분쟁의 소송을 위한 기록, 하천 및 수리구조물 설계, 유역의 유사량과 오염총량제 실시를 위한 기초자료로 이용된다. 그럼에도 불구하고 하천의 유량을 직접 측정하기 위해서는 많은 시간과 노력 및 비용이 소요되고, 특히 홍수에 대해서는 유량을 측정하기가 더욱 힘들다는 것은 잘 알려져 있다. 따라서 비교적 연속적으로 측정이 용이한 수위로부터 수위-유량관계곡선을 이용하여 유량으로 환산하여 사용하고 있다.

이러한 수위-유량관계식에 대해 국내에서 많은 연구(노재경 등, 2000; 이우석, 이길성, 2002; 박재영 등, 2006)가 진행되고 있으나, 수위-유량관계식의 근본적인 문제에 대해서는 소홀하였던 것도 사실이다. 따라서 본 연구에서는 과연 수위-유량관계식과 유량-수위관계식을 단순히 수학적 역함수관계로 사용될 수 있는지에 대한 의구심에서 출발하였다.

보통 관측 자료들을 사용하여 구해진 회귀식은 분명히 독립변수와 종속변수가 구분되어 유도되었음에도 불구하고 이 식을 사용함에 있어서는 간혹 그 구분을 무시하는 경향이 있다. 예를 들어 주어진 독립변수의 값을 회귀식에 대입하여 종속변수의 추정치를 구할 수 있다. 반대로 주어진 종속변수의 값을 회귀식의 역함수에 대입하여 독립변수의 추정치를 구할 수도 있다고 생각하지만, 실제로 자료들에서 독립변수와 종속변수가 바뀌게 되면 회귀식이 단순히 역함수관계가 아니라 다르게 나타난다는 사실에 유념해야 한다.

이렇게 독립·종속변수가 서로 바뀌면 회귀식이 서로 달라지는 이유는 최소자승법에서 자료점과 선형회귀식의 최단거리인 수직(perpendicular) 잔차들의 자승의 합을 최소화하는 것이 아니라 단순히 독립변수와 회귀식의 연직(vertical) 잔차들의 자승의 합을 최소화하는 형태로 회귀상수들이 유도되었기 때문이다. 따라서 본 연구에서는 선형회귀분석의 최소자승법에 있어 수직잔차의 자승의 합이 최소화되도록 하는 새로운 최소자승 회귀분석법을 수위-유량관계식에 적용하였다.

본 연구의 이론적 검토를 위해 김두섭 등(2000)의 단순 자료를 사용하여 분석하였고, 낙동강 유역의 사벌, 구미, 왜관, 성주, 고령교의 5개 수위표 지점에서 2005년 2월부터 2006년 8월까지의 수위와 유량의 측정 자료를 이용하여, 수위-유량의 관계를 나타내는 3가지 회귀

식이 유도되었다. 이들은 각각 수위(H)로부터 유량(Q)로(모형 1), 유량으로부터 수위로(모형2) 그리고 수위와 유량 양방향(모형3)으로 사용할 수 있도록 유도된 수위-유량관계식을 비교하여 그 동안 관행적으로 사용해온 문제점을 설명하고, 앞으로 독립·종속변수의 구분 없이 양방향으로 사용할 수 있는 수위-유량관계식을 유도하는 양방향 회귀모형을 제시하였다.

2. 최소자승 회귀분석의 검토

2.1 기존의 최소자승 회귀분석

선형회귀분석은 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ 와 같은 관측치에 직선을 적합(fitting)시키는 것이다.

따라서 관측된 y 값($y_{i,measured}$)과 회귀식으로 계산된 근사값($y_{i,model} = a_0 + a_1 x_i$) 사이의 잔차 ϵ_i 들의 자승의 합(S_r)을 최소화하는 것을 최소자승(least-square)법이라 한다(Carnahan 등, 1969, Ralston 등, 1978, Gerald 등, 1989). 이 경우 잔차를 자승하여 합을 구하므로 잔차가 큰 자료들의 영향이 더욱 크게 증가되어, 결국 시각적 접근법과 같은 결과를 나타내며 자료들의 최적 회귀식을 구하는데 도움이 된다. 이를 식으로 나타낸 것이 Eq. (1)이고 도식적으로 설명한 것이 Fig. 1이다.

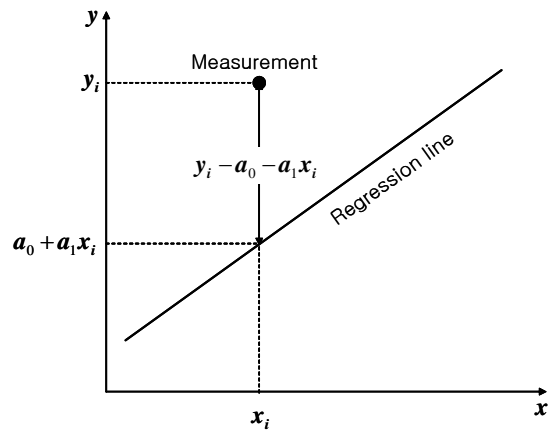


Fig. 2. Vertical Distance Between a Data Point and Regression Line

$$S_r = \sum_{i=1}^n \epsilon_i^2 = \sum_{i=1}^n (y_{i,measured} - y_{i,model})^2 \quad (1)$$

$$= \sum_{i=1}^n (y_i - a_0 - a_1 x_i)^2$$

여기서, i 는 자료의 번호이고, n 은 자료의 총 개수이다. Eq. (1)에서 회귀식의 절편 a_0 와 기울기 a_1 를 결정하기 위해서는 S_r 이 a_0 와 a_1 에 대해 최소값을 가져야 한다

는 점에 착안하여(정영진, 1982), a_0 와 a_1 에 대한 편미분 값이 0이 되어야 하므로 S_r 을 각각에 대해 편미분한 것이 Eqs. (2)와 (3)이다.

$$\frac{\partial S_r}{\partial a_0} = -2 \sum (y_i - a_0 - a_1 x_i) = 0 \quad (2)$$

$$\frac{\partial S_r}{\partial a_1} = -2 \sum [(y_i - a_0 - a_1 x_i) x_i] = 0 \quad (3)$$

여기서, \sum 는 i 가 1부터 n 까지 더한 것을 의미하며, 기호를 간편하게 하기 위한 것으로 다른 언급이 없는 경우 앞으로도 이렇게 적용한다. Eqs. (2)와 (3)이 0이 되도록 연립으로 풀면 a_0 와 a_1 을 Eq. (4)와 (5)와 같이 구할 수 있다(Chapra 등, 1998).

$$a_1 = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2} \quad (4)$$

$$a_0 = \bar{y} - a_1 \bar{x} \quad (5)$$

여기서, $\bar{y} = \sum y_i / n$ 이고 $\bar{x} = \sum x_i / n$ 으로서 관측자료 y 와 x 의 평균이다. 결과적으로 회귀식은 다음과 같이 나타낼 수 있으며, Eq. (6)에서 x_i 는 독립변수이고 \hat{y}_i 는 종속변수로서 회귀식에 의한 추정 값이다.

$$\hat{y}_i = a_0 + a_1 x_i \quad (6)$$

2.2 기존의 최소자승 회귀분석의 문제점

선형 회귀분석은 자료에 최적의 직선을 적합시키는 가장 보편적인 방법으로 사용되고 있다(박성현, 1998). 그러나 Eq. (6)의 회귀식은 독립변수 x 로부터 종속변수 y 의 반응결과를 \hat{y} 로 추정하는데 사용되는 것으로, y 로부터 x 의 반응결과를 나타내는데 사용할 수 없다는 문제점이 있다. 즉, x 로부터 y 의 관계를 나타내는 회귀식과, y 로부터 x 의 관계를 나타내는 회귀식이 일치하지 않는다는 것이다.

여기서 Table 1에 나타낸 임의의 자료(김두섭 등, 2000)를 이용하여 이들 관계식을 유도해 보기로 한다.

우선 Table 1 자료에서 독립변수를 x 로, 종속변수

를 y 로 하여 Eqs. (4)와 (5)로 유도한 회귀식은 다음과 같다.

$$\hat{y}_i = 25 + 0.6 x_i \quad (7)$$

또한 같은 자료로 Fig. 2의 개념에 따른 Eq. (8)을 사용하면, 독립변수를 y 로, 종속변수를 x 로 하여 S_r 을 최소화시킨 회귀식 Eq. (9)가 구해진다.

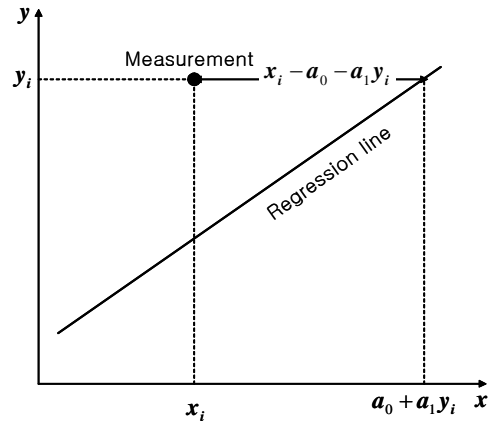


Fig. 3. Typical Drawing when y is the Independent Variable

$$S_r = \sum_{i=0}^n \epsilon_i^2 = \sum_{i=0}^n (x_i - a_0 - a_1 y_i)^2 \quad (8)$$

$$\hat{x}_i = -27.777 + 1.3436 y_i \quad (9)$$

여기서 $\text{hat}(y_i)$ 또는 \hat{x}_i 은 각 회귀식의 종속변수 추정 값을 나타낸다. Eq. (9)의 선형직선을 \hat{x}_i 의 함수로 변환하면 다음 식과 같아진다.

$$y_i = 20.674 + 0.7442 \hat{x}_i \quad (10)$$

Table 1의 자료와 Eqs. (7)과 (10)을 그림으로 나타내면 다음과 같다.

Fig. 3의 범례에서 y 는 자료의 값을 나타내고, $x \rightarrow y$ 는 x 를 독립변수로 회귀분석한 결과로서 Eq. (7)을 나타내며, $y \rightarrow x$ 는 y 를 독립변수로 회귀분석한 결과로서 Eq. (9) 또는 (10)을 의미한다.

Table 1. Data for Deriving Regression Equations

x	5	10	15	20	25	30	35	40	45	50	55
y	34	25	28	43	38	39	52	54	50	52	58

Fig. 3에서 나타난 바와 같이 자료 x 와 y 를 회귀분석함에 있어 독립변수를 x 또는 y 로 다르게 사용함으로써 그 회귀분석의 결과는 다르게 나타난다.

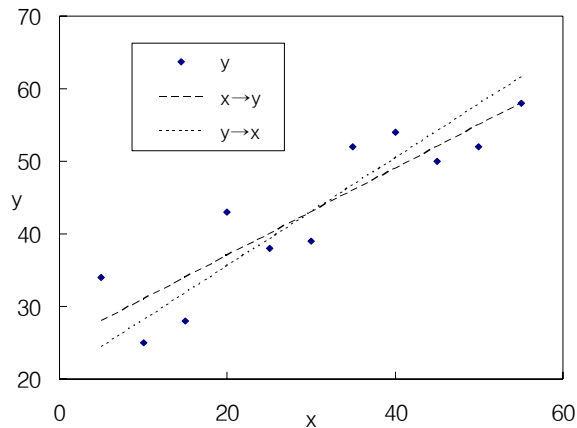


Fig. 4. Regression Equations Resulting from the Change of the Independent Variable

2.3 새로운 최소자승 회귀식의 유도

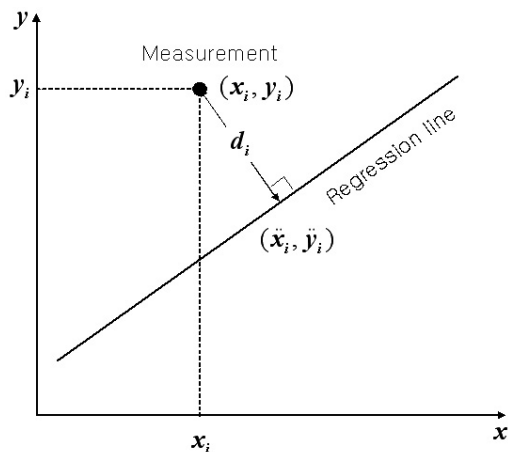


Fig. 5. Typical Drawing of Two-way Regression

Fig. 1과 2에서는 관측자료점과 회귀식의 연직(vertical) 또는 수평(horizontal)거리들을 잔차로하여 이들 잔차들의 자승합을 최소화하는 것이다. 그러나 Fig. 4는 관측점에서 회귀식에 수직(perpendicular)이 되는 최소거리 d_i 들의 자승합이 최소가 되도록 회귀식을 유도하는 새로운 개념의 최소자승법의 모식도이다. 따라서 Fig. 4에서의 회귀식을 Eq. (11)이라 할 때 점 (x_i, y_i) 에서 Eq. (11)에 수직으로 만나는 점을 $(\tilde{x}_i, \tilde{y}_i)$ 라 하면 이들 점간의 거리 d_i 는 Eq. (12)와 같이 유도된다.

$$y = a_0 + a_1x \quad (11)$$

$$d_i^2 = \frac{(y_i - a_0 - a_1x_i)^2}{a_1^2 + 1} \quad (12)$$

최소자승법의 정의에 따라 회귀식에 수직거리 d_i 의 자승의 합은 다음 식과 같다.

$$S_r = \sum_{i=1}^n \frac{(y_i - a_0 - a_1x_i)^2}{a_1^2 + 1} \quad (13)$$

여기서 Eq. (13)을 Eq. (1)과 비교하면 단지 $a_1^2 + 1$ 의 항이 분모에 추가된 점이 다를 뿐이다. 회귀상수 (a_0, a_1) 을 유도하기 위하여 Eq. (13)을 a_0 과 a_1 으로 각각 미분하고 미분결과가 각각 0이 되는 a_0 과 a_1 을 구하여야 한다.

$$\frac{\partial S_r}{\partial a_0} = \frac{-2}{a_1^2 + 1} \sum (y_i - a_0 - a_1x_i) = 0 \quad (14)$$

$$\frac{\partial S_r}{\partial a_1} = -2 \sum \left[\frac{(y_i - a_0 - a_1x_i)x_i}{a_1^2 + 1} + \frac{(y_i - a_0 - a_1x_i)^2 a_1}{(a_1^2 + 1)^2} \right] = 0 \quad (15)$$

그러나 미분결과가 Eqs. (14)와 (15)와 같이 복잡하게 나타나기 때문에 기존의 방법처럼 a_0 와 a_1 을 Eq. (4)와 (5)와 같이 유도하여 해석적으로 계산하는 것은 불가능하다. 이러한 이유로 인하여 지금까지 자료점과 회귀선까지의 수직거리 d_i 의 자승의 합이 최소가 되는 회귀분석법이 사용되지 못하였던 것으로 판단된다. 그러나 Eqs. (14)와 (15)를 해석적으로 구하지 않고, Eq. (13)을 수치해석적으로는 최소화시키는 것이 가능하다. 수학적으로 충분히 검증되어 사용되고 있는 최적화 기법인 Excel의 “해 찾기” 모형(이종혁, 1999; Orvis, 1996)을 사용하여 연직잔차들의 자승합인 Eq. (1)이 아닌 수직잔차들의 합인 Eq. (13)을 최소화시켜 이 문제를 해결할 수 있다.

Fig. 5에서 d^2 은 Eq. (12)를 계산하는 열을 나타내는 것이고, $\sum = 608.5$ 는 Eq. (13)을 의미하며 이 값을 최소화하는 “목표 셀”로 두고, 회귀상수 a_0 과 a_1 을 “값을 바꿀 셀”로 지정하여 해 찾기를 실행하여 회귀상수를 구한 결과가 Fig. 6이다.

Fig. 6에서 최소자승의 값은 608.5에서 171.9637로 최소화되었으며, 회귀상수는 $(a_0=23.79934, a_1=0.640022)$ 로

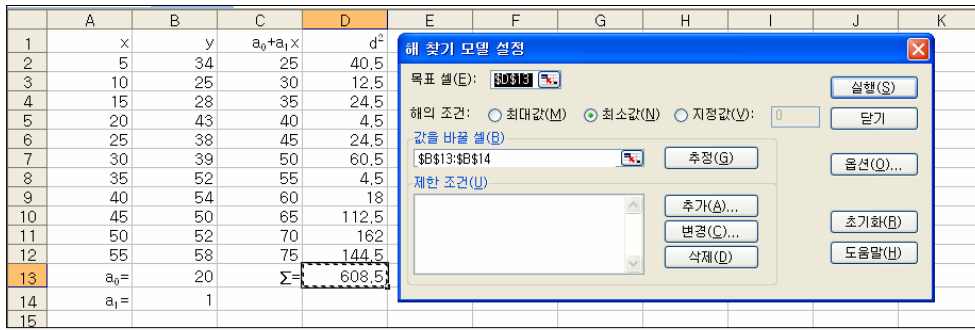


Fig. 6. Setting Excel Solver Model for the New Regression Model

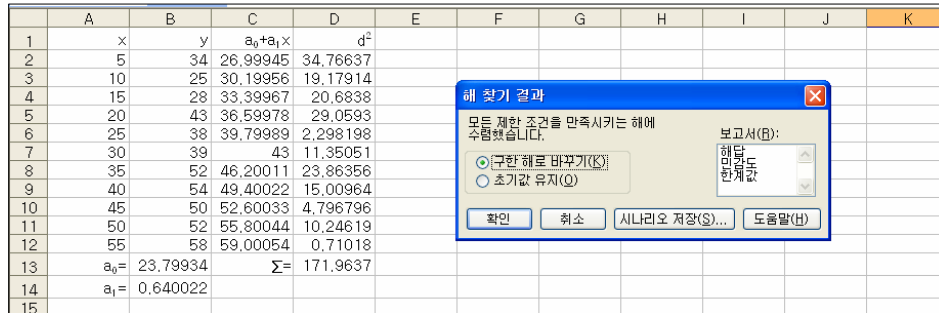


Fig. 7. Result of Excel Solver Using a New Regression Model

계산되었다. 이 결과를 다음과 같은 그림으로 나타낼 수 있다.

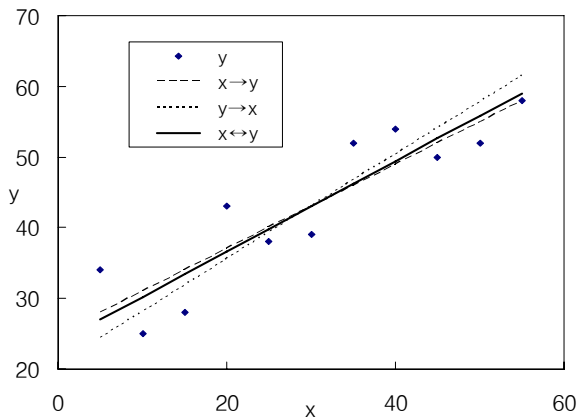


Fig. 8. Result of the New Regression Model

Fig. 7의 범례에서 y 와 $x \rightarrow y$ 및 $y \rightarrow x$ 는 앞에서 Fig. 3의 설명과 같고, $y \leftarrow x$ 는 본 논문에서 제시하고 있는 새로운 회귀분석방법으로 구한 회귀식을 나타낸다. 즉 독립변수와 종속변수를 상호 교환하여 사용할 수 있는 새로운 회귀식이 되는 것이다. 이러한 새로운 회귀식은 다음과 같다.

$$\hat{y}_i = 23.799 + 0.6400x_i \quad (16)$$

3. 양방향 회귀모형의 적용

본 연구에서 제안한 양방향 회귀모형을 수위-유량관계식에 적용하여 실제자료에 대하여 검토해 보았다. 사용된 자료는 낙동강 유역의 사벌, 구미, 왜관, 상주, 고령교 등 5개 수위표 지점의 2005년 2월부터 2006년 8월까지의 유량측정성과자료를 이용하였다(한강홍수통제소, 2007). 이 수위표 지점들의 자료를 사용한 것은 김삼은 등(2007)이 수위-유량관계식 검증에 이미 사용되어 자료의 신뢰도가 높았기 때문이다.

수위-유량관계식은 지수형과 포물선형 등 여러 가지의 형태로 제시되고 있으나, 수리학적 원리에 더 적합한 다음과 같은 지수형 함수를 많이 사용한다.

$$Q = A(H+z)^B \quad (17)$$

여기서, $-z$ 은 유량이 0이 되는 수위를 의미하고, A 와 B 는 회귀 상수이다.

Eq. (17)은 비선형식이므로 z 값을 가정하고 양변에 \log 를 취해 선형화시키고, 선형화된 회귀식의 기울기와 절편은 Eq. (4)와 (5)처럼 해석적으로 구할 수 있다. 이때 선형회귀식의 잔차들의 자승합 S_r 이 최소대가 되는 z 을 결정하기 위해 실무에서는 시행착오법으로 여러 번 반복 수행하는 번거로움이 있다. 따라서 몇 번의 시

도를 통해 z 을 포함한 Eq. (17)의 상수를 결정하므로 자료에 가장 적합된 결과라고 하기에 부족함이 있다. 이러한 이유 때문에 Table 2의 수문조사연보상의 z 값이 -0.200, 0.150, 0.100 등으로 나타나기도 한다.

이러한 문제를 해결하기 위하여 최적화 문제를 구성하였는데, z 을 결정변수로 Eq. (1)의 S_r 은 목적함수로 하면 회귀상수 A 와 B 는 종속적인 해석해로 결정된다. 이때 최적화 방법으로 황금분할법을 사용하여 B 가 비정상적으로 커져 발산하거나 $(H+z)$ 이 음수가 되어 $\log(H+z)$ 가 계산 불능의 상태가 되는 것을 막을 수 있었다. 이러한 회귀분석과 함께 황금분할법으로 최적화하는 방법은 수위를 독립변수로 하는 모형 1($H \rightarrow Q$)과 유량을 독립변수로 하는 모형 2($Q \rightarrow H$)에서 모두 사용할 수 있었다.

그러나 양방향 회귀분석인 모형 3($H \leftrightarrow Q$)의 경우 최적화 목적함수가 Eq. (13)이고, 결국 Eqs. (14)와 (15)는 해석적인 해를 구할 수 없으므로, A 와 B 및 z 등 3개가 최적화의 결정변수가 된다. 일반적으로 최적화 문제에서 3개 이상으로 많아지면 수렴에 어려움이 나타나게 된다. 수위-유량관계식에서도 A 는 작아지고 B 가 과도

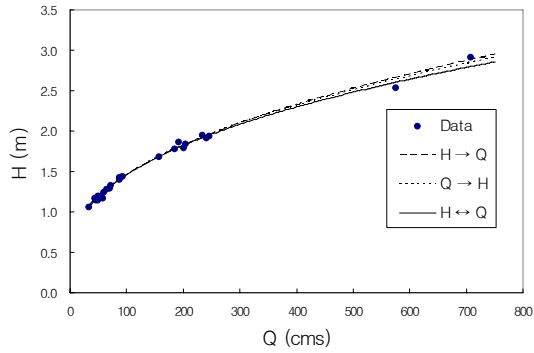
하게 커지면서 발산하는 경우가 종종 발생한다. 이를 해결하기 위하여 모형 1에서 결정된 회귀상수들을 모형 3의 초기 값으로 입력하고, 이때 계산된 Eq. (13)의 S_r 이 최적화 과정에서 25% 이상 감소하면 수렴하는 것으로 제한조건을 지정함으로써 불필요한 발산을 막으면서도 최적화의 목적을 달성할 수 있었다.

각 수위표 지점에는 수위구간 별로 3~4개의 수위-유량관계식을 사용하고 있으나, 본 연구에서는 각 수위표 별로 자료수가 가장 많은 수위 구간의 자료만을 대상으로 하였다. 또한 본 연구는 회귀분석의 새로운 접근 방법을 제시하는데 주안점을 두고 있으나, Table 2의 회귀식들을 산정하는 과정에서 사별, 구미, 고령교와 같이 저수위곡선인 경우는 유량이 0이 되는 수위라는 의미에 주의해서 z 을 결정해야 한다.

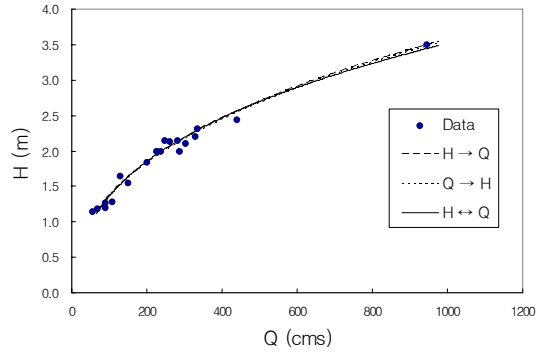
Table 2에서 첫 번째 열의 정보는 관측소 이름, 수위-유량관계식(두 번째 열)의 적용범위, 수위표상 수위(h)와 관계식에 대입하는 수위(H)의 조정치, 유의수준 $\alpha = 0.01$ 을 사용한 F 의 기각치 등을 나타내고 있다. 세 번째 열의 R^2 는 관계식의 결정계수이고, 네 번째 열의 F 는 관계식의 유의성 판정을 위한 분산분석의 결과로서

Table 2. Regression Result of Each Models and Stations

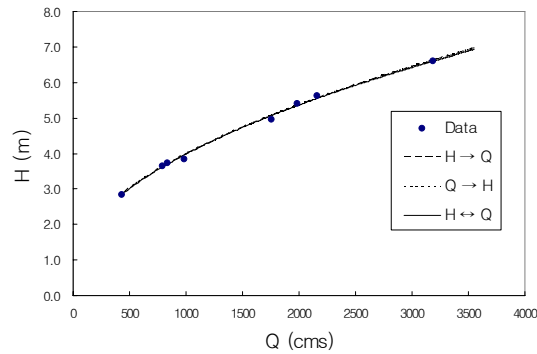
Water Level Station	Rating Curves	R^2	F	Remark
Sabeol $1.06 \leq H \leq 2.92$ $H = h + 1$ $F(0.01)=7.95$	$Q = 52.455(H - 0.200)^{2.660}$	0.986	3162	Annual report
	$Q = 102.382(H - 0.465)^{2.192}$	0.990	3470	$H \rightarrow Q$ (model 1)
	$Q = 88.813(H - 0.406)^{2.325}$	0.994	3653	$Q \rightarrow H$ (model 2)
	$Q = 69.042(H - 0.306)^{2.548}$	0.983	3588	$H \leftrightarrow Q$ (model 3)
Gumi $1.13 \leq H \leq 4.00$ $H = h + 3$ $F(0.01)=8.29$	$Q = 37.094(H + 0.150)^{2.380}$	0.951	454	Annual report
	$Q = 27.863(H + 0.259)^{2.662}$	0.986	611	$H \rightarrow Q$
	$Q = 39.191(H + 0.085)^{2.505}$	0.978	608	$Q \rightarrow H$
	$Q = 10.586(H + 0.666)^{3.178}$	0.983	593	$H \leftrightarrow Q$
Waegwan $2.85 \leq H \leq 6.61$ $H = h + 3$ $F(0.01)=13.75$	$Q = 25.164(H + 0.100)^{2.600}$	0.969	578	Annual report
	$Q = 105.089(H - 0.768)^{1.931}$	0.997	2333	$H \rightarrow Q$
	$Q = 121.513(H - 0.886)^{1.869}$	0.997	2398	$Q \rightarrow H$
	$Q = 89.163(H - 0.646)^{2.005}$	0.997	2220	$H \leftrightarrow Q$
Seongju $1.30 \leq H \leq 2.09$ $H = h + 2$ $F(0.01)=9.07$	$Q = 423.080(H - 0.950)^{2.529}$	0.949	266	Annual report
	$Q = 20.776(H - 0.248)^{6.120}$	0.961	412	$H \rightarrow Q$
	$Q = 20.153(H - 0.252)^{6.294}$	0.970	412	$Q \rightarrow H$
	$Q = 5.8314(H - 0.059)^{7.170}$	0.960	403	$H \leftrightarrow Q$
Goryeonggyo $1.43 \leq H \leq 3.45$ $H = h + 3$ $F(0.01)=8.40$	$Q = 19.341(H + 0.170)^{2.980}$	0.769	95	Annual report
	$Q = 152.221(H - 0.855)^{1.580}$	0.827	107	$H \rightarrow Q$
	$Q = 9.5871(H + 0.283)^{3.511}$	0.853	111	$Q \rightarrow H$
	$Q = 105.119(H - 0.694)^{1.973}$	0.816	101	$H \leftrightarrow Q$



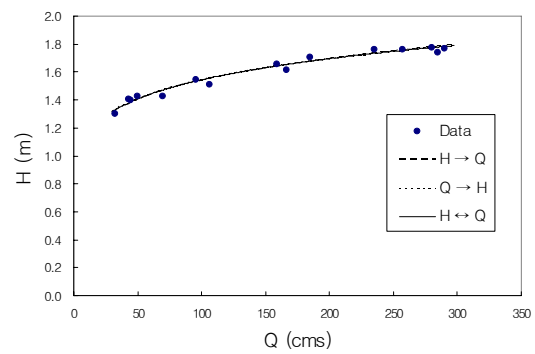
(a) Result of Sabeol station



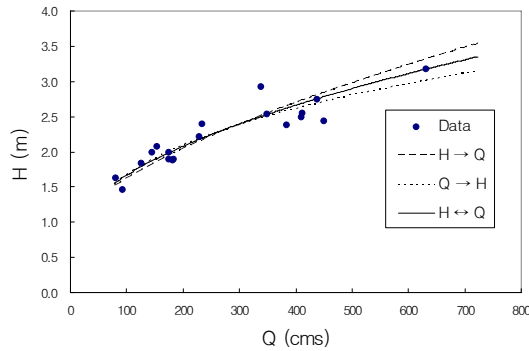
(b) Result of Gumi station



(c) Result at Waegwan station



(d) Result at Gumi station



(e) Result at Goryeonggyo station

Fig. 8. Comparison with Data and Rating Curves

각 모형별로 유도되는 수위-유량관계식이 매우 유의하다는 것을 보여주고 있다. 마지막 열의 Annual report는 수문조사연보상의 회귀식을 나타내고 있으며, 모형 1($H \rightarrow Q$)은 수위를 독립변수로 회귀분석한 것이며, 모형 2($Q \rightarrow H$)는 유량을 독립변수로 회귀분석한 것이고, 모형 3($H \leftrightarrow Q$)은 양방향 회귀분석을 의미한다.

Table 2의 2번째 열은 모형 1, 2, 3으로 도출한 수위-유량관계식을 나타내며 이 모형들의 프로그램은 Excel의 VBA로 작성하여 실무자들이 쉽게 접근이 가능토록 하였다. 단, 여기서 제시한 관계식은 기존의 방법과 본 연구의 방법들의 비교하기 위한 것이므로 실무에 적용

하고자 할 경우 계수들의 물리적 의미에 주의를 기울여 결정해야 한다.

낙동강유역의 5개 수위표 지점에서의 유량측정성과 자료와 Table 2에 나타난 각 모형별 수위-유량관계식을 도시한 것들이 Fig. 8이다. 결정계수 값들이 0.96 이상인 사벌과 구미, 왜관, 성주 수위표 지점에서는 모형들의 수위-유량관계식이 구분이 어려울 정도로 비슷한 결과가 나왔고(Fig. 8(a)~(d)), 고령교 수위표 지점에서는 모형들 간 관계식들의 차이를 Fig. 8(e)와 같이 육안으로도 쉽게 확인될 정도로 나타내고 있다.

현재까지 사용되고 있는 일반적인 수위-유량관계식

Table 3. Result of Traditional Regressions

(a) at Sabeol station

$H \rightarrow Q$		$Q \rightarrow H$		ΔH (m)
H (m)	Q (m ³ /s)	H (m)	Q (m ³ /s)	
1.1	37.84	1.099		0.001
1.7	162.61	1.703		-0.003
2.3	387.36	2.290		0.010
2.9	720.16	2.866		0.034

(b) at Goryeonggyo station

$H \rightarrow Q$		$Q \rightarrow H$		ΔH (m)
H (m)	Q (m ³ /s)	H (m)	Q (m ³ /s)	
1.5	76.13	1.521		-0.021
2.0	188.53	2.053		-0.053
2.6	366.87	2.541		0.059
3.2	585.19	2.942		0.258

은 수위에서 유량으로 회귀분석을 수행하여 모형 1 ($H \rightarrow Q$)에 해당한다. 따라서 수위측정치로부터 모형 1에 의해 구한 수위-유량관계식을 사용하여 유량을 환산하는 것은 타당하다. 그러나 홍수예보 모형 등에서 결과로 나온 유량을 모형 1의 수위-유량관계식으로부터 역산하여 수위로 환산하는 것은 회귀분석의 근본 원리 때문에 타당하지 못하다. 따라서 이 경우는 모형 2에서 유도된 유량-수위관계식을 사용하여 유량으로부터 수위로 환산하여야 한다.

이러한 수학적 근거를 바탕으로 계산한 결과들 중 상류와 하류의 사별과 고령교 수위표 지점의 결과들을 Table 3에 정리하였다. 유량이 작을 때보다 유량이 클 때 차이가 더 심해지는 경향을 보이고 있으며, 사별 수위표 지점의 경우 유량 720m³/s의 유량 정도에서 약 3.4cm의 수위차가 나타났으며, 고령교 지점에서는 585m³/s의 유량 정도에서 25.8cm 수위차가 나타났다.

이러한 혼란은 회귀분석에서 독립변수와 종속변수의 방향성에 따른 것으로, 연직잔차의 자승합을 최소화(모형 1과 2)할 것이 아니라 관측자료점과 회귀식의 실제적인 최소거리인 수직잔차의 자승합을 최소화시켜야 한다는 모형 3의 당위성을 보여주는 것이다. 물론 본 연구에서 제시한 새로운 양방향 회귀모형인 모형 3을 사용한다면, Table 3에 나타난 바와 같이 $\Delta H \neq 0$ 결과는 당연히 사라지게 된다.

수자원분야 연구에서 가정 어려움을 겪고 있는 것 중 하나는 양질의 수문자료가 부족하다는데 있다. 하천의 유량은 실시간으로 상시 관측되는 수위자료로부터 수위-유량관계식을 이용하여 얻고 있으므로 정확한 수위-유량관계식을 구하는 일은 매우 중요하다. 본 연구에서 제시한 양방향 회귀분석 모형은 수위-유량관계를 가역적으로 표현하는 모형이다. 이러한 모형의 결과가 일부 회귀 결정계수가 높은 지점에서 큰 개선 효과를 수치적으로 나타내기 어려울지라도 적용과정에서 발생할 수 있는 혼란이나 오차를 줄일 수 있는 역할을 할 수 있으리라 기대한다.

그 동안 중요성이 간과되었지만 유량은 모든 수자원

분야의 기초가 되는 매우 중요한 자료인 것이다. 정확한 수문자료의 구축을 위해 건교부에서 2007년 1월부터 유량조사사업단을 가동시킨 것은 다행스러운 일로 앞으로 수자원 분야 발전에 크게 기여할 것으로 판단된다.

4. 결 론

일련의 자료에 대하여 회귀분석을 할 때 독립변수와 종속변수가 완전히 구분되어 회귀식이 유도된다. 그렇기 때문에 독립변수로부터 회귀식을 통해 종속변수의 값으로 환산되는 것이고 역함수 관계는 성립하지 않는다. 그러나 이런 중요한 문제가 간과되어 쉽게 독립과 종속변수를 혼돈하여 사용해 오고 있다. 수자원 분야에서 가장 기본 자료인 유량을 다루는 수위-유량 관계식의 경우에도 이러한 잘못된 사용이 빈번하게 발생되고 있다.

본 연구에서는 이런 문제점을 부각시켰고, 그 과정에서 수위-유량관계식의 상수 z, A, B 를 결정하기 위해 선형회귀모형과 황금분할 최적화를 병행해서 사용하는 모형 1과 2를 개발하였다. 한편 연직잔차의 자승합을 최소화하는 기존의 회귀분석을 수직잔차의 자승합을 최소화하여 양방향 회귀분석이 가능한 새로운 모형 3을 제안하였다. 모형 3은 Excel의 “해 찾기”를 활용하여 최적화하였다. 이 모형들의 프로그램에는 Excel의 VBA를 사용함으로써 실무자들의 접근성을 용이하게 하였으며, 이들 모형을 낙동강 유역의 5개 수위표지점에 대하여 적용하여 그 타당성을 나타내 보였다.

본 연구에서 제시한 양방향 회귀모형은 수위-유량관계식에만 국한되는 것이 아니라 자연과학분야와 사회과학분야에도 적용범위가 넓어지기를 희망한다.

참 고 문 헌

김두섭, 강남준 (2000). 회귀분석 기초와 응용. 나남출판사.
 김삼은, 강규상, 정찬용, 임혁진, 김형섭, 정성원 (2007). “2006년도 낙동강분류 5개 지점 수위-유량관계곡선

- 식 검증.” **한국수자원학회 07 학술발표회 논문집**, 한국수자원학회, pp. 1814-1818.
- 노재경, 이우석, 이한구 (2000). “수위-유량 관계식의 예민도 분석.” **한국수자원학회 2000년 학술발표회 논문집**, 한국수자원학회, pp. 112-117.
- 박성현 (1998). 회귀분석(제3판). 민영사.
- 박재영, 오병동, 김재복, 채효석 (2006). “신개념의 수위-유량곡선식 개발과 적용 방향.” **한국수자원학회 06 학술발표회 논문집**, 한국수자원학회, pp. 144-148.
- 이우석, 이길성 (2002). “비선형 회귀모형에서 가중최소 제곱법에 의한 수위-유량곡선식 개발.” **한국수자원학회 02 학술발표회 논문집(I)**, 한국수자원학회, pp. 541-546.
- 이종혁 (1999). Microsoft 한글 엑셀 2000 Bible. 정보문화사.
- 정영진 (1982). 실용현대통계학. 선진문화사, pp. 44-66.
- 건설교통부 (2007). **한국수문조사연보(유량편) 2006**.
- Carnahan, B., Luther, H.A. and Wilkes, J.O. (1969). *Applied Numerical Methods*. Wiley.
- Chapra, S.C. and Canale, R.P. (1998). *Numerical Methods for Engineers: with Programming and Software Application, 3rd ed*. McGraw-Hill, pp. 330-565.
- Draper, N.R. and Smith, H. (1981). *Applied Regression Analysis, 2nd ed*. Wiley.
- Gerald, C.F. and Wheatley, P.O. (1989). *Applied Numerical Analysis, 4th ed*. Addison-Wesley.
- Orvis, W.J. (1996). *Excel for Scientists and Engineers, 2nd ed*. Sybex.
- Ralston, A. and Rabinowitz, P. (1978). *A First Course in Numerical Analysis, 2nd ed*. McGraw-Hill.

(논문번호:07-134/접수:2007.11.26/심사완료:2007.12.18)