의사우도추정법에 의한 분산함수를 고려한 수위-유량 관계 곡선 산정법 개선

Improvement of Rating Curve Fitting Considering Variance Function with Pseudo-likelihood Estimation

이우석*, 김상욱**, 정은성***,이길성****

Woo Seok Lee, Sang Ug Kim, Eun-Sung Chung, Kil Seong Lee

요 지

수위-유량 관계 곡선식에 포함되어져 있는 매개변수를 추정하기 위해 많이 사용되는 로그선형 회귀분석은 잔차의 비등분산성(heterocesdascity)을 고려하지 못하므로 본 연구에서는 의사우도추정법(Pseudo-likelihood Estimation, P-LE)에 의해 분산함수를 추정하고 이와 함께 회귀계수를 추정할 수 있는 방법을 제시하였다. 이 과정에서 제시된 회귀잔차를 최소화하기 위하여 SA(simulated annealing)이라는 전역 최적화 알고리즘을 적용하였다. 또한 수위-유량 관계 곡선식은 단면 등의 영향으로 인해 구간에 따라 각각 다르게 구축되어져야 하므로 이를 보다 객관적으로 판단하고 분리 위치를 정확히 추정하기 위하여 Heaviside 함수를 의사우도함수에 포함시켜 결과를 추정하도록 하였으며, 2개의 구간을 가지는 유량자료를 이용하여 제시된 방법의 합리성을 통계적으로 실험하였다. 이와 같이 통계적 실험을 통해 제시된 방법들이 기존 방법과 비교하여 가질 수 있는 장점을 파악하였으며, 제시된 방법들을 금강유역 5개 지점에서 대해 수행하여 효율성을 검증하였다.

핵심용어 : 수위-유량 관계 곡선식, 비등분산성, 의사우도추정법, 비선형회귀분석, Simulated Annealing, 수위-유량 관계 곡선식의 분리

1. 서 론

정확한 유량자료의 확보는 이수, 치수, 수질관리 등의 수자원관리에 있어서 가장 기본적이고 중요한 일이라 할 수 있다. 대부분의 선진국들은 이에 대해 많은 노력을 투자하고 있으며 국내에서도 최근 유량조사사업단이 발족되어 정확도 높은 유량자료의 취득과 관리를 위한 노력을 기울이고 있다. 그러나 대부분의 노력이 유량자료 획득을 위한 유량조사 사업, 실시간 모니터링 시스템, 자동화기기, 측정 장비 등에 집중되고 있을 뿐 가장 많이 사용하고 있는 수위-유량 관계곡선으로부터 산정되는 유량의 정확성과 관련된 기초 연구는 부족한 실정이다. 유량측정은 직접 측정하여 자료를 확보할 수도 있지만 연속된 자료를 지속적으로 확보하기 위해서는 시간과 비용이 막대하게 소요되므로 수위를 먼저 측정하고 기 작성된 수위-유량 관계곡선을 이용하여 유량을 환산하고 있다. 따라서 대부분의 국가에서는 측정된 수위를 이용하여 유량을 환산하기 위한 여러 가지 방법들을 개발하여 사용하고 있다(ISO, 1998; Mosley and McKerchar, 1993). 이 중 가장 보편적인 것은 Lambie (1978)가 제시한 Eq. (1)과 같은 비선형 관계식이다.

$$y = a(x-b)^{c} \tag{1}$$

여기서, y는 유량, x는 수위, a, b, c는 수위와 유량관계를 이용하여 추정되어져야 하는 매개변수들이다.

수위-유량 관계곡선식을 이용하여 유량을 산정하는 경우에는 관계식의 매개변수들을 추정해야하는 데, 일반적으로 로그선형 회귀분석이 이용된다. 이때 정확한 회귀 계수의 추정을 위해서는 회귀모형을 구성하는

* 정회원 · 한국수자원공사 조사기획처 차장 · E-mail : leews@kwater.or.kr

** 정회원 · 서울대학교 BK21 SIR 사업단 박사후 연구원 · E-mail : plethor1@snu.ac.kr

*** 정회원 · 서울대학교 공학연구소 선임연구원 · E-mail : cool77@snu.ac.kr

**** 정회원 · 서울대학교 공과대학 건설환경공학부 교수 · E-mail : kilselee7@snu.ac.kr

잔차의 특성을 분석해야한다. 즉 일반 최소자승법(ordinary least squares method, OLS)를 사용하기 위해서는 잔차가 등분산성(homoscedasticity)을 만족해야하며 그렇지 않은 경우는 다른 방법을 사용해야 한다. 특히 대부분의 수위-유량 관계에 있어서 유량이 증대함에 따라 오차가 증가되는 경향이 있으므로, 선형 회귀분석을 수행하는 경우 회귀잔차의 특성에 따라 회귀모형의 정확성이 크게 영향을 받을 수 있다. 이와 같은 문제점을 극복하기 위해 Seber and Wild(1989)는 잔차가 비등분산(heteroscedasticity)적인 경우에도 적용할 수 있는 일반화 최소자승법(Generalized Least Squares method, GLS)을 이용한 회귀분석을 적용한 바 있다. 따라서 본 연구에서는 비등분산적인 잔차를 고려하여 회귀분석을 수행할 수 있는 GLS 방법을 이용하되, 추정과정에 있어서 효율성을 증대시키기 위해 분산함수와 회귀계수를 추정할 때 의사우도추정법(pseudo-likelihood estimation method, P-LE)을 사용하였다. 그 결과를 잔차의 비등분산성을 고려하지 못하는 선형 회귀분석결과와 비교하였다. 또한 수위-유량 관계 곡선식의 분리 유무를 결정하기 위하여 Heaviside 함수를 의사우도함수(pseudo-likelihood function)에 포함시켜 분리위치를 함께 추정하도록 하였으며, 곡선의 객관적 분리 기준으로 Akaike(1974)가 제시한 AIC(Akaike Information Criterion)와 Schwartz(1979)가 제시한 BIC(Bayesian Information Criterion)를 이용하였다. 마지막으로 금강유역의 5개 수위관측지점의 유량자료(한국수자원공사, 2006; 한국수자원공사, 2007)를 이용하여 효율성을 검증하였다.

2. 일반화 최소자승법을 이용한 회귀분석과 의사우도추정법

일반적으로 OLS 방법은 회귀분석은 잔차의 등분산성이 만족되어야 정확성 있는 회귀계수를 추정할 수 있다. 그러나 수위와 유량관계에 있어서는 수위가 커짐에 따라 오차가 일정하지 않고 증가하므로 오차의 등분산 가정에 위배될 가능성이 크다. 이와 같이 잔차가 비등분산성을 가지는 경우에는 자기상관성(autocorrelation)을 고려하여야 하는데 잔차의 자기상관성이 없는 경우에는 WLS을 사용해도 무방하지만 그렇지 않은 경우에는 GLS를 사용하여야 보다 정확한 회귀계수를 얻을 수 있다. 만약 $E(\varepsilon_i)=0$이 성립된다면 종속변수의 평균값에 대하여 Eq. (2)가 성립될 것이며, 비등분산적인 성질을 고려한다는 것은 각각의 잔차에 가중치를 부여하는 것으로 해결할 수 있는데 이는 종속변수에 가중치를 부여하는 것과 궁극적으로 동일하다. 따라서 ε_i의 분산인 $\hat{\sigma}_i^2$은 종속변수 y_i의 분산과 같으며 이를 Eq. (3)과 같이 나타낼 수 있다.

$$E(y_i)=\mu_i(\beta)=f(x_i,\ \beta) \tag{2}$$

$$Var(y_i)=Var(\varepsilon_i)=\sigma_i^2=\sigma^2 g^2(\mu_i(\beta),\Theta)=\sigma^2/w_i \tag{3}$$

여기서, μ_i는 x_i에서의 종속변수의 참값에 대한 평균, β는 추정되어질 회귀계수 벡터, σ^2는 어떤 상수값으로 표현될 수 있는 분산, g는 분산의 비등분산성을 표현할 수 있는 분산함수(Variance function), Θ는 분산함수의 매개변수 벡터, w_i는 i번째 가중치를 나타낸다. 위의 식에서 가중치를 고려하지 않고 분산함수를 1로 취급하면 OLS가 되고, 모든 가중치가 임의의 기지값으로 사용될 수 있다면 WLS가 되며, 가중치가임의의 분산함수를 이용하여 추정되어져야 하는 미지의 값인 경우에는 GLS라 명명되어 사용된다. 추정된 분산함수를 이용하여 회귀계수를 추정하고자 하는 목적함수는 다음 Eq. (4)와 같이 나타낼 수 있으며, 가중치는 분산함수의 제곱에 반비례한다.

$$\min_{\beta} \quad \sum_{i=1}^{n} w_i (y_i - f(x_i,\ \beta))^2 \tag{4}$$

본 연구에서 사용된 의사우도추정법은 회귀계수를 기지의 값으로 하여 Θ와 β를 추정하게 되므로 최종적인 의사우도함수는 Eq. (5)와 같이 나타낼 수 있다.

$$\ln L_{PL}(\beta,\ \Theta)=-n\ln[\sigma_c(\widehat{\beta_*},\ \Theta)]-\sum_{i=1}^{n}\ln[g(\mu_i(\widehat{\beta_*}),\Theta)] \tag{5}$$

여기서, $\widehat{\beta_*}$는 기지의 값으로 이용된 회귀계수를 의미한다. 본 연구에서는 위와 같은 최적화 문제에서 전역해를 추정하기 위해서 SA(Simulated Annealing)을 사용하였다.

3. 수학적 방법에 의한 수위-유량 관계 곡선식의 분리방법

하천의 단면이 홍수 등으로 인하여 변하는 경우 한 지점으로부터 지속적으로 얻어진 수위와 유량간의 관계를 이용하여 수위-유량 관계 곡선식을 작성하기 위해서는 각 구간별로 다른 매개변수의 추정이 필요하며, 이는 곡선식의 분리에 의해 수행될 수 있다. 본 연구에서는 3개 이상의 곡선식의 분리는 고려하지 않고 2개를 최대 분리구간으로 하여 다음과 같이 나타내었다.

$$y_i = \begin{cases} f(x, \beta_1) = a_1 (x - b_1)^{c_1} & h_{min} \le x \le h_1 \\ f(x, \beta_2) = a_2 (x - b_2)^{c_2} & h_1 \le x \le h_{max} \end{cases} \tag{6}$$

여기서, 분리점 h_1은 측정된 수위-유량 관계를 분할하는 수위를 나타내며 h_{min}, h_{max}은 각각 수위의 최저값과 최고값을 나타낸다. 위와 같이 수위-유량 관계 곡선식을 2개로 분할하여 추정해야 하는 경우, Eq. (7)의 의사우도함수는 수위의 분할을 고려할 수 있는 Heaviside 함수를 이용하여 다음의 식으로 변환하여 나타낼 수 있다.

$$\ln L_{PL}(\beta, \Theta) = -n \ln [\sigma_c(\hat{\beta}_{1*}, \hat{\beta}_{2*}, \Theta)] - \sum_{i=1}^{n} \ln [g(\mu_i(\hat{\beta}_{1*}, \hat{\beta}_{2*}), \Theta)]$$

$$- w_p [H(f(h_1, \widehat{\beta_{1*}}) - f(_1 h \widehat{\beta_{2*}})] \tag{7}$$

4. 개발된 모형의 통계적 실험 및 실제 적용결과

본 연구에서는 사용된 P-LE에 의한 분산함수를 고려한 추정방법의 결과를 비교분석하기 위하여 수위-유량 관계 곡선식의 매개변수의 참값을 먼저 정하고 이로부터 산정된 유량에 6개 집합의 정규분포를 따르는 잡음(noise)을 발생시켜 최종적인 6개 셋의 수위와 유량관계를 산정한 후, 로그함수로 치환한 OLS 회귀분석(LT-LR), P-LE로 분산함수를 고려한 회귀분석(P-LE: CASE 1), P-LE로 분산함수를 고려한 회귀분석(P-LE: CASE 2)의 세 가지 방법에 따른 추정결과를 비교분석하였다. 그중 3개 셋은 매개변수의 보정에 사용되고 나머지 3개 셋은 모형의 검증에 사용되었다. 통계적 실험의 결과, 의사우도함수를 사용하면 MLE을 수행하기 위한 우도함수보다 간단한 우도함수를 구축할 수 있기 때문에 전역최적화 알고리즘을 수행하는데 있어서 효율적임을 알 수 있었으며 또한 기존 방법인 로그변환한 회귀방정식의 OLS에 의한 추정방법(LT-LR)과 분산함수의 매개변수가 커짐에 따라(즉, 유량에 의해 발생되는 잔차가 커짐에 따라) 잔차의 비등분산성을 고려할 수 있는 P-LE방법에 의한 추정결과를 비교하였는데 LT-LR방법보다 P-LE방법이 정확한 결과를 모의할 수 있음을 알 수 있었다. 이상의 통계적 실험결과의 실제 적용성을 확인하기 위해 금강유역의 천천, 동향, 송천, 옥천, 청성의 5개 지점에 적용하였다. 유량 측정자료는 한국수자원공사에서 수행한 2005년과 2006년의 유량 측정성과를 이용하였다(한국수자원공사, 2006; 2007). P-LE(CASE 1)과 Heaviside 함수를 이용하여 곡선식의 매개변수를 추정하고 또한 구간의 분리 유무에 따른 추정결과를 비교하여 분리여부를 판단하였다. 천천, 옥천의 2개 지점은 수위-유량 관계 곡선식을 분리하지 않고 사용하는 것이 타당한 것으로 판단되었으나, 천천 지점은 AIC와 BIC 중 BIC 값을 더욱 중요한 인자로 생각하여 BIC 값에 의하여 분리유무를 판단하였다. 또한 Fig. 1에는 5개 지점에서 P-LE(CASE 1)방법에 의해 추정된 변수들을 이용하고 구간의 분리가 있는 경우와 없는 경우에 대하여 유량을 산정한 후 이를 실제 측정된 유량과 비교하였다. 천천과 옥천의 경우에는 구간의 분리유무에 따른 유량의 차이가 크게 나지 않는 것을 알 수 있었으며, 나머지 청성, 동향, 송천지점에서는 구간을 분리한 경우의 수위-유량 관계 곡선식에 의한 유량 산정값이 분리하지 않은 경우의 수위-유량 관계 곡선식에 의한 유량 산정값보다 관측값과 더욱 근사한 유량을 산정할 수 있음을 알 수 있었다.

5. 결과

본 연구는 수위-유량 곡선 관계식의 매개변수를 추정하기 위하여 주로 사용되고 있는 OLS의 경우 발생되는 잔차의 비등분산성에 의한 오차의 증대문제가 발생하므로 GLS에 기반한 의사우도함수를 구축하여 분산함수의 매개변수와 회귀계수를 동시에 추정하는 P-LE 방법을 제시하였다.

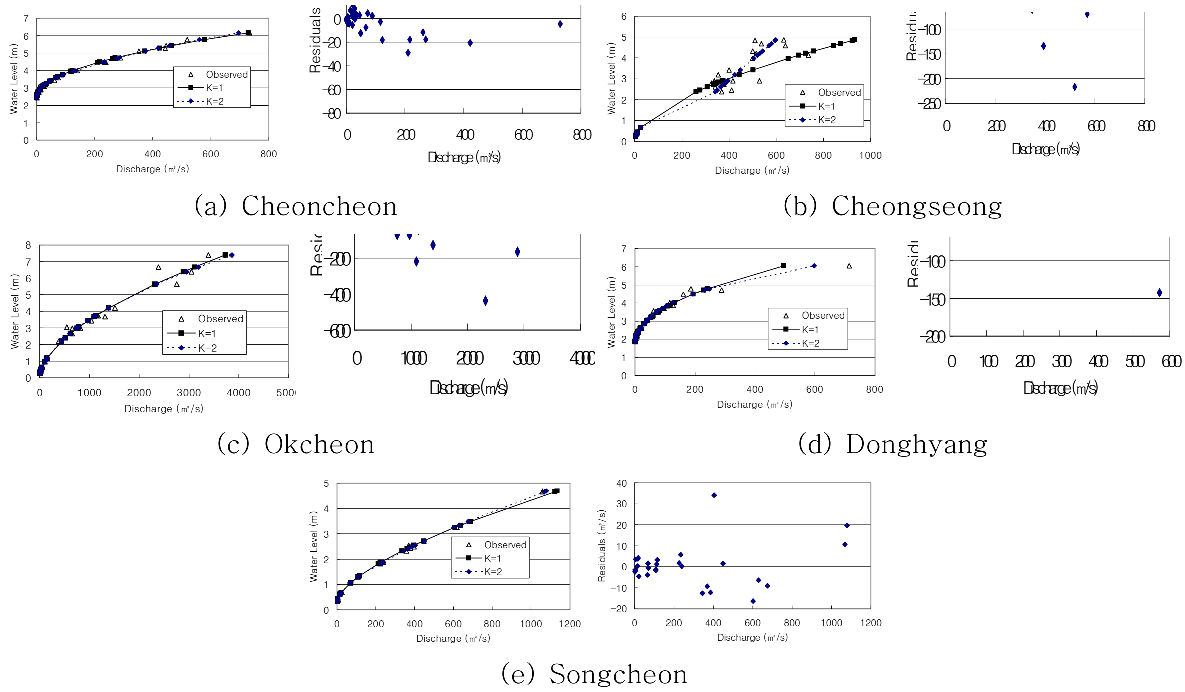

Fig. 4. Rating Curves and Residuals by P-LE Model at 5 Stations in Gum River Basin

P-LE방법으로 수위-유량 관계 곡선식의 매개변수를 추정하면 잔차의 비등분산성을 고려할 수 있으므로 추정결과에 있어서 오차를 감소시킬 수 있고, 이와 함께 유량의 증가에 따른 불확실성도 감소시킬 수 있다는 결론을 얻을 수 있었다. 특히 의사우도함수에 Heaviside 함수를 포함시켜 P-LE방법을 이용하면 보다 객관적인 측면에서 수위-유량 관계 곡선식의 분리 유무 및 분리 위치를 추정할 수 있음을 알 수 있었다. 그러므로 실무에 이용할 수 있는 수위-유량 관계 곡선식의 결정을 위한 의사결정시스템을 구성할 경우 보다 정확한 유량을 산정할 수 있을 것으로 기대된다.

감사의 글

본 연구는 21세기 프런티어 연구개발 사업인 수자원의 지속적 확보기술개발 사업단(과제번호 1-7-3)의 서울대학교 공학연구소를 통한 연구비 지원(60 %)과 서울대학교 BK21 안전하고 지속가능한 사회기반건설사업단의 연구비 지원(30 %) 및 한국수자원공사의 연구비 지원(10 %)에 의해 수행되었습니다. 연구비 지원에 심심한 감사의 뜻을 표합니다.

참고문헌

한국수자원공사 (2006). 2006 수문자료집.

한국수자원공사 (2007). 2007 수문자료집.

Akaike, H. (1974). "A new look at the statistical model identification." *IEEE Trans. Autom. Control AC-19*, pp. 716-723.

ISO (1998). "Determination of the stage-discharge relationship." *Measurement of liquid flow in open channels-Part 2*, ISO standard 1100-2, International Organization of Standards, pp. 133-153.

Lambie, J.C. (1978). "Measurement of flow-velocity-area methods." In: Hershcy RW, (Ed.), *Hydrometry: Principles and Practices,* Wiley, Chichester, Chapter 1.

Mosley, M.P., and McKerchar, A.I. (1993). "Streamflow." *Handbook of Hydrology*, Chap. 8, McGraw-Hill. N.Y.

Schwartz, G. (1979). "Estimating the dimension of a model." *Ann. Statist*. Vol. 6, pp. 461-464.

Seber, G.A.F. and Wild, C.J. (1989). *Nonlinear Regression*, John Wiley & Sons, Inc., N.Y.