

베이지안 회귀분석을 이용한 수위-유량 관계곡선의 불확실성 분석

Identification of Uncertainty in Fitting Rating Curve with Bayesian Regression

김 상 옥* / 이 길 성**

Kim, Sang Ug / Lee, Kil Seong

Abstract

This study employs Bayesian regression analysis for fitting discharge rating curves. The parameter estimates using the Bayesian regression analysis were compared to ordinary least square method using the t -distribution. In these comparisons, the mean values from the t -distribution and the Bayesian regression are not significantly different. However, the difference between upper and lower limits are remarkably reduced with the Bayesian regression. Therefore, from the point of view of uncertainty analysis, the Bayesian regression is more attractive than the conventional method based on a t -distribution because the data size at the site of interest is typically insufficient to estimate the parameters in rating curve. The merits and demerits of the two types of estimation methods are analyzed through the statistical simulation considering heteroscedasticity. The validation of the Bayesian regression is also performed using real stage-discharge data which were observed at 5 gauges on the Anyangcheon basin. Because the true parameters at 5 gauges are unknown, the quantitative accuracy of the Bayesian regression can not be assessed. However, it can be suggested that the uncertainty in rating curves at 5 gauges be reduced by Bayesian regression.

keywords : discharge rating curve, uncertainty, Bayesian regression, t -distribution, heteroscedasticity, ordinary least square

요 지

본 연구는 수위-유량 관계곡선식의 매개변수 추정을 수행하기 위하여 Bayesian 회귀분석을 적용하였다. 또한 불확실성측면에서의 효과를 탐색하기 위하여 Bayesian 회귀분석에 의한 추정치와 t 분포를 이용하여 산정한 일반 최소자승법(ordinary least square, OLS)에 의한 회귀분석의 추정치를 각각 산정하여 산정결과의 신뢰구간을 비교분석하였다. 등분산케이스의 통계적 실험결과 t 분포를 이용하여 산정된 평균 추정치와 Bayesian 회귀분석에 의한 평균 추정치는 크게 다르지 않았으나, 비등분산 케이스의 경우에는 Bayesian 회귀분석이 참값에 가까운 추정치를 산정함을 알 수 있었다. 또한 불확실성 측면에서 평가해 볼 때 신뢰구간의 상한추정치와 하한추정치의 차이는 Bayesian 회귀분석을 사용한 경우가 기존 방법을 사용한 경우보다 작은 것으로 나타났으며, 이로부터 수위-유량 관계곡선식의 매개변수를 추정하는 경우 Bayesian 회귀분석이 일반 회귀분석보다 불확실성을 표현하는데 있어서 우수하다는 결과를 얻을 수 있었다. 적용된 두 가지의 추정방법은 비등분산성을 고려한 통계적 실험을 통하여 장점과 단점이 비교되었으며, 안양천 지역의 5개 지점으로부터 얻어진 유량측정성과를 이용하여 적용성을 알아보았다. 현장 적용결과는

* 서울대학교 BK21 안전하고 지속가능한 사회기반건설 사업단 박사 후 연구원
Post-Doctor, SNU BK21 SIR Group, Seoul National University, Seoul, 151-744, Korea
(e-mail: plethor1@snu.ac.kr)

** 서울대학교 공과대학 건설·환경공학부 교수
Professor, Dept. of Civil and Environmental Engineering, Seoul National University, Seoul, 151-744, Korea

참값을 알지 못하므로 정량적 우수성은 평가할 수 없었으나, 기존에 사용되는 불확실성 산정방법보다 Bayesian 회귀 분석 불확실성은 감소시켜 나타냄을 알 수 있었다.

핵심용어 : 수위-유량 관계곡선, 불확실성, Bayesian 회귀분석, t 분포, 비등분산성, 일반최소자승법

1. 서론

직접적인 측정에 의해 얻어진 유량자료와 수위-유량 관계곡선(stage-discharge curve, rating curve)에 의해 산정된 유량자료는 모두 불확실성을 포함하고 있다. 직접적인 측정에 의해 얻어진 유량자료는 주로 측정 시 발생하는 계측기의 오차나 목측 시 발생하는 오차에 의해 불확실성이 발생하는 반면, 수위-유량 관계곡선을 이용하여 유량을 산정하는 경우에는 이들 외에도 수위-유량 관계곡선식의 매개변수를 추정하는 과정에서 발생하는 오차로 인해 많은 불확실성이 발생된다. 그러나 직접 측정에 의한 유량자료의 획득은 시간과 비용이 많이 소요되어 대부분의 국가에서는 수위를 먼저 측정하고 기 작성된 수위-유량 관계곡선을 이용하여 유량을 산정하는 일이 많으므로, 수위-유량 관계곡선식의 작성 시 발생하는 불확실성을 정확하게 산정하는 것은 유량의 정확성 및 불확실성의 규명을 위한 기본적인 과정이라 할 수 있다.

국내 또는 해외에서 흔히 사용되고 있는 수위-유량 관계곡선식 중 가장 보편적인 것은 Lambie (1978)가 제시한 Eq. (1)과 같은 비선형 관계식이다.

$$Y = A(X+B)^C \quad (1)$$

여기서, Y 는 유량(m^3/sec)이며 X 는 수위(m), A, B, C 는 수위와 유량관계를 이용하여 추정되어야 하는 매개변수들이다. 수위와 유량의 측정값은 필연적으로 오차를 포함하고 있는데, 주로 측정기기, 측정방법 등으로 인해 발생한다. 직접 측정의 경우 Herschy(1985)는 측정된 유속, 수심이 최종적인 유량에 미치는 영향을 정량화하면서 특히 단면 변화에 따른 유량의 변화를 가장 큰 오차의 원인으로 언급한 바 있으며 Sauer and Meyer(1992)는 측정장치, 평균유속의 산정 등에 있어서 발생하는 오차를 정량화하는 연구를 수행한 바 있다. 특히 위와 같은 수위-유량 관계곡선식을 이용하는 경우에는 관계식의 매개변수 추정과정에서 발생하는 오차도 중요한 요인으로 작용된다.

수위-유량 관계곡선식을 이용하여 유량을 산정하는 경우에는 관계식의 매개변수들을 추정해야 하는데, 일

반적으로 로그선형 회귀분석(log-transformed linear regression)이 이용된다. 이때 정확한 회귀계수의 추정을 위해서는 회귀모형을 구성하는 잔차(residual)의 특성을 분석해야 한다. 즉 일반 최소자승법(ordinary least square, OLS)을 사용하기 위해서는 잔차가 등분산성(homoscedasticity)을 만족해야 하며 그렇지 않은 경우는 다른 방법을 사용해야 한다. 특히 대부분의 수위-유량 관계에 있어서 유량이 증대함에 따라 오차가 증가되는 비등분산성(heteroscedasticity) 경향이 있으므로, 선형 회귀분석을 수행하는 경우 회귀잔차의 특성에 따라 회귀모형의 정확성이 크게 영향을 받을 수 있음을 주의하여 매개변수를 추정해야 한다. Seber and Wild(1989)는 잔차가 비등분산적인 경우에도 적용할 수 있는 일반화 최소자승법(generalized least squares method, GLS)을 이용한 회귀분석을 적용하면서, 가중 최소자승법(weighted least squares, WLS)은 수위-유량 관계곡선식의 매개변수를 추정하는 데 있어서 잘못된 분산함수(variance function)를 사용하는 경우 큰 오류를 범할 수 있음을 제시한 바 있다. Petersen-Øverleir(2004)는 비등분산적인 오차를 등분산적으로 변환할 수 있는 HMLE(heteroscedastic maximum likelihood estimation) 방법을 제안하고 이를 이용하여 수위-유량 관계곡선식의 매개변수를 추정한 결과를 비선형 회귀분석 결과와 비교함으로써 HMLE 방법의 우수성을 제안한 바 있다. 또한 Petersen-Øverleir and Reitan(2005)은 우도함수비를 이용하여 구축된 수위-유량 관계곡선식의 구간에 따른 분리 필요성을 객관적으로 정량화하였고 곡선의 분리 유무를 결정하는 방법을 제시한 바 있다.

일반적으로 OLS에 의한 회귀분석을 이용하여 곡선식의 매개변수를 추정한 이후에는 수위-유량 관계곡선식의 불확실성을 산정하기 위해 Herschy(1980)가 제안한 표준오차와 t 분포를 이용하여 전체오차 또는 개별 오차를 사용하는 방법을 사용하며, 국내에서도 일반적으로 이 방법을 사용하여 수위-유량 관계곡선식의 불확실성이 분석되고 있다(김원 등, 2001; 김치영 등, 2004; 서규우 등, 2005; 이길성과 이경호, 2004; 정성원 등, 2003; 차준호 등, 2001, 2002). 그러나 t 분포와 표준오차를 이용하여 불확실성을 산정하는 경우에는 오차의

정규성(normality), 선형성(linearity)등을 기본가정으로 사용하므로, 이를 만족시키지 않을 경우 산정된 불확실성의 범위가 과대 추정될 수 있으며, 또한 자료의 개수가 작은 경우 불확실성 추정에 큰 오차가 포함되어질 수 있다는 연구가 진행된 바 있다(Bickel and Doksum, 1977).

그러므로 본 연구에서는 기존 방법과는 달리 정규성이나 선형성의 가정을 사용하지 않고도 불확실성의 범위를 합리적으로 나타내는 것으로 알려진 Bayesian 회귀분석을 이용하여 수위-유량 관계 곡선식의 매개변수를 추정한 후, 유량의 불확실성을 나타내는 범위를 산정하여 기존 방법의 결과와 비교함으로써 적용방법의 장점 및 단점과 적용성을 파악하였다.

Bayesian 방법론의 초기 연구단계에서는 공액사전분포(conjugated prior distribution)을 적용하여 사후분포를 해석적으로 구하여 모수 및 불확실성을 나타내고자 하는 연구가 진행되었다(Vicens *et al.*, 1975; Wood and Rodriguez-Iturbe, 1975a, 1975b). 그러나 공액사전분포를 사용한 Bayesian 접근 방법의 수자원공학으로의 적용은 그 동안 많은 수의 추출문제와 관련된 계산능력의 부족으로 이론적인 부분만 주로 연구되고 적용범위를 넓히지 못하다가 최근 들어 발전한 계산 능력과 전체 탐색법 개념의 알고리즘의 개발로 인하여 90년대 후반부터 다시 빈도분석 분야(Madsen and Rosberg, 1997; Kuczera, 1999; O'Connell *et al.*, 2002; Reis *et al.*, 2005; Reis and Stedinger, 2005; Seidou *et al.*, 2006; Lee and Kim, 2007; 김상욱과 이길성, 2008a, 2008b)와 강우-유출모형의 보정 분야(Kuczera and Parent, 1998; Thiemann *et al.*, 2001; Vrugt *et al.*, 2003) 등에 활발히 사용되고 있다. 수위-유량 관계곡선식의 작성과 관련된 Bayesian 방법의 적용은 Moyeed and Clarke(2005)가 Bayesian MCMC를 이용하여 수행한 적이 있으나, 기존방법에 의한 불확실성의 비교평가부분에 대한 연구는 진행된 바 없다. 그러므로 본 연구에서는 먼저 Bayesian 회귀분석방법이 수위-유량 관계곡선식의 매개변수 추정과 유량의 불확실성을 합리적으로 나타내는지를 확인하기 위하여 참값을 알고 있는 수위-유량 관계곡선식으로부터 무작위 오차를 가하여 매개변수를 다시 추정하고 불확실성을 산정하는 통계적 실험을 수행하였으며, 기존 방법과의 장단점을 비교하기 위해 t 분포를 사용하여 불확실성을 산정하는 방법의 결과와 Bayesian 회귀분석 결과를 비교하였다. 또한 Bayesian 회귀분석의 적용성을 알아보기 위하여 유량자료가 비교적 잘 확보된 지점과 그렇지 못한 지점을 안양천 유역 내에서 선별하여 적용하고 그 산정결과를 평가하였다.

2. 이론적 배경

2.1 수위유량 관계곡선과 선형 회귀분석

일반적으로 수위-유량 관계곡선식은 Eq. (1)이 이용된다. 여기서 유량이 '0'인 수위를 나타내는 매개변수 B 는 도해적 방법 또는 통계적 방법을 이용하여 추정되거나 경험적으로 가정되어 사용될 수 있다. Eq. (1)의 매개변수들을 추정하기 위해 실무에서 사용되는 방법은 주로 곡선식의 양변에 로그를 취하여 선형 회귀식을 구성한 후 OLS를 적용하는 것이다. 이때 영(零)유량 수위 B 를 정확하게 측정하기는 어려우므로 다양한 기법에 의해 결정할 수 있다. USGS(United States Geological Survey)에서는 유량을 횡축으로 하고 수위를 종축으로 하는 양대수 그래프상에서 수위-유량 관측치의 형태가 직선화하도록 영유량 수위를 결정하는 방법을 제안하고 있다(Rantz, 1982). 한국수자원공사와 네델란드 Delft Hydraulics가 공동으로 개발한 수문자료 분석시스템에서는 위와 같은 방법으로 B 의 첫 번째 추정치를 구한 후, 첫 번째 추정치의 2 m 이내에서 변화시켜 가면서 오차제곱의 합을 최소화 하는 매개변수들을 추정하는 방법을 채택하고 있다(한국수자원공사, 2001). 또한 수자원분석시스템 구축기법에 관한 연구(이길성, 2001)에서 개발된 수위-유량 관계곡선 시범시스템에서는 영유량 수위인 B 를 1차원 최적화 기법중 하나인 황금비 분할법(golden section method)을 이용하여 추정한 후 비선형 최적화 기법을 사용하여 나머지 매개변수들을 결정하였다. 황금비 분할법은 비선형 최적화를 수행하기 위한 방법 중 하나로써 고대로부터 사용되는 황금비율을 이용하여 적절한 값을 선정하여 반복하여 목적함수를 산정함으로써 최적해를 추정하는 방법이다. 또 다른 방법으로 영유량 수위를 가정한 다음 회귀분석을 수행하여 가장 적은 오차를 보여주는 B 값을 최적으로 결정할 수 있다(Mosley and McKerchar, 1993).

또한 수위-유량 관계곡선식은 주어진 수위와 유량의 관계를 나타내는 데 있어서 곡선식을 분리하여 나타내야 할 필요성이 제기될 수 있다. 하천에서의 흐름은 저수위에서는 단면통제(section control)를 받을 수 있고 고수위에서는 하도통제(channel control)를 받게 되어 수위-유량 관계곡선식이 변화하게 된다. 자연하천에서 단면통제를 받는 유량관측지점은 쉽게 찾기 힘들므로 Parshall flume과 인공구조물을 설치하게 되는데, 저수위의 곡선식은 단면통제가 확실하지 않을 경우 하상 수생식물의 성장, 하상의 변동 등으로 인해 중·고수위 곡선식보다 오차가 크다고 할 수 있다. 대부분의 중·고수위 곡선식은 대수지에서 선형으로 나타낼 수 있으

며(WMO, 1994) 이러한 통제를 고려하여 수위에 따라 곡선식을 분할할 수 있는데, 수문자료 분석시스템은 유량 규모에 따라 5개의 구간으로 나누어 식을 작성할 수 있도록 구축되어져 있다(한국수자원공사, 2001). 또한 이길성(1996)은 저수위, 평수위, 홍수위에 대한 수위-유량 관계곡선을 별도로 작성하는 등 몇 가지의 관련 연구가 진행한 바 있으나, 곡선식 분할 위치를 구하는 방법은 대상하천에 대한 기술자의 판단과 눈대중(eye-fitting) 방법을 사용하였다. DeGagne *et. al.*(1996)은 수위-유량 관계곡선의 분석 및 활용을 위한 의사결정지원시스템을 개발하였다. 이 시스템은 하도의 안정 또는 불안정 상태를 통계적 기법을 사용하여 식별하고 이에 따라 수위-유량 관계곡선식을 개발할 수 있게 구성되어 있으며, 계절별로 서로 다른 관계식 유도 및 이상자료 검정 등이 가능하다. 곡선식 분할은 하천단면의 형상에 따라 사용자가 입력하게 되어 있고 곡선식 형태는 지수형태이며 OLS 방법에 의해 매개변수들이 추정된다.

본 연구에서 영유량 수위는 황금비분할법을 이용하여 추정하였다. 또한 본 연구는 불확실성 측면에서 기존 방법과 Bayesian 회귀분석 방법의 차이점, 장점 및 단점에 중심을 둔 연구이므로, 수위-유량 관계곡선식의 분리는 없는 것으로 가정하여 연구를 수행하였다. 그러나 실무적으로 수위-유량 관계선식의 분리는 최종적인 유량의 산정 값에 중요한 영향을 미칠 수 있으므로 향후에는 주관적인 방법에 의한 수위-유량 관계곡선식의 분리가 아니라 객관적이고 수학적 방법으로 수위-유량 관계곡선식의 분할을 함께 고려할 수 있는 연구가 필요하리라 판단된다. 위에서 살펴 본 바와 같이 수위-유량 관계곡선식의 매개변수들을 추정하는 일은 유량의 정확성 측면에서 매우 중요한 일이라 할 수 있으며, 주로 Eq. (2)와 같이 로그 변환된 수위-유량 관계곡선식에 대해 OLS를 이용한 선형 회귀분석을 수행함으로써 회귀계수를 추정하는 방법이 이용된다.

$$\ln Y_i = \ln A + C \ln(X_i + B) + \ln E_i$$

$$\text{or } y_i = a + cx_i + \epsilon_i \quad (2)$$

여기서, $y = \ln Y$, $a = \ln A$, $c = C$, $x = \ln(X+B)$, $\epsilon = \ln E$ 이다. Eq. (2)에서 i 는 전체 n 개의 자료 중 i 번째 값을 나타내며, Y_i 는 i 번째 관측유량, X_i 는 i 번째 관측수위이다. 또한 회귀계수 a 는 절편, c 는 기울기이다. 특히 ϵ_i 는 OLS를 이용하는 경우 평균이 0이고 일정한 분산 σ^2 을 갖는 정규분포로 가정되어 사용되어지며, 관측값 y_i 가 모집단 회귀식으로부터 ϵ_i 만큼 떨어져 있

음을 의미한다. 회귀분석에서는 표본을 이용하여 회귀계수를 추정하여 적합한 회귀식을 만들 수 있고 이 식에 의해 $x = x_i$ 에서 추정치 또는 예측치, \hat{y}_i 를 구할 수 있다. 실제 관측치와 예측치와의 차이를 잔차라 하며 다음 Eq. (3)과 같이 정의되고 회귀분석에서 잔차항 ϵ_i 는 관측할 수 없는 값이므로 다음과 같이 추정하여 사용된다.

$$\epsilon_i = y_i - \hat{y}_i, \quad i = 1, 2, \dots, n \quad (3)$$

각 오차의 제곱의 합을 최소로 하는 OLS를 이용하여 회귀식을 구하는 경우에는 정규방정식(normal equation)으로부터 회귀계수를 추정할 수 있으며 이를 회귀계수의 최소제곱추정량이라 한다. 이 때 최적의 영유량 수위를 나타내는 B 값은 회귀분석 오차가 가장 적을 때의 값이며 이는 하한값에서 상한값까지 0.01 씩을 증가시켜 회귀분석을 실시 후 가장 큰 결정계수(coefficient of determination)를 얻는 격자탐색법(grid search method)의 일종인 황금비분할법이 이용될 수 있다. OLS를 이용한 회귀분석에서 유의해야 할 점은 서론에서 언급한 바와 같이 잔차의 등분산성이 만족되는지의 여부이다. 잔차가 등분산적인 경우 잔차도(residual plot)는 주어진 범위 내에서 골고루 퍼져 있는 형상을 나타내게 되지만 비등분산적인 경우 증가 또는 감소하는 트럼펫 형태의 형상을 나타내는 것이 일반적이다. OLS를 사용한 회귀분석 결과의 잔차가 비등분산적인 경우는 추정된 회귀계수가 부정확할 가능성이 커지게 됨에 따라 결정계수가 낮아지고 회귀모형의 정확도가 감소된다.

2.2 t 분포를 이용한 불확실성의 산정

수위-유량 관계곡선의 불확실성을 결정하기 위해서는 대수변환을 통해서 얻어진 관계식으로부터 표준오차(S_e)와 t 분포를 이용하여 주어진 신뢰구간에 대한 전체오차를 산정하는 방법과 개별적인 수위에 대한 오차를 산정하는 방법을 고려할 수 있다. 표준오차는 다음의 식을 이용하여 산정할 수 있으며, 95 % 신뢰구간에서 전체오차($0.95S_e$)와 개별오차($2S_{mr}$)는 Eq. (5) 및 Eq. (6)을 이용하여 산정할 수 있다(ISO, 1998). 전체오차는 구축된 수위-유량 관계곡선식의 외삽구간까지도 포함하여 불확실성을 산정할 수 있는 방법인 반면 개별오차는 구축된 곡선 전체에 대한 오차가 아니라 측정된 점 하나하나로부터 발생하는 오차를 나타내는 것으로서 측정된 유량자료가 없는 구간에 대해서는 불확실성을 산정할 수 없는 특징이 있다.

$$S_e = \left[\frac{\sum (\ln Q - \ln Q_e)^2}{N-2} \right]^{1/2} \quad (4)$$

$$0.95S_e = \pm t_{(\alpha/2, N-p-1)} S_e \times 100\% \quad (5)$$

$$2S_{mr} = \pm t_{(\alpha/2, N-p-1)} S_e \left(\frac{1}{N} + \frac{[\ln(x-b) - \overline{\ln(x-b)}]^2}{\sum [\ln(x-b) - \overline{\ln(x-b)}]^2} \right)^{1/2} \times 100\% \quad (6)$$

위 식에서 Q 는 측정된 유량, Q_e 는 회귀식 등에 의해 산정된 예측유량, N 은 자료의 개수, α 는 유의수준, p 는 자유도를 나타내며, x 와 b 는 각각 수위와 영유량 수위이다. 위 식에서 95 % 신뢰구간의 평균에 대한 전체오차를 나타내는 $0.95S_e$ 와 개별적인 수위에 대한 오차를 나타내는 $2S_{mr}$ 을 산정하는데 있어서 t 분포를 이용하는데, Bickel and Doksum(1977)은 이와 같은 방법을 이용한 오차 하한 추정값과 상한 추정값은 자료의 개수가 감소함에 따라 과대 추정되는 경향이 있음을 제시한 바 있다. 또한 전체오차를 나타내는 $0.95S_e$ 는 측정 자료가 없는 수위-유량 관계곡선의 외삽구간에서 S_e 가 일정하다는 가정 하에 회귀식의 예측유량만 있으면 유량의 상한 및 하한 추정치를 산정할 수 있지만, 수위에 따른 개별오차를 산정하는 $2S_{mr}$ 은 측정된 자료가 있어야만 계산이 가능하므로 수위-유량 관계곡선식에서 자료가 없는 외삽구간의 경우 상한 및 하한 추정치를 산정할 수 없는 단점을 가지고 있다.

2.3 Bayesian 회귀분석과 불확실성의 산정

베이즈의 정리는 A가 먼저 발생하고 그 후에 B가 발생하는 두 개의 사건 A, B가 서로 종속적일 경우 A의 사건에 의해 B 사건의 확률이 달라진다는 것이다. 베이즈의 정리를 수식으로 나타내면 Eq. (7)과 같고, 여기서 각각의 확률 사건을 연속 확률밀도함수(probability density function)로 나타내면 베이즈의 정리는 Eq. (8)과 같이 표현될 수 있다(Sorensen and Gianola, 2002).

$$\Pr(B_j|A) = \frac{\Pr(B_j)\Pr(A|B_j)}{\sum_{j=1}^n \Pr(A|B_j)\Pr(B_j)} \quad (7)$$

$$\pi(\theta|x_1, x_2, \dots, x_n) = \frac{f(x_1|\theta) \cdots f(x_n|\theta)\pi(\theta)}{\int_{\theta} f(x_1|\theta) \cdots f(x_n|\theta)\pi(\theta)d\theta} \quad (8)$$

Eq. (8)에서 좌변의 $\pi(\theta|x_1, x_2, \dots, x_n)$ 는 사후분포 (posterior distribution), 우변 분자의 $\pi(\theta)$ 는 사전분포

(prior distribution)라 명명되며, 우변의 분모는 상수로서 주변분포(marginal distribution)이고, 우변 분자의 $f(x_1|\theta) \cdots f(x_n|\theta)$ 는 발생할 수 있는 모든 가능성을 고려한 우도함수 (likelihood function)이다. 분석하고자 하는 자료를 나타낼 수 있는 회귀모형이 결정되면 이로부터 우도함수를 유도할 수 있고, 적절한 사전분포를 부여함으로써 사후분포로부터 확률밀도함수의 매개변수를 추출하고 매개변수의 불확실성을 탐색할 수 있다. Bayesian 방법을 이용한 매개변수의 추정은 매개변수를 미지의 상수로 간주하는 것이 아니라 미지의 난수로 간주하게 됨으로써 추정의 관심이 되는 매개변수의 불확실성의 정도를 확률 모형을 이용하여 표현할 수 있게 된다. 결국 Bayesian 방법을 이용한 매개변수의 추정은 자료로부터 얻은 매개변수에 대한 정보와 매개변수에 대한 과거의 경험 또는 주관의 사전분포로 표현함으로써 보다 정확한 매개변수의 불확실성에 대한 탐색에 그 목적이 있다고 할 수 있다. Bayesian 회귀분석은 최소자승법을 회귀분석에 적용하는 과정에서 확률적 개념을 이용하는 것으로부터 시작된다. 회귀분석모형의 오차 Eq. (9)와 같이 평균 0과 분산 σ^2 을 가지는 정규분포 (normal distribution)를 따른다고 가정하면 설명변수와 회귀계수가 주어지는 경우 이에 대한 종속변수의 조건부 확률은 최소자승법의 특성과 Eq. (9)를 이용하여 Eq. (10)으로 나타낼 수 있고, 이에 대한 우도함수 (likelihood function), $L(\cdot)$ 를 구하면 Eq. (11)과 같다.

$$\epsilon_i \sim N(0, \sigma^2) \quad (9)$$

$$p(\mathbf{y} | a, c, \sigma^2) = \left(\frac{1}{2\pi\sigma^2} \right)^{1/2} \exp \left[-\frac{1}{2\sigma^2} (y_i - a - cx_i)^2 \right] \quad (10)$$

$$L(\mathbf{y} | a, c, \sigma^2) = \left(\frac{1}{2\pi\sigma^2} \right)^{n/2} \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - a - cx_i)^2 \right] \quad (11)$$

연속분포에 대한 베이즈의 정리를 주어진 변수 $\beta (= [a, c])$ 와 σ^2 에 대하여 다시 표현하면 다음과 같이 표현할 수 있으며, Eq. (12)에서 $\pi(\beta, \sigma^2)$ 가 사전분포이고 분모인 주변확률분포는 적분하여 임의의 상수로 표현될 수 있다.

$$\pi(\beta, \sigma^2 | \mathbf{y}) = \frac{L(\mathbf{y} | \beta, \sigma^2) \pi(\beta, \sigma^2)}{\int_{\sigma^2} \int_{\beta} L(\mathbf{y} | \beta, \sigma^2) \pi(\beta, \sigma^2) d\beta d\sigma^2} \quad (12)$$

Eq. (12)에서 사전분포를 적절히 선정하는 것은 Bayesian 방법을 이용하여 회귀분석을 수행하는데 있어서 가장 중요한 부분이라 할 수 있다. 사전분포는 크게 자료에 기반한 사전분포와 자료에 기반하지 않은 사전분포로 구분할 수 있다. 자료에 기반한 사전분포를 구성하기 위해서는 인근 유량자료로부터 분석하고자 하는 지점의 사전분포를 유도할 수 있으나, 본 연구의 회귀분석에 있어서는 회귀계수에 대한 인근 자료를 이용할 수가 없어 자료에 기반한 사전분포를 사용하는 것이 불가능하므로 Sorensen and Gianola(2002)가 제안한 다음과 같은 균일분포(uniform distribution)를 사용하였다. 이와 같은 균일분포는 회귀계수에 대한 사전정보를 전혀 알 수 없다는 것을 반영한 것으로서 사전분포가 모형의 분산에만 관련되어짐을 나타낸 것이다.

$$\pi(\beta, \sigma^2) = \frac{1}{\sigma^2} \quad (13)$$

Eq. (11)의 우도함수와 Eq. (13)의 사전분포를 Eq. (12)에 대입하면, 주변분포는 상수로 적분되므로 다음과 같이 우도함수와 사전분포의 곱이 사후분포에 비례한다는 식을 얻을 수 있다.

$$\pi(\beta, \sigma^2|y) \propto (\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - a - cx_i)^2\right] \sigma^{-2} \quad (14)$$

Eq. (14)에 포함되어 있는 항 $\sum_{i=1}^n (y_i - a - cx_i)^2$ 은 다음의 과정을 통하여 Eq. (15a)와 같이 정리되고 이를 다시 행렬표기로 나타내면 Eq. (15b)와 같다.

$$\sum_{i=1}^n (y_i - a - cx_i)^2 = (n-2)s^2 + n(a - \hat{a})^2 + (c - \hat{c})^2 \sum_{i=1}^n x_i^2 \quad (15a)$$

$$+ 2(a - \hat{a})(c - \hat{c}) \sum_{i=1}^n x_i = (n-2)s^2 + (\beta - \hat{\beta})^T (X^T X) (\beta - \hat{\beta}) \quad (15b)$$

단, 여기서 $X^T X = \begin{pmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{pmatrix}$ 이고,

$$s = \sum_{i=1}^n (x_i - \bar{x})^2 \text{ 이다.}$$

Eq. (15b)를 Eq. (14)에 대입하면 Eq. (16)으로 정리될 수 있으며, Eq. (16)이 균일분포가 적용된 사후분포라 할 수 있다. Eq. (16)의 우변의 전항을 $p(\sigma^2|y)$ 로 표기하고 후항을 $p(\beta, \sigma^2|y)$ 로 간략화하면 최종적인 사후분포는 Eq. (17)과 같이 정리될 수 있다.

$$\pi(\beta, \sigma^2|y) \propto (\sigma^2)^{-n/2} \exp\left[-\frac{(n-2)s^2}{2\sigma^2}\right] \cdot \sigma^{-2} \exp\left[-\frac{1}{2\sigma^2} (\beta - \hat{\beta})^T (X^T X) (\beta - \hat{\beta})\right] \quad (16)$$

$$\pi(\beta, \sigma^2|y) \propto p(\sigma^2|y) \cdot p(\beta, \sigma^2|y) \quad (17)$$

또한 매개변수 α, γ 를 가지는 역감마분포(inverse gamma distribution)는 분포에 있는 상수에 비례하는 것으로 간주하여 다음의 식으로 나타낼 수 있다.

$$f(x) \propto x^{-(\alpha+1)} \exp\left[-\left(\frac{\gamma}{x}\right)\right], \quad 0 < x < \infty \quad (18)$$

Eq. (18)에서 $x = \sigma^2$, $\alpha = (n-2)/2$, $\gamma = (n-2)s^2/2$ 로 두면, Eq. (16)의 우변의 전항은 Eq. (19)와 같이 역감마분포를 따르는 것을 알 수 있으며, 후항은 Eq. (20)과 같이 정규분포를 따르는 것을 알 수 있다.

$$\sigma^2|y \sim IG\left\{\frac{n-2}{2}, \frac{(n-2)s^2}{2}\right\} \quad (19)$$

$$\beta, \sigma^2|y \sim N(\hat{\beta}, \sigma^2(X^T X)^{-1}) \quad (20)$$

그러므로 구하고자 하는 회귀계수, β 를 추정하기 위해서는 먼저 $\hat{\beta}$, $(X^T X)^{-1}$, s^2 을 자료로부터 산정한 후 이 값들을 이용하여 Eq. (19)로부터 σ^2 을 생성하고, 생성된 σ^2 을 이용하여 Eq. (20)으로부터 최종적으로 β 를 생성시킴으로서 회귀계수를 얻을 수 있으며 이로부터 회귀계수의 평균추정치와 원하는 유의수준에서의 신뢰구간을 산정할 수 있다.

3. 통계적 실험을 통한 비교분석

일반적으로 OLS 방법은 회귀잔차의 등분산가정이 만족되는 경우와 만족되지 않는 경우를 나누어 회귀모형의 매개변수 추정방법에 따른 결과를 비교할 필요가 있다. 이론적으로 OLS를 사용한 회귀분석과 Bayesian 회귀분석의 경우 등분산가정만 만족되면 각 방법의 평균에 대한 유량 산정결과는 거의 유사한 결과로 산정되

어야 하며, 그렇지 않을 경우에는 각 방법의 평균에 대한 유량 산정결과부터 다른 결과를 보여야 한다. 또한 불확실성 측면에서 95 % 신뢰구간에 해당되는 하위 2.5 % 추정치와 상위 97.5 % 추정치에 대한 유량 산정결과는 각 방법의 결과를 통해 비교되어질 필요가 있으며, 이를 이용하여 OLS를 이용한 회귀분석과 Bayesian 회귀분석의 장점을 단점을 파악할 수 있다. 그러므로 본 연구에서는 실제 유량자료를 이용한 수위-유량 관계곡선식의 추정 이전에 참값을 알고 있는 매개변수의 값을 이용하여 각 방법의 비교분석을 위한 통계적 실험을 수행하였다. 참값이 정해진 수위-유량 관계곡선식은 Eq. (21)과 같다. 단, 영유량 수위는 황금분할법에 의해 사전에 -0.2로 추정하였으며, 수위-유량 관계곡선식의 분리는 없는 것으로 가정하였다. 주어진 수위 내에서 50개의 수위를 균일분포로부터 무작위적으로 발생시킨 후 수위-유량 관계식으로부터 유량을 산정하였다. 또한 등분산성의 경우와 비등분산성의 경우에 해당되는 유량을 생성하기 위해서 등분산케이스의 경우는 분산이 0.4로

일정한 정규분포로부터 50개의 난수를 발생시켰으며, 비등분산케이스의 경우는 분산이 0.1부터 0.4까지 0.06만큼 변화하도록 하여 정규분포로부터 50개의 난수를 발생시켜 각각의 발생된 오차를 유량에 포함하여 오차가 등분산 또는 비등분산적인 유량을 각각 발생하였다. 즉, 비등분산케이스의 분산은 유량에 대한 함수로써 최소 0.1, 최대 0.4의 범위에서 기울기가 0.06인 선형적인 관계를 가지도록 가정하여 사용하였다.

$$\tilde{y} = 300.00 \times (x - 0.20)^{1.50} \quad (21)$$

$$0.20 \leq x \leq 6.00$$

Fig. 1의 (a)와 (b)에는 등분산케이스의 경우 발생된 유량과 수위-유량 관계곡선, 그리고 OLS에 의한 회귀분석의 잔차를 나타낸 것이다. 또한 Fig. 2의 (a)와 (b)는 비등분산케이스의 경우 발생된 유량과 수위-유량 관계곡선, 그리고 OLS에 의한 회귀분석의 잔차를 나타낸 것이다. 잔차도에서 보면 등분산케이스의 경우 잔차는

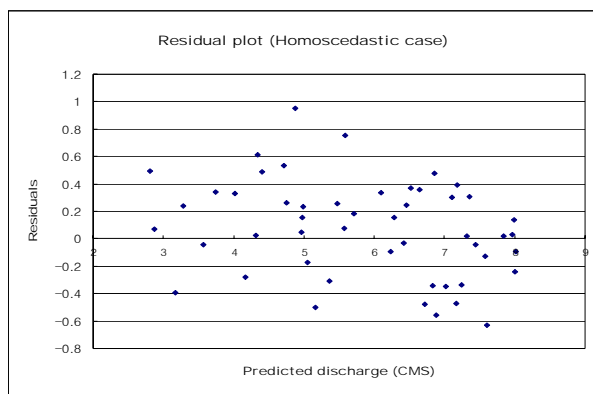
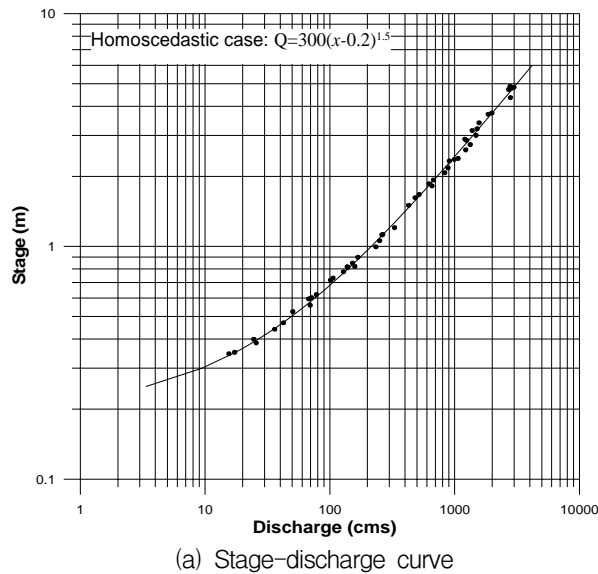


Fig. 1. Generated Discharge: Homoscedastic case

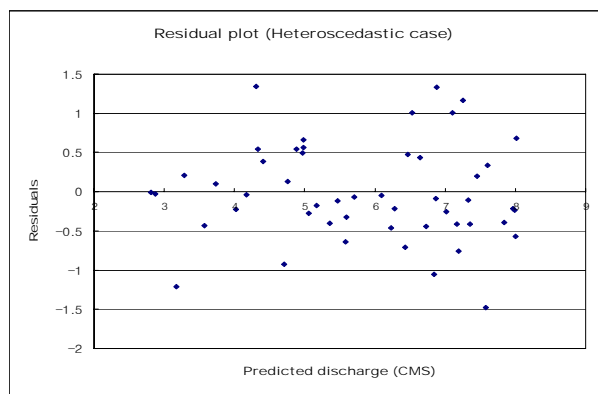
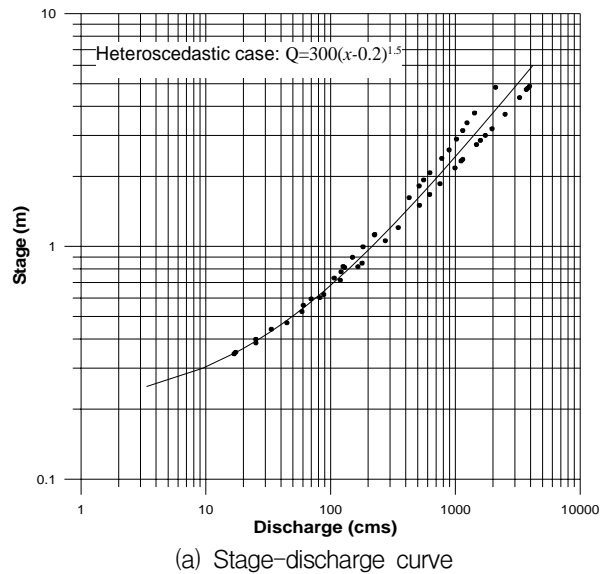


Fig. 2. Generated Discharge: Heteroscedastic case

Table 1. The Results of Regression Coefficients (homocedastic case)

Coeff.	OLS (2.5%) ①	OLS (50%) ②	OLS (97.5%) ③	diff. (③-①)	Bayesian (2.5%) ④	Bayesian (mean) ⑤	Bayesian (97.5%) ⑥	diff. (⑥-④)	diff. (⑤-②)
<i>a</i>	246.40	275.77	308.64	62.24	268.82	275.83	283.06	14.24	0.06
<i>c</i>	1.49	1.60	1.71	0.22	1.57	1.60	1.63	0.06	0.00

Table 2. The Results of Regression Coefficients (heterocedastic case)

Coeff.	OLS (2.5%) ①	OLS (50%) ②	OLS (97.5%) ③	diff. (③-①)	Bayesian (2.5%) ④	Bayesian (mean) ⑤	Bayesian (97.5%) ⑥	diff. (⑥-④)	diff. (⑤-②)
<i>a</i>	248.93	306.06	376.30	127.37	291.57	305.92	320.82	29.25	-0.14
<i>c</i>	1.37	1.58	1.84	0.47	1.48	1.53	1.53	0.05	-0.05

Table 3. Comparative Results to the True Parameter (heterocedastic case)

Coeff.	True ①	OLS (50%) ②	Bayesian (Mean) ③	difference	
				(①-②)	(①-③)
<i>a</i>	300.00	306.06	305.92	-6.06	-5.92
<i>c</i>	1.50	1.58	1.53	-0.08	-0.03

골고루 퍼져 있는데 반해서 비등분산케이스의 경우에는 몇 점을 제외하고는 유량이 증가할수록 잔차가 증가되는 트럼펫 형상을 나타내는 것을 볼 수 있다.

본 연구에서는 영유량 수위는 추정대상에서 제외하여 -0.2로 고정하고, Eq. (1)의 *a*와 *c*만을 추정대상으로 하여 OLS를 이용한 회귀분석과 Bayesian 회귀분석을 각각 수행하였다. OLS 회귀분석은 SPSS 통계 패키지를 이용하여 수행하였고, Bayesian 회귀분석은 Matlab과 통계프로그램의 일종인 R(Venables and Smith, 2008)을 이용하였다. 특히 Bayesian 회귀분석을 위해서는 각 추정대상 매개변수 당 10,000개씩을 샘플링하였으며, 초기 불안정한 추정값을 제외하기 위하여 처음 1,000개의 샘플링된 추정값은 제외하고 Bayesian 평균값 및 상하한값을 추정하였다.

등분산케이스의 경우에는 회귀계수를 이용하여 각 수위에서 유량을 산정하여 계산한 측정유량과 모의유량 사이의 결정계수가 OLS의 경우 0.9576, Bayesian 회귀분석의 경우 0.9654로 높은 결정계수가 계산되었다. 비등분산케이스의 경우에는 OLS의 경우 0.8592, Bayesian 회귀분석의 경우 0.9245로 등분산케이스에 비해 OLS의 경우 낮은 결정계수가 산정되었다.

Table 1과 Table 2에는 등분산케이스와 비등분산케이스에 대한 OLS를 이용한 회귀분석과 Bayesian 회귀분석에 의한 회귀계수의 95% 신뢰구간에 대한 추정결

과를 비교하였다. 등분산케이스의 경우 두 방법에 의한 회귀계수의 추정결과 *a*는 다소 과소추정되었고, *c*는 다소 과대추정된 것을 알 수 있다. 이는 참값에 포함된 오차가 회귀분석의 추정과정에서 발생시킨 오차로 판단된다. 또한 Table 1을 보면, OLS의 추정결과와 Bayesian 회귀분석의 평균에서의 추정결과는 거의 유사하게 나타나는 것을 알 수 있었으나, 불확실성을 나타내는 상한값과 하한값의 추정오차가 Bayesian 회귀분석의 경우 OLS보다 작게 산정되어 불확실성 측면에서 Bayesian 회귀분석이 OLS에 의한 회귀분석보다 우수한 결과를 보임을 알 수 있었다. 비등분산케이스의 경우에는 평균에 있어서 등분산케이스보다는 큰 차이를 나타내는 것으로 산정되었다. 그러나 OLS에 의한 추정결과는 특히 불확실성 측면에서 등분산케이스 보다 훨씬 증가된 범위를 보여준 반면 Bayesian 회귀분석의 결과는 오차의 특성이 비등분산적인 경우에도 불확실성이 크게 증가되지 않는 것을 확인할 수 있었다.

또한 Table 3에는 비등분산케이스의 경우 OLS 50% 추정값과 Bayesian 평균의 추정값을 통계적 실험에서 사용된 참값인 *a*=300.00, *c*=1.50과 비교하였다. 비교결과 OLS를 이용한 경우보다 Bayesian 회귀분석을 이용한 경우가 보다 참값에 근사한 값을 나타내는 것을 알 수 있었으며, 특히 최종적인 유량산정에 큰 영향을 미치는 매개변수인 *c*의 값은 두 기법에 의한 차이가 0.05

로써 이에 의한 유량산정에 있어서의 오차가 크게 발생됨을 알 수 있었다.

이와 같은 결과로부터 등분산케이스의 경우에는 OLS에 의한 회귀분석과 Bayesian 회귀분석은 평균에 의한 확정적인 값만을 이용할 경우 큰 차이를 보이지 않아 분석절차가 비교적 간단한 OLS에 의한 회귀분석을 수행하는 것이 유리할 수 있다는 결론을 얻을 수 있었다. 그러나 불확실성 측면에서 상한과 하한을 함께 추정해야 하는 경우에는 Bayesian 회귀분석에 의한 불확실성의 산정결과가 OLS에 비해 훨씬 감소되어 표현됨을 알 수 있었다. 비등분산케이스의 경우에는 참값과의 비교결과 OLS에 의한 추정값보다 Bayesian 회귀분석에 의한 추정값이 다소 참값에 가까운 결과를 제시하는 것을 알 수 있었으며, 불확실성의 감소에 있어서도 Bayesian 회귀분석이 훨씬 감소된 불확실성을 나타낼 수 있어 유리하다는 결론을 얻을 수 있었다. 회귀계수의 불확실성 측면에서 얻은 결과는 수위-유량 관계곡선에서 나타나는 불확실성에 대한 결과라기보다는 회귀분석 과정에서 발생하는 OLS 회귀분석과 Bayesian 회귀분석의 차이에 대한 비교평가라고 할 수 있다. 그러므로 수위-유량 관계곡선에서 나타나는 불확실성을 비교 평가하기 위해서는 OLS의 평균결과로부터 유량을 산정한 후 전체오차인 $0.95S_e$ 와 개별적인 오차인 $2S_{mr}$ 을 산정하고 이 결과를 Bayesian 회귀분석결과로부터 얻은 유량과 비교할 필요가 있다.

Fig. 3에는 등분산케이스와 비등분산케이스에 대한 수위-유량 관계곡선의 평균에 대한 유량과 이에 따른 불확실성을 산정한 결과를 나타내었으며, 측정유량이 존재하는 수위까지의 S_e 가 변하지 않는 것을 가정하여

수위-유량 관계곡선의 외삽구간까지 $0.95S_e$ 를 산정하였다. 등분산케이스의 경우 평균에 대한 OLS와 Bayesian 회귀분석 결과는 거의 일치하는 곡선으로 추정되었고 $2S_{mr}$ 에 의한 불확실성 산정결과는 Bayesian 회귀분석에 의한 불확실성 산정 결과보다 약간 증가되어 산정되는 것을 알 수 있었다. 그러나 개별오차를 나타내는 $2S_{mr}$ 에 의한 불확실성 산정결과는 측정유량이 존재하는 구간까지만 불확실성을 산정할 수 있기 때문에 불확실성이 크게 증가되어 지는 유량이 큰 구간에서의 비교는 불가능하였다. 또한 $0.95S_e$ 에 의한 불확실성의 산정결과는 유량이 증대됨에 따라 Bayesian 회귀분석에 의한 불확실성보다 훨씬 증가된 불확실성을 나타내는 것을 알 수 있었다. 비등분산케이스의 경우에도 평균에 대한 결과부터 다소 차이가 나는 것을 알 수 있었는데, 특히 유량이 증대됨에 따라 OLS의 결과는 유량을 과대추정하고 있음을 알 수 있었다. 또한 불확실성 측면에서 $2S_{mr}$ 과 $0.95S_e$ 에 의한 산정결과 모두 등분산케이스에 비하여 불확실성을 크게 증가시켜 표현하였다.

4. 현장 적용을 위한 비교 분석

이상의 결과에 대한 실제 적용성을 확인하기 위하여 안양천 유역의 안양, 고척교, 시흥, 학의천 출구, 목감천 출구의 5개 지점에서의 유량자료를 이용하여 개발된 모형을 적용하였다. 선정된 지점의 유량 측정성과는 한국수자원공사(2000, 2001, 2002, 2003, 2004, 2005, 2006)와 이길성(2007)의 유량측정성과 자료를 사용하였다. Table 4에는 각 지점의 자료의 개수 등에 대한 특징을

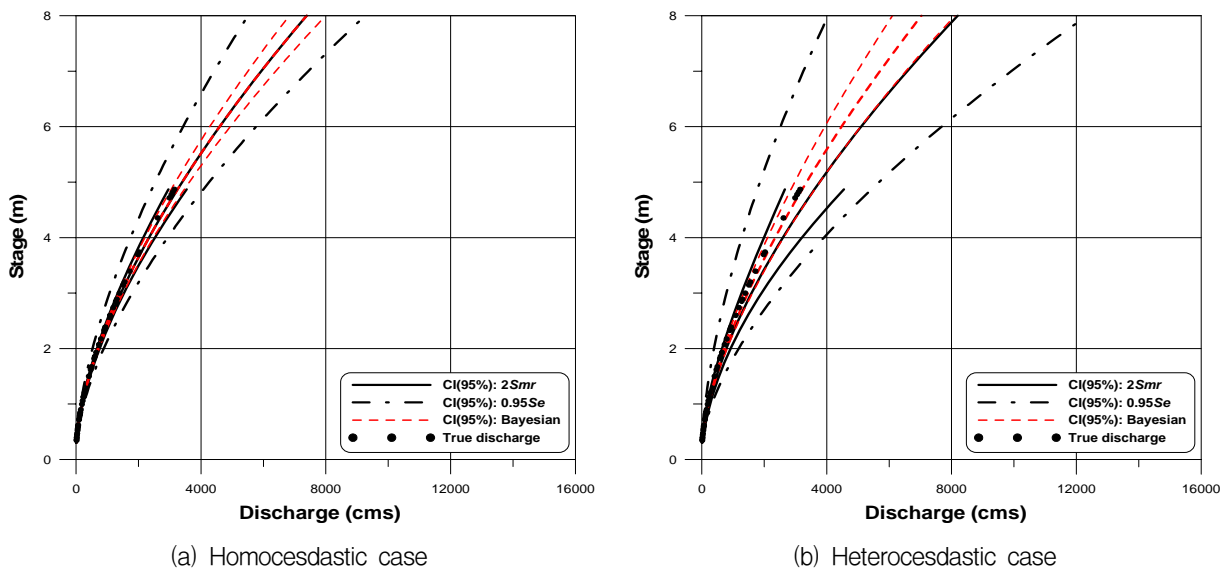


Fig. 3. Rating Curves Including Uncertainty

Table 4. The Characteristics of Selected Gauges

Name of gauge	Data year	Number of pairs (stage-discharge)
Anyang	2000 - 2006	154
Gocheoggyo	2004 - 2006	99
Siheung	2003 - 2005	49
Hakueicheon	2006	27
Mokgamcheon	2006	13

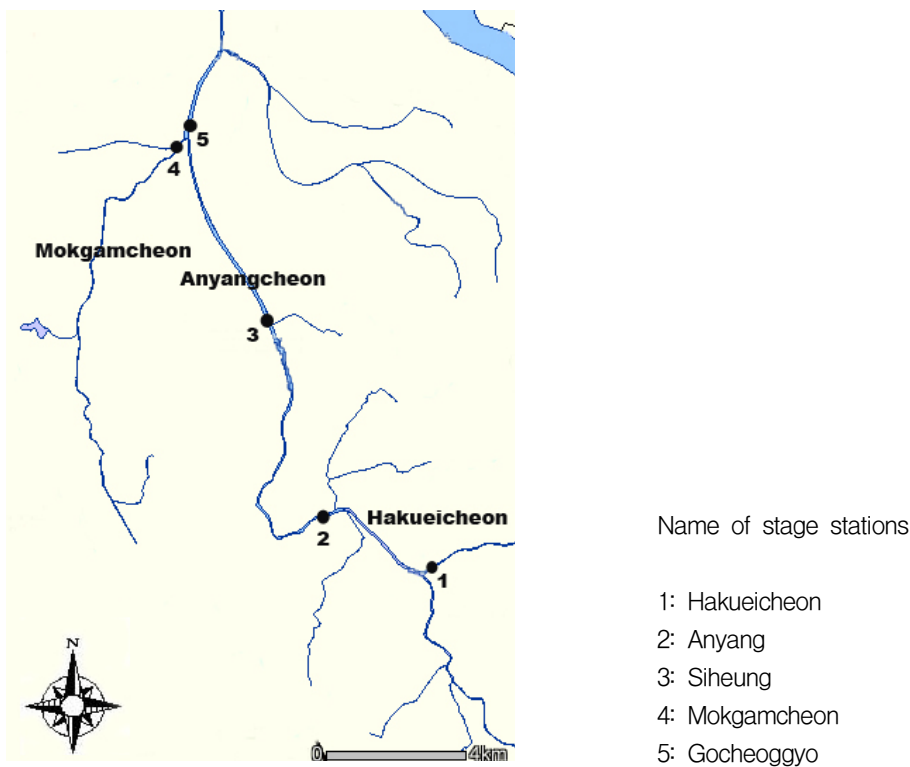


Fig. 4. Anyangcheon Basin and 5 Stage Stations

나타내었으며, 자료의 개수에 따른 적용결과를 분석하기 위하여 자료의 개수에 대한 차이가 있는 지점을 선정하여 개발된 모형을 적용하였다. Fig. 4는 안양천 유역도와 선정된 지점을 표시한 것이다.

잔차도를 비교분석 전에 알아봄으로써 등분산특성인 지점과 비등분산특성 지점을 구분하여 적용결과를 비교할 수 있고 이에 따른 결과를 얻을 수 있을 것이므로, OLS 회귀분석과 Bayesian 회귀분석의 비교를 위한 절차 이전에 선정된 각 기점으로부터 얻어진 유량측정성과 자료의 잔차특성을 알아보기 위하여 잔차도를 작성하여 Fig. 5에 나타내었다. 얻어진 잔차도를 보면 안양지점의 경우 뚜렷한 트랩렛 형태의 잔차도로서 비등분산성이 존재하는 것을 알 수 있으며, 안양 외에도 고척교, 시흥 지점 등 자료의 개수가 많은 지점에서는 비등분산성이 일정 정도 존재하는 것을 알 수 있었다. 또한

학의천과 목감천은 잔차의 특성이 등분산적이라고 할 수 있는 데, 이 지점에서도 자료의 개수가 증가한다면 잔차의 특성도 비등분산적으로 바뀌어 질 것으로 예측할 수 있어 대부분의 수위-유량 관계곡선에서는 유량이 증가함에 따라 오차가 증가되는 형태를 가지는 비등분산적인 특성을 가지는 것을 확인할 수 있었다.

적용과정에서도 통계적 실험과 같은 조건을 사용하여 OLS 회귀분석과 Bayesian 회귀분석을 수행하였다. 즉, 수위-유량 관계곡선식의 분리는 실무적으로 필히 고려해야 하는 문제이지만 본 연구에서는 두 가지 방법의 불확실성 측면의 비교평가에 초점을 맞추어 연구를 수행하였으므로 수위-유량 관계곡선식을 분리하지 않고 한 개의 식으로 추정하였다. 또한 Bayesian 회귀분석을 위한 샘플링 개수 및 초기 추정치의 제거수도 통계실험과 마찬가지로 각각 10,000개와 1,000개로 통일하

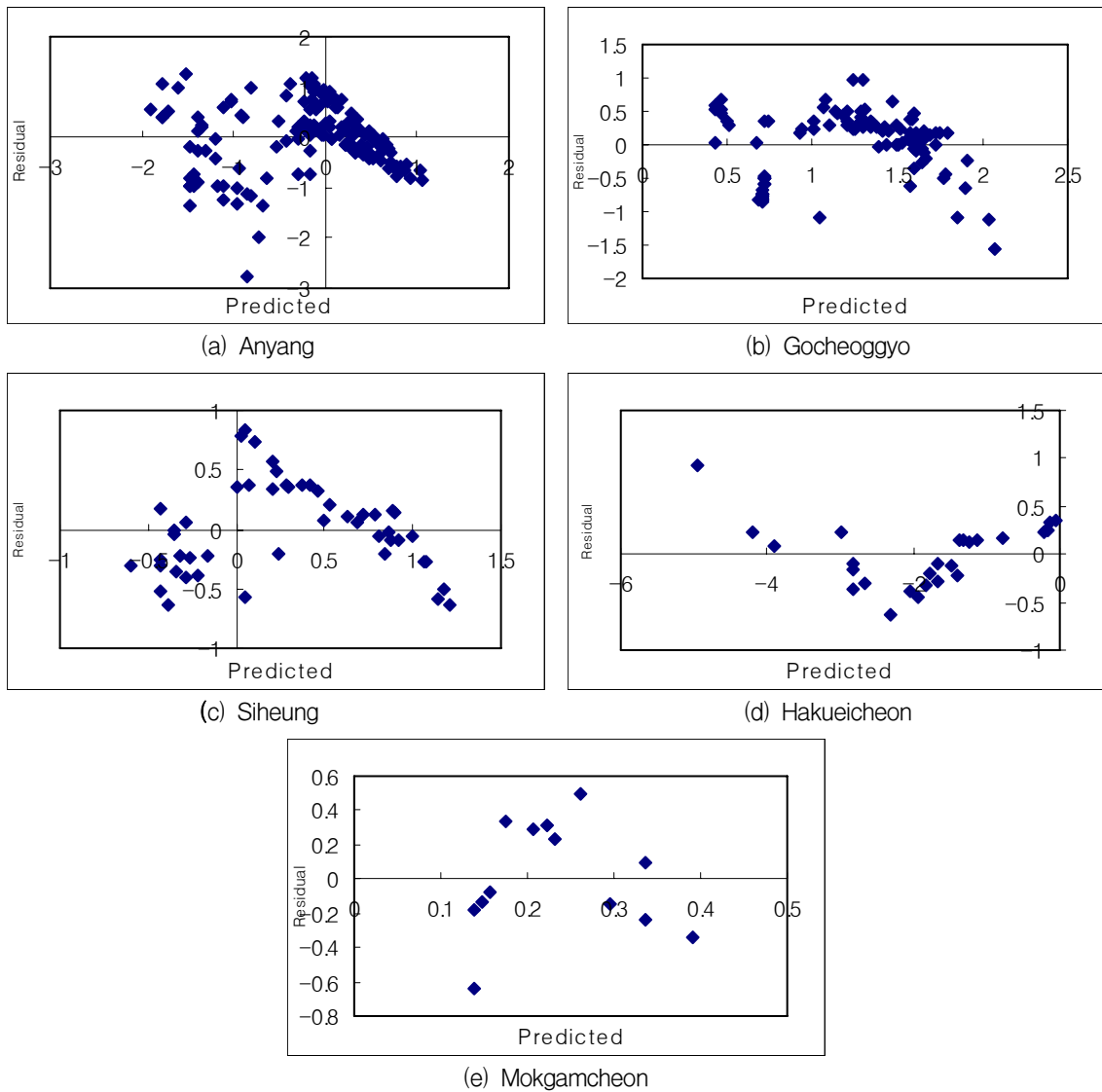


Fig. 5. Residual Plots for Predicted Variable at 5 Gauges

여 적용을 수행하였다. Table 5에는 각 지점의 수위-유량 관계곡선식을 추정한 추정결과이며, 여기서 영유량 수위는 통계적 실험과 같은 방법인 황금비분할법을 이용하여 미리 추정하여 사용하고, 두 가지 추정방법으로는 a 와 c 만 추정하였다.

Table 5는 통계적 실험결과와 유사한 결과를 나타내는 것을 알 수 있다. 즉 OLS와 Bayesian 회귀분석의 평균에 대한 결과는 거의 유사하게 추정되어지지만, 불확실성을 나타내는 상한과 하한 추정치는 OLS 회귀분석보다 Bayesian 회귀분석이 보다 감소되어 결과를 산정하게 됨을 알 수 있다. 또한 비등분산적 특성이 강했던 안양, 시흥 등의 지점에서 특히 Bayesian 회귀분석에 의한 불확실성의 산정결과가 많이 감소된 범위를 보여주는 것을 알 수 있었다. 그러나 참값을 모르는 현장 자료를 이용하는 데 있어서 회귀계수의 추정결과를 이용한 두 방법상의 결과의 비교는 신뢰구간의 감소만으

로는 OLS기법과 Bayesian 회귀분석 중 어떤 기법이 우수하다고 결론내리기 어려울 것으로 판단된다.

Fig. 6은 추정된 결과를 이용하여 각 지점에 대한 수위-유량 관계곡선에 대한 불확실성을 나타내는 결과이며, 회귀분석결과를 이용한 외삽으로 관측자료 보다 큰 유량에 대한 자료에 대해서도 수위-유량 관계곡선을 연장하였다. 불확실성의 산정은 통계적 실험의 결과 표시와 마찬가지로 $2S_{mr}$ 과 $0.95S_e$ 에 의한 불확실성 산정결과와 Bayesian 회귀분석에 의한 불확실성의 산정결과를 함께 도시하였다. 도시결과를 보면 비등분산성이 강하게 나타났던 안양지점을 제외한 모든 지점에서 OLS와 Bayesian 회귀분석의 평균에 대한 유량은 거의 유사하게 산정됨을 알 수 있었으며, 불확실성 측면에서는 Bayesian 회귀분석에 의한 산정결과가 가장 불확실성을 감소시켜 나타냄을 알 수 있었다.

불확실성 측면에서 대부분의 실측 자료들은 Bayesian

Table 5. The Results of Regression Coefficients (heteroscedastic case)

Gauge name		OLS (2.5%)	OLS (mean)	OLS (97.5%)	diff. (③-①)	Bayesian (2.5%)	Bayesian (mean)	Bayesian (97.5%)	diff. (⑥-④)	diff. (⑤-②)
		①	②	③		④	⑤	⑥		
I	a	36.412	41.303	46.852	10.440	39.453	41.309	43.307	3.854	0.006
	c	2.883	3.051	3.219	0.336	3.012	3.052	3.091	0.079	0.001
II	a	0.424	0.600	0.850	0.426	0.554	0.600	0.652	0.0098	0.000
	c	3.546	3.813	4.0880	0.534	3.751	3.814	3.875	0.124	0.001
III	a	18.778	21.665	24.997	6.220	20.937	21.651	22.357	1.420	-0.014
	c	2.535	2.769	3.004	0.469	2.713	2.769	2.823	0.109	0.000
IV	a	8.033	10.578	13.929	5.896	8.929	10.571	11.248	1.319	-0.008
	c	0.776	0.900	1.023	0.247	0.872	0.899	0.927	0.055	-0.001
V	a	0.056	0.118	0.248	0.192	0.101	0.118	0.138	0.037	0.000
	c	6.919	9.914	12.908	5.989	9.282	9.909	10.528	1.245	-0.005

* I : Anyang($b=-0.15$), II : Gocheogyo($b=-2.55$), III : Siheung($b=-0.55$)
 IV : Hakueicheon($b=-0.01$), V : Mokgamcheon($b=-0.02$)

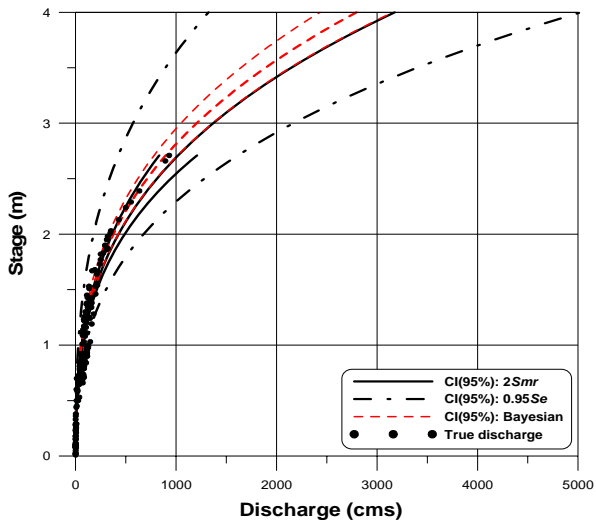
회귀분석의 하한과 상한 추정값 사이에 존재함으로써 Bayesian 회귀분석 결과가 실측 자료가 가지는 불확실성을 가장 합리적이고도 감소시켜 나타냄을 확인할 수 있었다. 그러나 본 연구에서 다루지 않은 수위-유량 관계곡선의 분리에 대한 문제를 고려하여 각 구간별 유량을 산정하고 이에 대한 불확실성을 산정하면, $2S_{mm}$ 과 $0.95S_c$ 에 의한 불확실성이 어느 정도 감소되어 질 것으로 예측되므로, 수위-유량 관계곡선의 분리와 이에 따른 불확실성의 감소 영향을 알아보는 것도 향후 중요한 연구과제라 할 수 있을 것이며, 이 때 Bayesian 회귀분석의 복잡한 절차를 시스템화 하여 일괄적으로 처리하기 위한 응용프로그램의 개발도 중요한 부분이라 할 수 있다.

5. 결론 및 향후 과제

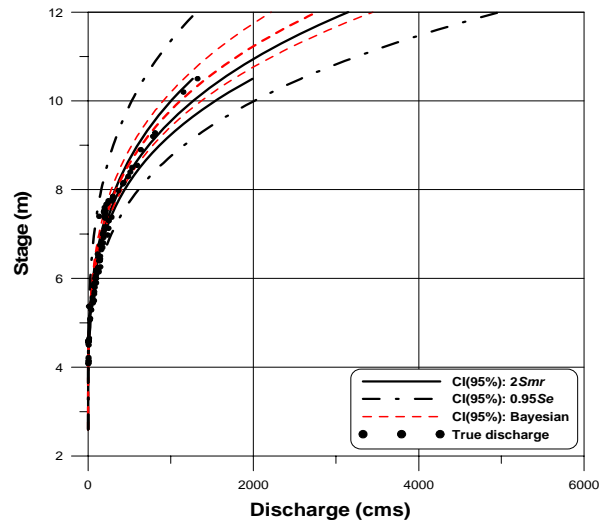
본 연구는 수위-유량 관계곡선식의 불확실성을 나타내기 위한 여러 가지 방법을 비교 평가함으로써 합리적으로 불확실성을 표현할 수 있는 방법을 제시하였다. 수위-유량 관계곡선식의 매개변수를 회귀분석으로 추정하는 데 있어서 가장 중요하게 고려해야 하는 회귀잔차의 등분산성과 비등분산성에 따른 OLS 추정결과를 통계적 실험을 통하여 분석해 보았으며, 비등분산성이 강한 경우에는 OLS 추정결과에 있어서 결정계수가 낮아지고, 불확실성도 과대 추정된다는 결론을 얻을 수 있었다. 이에 대한 대안으로 매개변수를 추정하는 데 있어서 선형성, 정규성 등의 가정이 필요없고 자료의 개수에 크게 영향을 받지 않는 Bayesian 회귀분석을 이

용하여 수위-유량 관계곡선식의 매개변수를 추정하고 그에 따른 불확실성을 산정하였다. 통계적 실험 결과 다음과 같은 결론을 얻을 수 있었다.

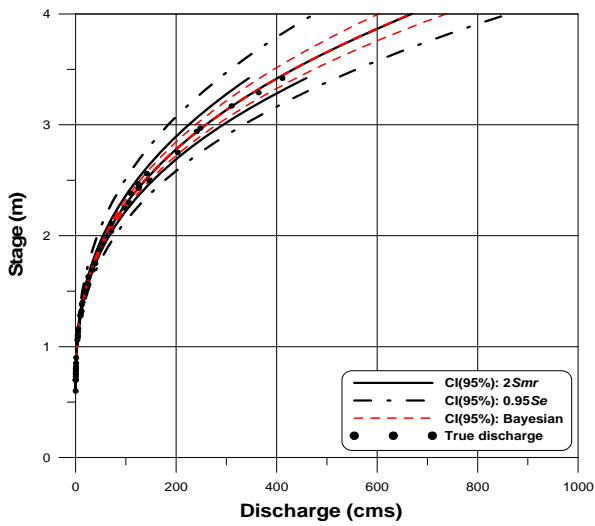
- 1) 회귀분석에 있어서 잔차가 등분산적인 특성을 가지는 경우에는 OLS에 의한 평균추정값과 Bayesian 회귀분석의 평균 추정값은 거의 동일한 결과를 보이므로, 확정적인 결론만을 필요로 하고 잔차가 등분산적인 경우에는 OLS에 의한 회귀분석을 사용하는 것이 분석을 위한 노력을 고려할 때 유리하다고 판단할 수 있었다.
- 2) 회귀분석에 있어서 잔차가 등분산적이라도 불확실성 측면의 결과를 이용하는 데 있어서는 OLS 회귀분석보다 Bayesian 회귀분석이 불확실성을 감소시켜 표현해 주기 때문에 유리하다고 할 수 있다.
- 3) 잔차의 특성이 비등분산적인 경우 OLS에 의한 회귀분석 결과는 잔차의 영향으로 인해 결정계수가 감소되고 최종적인 유량 산정결과를 과대추정할 수 있으므로 Bayesian 회귀분석을 사용하는 것이 확정적인 결론만을 필요로 하거나 불확실성 측면의 결과를 필요로 하는 경우 모두 합리적일 수 있다. 단, 확정적인 결론만을 필요로 하는 경우 통계적 측면에서 가중최소제곱법(WLS)이나 일반화최소제곱법(GLS)를 사용하면 OLS 적용 시 발생하는 오차를 감소시킬 수 있다. 특히 비등분산케이스의 경우 통계적 실험에서 사용된 참값과의 비교 결과 Bayesian 회귀분석이 OLS보다 참값에 보다 근사한 변수를 추정하는 것을 알 수 있었다.



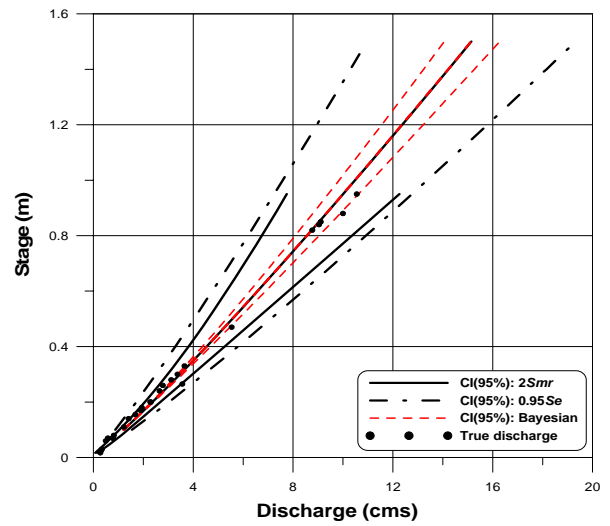
(a) Anyang



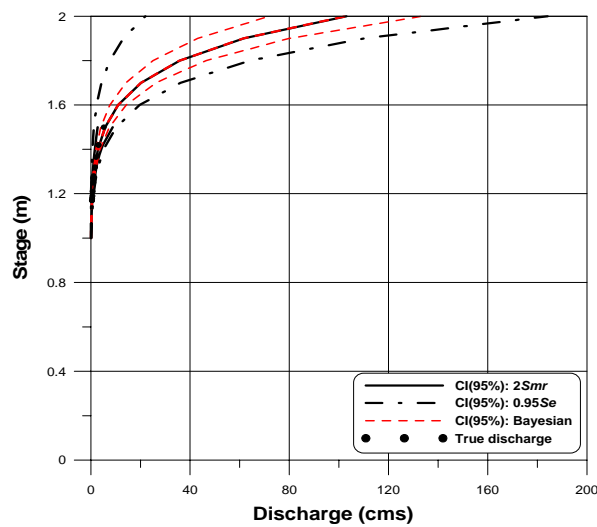
(b) Gocheoggyo



(c) Siheung



(d) Hakueicheon



(e) Mokgamcheon

Fig. 6. Rating Curves Including Uncertainty at 5 Gauges

- 4) 개별적인 수위에 대한 불확실성의 산정은 전체오차를 이용한 것보다 불확실성이 감소되어 산정되어 지지만, 외삽구간에 대한 불확실성을 산정할 수 없으므로 이를 위해서는 Bayesian 회귀분석을 수행하는 것이 유리하다.
- 5) 측정자료의 개수가 작은 경우 t 분포에 의한 오차의 산정은 과소 또는 과대추정의 위험이 있으나 Bayesian 회귀분석은 회귀분석 시 필요한 가정들과 자료의 개수에 대한 제한이 필요 없으므로 합리적으로 사용될 수 있고, 신규로 측정된 자료를 포함하여 다시 불확실성을 산정하는 경우 측정자료를 바로 반영할 수 있어 유용하게 사용될 수 있다.

위와 같은 통계적 실험의 결과가 실제 유량측정성과를 이용해서도 유사한 결과를 나타내는지 확인하기 위하여 안양천 유역의 안양, 고척교, 시흥, 학의천, 목감천 지점을 적용지점으로 선정하여 통계적 실험에서와 같은 방법을 적용하여 산정결과를 검증하였다. 현장자료는 참값을 모르므로 OLS와 Bayesian 회귀분석으로 추정된 회귀계수 중 어떤 기법에 의한 추정결과가 참값에 가까운지는 알 수 없어 어느 기법이 우수하다는 단적인 결론을 얻기는 어려웠다. 그러나 현장자료를 이용하는 경우에도 기존의 불확실성 산정 기법보다는 Bayesian 회귀분석에 의한 불확실성의 산정이 보다 불확실성을 감소시켜 나타내주므로 기존의 두 기법인 2 S_{mm} 과 $0.95S_e$ 보다는 Bayesian 회귀분석이 우수하다는 결론을 내릴 수 있었다.

본 연구에서는 본론에 언급한 것과 같이 수위-유량 관계곡선식의 작성에 있어서 필수적인 단계인 수위-유량 관계곡선의 분리에 대한 문제는 함께 다루지 않았다. 그러므로 향후 Bayesian 방법을 사용함에 있어서 수위-유량 관계곡선식을 분리해야 하는 경우 각 구간에 따라 Bayesian 회귀분석의 복잡한 절차를 간단하게 구성할 수 있는 사용자 시스템의 마련과 분리지점을 객관적으로 추정할 수 있는 방법도 함께 마련되어야 할 필요가 있으며, 이와 같은 시스템을 통하여 더욱 많은 지점에서의 불확실성 분석을 수행함으로써 본 연구에서 사용된 기법의 신뢰성 및 적용성을 제고할 수 있으리라 판단된다.

감사의 글

본 연구는 21세기 프런티어 연구개발 사업인 수자원의 지속적 확보기술개발 사업단(과제번호 1-7-3)의 서울대학교 공학연구소를 통한 연구비 지원(40%)과 서

울대학교 BK21 안전하고 지속가능한 사회기반건설사업단의 연구비 지원(60%)에 의해 수행되었습니다. 연구비 지원에 심심한 감사의 뜻을 포함합니다.

참고 문헌

김상욱과 이길성 (2008a). "Bayesian MCMC를 이용한 저수량 점 빈도분석: I. 이론적 배경과 사전분포의 구축." **한국수자원학회 논문집**, 한국수자원학회, 제41권, 제1호, pp. 35-47.

김상욱과 이길성 (2008b). "Bayesian MCMC를 이용한 저수량 점 빈도분석: II. 적용과 비교분석." **한국수자원학회 논문집**, 한국수자원학회, 제41권, 제1호, pp. 49-63.

김원, 김상호, 차준호, 김동구 (2001). "수위-유량관계의 정확도 분석." **대한토목학회 학술발표회 논문집**, 대한토목학회, pp. 1-4.

김치영, 차준호, 김동구, 김원 (2004). "하천 유량측정에서 제한된 측선수에 따른 불확실도." **한국수자원학회 학술발표회 논문집**, 한국수자원학회, pp. 1086-1091.

서규우, 김수현, 김대곤 (2005). "유량측정방법에 따른 형산강유량의 불확실도 분석." **한국수자원학회 학술발표회 논문집**, 한국수자원학회, pp. 1538-1542.

이길성 (1996). **낙동강 수계 실시간 최적 저수관리 시스템 개발 (분석모델 부문) 보고서**. 한국수자원공사.

이길성 (2001). **수자원 분석시스템 구축기법에 관한 연구보고서**. 한국수자원공사.

이길성, 이경호 (2004). "유량측정시 불확실성의 검토와 수위-유량곡선식의 개발." **한국수자원학회 학술발표회 논문집**, 한국수자원학회, pp. 1242-1246.

이길성 (2007). **안양천 유역의 물순환 건전화 기술개발**. 서울대학교, 과학기술부.

정성원, 김동희, 문장원, 김동필 (2003). "유량자료 품질 개선을 위한 정확도 제고방안(수위-유량 관계곡선)." **한국수자원학회 학술발표회 논문집(1)**, 한국수자원학회, pp. 121-124.

차준호, 김상호, 김원, 김동구 (2001). "유량측정자료의 불확실도." **대한토목학회 학술발표회 논문집**, 대한토목학회, pp. 5-8.

차준호, 김원, 윤광석, 김동구 (2002). "유량측정자료의 불확실도 분석." **한국수자원학회 학술발표회 논문집(2)**, 한국수자원학회, pp. 989-994.

한국수자원공사 (2001). **HYMOS**. pp. 8.8-8.12.

한국수자원공사 (2000, 2001, 2002, 2003, 2004, 2005,

- 2006). **수문자료집**.
- Bickel, P.J., and Doksum, K.A. (1977). *Mathematical statistics: basic ideas and selected topics*. Holden-Day, Inc., San Francisco, CA.
- DeGagne, M.P.J., Douglas, G.G., Hudson, H.R., and Simonovic, S.P. (1996). "A decision support system for the analysis and use of stage-discharge rating curves." *Journal of Hydrology*, Vol. 184, pp. 225-241.
- Herschey, R.W. (1980). *Manual on stream gauging: Computation of discharge*, Vol. 2, Report No. 13, WMO.
- Herschey, R.W. (1985). *Streamflow measurement*, Elsevier Applied Science Publishers. M.O.
- 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.
- Kuczera, G. (1999). "Comprehensive at-site flood frequency analysis using Monte Carlo Bayesian inference." *Water Resources Research*, Vol. 35, No. 5, pp. 1551-1557.
- Kuczera, G., and Parent E. (1998). "Monte Carlo assessment of parameter uncertainty in conceptual catchment models: The Metropolis algorithm." *Journal of Hydrology*, Vol. 211, pp. 69-85.
- Lambie, J.C. (1978). "Measurement of flow-velocity-area methods." In: Herschey RW, (Ed.), *Hydrometry: Principles and Practices*, Wiley, Chichester, Chapter 1.
- Lee, K.S., and Kim, S.U. (2007). "Identification of uncertainty in low flow frequency analysis using Bayesian MCMC method." *Hydrological Processes*, In press(on-line published).
- Madsen, H., and Rojsberg, H.D. (1997). "Generalized least squares and empirical Bayes estimation in regional partial duration series index flood modeling." *Water Resources Research*, Vol. 33, No. 4, pp. 771-781.
- Mosley, M.P., and McKerchar, A.I. (1993). "Streamflow." *Handbook of Hydrology*, Chap. 8, McGraw-Hill. N.Y.
- Moyeed R.A., and Clarke, R.T. (2005). "The use of Bayesian methods for fitting rating curves, with case studies." *Advances in Water Resources*, Vol. 28, pp. 807-818.
- O'Connell, D.R.H., Ostenaar, D.A., Levish, D.R., and Klinger, R.E. (2002). "Bayesian flood frequency analysis with paleohydrologic bound data." *Water Resources Research*, Vol. 38, No. 5, pp. 1-14.
- Petersen-Øverleir, A. (2004). "Accounting for heteroscedasticity in rating curve estimates." *Journal of Hydrology*, Vol. 292, pp. 173-181.
- Petersen-Øverleir, A., and Reitan, T. (2005). "Objective segmentation in compound rating curves." *Journal of Hydrology* Vol. 311, pp. 188-201.
- Rantz, S.E. (1982). "Measurement and computation of streamflow." *Vol. II. Computation of Discharge*. US Geological Survey Water Supply Paper 2175, Washington.
- Reis, D.S. Jr, and Stedinger, J.R. (2005). "Bayesian MCMC flood frequency analysis with historical information." *Journal of Hydrology*, Vol. 313, pp. 97-116.
- Reis, D.S. Jr, Stedinger, J.R., and Martins, E.S. (2005). "Bayesian generalized least squares regression with application to log Pearson type III regional skew estimation." *Water Resources Research*, Vol. 41, W10419.
- Sauer, V.B. and Meyer, R.W. (1992). *Determination of error in individual discharge measurements*. US Geological Survey Open-file Report, pp. 92-144.
- Seber, G.A.F. and Wild, C.J. (1989). *Nonlinear regression*, John Wiley & Sons, Inc., N.Y.
- Seidou, O., Ouarda, T.B.M.J., Barbet, M., Bruneau, P., and Bobee, B. (2006). "A parametric Bayesian combination of local and regional information in flood frequency analysis." *Water Resources Research*, Vol. 42, W11408, DOI: 10.1029/2005WR004397.
- Sorensen, D. and Gianola, D. (2002). *Likelihood, Bayesian, and MCMC methods in Quantitative Genetics*. Springer-Verlag, New York.
- Thiemann, M., Trosset, M., Gupta, H.V., and Sorooshian, S. (2001). "Bayesian recursive parameter estimation for hydrologic models." *Water*

- Resources Research*, Vol. 37, No. 10, pp. 2521-2535.
- Venables, W.N., and Smith, D.M. (2008). *An introduction to R*, R Development Core Team.
- Vicens, G.J., Rodriguez-Iturbe, I., and Schaake, J.C. Jr. (1975). "A Bayesian framework for the use of regional information in hydrology." *Water Resources Research*, Vol. 11, No. 3, pp. 405-414.
- Vrugt, J.A., Gupta, H.V., Bouten, W., and Sorooshian, S. (2003). "Shuffled complex evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters." *Water Resources Research*, Vol. 39, No. 8, SWC 1-16.
- Wood, E.F., and Rodriguez-Iturbe, I. (1975a). "Bayesian inference and decision making for extreme hydrologic events." *Water Resources Research*, Vol. 11, No. 4, pp. 533-542.
- Wood, E.F., and Rodriguez-Iturbe, I. (1975b). A Bayesian approach to analyze uncertainty among flood frequency models. *Water Resources Research*, Vol. 11, No. 6, pp. 839-843.
- WMO (1994). *Guide to Hydrological Practices*, 5th Ed., pp. 170-172.

(논문번호:08-19/접수:2008.02.13/심사완료:2008.07.15)