

비편향 회귀분석모형을 이용한 낙동강 본류 부유사량 산정방법의 신뢰도 향상

한수희 · 강두기* · 신현석* · 유재정** · 김상단*

부경대학교 환경시스템공학부

*부산대학교 토목공학과

**국립환경과학원 낙동강물환경연구소

Improvement of Suspended Solid Loads Estimation in Nakdong River Using Minimum Variance Unbiased Estimator

Han, Suhee · Kang, Du Kee* · Shin, Hyun Suk* · Yu, Jae-Jeong** · Kim, Sangdan*

Department of Environmental System Engineering, Pukyong National University

*Department of Civil Engineering, Pusan National University

**Nakdong River Water Environmental Research Center, National Institute of Environmental Research

(Received 15 January 2007, Accepted 26 February 2007)

Abstract

In this study three log-transformed linear regression models are compared with the focus of bias correction problem. The models are the traditional simple linear regression estimator (SL), the quasi maximum likelihood estimator (QMLE) and the minimum variance unbiased estimator (MVUE). Using such models, suspended solid loads can be estimated using the discharge - suspended solid data set that has been measured by NIER Nakdong River Water Environment Laboratory. As a result, SL shows negative bias for most values of the measured discharge range. QMLE is nearly unbiased for moderate values of the measured discharge range, but shows increasingly positive bias for either large or small value of the measured discharge range. MVUE is unbiased. It is also analyzed how the estimated regression coefficient and exponent are distributed along Nakdong river main stream.

keywords : Load estimation, Nakdong river, Regression, Suspended solid

1. 서론

하천을 통하여 이송되는 오염물질의 부하량을 추정하는 것과 관련하여 여러 가지 기법들이 연구되어 왔다(Ferguson, 1986; Koch et al., 1986; Richards et al., 1987; Young et al., 1988). 이러한 연구들은 일반적으로 하천의 특정 지점에서 이송되는 오염 부하량을 추정하기 위하여 수행되어왔으며, 이러한 경우 주로 사용되고 있는 기법은 관측된 유량 Q 와 농도 c 또는 오염 부하량 L 과의 관계를 근거로 한 단순 회귀분석기법인 rating curve($L = \alpha Q^{\beta}$)가 주로 이용되고 있다(김 등, 2004). 단순 회귀분석에 기초한 편향된 추정기법(biased estimator)은 식의 구조자체가 간단하다는 장점이 있으나 샘플링의 수적인 면이나 질적인 면에서 적절한 샘플링이 이루어졌을 경우에만 유효하다. 그러나 여러 경우에 있어 편향된 추정기법은 심각한 오차를 일으키는 것으로 알려져 있다(Lee, 1982). 특히 다수의 지류가 유입되는 대형 하천 본류의 총 오염 부하량을 추정해야 하는 경우 bias의 상대적인 중요성이 증가하는 경향이 있게 된

다. 즉, 각각의 지류에서의 부하량 추정 시 발생하는 무작위적 오차들은 부분적으로 상쇄되는 경향이 있는 반면에, 본류에서의 bias에 따른 오차는 하류방향으로 점진적으로 축적되는 경향을 보이게 되기 때문이다(Cohn et al., 1989).

Ferguson(1986, 1987) 및 Koch 등(1986)의 연구는 평균 오염 부하량을 추정하는 전통적인 단순 회귀분석기법이 크게 편향될 수 있으며, 오염 부하량이 심각하게 과소평가될 수도 있다는 사실에 초점을 맞추고 있다. 본 연구에서 bias는 회귀분석이 유도되는 “대수공간(log space)”에서 “실제공간(real space)”로 재변환(retransformation)하는 과정에서 발생하는 편차를 의미한다. Ferguson(1986)은 이러한 bias를 줄이기 위해 간단한 보정기법을 제안하고 있으나, 여러 특정한 상황에서는 적절하게 적용되기는 하지만, 그러한 보정기법은 bias를 근본적으로 제거하는 것은 아니며 경우에 따라서는 오염 부하량을 매우 과대평가할 수도 있는 것으로 알려져 있다(Gilroy et al., 1990).

이에 본 연구에서는 회귀분석 시 발생하는 bias가 보정되지 않은 추정기법과 bias가 보정된 추정기법의 특성을 정리하고, 보다 정확한 추정기법인 최소분산 비편향 추정기법(Finney, 1941; Bradu et al., 1970; Lee, 1982)을 이용하여 낙동강 본류 주요지점의 부유사 부하량 산정기법을 제안하

* To whom correspondence should be addressed.

skim@pknu.ac.kr

고자 한다.

2. 오염 부하량 산정을 위한 일반적인 회귀분석 모형

유량으로부터 오염 부하량을 추정하는데 사용되는 회귀 분석기법은 오염 부하량의 대수 $[\ln(L)]$ 와 유량의 대수 $[\ln(Q)]$ 사이에 어떤 관계식이 존재한다는 가정에서 출발하게 된다. 이러한 관계식은 일반적으로 다음과 같은 선형 모형으로 표현된다.

$$Y \equiv \ln(L) = \beta_0 + \beta_1 \ln(Q) + \varepsilon = \mu + \varepsilon \quad (1)$$

여기서 $\ln(\cdot)$ 은 자연대수함수, β_0 와 β_1 는 모형의 계수, μ 는 Q 가 주어졌을 때 Y 의 조건부 평균, ε 은 정규난수(normal random variable)이다. 실제 적용 시, 농도는 유량과 비교해보았을 때 상대적으로 간헐적인 측정(일반적인 경우 8일 간격 이상)이 이루어지게 되며, 따라서 각각의 샘플링에서의 ε_i 는 서로 독립임을 가정할 수 있다.

식 (1)은 승법모형의 경우로서 재변환하게 되면 다음과 같이 표현된다.

$$L = \exp[\mu]\eta = \exp[\beta_0]Q^{\beta_1}\eta \quad (2)$$

여기서 η 는 대수정규난수(lognormal random variable)이므로, 부하량 L 또한 확률변수가 된다. 유량 Q 가 주어졌을 때 부하량 L 의 조건부 중앙치(median)는 $\exp[\beta_0]Q^{\beta_1}$ 이 되며, 조건부 평균(mean)은 $\exp(\beta_0)Q^{\beta_1} \exp(\sigma_\varepsilon^2/2)$ 가 된다. 여기서 σ_ε^2 은 ε 의 분산이다(Aitchison et al., 1981).

부하량 L 과 유량 Q 에 따른 농도 C 는 다음과 같이 구할 수 있다.

$$C = K \frac{L}{Q} \quad (3)$$

여기서 K 는 단위환산계수이다. 따라서 농도 C 의 기대값은 다음과 같다.

$$E[C] = K \frac{E[L]}{Q} = K \frac{\exp(\beta_0)Q^{\beta_1}E[\eta]}{Q} = K \exp(\beta_0)Q^{\beta_1-1}E[\eta] \quad (4)$$

단위환산계수 K 와 유량 Q 는 결정론적 변수이므로 여기에 나타난 부하량 분석은 동일하게 농도 분석에도 적용될 수 있다. 예를 들어, 피센트로 표현된 부하량 추정치의 bias와 표준편차는 $\ln(c)$ 와 $\ln(Q)$ 에 관련된 모형에서와 동일하게 된다.

한 가지 중요한 언급사항으로는 본 연구에서 설명하고 있는 모든 과정은 오염 부하량이 특정한 선형모형을 따른

다는 대 전제 하에 이루어진 것이다. 이것은 연구 대상 구간인 낙동강 본류 하천의 유량-부유사량 관측 자료를 분석한 결과 타당한 것으로 나타났으나, 실제 적용하기 전에 항상 모형의 타당성의 평가가 우선되어야 할 것이다. 모형 구조의 선정이 잘못되게 된다면 그로부터 산출된 결과는 재변형에 따른 bias 효과보다 훨씬 큰 오차를 가져오게 될 것이다.

3. $\hat{\beta}$ 벡터 추정

오염 부하량 자료 $\underline{L} = [L_1, \dots, L_N]$ 과 이에 대응하여 동시에 관측된 유량자료 $\underline{Q} = [Q_1, \dots, Q_N]$ 가 주어지고, \underline{Y} 는 \underline{L} 의 대수로 이루어진 벡터, \underline{X} 는 첫 번째 열의 성분은 모두 1이며, 두 번째 열에는 \underline{Q} 의 대수로 이루어진 $N \times 2$ 행렬이라고 정의할 때, 회귀모형의 매개변수를 추정하기 위하여 σ_ε^2 에 대한 ordinary least squares estimates를 아래와 같이 구할 수 있다.

$$\hat{\underline{\beta}} = [\hat{\beta}_0, \hat{\beta}_1]' = (\underline{X}'\underline{X})^{-1}\underline{X}'\underline{Y} \quad (5)$$

여기서 \underline{X}' 은 행렬 \underline{X} 의 전치행렬(transposition)을 의미한다(Draper et al., 1981). 임의의 유량 Q^* 에 대응하는 부하량 L 의 대수의 비편향된 기대값 추정치 $\hat{\mu}$ 는 아래와 같다.

$$\hat{\mu} = \hat{\beta}_0 + \hat{\beta}_1 \ln(Q^*) = \underline{X}\hat{\underline{\beta}} \quad (6)$$

여기서 $\underline{X} = \{1, X^*\} = \{1, \ln(Q^*)\}$ 이다. $\hat{\mu}$ 는 평균 $\mu = \underline{X}\hat{\underline{\beta}}$ 와 분산 $V\sigma_\varepsilon^2$ 을 갖는 정규 확률변수이며, $V = \underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}'$ 이다. 본 연구의 경우 V 는 아래와 같이 간단하게 표현될 수 있다.

$$V = \left\{ \frac{1}{N} + \ln^2(Q^*) / \sum_{i=1}^N (\ln(Q_i) - \overline{\ln Q})^2 \right\} \quad (7)$$

여기서

$$\overline{\ln Q} = \sum_{i=1}^N \ln(Q_i) / N \quad (8)$$

부하량의 대수를 추정하는 문제는 식 (6)에 의해 별 다른 문제없이 해결된다. 그러나 부하량의 대수값을 구하는 것이 주된 목표가 아니라, 실제 공간에서의 부하량을 추정하는 것이 필요하게 되며, 추정치의 bias가 나타나는 것은 바로 이 때, 즉 대수 좌표에서 실제 좌표로 재변형할 때 발생하게 된다. 다음 절에서는 $E[L] = \exp[\mu + \sigma_\varepsilon^2/2]$ 를 추정하는 일반적인 기법에 대하여 소개할 것이다.

4. 부하량 산정 기법

4.1. 단순회귀분석기법(Simple Linear regression estimator, SL)

가장 널리 사용되는 재변형 기법은 $\hat{\mu}$ 에 지수를 취하는 것이다.

$$\hat{L}_{SL} = \exp[\hat{\mu}] = \exp[\underline{X}\hat{\beta}] = \exp[\hat{\beta}_0]Q^{\hat{\beta}_1} \quad (9)$$

단순히 Rating Curve라고도 불리는 기법으로 현재 가장 일반적으로 사용되고 있는 기법이다. Sichel(1952)과 Lane (1975), Landwehr(1987), DeLong(1982), Lee(1982), Ferguson (1986), DeLong 등(1987)에 따르면 식 (7)로부터의 \hat{L}_{SL} 는 $E[L]$ 에 대한 편향된 추정치라는 지적을 받아왔다. $\hat{\mu}$ 가 정규 분포 $N(\mu, V\sigma_r^2)$ 을 따르는 확률변수이기 때문에, $\exp[\hat{\mu}]$ 는 매개변수 $(\mu, V\sigma_r^2)$ 를 갖는 대수정규분포가 된다. 따라서 $\exp(\hat{\mu})$ 의 평균은 다음과 같이 주어진다(Aitchison et al., 1981).

$$E[\hat{L}_{SL}] = \exp[\mu] \exp[V\sigma_r^2/2] = E[L \exp[(V-1)\sigma_r^2/2]] \quad (10)$$

식 (10)로부터 대부분의 경우에 있어 \hat{L}_{SL} 는 아래쪽으로 편향되어 있음을 알 수 있다. 그러나 Rating Curve가 추정된 유량범위보다 훨씬 크거나 작은 경우에는 $E[\hat{L}_{SL}]$ 가 $E[L]$ 보다 크게 추정될 수도 있다. 이는 이후 보다 상세히 설명될 것이다(Fig. 5 참조).

이와 비슷하게 \hat{L}_{SL} 의 분산은 대수정규변수의 모멘트로부터 다음과 같이 계산할 수 있다.

$$\begin{aligned} Var[\hat{L}_{SL}] &= E[(\hat{L}_{SL})^2] - E^2[\hat{L}_{SL}] \\ &= \exp[2\mu + 2V\sigma_r^2] - \exp[2\mu + V\sigma_r^2] \\ &= \exp[2\mu + V\sigma_r^2](\exp[V\sigma_r^2] - 1) \end{aligned} \quad (11)$$

4.2 유사 최우도추정기법(Quasi Maximum Likelihood Estimator, QMLE)

Ferguson(1986)은 Rating Curve의 bias를 보정하기 위하여 $\exp(s^2/2)$ 를 \hat{L}_{RC} 에 곱하는 방법을 제시하였다. 여기서

$$s^2 \equiv \sum_{i=1}^N \frac{(Y_i - \hat{Y})^2}{N-k} \quad (12)$$

이 때, k 는 회귀모형의 매개변수의 수로서, 본 연구의 경우는 $k=2$ 이다. 따라서 아래와 같이 bias가 보정된 식을 구성할 수 있다.

$$\hat{L}_{QMLE} = \exp[\hat{\mu}] \exp[s^2/2] \quad (13)$$

V 가 작고 $\exp[s^2/2]$ 가 $\exp[\sigma_r^2/2]$ 에 대한 만족할 만한 추정치라면 식 (13)은 우수한 추정결과를 주는 것으로 알려져 있다.

\hat{L}_{QMLE} 의 기대값은 s^2 과 $\hat{\mu}$ 가 서로 독립이라는 사실로부터 구할 수 있고, 따라서 다음과 같다.

$$\begin{aligned} E[\hat{L}_{QMLE}] &= E[\exp[\hat{\mu}]]E[\exp[s^2/2]] \\ &= E[L] \exp[(V-1)\sigma_r^2/2] \left[1 - \frac{\sigma_r^2}{m} \right]^{-m/2} \end{aligned} \quad (14)$$

여기서 $m = N - k$ 는 오차분포에서 자유도(degrees of freedom)를 나타낸다.

\hat{L}_{QMLE} 의 분산은 위와 비슷한 과정을 통하여 유도되는데, 우선

$$E[(\exp(s^2/3))^2] = E[\exp(2s^2)] = \left[1 - \frac{2\sigma_r^2}{m} \right]^{-m/2} \quad (15)$$

이라 쓸 수 있으며, 따라서 분산은 아래와 같이 유도될 수 있다.

$$\begin{aligned} Var[\hat{L}_{QMLE}] &= E[(\hat{L}_{QMLE})^2] - E^2[\hat{L}_{QMLE}] \\ &= \exp[2\mu + 2V\sigma_r^2] \left[1 - \frac{2\sigma_r^2}{m} \right]^{-m/2} - \exp[2\mu + V\sigma_r^2] \left[1 - \frac{\sigma_r^2}{m} \right]^{-m} \\ &= \exp[2\mu + V\sigma_r^2] \left\{ \exp[V\sigma_r^2] \left[1 - \frac{2\sigma_r^2}{m} \right]^{-m/2} - \left[1 - \frac{\sigma_r^2}{m} \right]^{-m} \right\} \end{aligned} \quad (16)$$

4.3. 최소분산 비편향 추정기법 (Minimum Variance Unbiased Estimator, MVUE)

Finney(1941)의 연구를 기반으로 Bradu 등(1970)은 QMLE 추정기법보다 다소 복잡하기는 하지만 보다 적절한 특성을 가지는 추정기법인 최소분산 비편향 추정기법(Minimum Variance Unbiased Estimator)을 유도하였다. 최소분산 비편향 추정기법에서 부하량은 아래와 같이 정의된다.

$$\hat{L}_{MVUE} = \exp[\hat{\mu}] g_m \left(\frac{m+1}{2m} \{ (1-V)s^2 \} \right) \quad (17)$$

여기서

$$g_m(z) = \sum_{p=0}^{\infty} \frac{m^p (m+2p)}{m(m+2) \cdots (m+2p)} \left(\frac{m}{m+1} \right)^p \left(\frac{z^p}{p!} \right) \quad (18)$$

Bradu 등(1970)은 \hat{L}_{MVUE} 가 $E[L]$ 의 비편향 추정기법임을 보여준다. 최소 분산의 관점에서 보면, 샘플 크기에 관계없이 모든 비편향 추정기법 중에서 최소분산 비편향 추정기법이 가장 최적화된 기법임을 보인 바 있다.

\hat{L}_{MVUE} 의 분산은 Bradu 등(1970) 및 Likes(1980)에 의해서 아래와 같이 유도되어졌다.

$$Var[\hat{L}_{MVUE}] = \exp[2\mu] \{ \exp(2\sigma_r^2 V) \cdot G_m((1-V)\sigma_r^2/2) - \exp(\sigma_r^2) \} \quad (19)$$

여기서

$$G_m(z) = \sum_{h=0}^{\infty} \frac{\Gamma(m/2)}{\Gamma(m/2+h)} \binom{m+2h-2}{h} z^h \quad (20)$$

$$= \exp[2z] g_m \left[\frac{2(m+1)}{m^2} z^2 \right]$$

여기서 $\Gamma(\cdot)$ 는 Gamma 함수이다. 위 식들이 다소 복잡해 보일 수 있으나, 이것들은 이미 표의 형태로 만들어져 있으며(Bradu et al., 1970) 또한 인터넷 등을 통하여 Fortran 또는 Basic 등을 위한 subroutine 형태의 파일을 쉽게 획득할 수도 있다. 본 연구에서도 Matlab을 이용한 subroutine을 코딩하여 분석에 이용하였으며, 관련 파일을 Fig. 1에 수록하였다.

```
function GM2 = finney_gm(XM, ARG)

% *****
% FUNCTION TO COMPUTE FINNEY'S GMT)
%
% XM NUMBER OF DEGREES OF FREEDOM OF RESIDUALS
% ARG ARGUMENT TO FINNEY'S FUNCTION
%
% XM = NOBS - (# of unknown parameters)

TOL = 10^-7;
GM2 = -9051016.9051016;
if (abs(ARG) > 50.0)
    disp('MAGNITUDE OF ARG IS TOO LARGE (GM)')
    GM = 0;
else
    GM = 1;
    if (XM <= 0.0)
        disp('m <= 0, ERROR')
    end
    BT = ARG * XM^2 / (2.0 * (XM + 1));
    TERM = 1;
    for p=1:1000
        TERM = TERM * BT / ((XM/2 + p - 1) * p);
        GM = GM + TERM;
        if (p > 1 & abs(TERM) < TOL)
            GM2 = GM;
        end
    end
    if GM2 == -9051016.9051016
        disp('GM DID NOT CONVERGE')
    end
end
```

Fig. 1. The MATLAB function to evaluate the function $g_m(\cdot)$ for an argument in the range (-50,50).

5. 낙동강 본류에의 적용

국립환경과학원 낙동강물환경연구소에서는 2004년 하반기부터 낙동강 오염총량관리 단위유역 말단부를 대상으로 유량 및 수질을 8일 간격의 원칙으로 동시관측하고 있다. 본 연구에서는 관측자료의 분석결과 유량과 SS 관측자료의 상태가 양호한 2006년 1월부터 11월까지의 본류 지점(낙본 A, 낙본B, 낙본C, 낙본D, 낙본E, 낙본F, 낙본G, 낙본H, 낙본I, 낙본J, 낙본K, 낙본L, 낙본M)의 관측자료를 바탕으로 앞서 서술한 내용에 근거한 부유사량 산정공식을 제시하고자 한다. 관측회수는 30회이다.

유량측정방법은 수질오염공정시험법의 하천유량측정방법과 건설교통부 수문관측매뉴얼 등에 있는 일반적인 유량측정방법이 적용되었으며, 등유량 5%내의 측선배분, 최소 측정시간 40초 이상 등의 구체적인 유량측정지침을 최대한 준수하여 회전식유속계(Price USGS Type AA, LV와 USGS Pygmy 및 Valeport 002, 106)를 이용하여 도섭법과 추 내림을 이용한 교량법으로 측정하여 실측 유속과 단면적으로 유량을 산정하였다. 단, 조사지점의 수리적, 지형적 특성(접근성) 등으로 실측이 곤란한 지점에 대해서는 댐방류량의 활용, 물수지분석, 비유량법 등 간접유량법을 적용하여 산정하였는데, 예를 들면, 배수영향이 심한 낙본M지점은 하구언 방류량을, 낙본E지점은 실측유량자료를 기반으로 단위유역면적을 고려한 비유량법을, 낙본F, 낙본G, 낙본J 및 낙본L지점은 본류유입지천 유량을 고려한 물수지를, 낙본K지점은 본류유입지천 유량 및 취수장의 취수량을 고려한 물수지를 이용하여 유량을 산정하였다(낙동강물환경연구소, 2006). 낙동강 본류의 SS는 표층 1m 깊이에서 채수한 후 시료를 냉장 운반하여 수질오염공정시험방법(GF/C 여과 후, 여과전후의 중량 차)에 따라 수질분석을 실시하였다.

Table 1은 단순회귀분석기법에 의한 낙동강 본류 지점별 유량-부유사량 추정공식을 나타내었으며, Table 2에 유사 최우도추정기법에 의한 공식을 수록하였다. 최소분산 비편향추정기법의 경우에는 회귀계수 자체가 유량의 함수로 표

Table 1. SS estimation formular with simple linear regression estimator ($L = \alpha Q^{\beta}$)

| Unit Watershed | α | β | R^2 |
|----------------|----------|---------|-------|
| NB-A | 143.42 | 1.41 | 0.67 |
| NB-B | 80.60 | 1.55 | 0.92 |
| NB-C | 7.30 | 2.13 | 0.88 |
| NB-D | 23.95 | 1.78 | 0.89 |
| NB-E | 11.32 | 1.97 | 0.89 |
| NB-F | 303.41 | 1.34 | 0.67 |
| NB-G | 270.58 | 1.33 | 0.85 |
| NB-H | 631.55 | 1.21 | 0.95 |
| NB-I | 92.05 | 1.56 | 0.89 |
| NB-J | 376.90 | 1.33 | 0.82 |
| NB-K | 102.13 | 1.48 | 0.88 |
| NB-L | 211.75 | 1.34 | 0.88 |
| NB-M | 1,398.66 | 1.01 | 0.95 |

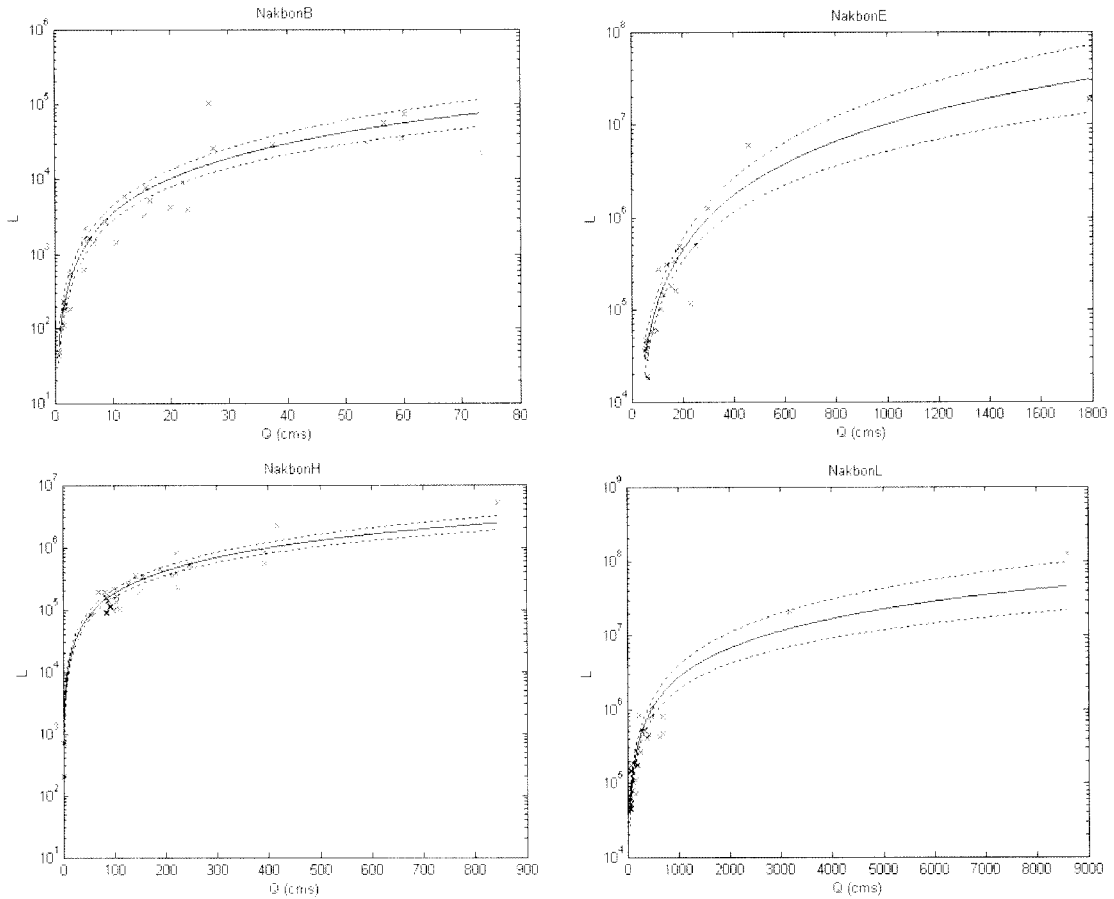


Fig. 2. Suspended solid load estimation results using MVUE.

Table 2. SS estimation formular with quasi maximum likelihood estimator ($L = \alpha Q^\beta$)

| Unit Watershed | α | β | R^2 |
|----------------|----------|---------|-------|
| NB-A | 229.75 | 1.41 | 0.67 |
| NB-B | 98.04 | 1.55 | 0.92 |
| NB-C | 8.47 | 2.13 | 0.88 |
| NB-D | 29.28 | 1.78 | 0.89 |
| NB-E | 13.54 | 1.97 | 0.89 |
| NB-F | 469.46 | 1.34 | 0.67 |
| NB-G | 308.14 | 1.33 | 0.85 |
| NB-H | 688.60 | 1.21 | 0.95 |
| NB-I | 108.30 | 1.56 | 0.89 |
| NB-J | 479.85 | 1.33 | 0.82 |
| NB-K | 128.63 | 1.48 | 0.88 |
| NB-L | 255.39 | 1.34 | 0.88 |
| NB-M | 1,468.32 | 1.01 | 0.95 |

현됨에 따라 단일한 공식으로 표현될 수 없다. Fig. 2에 주요지점에 대한 \hat{L}_{MVUE} 의 추정곡선과 95% 신뢰구간을 도시하였다. Fig. 2에서 Q는 유량(m^3)이며 L은 부유사 부하량(kg/day)이다.

본 연구에서 사용한 자료와 회귀분석 모형의 적합성을 평가하기 위해 추정값과 실측값의 간차(error)의 정규분포성은 Kolmogonov-Smirnov 검정법을 이용하여 적합성 검정을 실시하였으며, 간차의 독립성은 자기 계열상관도(autocorre-

lation coefficient, acf)를 작성하여 시각적으로 검정하였다. 검정 결과 추정기법에 상관없이 전 지점의 간차 자료들이 정규분포에 적합한 것으로 나타났으며(Fig. 3), 독립성 검정 부분은 낙본 H와 낙본 K에서 1차 계열상관계수가 다소 큰 값이 나왔으나, 전반적으로 독립성 여부도 양호한 것으로 나타났다(Fig. 4).

Fig. 5는 주요지점별 부하량 추정기법별 부하량 추정치의 편향정도를 대수 값을 취한 유량의 함수 값으로 나타내고 있다. 부유사 부하량의 추정은 회귀분석을 통하여 유량 자료로부터 유도되는 경우가 많으며 실무적으로 회귀분석 시에 원 자료에 대수를 취하여 시행하는 경우 설명변수로부터 도출된 종속변수는 재변환 과정을 반드시 거쳐야 할 필요가 있다. 이러한 재변환 과정에서 bias 보정 문제가 발생된다. 여기서 bias는 간차의 왜곡도를 평가하는 요소로서 회귀분석 모형의 적합성을 평가하는 기준이 된다. 앞서 예상한 바와 같이 \hat{L}_{SL} 는 대부분의 유량 조건에서 음의 편향, 즉 상대적으로 과소추정된 부하량을 제공하게 된다. 이러한 음의 편향성을 제거하기 위해서는 역-대수변환 전에 추정된 대수변환된 부하량에 어떤 보정계수를 곱해줌으로써 이를 해결할 수 있다. 다른 한편, Fig. 3은 유량 조건이 매우 크거나 작을 경우 \hat{L}_{SL} 는 양의 편향성을 보일 수도 있음은 또한 보여주고 있다. \hat{L}_{QMLK} 는 일반적인 유량 조건에서는 편향성을 거의 보이지 않지만, 유량 조건이 커지거나

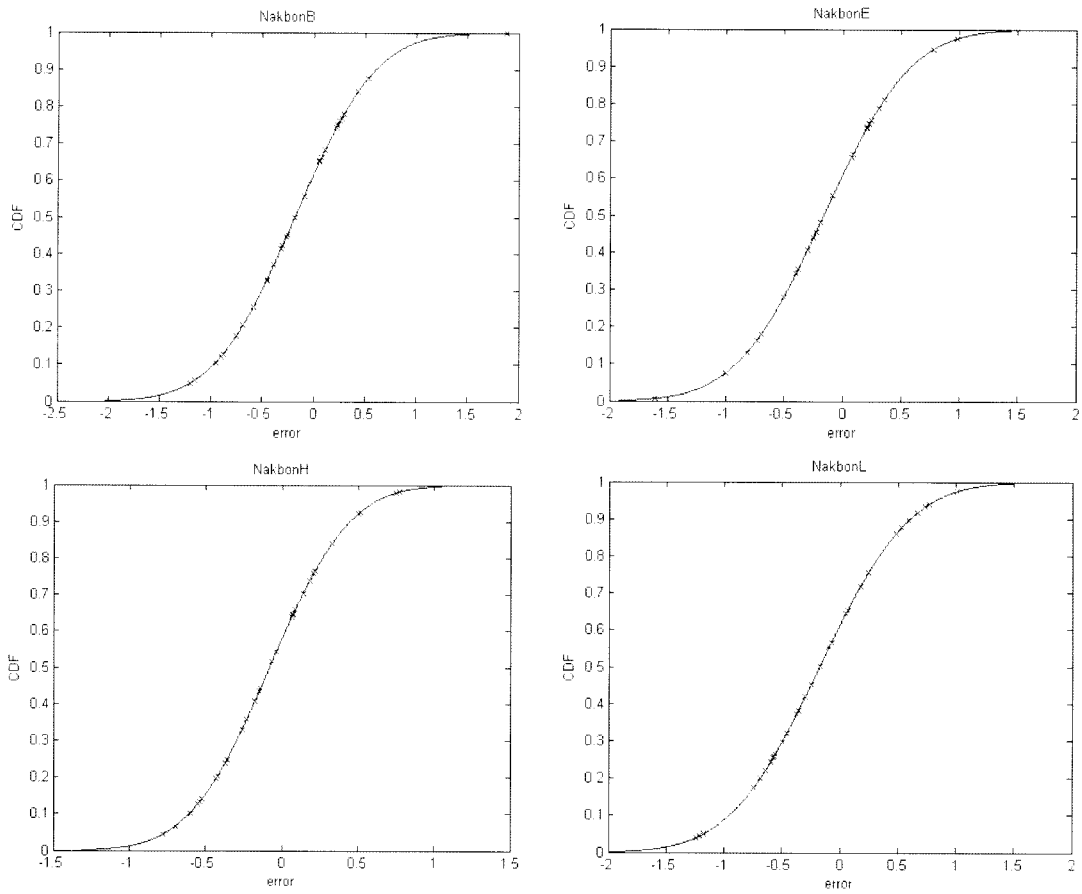


Fig. 3. Normality test results using MVUE.

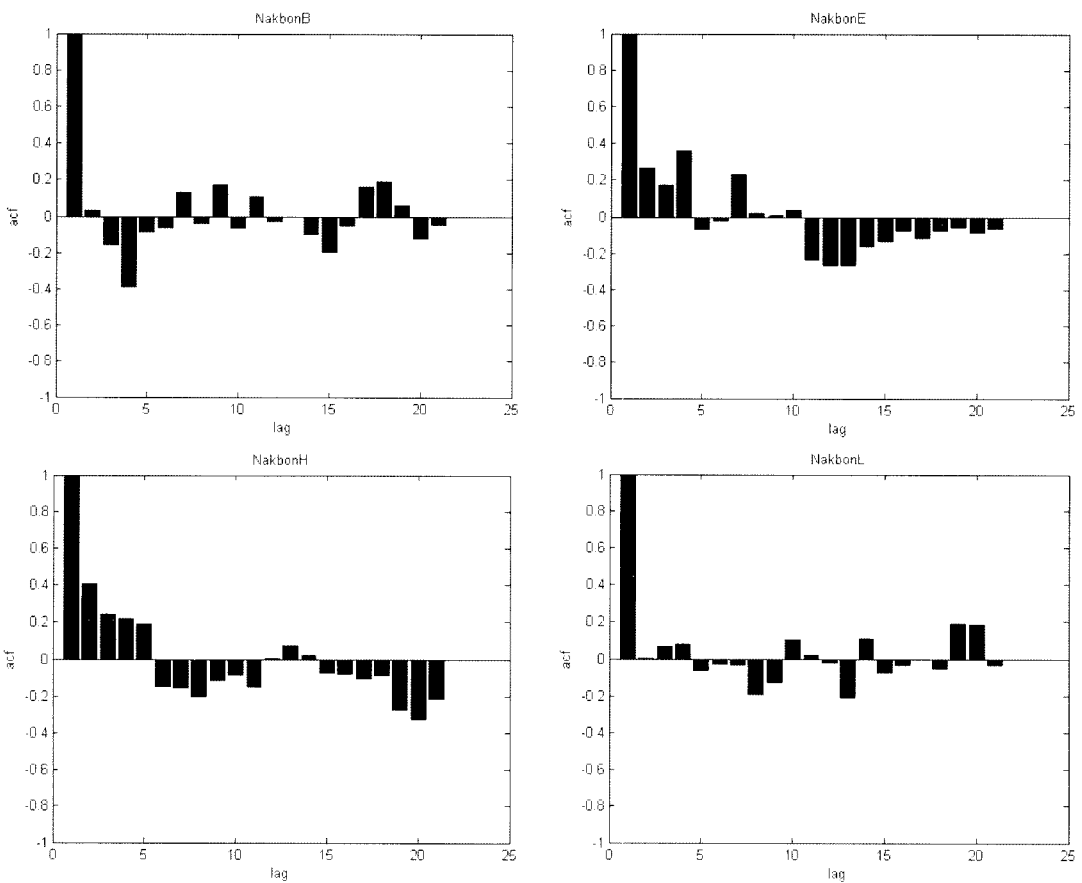


Fig. 4. Independency test results using MVUE.

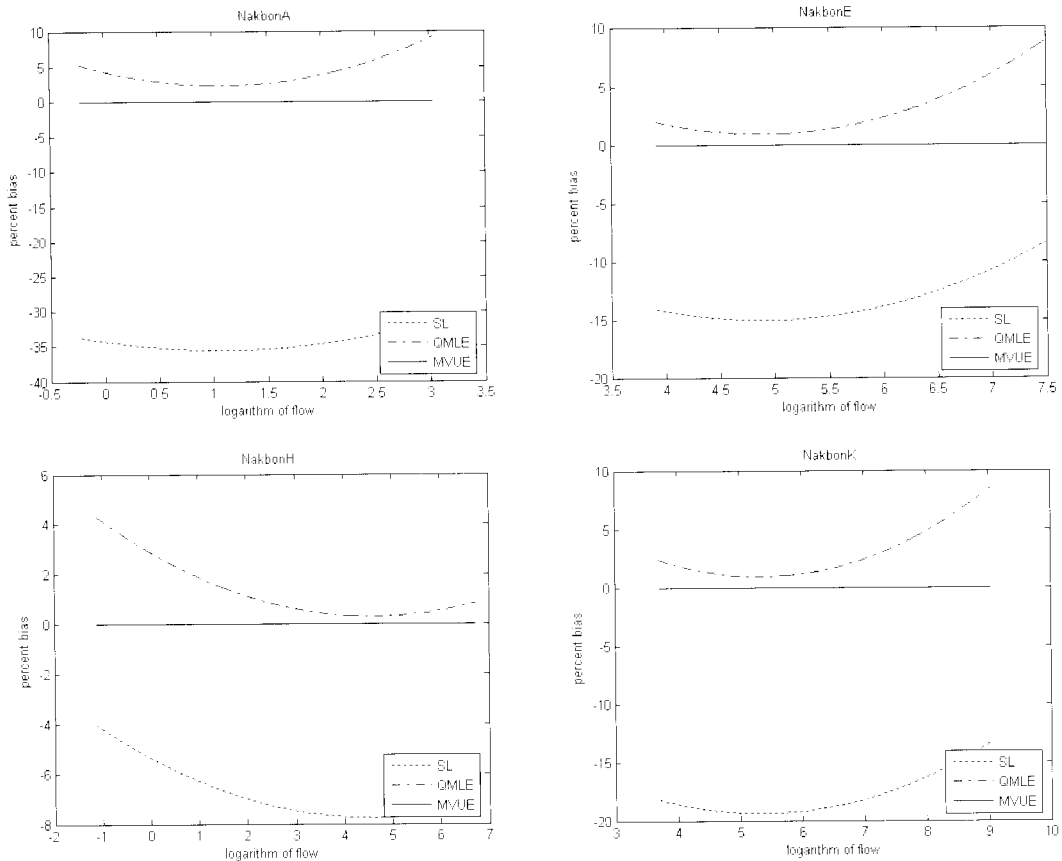


Fig. 5. Biases of SL, QMSE and MVUE as a function of the logarithm of the predictor variable Q.

작아질 경우에는 양의 편향성, 즉 부하량이 과대추정될 수도 있음을 보여주고 있다. 이러한 편향성은 유량 조건이 극한 값일 경우 매우 증폭될 수 있는 것으로 나타났다. 이는 QMLE의 경우 단순회귀분석 시의 편향성을 보정하기 위해 곱해진 보정계수가 너무 과다한 경향이 있음을 의미한다. 여기서 MVUE는 모든 경우에 비편향성을 나타내게 된다. 단, 이때 MVUE의 bias가 0인 것은 부하량 예측에 오차가 없다는 것을 의미하지 않음에 주의해야 한다. 다만, 잔차가 정규분포를 갖도록 조정된 것이며, 이에 따라 다른 기법에 비하여 상대적으로 정확한 예측을 수행할 수 있다는 것을 의미한다.

Fig. 6에 각 추정기법들로부터 도출된 추정곡선을 비교하여 나타내었다. 최소분산 비편향 추정기법과 비교해 볼 때, 단순회귀분석기법의 경우 부하량이 보다 작게 추정되고 있으며, Fig. 6에는 명확하게 나타나지는 않지만 Fig. 5의 결과로부터 유사 최우도추정기법의 경우 부하량이 보다 크게 추정되고 있음을 알 수 있다. 그러나 관측유량범위에서는 유사 최우도추정기법과 최소분산 비편향 추정기법 사이에 추정된 부하량 차이는 크지 않음을 또한 알 수 있다.

Fig. 7에 유사 최우도추정기법을 이용하여 낙동강 본류를 따라 추정된 α 값을 도시하였다. 최상류에서 중류부까지는 α 의 변동패턴은 감소하는 추세를 보이다가 중류(낙본F)에 이르러서는 갑자기 매우 큰 값으로 이동되는 현상을 보여 주고 있으며, 다시 낙본H에서 최하류까지는 감소되는 추세

를 보이다가 최종출구(낙본M)에서 급격하게 큰 값을 나타내고 있다. 부하량 추정공식 $L = \alpha Q^{\beta} \rightarrow \ln(L) = \ln(\alpha) + \beta \ln(Q)$ 의 구조를 살펴보면 α 의 역할은 하천을 따라 기본적으로 이동되는 부하량의 크기를 설명하는 매개변수임을 알 수 있다. 즉, 최상류부에서 낙동강 중류부까지는 이동되는 기본 부하량의 규모는 점차 작아지는 패턴을 보이다가 중류부에 이르러서 금호강의 유입 등 다량의 점·비점 오염 부하량이 유입됨에 따라 하천을 따라 이동되는 부하량이 급격히 증가되고 있으며, 이들은 영향은 하류로 유하하면서 점차 감소되는 것으로 설명될 수 있다. α 값이 다시 급격히 증가하는 최하류부 낙본M 지점의 경우는 조위 및 하구언 조절의 영향을 받는 구간으로 하천의 이송 메커니즘으로는 설명될 수 없는 지점인 것으로 판단된다.

Fig. 8에 유사 최우도추정기법을 이용하여 낙동강 본류를 따라 추정된 β 값을 도시하였다. 정도차이는 있으나 낙동강 본류 전 구간에 걸쳐 1보다 큰 값을 보이고 있으며, 이는 낙동강 본류의 경우 유량이 증가할수록 부유사 농도가 함께 증가함을 나타내는 것이다. 강우 시 유출되는 비점오염원의 유출특성을 볼 때, 낙동강 전 구간에 걸친 유량-부유사 농도 비례관계는 부유사 농도를 지표로 한 비점오염원 관리가 효과적일 수 있을 것이라는 추론을 가능하게 할 것이다. 낙동강 본류 상류와 하류 사이의 β 값의 공간분포는 또한 상류부에서 유량증가에 따른 부유사 농도증가패턴이 상대적으로 보다 뚜렷함을 보여주고 있다.

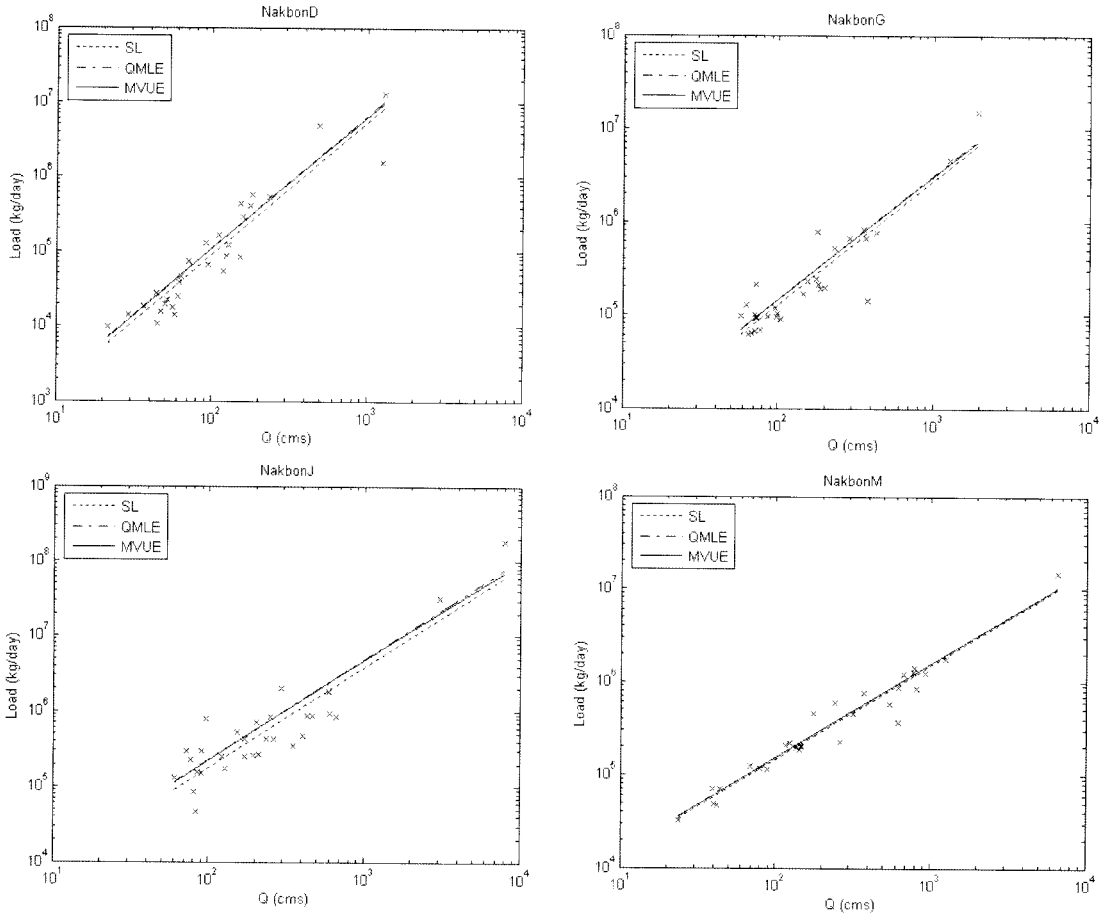


Fig. 6. Comparison of suspended solid loads estimated using three techniques.

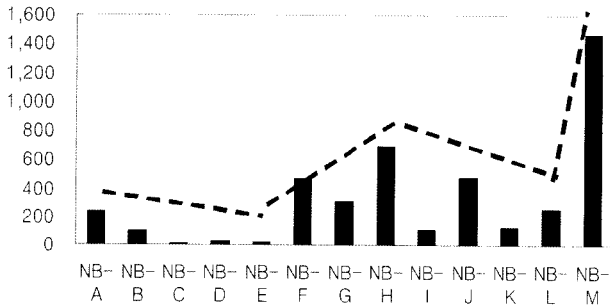


Fig. 7. Spatial distribution of coefficient α estimated using QMLE.

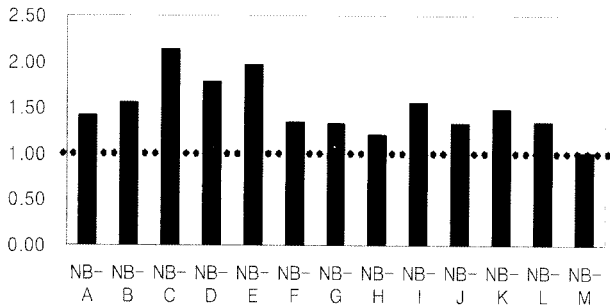


Fig. 8. Spatial distribution of exponent β estimated using QMLE.

6. 결론

본 연구에서는 대수변환 선형회귀분석모형과 함께 사용될 수 있는 세 가지 추정기법에서의 편향성과 분산을 소개하였다. 세 가지 부하량 추정기법을 이용하여 국립환경과학원 낙동강물환경연구소에서 2006년 1월부터 11월까지의 낙동강 본류 지점에 관측한 유량-부유사농도 자료에 대한 분석을 수행한 결과 단순회귀분석기법의 경우는 대부분의 유량 조건에서 과소추정된 부하량을 제공하고 있으며, 유사최우도 추정기법은 일반적인 유량 조건에서는 편향성을 거의 보이지 않지만, 유량 조건이 커지거나 작아질 경우에는 부하량이 과대추정될 수 있는 것으로 나타났다. 최소분산 비편향 추정기법은 모든 경우에 비편향성을 나타내게 되므로 다른 기법에 비하여 상대적으로 정확한 부하량 추정이 이루어질 수 있음을 살펴보았다. 또한 낙동강 본류를 따라 추정된 회귀분석모형의 계수 α 및 지수 β 를 도시해 본 결과, 최상류부에서 낙동강 중류부까지는 하천 규모에 대한 부하량의 규모는 점차 작아지는 패턴을 보이다가 중류부에 이르러서 금호강의 유입 등 다량의 점·비점 오염 부하량이 유입됨에 따라 부하량의 규모가 급격히 증가되고 있으며 이들은 영향은 다시 하류로 유하하면서 점차 감소되는 것으로 설명될 수 있었고, 낙동강 본류의 경우 유량이 증

가할수록 부유사 농도가 함께 증가하는 경향을 나타내며 상류부에서 유량증가에 따른 부유사 농도증가패턴이 상대적으로 보다 더 뚜렷함을 알 수 있었다. 강우 시 유출되는 비점오염원의 유출특성을 볼 때, 낙동강 전 구간에 걸친 유량-부유사 농도 비례관계는 부유사 농도를 지표로 한 비점오염원 관리가 효과적일 수 있을 것이라는 추론을 가능하게 한다.

사 사

본 연구는 낙동강 수계 환경기초조사사업(낙동강 하천수리특성분석 및 수리모의예측기법 개발)의 일환으로 수행되었음.

참고문헌

- 김상단, 송미영, 이기영, 이성룡, 단순회귀모형을 이용한 인구와 도시적 토지이용이 팔당호 수질에 미치는 영향분석, *수질보전 한국물환경학회지*, **20**, pp. 703-707 (2004).
- 낙동강물환경연구소, 2006년 유량측정사업, 국립환경과학원 (2006).
- Aitchison, J. and Brown, J. A. C., *The Lognormal Distribution With Special Reference to Its Uses in Economics*, Cambridge University Press, New York (1981).
- Bradu, D. and Mundlak, Y., Estimation in Lognormal Linear Models, *J. Am. Stat. Assoc.*, **65**(329), pp. 198-211 (1970).
- Cohn, T. A., DeLong, L. L. and Gilroy, E. J., Estimating Constituent Loads, *Water Resour. Res.*, **25**(5), pp. 937-942 (1989).
- DeLong, L. L., Water Quality of Streams and Springs, Green River Basin, Wyoming, U.S. Geol. Surv., *Water Resour. Invest. Rep.*, 82-4008, p. 36 (1982).
- DeLong, L. L. and Wells, D. K., Estimating Average Dissolved-solids Yields from Basins Drained by Ephemeral and Intermittent Streams-Green River Basin, Wyoming, U.S. Geol. Surv., *Water Resour. Invest. Rep.*, 87-4222, p. 29 (1987).
- Draper, N. and Smith, H., *Applied Regression Analysis*, 2nd ed., 709, John Wiley, New York (1981).
- Duan, N., Smearing Estimate: A Nonparametric Retransformation Method, *J. Am. Stat. Assoc.*, **78**(383), pp. 605-610 (1983).
- Ferguson, R. I., River Loads Underestimated by Rating Curves, *Water Resour. Res.*, **22**(1), pp. 74-76 (1986).
- Ferguson, R. I., Accuracy and Precision of Methods for Estimating River Loads, *Earth Surf. Processes Landforms*, **12**, pp. 95-104 (1987).
- Finney, D. J., On the Distribution of a Variate Whose Logarithm Normally Distributed, *J. R. Stat. Soc. Suppl.*, **7**, pp. 155-161 (1941).
- Gilroy, E. J., Hirsch, R. M. and Cohn, T. A., Mean Square Error of Regression-based Constituent Transport Estimates, *Water Resour. Res.*, **26**(9), pp. 2069-2077 (1990).
- Koch, R. W. and Smillie, G. M., Biased in Hydrologic Prediction Using Log-transformed Regression Models, *Water Resour. Bull.*, **22**(5), pp. 717-723 (1986).
- Landwehr, J. M., Some Properties of the Geometric Mean and Its Use in Water Quality Standards, *Water Resour. Res.*, **14**(3), pp. 467-473 (1978).
- Lane, W. L., Extraction of Information on Inorganic Water Quality, *Hydro. Pap.*, **73**, Colo. State Univ., Ft. Collins (1975).
- Lee, C. Y., Comparison of Two Correction Methods for the Bias due to the Logarithmic Transformation in the Estimation of Biomass, *Can. J. Forest Res.*, **12**, pp. 326-331 (1982).
- Likes, J., Variance of the MVUE for Lognormal Variance, *Technometrics*, **22**(2), pp. 253-258 (1980).
- Richards, R. P. and Holloway, J., Monte Carlo Studies of Sampling Strategies for Estimating Tributary Loads, *Water Resour. Res.*, **23**(10), pp. 1939-1948 (1987).
- Rukhin, A. L., Improved Estimation in Lognormal Models, *J. Am. Stat. Assoc.*, **81**(396), pp. 1046-1049 (1986).
- Sichel, H. S., New Methods in the Statistical Evaluation of Minor Sampling Data, *Inst. Min. Metall. Trans.*, **61**, pp. 261-288 (1952).
- Thomas, R. B., Estimating Total Suspended Sediment Yield with Probability Sampling, *Water Resour. Res.*, **21**(9), pp. 1381-1388 (1985).
- Young, T. C., DePinto, J. V. and Heidtke, T. H., Factors Affecting the Efficiency of Some Estimators of Fluvial Total Phosphorus Load, *Water Resour. Res.*, **24**(9), pp. 535-540 (1988).