



한강수계에서의 부유사 예측을 위한 LOADEST 모형의 회귀식의 평가

Evaluation of Regression Models in LOADEST to Estimate Suspended Solid Load in Hangang Waterbody

박윤식* · 이지민** · 정영훈** · 신민환** · 박지형*** · 황하선*** · 류지철*** · 박장호** · 김기성**,[†]

Park, Youn Shik · Lee, Ji Min · Jung, Younghun · Shin, Min Hwan · Park, Ji Hyung · Hwang, Hasun · Ryu, Jichul · Park, Jangho · Kim, Ki-sung

Abstract

Typically, water quality sampling takes place intermittently since sample collection and following analysis requires substantial cost and efforts. Therefore regression models (or rating curves) are often used to interpolate water quality data. LOADEST has nine regression models to estimate water quality data, and one regression model needs to be selected automatically or manually. The nine regression models in LOADEST and auto-selection by LOADEST were evaluated in the study. Suspended solids data were collected from forty-nine stations from the Water Information System of the Ministry of Environment. Suspended solid data from each station was divided into two groups for calibration and validation. Nash-Stucliff efficiency (NSE) and coefficient of determination (R_2) were used to evaluate estimated suspended solid loads. The regression models numbered 1 and 3 in LOADEST provided higher NSE and R_2 , compared to the other regression models. The regression modes numbered 2, 5, 6, 8, and 9 in LOADEST provided low NSE. In addition, the regression model selected by LOADEST did not necessarily provide better suspended solid estimations than the other regression models did.

Keywords: Load estimation; LOADEST; Regression model; Suspended solid

1. 서론

기후변화와 도시화는 유역특성의 변동성을 증가시키며, 연쇄적으로 인간과 생태계에 대한 하천의 기능에서 불안정한 요소로 작용한다 (Carey et al., 2011; Jung et al., 2014; Lee et al., 2014; Nam et al., 2014). 특히, 불투수층 면적과 강우강도의 증가는 하천유량의 증가와 함께 하천으로 비점오염원의 유입을 증가시키기 때문에 수생태 환경에 심각한 위협으로 다가올 수 있다 (Arnold and Gibbons, 1996; Shuster et al., 2005; Johnson et al., 2013). 이러한 심각성을 인지함에 따라 하천유량과 비점오염원에 대한 모니터링이 이루어지고 있다. 그러나 하천단면에서 하천유량과 오염원의 조합으로 이루어진 수질을 측정하는 것은 일반적으로 쉽지 않다 (Sun et al., 2013). 이러한 이유로 수질은 어느 특정한 주기를 가지며 측

정이 수행되며 (Robertson, 2003), 비점오염원에 대한 하천 수질은 강우사상의 영향을 많이 받기 때문에 강우사상 발생 시 하천유량의 변동성과 함께 측정된다.

비점오염원의 영향을 받는 하천의 수질은 하천 유량과 흐름조건에 대해 영향을 많이 받기 때문에 많은 하천유량의 장기적인 관측 자료를 이용하여 비연속성을 가지는 수질자료의 제한성을 극복할 수 있다 (Oh et al., 2013). 수문모형과 수질모형의 결합은 연속적인 유량을 이용하여 수문모형의 검보정을 수행하고 이와 더불어 비연속적인 수질자료를 이용하여 수질모형의 검보정을 수행하여 연속적인 수질모의를 가능하게 한다. 예를 들어 Hydrologic Simulation Program-Fortran (HSPF) (US Environmental Protection Agency; USEPA, 2001), Soil and Water Assessment Tool (SWAT) (Arnold et al., 1998), Generalized Watershed Loading Function (GWLf) (Haith et al., 1992)은 연속적인 수질모의 가능한 대표적 물리적 수문모형이다. 그러나 이러한 모형들은 수문 및 수질에 대한 각 과정을 물리적으로 해석하기 위하여 다양한 시공간적 입력 자료를 사용하기 때문에 넓은 지역이나 많은 지역들에 대한 수질모의를 수행하는 데 있어 자료수집, 자료구축, 계산 시간 등에서 많은 시간과 노력을 필요로 한다.

LOAD ESTimator (LOADEST; Runkel et al., 2004)는 최적 회귀모형 (best fit regression model)으로 위에서 언급한

* Department of Rural Construction Engineering, Kongju National University

** Regional Infrastructure Engineering, Kangwon National University

*** National Institute of Environmental Research

[†] Corresponding author

Tel.: +82-33-250-6466 Fax: +82-33-251-1518

E-mail: kskim@kangwon.ac.kr

Received: October 28, 2014

Revised: February 10, 2015

Accepted: February 11, 2015

물리적 기반의 모형들 보다는 상대적으로 간단하게 하천유량과 수질의 관계를 이용하여 수질을 모의할 수 있다. 또한, LOADEST는 특정기간에 대한 오염부하량을 산정하기 위한 2개의 회귀식과 모형 보정 및 오염부하량을 예측하기 위한 9개의 회귀식으로 구성되어 있다. 이러한 이유로 LOADEST는 다양한 수질관련 연구에 사용되어 왔으며, 많은 문헌에서 그 적용성이 입증되어 왔다. 최근 Oh et al. (2013)은 미국 남동부의 18개 지역에 대하여 K-nearest neighbor resampling 기법으로 날씨를 예측하여 일유량을 산정한 후 LOADEST에 적용하여 일단위의 총질소를 예측하였다. Sun et al. (2013)은 LOADEST를 이용하여 중국 Yangtze 강 상류와 Three Gorges 댐으로부터 유출되는 질소와 인의 부하량의 장기적인 추세를 파악하였다. 또한, Jha and Jha (2013)은 미국 North Carolina의 Neuse 강 유역에서 관측된 총인과 총질소에 대하여 LOADEST의 적용성을 평가하였으며, 총인과 총질소에 대한 예측결과에서 비교적 높은 정확성을 보여주었다. LOADEST는 다수의 회귀식을 가지고 있는데, Park and Engel (2014)은 LOADEST의 1, 3, 4, 7번 회귀식이 연단위의 인과 부유사를 산정하는데 있어 다른 회귀식보다 더 적합하다는 결과를 보여주었다. 또한 부유사(Domblaser and Striegl, 2009), 총질소(Oh and Sankarsubramanian, 2011) 등의 수질 자료에 이용된 바 있으며, 또한 아질산염질소, 질산염, 총인의 실측자료에 대해 적용한 결과 실측치에 가까운 모의치를 모의한 바 있다(Carey et al., 2011).

LOADEST는 수질자료의 미세측지점에 대한 예측에 널리 사용되고 있는 데에 비해, 모형에 대한 평가는 찾아보기 힘들며, 더욱이 한국의 수질자료에 대한 모형의 거동 및 LOADEST의 회귀식에 대한 평가가 필요한 것으로 생각된다. 따라서 본 연구의 목적은 부유사 예측에 대한 LOADEST의 회귀식 평가에 있다.

II. 연구방법

1. LOADEST 모형의 개요

LOADEST는 유량자료와 측정된 수질자료를 이용하여 회귀식 계수를 보정한 뒤, 보정된 회귀식에 의해 수질 자료의 미세측지점에 대한 수질 자료 혹은 오염부하량을 예측하기 위해 사용되고 있다. LOADEST는 11개의 회귀식(식 1-11)을 가지고 있으며, 이 중 한 개의 회귀식에 의해 오염부하량을 예측한다. 11개의 회귀식 중 회귀식 10(식 10)과 11(식 11)은 특정 기간에 대한 오염부하를 예측하기 위한 것이며, 회귀식 1부터 9(식 1-9)는 주어진 유량과 수질 자료에 대해 회귀식의

보정 및 오염부하를 예측하기 위한 것이다. 회귀식은 사용자에 의해 선택이 되거나 혹은 회귀식을 0으로 설정할 경우, 회귀식 1에서 9까지 9개의 회귀식 중 하나의 회귀식을 자동으로 선택한 후, 이 회귀식에 의해 오염부하를 예측하는 것이 가능하다.

LOADEST의 회귀식은 회귀식 계수와 함께 로그로 치환된 유량과 측정시간의 조합으로 이루어져 있다. 회귀식 1의 경우 로그로 치환된 유량과 두 개의 회귀식 계수로 이루어진 가장 간단한 형태의 회귀식으로, 이러한 형태의 회귀식은 유사량 자료에 대해 적용성이 높다고 보고된 바 있다(Crawford, 1991). 회귀식 9의 경우, 로그로 치환된 유량 및 로그로 치환된 유량의 제곱 등 7개의 회귀식 계수와 함께 7개의 항으로 이루어진 가장 복잡한 형태의 회귀식이며, 이러한 형태의 회귀식은 총인 자료에 대해 적용성이 높다고 보고된 바 있다(Cohn et al., 1992).

$$a_0 + a_1 \ln Q \tag{1}$$

$$a_0 + a_1 \ln Q + a_2 \ln Q^2 \tag{2}$$

$$a_0 + a_1 \ln Q + a_2 dtime \tag{3}$$

$$a_0 + a_1 \ln Q + a_2 \sin(2\pi dtime) + a_3 \cos(2\pi dtime) \tag{4}$$

$$a_0 + a_1 \ln Q + a_2 \ln Q^2 + a_3 dtime \tag{5}$$

$$a_0 + a_1 \ln Q + a_2 \ln Q^2 + a_3 \sin(2\pi dtime) + a_4 \cos(2\pi dtime) \tag{6}$$

$$a_0 + a_1 \ln Q + a_2 \sin(2\pi dtime) + a_3 \cos(2\pi dtime) + a_4 dtime \tag{7}$$

$$a_0 + a_1 \ln Q + a_2 \ln Q^2 + a_3 \sin(2\pi dtime) + a_4 \cos(2\pi dtime) + a_5 dtime \tag{8}$$

$$a_0 + a_1 \ln Q + a_2 \ln Q^2 + a_3 \sin(2\pi dtime) + a_4 \cos(2\pi dtime) + a_5 dtime + a_6 dtime^2 \tag{9}$$

$$a_0 + a_1 per + a_2 \ln Q + a_3 \ln Q per \tag{10}$$

$$a_0 + a_1 per + a_2 \ln Q + a_3 \ln Q per + a_4 \ln Q^2 + a_5 \ln Q^2 per \quad (11)$$

여기서 a_{0-6} 는 회귀식 계수, Q 은 유량, $dtime$ 연중 시간을 0-1의 소수로 환산한 값이며, per 는 사용자가 정의한 기간 (날짜)이다.

2. LOADEST 모형을 위한 유량 및 수질자료의 이용

LOADEST 모형의 회귀식의 평가를 위해 환경부의 물환경정보시스템 (<http://water.nier.go.kr/main/mainContent.do>)의 총량측정망 자료 중에, 한강 수계에 위치한 모든 지점인 49개 지점의 자료를 이용하였다(Fig. 1). 각 측정점으로부터 2004년부터 2014년까지의 부유사가 수집되었으며, 수집된 자료는 최소 83개 (진위 A; Jinwi A)부터 478개 (경안 B; Kyungan B)까지 다양했다. Table 1은 부유사 자료의 기간(Period), 기간 중 최소 농도 (min.; mg/l), 기간 중 최대 농도 (max.; mg/l), 기간 중 평균 농도 (mean; mg/l), 부유사 농도 자료의 개수

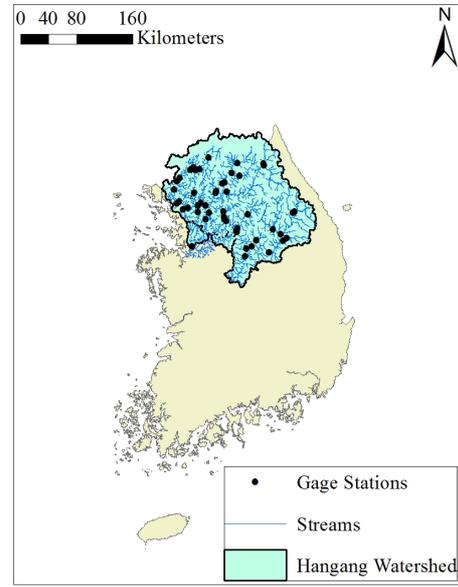


Fig. 1 Locations of streamflow and total suspended solid data for 49 stations

Table 1 Statistics of suspended solid data (mg/l) from 49 stations

Name	Period	Min.	Max.	Mean	Num.	Name	Period	Min.	Max.	Mean	Num.
Ahnyang A	2007.05.04-2014.05.29	2.0	389.0	17.3	319	Hangang I	2007.05.04-2014.05.29	2.7	280	20.2	319
Bukhan A	2004.08.04-2014.05.27	0.2	134.0	3.4	437	Hantan A	2007.05.02-2014.05.27	0.3	200	10.5	320
Bukhan B	2004.08.04-2014.05.27	0.3	49.2	3.4	437	Hantan B	2007.05.02-2014.05.27	1.7	546	15.7	320
Bukhan C	2004.08.03-2014.05.28	0.4	135.0	7.5	443	Inbuk A	2004.08.04-2014.05.27	0.4	343	9.3	441
Bukhan D	2004.08.03-2014.05.28	0.4	660.0	8.0	371	Imjin A	2007.05.03-2014.05.27	0.3	103	7.6	314
Bokha A	2004.08.10-2014.05.28	1.4	2073.3	41.8	373	Imjin B	2007.05.03-2014.05.28	1.0	896.7	16.6	309
Cheongmi A	2004.08.10-2014.05.28	0.7	1050.0	19.3	372	Jacheon A	2004.08.11-2014.05.27	0.2	311.3	7.6	372
Dalcheon A	2004.08.06-2014.05.27	0.2	126.5	8.3	444	Jojong A	2004.08.03-2014.05.28	0.4	263.9	7.6	443
Dalcheon B	2006.04.06-2014.05.27	1.0	143.3	9.0	370	Jeungrang A	2006.03.22-2014.05.29	2.2	988.6	23.6	371
Heukcheon A	2004.08.06-2014.05.28	0.2	336.7	8.2	445	Jucheon A	2004.08.05-2014.05.27	0.1	176	6.1	440
Kyungan A	2004.08.13-2014.05.29	0.7	727.0	16.2	475	Jinwi A	2012.01.11-2014.05.28	4.8	54.3	18.6	83
Kyungan B	2004.08.13-2014.05.29	0.8	690.0	14.2	478	Moosan A	2007.05.03-2014.05.29	4.6	1485	154.4	321
Golji A	2005.03.31-2014.05.27	0.1	204.5	7.2	415	Ohdae A	2004.08.05-2014.05.27	0.1	442	12.7	440
Kapyeong A	2004.08.03-2014.05.28	0.1	151.7	3.3	443	Pyeongchang A	2004.08.05-2014.05.27	0.2	150	7.4	372
Kongreung A	2012.01.19-2014.05.29	4.0	205.0	23.2	105	Shincheon A	2007.05.02-2014.05.28	4.7	1753.3	34.3	319
Gulpo A	2007.05.04-2014.05.29	3.8	168.0	16.2	320	Seomgang A	2004.08.25-2014.05.28	0.2	376	8.3	372
Hongcheon A	2004.08.03-2014.05.27	0.2	97.0	4.6	430	Seomgang B	2004.08.06-2014.05.28	0.8	950	11.0	463
Hangang A	2004.08.05-2014.05.28	0.2	485.5	8.0	441	Soyang A	2004.08.04-2014.05.27	0.1	281	5.5	441
Hangang B	2004.08.06-2014.05.28	0.1	421.5	7.7	437	Soyang B	2004.08.03-2014.05.28	0.1	166	6.0	456
Hangang C	2006.04.06-2014.05.27	0.6	61.0	3.7	372	Tancheon A	2006.03.30-2014.05.29	3.5	642	25.1	371
Hangang D	2004.08.06-2014.05.28	0.1	106.5	8.4	447	Okdong A	2004.08.05-2014.05.28	0.1	195	2.7	441
Hangang E	2004.08.06-2014.05.28	0.8	1546.7	18.6	463	Wangsuk A	2004.08.13-2014.05.29	1.8	482	14.2	374
Hangang F	2004.08.13-2014.05.29	1.8	250.0	11.0	445	Yanghwa A	2006.04.05-2014.05.28	1.7	1693.3	37.0	373
Hangang G	2005.01.20-2014.05.29	1.9	222.0	11.1	374	Youngpyeong A	2007.05.02-2014.05.27	1.2	572.5	14.5	320
Hangang H	2007.05.04-2014.05.29	1.2	192	13.1	336						

(num.)을 보여준다.

회귀식을 포함하여 수문 모형 등에 대해 적용성을 판단할 때에는 보정 및 검정 결과가 함께 분석되어야 하므로, LOADEST의 회귀식에 대해 보정 및 검정이 이루어졌다. 이를 위해서는 실측치를 분할할 필요가 있으며, 일반적으로는 자료의 기간에 의해 분할하여 보정과 검정 과정을 거치나, 회귀식은 보정된 자료의 기간을 벗어날 경우 유역 내 토지이용 현황 변화와 같은 유역의 장기기간에 걸친 변화나 자료 기간에 대한 외삽치 등이 회귀식의 거동에 영향을 줄 수 있다 (Runkel et al., 2004; Park, 2014). 따라서 회귀식의 보정 및 검정 과정 중 자료의 기간은 동일할 필요가 있었으며, 실측 자료의 기간과 무관하게 각 지점에서 자료의 홀수 번째 자료를 보정 자료로 사용하고, 짝수 번째 자료를 검정 자료로 활용하였다. 그리고 자료에 따라 회귀식의 보정 및 검정 자료에 대한 거동이 달라질 수도 있으므로, 짝수 번째 자료를 보정 자료로 사용하고, 홀수 번째 자료를 검정 자료로 활용하여 회귀식을 보정 및 검정하였다. 즉, 1 개의 지점에 대해 두 종류의 보정/검정 자료가 구축되었으며, 두 종류의 보정/검정 자료는 LOADEST의 9개의 회귀식 및 LOADEST의 자동 선택에 의한 회귀식에 의한 부유사 예측에 사용되었다. 따라서 1개 지점 자료에 대해 총 20회의 모의가 이루어졌으며, 49개 지점에 대해 총 980회의 부유사에 대한 예측이 이루어졌다.

3. LOADEST 모형에 의한 모의치 평가

총 49개의 측정지점자료를 홀수 번째 자료와 짝수 번째 자료로 분할하면서 총 98개의 유량/수질 자료가 구축이 되었으며, 이 98개의 유량/수질 자료는 LOADEST 모형의 회귀식 모형 번호를 0부터 9까지 적용하면서 총 980번의 LOADEST 모의가 이루어졌다. 그리고 980번의 LOADEST 모의에 대한 보정 및 검정 결과를 판단하기 위해 Nash-Stucliff Efficiency (NSE; 식 12)와 결정계수 (Coefficient of Determination; R^2 ; 식 13)가 사용되었다.

각 회귀식의 모의에 대한 NSE와 R^2 는 최솟값, 최댓값, 평균값과 백분위수 (Percentile)를 이용하여 표시하였다. NSE는 $-\infty$ 부터 1까지의 값을 가지기 때문에, 회귀식에 의한 NSE 값들의 범위가 큰 경우가 발생하였으며, 이에 NSE를 도식화함에 있어, 10 백분위수, 25 백분위수, 50 백분위수 (중앙값), 75 백분위수, 90 백분위수만을 이용하였다. 그리고 최솟값, 최댓값, 평균값은 표로 제시하였다.

$$NSE = 1 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (O_i - \bar{P})^2} \quad (12)$$

$$R^2 = \left[\frac{\sum_{i=1}^n (O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^n (O_i - \bar{O})^2} \sqrt{\sum_{i=1}^n (P_i - \bar{P})^2}} \right]^2 \quad (13)$$

여기서, O_i 는 측정시간 i 에 대한 실측치, P_i 는 측정시간 i 에 대한 모의치, \bar{O} 는 실측치의 평균, \bar{P} 는 모의치의 평균값을 의미한다.

III. 결과 및 고찰

1. LOADEST 모형의 보정 결과

LOADEST 모형의 회귀식 보정과정에서, NSE의 10 백분위수는 회귀식 2가 -0.1678로 최솟값을 보였으며, 회귀식 6이 0.2080으로 최댓값을 보였다 (Fig. 2). 90 백분위수는 회귀식 7이 0.7562로 최솟값을 보였으며, 회귀식 5가 0.9022로 최댓값을 보였다. R^2 의 10 백분위수는 회귀식 5가 0.3809로 최솟값을 보였으며, 회귀식 7이 0.4919로 최댓값을 보였다. 90 백분위수는 회귀식 7이 0.9339로 최솟값을 보였으며, 회귀식 2, 5, 6이 0.9799로 최댓값을 보였다 (Fig. 3). 모의 결과의 상위 10%와 하위 10%를 제외하면 보정과정에서 각 회귀식은 NSE와 R^2 의 범위에서 큰 차이를 보이지는 않았다. 그러나 보정과정의 전체 모의 결과 중 최솟값은 회귀식 2가 -6.6458의 NSE와 0.0301의 R^2 를 보였으며, 회귀식 2를 비롯하여 4, 5, 6, 7, 8, 9가 음수의 NSE값을 보였다 (Table 2). R^2 의 경우 다른 회귀식의 경우 0.1271 (회귀식 1) 보다 크게 나타났지만, 회귀식 2, 5, 6의 최솟값은 0.0301, 0.0541, 0.0791의 값을 보였다. 즉, 회귀식 1과 3이 다른 회귀식에 비해 NSE와 R^2 분포가 좁게 나타났다.

모형의 거동을 판단할 때, 모형에 의한 모의치가 얼마나 실측치에 가까운가를 판단하는 정확성 (Accuracy)과, 모형에 의한 모의치가 얼마나 분산되는가를 판단하는 정밀성 (Precision)을 모두 고려할 필요가 있다 (Park and Engel, 2014; Phillips et al., 1999; Preston et al., 1989). LOADEST의 회귀식들의 보정 결과, 회귀식 2와 5를 제외하고, 10 백분위수, 25 백분위수, 50 백분위수, 75 백분위수, 90 백분위수는 큰 차이를 보이지는 않았다. 그러나 전체 모의 결과 중 최솟값의 경우 회귀식 1

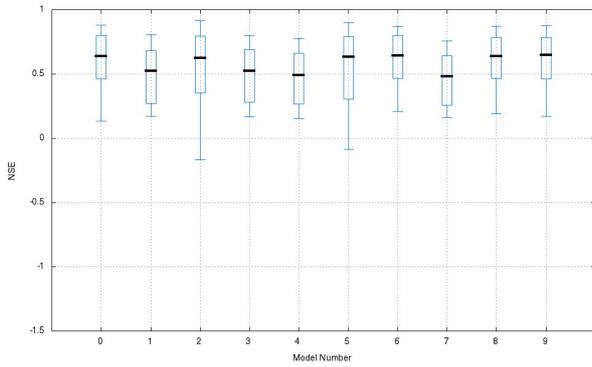


Fig. 2 Nash–Stucliff Efficiency of estimated sediment load by each model to measured sediment load in calibration

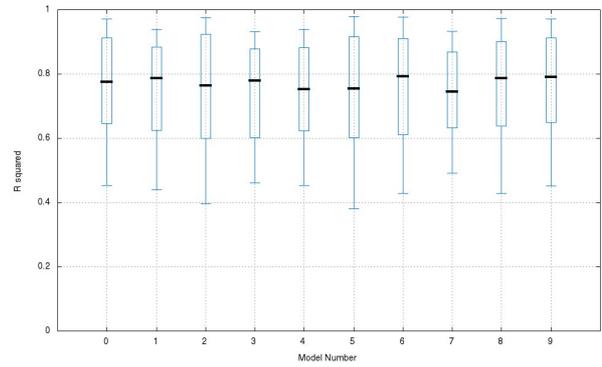


Fig. 3 Determination coefficient of estimated sediment load by each model to measured sediment load in calibration

Table 2 Minimum, maximum, and mean of NSE and R² in calibration

Model Numer	NSE			R ²		
	Min.	Max.	Mean	Min.	Max.	Mean
0	-3.8394	0.9933	0.5286	0.1338	0.9989	0.7415
1	0.0777	0.9819	0.4978	0.1271	0.9971	0.7256
2	-6.6548	0.9902	0.3630	0.0301	0.9995	0.7214
3	0.0819	0.9322	0.4950	0.1536	0.9978	0.7222
4	-0.7329	0.9856	0.4598	0.1539	0.9971	0.7298
5	-4.4152	0.9793	0.4016	0.0541	0.9995	0.7227
6	-5.8264	0.9766	0.4903	0.0791	0.9990	0.7350
7	-1.6054	0.9950	0.4454	0.1664	0.9972	0.7285
8	-3.8394	0.9972	0.5271	0.1320	0.9989	0.7385
9	-3.8697	0.9933	0.5295	0.1338	0.9989	0.7424

과 3에 비해 나머지 회귀식의 경우 상대적으로 넓은 범위의 NSE와 R²를 보였다. 즉, 회귀식 1과 3이 다른 회귀식에 비해 나은 정밀성을 보였다.

NSE의 평균값은 회귀식 2가 0.3630으로 최솟값을 보였으며, 회귀식 9가 0.5295로 최댓값을 보였다. 다른 회귀식에 비해 높은 정밀성을 보인 회귀식 3의 NSE 평균값은 0.4950을 보였다. R²의 평균값은 0.7214에서 0.7424로 큰 차이를 보이지는 않았다(Table 2). 즉 정확성은 회귀식 8과 9가 가장 높게 나타났지만, 정확성과 정밀성을 함께 고려할 때, 회귀식 1과 3이 다른 회귀식에 비해 실측치가 가까운 모의치를 모의 하는 것으로 나타났다.

LOADEST의 회귀식이 0일 경우, 9개의 회귀식 중에 가장 잘 모의하는 것으로 판단되는 회귀식을 자동으로 선택한다. LOADEST의 회귀식 0에 의한 모의 결과, NSE는 -3.8394부터 0.9933의 범위를 보였고, R²는 0.1338에서 0.9989의 범위를 보였다. NSE의 평균은 0.5286이었으며, R²의 평균은 0.7415였다. 즉, LOADEST의 회귀식 0에 의한 모의치가 특

정 회귀식을 선택하였을 때보다 모의치에 가깝다고 보기에는 힘든 것으로 나타났다. 특히, 회귀식 0에 의한 모의치를 회귀식 1과 3에 의한 모의치와 비교할 때 정확성에서는 큰 차이를 보이지 않으나, 정밀성에서는 낮은 것으로 나타났다.

2. LOADEST 모형의 검증

LOADEST 모형의 회귀식 검증과정에서, NSE의 10 백분위수는 회귀식 2가 -1.4283으로 최솟값을 보였으며, 회귀식 7이 0.1640으로 최댓값을 보였다(Fig. 4). 90 백분위수는 회귀식 1이 0.7886로 최솟값을 보였으며, 회귀식 4가 0.8425로 최댓값을 보였다. R²의 10 백분위수는 회귀식 2가 0.38으로 최솟값을 보였으며, 회귀식 7이 0.4890으로 최댓값을 보였다. 90 백분위수는 회귀식 3이 0.9243로 최솟값을 보였으며, 회귀식 6이 0.9992로 최댓값을 보였다. 모의 결과의 상위 10%와 하위 10%를 제외했을 때, 검증과정에서 회귀식 1, 3, 4, 7이 회귀식 2, 5, 6, 8, 9에 비해 좁은 NSE의 범위를 보였으며

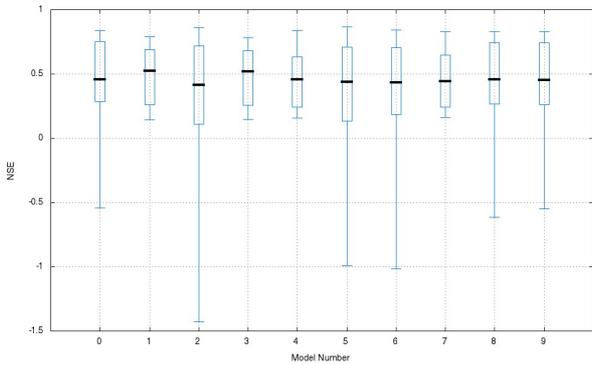


Fig. 4 Nash–Stucliff Efficiency of estimated sediment load by each model to measured sediment load in validation

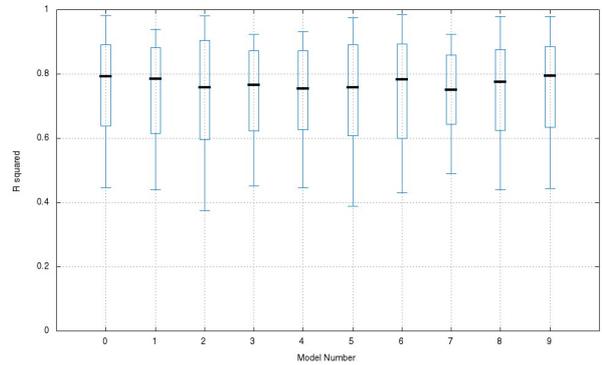


Fig. 5 Determination coefficient of estimated sediment load by each model to measured sediment load in validation

Table 3 Minimum, maximum, and mean of NSE and R² in validation

Model Number	NSE			R ²		
	Min.	Max.	Mean	Min.	Max.	Mean
0	-701,4280	0,9811	-7,1730	0,0933	0,9992	0,7372
1	-1,3265	0,9904	0,4694	0,1149	0,9968	0,7266
2	-320,3694	0,9762	-3,7563	0,0269	0,9995	0,7141
3	-1,7414	0,9918	0,4670	0,1350	0,9977	0,7226
4	-5,0527	0,9782	0,4177	0,1238	0,9971	0,7293
5	-328,9656	0,9872	-4,0884	0,0465	0,9995	0,7187
6	-701,4280	0,9830	-7,1177	0,0643	0,9992	0,7237
7	-6,3453	0,9819	0,4065	0,1421	0,9973	0,7275
8	-734,2746	0,9811	-7,4457	0,1064	0,9991	0,7301
9	-742,5269	0,9803	-7,6020	0,0933	0,9991	0,7345

R²는 큰 범위차이를 보이지는 않았다(Fig. 5). 검증과정의 전체 모의 결과를 볼 때, NSE의 경우 회귀식 1, 3, 4, 7은 최솟값의 범위가 -6.3453 (회귀식 7)에서 -1.3265 (회귀식 1)인데 비해, 회귀식 2, 5, 6, 8, 9의 경우 -742.5269 (회귀식 9)에서 -320.3694 (회귀식 2)이었다. NSE의 최댓값은 모든 회귀식이 0.9762 이상의 값을 보였다. 검증과정에서 NSE의 평균값은 회귀식 1, 3, 4, 7은 0.4065 (회귀식 7)에서 0.4694 (회귀식 1)를 보였지만, 회귀식 2, 5, 6, 8, 9는 -7.6020 (회귀식 9)에서 -3.7563 (회귀식 2)을 보였다. R²의 경우 최솟값의 범위는 0.0269 (회귀식 2)에서 0.1421 (회귀식 7)의 범위였으며, 평균값은 0.7141에서 0.7345의 범위를 보였다(Table 3).

R²의 범위에서는 큰 차이를 보이지 않았으나, NSE의 범위는 회귀식 1, 3, 4, 7보다 회귀식 2, 5, 6, 8, 9가 넓었기 때문에, 회귀식 1, 3, 4, 7이 상대적으로 높은 정밀성을 보인 것으로 나타났다. R²의 평균값에서는 큰 차이를 보이지 않았으나, NSE의 평균값을 볼 때, 회귀식 1, 3, 4, 7보다 회귀식 2, 5, 6, 8, 9가 작은 값들을 보였기 때문에, 정확성에 있어서도 회귀식 1, 3,

4, 7이 높은 것으로 나타났다.

LOADEST의 회귀식 0에 의한 모의 결과, NSE는 -701.4380부터 0.9811의 범위를 보였고, R²는 0.0933에서 0.9991의 범위를 보였다. NSE의 평균은 -7.1730이었으며, R²의 평균은 0.7372였다. 즉, LOADEST의 회귀식 0에 의한 모의치가 회귀식 1, 3, 4, 7에 의한 모의치보다 실측치에 가까울 가능성은 낮은 것으로 나타났다. 즉, LOADEST의 회귀식을 0으로 선택할 경우, 특정 회귀식을 선택하였을 때보다 모의치에 가깝다고 보기에는 힘든 것으로 나타났다.

3. LOADEST 모형에 의한 회귀식 자동선택 결과 분석

LOADEST의 회귀식은 로그로 치환된 유량과 두 개의 회귀식 계수로 이루어진 회귀식 1부터 로그로 치환된 유량이나 로그로 치환된 유량의 제곱 등과 7개의 회귀식 계수와 함께 6개의 항을 가진 회귀식 9까지 다양하다. 회귀식의 복잡성 혹은 항의 개수와 회귀식의 정확성 및 정밀성과의 관계를 찾아

Table 4 Percentage (%) at which LOADEST selects the best-fit regression model

Calibration		Validation	
NSE	R ²	NSE	R ²
11.2	14.3	7.1	12.2

볼 수 없었다. 그러나, 9개의 회귀식에 대해 로그로 치환된 유량의 제곱 항을 기준으로 회귀식 1, 3, 4, 7과 회귀식 2, 5, 6, 8, 9가 정확성 및 정밀성에서 차이가 있었다. 로그로 치환된 유량의 제곱 항을 포함하지 않은 회귀식 1, 3, 4, 7이 로그로 치환된 유량의 제곱 항을 포함하는 회귀식 2, 5, 6, 8, 9보다 높은 정확성 및 정밀성을 가지는 것으로 나타났다.

LOADEST에 의해 자동으로 선택된 회귀식에 의한 부유물질 모의치는 정확성 및 정밀성이 낮을 가능성이 크며, 이는 LOADEST에 의해 회귀식이 선택이 될 때, 정확성 및 정밀성이 낮은 회귀식 2, 5, 6, 8, 9가 선택이 된 경우 때문인 것으로 판단된다. 즉, 보정과정 중에 LOADEST에 의해 선택된 회귀식에 의한 모의치와 실측치의 NSE가, 선택되지 않은 회귀식에 의한 모의치와 실측치의 NSE 보다 큰 경우는 전체 98개 수질/유량자료 세트에 대해 11개 자료세트 (11.2%)였으며, R²의 경우는 14개 자료세트 (14.3%)인 것으로 나타났다. 다시 말해, NSE를 기준으로 하였을 때 LOADEST가 최적의 회귀식을 선택할 확률은 11.2%, R²를 기준으로 하였을 때는 14.3%로 낮게 나타났다 (Table 4). 검증과정을 볼 때, LOADEST에 의해 선택된 회귀식이 최적 회귀식인 경우는 전체 98개 수질/유량자료 세트 중, NSE를 기준으로 할 경우 7개 자료세트 (7.1%)이며 R²를 기준으로 할 경우 12개 자료세트 (12.2%)로 보정과정과 마찬가지로 낮게 나타났다. 각 지점에서 모의된 부하량의 실측된 부하량에 대한 NSE 중 가장 높은 값을 보인 회귀식에 대해서 Table 5에 표시하였다.

Fig. 6은 안양A (Ahnyang A) 지점의 부유물질부하량(ton) 모의치로, Fig. 6(a)의 경우 LOADEST에 의해 회귀식이 자동으로 선택되었으며 모의된 부하량이 실측된 부하량에 대해서 NSE는 0.1232였으며 R²는 0.7897이었다. Fig. 6(b)의 경우 회귀식 1에 의해 부하량이 모의되었으며, 실측된 부하량에 대해서 NSE는 0.8531이었으며, R²는 0.8807이었다. 안양A 지점에 대해서 LOADEST의 의해 선택된 회귀식은 6번 (식(6))이었는데, 실측치에 비해 모의치가 크게 산정된 경향이 있었으며 특히 2012년 8월 30일에 실측된 부하량은 2,698 ton (유량: 76.532 m³/s, 부유물질농도: 408 mg/L)이었는데 비해, 모의치는 5,806 ton으로 실측치의 2배를 초과하였다. 회귀식 1을 이용하였을 경우에는, 실측치 2007년 8월 29일에 1,687 ton (유량: 40.344 m³/s, 부유물질농도: 484 mg/L), 2010년 5

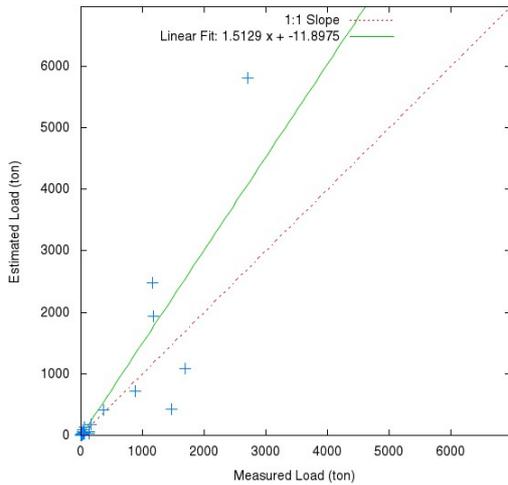
Table 5 Regression model which provided the highest NSE for each station

Name	Model number	Name	Model number
Ahnyang A	8	Hangang I	4
Bukhan A	6	Hantan A	8
Bukhan B	4	Hantan B	5
Bukhan C	8	Inbuk A	5
Bukhan D	1	Imjin A	4
Bokha A	1	Imjin B	5
Cheongmi A	4	Jaecheon A	3
Dalcheon A	8	Jojong A	1
Dalcheon B	7	Jeungrang A	4
Heukcheon A	4	Jucheon A	1
Kyungan A	1	Jinwi A	1
Kyungan B	1	Moosan A	5
Golji A	8	Ohdae A	5
Kapyeong A	4	Pyeongchang A	5
Kongreung A	1	Shincheon A	1
Gulpo A	1	Seomgang A	8
Hongcheon A	1	Seomgang B	7
Hangang A	8	Soyang A	1
Hangang B	1	Soyang B	4
Hangang C	5	Tancheon A	7
Hangang D	1	Okdong A	7
Hangang E	2	Wangsuk A	8
Hangang F	4	Yanhwa A	5
Hangang G	4	Youngpyeong A	5
Hangang H	8		

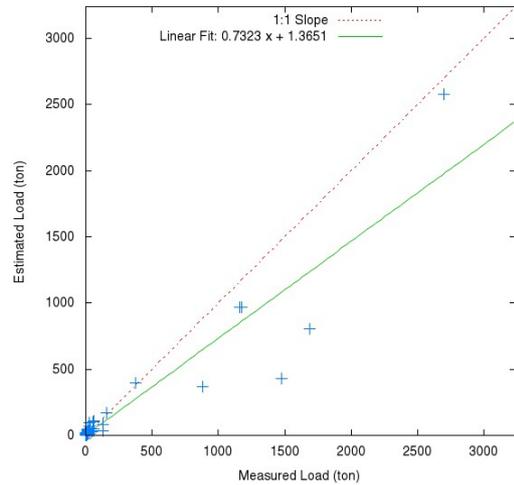
월 18일에 884 ton (유량: 26.236 m³/s, 부유물질농도: 390 mg/L), 2013년 9월 11일에 1,472 ton (유량: 28.778 m³/s, 부유물질농도: 592 mg/L)에 대해서, 각각 801 ton, 365 ton, 432 ton으로 낮게 예측했으나 LOADEST에 의해 선택된 회귀식 6에 의한 모의치보다 실측치에 가깝게 예측하였다 (Fig. 6). 즉, LOADEST에 의한 자동 선택보다는 특정 회귀식을 선택 하였을 때 실측치와 적은 차이를 가지는 모의치를 얻을 수 있었다.

IV. 결 론

본 연구에서는 환경부 물환경정보시스템의 한강 수계에 위치한 49개의 총량지점망의 부유사 자료를 사용하여, LOADEST의 9개의 회귀식 및 LOADEST에 의해 자동선택된 회귀식에



(a) Estimated loads by the regression model numbered 6 through auto-model selection of LOADEST



(b) Estimated loads by the regression model numbered 1

Fig. 6 Scatter plots of measured and estimated suspended solid loads from Ahnyang A station

의한 부유사 모의치를 통해 LOADEST 모형의 거동 평가가 이루어졌다. 49개의 지점에서 수집된 수질자료를 회귀식의 보정 및 검정을 위한 자료로 나누었으며, 또한 보정 및 검정에 사용된 자료를 교차하여 다시 보정 및 검정 자료로 활용하고, LOADEST의 회귀식 번호를 0번부터 9번까지 변환하면서, 총 980회의 부유사 예측이 이루어졌다.

9개의 회귀식의 복잡성과 실측치의 차이에는 경향이 없었다. 즉, 많은 항을 고려하는 것이 실측치에 가까운 모의치를 나타내지는 않았다. 그러나, 로그로 치환된 유량의 제공항의 유무에 따라 회귀식에 의한 모의치와 실측치에 대한 오차에는 차이가 있었으며, 로그로 치환된 유량의 제공항을 포함하지 않은 회귀식에 의한 모의치가 로그로 치환된 유량의 제공항을 포함한 회귀식에 의한 모의치보다 실측치와의 차이가 더 작은 것으로 나타났다. 특히, LOADEST의 회귀식 1과 3이 다른 회귀식에 비해 실측치가 가까운 모의치를 모의하는 것으로 나타났다. LOADEST에 의해 자동 선택된 회귀식은 최적회귀모형 (best-fit regression mode)으로 여겨지는 반면, 자동 선택된 회귀식이 최적회귀모형일 가능성은 NSE와 R² 및 5개의 백분위수를 기준으로 평가할 시 10 % 내외인 것으로 나타났다. 즉, LOADEST를 이용하여 부유사를 예측할 때, 모형의 자동선택에 의한 회귀식 보다는, 회귀식 1 혹은 3을 사용하는 것이 실측치에 더 가까운 모의치가 가능할 것으로 판단된다.

또한 LOADEST의 회귀식의 보정 결과 보다 검정 결과에서 넓은 백분위수의 범위를 보였기 때문에, 보정 시 모의치가 실측치에 가깝다고 해서 미세측지점에서의 모의치와 실측치의 차이가 작다고 보기는 힘든 것으로 판단된다. 따라서 회

귀식에 의한 부유사를 예측할 시, 실측 자료를 나누어 보정 및 검정 과정을 수행한 뒤, 보정 및 검정 결과가 만족스러운 것으로 판단될 때, 모의치를 신뢰할 수 있을 것으로 판단된다.

LOADEST는 9개의 회귀식을 자동으로 선택할 수 있는 기능을 제공하나, 자동으로 선택된 회귀식이 실측치와 가장 작은 차이를 가진다고 보기는 힘든 것으로 생각되기 때문, 사용자에게 의해 회귀식이 선택될 필요가 있는 것으로 판단된다. 본 연구 결과에서는 회귀식 1과 3이 다른 회귀식에 비해 작은 차이를 보였다. 따라서 유역에 LOADEST를 적용할 때, 회귀식 1과 3에 대한 보정 및 검정과정을 수행하고, 만약 실측치와 큰 차이를 보일 경우 다른 회귀식에 대한 보정 및 검정과정을 수행할 필요가 있는 것으로 판단된다.

사 사

이 논문은 2014년도 농림축산식품부와 강원도의 재원으로 강원도농촌활성화지원센터의 지원을 받아 수행된 연구 사업임 (2014GWRRSC03)

REFERENCES

1. Arnold C. L. and C. J. Gibbons, 1996. Impervious surface coverage: the emergence of a key environmental indicator. *Journal of the American Planning Association* 62: 243-257.
2. Arnold, J. G., R. Srinivasan, R. S. Muttiah, and J. R. Williams, 1998. Large area hydrologic modeling and assessment - Part 1:

- Model development. *Journal of the American Water Resources Association* 34(1): 73-89.
3. Carey, R. O., K. W. Migliaccio, and M. T. Brown, 2011. Nutrient discharges to Biscayne Bay, Florida: trends, loads, and a pollutant index. *Science of Total Environment* 409: 530-539.
 4. Cohn, T. A., D. L. Caulder, E. J. Gilroy, L. D. Zynjuk, and R. M. Summers, 1992. The validity of a simple statistical model for estimating fluvial constituent loads: an empirical study involving nutrient loads entering Chesapeake Bay. *Water Resources Research* 28(9): 2353-2463.
 5. Crawford, C. G., 1991. Estimation of suspended-sediment rating curve and mean suspended-sediment loads. *Journal of Hydrology* 129: 331-348.
 6. Dornblaser, M. M. and R. G. Striegl, 2009. Suspended sediment and carbonate transport in the Yukon river basin, Alaska: Fluxes and potential future responses to climate change. *Water Resource Research* 45, W06411, doi:10.1029/2008WR007546.
 7. Haith, D. A., R. Mandel, and R. S. Wu, 1992. GWLF, generalized watershed loading functions, version 2.0, user's manual; Dept. of Agricultural & Biological Engineering, Cornell University: Ithaca, NY, USA.
 8. Jha, B. and M. K. Jha, 2013. Rating Curve Estimation of Surface Water Quality Data Using LOADEST. *Journal of Environmental Protection* 4: 849-856.
 9. Johnson, R. C., H. S. JIN, M. M. Carreiro, and J. D. Jack, 2013. Macroinvertebrate community structure, secondary production and trophic-level dynamics in urban streams affected by non-point-source pollution. *Freshwater Biology* 58(5): 843-857.
 10. Jung, Y. H., J. P. Moon, S. O. Lee, and Y. S. Park, 2014. Assessment of the effect of urbanization on the watershed streamflow. *Journal of the Korean society of agricultural engineers* 56(1): 51-59 (in Korean).
 11. Lee G., Y. Shin, and Y. Jung. 2014 Development of Web-Based RECESS Model for Estimating Baseflow Using SWAT. *Sustainability* 6(4): 2357-2378.
 12. Nam, W. H., E. M. Hong, and J. Y. Choi, 2014. Uncertainty of water supply in agricultural reservoirs considering the climate change. *Journal of the Korean Society of Agricultural Engineers* 56(2): 11-23 (in Korean).
 13. Oh, J. and A. Sankarasubramanian, 2011. Interannual hydro-climatic variability and its influence on winter nutrients variability over the southeast United States. *Hydrology and Earth System Sciences* 8, 10935-10971.
 14. Oh, J., T. Sinha, and A. Sankarasubramanian, 2013. The role of retrospective weather forecasts in developing daily forecasts of nutrient loadings over the Southeast US. *Hydrology and Earth System Sciences Discussions* 10: 15625-15657.
 15. Park, Y. S., 2014. Estimation of pollutant load using genetic-algorithm and regression model. *Korean Journal of Environmental Agriculture* 33(1): 37-43.
 16. Park, Y. S. and B. A. Engel, 2014. Use of pollutant load regression models with various sampling frequencies for annual load estimation. *Water* 6: 1685-1697.
 17. Phillips, J. M., B. W. Webb, D. E. Walling, and L. Leeks, 1999. Estimating the suspended sediment loads of rivers in the LOIS study area using infrequent samples. *Hydrological Processes* 13: 1035-1050.
 18. Preston, S. D., V. J. Bierman, and S. E. Silliman, 1989. An evaluation of methods for the estimation of tributary mass loads. *Water Resource Research* 25: 1379-1389.
 19. Robertson, D. M., 2003. Influence of different temporal sampling strategies on estimating total phosphorus and suspended sediment concentration and transport in small streams, *Journal of the American Water Resources Association* 39: 1281-1308.
 20. Runkel, R. L., C. G. Crawford, T. A. Cohn, 2004. Load Estimator (LOADEST): A Fortran Program for Estimating Constituent Loads in Streams and Rivers; U.S. Geological Survey Techniques and Methods: Reston, VA, USA.
 21. Shuster W. C., J. Bonta, H. Thurston, E. Warnemuende, and D. R. Smith, 2005. Impacts of impervious surface on watershed hydrology: a review. *Urban Water Journal* 2, 263-275.
 22. Sun, C., Z. Shen, R. Liu, M. Xiong, F. Ma, O. Zhang, Y. Li, and L. Chen, 2013. Historical trend of nitrogen and phosphorus loads from the upper Yangtze River basin and their responses to the Three Gorges Dam. *Environmental Science and Pollution Research*, 20(12): 8871-8880.
 23. USEPA, 2001 Better Assessment Science Integrating Point and Nonpoint Sources (BASINS) v. 3.0 User's Manual; U. S. Environment Protection Agency: Washington, D.C., USA.