



# 머신러닝 기법을 활용한 유황별 LOADEST 모형의 적정 회귀식 선정 연구: 낙동강 수계를 중심으로

## Study of Selection of Regression Equation for Flow-conditions using Machine-learning Method: Focusing on Nakdonggang Waterbody

김종건\* · 박윤식\*\* · 이서로\* · 신용철\*\*\* · 임경재\* · 김기성\*,†

Kim, Jonggun · Park, Youn Shik · Lee, Seoro · Shin, Yongchul · Lim, Kyoung Jae · Kim, Ki-sung

### Abstract

This study is to determine the coefficients of regression equations and to select the optimal regression equation in the LOADEST model after classifying the whole study period into 5 flow conditions for 16 watersheds located in the Nakdonggang waterbody. The optimized coefficients of regression equations were derived using the gradient descent method as a learning method in Tensorflow which is the engine of machine-learning method. In South Korea, the variability of streamflow is relatively high, and rainfall is concentrated in summer that can significantly affect the characteristic analysis of pollutant loads. Thus, unlike the previous application of the LOADEST model (adjusting whole study period), the study period was classified into 5 flow conditions to estimate the optimized coefficients and regression equations in the LOADEST model. As shown in the results, the equation #9 which has 7 coefficients related to flow and seasonal characteristics was selected for each flow condition in the study watersheds. When compared the simulated load (SS) to observed load, the simulation showed a similar pattern to the observation for the high flow condition due to the flow parameters related to precipitation directly. On the other hand, although the simulated load showed a similar pattern to observation in several watersheds, most of study watersheds showed large differences for the low flow conditions. This is because the pollutant load during low flow conditions might be significantly affected by baseflow or point-source pollutant load. Thus, based on the results of this study, it can be found that to estimate the continuous pollutant load properly the regression equations need to be determined with proper coefficients based on various flow conditions in watersheds. Furthermore, the machine-learning method can be useful to estimate the coefficients of regression equations in the LOADEST model.

**Keywords:** LOADEST model, Machine-learning, Pollutant load, Flow conditions, Regression equation

### 1. 서 론

하천에는 청천시 연중 일정량의 기저유량과 하수처리장과 같은 시설물에서의 배출수가 유입되고 있으며 이로 인한 기저유출에 대한 비점오염원과 시설물에서의 점오염원이 유입되고 있다. 강우시 강우 유출수에 의한 비점오염물질이 하천으로 유입되어 수질오염원의 주요 원인이 되고 있다 (Lee et al., 2014). 또한, 기후변화로 인한 이상 기후에 따른 홍수 및

가뭄의 발생은 하천의 수량과 오염물질 발생에 영향을 미치고 있다. 현재 국내에서는 수질오염총량관리제를 시행하여 시, 군 구간별로 목표수질을 설정하고, 이를 달성하기 위해 허용부하량을 산정해 오염물질 배출량을 규제하는 데 노력을 기울이고 있다 (ME, 2004). 4대강 사업 이후 악화된 수질을 관리하기 위해 낙동강 수계에 대해서는 수계 내 지류를 대상으로 지류총량제를 적용하여 오염물질의 근본 원인을 파악하고 이를 저감할 수 있는 방안을 수립하고 있다. 지류총량제를 비롯한 수질오염총량관리제에서는 유역 내 오염부하량을 통해 수질오염 기준을 마련하고 수질 관리 대책을 수립하고 있다. 오염부하량은 유역 내 배출되는 유량과 그 유량 속에 포함되어 있는 오염물질 농도의 곱으로 산정되며, 오염부하 특성 분석을 통해 주요 오염원 파악 및 오염원의 기여도 분석이 가능하다. 이러한 오염부하 특성은 토지이용도, 유역 크기 및 지형 등 유역의 특성에 따라 다양하게 나타날 수 있다. 뿐만 아니라, 유황별 오염부하 특성을 통해 점오염원 및 비점오염원 등의 주요 오염원 파악이 가능하다. 그리하여 오염부하량의 정확한 산정은 유역 내 하천의 수질오염 분석에 있어 중요한 부

\* Department of Regional Infrastructures Engineering, Kangwon National University

\*\* Department of Rural Construction Engineering, Kongju National University

\*\*\* School of Agricultural Civil and Bio-Industrial Engineering, Kyungpook National University

† Corresponding author

Tel.: +82-33-250-6466 Fax: +82-33-259-5560

E-mail: [kskim@kangwon.ac.kr](mailto:kskim@kangwon.ac.kr)

Received: April 24, 2017

Revised: July 19, 2017

Accepted: July 20, 2017

분을 차지하고 있다. 또한, 유역 내 하천에서의 유량 및 수질 자료의 확보는 정확한 오염부하량 산정에 있어 큰 영향을 미치게 된다. 우리나라 수질오염 총량 지점에서는 8일 간격으로 유량 및 수질을 측정하고 있으며, 이를 활용하여 유황구간별 부하량을 평가한 연구가 다수 진행되었다 (Kim et al., 2007; NIER, 2013). 수질자료의 경우 다수의 결측으로 인해 충분한 자료 확보가 어려운 경우가 발생하고 있다.

비연속성을 가지는 수질자료의 제한성을 극복하기 위해 수문수질 모형을 활용한 모형의 검보정을 수행하고 이를 통해 연속적인 수질모의가 가능하다. 수문수질 모형에는 Soil and Water Assessment Tool (SWAT; Arnold et al., 1998), Hydrologic Simulation Program-Fortran (HSPF; US Environmental Protection Agency (USEPA), 2001), Generalized Watershed Loading Function (GWLF; Haith et al., 1992), Load Estimator (LOADEST; Runkel et al., 2004) 등이 널리 사용되고 있다. 이 중 LOADEST 모형은 최적 회귀모형으로 하천유량과 수질의 관계를 이용하여 결측된 수질자료를 모의할 수 있다. 또한, LOADEST 모형은 모형 내 다양한 회귀식을 제공하고 있으며, 유량뿐만 아니라 시계열 자료를 추가하여 연중 시간적 특성 반영이 가능하다. 이러한 이유로 국내외에서 LOADEST 모형을 활용한 수질예측 및 오염부하량 산정 연구가 다수 수행되어왔다 (Shin et al., 2008; Oh et al., 2013; Jha and Jha, 2013; Park and Engel, 2014; Park et al., 2015). 그러나 기존 LOADEST 모형을 활용한 연구에서는 전체 유량 자료에 대한 단일 회귀식만을 적용하여 수질 및 오염부하량을 산정하였다. 위에서 언급하였듯이 전체 유량 자료에 대해 유황 구간별 오염부하 특성이 상이하게 나타날 수 있다. 청천시와 강우시 유량이 조건이 달라지고 점오염원 및 비점오염원에 따른 수질 농도도 달라질 수 있다. 따라서 유황별 LOADEST 모형 내 적정 회귀식의 적용이 필요할 것으로 사료된다.

최근 머신러닝 (기계학습) 기법이 부각되면서 다양한 분야에서 적용되고 있다. 머신러닝은 컴퓨터가 학습할 수 있도록 하는 알고리즘과 기술을 개발하는 분야로 경험적인 데이터로부터 스스로 성능을 향상시키는 기술이다 (Marcin et al., 2016). 머신러닝의 다양한 알고리즘 중 최적화 알고리즘은 데이터를 학습하는 모형의 비용함수의 값을 최소화하는 방법으로 기존 알고리즘에 비해 보다 정확한 최적화가 가능하다.

따라서, 본 연구의 목적은 낙동강 수계 내 총량지점을 대상으로 머신러닝의 최적화 알고리즘을 활용한 유황별 부유사 예측에 대한 LOADEST 모형 내 적정 회귀식을 선정하고 적용성을 평가하는데 있다.

## II. 방 법

### 1. 연구대상지역

본 연구에서는 국내 4대강 주요 수계 중 낙동강 수계를 대상으로 머신러닝 기법을 활용한 유황별 LOADEST 모형 내 적정 회귀식 선정 연구를 수행하였다. 낙동강 수계에서는 수질오염총량제가 시행되어오고 있으며, 낙동강 수계 내 연속적인 오염부하량 예측 및 산정으로 정확한 오염총량의 평가가 필요하다. 그리하여 낙동강 수계 내에 위치한 16개의 총량

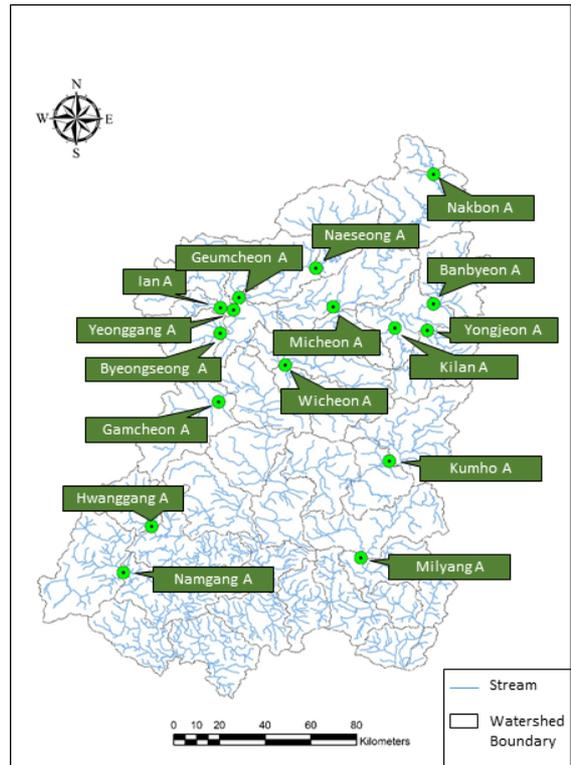


Fig. 1 Locations of streamflow and total suspended solid data in Nakdonggang waterbody

Table 1 Max. and Min. suspended solid data for 16 stations in Nakdonggang waterbody (mg/l)

Watersheds	Min	Max	Watersheds	Min	Max
Gamcheon A	1,2	203	Milyang A	0,2	131
Geumcheon A	0,2	504	Banbyeon A	0,1	79
Kumho A	0,9	260	Byeongseong A	0,8	356
Kilan A	0,1	209	Yeonggang A	0,3	283
Nakbon A	0,2	266	Yongjeon A	0,2	105
Namgang A	0,3	396	Wicheon A	0,2	505
Naeseong A	0,4	420	Ian A	0,3	393
Micheon A	0,2	122	Hwanggang A	0,3	373

측정지점에 대해 물환경정보시스템 (water.nier.go.kr)에서 제공하고 있는 유량 및 수질 자료를 수집하여 본 연구에 활용하였다. 각 측정별 2007년부터 2015년까지의 8일 간격 유량과 부유물질(SS) 실측자료를 수집하였다. Fig. 1은 본 연구의 대상지역 내 16개 총량측정 지점을 보여주고 있으며 Table 1은 각 지점에 대해 수집된 부유사의 최대, 최소 농도를 보여주고 있다.

## 2. LOADEST 모형의 개요

LOADEST 모형은 하천이나 강의 유량 및 수질에 대한 시계열 입력 자료를 활용하여 통계적인 분석을 통해 수질 미계측 날짜에 대한 오염 부하량을 예측하거나 특정 기간에 대한 연속적인 오염 부하량 예측을 위해 개발되었다(Runkel et al., 2004). 본 모형은 11가지의 회귀식을 제시하고 있으며, 유역의 특성에 맞게 유량과 수질 자료의 상관관계를 가장 잘 나타내고 있는 회귀식을 선정하여 연속적인 오염부하량을 산정한다. LOADEST 모형은 시계열 입력 자료를 기반으로 다양한 통계 분석을 통해 미계측 및 특정 기간에 대한 오염 부하량 예측이 가능하여 기존 다른 오염 부하량 산정 방법에 비해 보다 객관적이고 정확한 예측이 가능한 장점이 있다. 아래 식 (1)~(11)은 모형에서 제시하고 있는 11개의 회귀식을 보여주고 있다. 식 1번과 2번의 경우 유량에 대한 독립변수만을 고려한 회귀식으로 유량과 수질자료간의 상관관계가 클 경우 적용 가능하다. 반면, 식 3번부터 9번의 회귀식은 유량 자료와 더불어 연중 환산 시간을 고려하여 계절적인 특성 반영이 가능하다. 또한, LOADEST 모형에서는 9개의 회귀식 중 유사한 경향의 회귀식이 선정된다면 적용된 계수가 적은 회귀식(보통 1~3번)을 우선적으로 선정하는 내부 프로세스를 반영하고 있다.

연속적인 오염부하량 산정을 위한 회귀식은 유역의 특성에 따라 상이하게 적용될 수 있으며, 보다 정확한 예측을 위해서는 유역의 유황특성을 반영한 회귀식 산정이 필요하다. 그리하여 본 연구에서는 부유물질(SS)에 대한 연속적인 오염 부하량 산정을 위해 머신러닝 기법을 통한 유황별 적정 회귀식 선정 연구를 수행하였다. 특정기간에 대한 모의를 위해 적용될 수 있는 식 (10)번과 (11)번을 제외하고 식 (1)~식 (9)까지의 회귀식에 대해 적용하였다.

$$\ln y = a_0 + a_1 \ln Q \quad (1)$$

$$\ln y = a_0 + a_1 \ln Q + a_2 \ln Q^2 \quad (2)$$

$$\ln y = a_0 + a_1 \ln Q + a_2 dtime \quad (3)$$

$$\ln y = a_0 + a_1 \ln Q + a_2 \sin(2\pi dtime) + a_3 \cos(2\pi dtime) \quad (4)$$

$$\ln y = a_0 + a_1 \ln Q + a_2 \ln Q^2 + a_3 dtime \quad (5)$$

$$\ln y = a_0 + a_1 \ln Q + a_2 \ln Q^2 + a_3 \sin(2\pi dtime) + a_4 \cos(2\pi dtime) \quad (6)$$

$$\ln y = a_0 + a_1 \ln Q + a_2 \sin(2\pi dtime) + a_3 \cos(2\pi dtime) + a_4 dtime \quad (7)$$

$$\ln y = a_0 + a_1 \ln Q + a_2 \ln Q^2 + a_3 \sin(2\pi dtime) + a_4 \cos(2\pi dtime) + a_5 dtime \quad (8)$$

$$\ln y = a_0 + a_1 \ln Q + a_2 \ln Q^2 + a_3 \sin(2\pi dtime) + a_4 \cos(2\pi dtime) + a_5 dtime + a_6 dtime^2 \quad (9)$$

$$\ln y = a_0 + a_1 per + a_2 \ln Q + a_3 \ln Q per \quad (10)$$

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

여기서,  $y$ 는 오염부하량,  $Q$ 는 유량 ( $\ln Q$ 는 유량에서 중앙값을 뺀 값 후 로그를 취한 값)을 의미한다.  $dtime$ 은 연중 시간을 0~1의 소수로 환산한 값에서 중앙값을 뺀 값으로 연중 시간적 특성을 고려할 수 있는 변수이다 (decimal time - center of decimal time),  $per$ 은 사용자가 정의한 기간,  $a_0 \sim a_6$ 는 회귀식 계수를 의미한다. 회귀식에 사용된  $\sin$ 과  $\cos$  함수는 계절에 따른 변화 주기를 고려하기 위해서 적용되었다.

## 3. 머신러닝 기법을 활용한 유황별 적정 회귀식 선정

### 가. 텐서플로를 활용한 머신러닝 기술

머신러닝 기법은 주어진 입력 자료로부터 새로운 지식을 발견하고 습득하여 새로운 지식을 도출하는 알고리즘이다. 텐서플로(TensorFlow)는 머신러닝 엔진으로 구글에서 오픈 소스로 제공되고 있는 라이브러리이다. 텐서플로는 새로운 모형을 도출하고 수많은 정보를 학습하고 처리할 수 있는 능력을 가지고 있다. 초창기 연구자들에게 광대한 지식을 통해 정보를 간소화하고 일반화하고자 하는 요구에서 시작되어 현재 보다 빠른 학습 능력을 갖추고 새로운 기술 개발을 위해 널리 활용되고 있다 (Dean et al., 2012). 텐서플로는 개별적인 연산과정, 매개변수 업데이트, 입력자료의 사전처리과정을 포함한 모든 연산과정 처리를 위해 Fig. 2에서 보이는 바와 같

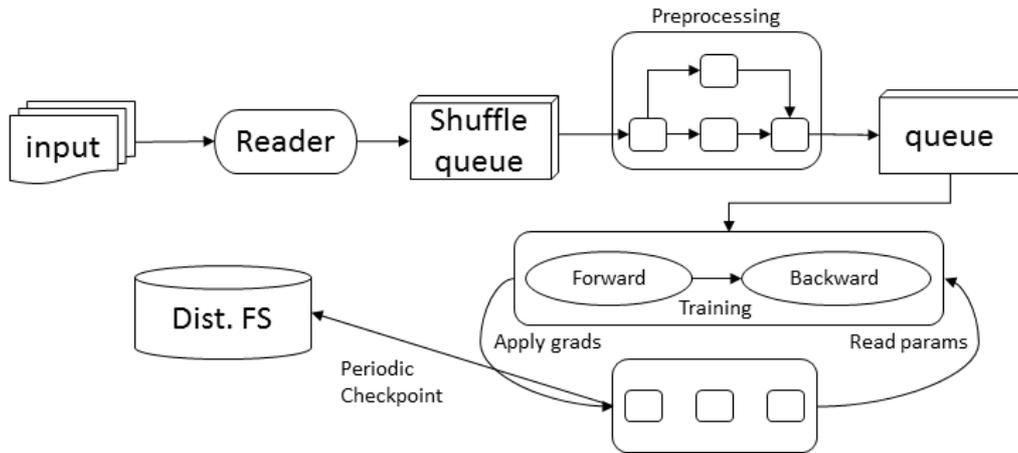


Fig. 2 TensorFlow dataflow graph

이 단일 데이터 흐름 그래프를 사용한다.

머신러닝에서 최적화 알고리즘은 데이터를 학습하는 모델의 비용함수의 값을 최소화하는 방법이다. 모델의 비용함수의 값은 학습 모델의 결과와 실제 결과 값의 차이 (error)를 의미한다. 비용함수의 최소화를 위해 기울기를 구하고 오류가 적어지는 방향을 따라 점진적으로 파라미터 값을 업데이트 해주어 찾는 방법이다. 본 연구에서는 머신러닝 엔진인 텐서플로의 다양한 학습방법 중 경사하강법을 활용하여 LOADEST 모형 내 다양한 회귀식을 분석하고 최적 회귀식 계수를 산정하였다. 경사하강법은 반복적으로 계수를 변경·적용하여 손실함수가 최소가 되도록 학습하는 알고리즘이다. Fig. 2에서 보여주는 텐서플로 알고리즘에서 기존 일차 함수가 아닌 LOADEST 모형 내 회귀식으로 적용하여 손실함수가 최소일 때의 변수를 찾도록 수정하였다. 따라서 텐서플로의 경사하강법을 적용하여 낙동강 수계 내 16개 유역을 대상으로 부유물질 (SS)에 대한 연속적인 오염부하량 산정을 위해 유황별 적정 오염부하량 회귀식 선정 및 해당 계수를 산정하였다.

나. 유황곡선 (FDC)를 이용한 유역별 유황구간 선정

오염부하 특성은 유역의 특성에 따라 상이하게 나타날 수 있으며, 유황의 변화(갈수기, 저수기, 평수기, 풍수기, 홍수기)는 정확한 오염부하 예측을 위해 중요한 유역 특성 중 하나이다. 유황곡선 (Flow Duration Curve)은 유역 내 유황변화 분석에 널리 활용되고 있으며, 분할된 유황구간을 통해 해당 유역의 특성과 기후 경향의 설명이 가능하고 특정 유역에 대한 유량의 크기 및 빈도 분석이 가능하다. 본 연구에서 유황구간은 연속적인 시계열 유량 자료에 대해 고유량에서부터 저유량까지 재배열 후 전체 구간을 0~10% (홍수기), 10~40% (풍수기), 40~60% (평수기), 60~90% (저수기), 90~100% (갈

수기)로 분할하여 유황구간별 오염부하 예측을 위한 회귀식을 판별하였다. 유황곡선 산정 방법은 아래 식과 같다.

$$\begin{aligned} & \text{Percent of Days Flow Exceeded}(\%) \\ & = \text{Rank} / \text{Number of Data} \times 100 \end{aligned} \tag{12}$$

다. 회귀식 검증을 위한 통계분석

머신러닝 기법을 활용한 유황별 선정된 LOADEST 모형의 회귀식 검증을 위해 식 (13)부터 (15)까지의 3가지 통계분석 방법을 적용하였다. NSE (Nash-Sutcliffe model Efficiency coefficient)와 R2 (coefficient of determination)는 1에 가까울수록 예측치가 실측치와 상관관계가 높다는 것을 의미하며, RMSE (Root Mean Square Error)는 0에 가까울수록 상관관계가 높다는 것을 의미한다.

$$NSE = 1 - \frac{\sum_{i=1}^n (S_i - O_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2} \tag{13}$$

$$R^2 = \left( \frac{\sum_{i=1}^n (O_i - \bar{O})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^n (O_i - \bar{O})^2} \sqrt{\sum_{i=1}^n (S_i - \bar{S})^2}} \right)^2 \tag{14}$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^n (S_i - O_i)^2} \tag{15}$$

여기서, N은 자료의 개수, S는 예측치, O는 실측치,  $\bar{S}$ 와  $\bar{O}$ 는 예측치와 실측치의 평균값을 의미한다.

### III. 결과 및 고찰

#### 1. 유황곡선(Flow duration curve)을 이용한 유황구간 분할

본 연구에서는 유황별 부유물질(SS)에 대한 오염부하량 산정 회귀식 및 해당 계수 산정을 위한 기본 입력 자료로 낙동강 수계 내 16개의 총량지점을 대상으로 과거 9년치의 유황 자료를 확보하였으며, 이를 5개의 유황구간으로 분할하였다. 유황구간으로는 0~10% (홍수기), 10~40% (풍수기), 40~60% (평수기), 60~90% (저수기), 90~100% (갈수기)로 설정하였으며 Table 2에서는 유역별 각 유황구간에 대한 평균 유량을 보여주고 있다. 유역별 유량의 차이는 유역 및 하천의 크기와 연관될 수 있으며, 특히 고유량(High flow)에서는 유역 내 강우 발생과 직접적인 연관이 있다. 유황구간별 평균 유량을 비교한 결과 유역마다 차이를 보이고 있으며, 특히 평수기, 저수기, 갈수기에 비해 홍수기와 풍수기에 유량 변동이 크게 나타나는 것을 알 수 있다. 이는 우리나라 기후 특성상 여름철 집중되는 강우 패턴과 하상계수가 큰 유역 특성으로 인해 발생될 수 있다. Fig. 3은 16개의 대상유역에 대한 유황곡선을 보여주고 있다. 그림에서 나타난 바와 같이 길안 A, 미천 A, 반변 A, 용전 A, 위천 A, 이안 A 유역의 경우 고유량에서 저유량으로 갈수록 유황곡선이 급격히 변화하면서 유량의 변동이 크게 나타난 것을 알 수 있다. 이러한 유황구간별 큰 유량 변동은 오염부하량 산정에 직접적인 영향을 미칠 수 있기 때문에 오염부하 특성 분석시 유황구간에 따른 분할 분석이 필요하다. 또한, 유황구간 분할을 통해 유역 내 발생하는 점오염 및 비점오염의 영향 평가가 가능하며 향후 오염부하량 산정 모형의 개선 방향에 대한 모색이 가능할 것이다. 따라서 본 연구에서는 유역별 유황구간에 따른 LOADEST 모형을 활용한 오염부하량 산정 회귀식을 평가하였다.

#### 2. LOADEST 모형을 이용한 유역별 회귀식 선정

LOADEST 모형은 위에서 언급한 11개 회귀식(특정기간에 대한 적용 회귀식 2개 포함)을 통해 오염부하량을 산정한다. 사용자의 정의에 따라 9개 중 특정 회귀식을 이용하여 오염부하량을 산정할 수 있으며, 뿐만 아니라 9개의 회귀식 분석을 통해 최적의 회귀식 도출이 가능하다. 본 연구에서는 수집된 9년간의 유량과 수질 자료를 이용하여 LOADEST 모형을 통해 전체 모의 기간에 대한 최적의 회귀식을 추출하였다. Table 3은 16개의 대상 유역에 대해 LOADEST 모형을 통해 선정된 최적 회귀식과 실측 자료와의 상관계수를 보여 주고 있다. 분석 결과 16개의 유역에서 다양한 회귀식(3, 6, 8, 9번 회귀식)이 최적의 회귀식으로 선정되었으며, 상관계수( $R^2$ )는 0.55~0.92로 나타났다. 이 중 7개의 계수(coefficient)를 가지는 9번 회귀식이 주로 높은 모의 상관성을 나타내는 것으로 분석되었다. 9번 회귀식의 경우 다른 회귀식에 비해 많은 계수를 가지고 있는 만큼 다양한 오염부하 패턴을 고려할 수 있는 장점이 있다. 반면, 금촌 A 유역의 경우  $R^2$  0.82로 3번 회귀식이 선정되었다. 이는 LOADEST 모형에서 유사한 경향의 회귀식이 있을 경우 적용 계수가 적은 회귀식을 선정하는 내부 프로세스로 인한 영향을 받은 것으로 판단될 수 있다.

본 연구 결과를 통해 LOADEST 모형에서 선정된 회귀식은 높은  $R^2$ 로 그 정확성이 높은 것으로 볼 수 있으나, 이는 특정 기간 고유량에 의해 높게 산정되는 오염부하량에 크게 영향을 받을 수 있다(Kang et al., 2010). 위에서 언급한 바와 같이 16개의 대상 유역은 유황구간 별 큰 유량 변동을 보이고 있기 때문에 이는 오염부하량 산정에도 큰 영향을 줄 수 있을 것으로 판단된다. 그리하여 오염부하량 산정을 위한 최적의 회귀식 선정시 이러한 유황 특성의 고려가 필요할 것으로 판단된다.

Table 2 Mean flow amount for each flow condition in the study watersheds

Watersheds	Hi*	Mo*	Mi*	Lo*	Dr*	Watersheds	Hi*	Mo*	Mi*	Lo*	Dr*
Gamcheon A	73,5	8,3	3,9	2,3	1,0	Milyang A	104,1	8,4	4,1	2,5	0,9
Geumcheon A	25,8	2,8	1,1	0,6	0,3	Banbyeon A	68,7	5,3	1,5	0,7	0,3
Kumho A	87,9	8,1	4,8	3,0	1,3	Byeongseong A	29,4	5,0	2,0	1,2	0,6
Kilan A	40,4	2,9	0,9	0,4	0,1	Yeonggang A	85,7	11,8	4,5	2,5	0,9
Nakbon A	30,4	4,9	2,2	1,4	0,8	Yongjeon A	34,9	3,0	0,8	0,4	0,2
Namgang A	125,8	19,8	9,9	5,8	2,7	Wicheon A	60,3	5,0	2,1	1,0	0,3
Naeseong A	93,4	11,9	5,2	3,3	1,4	Ilan A	19,2	2,6	1,0	0,5	0,2
Micheon A	27,8	2,1	0,6	0,3	0,1	Hwanggang A	85,7	8,3	3,77	2,2	1,0

\*Hi, Mo, Mi, Lo, and Dr mean High, Moist, Mid, Low, and Dry flow conditions, respectively (unit: m<sup>3</sup>/s)

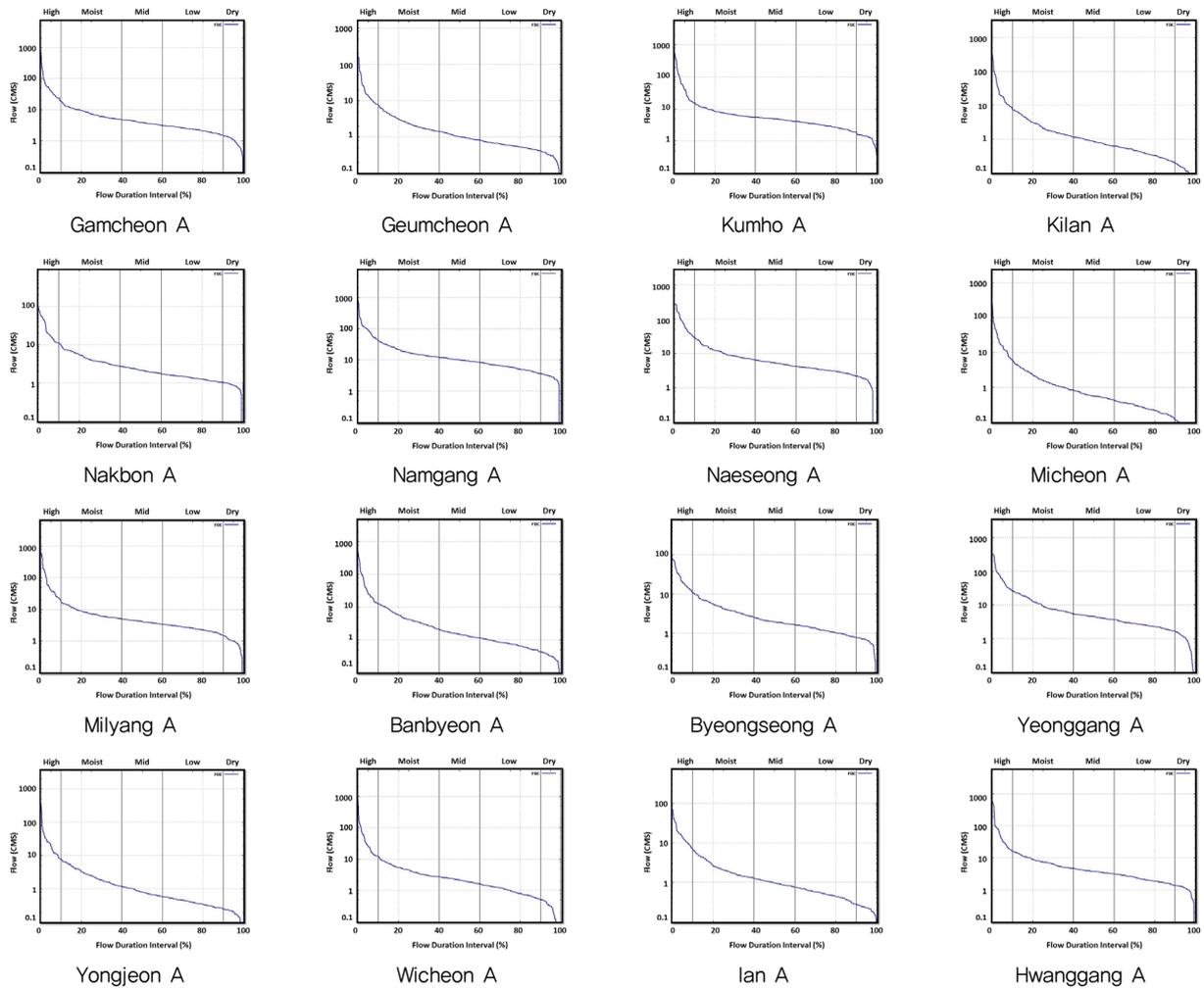


Fig. 3 Flow duration curves for 16 watersheds

Table 3 Selected regression equation and correlation coefficient for 16 watersheds using the LOADEST model

Watershed	# Equation	R <sup>2</sup>	Watershed	# Equation	R <sup>2</sup>
Gamcheon A	6	0.82	Milyang A	9	0.89
Geumcheon A	3	0.82	Banbyeon A	6	0.92
Kumho A	9	0.83	Byeongseong A	9	0.87
Kilan A	8	0.92	Yeonggang A	6	0.83
Nakbon A	9	0.56	Yongjeon A	9	0.87
Namgang A	8	0.55	Wicheon A	9	0.77
Naeseong A	9	0.88	Ian A	8	0.83
Micheon A	9	0.90	Hwanggang A	8	0.86

### 3. 머신러닝 기법을 이용한 유황별 회귀식 및 계수 선정

유황구간별 LOADEST 모형 내 적정 회귀식 선정 및 계수 산정을 위해 최근 널리 활용되고 있는 머신러닝 기법을 적용하였다. 머신러닝 엔진인 텐서플로의 다양한 분석 기법 중 경

사하강법을 활용하여 각 회귀식의 계수를 선정하고 이를 통해 유황구간별 선정된 회귀식을 비교 분석하였다. 경사하강법은 각 회귀식에서의 계수를 반복적으로 업데이트하여 손실 함수가 최소가 되도록 학습하는 기법이다. 본 연구에서 적용

된 머신러닝 기법은 텐서플로의 경사하강법을 통해 계산된 손실함수를 바탕으로 최적의 회귀식 계수를 찾아가는 과정으로 계수의 개수에 따라 달라질 수 있지만 빠른 시간 내 최적의 해를 찾을 수 있는 장점이 있다.

머신러닝 기법을 활용하여 유역별 유황구간별 최적의 회귀식을 도출한 결과 모두 7개의 계수를 가지고 있는 9번 회귀식이 선정되었다. 기존의 LOADEST 모형 적용 방법 결과 9

번 회귀식이 다수 선정된 결과와 유사하게 나타났다. 본 연구에서는 기존 LOADEST 모형 적용 방법과는 달리 유황구간별로 입력 자료를 분할하여 각 유황구간에 대해 적정 회귀식(9번 회귀식)을 도출하였으며, 각 회귀식의 최적 계수를 산정하였다. Table 4는 16개의 유역에 대해 머신러닝 기법 적용을 통해 산정된 유황별 7개의 계수를 보여주고 있다. 표에서 나타난 바와 같이 유황별 산정된 계수는 큰 차이를 보이고 있다.

Table 4 Coefficients of regression equation for flow duration conditions in the study watersheds

Watersheds	Cond.	a <sub>0</sub>	a <sub>1</sub>	a <sub>2</sub>	a <sub>3</sub>	a <sub>4</sub>	a <sub>5</sub>	a <sub>6</sub>	Watersheds	Cond.	a <sub>0</sub>	a <sub>1</sub>	a <sub>2</sub>	a <sub>3</sub>	a <sub>4</sub>	a <sub>5</sub>	a <sub>6</sub>
Gamcheon A	High	12,892	-8,630	5,150	-0,098	0,494	0,097	-0,014	Milyang A	High	11,99	-0,635	1,283	-0,491	-0,414	0,078	-0,022
	Moist	8,917	5,229	-1,725	-0,173	-0,259	0,001	0,002		Moist	7,444	-3,132	2,246	-0,137	0,480	0,039	-0,007
	Mid	7,582	-3,741	2,292	-0,323	0,123	0,012	0,017		Mid	6,580	-5,379	3,943	-0,249	0,301	0,010	-0,011
	Low	5,479	0,853	0,441	0,056	0,282	-0,172	-0,008		Low	4,966	8,675	-3,556	-0,495	0,287	0,018	-0,124
	Dry	6,811	4,391	-1,478	-0,088	-0,027	-0,022	0,031		Dry	5,733	8,600	-3,694	-0,512	0,243	0,028	-0,002
Geumcheon A	High	11,496	-3,766	2,888	0,367	-0,225	0,237	0,042	Banbyeon A	High	10,99	7,171	-2,721	0,006	0,110	0,044	0,028
	Moist	7,352	3,722	-1,148	0,221	0,853	0,038	0,006		Moist	6,890	5,963	-2,364	0,028	1,088	0,038	0,007
	Mid	5,799	0,586	0,825	-0,923	-0,153	0,040	0,008		Mid	5,515	9,709	-4,349	-0,424	1,210	-0,003	-0,002
	Low	5,151	3,043	-1,024	-0,706	0,119	-0,173	-0,067		Low	3,073	6,581	-2,550	0,499	-0,856	0,076	0,064
	Dry	5,227	1,484	-0,325	-0,083	0,828	-0,070	-0,011		Dry	4,934	-5,887	3,356	1,138	0,547	-0,041	-0,014
Kumho A	High	11,668	1,621	0,090	-0,417	-0,238	0,113	0,002	Byeongseong A	High	11,24	9,549	-3,885	-0,389	0,578	0,111	0,019
	Moist	8,069	9,557	-4,049	-0,340	0,250	0,036	0,011		Moist	8,524	4,834	-1,408	-0,342	-0,284	0,078	-0,013
	Mid	7,453	6,888	-2,349	-0,073	0,152	0,034	-0,009		Mid	7,094	2,819	-1,275	-0,487	0,121	0,015	-0,022
	Low	5,496	-6,596	3,822	-0,407	0,007	-0,066	0,021		Low	4,882	-2,403	1,700	-0,163	0,265	-0,023	0,001
	Dry	7,219	-2,815	2,005	-0,153	0,012	0,006	-0,038		Dry	5,967	1,848	-0,061	-0,023	0,428	0,003	-0,005
Kilan A	High	10,307	-0,425	1,073	0,229	0,150	0,058	-0,025	Yeonggang A	High	11,77	8,361	-3,322	-0,482	-0,282	0,068	-0,001
	Moist	5,611	-5,587	3,363	-0,556	0,705	0,098	0,012		Moist	8,242	-0,025	0,492	0,281	0,495	0,019	-0,003
	Mid	4,313	5,457	-1,980	0,524	0,887	0,072	0,016		Mid	7,240	-10,77	6,534	0,395	0,680	0,021	-0,024
	Low	2,134	-2,287	1,733	0,924	0,507	-0,018	0,002		Low	4,712	-2,338	1,687	-0,847	0,254	0,066	0,009
	Dry	3,557	6,875	-2,833	-0,673	0,613	0,006	0,009		Dry	6,627	-4,727	2,694	-0,816	0,090	0,020	-0,006
Nakbon A	High	10,267	-2,292	2,281	0,007	-0,157	-0,057	-0,005	Yongjeon A	High	10,73	6,961	-2,507	0,465	0,556	0,009	0,004
	Moist	6,743	3,834	-1,025	0,048	0,891	-0,011	0,031		Moist	6,325	3,202	-1,160	-0,527	0,149	-0,028	-0,040
	Mid	6,099	-1,363	1,390	-0,476	0,245	-0,016	-0,007		Mid	5,127	0,043	0,344	-0,459	0,610	0,004	-0,024
	Low	4,288	-9,704	6,028	0,048	-0,695	0,018	-0,029		Low	3,728	2,320	-0,519	-0,480	-0,666	-0,019	-0,073
	Dry	5,581	-2,575	1,647	-0,251	0,264	-0,012	-0,010		Dry	4,451	3,998	-1,614	0,168	0,684	0,012	-0,014
Namgang A	High	12,591	-2,568	2,329	-0,446	0,367	0,055	0,025	Wicheon A	High	11,74	-3,065	2,484	-0,173	-0,186	0,046	-0,011
	Moist	8,956	-8,544	5,008	-0,548	-0,183	0,037	-0,008		Moist	7,164	9,703	-4,181	-0,154	0,320	-0,021	0,001
	Mid	7,953	-10,40	5,309	-0,758	-0,271	0,020	-0,012		Mid	6,179	-5,048	3,000	-0,590	0,077	-0,049	-0,004
	Low	6,798	-5,504	2,983	-0,755	-0,599	0,065	0,002		Low	3,247	4,377	-1,705	-0,116	-0,316	-0,090	-0,030
	Dry	7,440	-3,979	2,368	0,188	0,951	0,010	-0,016		Dry	5,236	2,657	-0,768	-0,383	0,243	-0,056	-0,003
Naeseong A	High	11,872	9,076	-3,545	0,333	0,831	0,128	0,013	Ilan A	High	10,35	3,784	-1,034	0,529	-0,252	0,168	0,004
	Moist	9,628	1,062	0,441	-0,453	-0,044	0,015	-0,002		Moist	6,300	9,409	-4,022	0,432	0,788	0,030	0,047
	Mid	8,084	7,711	-2,892	-0,306	-0,327	0,003	-0,012		Mid	5,486	-1,966	1,632	1,047	0,458	0,051	0,003
	Low	6,187	-0,631	0,853	-0,121	0,271	-0,023	0,025		Low	2,565	-0,143	0,278	1,531	-0,489	0,426	0,026
	Dry	7,155	-1,464	1,608	0,108	-0,279	-0,046	-0,011		Dry	4,762	3,428	-1,030	-1,080	0,016	0,023	-0,005
Micheon A	High	10,938	-2,167	1,886	0,207	0,295	0,030	-0,055	Hwanggang A	High	12,48	2,270	-0,126	-0,984	0,084	0,081	0,034
	Moist	6,839	2,474	-0,738	-0,769	-0,307	-0,001	-0,053		Moist	8,109	6,485	-2,559	-0,344	0,190	0,063	0,008
	Mid	5,610	7,480	-3,232	-0,630	0,680	0,060	-0,053		Mid	7,132	-4,344	2,527	-0,588	0,273	0,073	-0,014
	Low	2,789	3,760	-1,259	-0,714	-0,523	-0,131	-0,091		Low	5,117	0,222	0,538	1,008	-0,274	0,010	0,009
	Dry	4,503	7,903	-3,390	0,863	0,123	0,064	-0,039		Dry	6,520	1,969	-0,550	-0,476	0,446	0,113	-0,006

이를 통해 유황구간별 오염부하 특성이 상이하게 나타날 수 있으며, 전체 모의 기간에 대해 동일한 계수의 적용은 정확한 오염부하 산정에 있어 다소 불확실성을 내포할 수 있을 것으

로 판단된다. 동일한 회귀식(9번 회귀식)을 적용한다고 할지라도 적용된 계수의 차이는 오염부하 산정에 큰 영향을 미칠 수 있다. Fig. 4는 남강 A 유역을 대상으로 산정된 유황구간에

**Table 5** Comparisons of simulated and observed SS load for each flow condition in the study watersheds

Watershed		R <sup>2</sup>	NSE	RMSE (kg)	Watershed		R <sup>2</sup>	NSE	RMSE (kg)
Gamcheon A	High	0.83	0.83	0.60	Milyang A	High	0.84	0.84	0.80
	Moist	0.44	0.44	0.74		Moist	0.47	0.46	0.61
	Mid	0.11	0.11	0.70		Mid	0.31	0.31	0.58
	Low	0.46	0.46	0.65		Low	0.34	0.34	0.60
	Dry	0.17	0.17	0.71		Dry	0.36	0.35	0.53
Geumcheon A	High	0.65	0.65	1.08	Banbyeon A	High	0.90	0.90	0.51
	Moist	0.38	0.37	0.95		Moist	0.48	0.48	0.75
	Mid	0.46	0.46	0.74		Mid	0.34	0.34	0.80
	Low	0.12	0.12	0.81		Low	0.31	0.30	0.93
	Dry	0.13	0.13	0.97		Dry	0.42	0.42	0.86
Kumho A	High	0.88	0.88	0.63	Byeongseong A	High	0.64	0.64	0.77
	Moist	0.33	0.33	0.60		Moist	0.55	0.55	0.80
	Mid	0.09	0.09	0.73		Mid	0.11	0.11	0.77
	Low	0.26	0.25	0.55		Low	0.23	0.22	0.49
	Dry	0.16	0.16	0.65		Dry	0.41	0.41	0.55
Kilan A	High	0.88	0.88	0.62	Yeonggang A	High	0.57	0.57	0.97
	Moist	0.47	0.47	0.74		Moist	0.31	0.31	0.84
	Mid	0.48	0.48	0.73		Mid	0.31	0.31	0.71
	Low	0.31	0.30	0.57		Low	0.71	0.70	0.57
	Dry	0.43	0.43	0.79		Dry	0.18	0.17	0.73
Nakbon A	High	0.87	0.87	0.57	Yongjeon A	High	0.79	0.78	0.83
	Moist	0.56	0.56	0.73		Moist	0.28	0.27	0.82
	Mid	0.33	0.33	0.51		Mid	0.26	0.26	0.62
	Low	0.28	0.26	0.51		Low	0.14	0.14	0.76
	Dry	0.09	0.09	0.80		Dry	0.29	0.29	0.63
Namgang A	High	0.74	0.74	0.85	Wicheon A	High	0.94	0.94	0.44
	Moist	0.51	0.51	0.64		Moist	0.59	0.59	0.50
	Mid	0.32	0.31	0.62		Mid	0.31	0.30	0.53
	Low	0.51	0.51	0.65		Low	0.69	0.69	0.69
	Dry	0.34	0.34	0.66		Dry	0.33	0.33	0.57
Naeseong A	High	0.83	0.83	0.61	Ian A	High	0.51	0.51	0.98
	Moist	0.63	0.63	0.66		Moist	0.38	0.38	0.98
	Mid	0.20	0.20	0.67		Mid	0.32	0.32	0.88
	Low	0.23	0.22	0.53		Low	0.29	0.29	1.02
	Dry	0.22	0.22	0.53		Dry	0.41	0.40	0.87
Micheon A	High	0.86	0.86	0.56	Hwanggang A	High	0.82	0.82	0.87
	Moist	0.49	0.49	0.76		Moist	0.31	0.31	0.78
	Mid	0.20	0.19	0.78		Mid	0.26	0.25	0.59
	Low	0.51	0.51	0.77		Low	0.38	0.38	0.67
	Dry	0.35	0.35	0.86		Dry	0.29	0.28	0.67

따른 각 계수의 편차를 보여주고 있으며,  $a_0, a_1, a_2$ 의 경우 큰 변동을 나타내고 있다. 또한 9번 회귀식에 대해 각 계수의 민감도 분석을 수행한 결과  $a_0, a_1, a_4, a_5$ 는 오염부하량 산정에 있어 민감하게 반응하는 것으로 나타났으며, 이로 인해 9번 회귀식 적용시 위 4개의 계수에 의해 크게 상이한 모의 결과가 산정될 수 있다. Table 5는 유역별 유황구간별 9번 회귀식과 머신러닝 기법을 통해 산정된 각 계수를 적용하여 예측된 SS

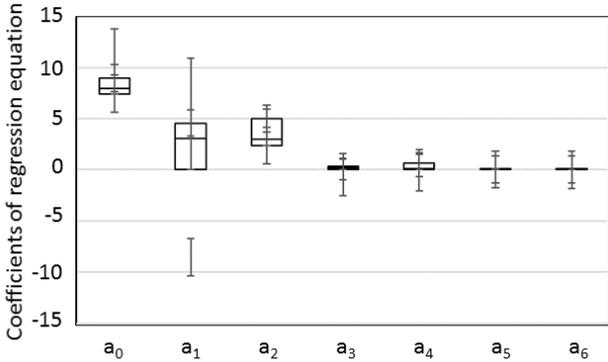


Fig. 4 Variability of coefficients of regression equation (#9) according to flow conditions for Namgang watershed

오염부하량과 실측값을 비교 분석한 결과이다. 대부분의 유역에서 고유량에 대한 분석 결과는 실측값과 유사하게 모의되는 경향 ( $R^2$ 와 NSE 0.51~0.94)을 확인할 수 있었다. 반면 저유량 구간에 대해서는 다소 실측값과 차이를 보이는 경향이 있으나, 남강 A, 미천 A, 영강 A, 위천 A 유역과 같은 경우  $R^2$ 와 NSE가 0.4 이상을 보였다. Ramanarayanan et al. (1997)에 따르면 결정계수 ( $R^2$ )와 모형효율지수 (EI)가 각각 0.5와 0.4 이상이면 모형이 자연현상을 잘 모의한다고 명시한바 있다. 본 연구 결과 대체로 저유량에서는 다소 차이를 보이고 있으나, 남강 A, 미천 A, 영강 A, 위천 A 유역의 경우 저유량에서도 실측과 유사하게 나타난 것을 볼 수 있다. 또한, Fig. 5는 대표적으로 감천 A, 남강 A, 반변 A 유역에 대한 실측과 모의된 SS 농도 결과를 시계열로 비교한 그림을 보여주고 있다. 세 유역에서 전반적으로 예측된 SS 농도가 실측 농도와 유사한 경향을 보이고 있어 본 연구에서 산정된 유황별 회귀식 계수의 활용이 가능할 것으로 판단된다.

따라서 본 연구의 결과, 고유량 부분에서는 강우 패턴의 영향이 직접적으로 고려되어 관련 매개변수의 적용으로 실측과 유사한 경향을 보일 수 있으나, 저유량 부분에서는 다소 차이가 발생하였다. 그 이유로는 저유량에 대해서는 강우에 의한

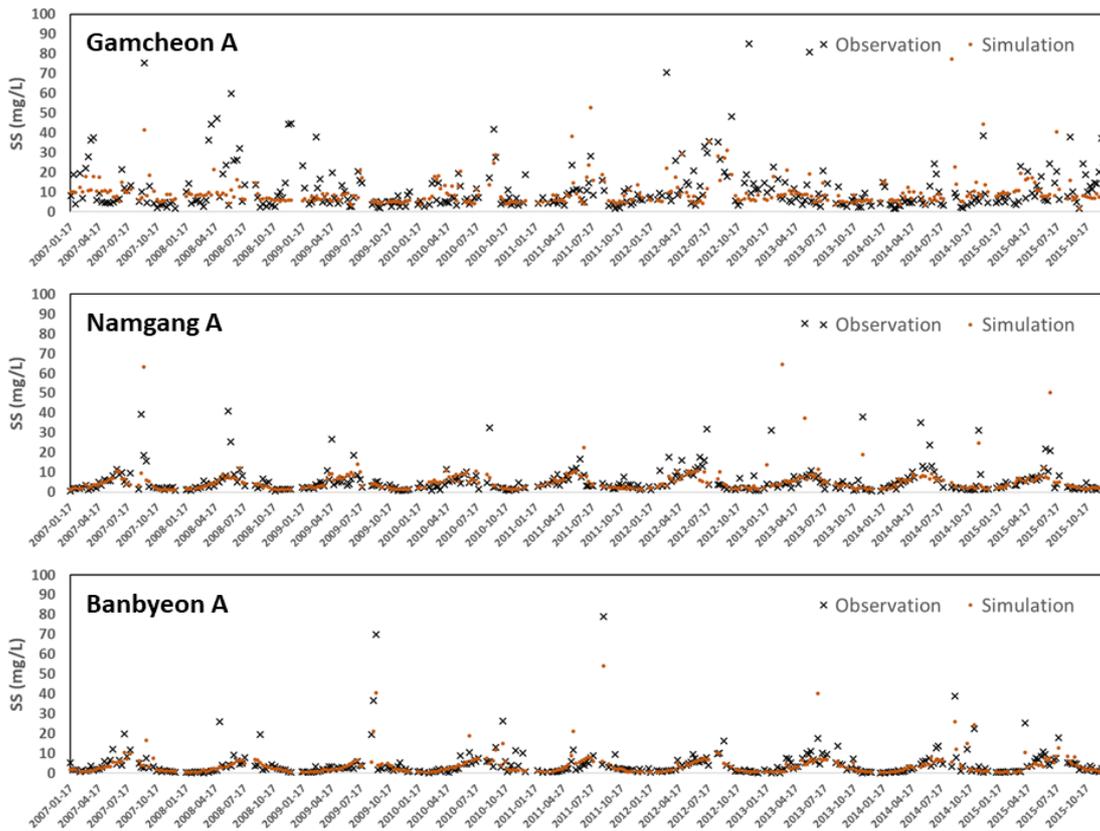


Fig. 5 Time series comparisons of observed and simulated SS (mg/L) for Gamcheon A, Namgang A, and Banbyeon A watersheds

영향보다 기저유출이나 다양한 점오염원에 의해 영향을 크게 받을 수 있기 때문인 것으로 판단된다. 따라서 본 연구 결과를 통해 LOADEST 모형을 활용한 오염부하량 산정시 유황구간에 따른 정밀 분석이 필요할 것이며, 향후 보다 정확한 오염부하량 산정을 위해서는 강우의 패턴뿐만 아니라 기저유출이나 점오염원에 의한 영향도 반영해야 할 필요가 있다.

#### IV. 결 론

본 연구는 기존 LOADEST 모형의 단점을 보완하기 위해 최근 널리 활용되고 있는 머신러닝 기법을 적용하여 적정 회귀식 선정 및 각 회귀식의 계수를 산정하였다. 머신러닝 기법은 주어진 문제에 대해 도출한 값의 미분을 통해 최적의 해를 찾고 학습 계수를 통한 수많은 반복을 통해 정확한 결과의 도출이 가능하여 기존 LOADEST 모형 내 적용되고 있는 방법에 비해 보다 정확한 회귀식 계수의 산정이 가능하다. 또한 우리나라와 같이 연간 유량 변동이 크고 여름철 강우가 집중되는 기후 특성에 있어서는 유황별 오염부하 특성이 달라질 수 있다. 따라서 본 모형을 이용한 오염부하량 산정시 전체 모의기간에 대해 회귀식을 선정하는 기존 방식과는 달리 전체 모의기간을 유황별로 분할하여 각각에 대한 최적 회귀식 계수 산정 및 적정 회귀식을 도출하였다. 본 연구 결과에서 나타났듯이 모든 대상 유역과 유황구간에 대해서 9번 회귀식이 적정 회귀식으로 선정되었으며, 9번 회귀식은 유량 특성뿐만 아니라 계절적인 특성 반영이 가능하다. 또한 머신러닝 기법을 통해 유황구간별 9번 회귀식의 계수를 산정한 결과 민감도가 크게 나타난 계수의 경우 그 값의 차이가 크게 발생한 것을 알 수 있었다. 이로 인해 동일한 회귀식을 적용하더라도 적용된 계수에 따라 오염부하량이 크게 차이가 날 수 있다. 또한, 본 연구에서의 모의치를 실측값과 비교 분석한 결과 고유량에서는 비교적 유사한 경향을 보였으나 저유량에서는 유사한 경향을 보이는 유역도 있었으나 다소 차이가 발생한 것을 볼 수 있었다. 이는 모형 결과의 정확성 분석시 통계방법이 고유량 또는 상대적으로 높은 오염부하량 값에 크게 영향을 받을 수 있기 때문에 고유량 부분에서 높은 정확도를 보인 것으로 판단된다. 반면, 저유량에서는 실측된 오염부하량이 회귀식 내 고려되지 않는 기저유출 및 점오염원에 의한 영향을 크게 받기 때문에 상대적으로 정확도가 떨어지는 것으로 판단된다. 따라서 저유량 부분에서 보다 정확한 예측을 위해 향후 모형의 내부 프로세스 및 적용 방법에 대한 개선이 필요할 것으로 판단된다. 결론적으로 본 연구의 결과를 통해 연속적인 오염부하량 예측 시 전체 모의기간보다 유황구간별로 분할 분석으로 보다 정확한 오염부하량 산정이 가능하며, 향후 갈수기 오염

부하량 평가에도 활용될 수 있을 것이라 판단된다. 또한, 기존 연구에서는 LOADEST 모형 이용시 유황구간에 따른 적용사례가 없기 때문에 본 연구를 통해 향후 모형 개선에 있어 그 기여가 클 것으로 판단된다.

#### 사 사

본 연구는 국토교통부/국토교통과학기술진흥원의 지원으로 수행되었음 (과제번호 17AWMP-B083066-04).

#### REFERENCES

1. 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.
2. Dean, J., G. S. Corrado, R. Monga, K. Chen, M. Devin, Q. V. Le, M. Z. Mao, M. Ranzato, A. Senior, P. Tucker, K. Yang, and A. Y. Ng, 2012. Large scale distributed deep networks. *In Proceedings of NIPS*, pp. 1232-1240.
3. 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.
4. Jha, B. and M. K. Jha, 2013. Rating Curve Estimation of Surface Water Quality Data Using LOADEST. *Journal of Environmental Protection* 4: 849-856.
5. Kang, H. W., Y. S. Park, J. Kim, W. S. Jang, J. C. Ryu, N. W. Kim, D. S. Kong, and K. J. Lim, 2010. Enhancement of SWAT Auto-calibration using K-means Clustering Algorithm. *2010 Autumn Conference of the Korean Society on Water Quality & Korean Society of Water and Wastewater*, pp. 197-198. [Korean Literature]
6. Kim, S. D., D. K. Kang, M. S. Kim, and H. S. Shin, 2007. The Possibility of Daily Flow Data Generation from 8-Day Intervals Measured Flow Data for Calibrating Watershed Model, *Journal of Korean Society on Water Environment* 23(1): 64-71. [Korean Literature]
7. Lee, J., H. Kwon, and H. Choi, 2014. Evaluation of pollution level attributed to nonpoint sources in Nakdonggan Basin, Korea, *Journal of Environmental Impact Assessment* 23(5): 393-405. [Korean Literature]
8. Marcin Andrychowicz, Misha Denil, Sergio Gómez Colmenarejo, Matthew W. Hoffman, David Pfau, Tom Schaul, Brendan Shillingford, Nando de Freitas, 2016. Learning to learn by gradient descent. 30th Conference on Neural Information

- Processing Systems (NIPS 2016), Barcelona, Spain.
9. Ministry of Environment (ME), 2004. Total Maximum Daily Loads Handbook. Ministry of Environment, Ministry of Environment (ME), pp. 1-69. [Korean Literature]
  10. National Institute of Environment Research (NIER), 2013. The Study on the Optimum Assessment Methods for Achievement of Target Water Quality and Estimation of Allocation Loads Using a Dynamic Model, NIER-RP2013-274, *National Institute of Environment Research*, pp. 1-31. [Korean Literature]
  11. 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.
  12. 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.
  13. Park, Y. S., J. M. Lee, Y. Jung, M. H. Shin, J. H. Park, H. Hwang, J. Ryu, J. Park, and K. Kim, 2015. Evaluation of regression models in LOADEST to estimate suspended solid load in Hangang waterbody. *Journal of the Korean Society of Agricultural Engineers* 57(2): 37-45.
  14. Shin, M. H., C. H. Won, Y. H. Choi, K. C. Kim, J. Y. Seo, K. J. Lim, and J. D. Choi, 2008. Estimation of LOADEST Model application for NPS pollutants from agricultural watershed. the 2008 KSAE Annual Conference 24 October, poster No. 6.
  15. 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.
  16. Ramanarayanan, T. S., J. R. Williams, W. A. Dugas, L. M. Hauck, and A. M. S. McFarland, 1997. Using APEX to identify alternative practices for animal waste management, Minneapolis, MN. Paper No. 97-2209.
  17. Runkel, R. L., C. G. Crawford, and 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 Book 4, Chapter A5, pp. 69.