

ENGINEERING

# Flood prediction in the Namgang Dam basin using a long short-term memory (LSTM) algorithm

Seungsoo Lee<sup>1</sup>, Hyunuk An<sup>2\*</sup>, Youngteck Hur<sup>3</sup>, Yeonsu Kim<sup>3</sup>, Jisun Byun<sup>4\*</sup>

<sup>1</sup>Division for Integrated Water Management, Korea Environmental Institute, Sejong 30147, Korea

<sup>2</sup>Department of Agricultural and Rural Engineering, Chungnam National University, Daejeon 34134, Korea

<sup>3</sup>K-water Research Institute, Daejeon 34045, Korea

<sup>4</sup>Department of Civil Engineering, Chungnam National University, Daejeon 34134, Korea

\*Corresponding author: [hyunuk@cnu.ac.kr](mailto:hyunuk@cnu.ac.kr), [byun.jisun5@gmail.com](mailto:byun.jisun5@gmail.com)

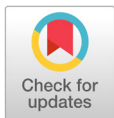
## Abstract

Flood prediction is an important issue to prevent damages by flood inundation caused by increasing high-intensity rainfall with climate change. In recent years, machine learning algorithms have been receiving attention in many scientific fields including hydrology, water resources, natural hazards, etc. The performance of a machine learning algorithm was investigated to predict the water elevation of a river in this study. The aim of this study was to develop a new method for securing a large enough lead time for flood defenses by predicting river water elevation using the a long- short-term memory (LSTM) technique. The water elevation data at the Oisong gauging station were selected to evaluate its applicability. The test data were the water elevation data measured by K-water from 15 February 2013 to 26 August 2018, approximately 5 years 6 months, at 1 hour intervals. To investigate the predictability of the data in terms of the data characteristics and the lead time of the prediction data, the data were divided into the same interval data (group-A) and time average data (group-B) set. Next, the predictability was evaluated by constructing a total of 36 cases. Based on the results, group-A had a more stable water elevation prediction skill compared to group-B with a lead time from 1 to 6 h. Thus, the LSTM technique using only measured water elevation data can be used for securing the appropriate lead time for flood defense in a river.

**Keywords:** deep-learning, flood forecasting, Keras, long- short-term memory (LSTM), machine learning

## Introduction

기후변화로 인해 최근 국내 연평균 강수량과 태풍의 발생 빈도가 증가하고 있다. 기상청 자료에 따르면 1970년대까지 1일 강우량이 200 mm 이상인 경우가 연평균 1회에 불과하였으나 1998년에는 5회로 증가하였으며, 2002년에는 태풍 ‘루사’로 인해 강릉지방의 일강우량이 과거 최대치인 장흥지방의 547.4 mm의 약 1.6배에 달하는 870.5 mm를 기록하였다(KMA, 2012). 이



### OPEN ACCESS

**Citation:** Lee S, An H, Hur Y, Kim Y, Byun J. 2020. Flood prediction in the Namgang Dam basin using a long short-term memory (LSTM) algorithm. Korean Journal of Agricultural Science 47:471-483. <https://doi.org/10.7744/kjoas.20200036>

**Received:** April 23, 2020

**Revised:** July 13, 2020

**Accepted:** July 14, 2020

**Copyright:** © 2020 Korean Journal of Agricultural Science



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

러한 극한 강우의 증가는 치수목적으로 설계된 수공 구조물의 안전성을 위협하고 있으며 하천의 홍수에 많은 어려움을 초래하고 있다. 홍수는 기후변화에 따른 집중호우와 태풍으로 인한 피해를 예방하거나 줄이기 위하여 실시하고 있으며 ‘하천법시행규칙 제 23조 2항’에 의거하여 홍수예보발령지점과 해당 지점의 주의보수위와 경보수위를 사전에 정하여 고시하고 있다. 이러한 홍수예보의 목적을 달성하기 위해서는 정확도 높은 홍수예보를 신속하게 수행하여 홍수주의보 및 경보에 도달하는 시간과 수위를 관계기관 및 하류지역 주민에게 알려 사전에 필요한 조치를 수행할 수 있도록 하는 것이 필요하다. 홍수예보 방법으로는 기상법, 수위법, 강우-유출법 등을 활용하고 있다. 기상법은 강우량을 정량적으로 예측하여 홍수 유발량 이상이면 경보를 발령하는 방법이며, 수위법은 상류의 수위관측소에서 계측된 수위를 기반으로 하여 하류의 수위를 예측하는 방법을 말한다. 강우-유출법은 강우유출 모형모형을 이용하여 해당 지점에 대한 홍수위와 도달시간을 예측하는 방법을 의미하며, 현재 4대강 홍수통제소에서 홍수예보를 위해 활용하고 있는 방법으로써 상·하류 수위관측소의 계측수위자료가 동시에 필요하지 않다는 점과 상대적으로 빠르게 홍수위와 도달시간을 산정할 수 있다는 장점을 가지고 있으나 모형상수를 결정하기 위해서는 양질의 수문자료가 요구되며 모형 운영시 해당유역에 대한 수문학적 지식과 경험이 대단히 중요하다는 특징을 가지고 있다. 유역내의 강우유출 현상은 강우패턴, 지속시간, 토질, 토지피복도, 하천의 형상, 유역 경사, 토양수분량, 기온 등과 같은 변수들에 의해 복합적으로 영향을 받는 매우 비선형적인 현상으로써 개념적(conceptual) 모형으로만 표현하기에는 한계가 있으며, 매우 정교한 물리적 기반(physical based) 수문모형을 이용할지라도 매개변수화(parameterization)를 통한 검보정이 필수적이다. 또한 물리적 기반 모델의 경우 모형이 정교해질수록 고려해야 하는 변수의 양이 늘어나고 관측자료를 이용한 모형의 검보정에 많은 시간이 소모될 뿐만 아니라 증가된 매개변수가 모형의 불확실성을 유발할 수 있는 한계를 가지고 있어 홍수 예보를 위한 실무에 활용되기에는 한계가 있다. 인공지능 경망은 McCulloch et al. (1943)에 의해 최초 기초논리가 연구되었으며 1980년에 전문가 시스템(expert system, ES)을 거쳐 Rumelhar et al. (1986)이 오차역전파기법(backpropagation)기법을 제안한 이후 Hinton et al. (2006)이 다계층 인공신경망의 학습을 소개한 이후 다층퍼셉트론 (multi-perceptron), 순환신경망(recurrent neural network, RNN), 합성곱 신경망(convolutional neural network, CNN) 등의 기법이 개발되어 다양한 분야에서 괄목할만한 성과를 나타내고 있다. LSTM (long- short-term memory)을 이용한 수문분야에 관한 선행 연구를 살펴보면 Sheng and Lewis (2011)은 LSTM을 이용하여 진동수조(oscillating water column, OWC)의 수위를 예측하는 연구를 수행하여 근 미래의 진동수위를 예측하는 효과적임을 보인바 있으며, Liang et al. (2018)은 LSTM 모델에 7개의 구성요소(수위, 5개의 유량, 강수량)를 입력자료로 하여 Dongting Lake의 수위를 예측하였으며 SVM (support vector model)모델 보다 예측성이 우수함을 보인바 있다. Zhang et al. (2018)은 사물인터넷(internet of things, IoT) 기술을 통해 수집된 하수관로의 수위를 예측하기 위하여 다층퍼셉트론(multilayer perceptron, MLP), 웨이블릿 신경망(wavelet neural network, WNN), LSTM 그리고 GRU (gated recurrent unit) 모델 결과를 비교 분석하여 LSTM의 결과가 가장 우수함을 보였다. 특히 수리·수문 분야에서 인공신경망을 이용한 연구는 Yeo et al. (2010)은 유전자알고리즘을 인공신경망에 결합하여 유전자 알고리즘을 사용하여 신경망을 학습시킨 모형이 그렇지 않은 모형에 비해 우수한 결과를 보여줌을 증명하였으며 입력 자료로서 강우와 수위를 함께 사용한 경우가 수위만을 사용한 모형에 비해 더 우수한 결과를 나타냄을 보인바 있다. Ryu et al. (2012)는 비선형적 모델인 웨이블릿-인공신경망을 적용하여 충주댐 유역의 일유입량을 예측하였으며 일반적인 인공신경망에 비하여 웨이블릿 변환을 통한 디노이징 기법의 적용이 잡음이 포함되어 있는 원자료의 사용보다 더 정확한 예측을 하는 것을 보인바 있다. Li et al. (2016)이 SDF (satellite-derived flow) 신호를 입력자료로 임진교에서의 유량을 예측하는 연구에서 1일 전 유량자료가 학습모델의 입력 자료로 추가되는 경우 예측성이 향상됨을 보인바 있다. Seo et al. (2017)은 ANN (artificial neural network), GRNN (general regression neural network), ANFIS (adaptive-network-based fuzzy inference system), RF (random forest)를 이용하여 저수지 수위예측 능력을 비교하였으며 ANN이 다른 기법에 비하여 우수한 성능을 가지고 있음을 보인바 있다. Yoon et al. (2018)은 관측 강우량과 유량자료를 이용하여 도달

시간을 산정하고 수문학적 인자를 입력자료로 학습된 인공지능망의 결과와 비교하여 신뢰성을 평가하였다. Jung et al. (2018b)은 구글의 Tensorflow 모듈을 이용하여 다중회귀 분석기법과 LSTM 기법을 이용하여 대청댐 유역의 옥천 수위 관측소의 1시간 후 수위예측 성능을 평가하고 다중회귀 분석 기법에 비하여 LSTM 기법의 예측성이 우수함을 보인 바 있다. 앞서 살펴본 바와 같이 인공지능망을 이용하여 하천 수위와 유량을 예측하기 위한 다양한 연구가 수행되었다. Jung et al. (2018a)은 LSTM 기법을 감조하천에 적용하여 수위를 예측하였으며, Lee et al. (2018)은 LSTM 기법을 매공강 유출모의에 적용하여 물리기반의 모형과 비교하였다. Lim et al. (2019)는 LSTM 기법을 일단위 수질 예측에 적용하여 그 적용성을 평가하였다.

그러나 하천의 수위만을 단독으로 이용하여 관측자료의 관측 시간간격과 선행시간에 따른 인공지능망의 예측성에 관한 연구는 미진한 상태이다. 따라서 본 연구에서는 시계열 자료의 예측에 최적화된 LSTM 기법을 이용하여 하천수위의 관측 시간별 선행시간에 따른 예측성능을 비교분석하는 연구를 수행한다. 대상유역은 댐의 방류 등과 같은 인위적인 간섭이 가장 적은 남강댐 상류지역의 외송 수위관측소에 관측된 자료를 이용하였으며, 한국수자원공사에서 관측한 1시간 단위 관측수위를 1, 2, 3, 6, 12, 24시간 간격으로 재가공하여 동일 간격 수위(group-A)와 시간 평균 수위(group-B)로 입력자료를 구축하였다. 재가공된 수위자료는 LSTM 모형의 입력자료로 이용되었으며 선행시간 1시간부터 72시간까지의 예측성능에 대해 알아보고 홍수예측을 위한 실무적용 가능성에 대해 평가하였다.

## Materials and Methods

### Theoretical background기계학습과 인공지능망

인공지능망은 기계학습의 한 분야로서 인간 뇌의 뉴런구조를 수학적으로 재구성한 정보처리 시스템을 의미하며 다층의 뉴런과 각 뉴런을 연결하는 가중치로 구성되어 있다. 딥러닝(deep-learning)은 Rosenblatt가 1957년에 제안한 퍼셉트론(perceptron) 이론에 기반하여 발전된 인공지능망의 한 종류이며 일반적인 기계학습 기법에 활용되는 학습층(layer)을 깊게(deep) 하여 학습(learning)한다 하여 딥러닝으로 불리운다(Lee et al., 2016). 인공지능망은 기본적으로 입력층(input layer)과 은닉층(hidden layer) 그리고 출력층(output layer)로 구성되어 있으며 각 층은 가중치를 조절하는 한 개 이상의 노드와 하나의 편향으로 구성되어 있으며 가중치와 편향을 조절하여 원하는 값을 출력할 수 있도록 한다. 단순 신경망은 은닉층 노드의 개수와 상관없이 은닉층이 한 개 인 것을 의미하며 딥러닝은 두 개 이상의 은닉층을 가지는 인공지능망을 의미한다.

### Keras

케라스(<https://keras.io>)는 구글 엔지니어인 프랑소와 쵸레(Francois Chollete)에 의해 개발되어 유지보수되고 있는 공개 API (application programming interface)기반 딥러닝 라이브러리이다. 케라스는 딥러닝 비전문가도 해당 분야에 딥러닝 모델을 개발하고 활용할 수 있도록 직관적인 API를 제공하고 있으며 내부적으로는 텐서플로우(TensorFlow), 티아노(Theano), CNTK (computational network toolkit)등의 딥러닝 전용 엔진으로 구동되며 프로그래밍 언어로는 파이썬(Python)을 채택하고 있다. 케라스는 직관적인 API로 손쉽게 다층퍼셉트론(multi-perceptron), 컨볼루션 신경망(CNN), 순환신경망(RNN) 등의 모델을 구성할 수 있으며 단독 모델 구성뿐만 아니라 다중 입력 또는 다중 출력을 통한 모델 구성도 가능하다.

## LSTM 기법

RNN은 가중치 학습과정에서 전단계에 학습된 가중치의 영향을 고려할 수 있어 시계열자료를 학습하는 경우 정확도 높은 예측 결과를 생산하는 모델로 알려져 있다(Jung et al., 2018a). Fig. 1에서는 RNN을 이용한 학습 방식을 표현하고 있다. RNN은 학습이 순차적으로(1→2→3→4) 진행됨에 따라서 입력층(input layer)에서 입력받은 값이 은닉층(hidden layer)의 가중치에 영향을 주며 모든 학습단계(여기에서는 1에서 4까지)에서 입력받은 입력값의 영향을 반영하여 가중치를 학습하게 된다. 따라서 시계열 자료와 같이 연속된 자료에서 과거의 값이 미래의 값에 영향을 주는 자료의 학습에 강점을 가지고 있으나 가중치소실(vanishing gradient)이라는 단점을 가지고 있다. 가중치소실이란 입력자료의 길이가(여기에서는 학습 단계를 의미) 길어짐에 따라 초기에 입력받은 값이 가중치에 미치는 영향이 상실됨을 의미한다.

이러한 가중치소실 문제를 해결하기 위하여 Hochreiter and Schmidhuber (1997)은 LSTM 기법을 제안하였다. LSTM 기법은 RNN에서 발생하는 가중치 손실을 방지하기 위하여 Fig. 2에서와 같이 전통적인 RNN 기법에 컨베이어 벨트와 같이 cell state라는 구조를 배치하여 과거의 정보를 유지할 수 있도록 하였다. LSTM은 전 단계로부터 출력된  $h_{t-1}$ 와 현재의 입력값  $x_t$ 중 무엇을 버릴지 결정하는 망각게이트(forget gate,  $f_t$ )를 통하여 cell state를 통하여 유지되어야 하는 정보를 선택하며 그 수식은 다음과 같다.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

여기서,  $\sigma$ 는 활성화 함수(sigmoid 함수),  $W_f$ 는 가중치,  $b_f$ 는 편향을 나타낸다.

망각게이트를 통하여 cell state에 유지되어야 하는 정보를 선별한 후에는 식(2)와 (3)을 통하여 계산된 벡터값을 이용하여 다시 입력게이트(input gate,  $i_t$ )에서 어떤 정보를 새롭게 cell state에 더할지를 결정한다.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (3)$$

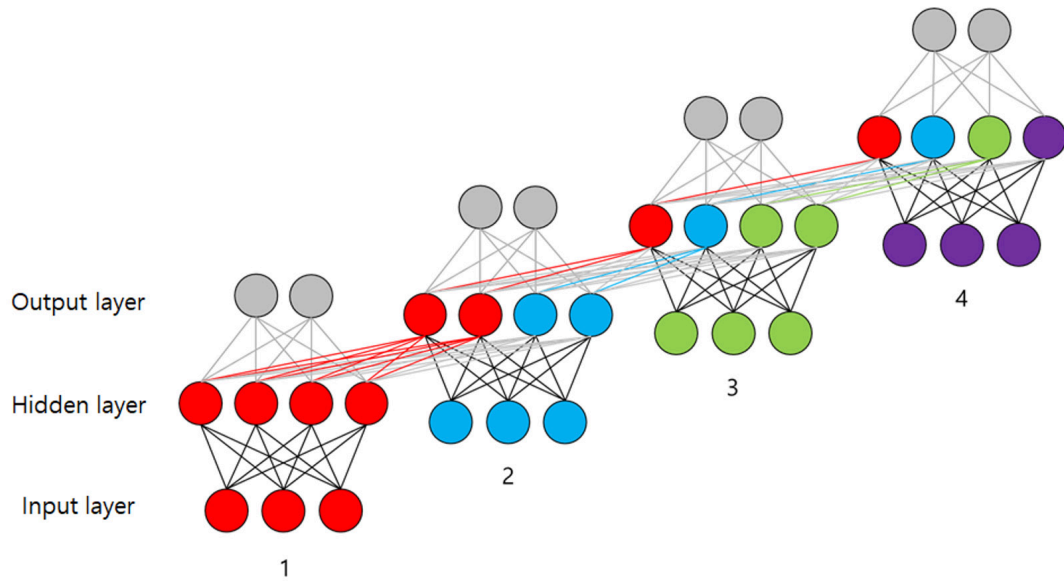
여기서,  $\tilde{C}_t$ 는 새롭게 업데이트된 셀의 상태,  $W_i$ ,  $W_c$ 는 각각 함수의 가중치,  $b_i$ ,  $b_c$ 는 각각 함수의 편향을 의미한다.

최종적으로 망각게이트를 통과한 셀의 상태는 시그모이드 활성화 함수를 통과하게 됨으로써 0과 1사이의 값을 갖게 되고 cell state에 곱해지게 되므로 0의 값을 갖는 경우 소멸되며 0이외의 값을 갖게 되는 경우 상태량이 유지된다. 입력게이트를 통과한 상태량은 다시 tanh를 통과하게 되므로 -1과 1사이의 값을 가지게 되고, cell state에 더해져 다음단계로 학습이 진행되게 된다. 마지막 출력단계에서는 출력게이트(output gate,  $o_t$ )를 통과하여 출력값이 결정되게 되며 그 식은 다음과 같다.

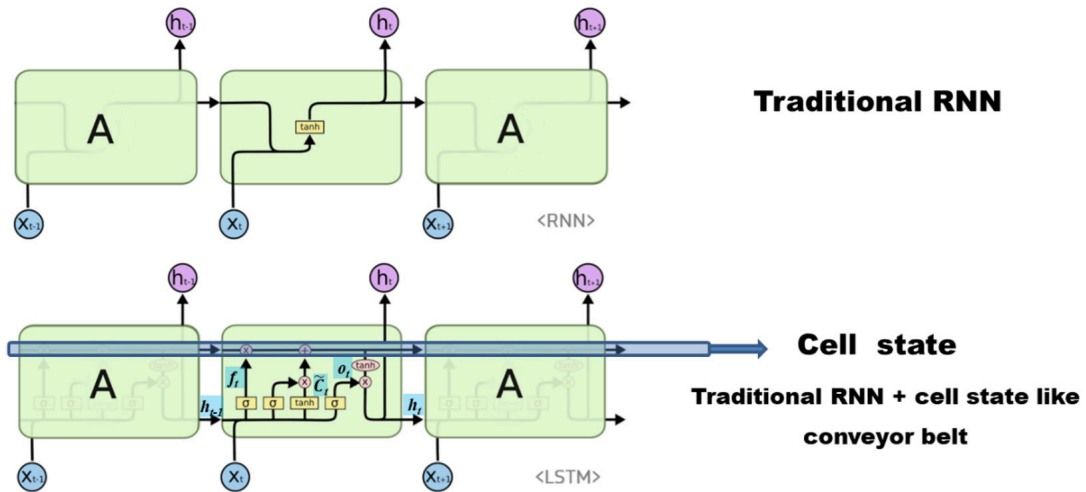
$$o_t = \sigma(W_o [h_{t-1}, x_t] + b_o) \quad (4)$$

$$h_t = o_t \cdot \tanh(C_t) \quad (5)$$

여기서  $h_t$ 는 다음 단계로 넘어가는 현재의 출력값을 의미한다.



**Fig. 1.** Recurrent neural network. Red, input values in training step 1; blue, input values in training step 2; green, input values in training step 3; purple, input values in training step 4; grey, output values.



**Fig. 2.** Comparison between traditional recurrent neural network (RNN) and long- short-term memory (LSTM) (Olah, 2015).

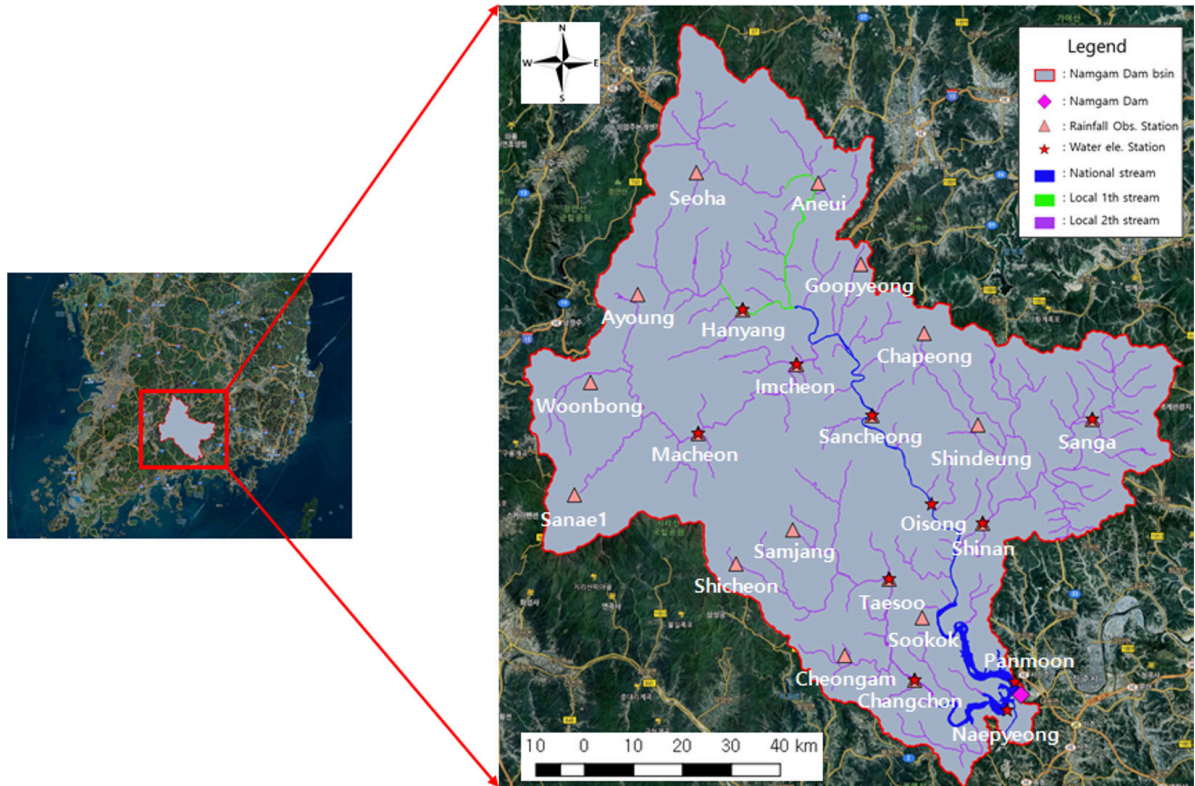
### Methods 대상유역 모형의 구성

본 연구에서는 LSTM 모형을 이용하여 남강댐에 유입되는 국가하천 중 가장 하류단에 위치하고 있는 외송 수위 관측소의 수위 예측성능을 평가한다. 대상유역과 수위관측소의 위치는 Fig. 3과 같으며, 관측수위 자료는 수자원공사에서 운영하고 있는 외송 수위관측소의 수위자료를 이용하였다. 수위관측소의 관측기간은 2013년 2월 15일부터 2018년 8월 26일까지 약 5년 6개월이며 관측 시간 간격은 1시간 단위의 자료를 이용하였다. 관측자료의 시간 간격에 따른 예측 성능을 알아보기 위하여 시간 단위 관측수위 자료를 1, 2, 4, 6, 12, 24시간 간격으로 재구성하였으며 동일

간격 수위로 구성한 자료(group-A)와 시간 평균 수위(group-B)로 구성된 자료로 나누어 예측성능을 평가하고 비교분석하였다. 시간 평균 수위자료는 아래의 식을 이용하여 구성하였다.

$$h_n = \frac{\sum_{i=1}^n h_i}{n} \quad (6)$$

여기서  $h_n$ 은 시간 평균 수위(m),  $n$ 은 시간 간격,  $h_i$ 는 1시간 단위 수위(m)이다.



**Fig. 3.** Study area and location of water elevation gauge station.

LSTM 모형 구축시 전체 자료의 80%를 학습에 이용하도록 하였으며 그중 20%를 임의로 추출하여 학습검증에 활용하도록 하였다. 따라서 학습된 모델의 적용성 평가는 전체 데이터의 후반부 20%에 대해서 적용성을 평가하게 된다. 입력노드는 12개로 설정하여 선행시간 별 수위값을 예측할 수 있도록 입력자료를 구성하였으며 은닉층 개수는 사전 테스트를 통해 15개로 선정하였다. 학습 반복시행 횟수는 모든 경우에 대해서 균일하게 1,000회로 설정하였으며 반복 학습시 검증손실(validation loss)이 감소하지 않는 경우가 20회(patience = 20)에 도달하게 되면 강제로 학습을 종료할 수 있도록 설정하여 모델 학습에 소요되는 시간을 줄일 수 있도록 설계 하였다. 활성화 함수로는 최근 LSTM 예측 모델에서 우수한 성능을 보여주고 있는 것으로 알려진 ReLU (rectified linear unit) 함수를 사용하였으며 최적화기법(optimizer)으로는 adam을 사용하였다. 또한 시계열 자료의 특징을 최대한 반영할 수 있도록 반복 학습(epoch)을 수행할 때 마다 전 단계에 학습된 가중치의 상태량을 유지할 수 있도록 statueful = true로 옵션을 설정하여 학습을 진행하였다.

## 결과 분석 방법

본 연구에서는 LSTM모형을 이용한 예측결과의 재현성을 비교·분석하기 위한 평가지수로 평균제곱근 오차 (root mean square error, RMSE)와 결정계수(coefficient of determination,  $R^2$ )를 이용하였다.

$$\text{평균제곱근 오차: RMSE} = \sqrt{\frac{\sum_{n=1}^N (\hat{y}_n - y_n)^2}{N}} \quad (7)$$

$$\text{결정계수: } R^2 = 1 - \frac{\sum_{n=1}^N (y_n - \bar{y}_n)^2}{\sum_{n=1}^N (\hat{y}_n - \bar{y}_n)^2} \quad (8)$$

여기서,  $N$ 은 데이터의 개수,  $\hat{y}_n$ 은 예측값,  $y_n$ 은 실측값,  $\bar{y}_n$ 은 실측값의 평균이다.

## Results and Discussion

본 연구에서는 전술한 바와 같이 관측자료의 시간 간격에 따른 예측 성능을 알아보기 위하여 먼저 6개의 시간간격(1, 2, 4, 6, 12, 24시간) 자료에 대해서 동일 간격 자료와(group-A) 시간 평균 자료(group-B)로 재구성하였다. 여기서 동일 시간 간격이란 1시간 단위의 원자료에서 일정 시간 간격으로 데이터를 재구성하였다는 의미이며, 시간 평균 자료는 다음 식(9)에 의해 재구성 자료를 의미한다.

$$x_n = \frac{\sum_{i=1}^n h_i}{n} \quad (9)$$

여기서,  $x_n$ 은 시간 평균 자료,  $n$ 은 시간 평균( $n = 1, 2, 4, 6, 12, 24$ ),  $h_i$ 는 1시간 단위 관측자료를 의미한다.

재구성된 자료는 각각 3단계의 미래값(1시간 간격자료의 경우 1, 2, 3시간 후의 값을, 24시간 간격자료의 경우 24, 48, 72시간 후의 값을)을 예측하도록 총 36개의 예측모형을 구성하였다. 학습된 모형을 이용한 예측성능을 알아보기 위하여 Table 1에 개별 모델에 대한 RMSE와  $R^2$ 값을 표시하였다.

먼저 group-A의 예측결과 case-1 (1시간 간격자료를 사용하여 1시간 후 수위 예측)에서  $R^2$ 가 0.9867, RMSE는 0.035로 가장 우수한 예측 성능을 보여주었으며 case-18 (24시간 간격자료를 이용하여 72시간 후 수위 예측)에서  $R^2$ 는 0.2586, RMSE는 0.2515로 가장 저조한 예측 성능을 보여주었다. 전반적으로 예측시간이 길어질수록 예측 정확도가 떨어지는 경향을 확인할 수 있다. 본 연구에서는 6시간 이내의 경우  $R^2$  0.85 이상의 높은 정확도로 수위 예측이 가능하나 그 이상의 경우 예측정확도가 떨어지는 경향이 있는 것으로 판단되었다. 이는 시계열 자료의 학습에 우수한 성능을 보여주는 LSTM의 특성상 원자료의 간격이 길어짐에 따라 자료 고유의 시계열성이 사라짐에 따라 학습이 잘 이루어지지 않는 것에 기인한 것으로 판단된다. 또한 예측시간이 긴 경우 시간간격이 짧은 데이터를 활용하는 경우 보다 긴 시간간격의 데이터를 그대로 학습자료로 활용한 경우가 예측 정확도가 높은 것으로 나타났다. 예를 들어 12시간 뒤의 수위예측을 하는 경우 12시간 시간간격 자료를 사용한 case-13가 6시간과 4시간 시간간격 자료를 사용한 case-11, case-9보다 예측정확도가 높은 것으로 나타났다. 이는 6시간 예측의 경우에도 마찬가지로 case-6보다 case-10이 예측정확도가 높은 것으로 나타났다. 그러나 case-6과 case-10에서처럼 6시간 이후의 수위예측 결과  $R^2$ 가 각각 0.8453, 0.8677, RMSE는 각각 0.1191, 0.1082를 보여주며 고정확도의 수위 예측 성능을 가지고 있는 것으로 판단된다. Group-B의 예측결과 역시 case-19에서 가장 우수한 예측성능( $R^2$ : 0.9850, RMSE: 0.0371)을 보여주었으며 case-35에서 가장 저조한 예측성능( $R^2$ : 0.1038, RMSE: 0.2707)을 보여주었으며 전반적으로 group-A와 대동소이한 결과를 나타내었다.

**Table 1.** Contents of Korea-Mongolia Greenbelt Plantation Project.

Index	Type	Interval time (h)	Lead time (h)	R <sup>2</sup>	RMSE (m)
Case-1	Time interval value (group-A)	1	1	0.9867	0.0350
Case-2			2	0.9634	0.0580
Case-3			3	0.9381	0.0754
Case-4		2	2	0.9603	0.0604
Case-5			4	0.8975	0.0969
Case-6			6	0.8453	0.1191
Case-7		4	4	0.9055	0.0925
Case-8			8	0.7958	0.1360
Case-9			12	0.6977	0.1654
Case-10		6	6	0.8677	0.1082
Case-11			12	0.7218	0.1569
Case-12			18	0.6267	0.1818
Case-13	12	12	0.7453	0.1436	
Case-14		24	0.5588	0.1890	
Case-15		36	0.3396	0.2312	
Case-16	24	24	0.5620	0.1933	
Case-17		48	0.3406	0.2372	
Case-18		72	0.2586	0.2515	
Case-19	Time averaged value (group-B)	1	1	0.9850	0.0371
Case-20			2	0.9629	0.0583
Case-21			3	0.9388	0.0750
Case-22		2	2	0.9715	0.0510
Case-23			4	0.9221	0.0843
Case-24			6	0.8635	0.1116
Case-25		4	4	0.9398	0.0738
Case-26			8	0.8285	0.1246
Case-27			12	0.7164	0.1602
Case-28		6	6	0.7667	0.1445
Case-29			12	0.7195	0.1585
Case-30			18	0.6313	0.1817
Case-31		12	12	0.7030	0.1620
Case-32			24	0.5595	0.1972
Case-33			36	0.4303	0.2243
Case-34		24	24	0.5735	0.1867
Case-35			48	0.1038	0.2707
Case-36			72	0.3124	0.2371

R<sup>2</sup>, coefficient of determination; RMSE, root mean square error.



Fig. 4에는 group-A와 group-B의 예측성능을 비교하기 위하여 예측값과 관측값의 시계열 비교 그림을 표시하였다. 지면의 제약으로 인하여 각 시간 간격에 대해서 1단계 예측에 대한 그림만을 표시하였으며 좌측에는 group-A의, 우측에는 group-B의 예측결과를 표시하였다. 육안으로 확인할 수 있듯이 1시간(case-1, case-19)에서 4시간 간격(case-7, case-25)까지의 자료를 이용하는 경우에 group-A와 group-B의 예측결과가 유사한 것을 확인할 수 있으나 6시간 간격 자료를 이용한 경우부터는 group-A에 비해 group-B의 예측성이 저하되는 것을 확인할 수 있다. 그러나 한 가지 흥미로운 사실은 비록 저수위의 경우 group-A의 경우에 비해 group-B 모델의 예측성이 낮게 나타났으나 고수위의 경우 group-B가 group-A보다 정확하게 수위를 예측했다는 점이다. 이를 확인하기 위하여 Fig. 5에는 수위에 따른 예측값과 관측값의 1:1 비교가 가능하도록 scatter plot 그래프를 나타내었다.

Fig. 5에서 볼 수 있듯이 group-A의 경우 자료의 시간 간격이 증가함에 따라 4.0 m 이상의 수위를 낮게 예측하는 반면 group-B의 경우에는 group-A보다 높게 예측하는 것을 확인할 수 있다. 이러한 결과를 나타내는 원인에 대해서는 향후 추가적인 연구를 통해 두 그룹이 가진 자료의 특성에 대한 분석이 수행되어야 할 것으로 판단된다.

## Conclusion

본 연구에서는 딥러닝 학습기법 중 시계열 자료 학습에 특화되어 개발된 LSTM기법을 이용하여 남강댐 유역에 위치하고 있는 외송 수위관측소의 수위 예측의 활용성을 평가하기 위한 연구를 수행하였다. 먼저 대상유역의 수위관측자료는 한국수자원공사에서 운영하고 있는 1시간 단위 관측수위를 사용하였으며 관측기간은 2013년 2월 15일부터 2018년 8월 26일까지 약 5년 6개월간의 자료를 활용하였다. 자료의 특성과 예측자료의 선행시간에 따른 예측성능을 알아보기 위하여 1시간에서 24시간까지 동일 간격 자료(group-A)와 시간 평균 자료(group-B)로 구분하여 각각 1단계에서 3단계 후까지의 자료를 예측하도록 총 36개의 모델을 구성하여 예측성능을 평가하였으며 본 연구에서 도출한 결론은 다음과 같다.

1) LSTM을 이용하여 하천수위 예측 모델을 구성하는 경우 선행시간 1시간(case-1  $R^2$ : 0.9867, RMSE: 0.035)에서 6시간(case-10  $R^2$ : 0.8677, RMSE: 0.1082)까지 안정적인 수위 예측 성능을 보여주었으며 이는 시계열 수위자료만을 이용하여 예측 모델을 구축하여도 홍수예보를 위한 실무에서 활용 가능한 수준으로 판단된다.

2) 동일 간격 자료(group-A)를 이용하는 경우가 시간 평균 자료(group-B)를 이용하는 경우에 비해 안정적인 수위 예측 결과를 보여주었다. 그러나 선행시간이 길어짐에 따라 group-A는 저수위(4.0 m 미만)에서 높은 정확도를 보여주었으며 group-B는 상대적으로 고수위(4.0 m 이상)에서 높은 정확도를 보여주었다.

3) 예측 선행시간 12시간 이후부터는 group-A와 group-B 모두에서 예측성능이 급격히 떨어지는 현상이 발생하였다. 이는 수위변동에 영향을 미치는 유역의 도달시간이 짧은 이유에 기인한 것으로 예상되나 이는 향후 추가적인 연구를 통해 과학적인 원인 분석이 필요할 것으로 판단된다. 따라서 복잡하고 다양한 입력자료를 필요로 하고 모형의 수행에 소요되는 시간이 상대적으로 긴 분포형 수문 모델에 기반한 하천수위 예측 대신 수위 관측자료가 충분히 확보된 유역의 경우 향후 상대적으로 간단한 입력자료와 짧은 모의 시간이 소요되는 LSTM 모델을 활용하여 수위 예측을 수행하는 경우 빠르고 정확한 실시간 수위 예측이 가능하여 홍수방어에 필요한 충분한 선행시간을 확보하는 것이 가능할 것으로 판단된다.

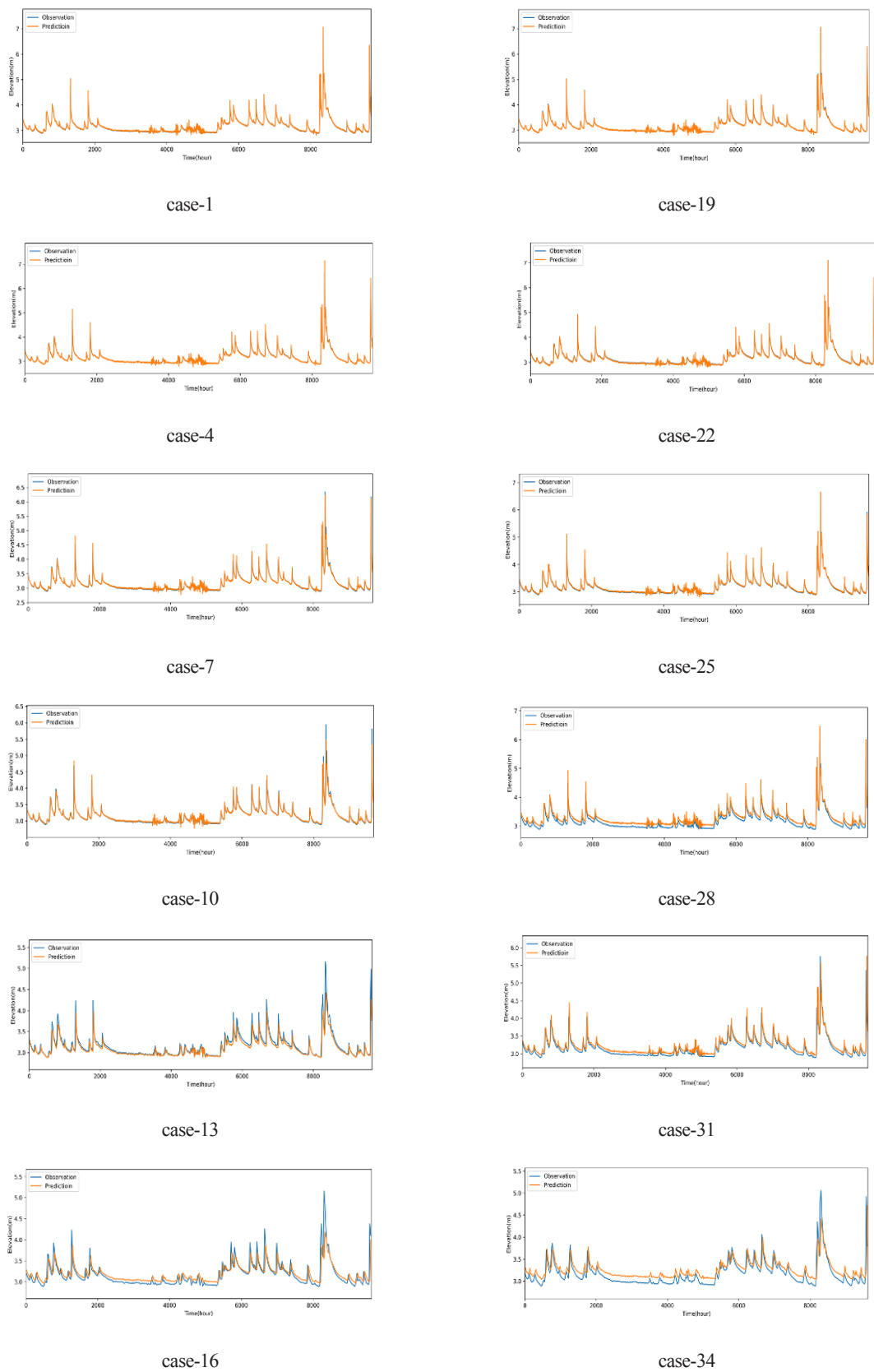
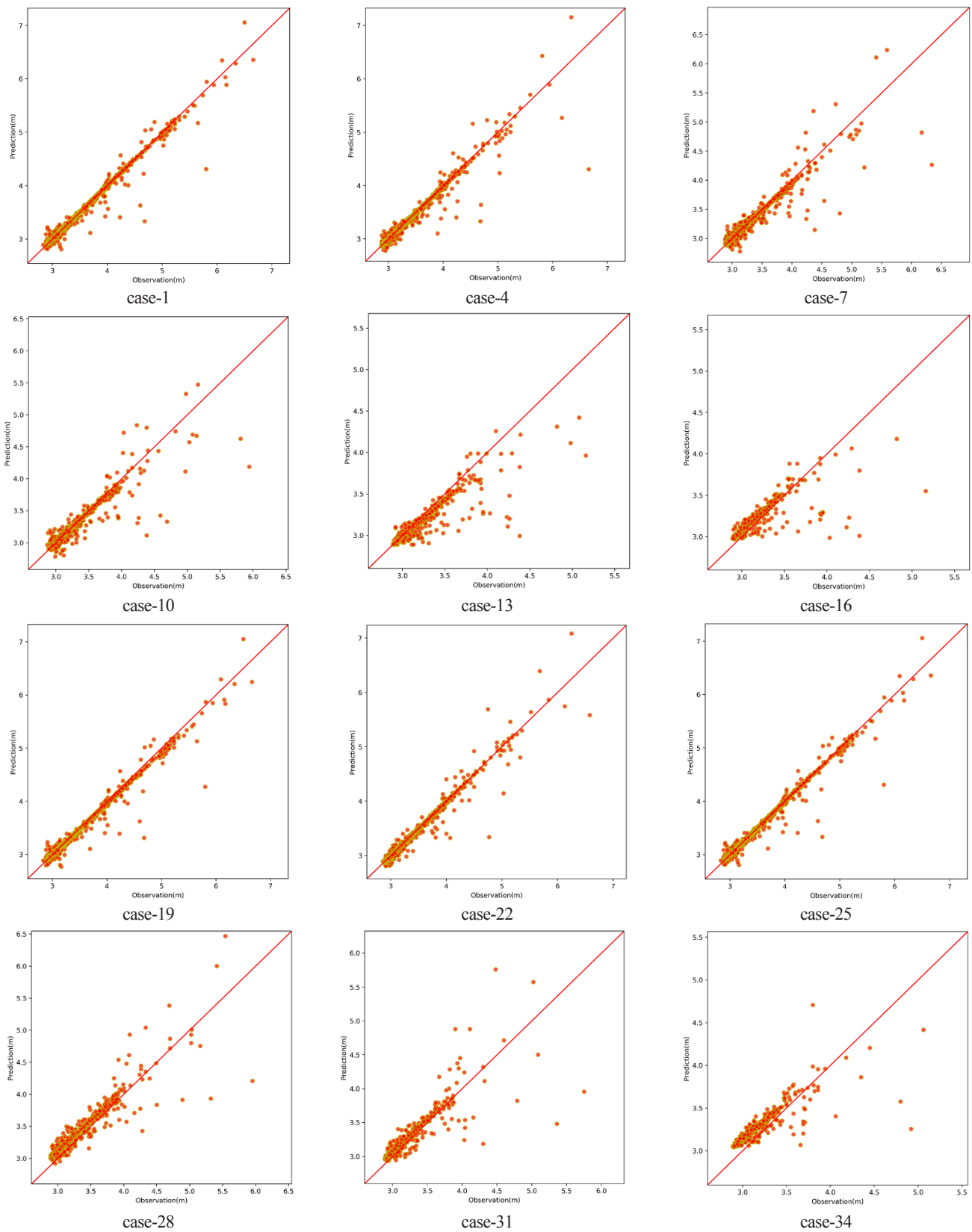


Fig. 4. Comparison of simulation results between group-A (left-side) and group-B (right-side).



**Fig. 5.** Comparison of simulation results between group-A (left-side) and group-B (right-side).

## Acknowledgements

본 결과는 농림축산식품부의 재원으로 농림식품기술기획평가원의 첨단생산기술개발사업의 지원을 받아 연구되었음(과제번호: 318060-3).

## Authors Information

Seungsoo Lee, <https://orcid.org/0000-0002-1537-5417>

Hyunuk An, <https://orcid.org/0000-0002-4566-5159>

Youngteck Hur, <https://orcid.org/0000-0003-1147-5776>

Yeonsu Kim, <https://orcid.org/0000-0003-0893-672X>

Jisun Byun, Chungnam National University, Postdoctoral researcher

## References

- Hinton G, Osindero S, Teh Y. 2006. A fast learning algorithm for deep belief nets. *Neural computation* 18:1527-1554.
- Hochreiter S, Schmidhuber J. 1997. Long short-term memory. *Neural computation* 9:1735-1780.
- Jung S, Cho H, Kim J, Lee G. 2018a. Prediction of water level in a tidal river using a deep-learning based LSTM model. *Journal of Korea Water Resources Association* 51:1207-1216. [in Korean]
- Jung S, Lee D, Lee K. 2018b. Prediction of river water level using deep learning open library. *Journal of the Korean Society of Hazard Mitigation* 18:1-11. [in Korean]
- KMA (Korean Meteorological Administration). 2012. Korean climate change assessment report. KMA, Seoul, Korea.
- Lee G, Jung S, Lee D. 2018. Comparison of physics-based and data-driven models for streamflow simulation of the Mekong river. *Journal of Korea Water Resources Association* 51:503-514. [in Korean]
- Lee H, Jung S, Choi E. 2016. A case study on improving performance of machine learning application and learning algorithm. *Digital fusion research* 14:245-258. [in Korean]
- Liang C, Li H, Lei M, Du Q. 2018. Dongting Lake water level forecast and its relationship with the three Gorges Dam based on a long short-term memory network. *Water* 10:1389.
- Li L, Kim H, Jun K, Choi M. 2016. Estimation of river discharge using satellite-derived flow signals and artificial neural network model: Application to imjin river. *Journal of Korea Water Resources Association* 49:589-597. [in Korean]
- Lim H, An H, Kim H, Lee J. 2019. Prediction of pollution loads in the Geum River upstream using the recurrent neural network algorithm. *Korean Journal of Agricultural Science* 46:67-78. [in Korean]
- McCulloch, Warren S, Walter Pitts. 1943. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics* 5:115-133.
- Olah C. 2015. Understanding lstm networks. GITHUB blog. Accessed in <http://colah.github.io/posts/2015-08-Understanding-LSTMs/> on 1 March 2020.
- Rumelhar D, Geoffrey E, Ronald J. 1986. Learning representations by back-propagating errors. *Nature* 323:533.
- Ryu Y, Shin J, Nam W, Heo J. 2012. Forecast of the daily inflow with artificial neural network using wavelet transform at Chungju Dam. *Journal of Korea Water Resources Association* 45:1321-1330. [in Korean]
- Seo Y, Choi E, Yeo W. 2017. Reservoir water level forecasting using machine learning Models. *Journal of The Korean Society of Agricultural Engineers* 59:97-110. [in Korean]
- Sheng W, Lewis A. 2011. Short-term prediction of an artificial neural network in an oscillating water column. *International Journal of Offshore and Polar Engineering* 21.

- Yeo W, Seo Y, Lee S, Jee H. 2010. Study on water stage prediction using hybrid model of artificial neural network and genetic algorithm. *Journal of Korea Water Resources Association* 43:721-731. [in Korean]
- Yoon E, Park J, Lee J, Shin H. 2018. Reliability evaluations of time of concentration using artificial neural network model-focusing on Oncheoncheon basin. *Journal of Korea Water Resources Association* 51:71-80. [in Korean]
- Zhang D, Lindholm G, Ratnaweera H. 2018. Use long short-term memory to enhance Internet of Things for combined sewer overflow monitoring. *Journal of Hydrology* 556:409-418.