

합리적인 하천수 관리체계 구축을 위한 자료기반 방류량 예측모형 개발

유형주¹ · 이승오² · 최서혜³ · 박문형^{4*}

¹홍익대학교 토목공학과 박사과정, ²홍익대학교 토목공학과 교수, ³한국건설기술연구원 국토보전연구본부 선진연구원,

⁴한국건설기술연구원 국토보전연구본부 수석연구원

Development of a Data-Driven Model for Forecasting Outflow to Establish a Reasonable River Water Management System

Hyung Ju Yoo¹, Seung Oh Lee², Seo Hye Choi³, and Moon Hyung Park^{4*}

¹Ph.D Student, Dept. of Civil Engineering, Hongik University

²Professor, Dept. of Civil Engineering, Hongik University

³Researcher, Korea Institute of Civil Engineering and Building Technology

⁴Senior Researcher, Korea Institute of Civil Engineering and Building Technology

요약

일반적으로 물수지 분석 시 공급에 해당되는 회귀수량의 경우 용수별 회귀율을 일률적으로 정하여 산정하는 방법을 채택하고 있어 정확한 가용유량을 산정하지 못하는 한계를 갖고 있다. 이에 본 연구에서는 회귀수 중 하-폐수에 초점을 두었고 인공지능망 등의 기계학습 모형을 적용하여 하수종말처리장의 방류량 예측 모형을 개발하였다. 시계열 자료예측 시 사용되는 주요 기계학습 모형인 LSTM (Long Short-Term Memory), GRU (Gated Recurrent Units), SVR (Support Vector Regression) 모형을 적용하였으며 관측 값과 예측 값을 비교하는 오차지표를 통하여 방류량 예측의 최적의 모형을 선정하였다. 모형 적용 결과, GRU 모형의 평균제곱근 오차(Root Mean Square Error, RMSE)는 LSTM 모형과 SVR 모형보다 작으며 Nash-Sutcliffe 계수(NSE)는 LSTM 모형과 SVR 모형보다 큰 것을 확인하였고, 이를 근거로 하수종말처리장의 방류량 예측에 최적모형은 GRU 모형이라고 판단하였다. 다만, 극값에서는 예측 값이 과소 및 과대 산정되는 경향을 보여 추후 예측 정확도 향상을 위해서는 극한사상에 대한 추가자료 구축 및 입력 자료의 최소시간단위를 축소하는 것이 필요할 것으로 판단되었다. 또한, 예측하고자 하는 대상지의 용수이용량을 검토하고 계절적 영향을 반영할 수 있는 추가인자를 고려하게 되면 기후변동성에 대비하여 정확한 방류량 예측이 가능하며 예측 결과를 토대로 종합적인 하천수 사용관리 및 물이용 계획 수립을 위한 기초자료로 활용될 수 있을 것으로 기대된다.

핵심용어: 회귀수, 하천수 관리체계, LSTM, GRU, SVR

ABSTRACT

In most cases of the water balance analysis, the return flow ratio for each water supply was uniformly determined and applied, so it has been contained a problem that the volume of available water would be incorrectly calculated. Therefore, sewage and wastewater among the return water were focused in this study and the data-driven model was developed to forecast the outflow from the sewage treatment plant. The forecasting results of LSTM (Long Short-Term Memory), GRU (Gated Recurrent Units), and SVR (Support Vector Regression) models, which are mainly used for forecasting the time series data in most fields, were compared with the observed data to determine the optimal model parameters for forecasting outflow. As a result of applying the model, the root mean square error (RMSE) of the GRU model was smaller than those of the LSTM and SVR models, and the Nash-Sutcliffe coefficient (NSE) was higher than those of others. Thus, it was judged that the GRU model could be the optimal model for forecasting the outflow in sewage treatment plants. However, the forecasting outflow tends to be underestimated and overestimated in extreme sections. Therefore, the additional data for extreme events and reducing the minimum time unit of input data were necessary to enhance the accuracy of forecasting. If the water use of the target site was reviewed and the additional parameters

*Corresponding author: Moon Hyung Park, moon@kict.re.kr

Received: 15 September 2020, Revised: 6 October 2020, Accepted: 24 October 2020



that could reflect seasonal effects were considered, more accurate outflow could be forecasted to be ready for climate variability in near future. And it is expected to use as fundamental resources for establishing a reasonable river water management system based on the forecasting results.

Keywords: Return water, River water management system, LSTM, GRU, SVR

1. Introduction

최근 기후변화로 인해 국내에서는 가뭄 및 홍수피해가 빈번히 발생하고 있고, 가뭄빈도의 경우 여름철은 감소한 반면, 겨울과 봄철에는 가뭄이 오히려 심화되는 추세를 보이는 등 강수량이 시간적·지역적으로 편중되는 경향으로 변화하고 있다(KMA, 2020). 최근에는 가뭄·갈수로 인해 심화되고 있는 용수부족 현상에 대한 대비가 필요하며 종합적인 하천수 사용관리 및 물이용 계획 수립이 필요한 실정이다. 이에 국내에서는 수자원장기종합계획을 통해 물수지 분석을 실시하여 물의 과부족 또는 가용한 물을 정량적으로 평가한다(Lee et al., 2005). 물수지 분석은 하천의 불규칙한 과거 자연유량과 장래 용수 수요의 시기별 분포를 비교, 검토하여 물 부족 여부를 판단하는 계산 과정으로 물수지 분석 시 공급은 자연유량과 댐에서의 용수공급량, 그리고 용수공급 후의 회귀수량으로 이루어지며, 물의 수요는 생활, 공업, 농업, 그리고 하천유지 용수량으로 구성된다(MOLIT, 2016). 여기서, 회귀수량은 물 수요를 충족시킨 후 다시 하천으로 유입되는 물의 양으로 일반적으로 생활·공업용수의 회귀율은 65%, 농업용수의 회귀율은 35%로 하천수 사용관리 측면에서 회귀수량의 정확한 예측은 중요하다(MOLIT, 2016). 그러나 일반적인 회귀수량은 용수별 일률적인 회귀율을 적용하여 산정하는 문제점이 있다(Choi et al., 2018; Oh et al., 2019). 따라서 본 연구에서는 회귀수량 중 하수종말처리장의 방류량에 대하여 정확한 예측을 수행하기 위하여 인공신경망 등의 기계학습 모형을 적용하였다.

기계학습 모형은 국내·외 다양한 분야에서 적용되고 있으며, 수자원 분야에서는 시계열 특성을 갖는 수위, 취수량, 유입량 등의 예측에 사용되고 있다(Yoo et al., 2020; Choi et al., 2019; Lee et al., 2020; Lee et al., 2018; Yoo et al., 2019). 시계열 특성을 갖는 자료 예측에 사용되는 기계학습 모형에는 다양한 모형이 있으며, 국내의 경우, Yeo et al.(2010)은 ANN(Artificial Neural Network) 모형을 이용하여 수위예측을 수행하였고, Tran and Song(2017)은 미국 미국 텍사스 트리니티 강의 수위예측을 수행하기 위하여 RNN(Recurrent Neural Network), RNN-BPTT(Backpropagation Through Time), LSTM(Long Short-Term Memory) 모형을 사용하였으며, 이 중 LSTM 모형이 좋은 성능을 나타냈다고 하였다. Choi et al.(2019)는 SVM(Support Vector Machine)모형을 적용하여 취수량 예측을 하였으며, 1~2달에 대해 중·단기 예측을 수행하였을 경우 예측 정확도가 증가하는 것을 확인하였다. 국외의 경우, Chen et al.(2012)은 하천에서 ANN 기반의 예측 모형을 구축하여 하천의 수위예측을 수행하고 예측정확도를 검토하였다. Zhang et al.(2018)은 LSTM 모형을 적용하여 장기간에 걸친 수위 예측 시, 기존의 FFNN(Feed-Forward Neural Network)모형 보다 예측성능이 뛰어나다는 것을 확인하였다. Zhang et al.(2018)은 하수도 수위 예측을 위하여 순환신경망의 하나인 LSTM 과 GRU(Gated Recurrent Unit) 모형을 적용하였고 GRU 모형의 성능이 우수함을 확인하였다. 선행연구 조사를 통하여 시계열 자료예측에 사용되는 인공신경망의 경향이 ANN 및 RNN 모형에서 최근에는 LSTM, SVM, GRU 모형을 많이 사용하는 것으로 판단되었다. 그러나 LSTM 모형 및 GRU 모형은 극값에서 과소 또는 과대추정되는 경향이 있고(Hochreiter, 1998; Pascanu et al., 2012; Agarap, 2018; Yoo et al., 2020) SVM 모형은 학습에 사용된 자료의 수에 대한 의존도가 높아 예측기간이 증가할수록 예측 정확도가 떨어진다는(Choi et al., 2019)는 한계점이 있어 예측하고자 하는 변수의 특성에 따라 적합한 모형 선정이 필요하다.

따라서 본 연구에서는 하수종말처리장의 방류량 예측에 적합한 모형을 선정하기 위하여 최근에 많이 사용되는 LSTM, SVM, GRU 모형을 적용하였다. 청평댐 유역에 있는 하수종말처리장의 2012~2018년의 방류량자료 및 수문자료를 활용하

였고, 오차 분석을 통하여 방류량 예측의 적합한 모형을 선정하였다.

2. Methodology

2.1 Support Vector Regression Model

Vapnik(1995)에 의해 제안된 서포트 벡터 머신(Support Vector Machine, SVM)은 결정 경계(decision boundary)를 통하여 자료를 분류하는 지도 학습(supervised learning) 모형이다. SVM 모형은 선형 분류와 비선형 분류에서 사용되고, 비선형 분류를 수행하기 위해서 주어진 데이터에 Kernel 함수를 이용하여 데이터의 오분류를 최소화 하는 최적의 결정경계를 선정한다. SVM에 주로 사용되는 Kernel 함수의 종류로는 linear, polynomial, rbf(radial basis function), sigmoid 등이 있으며, 각 데이터 분포 특성에 따라 적합한 Kernel 함수 적용을 통해 비선형 분류를 수행 할 수 있다.

SVR(Support Vector Regression)은 SVM 모형을 회귀 문제에 적용한 방법으로, 주어진 데이터를 Kernel 함수를 통해 결정경계를 구하고 이를 통해 예측하는 기법이다. SVR 모형을 학습하기 위한 기본 식은 아래 Eq. (1) 및 (2)와 같다.

$$f(x) = w^T x + b \quad (1)$$

$$\text{Min. } \frac{1}{2} |w|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*), \text{ s.t. } y_i - wx_i - b < \epsilon + \xi_i \quad (2)$$

여기서, x 는 입력, y 는 출력, $f(x)$ 는 모든 입력 x 에 대해 출력 y 를 가장 적합하게 예측하는 함수를 의미하며, ξ, ξ^* 는 여유 변수, C 는 계수, ϵ 은 실제 값과 예측 값의 허용 오차를 의미한다.

2.2 Long Short-Term Memory Model

LSTM(Long Short-Term Memory) 모형은 Hochreiter와 Schmidhuber에 의하여 제안되었다. Recurrent Neural Network (RNN) 모형에서 발생하는 최적화 오류(optimization hurdle) 및 오차경사의 기울기 소실(vanishing gradient)문제를 해결하고 장기적인 시간에 따른 종속성을 파악하는데 효과적이어서 시계열 자료 예측에 사용되고 있고 다양한 분야에서 연구되고 있다(Hochreiter and Schmidhuber, 1998; Olah, 2018).

LSTM 모형은 RNN 모형과 달리 핵심요소는 시간에 따라 상태(state)를 유지할 수 있는 메모리이동 셀(cell)과 셀 안과 밖으로 데이터의 이동을 조절하는 세 개의 게이트이다(Greff et al., 2016). LSTM 모형은 특정 시간의 상태(h_t)를 갱신하기 위해 셀(cell)의 개념을 도입하여 입력과 현재까지의 상태를 이용하여 내부에 가지고 있는 정보의 갱신여부를 판단한다. 그리고 이 셀의 데이터 이동을 조절하기 위한 게이트의 종류로는 입력게이트(input gate, i_t), 망각게이트(forget gate, f_t) 그리고 출력게이트(output gate, o_t)로 구성되어 있다. 망각 게이트 f_t 는 아래의 Eq. (3)으로 나타낼 수 있으며, 이전 cell의 출력인 h_{t-1} 과 현재의 입력인 x_t 를 시그모이드 함수(sigmoid function)에 적용해 0과 1사이의 값을 얻어 이 값을 현재 상태와 요소끼리 곱하게 되며, 그 과정에서 입력정보의 유지여부를 선택한다.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (3)$$

여기서, σ 는 시그모이드 활성화함수를 나타내며, W_f 는 망각게이트의 가중치, b_f 는 망각게이트의 편향(bias)을 나타낸다.

입력게이트 i_t 는 Eq. (4)와 같이 표현하며, 어떤 입력정보를 셀에 저장할 것인지를 결정하게 되며 시그모이드 함수를 이용하여 어떤 정보를 업데이트할지 결정하고 쌍곡탄젠트 함수(hyperbolic tangent function)를 이용하여 새로운 셀 상태를 업데이트 시 사용되는 후보 셀(N_t)을 아래의 Eq. (5)를 통하여 생성하게 된다.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (4)$$

$$N_t = \tanh(W_n \cdot [h_{t-1}, x_t] + b_n) \quad (5)$$

여기서, W_i, W_n 는 각각 입력게이트와 후보 셀의 가중치, b_i, b_n 는 입력게이트와 후보 셀의 편향을 나타낸다.

다음으로 과거의 셀 상태(C_{t-1})와 후보 셀(N_t)을 아래 Eq. (6)과 같이 조합하여 현재의 셀 상태(C_t)를 업데이트 하게 된다.

$$C_t = f_t \cdot C_{t-1} + i_t \cdot N_t \quad (6)$$

최종적으로 출력게이트 o_t 는 아래의 Eqs. (7) 및 (8)과 같이 시그모이드 함수를 이용해서 셀 상태의 어느 부분을 출력하고자 하는지 결정한다. 최종적으로는 쌍곡탄젠트 함수를 이용하여 활성화된 셀 상태(C_t)와의 곱을 통해 특정 시점의 상태(h_t)를 업데이트 한다.

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (7)$$

$$h_t = o_t \cdot \tanh(C_t) \quad (8)$$

여기서, W_o 는 출력게이트의 가중치, b_o 는 출력게이트의 편향을 나타낸다.

2.3 Gated Recurrent Units Model

GRU(Gated Recurrent Units)모형은 Cho et al.(2014)에 의하여 소개되었으며, LSTM 모형과 같이 게이트 개념을 적용한 순환 신경망의 한 종류이나 게이트의 제어 방식에서 LSTM 모형과 차이가 있다(Cho et al., 2014). GRU 모형은 두 개의 게이트를 이용하여 입력을 연산하며 각각의 게이트는 갱신게이트(update gate, u_t)와 초기화게이트(reset gate, r_t)로 구성되어 있다. 갱신게이트는 과거의 셀 상태와 현재의 셀 상태 중 어느 값에 더 가중치를 주는가를 결정해 주며 Eq. (9)로 나타낼 수 있다.

$$u_t = \sigma(W_u \cdot [C_{t-1}, x_t] + b_u) \quad (9)$$

여기서, σ 는 시그모이드 활성화함수, W_u 는 갱신게이트의 가중치, C_{t-1} 은 과거 시점의 셀 상태, x_t 는 현재 시점의 입력 값, b_u 는 갱신게이트의 편향을 의미한다.

이 후 초기화게이트를 통하여 과거 시점의 셀 상태를 새로운 현재 시점의 출력 후보 셀에 얼마나 포함시킬 것인지 결정한

다. 초기화게이트는 Eq. (10) 및 (11)로 나타낼 수 있다.

$$N_t = \tanh(W_c \cdot [r_t \odot C_{t-1}, x_t] + b_c) \quad (10)$$

$$r_t = \sigma(W_r \cdot [C_{t-1}, x_t] + b_r) \quad (11)$$

여기서, N_t 는 현재 시점의 출력 후보 셀 상태, W_c 와 b_c 는 가중치 및 편향을 나타내고, r_t 는 초기화게이트 값, \odot 는 행렬의 성분 곱 연산을 의미한다. W_r 과 b_r 은 초기화 게이트의 가중치 및 편향을 의미한다.

최종적으로 연산한 결과를 토대로 현재시점의 출력 셀 상태를 결정하며 Eq. (12)로 나타낸다.

$$C_t = u_t N_t + (1 - u_t) C_{t-1} \quad (12)$$

갱신게이트 값 u_t 가 현재 시점의 출력 후보 셀 상태와 과거 시점의 출력 셀 상태의 반영 비율을 결정하는 것을 확인할 수 있으며, u_t 값은 시그모이드 함수의 출력 값이기 때문에 범위가 0~1 사이임을 확인할 수 있다.

2.4 Model Evaluation Criteria

본 연구에서는 구축된 입력자료를 개발된 방류량예측모형을 통하여 학습시킨 후 예측 결과 자료에 대한 통계분석을 수행하여 모형의 정확도를 비교하고자 한다. 모형의 정확도 판단 지표로는 평균제곱근오차(Root Mean Square Error, RMSE)와 Nash-Sutcliffe 계수(NSE)를 사용하여 비교하였다. 각각의 오차 및 계수 산정식은 Eqs. (13) 및 (14)와 같다.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (O_i - P_i)^2}{n}} \quad (13)$$

$$NSE = \left(1 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (O_i - \bar{O}_i)^2} \right) \times 100 \quad (14)$$

여기서 O_i 와 P_i 는 각각 시간 t 에 대하여 관측방류량과 예측방류량을 나타내며, \bar{O}_i 는 관측 방류량의 평균 값, n 은 총 관측 자료의 수를 의미한다.

RMSE의 경우 값이 0에 가까워질수록 예측 값과 관측 값이 잘 일치한다고 판단할 수 있으며, NSE의 경우 값이 1에 가까워질수록 예측 값과 관측 값이 비교적 잘 맞는 것을 확인할 수 있다.

2.5 Autocorrelation, Extended Autocorrelation and Partial Autocorrelation

시계열 자료의 경우, 연속적인 일련의 관측 값들이 서로 상관되어 있을 수 있으며, 시계열의 시차 값(lagged values) 사이의 선형 관계를 측정하는 지표에는 자기상관계수, 편자기상관계수 및 확장자기상관계수가 있다. 자기상관계수는(a_k) 임의

시간 t 에서 특정 시차 k 를 두고 두 관측치, y_t 와 y_{t-k} 가 얼마나 관련 있는 지를 나타내는 척도로 Eq. (15)와 같다.

$$a_k = \frac{\sum_{t=k+1}^n (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^n (y_t - \bar{y})^2} \quad (15)$$

여기서, t 는 임의의 시간, k 는 자기상관함수의 시차, n 은 관측 값의 수, \bar{y} 는 관측 값의 평균값을 의미한다.

편자기상관계수(ϕ_{kk})는 특정 시차 k 이외의 모든 시차를 갖는 관측치 들로부터의 영향력을 배제한 가운데 특정한 두 관측치, y_t 와 y_{t-k} 가 얼마나 관련이 있는 지를 나타내는 척도로 Eq. (16)과 같다.

$$\phi_{kk} = \frac{a_k - \sum_{j=1}^{k-1} \phi_{k-1,j} a_{k-j}}{1 - \sum_{j=1}^{k-1} \phi_{k-1,j} a_j} \quad (16)$$

여기서, j 는 임의의 시간, k 는 자기상관함수의 시차, $\phi_{k,j} = \phi_{(k-1),j} - \theta_{kk} \phi_{(k-1),(k-j)}$ 를, $\theta_{kk} = \text{cor}(\hat{e}_t, \hat{e}_{t-k})$, \hat{e}_t, \hat{e}_{t-k} 는 시차 k 에 대해서 $y_{t-1}, \dots, y_{t-(k-1)}$ 로 y_t 및 y_{t-k} 를 회귀분석 한 잔차(residual)를 의미한다.

확장자기상관계수($ea_{t,k,j}$)는 ARMA 모형 적용 시 ARMA 모형의 가역성(reversibility) 문제를 회피하기 위한 방법으로 Eq. (17) 및 (18)과 같다.

$$ea_{t,k,j} = y_t - \tilde{\zeta}_1 y_{t-1} - \dots - \tilde{\zeta}_k y_{t-k} \quad (17)$$

$$ea_{t,k,j} = \hat{e}_t - \sum_{i=1}^j \lambda_i \hat{e}_{t-i} \quad (18)$$

여기서, t 는 임의의 시간, k, j 는 시차, y_t, y_{t-k} 는 임의의 시간 t 및 특정 시차 k 의 관측 값, $\tilde{\zeta}_k, \lambda_i$ 는 회귀분석을 통한 추정회귀계수, \hat{e}_t, \hat{e}_{t-k} 는 시차 k 에 대해서 $y_{t-1}, \dots, y_{t-(k-1)}$ 로 y_t 및 y_{t-k} 를 회귀분석 한 잔차(residual)를 의미한다.

일반적으로 자기상관계수가 값이 지속적으로 큰 값을 유지할 경우, 해당 시계열 자료는 시간에 따라 값이 일정한 정상성(stationary)의 특징을 갖는다고 간주할 수 있다. 따라서 자기상관 및 편자기상관의 경우 95% 신뢰구간 내에서는 유의수준 5%에서 자기상관계수 및 편자기상관계수가 0이라고 가정을 하는 신뢰구간 개념을 사용하며, 95% 신뢰구간은 Eq. (19)과 같다.

$$95\% \text{ CI} = \pm 1.96 \frac{1}{\sqrt{n}} \quad (19)$$

여기서, CI 는 신뢰구간으로 p -value 기준으로 -1.96 에서 $+1.96$ 사이를 의미한다.

자기상관계수는 신뢰구간 내 값이 분포하면 해당되는 시차 k 전까지의 관측 값이 임의시간 관측 값 y_t 와 상관이 있음을 의미하며, 편자기상관계수의 경우, 신뢰구간 내 값이 분포 하면 해당되는 시차 k 전까지의 관측 값이 임의시간 관측 값 y_t 에 강한 상관관계 보이는 것을 의미한다. 확장자기상관계수는 결과가 표로 제시되며, AR(Autocorrelation) 및 MA(Moving Average)의 정확한 차수(order)를 확인할 수 있다. 일반적으로 각각의 계수들은 통계학, 경제학과 컴퓨터 I/O 패턴 분석, 수질 예측(수온 및 용존 산소 농도), 전력 공급 및 수요 예측 등 ARIMA(Autoregressive Integrated Moving Average) 모형을 이용한 시계열 분석에 널리 사용되고 있다.

3. Study Area and Data Collection

3.1 Study Area

본 연구 대상지인 청평댐유역은 북위 37도32분~37도52분, 동경 127도15분~127도26분에 걸쳐 한반도 중앙부에 위치하고 있으며, 유역면적은 760.61 km²이고, 유로연장은 36.69 km로 청평댐 하류부에는 3개의 주요 하수종말처리장이 건설되어 있다(Fig. 1 참조).

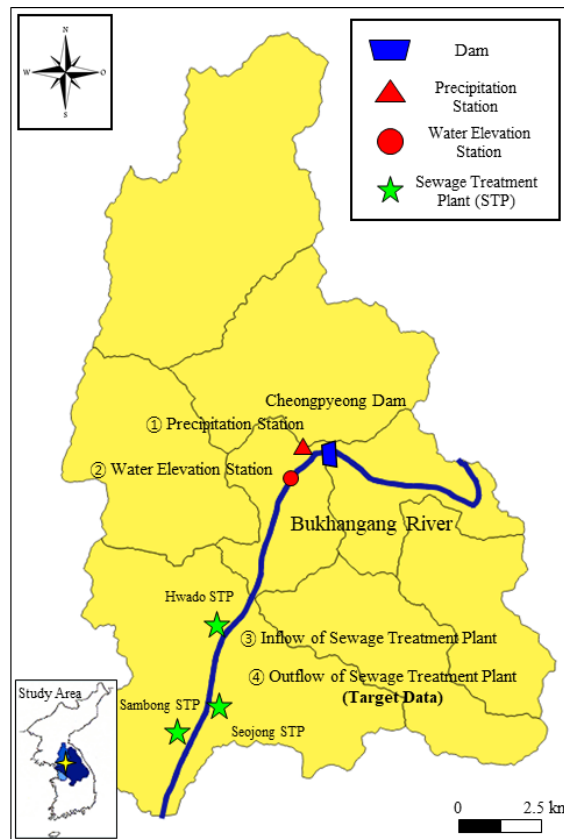


Fig. 1. Study Area

3.2 Input Data

본 연구는 시계열 특성을 가지고 있는 하수종말처리장의 방류량 예측을 위한 최적의 기계학습 모형을 선정하기 위하여 시계열 자료 예측에 사용되는 주요 기계학습 모형(SVR, LSTM, GRU) 별 예측정확도(accuracy of forecasting)를 평가하는 것에 초점을 두었다. 또한 방류량 예측 시 사용하게 될 입력자료의 종류, 입력자료의 양, 입력자료 간의 상관관계를 고려하였다.

방류량 예측 시 사용된 입력자료는 강수량, 신청평대교 수위, 하수종말처리장의 유입량, 방류량이며, 각각 기상청, 한강홍수통제소 및 국립환경과학원에서 일 단위로 2012년부터 2018년까지 7년간 측정된 자료를 이용하였다. 본 연구에서 수집된 자료의 관측소 위치 및 자료 보유 기간 등에 대해서 Table 1에 제시하였으며, 측정된 강수량의 단위는 (mm), 하수종말처리장 유입량 및 방류량은 (m^3/day), 수위는 (m)이다.

Table 1. Information about hydrological stations and sewage treatment plants (STP)

No	Stations	Items*	Latitude	Longitude	Period
1	SinCheongpyeongdaegyo (Way Bridge)	P	37 ° 42'	127 ° 23'	7 years (2012-2018)
2	SinCheongpyeongdaegyo (Way Bridge)	W	37 ° 42'	127 ° 23'	
3	Hwado, Seojong, Sambong STP	I	37 ° 37'	127 ° 20'	
			37 ° 35'	127 ° 20'	
4	Hwado, Seojong, Sambong STP	O	37 ° 34'	127 ° 18'	
			37 ° 37'	127 ° 20'	
			37 ° 35'	127 ° 20'	
			37 ° 34'	127 ° 18'	

* P = Precipitation; W = Water Surface Elevation; I = Inflow; O = Outflow

3.3 *t*-test(*p*-value) and pearson correlation analysis

수집된 자료가 방류량 예측 시 필요한 주요한 인자인지 판단하기 위하여, 구축한 입력자료 별 *t*-test 분석 및 *p*-value를 도출하였다. *t*-test 분석은 두 집단 간의 평균의 차이가 통계적으로 유의한지를 파악할 때 필요한 통계적 기법으로 *t*-test 분석 시 다음의 가정 사항을 만족해야한다.

- ① 자료는 모두 동일 간격을 가진 연속형 수치여야 한다(identical interval and continuity).
- ② 두 집단은 서로 독립적 이어야 한다(independence).
- ③ 자료의 수치는 정규성을 가져야 한다(normality).

또한 예측인자인 방류량과 입력자료 간의 선형적 관계를 분석하기 위하여 상관성 분석을 수행하였다. 입력 자료인 강수량, 신청평대교 수위 및 하수종말처리장의 유입량 및 방류량에 대하여 *t*-test 분석 및 *p*-value, 피어슨 상관계수(Pearson correlation coefficient)를 도출한 결과는 Table 2와 같다.

Table 2. *t*-test analysis and correlation analysis for input data

Input Data*	<i>t</i> -score	<i>p</i> -value	effect size	PCC**	Significant Factor***
P	438.24	~0.00	12.25	0.23	○
W	438.26	~0.00	12.25	0.28	○
I	9.30	~0.00	0.26	0.96	◎
O	9.80	~0.00	0.27	0.91	◎

* P = Precipitation; W = Water Surface Elevation; I = Inflow; O = Outflow

** PCC = Pearson Correlation Coefficient

*** Weight : ○ = 1.0, ◎ = 2.0

일반적으로 p -value의 값이 0.05 이하의 값을 가지면 그 인자는 유의미한 인자로 판단할 수 있으나, 두 집단의 평균 차이를 표현하는 효과크기(effect size)는 p -value의 값에 영향을 크게 미치지 때문에 효과크기의 값을 고려하여 유의미한 인자를 판단해야 한다(Ruxton, 2006). t -test 분석 및 p -value 도출 결과 강수량 및 수위는 하수종말처리장의 유입량 및 방류량에 비해 효과크기 값이 크기 때문에 p -value 값이 작게 산정되었다고 판단되어 방류량 예측 시 유입량, 방류량 자료에 비하여 가중치를 작게 설정하였다.

4. Model Setup

본 연구에서는 기계학습 모형의 실행을 위해 오픈소스 라이브러리를 이용하였다. 컴퓨터 언어는 Python(version 3.7, Anaconda)을 사용하였고, Python 내부에 있는 Numpy(version 1.14.2), Pandas(version 0.22.0) 라이브러리를 이용하여 모형을 실행하기 위한 자료 관리를 수행하였다. 또한 Google에서 오픈소스 라이브러리로 제공하는 TensorFlow(version 1.13.1)를 이용하여 SVR, LSTM, GRU 알고리즘을 활용한 방류량예측 모형을 구축하였다. TensorFlow는 data flow graph 구조로 구성되어 있으며, 기계학습(machine learning) 및 딥 러닝(deep learning)의 다양한 연구 분야에서 활용되어 본 연구에서도 활용이 가능할 것이라고 판단되었다. 또한 방류량예측은 Inter Core i5-6600 CPU(Central Processing Unit), 8.00 GB of RAM(Random Access Memory), 256 GB of SSD(Solid State Drive) 제원을 갖춘 컴퓨터에서 수행되었다.

4.1 Sensitivity Analysis

본 연구에서는 다양한 방류량 예측모형의 성능을 평가하기 위해 앞서 SVR 모형, LSTM 및 GRU 모형 내에 있는 여러 인자에 대하여 민감도 분석을 수행하였다. 각 모형 내에서 설정 가능한 인자는 총 4가지로 각 인자별 민감도 분석을 수행할 Case를 Table 3에 정리하였다. 입력 자료는 1년간의 수문자료 및 하수종말처리장의 유입량 방류량을 사용하였으며, 예측 데이터는 향후 실시간 방류량을 예측하기 위하여 1일 후의 방류량으로 설정하였다. 각 인자별로 30번의 반복과정을 통하여 인자별 불확실성을 표준편차로 나타내었고, CPU Time을 평균 내어 검토하였다.

Table 3. Sensitivity analysis for LSTM, GRU & SVR for forecasting outflow (*: reference value)

Parameter	Model	Value	Evaluation
Hidden Layer	LSTM GRU	1*,3,5,10,20	Standard deviation (Uncertainty) and CPU Time
Learning Rate		0.005,0.01*,0.02,0.05,0.1	
Sequence Length		1*,2,3,5,10	
Iterations		10,100*,1000,5000,10,000	
C	SVR	1, 3*, 5, 10, 100	Standard deviation (Uncertainty) and CPU Time
epsilon		0.01,0.05,0.1*,0.5,1.0	
kernel		linear, poly, rbf*, sigmoid	
tolerance		0.0000001,0.00001,0.001*,0.1	

각 인자별로 최적의 설정 값을 도출하기 위하여 표준편차 값과 CPU Time을 고려하였다. 각 인자별 표준편차 및 CPU Time에 대하여 Fig. 2에 나타내었고, 결과를 Table 4에 요약하여 정리하였다.

민감도 분석 결과 본 연구에서 예측 모형별 선정한 인자별 최적의 값은 LSTM 모형의 경우, Hidden Layer 개수 10개, Learning Rate 0.05, 반복계산 횟수 1000번, 시퀀스 길이는 2일이며, GRU 모형의 경우, Hidden Layer 개수 20개, Learning Rate 0.10, 반복계산 횟수 1000번, 시퀀스 길이는 2일, SVR 모형의 경우, Coefficient 1.0, Epsilon 0.10, Kernel은 radial

basis function, Tolerance 0.001 이다.

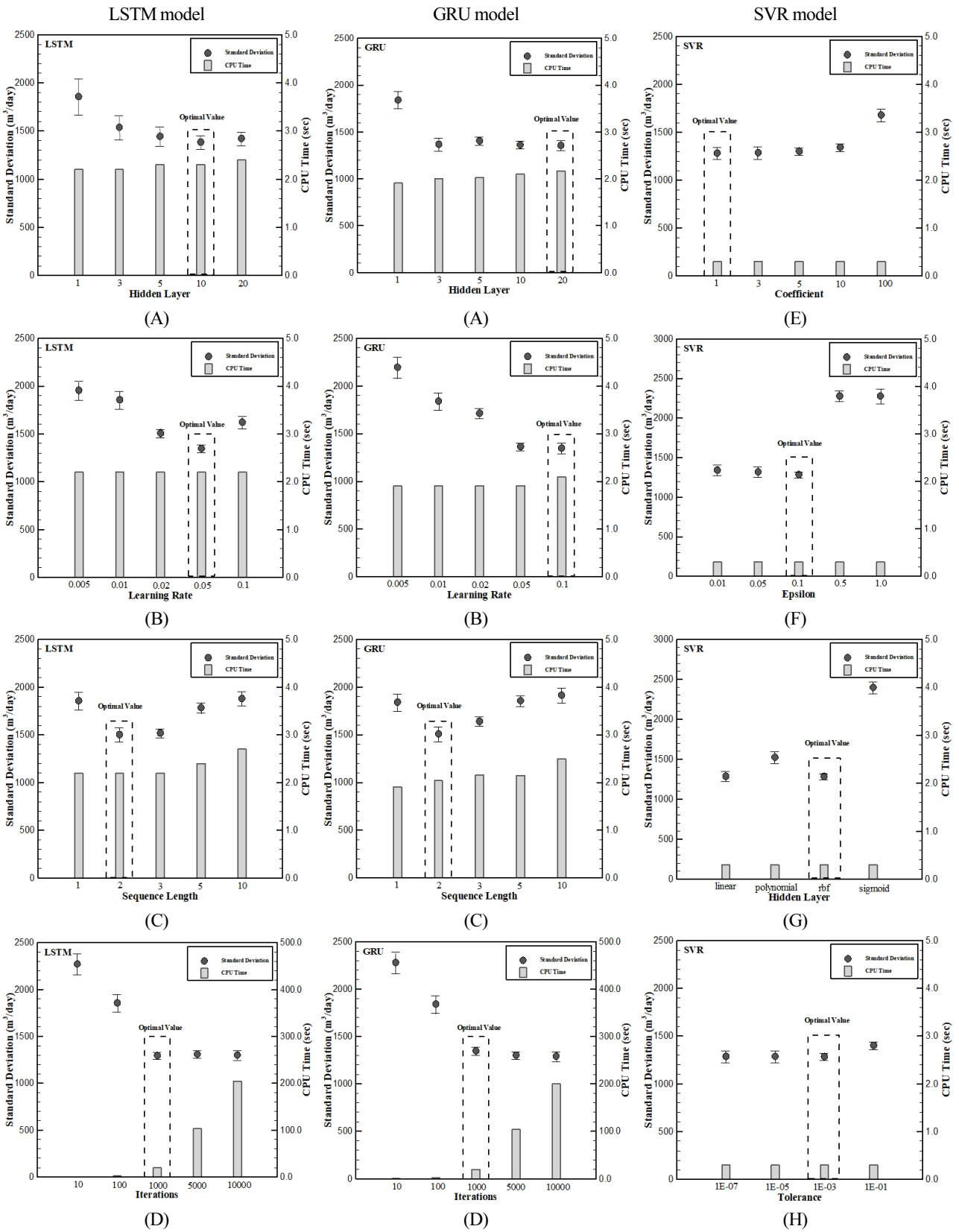


Fig. 2. Results of Sensitivity Analysis (A: Hidden Layer, B: Learning Rate, C: Sequence Length, D: Iterations, E: Coefficient, F: Epsilon, G: Kernel, H: Tolerance)

Table 4. Summary of Sensitivity Analysis Results (LSTM and GRU)

Parameter	Value	Uncertainty (m ³ /day)		CPU Time (s)	
		LSTM	GRU	LSTM	GRU
Hidden Layer	1	1858.29	1842.39	2.2	1.9
	3	1539.92	1368.80	2.2	2.0
	5	1444.92	1405.67	2.3	2.0
	10*	1385.71	1362.85	2.3	2.1
	20**	1424.09	1357.50	2.4	2.2
Learning Rate	0.005	1956.74	2195.74	2.2	1.9
	0.01	1858.29	1842.39	2.2	1.9
	0.02	1507.38	1714.72	2.2	1.9
	0.05*	1348.23	1366.06	2.2	1.9
	0.10**	1622.78	1350.09	2.2	2.0
Sequence Length	1	1858.29	1842.39	2.2	1.9
	2*,**	1503.39	1509.37	2.2	2.0
	3	1520.32	1641.64	2.2	2.1
	5	1785.88	1857.33	2.4	2.1
	10	1880.99	1915.30	2.7	2.5
Iterations	10	2274.08	2280.95	0.3	0.2
	100	1858.29	1842.39	2.2	1.9
	1000*,**	1297.30	1293.21	20.1	19.7
	5000	1311.20	1315.21	102.9	101.6
	10000	1301.32	1303.72	204.5	199.9

*: Optimal value for outflow forecasting model (LSTM)

**: Optimal value for outflow forecasting model (GRU)

Table 5. Summary of Sensitivity Analysis Results (SVR)

Parameter	Value	Uncertainty (m ³ /day)	CPU Time (s)
Coefficient	1*	1283.56	0.3** (Each parameters did not affect CPU Time in SVR model)
	3	1286.85	
	5	1303.10	
	10	1345.39	
	100	1681.27	
Epsilon	0.01	1342.43	
	0.05	1322.31	
	0.10*	1286.85	
	0.50	2279.90	
Kernel	1.00	2279.90	
	linear	1287.85	
	polynomial	1526.09	
	rbf*	1286.85	
	sigmoid	2399.2	
Tolerance	0.000001	1287.75	
	0.00001	1287.74	
	0.001*	1286.85	
	0.1	1402.75	

*: Optimal value for outflow forecasting model (SVR)

4.2 Model Cases

앞서 민감도 분석을 통하여 각 예측 모형별 도출한 매개변수의 최적 값을 적용하였다. 본 연구는 각각의 기계학습 모형별 하수종말처리장의 방류량 예측 정확도를 비교하여 최적의 모형을 선정하기 위하여 모의 Case를 설정하였고 Table 6에 정리하였다. 모의에 사용한 입력 자료는 앞서 언급하였던 강수량, 신청평대교 수위, 각 하수종말처리장 유입량 및 방류량이며 각각 총 7년간의 자료를 연구대상지에 있는 관측소에서 확보하였다.

Table 6. Cases for application of forecasting model

Case	STP	Model
I -(1)		SVR
I -(2)	Hawdo	LSTM
I -(3)		GRU
II -(1)		SVR
II -(2)	Sambong	LSTM
II -(3)		GRU
III -(1)		SVR
III -(2)	Seojong	LSTM
III -(3)		GRU

시나리오에 따라 모의를 수행하기 위하여 총 7년간의 입력자료 중 6년간의 자료(2009년 ~ 2017년)를 학습 데이터 (training data set)로 사용하였으며, 1년간의 자료(2018년)를 평가 데이터(test data set)로 설정하였다. 방류량예측 모형의 정확도 및 학습속도의 향상을 위하여 입력 자료의 정규화(normalization)를 수행하였다. 정규화 방법은 MinMax 정규화를 사용하였으며 이를 통하여 입력자료의 rescaling(0~1)을 수행하였고 본 모의에 적용하였다.

5. Results and Analysis

본 연구에서 개발한 모형에 대하여 학습 데이터와 평가 데이터는 독립적으로 적용되었다. 학습 데이터를 통하여 학습 후 평가데이터 적용을 통해 모형의 예측결과와 관측결과를 비교하여 Case 별 방류량 예측모형의 성능평가를 RMSE 및 NSE로 비교하여 Table 7에 결과를 정리하였다. 또한 모형별 예측결과와 관측결과를 비교하여 Fig. 3과 Fig. 4에 제시하였다.

Table 7. Summary of the results according to network models

Case	RMSE (m ³ /day)	NSE
I -(1)	1,425	0.61
I -(2)	830	0.87
I -(3)	804	0.88
II -(1)	89	0.54
II -(2)	39	0.91
II -(3)	37	0.92
III -(1)	158	0.41
III -(2)	93	0.80
III -(3)	80	0.85

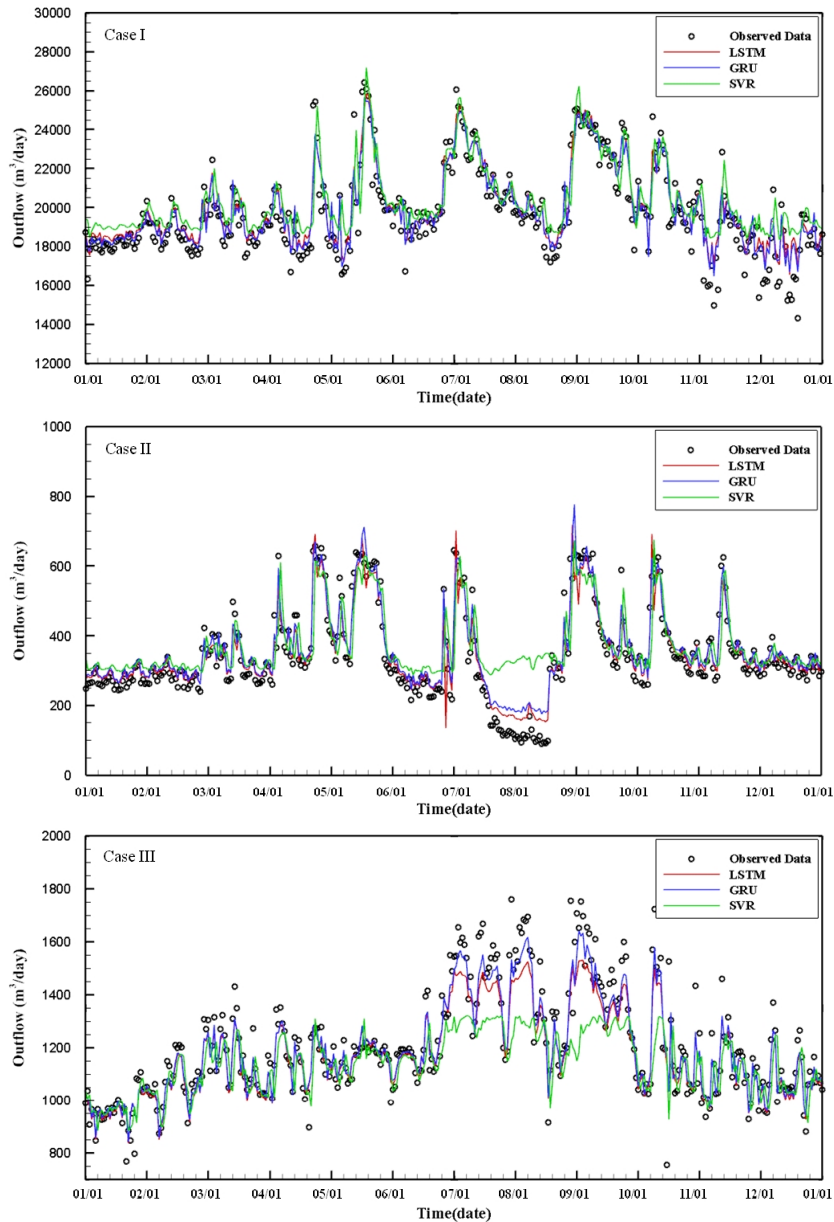
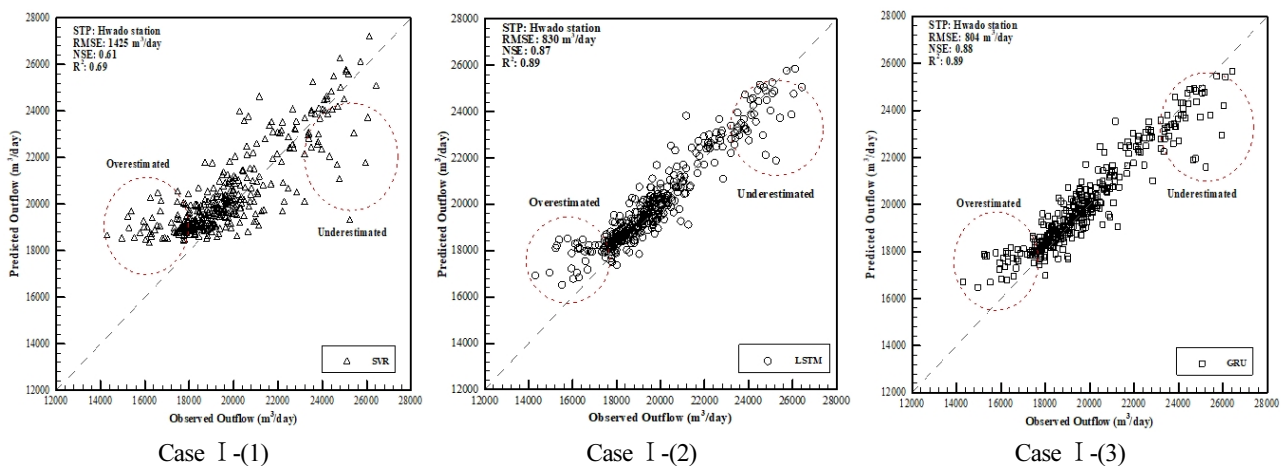


Fig. 3. Results of the outflow forecasting according to network models



Case I-(1)

Case I-(2)

Case I-(3)

Fig. 4. Scatter plot according to network models

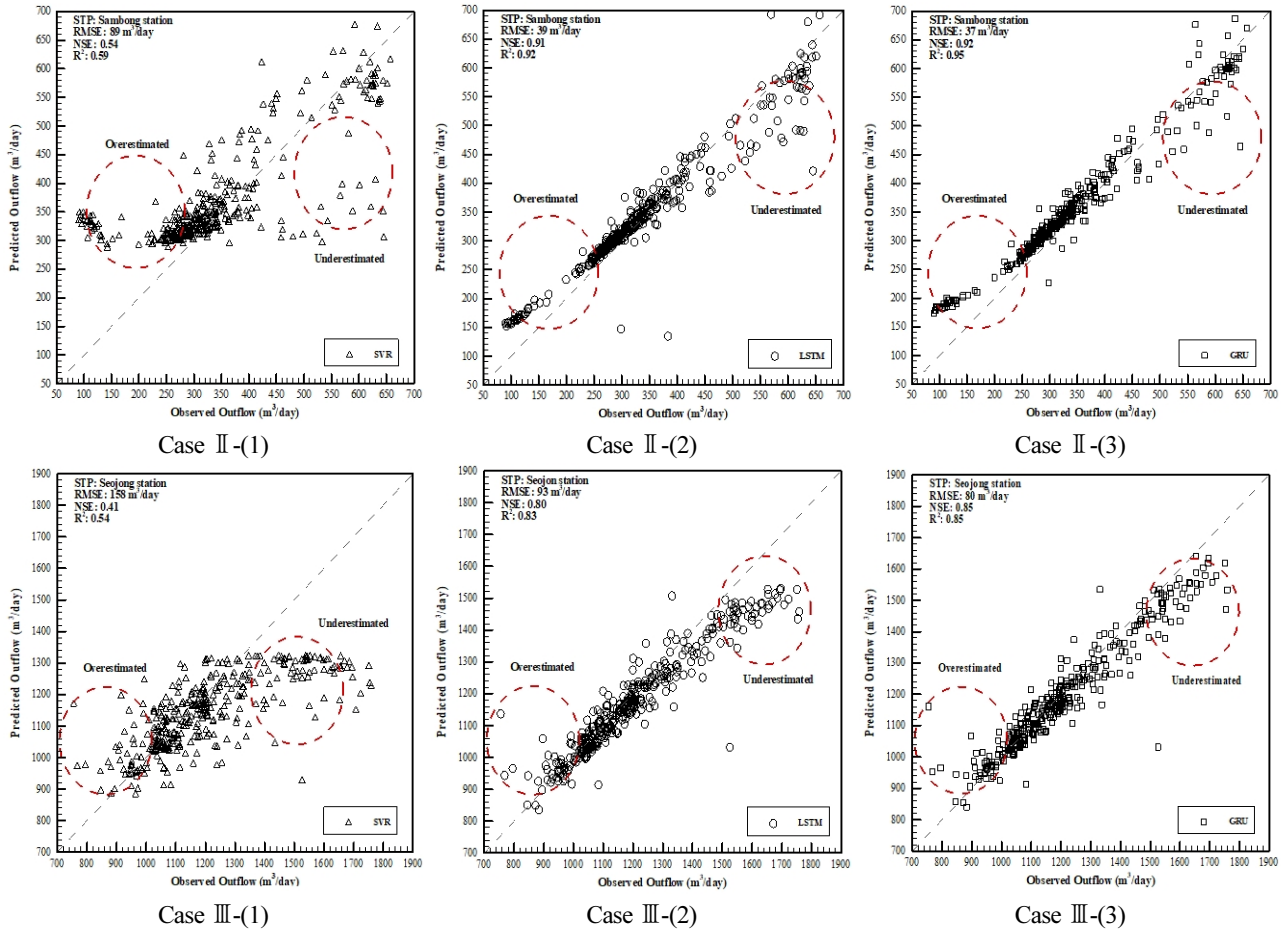


Fig. 4. Scatter plot according to network models (Continue)

SVR, LSTM, GRU 모형의 성능을 평가하기 위하여 예측방류량 결과에 대하여 평가지표(RMSE 및 NSE)를 검토하였으며, SVR 모형은 화도, 삼봉, 서중 하수처리장의 방류량에 대한 RMSE가 각각 1,425, 89, 158 m³/day, NSE가 0.61, 0.54, 0.41로 나타났고, LSTM 모형은 RMSE가 각각 830, 39, 93 m³/day, NSE가 0.87, 0.91, 0.80로 나타났다. GRU 모형의 경우, RMSE가 각각 804, 37, 80 m³/day, NSE가 0.88, 0.92, 0.85로 방류량 예측 정확도 측면에서 가장 우수한 것으로 확인하였다. LSTM과 GRU 모형의 오차가 SVR 모형에 비하여 작은 원인은 방류량의 자기상관성 분석 시 2일 전까지의 방류량이 유의한 상관성이 있는 것으로 분석되었으며(Fig. 5 참고) 이후의 선행시간을 고려할 경우, 예측 정확도가 떨어질 것으로 판단되는데 LSTM과 GRU 모형의 경우, 각각 게이트(LSTM의 경우 input, forget, output, GRU의 경우, reset, update)를 통하여 시퀀스를 효율적으로 기억해줄 수 있기 때문에 모형의 성능이 뛰어난 것으로 판단된다. 반면, SVR 모형의 경우 LSTM, GRU 모형처럼 과거자료의 비중이 적용되지 않고 학습에 사용된 전체 자료를 통해 최적의 예측모형을 구현하기 때문에 오차가 크게 발생하는 것으로 판단된다.

GRU 모형이 LSTM 모형 보다 예측정확도가 높은 원인은 LSTM 모형의 경우 출력게이트(output gate)를 통하여 현재 예측이 다음 예측에 얼마나 영향을 줄 것인지 정하게 되고, 학습에서 발생된 오차가 중첩되어 예측값의 오차가 커질 수 있기 때문이다(Fig. 4 참고). 반면, GRU 모형은 출력게이트가 존재하지 않기 때문에 과거의 예측 값에만 영향을 받게 되고 또한 본 연구에서 구축된 입력 자료의 수가 적기 때문에 적은 입력자료 수에서도 학습이 가능한 GRU 모형의 예측정확도가 LSTM 모형보다 우수하다고 판단된다(Chung et al., 2014; Jozefowicz et al., 2015).

그러나 본 연구에서 적용된 모든 예측 모형은 극값 주변에서 과소 및 과대 산정되는 결과를 확인하였다. 이는 현재 구축되

어진 LSTM 및 GRU 모형의 일반적인 기울기 손실(vanishing gradient)로 인한 문제이며, 이에 대한 문제를 해결하기 위해서는 극한사상(event)에 대한 입력 자료의 수가 많이 구축되어야 하며, 현재 구축된 자료의 최소 시간 단위를 일 단위에서 시 단위 또는 분 단위로 축소시켜야 될 것으로 판단된다. 또한 본 연구 대상지의 경우, 국가수자원관리종합정보시스템(WAMIS)에서 제공하는 용수이용량을 검토 시, 생활용수 및 농업용수의 사용량이 전체 용수이용량의 94.7%(농업용수 - 53.7%, 생활용수 - 41.0%)에 해당되기 때문에 계절적 영향을 반영할 수 있는 인자를 추가적으로 고려하게 된다면 방류량 예측 정확도를 향상시킬 수 있을 것으로 판단된다.

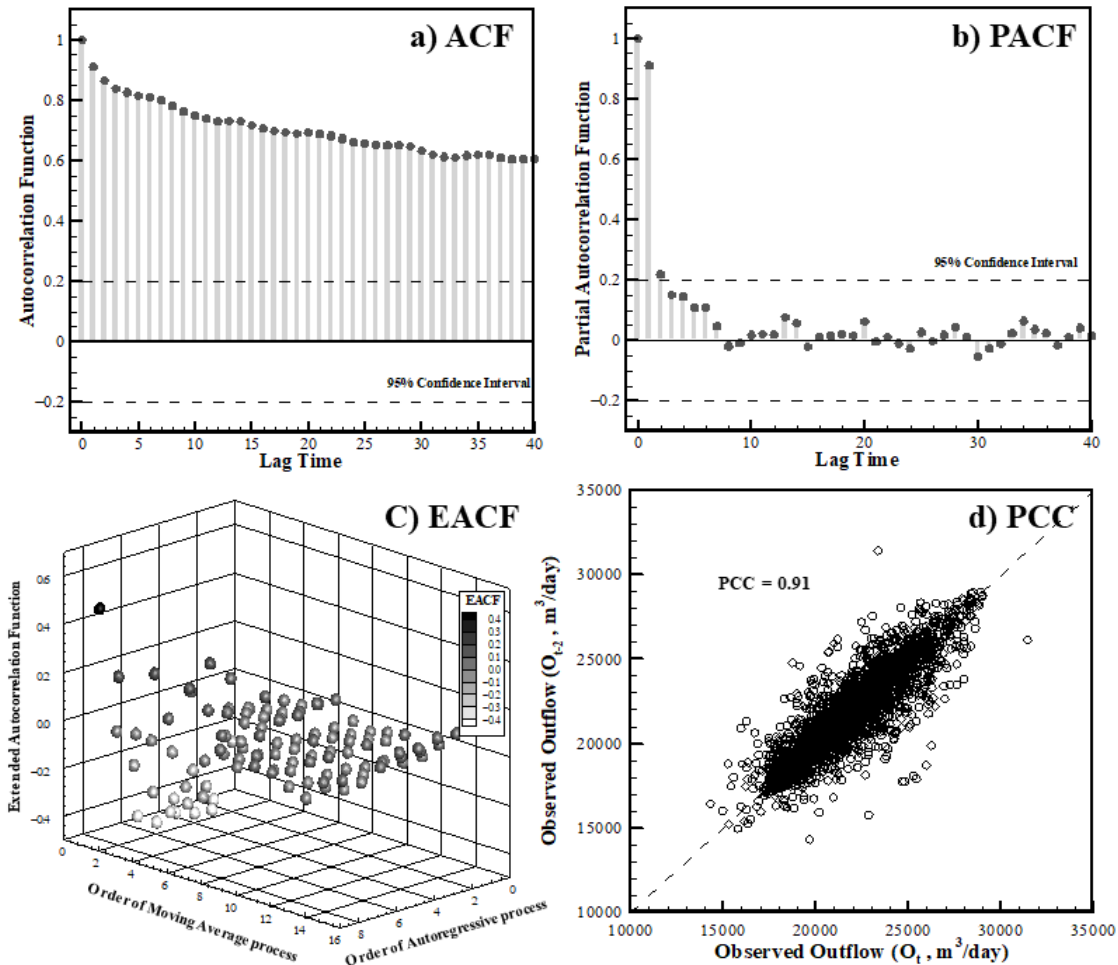


Fig. 5. a) Autocorrelation, b) Partial Autocorrelation, c) Extended Autocorrelation, d) Pearson Correlation of outflow at sewage treatment plant

6. Conclusions

본 연구에서는 하천수의 합리적 사용을 위한 관리체계 개선의 일환으로 하천 회귀수 중 하·폐수에 초점을 맞추어 하수종말처리장의 정확한 방류량 예측을 위한 최적의 기계학습 모형 선정을 목적으로 하고 있으며, 이를 위해 청평댐 유역에 존재하는 하수종말처리장을 대상으로 딥러닝 오픈소스 소프트웨어 라이브러리인 TensorFlow를 활용하여 방류량예측 모형을 개발하였다.

기계학습 모형은 시계열 자료의 예측에 주로 사용되는 LSTM, GRU 및 SVR 모형을 사용하였고 2012년부터 2018년까지의 1일 단위의 강수량, 신청평대교 수위, 화도, 삼봉, 서중 하수종말처리장의 유입량과 방류량 자료를 이용하였다. LSTM,

GRU 및 SVR 모형의 매개변수는 민감도분석을 통하여 LSTM 모형의 경우, 은닉층의 개수는 10개, 학습속도는 0.05, 학습회수는 1,000번, 시퀀스 길이는 2일, GRU 모형의 경우, 은닉층의 개수는 20개, 학습속도는 1.0, 학습회수는 1,000번, 시퀀스 길이는 2일, SVR 모형의 경우, 계수(C)는 1.0, 예측허용 오차는 0.1, Kernel 함수는 radial basis function, 공차는 0.001로 결정하여 모의를 수행하였다. 최종적으로 각 하수종말처리장에 대해 기계학습 모형에 따른 방류량 예측 정확도를 평가하기 위해 관측 방류량과 예측 방류량을 비교·분석을 수행한 후 다음과 같은 결과를 도출하였다.

첫째, 방류량의 자기상관성 분석을 통하여 2일 전까지의 방류량이 현재 방류량을 예측하는데 유의하다고 판단하여 과거의 2일 전까지의 방류량을 기억하도록 시퀀스를 설정하였다. 그 결과 GRU 모형 적용 시 화도, 삼봉, 서종 방류량에 대한 예측 정확도는 RMSE가 각각 804, 37, 80 m³/day, NSE가 0.88, 0.92, 0.85로 SVR 모형과 비교하여 RMSE는 각각 621, 52, 78 m³/day 감소하였으며, NSE는 0.27, 0.38, 0.44 향상하였고 LSTM 모형과 비교하여 RMSE는 각각 26, 2, 13 m³/day 감소하였으며, NSE는 각각 0.01, 0.01, 0.05 정확도 향상 효과를 확인하여 하수처리장의 방류량 예측 시에는 GRU 모형이 적합하다고 판단하였다. 이는 SVR 모형의 경우, 학습의 사용된 전체 입력 자료를 통해 최적의 예측모형을 구현하기 때문에 오차가 크게 발생한 것이고, LSTM 모형의 경우 모형 구조 상 출력 게이트의 가중치에 따라 발생 오차가 중첩될 수 있기 때문에 오차가 GRU 모형에 비해 크게 발생한 것으로 판단된다.

둘째, GRU 모형 및 LSTM 모형은 일반적인 기율기 손실 문제로 인하여 극값 주변에서 과소 및 과대 산정되는 결과를 확인하였으며 이는 모형의 특성 상 극한사상(event)에 대한 입력 자료의 수가 많이 구축하고, 구축된 자료의 최소시간 단위를 축소시키면 해결될 것으로 판단된다.

기존의 물수지 분석 시 용수별 회귀율을 통하여 일률적으로 회귀수량을 산정하는 방법과 비교하여 기계학습을 통하여 본 연구에서 구축한 방류량 예측 모형을 적용할 경우 상대적으로 정확한 방류량 예측을 통하여 합리적인 하천수 관리체계 구축이 가능할 것으로 판단된다. 향후에는 예측하고자 하는 대상지의 용수이용량 검토 및 계절적 영향을 반영할 수 있는 인자를 추가적으로 고려하고 자료의 최소시간 단위를 축소시켜 본 연구에서 제안한 모형의 예측 결과를 이용하면, 기후변동성에 대비하여 실시간 가용수량을 보다 정확히 확인할 수 있으며, 최종적으로 종합적인 하천수 사용관리 및 물이용 계획 수립을 위한 기초자료로 활용될 수 있을 것으로 기대된다. 이외에도 SNS 등을 통하여 국민들에게 신속하게 정보를 전달한다면 향후 물 부족에 따른 문제점을 완화시킬 수 있을 것으로 판단된다.

감사의 글

본 연구는 국토교통부/국토교통과학기술진흥원의 지원으로 수행되었음(과제번호 20AWMP-C140010-03).

References

- Agarap, A. F. (2018). Deep Learning using Rectified Linear Units (RELU). arXiv preprint arXiv:1803.08375.
- Chen, W. B., Liu, W. C., and Hsu, M. H. (2012). Comparison of ANN Approach with 2D and 3D Hydrodynamic Models for Simulating Estuary Water Stage. *Advances in Engineering Software*. 45(1): 69-79.
- Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., and Bengio, Y. (2014). Learning Phrase Representations using RNN Encoder-decoder for Statistical Machine Translation. arXiv Preprint arXiv:1406.1078.
- Choi, S., Kang, S., Lee, D., and Kim, J. (2018). A Study on Water Supply and Demand Prospects for Water Resources Planning. *Journal of the Korean Society of Hazard Mitigation*. 18(7): 589-596.
- Choi, S. H., Kwon, H. H., and Park, M. (2019). Prediction on the Amount of River Water use using Support Vector Machine with Time Series Decomposition. *Journal of Korea Water Resources Association*. 52(12): 1075-1086.

- Chung, J., Gulcehre, C., Cho, K., and Bengio, Y. (2014). Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. arXiv preprint arXiv: 1412.3555.
- Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., and Schmidhuber, J. (2016). LSTM: A Search Space Odyssey. *IEEE Transactions on Neural Networks and Learning Systems*. 28(10): 2222-2232.
- Hochreiter, S. (1998). The Vanishing Gradient Problem During Learning Recurrent Neural Nets and Problem Solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*. 6(02): 107-116.
- Jozefowicz, R., Zaremba, W., and Sutskever, I. (2015, June). An Empirical Exploration of Recurrent Network Architectures. In *International Conference on Machine Learning* (pp. 2342-2350).
- Korea Meteorological Administration (KMA). (2020). Korea Climate Change Assessment Report 2020. Seoul: KMA.
- Lee, G., Jung, S., and 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(6): 503-514.
- Lee, J. Y., Kim, H. I., and Han, K. Y. (2020). Linkage of Hydrological Model and Machine Learning for Real-time Prediction of River Flood. *Journal of The Korean Society of Civil Engineers*. 40(3): 303-314.
- Lee, S. W., Kim, Y. O., and Lee, D. R. (2005). Quantifying Uncertainty for the Water Balance Analysis. *Journal of Korea Water Resources Association*. 38(4): 281-292.
- Ministry of Land Infrastructure and Transport (MOLIT). (2016). National Water Resources Plan (2011~2020) (3rd revision). Sejong: MOLIT.
- Oh, J. H., Ryu, K. S., Bok, J. S., Jang, Y. S., Bae, Y. D., and Lee, B. G. (2019). Water Supply-and-Demand Analysis Considering the Actual Water-Use System in the East Basin of Han River. *Journal of the Korean Society of Hazard Mitigation*. 19(7): 529-543.
- Olah, C. (2018). Understanding Lstm Networks, August 2015. URL <https://colah.github.io/posts/2015-08-Understanding-LSTMs>.
- Pascanu, R., Mikolov, T., and Bengio, Y. (2012). Understanding the Exploding Gradient Problem. *CoRR*, abs/1211.5063, 2, 417.
- Ruxton, G. D. (2006). The Unequal Variance t-test is an Underused Alternative to Student's t-test and the Mann – Whitney U test. *Behavioral Ecology*. 17(4): 688-690.
- Tran, Q. K. and Song, S. K. (2017). Water Level Forecasting based on Deep Learning: A use Case of Trinity River-Texas-The United States. *Journal of KIISE*. 44(6): 607-612.
- Vapnik, V., Guyon, I., and Hastie, T. (1995). Support Vector Machines. *Mach. Learn.* 20(3): 273-297.
- Yeo, W. K., Seo, Y. M., Lee, S. Y., and Jee, H. K. (2010). Study on Water Stage Prediction using Hybrid Model of Artificial Neural Network and Genetic Algorithm. *Journal of Korea Water Resources Association*. 43(8): 721-731.
- Yoo, H. J., Kim, D. H., Kwon, H. H., and Lee, S. O. (2020). Data Driven Water Surface Elevation Forecasting Model with Hybrid Activation Function—A Case Study for Hangang River, South Korea. *Applied Sciences*. 10(4): 1424.
- Yoo, H. J., Lee, S. O., Choi, S., and Park, M. (2019). A Study on the Data Driven Neural Network Model for the Prediction of Time Series Data: Application of Water Surface Elevation Forecasting in Hangang River Bridge. *Journal of Korean Society of Disaster and Security*. 12(2): 73-82.
- Zhang, J., Zhu, Y., Zhang, X., Ye, M., and Yang, J. (2018). Developing a Long Short-Term Memory (LSTM) based Model for Predicting Water Table Depth in Agricultural Areas. *Journal of Hydrology*. 561: 918-929.
- Zhang, D., Lindholm, G., and Ratnaweera, H. (2018). Use Long Short-term Memory to Enhance Internet of Things for Combined Sewer Overflow Monitoring. *Journal of Hydrology*. 556: 409-418.

Korean References Translated from the English

- 국토교통부 (2016). 수자원장기종합계획(2001~2020) 제3차 수정계획. 세종: 국토교통부.
- 기상청 (2020). 한국 기후변화 평가보고서 2020 – 기후변화 과학적 근거 –. 서울: 기상청.
- 여운기, 서영민, 이승윤, 지홍기 (2010). 인공신경망과 유전자알고리즘의 결합모형을 이용한 수위예측에 관한 연구. *한국수자원학회논문집*. 43(8): 721-731.

- 오지환, 류경식, 복정수, 장연석, 배영대, 이봉국 (2019). 실제 물이용 체계를 고려한 한강 동해 권역 물수급 평가. 한국방재학회 논문집. 19(7): 529-543.
- 유형주, 이승오, 최서혜, 박문형 (2019). 시계열 자료의 예측을 위한 자료 기반 신경망 모델에 관한 연구: 한강대교 수위예측 적용. 한국방재안전학회논문집. 12(2): 73-82.
- 이기하, 정성호, 이대엽 (2018). 메콩강 유출모의를 위한 물리적 및 데이터 기반 모형의 비교·분석. 한국수자원학회논문집. 51(6): 503-514.
- 이승욱, 김영오, 이동률 (2005). 물수지 분석을 위한 불확실성 정량화. 한국수자원학회논문집. 38(4): 281-292.
- 이재영, 김현일, 한건영 (2020). 수문모형과 기계학습을 연계한 실시간 하천홍수 예측. 대한토목학회논문집. 40(3): 303-314.
- 최서혜, 권현한, 박문형 (2019). TDSVM을 이용한 하천수 취수량 예측. 한국수자원학회논문집. 52(12): 1075-1086.
- 최시중, 강성규, 이동률, 김중훈 (2018). 수자원 계획 수립을 위한 물 수급 전망 개선 방안 연구. 방재학회논문집. 18(7): 589-596.
- 트란 광 카이, 송사광 (2017). 딥러닝 기반 침수 수위 예측: 미국 텍사스 트리니티강 사례연구. 정보과학회 논문지. 44(6): 607-612.