

Artificial Intelligence Technique in Hydrological Forecasts Supporting for Water Resources Management of a Large River Basin in Vietnam

Truong Van Anh

Faculty of Meteorology and Hydrology, Hanoi University of Natural Resources and Environment, Hanoi, Vietnam

Email: tvanh@hunre.edu.vn

How to cite this paper: Anh, T.V. (2023) Artificial Intelligence Technique in Hydrological Forecasts Supporting for Water Resources Management of a Large River Basin in Vietnam. *Open Journal of Modern Hydrology*, 13, 246-258.
<https://doi.org/10.4236/ojmh.2023.134014>

Received: August 6, 2023

Accepted: October 22, 2023

Published: October 25, 2023

Copyright © 2023 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Hydrological forecasting plays an important role in water resource management, supporting socio-economic development and managing water-related risks in river basins. There are many flow forecasting techniques that have been developed several centuries ago, ranging from physical models, physics-based models, conceptual models, and data-driven models. Recently, Artificial Intelligence (AI) has become an advanced technique applied as an effective data-driven model in hydrological forecasting. The main advantage of these models is that they give results with compatible accuracy, and require short computation time, thus increasing forecasting time and reducing human and financial effort. This study evaluates the applicability of machine learning and deep learning in Hanoi water level forecasting where it is controlled for flood management and water supply in the Red River Delta, Vietnam. Accordingly, SANN (machine learning algorithm) and LSTM (deep learning algorithm) were tested and compared with a Physics-Based Model (PBM) for the Red River Delta. The results show that SANN and LSTM give high accuracy. The R-squared coefficient is greater than 0.8, the mean squared error (MSE) is less than 20 cm, the correlation coefficient of the forecast hydrology is greater than 0.9 and the level of assurance of the forecast plan ranges from 80% to 90% in both cases. In addition, the calculation time is much reduced compared to the requirement of PBM, which is its limitation in hydrological forecasting for large river basins such as the Red River in Vietnam. Therefore, SANN and LSTM are expected to help increase lead time, thereby supporting water resource management for sustainable development and management of water-related risks in the Red River Delta.

Keywords

Hydrological Forecast, Water Resources Management, Machine Learning, Deep Learning, Red River Basin

1. Introduction

An accurate hydrological forecast plays an important role in flood and water supply management of a river delta [1] [2]. In the traditional way, physical-based models have been applied to simulate and forecast the flow of the river delta which is usually a very complicated system with dense rivers and canals [3] [4]. However, this approach has posed two difficulties: the first one is the highly required computation time leads to reduce the forecasted time then affect the end user who based on this information takes action; and the second one is it requires many data and information to develop the model then lead to the problem of financial [5] [6] [7].

Nowadays, AI techniques can improve the problem that physical-based model pose via machine learning or deep learning algorithms [8] [9] [10] [11]. This kind of technique can be classified as a data-driven model in which the relationship between inputs and outputs is built through the set of neurons in different layers so-called hidden layers. The parameters of the model have been tuned through the network to get their optimal value during the training process. After that, this algorithm can be used to predict the flow of a control section in the river with a very fast response. Usually, it takes a few minutes to estimate the output. Due to this advantage, many researchers developed machine learning or deep learning algorithms in hydrology.

The Radial basic function (RBF) has been applied to forecast the river flow at High Aswan DAM on the Nile River [12]. They use a Genetic algorithm for input selections that will feed to RBF to predict the flow. The results are good in terms of the coefficient of determination (R^2) in all cases around 0.9 and the good performance of the forecasted flows when they were compared with the observed ones. Another application developed a newly based AI Flow Difference Model (FDM) to forecast the flow in a cold-regional river. The results have been compared with the Regression Model (RM) and Base Difference Model (BDM) for better performance. All evaluated criteria such as R^2 , Nash coefficient, and RMSE give good values. A comparison between physical-based model and machine learning has been tested by forecasting the flow in the Bogota River basin in Colombia using HEC-RAS and ANN [13]. The results can help to make a better mark for ANN when comparing the performances of the forecast leveled with the observed one in terms of R^2 , MAE, and RMSE.

Red River Delta is a dense and complex system combined with rivers and canals feeding the 13 integrated irrigation districts [14]. The dynamic of the system is based on the inflows coming from the release of three big reservoirs and some natural flow from some uncontrolled rivers in the upstream and the tide coming

from the East Sea [15]. The water level of Red River in Hanoi usually used to monitor the water supply, navigation and flood control of the delta. Therefore, it needs to be forecast in advantage to inform the current state of the system and help the stakeholder make the right decision in their action of water usable.

Besides, at the basin management level, to find out the best policy to apply in the basin, it needs to run an optimal algorithm which in turn generates a set of suitable policy for the choices of decision-makers and stakeholders [16]. In this case, using physical-based model will pose a big problem of computation effort that cannot be helped to simulate all sets of policy for the race of policy selection. Therefore, an emulator as a data-driven model will be very helpful and reasonable to use in this case.

The main objective of this study was to develop an emulator that could be used to make short-term river level forecasting at Hanoi in Red River Delta. A machine learning algorithm SANN and a deep learning algorithm LSTM have been tested and compared with the observation data and the result of physical-based model MIKE 11 of the system. In addition, this research uses the input interactive feature ranking (IIFS) algorithm to select the candidates as the inputs for the network.

2. Case Study and Methodology

2.1. Case Study

Red river basin is the second biggest transboundary basin in Vietnam. Its area is about 169,000 km² cover a part of China, Lao and Vietnam before reaching to East Sea (Figure 1).



Figure 1. Red river basin from HydroSHEDS.

Vietnam's part is located at the downstream of Red-Thai Binh basin and accounts for about 51% of its total area. In this area, there is a dense river network which feeds by three big tributaries including Da River, Thao River and Lo-Gam River. There are four big strategic reservoirs (Hoa Binh, Son La, Thac Ba, Tuyen Quang) among these others located in this area which has the role of managing the water resources for the entire delta.

In the delta, there are 13 irrigation districts and 11 provinces including Hanoi. Therefore, water resources are one of the biggest issues in this region. The water resources of the delta can be evaluated via the water level in the Red River at Hanoi Hydrological station (see **Figure 2**). The dynamic flow in this section is influenced mainly by the inflow from the upper part (from left to right in **Figure 2** including Hoa Binh Reservoir, Thao River, Thac Ba Reservoir, Lo River, Tuyen Quang Reservoirs), the flow in Thai Binh river system connected by Duong river, and the tides from 9 estuaries from East sea.

2.2. Methodology

In this study, a Shadow Artificial Newron (SANN) and Long Short Term Memories (LSTM) have been applied to forecast the water levels at Hanoi in the Red River. These models' performance was also compared with the result from physical-based model MIKE 11 which has been setup for the whole Red River delta.

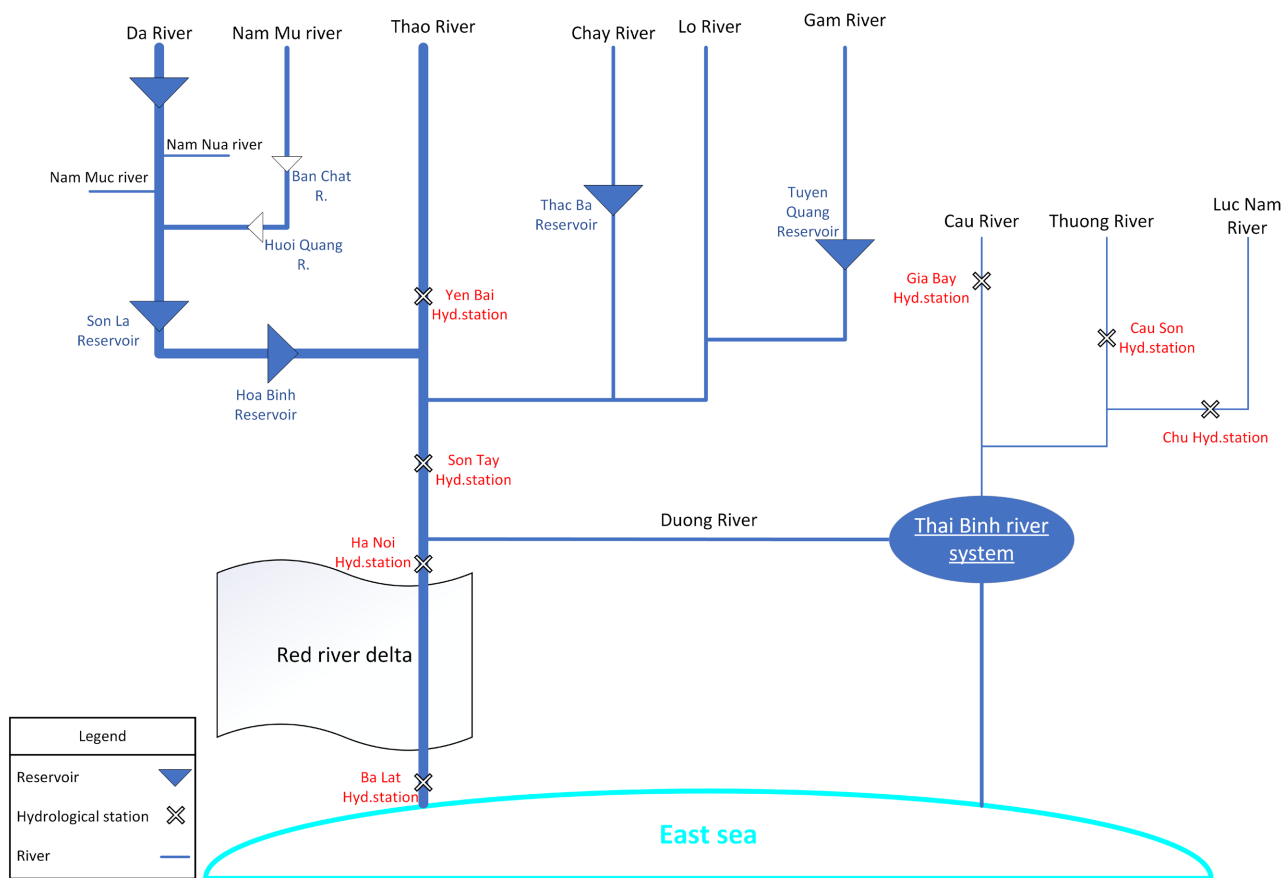


Figure 2. The diagram of the Red River system in Vietnam part.

SANN is a simple machine-learning algorithm inspired by the structural and functional human brain. SANN is composed of one input layer, one hidden layer and one output layer. Each layer contains a number of neurons (nodes) which is the core of the system. Each neuron in the hidden layer presents a linear classifier through the input via function.

$$n_j = \sum_{i=1}^k w_{ij}x_i + b_j \tag{1}$$

In which:

- x_i is the i^{th} input in the input layer;
- n_j is the j^{th} neuron of the hidden layer;
- w_{ij} is the weight values of the i^{th} inputs through the j^{th} neuron;
- b_j is the bias values of the j^{th} neuron.

Then, n_j will pass through an activation function to represent the non-linear relationship between input and output and give the estimated output. In this case, the tansig function has been applied.

$$y = \text{tansig}(n) = \frac{2}{1 + \exp(-2 * n)} - 1 \tag{2}$$

Due to the fact that in hydrology, an input is the sequence values of a variable. Therefore, they usually have an auto correlation with its previous values. In this case, SANN takes the previous target values in the target sequences into the input vector to keep the memory of the sequences. However, this will give the challenge of optimizing the lag time steps of the sequence. That is why in this research, Long Short-Term Memory (LSTM) has been tested and made a comparison with the SANN to evaluate the importance of the long-term memory in the hydrological sequences. LSTM is a deep learning algorithm as it considers the memory of the previous time steps of input and output sequences. In LSTM architecture, there are three gates namely input gate, forget gate, and output gate is going in one LSTM cell to produce the output of cell, as shown in **Figure 3**.

In this case, gates activation functions are the logistic sigmoid function as below:

$$\sigma(x) = \frac{2}{1 + \exp(-x)}$$

and the output activation function is the hyperbolic tangent function

$$c(x) = h(x) = \tanh(x)$$

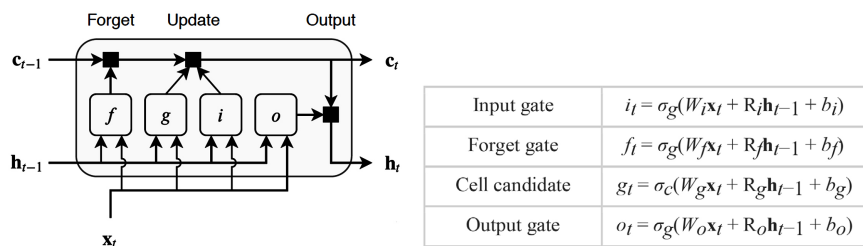


Figure 3. LSTM architecture modified from Matlab documentation.

The number of inputs nodes was defined by an input selection algorithm namely Interactive Input Selection (IIS) [17]. This algorithm helps to choose the first set of input candidates based on a nonlinear statistical measure of significant and then they will be test the correlation with the target by mean of the Single Input-Single Output model to choose a further set of inputs which then feed into the Multi Inputs-Single Output model to select the final set of input for modeling.

To evaluate the performance of the algorithms, statistical indicators including Root Mean Square Error (RMSE-S), Mean Error (ME), Absluted Mean Error (AME), Maximum Absluted Error (MAE), the determination coefficient (R^2). If we call the square deviation of the recorded is σ , the assurance level of forecast (P) which estimates the numbers of forecasts matched actual values over the forecasts considering the allowable error Scp equal to 0.2σ , and the forecast bias estimated by the fraction of RMSE over standard deviation of recorded $\frac{S}{\sigma}$, as shown in Table 1.

2.3. Data Set

The data was collected from 1994, when the Hoa Binh reservoir started in operation, to 2019. The Thac Ba reservoir was in operation in the 1970s, Tuyen Quang reservoir in late 2008 and Son La reservoir in late 2012. For that reason, we have had very low water level since 2011 (Figure 4). Therefore, the training dataset has been created carefully which should include “almost” the situation that happens in the “current condition”. To generate those datasets, the last three years (2017-2019) have been exchanging with “three more natural years” (1997-1999). Then, we choose the first 17 years for training, the next 5 years for validating and the last three year for testing the designed network.

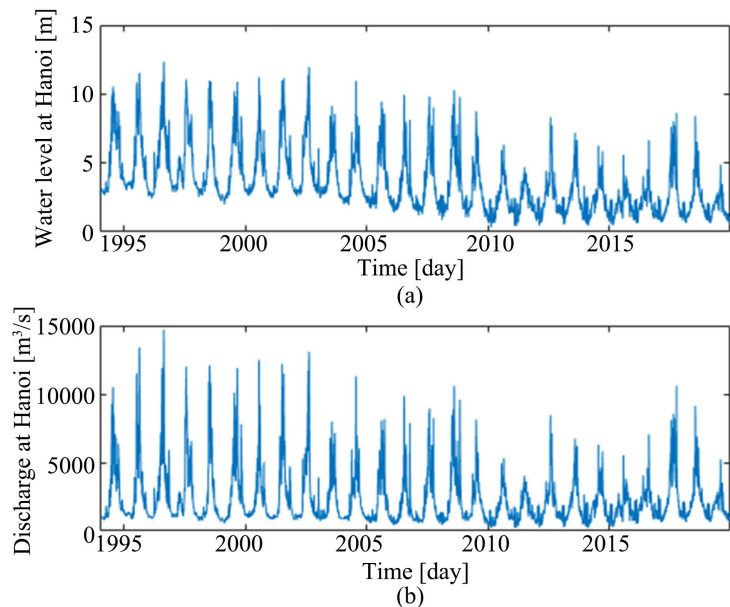


Figure 4. The Hanoi water level (a) and discharge (b) from 1994-2019.

Table 1. The evaluated indicators for forecast assessment.

Accuracy level	S/σ	R^2	$P(\%)$
Good	<0.4	>0.9	>90
Medium	0.4 - 0.6	0.8 - 0.9	75 - 90
Low level	0.6 - 0.8	0.6 - 0.8	60 - 75
Not achieved	>0.8	<0.6	<60

3. Result and Discussion

3.1. Input Variable Selection

There are fifty-four input candidates for the system. They are three releases from three strategic reservoirs in the upstream part of the system including Hoa Binh, Thac Ba and Tuyen Quang; the inflows from other “natural tributeries” such as Thao river, Lo river, Pho day river, Cau river, Thuong river, Luc Nam river; the discharge in the delta at Son Tay and Ha Noi at current states and previous states; the tides from East sea via 9 estuarine of Red—Thai Binh system; and 40 lateral flows going into the system. Applying an input variable selection (IIS) [17] show the effect of 40 lateral flows on the water level in Hanoi is minor. Therefore, the first step we choose first 20 remaining candidates fed to LSTM to train the system. We got further filter to 7 candidates with a very high correlation with the output. These candidates will be the input to the SANN algorithm. The evaluation of IIS algorithm will be postponed after we see the result of all AI algorithms tested on the system.

3.2. SANN

Based on the number of input candidates, the form of SANN applied in this study is the following

$$h_{t+1}^{HN} = f(h_t^{HN}, R_t^{HB}, R_t^{TQ}, R_t^{TB}, Q_t^{YB}, Q_t^{HY}, \tau_t)$$

where

- h_t^{HN} : is the water level at Hanoi section on Red River in day t ;
- R_t^{HB} : is the release from Hoa Binh reservoir on Da river in day t ;
- R_t^{TQ} : is the release from Tuyen Quang reservoir on Gam river in day t ;
- R_t^{TB} : is the release from Thac Ba reservoir on Lo river in day t ;
- Q_t^{YB} : is the discharge at Yen Bai on Thao river in day t ;
- Q_t^{HY} : is the discharge at Ham Yen on Gam river in day t ;
- τ_t : is the maximum tide level at Balat in day t .

The total number of input and output is 8, therefore 10 neurons have been used in the hidden layer. The performance reported in **Table 2** and **Figure 5** clearly shows a model for predicting the water level at Hanoi section with the leading time of 1 day ahead.

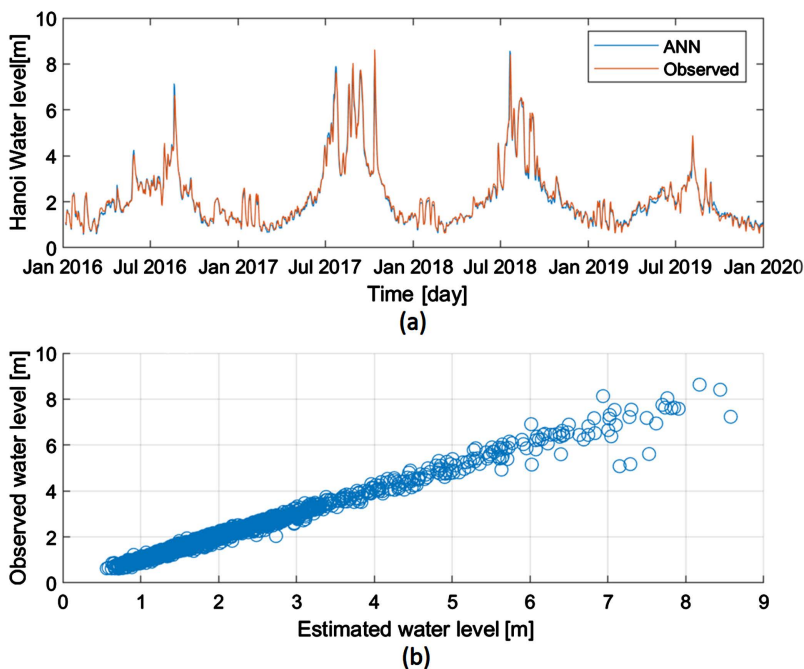


Figure 5. The Hanoi water level estimated by SANN comparing with observed one in the test dataset: (a) for hydrograph and (b) for scatter plot.

Table 2. The evaluated indicators of SANN model.

Dataset	RMSE	ME	AME	MAE	σ	Scp	S/σ	R^2	P
Training	0.17	0.00	0.11	3.07	2.18	0.27	0.08	1.00	90.73
Testing	0.17	-0.03	0.11	2.12	1.37	0.27	0.13	0.99	95.05

3.3. LSTM

Taking advantage of LSTM, the 19 inputs have been used. They are the release from three reservoirs, two inflows from Thao and Lo river, the water level at Hanoi at the previous time step, 9 tides, three lateral flows to Thai Binh river system including inflows to Cau river (Gia Bay gauged station), Thuong river (Cau Son gauged station) and Luc Nam river (Chu gauged station). LSTM can keep both long-term and short-term memory of previous states through the hidden units (neurons) to the current state. Therefore, they can take advantage of all available data and decide which information will go further and which will be left at the forget gates. We used 30 neurons for 19 inputs and one output network. The selection of the final neuron network using the training iteration with the lowest validation loss. This option gave a “perfect” validation result as shown in **Figure 6**.

The training and testing results are better when compared with SANN (**Table 3** and **Figure 7**). Note that LSTM seems to be more sensitive to the dataset. Changing the normalization method will give different results from “bad” to “good”. In this study, we recommend z-score’s normalization which gives us the best network of LSTM for predicting the water level at Hanoi.

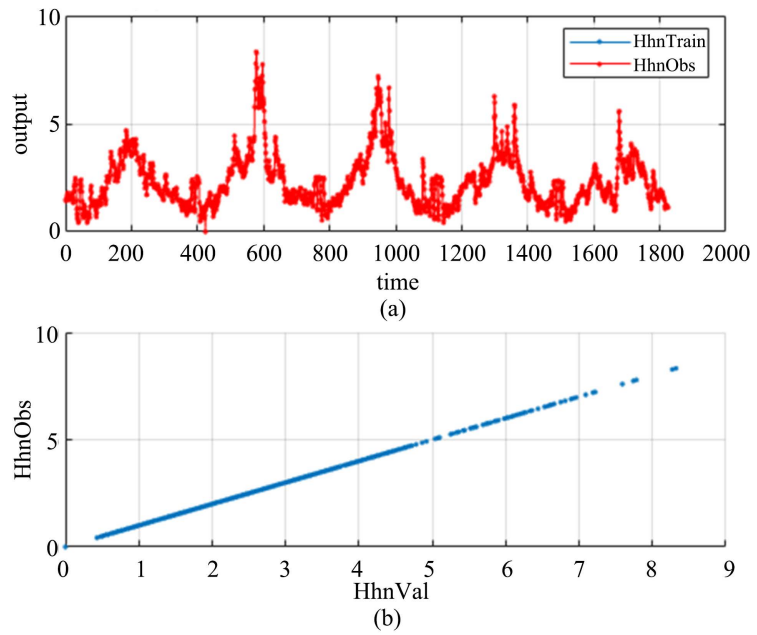


Figure 6. The Hanoi water level estimated by LSTM compared with observed one in the validated dataset: (a) for hydrograph and (b) for scatter plot.

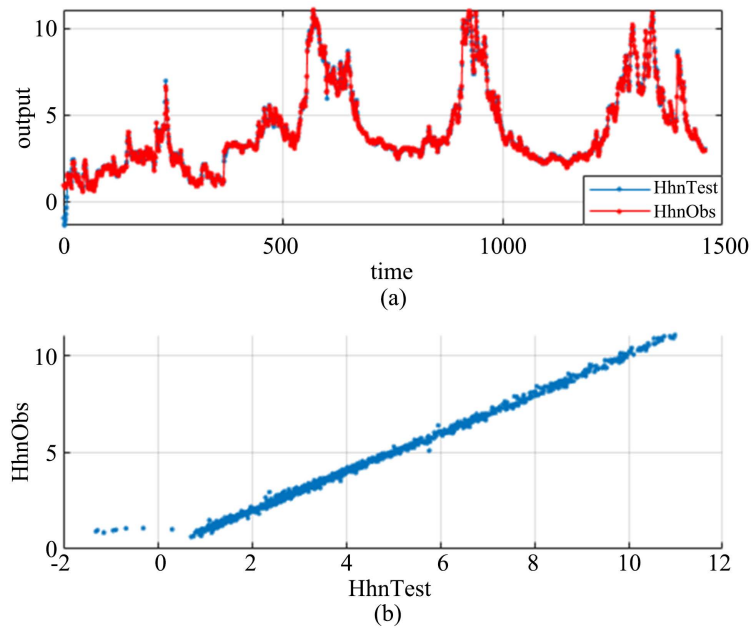


Figure 7. The Hanoi water level estimated by LSTM comparing with observed one in the test dataset: (a) for hydrograph and (b) for scatter plot.

Table 3. The evaluated indicators of LSTM model.

Dataset	RMSE	ME	AME	MAE	σ	Scp	S/σ	R^2	P
Training	0.11	-0.01	0.07	2.97	2.20	0.27	0.05	1.00	99.61
Validating	0.00	0.00	0.00	0.00	1.21	0.27	0.00	1.00	100.00
Testing	0.14	-0.01	0.08	1.59	2.33	0.27	0.06	1.00	99.11

3.4. Comparison and Discussion

The main purpose of this study is to compare three types of models: physical-based model (MIKE 11), machine learning model (SANN) and deep learning model (LSTM). The results show that the best one of three is LSTM. Since LSTM can take the states from previous steps, it shows the advantage of simulating the current situation while looking back the history. That is why the estimated water level is perfectly matched the observed one in whole simulation period. The second one is SANN, the estimated profiles matched the recorded ones. However, due to the form of the network, it cannot give the memory from a long previous time to the current situation. This led to the delay one-time step at almost high peaks as shown in **Figure 8**.

Finally, the physical based model MIKE 11 seems simulate the system very well in recently time but not in the past (see the purple line in **Figure 9**). In the first years from 1994 to 2008, MIKE 11 giving the under-estimated result. This is due to the fact that the data set of cross sections collected in 2010 cannot mimic the system in the past which actually should have different set of cross sections. In this case study, after the big reservoir went to operation, the situation of river bed intrusion occurs with very high speed. This make the water level reduce sharply. Therefore, the result of MIKE 11 in the year near 2010 is better and comparable with other models.

Therefore, in the similar situation, data driven model, especially machine learning and deep learning, have their advantage in simulating the changing-system and give a better forecast while the physical based model need more data and time to correct the mimic system for the prediction problem. After this test, we recommend to use LSTM in the hydrological forecast since it can keep the advance of nonlinear correlated depending variables of a changing system.

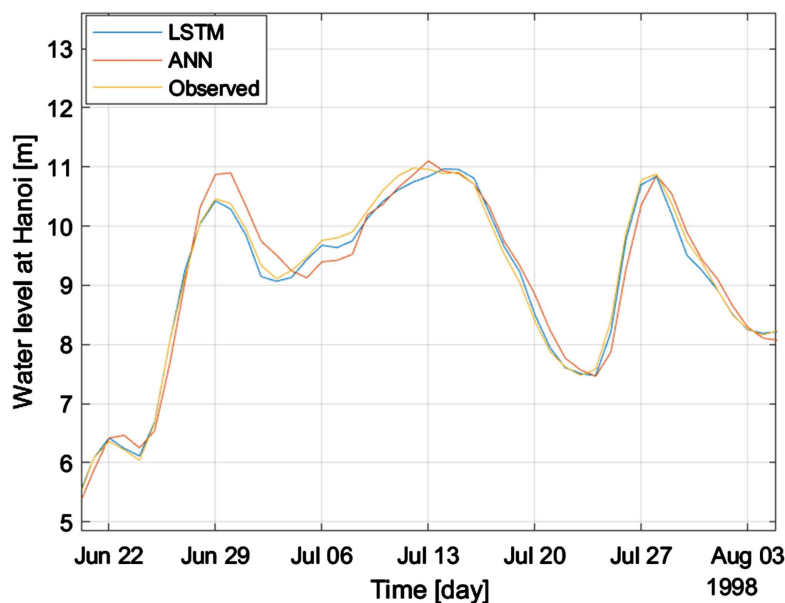


Figure 8. Comparison between SANN and LSTM.

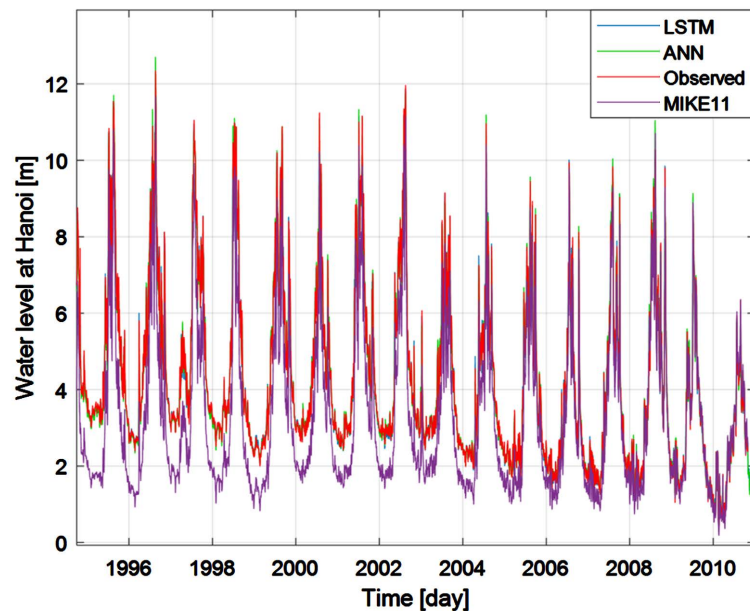


Figure 9. Comparison among LSTM, SANN and MIKE 11 versus observation.

4. Conclusions

The process of flow routing in the dense river delta always needs to be monitored and forecasted for the better management of water resources in such areas. However, this process is a complex problem for hydraulic modeling. River bed and morphology change lead to the performances of physically based models and conceptual models were badly used as the case analysed in this study with R^2 got only 0.85 and the forecasted accuracy level P got only 23.6%. Recently, machine learning and deep learning have been applied for various simulation and prediction problems in hydrology. In this study, we have tested the simple SANN and a popular deep learning algorithm LSTM to forecast the water level at the control station of the Red River system, the second biggest basin in Vietnam; and compared it against the result of a physical-based model MIKE 11. In general, LSTM model is the best among SANN and MIKE 11 due to its characteristic of long memory applied in the high correlation system as in the usual hydrology case. However, SANN model is still a good performance and much better than the physical-based model in this case study where the upstream reservoirs system made change of riverbeds downstream and let the reduction of water level over time. This should happen in almost all intervention river basins in the world and pose the challenge to hydrologists to simulate or forecast the hydrological variables there. It is the advancement of AI-based techniques that has created revolution in hydrology [18].

The result shows that, LSTM can predict correctly over 99% of test cases, SANN got 98% for the training and 97% for testing while MIKE 11 got only 23.5%. Other evaluation indicators such as RMSE, ME, AME of both SANN and LSTM models are smaller than 10% of the prediction and mean; and R^2 is nearly 1 in both models compared to 0.85 of MIKE 11.

In conclusion, we recommend using LSTM and other deep-learning techniques in hydrology. LSTM can use all type of data and have the advantage of exploiting all relevant information for them to train the network for a certain problem. In the next study, we would like to try it with more complex situations such as a longer lead time in the forecasting and other studies in different fields of hydrological and hydraulic systems.

Acknowledgements

This research received funding from the Vietnam Ministry of Natural Resources and Environment entitled “Applying Artificial Intelligence (AI) for flow routing to support water resources allocation in the river basin, testing on the Red River basin”, Code number: TNMT.2021.04.05. The author would like to express their sincere thanks to anonymous reviewers for their helpful comments and review of the manuscript.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- [1] Chen, H., Xu, C.Y. and Guo, S. (2012) Comparison and Evaluation of Multiple GCMs, Statistical Downscaling and Hydrological Models in the Study of Climate Change Impacts on Runoff. *Journal of Hydrology*, **434-435**, 36-45. <https://doi.org/10.1016/j.jhydrol.2012.02.040>
- [2] Chang, F.-J. and Guo, S. (2020) Advances in Hydrologic Forecasts and Water Resources Management. *Water*, **12**, 1819. <https://doi.org/10.3390/w12061819>
- [3] Bidwell, V.J. and Griffiths, G.A. (1994) Adaptive Flood Forecasting: An Application to the Waimakariri River. *Journal of Hydrology (New Zealand)*, **32**, 1-15. <https://www.jstor.org/stable/43944722>
- [4] Mai, D.T. and De Smedt, F.A. (2017) A Combined Hydrological and Hydraulic Model for Flood Prediction in Vietnam Applied to the Huong River Basin as a Test Case Study. *Water*, **9**, 879. <https://doi.org/10.3390/w9110879>
- [5] Ji, J., Choi, C., Yu, M. and Yi, J. (2012) Comparison of a Data-Driven Model and a Physical Model for Flood Forecasting. In Press, W., Ed., *Flood Recovery Innovation and Response XI*, **159**, 133-142. <https://doi.org/10.2495/FRIAR120111>
- [6] Cordeiro, M., Wilson, H., Vanrobaeys, J., Pomeroy, J. and Fang, X. (2017) Simulating Cold-Region Hydrology in an Intensively Drained Agricultural Watershed in Manitoba, Canada, Using the Cold Regions Hydrological. *Hydrology and Earth System Sciences*, **21**, 3483-3506. <https://doi.org/10.5194/hess-21-3483-2017>
- [7] Aygün, O., Kinnard, C. and Campeau, S. (2020) Impacts of Climate Change on the Hydrology of Northern Midlatitude Cold Regions. *Progress in Physical Geography: Earth and Environment*, **2**, 338-375. <https://doi.org/10.1177/0309133319878123>
- [8] Hu, C., Wu, Q., *et al.* (2018) Deep Learning with a Long Short-Term Memory Networks Approach for Rainfall-Runoff Simulation. *Water*, **10**, 1543. <https://doi.org/10.3390/w10111543>
- [9] Wee, W.Z. (2021) A Review of Models for Water Level Forecasting Based on Ma-

- chine Learning. *Earth Science Informatics*, **14**, 1707-1728.
<https://doi.org/10.1007/s12145-021-00664-9>
- [10] Hastie, T., Tibshirani, R. and Friedman, J. (2008) The Elements of Statistical Learning. Springer, Stanford, 745. <https://hastie.su.domains/Papers/ESLII.pdf>
- [11] Irving, K., Kuemmerlen, M., *et al.* (2018) A High-Resolution Streamflow and Hydrological Metrics Dataset for Ecological Modeling Using a Regression Model. *Scientific Data*, **5**, 180224. <https://doi.org/10.1038/sdata.2018.224>
- [12] Afan, H.A., Allawi, M.F., El-Shafie, A., *et al.* (2020) Input Attributes Optimization Using the Feasibility of Genetic Nature Inspired Algorithm: Application of River Flow Forecasting. *Scientific Report*, **10**, 4684.
<https://doi.org/10.1038/s41598-020-61355-x>
- [13] Agudelo-Otálora, L.M., Moscoso-Barrera, W.D., *et al.* (2020) Comparison of Physical Models and Artificial Intelligence for Prediction of Flood Levels. *SciELO Mexico*, **9**, 209-236.
- [14] Devienne, S. (2006) Red River Delta: Fifty Years of Changes. *Moussons*, **9-10**, 255-280. <https://doi.org/10.4000/moussons.2042>
- [15] Giuliani, M., Castelletti, A. and Pianosi, E., *et al.* (2016) Curses, Tradeoffs, and Scalable Management: Advancing Evolutionary Multi-Objective Direct Policy Search to Improve Water Reservoir Operations. *Journal of Water Resources Planning and Management*, **142**, 04105050.
[https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0000570](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000570)
- [16] Soncini-Sessa, R., Weber, E. and Castelletti, A. (2007) Integrated and Participatory Water Resources Management-Theory. Elsevier Science.
<https://shop.elsevier.com/books/integrated-and-participatory-water-resources-management-theory/soncini-sessa/978-0-444-53013-4>
- [17] Galelli, S. and Castelletti, A. (2013) Tree-Based Iterative Input Variable Selection for Hydrological Modeling. *Water Resources Research*, **49**, 4295-4310.
<https://doi.org/10.1002/wrcr.20339>
- [18] Meng, C., Zhou, J., Tayyab, M., Zhu, S. and Zhang, H. (2016) Integrating Artificial Neural Networks into the VIC Model. *Water*, **8**, 407.
<https://doi.org/10.3390/w8090407>