Predicting Inflow Flow in Hydraulic Dams Using Artificial Neural Networks

Alberto Fernandez Sanchez, Juan Ramon Rabuñal Dopico, Daniel Rivero Cebrian, Alejandro Celestino Pazos Sierra, Marcos Gestal Pose, and Luis Cea Gomez

RNAsa-IMEDIR, Faculty of Computer Science, Universidade da Coruña, 15071 A Coruña, Spain
CITEEC Research Center, Universidade da Coruña, 15071 A Coruña, Spain
Correspondence: alberto.fsanchez@udc.es

DOI: *https://doi.org/10.17979/spudc.000024.15*

Abstract: Accurate prediction of inflow in dams plays a crucial role in water resource management Kim et al. (2019); Vargas-Garay et al. (2018); Zhong et al. (2018) and risk mitigation Costabile et al. (2020); Rabuñal et al. (2007). This study focuses on the Portodemouros dam (located between the provinces of A Coruña and Pontevedra), where a model based on a Long Short-Term Memory (LSTM) artificial neural network has been implemented to predict dam inflow.
The results demonstrate the well-established effectiveness of the LSTM network in flow prediction Dongkyuna and Seokkoob (2021); Jo and Jung (2023); Li et al. (2020) applied to the Portodemouros dam compared to other models. This comparison has already been performed in other studies with both mathematical models Amirreza et al. (2022); Ansori and Anwar (2022); A.R1 et al. (2018); Beck et al. (2017); Ciabatta et al. (2016); Costabile et al. (2020); Fan et al. (2013); Heřmanovský et al. (2017); Kim et al. (2019); Vargas-Garay et al. (2018); Zhong et al. (2018), genetic programming Aytek et al. (2008); Havlíček et al. (2013); Heřmanovský et al. (2017); Rabuñal et al. (2007) and other machine learning algorithms Jo and Jung (2023). Combining precipitation data from multiple regions and meteorological forecasts significantly enhances the model's ability to anticipate variations in dam inflow. This improved accuracy is essential for early flood detection and informed decision-making in dam operation.
This study forms part of the Marine Science programme (ThinkInAzul) supported by Ministerio de Ciencia e Innovación and Xunta de Galicia with funding from European Union NextGenerationEU (PRTR-C17.I1) and European Maritime and Fisheries Fund.

1 Introduction

Predicting the variation in surface water flows (runoff) based on local precipitation (rainfall-runoff models) has been addressed from the 19th century to the present day using mathematical models and later machine learning models separately or with comparative studies.

2 Dataset

2.1 Sources

For the training of the neural networks, a dataset of parameters from the Portodemouros dam, located in the Ulla River basin (42°50′47″N 8°08′26″W) with a total capacity of 297 hm^3, which includes a set of 4554 daily records of dam inflow, precipitation, and 1, 2, and 3-day predictions provided by meteogalicia, was used. See Figure 1.

Figure 1: Dam and pluviometer stations situation map

To prepare the complete dataset, several sources were used, including records from the dam, which include filling level, dam inflow, and dam flow outputs. These data will allow us to link this study to a subsequent one on human behavior modeling.

2.2 Preliminary Analysis: Discharge Time

To establish a rainfall-runoff model, it is necessary to estimate the time it takes for precipitation at each measuring station to affect the dam inflow. To accomplish this, the correlation between precipitation in several areas and changes in the stream was assessed. Four pluviometer station records were used to measure the aforementioned correlation: Arzua, Melide, Serradofaro, and Olveda. These stations were selected due to their proximity to the upper part of the Ulla River, demonstrating that all basins take less than 24 hours (1 day) to reach the dam, regardless of the soil water saturation state.. See Figure 1.

Table 1: Maximum correlation between precipitation and dam inflow (in days)

Soil Moisture	Arzua	Olveda	Serradofaro	Melide
Dry	1	1	1	1
Half-wet	1	1	1	1
Wet	1	1	1	1
Very wet	1	1	1	1
Saturated	1	1	1	1

2.3 Cross-Validation

The dataset has been split into subsets based on the year of each sample, resulting in 13 subsets (spanning from 2010 to 2022). Once these subsets were separated, 13 different sets of training, validation, and test data were prepared, such that the test set covers a complete year, the validation set the following complete year (or the first year if the test set includes the last year), and the training set includes the remaining years. This approach helps mitigate the impact of year-to-year variability on both model training and evaluation metrics. Subsequently, each of these sets was normalized by subtracting the mean of each column and dividing the result by the standard deviation of that column Li et al. (2020). The measurements were calculated by averaging across all k-fold subsets.

To compare the model's performance, other machine learning models and a naive model have been used:

- Linear Regression Tallarida and Murray (1987)
- SVM (Support Vector Machine for Regression) C. and V. (1995)
- Naive model (using the last known dam inflow value as the prediction for each sample)

3 Results

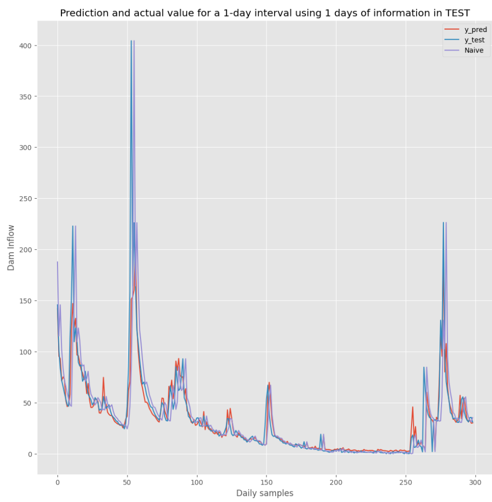

Figure 2: LSTM predictions (y_pred), Naive predictions (Naive) and real values (y_test) of dam inflow

Below are the results using different metrics commonly used as a basis in most rainfall-runoff studies, such as the coefficient of determination (R2) Ansori and Anwar (2022); Aytek et al. (2008); Jo and Jung (2023); Zhong et al. (2018) and the Nash-Sutcliffe Efficiency coefficient (NS or NSE) Ansori and Anwar (2022); A.R1 et al. (2018); Ciabatta et al. (2016); Dongkyuna and Seokkoob (2021); Havlíček et al. (2013); Heřmanovský et al. (2017); Jo and Jung (2023); Li et al. (2020).

Nash-Sutcliffe Efficiency (NS) is a widely used metric in hydrology to assess the accuracy of runoff models or hydrological forecasts in comparison to observed data McCuen et al. (2006). It was proposed by John R. Nash and James C. Sutcliffe in 1970 and is considered an effective measure to evaluate the overall performance of a model in simulating runoff events over time.

Model output metrics

Table 2: Prediction Evaluation

Model	R2 train	R2 test	NS train	NS test
LR	0.849	**0.816**	0.822	**0.782**
SVM	0.877	0.76	0.795	0.494
Naive	0.777	0.717	0.776	0.715
LSTM	**0.881**	0.8	**0.845**	0.709

These metrics were obtained using a 13-fold cross-validation with complete years as subsets. Each value is computed by a mean of each subset metric giving a better estimation of each measure avoiding any bias by extreme climatic difference between dataset years range.

4 Conclusions

As can be observed, the LSTM network achieves a training score of 0.845 and a testing score of 0.709 for NS, which measures the goodness of rainfall-runoff models, regardless of the dataset used. These NS values are very similar to those found in other studies, such as 0.73 in training by Ansori and Anwar (2022) or 0.782 in testing by Jo and Jung (2023).

LR has also demonstrated strong performance in the test dataset, achieving the highest values of R2 and NS in a subset of 365 samples, which makes it a reliable estimator for the rainfall-runoff model in this specific dataset.

This results confirms the robustness of LSTM algorithm in this dataset and establishes it as a valid foundation for further fine-tuning or complementation with other types of algorithms. For example, it can be used to model dam operator behavior, allowing predictions of dam overflow based on inflow flow forecasted by an algorithm like LSTM, along with another model predicting the orderly dam release by the operator.

In addition to the rainfall-runoff model, analyzing the behavior of the dam operator when managing dam releases is crucial to enhance the robustness of an early flood and overflow alarm system. This helps minimize operational risks while maximizing effectiveness in complex situations. To model their behavior, we have access to a historical record of actions based on the variables previously used in the rainfall-runoff model.

Bibliography

M. Amirreza, D. Amirhossein, S. Gerrit, and T. Massoud. Daily reservoir inflow forecasting using weather forecast downscaling and rainfall-runoff modeling. *Journal of Hydrology: Regional Studies*, 44(101228):1–20, 2022.

M. B. Ansori and N. Anwar. The trmm rainfall-runoff transformation model using gr4j as a prediction of the tugu dam reservoir inflow. *GEOMATE Journal*, 23(97):45–52, 2022.

A. R. A.R1, S. L.M, B. H., A. J.L., J. K., and S. S.K. Reservoir sediment inflow prediction using integrated rainfall-runoff and discharge–sediment model. *International Journal of Engineering and Technology*, 7(4):917,923, 2018.

A. Aytek, M. Asce1, and M. Alp. An application of artificial intelligence for rainfall–runoff modeling. *Journal of Earth System Science*, 117(2):145–155, 2008.

H. E. Beck, A. I. J. M. van Dijk, A. de Roo, E. Dutra, G. Fink, R. Orth, and J. Schellekens. Global evaluation of runoff from 10 state-of-the-art hydrological models. *Hydrology and Earth System Sciences*, 21(6):2881–2903, 2017.

C. C. and V. V. Support-vector networks. *Machine Learning*, 1(20):273–297, 1995.

L. Ciabatta, L. Brocca, C. Massari, T. Moramarco, S. Gabellani, S. Puca, and W. Wagner. Rainfall-runoff modelling by using sm2rain-derived and state-of-the-art satellite rainfall products over italy. *International Journal of Applied Earth Observation and Geoinformation*, 48:163–173, 2016.

P. Costabile, C. Costanzo, D. Ferraro, F. Macchione, and G. Petaccia. Performances of the new hec-ras version 5 for 2-d hydrodynamic-based rainfall-runoff simulations at basin scale: Comparison with a state-of-the art model. *Water*, 12(9), 2020.

K. Dongkyuna and K. Seokkoob. Data collection strategy for building rainfall-runoff lstm model predicting daily runoff. *Journal of Korea Water Resources Association*, 54(10):795–805, 2021.

F. Fan, Y. Deng, X. Hu, and Q. Weng. Estimating composite curve number using an improved scs-cn method with remotely sensed variables in guangzhou, china. *Remote Sensing*, 5(3): 1425–1438, 2013.

V. Havlíček, M. Hanel, P. Máca, M. Kuráž, and P. Pech. Incorporating basic hydrological concepts into genetic programming for rainfall-runoff forecasting. *Springer-Verlag*, 95(1):363–380, 2013.

M. Heřmanovský, V. Havlíček, M. Hanel, and P. Pech. Regionalization of runoff models derived by genetic programming. *Journal of Hydrology*, 547:544–556, 2017.

Y. Jo and K. Jung. Comparative study of machine learning and deep learning models applied to data preprocessing methods for dam inflow prediction. *GeoAI Data Society*, 5(2):92–102, 2023.

S.-H. N. Kim, W.-S. Bae, and Deg-Hyoc. An analysis of effects of seasonal weather forecasting on dam reservoir inflow prediction. *Journal of Korea Water Resources Association*, 52(7):451–461, 2019.

W. Li, A. Kiaghadi, and C. N. Dawson. High Temporal Resolution Rainfall Runoff Modelling Using Long-Short-Term-Memory (LSTM) Networks. *arXiv e-prints*, art. arXiv:2002.02568, 2020.

R. H. McCuen, Z. Knight, and A. G. Cutter. Evaluation of the nash-sutcliffe efficiency index. *Journal of Hydrologic Engineering*, 11(6):597–602, 2006.

J. R. Rabuñal, J. Puertas, J. Suárez, and D. Rivero. Determination of the unit hydrograph of a typical urban basin using genetic programming and artificial neural networks. *Hydrological Processes*, 21(4):476–485, 2007.

R. J. Tallarida and R. B. Murray. *Linear Regression I*. Springer, New York, NY, 1987.

L. Vargas-Garay, O. D. Torres-Goyeneche, and G. A. Carrillo-Soto. Evaluation of scs - unit hydrograph model to estimate peak flows in watersheds of norte de santander. *Respuestas journal of engineering sciences*, 24(1):6–15, 2018.

W. Zhong, R. Li, Y. Q. Liu, and J. Xu. Effect of different areal precipitation estimation methods on the accuracy of a reservoir runoff inflow forecast model. *IOP Conference Series: Earth and Environmental Science*, 208(1):012043, dec 2018.