CIÊNCIA e NATURA
UFSM
ISSN 2179-460X
OPEN ACCESS

Ci. e Nat., Santa Maria, v. 46, n. spe. 1, e87129, 2024 • https://doi.org/10.5902/2179460X87129
Submitted: 04/11/2024 • Approved: 08/28/2024 • Published: 11/07/24

Mathematics

Hydrological modeling using artificial neural networks for flood event forecasting. Case study: Pomba river in Santo Antônio de Pádua-RJ

Modelagem hidrológica utilizando redes neurais artificiais para a previsão de eventos de cheias. Estudo de caso: Rio Pomba em Santo Antônio de Pádua-RJ

Rennan Mendes de Moraes dos Santos Dias [I], **Wagner Rambaldi Telles** [II], **Antônio José da Silva Neto** [I],

[I] Polytechnic Institute, Rio de Janeiro State University, Nova Friburgo, RJ, Brazil
[II] Northwest Fluminense Institute of Higher Education, Federal Fluminense University, Santo Antônio de Pádua, RJ, Brazil

ABSTRACT

Flood prediction through hydrological modeling of watersheds remains an emerging need in society, particularly in regions highly affected by these extreme events. Models based on artificial neural networks have demonstrated significant potential for addressing this issue due to their simplicity and agility. In this study, a model was developed using a multilayer perceptron network for predicting river discharge and water level based on the previous day's river state and precipitation forecast. The Pomba river in the city of Santo Antônio de Pádua-RJ was investigated due to its regular occurrence of flood events that impact the entire population. Metric and graphical results showed the model's strong ability to estimate discharge and water levels throughout the year at a station with limited data. On the other hand, the model encountered difficulties in accurately estimating peak values.

Keywords: Artificial neural networks; Hydrological modeling; Flood event; Multilayer perceptron

RESUMO

A previsão de enchentes, através da modelagem hidrológica de bacias hidrográficas, continua sendo uma necessidade emergente na sociedade, principalmente em regiões muito afetadas por esses eventos extremos. Modelos baseados em redes neurais artificiais têm apresentado significativo potencial para esta problemática devido a sua simplicidade e agilidade. Neste trabalho, produziu-se um modelo utilizando uma rede perceptron multicamadas para a previsão de vazão e cota de um rio com base no estado do deste no dia anterior e na previsão da precipitação. Estudou-se o rio Pomba na cidade de Santo Antônio de Pádua-RJ por este apresentar, ordinariamente, eventos de cheias que afetam toda a população. Os resultados métricos e gráficos mostraram uma boa capacidade do modelo em estimar as vazões e cotas ao longo de todo um ano em uma estação com poucos dados. Por outro lado, o modelo apresentou dificuldades na estimação precisa dos picos.

Palavras-chave: Redes Neurais Artificiais; Modelagem Hidrológica; Eventos de Cheias; Perceptron Multicamadas

1 INTRODUCTION

Since the 1990s, the successful application of Artificial Neural Network (ANN) models to various areas of hydrology has been documented (ASCE Task Committee on Application of Artificial Neural Networks in Hydrology, 2000). It is noteworthy that ANNs have proven to be a powerful tool in modeling several nonlinear hydrological problems, for instance rainfall-runoff modeling (Tokar and Markus, 2000), precipitation forecasting (French et al., 1992), water quality (Kalin et al., 2010), and water management (Kralisch et al., 2003).

Regarding rivers that flow through cities, the hazards caused by extreme events, such as floods, include traffic disturbances, communication network problems, damage to structures and infrastructure, agricultural losses, all posing risks to human health. Therefore, the implementation of policies for prevention and protection of people and properties is necessary, aiming to reduce vulnerability to events of this nature (Elsafi, 2014).

Although hydrodynamic models provide a solid physical foundation and are capable of simulating flow for large areas, these models require rigorous data collection regarding river characteristics, which are not always available in various locations. Additionally, it has not been possible to directly implement observed data from a desired small-scale location into these types of models in order to improve results for that region (Elsafi, 2014).

Furthermore, for the successful implementation of a flood prevention policy and risk reduction, information from different sources, such as river and rainfall gauge stations, must be integrated and used to obtain a quick and accurate forecast of the river water level elevation (Campolo et al., 2003). Thus, there has been a growing interest in recent decades in ANN models capable of producing reasonably accurate results in a short period of time, leveraging a reduced dataset subject to errors, such as real-time datasets (Kim and Barros, 2001).

Hence, in this study, a hydrological model with artificial neural networks was developed for predicting the water level and discharge of a water body, one day ahead, based on the forecasted rainfall and the current state of discharge, water level, and daily accumulated precipitation. For this research, the Pomba river at a station in the center of Santo Antônio de Pádua-RJ, a municipality that frequently experiences floods, was selected as the study area.

2 CASE STUDY

The Pomba river, which originates in the municipality of Santa Bárbara do Tugúrio-MG, at an altitude of 1,182 meters, flows into the South Paraíba river in the city of Aperibé-RJ, at an altitude of 55 meters. This river basin encompasses a drainage area of 8,616 km², covering approximately 35 municipalities in the state of Minas Gerais upstream and 3 municipalities in the state of Rio de Janeiro downstream (AGEVAP, 2017).

Among the municipalities within the basin, Santo Antônio de Pádua-RJ (the region of interest) is one of the most populated, with over 40,000 inhabitant according to the latest Brazilian census (IBGE – Instituto Brasileiro de Geografia e Estatística, 2023). This municipality is bisected by the Pomba river from end to end and experiences annual flood events, including notable floods in 2008, 2010, 2020 and 2022 (figure 1).

Figure 1 – Square of the Parish of Santo Antônio de Pádua

Source: Terceira Via Newspaper (Photo: Felipe Sião)

Caption: The Square of the Parish of Santo Antônio de Pádua completely flooded, center of the city, in 2022

3 METHODOLOGY

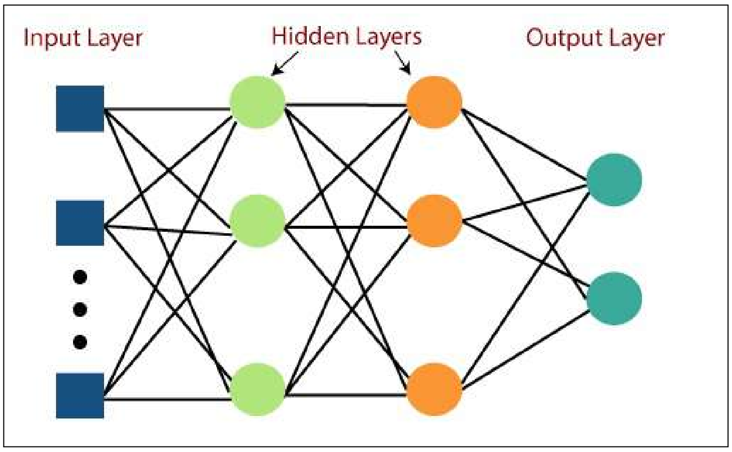
In this section, we present the processes used in the development of the ANN (Subsection 3.1), the collection of observed data (Subsection 3.2), and the evaluation of the model through efficiency metrics (Subsection 3.3).

3.1 Artificial neural network

Artificial neural networks with different configurations have been used to create mathematical models for a variety of problems. Among these network architectures, remarkable examples include the use of convolutional neural networks (CNN), recurrent neural networks (RNN), autoencoders, and feedforward networks, where the multilayer perceptron (MLP) resides.

Furthermore, there is a growing utilization of MLP-type ANN models for flood prediction (Mosavi et al., 2018). This class of network employs supervised learning known as backpropagation (BP) during training to optimize the parameters of interconnected neurons across multiple layers. Simplicity, non-linear activation functions, and a high number of layers are characteristics of the multilayer perceptron network (figure 2).

Figure 2 – Diagram of a multilayer perceptron neural network



Source: Java T Point website
Caption: Diagram of a general multilayer perceptron neural network. In the illustrated case, with at least 3 inputs, 2 hidden layers, and 2 outputs

In this paper, an MLP-type ANN was used where the model (F) output was expected to be the values of discharge (Q_i) in m^3/s and river level (h_i) in cm, one day ahead, based on the input values of discharge ($Q_{i-1}[m^3/s]$), level ($h_{i-1}[cm]$), and precipitation ($p_{i-1}[mm]$) for the current day, as well as the rainfall forecast for the following day ($p_i[mm]$), that is, $F(p_i, p_{i-1}, Q_{i-1}, h_{i-1}) = (Q_i, h_i)$.

The number of layers and the number of neurons per hidden layer were varied to find the best configuration for the network in modeling the addressed problem. The remaining hyper-parameters remained constant values after a brief effectiveness analysis based on trial and error.

Regarding the hyper-parameters, the following values and definitions were used:

- The Rectified Linear Unit with Leaky Slope (LeakyReLU) as the activation function for all hidden layers of the network;
- Mean Squared Error (MSE) as the loss function, where specifically, it was defined as $loss = loss1 + 3 * loss2$, where $loss1$ and $loss2$ represent the loss related to discharge and level, respectively, thus emphasizing the river level values, which are considered the main flood alert indicator;
- Adaptive Moment Estimation (Adam) as the optimizer;
- $batchsize = 84$, corresponding to a dataset of 12 weeks;
- Early stopping with a tolerance of 1000 epochs and no epoch limit.

To generate the results in this work, the Google Colab environment (with GPU usage) was utilized, where the ANN models were implemented in Python, with the assistance of the following libraries:“pandas” for reading the available data in Microsoft Excel 2010; “torch” for implementing the ANN; “numpy” for general mathematical operations; “matplotlib.pyplot” for plotting the graphs; “time” for measuring the computational time spent.

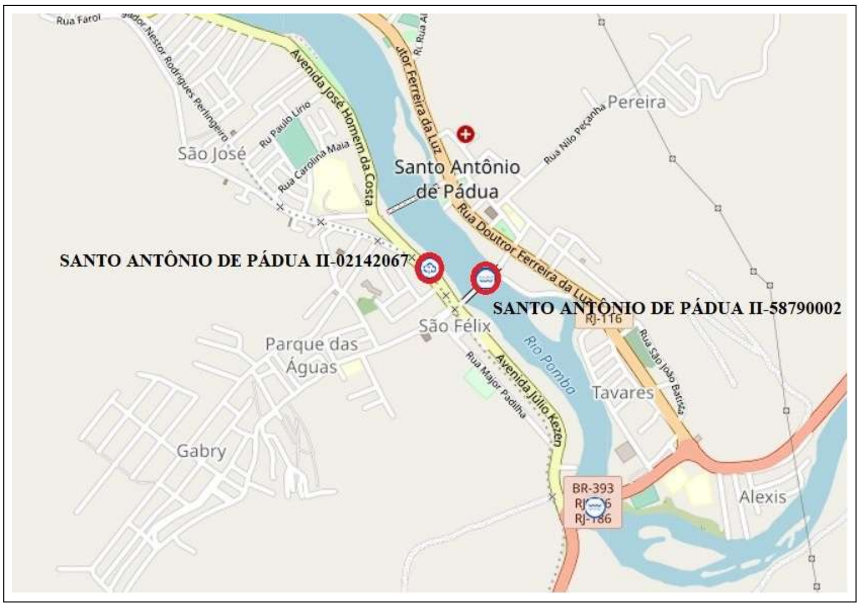
3.2 Database

The acquisition of observed data in the region of interest for setting the training patterns of the artificial neural network was carried out through the website of the National Water and Basic Sanitation Agency (ANA) via the HidroWeb Portal. In summary, the HidroWeb Portal is described as “a tool integrated into the National System of Information on Water Resources (SNIRH) that provides access to the database containing all the information collected by the National Hydrometeorological

Network (RHN), gathering data on river levels, discharge, rainfall, climatology, water quality, and sediments."

For this purpose, a pluviometric station (SANTO ANTÔNIO DE PÁDUA II-02142067) and a fluviometric station (SANTO ANTÔNIO DE PÁDUA II-58790002) located in the center of Santo Antônio de Pádua-RJ were selected (figure 3), from where rainfall, river levels, and discharge data were collected.

Figure 3 – Pluviometric and fluviometric stations used



Source: HidroWeb-ANA Portal
Caption: Location of the pluviometric and fluviometric stations used in Santo Antônio de Pádua-RJ

Data was collected from January 1, 2013, to December 31, 2020, for training, from January 1, 2021, to December 31, 2021, for validation, and from January 1, 2022, to December 31, 2022, for testing. In other words, the model was trained using data from 2013 to 2020 based on the loss of the validation data from 2021, and then tested on the 2022 dataset. It is worth mentioning that a rescaling process was performed on the data, scaling it to the interval $[0, 1]$, for faster and more accurate training.

3.3 Evaluation metrics

It is worth noting the use of efficiency evaluation metrics when quantifying the results obtained by different models. For this purpose, the Mean Squared Error (MSE) function was used during the training of the ANN, while the Root Mean Squared Error (RMSE), Nash-Sutcliffe Efficiency (NSE), and Mean Error or Bias (BIAS) were utilized for assessing the validation group.

The MSE function is sensitive to larger errors as it squares individual differences. It always returns positive values, with a value of 0 indicating a perfect simulation (Hallak and Pereira Filho, 2011). It can be expressed by as:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(X_{S_i} - X_{O_i})^2 \tag{1}$$

where N is the number of training indexes, X_{S_i} is the simulated variable, either river level (cm) or discharge (m^3/s), based on index i, and X_{O_i} is the observed variable of index i.

The RMSE function, commonly used to assess the accuracy of numerical results, displays values in the same dimensions as the analyzed variable (Hallak and Pereira Filho, 2011). A value of 0 also indicates a perfect simulation. It can be defined as follows:

$$RMSE = \left(\frac{1}{N}\sum_{i=1}^{N}(X_{S_i} - X_{O_i})^2\right)^{\frac{1}{2}} \tag{2}$$

On the other hand, the Nash-Sutcliffe Efficiency (NSE) criterion (Celeste and Chaves, 2014) indicates how accurate the model predictions are relative to the mean of the experimental data. A value of 1 corresponds to a perfect fit, while a value of 0 indicates that the simulation is as accurate as the mean of the observed data. Negative values occur when the observed mean exceeds the model. It can be calculated as:

$$NSE = 1 - \frac{\sum_{i=1}^{N}\left(X_{O_i} - X_{S_i}\right)^2}{\sum_{i=1}^{N}\left(X_{O_i} - \overline{X_O}\right)^2} \tag{3}$$

where $\overline{X_O}$ is the mean value of the observed variable across all indexes i.

Nevertheless, the Bias (BIAS) indicates the tendency of a model to underestimate (negative value) or overestimate (positive value) the simulated variable compared to the observed one. A value of 0 represents no bias. In this work, BIAS was calculated as a percentage as follows:

$$BIAS = \frac{\frac{1}{N}\sum_{i=1}^{N}\left(X_{S_i} - X_{O_i}\right)}{\overline{X_O}} * 100 \tag{4}$$

4 RESULTS AND DISCUSSION

In this section, the results obtained over the 5 executions for each of the 14 models studied and described in Section 3 were presented and discussed. The implementation of the multiple ANN configurations were evaluated in comparison to each other.

The average training time for each of the configurations was 15 minutes, where the initial guess for the network parameters was varied 5 times, using the five same different seeds for generating random values for all tested architectures, and the best result among these 5 iterations was taken. These configurations were ranked and selected based on the NSE values for peak level (floods) and for the entire testing period. The best results are shown in blue, while the worst are shown in red.

Table 1 – Results obtained simulating the year 2022

		Peak level		RMSE		NSE	
Neurons per hidden layer	Best epoch	NSE	BIAS (%)	Discharge	Level	Discharge	Level
10, 10, 10	260	-0.073	-8.749	48.337	17.203	0.911	0.923
20, 20, 20	133	-0.118	-5.503	46.587	17.677	0.918	0.919
40, 40, 40	46	0.052	-7.599	48.046	17.470	0.912	0.921
80, 80, 80	8	-0.050	-7.952	53.119	17.379	0.893	0.922
40, 40, 40, 40	28	0.088	-6.504	44.710	17.248	0.924	0.923
40, 40, 40, 40, 40	257	-0.085	-5.762	53.476	17.635	0.891	0.919
40, 40, 40, 40, 40, 40	59	0.142	-3.896	45.342	17.302	0.922	0.923
40, 40, 40, 40, 40, 40, 40	70	0.076	-6.428	50.230	16.919	0.904	0.926
40, 20, 10	176	-0.065	-5.358	48.356	17.661	0.911	0.919
60, 40, 20	29	-0.080	-8.305	49.464	17.351	0.907	0.922
40, 20, 10, 5	244	-0.110	-9.789	48.851	17.864	0.909	0.917
60, 40, 20, 10	25	-0.156	-2.528	77.929	18.085	0.769	0.915
60, 50, 40, 30, 20, 10	36	-0.024	-5.356	46.917	17.497	0.916	0.921
40, 20, 40, 20, 40, 20	12	-0.196	-2.577	61.920	18.266	0.854	0.914

Caption: Results obtained for the different networks tested in simulating the discharge and water level for the year 2022.

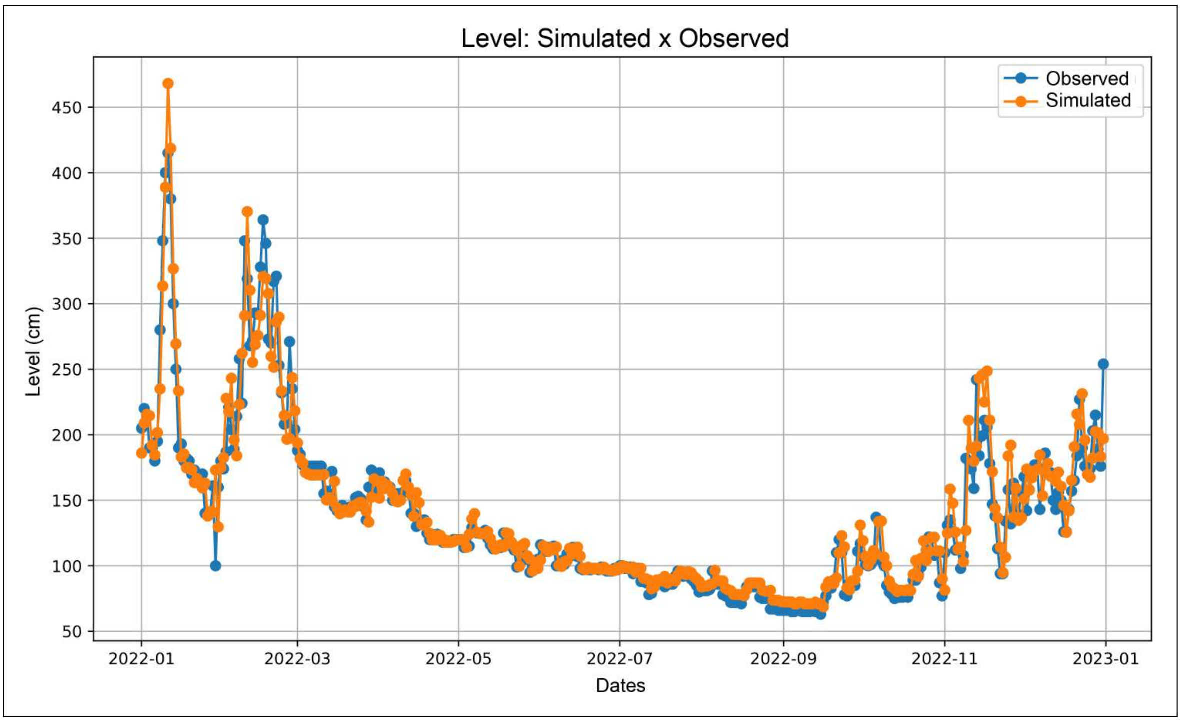
Initially, observing the column of epochs that generated the best results (Best epoch), based on the low values presented (less than 300), it can be inferred that these models exhibited a significant dependence on the initial guess, often susceptible to local minima. Additionally, the BIAS (%) column shows that, in general, all models underestimated the water levels during flood events.

Looking over the best and worst NSE values for peak level and overall level, it can be observed that ANNs with 40 neurons in all layers generally yield more satisfactory results than networks with a ”pyramid” structure, where the number of neurons decreases across layers. It is also noticeable that all networks struggled to accurately simulate flood values, although overall, they achieved very good values (NSE above 0.9) for the water level. This is due to the lack of available information for model training, i.e., a short time series with few flood events compared to the overall dataset, and the use of only one pluviometric and one fluviometric station for forecast at a point along the river that drains a large region.

As observed in the study published by Aghelpour and Varshavian (2020), MLP models for predicting channel discharge based on the current state of the river ($Q(t-1)$ as example) faced challenges during flood periods and yielded good results during drier periods, consistent with the findings of this study. Additionally, there is evidence of model convergence to the optimum with low numbers of epochs (less than 100), as suggested in the best cases (blue names).

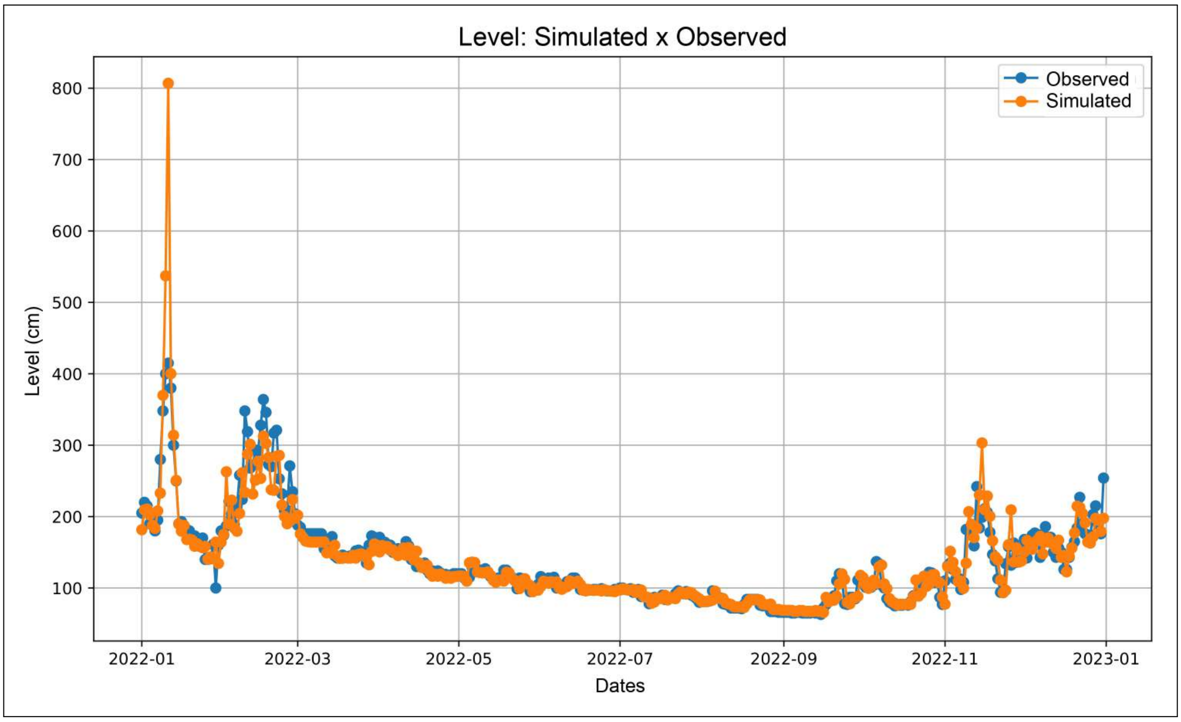
Based on the results early presented, was chosen to display the graphs of simulated water level versus observed water level for the best (figure 4) and worst (figure 5) configurations, where the good fit of the simulations is evident in both cases.

Figure 4 – Simulated level (best case)



Caption: Simulated level versus observed level graph for the configuration with the best result

Figure 5 – Simulated level (worst case)



Caption: Simulated level versus observed level graph for the configuration with the worst result

The lack of precision in simulating the peaks, underestimating them, is evident even in the best result. Thus, a revision of the training process is estimated to emphasize flood values. Despite this quantity inaccuracy, the water level curves (figure 4 and figure 5), both simulated and observed, develop a similar profile, suggesting that this model correctly indicates the dates when flood events will occur.

The accuracy of ANN models, with NSE values above 0.9 when considering predictions over a one-year period encompassing both wet and dry periods, indicates the effectiveness of the methodology, as well as observed in the studies by Dalkiliç and Hashimi (2020) and Kumar et al. (2020), which demonstrated that techniques employing various configurations of neural networks and data-driven models yielded the same consistent results, even across different research areas and slightly different approaches.

Therefore, this research concluded that the methodology using an MLP-type neural network for flood event prediction based on real observed data available on government platforms of the river's current state yielded satisfactory results based on the efficiency metrics commonly used in this type of problem.

Despite the difficulties encountered in simulating the peaks, it was inferred that all configurations were able to adequately represent the profile of the discharge and level curves for the year 2022. This leads to considering the complexity involved in estimating two different variables in a single model.

Thus, given the challenges regarding the problem of flood prediction in regions such as the one studied in this work, the following future work is estimated: Aggregating information from other stations along the watershed in question; Creation of separate neural network models, one for discharge prediction and another for level prediction, connected as necessary; Implementation of a neural network with Physics-informed neural networks (PINN) by the Saint-Venant equations, adding a physical bias to the network training; Developing techniques, such as cross-validation and oversampling, to emphasize flood events during model training; Increasing input information, providing a better information base about the current state of the region for the neural network; Implementing a model to make hourly predictions; Implementing a model to predict larger intervals; Coupling the neural network model with a traditional numerical model.

ACKNOWLEDGEMENTS

The authors acknowledge the financial support provided by the following Brazilian agencies: FAPERJ, Carlos Chagas Filho Foundation for Research Support of the State of Rio de Janeiro; CNPq, National Council for Scientific and Technological Development; and CAPES, Coordination for the Improvement of Higher Education Personnel (Finance Code 001).

REFERENCES

AGEVAP (2017). *Plano de Recursos Hídricos da Bacia do Rio Paraíba do Sul-Resumo*. Fundação COPPETEC Laboratório de Hidrologia e Estudos de Meio Ambiente, relatório contratual r-10 edition.

Aghelpour, P. and Varshavian, V. (2020). Evaluation of stochastic and artificial intelligence models in modeling and predicting of river daily flow time series. *Stochastic Environmental Research and Risk Assessment*, 34(1):33–50.

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology (2000). Artificial neural networks in hydrology. ii: Hydrologic applications. *Journal of Hydrologic Engineering*, 5(2):124–137.

Campolo, M., Soldati, A., and Andreussi, P. (2003). Artificial neural network approach to flood forecasting in the river arno. *Hydrological sciences journal*, 48(3):381–398.

Celeste, A. and Chaves, V. S. (2014). Avaliação de algoritmos de otimização e funções objetivo para calibração automática do modelo chuva-vazão tank model. *Ciência e Natura*, 36(3):527–537.

Dalkiliç, H. Y. and Hashimi, S. A. (2020). Prediction of daily streamflow using artificial neural networks (anns), wavelet neural networks (wnns), and adaptive neuro-fuzzy inference system (anfis) models. *Water Supply*, 20(4):1396–1408.

Elsafi, S. H. (2014). Artificial neural networks (anns) for flood forecasting at dongola station in the river nile, sudan. *Alexandria Engineering Journal*, 53(3):655–662.

French, M. N., Krajewski, W. F., and Cuykendall, R. R. (1992). Rainfall forecasting in space and time using a neural network. *Journal of hydrology*, 137(1-4):1–31.

Hallak, R. and Pereira Filho, A. J. (2011). Metodologia para análise de desempenho de simulações de sistemas convectivos na região metropolitana de são paulo com o modelo arps: sensibilidade a variações com os esquemas de advecção e assimilação de dados. *Revista Brasileira de Meteorologia*, 26:591–608.

IBGE – Instituto Brasileiro de Geografia e Estatística (2023). *Censo Brasileiro de 2022*. Governo Federal, Rio de Janeiro, Brasil.

Kalin, L., Isik, S., Schoonover, J. E., and Lockaby, B. G. (2010). Predicting water quality in unmonitored watersheds using artificial neural networks. *Journal of environmental quality*, 39(4):1429–1440.

Kim, G. and Barros, A. P. (2001). Quantitative flood forecasting using multisensor data and neural networks. *Journal of Hydrology*, 246(1-4):45–62.

Kralisch, S., Fink, M., Flügel, W.-A., and Beckstein, C. (2003). A neural network approach for the optimisation of watershed management. *Environmental Modelling & Software*, 18(8-9):815–823.

Kumar, M., Kumari, A., Kushwaha, D. P., Kumar, P., Malik, A., Ali, R., and Kuriqi, A. (2020). Estimation of daily stage–discharge relationship by using data-driven techniques of a perennial river, india. *Sustainability*, 12(19):7877.

Mosavi, A., Ozturk, P., and Chau, K.-w. (2018). Flood prediction using machine learning models: Literature review. *Water*, 10(11):1536.

Tokar, A. S. and Markus, M. (2000). Precipitation-runoff modeling using artificial neural networks and conceptual models. *Journal of Hydrologic Engineering*, 5(2):156–161.

Author contributions

1 – Rennan Mendes de Moraes dos Santos Dias (Corresponding Author)
MSc Computational Modelling
https://orcid.org/0000-0002-8593-7630 • rennan.dias@iprj.uerj.br
Contribution: Conceptualization; Data curation; Formal Analysis; Funding acquisition; Investigation; Methodology; Project administration; Resources; Software; Supervision; Validation; Visualization; Writing – Original Draft Preparation; Writing – Review & Editing

2 – Wagner Rambaldi Telles
PhD Computational Modelling
https://orcid.org/0000-0002-6032-3405 • wtelles@id.uff.br
Contribution: Conceptualization; Funding acquisition; Project administration; Writing – Review & Editing

3 – Antônio José da Silva Neto
PhD Mechanical Engineering
https://orcid.org/0000-0002-9616-6093 • ajsneto@iprj.uerj.br
Contribution: Conceptualization; Funding acquisition; Project administration; Writing – Review & Editing

How to cite this article

Dias, R. M. M S., Telles, W. R., Silva Neto, A. J. (2024). Hydrological modeling using artificial neural networks for flood event forecasting. Case study: Pomba river in Santo Antônio de Pádua-RJ. *Ciência e Natura*, Santa Maria, v. 46, spe. 1, e87129. https://doi.org/10.5902/2179460X87129