

Article

Jucazinho Dam Streamflow Prediction: A Comparative Analysis of Machine Learning Techniques

Erickson Johny Galindo da Silva , Artur Paiva Coutinho * , Jean Firmino Cardoso 
and Saulo de Tarso Marques Bezerra 

Agreste Campus, Federal University of Pernambuco, Av. Marielle Franco, Km 59, Caruaru 55014-900, Brazil; erickson.galindo@ufpe.br (E.J.G.d.S.); jean.firmino@ufpe.br (J.F.C.); saulo.tarso@ufpe.br (S.d.T.M.B.)

* Correspondence: arthur.coutinho@ufpe.br

Abstract: The centuries-old history of dam construction, from the Saad el-Kafara Dam to global expansion in the 1950s, highlights the importance of these structures in water resource management. The Jucazinho Dam, built in 1998, emerged as a response to the scarcity of water in the Agreste region of Pernambuco, Brazil. After having less than 1% of its water storage capacity in 2016, the dam recovered in 2020 after interventions by the local water utility. In this context, the reliability of influent flow prediction models for dams becomes crucial for managers. This study proposed hydrological models based on artificial intelligence that aim to generate flow series, and we evaluated the adaptability of these models for the operation of the Jucazinho Dam. Data normalization between 0 and 1 was applied to avoid the predominance of variables with high values. The model was based on machine learning and employed support vector regression (SVM), random forest (RF) and artificial neural networks (ANNs), as provided by the Python Sklearn library. The selection of the monitoring stations took place via the Brazilian National Water and Sanitation Agency's (ANA) HIDROWEB portal, and we used Spearman's correlation to identify the relationship between precipitation and flow. The evaluation of the performance of the model involved graphical analyses and statistical criteria such as the Nash–Sutcliffe model efficiency coefficient (NSE), the percentage of bias (PBIAS), the coefficient of determination (R^2) and the root mean standard deviation ratio (RSR). The results of the statistical coefficients for the test data indicated unsatisfactory performance for long-term predictions (8, 16 and 32 days ahead), revealing a downward trend in the quality of the fit with an increase in the forecast horizon. The SVM model stood out by obtaining the best indices of NSE, PBIAS, R^2 and RSR. The graphical results of the SVM models showed underestimation of the flow values with an increase in the forecast horizon due to the sensitivity of the SVM to complex patterns in the time series. On the other hand, the RF and ANN models showed hyperestimation of the flow values as the number of forecast days increased, which was mainly attributed to overfitting. In summary, this study highlights the relevance of artificial intelligence in flow prediction for the efficient management of dams, especially in water scarcity and data-scarce scenarios. A proper choice of models and the ensuring of reliable input data are crucial for obtaining accurate forecasts and can contribute to water security and the effective operation of dams such as Jucazinho.



Citation: Silva, E.J.G.d.; Coutinho, A.P.; Cardoso, J.F.; Bezerra, S.d.T.M. Jucazinho Dam Streamflow Prediction: A Comparative Analysis of Machine Learning Techniques. *Hydrology* **2024**, *11*, 97. <https://doi.org/10.3390/hydrology11070097>

Academic Editor: Andrea Petroselli

Received: 3 April 2024

Revised: 11 June 2024

Accepted: 11 June 2024

Published: 4 July 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: support vector machine; random forest; artificial neural network; hydrological modeling; rainfall; flow; forecasting

1. Introduction

Many archaeologists consider the Saad el-Kafara dam to be one of the first in the world. It was probably built during the reign of Khufu, who was the king of Egypt around 2900–2877 B.C. [1]. Since then, there has been a long tradition of dam construction that spans millennia and with various purposes, such as flood control, providing water for human consumption, irrigation and animal thirst and, more recently in the history of dams, for industrial purposes and electricity generation.

As we arrived in the 1950s, which was a period of great expansion of global populations and economies, dams began to be increasingly considered as a solution to meet the growing demands for water and energy. Since then, according to the World Commission on Dams (WCD) [2], at least 45,000 large dams have been built worldwide, with almost half of the world's rivers having at least one large dam in their course.

In the Brazilian city of Surubim in the state of Pernambuco, the Jucazinho Dam was inaugurated in 1998; it is located in the Capibaribe watershed and bars the river that is also called the Capibaribe. It was built to minimize the water scarcity in the rural region of Pernambuco and to control floods on the Capibaribe River.

According to the Pernambuco Water and Climate Agency (APAC) [3], the dam has an accumulation capacity of 204.82 million cubic meters of water at an elevation of 292 m from the main spillway crest, and, as recorded by the Brazilian National Water and Sanitation Agency (ANA) [4], it has 92% of the total withdrawal demand for water for human supply, 7% for animal thirst, and 1% for irrigation. This amount of water is relevant due to the large number of municipalities it serves: a total of 15 cities.

The Hydroenvironmental Plan for the Capibaribe Hydrographic Basin of the Secretariat of Water Resources of the State of Pernambuco highlights the economic potential of the municipalities served by the dam, nine of which belong to the second largest textile and clothing pole in Brazil, three with relevant agricultural activities, two belonging to the furniture and tourism pole and one with a strong thermal tourism industry.

Even with the efforts of the Pernambuco Sanitation Company (COMPESA), which is responsible for the operation of the Jucazinho Dam, to combat the prolonged drought, the dam's water level dropped to below 0.01% in 2016 and only recovered four years later, in 2020, surpassing 1% [5]. At the height of the water crisis, the company implemented a rotation in the supply, providing the population with potable water only seven days per month [6].

To address such challenges and improve the management of water resources, researchers have increasingly turned to advanced computational methods. In recent decades, several artificial intelligence models have been used by researchers to predict influentary flows due to the great accuracy and flexibility of considering physical processes with all their characteristics. Among the methods used, artificial neural network (ANN) [7–9], random forest (RF) [10–12] and support vector machine (SVM) [13–15] models stand out.

Based on the literature, there are models that are better suited depending on the watershed. Adnan et al. [13] conducted an evaluation of various models, including the Optimally Pruned Extreme Learning Machine (OP-ELM), Least Square Support Vector Machine (LSSVM), Multivariate Adaptive Regression Splines (MARS) and M5 Model Tree (M5Tree), for modeling monthly streamflows with precipitation and temperature inputs. Their findings highlighted the superiority of LSSVM and MARS-based models for streamflow prediction without the need for local input data, surpassing the OP-ELM and M5Tree models. Parisouj et al. [8] investigated the predictive accuracy of three renowned machine learning algorithms—Support Vector Regression (SVR), ANN and Extreme Learning Machine—for monthly and daily streamflows across four rivers in the United States. The study identified SVR as the most effective model at both the monthly and daily scales, whereas the ANN model exhibited the least satisfactory performance. Meshram et al. [9] compared the efficacy of three AI techniques—Adaptive Neuro Fuzzy Inference System (ANFIS), Genetic Programming and ANN—for forecasting streamflow within India's Shakkur watershed. The findings underscored that models incorporating cyclic terms outperformed those that did not consider periodicity and relied solely on previous streamflow data. Sousa Jr. et al. [12] assessed K-Nearest Neighbor, SVM, RF and ANN models for daily streamflow prediction in a transitional region between the Savanna and Amazon biomes in Brazil. The results demonstrated that these models achieved promising streamflow predictions for up to three days ahead, even in basins with scarce hydrological data.

The motivation of this study is the delivery of a hydrological model for reservoir management based on emerging artificial intelligence techniques, adding to the literature machine learning adaptations for the prediction of tributary flows and optimizing the operation of water reservoirs.

Therefore, the objective of the present study is to generate synthetic series of tributary flows for the Jucazinho Dam, which is located in the Agreste region of Pernambuco, from stochastic models. By evaluating the adaptability of machine learning models using SVM, RF and ANN to generate synthetic series of tributary flows to the dam, analyzing the influence on the quality of adjustments with the increase in the number of forecast days and verifying the quality of the adjustments based on statistical metrics, we determine the best model for the hydrological variables under study.

2. Materials and Methods

2.1. Study Area

Located in the Brazilian city of Surubim in the state of Pernambuco in the Capibaribe watershed, the Jucazinho Dam bars the river that is also called the Capibaribe, as we can see in the map of the situation in Figure 1. Its construction began in 1995 and was completed in 1998, and at the time, there was the expropriation of more than two thousand hectares in the riverside areas, involving about 5000 people [16].

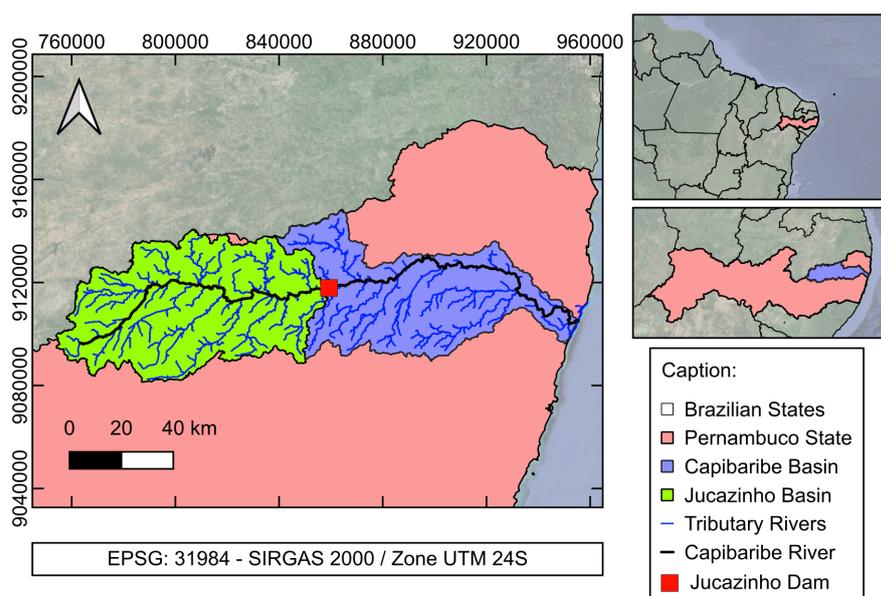


Figure 1. Situation map.

One of the reasons for the construction of the dam was the scenario of scarcity of water supply in the rural region of Pernambuco. With the construction of Jucazinho, which has a maximum storage capacity of 245.26 million cubic meters of water at the maximum maximum elevation of 295 m, 21 municipalities could be served, which impacted the lives of approximately 800 thousand inhabitants. As stated in Figure 2, 92% of the dam's use is for water for human supply. In addition, the dam is used for fish farming, livestock and agricultural.

The other reason for the construction of the dam was for flood control planning in Capibaribe, which involved the construction of several dams in order to protect the metropolitan region of Recife (about 135 km away) from historical floods such as those that occurred between the years 1960 and 1980. The dam has a flood control volume of 10^6 m³.

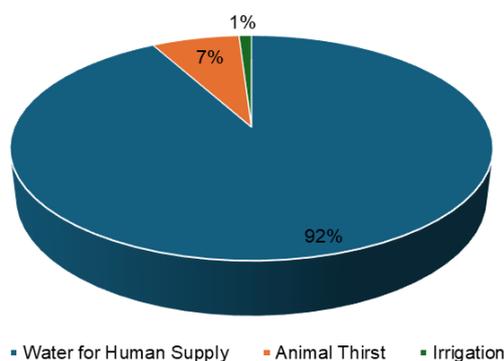


Figure 2. Total withdrawal demands.

Its construction type is gravity with roller-compacted concrete, and it has a central stepped spillway with a ski-jump-type dissipation basin. There are also two side spillways connected to the dam's abutment. It contains a gallery with access at the two abutments and which extends over the entire embankment of the dam. Above the central spillway is a bridge that connects the abutments. It has a water intake for supply and another for the release of the ecological flow downstream, both with a pipe of 2.0 m in diameter and a reduction to 1.5 m. In Table 1, we provide data from the dam's technical file, with information provided by Department of Water Resources of Pernambuco and data measured by Neves et al. [17].

2.2. Input Data

The “garbage in, garbage out” principle refers to the fundamental idea that the quality of the output of a data processing system is directly influenced by the quality of the input data. In other words, if inaccurate, incomplete or inadequate information is fed into a system, it is inevitable that the resulting output will also be inaccurate or of poor quality. This principle highlights the critical importance of reliable and high-quality data entry to ensure accurate and useful results in any computing or decision-making process.

The data for this study were obtained from the HIDROWEB portal: an online platform by ANA that offers information on Brazil's water resources. The portal provides real-time data from a vast network of monitoring stations across the country, covering hydrometeorological, hydrographic and water quality aspects. In the Jucazinho Dam's basin, 32 rainfall monitoring stations and two river flow monitoring stations were identified, as shown in Figure 3.

Table 1. Technical data sheet of the Jucazinho Dam.

Name	Dimension	Unit
Embankment		
Latitude	07°57'59.39" S	-
Longitude	35°44'3.16" W	-
Incremental drainage area	2865.60	km ²
Total drainage area	4149.90	km ²
Maximum volume	204.82	hm ³
Minimum volume	0.29	hm ³
Usable volume	326.75	hm ³
Maximum operating water level	292.00	m
Minimum operating water level	253.00	m
Elevation of the bottom of the lake	238.00	m

Table 1. *Cont.*

Name	Dimension	Unit
Crest		
Length	442.00	m
Width	8.00	m
Elevation	299.00	m
Main Spillway		
Length	170.00	m
Crest elevation	292.00	m
Distance between spillway and embankment crest	7.00	m
Maximum flow rate	5446.69	m ³ ·s ⁻¹
Side Spillways		
Length	57.00	m
Crest elevation	295.00	m
Maximum flow rate	1291.30	m ³ ·s ⁻¹
Gallery		
Length	2.00	m
Width	2.00	m
Elevation	250.00	m
Maximum flow rate	2.72	m ³ ·s ⁻¹

To choose stations for the study, those with records within the same time frame were initially selected. Fluviometric stations 39100000 and 39130000 had records matching with pluviometric stations 735159, 736040, 736041, 736042 and 836092 and were the most recently updated and were thus chosen for the study.

A key aspect of hydrological modeling is the correlation between rainfall and flow data. Rainfall drives surface runoff and groundwater recharge, directly affecting river flow levels. Spearman's correlation (ρ), a robust non-parametric measure, was used to assess this correlation, as shown in Equation (1). It ranges from -1 to 1 and indicates negative ($\rho < 0$), positive ($\rho > 0$) or no correlation ($\rho = 0$).

$$\rho = 1 - \left(\frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)} \right) \quad (1)$$

where d_i and n are, respectively, the difference in ranks between the original series and the series sorted in ascending order for the i -th observation and the total number of observations.

The highest correlation was observed between pluviometric station 736042 and fluvio-metric station 39130000, as demonstrated in Figure 4, with a correlation coefficient of 0.18, leading to their selection for the study.

Despite the low correlation, this research addresses a real situation where the case study (Jucazinho Dam) was selected by the study funder (COMPESA). The choice was due to its importance to the region, its history of "collapse", and its operation based on the technicians' empiricism. Precisely due to data limitations, this study can contribute to the literature by evaluating artificial intelligence models in real situations with scarce data.

The general data of both stations are contained in Table 2. Due to the beginning and end of both series, the records from 1 January 1986 to 1 June 2023 were used to develop the hydrological model.

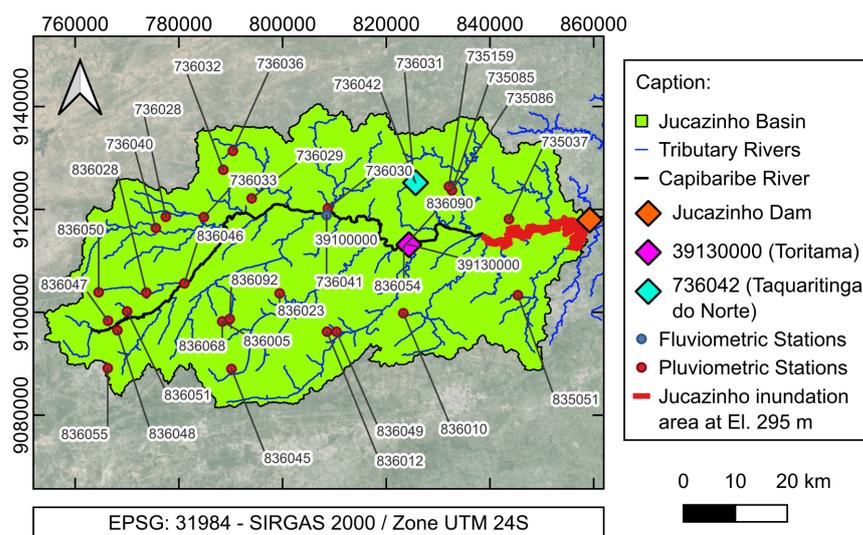


Figure 3. Pluviometric and fluvimetric stations in the Jucazinho Dam’s catchment area.

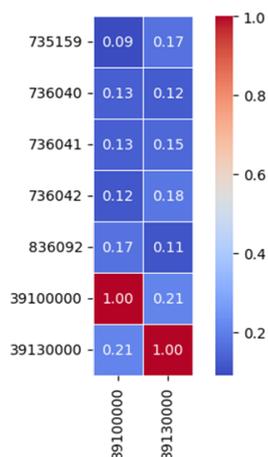


Figure 4. Correlation matrix graph.

Regarding the flow data, to fill the faults of station 39130000, data from station 39100000 were used and were multiplied by a factor of 1.57, referring to the ratio between the drainage area of 2450 km² of station 39130000 and the drainage area of 1560 km² of station 39100000, which were obtained through data from ANA’s HIDROWEB portal.

It is noteworthy that both stations are located on the Capibaribe River, which is the main river of the Capibaribe Hydrographic Basin, which is barred by the Jucazinho Dam, and that the number of faults is insignificant when compared to the total series, allowing, without major damage to the model, this type of fault filling.

The rest of the missing data, both flow and precipitation, for stations 736042 and 39130000, were interpolated linearly. The series with the gaps filled is presented in Figure 5 and covers a total of 13,668 days. Initially, stations 736042 and 39130000 presented, respectively, 0.61% and 5.38% of failures in this number of days. After using the data from station 39100000 multiplied by the factor of 1.57, the percentage of failures from station 39130000 dropped to 4.15%. Finally, both series presented 13,668 days of records without failures.

It should be noted that linear interpolation is not the appropriate methodology for filling in daily failures of a precipitation or flow series, but specifically because we were filling for a period with low precipitation and flow, with values equal to zero or close to it, it was acceptable to apply the method without the association of significant errors in the results of the filled series.

Table 2. General data of the pluviometric and fluviometric stations used.

Information	Pluviometric Station	Fluviometric Station
Station Name	Taquaritinga do Norte	Toritama
Code	736042	39130000
Basin	3—Atlantic, NW/NE section	3—Atlantic, NW/NE section
Sub-basin	39—Capibaribe, Ipojuca, Una, Goiana, Mundaú, Paraíba do Meio, Coruripe, Sirinhaém, São Miguel and Camaragibe Rivers	39—Capibaribe, Ipojuca, Una, Goiana, Mundaú, Paraíba do Meio, Coruripe, Sirinhaém, São Miguel and Camaragibe Rivers
City	Taquaritinga do Norte	Santa Cruz do Capibaribe
State	Pernambuco	Pernambuco
Accountable	ANA	ANA
Operator	Geological Survey of Brazil (CPRM)	CPRM
Latitude	−7.9039	−8.0128
Longitude	36.0469	−36.0578
Elevation (m)	785	376
Drainage area (km ²)	-	2450
Distance to Jucazinho Dam (km)	34.31	35.16
Start of the series	1 January 1986	1 January 1973
End of the series	30 June 2023	1 June 2023
Series size (years)	36.5	49.5

It also should be noted that fluviometric station 391300000 is about 35.16 km from the Jucazinho Dam embankment and 14.64 km from the Jucazinho inundation area, and side spillway crest elevations are about 295 m. Therefore, the flow series generated by the artificial intelligence models trained from the data presented in Figure 5 should be multiplied by a factor of 1.69 for a practical application of reservoir management. This factor refers to the ratio between the drainage area of 4149.90 km² of the Jucazinho Dam and the drainage area of 2450.00 km² of station 39130000.

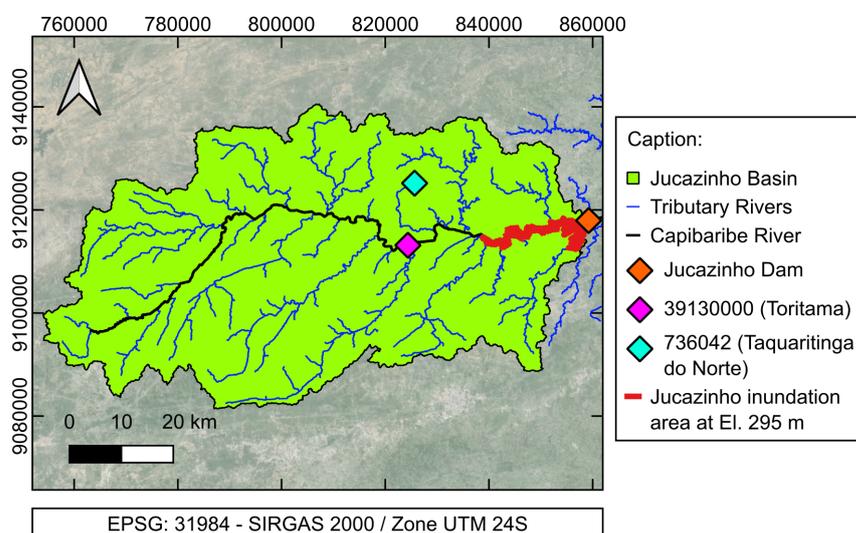
**Figure 5.** Data series used in this study.

Figure 6 presents the average of the monthly accumulations of all the years of the catchment basin. It is verified that there is little rainfall in the area, with approximately four months of rain and eight months of drought, indicating that the Capibaribe watershed has no hydrological memory.

The hydrological memory of a watershed represents its ability to store and release water over time in response to climatic conditions. It is influenced by factors such as geology and land use, and basins with permeable soils tend to have greater hydrological memory.

Hydrological models face challenges in basins without hydrological memory because they may have less predictable responses, impairing the model's ability to capture anomalous climate events, given that temporal variability in water retention and release directly influences the hydrological response.

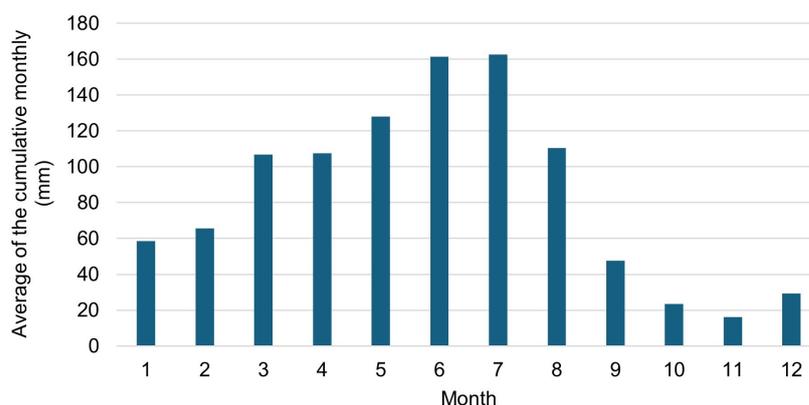


Figure 6. Average of the cumulative monthly index of all the years in the series.

2.3. Model Construction

For the development of the model, the input variables are presented in Table 3; the model predicts a sequence of next steps from a sequence of past observations. As $Q_{(t-1)}$ represents the flow rate for the time prior to $Q_{(t)}$, the delayed flows of 1, 2, 3, ..., 32 days with respect to t are called, respectively, $Q_{(t-1)}, Q_{(t-2)}, \dots, Q_{(t-32)}$; likewise, the flows with an advance of 1, 2, 3, ..., 32 days in relation to t are called, respectively, $Q_{(t+1)}, Q_{(t+2)}, \dots, Q_{(t+32)}$. The same nomenclature logic is used for precipitation.

Thus, in the first scenario, the prediction of the flow for the next day was made from the previous data of one day of flow and precipitation (C-1). In the second scenario, previous data from two days of flow and precipitation were considered in order to predict the next two days of flow (C-2). In the other scenarios, the same logic was used but with 4 (C-4), 8 (L-8), 16 (L-16) and 32 days of data (L-32). The models were grouped into:

- (Group C): Short-term prediction for 1 (C-1), 2 (C-2) and 4 (C-4) days of prediction;
- (Group L): Long-term prediction for 8 (L-8), 16 (L-16) and 32 days (L-32) of prediction.

The models were constructed in an orderly manner without mixing the chronological order of the pairs containing the input and output variables and with recursive, repeating values in these pairs. Using model C-1 as an example, we show the pairs in Table 4.

Note that $P_{(t-2)}, Q_{(t-2)}$ and $Q_{(t+1)}$ are repeated in both the input and output sets. In this way, historical values are used both to predict future flow values and to provide information about past patterns that influence these predictions.

In a short-term context, usually covering periods of up to a week, flow prediction is essential for immediate decision-making. This includes real-time control of water flow, flood prevention, and reservoir water level management. The ability to anticipate intense weather events or sudden changes in hydrological conditions allows for rapid responses, such as the controlled release of water to prevent flooding or the immediate adjustment of the stored volume to meet current demand.

Table 3. Structure of the model for predicting flows.

Template	Input	Output
C-1	$P_{(t-1)}, Q_{(t-1)}$	$Q_{(t)}$
C-2	$P_{(t-1)}, P_{(t-2)}, Q_{(t-1)}, Q_{(t-2)}$	$Q_{(t)}, Q_{(t+1)}$
C-4	$P_{(t-1)}, P_{(t-2)}, \dots, P_{(t-4)}, Q_{(t-1)}, Q_{(t-2)}, \dots, Q_{(t-4)}$	$Q_{(t)}, Q_{(t+1)}, Q_{(t+2)}, Q_{(t+3)}$
L-8	$P_{(t-1)}, P_{(t-2)}, \dots, P_{(t-8)}, Q_{(t-1)}, Q_{(t-2)}, \dots, Q_{(t-8)}$	$Q_{(t)}, Q_{(t+1)}, \dots, Q_{(t+7)}$
L-16	$P_{(t-1)}, P_{(t-2)}, \dots, P_{(t-16)}, Q_{(t-1)}, Q_{(t-2)}, \dots, Q_{(t-16)}$	$Q_{(t)}, Q_{(t+1)}, \dots, Q_{(t+15)}$
L-32	$P_{(t-1)}, P_{(t-2)}, \dots, P_{(t-32)}, Q_{(t-1)}, Q_{(t-2)}, \dots, Q_{(t-32)}$	$Q_{(t)}, Q_{(t+1)}, \dots, Q_{(t+31)}$

Table 4. Ordering and recursion of models, using model C-1 as an example.

Pair	Input	Output
1	$P_{(t-1)}, P_{(t-2)}, Q_{(t-1)}, Q_{(t-2)}$	$Q_{(t)}, Q_{(t+1)}$
2	$P_{(t)}, P_{(t-1)}, Q_{(t)}, Q_{(t-1)}$	$Q_{(t+1)}, Q_{(t+2)}$
3	$P_{(t+1)}, P_{(t)}, Q_{(t+1)}, Q_{(t)}$	$Q_{(t+2)}, Q_{(t+3)}$
...
n	$P_{(t+n-2)}, P_{(t+n-3)}, Q_{(t+n-2)}, Q_{(t+n-3)}$	$Q_{(t+n-1)}, Q_{(t+n)}$

On the other hand, from a long-term perspective, usually covering periods longer than a week, flow forecasting is crucial for strategic planning. This allows for gradual adaptation to seasonal hydrological conditions, contributing to the sustainable management of water resources over time.

Regarding their classification, hydrological models based on artificial intelligence can be adapted to generate both probabilistic and deterministic predictions. Because the model generates unique and specific predictions based on initial conditions and defined parameters, it is reasonable to classify it as deterministic.

The input and output data were normalized between 0 and 1 to prevent the predominance of variables with high values, which is common in machine learning models. It is noteworthy that the normalization was done by variable: for example, the normalization of $Q_{(t-1)}$ considers the data series only of $Q_{(t-1)}$. For this process, the MinMaxScaler function of the Sklearn library was used, which performs the transformation through Equation (2):

$$x'_n = \frac{x_n - x_{min}}{x_{max} - x_{min}} \quad (2)$$

where x_n is the n th value of the data series, x'_n is the value after normalization, and x_{min} and x_{max} are, respectively, the minimum and maximum values of the data series.

To apply the machine learning models, the data were divided into training and testing using the `traintestsplit` function from the Sklearn library, and a sensitivity analysis was performed considering the following proportions:

- 65% for training and 35% for testing (65–35);
- 70% for training and 30% for testing (70–30);
- 75% for training and 25% for testing (75–25);
- 80% for training and 20% for testing (80–20).

2.3.1. Defining Hyperparameters

Hyperparameters are external parameters that are not learned directly by the model during training but need to be specified before the training process begins. These parameters influence the overall behavior of the model and affect training performance and

effectiveness. In contrast, model parameters are the weights that the model adjusts during training to make predictions based on the data.

Because the time series is large-scale, which significantly increases the computational cost, and we aim to obtain an initial reference point to evaluate the adaptability of machine learning models, the present work used the standard values for the hyperparameters. These values, because they are generic and work well in a variety of situations, provide computational efficiency, avoiding the initial need for extensive experimentation.

2.3.2. Support Vector Machine

In the SVM model, the SVR function of the Sklearn library was used, for which the list of hyperparameters with their respective types of variables and default values are provided in Table 5.

The SVR machine learning model is an extension of the SVM algorithm for regression tasks. The original SVM was initially developed for classification problems, but the SVR variant has been adapted to handle the prediction of numerical values instead of classes.

The core logic behind SVR is the same as for SVM and involves searching for an optimal hyperplane that best fits the training data. However, unlike SVM classification, where the goal is to find a hyperplane that separates classes efficiently, SVR seeks a hyperplane that optimizes the prediction of continuous values.

Table 5. Hyperparameters of the SVR function from the Sklearn library.

Parameter	Variable Type	Default Value
Kernel	{'linear', 'poly', 'rbf', 'sigmoid', 'precomputed'} or callable	'rbf'
Degree	int	3.0
Gamma	{'scale', 'auto'} or float	Scal
coef0	float	0.0
tol	float	0.001
C	float	1.0
Epsilon	float	0.1
Shrinking	bool	true
cache_size	float	200
Verbose	bool	false
max_iter	int	−1.0

2.3.3. Random Forest

In the RF model, the RandomForestRegressor function of the Sklearn library was used, for which the list of hyperparameters with their respective types of variables and default values is provided in Table 6.

The RandomForestRegressor machine learning model resides in the construction of an ensemble of decision trees, which are a fundamental component of RF. Each tree is trained independently using random sampling of both the dataset instances and the characteristics in each node split. This random approach contributes to the diversity among the trees and, consequently, to the robustness of the model.

During the training process, each tree makes individual predictions for the dataset instances according to the decisions made in their structures. The final RandomForestRegressor prediction is obtained by averaging these predictions, resulting in a more stable estimate that is less susceptible to overfitting.

Table 6. Hyperparameters of the RandomForestRegressor function of the Sklearn library.

Parameter	Variable Type	Default Value
n_estimators	int	100
criterion	string: {"squared_error", "absolute_error", "friedman_mse", "poisson"}	squared_error
max_depth	int	None
min_samples_split	int or float	2.0
min_samples_leaf	int or float	1.0
min_weight_fraction_leaf	float	0.0
max_features	string: {"sqrt", "log2", None}, int or float	1.0
max_leaf_nodes	int	None
min_impurity_decrease	float	0.0
bootstrap	bool	true
oob_score	bool	false
n_jobs	int	None
random_state	int, RandomState or None	None
verbose	Int	0.0
warm_start	bool	False
ccp_alpha	non-negative float	0.0
max_samples	int or float	None

Overfitting is a common phenomenon in machine learning in which a model over-adapts to the specific details of the training data, losing the ability to generalize to new data. This occurs when the model is too complex relative to the inherent complexity of the data, capturing irrelevant patterns, noise, or specific variations of the training set.

As a result, the model's performance may be excellent on the training data but may fail to tackle new data, hindering its ability to make accurate predictions in real-world situations.

2.3.4. Artificial Neural Network

In the ANN model, the MLPRegressor function from the Sklearn library was used, for which the list of hyperparameters with their respective types of variables and default values is provided in Table 7.

The MLPRegressor machine learning model belongs to the category of ANNs known as MLPs. This model is specifically designed for regression tasks as it is able to perform predictions of numerical values based on input data.

The fundamental structure of MLPRegressor is composed of multiple layers of neurons, with each layer connected to its adjacent layers. This architecture allows the model to capture complex relationships between input and output variables. Unlike simple linear models, MLPRegressor is able to learn nonlinear patterns in data.

During training, MLPRegressor uses an iterative process known as backpropagation. This process involves forward-passing inputs through the network to generate predictions and then comparing those predictions with the actual values to calculate the error. The error is then backpropagated through the network, and the weights of the connections between neurons are adjusted to minimize the error in the next iteration.

Table 7. Hyperparameters of the MLPRegressor function from the Sklearn library.

Parameter	Variable Type	Default Value
hidden_layer_sizes	array-like of shape(n_layers-2,)	100
activation	{'identity', 'logistic', 'tanh', 'relu'}	'relu'
solver	{'lbfgs', 'sgd', 'adam'}	'adam'
alph	float	0.0001
batch_size	int	auto
learning_rate	{'constant', 'invscaling', 'adaptive'}	'constant'
learning_rate_init	float	0.001
power_t	float	0.5
max_iter	int	200
shuffle	bool	True
random_state	int	None
tol	float	0.0001
verbose	bool	False
fwarm_start	bool	False
momentum	float	0.9
nesterovs_momentum	bool	True
early_stopping	bool	False
validation_fraction	float	0.1
beta_1	float	0.9
beta_2	float	0.999
epsilon	float	0.00000001
n_iter_no_change	int	10
max_fun	int	15,000

2.4. Model Evaluation

The models were evaluated for their performance by means of a graphical analysis and statistical criteria. The purpose of the evaluation was to verify the quality of the calibration and validation results by comparing the flow data simulated by the models with the actual observed data.

The adopted graphical analysis sought to verify the trend of the correlation of the observed and predicted data along the increase in the flow and also with the increase in days according to the models.

The statistical criteria adopted were the Nash–Sutcliffe model efficiency coefficient [18] (NSE), the percentage of bias (PBIAS), the coefficient of determination (R^2) and the root mean standard deviation ratio (RSR). The NSE ranges from $-\infty$ to 1, with 1 being representative of the optimal value. Values between 0 and 1 are seen as acceptable performance levels; however, values ≤ 0 indicate that the observed mean is a better predictor than the simulated value, indicating poor performance of the model.

The optimal value of the PBIAS is 0, and positive or negative values with low magnitudes represent good performance. Positive values indicate that the model underestimated measured values, while negative values indicate that the model overestimated measured values.

The R^2 estimates the correlation between the measured and simulated values and ranges from 0 to 1, with a value of 1 representing perfect agreement. The RSR ranges from the optimal value of 0, which indicates that the root mean square error (RMSE) is zero, to a large positive value. Therefore, the lower the RSR, the lower the RMSE and the better the simulation performance.

In the quantitative analysis, the Moriasi et al. [19] classification presented in Table 8 was used. The classification used the evaluation of a monthly model, and therefore, for simulations with daily values, it can be considered that NSE values above 0.36 are still satisfactory.

Table 8. Classification of modeling efficiency coefficients.

Classification	NSE	PBIAS
Very Good	$0.75 < \text{NSE} \leq 1.00$	$\text{PBIAS} \leq \pm 10$
Good	$0.65 < \text{NSE} \leq 0.75$	$\pm 10 < \text{PBIAS} \leq \pm 15$
Satisfactory	$0.50 < \text{NSE} \leq 0.65$	$\pm 15 < \text{PBIAS} \leq \pm 25$
Unsatisfactory	$\text{NSE} \leq 0.50$	$\text{PBIAS} \geq \pm 25$
Classification	R^2	RSR
Very Good	$0.8 < R^2 \leq 1.0$	$0.0 < \text{RSR} \leq 0.5$
Good	$0.7 < R^2 \leq 0.8$	$0.5 < \text{RSR} \leq 0.6$
Satisfactory	$0.6 < R^2 \leq 0.7$	$0.6 < \text{RSR} \leq 0.7$
Unsatisfactory	$R^2 \leq 0.6$	$\text{RSR} \geq 0.6$

Graphical analyses were performed using scatter plots for all models and graphs of readings over time for models C-1 and L-32. The graph of readings over time for the L-32 model was created considering only the first, fifteenth and last day of the pairs, considering that the values are repeated. Both graphs depict the relationships between two variables visually.

The scatter plot displays observations comparing the independent variable on the x -axis to the dependent variable on the y -axis. The pattern of points on this graph offers insights into the relationship's nature: indicating whether it follows a linear trend and assessing the model's fit quality. The second graph shows readings over time, with each point representing observations comparing dates on the x -axis to flows measured at the pluviometric station and predicted by the model on the y -axis. The closer the generated curves, the better the model fit. This graphical representation allows for the identification of difficulties with predicting highs or lows in the series and detecting any delays between the graphs.

3. Results and Discussion

The results of the statistical efficiency coefficients of the models are shown in Table 9 for the test data. It is observed that, in general, all models showed unsatisfactory performance for long-term prediction, which includes 8 (L-8), 16 (L-16) and 32 (L-32) days for every training and testing division (65–35, 70–30, 75–25 and 80–20, where the first number is the percentage of data used for training, and the second number is the percentage of data used for testing, as described in the methodology). In addition, with the increase in prediction days, there is a clear tendency to decrease the quality of the fit.

It is also important to note that, despite the low correlation between the pluviometric and fluviometric stations used in the study, the artificial intelligence models SVM, RF and ANN presented acceptable results for the models for short-term prediction, which includes one day (C-1), two days (C-2) and four days (C-4).

Cheng et al. [20] adopted an ANN and long short-term memory (LSTM) to predict flow rates at daily and monthly scales. For the monthly flow forecast, a recursive prediction approach was used. The two models were trained and validated using precipitation and flow datasets collected in the Nan and Ping river basins in Thailand, covering the period from 1974 to 2014. The main findings of the study highlight that both ANN and LSTM models can provide accurate daily predictions of up to 20 days.

It is not uncommon for the performance of flow prediction models using artificial intelligence to decrease with increasing prediction horizons. This comparison suggests that the accuracy of long-term flow forecasting can vary significantly depending on the modeling method and the approach taken.

The decrease in the quality of adjustments with the increase in the number of forecast days is attributed to increasing uncertainty regarding future conditions and the accumulation of errors over the forecast horizon. According to the statistical criteria, the model that best adapted to the series was SVM 80-20 (support vector machine using 80% for training and 20% for testing), which resulted in the best statistical indices of NSE, PBIAS, R² and RSR according to Table 10.

Al-Mukhtar [21] investigated the modeling and prediction of sediment in the Tigris River in Baghdad, which is an influential parameter for the pollution of water bodies. Three artificial intelligence methods (RF, SVM and ANN) were employed using observed flow (m³/s) and suspended sediment concentration (mg/L) data collected between 1962–1981 and 2000–2010. The predictive results of the three methods evaluated were analyzed based on the coefficients R², RMSE and NSE and indicated that RF presented the best performance.

These findings highlight the importance of considering the variation in performance between different forecasting methods as well as the need to properly assess the uncertainties associated with models, especially for long-term forecasting scenarios.

Table 9. Statistical criteria of training.

NSE												
	SVM				RF				ANN			
	65–35	70–30	75–25	80–20	65–35	70–30	75–25	80–20	65–35	70–30	75–25	80–20
C-1	0.54	0.83	0.91	0.94	0.70	0.87	0.85	0.82	0.73	0.92	0.92	0.93
C-2	0.40	0.75	0.85	0.91	0.66	0.85	0.80	0.69	0.71	0.87	0.88	0.79
C-4	0.27	0.63	0.78	0.86	0.57	0.63	0.72	0.56	0.64	0.76	0.77	0.71
L-8	0.16	0.47	0.68	0.80	0.50	0.49	0.62	0.48	0.53	0.60	0.70	0.59
L-16	0.08	0.30	0.52	0.70	0.37	0.35	0.50	0.20	0.41	0.37	0.57	0.38
L-32	0.01	0.17	0.41	0.57	0.21	0.05	0.06	−1.81	0.24	0.15	0.42	0.00
PBIAS												
	SVM				RF				ANN			
	65–35	70–30	75–25	80–20	65–35	70–30	75–25	80–20	65–35	70–30	75–25	80–20
C-1	−27.73	−10.17	−0.09	5.52	−9.63	1.15	7.23	9.04	−20.12	8.42	−11.74	20.32
C-2	−37.38	−16.13	−4.18	2.61	−5.61	1.27	9.64	13.25	−11.94	3.61	−15.1	6.78
C-4	−48.06	−23.69	−7.71	0.94	−13.05	5.97	16.49	20.56	−19.35	6.67	61.92	37.23
L-8	−56.63	−32.24	−13.74	−1.23	−22.89	8.78	23.27	31.75	−16.22	−0.04	−1.77	55.23
L-16	−62.5	−41.70	−23.98	−6.69	−34.88	15.48	34.72	49.17	−24.01	15.22	39.53	22.71
L-32	−70.99	−51.05	−30.33	−14.68	−44.57	23.63	82.32	128.36	−35.57	12.38	53.87	63.46

Table 9. Cont.

R ²												
	SVM				RF				ANN			
	65–35	70–30	75–25	80–20	65–35	70–30	75–25	80–20	65–35	70–30	75–25	80–20
C-1	0.54	0.83	0.91	0.94	0.70	0.87	0.85	0.82	0.73	0.92	0.92	0.93
C-2	0.40	0.75	0.85	0.91	0.66	0.85	0.80	0.69	0.71	0.87	0.88	0.79
C-4	0.27	0.63	0.78	0.86	0.57	0.63	0.72	0.56	0.64	0.76	0.77	0.71
L-8	0.16	0.47	0.68	0.80	0.50	0.50	0.62	0.48	0.53	0.60	0.70	0.59
L-16	0.08	0.30	0.52	0.70	0.37	0.35	0.50	0.20	0.41	0.37	0.57	0.38
L-32	0.01	0.17	0.41	0.57	0.21	0.05	0.03	−1.81	0.24	0.15	0.41	0.00

RSR												
	SVM				RF				ANN			
	65–35	70–30	75–25	80–20	65–35	70–30	75–25	80–20	65–35	70–30	75–25	80–20
C-1	0.67	0.41	0.30	0.25	0.55	0.36	0.38	0.42	0.52	0.28	0.28	0.27
C-2	0.78	0.50	0.38	0.30	0.58	0.38	0.44	0.55	0.54	0.35	0.34	0.46
C-4	0.86	0.61	0.47	0.37	0.66	0.61	0.53	0.66	0.60	0.49	0.48	0.54
L-8	0.92	0.73	0.56	0.44	0.70	0.71	0.61	0.72	0.69	0.63	0.55	0.64
L-16	0.96	0.84	0.69	0.55	0.80	0.80	0.71	0.90	0.77	0.79	0.65	0.79
L-32	0.99	0.91	0.77	0.66	0.89	0.97	0.97	1.68	0.87	0.92	0.76	1.00

Table 10. Better statistical criteria for training.

C-N	NSE	PBIAS	R ²	RSR
C-1	SVM 80-20	SVM 75-25	SVM 80-20	SVM 80-20
C-2	SVM 80-20	RF 70-30	SVM 80-20	SVM 80-20
C-4	SVM 80-20	SVM 80-20	SVM 80-20	SVM 80-20

The results of the models are presented in Figures 7–12. The models took 90% of the code execution time to train, which means that the computational cost of training is significantly higher than what is required to use the trained model.

In the SVM models, with the increase in the number of days, there was underestimation of the flow forecast values (points approaching the x -axis). Despite being effective at modeling nonlinear relationships, the SVM model is sensitive to complex patterns in time series. For this reason, as the forecast horizon increased, the model was unable to predict the significant changes in flow patterns, as can be seen in Figure 7.

The AF and ANN models had the opposite results of the SVM model: with the increase in days, there was an overestimation of the flow prediction values (points approaching the y -axis), as can be seen in Figures 8 and 9. This is mainly due to overfitting, where the models have adjusted too much to the training data and incorporated transient noise and patterns that are not representative of the actual flow behavior.

Han et al. [22] applied SVM to the Bird Creek watershed for flood prediction and found that, like ANN models, SVM also suffers from underfitting and overfitting problems, with overfitting being more harmful than underfitting. The study also reveals an interesting result in the response of the SVM to different rainfall inputs, where lighter rains generated very different responses than more intense rainfall, similar to what occurred in the present work.

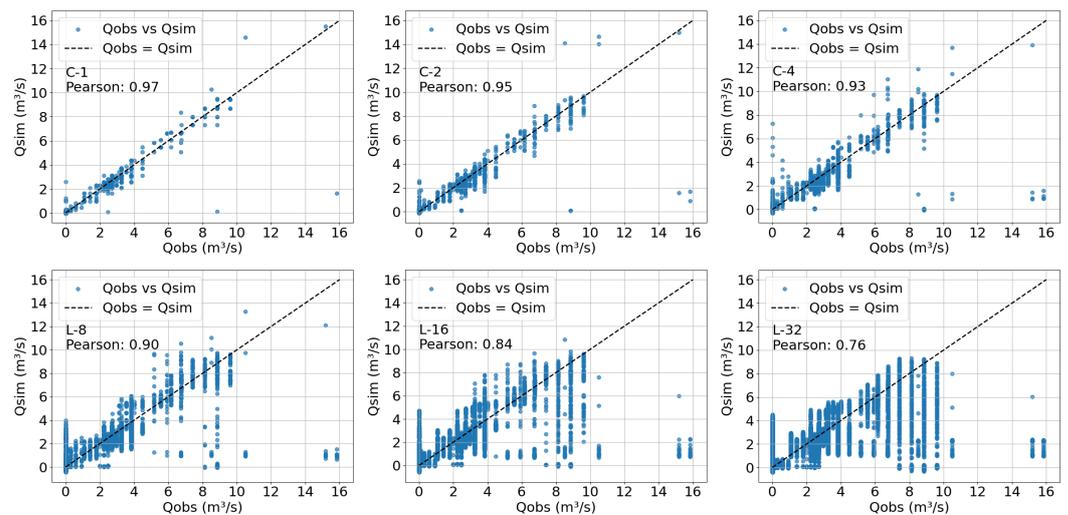


Figure 7. Scatter plots of the tests for SVM 80-20.

On the graph of the readings over time shown in Figure 10, it is possible to verify the underestimation of the peaks mentioned above for the SVM models. The C-1 model was able to predict peaks better than the L-32 model. It is also verified that there is a delay between the graphs of the observed and measured flow for the L-32 model during the increase in the number of pairs of input and output data points. This delay explains the low statistical coefficients because the accumulated error is summed.

According to Figures 11 and 12, the RF models presented similar results as those of the ANN: generating noise that oscillated much above the measured flow. The underlying relationship between the features and the target variable is highly nonlinear and complex; that is why both RF and ANN struggled to capture it effectively, leading to instability and noisy predictions. As with the SVM model, there is also a delay between the graphs of the observed and measured flow for the L-32 model during the increase in the number of pairs of input and output data points.

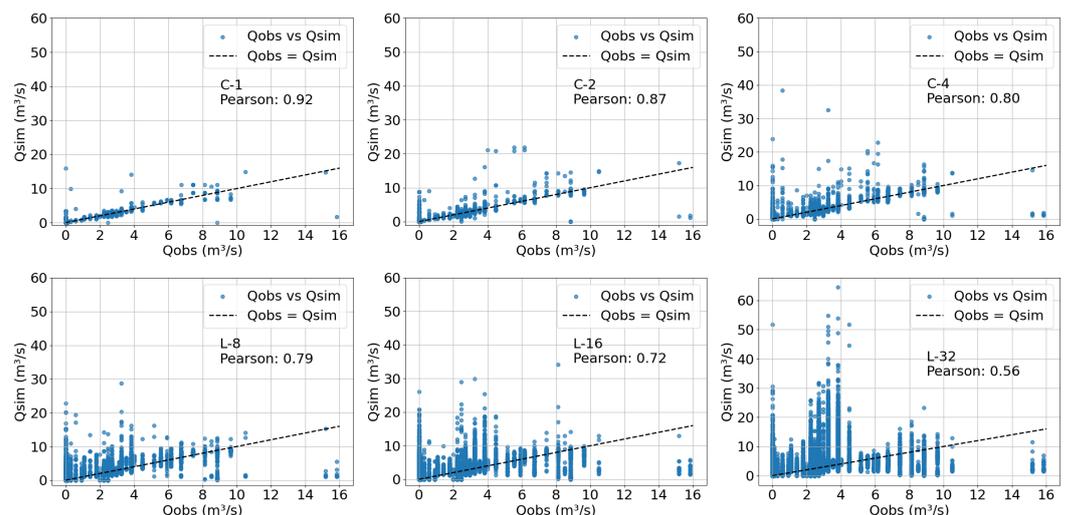


Figure 8. Scatter plots of tests for RF 80-20.

The ANN models were able to represent the flow peaks well, as Figure 12 shows. Above all, generated noise oscillated the values that should be zero between values slightly higher or lower than zero, even predicting negative flow values. As with both models presented above, there is also a delay between the graphs of the observed and measured flow for the L-32 model during the increase in the number of pairs of input and output data points.

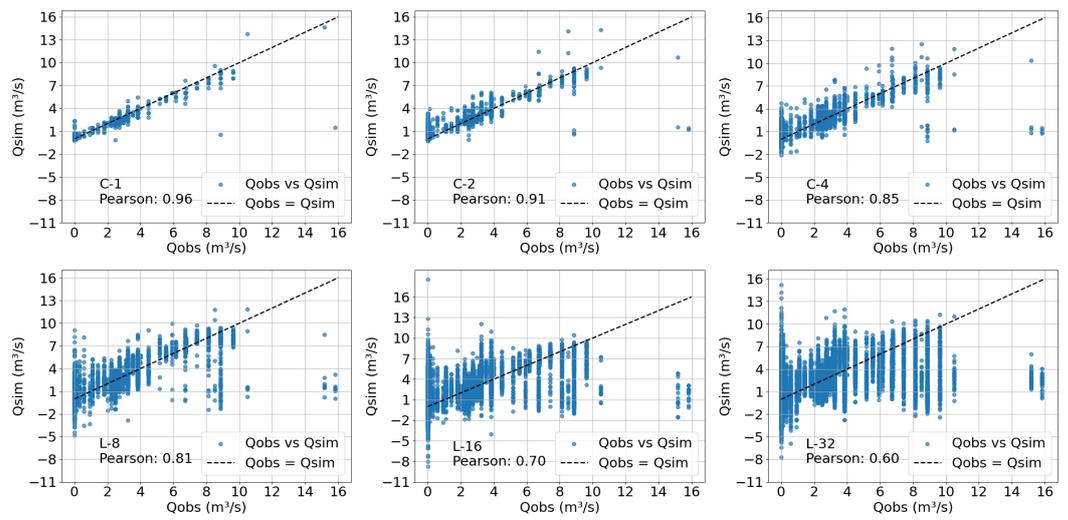


Figure 9. Scatter plots of tests for ANN 80-20.

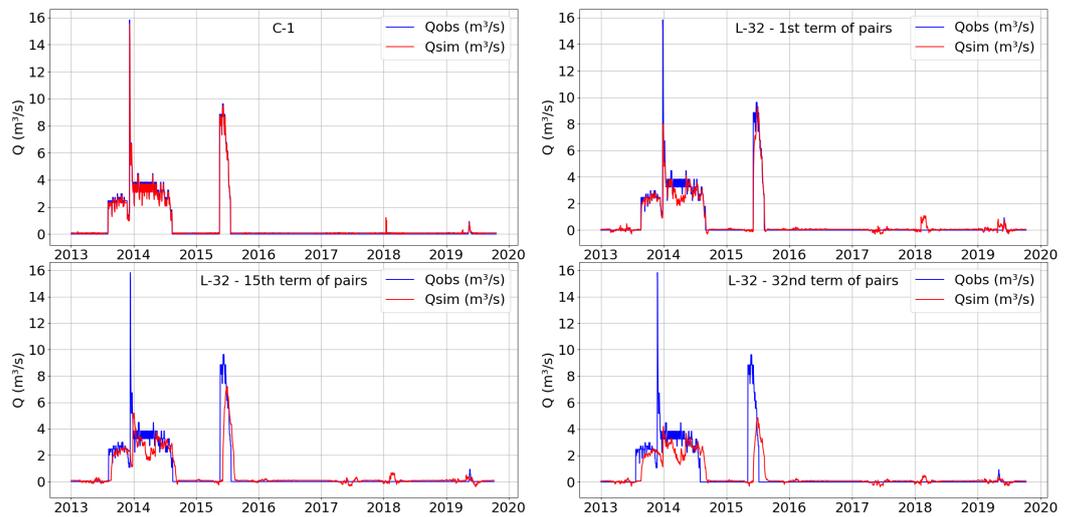


Figure 10. Comparison between observed and simulated flows for SVM 80-20.

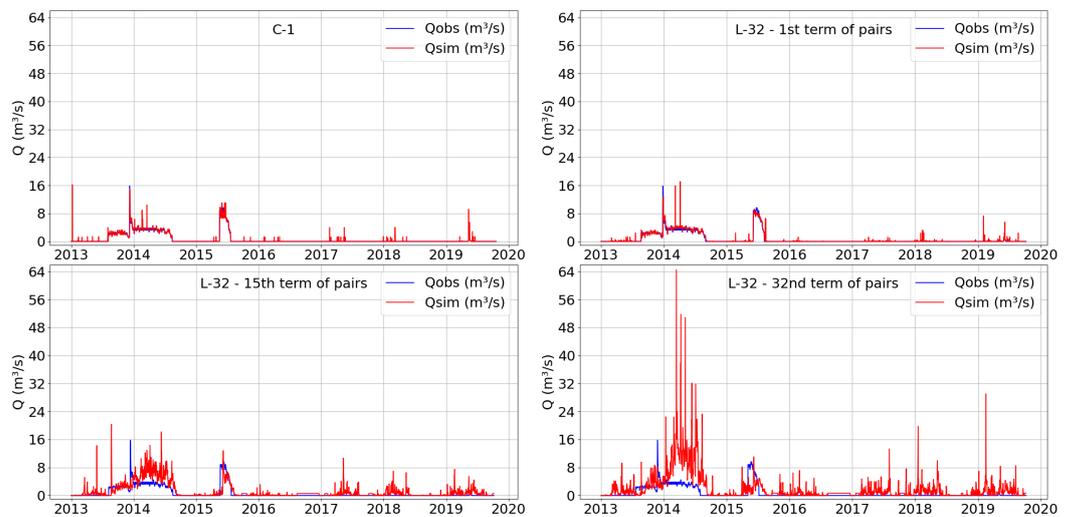


Figure 11. Comparison between observed and simulated flows for RF 80-20.

The results confirm that the Capibaribe watershed does not have hydrological memory. It is recommended in future works using the application of SVM for different watersheds to verify the adaptability of the model according to the hydrological memory of each basin.

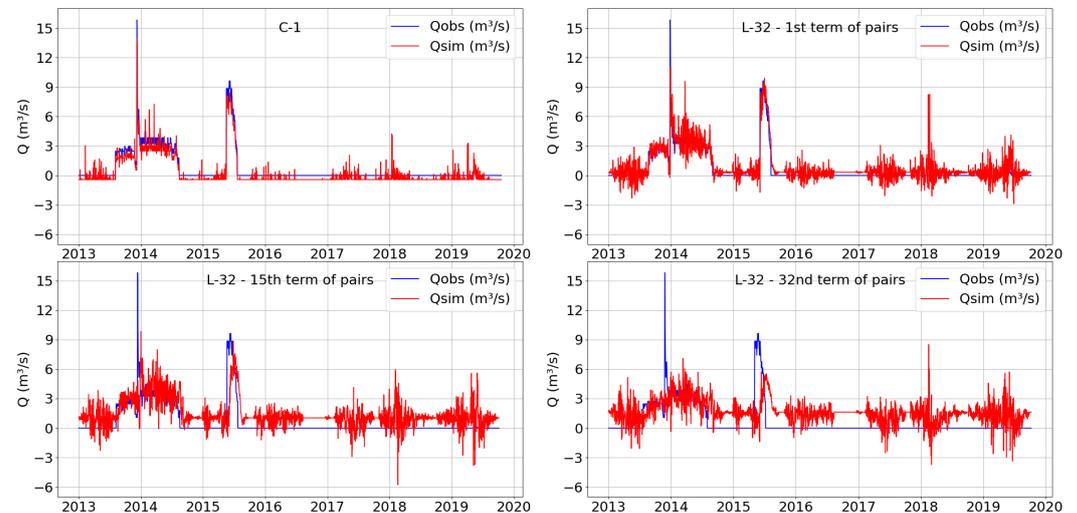


Figure 12. Comparison between observed and simulated flows for ANN 80-20.

4. Conclusions

The general objective of this research was to generate synthetic series of tributary flows to the Jucazinho Dam, which is located in the Agreste region of Pernambuco, based on stochastic models. The specific objectives were to evaluate the adaptability of SVM, RF and ANN machine learning models to generate the synthetic flow series, to analyze the influence of the quality of the adjustments with the increase in the number of days of the forecast, and to verify the quality of the adjustments made by the models using statistical metrics. Based on the results obtained and the discussions presented, it is concluded that:

- All models showed satisfactory performance for short-term prediction, which includes 1, 2 and 4 days, and unsatisfactory for long-term prediction, which includes 8, 16 and 32 days.
- The graphical results of the SVM models showed underestimation of the flow values with an increase in the forecast horizon due to the sensitivity of the SVM to complex patterns in the time series.
- On the other hand, the RF and ANN models showed hyperestimation of the flow values as the number of forecast days increased, which was mainly attributed to overfitting.
- For all models, an increase in the number of prediction days led to a tendency to decrease the quality of the adjustment; this was mainly justified as due to the delay in the predictions, which generated an accumulation of errors.
- According to the statistical criteria, the model that best adapted to the series was SVM, which resulted in the best statistical indices of NSE, PBIAS, R^2 and RSR.
- Even in situations where data are scarce, artificial intelligence models SVM, RF and ANN have the potential to be applied for short-term prediction.
- The Capibaribe watershed does not have hydrological memory, which impacted model training. It is recommended in future works using the application of ANNs in different watersheds to verify the adaptability of the model according to the hydrological memory of each basin.

Author Contributions: Conceptualization, E.J.G.d.S. and A.P.C.; Formal analysis, A.P.C. and S.d.T.M.B.; Investigation, E.J.G.d.S.; Methodology, E.J.G.d.S.; Project administration, S.d.T.M.B.; Software, E.J.G.d.S. and J.F.C.; Supervision, A.P.C. and S.d.T.M.B.; Validation, A.P.C.; Writing—original draft, E.J.G.d.S.; Writing—review and editing, E.J.G.d.S. and S.d.T.M.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Academic Master’s and Doctoral Program for Innovation (MAI/DAI), which was promoted by the Brazilian National Council for Scientific and Technological Development (Conselho Nacional de Desenvolvimento Científico e Tecnológico—CNPq, Brazil), in a partnership between the Federal University of Pernambuco (Universidade Federal de Pernambuco—UFPE, Brazil) and the Pernambuco Sanitation Company (Companhia Pernambucana de Saneamento—COMPESA, Brazil).

Data Availability Statement: The data presented in this study are available on request from the corresponding authors.

Acknowledgments: The authors would like to thank Vinnycius Luz and Milton Melo Neto from the Pernambuco Sanitation Company (Companhia Pernambucana de Saneamento—COMPESA, Brazil), the National Council for Scientific and Technological Development (Conselho Nacional de Desenvolvimento Científico e Tecnológico—CNPq, Brazil) for the productivity scholarships for Artur Coutinho [process 315927/2021-6] and Saulo Bezerra [process 308202/2022-8], the Foundation for Support of Science and Technology of the State of Pernambuco (Fundação de Amparo à Ciência e Tecnologia de Pernambuco—FACEPE, Brazil) [process APQ-1767-3.01/22], and the Coordination for the Improvement of Higher Education Personnel (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—CAPES, Brazil) [Finance Code 001].

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Jansen, R.B. *Dams and Public Safety: A Water Resources Technical Publication*; United States Printing Office: Denver, CO, USA, 1980.
2. World Commission on Dams. *Dams and Development: A New Framework for Decision-Making: The Report of the World Commission on Dams*; Earthscan: Oxford, UK, 2000.
3. Agência Pernambucana de Águas e Clima-Apac. Available online: <https://acesse.one/hyipu> (accessed on 20 November 2022).
4. Agência Nacional das Águas-Ana. *Reservatórios do Semiárido Brasileiro: Hidrologia, Balanço Hídrico e Operação*; ANA: Brasília, Brazil, 2017; Volume Anexo E; 178 p.
5. Companhia Pernambucana de Saneamento-Compesa. Available online: <https://11nq.com/RCLDM> (accessed on 20 November 2022).
6. Santana, R.A.; Bezerra, S.T.M.; Santos, S.M.; Coutinho, A.P.; Coelho, I.C.L.; Pessoa, R.S.V. Assessing alternatives for meeting water demand: a case study of water resource management in the Brazilian Semi-arid region. *Util. Policy* **2019**, *61*, 100974.
7. Ali, S.; Shahbaz, M. Streamflow forecasting by modeling the rainfall–streamflow relationship using artificial neural networks. *Model. Earth Syst. Environ.* **2020**, *6*, 1645–1656. [[CrossRef](#)]
8. Parisouj, P.; Mohebzadeh, H.; Lee, T. Employing machine learning algorithms for streamflow prediction: a case study of four river basins with different climatic zones in the United States. *Water Resour. Manag.* **2020**, *34*, 4113–4131. [[CrossRef](#)]
9. Meshram, S.G.; Meshram, C.; Santos, C.A.G.; Benzougagh, B.; Khedher, K.M. Streamflow prediction based on artificial intelligence techniques. *Iran. J. Sci. Technol. Trans. Civ. Eng.* **2022**, *46*, 2393–2403. [[CrossRef](#)]
10. Sun, N.; Zhang, S.; Peng, T.; Zhang, N.; Zhou, J.; Zhang, H. Multi-variables-driven model based on random forest and Gaussian process regression for monthly streamflow forecasting. *Water* **2022**, *14*, 1828. [[CrossRef](#)]
11. Islam, K.I.; Elias, E.; Carroll, K.C.; Brown, C. Exploring random forest machine learning and remote sensing data for streamflow prediction: An alternative approach to a process-based hydrologic modeling in a snowmelt-driven watershed. *Remote Sens.* **2023**, *15*, 3999. [[CrossRef](#)]
12. de Sousa, M.F., Jr.; Uliana, E.M.; Aires, R.V.; Rápalo, L.M.; da Silva, D.D.; Moreira, M.C.; Lisboa, L.; da Silva Rondon, D. Streamflow prediction based on machine learning models and rainfall estimated by remote sensing in the Brazilian Savanna and Amazon biomes transition. *Model. Earth Syst. Environ.* **2024**, *10*, 1191–1202. [[CrossRef](#)]
13. Adnan, R.M.; Liang, Z.; Heddam, S.; Kermani, M.; Kisi, O.; Li, B. Least square support vector machine and multivariate adaptive regression splines for streamflow prediction in mountainous basin using hydro-meteorological data as inputs. *J. Hydrol.* **2020**, *586*, 124371. [[CrossRef](#)]
14. Essam, Y.; Huang, Y.F.; Ng, J.L.; Birima, A.H.; Ahmed, A.N.; El-Shafie, A. Predicting streamflow in Peninsular Malaysia using support vector machine and deep learning algorithms. *Sci. Rep.* **2022**, *12*, 3883. [[CrossRef](#)] [[PubMed](#)]
15. Ikram, R.M.A.; Hazarika, B.B.; Gupta, D.; Heddam, S.; Kisi, O. Streamflow prediction in mountainous region using new machine learning and data preprocessing methods: a case study. *Neural Comput. Appl.* **2023**, *35*, 9053–9070. [[CrossRef](#)]

16. Girão, L.C.P. Uma Análise da Contribuição dos Programas Básicos Ambientais Como Instrumento de Gestão Ambiental Para a Barragem de Jucazinho Localizada no Município de Surubim/PE. Master's Thesis, Universidade Federal de Pernambuco-UFPE, Recife, Brazil, 2004.
17. Neves, Y.T.; Rodrigues, A.; Cabral, J.J.S.P. Modelagem computacional do rompimento hipotético da barragem de Jucazinho no estado de Pernambuco (Brasil). *Rev. DAE* **2021**, *69*, 167–182. [[CrossRef](#)]
18. Nash, J.E.; Sutcliffe, J.V. River flow forecasting through conceptual models part I-A discussion of principles. *J. Hydrol.* **1970**, *10*, 282–290. [[CrossRef](#)]
19. Moriasi, D.N.; Arnold, J.G.; van Liew, M.W.; Bingner, R.L.; Harmel, R.D.; Veith, T.L. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Am. Soc. Agric. Biol. Eng.* **2007**, *50*, 885–900.
20. Cheng, M.; Fang, F.; Kinouchi, T.; Navon, I.M.; Pain, C.C. Long lead-time daily and monthly streamflow forecasting using machine learning methods. *J. Hydrol.* **2020**, *590*, 125376. [[CrossRef](#)]
21. Al-Mukhtar, M. Random forest, support vector machine, and neural networks to modelling suspended sediment in Tigris River-Baghdad. *Environ. Monit. Assess.* **2019**, *191*, 673. [[CrossRef](#)] [[PubMed](#)]
22. Han, D.; Chan, L.; Zhu, N. Flood forecasting using support vector machines. *J. Hydroinform.* **2007**, *9*, 267–276. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.