

## Estimating flow rates in a tropical basin using artificial neural modeling based on hydrological memory

### Estimación de caudales en una cuenca tropical mediante modelación neuronal artificial basada en memoria hidrológica

León Darío Blandón-López<sup>1</sup>   Luis Octavio González-Salcedo<sup>1</sup> 

<sup>1</sup>Corporación Autónoma Regional del Valle del Cauca, Palmira Colombia,

<sup>2</sup>Departamento de Ingeniería, Facultad de Ingeniería y Administración, Universidad Nacional de Colombia.

#### How to cite?

Blandón-López LD, González-Salcedo LO. Estimating flow rates in a tropical basin using artificial neural modeling based on hydrological memory. Ingeniería y Competitividad, 2026, 28(1)e-20415199

<https://doi.org/10.25100/iyv.v28i1.15199>

Received: 31/08/25

Reviewed: 05/11/25

Accepted: 21/01/26

Online: 18/02/26

#### Correspondence

lblandon@unal.edu.co

## Abstract

**Introduction:** Hydrological systems have been severely impacted by the increase in extreme events such as floods and droughts, intensified by climate change. The Tuluá River basin, located in the Cauca Valley of Colombia, faces additional challenges due to deforestation and limited monitoring coverage, complicating efficient water resource management.

**Objectives:** Develop a neural model based on LSTM networks to predict the flow rate of the Tuluá River from hydro-meteorological variables, for the anticipation of extreme scenarios and support in decision-making in regional water management.

**Methodology:** 6875 daily records for five variables (precipitation, temperature, evaporation, relative humidity, and streamflow) from seven stations were consolidated. To fill missing values, data preprocessing was applied using KNN and Pearson correlation analysis. The LSTM model was configured and trained, and evaluated for the training, test, and validation subsets using standard neural network metrics.

**Results:** The neural model demonstrated high predictive capacity with correlation coefficients  $R$  (0.75–0.98). It was evident that the model effectively captures temporal dependencies and nonlinear patterns, in addition to identifying the influence of the ENSO phenomenon on local hydrological dynamics.

**Conclusions:** The LSTM model is an effective tool for estimating flows for the basin under study, facilitating both water management and the prevention of risks associated with climate variability. Future developments can enhance its accuracy and usefulness if they include global climate indices.

**Keywords:** Artificial intelligence, climate change, evapotranspiration, flooding, precipitation, resource management, river basin, simulation model, sustainable development, temperature.



Spanish version

## Resumen

**Introducción:** Los sistemas hidrológicos han sido severamente impactados por el aumento de eventos extremos como inundaciones y sequías, a su vez intensificados por el cambio climático. La cuenca del río Tuluá, ubicada en el Valle del Cauca – Colombia, enfrenta retos adicionales por la deforestación y limitada cobertura de monitoreo, complicando la gestión eficiente del recurso hídrico.

**Objetivos:** Desarrollar un modelo neuronal basado en redes LSTM para predecir el caudal del río Tuluá a partir de variables hidrometeorológicas, para la anticipación de escenarios extremos y el apoyo en la toma de decisiones en la gestión hídrica regional.

**Metodología:** 6875 registros diarios para cinco variables (precipitación, temperatura, evaporación, humedad relativa y caudal) de siete estaciones, fueron consolidadas. Para completar valores faltantes se aplicó preprocesamiento de datos usando KNN y análisis de correlación de Pearson. Se configuró y se entrenó el modelo LSTM, el cual se evaluó para los subconjuntos de entrenamiento, prueba y validación mediante métricas usuales para redes neuronales.

**Resultados:** Una alta capacidad predictiva con coeficientes de correlación  $R$  (0,75 – 0,98) mostró el modelo neuronal. Se evidenció que el modelo captura efectivamente las dependencias temporales y patrones no lineales, además de identificar la influencia del fenómeno ENSO en la dinámica hidrológica local.

**Conclusiones:** El modelo LSTM es una herramienta efectiva en la estimación de caudales para la cuenca en estudio, lo cual facilita tanto la gestión hídrica como la prevención de los riesgos que se asocian a la variabilidad climático. Futuros desarrollos pueden potenciar su precisión y utilidad, si consideran la inclusión de índices climático globales.

**Palabras clave:** Inteligencia artificial, cambio climático, evapotranspiración, inundación, precipitación, gestión de recursos, cuenca fluvial, modelo de simulación, desarrollo sostenible, temperatura.



### Why was it done?

A reliable predictive model, based on artificial neural network modeling, was developed to estimate the flow rate of the Tuluá River, whose basin is vulnerable to hydrological variability and climate change, and also suffers from limited data and environmental pressures. Its objective is to improve the anticipation of extreme events and strengthen regional water management through the use of artificial intelligence techniques.

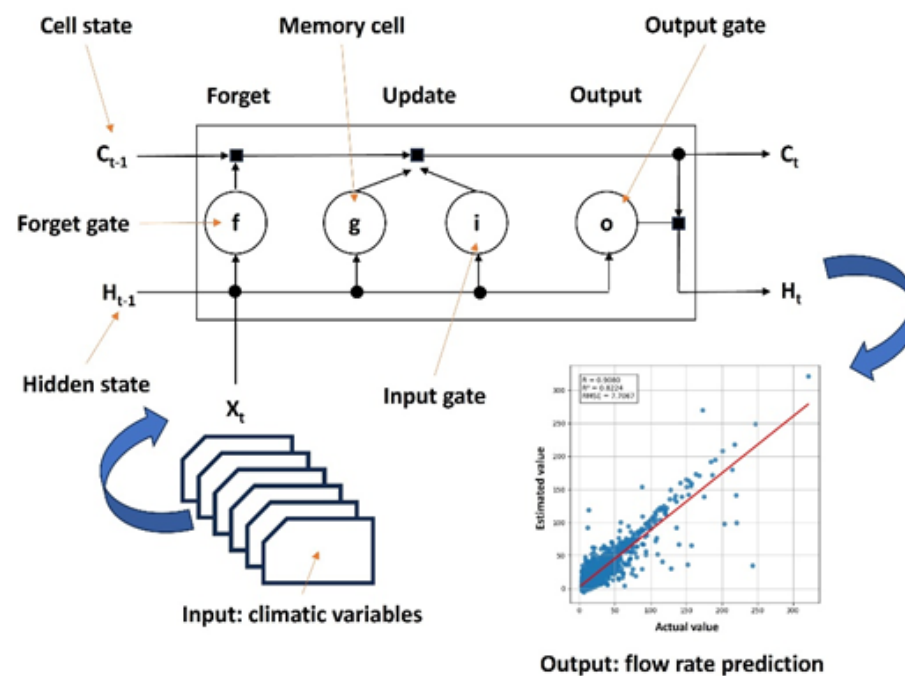
### What were the most relevant results?

The artificial neural model showed high accuracy in estimating flow rate according to the evaluation metrics ( $R = 0.75 - 0.98$ ), which demonstrates the capture of non-linear patterns and the complex temporal dependencies of hydroclimatic data, confirming the significant influence of the ENSO phenomenon on flow rate dynamics.

### What do they contribute?

A robust and adaptable tool for hydrological prediction in areas with limited data and high climate variability.

## Graphical Abstract



## Introduction

A significant increase, globally, in the frequency and impact of extreme events such as floods and droughts has been caused by hydrological variability, which is in turn intensified by climate change and human activities (1, 2). Approximately 48% of natural disasters recorded during the period 1995–2015 were linked to hydrometeorological phenomena (3, 4), affecting millions of people and causing significant economic losses (5). Population growth and unplanned urban expansion, which put pressure on global water resources, exacerbate this situation (6, 7).

The hydrological systems of Latin America have suffered severe consequences due to the increase in average temperature, glacial retreat, and alteration of precipitation patterns (8, 9). Likewise, phenomena such as El Niño and La Niña affect water availability, in the case of Colombia, producing significant fluctuations in river flows (10). According to Ávila et al., 2019 (11), in some river beds reductions of more than 75% can occur due to these variations, thus demonstrating that water systems are vulnerable to climatic and anthropogenic changes (12, 13).

The Valle del Cauca is characterized as a region of high hydroclimatic diversity (14, 15); however, its basins face increasing pressure, further exacerbated by deforestation, land-use change, and limited monitoring station coverage. A representative example where the lack of historical data and the scarcity of time-series analysis hinder efficient water resource management is the Tuluá River basin.

This area, of vital importance to the center of the Valle del Cauca department, with characteristics of Andean and sub-Andean Forest, provides water for agriculture, consumption and livestock, and currently faces challenges related to water quality due to pollution and deforestation, which leads to focusing interest on it from an environmental management that preserves its regulatory and ecological function.

This gives rise to the need to use advanced analytical tools that allow the identification of nonlinear patterns in hydrological dynamics. Recurrent neural networks (RNNs), and in particular Long Short-Term Memory (LSTM) networks, are specialized for training time-series data (16). In the case of Colombia, they have been used to predict water levels in the Atrato River (17), specifically to predict flow rates at the Calamar station on the Magdalena River (18). Performance was evaluated using an error statistic, and the results showed that the model developed does not adequately capture the actual hydrological dynamics of the basin under study. Therefore, there is a limited study of the use of these models in Colombia, and particularly in basins of the Valle del Cauca.

Based on these premises, the development of an artificial neural network model using Long Short-Term Memory (LSTM) networks is proposed for estimating the flow rate of the Tuluá River using hydrometeorological variables recorded by local stations. The proposed tool will allow for the identification of subtle patterns, the anticipation of extreme scenarios, and the generation of useful inputs for decision-making in public environmental management. In this way, the integration of artificial intelligence with hydrological analysis (19, 20) aims to close information gaps, strengthen water resilience, and improve the capacity to respond to extreme events in this region of southwestern Colombia.

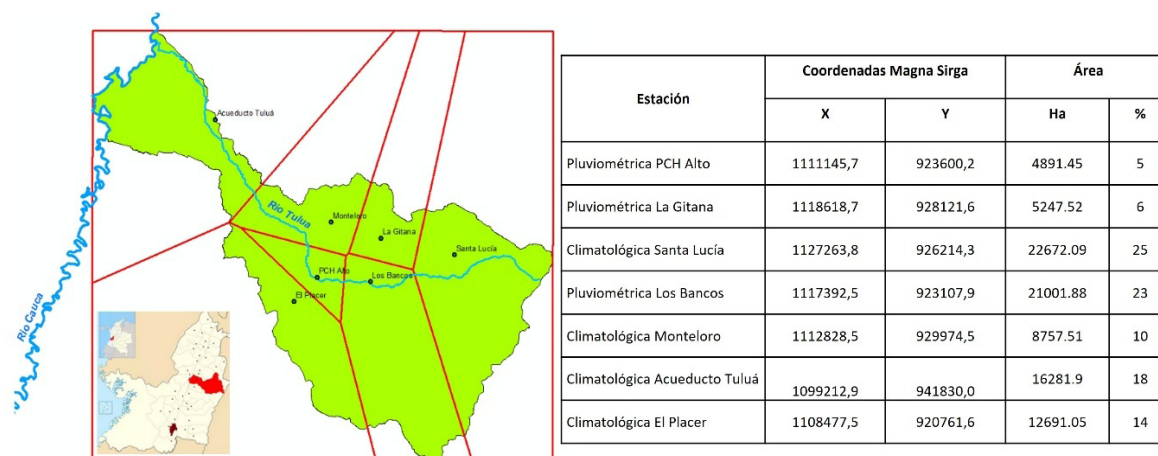
## Materials and methods

### Study area and data

The target area of this study is the Tuluá River watershed (Figure 1). It originates in Las Herosas National Park at an altitude of 4100 meters above sea level, and its main channel flows for 72 km through the municipality of the same name towards the Cauca River at 900 meters above sea level, receiving tributaries such as the Cofre, Loro, San Antonio, San Marcos, and Nogales rivers, among others (35). The watershed covers an area of 91485–91495 hectares and includes Andean and sub-Andean forests with significant biodiversity. The Andean Forest ecosystems contribute to water regulation, while the sub-Andean ecosystems support extensive livestock farming. Land use determined by human activity consists of agriculture and livestock farming, as well as population settlements.

Its importance lies in its role as a source of water for domestic use, agricultural irrigation, livestock farming, and industrial activities. Likewise, for the municipality of Tuluá, it acts as a spatial and environmental connector, integrating rural and urban areas (36). Currently, it faces environmental problems due to pollution in the lower reaches caused by sewage discharges and alterations to the riverbed, and deforestation due to the replacement of native forests with pastures and crops (35).

For this basin (Figure 1), the Corporación Autónoma Regional del Valle del Cauca – CVC operates seven hydroclimatic monitoring stations (Acueducto Tuluá – AT, Monteloro – ML, El Placer – EP, Santa Lucía – SL, La Gitana – LG, Los Bancos – LB and Mateguadua – MG), from where daily records were collected, and whose information in all stations presented missing data for all stations and for all variables.



**Figure 1.** Study area in the Tuluá River basin and location of the hydroclimatic stations.

Regarding the variables, the missing data corresponded to 5,6–11,1% for daily evaporation, 7,3% for flow rate, 8,6–61,4% for relative humidity, 0,5–3,9% for precipitation, and 7,1–58,0% for temperature. Monteloro was identified as the most critical station in this situation, with the highest missing data values for most of the variables considered.

The data comprise five main hydroclimatic variables (precipitation, temperature, evaporation, relative humidity, and flow rate), Table 1, totaling 6785 daily records after the data preprocessing procedure, which involved structuring the database, cleaning, and completing it with linear and nonlinear correlations. The information was organized into a file with 16 columns, in which the output variable is the flow rate of the Tuluá River, while the others correspond to input variables, and was used both for statistical analysis and for training the predictive model.

## Preliminary statistical analysis

### Descriptive statistics

Measures of central tendency (mean  $\bar{M}_e$ , median  $\mu$ , mode  $M_o$ ), dispersion (variance  $\sigma^2$ , standard deviation  $\sigma$ ), and quartile ranges (Q1, Q2), as defined in Abubaker, 2024 (21), were calculated for each variable and selected station (Table 1, Figure 2). This allowed for the identification of characteristic patterns, including greater variability at the Acueducto Tuluá station. Regarding flow rate, records were concentrated at the Mateguadua station, strategically located between the water production and consumption zones. For precipitation, the Acueducto Tuluá, El Placer, La Gitana, Los Bancos, Montero, and Santa Lucía stations were considered. For evaporation, the Acueducto Tuluá, El Placer, Monteloro, and Santa Lucía stations were included. Finally, for temperature, the Acueducto Tuluá, El Placer, Monteloro, and Santa Lucía stations were used.

### Test of normality

The Shapiro-Wilk normality test was applied to each variable at each station (Table 2), with W statistic and P value; for the test,  $W \sim 1,0$  indicates normal distribution and  $P < 0,05$  suggests rejection of normality (22).

### Trend analysis

The non-parametric Mann-Kendall test was applied (Table 2), in order to detect significant trends in the time series and using Kendall's Tau coefficient  $\tau$  and the P value; for the test,  $-1 < \tau < 1$  indicates direction and strength of the trend while,  $P < 0,05$  indicates a statistically significant trend (23).

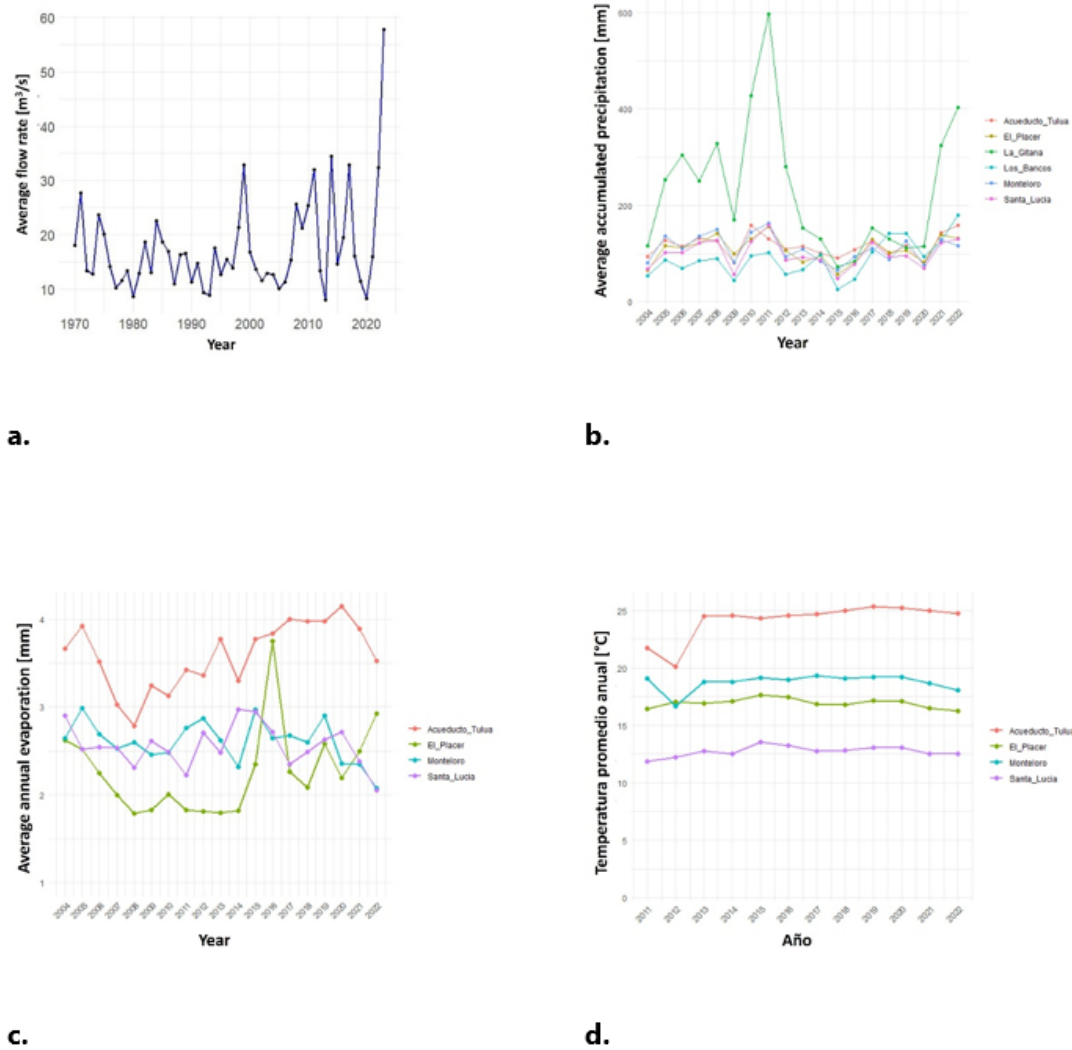
**Table 1.** Summary of descriptive statistics of hydroclimatic variables by station.

Flow rate									
Station	Minimum	Q1	Me	$\mu$	Mo	Q3	Maximum	$\sigma$	$\sigma^2$
	m <sup>3</sup> /s	m <sup>3</sup> /s	m <sup>3</sup> /s	m <sup>3</sup> /s	m <sup>3</sup> /s	m <sup>3</sup> /s	m <sup>3</sup> /s	m <sup>3</sup> /s	
MG	-	0,93	8,3	13,3	8,0	20,9	321,16	15,06	226,83
Precipitation									
Station	Minimum	Q1	Me	$\mu$	Mo	Q3	Maximum	$\sigma$	$\sigma^2$
	mm	mm	mm	mm	mm	mm	mm	mm	
AT	4,0	64,0	103,0	116,2	113,0	157,0	408,0	69,51	4831,54
EP	0,0	55,0	98,0	109,5	42,0	148,0	323,0	70,81	5014,13
LG	0,0	62,0	150,0	235,7	8,0	336,0	1643,0	252,79	252,79
LB	0,0	34,0	66,0	88,88	42,0	119,0	550,0	82,25	42,0
ML	0,0	54,0	101,3	113,3	44,0	156,6	344,0	72,28	5224,15
SL	0,0	44,0	89,0	100,3	106,0	139,0	323,0	70,46	4965,2
Evaporation									
Station	Minimum	Q1	Me	$\mu$	Mo	Q3	Maximum	$\sigma$	$\sigma^2$
	mm	mm	mm	mm	mm	mm	mm	mm	
AT	2,2	3,2	3,6	3,58	3,4	4,0	5,5	0,597	0,357
EP	1,0	1,7	2,1	2,22	1,7	2,5	8,1	0,790	0,624
ML	1,4	2,3	2,6	2,61	2,6	2,9	5,5	0,519	0,269
SL	1,3	2,1	2,5	2,54	2,5	2,9	4,4	0,607	0,368
Temperature									
Station	Minimum	Q1	Me	$\mu$	Mo	Q3	Maximum	$\sigma$	$\sigma^2$
	°C	°C	°C	°C	°C	°C	°C	°C	
AT	19,3	24,3	24,7	24,06	24,5	25,05	25,9	1,741	3,0
EP	15,6	16,4	16,9	16,95	17,2	17,45	19,0	0,722	0,5
ML	15,6	18,45	18,9	18,76	18,6	19,4	24,1	1,084	1,2
SL	11,0	12,2	12,8	12,75	13,3	13,3	14,5	0,736	1,084

## Development and evaluation of the predictive model

### Data preparation

The following data preprocessing was performed: 1.- replacement of missing values using the k-Nearest Neighbors (KNN) method, in order to ensure temporal continuity (24), 2.- identification of significant relationships between variables by Pearson correlation analysis (25), and 3.- shaping of the dataset into four subsets (55% training, 20% testing, 25% validation and 100% simulation) for training and evaluation of model performance.



**Figure 2.** Behavior of the annual average of climatic variables: a.- Flow rate, b.- Precipitation, c.- Evaporation, and d.- Temperature.

### Artificial neural model architecture

A specialized variant of recurrent neural networks, specifically a Long Short-Term Memory (LSTM) type, described in Ayadi et al., 2024 (26), was implemented for the model. This variant was selected for its ability to capture the complex temporal dependencies in hydrological series. The structure included: 1.- an input layer for precipitation, temperature, evaporation, and flow variables; 2.- hidden LSTM layers with gate mechanisms (input, forget, and output) for controlling information flow and retaining long-term dependencies; and 3.- a dense output layer with a linear activation function to estimate flow rate. The model was trained using the Adam optimizer and the mean squared error (MSE) loss function.

**Table 2.** Results of the Shapiro-Wilk normality test and the Mann-Kendall test

Variable	Station	Shapiro-Wilk test		Mann-Kendall test	
		W	P value	$\tau$	P value
Flow rate	MG	0,62138	< 2,2E-16	-0,050570093	1,40E-04
	AT	0,48556	< 2,2E-16	0,020568889	1,71E-01
	EP	0,57559	< 2,2E-16	-0,004470214	7,63E-01
Precipitation	SL	0,58548	< 2,2E-16	-0,026856607	7,05E-02
	ML	0,55364	< 2,2E-16	0,018138796	2,21E-01
	LG	0,63872	< 2,2E-16	-0,059117045	4,23E-05
	LB	0,51017	< 2,2E-16	-0,037041675	1,38E-02
Evaporation	AT	0,94273	< 2,2E-16	0,082426094	0,00E+00
	EP	0,81697	< 2,2E-16	0,051181767	1,39E-04
	ML	0,96758	< 2,2E-16	-0,101046145	5,63E-14
	SL	0,93201	< 2,2E-16	-0,022796048	8,91E-02
Temperature	AT	0,78094	< 2,2E-16	0,390442461	0,00E+00
	ML	0,74612	< 2,2E-16	0,082547799	0,00E+00
	EP	0,99503	< 2,2E-16	-0,039089784	3,43E-03
Relative humidity	AT	0,91579	< 2,2E-16	-0,317095846	0,00E+00
	ML	0,99341	< 2,2E-16	-0,058222272	1,28E-05
	EP	0,92717	< 2,2E-16	0,139304444	0,00E+00

### Evaluation metrics

The model's performance was evaluated using the following metrics (Equations (1)-(3)): 1.- root mean square error (RMSE), which measures the average magnitude of the error, penalizing larger errors more, 2.- the correlation coefficient (R), which indicates the strength and direction of the linear relationship in this case between the actual value and the predicted value, and 3.- the coefficient of determination ( $R^2$ ), which indicates the proportion of the variability between the dependent variable that is explained by the independent variables, corresponds to the square of R, and essentially indicates the goodness of fit of the model.

$$RMSE = \sqrt{\frac{1}{N} \sum (X_{predicho} - X_{observado})^2} \quad \text{Ec. (1)}$$

$$R = \sqrt{1 - \frac{\sigma_y^2}{\sigma_r^2}} \quad \text{Ec. (2)}$$

$$R^2 = 1 - \frac{\sigma_y^2}{\sigma_r^2} \quad \text{Ec. (3)}$$

### Computational tools

Using the Python programming language (37), a computational code is developed whose architecture implements the LSTM neural network for regression purposes (Figure 3), and the script is executed for modeling on the Google Colab platform, using libraries of the aforementioned language that allowed the implementation and training of the LSTM network architecture (NumPy, Pandas, Scikit-learn, TensorFlow and Keras).

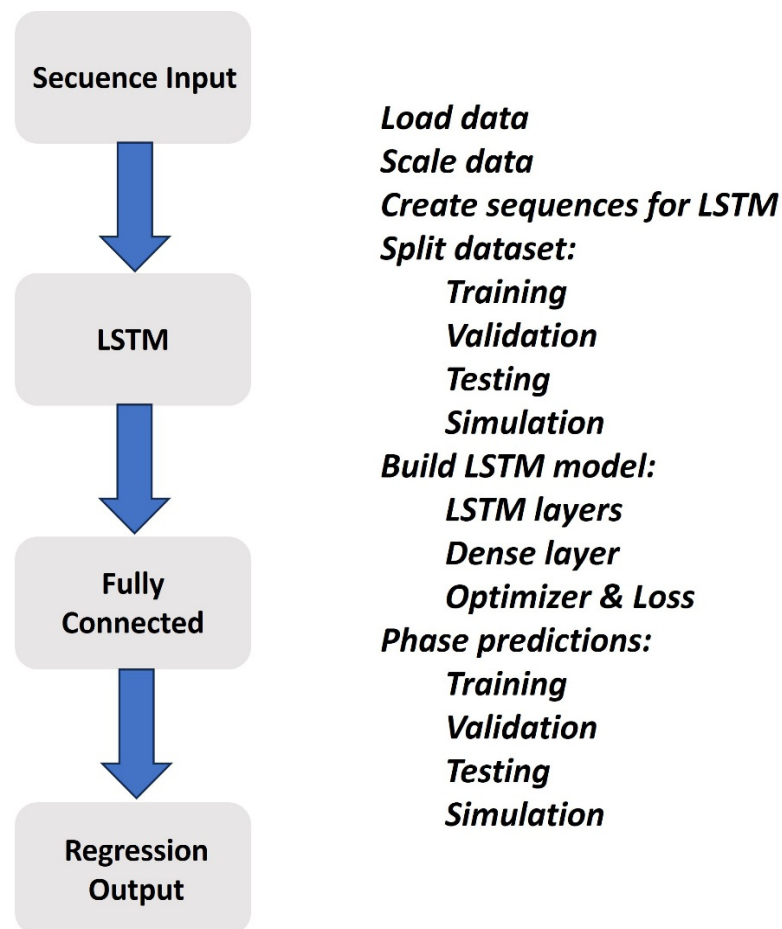
## Results

### Statistical analysis of hydroclimatic variables

The statistical analysis of the hydroclimatic variables revealed a symmetrical distribution in the analyzed stations, given that the mean and median values showed a high degree of consistency. The greatest thermal variability was observed at the Acueducto Tuluá station in the range of 19.3-25.05 °C in the third quartile, which positions it as the warmest area of the basin. Conversely, the lowest temperatures, ranging between 11.0-13.3 °C, were recorded at the Santa Lucía station. The highest variance and standard deviation were observed at the Acueducto Tuluá station, 3.0°C and 1.741°C respectively, indicating a greater thermal dispersion of this station in relation to the others.

### Test of normality

Although several  $W$  se acercaron a 1, coefficients approached 1, the associated  $P$  values were less than 0,05. In particular, the evaporation and relative humidity variables showed high  $W$  coefficients; however, with  $P < 2,2E^{-16}$ , they confirm that normality is not met for these stations.



**Figure 3.** Architecture for regression (left) and pseudocode (right) for training the LSTM network.

Source: own elaboration in accordance with the methodology and application examples described in (37).

## Trend analysis

- Based on the Mann-Kendall coefficient, trend analysis revealed differentiated behaviors in the hydroclimatic variables, highlighting the following:
- The flow rate at the Mateguadua station showed a significant decrease, with  $\tau = 0,051$  and  $P = 0,00014$ .
- Evaporation showed combined trends; on the one hand, at the El Placer station there was a significant increase with  $\tau = 0,051$ , while at the Monteloro station there was a significant decrease with  $\tau = -0,051$  and  $P \sim 0$ .
- Rainfall at the La Gitana and Los Bancos stations showed a slight decreasing trend.
- Relative humidity decreased at the Monteloro station and increased at the El Placer station, an expected behavior given the inverse relationship between temperature and humidity.
- Regarding temperature, significant increases were observed at the Acuedcto Tuluá and Monteloro stations, and a slight decrease was observed at the El Placer station.

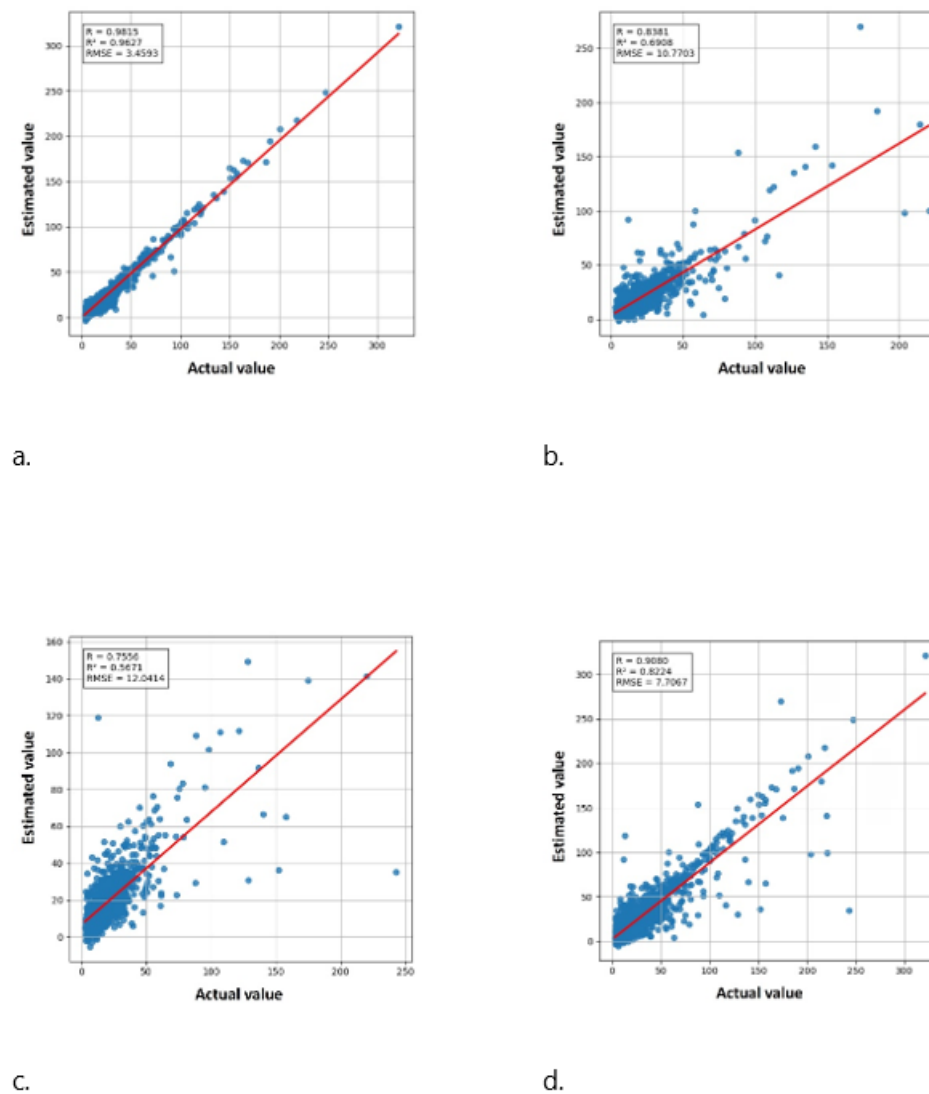
## Performance of the predictive model based on the LSTM network

The graphical comparison between the actual and estimated values at each stage (Figure 4) demonstrates the model's good fit and its usefulness for practical applications in water management. In this regard, the evaluation metrics of the LSTM neural network confirm the high performance graphically displayed in predicting the flow rate of the Tuluá River at the different stages of the modeling process.

As shown in Figure 4, the model fit is excellent during the training phase, with 96,27% of the flow rate variability explained by the input variables. During the testing phase, the model maintains good predictive capacity outside the training set, explaining 69,08% of the variability. During the validation phase, performance remains adequate, with 56,71% of the variability explained by the input variables. Finally, in the overall simulation with all the data, the model's generalizability and accuracy are confirmed, achieving 82,24% of the variability explained.

## Discussion

In general, this architecture demonstrates its ability to handle complex and nonlinear temporal dependencies, allowing it to accurately capture the seasonal and lagged dynamics present in hydrological variables such as precipitation, temperature, and particularly streamflow, which is the objective of the prediction. In this regard, Waqas & Humphries, 2024 (27), highlight the effectiveness of LSTM networks in modeling nonlinear and time-varying hydrological systems, overcoming the limitations of other networks in modeling long-term dependencies, attributed to the use of memory cells and gate mechanisms.



**Figure 4.** Graphical comparison between the actual values and the estimated values in each of the phases: a.- Training, b.- Testing, c.- Validation, and d.- Simulation.

The initial exploratory analysis confirmed the presence of atypical variations and scattered records in the studied variables, suggesting relevant climatic alterations in the region, as inferred in other studies (28, 29). Applying Pearson's correlation coefficient in combination with the KNN algorithm for completing missing data strengthened the database, ensuring the continuity of the time series and increasing the number of daily records per variable (in the most critical case of missing data, temperature in ML, from 4293 to 6785), which was fundamental for effective model training.

The Shapiro-Wilk normality test indicated that the time series did not meet the normality assumption (see Table 2), justifying the choice of a non-parametric technique for subsequent statistical analysis. This was complemented by trend identification using the Mann-Kendall coefficient, which revealed significant patterns in the study variables, such as decreases in flow rate in specific areas and mixed patterns regarding evaporation and relative humidity.

The above statistical findings highlight the need for climate monitoring segmented by sub-zones of the basin due to the presence of significant heterogeneities in the hydrological response.

Finally, the LSTM network showed robust performance with correlation coefficients greater than 0.75 in all stages, and an overall value of 0,9080 in the simulation with all the data, thus providing a reliable level of accuracy that confirms its potential for hydrological flow prediction and its use as a tool in regional water management. De La Fuente et al., 2024 (30), highlight the ability of LSTM-based modeling to overcome traditional methods of modeling based on grouped processes both temporally and spatially, for flow rate prediction; however, they recognize the structural complexity of the network in question, which includes control operations and sequential data processing, as well as the difficulties in interpreting the internal processes and weightings of the model. Therefore, the above reaffirms the approach followed for the work carried out in the Tuluá River basin; likewise, the performance obtained is comparable with that reported in previous studies, as summarized in Table 3.

Table 3 presents the performance results (R correlation coefficient) reported in previous studies and in the present research. It includes different neural network models applied to various watersheds, the phases evaluated, and the metrics obtained. The last row corresponds to the present study, showing evaluation metrics by phase for the LSTM model applied to the Tuluá River basin, demonstrating its ability to capture hydrological dynamics and its competitiveness compared to other models reported in the literature.

**Table 3.** Comparative evaluation of LSTM models for flow rate prediction: previous studies and present work.

Research	Model	Basin / Region	Phases evaluated	Metrics (R)	Observations
Li et al., 2022 (31)	CNN-LSTM	Río Hun, China	Training / Validation	All year round: 0.894 / 0.886; Rainy season: 0.888 / 0.884; Dry season: 0.799 / 0.721	Hybrid model, evaluation by periods
Wilbran et al., 2023 (32)	Hybrid Multiscale Time Models-LSTM	> 500 basins from the CAMEL-US database	It does not specify.	Average slope: 0.68; Low precipitation frequency: 0.62; Saturated hydraulic conductivity: 0.63	Assessment according to basin characteristics
Sahoo et al., 2019 (33)	LSTM / Conventional RNN	Mahanadi River, India	Training	LSTM: 0,971; RNN: 0,967	Low flow rate basins

Girihagama et al., 2022 <a href="#">(34)</a>	LSTM / LSTM + attention mechanisms	10 sub-basins of the Ottawa River, Canada	Training	LSTM: 0,992; LSTM + attention mechanisms: 0,977	Comparison between conventional LSTM and LSTM with attention mechanisms
<b>This research</b>	Conventional RNN	Tuluá River, Colombia	Training / Testing / Validation / General Simulation	Training: 0.9815; Testing: 0.8381; Validation: 0.7556; Overall simulation: 0.9080	Phased evaluation; capturing seasonal and delayed dynamics

## Conclusions

The artificial neural model based on the LSTM network proved to be a robust and reliable tool for modeling and predicting flow behavior in the Tuluá River basin, showing high levels of fit and generalization.

The use of non-parametric statistical methods and deep learning techniques enabled us to approach the complexity and non-linearity inherent in hydroclimatic time series.

The patterns of variability in flow rate, evaporation, and relative humidity reflect the influence of global climatic phenomena such as ENSO, indicating that their incorporation into future models could improve predictive accuracy.

The model's performance in different phases and seasons demonstrates its usefulness as a tool for the sustainable management of water resources in the Tuluá River basin, supporting planning and adaptation strategies in the face of variability and climate change.

### CreditT Authorship Contribution Statement

Conceptualization – Ideas, Data Curation, Formal Analysis, Research, Methodology, and Resources: León Blandón. Project Management, Software, and Supervision: Luis González. Validation, Drafting – Original Draft – Elaboration, and Drafting – Review and Editing – Elaboration: León Blandón, Luis González.

Funding: not declared. Conflict of interest: not declared. Ethical considerations: not declared.

## References

1. Pizzorni M, Innocenti A, Tollin N. Droughts and floods in a changing climate and implications for multi hazard urban planning: A review. *City Environ Interact*. 2024;24:100169.

<https://doi.org/10.1016/j.cacint.2024.100169>

2. Furtak K, Wolińska A. The impact of extreme weather events as a consequence of climate change on the soil moisture and on the quality of the soil environment and agriculture - A review. *Catena*. 2023;231:107378

<https://doi.org/10.1016/j.catena.2023.107378>

3. Chaudhary MT, Piracha A. Natural Disasters-Origins, Impacts, Management. *Encyclopedia*. 2021;1:1101 31.

<https://doi.org/10.3390/encyclopedia1040084>

4. Skilodimou HD, Bathrellos GD. Natural and technological hazards in urban areas: assessment, planning and solutions. *Sustainability*. 2021;13(15):8301. doi: <https://doi.org/10.3390/su13158301>

5. Fischer G. El aumento de las inundaciones generado por el cambio climático afectará nuestros cultivos. *Rev Fac Nac Agron Medellín*. 2021;74(3):9619 20. [http://www.scielo.org.co/scielo.php?script=sci\\_arttext&pid=S0304-28472021000309619&lng=en&tlng=es](http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0304-28472021000309619&lng=en&tlng=es)

6. Molina JE, Arboit ME, Maglione DS, Sedevich A, Mutani G. Estudio de expansión urbana, crecimiento poblacional, consumos energéticos e índices de vegetación en el Área Metropolitana de Mendoza. *AREA*. 2020;26(1):1 21. <https://dialnet.unirioja.es/servlet/articulo?codigo=8213842>

7. Zhungo Carpio JP, Baquedano Rivera PR. Impacto Ambiental de la Expansión Urbana no Planificada en los Recursos hídricos del Cantón Santo Domingo: Un Análisis crítico. *Rev Científica Salud Desarro Humano*. 2025;6(1):1600 17.

<https://doi.org/10.61368/r.s.d.h.v6i1.542>

8. Reyer C, Adams S, Albrecht T, et al. Climate change impacts in Latin America and the Caribbean and their implications for development. *Reg Environ Change*. 2017;17:1601 21.

<https://doi.org/10.1007/s10113-015-0854-6>

9. Restrepo JC, Higgins A, Escobar J, Ospino S, Hoyos N. Contribution of low frequency climatic-oceanic oscillations to streamflow variability in small, coastal rivers of the Sierra Nevada de Santa Marta (Colombia). *Hydrol Earth Syst Sci*. 2019;23:2379 400.

<https://doi.org/10.5194/hess-23-2379-2019>

10. Rodriguez Espinosa CA, Valencia CF, Ramirez CG, Valencia CD. Dynamic effect of climate change on flood damage cost in the Andean region of Colombia using an ARDL ECM model and climate change projections. *Sustainable Cities Soc*. 2025;125:106309.

<https://doi.org/10.1016/j.scs.2025.106309>

11. Ávila Á, Guerrero FC, Escobar YC, Justino F. Recent precipitation trends and floods in the Colombian Andes. *Water*. 2019;11(2):379.

<https://doi.org/10.3390/w11020379>

12. Flores P, Navarro Y. Efectos del cambio climático en los recursos hídricos: un análisis de la experiencia en Venezuela. *Novum Scientiarum*. 2016;2(4):45 55.

13. Amparo Salcedo M, Pérez Gimeno A, Navarro Pedreño J. Water security under climate change: challenges and solutions across 43 countries. *Water*. 2025;17(5):633.  
<https://doi.org/10.3390/w17050633>
14. Carvajal Y, Enciso A, Sandoval MC. Hydrological analysis of historical floods in the upper valley of Cauca river. *Inycomp* [Internet]. 2016 Jan 19 [cited 2025 Aug 19];18(1):46-57.  
<https://doi.org/10.25100/iyc.v18i1.2176>
15. González López N, Carvajal Escobar Y. Characterization of hydrological drought in the Cauca river high valley. *Tecnol Cienc Agua*. 2020;11(1):235-65.  
<https://doi.org/10.24850/j-tyca-2020-01-06>
16. Yunita A, Pratama MI, Almuzakki MZ, Ramadhan H, Akashah PAE, Mansur ABF, et al. Performance analysis of neural network architectures for time series forecasting: a comparative study of RNN, LSTM, GRU, and hybrid models. *MethodsX*. 2025;15:103462.  
<https://doi.org/10.1016/j.mex.2025.103462>
17. Renteria Mena JB, Plaza D, Giraldo E. Multivariate hydrological modeling based on long short term memory networks for water level forecasting. *Information*. 2024;15(6):358.  
<https://doi.org/10.3390/info15060358>
18. Montoya Zuluaga LM. Predicción de caudales en el sistema Magdalena Cauca utilizando modelos estadísticos y redes neuronales recurrentes [thesis]. Medellín: Universidad de Antioquia; 2023 <https://bibliotecadigital.udea.edu.co/handle/10495/37887>
19. Morante Carballo F, Arcentales Rosado M, Caicedo Potosí J, Carrión Mero P. Artificial intelligence applications in hydrological studies and ecological restoration of watersheds: a systematic review. *Watershed Ecol Environ*. 2025;7:230-48.  
<https://doi.org/10.1016/j.wsee.2025.05.004>
20. Volpi E, Kim JS, Jain S, Shrestha S. Editorial: artificial intelligence in hydrology. *Hydrol Res*. 2023;54(6):iii-iv. <https://doi.org/10.2166/nh.2023.102>
21. Abubaker Blbas HT. Descriptive statistics [Internet]. In: *Recent Advances in Biostatistics*. IntechOpen; 2024 <https://doi.org/10.5772/intechopen.1002179>
22. Das KR, Imon AHM. A brief review of tests for normality. *Am J Theor Appl Stat*. 2016;5(1):5-12.  
<https://doi.org/10.11648/j.ajtas.20160501.12>
23. Liu S, Xie Y, Fang H, Du H, Xu P. Trend test for hydrological and climatic time series considering the interaction of trend and autocorrelations. *Water*. 2022;14(19):3006.  
<https://doi.org/10.3390/w14193006>
24. Seoane Santos M, Henriques Abreu P, Wilk S, Santos J. How distance metrics influence missing data imputation with k-nearest neighbours. *Pattern Recognit Lett*. 2020;136:111-9.  
<https://doi.org/10.1016/j.patrec.2020.05.032>
25. Hernández Lalinde JD, Espinosa Castro F, Rodríguez JE, Chacón Rangel JG, Toloza Sierra CA, Arenas Torrado MK, et al. Sobre el uso adecuado del coeficiente de correlación de Pearson: definición, propiedades y suposiciones. *Arch Venez Farmacol Ter*. 2018;37(5).  
[https://www.revistaavft.com/images/revistas/2018/avft\\_5\\_2018/20sobre\\_uso\\_adecuado\\_coeficiente.pdf](https://www.revistaavft.com/images/revistas/2018/avft_5_2018/20sobre_uso_adecuado_coeficiente.pdf)



26. Ayadi A, Truong TMD, Frydman C, Wemmert C. Explaining LSTM neural networks: a DEVS-based approach. In: Proceedings of the 6th European Modeling & Simulation Symposium, 21st International Multidisciplinary Modeling & Simulation Multiconference. 2024.  
<https://doi.org/10.46354/i3m.2024.emss.035>
27. Waqas M, Wannasingha Humphries U. A critical review of RNN and LSTM variants in hydrological time series predictions. *MethodsX*. 2024;13:102946.  
<https://doi.org/10.1016/j.mex.2024.102946>
28. Arrieta Castro M, Donado Rodríguez A, Acuña GJ, Canales FA, Teegavarapu RSV, Kaźmierczak B. Analysis of streamflow variability and trends in the Meta River, Colombia. *Water*. 2020;12(5):1451.  
<https://doi.org/10.3390/w12051451>
29. Umar DA, Ramli MF, Aris AZ, Jamil NR, Aderemi AA. Evidence of climate variability from rainfall and temperature fluctuations in semi arid region of the tropics. *Atmos Res*. 2019;224:52-64. <https://doi.org/10.1016/j.atmosres.2019.03.023>
30. De la Fuente LA, Ehsani MR, Gupta HV, Condon LE. Toward interpretable LSTM-based modeling of hydrological systems. *Hydrol Earth Syst Sci*. 2024;28:945-71. doi:10.5194/hess-28-945-2024  
<https://doi.org/10.5194/hess-28-945-2024>
31. Li X, Xu W, Ren M, Jiang Y, Fu G. Hybrid CNN-LSTM models for river flow prediction. *Water Supply*. 2022;22(5):4902-11. <https://doi.org/10.2166/ws.2022.170>
32. Wilbrand K, Taormina R, ten Veldhuis M-C, Visser M, Hrachowitz M, Nuttall J, et al. Predicting streamflow with LSTM networks using global datasets. *Front Water*. 2023;5:1166124. <https://doi.org/10.3389/frwa.2023.1166124>
33. Sahoo BB, Jha R, Singh A, et al. Long short-term memory (LSTM) recurrent neural network for low-flow hydrological time series forecasting. *Acta Geophys*. 2019;67:1471-81. <https://doi.org/10.1007/s11600-019-00330-1>
34. Giriagama L, Khaliq MN, Lamontagne P, et al. Streamflow modelling and forecasting for Canadian watersheds using LSTM networks with attention mechanism. *Neural Comput Appl*. 2022;34:19995-20015. <https://doi.org/10.1007/s00521-022-07523-8>
35. Bustamante CA, Naranjo JF, Ahumada JM. Determinación de la calidad ecológica del río Tuluá Valle del Cauca. *Mutis*. 2014;4(2):35-43. <https://doi.org/10.21789/22561498.955>
36. López, JP. El río como hilo conector: caso Tuluá, Valle del Cauca [Internet]. *Sociedad-Espacio-Naturaleza*; 2022. <https://sociedadespacionaturaleza.wordpress.com/2021/09/03/el-rio-como-hilo-conector-caso-tuluá-valle-del-cauca/>
37. TechwithJulles. Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) - Creating an LSTM Model in Python Using TensorFlow and Keras [Internet]. 2024 Oct 12 [citado 2025 Dic 05 ]; Disponible en: <https://medium.com/@techwithjulles/recurrent-neural-networks-rnns-and-long-short-term-memory-lstm-creating-an-lstm-model-in-13c88b7736e2>