Use of nonlinear principal components of CHIRPS precipitation data and ocean-atmospheric variables for streamflow forecasting in an area of scarce data. Case study, Tocaría river basin - Orinoquia Colombiana

Sarria-Ospina, Jhon Derly[1]; Ocampo-Marulanda, Camilo[2,3]; Ceron-Aramburo, Lina Maria[1]; Canchala, Teresita[4]; Ferreira, Tiago Alessandro[2]

[1]Research Group TERRANARE, Faculty of Natural Sciences and Engineering, Fundación Universitaria de San Gil, Yopal, 850001, Colombia
[2]Programa de Pós-graduação em Biometria e Estatística Aplicada, Departamento de Estatística e Informática, Universidade Federal Rural de Pernambuco, Recife, 52171-900, Brasil
[3]Water Resources, Engineering and Soil Research Group (IREHISA), School of Natural Resources and Environmental Engineering, Universidad del Valle, 760032, Cali, Colombia
[4]Environmental Engineering Program, Faculty of Engineering, Universidad Mariana, Pasto, 520002, Colombia

Correspondence to: Sarria-Ospina, Jhon (jsarria.sgryopal@unisangil.edu.co)

Abstract. Accurate streamflow forecasting is essential for mitigating hydrological extremes and supporting sustainable water-resources management, particularly in data-scarce tropical catchments. This study proposes a hybrid forecasting framework that integrates Seasonal Autoregressive Integrated Moving Average models with exogenous inputs (SARIMAX), nonlinear principal components (NLPCs) derived from CHIRPS precipitation data, and large-scale ocean–atmosphere indices (macroclimatic variables, MVs). Four monthly models were developed and evaluated for the Tocaría River basin in the Colombian Orinoquía region: (1) a SARIMA $(4,0,4)(0,0,3)_{12}$ model; (2) SARIMAX with MVs; (3) SARIMAX with NLPCs; and (4) a hybrid SARIMAX combining both MVs and NLPCs. All models were assessed using both raw and log-transformed streamflow data and benchmarked against simple baseline approaches. The hybrid model fitted to log-transformed streamflow showed consistent skill in both validation and testing ($R^2 = 0.49$ and 0.55, respectively), and providing the most accurate reproduction of peak-flow conditions (lowest RMSE = 60.8 m³/s). Skill remained over longer lead times, for forecasts up to 24 months ahead, RMSE stayed below 50 m³/s and R^2 over 0.80 across the forecast horizon. These results underscore the effectiveness of integrating nonlinear precipitation patterns, macroclimatic variability, and temporal dependence to improve forecast accuracy under data-scarce conditions, particularly for peak flows and anomalously wet periods that are not adequately captured by the climatological benchmark. The proposed methodology

offers a transferable approach for operational forecasting in ungauged or sparsely monitored basins, contributing to early warning systems, drought preparedness, and adaptive water governance in vulnerable tropical regions.

Keywords: monthly forecasting, macroclimatic variables, NLPC, SARIMA, exogenous variables

1 INTRODUCTION

Sustainable management of water resources in tropical basins critically depends on the ability to accurately anticipate streamflow dynamics amid complex hydroclimatic variability. Streamflow integrates precipitation, evapotranspiration, infiltration, and climatic anomalies across multiple spatial and temporal scales, yet sparse hydrometeorological monitoring limits the characterization of these processes (Tootle et al., 2008; Yao et al., 2020).

Conventional statistical models such as Autoregressive Integrated Moving Average (ARIMA) and its seasonal extension (SARIMA) have been widely applied for streamflow prediction due to their capacity to model temporal dependencies and seasonality under stationarity assumptions (Sirisha et al., 2022; Valipour, 2015). However, their performance diminishes in capturing abrupt, nonlinear fluctuations characteristic of tropical catchments influenced by exogenous climate drivers (Moeeni and Bonakdari, 2017). Previous studies have shown that incorporating exogenous variables into forecasting models significantly enhances streamflow forecasting skill, particularly at seasonal to interannual timescales, by explicitly accounting for external climate forcing mechanisms (Wong et al., 2007).

In regions like the Colombian Orinoquía, the interplay of large-scale ocean–atmosphere phenomena; especially the El Niño–Southern Oscillation (ENSO); introduces pronounced nonstationarity and nonlinearity, challenging traditional time series forecasting methods (Chiew and McMahon, 2002; Pasquini et al., 2007; Yap and Musa, 2023). Models incorporating macroclimatic indices; such as sea surface temperature anomalies and ENSO-related indices; have demonstrated substantial improvements in streamflow forecasting accuracy for Andean and Pacific basins (Cárdenas et al., 2022; Mesa et al., 2001; Poveda et al., 2002; Velásquez et al., 2010). Nevertheless, similar studies remain scarce for the Orinoquía region, despite its hydrological sensitivity to both Pacific and Atlantic climatic influences (Builes-Jaramillo et al., 2022). The Tocaría River basin (Colombian Orinoquia), exemplifies this knowledge gap; it is characterized by limited hydrometeorological instrumentation and experiences seasonal extremes of drought and flooding that impact local agriculture and ecosystems (Corporinoquia and Corpoboyacá, 2015a).

Building on previous work that successfully employed Nonlinear Principal Component Analysis (NLPCA) combined with satellite-derived Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) precipitation data to impute missing monthly streamflow records in the Tocaría basin; demonstrating high accuracy across stations with substantial data gaps (Ocampo-Marulanda et al., 2025); this study advances the development of a parsimonious hybrid forecasting framework. The proposed model, SARIMAX, integrates a SARIMA component and incorporates NLPC derived from CHIRPS data alongside MVs representing large-scale atmospheric–oceanic drivers.

The research aims to provide an accurate monthly streamflow forecasting tool tailored to the Tocaría River's hydrological regime, supporting adaptive water resource management in a data-limited and climate-vulnerable tropical basin. By addressing key gaps in scientific knowledge and practical decision-making tools, the framework offers scalable potential for application in other similarly under-monitored tropical catchments subject to complex climatic forcing.

2 METHODOLOGY

This section outlines the methodological framework developed for monthly streamflow forecasting in poorly instrumented basins, using the Tocaría River (Colombian Orinoquia) as a case study. The approach integrates seasonal autoregressive modeling with exogenous predictors, nonlinear dimensionality reduction, and multivariate correlation analysis.

As illustrated in Fig. 1, the methodological workflow integrates input data preprocessing, SARIMA order selection, and the determination of the predictor pool. Subsequently, four SARIMA-based streamflow forecasting models were independently trained, validated, and tested. All models were evaluated using both raw and log-transformed streamflow data and benchmarked against simple climatology and persistence baselines. The statistical evaluation of the predicted values was assessed using the Root Mean Square Error (RMSE), the coefficient of correlation (R^2), the Akaike Information Criterion (AIC), and the Bayesian Information Criterion (BIC). Finally, for the best-performing model, forecast-error degradation across increasing prediction horizons was analyzed, and a comprehensive residual analysis was conducted.

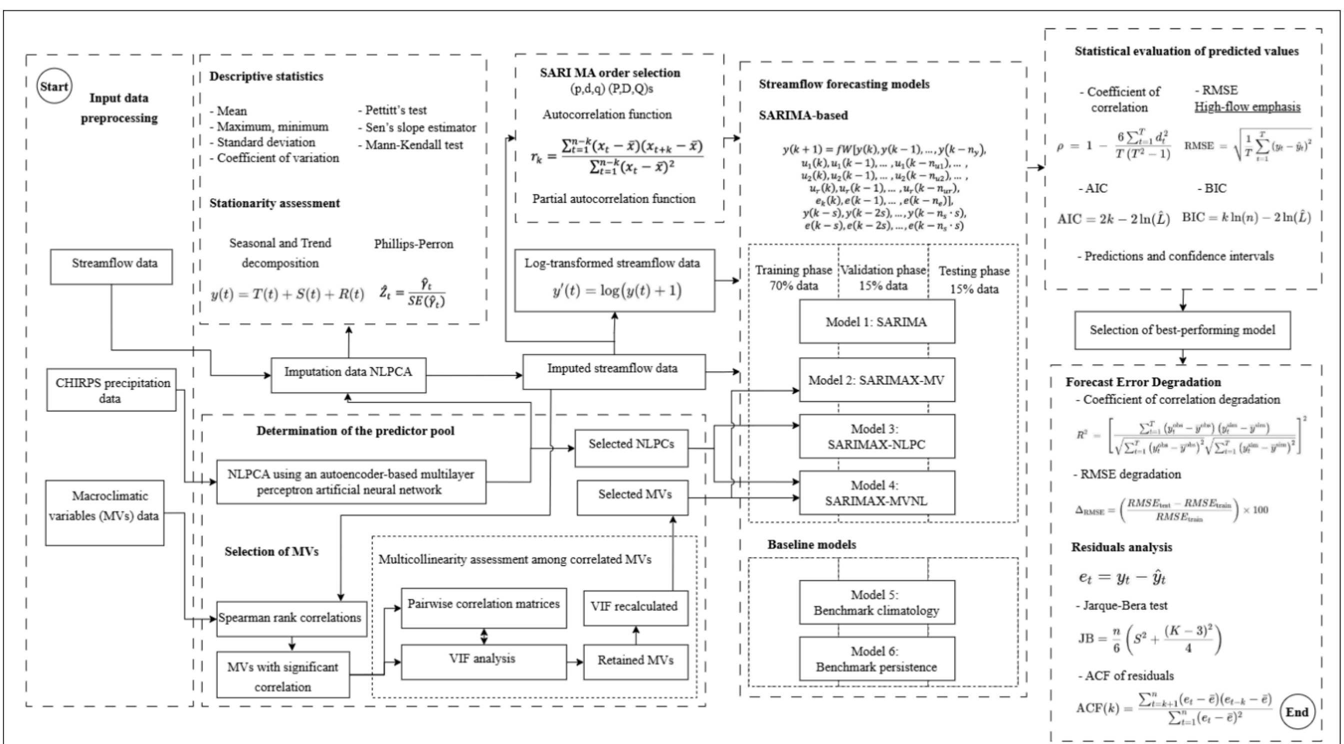

Figure 1. Schematic overview of the methodological workflow, including input data preprocessing, determination of the predictor pool, streamflow forecasting, and performance evaluation of the best-performing model.

2.1 Study area

The Tocaría River originates at 3,200 m above sea level in the Guevarrica hill and drains a watershed of 2,223 km² on the eastern flank of the eastern cordillera. The river extends over 127.5 km with an average longitudinal slope of 5%. The basin is part of the Cravo Sur system, a tributary to the Meta River within the Orinoco macro-basin (Corporinoquia, 2010; Corporinoquia and Corpoboyacá, 2015b) (Fig. 2).

The basin experiences a monomodal rainfall regime with a distinct wet season from April to November and a dry season from December to March. Although 90% of the basin lies within the Andean region, its discharge contributes to the Orinoquía domain. Mean annual precipitation is approximately 2,031 mm, with peak streamflows during the wet season and minimum flows in February (Ruíz-Ochoa et al., 2022; Urrea et al., 2016).

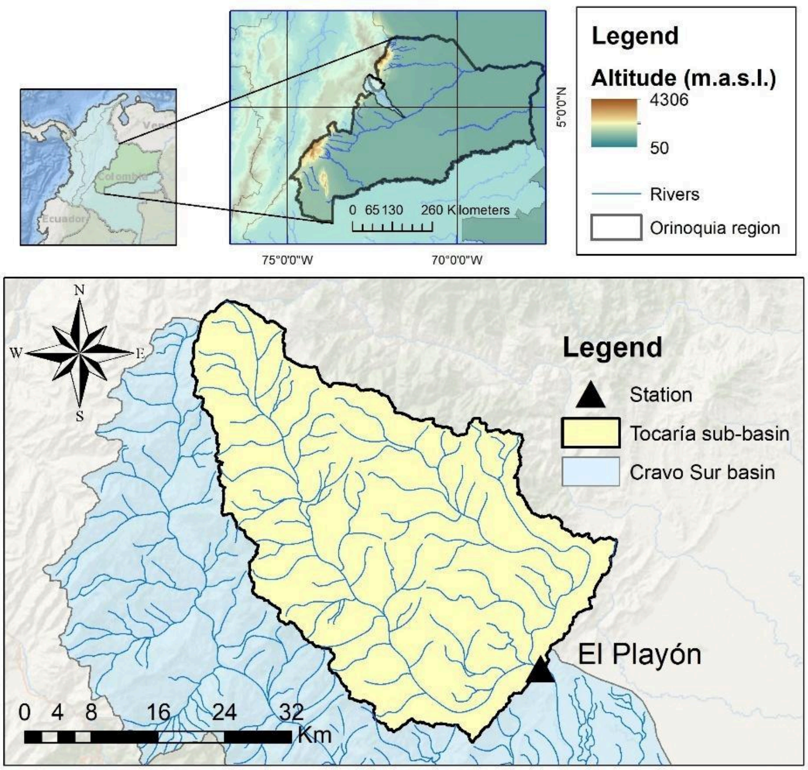

Figure 2: Geographic location of the Tocaría River basin and El Playón hydrometric station. Digital elevation model, political boundaries, watershed boundaries, and rivers extracted from GEOPORTAL IDEAM: https://archivo.ideam.gov.co/geoportal

2.2 Data

Three main datasets were used for model development: observed streamflow, CHIRPS precipitation, and macroclimatic variables.

Streamflow data were obtained from the El Playón station, operated by Colombia's national hydrometeorological institute (IDEAM), covering 1983–2019 with approximately 5% missing values. Data were sourced from the DIHME geoportal (http://dhime.ideam.gov.co) and the missing values were imputed using NLPCA approach, incorporating CHIRPS precipitation data as an exogenous input, following the approach described with more details by Ocampo-Marulanda et al. (2025).

Mean monthly discharge averaged 92.4 m^3/s, with strong seasonal variability: February averaged 17 m^3/s (dryest month), while July peaked at 177 m^3/s (Fig. 3). These fluctuations highlight the necessity of modeling strategies that capture both hydroclimatic seasonality and interannual variability.

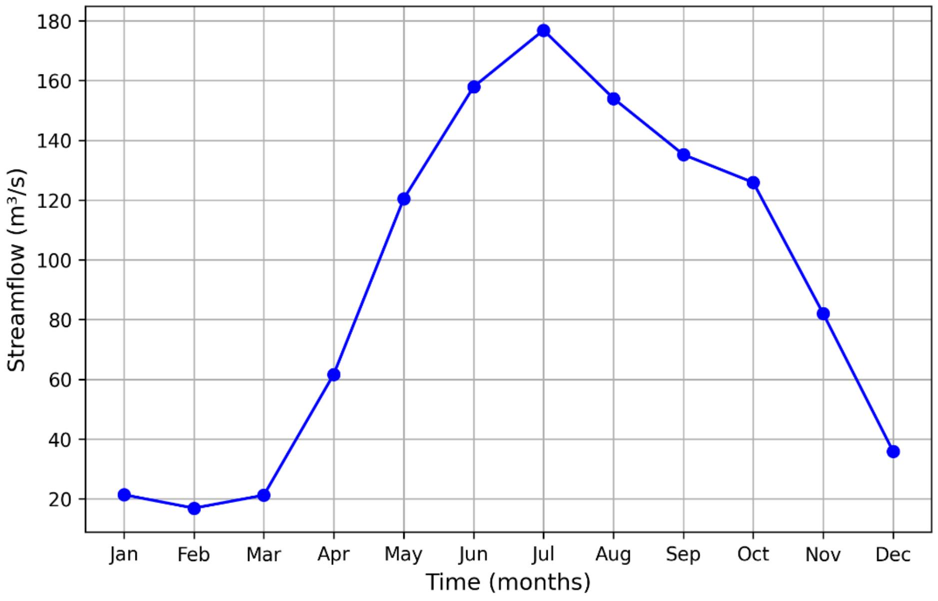

Figure 3: Monthly streamflow climatology at the El Playón station, Tocaría River.

Satellite-based Precipitation data were sourced from the CHIRPS, which integrates satellite and in situ observations. CHIRPS provides global coverage at 0.05° spatial resolution from 1981 to present. Its performance in Colombia is well documented (Funk et al., 2014; Urrea et al., 2016; Ocampo-Marulanda et al., 2022). In this study, were used 81 time series CHIRPS corresponding to 81 grid cells located within the Tocaría River basin.

Macroclimatic variables were retrieved from the National Oceanic and Atmospheric Administration (NOAA, https://psl.noaa.gov/gcos_wgsp/). Twenty-one indices representing sea surface temperature anomalies, atmospheric pressure, and wind patterns were selected based on prior demonstrated correlations with Colombian streamflow regimes (Canchala et al., 2020a; Cerón et al., 2020; Poveda et al., 2011). Complete descriptions of selected variables are available in Supplementary Table S1.

2.3 Data preprocessing

This subsection details statistical preprocessing applied to streamflow data prior to model development, including variability characterization, trend detection, structural change identification, and stationarity assessment. All analyses were conducted using the Python programming language. Data processing and manipulation were carried out using pandas and NumPy. Statistical and time series analyses were performed using SciPy and statsmodels. The NLPCA analysis was implemented using scikit-learn and TensorFlow.

2.3.1 Descriptive statistics and trend analysis

Descriptive statistics calculated included mean (μ), maximum, standard deviation (σ), and coefficient of variation (CV), providing an initial quantification of streamflow magnitude and variability. Structural changes were identified via Pettitt's test (Pettitt, 1979), revealing a significant breakpoint in the time series mean. Sen's slope estimator (Sen, 1968) quantified long-term trends, while the Mann-Kendall test (Mann, 1945) evaluated trend significance.

2.3.2 Stationarity assessment

Stationarity was explored through additive time series decomposition using classical seasonal decomposition based on moving averages, which separates the series into trend, seasonal, and residual components, enabling visual inspection of long-term behavior, seasonal structure, and residual stability. Subsequently, stationarity was formally evaluated using the Phillips–Perron (PP) test, which tests the null hypothesis of a unit root and is robust to autocorrelation and heteroskedasticity in the residuals.

2.3.3 Autocorrelation and Partial Autocorrelation

Temporal dependencies were examined via autocorrelation function (ACF) and partial autocorrelation function (PACF) analyses (Chatfield, 1989). The ACF quantifies linear correlation at lag *k*, defined as:

$$r_k = \frac{\sum_{t=1}^{n-k} (x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^{n-k} (x_t - \bar{x})^2} \quad (1)$$

where x_t is the series value at time *t*, $\bar{x}$ is the mean, and *n* the number of observations.

PACF measures correlation at lag k controlling for intermediate lags, informing autoregressive and moving average term selection in SARIMA models (Box et al., 2015). ACF and PACF plots were examined for seasonality and persistence, using 95% confidence intervals to identify significant lags.

2.4 Selection of possible correlators

2.4.1 CHIRPS precipitation as a predictor

The hydrological relationship between precipitation and streamflow in the Cravo Sur basin has been previously demonstrated (Ocampo-Marulanda et al., 2025), supporting the use of CHIRPS precipitation data as potential predictors. In this study, 81 precipitation time series corresponding to CHIRPS grid cells within the Tocaría sub-basin were analyzed and for reduce the dimensions and to extract the main information of this dataset (Xn) we used NLPCA approach. The Non-linear Principal Components (NLPCs) obtained in the bottleneck of the NLPCA approach (ym) depict the main modes precipitation variability in this region. For this purpose, we use Multi-Layer Perceptron (MLP) neural network which has a structure with an input layer, single or multiple hidden layers, and an output layer. For the training, we use a learning algorithm to find the best combination of synapse weights with the least error using the back-propagation algorithm to update the weight and bias values. We evaluated several network topologies and found that the [81–2–81] architecture produced the best results, and the components were estimated in hierarchical order. We used the linear functions for the hidden and output layers activation. The best results were obtained using 100 epochs in the training.

The Non-linear principal components were estimated using only the 70% subset of precipitation data corresponding to the SARIMAX model training period (1985-2008), as defined in Section 2.5. Implemented via an autoencoder with bottleneck architecture, NLPCA minimizes reconstruction error between inputs and outputs, capturing nonlinear relationships beyond traditional PCA capabilities (Canchala et al., 2019; Scholz et al., 2007). Validation of extracted components included reconstruction accuracy and latent space visualization, confirming their efficacy in summarizing precipitation variability.

2.4.2 Macroclimatic variables as large-scale predictors

MVs comprising sea surface temperature anomalies, sea-level pressure, and wind indices known to influence Colombian hydrology (Canchala et al., 2020a,b; Cerón et al., 2020; Poveda et al., 2001) were evaluated.

Spearman rank correlations between each MV and streamflow were calculated for lags up to 14 months to capture delayed effects. Variables exhibiting statistically significant correlation at lags $\geq$ 6 months were retained, reflecting hydroclimatic memory. Only predictors with demonstrated hydrological relevance and minimal intercorrelation were preserved to ensure model parsimony and stability.

2.4.3 Multicollinearity diagnosis and variable selection

A two-step screening procedure was applied to select the exogenous predictors used in the forecasting models. First, pairwise correlation matrices visualized as heatmaps were used to identify groups of highly correlated MVs, thereby detecting potential redundancy among predictors and identifying those with the strongest association with streamflow (Dormann et al., 2013). Second, the Variance Inflation Factor (VIF) was computed to quantify multivariate collinearity among the candidate predictors (Katrutsa and Strijov, 2017).

Predictors with VIF values greater than 10 were considered excessively collinear and were iteratively removed. When several variables conveyed similar large-scale climatic information, selection was guided not only by the reduction of VIF values but also by the strength of their absolute correlation with streamflow, in order to retain the predictor showing the most hydrologically relevant signal while preserving model parsimony. After each elimination step, VIF values were recalculated until the reduced predictor set showed acceptable multicollinearity levels.

2.5 Forecasting model: SARIMAX architecture

The forecasting approach proposed herein is based on a hybrid SARIMAX framework that integrates nonlinear temporal dynamics, exogenous climatic forcings, and seasonal variability within a unified predictive architecture.

The forecasted streamflow vector, denoted as $\hat{y}(k) \in R^m$, is modeled as a nonlinear function of multiple lagged sequences:

Streamflow autoregressive terms:

$$y(k),\ y(k-1),\ \ldots,\ y\left(k-n_y\right) \tag{2}$$

Exogenous forcings (e.g., ocean-atmospheric indices):

$$u_1(k),\ u_1(k-1),\ \ldots,\ u_1\left(k-n_{u1}\right),\ \ldots;\ u_r(k),\ u_r(k-1),\ \ldots,\ u_r\left(k-n_{ur}\right) \tag{3}$$

Forecast residuals:

$$e_k(k),\ e(k-1),\ \ldots,\ e\left(k-n_e\right) \tag{4}$$

Seasonal lags of streamflow and residuals:

$$y(k-s),\ y(k-2s),\ \ldots,\ y\left(k-n_s \cdot s\right), \tag{5}$$

$$e(k-s), e(k-2s),\ \ldots,\ e(k-n_s \cdot s) \tag{6}$$

The model's general transfer function is expressed as Ec. (7):

$$y(k+1) = fW\left[y(k),\ y(k-1),\ \ldots,\ y\left(k-n_y\right);\ u_1(k),\ u_1(k-1),\ \ldots,\ u_1\left(k-n_{u1}\right);u_r(k),\ u_r(k-1),\ \ldots,\ u_r\left(k-n_{ur}\right) e_k(k),\ e(\right. \tag{7}$$

Where n_y, n_{ur}, n_e, and n_s represent the autoregressive, exogenous, residual, and seasonal lag orders, respectively; s denotes seasonal periodicity; and fW is a nonlinear mapping function parameterized by model weights W.

This hybrid configuration enables simultaneous modeling of high-order dependencies, memory effects, and seasonal recurrence. Incorporating lagged residuals and seasonal components allows the model to capture patterns not accounted for by direct input–output relations, enhancing its capacity to learn both deterministic and stochastic dynamics.

An open-loop configuration was adopted during training, feeding observed streamflow values back to stabilize learning. For validation and operational forecasting, a closed-loop mode recursively propagates forecasted values. Residual sequences were standardized during validation and testing to ensure numerical stability.

The structure of the SARIMA model was determined through visual inspection of autocorrelation (ACF) and partial autocorrelation (PACF) plots of the monthly streamflow series. The orders (p,d,q) and (P,D,Q) were selected to reflect the presence of short-term dependencies and seasonal dynamics. Differencing was not applied, as the series appeared stationary in both its level and seasonal pattern. To support this assessment, the Phillips–Perron test was applied to evaluate the presence of unit roots and inform the choice of d=0 and D=0. A seasonal period of s=12 was defined based on the observed annual cycle in streamflow behavior.

Data were partitioned into 70% training (1985-2008), 15% validation (2009-2013), and 15% testing (2014-2019) subsets, preserving temporal ordering to avoid information leakage. Alongside Bayesian regularization, early stopping was implemented by monitoring validation loss with a patience threshold of 100 epochs, halting training upon stagnation in performance.

2.6 Logarithmic transformation of streamflow

Monthly streamflow series often present positive skewness, heteroscedastic variance, and pronounced extremes associated with flood events, which challenge the assumptions of linear stochastic models such as SARIMA and SARIMAX. To address these issues, a logarithmic transformation was applied to the observed streamflow prior to model calibration:

$$y_t^{log} = log(y_t + 1) \tag{8}$$

This transformation stabilizes variance and improves the suitability of the series for linear time-series modeling. All SARIMA estimation and forecasting were performed in the log-transformed space. Forecasts were then back-transformed to the original scale to ensure physical interpretability:

$$\hat{y}_t = e^{(\hat{y}_t^{log})} - 1 \tag{9}$$

The transformation enhanced the model's ability to represent both low- and high-flow conditions, particularly under strong seasonal variability.

2.7 Evaluation of the models

2.7.1 Evaluation with metrics and baselines

Model performance was first assessed using R^2 and RMSE for the training, validation, and test periods, alongside AIC and BIC to evaluate goodness of fit. These metrics provide a quantitative overview of how well each model reproduces the general behavior of the streamflow series.

Forecasts were then compared against simple baselines: monthly climatology, computed as the average flow for each month in the training period, and persistence, which assumes the flow at time $t + 1$equals the flow at time t. This comparison allows us to contextualize the added value of the SARIMAX model relative to straightforward, intuitive approaches.
Special attention was given to the evaluation of high flows, defined as events above the 90th percentile, by computing RMSE for these peaks. Since the SARIMAX model is designed to capture extreme events, this peak-focused analysis was prioritized to assess its skill in predicting the most critical conditions for water management.
Finally, 95% confidence intervals were estimated for all forecasts. After fitting the model on log-transformed data with exogenous variables, predictions and confidence intervals were obtained in log scale and back-transformed to the original units. These intervals illustrate the uncertainty around the forecasted values, providing a clear sense of the expected range for future streamflow.

2.7.2 Forecast error degradation across prediction horizons

Forecasting accuracy typically decreases as the prediction horizon extends, a well-documented phenomenon referred to as forecast error degradation (Xiang et al., 2024). To assess this behavior, the best-performing SARIMAX model was evaluated over multiple lead times using RMSE and R^2 as performance metrics. Forecasts were produced for 2 to 24 months ahead in increments of two months, enabling the quantification of predictive accuracy across short, medium, and long term horizons.
To capture the evolution of forecast quality, RMSE and R^2 values were analyzed in pairs of consecutive horizons. This pairing facilitated the visualization of degradation trends through line plots, offering insights into the temporal limits of model effectiveness.

2.7.3 Residual analysis

Residual diagnostics are essential to evaluate model adequacy by verifying whether prediction errors conform to standard statistical assumptions. Residuals $e_t = y_t - \hat{y}_t$ were computed as the difference between observed and forecasted streamflow values, forming a time series subjected to rigorous statistical and graphical analysis.
Normality was tested using the Jarque–Bera test (Jarque and Bera, 1980), assessing whether residuals follow a Gaussian distribution, a prerequisite for valid inference and uncertainty quantification. Autocorrelation was examined through the Durbin–Watson statistic (Durbin and Watson, 1950), where values near 2 suggest the absence of significant serial dependence, thus confirming that temporal structures have been effectively captured. Heteroscedasticity was assessed via the Breusch–Pagan test (Breusch and Pagan, 1979); significant p-values indicated non-constant residual variance, potentially undermining the model's generalizability.
Complementing these tests, graphical diagnostics were employed to detect patterns indicative of model misspecification. Histograms with kernel density estimation (Silverman, 1986) visually assessed normality, while residuals versus fitted values plots revealed heteroscedastic patterns or structural biases. Quantile–quantile plots (Wilk and Gnanadesikan, 1968) further

tested alignment between empirical and theoretical residual quantiles. Finally, ACF plots identified lagged dependencies, where the absence of significant peaks outside confidence bounds supported the white noise assumption.

3 RESULTS AND DISCUSSIONS

The experimental results are organized into three subsections: (1) Description of the streamflow data, (2) Determination of the pool predictors, and (3) Forecasted streamflow series. Each subsection presents and discusses relevant findings, emphasizing the identification of the most suitable forecasting model based on accuracy metrics. A comparative analysis of the four evaluated models highlights their strengths and limitations, guiding the selection of the most effective model for streamflow prediction.

Long-range streamflow forecasting is a valuable tool for water resource managers, enabling proactive decision-making and strategic planning. However, excessively extending the forecast lead-time tends to degrade model performance, often resulting in diminished predictive accuracy. Consequently, this study adopts a lead-time of 24 months to evaluate and compare forecasting models. This horizon provides a balance between planning requirements and achievable accuracy for watershed management stakeholders.

3.1 Description of the streamflow data

3.1.1 Statistical and graphical description

The streamflow time series at El Playón station exhibits pronounced interannual and seasonal variability, with values ranging from 2.5 m^3/s to 280 m^3/s and a mean of 92.4 m^3/s. This amplitude is consistent with the monomodal rainfall regime characteristic of the region (Urrea et al., 2016). A standard deviation of 66.4 m^3/s and a coefficient of variation of 71.8% confirm the high dispersion and temporal heterogeneity of the data, which poses a challenge for accurate modeling and prediction.

Trend detection tests were also applied. A visual inspection suggested a potential breakpoint in January 2019; however, the Pettitt test yielded a p-value of 1.00, indicating that this change is not statistically significant. Likewise, Sen's slope and the Mann-Kendall test indicate a slight upward trend in discharge, but without statistical significance ($p = 0.3445$), suggesting the absence of a persistent monotonic trend (see Table 1 and Fig. 4).

Table 1: Streamflow characteristics at El Playón station

Mean m^3/s	Min m^3/s	Max m^3/s	Standard deviation m^3/s	Coefficient of variation %	Break poin	Pettitt (p-value)	Sen slope	Mann-Kendall (p-value)
92.4	2.5	280	66.4	71.8	Jan/2019	1.00	0.0188	0.3445

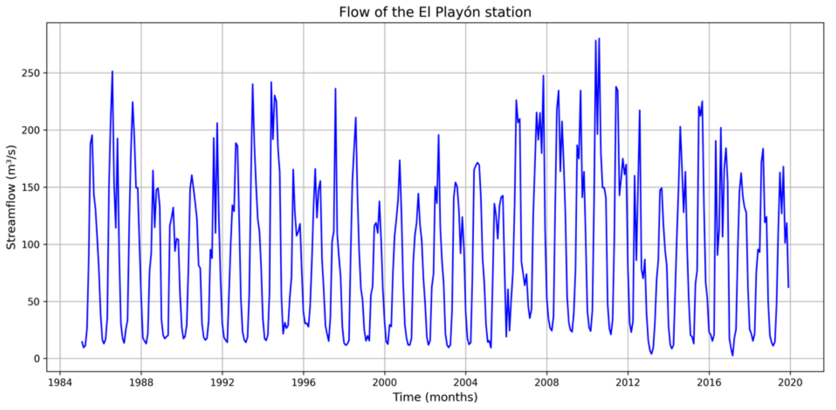

Figure 4: Temporal behavior of streamflow at the Tocaría River.

These characteristics highlight the complexity of the time series and justify the need for rigorous preprocessing, including stationarity and autocorrelation analysis.

3.1.2 Stationarity analysis

Stationarity was assessed through a combination of exploratory decomposition and formal statistical testing. Figure 5 presents the seasonal decomposition results. The trend component does not exhibit a persistent increasing or decreasing pattern over the analyzed period; instead, it shows multi-year fluctuations characterized by alternating phases of higher and lower streamflow. These variations suggest low-frequency hydroclimatic variability rather than a clear long-term trend, indicating that the mean discharge remains broadly stable over time, although modulated by interannual and decadal variability.

The seasonal component displays a clear and regular periodicity, reflecting a 12-month seasonal cycle in streamflow. This pronounced seasonality is consistent with the regional hydroclimatic regime, dominated by the alternation between wet and dry seasons, and confirms that streamflow variability in the Tocaría River is largely controlled by seasonal precipitation patterns.

The residual component exhibits a random distribution centered around zero, with no discernible systematic patterns or trends. It captures short-term variability, extreme events, and noise not explained by the trend or seasonal components, supporting the adequacy of the seasonal decomposition and the assumption of weak stationarity after seasonal removal.

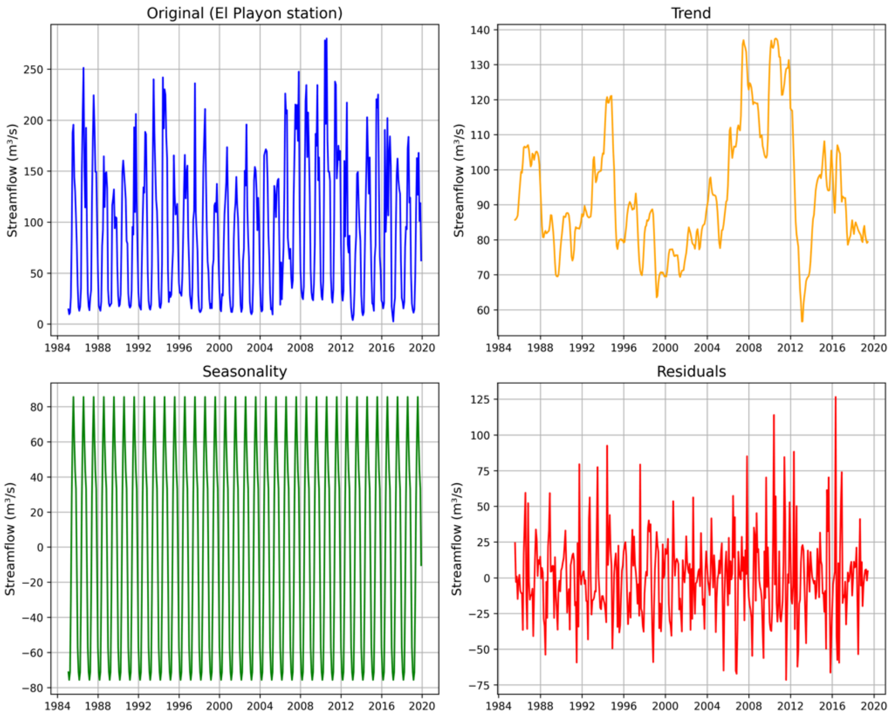

Figure 5: Decomposition of El Playón streamflow time series into trend, seasonality, and residual components.

The Phillips-Perron test further confirmed stationarity, yielding a test statistic of -6.5005 and a p-value < 0.0001, allowing rejection of the null hypothesis of a unit root. These results justify modeling without differencing and the selection of $d = 0$ in the SARIMA model.

3.1.3 Autocorrelation and partial autocorrelation structure

The ACF displays significant periodic peaks at regular intervals, indicative of strong seasonal periodicity and persistent temporal dependencies beyond short lags (Fig. 6). This cyclicality aligns with hydrological patterns driven by seasonal precipitation and runoff dynamics, underscoring the appropriateness of seasonal forecasting models such as SARIMA.

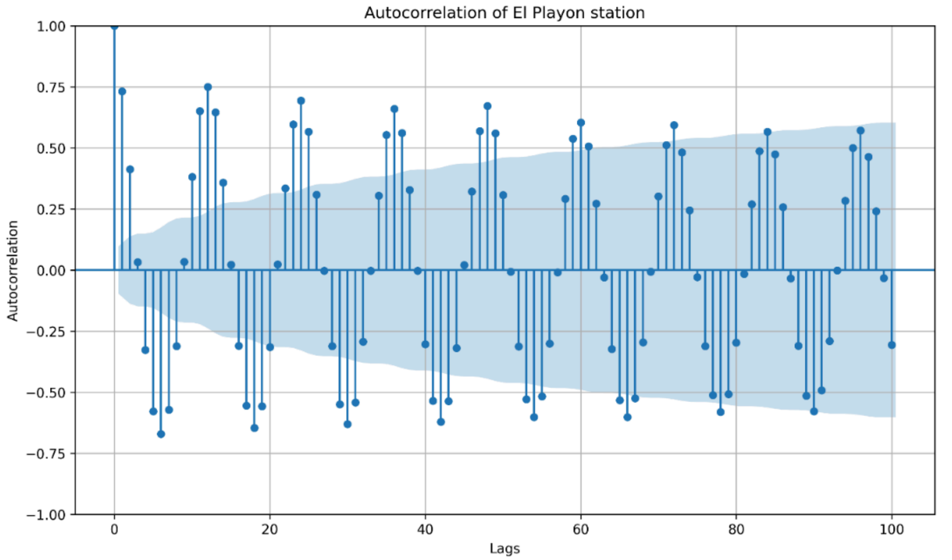

Figure 6: Autocorrelation of the time series of El Playón streamflow for 100 months of lags.

The PACF shows a prominent spike at lag 1 followed by rapid decay to values within the confidence bounds (Fig. 7), consistent with an autoregressive process of order 1 (AR(1)) though the selected SARIMA model incorporates higher-order dynamics to capture the longer memory and seasonal structure identified in the ACF. The lack of significant PACF values at higher lags indicates limited benefit from additional autoregressive terms.

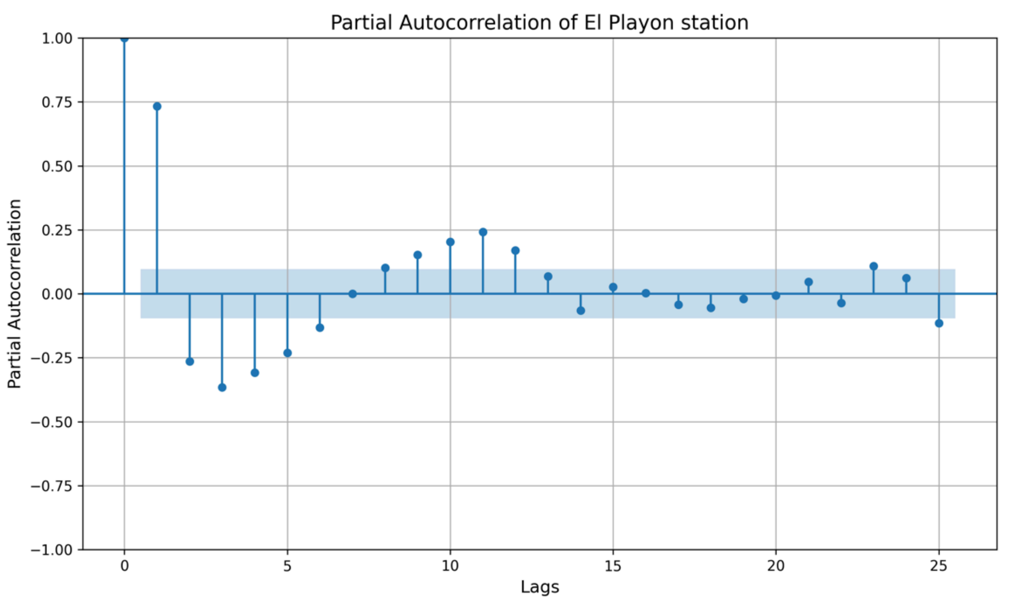

Figure 7: Partial autocorrelation function of El Playón streamflow series over 25 monthly lags.

Based on combined ACF and PACF analysis, a SARIMA $(4,0,4)(0,0,3)_{12}$ model was selected. This model configuration incorporates the identified seasonal periodicity (season length s = 12), confirms stationarity (d = 0), and captures short-term autoregressive and moving average dynamics. Residual diagnostics validated the model's adequacy, demonstrating no significant autocorrelation.

3.2. Determination of the pool predictors

3.2.1 Selection of precipitation NLPCs as exogenous variables

The 81 CHIRPS precipitation time series for the Tocaría River basin were compressed into a two-dimensional nonlinear principal components (NLPC1 and NLPC2) using an autoencoder-based NLPCA. The resulting latent scores show variances of 17.64 (NLPC1) and 6.30 (NLPC2), corresponding to 73.69% and 26.31% of the variance within the two-dimensional latent space, respectively. These percentages describe the relative contribution of each latent coordinate. The two NLPCs were subsequently used as exogenous covariates in the streamflow forecasting model, providing a parsimonious representation of basin-scale precipitation variability. This approach is consistent with Ocampo-Marulanda et al. (2025), who used CHIRPS-derived nonlinear precipitation components as compact predictors for streamflow reconstruction in data-scarce basins.

3.2.2 Selection of macroclimatic variables as exogenous variables

The influence of 21 MVs on the streamflow variability of the Tocaría River was assessed using a lagged Spearman correlation analysis, considering time lags from 0 to 14 months. This asynchronous framework allowed for identification of delayed relationships between climate anomalies and hydrological responses, which are crucial for predictive modeling. This lagged correlation framework enables detection of delayed hydroclimatic teleconnections, which are essential for long-lead streamflow forecasting.

Statistical significance was evaluated using two-tailed tests with confidence levels between 90% and 99%. Figure 8 displays the lagged correlations between the MVs and the streamflow at the El Playón station. Several variables demonstrated statistically significant correlations across multiple lags, exhibiting a quasi-periodic structure that reflects the hydroclimatic seasonality modulated by the Intertropical Convergence Zone (ITCZ).

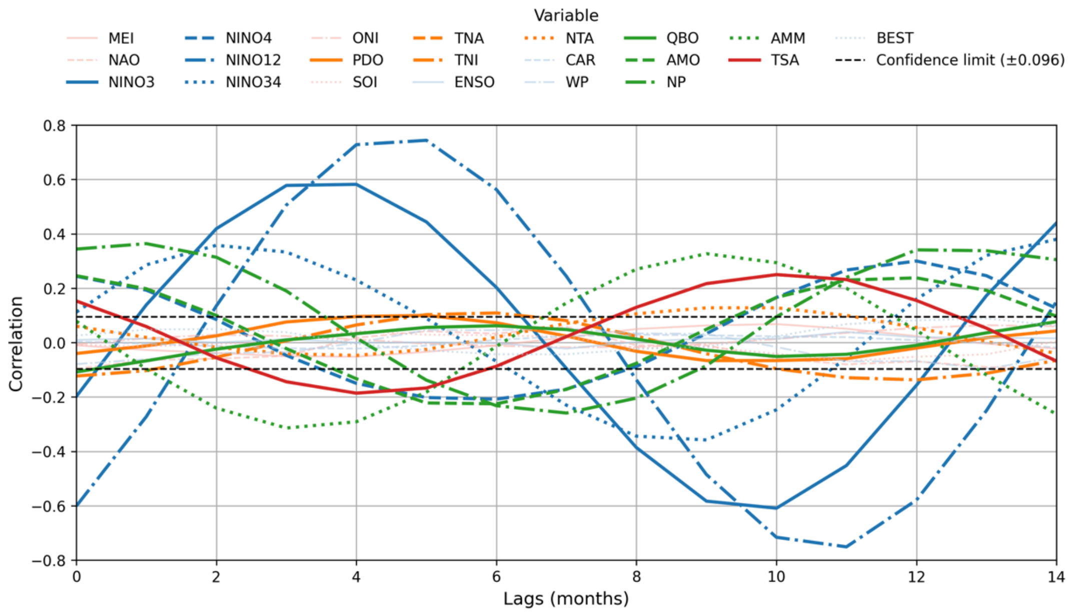

Figure 8: Lagged Spearman correlations between MVs and streamflow at El Playón station.

Key Pacific-related indices—namely NINO12 (ρ = 0.75) and NINO3 (ρ = 0.60)—emerged as the strongest predictors, underscoring the robust influence of ENSO-related sea surface temperature (SST) anomalies on the Tocaría River streamflow. These findings are consistent with previous studies across Colombia (e.g., Cadavid and Salazar, 2008; Canchala, 2020c; Poveda et al., 2002), which reported comparable lagged correlations between ENSO indices and regional streamflow, often with cyclic behavior.

In addition, Atlantic-related variables such as AMM (ρ = 0.32) and TSA (ρ = 0.25) showed statistically significant, albeit weaker, correlations. These results corroborate the relevance of Atlantic SST variability in eastern Colombia. The Llanos low-level jet and moisture transport from the tropical Atlantic—modulated by indices such as AMM and TSA—have been highlighted as key controls of precipitation and streamflow in the Orinoquía region (Correa et al., 2024; Builes-Jaramillo et al., 2022; Labat et al., 2012; Nieto et al., 2008).

The set of significantly correlated predictors spans both Pacific and Atlantic basins: NINO3, NINO4, NINO12, NINO34, PDO, TNI, NP, QBO (Pacific-related), and AMO, NTA, TNA, NAO, AMM, TSA (Atlantic-related). This reinforces the complexity of the climate–hydrology interactions in the Tocaría River basin and emphasizes the need to account for both direct and lagged macroclimatic signals in predictive modeling.

3.2.3 Assessing multicollinearity and variable selection for improved time series forecasting

The results of the correlation heatmap (Fig. 9), reveal substantial interdependence among several MVs, indicating that the inclusion of multiple closely related predictors may introduce redundancy and model instability. For instance, the ENSO indices NINO3, NINO4, NINO12, and NINO34 exhibit strong mutual correlations, suggesting that a single representative index is sufficient to capture ENSO-related variability. Similarly, the high correlations observed among the TNA, NTA, and TSA indices indicate overlapping information content; therefore, only one of these indices was retained in the final predictor set to mitigate multicollinearity.

Similarly, the high correlations observed among the TNA, NTA, and TSA indices indicate overlapping information content; therefore, only one of these indices was retained in the final predictor set to mitigate multicollinearity. In contrast, variables such as NAO, PDO, TNI, and QBO showed weaker absolute correlations with streamflow and, therefore, were considered less informative for forecasting purposes than the predictors retained in the final subset.

To ensure model robustness and reduce the risk of overfitting, multicollinearity among the candidate predictors was quantified using the VIF. Table 2 summarizes the VIF values for all variables considered. While most predictors exhibited acceptable multicollinearity levels (VIF < 10), several variables—particularly those associated with ENSO—showed extreme multicollinearity. Specifically, the NINO3, NINO4, NINO12, and NINO34 indices presented VIF values ranging from 2,492 to 81,825. In addition, the TNA and TSA indices yielded infinite VIF values, indicating perfect collinearity with other predictors, most likely due to structural redundancy among Atlantic SST indices. This behavior is consistent with the strong interdependencies observed in the correlation heatmap.

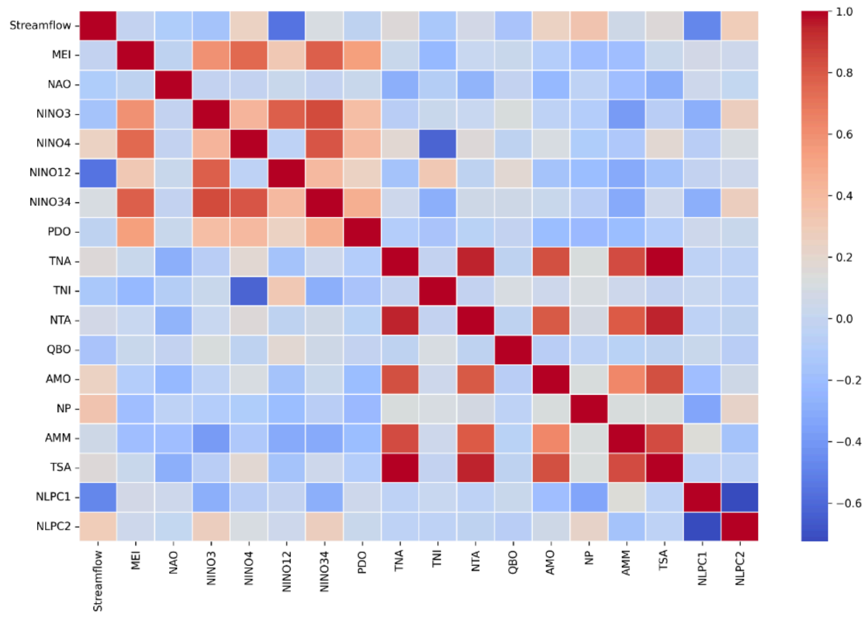

Figure 9: Correlation heatmap of predictor variables.

Table 2: Variance Inflation Factor Analysis for predictor variables

Variable	VIF with all variables	VIF with selected variables
NAO	2	N.A
NINO3	38,899	N.A
NINO4	29,126	N.A
NINO12	2,492	3.2
NINO34	81,825	N.A
PDO	2	N.A
TNA	inf	8.9
TNI	4	N.A
NTA	16	N.A
QBO	1	N.A
AMO	8	3.8
NP	17,75	N.A
AMM	8	4.2
TSA	INF	N.A
PNLPCA_PC1	4	2.5
PNLPCA_PC2	4	3.7

To address this, various combinations of variables were tested through an iterative model refinement process, prioritizing predictors with strong individual relationships to streamflow and minimal mutual collinearity. This selection strategy balances predictive power with model parsimony, mitigating multicollinearity-induced variance inflation and enhancing generalizability.

Ultimately, the optimal subset of exogenous predictors was determined to be:

PNLPCA_PC1, PNLPCA_PC2, NINO12, TNA, AMO and AMM.

This final configuration minimized multicollinearity, preserved critical information from both precipitation and climate variability, and improved the interpretability and reliability of the streamflow forecasting model. The final pool of predictors; comprising key macroclimatic indices and the first two NLPCs; provided the input for the forecasting models evaluated in the subsequent section. The next subsection presents the performance comparison of candidate models and the resulting streamflow forecasts.

3.3 Forecasted streamflow series

Models were trained, validated, and tested on independent datasets to ensure robust generalization. Table 3 summarizes the corresponding performance metrics. Overall, all model configurations exhibited adequate skill in forecasting monthly streamflow with lead times of up to 24 months.

Among the SARIMAX configurations, models based on log-transformed streamflow consistently showed good performance during both the validation and test phases, with stable R^2 and RMSE values, indicating robust generalization. These models achieved the best SARIMAX performance in the test period, with R^2 values ranging from 0.55 to 0.64 and RMSE values between 39.3 and 43.9, outperforming their counterparts using non-transformed data.

Across the validation and test periods, the monthly climatology model yielded the lowest overall RMSE (33.9) and the highest R^2 (0.73), thus outperforming the SARIMAX-based models in terms of average predictive accuracy. This result is expected because climatology effectively reproduces the dominant seasonal cycle of monthly streamflow, which explains a large proportion of the variance under typical conditions. However, when model performance was specifically evaluated under high-flow conditions, the LOG-NNSARIMAX-MVNL configuration achieved the lowest RMSE (60.8), indicating superior skill in reproducing peak flows. This improved performance likely arises from the model's ability to incorporate nonlinear precipitation patterns, macroclimatic variability, and temporal dependence, which provide additional predictive information during months that deviate markedly from the historical seasonal mean. In contrast, the climatological benchmark, by construction, cannot respond to anomalously wet conditions and therefore tends to smooth or underestimate peak-flow events.

In terms of model complexity, AIC values were similar (>2000) among SARIMAX models calibrated using non-transformed data, whereas substantially lower AIC values (>200) were obtained for models based on log-transformed inputs. A comparable pattern was observed for BIC, with lower values for log-transformed models (>200) relative to non-transformed models (>2500), indicating a more parsimonious model structure when logarithmic transformation was applied.

Table 3: Forecast performance metrics for the four streamflow forecasting models

Model	Training		Validation		Test		RMSE test	AIC	BIC
	R^2	RMSE m^3/s	R^2	RMSE m^3/s	R^2	RMSE m^3/s	Peak m^3/s		
SARIMA (4,0,4) $(0,0,3)_{12}$	0.70	35.3	0.64	108.4	0.46	48.4	108.4	2580	2598
SARIMAX-MV	0.72	34.2	0.58	81.1	0.30	54.9	81.1	2584	2616
SARIMAX-NLPC	0.70	35.3	0.63	109.0	0.44	49.2	109.0	2582	2607
SARIMAX-MVNL	0.72	34.3	0.59	93.8	0.35	52.8	93.8	2585	2624
LOG-SARIMA (4,0,4) $(0,0,3)_{12}$	0.57	42.3	0.58	47.0	0.64	39.3	68.0	209	227
LOG-SARIMAX-MV	0.57	42.4	0.52	50.1	0.58	42.6	65.2	212	244

LOG-SARIMAX-NLPC	0.57	42.2	0.53	49.4	0.63	39.9	61.9	201	226
LOG-SARIMAX-MVNL	0.57	42.6	0.49	52.0	0.55	43.9	60.8	206	246
Benchmark climatology	0.77	31.0	0.62	44.7	0.73	33.9	76.4	N.A.	N.A.
Benchmark persistence	0.51	45.1	0.41	55.8	0.32	54.0	98.1	N.A.	N.A.

N.A. Not applicable. AIC and BIC are information criteria that balance model goodness-of-fit and complexity. Lower values indicate a more parsimonious model, with BIC imposing a stronger penalty for model complexity than AIC.

These results are consistent with previous findings, Table 4, in the literature. Reported R^2 values for monthly streamflow prediction range from moderate (~0.65) to near-perfect (~0.99), depending on lead time, predictor variables, and catchment complexity.

For instance, Hosseinzadeh et al. (2023) reported $R^2 = 0.92$ using a multivariate SARIMAX model with temperature and precipitation as exogenous inputs in the Colorado River Basin. Cheng et al. (2020) achieved $R^2 = 0.95$ using an LSTM model, although this performance was limited to one-month-ahead forecasts and dropped significantly ($R^2 < 0.20$) at longer lead times.

Several hybrid or AI-based approaches have yielded high predictive accuracy ($R^2 \approx 0.99$), such as those reported by Fathian et al., (2019), Moeeni and Bonakdari (2017), and Dariane and Behbahani (2024). However, these results often stem from short-term horizons, multi-source calibration, or highly tuned model configurations prone to overfitting. Ghorbani et al., (2016) obtained R^2 between 0.77 and 0.84 using ANN and SVM models, while Gómez et al., (2010) reported a more moderate $R^2 \approx 0.64$ for the Bogotá River basin.

In comparison, the performance of our models; particularly the LOG-SARIMAX-MVNL; stands competitively within this broader context. Despite relying on a single hydrometric station and forecasting beyond the one-month horizon, our approach demonstrates robust skill in capturing the high temporal variability and structural complexity of streamflow in the Tocaría River Basin.

Table 4: Validation R^2 of selected monthly streamflow forecasting models from literature

Best model	Description	Validation R^2	References
SARIMAX	24-month forecasts with temperature and precipitation	0.92	Hosseinzadeh et al., 2023
LSTM	One-month-ahead forecasts	0.95	Cheng et al., 2020
MARS1-SETAR	Hybrid nonlinear time series and AI modeling	0.99	Fathian et al., 2019

SARIMA-ANFIS	SARIMA combined with neuro-fuzzy and ANN systems	0.94	Moeeni and Bonakdari, 2017
Multistep data-driven	Neural-fuzzy hybrid with input selection	0.97	Dariane and Behbahani, 2017
ELM	Extreme Learning Machine approach	0.82	Yassen et al., 2016
RBF	MLP, RBF, and SVM models for river flow prediction	0.87	Ghorbani et al., 2016
PMC 20 NCO	Neuro-fuzzy vs neural networks for Bogotá River	0.64	Gómez et al., 2010

3.3.1 Statistical evaluation of the best model for streamflow forecasting

Figure 10 shows the streamflow prediction results for the El Playón station using the LOG-SARIMAX-MVNL model, which integrates exogenous variables including MV's and NLPCs derived from local precipitation data. The figure includes three key elements: observed streamflow, model predictions for the training period, forecasts for the validation period, and test period.

The observed streamflow series reveals pronounced seasonality, characterized by recurrent peaks and troughs consistent with the hydroclimatic regime of the basin. During the training phase, the model reproduces the seasonal dynamics and intra-annual variability with high accuracy, showing minimal deviations. In the validation and test period, although some differences appear; particularly underestimations of extreme highs or delays in response; the general flow patterns, including the timing and intensity of seasonal peaks, are well preserved. These results highlight the model's capacity to capture both historical and prospective streamflow behaviors based on relevant exogenous predictors.

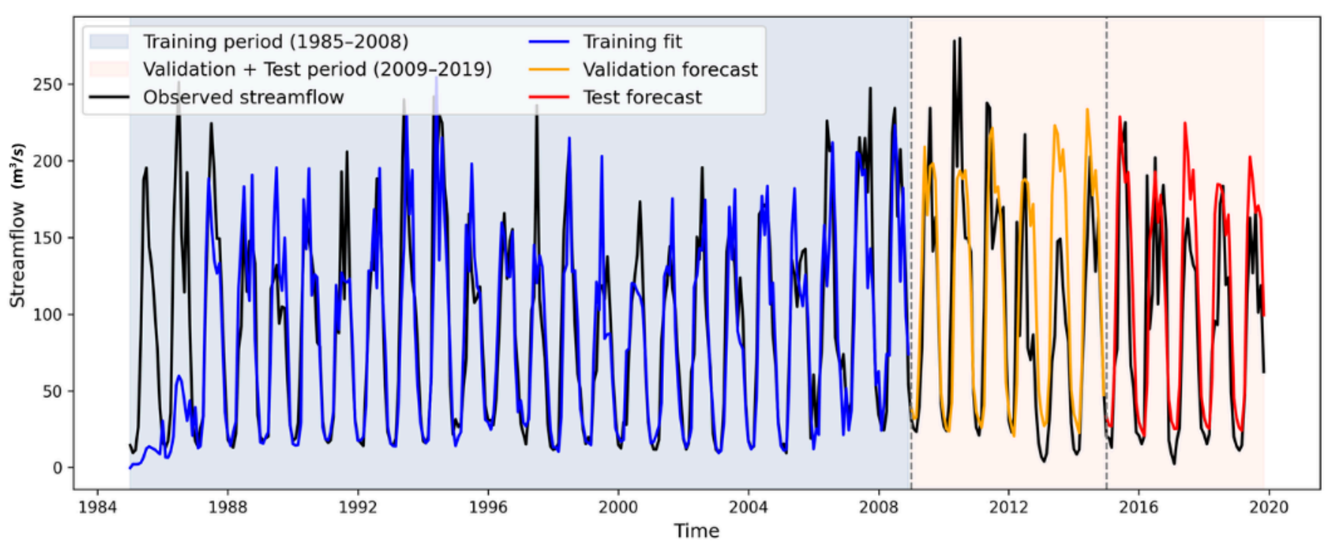

Figure 10: Observed and predicted streamflow for training and validation using the LOG-SARIMAX-MVNL model

The figure 11 shows the performance of the log-transformed SARIMAX model for river flow forecasting. Observed flows are in black, the model fit during training in blue, and the forecast mean in red. Shaded areas indicate uncertainty: the lighter band is the 95% prediction interval, and the darker band is the 95% confidence interval. The model captures the overall trend and seasonal patterns well, though extreme peaks tend to be underestimated, and uncertainty increases for higher flows, reflecting their inherent variability. Overall, the figure highlights both the model's accuracy and the reliability of its forecasts.

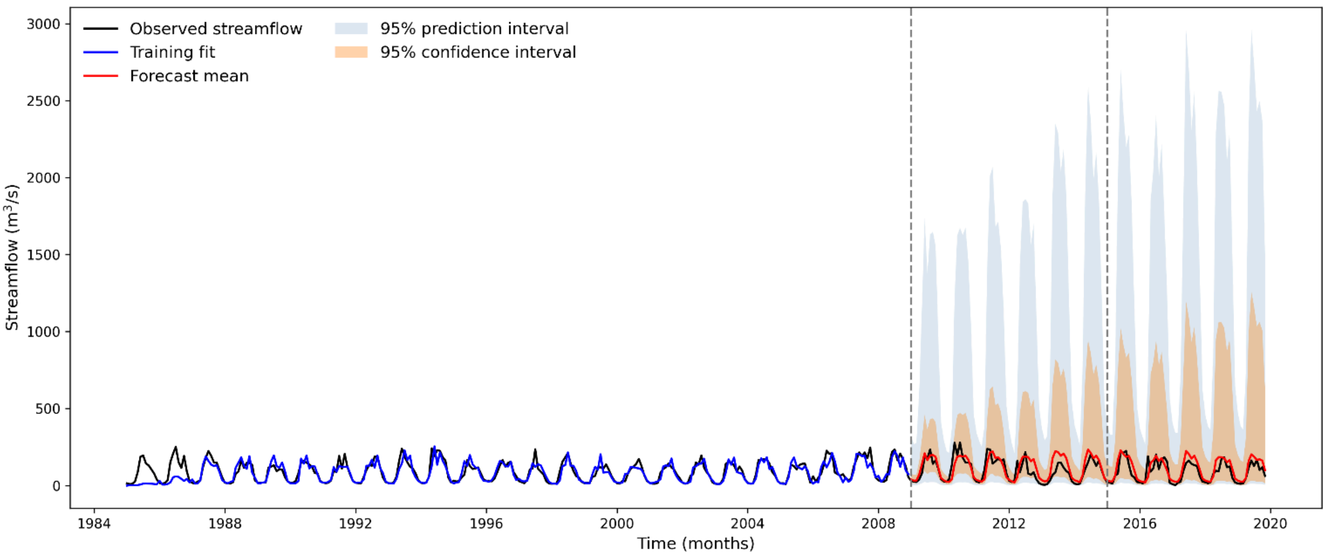

Figure 11. Forecast of log-transformed streamflow using SARIMAX with 95% confidence and prediction intervals

To assess the model's temporal generalization capability, streamflow forecasts were generated for lead times of up to 24 months (Fig. 12). RMSE values remained below 30 m³/s during the first eight forecast months, which is low relative to the standard deviation of the observed series ($\sigma \approx 66$ m³/s). Beyond month 14, RMSE increased progressively before stabilizing at values below 50 m³/s. The coefficient of determination (R^2) remained above 0.92 up to month 12, after which it gradually declined, reaching 0.82 by month 18 and remaining above 0.80 throughout the entire 24-month forecast horizon.

These results confirm the robustness of the LOG-SARIMAX-MVNL model for medium- and long-term forecasting, maintaining a favorable balance between accuracy and lead time. The performance is suitable for practical decision-making applications in water resource planning, including drought early warning and integrated watershed management in the Tocaría River basin.

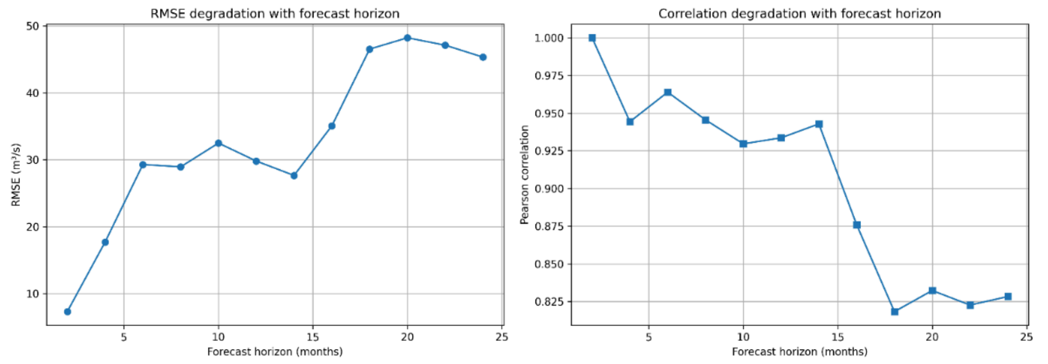

Figure 12: Forecast error degradation across prediction horizons based on RMSE and R^2 metrics

The residuals are not perfectly normal (Shapiro–Wilk W = 0.913, p = 0.0004; Jarque–Bera JB = 58.951, p < 0.001), show some positive autocorrelation (Durbin–Watson = 1.243), and a slight change in variance over time (Breusch–Pagan BP = 3.987, p = 0.0459). These patterns offer useful insights into the series' dynamics and provide additional information for interpreting its temporal behavior.

The diagnostic plots show that the residuals generally follow the expected distribution (see Figure 13), with slight deviations in the tails (Q–Q plot). Residuals tend to spread more at higher fitted values, indicating mild heteroscedasticity (Residuals vs Fitted). The ACF shows mostly no significant autocorrelation, though the first lags suggest weak positive dependence. Overall, the model captures the main dynamics, with residuals showing only minor deviations.

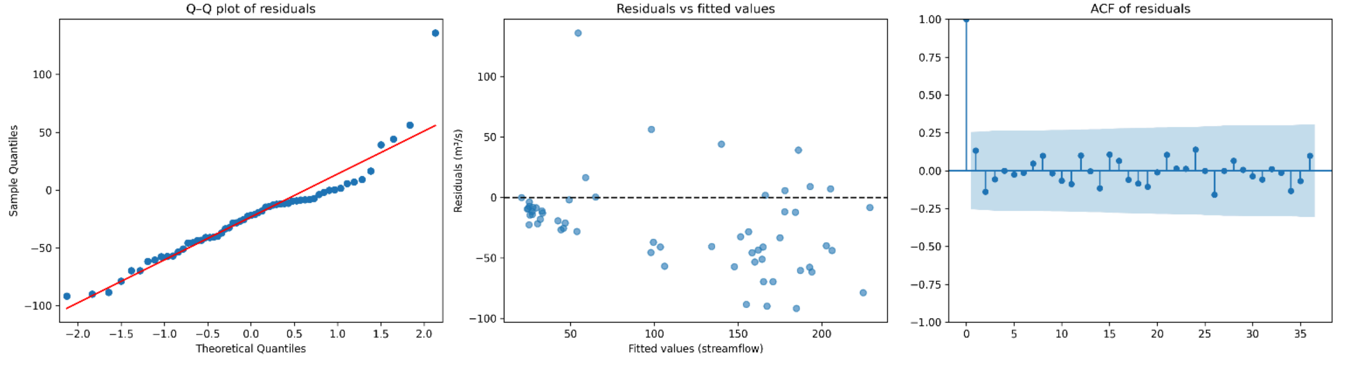

Figure 13: Graphical analysis of normality (Q-Q-Plot of residuals), heteroscedasticity (scatter plot, Residuals vs Fitted values), and ACF of residuals

In summary, the statistical analysis indicates that the LOG-SARIMAX-MVNL model provides reliable streamflow forecasts under peak-flow conditions; however, its ability to accurately reproduce high-magnitude extreme flows remains limited, likely due to complex hydrometeorological processes that are not fully captured by the selected predictors.

4 CONCLUSIONS

This study conducted a comparative assessment of four monthly streamflow forecasting models for a data-scarce basin, all based on the SARIMA framework and progressively enhanced through the inclusion of exogenous predictors. While all configurations showed satisfactory performance, the LOG-NNSARIMAX-MVNL model provided the greatest added value under hydrologically critical conditions. Specifically, when performance was evaluated for high-flow events, this model achieved the lowest RMSE, demonstrating superior skill in reproducing peak flows. This advantage likely reflects its capacity to incorporate nonlinear precipitation patterns, macroclimatic variability, and temporal dependence, thereby capturing anomalous wet conditions that are not represented by the climatological benchmark. The robust performance of the hybrid model, including at lead times of up to 24 months, highlights its potential for anticipatory water-resources management and hydroclimatic risk mitigation in data-limited basins.

The selection of exogenous predictors was grounded in hydroclimatic relevance. At the local scale, precipitation data derived from the CHIRPS dataset were incorporated via NLPCA, allowing the model to account for spatially distributed and nonlinear hydrometeorological influences. The MVs included indices such as NINO12, TNA, AMO and AMM, which collectively capture dominant modes of Pacific and Atlantic ocean–atmosphere variability affecting the Tocaría River basin.

5 AUTHOR CONTRIBUTION

Conceptualization: JDSO, COM, LMCA, TC, and TAFE; data curation: JDSO and COM; formal analysis: JDSO, COM; LMCA, TC, and TAFE; Funding acquisition: LMCA; Methodology: COM, TC, and TAFE; Project administration: LMCA; Resources: LMCA and TAFE; Software: JDSO, COM; Supervision: LMCA, TC and TAFE; Validation: TC and TAFE; Visualization: JDSO and COM; Writing (original draft preparation): JDSO, COM, LMCA and TC; and Writing (review and editing): JDSO, COM; LMCA, TC, and TAFE.

6 COMPETING INTERESTS

The authors also have no other competing interests to declare.

7 ACKNOWLEDGEMENTS

The authors would like to thank to the Research Group TERRANARE of the Fundación Universitaria de San Gil, the Interdisciplinary Forecasting Research Oriented Group - IFROG of the Federal Rural University of Pernambuco and the Environmental Engineering Group - GIA of the Universidad Mariana for their contributions to this research work. Finally, thanks to IDEAM, the United States Geological Survey USGS, and the University of California, Santa Barbara (UCSB) for providing the database containing flow data for the Colombian Orinoquía and precipitation data from CHIRPS.

8 FINANCIAL SUPPORT

The project "Fortalecimiento de los sistemas de información de la calidad y valoración de los servicios ambientales del agua para contribuir al desarrollo sostenible del sector agroindustrial del departamento de Casanare. Yopal – Nunchía" (BPIN Code: 2020000100435) was funded with resources from the Ministerio de Ciencia, Tecnología e Innovación de Colombia – CTeI, through the fund of the Sistema General de Regalías – SGR.

9 REFERENCES

Builes-Jaramillo, A., Yepes, J., Salas, H.D.: The Orinoco low-level jet and its association with the hydroclimatology of northern South America, J. Hydrometeorol., 23, 209–223, https://doi.org/10.1175/JHM-D-21-0073.1, 2022.

Box, G.E.P., Jenkins, G.M., Reinsel, G.C., Ljung, G.M.: Time Series Analysis: Forecasting and Control, 5 ed., John Wiley and Sons, 2015.

Breusch, T.S., Pagan, A.R.: A Simple Test for Heteroscedasticity and Random Coefficient Variation, Econometrica, 47, 1287-1294, http://dx.doi.org/10.2307/1911963, 1979.

Cadavid, J.M., Salazar, J.E.: Generación de series sintéticas de caudales usando un modelo Matalas con medias condicionadas, Av. Recur. Hídricos, 17, 2008.

Canchala-Nastar, T., Carvajal-Escobar, Y., Alfonso-Morales, W., and Caicedo-Bravo, E.: Estimation of missing data of monthly rainfall in Southwestern Colombia using artificial neural networks. Data in Brief, 26, 104517, https://doi.org/10.1016/j.dib.2019.104517, 2019.

Canchala, T., Cerón, W. L., Frances, F., Carvajal-Escobar, Y., Andreoli, R. V., Kayano, M. T., Alfonso-Morales, W., Caicedo-Bravo, E., Ferreira de Souza, R. A.: Streamflow Variability in Colombian Pacific Basins and Their Teleconnections with Climate Indices. Water, 12, 526, https://doi.org/10.3390/w12020526, 2020a.

Canchala, T., Alfonso-Morales, W., Cerón, W. L., Carvajal-Escobar, Y., Caicedo-Bravo, E.: Teleconnections between Monthly Rainfall Variability and Large-Scale Climate Indices in Southwestern Colombia. Water, 12, 1863, https://doi.org/10.3390/w12071863, 2020b.

Canchala, T., Alfonso-Morales, W., Carvajal-Escobar, Y., Cerón, W. L., Caicedo-Bravo, E.: Monthly Rainfall Anomalies Forecasting for Southwestern Colombia Using Artificial Neural Networks Approaches. Water, 12, 2628, https://doi.org/10.3390/w12092628, 2020c.

Cárdenas-Rodríguez, S., Vides-Herrera, A., Pardo-García, A.: Sistema de alerta temprana de inundaciones para el río Arauca basado en técnicas de inteligencia artificial, Rev. Investig. Desarr. Innov., 12(2), 315–326, 2022.

Cerón, W. L., Andreoli, R. V., Kayano, M. T., Ferreira de Souza, R. A., Jones, C., Carvalho, L. M. V.: The Influence of the Atlantic Multidecadal Oscillation on the Choco Low-Level Jet and Precipitation in Colombia. Atmosphere, 11, 174, https://doi.org/10.3390/atmos11020174, 2020.

Chatfield, C.: The Analysis of Time Series: An Introduction, 4 ed., Chapman and Hall, 1989.

Cheng, M., Fang, F., Kinouchi, T., Navon, I.M, Pain, C.C.: Long lead-time daily and monthly streamflow forecasting using machine learning methods. Journal of Hydrology, 590, https://doi.org/10.1016j.jhydrol.2020.125376. 2020.

Chiew, F.H.S., McMahon, T.A.: Teleconnections between the ENSO phenomenon and streamflow, streamflow forecasting, and interannual variability, Hydrol. Sci. J., 47(3), 505–522, https://doi.org/10.1080/02626660209492950, 2002.

Corporinoquia: Resolución No. 200.41‑10.1402 — Por medio de la cual se regula el uso y aprovechamiento del recurso hídrico en el río Tocaría, https://www.corporinoquia.gov.co (últ. acc. 20 jun 2025), 2010.

Corporinoquia, Corpoboyacá: Evaluación de la zonificación de las amenazas en la cuenca del río Cravo Sur (Cap. 6), en: Plan de Ordenación y Manejo de la Cuenca del Río Cravo Sur – POMCA, 1–20, 2015a.

Corporinoquia, Corpoboyacá: Componente hidrológico (Cap. 3), en: Plan de Ordenación y Manejo de la Cuenca del Río Cravo Sur – POMCA, 1–22, 2015b.

Correa, I., Arias, P., Vieira, S., Martínez, A.: A drier Orinoco basin during the twenty‑first century: The role of the Orinoco low‑level jet, Clim. Dyn., 62, 2369–2390, https://doi.org/10.1007/s00382‑023‑07028‑7, 2024.

Dariane, A.B., Behbahani, M.M.: Maximum energy entropy: A novel signal preprocessing approach for data‑driven monthly streamflow forecasting, Ecol. Inform., 79, 102452, https://doi.org/10.1016/j.ecoinf.2023.102452, 2024.

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Münkemüller, T.: Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography, 36, 27–46, https://doi.org/10.1111/j.1600-0587.2012.07348.x, 2013.

Durbin, J.; Watson, G. S.:Testing for Serial Correlation in Least Squares Regression, I", Biometrika, **37** , 409–428., doi:10.1093/biomet/37.3-4.409. JSTOR 2332391, 1950.

Fathian, F., Mehdizadeh, S., Kozekalani Sales, A., and Safari, M. J. S.: Hybrid models to improve the monthly river flow prediction: Integrating artificial intelligence and non-linear time series models. J. Hydrol., 575, 1200–1213, https://doi.org/10.1016/j.jhydrol.2019.06.025, 2019.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Rowland, J., Michaelsen, J.: A quasi‑global precipitation time series for drought monitoring: CHIRPS 1981–2012, Clim. Data Rec., 35(23), 121–124, https://doi.org/10.1175/JCLI‑D‑13‑00353.1, 2014.

Ghorbani, M. A., Ahmad Zadeh, H., Isazadeh, M., and Terzi, O.: A comparative study of artificial neural network (MLP, RBF) and support vector machine models for river flow prediction. Environ. Earth Sci., 75, 5096, https://doi.org/10.1007/s12665-015-5096-x, 2016.

Gómez, E., Obregón, N., Socarras, V. Aplicación del modelo neurodifuso ANFIS vs redes neuronales, al problema predictivo de caudales medios mensuales del río Bogotá en Villapinzón. Tecnura, vol. 14, núm. 27, pp. 18-29. 2010

Hosseinzadeh, P., Nassar, A., Boubrahimi, S. F., & Hamdi, S. M.. ML-Based Streamflow Prediction in the Upper Colorado River Basin Using Climate Variables Time Series Data. *Hydrology*, *10*(2), 29. https://doi.org/10.3390/hydrology10020029, 2023

Jarque, C.M. and Bera, A.K.: Efficient Tests for Normality, Homoscedasticity and Serial Independence of Regression Residuals. Economics Letters, 6, 255-259, https://doi.org/10.1016/0165-1765(80)90024-5, 1980.

Katrutsa, A., and Strijov, V.: Comprehensive study of feature selection methods to solve multicollinearity problem according to evaluation criteria. Expert Syst. Appl., 76, 1–11, https://doi.org/10.1016/j.eswa.2016.11.005, 2017.

Labat, D., Espinoza, J.C., Ronchail, J., Cochonneau, G., de Oliveira, E., Doudou, J.C., Guyot, J.L.: Fluctuations in the monthly discharge of Guiana Shield rivers related to Pacific and Atlantic climate variability, Hydrol. Sci. J., 57, 1081–1091, https://doi.org/10.1080/02626667.2012.695074, 2012.

Mann, H.B.: Nonparametric tests against trend, Econometrica, 13(3), 245–259, https://doi.org/10.2307/1907187, 1945.

Mesa, O.J., Poveda, G., Hoyos, C.D., Mejía, F., Carvajal, L.F., Cuartas, A., Barco, J.: Predicción no lineal de los caudales medios mensuales del río Nare, Antioquia, Colombia, Av. Recur. Hídricos, 8, 2001.

Moeeni, H., Bonakdari, H.: Forecasting monthly inflow with extreme seasonal variation using the hybrid SARIMA-ANN model, Stoch. Environ. Res. Risk Assess., 31, 1997–2010, https://doi.org/10.1007/s00477-016-1273-z, 2017.

Nieto, R., Gallego, D., Trigo, R., Ribera, P., Gimeno, L.: Dynamic identification of moisture sources in the Orinoco basin in equatorial South America, Hydrol. Sci. J., 53, 602–617, https://doi.org/10.1623/hysj.53.3.602, 2008.

Ocampo-Marulanda, C., Fernández-Álvarez, C., Cerón, W. L., Canchala, T., Carvajal-Escobar, Y., Alfonso-Morales, W.: A spatiotemporal assessment of the high-resolution CHIRPS rainfall dataset in southwestern Colombia using combined principal component analysis. Ain Shams Eng. J., 13, 101739, https://doi.org/10.1016/j.asej.2022.101739, 2022.

Ocampo-Marulanda, C., Canchala Nastar, T., Carvajal Escobar, Y., Cerón-Arámburo, L.: Estimation of monthly flow data with artificial intelligence in basins with scarce information in the Colombian Orinoquía, Rev. Bras. Recur. Hídricos, 30(2), e14, https://doi.org/10.1590/2318-0331.302520240072, 2025.

Pasquini, A.I., Depetris, P.J.: Discharge trends and flow dynamics of South American rivers draining the southern Atlantic seaboard: An overview, J. Hydrol., 333(2–4), 385–399, https://doi.org/10.1016/j.jhydrol.2006.09.005, 2007.

Poveda, G., Jaramillo, A., Gil, M.M., Quiceno, N., Mantilla, R.I.: Seasonality in ENSO-related precipitation, river discharges, soil moisture, and vegetation index in Colombia, Water Resour. Res., 37, 2169–2178, https://doi.org/10.1029/2000WR900395, 2001.

Poveda, G., Mesa, O., Carvajal, L., Hoyos, C., Mejía, J., Cuartas, A., Pulgarin, A.: Predicción de caudales medios mensuales en ríos colombianos usando métodos no lineales, Meteorol. Colomb., 6, 1–8, 2002.

Poveda, G., Álvarez, D.M., Rueda, Ó.A.: Hydro-climatic variability over the Andes of Colombia associated with ENSO: A review of climatic processes and their impact on one of the Earth's most important biodiversity hotspots, Clim. Dyn., 36, 2233–2249, https://doi.org/10.1007/s00382-010-0931-y, 2011.

Ruíz-Ochoa, M.A., Vargas-Corredor, Y.A., Orduz-Amaya, L.P., Torres-Corredor, J.S.: Climate variability in water planning in the Cravo Sur River basin (Casanare, Colombia), Inf. Tecnol., 33(4), 117–124, https://doi.org/10.4067/S0718-07642022000400117, 2022.

Scholz, M., Fraunholz, M., and Selbig, J.: Nonlinear principal component analysis: neural network models and applications. En A. N. Gorban, B. Kégl, D. C. Wunsch, and A. Zinovyev (Eds.), Principal manifolds for data visualization and dimension reduction, Springer, Berlin Heidelberg, 58, 44–67, https://doi.org/10.1007/978-3-540-73750-6_2, 2007.

Sen, P.K.: Estimates of the regression coefficient based on Kendall's tau, J. Am. Stat. Assoc., 63(324), 1379–1389, https://doi.org/10.2307/2285891, 1968.

Silverman, B. W.: Density Estimation for Statistics and Data Analysis. Chapman and Hall/CRC Monographs on Statistics and Applied Probability, 1st edn., Vol. 26, Chapman and Hall, London/New York, 175 pp., https://doi.org/10.1201/9781315140919, 1986

Sirisha, U.M., Belavagi, M.C., Attigeri, G.: Profit prediction using ARIMA, SARIMA and LSTM models in time-series forecasting: A comparison, IEEE Access, 10, 124715–124727, https://doi.org/10.1109/ACCESS.2022.3224938, 2022.

Tootle, G.A., Piechota, T.C., Gutiérrez, F.: The relationships between Pacific and Atlantic Ocean sea-surface temperatures and Colombian streamflow variability, J. Hydrol., 349(3–4), 2008.

Urrea, V., Ochoa, A., Mesa, O.: Validación de la base de datos de precipitación CHIRPS para Colombia a escala diaria, mensual y anual (1981–2014), XXVII Congr. Latinoam. Hidráulica, Lima, Perú, https://www.researchgate.net/publication/310844678, 2016.

Valipour, M.: Long-term runoff study using SARIMA and ARIMA models in the United States, Meteorol. Appl., 22(3), 592–598, https://doi.org/10.1002/met.1491, 2015.

Velásquez, N., Sánchez, J., Carvajal, L.F.: Predicción de caudales medios mensuales en ríos de Colombia usando RNA y MARS, XXIV Congr. Latinoam. Hidráulica, Punta del Este, Uruguay, 2010.

Wong, H., Ip, W.C., Zhang, R.Q., Xia, J. Non-parametric time series models for hydrological forecasting. Journal of Hydrology 332 (3–4), 337–347. 2007

Xiang, X., Guo, S., Cui, Z., Wang, L., Xu, C. Y.: Improving flood forecast accuracy based on explainable convolutional neural network by Grad-CAM method. Journal of Hydrology, 642, 131867, https://doi.org/10.1016/j.jhydrol.2024.131867, 2024.

Yao, L., Libera, D.A., Kheimi, M., Sankarasubramanian, A., Wang, D.: The roles of climate forcing and its variability on streamflow at daily, monthly, annual, and long-term scales, Water Resour. Res., 56(7), https://doi.org/10.1029/2020WR027111, 2020.

Yap, Z.N., Musa, S.: Stream-flow forecasting on Pahang River by time-series models ARMA, ARIMA and SARIMA, Recent Trends Civ. Eng. Built Environ., 4(1), 331–341, https://publisher.uthm.edu.my/periodicals/index.php/rtcebe/article/view/5621, 2023.

Yaseen, Z.M., Jaafar, O., Deo, R.C., Kisi, O., Adamowski, J., Quilty, J. and El-Shafie, A. Stream-flow forecasting using extreme learning machines: a case study in a semi-arid region in Iraq. Journal of Hydrology, 542, pp.603-614, 2016

Wilk,M.B. and Gnanadesikan,R.: Probability plotting methods for the analysis of data. Biometrika, 55, 1–17, https://doi.org/10.1093/biomet/55.1.1, 1968.