REMUNOM
ISSN 2178-6925

Received: 18/05/2025 - **Accepted:** 24/06/2026
Vol: 13.14
DOI: 10.66104/eyzs1565
Pages: 1-37

WHEN THE RIVER PREDICTS ITSELF: LOCAL HYDROLOGICAL PERSISTENCE AND SEASONAL FLOODS IN THE NEGRO RIVER

QUANDO O RIO PREVÊ A SI MESMO: PERSISTÊNCIA HIDROLÓGICA LOCAL E CHEIAS SAZONAIS NO RIO NEGRO

CUANDO EL RÍO SE PREDICE A SÍ MISMO: PERSISTENCIA HIDROLÓGICA LOCAL Y CRECIDAS ESTACIONALES EN EL RÍO NEGRO

Matheus Frota Guimarães de Melo
High School Student, Laviniense, Brazil
E-mail: matheeusfgmelo@gmail.com

Clodoaldo Matias da Silva
Master's Student in Social Anthropology, UFAM, Brazil
E-mail: cms.1978@hotmail.com

Abstract

This research analyses the prediction of seasonal floods in the Negro River using Random Forest and Support Vector Regressor algorithms, considering hydroclimatic variables associated with river levels and ocean atmospheric indices from the Equatorial Pacific. The study aims to evaluate the statistical performance of supervised models applied to Amazonian seasonal hydrological forecasting, using different temporal horizons before the annual flood peak. The methodology uses historical series between 1951 and 2024, fluviometric data from the Port of Manaus, SST and SOI climatic indices, scalar feature selection, leave one out validation, and statistical metrics based on Pearson correlation, MAE, RMSE, and confusion matrices. The results demonstrate that the Random Forest model presents superior performance compared with the Support Vector Regressor in all analysed scenarios, achieving high predictive accuracy even in forecasts produced four months before the flood peak. The research demonstrates that local hydrological variables exert stronger predictive influence on Negro River floods than isolated ocean atmospheric indices, reinforcing artificial intelligence applications focused on Amazonian hydrological monitoring systems.

Keywords: Amazon. Hydrological modelling. Random Forest. Seasonal floods. Supervised learning.

Resumo

Esta pesquisa analisa a previsão das cheias sazonais do Rio Negro por meio dos algoritmos Random Forest e Support Vector Regressor, considerando variáveis hidroclimáticas associadas aos níveis fluviométricos e aos índices oceânico atmosféricos do Pacífico Equatorial. O estudo objetiva avaliar o desempenho estatístico de modelos supervisionados aplicados à previsão hidrológica sazonal amazônica, utilizando diferentes horizontes temporais de antecedência em relação ao pico anual das cheias. A metodologia utiliza séries históricas compreendidas entre 1951 e 2024, dados fluviométricos do Porto de Manaus, índices climáticos SST e SOI, seleção escalar de características, validação leave one out e métricas estatísticas baseadas em coeficiente de Pearson,

REMUNOM
ISSN 2178-6925

Received: 18/05/2025 - **Accepted:** 24/06/2026
Vol: 13.14
DOI: 10.66104/eyzs1565
Pages: 1-37

MAE, RMSE e matrizes de confusão. Os resultados demonstram que o modelo Random Forest apresenta desempenho superior ao Support Vector Regressor em todos os cenários analisados, alcançando elevada precisão preditiva mesmo em previsões realizadas com quatro meses de antecedência. A pesquisa demonstra que variáveis hidrológicas locais apresentam maior influência preditiva sobre as cheias do Rio Negro do que índices oceânico atmosféricos isolados, fortalecendo aplicações de inteligência artificial voltadas ao monitoramento hidrológico amazônico.

Palavras-chave: Amazônia. Aprendizado supervisionado. Cheias sazonais. Modelagem hidrológica. Random Forest.

Resumen

Esta investigación analiza la previsión de las crecidas estacionales del Río Negro mediante los algoritmos Random Forest y Support Vector Regressor, considerando variables hidroclimáticas asociadas a los niveles fluviométricos y a los índices oceánico atmosféricos del Pacífico Ecuatorial. El estudio tiene como objetivo evaluar el desempeño estadístico de modelos supervisados aplicados a la previsión hidrológica estacional amazónica, utilizando diferentes horizontes temporales de anticipación respecto al pico anual de las crecidas. La metodología utiliza series históricas comprendidas entre 1951 y 2024, datos fluviométricos del Puerto de Manaos, índices climáticos SST y SOI, selección escalar de características, validación leave one out y métricas estadísticas basadas en correlación de Pearson, MAE, RMSE y matrices de confusión. Los resultados demuestran que el modelo Random Forest presenta desempeño superior al Support Vector Regressor en todos los escenarios analizados, alcanzando elevada precisión predictiva incluso en previsiones realizadas con cuatro meses de anticipación. La investigación demuestra que las variables hidrológicas locales presentan mayor influencia predictiva sobre las crecidas del Río Negro que los índices oceánico atmosféricos aislados, fortaleciendo aplicaciones de inteligencia artificial dirigidas al monitoreo hidrológico amazónico.

Palabras clave: Amazonia. Aprendizaje supervisado. Crecidas estacionales. Modelación hidrológica. Random Forest.

1. Introduction

Climate change has increased the frequency and intensity of extreme hydrological events in different regions of the planet, causing significant changes in the seasonal behavior of large river systems and raising the levels of socio-environmental vulnerability in urban and riverside areas. In the Amazonian context, the Negro River presents a complex hydrological dynamic due to the size of its drainage basin, the low slope of its channels, and the strong influence exerted by oceanic-atmospheric variables associated with climatic anomalies in the Equatorial Pacific. In this scenario, this research analyzes the seasonal forecasting of the Negro River floods using the Random Forest (RF) and Support Vector Machines (SVM) algorithms, considering different lead-time horizons and hydroclimatic variables related to Amazonian seasonal variability.

REMUNOM
ISSN 2178-6925

Received: 18/05/2025 - **Accepted:** 24/06/2026
Vol: 13.14
DOI: 10.66104/eyzs1565
Pages: 1-37

The academic relevance of this investigation lies in expanding the applications of artificial intelligence in hydrological studies aimed at forecasting extreme events, especially in tropical river systems marked by high statistical nonlinearity and strong temporal variability. Scientifically, the study contributes to the development of predictive models capable of integrating hydrological variables and climate indices into computational architectures directed toward highly complex environmental modeling, broadening the debate on machine learning applied to Amazonian hydrology. Socially, the research assumes strategic importance in view of the impacts produced by major floods on urban infrastructure, mobility, fishing activities, agricultural production, riverside housing, and the planning of preventive actions developed by public agencies responsible for hydrological monitoring.

The general objective of this research is to evaluate the performance of RF and SVM models in the seasonal forecasting of Negro River floods, considering different lead times in relation to the annual flood peak recorded in the city of Manaus. As specific objectives, the study seeks to identify the most relevant hydrological and climatic variables for predictive modeling, analyze the statistical capacity of the algorithms across different time horizons, compare the levels of computational performance between the models used, and examine the operational applicability of forecasts in the context of Amazonian hydrological management. It also seeks to understand how contemporary machine learning techniques can contribute to strengthening seasonal forecasting systems used in the monitoring of extreme hydrological events in the Negro River basin.

In view of the intensification of extreme floods recorded in recent decades in Central Amazonia, the following question is posed: To what extent do Random Forest and SVM models present statistical and computational capacity to forecast, with satisfactory lead time and accuracy, the seasonal flood peaks of the Negro River based on the use of historical hydroclimatic variables? The study is based on the hypothesis that supervised machine learning algorithms, when fed with time series associated with river levels and oceanic-atmospheric climate indices, present high predictive potential for modeling the seasonal behavior of Amazonian floods. It is also considered that the RF model tends to present greater predictive stability

REMUNOM
ISSN 2178-6925

Received: 18/05/2025 - **Accepted:** 24/06/2026
Vol: 13.14
DOI: 10.66104/eyzs1565
Pages: 1-37

when compared to the Support Vector Regressor (SVR), especially in hydrological datasets characterized by high temporal variability and nonlinear relationships among the analyzed variables.

Recent advances in the applications of artificial intelligence in computational hydrology have shown that models based on supervised learning present a high capacity for adaptation to complex environmental systems, especially in scenarios characterized by multiple interdependent climatic variables and strong seasonal oscillation. In the case of the Negro River, the integrated use of oceanic-atmospheric indices and historical records of river level allows for a broader understanding of the relationships between climate variability and hydrological behavior, favoring the development of predictive systems aimed at anticipating extreme events. Thus, the incorporation of techniques such as RF and SVM represents an important advance in the field of environmental modeling applied to Amazonian hydrological dynamics.

In addition to regional climatic complexity, forecasting the floods of the Negro River requires an analytical approach capable of dealing with nonlinear statistical relationships, high temporal variability, and the simultaneous influence of atmospheric, hydrological, and oceanic factors that act upon the seasonal dynamics of the Amazon basin. In this context, machine learning algorithms have become increasingly relevant because they enable a greater capacity for pattern recognition in extensive time series, allowing predictive estimates with different hydrological lead-time horizons. Therefore, the comparative analysis between RF and SVM becomes fundamental for understanding the limits, potentialities, and levels of efficiency of these computational models applied to the seasonal forecasting of Amazonian floods.

Structurally, the article is organized into five interconnected parts, beginning with the present introduction, which is responsible for the climatic, hydrological, and computational contextualization related to the seasonal forecasting of Negro River floods in Central Amazonia. Subsequently, the theoretical framework discusses the impacts of climate change on extreme hydrological events, the mathematical foundations of RF and SVM, as well as the main contemporary approaches to

REMUNOM
ISSN 2178-6925

Received: 18/05/2025 - **Accepted:** 24/06/2026
Vol: 13.14
DOI: 10.66104/eyzs1565
Pages: 1-37

predictive modeling applied to computational hydrology. Later, the results, analytical discussions, conclusion, and references are presented, seeking to contribute to the scientific strengthening of research focused on seasonal hydrological forecasting and the mitigation of the impacts produced by major Amazonian floods.

Based on this analytical structure, the bibliographic development of the study proceeds to examine the hydrological, climatic, and computational foundations that support the application of artificial intelligence techniques in the forecasting of extreme events in large Amazonian river systems, considering the growing environmental complexity associated with contemporary climate change. The relationships between climate variability, hydrological time series, and supervised statistical modeling will also be discussed, emphasizing the role of machine learning algorithms in the construction of predictive systems applied to the monitoring of seasonal floods in the Negro River. In this sense, the following section deepens the theoretical foundations necessary for understanding the climatic variables, computational models, and hydrological parameters used throughout the research.

2. Methodology

2.1 Materials

The predictor variables used in this research consisted of the hydrological levels of the Negro River and climate indices associated with oceanic-atmospheric oscillations in the Equatorial Pacific, considering a continuous time series corresponding to the period from 1951 to 2024 (Figure 1). The records of river gauge levels were obtained from the official database of the Port of Manaus, allowing the construction of a historical dataset representative of the seasonal variations associated with flood events observed over approximately seventy-four years of hydrological monitoring. This aspect is considered fundamental for seasonal forecasting studies in large Amazonian river systems (Port of Manaus, 2024). The selected set of variables was structured based on hydrological and computational approaches applied to flood forecasting, prioritizing climatic and hydrological parameters capable of representing the atmospheric and fluvial

REMUNOM
ISSN 2178-6925

Received: 18/05/2025 - **Accepted:** 24/06/2026
Vol: 13.14
DOI: 10.66104/eyzs1565
Pages: 1-37

dynamics of Central Amazonia, as discussed by Schöngart and Junk (2007) and Marães, Costa, and Costa Filho (2019).

The climate indices employed were Sea Surface Temperature (SST) and the Southern Oscillation Index (SOI), both recognized in the scientific literature for their high capacity to represent atmospheric and oceanic patterns associated with climate anomalies that influence rainfall and hydrological regimes in the Amazon (Schöngart; Junk, 2007). The SST index corresponds to the difference between the temperature observed in a given month and the historical average recorded for the same oceanic region of the Equatorial Pacific, making it possible to identify thermal variations associated with El Niño and La Niña events in different oceanic sectors internationally monitored by the National Centers for Environmental Information (2024). As illustrated in Figure 2, the analyzed regions include the El Niño 1+2, El Niño 3, El Niño 3+4, and El Niño 4 sectors, which present distinct thermal behaviors and high relevance for climate and hydrological forecasting studies on an intercontinental scale.

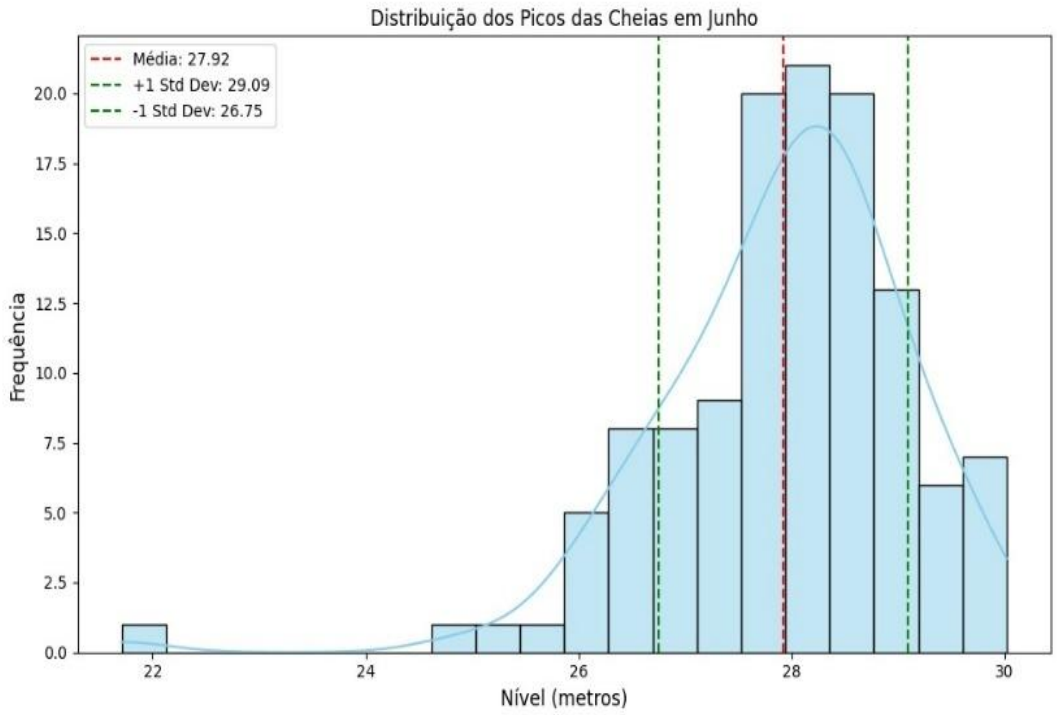


Figure 1. Distribution of the Negro River flood peaks from 1951 to 2024.
Prepared by: Melo (2026).

REMUNOM
ISSN 2178-6925

Received: 18/05/2025 - **Accepted:** 24/06/2026
Vol: 13.14
DOI: 10.66104/eyzs1565
Pages: 1-37

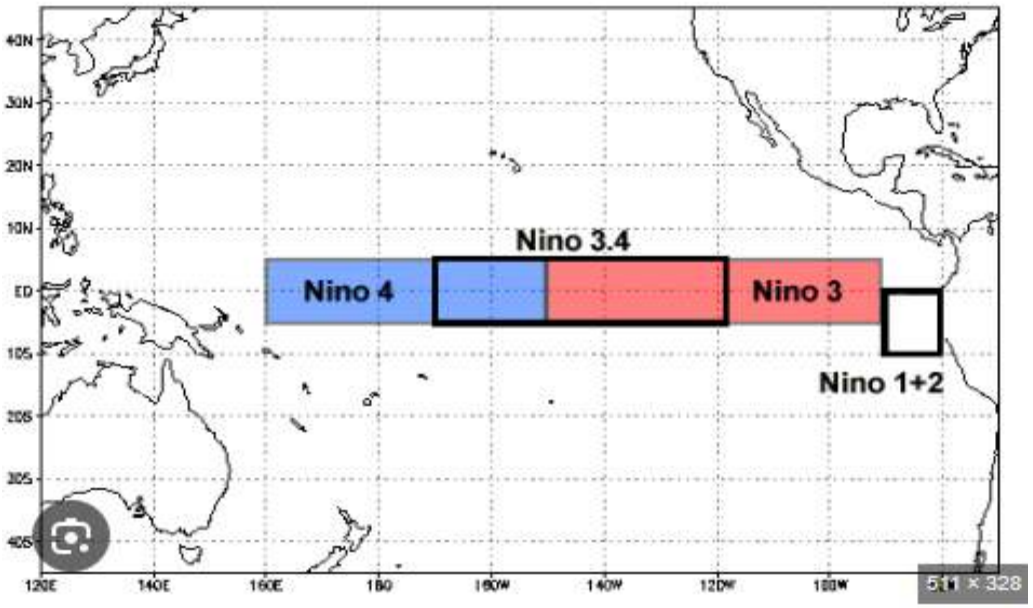


Figure 2. Regions of the Equatorial Pacific where temperature measurements are taken for SST calculation.
Prepared by: Melo (2026).

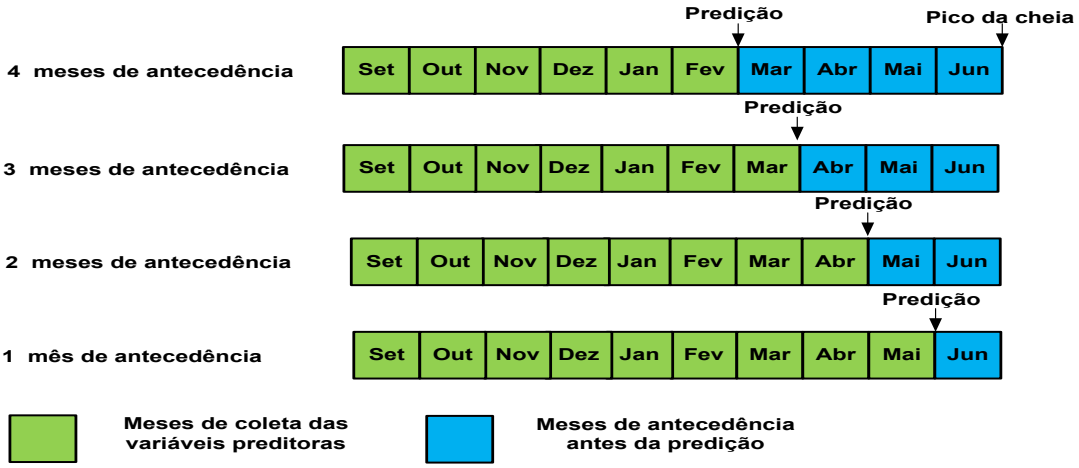


Figure 3. Prediction scheme.
Prepared by: Melo (2026).

The SOI index, in turn, consists of an atmospheric indicator based on the difference in sea-level pressure between the meteorological stations of Tahiti and Darwin, Australia, and is widely used to assess the intensity and variability of atmospheric oscillations related to the ENSO phenomenon. The simultaneous incorporation of the SST and SOI indices made it possible to expand the analytical capacity of the set of predictor variables, considering both oceanic thermal patterns

REMUNOM
ISSN 2178-6925

Received: 18/05/2025 - **Accepted:** 24/06/2026
Vol: 13.14
DOI: 10.66104/eyzs1565
Pages: 1-37

and atmospheric behaviors associated with large-scale climate changes that influence the Amazonian hydrological regime, an aspect also highlighted by Kim et al. (2023) in research focused on artificial intelligence-based hydrological forecasting. Thus, the integrated use of these variables enabled the construction of a consistent hydroclimatic database, suitable for the application of machine learning algorithms aimed at the seasonal forecasting of Negro River floods.

The values corresponding to the climate indices were collected from the official database of the Climate Prediction Center (2024), ensuring statistical standardization and scientific reliability for the time series used in the computational modeling developed throughout the research. Considering that the flood peak of the Negro River occurs predominantly at the end of June, a temporal scheme for extracting climatic and hydrological variables was defined according to different predictive lead-time horizons in relation to the maximum annual flood event, a methodological strategy widely employed in contemporary studies of seasonal hydrological forecasting (Hadi et al., 2024). For forecasts carried out four months in advance, the variables were extracted between September of the previous year and February of the current year, while forecasts with three, two, and one month of lead time progressively used hydroclimatic information up to March, April, and May of the respective year analyzed.

The temporal scheme adopted made it possible to evaluate the gradual influence of incorporating new hydrological and climatic information on the predictive performance of the algorithms used in the research, especially in scenarios of high seasonal variability associated with Amazonian fluvial dynamics. This approach allowed the statistical behavior of the predictor variables to be analyzed at different temporal scales, making it possible to identify correlation patterns between oceanic-atmospheric indices and Negro River levels in the months preceding the annual flood peak, an aspect that is fundamental for machine learning applications in extreme event forecasting. In this context, Figure 3 summarizes the general prediction scheme employed in the study, temporally organizing the variables used in the different seasonal forecasting scenarios analyzed throughout the computational modeling.

REMUNOM
ISSN 2178-6925

Received: 18/05/2025 - **Accepted:** 24/06/2026
Vol: 13.14
DOI: 10.66104/eyzs1565
Pages: 1-37

2.2 Methods

The block diagram presented in Figure 4 describes the methodological steps used in the training and evaluation of the RF and SVR predictive models, employed in the seasonal forecasting of Negro River floods based on previously selected hydroclimatic variables. The methodological structure was organized into sequential stages of normalization, scalar feature selection, supervised training, and statistical validation of the predictors, following approaches widely used in classification and regression problems associated with machine learning (Theodoridis; Koutroumbas, 2009; Vapnik, 2000). In the blue-highlighted blocks in Figure 3, the initial set of variables comprises the SST and SOI climate indices and the hydrological levels corresponding to the months used in each forecasting time horizon analyzed throughout the computational modeling.

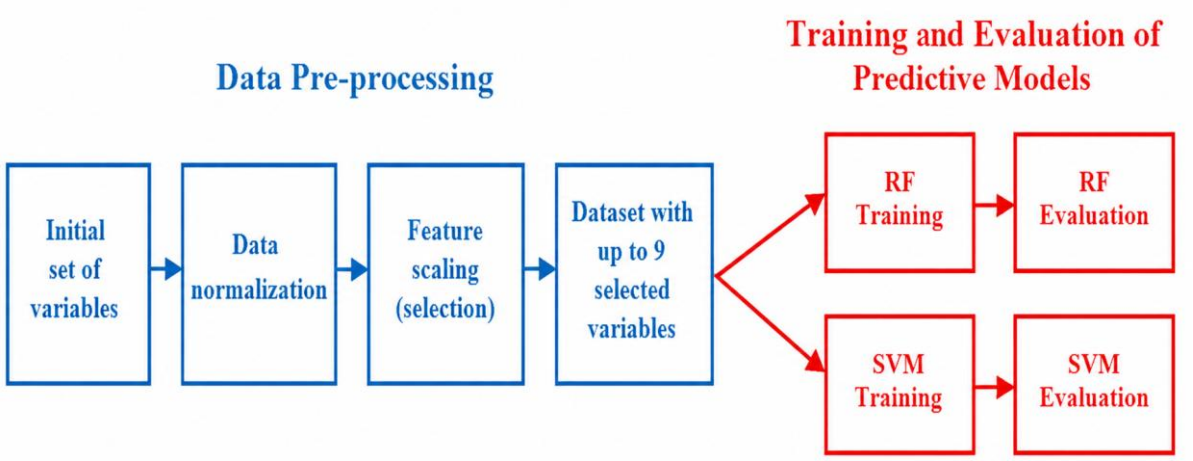


Figure 4. Methodology used for training the predictors.
Prepared by: Melo (2026).

After defining the initial set of variables, the data normalization stage was carried out with the objective of reducing statistical distortions associated with the different numerical scales present in the hydroclimatic series used during the training of the supervised algorithms. Subsequently, the scalar feature selection method modified by Rodrigues, Costa, and Costa Filho (2015) was applied, being responsible for extracting and ranking the nine variables with the greatest statistical

REMUNOM
ISSN 2178-6925

Received: 18/05/2025 - **Accepted:** 24/06/2026
Vol: 13.14
DOI: 10.66104/eyzs1565
Pages: 1-37

relevance for each seasonal forecasting scenario analyzed in the research. As discussed by Theodoridis and Koutroumbas (2009), feature selection methods are highly important in pattern recognition systems, especially in datasets characterized by informational redundancy and high statistical dimensionality.

Order	Variables used for forecasting according to the lead-time period in months			
	4 moths	**3 moths**	**2 moths**	**1 moth**
1st	February river level	March river level	April river level	May river level
2nd	SOI in January	January river level	February river level	March river level
3rd	SOI in September	SOI in November	SOI in January	January river level
4th	January river level	February river level	March river level	April river level
5th	SOI in November	SOI in January	January river level	February river level
6th	SOI in December	El Niño 3+4 in December	El Niño 3+4 in February	El Niño 3+4 in January
7th	El Niño 3+4 in January	SOI in September	SOI in November	SOI in January
8th	El Niño 3+4 in February	El Niño 3+4 in January	SOI in September	SOI in November
9th	El Niño 3+4 in December	El Niño 3+4 in February	El Niño 3+4 in January	SOI in September

Table 1. Set of variables used for training the predictors.
Prepared by: Melo (2026).

Table 1 presents the final set of variables used in the training of the predictors according to the lead-time horizon of the forecast, organized according to the degree of importance established by the scalar selection method applied during the computational processing of the data. A predominance of variables associated with the hydrological levels of the Negro River and with the SOI and El Niño 3+4 indices is observed, indicating a strong relationship between oceanic-atmospheric variability and the seasonal behavior of Amazonian floods, an aspect also identified by Schöngart and Junk (2007) in hydrological studies focused on Central Amazonia. In the red-highlighted blocks in Figure 3, the training, adjustment, and statistical

REMUNOM
ISSN 2178-6925

Received: 18/05/2025 - **Accepted:** 24/06/2026
Vol: 13.14
DOI: 10.66104/eyzs1565
Pages: 1-37

evaluation stages of the RF and SVR models are performed, making it possible to compare the computational performance of the algorithms in different predictive scenarios.

The leave-one-out methodology (Figure 5) was used during the training and validation of the predictive models due to the reduced size of the database, composed of seventy-four annual hydrological samples corresponding to the period from 1951 to 2024. As highlighted by Sonka and Fitzpatrick (2000), this cross-validation strategy presents high statistical efficiency in small datasets, allowing maximum use of the available samples without compromising the generalization capacity of the trained computational models. In this procedure, a given year k is temporarily removed from the set of N observations for training with the remaining N−1 samples, and is subsequently used as a test element to evaluate the predictive performance of the developed model.

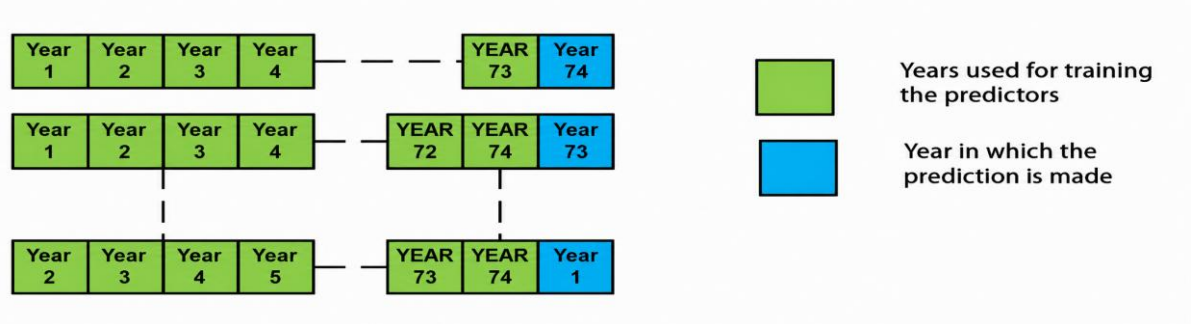


Figure 5. Leave-One-Out methodology for predicting Negro River floods.
Prepared by: Melo (2026).

The successive repetition of this procedure for all years in the historical series made it possible to obtain individual forecasts for each hydrological event recorded throughout the analyzed period, increasing the statistical robustness of the validation process of the RF and SVR models employed in the research. The use of the leave-one-out methodology proved to be particularly suitable for hydrological applications involving reduced time series and high seasonal variability, especially in complex environmental systems associated with Amazonian fluvial dynamics. In this context, Figure 5 illustrates the general operation of the leave-one-out methodology applied to the seasonal forecasting of Negro River floods,

REMUNOM
ISSN 2178-6925

Received: 18/05/2025 - **Accepted:** 24/06/2026
Vol: 13.14
DOI: 10.66104/eyzs1565
Pages: 1-37

summarizing the operational flow adopted during the computational training and validation stages of the predictors used in the study.

2.2.1 Support Vector Regression (SVR)

SVRs are supervised learning models widely used in classification and regression problems due to their high statistical generalization capacity in nonlinear and high-dimensional datasets (Vapnik, 2000). When applied to regression, these computational structures are referred to as SVR and are responsible for predicting continuous values through the construction of mathematical functions capable of simultaneously minimizing predictive error and the structural complexity of the model. As discussed by Theodoridis and Koutroumbas (2009), SVR models present high efficiency in tasks involving the recognition of hydrological and environmental patterns, especially in scenarios characterized by complex temporal variability and nonlinear relationships between predictor variables and response variables.

The fundamental principle of the SVR algorithm consists of determining an optimal hyperplane capable of representing the relationship between input data and predicted values within a tolerance margin previously defined during the supervised training of the model (Vapnik, 2000). Support vectors correspond to the samples positioned close to the limits of this margin, assuming a central role in the process of defining the regression function and adjusting the mathematical structure responsible for the predictive capacity of the algorithm. Depending on the number of predictor variables used in training, the hyperplane may assume different geometric dimensions, being represented as a line in two-dimensional spaces, a plane in three-dimensional spaces, or a hyper structure in multivariate spaces of high computational dimensionality.

The use of kernel functions allows the SVR model to perform nonlinear transformations in the feature space, expanding its capacity to adapt to complex patterns frequently observed in hydrological systems associated with climatic and environmental variability. According to Vapnik (2000), the radial basis function (RBF) kernel presents high efficiency in regression problems involving nonlinear

REMUNOM
ISSN 2178-6925

Received: 18/05/2025 - **Accepted:** 24/06/2026
Vol: 13.14
DOI: 10.66104/eyzs1565
Pages: 1-37

relationships, especially in hydrological forecasting applications and environmental time series modeling. In this context, recent research focused on flood forecasting using machine learning demonstrates that SVR models present satisfactory performance in scenarios characterized by strong hydroclimatic variability and temporal dependence, as observed by Kim et al. (2023) and Hadi et al. (2024).

In the present research, the SVR model was trained using the RBF kernel, degree 3, and regularization parameter C equal to 1.0, a configuration adopted due to the model's capacity to represent nonlinear statistical relationships between climate indices and the hydrological levels of the Negro River. The definition of these parameters sought to optimize the balance between generalization capacity and minimization of predictive error during the supervised training of the algorithm, an aspect considered fundamental in computational applications aimed at the seasonal forecasting of extreme hydrological events. Figure 6 presents the main geometric elements associated with the operation of SVMs in a two-dimensional space, illustrating the relationship between the hyperplane, separation margins, and support vectors used in the computational modeling developed throughout the research.

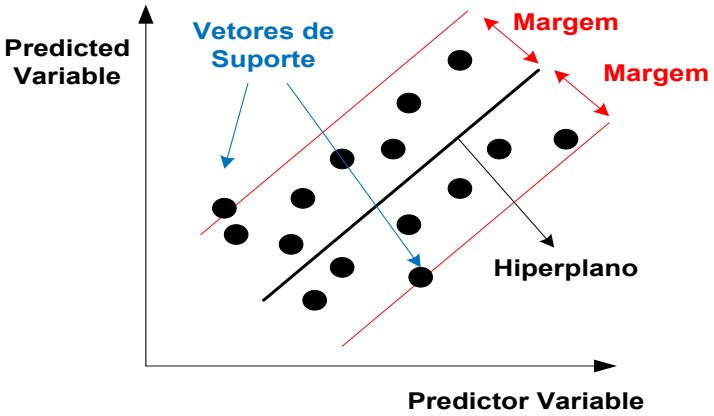


Figure 6. Illustration of a support vector machine.
Prepared by: Melo (2026).

The incorporation of the SVR model in this research is based on the need to employ algorithms capable of dealing with complex and nonlinear statistical relationships between oceanic-atmospheric climate variables and the seasonal

REMUNOM
ISSN 2178-6925

Received: 18/05/2025 - **Accepted:** 24/06/2026
Vol: 13.14
DOI: 10.66104/eyzs1565
Pages: 1-37

hydrological behavior of the Negro River. In addition to their high statistical generalization capacity, SVM-based models present computational robustness in scenarios characterized by reduced databases and high temporal variability, a condition frequently observed in Amazonian hydrological applications (Theodoridis; Koutroumbas, 2009). Thus, the use of SVR made it possible to expand the analytical capacity of the predictive modeling developed in the study, establishing consistent computational foundations for subsequent comparison with the results obtained by the RF algorithm.

2.2.2 Random Forest (RF)

Random Forest (RF) (Figure 7) is a supervised learning model widely used in classification and regression tasks due to its high statistical generalization capacity and computational robustness when dealing with datasets characterized by high variability and nonlinearity (Breiman, 2001). This model is structured through the combination of multiple independent decision trees, constructed from random subsets of data and predictor variables, forming a committee-based learning system capable of reducing problems associated with overfitting and statistical instability present in individual trees. As discussed by Breiman (1996), the use of bagging strategies makes it possible to significantly increase the predictive accuracy of computational models, especially in applications involving high variability in input data.

The operational structure of RF is based on the successive construction of decision trees responsible for producing individual forecasts from different sample subsets randomly obtained during the supervised training of the model. According to Ho (1998), the use of distinct subsets of attributes in each tree promotes greater statistical diversity among classifiers, expanding the algorithm's capacity to recognize complex patterns present in multivariate databases. In this context, the final forecast produced by the forest results from the combination of the individual responses of all constructed trees, generally through the arithmetic mean of the predictive outputs in the case of hydrological regression problems.

REMUNOM
ISSN 2178-6925

Received: 18/05/2025 - **Accepted:** 24/06/2026
Vol: 13.14
DOI: 10.66104/eyzs1565
Pages: 1-37

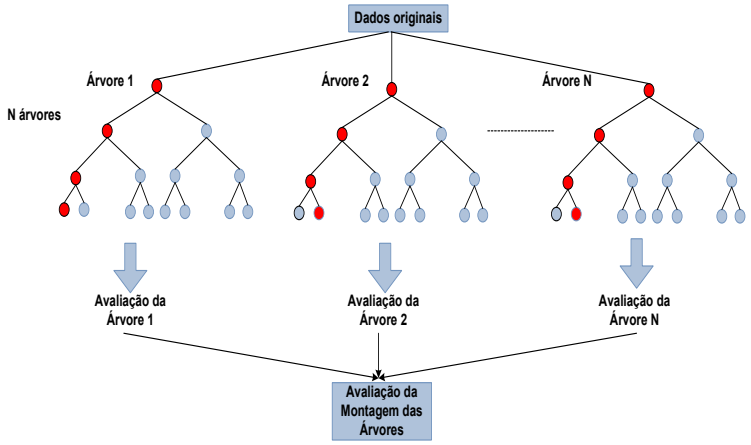


Figure 7. Illustration of a Random Forest.
Prepared by: Melo (2026).

Each decision tree is formed by internal nodes responsible for performing conditional tests associated with the predictor variables used during the computational training of the model. From each decision node, the analytical flow may follow different paths according to the values assumed by the analyzed hydroclimatic variables, allowing the dataset to be segmented into statistically more homogeneous groups during the supervised learning process (Breiman, 2001). In hydrological applications, for example, a node may represent a specific threshold associated with river level or with the behavior of oceanic-atmospheric climate indices, guiding the decision-making process according to hydrological patterns historically observed in the time series used for training the algorithms.

Although individual decision trees are highly sensitive to the samples used in training and to the variables selected during data processing, the use of RF significantly reduces this dependence through the construction of multiple trees based on distinct subsets of data and predictor attributes (Breiman, 2001). As demonstrated by Dietterich (2000), ensemble methods present greater statistical stability and generalization capacity when compared to models based on single classifiers, especially in applications involving high complexity in environmental data. Recent research, such as the study by Hadi et al. (2024) focused on flood forecasting using artificial intelligence, also indicates that RF models present high performance in hydrological scenarios marked by strong temporal variability and nonlinear relationships between climatic and hydrological variables.

REMUNOM
ISSN 2178-6925

Received: 18/05/2025 - **Accepted:** 24/06/2026
Vol: 13.14
DOI: 10.66104/eyzs1565
Pages: 1-37

In the present research, the RF model was trained using one hundred decision trees, a min_samples_split parameter equal to 2, and a min_samples_leaf parameter equal to 1, a configuration adopted due to the algorithm's high capacity to represent complex statistical patterns associated with the seasonal dynamics of Negro River floods. The definition of these hyper parameters sought to optimize the balance between predictive accuracy, computational stability, and the generalization capacity of the models trained throughout the research, especially in hydrological scenarios characterized by high interannual variability. Figure 7 presents a schematic representation of the general operating structure of RF, illustrating the organization of the decision trees used in the computational modeling applied to the seasonal forecasting of Amazonian floods.

The incorporation of the RF model in this research is based on the need to employ algorithms capable of dealing with highly complex statistical relationships between hydroclimatic variables and the seasonal behavior of the Negro River, especially in scenarios influenced by large-scale oceanic-atmospheric oscillations. In addition to its high computational robustness, decision-tree-based models with ensemble learning present an expanded capacity for statistical interpretation of predictor variables, allowing the identification of relevant patterns associated with Amazonian hydrological variability (Breiman, 2001). Thus, the use of RF made it possible to increase the analytical consistency of the predictive modeling developed in the study, establishing fundamental comparative parameters in relation to the performance obtained by the SVR model.

2.3 Evaluation of Predictor Performance

2.3.1 Pearson Correlation Coefficient

The first strategy used to evaluate the statistical performance of the predictive models consisted of applying the Pearson correlation coefficient (rP), widely employed in hydrological studies to measure the degree of linear association between observed values and values predicted by computational models. This dimensionless coefficient presents values ranging from −1 to 1, allowing the

REMUNOM
ISSN 2178-6925

Received: 18/05/2025 - **Accepted:** 24/06/2026
Vol: 13.14
DOI: 10.66104/eyzs1565
Pages: 1-37

simultaneous evaluation of the intensity and direction of the statistical relationship established between the actual flood data and the values estimated by the RF and SVR algorithms used in the research. As discussed by Theodoridis and Koutroumbas (2009), correlation metrics are highly relevant in pattern recognition systems and supervised modeling, especially in applications involving environmental time series and hydroclimatic variability.

Values of rP close to 1 indicate a high positive correlation between observed and predicted data, demonstrating the model's high capacity to reproduce the seasonal behavior of Negro River floods throughout the analyzed historical series. In contrast, values close to −1 indicate a perfect negative correlation between the compared variables, while values close to zero suggest the absence of significant linear dependence between the datasets analyzed during the statistical validation process of the predictors. However, as highlighted by Vapnik (2000), results associated with low levels of linear correlation do not necessarily indicate the absence of a statistical relationship between variables, since nonlinear patterns may exist and can be identified by supervised learning algorithms applied to hydrological modeling.

2.3.2 Mean Absolute Error (MAE) and Root Mean Square Error (RMSE)

In addition to the Pearson correlation coefficient, the Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) metrics were used, both widely employed in the quantitative evaluation of predictive models applied to hydrological forecasting and environmental modeling based on machine learning. MAE corresponds to the mean of the absolute values of the differences between observed and predicted values, allowing the evaluation of the average magnitude of the errors produced by the algorithms used in the research without considering the direction of the statistical discrepancies. RMSE, in turn, is based on the square root of the mean of squared errors, assigning greater statistical penalization to forecasts with large deviations in relation to the actual flood values analyzed.

REMUNOM
ISSN 2178-6925

Received: 18/05/2025 - **Accepted:** 24/06/2026
Vol: 13.14
DOI: 10.66104/eyzs1565
Pages: 1-37

$$MAE = \frac{\sum_{i=1}^{n} |P_i - R_i|}{n} \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} |P_i - R_i|}{n}} \quad (2)$$

Where:
P_i – predicted value for sample i
R_i – actual value for sample i
n – number of samples

Equations 1 and 2 present, respectively, the mathematical formulations used to calculate MAE and RMSE throughout the evaluation stages of the Random Forest and Support Vector Regressor models employed in the research. As discussed by Hadi et al. (2024), metrics based on absolute and squared errors present high efficiency in analyzing the performance of algorithms applied to flood forecasting, especially in scenarios characterized by high seasonal variability and the occurrence of extreme hydrological events. In this context, the simultaneous use of MAE and RMSE made it possible to evaluate more robustly the capacity of the computational models to reproduce the seasonal behavior of Negro River floods across different lead-time horizons.

2.3.3 Flood Classification Ranges

In addition to traditional statistical metrics, a hydrological classification of Negro River floods was proposed based on the historical distribution of the maximum peaks recorded between 1903 and 2024, allowing flood events to be categorized according to different levels of hydrological intensity. The classification procedure considered that the statistical distribution of annual peaks presents an approximate behavior of a normal distribution, whose mean value corresponds to M and whose standard deviation corresponds to σ, a methodology also employed in previous studies focused on the analysis of Amazonian floods. Based on this

REMUNOM
ISSN 2178-6925

Received: 18/05/2025 - **Accepted:** 24/06/2026
Vol: 13.14
DOI: 10.66104/eyzs1565
Pages: 1-37

criterion, four distinct hydrological categories were defined: low flood, medium-low flood, medium-high flood, and high flood, as presented in Table 2.

Range	Classification criterion	Value	Classification
1st	$PC_p < (M-\sigma)$	$PC_p < 26,72m$	Low
2nd	$(M-\sigma) < PC_p < M$	$26,72m < PC_p < 27,87m$	Medium-low
3rd	$M < PC_p < (M+\sigma)$	$27,87m\ PC_p < 29,02m$	Medium-high
4th	$PC_p > (M+\sigma)$	$PC_p > 29,02m$	High

Table 2. Flood ranges and classes according to the mean and standard deviation of flood peaks from 1903 to 2017 (M = 27.87 m and σ = 1.15 m).
Prepared by: Melo (2026).

Events whose maximum values fell within the interval M ± σ were considered medium floods, corresponding to approximately 68.26% of the historical records analyzed throughout the time series used in the research. Within this intermediate class, two additional subclasses were established, named medium-low and medium-high, allowing greater statistical detailing of the seasonal behavior of Negro River floods in relation to the historical variability observed in the river gauge records. Events whose peaks were below (M − σ) were classified as low floods, while values above (M + σ) were classified as high floods, representing approximately 15.87% of the occurrences at each end of the analyzed statistical distribution.

2.3.4 Error Indices

The third strategy used to evaluate the performance of the predictive models consisted of applying the error indices proposed by Schöngart and Junk (2007), allowing the quantitative classification of the degree of accuracy of the forecasts produced by the Random Forest and Support Vector Regressor algorithms. These indices were named Ind_1, Ind_2, Ind_3, and Ind_4, and were structured based on the absolute difference between the observed actual values and the values predicted by the computational models throughout the hydrological series analyzed in the research. As discussed by Kim et al. (2023), metrics based on error intervals have high relevance in hydrological applications because they allow the operational

REMUNOM
ISSN 2178-6925

Received: 18/05/2025 - **Accepted:** 24/06/2026
Vol: 13.14
DOI: 10.66104/eyzs1565
Pages: 1-37

interpretation of the accuracy levels associated with forecasts of extreme events. The error indices, named Ind_1, Ind_2, Ind_3, and Ind_4, are described below:

• Ind_1: number of forecasts in which the absolute value of the difference between the predicted and actual value is less than 0.5 m;

• Ind_2: number of forecasts in which the absolute value of the difference between the predicted and actual value is between 0.5 m and 1 m;

• Ind_3: number of forecasts in which the absolute value of the difference between the predicted and actual value is greater than 1 m and less than 1.5 m;

• Ind_4: number of forecasts in which the absolute value of the difference between the predicted and actual value is greater than 1.5 m.

3. Results

3.1 Statistical Performance of the RF and SVR Predictors

The RF and SVR algorithms were implemented using the Python programming language, with support from the TensorFlow 2.18.0 and Keras 3.8.0 libraries, while training and validation were carried out in a computational environment equipped with 12 GB of RAM and a T4 GPU with 15 GB of dedicated memory. The use of these libraries made it possible to structure supervised models aimed at the seasonal forecasting of Negro River floods across different lead-time horizons, considering hydroclimatic variables associated with Amazonian oceanic-atmospheric dynamics and historical river level records obtained throughout the analyzed period. As discussed by Theodoridis and Koutroumbas (2009), computational environments based on supervised learning present high efficiency in processing multivariate time series associated with complex environmental patterns.

The statistical results obtained for the RF and SVR models were evaluated through the Pearson correlation coefficient, the Ind_1 to Ind_4 error indices, the MAE and RMSE metrics, and the confusion matrices, considering forecasts carried out four, three, two, and one month in advance of the seasonal flood peak of the Negro River. Tables 3 and 4 present the Pearson coefficients and error indices associated with the RF and SVR models, respectively, while Table 5 gathers the

REMUNOM
ISSN 2178-6925

Received: 18/05/2025 - **Accepted:** 24/06/2026
Vol: 13.14
DOI: 10.66104/eyzs1565
Pages: 1-37

mean absolute error and root mean square error values obtained in each seasonal forecasting scenario analyzed throughout the research. As highlighted by Hadi et al. (2024), statistical metrics based on linear correlation and absolute errors have high relevance in the validation of machine learning models applied to flood forecasting and extreme hydrological events.

Months	Variables	Pearson	Ind_1	Ind_2	Ind_3	Ind_4
1	9	0.9578	71	3	1	0
2	9	0.9182	58	16	1	0
3	9	0.8489	49	23	2	1
4	9	0.6764	30	31	12	2

Table 3. Pearson coefficient and error indices for the RF predictor, for 4, 3, 2, and 1 month of lead time.
Prepared by: Melo (2026).

Months	Variables	Pearson	Ind_1	Ind_2	Ind_3	Ind_4
1	9	0.9391	68	4	3	0
2	9	0.8914	58	14	2	1
3	9	0.7451	36	33	3	3
4	9	0.6163	25	36	12	2

Table 4. Pearson coefficient and error indices for the SVR predictor, for 4, 3, 2, and 1 month of lead time.
Prepared by: Melo (2026).

1 month			2 months		
Model	**MAE**	**RMSE**	**Model**	**MAE**	**RMSE**
RF	0.2284	0.2915	RF	0.3073	0.4001
SVR	0.2871	0.4002	SVR	0.3705	0.4722
3 months			**4 months**		
Model	**MAE**	**RMSE**	**Model**	**MAE**	**RMSE**
RF	0.4372	0.5364	RF	0.6318	0.7500
SVR	0.5690	0.6766	SVR	0.6742	0.8010

Table 5. MAE and RMSE errors for the RF and SVR predictors, for 4, 3, 2, and 1 month of lead time.
Prepared by: Melo (2026).

REMUNOM
ISSN 2178-6925

Received: 18/05/2025 - **Accepted:** 24/06/2026
Vol: 13.14
DOI: 10.66104/eyzs1565
Pages: 1-37

Actual values	Confusion matrix for range classification using RF				
		1	**2**	**3**	**4**
	1	6	1	0	0
	2	2	14	3	0
	3	0	2	30	3
	4	0	0	3	10
	Predicted values				

Table 6. Confusion matrix for the RF predictor.
Prepared by: Melo (2026).

Actual values	Confusion matrix for classification into ranges using SVR				
		1	**2**	**3**	**4**
	1	2	5	0	0
	2	1	15	3	0
	3	0	3	29	3
	4	0	0	3	11
	Predicted values				

Table 7. Confusion matrix for the SVR predictor.
Prepared by: Melo (2026).

Tables 6 and 7 present the confusion matrices produced by the RF and SVR models, allowing the evaluation of the predictors' capacity to correctly classify hydrological events according to the flood ranges previously defined in the research. The use of confusion matrices in supervised learning applications makes it possible to interpret not only the quantitative accuracy of the forecasts, but also the operational capacity of the models to recognize patterns associated with different levels of hydrological intensity. As discussed by Kim et al. (2023), classification methods based on artificial intelligence have high importance in hydrological risk monitoring systems, especially in scenarios marked by high seasonal variability and recurrent occurrence of extreme events.

In general, it is observed that both algorithms presented high capacity to reproduce the seasonal behavior of Negro River floods, although the RF model demonstrated superior statistical performance in practically all time horizons analyzed throughout the research. As demonstrated by Breiman (2001), RF-based models present high computational robustness due to the use of multiple decision

REMUNOM
ISSN 2178-6925

Received: 18/05/2025 - **Accepted:** 24/06/2026
Vol: 13.14
DOI: 10.66104/eyzs1565
Pages: 1-37

trees constructed from random subsets of data and predictor variables, significantly reducing problems associated with overfitting and the statistical variability of individual classifiers. This characteristic was also discussed by Dietterich (2000), when highlighting that ensemble methods present greater predictive stability when compared to models based on isolated decision structures.

The statistical superiority of the RF model becomes particularly evident in the results associated with forecasts carried out one month in advance, in which the Pearson correlation coefficient reached 0.9578, while the SVR model presented a coefficient of 0.9391 for the same forecasting time horizon. Also according to Tables 3 and 4, the Ind_1 index, responsible for counting the number of forecasts whose absolute error remained below 0.5 meter, reached seventy-one occurrences in the RF model, while the SVR model presented fifty-eight occurrences under the same statistical condition. These results corroborate the discussions presented by Breiman (1996) and Ho (1998), according to which methods based on ensemble learning and random subspace present greater capacity for statistical generalization in problems characterized by high variability and nonlinear dependence among environmental variables.

The results presented in Table 5 also demonstrate that the RF model presented lower MAE and RMSE values when compared to SVR, especially in the four-month-ahead forecasting scenarios, indicating greater quantitative accuracy in the hydrological estimates produced by the decision-tree-based algorithm. For the four-month horizon, the RF MAE and RMSE errors were 0.6318 and 0.7500 meters, respectively, while the SVR model presented values of 0.6742 and 0.8010 meters for the same statistical metrics. As discussed by Vapnik (2000), SVR models present high generalization capacity in nonlinear scenarios; however, they may present limitations associated with the reduced size of the available samples and the high statistical complexity of the environmental data used during supervised training.

The results obtained in this research also demonstrate advances in relation to previous studies developed for the seasonal forecasting of Negro River floods, especially when compared to the results presented by Marães, Costa, and Costa

REMUNOM
ISSN 2178-6925

Received: 18/05/2025 - **Accepted:** 24/06/2026
Vol: 13.14
DOI: 10.66104/eyzs1565
Pages: 1-37

Filho (2019), who used artificial neural networks associated with decision trees and committee-based learning. In that study, forecasts carried out one month in advance presented errors below 0.5 meter in approximately 85.07% of the analyzed occurrences, while in the present research the RF model reached approximately 94.67% under the same statistical condition. These results indicate that the use of ensemble methods based on Random Forest expanded the predictive capacity of computational models applied to the seasonal forecasting of Amazonian floods.

The analysis of the correlation coefficients presented in Figure 6 demonstrates that the variables associated with the hydrological levels of the Negro River in the months preceding the flood peak presented the highest Pearson correlation values in relation to the other hydroclimatic predictors used in the research. This behavior was also identified by the scalar feature selection technique applied during the methodological stage, which primarily selected river levels as the main predictor variables in the different seasonal forecasting scenarios analyzed throughout the computational modeling. As discussed by Schapire et al. (1998) and Freund and Schapire (1996), supervised algorithms based on statistical learning tend to assign greater relevance to variables with higher discriminatory capacity in the recognition of complex patterns associated with the training sets used during the computational modeling process.

3.2 Evaluation of Error Indices, MAE, RMSE, and Flood Classification

The quantitative analysis of the error indices demonstrates that the RF model presented superior statistical performance compared to SVR across all forecasting time horizons evaluated throughout the research, especially in scenarios with one and two months of lead time in relation to the seasonal flood peak of the Negro River. The Ind_1 index, responsible for accounting for forecasts whose absolute error remained below 0.5 meter, presented the best results for the RF model, reaching seventy-one occurrences in the one-month-ahead forecasting scenario, while the SVR model presented fifty-eight occurrences under the same statistical condition. As discussed by Hadi et al. (2024), error indices associated with small

REMUNOM
ISSN 2178-6925

Received: 18/05/2025 - **Accepted:** 24/06/2026
Vol: 13.14
DOI: 10.66104/eyzs1565
Pages: 1-37

absolute discrepancies have high operational relevance in hydrological forecasting systems because they allow direct assessment of the practical accuracy level of computational models used in environmental applications.

From an operational perspective, errors below 0.5 meter represent an important reliability indicator for hydrological monitoring systems applied to the seasonal forecasting of Amazonian floods, especially in regions marked by high socio-environmental vulnerability and dependence on fluvial dynamics. The capacity of the RF model to maintain a high number of forecasts within this error range demonstrates significant statistical stability even in scenarios characterized by high interannual hydroclimatic variability. As observed by Kim et al. (2023), supervised learning models that present low absolute errors in the forecasting of hydrological levels have high applicability in risk management systems and preventive monitoring of extreme events.

The results associated with the MAE and RMSE metrics reinforce the statistical superiority observed in the RF model when compared to SVR, especially in forecasts performed four months in advance. As presented in Table 5, the RF model showed MAE and RMSE values equal to 0.6318 and 0.7500 meters, respectively, while the SVR model presented errors of 0.6742 and 0.8010 meters for the same statistical metrics. In hydrological terms, these results indicate that the RF algorithm was able to forecast the seasonal behavior of Negro River floods with a mean error below one meter even in forecasts carried out with considerable temporal lead time, an aspect considered highly relevant for preventive applications associated with the management of extreme hydrological events.

The possibility of forecasting the seasonal flood peak approximately four months in advance represents a significant advance in relation to traditional hydrological approaches previously used in Central Amazonia. Studies developed by Rodrigues, Costa, and Costa Filho (2015) and Schöngart and Junk (2007) performed forecasts predominantly concentrated on shorter time horizons, especially in the context of the relationships between oceanic-atmospheric variability and Amazonian hydrological behavior. In this sense, the use of RF and SVR models made it possible to expand the anticipation capacity of hydrological

REMUNOM
ISSN 2178-6925

Received: 18/05/2025 - **Accepted:** 24/06/2026
Vol: 13.14
DOI: 10.66104/eyzs1565
Pages: 1-37

forecasts, favoring operational applications related to the mitigation of environmental, social, and economic impacts produced by major Negro River floods.

The confusion matrices presented in Tables 6 and 7 demonstrate that the computational models used in the research presented high efficiency in classifying floods according to the hydrological ranges previously defined in the study. The use of classification categories based on the mean and standard deviation of historical peaks made it possible to transform continuous numerical forecasts into hydrological classes that are more operationally interpretable for environmental monitoring systems and governmental decision-making. As discussed by Marães, Costa, and Costa Filho (2019), classification systems associated with seasonal forecasting of Amazonian floods significantly expand the practical applicability of computational models in contexts of territorial management and preventive planning.

The classification of floods into low, medium-low, medium-high, and high ranges made it possible to interpret the predictive results in a more accessible way for public agencies responsible for hydrological monitoring of the Amazon basin. Instead of simply disseminating absolute numerical values related to the flood peak, the models developed in this research allow hydrological events to be categorized according to intensity levels associated with potential territorial, economic, and social impacts on riverside populations and vulnerable urban areas. This approach has high operational relevance for institutions such as Civil Defense, state environmental monitoring systems, and urban planning agencies associated with the management of extreme events in the Amazon.

The results obtained also demonstrate advances in relation to previous research developed for the seasonal forecasting of Negro River floods using artificial intelligence applied to Amazonian hydrology. In the study conducted by Marães, Costa, and Costa Filho (2019), approximately 85.07% of the forecasts performed one month in advance presented errors below 0.5 meter, while in the present research the RF model reached an approximate percentage of 94.67% under the same statistical condition. These results show that the use of methods

REMUNOM
ISSN 2178-6925

Received: 18/05/2025 - **Accepted:** 24/06/2026
Vol: 13.14
DOI: 10.66104/eyzs1565
Pages: 1-37

based on decision trees with ensemble learning significantly expanded the predictive capacity of the hydrological models used in the seasonal monitoring of Amazonian floods.

An important limitation observed during the development of the research is related to the reduced size of the database available for training, validation, and testing of the supervised models employed in the seasonal forecasting of Negro River floods. Despite the use of historical series from 1951 to 2024, the total set of samples remained relatively limited when compared to the statistical requirements normally associated with the training of supervised learning algorithms in complex environmental applications. Due to this limitation, the leave-one-out methodology was adopted, making it possible to maximize the statistical use of the hydrological observations available throughout the analyzed historical series.

The use of the leave-one-out method proved to be particularly suitable for hydrological applications characterized by small sample size and high temporal variability of environmental observations used during the supervised training of computational models. As highlighted by Sonka and Fitzpatrick (2000), cross-validation techniques based on leave-one-out make it possible to increase the statistical robustness of predictive systems by maximizing the use of available samples during the training and validation stages. In this context, the adoption of this methodological strategy contributed to increasing the statistical generalization capacity of the RF and SVR models applied to the seasonal forecasting of Negro River floods.

3.3 Hydroclimatic Variability and Interpretation of Seasonal Forecasts

The analysis of the Pearson correlation coefficients presented in Figure 8 demonstrates that the variables associated with the hydrological levels of the Negro River in the months preceding the seasonal flood peak presented the highest levels of statistical association with the maximum annual values recorded throughout the period from 1951 to 2024. This behavior indicates a strong temporal dependence between the river gauge levels observed in the months immediately before the flood

REMUNOM
ISSN 2178-6925

Received: 18/05/2025 - **Accepted:** 24/06/2026
Vol: 13.14
DOI: 10.66104/eyzs1565
Pages: 1-37

and the maximum hydrological behavior subsequently recorded at the Port of Manaus. According to the historical records made available by the Port of Manaus (2024) and previously analyzed by Nunes de Mello and Barros (2001), the Negro River presents a relatively regular seasonal behavior in hydrological terms, although it is subject to intensifications associated with interannual climate variability.

The results presented in Figure 8 demonstrate that river levels in the months preceding the flood occupied the first positions in statistical importance during the scalar feature selection process used in the computational modeling. This behavior indicates that local hydrological dynamics exert a more significant predictive influence on the seasonal flood peak than part of the oceanic-atmospheric indices used in the research. The first climate variable associated with the ENSO phenomenon with the highest Pearson coefficient was the SOI-JAN index, which occupied only the sixth position of statistical relevance among all variables analyzed during the training of the supervised models applied to the seasonal forecasting of the Negro River.

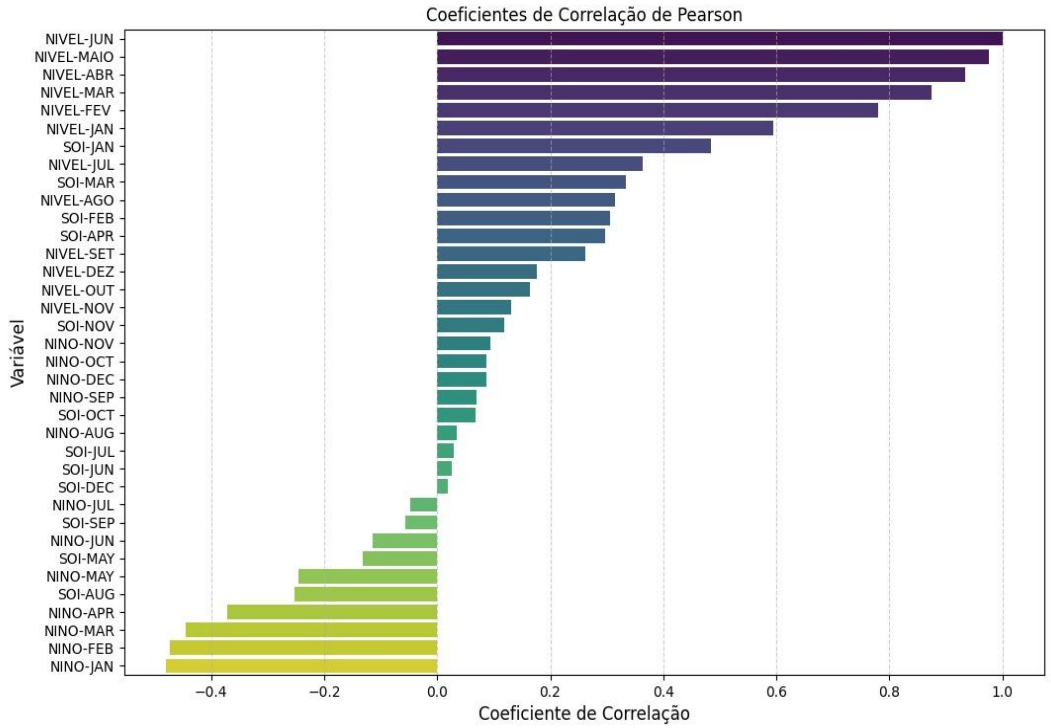


Figura 8. Pearson correlation coefficients between each variable shown in Table 2 and the flood peak value, for the period from 1951 to 2024.

REMUNOM
ISSN 2178-6925

Received: 18/05/2025 - **Accepted:** 24/06/2026
Vol: 13.14
DOI: 10.66104/eyzs1565
Pages: 1-37

Prepared by: Melo (2026).

Although the oceanic-atmospheric indices presented lower correlation coefficients than the local hydrological levels, the results demonstrate that these variables continue to exert significant influence on the seasonal behavior of Amazonian floods at interannual climate scales. As discussed by Schöngart and Junk (2007), anomalies associated with the ENSO phenomenon directly affect rainfall regimes in Central Amazonia, altering precipitation patterns, river discharge, and the intensity of seasonal floods observed in the Negro River basin. In this context, the SST and SOI indices used in the research made it possible to incorporate into the computational models information related to the oceanic-atmospheric variability of the Equatorial Pacific, expanding the hydroclimatic interpretation capacity of the predictive scenarios analyzed throughout the study.

The oceanic regions used to calculate SST thermal anomalies correspond to the El Niño 1+2, El Niño 3, El Niño 3+4, and El Niño 4 sectors, which are continuously monitored by international institutions specialized in climatology and ocean monitoring. According to the National Centers for Environmental Information (2024), these regions are highly relevant for identifying patterns associated with El Niño and La Niña events, which directly influence the atmospheric systems responsible for rainfall distribution in different sectors of the Amazon. The climate indices used during the training of the RF and SVR models were obtained from the official database of the Climate Prediction Center (2024), ensuring statistical standardization and scientific reliability for the oceanic-atmospheric variables incorporated into the hydroclimatic database developed in the research.

The hydroclimatic interpretation of the results demonstrates that the seasonal variability of Negro River floods does not depend exclusively on global oceanic climate conditions, but also on the hydrological persistence of the Amazonian river system itself over the months preceding the annual flood peak. The high correlation observed between river levels in previous months and the maximum seasonal peak indicates the cumulative behavior of regional hydrological dynamics, strongly influenced by the large territorial extension of the Amazon basin and the low slope of the river channels associated with the Negro River. This behavior favors slower and relatively more predictable hydrological responses when compared to smaller

REMUNOM
ISSN 2178-6925

Received: 18/05/2025 - **Accepted:** 24/06/2026
Vol: 13.14
DOI: 10.66104/eyzs1565
Pages: 1-37

hydrographic systems subject to rapid responses associated with extreme precipitation events.

The results obtained also show that the computational models used in the research presented high capacity to recognize hydroclimatic patterns associated with Amazonian seasonal variability, even in view of the reduced size of the historical database used during supervised training. The simultaneous use of local hydrological variables and oceanic-atmospheric climate indices made it possible to expand the analytical robustness of the forecasts produced by the RF and SVR algorithms, favoring more consistent interpretations of the factors associated with the seasonal behavior of Amazonian floods. As discussed by Silva (2025), extreme hydrological events in the Amazon have shown increasing climatic complexity in recent decades, especially due to the intensification of atmospheric oscillations and the expansion of regional socio-environmental vulnerabilities.

The high accuracy obtained in forecasts carried out up to four months in advance has important territorial and socioeconomic relevance for populations located in the Negro River basin, especially in riverside communities strongly dependent on Amazonian fluvial dynamics. Major floods directly affect fishing activities, agriculture, fluvial mobility, and urban occupation in vulnerable areas of Manaus and municipalities in the interior of Amazonas, increasing economic impacts associated with the extreme seasonality of the regional hydrological system. As discussed by Melo et al. (2026), extreme flood events produce significant effects on the Amazonian informal economy, especially in contexts marked by strong territorial dependence on seasonal fluvial dynamics.

The integrated interpretation of the statistical and hydroclimatic results obtained in this research demonstrates that the Random Forest and Support Vector Regressor models presented high capacity to represent the seasonal dynamics of Negro River floods, especially when fed with local hydrological variables associated with the months preceding the annual flood peak. In addition to expanding predictive capacity in relation to previous studies developed in the Amazonian context, the results highlight the importance of integrating computational hydrology, supervised learning, and oceanic-atmospheric climate variability in the construction of predictive

REMUNOM
ISSN 2178-6925

Received: 18/05/2025 - **Accepted:** 24/06/2026
Vol: 13.14
DOI: 10.66104/eyzs1565
Pages: 1-37

systems applied to environmental monitoring in the Amazon. In this context, the results obtained establish important analytical foundations for the final discussion regarding the potentialities, limitations, and operational applications of the models developed throughout the research.

4. Discussions

The results obtained throughout the research show that the RF model presented superior statistical performance compared to SVR in all analyzed time horizons, especially in forecasts carried out one and two months in advance of the seasonal flood peak of the Negro River. This superiority became more evident in the Ind_1 indices, Pearson coefficients, and MAE and RMSE values, demonstrating greater statistical stability and generalization capacity of the model based on decision trees with ensemble learning. As discussed by Breiman (2001) and Dietterich (2000), ensemble models present high computational robustness in scenarios characterized by nonlinear relationships and high statistical variability, a condition strongly observed in the Amazonian hydroclimatic series used during the supervised training of the predictors.

The interpretation of the results also demonstrates that the hydrological levels of the Negro River in the months preceding the seasonal peak exerted a more significant predictive influence than the oceanic-atmospheric indices associated with the ENSO phenomenon. Although the SST and SOI indices contributed to expanding the analytical capacity of the computational models, the highest correlation coefficients observed in Figure 8 remained associated with local hydrological variables, indicating strong temporal persistence in the seasonal behavior of the Amazonian river system. This behavior corroborates the discussions presented by Schöngart and Junk (2007), according to whom the dynamics of Amazonian floods are strongly dependent on hydrological conditions accumulated over the months preceding the annual peak, especially in basins with large territorial extension and low slope, such as the Negro River basin.

REMUNOM
ISSN 2178-6925

Received: 18/05/2025 - **Accepted:** 24/06/2026
Vol: 13.14
DOI: 10.66104/eyzs1565
Pages: 1-37

When comparing the results obtained in this research with previous studies developed in the Amazonian context, a relevant advance is observed in the predictive capacity of supervised models applied to the seasonal forecasting of Negro River floods. Marães, Costa, and Costa Filho (2019), using artificial neural networks associated with decision trees, obtained lower percentages than those observed in the present study for forecasts with errors below 0.5 meter, while Rodrigues, Costa, and Costa Filho (2015) had already highlighted the potential of integrating hydroclimatic variables and artificial intelligence in Amazonian seasonal forecasting systems. In this context, the results presented in this research demonstrate that the incorporation of RF and SVR models made it possible to increase the statistical accuracy of hydrological forecasts even in scenarios characterized by reduced sample size and high climatic complexity.

Despite the high statistical efficiency observed in the developed models, some methodological limitations remain associated with the reduced size of the historical database available for supervised training and validation of the algorithms used throughout the research. Although the time series from 1951 to 2024 represents an important historical set of Amazonian hydrological observations, the total number of samples still remains relatively limited when compared to the statistical requirements normally associated with advanced applications of machine learning. As discussed by Theodoridis and Koutroumbas (2009) and Vapnik (2000), supervised models applied to complex environmental time series may present statistical sensitivity when using small sample bases, an aspect partially minimized in this research through the use of the leave-one-out methodology during the computational training and validation stages.

In addition to the methodological and computational contributions, the results obtained present important territorial and socioeconomic relevance for populations located in the Negro River basin, especially in contexts marked by high socio-environmental vulnerability and strong dependence on Amazonian fluvial dynamics. The possibility of forecasting the seasonal behavior of floods up to four months in advance significantly expands the operational capacity of public agencies responsible for hydrological monitoring, urban planning, and mitigation of impacts

REMUNOM
ISSN 2178-6925

Received: 18/05/2025 - **Accepted:** 24/06/2026
Vol: 13.14
DOI: 10.66104/eyzs1565
Pages: 1-37

associated with extreme events in the Amazon. As observed by Melo et al. (2026), major floods produce direct impacts on economic activities, fluvial mobility, supply systems, and territorial occupation in Amazonian communities dependent on regional water seasonality.

In general, the results presented demonstrate that the integration of computational hydrology, supervised learning, and oceanic-atmospheric hydroclimatic variability constitutes a promising approach for the development of predictive systems applied to Amazonian environmental monitoring. In addition to expanding the statistical capacity of seasonal forecasts for the Negro River in relation to previously developed studies, the RF and SVR models demonstrated high potential for future applications involving recurrent neural networks, sequential processing of time series, and automated hydrological warning systems. In this context, the research contributes to strengthening the applications of artificial intelligence in Amazonian hydrological modeling, establishing important analytical foundations for future investigations aimed at forecasting extreme events and territorial management in scenarios of increasing climate variability.

5. Final Considerations

The results obtained throughout the research demonstrated that the RF and SVR models presented high predictive capacity in estimating the seasonal floods of the Negro River, especially when fed with hydroclimatic variables associated with river gauge levels preceding the annual flood peak. The statistical analysis carried out through Pearson coefficients, error indices, and MAE and RMSE metrics showed superior performance of the Random Forest model across all evaluated time horizons, especially in forecasts performed one and two months in advance. In this context, the results confirm that supervised algorithms based on ensemble learning have high efficiency in the seasonal hydrological modeling of complex river systems associated with Amazonian dynamics.

The initially established hypothesis was confirmed by demonstrating that computational models based on artificial intelligence present sufficient statistical

REMUNOM
ISSN 2178-6925

Received: 18/05/2025 - **Accepted:** 24/06/2026
Vol: 13.14
DOI: 10.66104/eyzs1565
Pages: 1-37

capacity to forecast the seasonal behavior of Negro River floods with high accuracy and significant lead time. The results showed that the RF model was able to produce forecasts with a mean error below one meter even in scenarios with four months of lead time, a condition considered operationally relevant for preventive applications associated with Amazonian hydrological monitoring. Thus, the research satisfactorily answered the central question of the study by showing that local hydroclimatic variables and oceanic-atmospheric indices can be efficiently integrated into computational systems aimed at the seasonal forecasting of Amazonian floods.

The results also demonstrated that the hydrological levels observed in the months preceding the seasonal peak exerted greater predictive influence on the computational models than the climate indices associated with the ENSO phenomenon. This behavior indicates strong temporal persistence in the hydrological dynamics of the Negro River and reinforces the importance of local variables in the construction of predictive systems applied to Amazonian seasonal modeling. Although the SST and SOI indices showed secondary statistical relevance when compared to river gauge variables, their incorporation into the models contributed to expanding the hydroclimatic robustness of the forecasts and strengthening the analytical capacity associated with regional oceanic-atmospheric variability.

The classification of floods into hydrological ranges constituted an important methodological contribution of the research by allowing predictive results to be interpreted in a more operational and territorially applicable manner by public agencies responsible for Amazonian environmental monitoring. The use of confusion matrices associated with the low, medium-low, medium-high, and high categories made it possible to broaden the practical interpretation of the forecasted hydrological scenarios, favoring applications directed toward Civil Defense, urban planning, fluvial logistics, and mitigation of socioeconomic impacts associated with major Amazonian floods. In this sense, the research expands the possibilities of integration between computational hydrology, data science, and territorial management applied to the Amazonian context.

REMUNOM
ISSN 2178-6925

Received: 18/05/2025 - **Accepted:** 24/06/2026
Vol: 13.14
DOI: 10.66104/eyzs1565
Pages: 1-37

From a theoretical and computational perspective, the research contributes to strengthening the applications of supervised learning in complex environmental systems characterized by high temporal variability and reduced sample size. The integrated use of scalar feature selection techniques, leave-one-out validation, and supervised algorithms made it possible to structure statistically consistent predictive models even in view of the limitations inherent to the historical hydrological series available for the Negro River. In addition, the results obtained reinforce the potential of artificial intelligence applications as strategic tools for the development of automated hydrological forecasting systems in regions subject to the intensification of extreme climate events.

Finally, it is recommended that future research expand the use of sequential computational architectures based on recurrent neural networks, hybrid models, and bidirectional systems capable of incorporating more complex temporal dependencies associated with Amazonian hydroclimatic dynamics. It is also recommended to expand the historical databases used in the supervised training of the models, as well as to incorporate new environmental variables related to regional precipitation, evapotranspiration, and large-scale atmospheric circulation. In this way, it will be possible to progressively increase the predictive capacity of computational hydrological systems applied to the Negro River and strengthen territorial adaptation strategies in view of the increasing climate variability observed in the Amazon.

6. Referências

BREIMAN, L. *Bagging predictors*. **Machine Learning**, New York, v. 24, n. 2, p. 123-140, 1996.

BREIMAN, L. *Random forests*. **Machine Learning**, New York, v. 45, n. 1, p. 5-32, 2001.

CLIMATE PREDICTION CENTER. **NOAA Center for Weather and Climate Prediction**. Washington, 2024. Disponível em: <Climate Prediction Center>. Acesso em: 8 jul. 2025.

REMUNOM
ISSN 2178-6925

Received: 18/05/2025 - **Accepted:** 24/06/2026
Vol: 13.14
DOI: 10.66104/eyzs1565
Pages: 1-37

DIETTERICH, T. G. *An experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting and randomization*. **Machine Learning**, New York, v. 40, n. 2, p. 139-157, 2000.

FREUND, Y.; SCHAPIRE, R. E. ***Experiments with a new boosting algorithm***. In: International Conference on Machine Learning, 13., 1996, Bari. Proceedings [...] Bari: ICML, 1996.

HADI, F. A. A. et al. *Machine learning techniques for flood forecasting*. **Journal of Hydroinformatics**, London, v. 26, n. 4, p. 779-799, 2024. DOI: 10.2166/hydro.2024.208.

HO, T. K. *The random subspace method for constructing decision forests*. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, New York, v. 20, n. 8, p. 832-844, 1998.

KIM, D. et al. *Application of AI-based models for flood water level forecasting and flood risk classification*. ***KSCE Journal of Civil Engineering*, Seoul**, v. 27, n. 7, p. 3163-3177, 2023.

MARÃES, P.; COSTA, M. G. F.; COSTA FILHO, C. F. F. **Previsão de cheias sazonais para o Rio Negro usando redes neurais artificiais e árvores de decisão com aprendizado baseado em comitê**. *In*: Workshop de Computação Aplicada à Gestão do Meio Ambiente e Recursos Naturais (WCAMA), 2019, Belém. *Anais [...]* Belém: SBC, 2019.

MELO, M. F. G.; SILVA, C. M.; LOPES, F. H. P.; AGUIAR, D. M. *Floods as economic shocks: a theoretical analysis of the impacts of Rio Negro flooding on the Amazonian informal economy*. **REMUNOM**, *[S. l.]*, v. 13, n. 5, p. 1-19, 2026.

NATIONAL CENTERS FOR ENVIRONMENTAL INFORMATION. **ENSO Sea Surface Temperature Regions**. Washington, 2024. Disponível em: <National Centers for Environmental Information>. Acesso em: 23 jul. 2025.

NUNES DE MELLO, J. A.; BARROS, W. G. Enchentes e vazantes do Rio Negro medidas no Porto de Manaus, Amazonas, Brasil. **Acta Amazônica**, Manaus, v. 31, n. 2, p. 343-352, 2001.

PORTO DE MANAUS. Níveis máximo e mínimo do Rio Negro. Manaus, 2024. Disponível em: <Porto de Manaus>. Acesso em: 8 jul. 2025.

RODRIGUES, M.; COSTA, M.; COSTA FILHO, C.F.F. **Proposta de um método para previsão de cheias sazonais utilizando redes neurais artificiais**: uma aplicação ao Rio Amazonas. *In*: Workshop de Computação Aplicada à Gestão do Meio Ambiente e Recursos Naturais (WCAMA), 2015, Recife. *Anais [...]* Recife: SBC, 2015.

REMUNOM
ISSN 2178-6925

Received: 18/05/2025 - **Accepted:** 24/06/2026
Vol: 13.14
DOI: 10.66104/eyzs1565
Pages: 1-37

SCHAPIRE, R. E.; FREUND, Y.; BARTLETT, P.; LEE, W. S. *Boosting the margin: a new explanation for the effectiveness of voting methods*. **The Annals of Statistics**, Hayward, v. 26, n. 5, p. 1651-1686, 1998.

SCHÖNGART, J.; JUNK, W. J. *Forecasting the flood-pulse in Central Amazonia by ENSO-indices*. **Journal of Hydrology**, Amsterdam, v. 335, n. 1-2, p. 124-132, 2007. DOI: 10.1016/j.jhydrol.2006.11.005.

SILVA, C. M. da. Entre rios secos e comunidades resistentes: as estiagens no Amazonas como desafio climático e territorial. **Revista Ibero-Americana de Humanidades**, *Ciências e Educação*, [S. l.], v. 11, n. 9, p. 3900-3916, 2025.

SONKA, M.; FITZPATRICK, J. M. **Medical image processing and analysis**. Washington: SPIE Press, 2000.

THEODORIDIS, S.; KOUTROUMBAS, K. **Pattern recognition**. 4. ed. Amsterdam: Elsevier, 2009.

VAPNIK, V. N. **The nature of statistical learning theory**. 2. ed. New York: Springer, 2000.