





<https://doi.org/10.1590/2318-0331.292420230105>

## Combining traditional hydrological models and machine learning for streamflow prediction

*Combinando modelos hidrológicos tradicionais e aprendizado de máquina para previsão de vazão*

Antonio Duarte Marcos Junior<sup>1</sup> , Cleiton da Silva Silveira<sup>2</sup> , José Micael Ferreira da Costa<sup>3</sup>  & Suellen Teixeira Nobre Gonçalves<sup>2</sup> 

<sup>1</sup> Universidade Federal do Ceará, Redenção, CE, Brasil

<sup>2</sup> Universidade Federal do Ceará, Fortaleza, CE, Brasil

<sup>3</sup> Universidade Federal do Ceará, Acarape, CE, Brasil

E-mails: duarte.jr105@gmail.com (ADMJ), cleitonsilveira@ufc.br (CSS), jmicaelcosta@gmail.com (JMFC), suellen.nobre@gmail.com (STNG)

Received: September 25, 2023 - Revised: December 30, 2023 - Accepted: February 22, 2024

### ABSTRACT

Traditional hydrological models have been widely used in hydrologic studies, providing credible representations of reality. This paper introduces a hybrid model that combines the traditional hydrological model Soil Moisture Accounting Procedure (SMAP) with the machine learning algorithm XGBoost. Applied to the Sobradinho watershed in Brazil, the hybrid model aims to produce more precise streamflow forecasts within a three-month horizon. This study employs rainfall forecasts from the North America Multi Model Ensemble (NMME) as inputs of the SMAP to produce streamflow forecasts. The study evaluates NMME forecasts, corrects bias using quantile mapping, and calibrates the SMAP model for the study region from 1984 to 2010 using Particle Swarm Optimization (PSO). Model evaluation covers the period from 2011 to 2022. An XGBoost model predicts SMAP residuals based on the past 12 months, and the hybrid model combines SMAP's streamflow forecast with XGBoost residuals. Notably, the hybrid model outperforms SMAP alone, showing improved correlation and Nash-Sutcliffe index values, especially during periods of lower streamflow. This research highlights the potential of integrating traditional hydrological models with machine learning for more accurate streamflow predictions.

**Keywords:** Hydrological models; Machine learning; Streamflow forecast; SMAP; XGBoost; Rainfall forecasting.

### RESUMO

Os modelos hidrológicos tradicionais têm sido amplamente utilizados em estudos hidrológicos, fornecendo representações credíveis da realidade. Este artigo introduz um modelo híbrido que combina o modelo hidrológico tradicional Soil Moisture Accounting Procedure (SMAP) com o algoritmo de aprendizado de máquina XGBoost. Aplicado à bacia de Sobradinho no Brasil, o modelo híbrido tem como objetivo produzir previsões de vazão mais precisas em um horizonte de três meses. Este estudo utiliza previsões de chuvas do North America Multi Model Ensemble (NMME) como entradas do SMAP para produzir previsões de vazão. O estudo avalia as previsões do NMME, corrige vies usando mapeamento de quantis e calibra o modelo SMAP para a região de estudo de 1984 a 2010 usando a Otimização por Enxame de Partículas (PSO). A avaliação do modelo abrange o período de 2011 a 2022. Um modelo XGBoost prevê os resíduos do SMAP com base nos últimos 12 meses, e o modelo híbrido combina a previsão de vazão do SMAP com os resíduos do XGBoost. Notavelmente, o modelo híbrido supera o SMAP sozinho, mostrando melhor correlação e valores do índice Nash-Sutcliffe, especialmente durante períodos de menor vazão. Esta pesquisa destaca o potencial da integração de modelos hidrológicos tradicionais com aprendizado de máquina para previsões de vazão mais precisas.

**Palavras-chave:** Modelos hidrológicos; Aprendizado de máquina; Previsão de vazão; SMAP; XGBoost; Previsão de chuvas.



## INTRODUCTION

Reliable streamflow forecasts in watersheds are crucial for effective water resources management, particularly in regions affected by extreme climate conditions such as severe droughts (Block et al., 2009; Parisouj et al., 2020). However, generating more accurate streamflow forecasts remains a challenge due to the intricate relationship between climatic variables and the non-linearity that exists in the transformation process of precipitation into streamflow (Adnan et al., 2018; Adnan et al., 2019; Niu et al., 2019).

Given the importance of water management over time, numerous streamflow forecasting methods have been developed for various time scales, including monthly and daily streamflow (Hadi & Tombul, 2018). Streamflow forecasting models can be broadly categorized into physical-based models and data-driven models (Parisouj et al., 2020). Physical-based models include conceptual approaches like the Soil Moisture Accounting Procedure (SMAP) (Cheng et al., 2006), whereas machine learning models fall under the data-driven category (Belayneh et al., 2014).

Physical-based models simplify physical processes but demand extensive data (Meng et al., 2019), while machine learning models operate statistically and typically require less data, gaining popularity for accurate streamflow forecasts (Yaseen et al., 2019; Cheng et al., 2020; Adnan et al., 2022). Machine learning's ease of implementation, relying on historical data statistics, contributes to its widespread use across domains (Liu et al., 2015).

In this context, the combination of traditional hydrological models with machine learning techniques presents an alternative to enhance prediction accuracy and reduce uncertainties associated with these models. Some of these uncertainties are associated with the model parameters, model structure and meteorological inputs (Li et al., 2012). Hydrological models are limited by the knowledge about the hydrological process and their structure is not able to describe perfectly all the physical processes that occur in a watershed (Yang et al., 2020). Several studies indicate that the fusion of diverse models yields more accurate predictions compared to relying solely on a single model (Akbarian et al., 2023; Block et al., 2009; Regonda et al., 2006).

Considering this, we propose, in this work, a hybrid model that combines a traditional methodology, the SMAP model, with a new machine learning model, XGBoost. The idea is to use XGBoost as a SMAP correction model. It will be coupled to the SMAP output seeking to improve the accuracy of the flow forecasts generated. The study area of this work covers the drainage area of the Sobradinho reservoir.

SMAP, developed by Lopes et al. (1982), is widely used in Brazilian hydrological studies. Studies by Cavalcante et al. (2020) and Maciel et al. (2020) demonstrated SMAP's effectiveness in flood prediction, with Maciel et al. (2020) combining SMAP with a Conv3D-LSTM model for superior streamflow predictions. Silva et al. (2019) utilized SMAP with precipitation data from the RegCM model for streamflow prediction in the Três Marias hydroelectric reservoir.

In recent studies, Extreme Gradient Boosting (XGBoost) proposed by Chen & Guestrin (2016) has been employed for streamflow prediction across various locations and temporal scales. Ni et al. (2020) combined XGBoost with a Gaussian mixture model for monthly streamflow prediction in the Yangtze River.

Szczepanek (2022) assessed three tree-based models for daily streamflow prediction in mountainous regions, with XGBoost delivering some of the most promising outcomes. Liu et al. (2022) achieved the best Nash-Sutcliffe performance using XGBoost for streamflow prediction one month ahead in watersheds across the United States. Akbarian et al. (2023) explored the applicability of different machine learning models for monthly streamflow prediction in Iran and highlighted XGBoost's favorable performance.

## METHODOLOGY

### Study area

This study focuses on the Sobradinho watershed in the São Francisco River, spanning the Northeast and Midwest regions of Brazil, covering Minas Gerais and Bahia (Figure 1). The watershed has a main riverbed length of 1,892 km and an area of 147,248 km<sup>2</sup>. The Sobradinho reservoir, located in Bahia, is 320 km long, with a water surface of 4,214 km<sup>2</sup> and a storage capacity of 34.1 billion cubic meters. The reservoir houses a hydropower plant with 6 generating units and an installed capacity of 1,050,300 kW (Eletrobras Chesf, 2023).

According to Silva (2018), Sobral et al. (2018), and Pereira (2004), this reservoir holds significant importance in various areas, including:

- a) Hydropower Generation: The Sobradinho hydropower plant contributes to both local and national electricity demand as part of the National Interconnected System (ONS);
- b) Water Regulation: The reservoir plays a crucial role in regulating water resources, serving as a primary water source for Juazeiro (Bahia) and Petrolina (Pernambuco);
- c) Agriculture: The reservoir supports irrigation, boosting agricultural productivity in one of Brazil's most arid regions.

The Sobradinho watershed confronts challenges in sustainable water management, biodiversity preservation, pollution control, and mitigating the impact of drought periods. Continuous efforts are essential for the conservation and responsible utilization of this critical water source (Sobral et al., 2018).

### Stages of the study

The study's flowchart, depicted in Figure 2, outlines key stages. It initiates with data acquisition, encompassing rainfall, evapotranspiration, and streamflow data. Rainfall data is sourced from both observations and North America Multi Model Ensemble (NMME) forecasts. Calibration and evaluation of the Soil Moisture Accounting Procedure (SMAP) hydrological model follow. Bias correction is applied to NMME rainfall forecasts, used as inputs for SMAP to generate initial streamflow forecasts. Residual analysis guides the training of the XGBoost model, predicting SMAP forecast residuals. The hybrid model integrates XGBoost and SMAP forecasts, aiming for improved prediction accuracy, and its forecasts are compared with standalone SMAP forecasts in the final assessment.

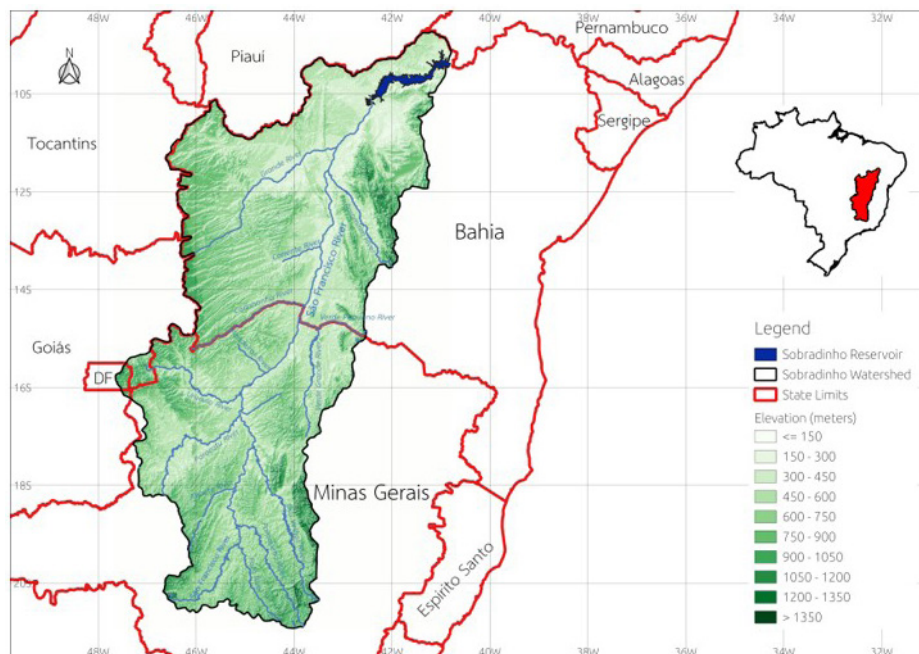


Figure 1. Study area.

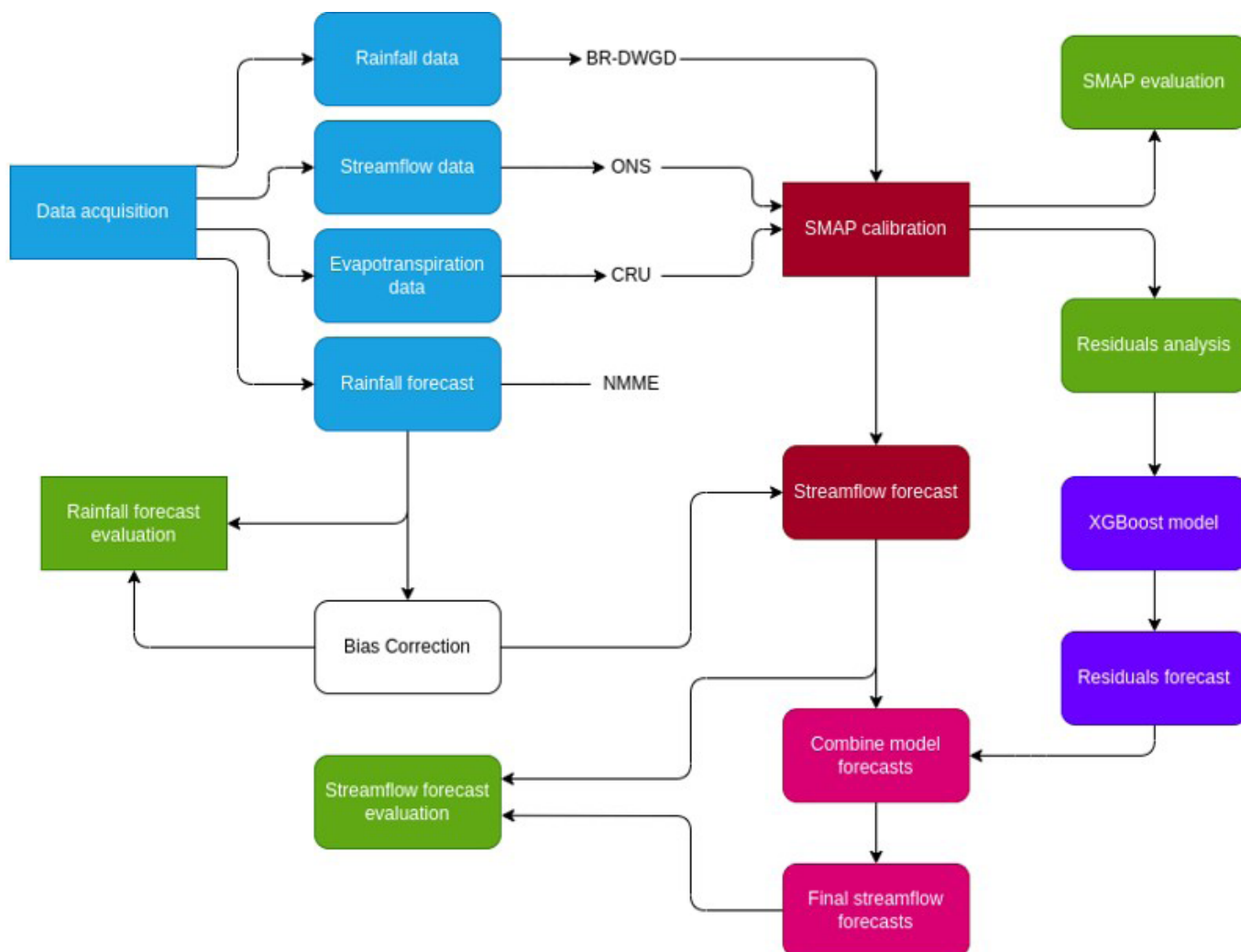


Figure 2. Flowchart depicting the study's stages.

## Databases description

### Brazilian Daily Weather Gridded Data (BR-DWGD)

The BR-DWGD, introduced in 2016 and updated to 2022, is a meteorological database in Brazil with improved spatial resolution (Xavier et al., 2016, 2022). Covering January 1st, 1961, to December 31st, 2022, it provides data on rainfall, temperature, solar radiation, wind speed, and humidity from 11,473 rain gauges and 1,252 meteorological stations (Xavier et al., 2022). Widely utilized in studies, positive evaluations highlight its strong correlation with observed data (Bender & Sentelhas, 2018; Battisti et al., 2019; Duarte & Sentelhas, 2020). This study extracted monthly rainfall timeseries from BR-DWGD for the Sobradinho reservoir's drainage area from January 1984 to December 2022.

### Climatic Research Unit (CRU)

Evapotranspiration data from January 1994 to December 2022 were obtained from the CRU database, featuring a spatial resolution of  $0.5^\circ \times 0.5^\circ$ . This global climatic data collection, in its fourth version (CRU TS v4.07), is based on interpolations from meteorological station data (Harris et al., 2020). Updated annually, it spans from January 1901 to December 2022, and its reliability is acknowledged in various academic studies, including climatic (Wang et al., 2013) and hydrological research (Vollmer et al., 2005), and bias correction in climatic models (Miao et al., 2016). Mutti et al. (2020) specifically found a good correlation with observed rainfall and evapotranspiration data in the São Francisco River basin, Brazil.

### National Operator of the Electrical System (ONS)

The Sobradinho reservoir's streamflow data (January 1984 to December 2022) was acquired from ONS databases, operating under Brazil's SIN overseen by Aneel. ONS disseminates reports on reservoir volumes, and this study utilized the Natural Affluent Streamflow, representing river flow without human intervention. Calculated by ONS, it considers incoming water volume, subtracts human activity water use, and applies regression functions. Additional details are in the "Sub-module 13.5 ONS Network Procedures" manual, accessible on their website (Brasil, 2017).

### North America Multi Model Ensemble (NMME)

The NMME is a climatic forecasting project with models from various centers, initiated in 2010 and updated monthly since August 2011 (Kirtman et al., 2014). The centers follow protocols, including a climatology with initial conditions for each month, producing forecasts for a minimum of 9 months ahead, and using required monthly fields like temperature at 2 meters, precipitation rate, and sea surface temperature.

This study utilized forecasts from four NMME models: COLA-RSMAS-CCSM4, GFDL-SPEAR, NASA-GEOS2S, and NCEP-CFSv2. These models were selected due to their extended

time series availability and data coverage until 2022. The time series used spanned from 1994 to 2022. Hindcast data from the models were used for the period from 1984 to 2011, while forecasts were used for the period from 2012 to 2022. Table 1 presents some characteristics of the models used.

The decision to use NMME was influenced by its novelty and limited exploration in hydroclimatic studies in Brazil. While global studies have shown positive evaluations (Li et al., 2011; Mo & Lyon, 2015; Kirtman et al., 2014; Shukla et al., 2015), Brazil-specific assessments are scarce. Studies by Flores (2021), Rocha Júnior et al. (2021), and Andrian et al. (2023) in Brazil demonstrated NMME's effectiveness, showcasing better forecasting abilities across regions, particularly in the Northeast.

## Data analysis and processing

### Bias correction

Bias correction aligns global models with observed data at local scales. The mapping quantile technique described by Bárdossy & Pegram (2011) and Abdolmanafi et al. (2021), chosen in this study, compares Cumulative Distribution Functions (CDFs) of observed and modeled data. The technique is routinely used to correct biases of regional climate models (Maraun, 2013). It involves modeling data distributions using known distribution functions. The Gamma Distribution was chosen, as shown by Silveira et al. (2017), to accurately represent rainfall distribution in the Northeast region of Brazil.

In this technique, CDFs of observed and hindcast data were determined for the historical period (1984 to 2010). These CDFs were adjusted to a Gamma distribution function, commonly used in Brazilian hydrological studies (Billerbeck et al., 2021; Santos et al., 2019; Gondim et al., 2018). Bias correction was then applied to rainfall forecasts of NMME models for the years 2011 to 2022. The method involves comparing the probability of occurrence in the model's CDF with the observed data's CDF and adjusting the forecasted values based on this comparison, ensuring more accurate predictions.

### Yeo-Johnson transformation

Yeo-Johnson transformation of SMAP residuals utilizes a Box-Cox-based method to address distribution asymmetry (Yeo & Johnson, 2000; Box & Cox, 1964). For positive  $y$ , it resembles Box-Cox on  $(y+1)$ . When  $y$  is always negative, it's like Box-Cox on  $(-y+1)$  with power  $2 - \lambda$ . Series with both positive and negative values combine these cases using distinct powers for positive and negative values.

$$\Psi(\lambda, y) = \begin{cases} \frac{(y+1)^\lambda - 1}{\lambda} & \text{if } \lambda \neq 0, y \geq 0 \\ \log(y+1) & \text{if } \lambda = 0, y \geq 0 \\ -\frac{[(-y+1)^{(2-\lambda)} - 1]}{(2-\lambda)} & \text{if } \lambda \neq 2, y < 0 \\ -\log(-y+1) & \text{if } \lambda = 2, y < 0 \end{cases} \quad (1)$$



**Table 1.** Characteristics of the NMME model forecasts used.

Model	Nº of members	Maximum Lead Time (months)	Spatial Resolution
COLA-RSMAS-CCSM4	10	12	1°x1°
GFDL-SPEAR	30	12	1°x1°
NCEP-CFSv2	32	10	1°x1°
NASA-GEOSS2S	10	9	1°x1°

## Autocorrelation

Autocorrelation was used to assess time dependency in streamflow forecast residuals, indicating the correlation between a time series and itself with a specified time lag. Small autocorrelations in residuals suggest the model lacks significant fit issues. Mathematically, autocorrelation is defined by Silva Filho (2014):

$$\rho(k) = \frac{\sum_{t=1}^{n-k} (x_t - \mu)(x_{t+k} - \mu)}{\sigma^2} \quad (2)$$

where:  $\mu$  is the average of the variable;  $x_t$  is the variable value in the time lag  $t$ ;  $\sigma^2$  is the variance of the variable.

## Ljung-Box test

The Ljung-Box test (Ljung & Box, 1978) assessed autocorrelation in streamflow forecast residuals to evaluate the model's fit to observed data. The Ljung-Box test is defined as:

$$Q = n(n+2) \sum_{k=1}^m \frac{\hat{\rho}_k^2}{n-k} \quad (3)$$

In which the  $\hat{\rho}_k$  is the autocorrelation in a lag  $k$  and  $m$  is the number of lags that are used.

## Hydrological modeling

### SMAP model

SMAP is a conceptual hydrological model developed by Lopes et al. (1982). It is based in the Soil Conservation Service (SCS) method (USDA - United States Department of Agriculture, 1986) and was initially designed for daily scale simulations, later adapted for monthly scale (Miranda et al., 2017).

SMAP models water flow in the soil with two reservoirs (soil and subsurface), using transfer functions for reservoir recharge and surface runoff processes. Equations are derived from observed precipitation and streamflow data, adjusted for the watershed's hydrological cycle. The monthly version utilizes precipitation, potential evapotranspiration, and watershed area, with internal parameters like K, Tuin, Pes, Sat, Crec, and Ebin. More technical details can be found in Lopes et al. (1982).

SMAP models water flow in soil with two reservoirs, employing transfer functions for recharge and surface runoff. Equations derive from observed precipitation and streamflow data, adapted for the watershed's hydrological cycle. Monthly version uses precipitation, potential evapotranspiration, and watershed area, with internal parameters

like K, Tuin, Pes, Sat, Crec, and Ebin. Further details in Lopes et al. (1982). The Nash-Sutcliffe Efficiency (NSE) assesses the fit between a model and observed series, with a perfect fit having an NSE value of 1. Widely used in hydrological studies, it demonstrates flexibility as a fitting statistic studies (Birikundavyi et al., 2002; Johnson et al., 2003; Downer & Ogden, 2004) and recommended by the ASCE (Water Management Committee) for continuous moisture accounting models (American Society of Civil Engineers, 1993).

Streamflow forecasts with SMAP require forecasted precipitation and evapotranspiration data. NMME precipitation forecasts were used as inputs for SMAP from 2011 to 2022. Evapotranspiration input was the monthly climatological average calculated between 1984 and 2011, given the absence of forecast models for this variable.

### SMAP calibration using PSO

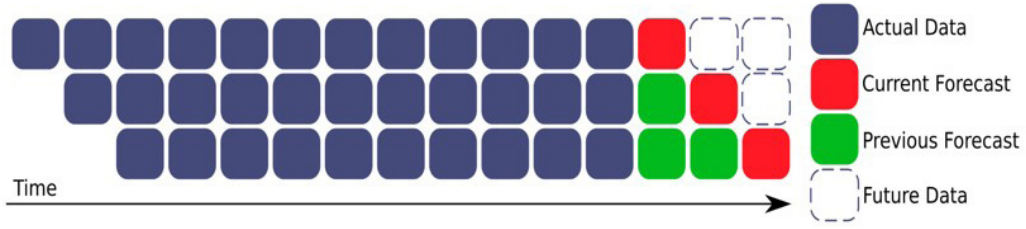
PSO, inspired by swarm behavior, generates a swarm of potential solutions to optimize a specific objective function. It explores the solution space to identify the optimal solution, minimizing or maximizing the objective function value (Marini & Walczak, 2015). In this study, PSO determined SMAP parameters to maximize NSE, aiming for accurate simulations. PSO, as per Kachitvichyanukul (2012), begins by generating a swarm of potential solutions. Each particle is evaluated to find the one closest to the maximum of the objective function. The best global and individual solutions are retained. The process continues until a stopping condition, such as a specified number of iterations or proximity to the maximum value, is met.

### The hybrid model SMAP-XGBoost

XGBoost, by Chen & Guestrin (2016), is a versatile ensemble learning algorithm used for regression and classification, gaining popularity in hydrological series (Ma et al., 2021). It combines multiple shallow trees, correcting errors sequentially and minimizing a loss function to enhance accuracy. XGBoost incorporates regularization techniques for improved generalization. XGBoost utilizes optimized gradient descent for faster and more accurate convergence, surpassing traditional gradient boosting (Chen & Guestrin, 2016). It also employs parallel computation for increased speed. Mathematically, XGBoost is defined as follows:

$$\hat{x}_i = \phi x_i = \sum_{k=1}^{\kappa} f_k(x_i), f_k \in F \quad (5)$$

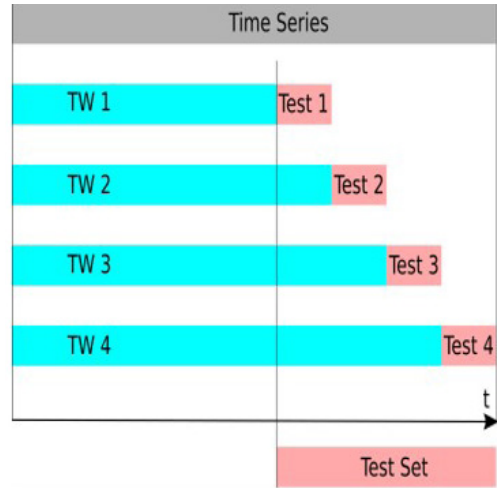
In which:  $f_k(x_i)$  is the predicted value by the  $k$ -th tree;  $\kappa$  is the total of trees;  $x_i$  is the value of the  $x$ -th model input;  $F$  are all the tree developed.



**Figure 3.** Multi-Step ahead forecast.

The XGBoost model was used to model the time series of SMAP residuals. It utilized the residuals from the preceding 12 months as input variables for the month under prediction. A recursive strategy was employed for forecasting residuals across different time horizons, as shown in Figure 3. This strategy involves using the model's past predictions as inputs for predictions beyond one month. For example, to forecast the residual for 2 months ahead, the model uses observed residuals from the 11 months prior to the prediction date along with the prediction made for 1 month ahead.

The model was trained and forecasted using a nested cross-validation strategy by Ratanamahatana et al. (2009), as depicted in Figure 4. Initially, the model trained on a 70% portion of the residual time series (January 1984 to December 2010), denoted as the training window (TW). The model then forecasted a three-month test set in each iteration. The TW size increased by 1 month in each iteration using the expanding window method, covering the entire data series and minimizing uncertainties related to the model's initial conditions. Streamflow forecasts from the hybrid SMAP-XGBoost model result from combining streamflow predictions from SMAP with the residual forecasts generated by XGBoost. These forecasts were conducted for the period 2011-2022.



**Figure 4.** Nested expanding windows strategy.

## Evaluation metrics

The performance of the models was evaluated using the following metrics.

Pearson correlation

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (6)$$

Nash-Sutcliffe index (Nash & Sutcliffe, 1970)

$$NSE = 1 - \frac{\sum_{i=1}^n (x_i - y_i)^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (7)$$

Bias

$$Bias = \frac{\sum_{i=1}^n (y_i - x_i)}{n} \quad (8)$$

Root Mean Squared Error

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - x_i)^2}{n}} \quad (9)$$

Mean Absolute Error

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - x_i| \quad (10)$$

Mean Absolute Percentage Error

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{(y_i - x_i)}{x_i} \right| \quad (11)$$

where:  $x_i$  is the observed value;  $\bar{x}$  is the average of the observed values;  $y_i$  is the forecast value;  $\bar{y}$  is the average of the forecast values;  $n$  is the number of samples.

Monthly forecasts until to three months were used to calculate metrics. NMME precipitation forecasts were compared with observed data from BR-DWGD, while streamflow forecasts from SMAP and SMAP-XGBoost were compared with naturalized streamflow series from ONS.

The correlation coefficient ( $r$ ) assesses the relationship between forecasts and observed time series, ranging from -1 to 1. A value of 1 indicates a perfect positive correlation, 0 indicates no correlation, and -1 indicates a perfect negative correlation. A correlation coefficient closer to 1 signifies better performance

in capturing observed data. The NSE index reflects model fit, bias measures tendency, MAE gauges average error, RMSE is sensitive to outliers, and MAPE penalizes errors during low streamflow. Lower values in these metrics indicate better model performance.

## RESULTS

### Evaluation of rainfall forecasts from the NMME models

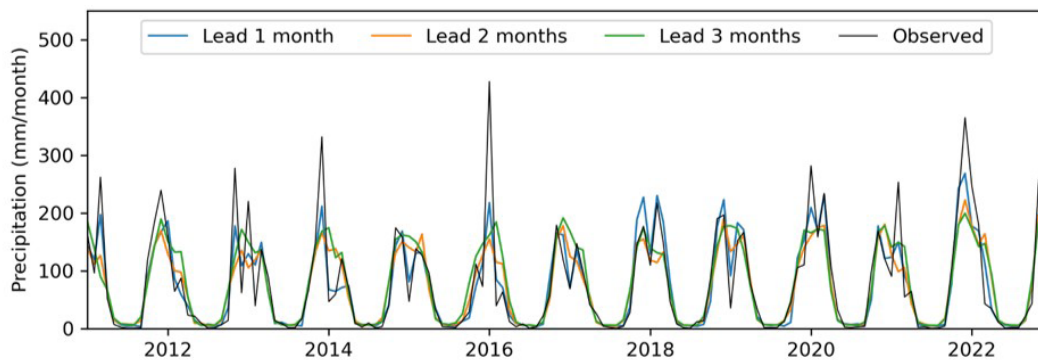
Figure 5 compares monthly rainfall forecasts from the average NMME ensemble with observed data for a three-month forecast horizon from 2011 to 2022. The NMME-Ensemble represents the mean of the NMME models discussed earlier.

The rainfall forecast effectively captured the seasonality in the observed series, representing both maximum and minimum rainfall cycles. Accuracy was highest for the 1-month forecast, followed by the 2-month forecast, and finally, the 3-month forecast. As the forecast horizon increased, there was a noticeable decrease in the magnitude of peak rainfall forecast values. In certain notably atypical years, such as January 2016, the models struggled to accurately replicate the observed time series. During this month, which was the rainiest in the assessed series, the observed values significantly exceeded the average observed in other years.

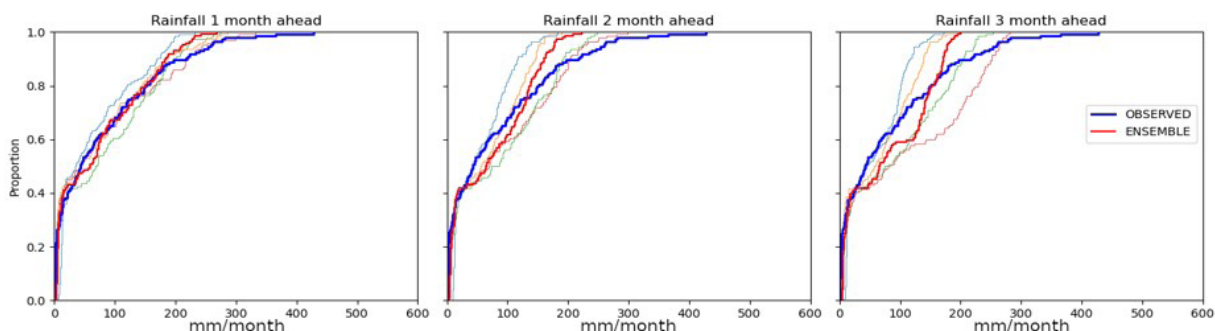
Figure 6 compares the cumulative distribution functions (CDFs) for 1-month, 2-month, and 3-month-ahead forecast horizons. The NMME-Ensemble’s CDF aligns closely with the observed series for the 1-month horizon, with a slight tendency to underestimate extreme maximum values. However, the gap widens for the 2-month horizon, where the ensemble tends to overestimate intermediate values and underestimate maximum values. This trend persists for the 3-month horizon, indicating a consistent tendency to underestimate extreme values across all forecast horizons, with predictive skill diminishing over longer horizons.

Table 2 summarizes the evaluation metrics used. Among all evaluated models, the NMME-Ensemble performed the best across the metrics. In comparison to this model, the others exhibited lower correlation and NSE values, as well as higher bias, RMSE, and MAE values. This indicates that the other models are less accurate than the NMME-Ensemble, resulting in increased forecast uncertainties.

Across all forecast horizons, the models exhibited correlations above 0.7, indicating satisfactory results. Regarding the NSE metric, the ensemble consistently achieved high values, surpassing 0.6 and even reaching above 0.8 for the 1-month horizon. In contrast, the COLA-RSMAS-CCSM4 model displayed the poorest performance, with NSE values falling below 0.5 in two forecast horizons, despite the lowest individual value being observed in NCEP-CFSv2 for the 3-month horizon. As for Bias,



**Figure 5.** Comparing the forecasted rainfall from the NMME-ENSEMBLE model for 1, 2, and 3 months ahead with the average observed rainfall in the Sobradinho reservoir’s drainage area.



**Figure 6.** Cumulative Distribution Function (CDF) analysis, comparing NMME climate forecasts with average rainfall at the Sobradinho drainage area from 2011 to 2022.

all models displayed relatively low values, with the smallest value recorded for NCEP-CFSv2 in the 1-month forecast horizon. In terms of RMSE and MAE metrics, all models had relatively close values, yet the ensemble mean outperformed the others.

### Evaluation of NMME models after bias correction

Figure 7 depicts cumulative distribution functions (CDFs) after bias correction. Notably, the 1-month forecast horizon shows improved alignment with the observed distribution, particularly in maximum values. The 2 and 3-month horizons exhibit slight enhancements, with the 3-month horizon showing improved distribution in average rainfall values compared to the series without bias correction (Figure 6).

Table 3 presents the same metrics as shown in Table 2, but for the data after bias correction. Focusing on the ensemble, which had the best metrics in the previous evaluation, we can observe improvements in almost all the metrics. For the ensemble, both correlation and NSE showed slightly higher values compared to those obtained before bias correction. Bias was the metric with

the most significant changes, with reduction in bias observed in all model and for all forecast horizons, with the exception of the GFDL-SPEAR model in its 1-month ahead forecast. RMSE saw changes in values for most models, with only the NCEP-CFSv2 and COLA-RSMAS-CCSM4 models consistently decreasing and increasing their RMSE values, respectively, across all forecast months. The other models exhibited variations in RMSE values, but without significant changes. MAE also showed a reduction in its values for all forecast horizons. Similar behavior was observed in the other models, mirroring the improvements seen in the ensemble mean.

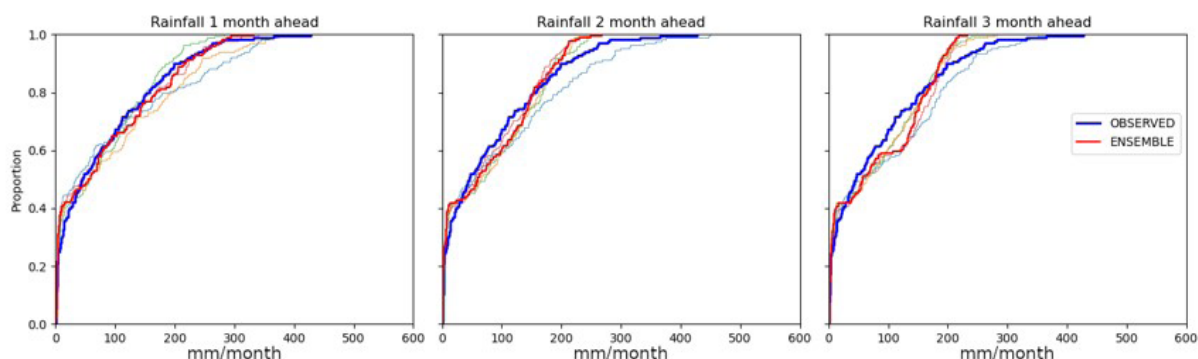
Bias correction significantly improved precipitation forecasts, and consequently, forecasts from the Ensemble mean after bias correction will be utilized in subsequent stages of the work due to its superior performance in the evaluation.

### Calibration and evaluation of the SMAP model

This section presents the results of the calibration and evaluation of the SMAP model using observed rainfall and

**Table 2.** Evaluation metrics for the NMME rainfall forecast for the period from 2011-2022.

Model	Lead Time (months)	Correl.	NSE	Bias (mm/month)	RMSE (mm/month)	MAE (mm/month)
COLA-RSMAS-CCSM4	1	0.85	0.67	-15.38	50.02	31.67
COLA-RSMAS-CCSM4	2	0.78	0.46	-24.57	63.71	38.68
COLA-RSMAS-CCSM4	3	0.74	0.44	-21.82	65.48	41.86
GFDL-SPEAR	1	0.87	0.76	-5.71	42.78	26.18
GFDL-SPEAR	2	0.8	0.59	-14.94	55.62	33.15
GFDL-SPEAR	3	0.77	0.54	-16.25	59.15	35.45
NASA-GEOSS2S	1	0.86	0.73	5.26	45.45	29.41
NASA-GEOSS2S	2	0.81	0.64	9.31	52.48	33.37
NASA-GEOSS2S	3	0.76	0.52	12.26	60.24	37.3
NCEP-CFSv2	1	0.89	0.78	0.88	40.84	24.87
NCEP-CFSv2	2	0.8	0.59	11.62	55.5	33.4
NCEP-CFSv2	3	0.77	0.31	31.41	71.88	45.95
ENSEMBLE	1	0.91	0.82	-3.74	36.92	22.53
ENSEMBLE	2	0.82	0.67	-4.86	50.11	29.81
ENSEMBLE	3	0.78	0.61	1.07	54.41	32.74



**Figure 7.** CDF analysis, comparing NMME climate and average rainfall in the Sobradinho drainage area from 2011 to 2022, after bias correction.



streamflow time series as inputs. The generated streamflow is not a forecast but a simulation to assess the model’s performance when actual values are known. Streamflow forecasts will be evaluated in the subsequent section.

Figure 8 illustrates the hydrograph generated during the calibration period of the SMAP model. The streamflow and rainfall series are in phase, aligning with coinciding periods of maximums and minimums. SMAP tends to overestimate observed streamflow values during low flow periods, successfully capturing the rising and falling curves in the streamflow with some peak occurrences.

Table 4 provides the SMAP parameters utilized for generating the flow series shown in Figure 8. The estimated value for Tuin is 63%, fitting the expected range as the model initiates

in January during the watershed’s rainy season. The Ebin value corresponds to the minimum observed flow in the period. A “K” value close to zero suggests a concentration time in the watershed of less than 1 month.

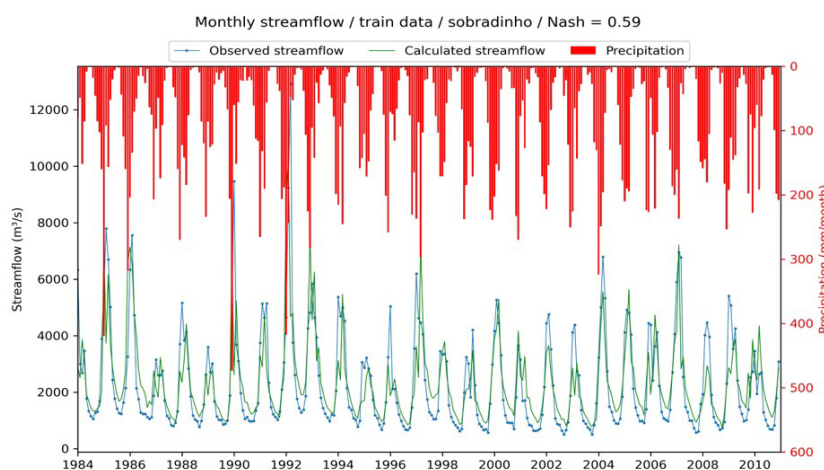
The evaluation metrics related to the calibration of SMAP are displayed in Table 5. Considering that the maximum correlation possible is 1, the obtained value in calibration (0.79) is a good result. It indicates that the model answer captures most of the variability in the original series as shown in Figure 8, in which it is possible to see that the model represents the seasonality of the observed series. Considering the NSE models with  $NSE > 0.5$  are considered acceptable (Knoben et al., 2019) and how was obtained a NSE of 0.62 in calibration the model can be considered acceptable. Considering

**Table 3.** Evaluation metrics for the NMME rainfall forecast for the period from 2011-2022 after Bias Correction.

Model	Lead Time (months)	Correl.	NSE	BIAS (mm/month)	RMSE (mm/month)	MAE (mm/month)
COLA-RSMAS-CCSM4	1	0.84	0.57	8.38	58.32	36.05
COLA-RSMAS-CCSM4	2	0.78	0.38	18.26	69.97	43.37
COLA-RSMAS-CCSM4	3	0.75	0.42	15.39	67.74	42.61
GFDL-SPEAR	1	0.89	0.71	13.23	47.63	29.34
GFDL-SPEAR	2	0.81	0.65	1.31	52.57	30.95
GFDL-SPEAR	3	0.77	0.58	-0.98	57.67	34.82
NASA-GEOSS2S	1	0.87	0.75	-3.98	44.47	27.28
NASA-GEOSS2S	2	0.8	0.64	1.73	53.17	33.35
NASA-GEOSS2S	3	0.76	0.58	-1.99	57.95	33.29
NCEP-CFSv2	1	0.91	0.81	-0.52	38.58	23.3
NCEP-CFSv2	2	0.81	0.65	-5.22	52.31	30.14
NCEP-CFSv2	3	0.77	0.59	5.21	56.77	35.02
ENSEMBLE	1	0.92	0.85	1.76	34.4	22.15
ENSEMBLE	2	0.83	0.69	-1.14	49.74	28.63
ENSEMBLE	3	0.79	0.62	0.58	55.1	31.97

**Table 4.** SMAP parameters obtained during the calibration process.

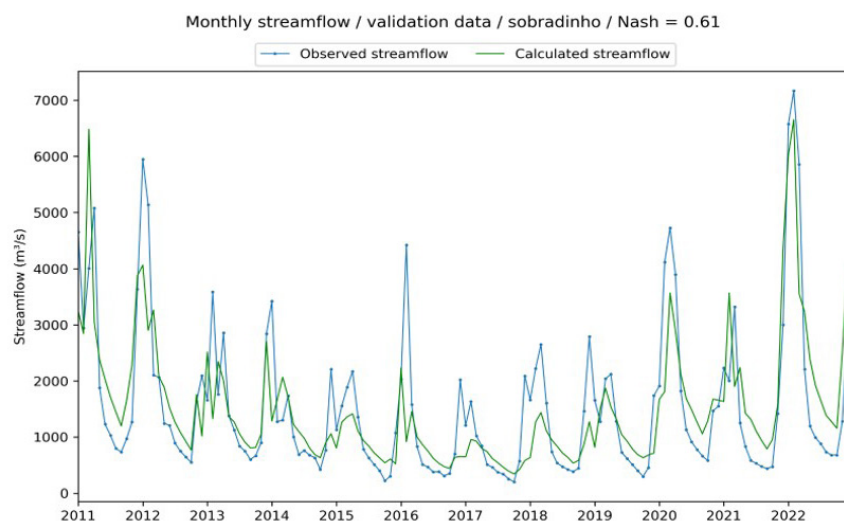
Sat (mm)	Pes	Crec	K (months)	Tuin (%)	Ebin (m <sup>3</sup> /s)
3216	7.1	1.6	0.02	63	216



**Figure 8.** Monthly hydrograph for the calibration period of the SMAP model in the Sobradinho watershed.

**Table 5.** Evaluation metrics for SMAP for calibration and validation.

	Correlation	NSE	BIAS (m <sup>3</sup> /s)	RMSE (m <sup>3</sup> /s)	MAE (m <sup>3</sup> /s)	MAPE (%)
Calibration	0.79	0.62	13.04	1142.09	722.52	32.51
Validation	0.78	0.61	-18.77	846.19	600.3	45.19


**Figure 9.** Monthly streamflow for the SMAP test period in the Sobradinho watershed.

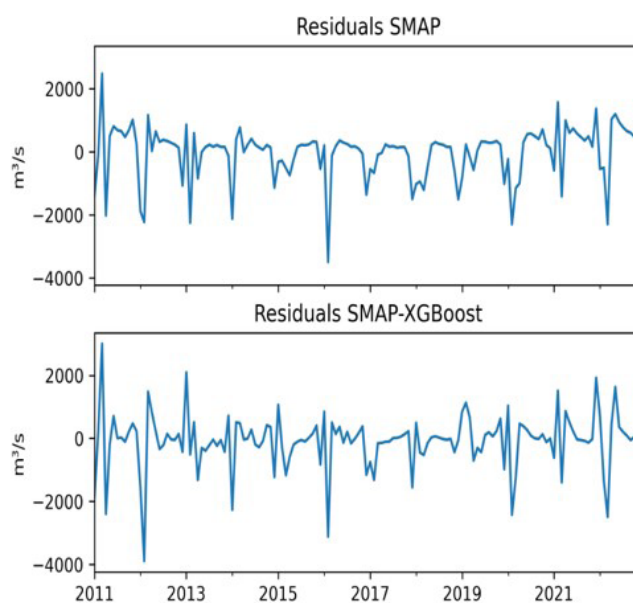
the average flow during the period, which is 2,531.65 m<sup>3</sup>/s, the bias value obtained indicates that the model has virtually no bias. RMSE is a metric that penalizes errors in flow peaks, so it is expected to have higher values than MAE, which evaluates in terms of averages. The obtained MAP value can also be considered acceptable.

In Figure 9, the streamflow for the SMAP test series is depicted. During the test period, SMAP demonstrated a behavior similar to the calibration period, tending to overestimate values in low-flow periods. However, SMAP faced challenges in capturing peaks during this test period, especially between 2013 and 2019, which experienced the lowest monthly precipitation averages in the available historical data. This unusual behavior may have influenced the model's performance.

The performance metrics for the validation period are also presented in Table 5. Overall, the model obtained results very close to those recorded in the calibration period, especially for the metrics: correlation, NSE, and bias. This indicates that the model did not overfit, which is when the model is overly adjusted to the training series but cannot generalize to the test series and ends up with much lower performance. RMSE and MAE were metrics that showed slightly better values in the test period compared to the calibration period. However, in relation to MAPE, which penalizes error in minimum flows more, its value increased. In other words, the model in the test period had better performance for modeling maximum and medium flows, but worse for minimum flows.

### Analysis of the SMAP-XGBoost hybrid model

Since the proposed hybrid model aims to use a machine learning model for model the SMAP residuals, part of this section focuses on their analysis. The proposed machine learning model

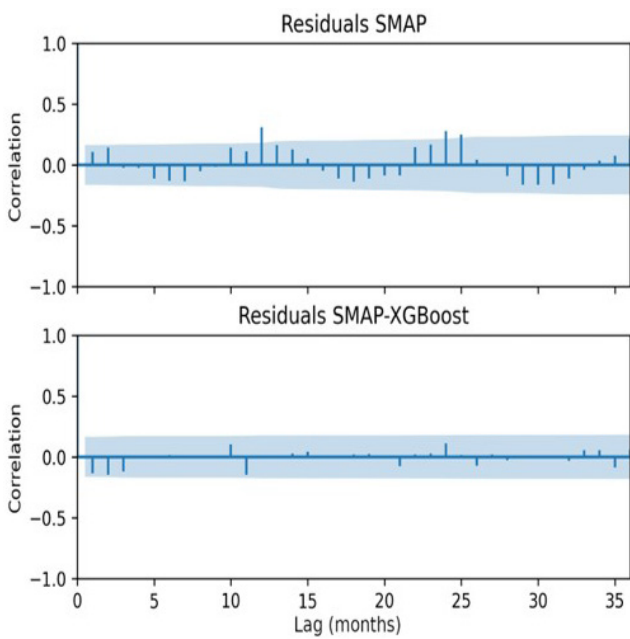

**Figure 10.** Comparison between the SMAP residual series and the hybrid SMAP-XGBoost model residuals.

was applied as a time series model. Ideally, the residuals of a time series model should exhibit the characteristics of white noise, i.e., they should not show a lack of temporal dependence, follow a normal distribution, and have a mean of zero.

In Figure 10, a comparison between the SMAP residual series and the hybrid SMAP-XGBoost model residuals for the period from 2011 to 2022 is presented. It can be observed that the SMAP residuals contain systematic errors with seasonality,

indicating temporal dependence. In contrast, the residual series of the SMAP-XGBoost model shows reduced seasonal patterns, suggesting that there is no temporal dependence in the residuals. The observed behavior in the SMAP residual series is partly due to the seasonality of the streamflow series, as the periods of higher errors coincide with periods of higher streamflow.

The autocorrelation analysis, as depicted in Figure 11, reveals temporal trends in the SMAP residual series. Statistically significant autocorrelation values at lags of 12 and 24 months indicate a connection between SMAP values at a given time and observed data from 12 and 24 months ago. This temporal



**Figure 11.** Autocorrelation of the residual series for the SMAP and SMAP-XGBoost models for the period from 2011 to 2022.

dependence is confirmed by the Ljung-Box test, with a p-value of 0.001 at lag 12, falling below the significance threshold of 0.05.

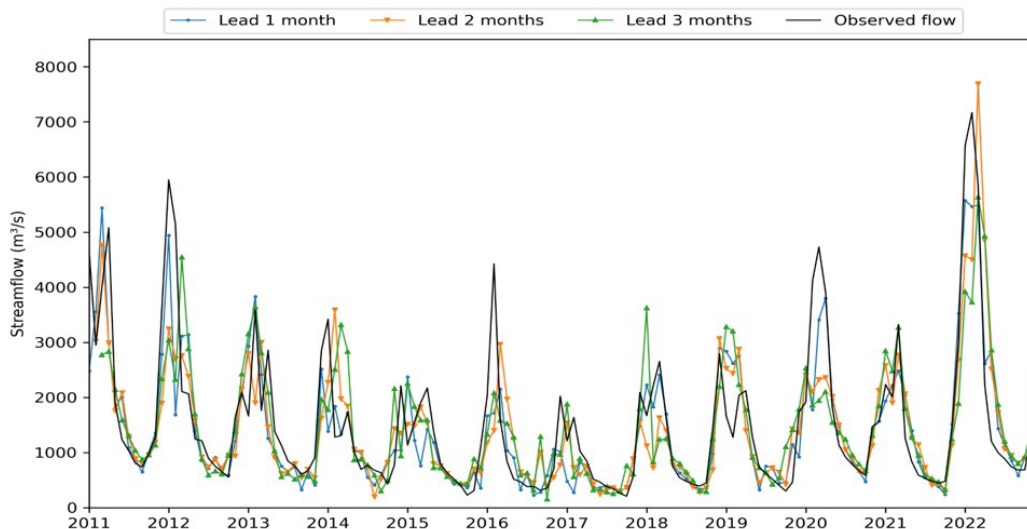
Regarding the removal of temporal dependence in the residuals, the SMAP-XGBoost model performed satisfactorily. Because the SMAP-XGBoost residual series, there was no statistically significant autocorrelation at any of the lags shown in Figure 12. The Ljung-Box test statistic obtained a p-value of 0.360. In this case, since the p-value is greater than the 0.05 threshold, we cannot reject the hypothesis that the data are independently distributed.

On the other hand, Figure 13 shows histograms with the distribution of residuals from both models: SMAP and SMAP-XGBoost. Although neither of them follows a normally distributed series, the residual series of SMAP-XGBoost showed a distribution closer to the standards of a normal distribution. In the distribution of SMAP-XGBoost residuals, it is observed that the median and mean are not as close to each other, which is characteristic of a normal distribution, compared to the distribution of SMAP residuals. Thus, although SMAP-XGBoost had a higher mean error than SMAP, considering the distribution of residuals, it is observed that the model has less tendency in the distribution of errors.

The Table 6 presents performance metrics for the SMAP-XGBoost model. Compared to the results for the SMAP model in Table 5, it can be observed that the correlation values were maintained. The NSE had a minimal reduction of only 0.04 hundredths. The higher RMSE indicates that the SMAP-XGBoost model has more difficulty representing peak flows. However, in percentage terms, the difference in RMSE between the models is only 5%.

With regard to MAE, the result obtained by SMAP-XGBoost is about 8% lower than that obtained by SMAP. The biggest difference between the metrics was obtained for MAPE, with a 25% reduction compared to the value obtained for SMAP.

The Figure 14 contains the comparison between the series generated by SMAP-XGBoost and the observed flow data. As suggested by the previous metrics, the SMAP-XGBoost model



**Figure 12.** Comparison between the predicted and observed streamflow series using SMAP-XGBoost.

tends to represent the flows better during its low-flow periods than the original SMAP, as seen in Figure 9. The graph also shows that the model generally generates higher peak flows than the

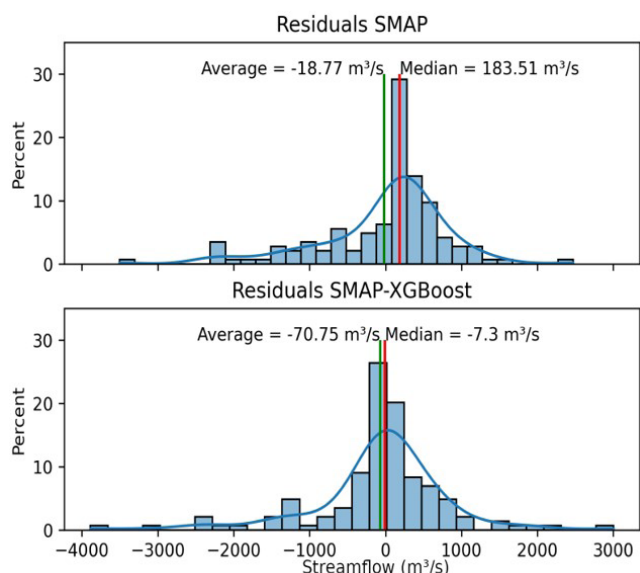
previous version. However, some peaks are delayed compared to the observed peaks and overestimate them mode, while the SMAP (Figure 6) showed a greater tendency to underestimate peak flows.

### Streamflow forecast using SMAP and NMME

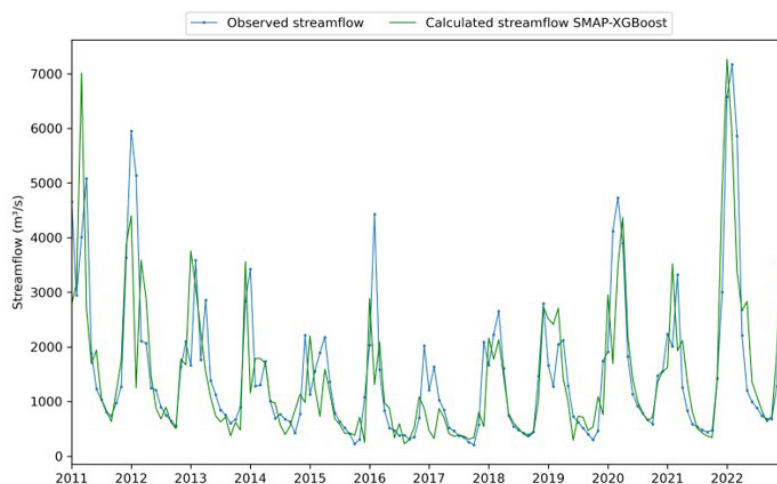
The performance metrics for streamflow forecast using SMAP and rainfall forecast from the NMME mean ensemble are shown in Table 7. In all metrics and forecast horizons, the model exhibits value similar to those obtained during its validation period. As expected, the model's performance decreases as the forecast horizon increases, as the uncertainty about the influence of initial conditions also increases.

In terms of correlation, the model consistently achieved high values across all forecast horizons, indicating good performance. The average bias was relatively low, almost negligible. However, the MAPE, particularly for the 3-month horizon, approached 50%, signifying a weakness in accurately representing periods of low streamflow. RMSE and MAE exhibited values within the expected range, similar to those obtained during the validation period.

Figure 15 illustrates the forecast series at three-time horizons compared to the observed streamflow series. The metrics in Table 7, particularly MAPE, indicate a tendency for the model to overestimate observed values at all forecast horizons. The model faces challenges in capturing certain peaks but generally performs



**Figure 13.** Distribution of residuals from the SMAP and SMAP-XGBoost models.



**Figure 14.** Comparison between the observed and generated flow series by SMAP-XGBoost.

**Table 6.** Performance metric for SMAP-XGBoost during the testing period.

Correlation	NSE	BIAS (m <sup>3</sup> /s)	RMSE (m <sup>3</sup> /s)	MAE (m <sup>3</sup> /s)	MAPE (%)
0.78	0.57	-70.75	885.28	550.59	33.93

**Table 7.** Performance metrics for streamflow forecast using SMAP with bias-corrected rainfall forecasts from the NMME-Ensemble.

Lead time (months)	Correlation	NSE	BIAS (m <sup>3</sup> /s)	RMSE (m <sup>3</sup> /s)	MAE (m <sup>3</sup> /s)	MAPE (%)
1	0.86	0.73	-9.87	702.94	517.48	41.06
2	0.77	0.59	-2.53	854.46	600.84	45.13
3	0.73	0.53	-18.99	915.90	628.37	47.17



well in representing the seasonality of the series, encompassing both high and low streamflow periods.

### Streamflow forecasts from the SMAP-XGBoost model.

Table 8 present streamflow forecast results from the SMAP-XGBoost model using NMME rainfall forecasts. For the 1-month horizon, the model's metrics were similar to the test data (Table 6), with correlation and NSE values close to SMAP, higher bias and RMSE, and lower MAE and MAPE. At the 2-month horizon, only bias was worse than SMAP, while other metrics were better. However, for the 3-month horizon, SMAP-XGBoost performed worse in all metrics compared to SMAP.

As SMAP-XGBoost forecast better capture minimum streamflow levels compared to those made by SMAP (Figure 12). However, at a 3-month forecast horizon, there is increased noise in the series, resulting in worsened evaluation metrics. While SMAP better captured some peaks, it also showed higher values in others (Figure 9), although they were delayed compared to the observed data.

In the Table 9 is showed the evaluation by quarter. Flow forecasts were reviewed and separated into quarters in this assessment; the complete list can be found in the appendix. The largest flows in the study region occur between December and May. According to the measurements, SMAP operates better during periods of larger flows, particularly during the March to May period (MAM). SMAP-XGBoost, on the other hand, provides measures with better values, reduced errors, and higher correlation

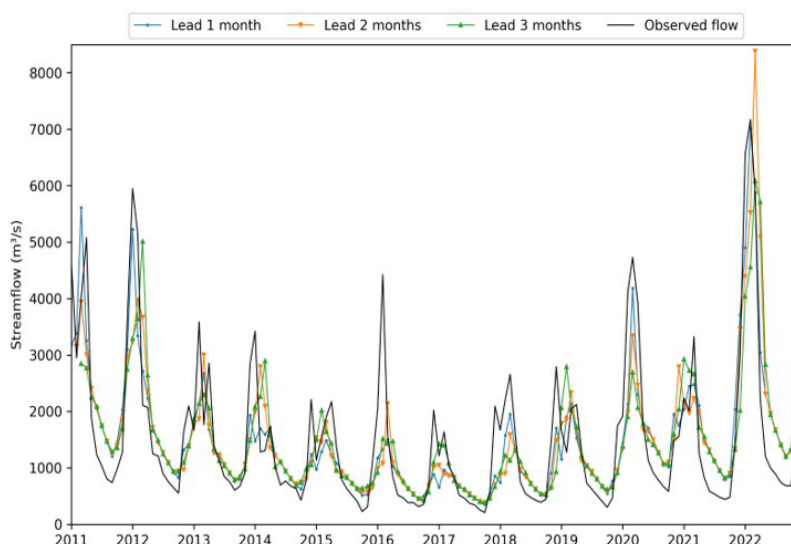


Figure 15. Comparison between the forecasted and observed streamflow series using SMAP-NMME.

Table 8. Performance metrics for streamflow predictions using SMAP-XGBoost fed with bias-corrected precipitation forecasts from the NMME ensemble.

Lead time (months)	Correlation	NSE	BIAS (m <sup>3</sup> /s)	RMSE (m <sup>3</sup> /s)	MAE (m <sup>3</sup> /s)	MAPE (%)
1	0.84	0.70	-110.3	746.23	466.58	31.22
2	0.79	0.62	-103.45	834.4	565.59	43.91
3	0.56	0.26	-53.42	1165.76	872.24	85.77

Table 9. Performance metrics for streamflow predictions of SMAP and SMAP-XGBoost divided in quarters.

Model	Quarter	Correl	NSE	BIAS (m <sup>3</sup> /s)	RMSE (m <sup>3</sup> /s)	MAE (m <sup>3</sup> /s)	MAPE (%)
SMAP	DJF	0.73	0.14	-934.47	1459.98	1121.94	39.20
SMAP-XGBoost	DJF	0.62	0.26	-546.48	1353.19	1039.66	40.64
SMAP	MAM	0.74	0.51	8.96	895.72	642.21	35.35
SMAP-XGBoost	MAM	0.72	0.46	9.56	936.68	719.70	41.00
SMAP	JJA	0.86	-2.76	407.77	473.76	407.77	58.19
SMAP-XGBoost	JJA	0.78	0.06	35.71	236.98	183.31	24.49
SMAP	SON	0.72	0.07	226.15	367.48	313.72	55.75
SMAP-XGBoost	SON	0.66	0.20	74.99	342.26	206.68	38.47

Obs.: Descriptions of the quarters: DJF (December, January, and February), MAM (March, April, and May), JJA (June, July, and August), and SON (September, October, and November).

and NSE between June and November. The best prediction period for SMAP-XGBoost among the mentioned quarters was between June and August (JJA), with a correlation of 0.78 and MAPE of 24.49%. The model also obtained its lowest error metric values throughout this time period. The best evaluation period for SMAP was the March to May quarter (MAM), with a correlation of 0.72 and a MAPE of 35.35%. This quarter also featured the lowest BIAS value and the highest NSE in the model.

In conjunction with Figure 15, the table findings show that SMAP-XGBoost has a stronger predicting ability for low-flow periods in the basin. During periods of higher flows, the SMAP model performs just marginally better than the SMAP-XGBoost model, particularly in the December to February quarter.

## CONCLUSIONS

According to the results obtained in this study, the precipitation forecasts, considering the ensemble mean of the NMME models, showed a good fit to the observed series in the Sobradinho reservoir region. The model is able to capture both the seasonality of the series and the maximum and minimum precipitation events. Only in some atypical months with precipitation much higher than expected did the model have difficulty capturing the signal. The bias correction proved to be satisfactory in improving the overall performance of the NMME model for the study region and period analyzed.

The hydrological model SMAP, which is already widely used in Brazil, demonstrated its ability to represent the hydrological regime of the watershed. Simulations conducted by the model for the period from 2011 to 2022 with observed data showed that the model has the capability to represent the hydrological regime of the watershed, even under unusual conditions. During the mentioned period, there was a severe drought, which a condition not observed during the parameter calibration period, yet the model obtained appropriate values for the analyzed metrics.

The analysis of residuals showed that although SMAP had a low bias (close to zero) for the data related to the Sobradinho watershed, the distribution of residuals had significant left-skewness. This skewness indicates that the data generated by the model tends to underestimate the observed flow values. The application of the hybrid model SMAP-XGBoost proved to be capable of reducing this skewness.

The hybrid model SMAP-XGBoost obtained better results than the SMAP model in terms of MAE and MAPE, specially to forecasts until 2-month ahead. Considering the average to these metrics the improvement was near 10% to 1-month ahead and 5% to 2-month ahead. However, the main improvement occurred for flow-flow periods, where the SMAP model tends to overestimate observed values. This model stands out for forecasts made between the months of June and November, it is during this period that it has its best metrics and evaluation, surpassing the results obtained with SMAP.

The results obtained demonstrate that a hybrid approach, combining well-established hydrological models with machine learning models, can generate better results than traditional standalone models. Considering the many decisions related to water resources management are made during low-flow periods, as well as for other flow regimes, having more accurate forecasts

for these periods can assist in better reservoir and watershed management as a whole.

## ACKNOWLEDGEMENTS

We thank CAPES for funding the scholarships and POSDEHA from UFC for institutional support.

## REFERENCES

- Abdolmanafi, A., Saghafian, B., & Aminyavari, S. (2021). Evaluation of global ensemble prediction models for forecasting medium to heavy precipitations. *Meteorology and Atmospheric Physics*, 133(1), 15-26. <http://doi.org/10.1007/s00703-020-00731-8>.
- Adnan, R. M., Liang, Z., Trajkovic, S., Zounemat-Kermani, M., Li, B., & Kisi, O. (2019). Daily streamflow prediction using optimally pruned extreme learning machine. *Journal of Hydrology*, 577, 123981. <http://doi.org/10.1016/j.jhydrol.2019.123981>.
- Adnan, R. M., Mostafa, R. R., Elbeltagi, A., Yaseen, Z. M., Shahid, S., & Kisi, O. (2022). Development of new machine learning model for streamflow prediction: case studies in Pakistan. *Stochastic Environmental Research and Risk Assessment*, 36(4), 999-1033. <http://doi.org/10.1007/s00477-021-02111-z>.
- Adnan, R. M., Yuan, X., Kisi, O., Adnan, M., & Mehmood, A. (2018). Stream flow forecasting of poorly gauged mountainous watershed by least square support vector machine, fuzzy genetic algorithm and M5 model tree using climatic data from nearby station. *Water Resources Management*, 32(14), 4469-4486. <http://doi.org/10.1007/s11269-018-2033-2>.
- Akbarian, M., Saghafian, B., & Golian, S. (2023). Monthly streamflow forecasting by machine learning methods using dynamic weather prediction model outputs over Iran. *Journal of Hydrology*, 620, 129480. <http://doi.org/10.1016/j.jhydrol.2023.129480>.
- American Society of Civil Engineers – ASCE. (1993). Criteria for evaluation of watershed models. *Journal of Irrigation and Drainage Engineering*, 119(3), 429-442. [http://doi.org/10.1061/\(ASCE\)0733-9437\(1993\)119:3\(429\)](http://doi.org/10.1061/(ASCE)0733-9437(1993)119:3(429)).
- Andrian, L. G., Osman, M., & Vera, C. S. (2023). Climate predictability on seasonal timescales over South America from the NMME models. *Climate Dynamics*, 60(11-12), 3261-3276. <http://doi.org/10.1007/s00382-022-06506-8>.
- Bárdossy, A., & Pegram, G. (2011). Downscaling precipitation using regional climate models and circulation patterns toward hydrology. *Water Resources Research*, 47(4), 2010WR009689. <http://doi.org/10.1029/2010WR009689>.
- Battisti, R., Bender, F. D., & Sentelhas, P. C. (2019). Assessment of different gridded weather data for soybean yield simulations in Brazil. *Theoretical and Applied Climatology*, 135(1-2), 237-247. <http://doi.org/10.1007/s00704-018-2383-y>.

- Belayneh, A., Adamowski, J., Khalil, B., & Ozga-Zielinski, B. (2014). Long-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet neural network and wavelet support vector regression models. *Journal of Hydrology*, 508, 418-429. <http://doi.org/10.1016/j.jhydrol.2013.10.052>.
- Bender, F. D., & Sentelhas, P. C. (2018). Solar radiation models and gridded databases to fill gaps in weather series and to project climate change in Brazil. *Advances in Meteorology*, 2018, 1-15. <http://doi.org/10.1155/2018/6204382>.
- Billerbeck, C., Silva, L. M. D., Marcellini, S. S., & Mélo Junior, A. (2021). Multi-criteria decision framework to evaluate bias corrected climate change projections in the Piracicaba River Basin. *Revista Brasileira de Meteorologia*, 36(3), 339-349. <http://doi.org/10.1590/0102-77863630068>.
- Birikundavyi, S., Labib, R., Trung, H. T., & Rousselle, J. (2002). Performance of neural networks in daily streamflow forecasting. *Journal of Hydrologic Engineering*, 7(5), 392-398. [http://doi.org/10.1061/\(ASCE\)1084-0699\(2002\)7:5\(392\)](http://doi.org/10.1061/(ASCE)1084-0699(2002)7:5(392)).
- Block, P. J., Souza Filho, F. A., Sun, L., & Kwon, H.-H. (2009). A streamflow forecasting framework using multiple climate and hydrological models. *Journal of the American Water Resources Association*, 45(4), 828-843. <http://doi.org/10.1111/j.1752-1688.2009.00327.x>.
- Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, 26(2), 211-243. <http://doi.org/10.1111/j.2517-6161.1964.tb00553.x>.
- Brasil. Operador Nacional do Sistema Elétrico – ONS. (2017). *Submódulo 23.5: critérios para estudos hidrológicos*. Brasília: ONS. Retrieved in 2023, August 23, from <https://www.ons.org.br>
- Cavalcante, M. R. G., Cunha Luz Barcellos, P., & Cataldi, M. (2020). Flash flood in the mountainous region of Rio de Janeiro state (Brazil) in 2011: part I—calibration watershed through hydrological SMAP model. *Natural Hazards*, 102(3), 1117-1134. <http://doi.org/10.1007/s11069-020-03948-3>.
- Chen, T., & Guestrin, C. (2016). XGBoost: a scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785-794). San Francisco: ACM. <http://doi.org/10.1145/2939672.2939785>.
- Cheng, C.-T., Zhao, M.-Y., Chau, K. W., & Wu, X.-Y. (2006). Using genetic algorithm and TOPSIS for Xinanjiang model calibration with a single procedure. *Journal of Hydrology*, 316(1-4), 129-140. <http://doi.org/10.1016/j.jhydrol.2005.04.022>.
- Cheng, M., Fang, F., Kinouchi, T., Navon, I. M., & Pain, C. C. (2020). Long lead-time daily and monthly streamflow forecasting using machine learning methods. *Journal of Hydrology*, 590, 125376. <http://doi.org/10.1016/j.jhydrol.2020.125376>.
- Downer, C. W., & Ogden, F. L. (2004). GSSHA: model to simulate diverse stream flow producing processes. *Journal of Hydrologic Engineering*, 9(3), 161-174. [http://doi.org/10.1061/\(ASCE\)1084-0699\(2004\)9:3\(161\)](http://doi.org/10.1061/(ASCE)1084-0699(2004)9:3(161)).
- Duarte, Y. C. N., & Sentelhas, P. C. (2020). NASA/POWER and DailyGridded weather datasets: how good they are for estimating maize yields in Brazil? *International Journal of Biometeorology*, 64(3), 319-329. <http://doi.org/10.1007/s00484-019-01810-1>.
- Eletrobras Chesf. (2023). *Descrição do aproveitamento de Sobradinho*. Retrieved in 2023, August 23, from <https://www.chesf.com.br/SistemaChesf/Pages/SistemaGeracao/Sobradinho.aspx>
- Flores, J. P. O. (2021). *Avaliação da previsão sazonal de precipitação do projeto North America Multi-Model Ensemble (NMME) sobre o Brasil* (Dissertação de mestrado). Universidade de São Paulo, São Paulo. <http://doi.org/10.11606/D.14.2021.tde-31052021-115217>.
- Gondim, R., Silveira, C., Souza Filho, F., Vasconcelos, F., & Cid, D. (2018). Climate change impacts on water demand and availability using CMIP5 models in the Jaguaribe basin, semi-arid Brazil. *Environmental Earth Sciences*, 77(15), 550. <http://doi.org/10.1007/s12665-018-7723-9>.
- Hadi, S. J., & Tombul, M. (2018). Forecasting daily streamflow for basins with different physical characteristics through data-driven methods. *Water Resources Management*, 32(10), 3405-3422. <http://doi.org/10.1007/s11269-018-1998-1>.
- Harris, I., Osborn, T. J., Jones, P., & Lister, D. (2020). Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Scientific Data*, 7(1), 109. <http://doi.org/10.1038/s41597-020-0453-3>.
- Johnson, M. S., Coon, W. F., Mehta, V. K., Steenhuis, T. S., Brooks, E. S., & Boll, J. (2003). Application of two hydrologic models with different runoff mechanisms to a hillslope dominated watershed in the northeastern US: a comparison of HSPF and SMR. *Journal of Hydrology*, 284(1-4), 57-76. <http://doi.org/10.1016/j.jhydrol.2003.07.005>.
- Kachitvichyanukul, V. (2012). Comparison of three evolutionary algorithms: GA, PSO, and DE. *Industrial Engineering and Management Systems*, 11(3), 215-223. <http://doi.org/10.7232/iems.2012.11.3.215>.
- Kirtman, B. P., Min, D., Infanti, J. M., Kinter III, J. L., Paolino, D. A., Zhang, Q., van den Dool, H., Saha, S., Mendez, M. P., Becker, E., Peng, P., Tripp, P., Huang, J., DeWitt, D. G., Tippet, M. K., Barnston, A. G., Li, S., Rosati, A., Schubert, S. D., Rienecker, M., Suarez, M., Li, Z. E., Marshak, J., Lim, Y.-K., Tribbia, J., Pegion, K., Merryfield, W. J., Denis, B., & Wood, E. F. (2014). The North American Multimodel Ensemble: phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction. *Bulletin of the American Meteorological Society*, 95(4), 585-601. <http://doi.org/10.1175/BAMS-D-12-00050.1>.
- Knoben, W. J. M., Freer, J. E., & Woods, R. A. (2019). Technical note: inherent benchmark or not? Comparing Nash-Sutcliffe and Kling-Gupta efficiency scores. *Hydrology and Earth System Sciences*, 23(10), 4323-4331. <http://doi.org/10.5194/hess-23-4323-2019>.



- Li, M., Yang, D., Chen, J., & Hubbard, S. S. (2012). Calibration of a distributed flood forecasting model with input uncertainty using a Bayesian framework. *Water Resources Research*, 48(8), 2010WR010062. <http://doi.org/10.1029/2010WR010062>.
- Li, W., Li, L., Fu, R., Deng, Y., & Wang, H. (2011). Changes to the North Atlantic subtropical high and its role in the intensification of summer rainfall variability in the southeastern United States. *Journal of Climate*, 24(5), 1499-1506. <http://doi.org/10.1175/2010JCLI3829.1>.
- Liu, J., Ren, K., Ming, T., Qu, J., Guo, W., & Li, H. (2022). Investigating the effects of local weather, streamflow lag, and global climate information on 1-month-ahead streamflow forecasting by using XGBoost and SHAP: two case studies involving the contiguous USA. *Acta Geophysica*, 71(2), 905-925. <http://doi.org/10.1007/s11600-022-00928-y>.
- Liu, Z., Zhou, P., Chen, X., & Guan, Y. (2015). A multivariate conditional model for streamflow prediction and spatial precipitation refinement. *Journal of Geophysical Research. Atmospheres*, 120(19), <http://doi.org/10.1002/2015JD023787>.
- Ljung, G. M., & Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2), 297-303. <http://doi.org/10.1093/biomet/65.2.297>.
- Lopes, J. E. G., Braga, B. P. F., & Conejo, J. G. L. (1982). SMAP: a simplified hydrological model. In V. P. Singh (Ed.), *Applied modeling in catchment hydrology*. Littleton, CO: Water Resources Publications.
- Ma, M., Zhao, G., He, B., Li, Q., Dong, H., Wang, S., & Wang, Z. (2021). XGBoost-based method for flash flood risk assessment. *Journal of Hydrology*, 598, 126382. <http://doi.org/10.1016/j.jhydrol.2021.126382>.
- Maciel, G. M., Cabral, V. A., Marcato, A. L. M., Júnior, I. C. S., & Honório, L. D. M. (2020). Daily water flow forecasting via coupling between SMAP and deep learning. *IEEE Access: Practical Innovations, Open Solutions*, 8, 204660-204675. <http://doi.org/10.1109/ACCESS.2020.3036487>.
- Maraun, D. (2013). Bias correction, quantile mapping, and downscaling: revisiting the inflation issue. *Journal of Climate*, 26(6), 2137-2143. <http://doi.org/10.1175/JCLI-D-12-00821.1>.
- Marini, F., & Walczak, B. (2015). Particle Swarm Optimization (PSO): a tutorial. *Chemometrics and Intelligent Laboratory Systems*, 149, 153-165. <http://doi.org/10.1016/j.chemolab.2015.08.020>.
- Meng, E., Huang, S., Huang, Q., Fang, W., Wu, L., & Wang, L. (2019). A robust method for non-stationary streamflow prediction based on improved EMD-SVM model. *Journal of Hydrology*, 568, 462-478. <http://doi.org/10.1016/j.jhydrol.2018.11.015>.
- Miao, C., Su, L., Sun, Q., & Duan, Q. (2016). A nonstationary bias-correction technique to remove bias in GCM simulations: bias-correction in the GCM simulation. *Journal of Geophysical Research. Atmospheres*, 121(10), 5718-5735. <http://doi.org/10.1002/2015JD024159>.
- Miranda, N. M., Cataldi, M., & Silva, F. N. R. (2017). Simulação do regime hidrológico da cabeceira do rio São Francisco a partir da utilização dos modelos SMAP e RegCM. *Anuário do Instituto de Geociências*, 40(3), 328-339.
- Mo, K. C., & Lyon, B. (2015). Global meteorological drought prediction using the North American multi-model ensemble. *Journal of Hydrometeorology*, 16(3), 1409-1424. <http://doi.org/10.1175/JHM-D-14-0192.1>.
- Mutti, P. R., Dubreuil, V., Bezerra, B. G., Arvor, D., Oliveira, C. P., & Santos e Silva, C. M. (2020). Assessment of gridded CRU TS data for long-term climatic water balance monitoring over the São Francisco Watershed, Brazil. *Atmosphere*, 11(11), 1207. <http://doi.org/10.3390/atmos11111207>.
- Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I: a discussion of principles. *Journal of Hydrology*, 10(3), 282-290. [http://doi.org/10.1016/0022-1694\(70\)90255-6](http://doi.org/10.1016/0022-1694(70)90255-6).
- Ni, L., Wang, D., Wu, J., Wang, Y., Tao, Y., Zhang, J., & Liu, J. (2020). Streamflow forecasting using extreme gradient boosting model coupled with Gaussian mixture model. *Journal of Hydrology*, 586, 124901. <http://doi.org/10.1016/j.jhydrol.2020.124901>.
- Niu, W., Feng, Z., Zeng, M., Feng, B., Min, Y., Cheng, C., & Zhou, J. (2019). Forecasting reservoir monthly runoff via ensemble empirical mode decomposition and extreme learning machine optimized by an improved gravitational search algorithm. *Applied Soft Computing*, 82, 105589. <http://doi.org/10.1016/j.asoc.2019.105589>.
- Parisouj, P., Mohebzadeh, H., & Lee, T. (2020). Employing machine learning algorithms for streamflow prediction: a case study of four river basins with different climatic zones in the United States. *Water Resources Management*, 34(13), 4113-4131. <http://doi.org/10.1007/s11269-020-02659-5>.
- Pereira, S. B. (2004). *Evaporação no lago de Sobradinho e disponibilidade hídrica no rio São Francisco* (Tese de doutorado). Universidade Federal de Viçosa, Viçosa. Retrieved in 2023, August 23, from <http://www.locus.ufv.br/handle/123456789/9701>
- Ratanamahatana, C. A., Lin, J., Gunopulos, D., Keogh, E., Vlachos, M., & Das, G. (2009). Mining time series data. In O. Maimon, & L. Rokach (Eds.), *Data mining and knowledge discovery handbook* (pp. 1049-1077). Boston: Springer US. [http://doi.org/10.1007/978-0-387-09823-4\\_56](http://doi.org/10.1007/978-0-387-09823-4_56).
- Regonda, S. K., Rajagopalan, B., Clark, M., & Zagona, E. (2006). A multimodel ensemble forecast framework: application to spring seasonal flows in the Gunnison River Basin. *Water Resources Research*, 42(9), 2005WR004653. <http://doi.org/10.1029/2005WR004653>.
- Rocha Júnior, R. L., Pinto, D. D. C., Silva, F. D. S., Gomes, H. B., Barros Gomes, H., Costa, R. L., Pereira, M. P. S., Peña, M., Coelho, C. A. S., & Herdies, D. L. (2021). An empirical seasonal



- rainfall forecasting model for the northeast region of Brazil. *Water*, 13(12), 1613. <http://doi.org/10.3390/w13121613>
- Santos, C., Rocha, F., Ramos, T., Alves, L., Mateus, M., Oliveira, R., & Neves, R. (2019). Using a hydrologic model to assess the performance of regional climate models in a semi-arid watershed in Brazil. *Water*, 11(1), 170. <http://doi.org/10.3390/w11010170>.
- Shukla, S., Safeeq, M., AghaKouchak, A., Guan, K., & Funk, C. (2015). Temperature impacts on the water year 2014 drought in California. *Geophysical Research Letters*, 42(11), 4384-4393. <http://doi.org/10.1002/2015GL063666>.
- Silva Filho, A. M. (2014). *Autocorrelação e correlação cruzada: teorias e aplicações* (Tese de doutorado). SENAI-CIMATEC, Salvador.
- Silva, F. D. N. R., Alves, J. L. D., & Cataldi, M. (2019). Climate downscaling over South America for 1971-2000: application in SMAP rainfall-runoff model for Grande River Basin. *Climate Dynamics*, 52(1-2), 681-696. <http://doi.org/10.1007/s00382-018-4166-7>.
- Silva, J. F. (2018). *Análise espaço-temporal das áreas inundáveis do Reservatório de Sobradinho na Bacia Hidrográfica do Rio São Francisco* (Dissertação de mestrado). Universidade Federal de Pernambuco, Recife. Retrieved in 2023, August 23, from <https://repositorio.ufpe.br/handle/123456789/31942>
- Silveira, C. D. S., Souza Filho, F. D. A. D., & Vasconcelos Júnior, F. D. C. (2017). Streamflow projections for the Brazilian hydropower sector from RCP scenarios. *Journal of Water and Climate Change*, 8(1), 114-126. <http://doi.org/10.2166/wcc.2016.052>.
- Sobral, M. C. M., Assis, J. M. O., Oliveira, C. R., Silva, G. M. N., Moraes, M., & Carvalho, R. M. C. (2018). Impacto das mudanças climáticas nos recursos hídricos no submédio da bacia hidrográfica do Rio São Francisco – Brasil. *Revista Eletrônica do PRODEMA*, 12(3), 95-106. <https://doi.org/10.22411/rede2018.1203.10>.
- Szczepanek, R. (2022). Daily streamflow forecasting in mountainous catchment using XGBoost, LightGBM and CatBoost. *Hydrology*, 9(12), 226. <http://doi.org/10.3390/hydrology9120226>.
- United States Department of Agriculture – USDA. (1986). *Urban hydrology for small watersheds* (2nd ed). Washington, D.C.: U.S. Department of Agriculture, Soil Conservation Service, Engineering Division.
- Vollmer, M. K., Bootsma, H. A., Hecky, R. E., Patterson, G., Halfman, J. D., Edmond, J. M., Eccles, D. H., & Weiss, R. F. (2005). Deep-water warming trend in Lake Malawi, East Africa. *Limnology and Oceanography*, 50(2), 727-732. <http://doi.org/10.4319/lo.2005.50.2.0727>.
- Wang, J., Yang, B., Ljungqvist, F. C., & Zhao, Y. (2013). The relationship between the Atlantic Multidecadal Oscillation and temperature variability in China during the last millennium. *Journal of Quaternary Science*, 28(7), 653-658. <http://doi.org/10.1002/jqs.2658>.
- Xavier, A. C., King, C. W., & Scanlon, B. R. (2016). Daily gridded meteorological variables in Brazil (1980-2013). *International Journal of Climatology*, 36(6), 2644-2659. <http://doi.org/10.1002/joc.4518>.
- Xavier, A. C., Scanlon, B. R., King, C. W., & Alves, A. I. (2022). New improved Brazilian daily weather gridded data (1961-2020). *International Journal of Climatology*, 42(16), 8390-8404. <http://doi.org/10.1002/joc.7731>.
- Yang, S., Yang, D., Chen, J., Santisirisomboon, J., Lu, W., & Zhao, B. (2020). A physical process and machine learning combined hydrological model for daily streamflow simulations of large watersheds with limited observation data. *Journal of Hydrology*, 590, 125206. <http://doi.org/10.1016/j.jhydrol.2020.125206>.
- Yaseen, Z. M., Sulaiman, S. O., Deo, R. C., & Chau, K.-W. (2019). An enhanced extreme learning machine model for river flow forecasting: state-of-the-art, practical applications in water resource engineering area and future research direction. *Journal of Hydrology*, 569, 387-408. <http://doi.org/10.1016/j.jhydrol.2018.11.069>.
- Yeo, I.-K., & Johnson, R. A. (2000). A new family of power transformations to improve normality or symmetry. *Biometrika*, 87(4), 954-959. <http://doi.org/10.1093/biomet/87.4.954>.

### Authors contributions

Antonio Duarte Marcos Junior: Main author.

Cleiton da Silva Silveira: Review and research advisor.

José Micael Ferreira da Costa: Text review and collaboration in editing figures.

Suellen Teixeira Nobre Gonçalves: Text review and results.

**Editor-in-Chief:** Adilson Pinheiro

**Associated Editor:** Fernando Mainardi Fan