

<https://doi.org/10.1590/2318-0331.282320230120>

Enhancing streamflow forecasting for the Brazilian electricity sector: a strategy based on a hyper-multimodel

*Aprimorando a previsão de vazão para o setor elétrico brasileiro:
uma estratégia baseada em um hiper-multimodelo*

Francisco de Assis de Souza Filho¹ , Renan Vieira Rocha² , Ályson Brayner Estácio^{1,2} , Larissa Zaira Rafael Rolim¹ ,
João Dehon de Araújo Pontes Filho² , Victor Costa Porto¹  & Sullyandro Oliveira Guimarães^{3,4,5} 

¹Universidade Federal do Ceará, Fortaleza, CE, Brasil

²Fundação Cearense de Meteorologia e Recursos Hídricos, Fortaleza, CE, Brasil

³Potsdam-Institut für Klimafolgenforschung, Potsdam, Deutschland

⁴Leibniz-Gemeinschaft, Potsdam, Deutschland

⁵Universität Potsdam, Potsdam, Deutschland

E-mails: assis@ufc.br (FASF), renan.rocha@funceme.br (RVR), alyson.estacio@funceme.br (ÁBE), larissazairarr@gmail.com (LZRR), joao.pontes@funceme.br (JDAPF), victorporto@gmail.com (VCP), sullyandro@pik-potsdam.de (SOG)

Received: October 06, 2023 – Revised: October 13, 2023 – Accepted: October 14, 2023

ABSTRACT

Streamflow forecasting plays an important role in ensuring the reliable supply of electricity in countries heavily reliant on hydropower. This paper proposes a novel framework that integrates various hydrological models, climate models, and observational data to develop a comprehensive forecasting system. Three families of models were employed: seasonal forecasting climate models integrated with hydrological rainfall-runoff models; stochastic or machine learning models utilizing endogenous variables, and stochastic or machine learning models that consider exogenous variables. The hyper-multimodel framework could successfully increase the overall performance of the scenarios generated through the use of the individual models. The quality of the final scenarios generated was directly connected to the performance of the individual models. Therefore, the proposed framework has potential to improve hydrological forecast for the Brazilian electricity sector with the use of more refined and calibrated individual models.

Keywords: Hydrological modeling; Machine learning; Periodic autoregressive models; Flow scenario generation.

RESUMO

A previsão de vazão desempenha um papel importante para garantir o fornecimento confiável de eletricidade em países altamente dependentes da energia hidrelétrica. Este artigo propõe uma abordagem inovadora que integra vários modelos hidrológicos, modelos climáticos e dados observacionais para desenvolver um sistema abrangente de previsão. Três famílias de modelos foram empregadas: modelos climáticos de previsão sazonal integrados com modelos hidrológicos chuva-vazão; modelos estocásticos ou de aprendizado de máquina que utilizam variáveis endógenas e modelos estocásticos ou de aprendizado de máquina que consideram variáveis exógenas. A abordagem do hiper-multimodelo conseguiu aumentar com sucesso o desempenho geral dos cenários gerados com uso dos modelos individuais. A qualidade dos cenários finais gerados estava diretamente relacionada ao desempenho dos modelos individuais. Portanto, a estrutura proposta tem o potencial de melhorar a previsão hidrológica para o setor de eletricidade brasileiro com o uso de modelos individuais mais refinados e calibrados.

Palavras-chave: Modelagem hidrológica; Aprendizado de máquina; Modelos autoregressivos periódicos; Geração de cenários de vazão.

INTRODUCTION

Hydrological forecasting, particularly medium and long-term streamflow forecasting, is vital for effective water resource management. Forecasts over varying time scales provide critical insights into flood control, power generation, water supply, and drought mitigation (Kim et al., 2001; Jiang et al., 2018; Wang et al., 2019). To enhance the accuracy of these forecasts, researchers have explored various forecasting models and factors.

In recent years, a range of strategies have been developed to improve the accuracy of streamflow predictions, incorporating physical-based methodologies, mathematical statistical analyses, and artificial intelligence techniques (Zhang et al., 2018). However, these methods often fall short in predicting hydrological extreme events, due to propagating uncertainties, associated with hydrology and land processes (Chevuturi et al., 2023).

Hydrological models, which can be dynamical, empirical, or data-driven, largely rely on detailed local data, such as soil characteristics and geological features. In regions where this data is sparse, these models face uncertainties because they have to depend on broader global datasets that may not accurately represent the local conditions. However, research suggests that, in certain situations, averaging the outputs of several models can reasonably approximate a locally calibrated model (Chevuturi et al., 2023).

Research indicates that multi-model hydrological forecasting is a viable approach for reliable predictions in various regions, according to numerous studies (e.g., Velazquez et al., 2011; Wanders & Wood, 2016). Combining forecasts from multiple models, a strategy that has proven effective regionally (e.g., Ajami et al., 2006), can potentially offer comprehensive global hydrological predictions. Several multi-model blending techniques exist, each striving to enhance forecast accuracy by capitalizing on the strengths of some models while ignoring the weaknesses of others (e.g., Shamseldin et al., 1997; Roy et al., 2020).

A fundamental approach involves averaging all model outputs, sometimes incorporating model simulation standardization to negate forecast biases (Georgakakos et al., 2004; Ajami et al., 2006). This method, however, does not maximize the potential benefits derived from the high-performing models selectively. To overcome this, weighted averaging methods can be employed, wherein weights are assigned to different models based on a variety of estimation techniques, such as multiple linear regression (Wanders & Wood, 2016) and machine learning methods (Zaherpour et al., 2019), among others. These techniques essentially reward proficient models while penalizing less effective ones (Arsenault et al., 2015).

In countries heavily reliant on hydropower, like Brazil, hydrological forecasting plays an important role in ensuring the reliable supply of electricity and meeting the ever-growing electricity demands. However, the sustainable operation and management of these hydropower facilities depend on accurate streamflow forecasting, which, in turn, is influenced by the complex interactions of climate and hydrological processes.

This paper addresses the critical need to enhance streamflow forecasting for the Brazilian electricity sector, aiming to provide improved tools and insights for decision-makers and stakeholders involved in energy production and management. The proposed strategy is grounded in a hyper-multimodel forecasting framework, which harnesses the power of cutting-edge modeling

techniques and climate data to bolster the accuracy and reliability of streamflow predictions.

Brazil's geographical vastness and climatic diversity pose a significant challenge when it comes to forecasting streamflow patterns. The country spans multiple climate regions, from the Amazon rainforest to the semi-arid Northeast, each with its unique hydrological characteristics and sensitivities to climate variability and change. Furthermore, the irregular occurrence of extreme climate events, such as droughts and heavy rainfall, further complicates the use of a unique modeling strategy for streamflow forecasts.

In response to these challenges, the Brazilian electrical sector has been working, in collaboration with academic institutions and research centers, to improve the accuracy of models used for the prediction and generation of streamflow scenarios. It was established as an activity conducted within the Technical Committee of PMO/PLD (CT PMO/PLD), by the Hydrological Scenario Representation Working Group (GT-CH), coordinated by the National Operator of Electrical System (ONS) and the Chamber of Electric Energy Commercialization (CCEE). This activity aimed to investigate advanced streamflow scenario generation models for horizons of up to one year.

Therefore, this paper, as part of this initiative, proposes a novel approach that integrates various hydrological models, climate models, and observational data to develop a comprehensive forecasting system. Through this research, we aim to contribute to improving streamflow forecasting in complex regions such as Brazil.

METHODOLOGY

The proposal is to employ a hyper-multimodel that facilitates the combination of the strengths of different paradigms of river flow modeling. This strategy is built on three families of models, aiming to create synthetic flow scenarios. The three families of models are: seasonal forecasting climate models that are integrated with hydrological rainfall-runoff models; stochastic or machine learning models utilizing endogenous variables, which employ flow data to forecast future flows; and stochastic or machine learning models that consider exogenous variables, leveraging both flow data and climate indices for forecasting flows. This approach strives to harmonize various modeling strategies to enhance the reliability and precision of the predicted flow outputs. The subsequent sections offer a brief overview of the different approaches adopted for streamflow forecasting, the data applied, the details of the model families utilized in this study, and the hyper-multi-model methodology.

Overview

Rainfall-runoff modeling approaches used in forecasting streamflow generally fall into two categories: physical or process-based models, and empirical or statistical models. Physical or process-based models strive to incorporate the relevant physical laws that control watershed responses and the generation of streamflow, leveraging a substantial amount of observed data. Although these models are generally reliable and can offer credible results, they also entail a great deal of uncertainty, stemming from variable input data and complex numerical techniques. On the

other hand, empirical or statistical models aim to replicate the relationships between inputs, such as precipitation, and outputs, such as streamflow, without delving into an understanding of the internal processes involved. Despite being relatively simpler to develop and capable of providing reliable forecasts when grounded in robust data, these models face their own set of challenges, including identifying vital input variables and navigating parameterization issues.

In scenarios where only univariate time series data are accessible, it is necessary to rely on historical data to forecast streamflow. The scientific interest in utilizing historical data to generate statistical predictions for univariate time series, or “forecasting”, has been a long-standing topic in the field of hydrology (Yevjevich, 1987). The complexity of these time series, coupled with our limited understanding of the exact governing equations required in deterministic models, led to a shift towards the development and application of methods rooted in probability and statistics for modelling and prediction of these time series. Stochastic methods came into prominence as they aim at probabilistic prediction or estimation of data, emphasizing statistical characteristics of the data like mean, standard deviation, and variance, while also accounting for uncertainty in these predictions. A significant evolution in stochastic methods occurred in the 1970s, largely influenced by Box & Jenkins (1970). Models like the autoregressive (AR), moving average (MA), autoregressive moving average (ARMA), Markov chain, and point process models have seen extensive use in many fields, including hydrology. In order to integrate periodic patterns such as seasonality in rainfall and streamflow into models, several periodic versions of models have been proposed, as detailed extensively in works by Brockwell & Davis (1991), and Salas et al. (1980). Notably, periodic variants of AR (PAR), ARMA (PARMM), and GAR (PGAR) models have gained significant attention. These models have seen broad use in hydrological applications, as elaborated by Loucks et al. (1981).

Contrasting traditional models, the realm of nonlinear models, specifically Machine Learning (ML) algorithms, often referred to as “black-box models”, introduces a different perspective in statistical modelling. They embody what is known as the algorithmic modelling culture (Breiman, 2001). While the traditional, or data modelling culture, hinges on the assumption that the data generation process is underpinned by an analytically formulated stochastic model, the algorithmic modelling culture operates on a fundamentally different principle. It presumes the underlying process to be complex and potentially unknown, not necessarily requiring an analytically defined model. The primary concern here is that the algorithmic model must be able to deliver high forecast accuracy, rather than necessarily understanding or

accurately representing the future behavior of a process. In this context, the principles of understanding, modelling, and predicting a process’s behavior, which are crucial in the data modelling culture, become less important. This approach emphasizes the algorithm’s ability to accurately predict outcomes, regardless of the complexity or understandability of the process itself.

Data

This case study used the natural monthly streamflow series from 146 hydropower plants (HPPs) that were provided by ONS for the period of 1931 to 2021. The natural monthly streamflow series were transformed to incremental series.

Seasonal forecasting climate models that are integrated with hydrological rainfall-runoff models

The dynamic seasonal streamflow forecast approach where climate and hydrological process are coupled by the forcing of hydrological models with numerical predictions from Atmospheric General Circulation Models (GCMs) (Block et al., 2009; Kwon et al., 2012; Ávila et al., 2023; Greuell & Hutjes, 2023) differs from the statistical approaches and offers the benefit of integrating climate prediction and its efficacy improves with advancements in our comprehension of climate processes and the development of atmospheric modeling (Block et al., 2009).

In this work, the seasonal rainfall forecasts from the GCMs that compose the North American Multi-Model Ensemble (NMME) were considered and are detailed in Table 1. The forecasts are publicly available on a global scale in the NMME project webpage, represented on a grid with a resolution of 1° latitude by 1° longitude. The lead times of these forecasts range from 0.5 to 11.5 months, providing valuable information for medium to long-term climate predictions.

The NMME seasonal rainfall forecasts were biased corrected by a gamma quantile mapping for each month and then were interpolated into catchment scale by the Thiesen polygon Method. After this statistical treatment, the rainfall forecasts are used to force the Soil Moisture Accounting Procedure (SMAP) hydrological model to generate streamflow forecasts.

The SMAP model was previously calibrated for each catchment with observed streamflow and rainfall data by using the DREAM (DiffeREntial Evolution Adaptive Metropolis) algorithm (Vrugt et al., 2009) to minimize the Root Mean Squared Error (RMSE). The hydrological model showed good skill in most of the basins for both the calibration and validation sets.

Table 1. Summary of the active NMME models.

Model	Institution	Hindcast Period	Lead time (months)	Members
CanCM4i-IC3	CMC	[Jan 1980, Nov 2021]	0.5-11.5	10
GEM5-NEMO	CMC	[Jan 1980, Nov 2021]	0.5-11.5	10
GFDL-SPEAR	GFDL	[Jan 1991, Nov 2021]	0.5-11.5	30
NASA-GEOSS2S	NASA	[Feb 1981, Sep 2017]	0.5-9.5	4
COLA-RSMAS-CCSM4	NCAR	[Jan 1991, Nov 2021]	0.5-11.5	10
NCEP-CFSv2	NCEP	[Jan 1982, Sep 2011]	0.5-9.5	24

Stochastic or machine learning models utilizing endogenous variables

Generally, hydrological series with a duration shorter than a year, like monthly series, primarily exhibit periodic behavior in their probabilistic properties, including mean, variance, skewness, and autocorrelation structure. Thus, the stochastic models, namely the periodic auto-regressive model (PAR), the periodic auto-regressive model with an annual component (PAR-A), and the wavelet autoregressive model (WARM) were applied in this paper. The PAR(p) stochastic model, as referenced by Salas et al. (1980) can be written as:

$$\left(\frac{Z_t - \mu_m}{\sigma_m}\right) = \phi^m \left(\frac{Z_{t-1} - \mu_{m-1}}{\sigma_{m-1}}\right) + \dots + \phi_{p_m}^m \left(\frac{Z_{t-p_m} - \mu_{m-p_m}}{\sigma_{m-p_m}}\right) + \alpha_t \quad (1)$$

Let Z_t (where $t=1,2,\dots$ represent the seasonal time series with a period s . In this context, s equals to 12 since we are working with a monthly time series. N denotes the number of years and t is a function of the year T (where $T=1,\dots,N$) and the season m ($m=1,\dots,s$). The term p_m represents the order of the model, and $\phi_{p_m}^m$ is the auto-regressive coefficient of order p_m . Furthermore, μ_m is the mean of the period m , σ_m is the standard deviation of the period m , and α_t is the time that is uncorrelated with zero mean and variance $\sigma_a^2(m)$.

The initial step in fitting a PAR(p) model entails determining the optimal order, p_m , for the auto-regressive operators associated with each period. The order of the model were selected using a stepwise regression. To accomplish this, we employed stepwise regression as a method to select the order of the model. Stepwise regression is a systematic method for adding or removing terms from a multilinear model based on their statistical significance in explaining the dependent variable. Once the relevant months were identified through this process, their corresponding coefficients were derived using linear regression.

The PAR(p)-A methodology (Treisman et al., 2020) proposes the inclusion of a new term in the periodic autoregressive model PAR(p), referring to the average of the last inflows until completing one year (A_{t-1}). The Periodic Autoregressive Model with Annual Component, here denoted as PAR-A(p), and can be written as:

$$\left(\frac{Z_{t,i} - \mu_{m,i}}{\sigma_{m,i}}\right) = \sum_{j=1}^{p_{m,i}} \phi_{t,j,i}^m \left(\frac{Z_{t-j,i} - \mu_{m-j,i}}{\sigma_{m-j,i}}\right) + \psi_{t,i}^m \left(\frac{A_{t-1,i} - \mu_{m-1,i}^A}{\sigma_{m-1,i}^A}\right) + \alpha_{t,i} \quad (2)$$

Where ψ^m is the annual component coefficient, μ_m^A and σ_m^A are the mean and standard deviation of the considered year, respectively.

The WARM employs an autoregressive model in conjunction with wavelet decomposition to simulate time series. Through wavelet transform, a time series y_t is decomposed by convolving the series with “daughter wavelets”, derived from translating the foundational “mother wavelet” over a time step τ and at a scale s . The component of the time series signal is identified through a significance test of 90-95%, using white noise as the null hypothesis. The significant component recognized between periods J_1 and J_2 is extracted from the original time series using the wavelet reconstruction function. The reconstruction of the original time series over a range of periods from J_1 and J_2 can be achieved as:

$$y_n' = \frac{\delta_j \delta_t^{1/2}}{C_\delta \psi_0(0)} \sum_{j=J_1}^{J_2} \frac{R[W_n(a_j)]}{a_j^{1/2}} \quad (3)$$

Where C_δ is the reconstruction factor; δ_j and δ_t are the scale and time factors, respectively; $\psi_0(0) = \pi^{-1/4}$ is the factor that removes the energy scale for the chosen mother wavelet function; $R[W_n(a_j)]$ represents the real part of the wavelet transform; a_j is the scale parameter; and J_1 and J_2 are the lower and upper periods, respectively, encompassing the desired range of scales. Further details about the wavelet transform method can be found in the works of Torrence & Compo (1998) and Kwon et al. (2007). For each reconstructed component an autoregressive model was adjusted to simulate the time series. Then, the simulated series were added. Note that this is a very simple additive model with no interactions across the noise or the reconstructed signals.

Additionally, machine learning models including Ridge regression and Multilayer Perceptron applied historical data to enhance streamflow forecasting. Ridge regression (Hastie et al., 2001), often known as Tikhonov, is a machine learning technique frequently employed for regression analysis within supervised learning scenarios. This method aims to reduce the effective number of parameters, primarily to prevent overfitting. Overfitting arises when a model becomes too attuned to the training data, compromising its performance on unfamiliar, unseen data. Ridge regression is often utilized in regression analysis to address challenges related to multicollinearity. Building upon the foundation of ordinary least squares (OLS) regression, it augments the loss function with a penalizing element. The mathematical representation of Ridge regression is as provided below:

$$\text{minimize} = \|Y - X\beta\|^2 + \lambda \|\beta\|^2 \quad (4)$$

Where Y represents the target variable, while X signifies the predictor variables. The coefficients are represented by β , and the regularization parameter, denoted by λ , determines the extent of shrinkage applied to the coefficients. The Euclidean norm is indicated by $\|\beta\|$. Ridge regression aims to minimize the squared differences between the predicted and actual values ($Y - X$), while simultaneously imposing a penalty on the magnitude of the coefficients ($\|\beta\|^2$) (Kumar et al., 2023).

The Multilayer Perceptron (MLP) is a type of artificial neural network comprising several layers of interconnected nodes, commonly referred to as neurons. This feed-forward neural network structure ensures data progresses from the input layer, through the hidden layers, and finally culminates at the output layer. At its core, each neuron within the MLP carries out a weighted sum of its input values, subjects that sum to an activation function, and then forwards the resulting output to subsequent layer neurons (Gardner & Dorling, 1998; Kumar et al., 2023). Situated between the input and output layers, the hidden layer processes data by forwarding it through its neurons. Unlike the input and output layers, the hidden layer’s processes aren’t directly accessible. The output of a given neuron j in the hidden layer can be mathematically represented as:

$$y_j = f\left(\sum_{i=1}^n w_{ij}x_i + b_j\right) \quad (5)$$

Where w_{ij} and b_j are the weights and biases associated with the neurons of the hidden layer. $f(\cdot)$ denotes a non-linear activation function, specifically the hyperbolic tangent sigmoid transfer function (tansig). The use of the hyperbolic tangent sigmoid transfer function, or ‘tansig’, as the activation function adds a non-linear twist to the proceedings. This ensures that the network can capture and represent intricate relationships in the data, which a linear function might miss.

Stochastic or machine learning models that consider exogenous variables

Furthermore, prominent climate indices have been utilized as input variables in distinct stochastic and machine learning models, notably PARX and Ridge-X. Detailed descriptions of the applied climate indices can be found in Table 2.

Hyper-multimodel

Figure 1 outlines our methodology for creating hyper-multimodel scenarios using results from three family models. We start by fitting each individual model for the period of 2001 to 2010, setting aside the period from 2011 to 2021 for evaluation. In multimodel modeling, relying on a single performance metric, such as RMSE or Nash-Sutcliffe, may not yield optimum results for all family models. Therefore, we allow for a non-uniform fitting scheme, enabling model experts to choose a performance metric that best suits the conceptual assumptions and specific attributes of the models.

The second step involves generating scenarios for the various models. To create scenarios derived from both endogenous and exogenous stochastic models, along with machine learning models, we applied a multivariate normal distribution to the residuals. This approach was chosen to ensure spatial correlation. Regarding rainfall-runoff scenarios, we assumed that spatial correlation was inherently preserved by the spatial characteristics of the forecasted precipitation. It is important to highlight that the number of scenarios in the rainfall-runoff method is constrained by the number of members in the precipitation forecasting model.

The third step consists in evaluating the performance of the individual models scenarios and removing those with low performance. The likelihood ratio of the forecasted scenarios of each model across all HPPs was evaluated for one-month lead time and models whose median likelihood value was below zero were removed from subsequent steps of the hyper-multi-model process. It was adopted a discrete calculation of the likelihood:

1. At each location, the naturalized flow series were categorized into five classes based on 20, 40, 60, and 80 percentiles (e.g., very dry, dry, normal, rainy and very rainy) for every month of the year.
2. For each month with a forecast, we calculated the likelihood of each flow class occurring, according to all utilized models.
3. During each monthly evaluation, we checked how well the observed flows matched predictions from different models.

The fourth step consists in assigning weights to the remaining models to generate the hyper-multimodel. The individual weight

Table 2. Climate indices used as input variables.

Name	Description
Dipolo	The difference between TSA (Tropical Southern Atlantic Index) and TNA (Tropical Northern Atlantic Index)
ENSO	ENSO Precipitation Index. ENSO Indices Based on Patterns of Satellite-Derived Precipitation
AMO, unsmoothed	Atlantic Multidecadal Oscillation Long Version. Enfield, D.B., A. M. Mestas-Nunez and P.J. Trimble, 2001: The Atlantic multidecadal oscillation and it's relation to rainfall and river flows in the continental U.S. Geophysical Research Letters, Vol. 28, 2077-2080.
TNI	TNI (Trans-Niño Index). Indices of El Niño Evolution Calculated at PSL.
BEST	Bivariate ENSO Timeseries Calculated from combining a standardized SOI and a standardized Niño3.4 SST timeseries.
QBO	Quasi-Biennial Oscillation Calculated at PSL (from the zonal average of the 30mb zonal wind at the equator as computed from the NCEP/NCAR Reanalysis).
PNA	Pacific North American Index: From NOAA Climate Prediction Center (CPC)
SOI	Southern Oscillation Index: From NOAA Climate Prediction Center (CPC)
TSA	Tropical Southern Atlantic Index Anomaly of the average of the monthly SST from Eq-20S and 10E-30W. HadISST and NOAA OI 1x1 datasets are used to create index. Climatology is 1971-2000.
NAO	North Atlantic Oscillation: From NOAA Climate Prediction Center (CPC)
PDO	Pacific Decadal Oscillation PDO is the leading PC of monthly SST anomalies in the North Pacific Ocean.
NINO12	Extreme Eastern Tropical Pacific SST (0-10S, 90W-80W): From CPC CPC uses the NOAA ERSST V5 anomalies.

ENSO: El Niño-Southern Oscillation; PSL: Physical Sciences Laboratory; SOI: Southern Oscillation Index; SST: Sea Surface Temperature; NCEP/NCAR: National Centers for Environmental Prediction/National Center for Atmospheric Research; NOAA: National Oceanic and Atmospheric Administration; PC: Principal Component; ERSST: Extended Reconstructed Sea Surface Temperature.

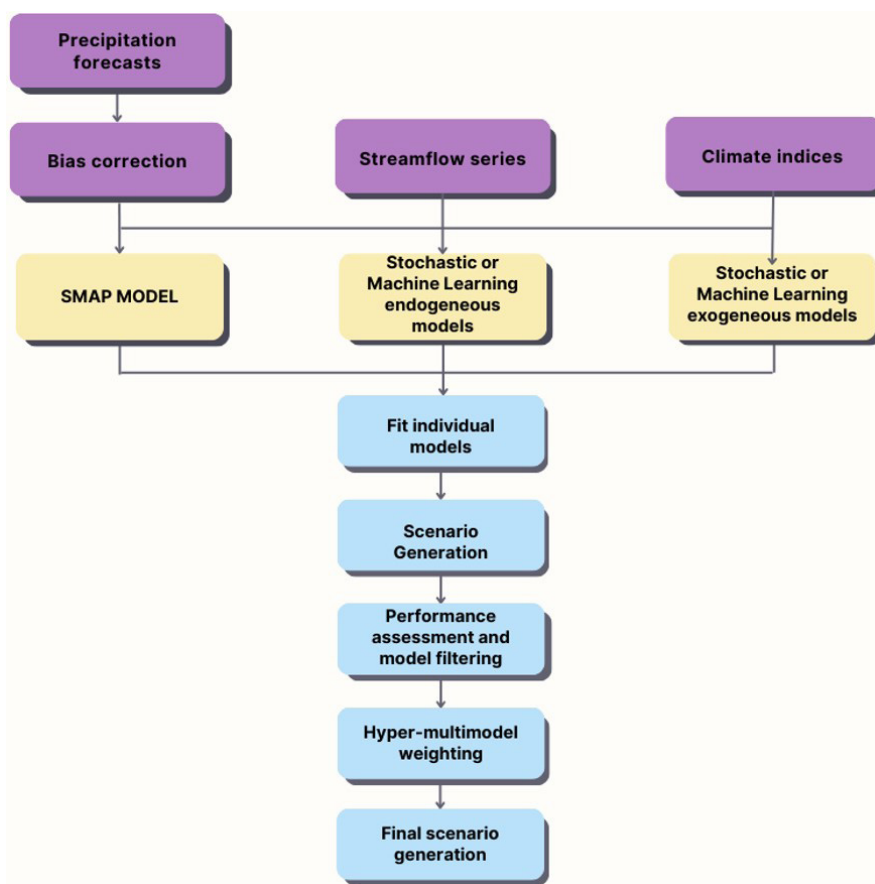


Figure 1. Hyper-multimodel methodology.

of each model (w_m) is obtained by maximizing the likelihood through the following equation:

$$L = \prod_{i=1}^T \left\{ \sum_{m=1}^Z [w_m P_{m,t}] \right\} \tag{6}$$

Where Z is the number of models, T is the quantity of periods considered during the evaluation, and $P_{m,t}$ is the probability of the observed class in the model m for the period of simulation t . Since the forecasts range from one to twelve months of lead time, we maximized the mean value of the likelihood across the twelve forecasting horizons for each HPP using the PSO (*Particle Swarm Optimization*) algorithm.

The final step* (in concordance to the results submitted to the GT-CH activity, for further details, refer to section 4) consists in the pooling of n_m members of each model according to its weights ($w_m \times 200$), selected by the n_m equidistant quantiles in the probability distribution of the members from each model.

The enhancement of the hyper-multimodel scenarios in comparison to the individual models was assessed using the Normalized Continuous Ranked Probabilistic Score (NCRPS) metric in coherence to the evaluation framework proposed by Treistman et al. (2023). Four HPP were chosen for a more detailed analysis of the results, they were selected due to its high energy production and local importance: Furnas, Itaipu, Sobradinho and Tucuruí.

RESULTS

Figures 2 and 3 present the boxplot of the likelihood ratio of the forecasted scenarios during the calibration period for the different models, with lead times ranging from one to twelve months. Notably, the dynamical models exhibited very low forecasting performance, with none meeting the established criteria for inclusion in the hyper-multimodel. However, the boxplot presented outliers with high likelihood values, suggesting that this forecasting approach may have a outstanding performance for at least one station.

Almost all stochastic and machine learning models met the criteria established to be included in the hyper-multimodel. However, the PARX and WPAR models presented low performance across all analyzed lead times. Additionally, both the endogenous and exogenous ridge regression models stood out for their high performance at a one-month lead time.

The low performance of both WPAR and PARX may be related to the fitting mechanism used, which could not deal with the higher degree of freedom imposed by the addition of external variables and the decomposition of the time series. Furthermore, it is worth noting that calibrating all the models using incremental streamflow might have introduced potential distortions in the representation of inherent natural periodic patterns, which are critical for achieving optimal performance with the WPAR model.

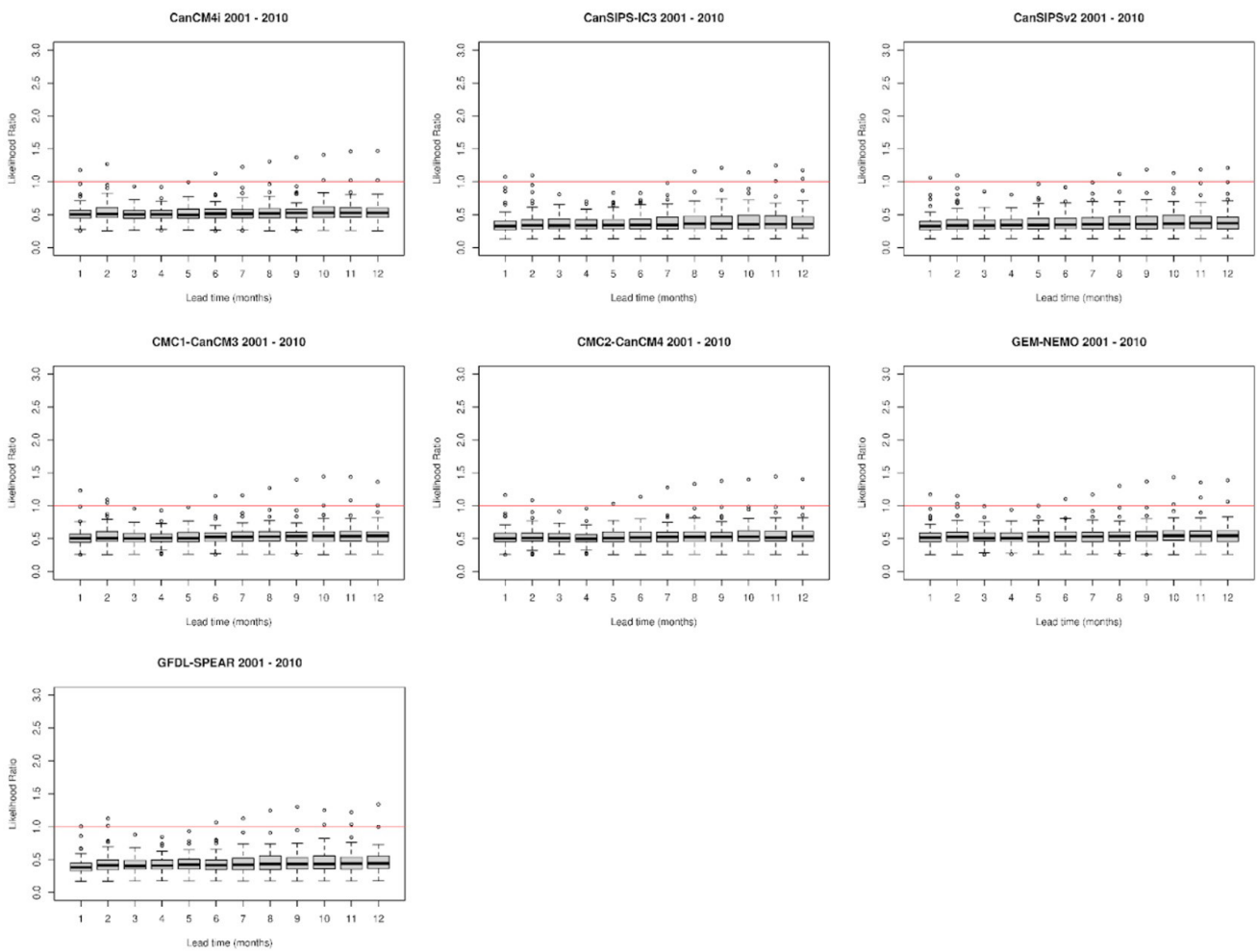


Figure 2. Boxplot of likelihood ratio of the forecasted scenario for the SMAP model coupled with different NMME models with lead times of one to twelve months.

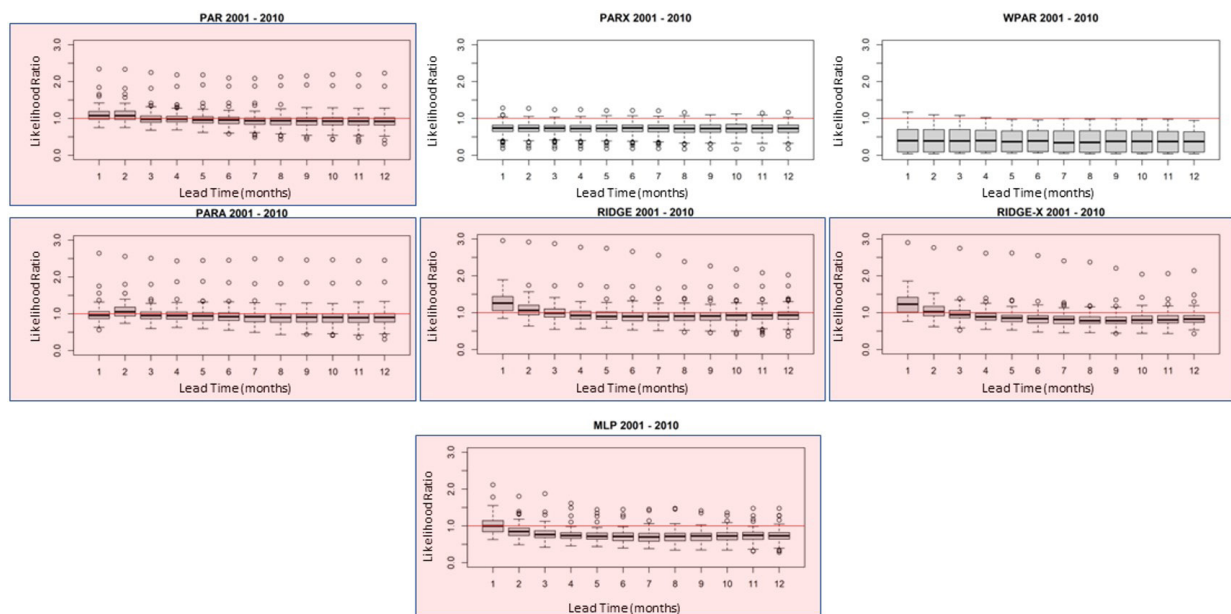


Figure 3. Boxplot of likelihood ratio of the forecasted scenario for the stochastic and machine. The highlighted models were selected to be used in further steps of the hyper-multimodel.

Figure 4 illustrates the spatial correlation of the generated scenarios from both the individual models and the hyper-multimodel. While all individual models successfully replicated the observed spatial correlation, the hyper-multimodel scenarios failed to do so.

Figure 5 displays the boxplot of the autocorrelation of the observed values, along with the scenarios from the individual models and hyper-multimodel, for lead times ranging from one to twelve months for four HPP selected for this analysis. The autoregressive models closely followed the observed autocorrelation patterns. However, the scenarios from the other models, hyper-multimodel included, could not reproduce this characteristic. These results were expected for the remaining individual models since the autocorrelation representation is not embedded in their conceptual formulation. The hyper-multimodel presented unusual autocorrelation results; section 4 discusses the operational reasons

behind the observed problems in the spatial-temporal correlation of the generated scenarios.

Figure 6 presents the boxplot of the NCRPS percentage enhancement compared to the results of individual models for the calibration period across all analyzed lead times. In general, the hyper-multimodel yielded superior results in the metric for all stations and individual models. The most significant improvement across all lead times was observed in comparison to the MLP endogenous results, where the median enhancement for the NCRPS was higher than 15% and the outliers were around 40%. The substantial improvement for the one-month lead time when compared to the PAR-A results was also notable. The smallest enhancement was observed when compared to the endogenous ridge model. This result was expected given that this model presented high likelihood ratio values in the previous analysis.

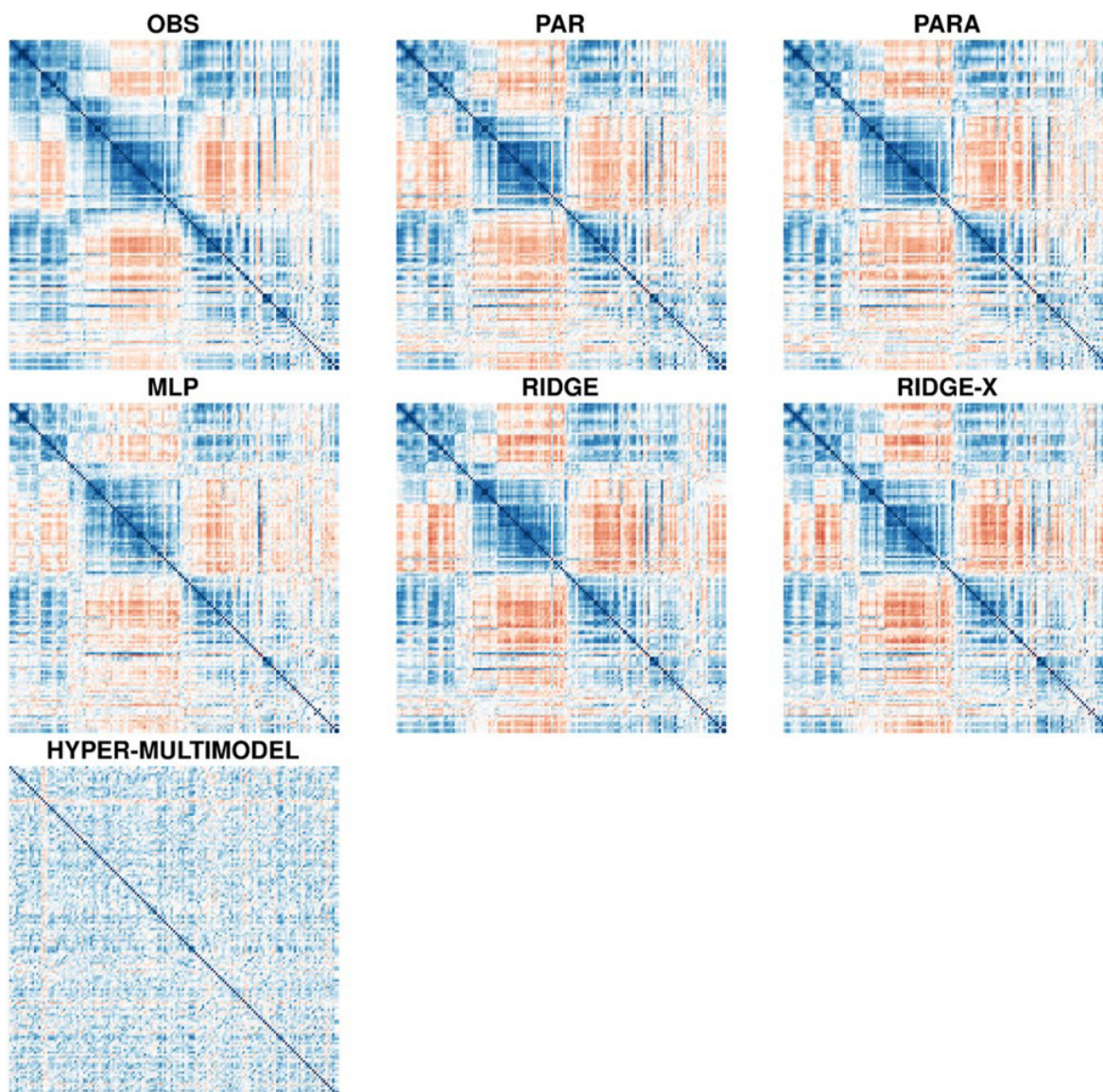


Figure 4. Spatial correlation of the observed values (OBS) for the month of January and the ensemble forecast for January of 2001 of the selected models: periodic auto-regressive model (PAR); periodic auto-regressive model with an annual component (PARA); multilayer perceptron (MLP); ridge regression with endogenous variables (RIDGE) and exogenous variables (RIDGE-X) and the hyper-multimodel.

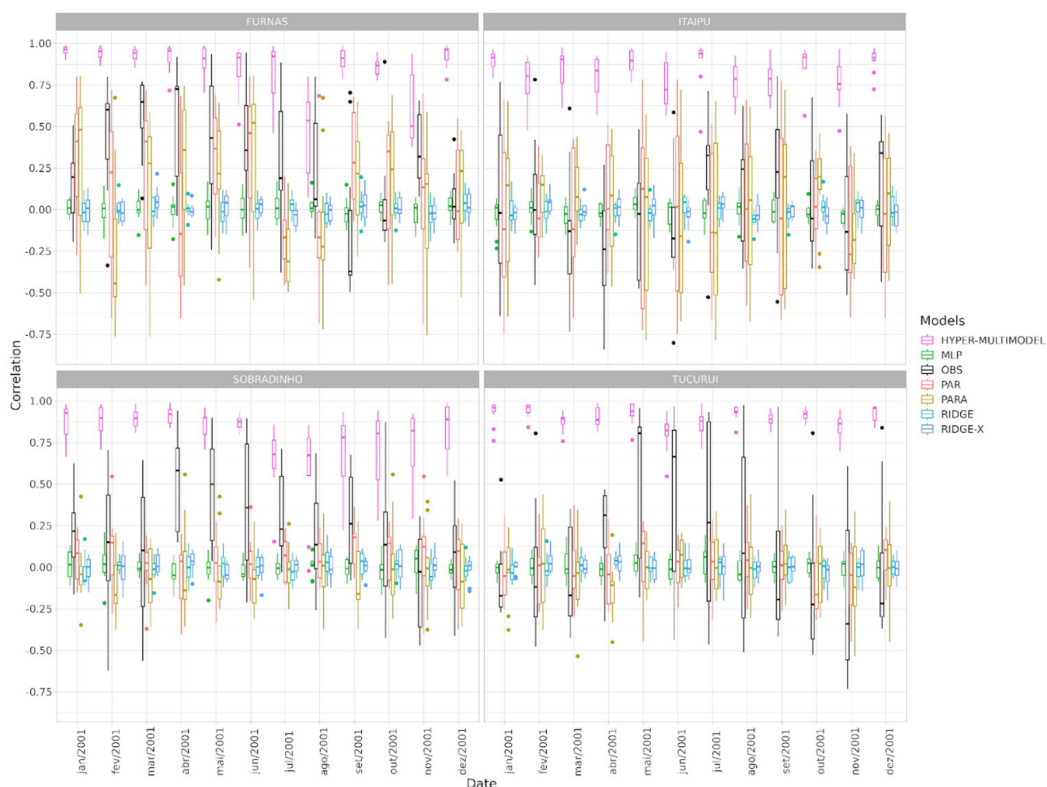


Figure 5. Boxplot of the autocorrelation of the observed values and for the ensemble forecasts of 2001 of the selected models and hyper-multimodel for four key stations (Furnas, Itaipu, Sobradinho and Tucuru) for the lead time ranging from one to twelve months.

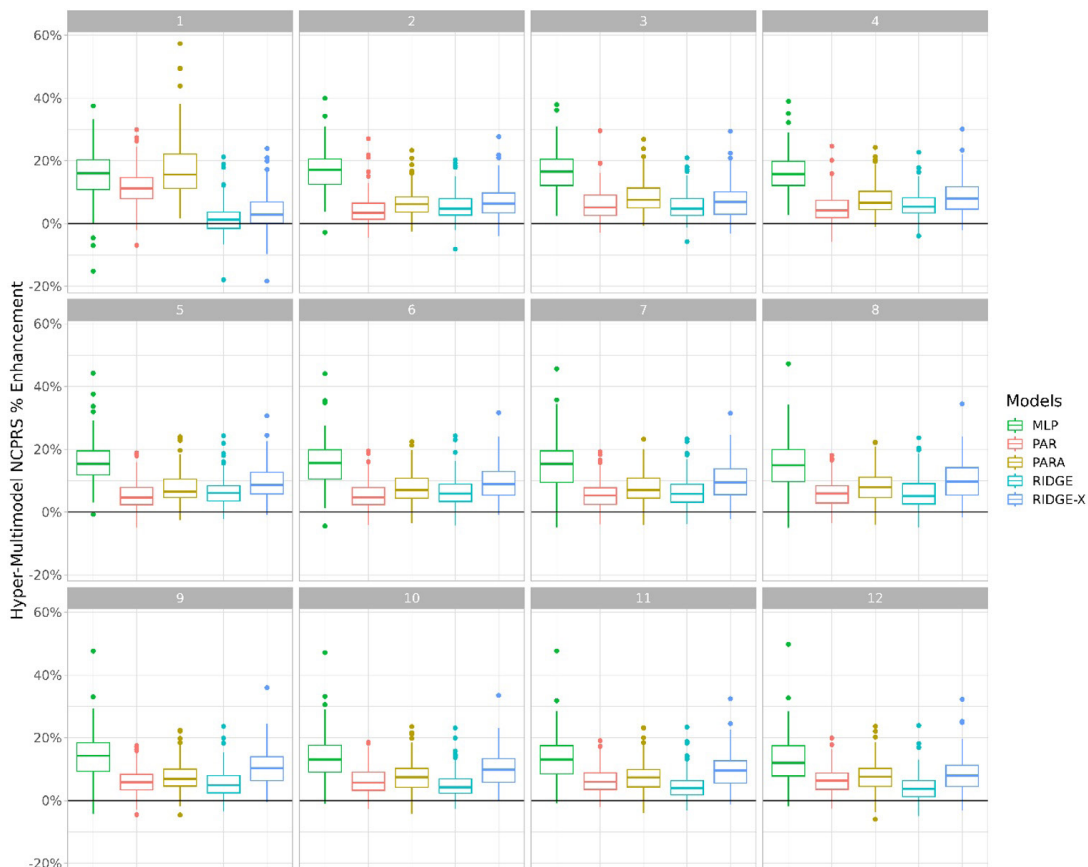


Figure 6. Hyper-Multimodel Normalized Continuous Ranked Probabilistic Score (NCRPS) percentage enhancement in comparison to individual models for lead times from 1 to 12 months - Calibration Period (2001 to 2010).

Figure 7 corroborates similar conclusions regarding the enhancement achieved through the implementation of the hyper-multimodel scheme compared to the individual models. However, it is evident how the enhancement is not uniform across the different lead, with the hyper-multimodel delivering slightly inferior results than individual models during certain specific lead times. This observation is in concordance to the data presented in Figure 6, where it is noticeable that the metric was not improved for a small subset of stations, especially for shorter lead times. These results can be explained by the maximization of the mean likelihood value of all lead times used.

Figures 8 and 9 show that the conclusions drawn from the calibration period hold true for the verification period. However, it is important to note an increase in the number of HPP where the hyper-multimodel could not enhance the metric during this period. The results for Sobradinho (Figure 9) also highlight this aspect, showing that performance did not enhance for the endogenous and exogenous ridge models for lead times higher than two months.

DISCUSSIONS

This section provides further discussions regarding some methodological decisions made along the conceptualization of

the hyper-multimodel, its possible impacts, and alternatives. To propose a new and relatively complex methodology involving several models is an ambitious endeavor, and implementing this new methodology within a time-limited such as the one involved in the GT-CH is an extremely challenging process. After the submission of the results and evaluation by the GT-CH, it was noticed some operational/methodological issues that required some clarifications. However, an agreement was made among the involved institutions stipulating that no further results would be submitted if they entailed significant changes to the initial results and methods, to maintain fair competition. Therefore, although we understood that further improvements are necessary, the results presented in this paper are the same as presented for the GT-CH.

A note about the results submitted to the GT-CH

The hyper-multimodel results presented problems in its spatial-temporal correlation due to operational divergences that were overlooked and that may also have affected the final performance. The first and major problem originated from a sorting step to calculate the likelihood of the five used classes; this mechanism was also used to pool the resulting weighted members from different models. Consequently, the predicted



Figure 7. Normalized Continuous Ranked Probabilistic Score (NCRPS) percentage enhancement of the Hyper-Multimodel in comparison to individual models for lead times from 1 to 12 months for the 4 key stations - Calibration Period (2001 to 2010).

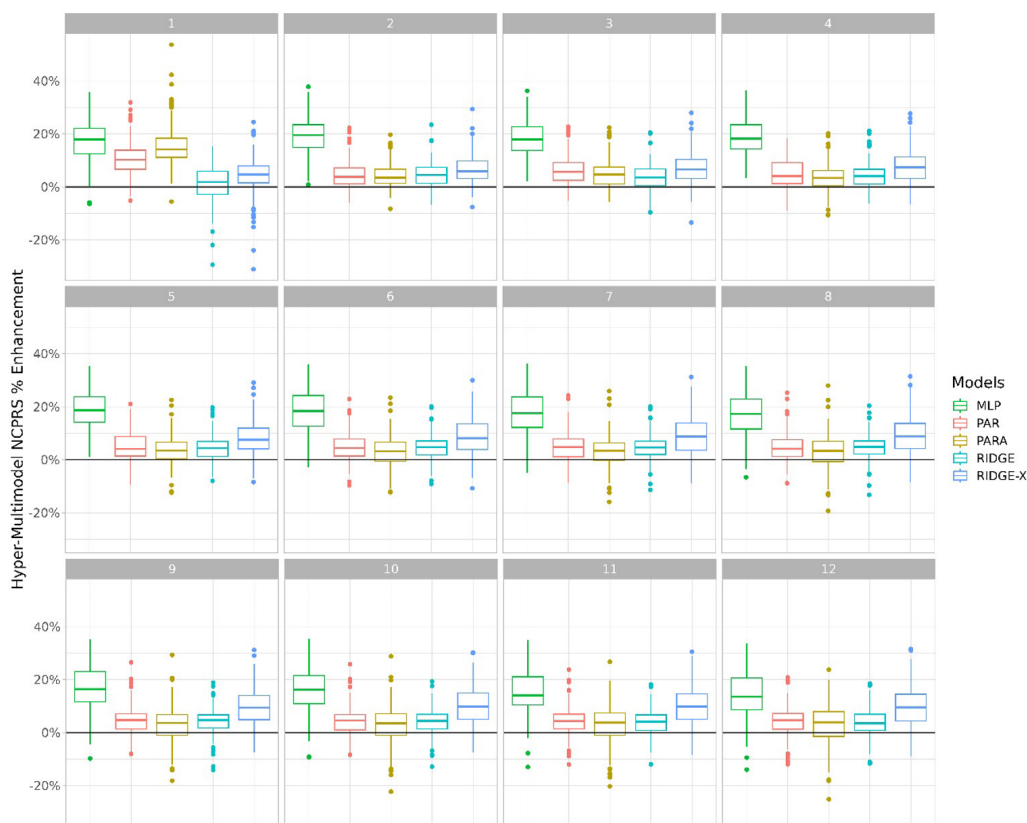


Figure 8. Hyper-Multimodel Normalized Continuous Ranked Probabilistic Score (NCRPS) percentage enhancement in comparison to individual models for lead times from 1 to 12 months - Verification Period (2011 to 2021).

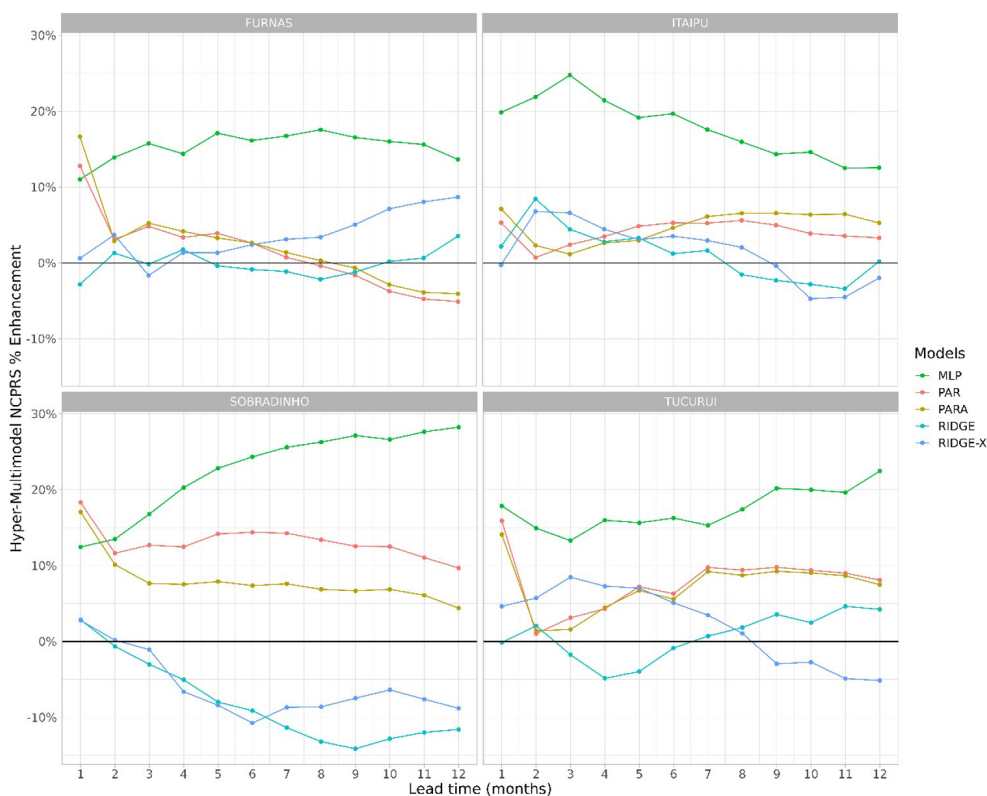


Figure 9. Normalized Continuous Ranked Probabilistic Score (NCRPS) percentage enhancement of the Hyper-Multimodel in comparison to individual models for lead times from 1 to 12 months for the 4 key stations - Verification Period (2011 to 2021).

members were mostly always from the same classes, leading to an artificial elevation of the scenarios' autocorrelation. The second problem was the use of the scenarios from individual models to generate the hyper-multimodel scenarios directly, as outlined in section 2.6. The pooling mechanism used failed to guarantee spatial correlation, as it pools a different number of scenarios from the selected models for each station. The appropriate procedure would involve fitting the weighted individual model (hyper-multimodel) and correlating the errors to preserve spatial coherence while avoiding combining the errors factors derived from different individual models (included for the scenario generation). This latter aspect may also have affected the hyper-multimodel performance, however, this affirmation requires testing and proofing.

Unfortunately, these divergences could not be identified during the assessment of the overall median performance of the scenarios generated and its overall dispersion, but only when evaluating the scenarios individually. The resolution of the stated problems required the re-run of all individual models and the fitting and scenario generation of the hyper-multimodel and could not be done within the time available in the GT-CH activity due to the large number of models involved and the extensive processing time required. Moreover, undertaking these steps could introduce substantial alterations to the initially submitted result.

Further methodological discussions

The decision of selecting models based on their median likelihood performance in all HPPs can result in the removal of models that perform exceptionally well in specific regions. To maximize the performance of the hyper-multi-model, it might be more beneficial to undertake this step individually for each station.

The use of multiple models to generate weighted or ensemble scenarios may present problems in the reproduction of the autocorrelation due to the resulting high degree of freedom involved and the use of families of models that do not intend to reproduce this characteristic directly. This particularity could not be analyzed in this study due to the mentioned problems and is a scientific question that requires further studies.

The development of a more operational version of this proposed methodology still faces some scientific challenges, to ensure, for instance, the autocorrelation characteristic of the scenarios. This characteristic is desirable to keep the coherency of the scenarios, and it may not be met when using models that do not have this characteristic embedded in its conceptualization. Also, the compatibility of different methodologies requires further steps and cautions. As shown by the WPAR results, the decision of the data used to fit the model needs to fit the conceptualization of the model.

CONCLUSIONS

In this research, we undertook the critical task of enhancing streamflow forecasting within the Brazilian electricity sector with the objective to aid decision-makers and stakeholders involved in energy production and management. To achieve this, the paper proposed a complex hyper-multimodel forecasting approach, built

upon multiple state-of-the-art modeling techniques and detailed climate data analysis, to predict streamflow.

Our analyses revealed discernible performance disparities among the diverse models applied, especially regarding dynamical models, which mostly fell short of the criteria to be integrated into the hyper-multimodel. The results obtained showed that the proposed methodology could successfully increase the overall performance of the scenarios generated through the use of the individual models. However, it is essential to underscore the non-uniform enhancement across different lead times; specific periods experienced decreased performance compared to individual models, a phenomenon significantly influenced by the optimization of mean likelihood value across all employed lead times.

The quality of the final scenarios generated is directly connected to the performance of the individual models. In this study, the individual models have a wide margin of improvement, as they were mostly fitted automatically and not fully refined on an individual basis. This aspect was expected due to the challenge that the hyper-multimodel proposes to embrace, undertaking different model families, large number of locations and forecast timeframe. Different steps in a short period were needed, such as organizing the data, run the different models, developing uniform methodologies to integrate these results, organizing, and evaluating the large number of results generated to finally develop the hyper-multimodel. Therefore, there is potential to further enhance the performance of the final scenarios with the use of more refined and calibrated individual models, such as the ones that produced good scenarios in this activity, as shown by Treistman et al. (2023). Another aspect that may improve the final scenarios is the removal of the models individually for each HPP, to allow the use of models with high region-specific performance.

Despite the short time available, participating in an activity such as the GT-CH was engaging and boosted the technological development of the hyper-multimodel. The integration with different research groups and the interaction with the ONS and CCEE personnel improved the discussions and results analysis. Furthermore, the hyper-multimodel framework showed potential to improve hydrological forecast for the Brazilian electricity sector, especially when using well calibrated individual models.

ACKNOWLEDGMENTS

This work has been supported by the National Council for Scientific and Technological Development – CNPq through scholarship grants and the MODASH project (421688/2021-1). This work was also supported by FINEP. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

REFERENCES

Ajami, N. K., Duan, Q., Gao, X., & Sorooshian, S. (2006). Multimodel combination techniques for analysis of hydrological simulations: application to distributed model intercomparison project results. *Journal of Hydrometeorology*, 7(4), 755-768. <http://dx.doi.org/10.1175/JHM519.1>.

- Arsenault, R., Gatién, P., Renaud, B., Brissette, F., Martel, J.-L. (2015). A comparative analysis of 9 multi-model averaging approaches in hydrological continuous streamflow simulation. *Journal of Hydrology*, 529, 754-767. <http://dx.doi.org/10.1016/j.jhydrol.2015.09.001>.
- Ávila, L., Silveira, R., Campos, A., Rogiski, N., Freitas, C., Aver, C., & Fan, F. (2023). Seasonal streamflow forecast in the Tocantins River Basin, Brazil: an evaluation of ECMWF-SEAS5 with multiple conceptual hydrological models. *Water*, 15(9), 1695. <http://dx.doi.org/10.3390/w15091695>.
- Box, G., & Jenkins, G. (1970). *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day.
- Block, P. J., Souza Filho, F. A., Sun, L., & Kwon, H. H. (2009). A streamflow forecasting framework using multiple climate and hydrological models 1. *Journal of the American Water Resources Association*, 45(4), 828-843. <http://dx.doi.org/10.1111/j.1752-1688.2009.00327.x>.
- Brockwell, P., & Davis, R. (1991). *Time Series: Theory and Methods*. New York: Springer. <http://dx.doi.org/10.1007/978-1-4419-0320-4>.
- Breiman, L. (2001). Statistical modeling: the two cultures (with comments and a rejoinder by the author). *Statistical Science*, 16(3), 199-231. <http://dx.doi.org/10.1214/ss/1009213726>.
- Chevuturi, A., Tanguy, M., Facer-Childs, K., Martínez-de la Torre, A., Sarkar, S., Thober, S., Samaniego, L., Rokovec, O., Kelbling, M., Sutanudjaja, E. H., Wanders, N., & Blyth, E. (2023). Improving global hydrological simulations through bias-correction and multi-model blending. *Journal of Hydrology*, 621, 129607. <http://dx.doi.org/10.1016/j.jhydrol.2023.129607>.
- Gardner, M. W., & Dorling, S. R. (1998). Artificial neural networks (the multilayer perceptron): a review of applications in the atmospheric sciences. *Atmospheric Environment*, 32(14-15), 2627-2636. [http://dx.doi.org/10.1016/S1352-2310\(97\)00447-0](http://dx.doi.org/10.1016/S1352-2310(97)00447-0).
- Georgakakos, K. P., Seo, D. J., Gupta, H., Schaake, J., & Butts, M. B. (2004). Towards the characterization of streamflow simulation uncertainty through multimodel ensembles. *Journal of Hydrology*, 298(1-4), 222-241. <http://dx.doi.org/10.1016/j.jhydrol.2004.03.037>.
- Greuell, W., & Hutjes, R. W. (2023). Skill and sources of skill in seasonal streamflow hindcasts for South America made with ECMWF's SEAS5 and VIC. *Journal of Hydrology*, 617, 128806. <http://dx.doi.org/10.1016/j.jhydrol.2022.128806>.
- Hastie, T., Friedman, J., & Tibshirani, R. (2001). *The elements of statistical learning*. New York: Springer. Springer Series in Statistics. <https://doi.org/10.1007/978-0-387-21606-5>.
- Jiang, Z., Li, R., Li, A., & Ji, C. (2018). Runoff forecast uncertainty considered load adjustment model of cascade hydropower stations and its application. *Energy*, 158, 693-708. <http://dx.doi.org/10.1016/j.energy.2018.06.083>.
- Kim, Y. O., Jeong, D. I., & Kim, H. S. (2001). Improving water supply outlook in Korea with ensemble streamflow prediction. *Water International*, 26(4), 563-568. <http://dx.doi.org/10.1080/02508060108686957>.
- Kumar, V., Kedam, N., Sharma, K. V., Mehta, D. J., & Caloiero, T. (2023). Advanced machine learning techniques to improve hydrological prediction: a comparative analysis of streamflow prediction models. *Water*, 15(14), 2572. <http://dx.doi.org/10.3390/w15142572>.
- Kwon, H. H., Lall, U., & Khalil, A. F. (2007). Stochastic simulation model for nonstationary time series using an autoregressive wavelet decomposition: applications to rainfall and temperature. *Water Resources Research*, 43(5), W05407. <http://dx.doi.org/10.1029/2006WR0052>.
- Kwon, H. H., Souza Filho, F. A., Block, P., Sun, L., Lall, U., & Reis Junior, D. S. (2012). Uncertainty assessment of hydrologic and climate forecast models in Northeastern Brazil. *Hydrological Processes*, 26(25), 3875-3885. <http://dx.doi.org/10.1002/hyp.8433>.
- Loucks, D. P., Stedinger, J. R., & Haith, D. A. (1981). *Water resources system planning and analysis*: Prentice-Hall Inc.
- Roy, T., Valdés, J. B., Serrat-Capdevila, A., Durcik, M., Demaria, E. M., Valdés-Pineda, R., & Gupta, H. V. (2020). Detailed overview of the multimodel multiproduct streamflow forecasting platform. *Journal of Applied Water Engineering and Research*, 8(4), 277-289. <http://dx.doi.org/10.1080/23249676.2020.1799442>.
- Salas, J. D., Delleur, J. W., Yevjevich, V., & Lane, W. L. (1980). *Applied modeling of hydrologic time series*. Littleton, Colorado: Water Resources Press Publications.
- Shamseldin, A. Y., O'Connor, K. M., & Liang, G. C. (1997). Methods for combining the outputs of different rainfall-runoff models. *Journal of Hydrology*, 197(1-4), 203-229. [http://dx.doi.org/10.1016/S0022-1694\(96\)03259-3](http://dx.doi.org/10.1016/S0022-1694(96)03259-3).
- Torrence, C., & Compo, G. P. (1998). A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society*, 79(1), 61-78. [http://dx.doi.org/10.1175/1520-0477\(1998\)079<0061:APGTWA>2.0.CO;2](http://dx.doi.org/10.1175/1520-0477(1998)079<0061:APGTWA>2.0.CO;2).
- Treistman, F., Maceira, M. E. P., Damázio, J. M., & Cruz, C. B. (2020, August, 18-21). Periodic time series model with annual component applied to operation planning of hydrothermal systems. In Institute of Electrical and Electronics Engineers (Org.), *2020 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)* (pp. 1-6). New York, United States: IEEE. <https://doi.org/10.1109/PMAPS47429.2020.9183472>.
- Treistman, F., Penna, D. D. J., Khenayfis, L. S., Cavalcante, N. B. R., Souza Filho, F. A., Rocha, R. V., Estacio, A. B. S., Rolim, L. R.

- Z., Pontes Filho, J. D. A., Porto, V. C., Guimarães, S. O., Pessanha, J. F. M., Almeida, V. A., Chan, P. D. S., Lappicy, T., Lima, C. H. R., Detzel, D. H. M., & Bessa, M. R. (2023). A framework to evaluate and compare synthetic streamflow scenario generation models. *Brazilian Journal of Water Resources*, (under review)
- Velazquez, J. A., Anctil, F., Ramos, M. H., & Perrin, C. (2011). Can a multi-model approach improve hydrological ensemble forecasting? A study on 29 French catchments using 16 hydrological model structures. *Advances in Geosciences*, 29, 33-42. <http://dx.doi.org/10.5194/adgeo-29-33-2011>.
- Vrugt, J. A., Ter Braak, C. J., Gupta, H. V., & Robinson, B. A. (2009). Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling? *Stochastic Environmental Research and Risk Assessment*, 23, 1011-1026. <http://dx.doi.org/10.1007/s00477-008-0274-y>.
- Wanders, N., & Wood, E. F. (2016). Improved sub-seasonal meteorological forecast skill using weighted multi-model ensemble simulations. *Environmental Research Letters*, 11(9), 094007. <http://dx.doi.org/10.1088/1748-9326/11/9/094007>.
- Wang, L., Li, X., Ma, C., & Bai, Y. (2019). Improving the prediction accuracy of monthly streamflow using a data-driven model based on a double-processing strategy. *Journal of Hydrology*, 573, 733-745. <http://dx.doi.org/10.1016/j.jhydrol.2019.03.101>.
- Yevjevich, V. (1987). Stochastic models in hydrology. *Stochastic Hydrology and Hydraulics*, 1, 17-36. <http://dx.doi.org/10.1007/BF01543907>.
- Zaherpour, J., Mount, N., Gosling, S. N., Dankers, R., Eisner, S., Gerten, D., Liu, X., Masaki, Y., Schmied, H. M., Tang, Q., & Wada, Y. (2019). Exploring the value of machine learning for weighted multi-model combination of an ensemble of global hydrological models. *Environmental Modelling & Software*, 114, 112-128. <http://dx.doi.org/10.1016/j.envsoft.2019.01.003>.
- Zhang, Z., Zhang, Q., Singh, V. P., & Shi, P. (2018). River flow modelling: comparison of performance and evaluation of uncertainty using data-driven models and conceptual hydrological model. *Stochastic Environmental Research and Risk Assessment*, 32, 2667-2682. <http://dx.doi.org/10.1007/s00477-018-1536-y>.

Authors contributions

Francisco de Assis de Souza Filho: Hyper-multimodel methodology conceptualization and discussions, activities coordinator, text revision.

Renan Vieira Rocha: Methodology discussions, assistant coordinator of the activities, generation and consolidation of the results, hyper-multimodel performance analysis, text writing and revision.

Ályson Brayner Estácio: Methodology discussions, assistant coordinator of the activities, hyper-multimodel optimization, machine learning results generation and analysis, data organization, text writing.

Larissa Zaira Rafael Rolim: Methodology discussions, endogenous and exogenous stochastic results generation and analysis, text writing and revision.

João Dehon de Araújo Pontes Filho: Methodology discussions, generation and organization of the results, performance analysis, text writing and revision.

Victor Costa Porto: Methodology discussions, coupled hydrological models results generation and analysis, text writing.

Sullyandro Oliveira Guimarães: Data organization of exogenous variables and NMME models, aid in the run of coupled hydrological models, climate indices discussions.

Editor-in-Chief: Adilson Pinheiro

Associated Editor: Adilson Pinheiro