Revista Brasileira de Recursos Hídricos
Brazilian Journal of Water Resources
Versão On-line ISSN 2318-0331
RBRH, Porto Alegre, v. 28, e43, 2023
Special Session

https://doi.org/10.1590/2318-0331.282320230115

A framework to evaluate and compare synthetic streamflow scenario generation models

Procedimento para avaliar e comparar modelos de geração de cenários sintéticos de vazões

Felipe Treistman[1], Débora Dias Jardim Penna[1], Lucas de Souza Khenayfis[1], Nelson Bernardo Rodrigues Cavalcante[2], Francisco de Assis de Souza Filho[3], Renan Vieira Rocha[4], Ályson Brayner Estácio[3,4], Larissa Zaira Rafael Rolim[3], João Dehon de Araújo Pontes Filho[4], Victor Costa Porto[3], Sullyandro Oliveira Guimarães[5,6], José Francisco Moreira Pessanha[7], Victor Andrade de Almeida[7], Priscilla Dafne Shu Chan[7], Thiago Lappicy[8], Carlos Henrique Ribeiro Lima[8], Daniel Henrique Marco Detzel[9] & Marcelo Rodrigues Bessa[9]

[1]Operador Nacional do Sistema Elétrico, Rio de Janeiro, RJ, Brasil
[2]Câmara de Comercialização de Energia Elétrica, São Paulo, SP, Brasil
[3]Universidade Federal do Ceará, Fortaleza, CE, Brasil
[4]Fundação Cearense de Meteorologia e Recursos Hídricos, Fortaleza, CE, Brasil
[5]Potsdam-Institut für Klimafolgenforschung, Potsdam, Deutschland
[6]Universität Postdam, Potsdam, Deutschland
[7]Centro de Pesquisas de Energia Elétrica, Rio de Janeiro, RJ, Brasil
[8]Universidade de Brasília, Brasília, DF, Brasil
[9]Universidade Federal do Paraná, Curitiba, PR, Brasil

E-mails: felipe.treistman@ons.org.br (FT), lkhenayfis@ons.org.br (LSK), debora.jardim@ons.org.br (DDJP), ncavalcante@ccee.org.br (NBRC), assis@ufc.br (FASF), renan.rocha@funceme.br (RVR), alyson.estacio@funceme.br (ÁBE), larissazairarr@gmail.com (LZRR), joao.pontes@funceme.br (JDAPF), victorcporto@gmail.com (VCP), sullyandro@pik-potsdam.de (SOG), francisc@cepel.br (JFMP), andrade@cepel.br (VAA), pchan@cepel.br (PDSC), lappicy@gmail.com (TL), chrlima@unb.br (CHRL), detzel@ufpr.br (DHMD), marcelo.bessa@ufpr.br (MRB)

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

ABSTRACT

Synthetic streamflow scenario generation is particularly important in countries like Brazil, where hydroelectric power generation plays a key role and properly handling the uncertainty of future streamflow is crucial. This paper showcases a collaborative effort within the Brazilian electrical sector to enhance streamflow scenario models, focusing on horizons up to one year. Five institutions proposed diverse methodologies, and their effectiveness was evaluated using a comparative framework. The results reveal the strengths and areas for improvement in each model. GHCen emerged as the top performer, excelling in both short-term and moving average analyses, while the PARX model demonstrated superior performance in specific regions. The PAR(p)-A, which is the official methodology in Brazil, was the second-best model in the moving average analysis. This research offers valuable insights for countries facing similar hydrothermal scheduling and scenario generation challenges.

Keywords: Comparative evaluation framework; Synthetic streamflow scenario generation; Multivariate multistage probabilistic forecast.

RESUMO

A geração de cenários sintéticos de vazões afluentes é particularmente importante em países como o Brasil, onde a geração de energia hidrelétrica desempenha um papel fundamental e é crucial lidar adequadamente com a incerteza da vazão futura. Este artigo apresenta um esforço colaborativo do setor elétrico brasileiro para aprimorar os modelos de geração de cenários de vazão, com foco em horizontes de até um ano. Cinco instituições propuseram metodologias diversas e a sua eficácia foi avaliada através de uma avaliação comparativa estruturada. Os resultados revelam os pontos fortes e as áreas de possíveis aprimoramentos em cada modelo. O modelo GHCen apresentou o melhor desempenho, destacando-se nas análises de curto prazo e de média móvel, enquanto o modelo PARX demonstrou desempenho superior em regiões específicas do Brasil. O PAR(p)-A, metodologia oficial no Brasil, foi o segundo melhor modelo na análise de médias móveis. Este artigo oferece informações valiosas para países que enfrentam desafios semelhantes de programação hidrotérmica e geração de cenários.

Palavras-chave: Procedimento para avaliação comparativa; Geração de cenários sintéticos de afluências; Previsão probabilística multivariada multiestágio.

INTRODUCTION

The increasing integration of renewable sources into the electricity generation matrix brings significant challenges and opportunities to the electrical sector. One of the main challenges is to enhance the models for generating synthetic scenarios, especially in the Brazilian context, where the generation of synthetic streamflow scenarios plays a fundamental role in forecasting the availability of water resources and supporting strategic decisions in the mid- and long-term energy planning process.

In this context, in collaboration with academic institutions and research centers, the Brazilian electrical sector has been working to improve the accuracy of models used to predict and generate streamflow scenarios. A recent example of this improvement effort was the establishment of 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.

This initiative aimed to meet the need for increasingly accurate projections, considering the dynamics of climate changes and variations, and seasonal fluctuations. The initiative involved the participation of five institutions that proposed and developed methodologies with different approaches, detailed in the articles presented in this special edition of RBRH. To assess the effectiveness of these models, a comparative evaluation framework was proposed, wherein various aspects were analyzed, including the ability to generate monthly streamflow scenarios consistent with the most critical scenarios observed in recent history for different time horizons (from 1 to 12 months ahead).

This work aims to present the proposed testing framework to evaluate the quality of synthetic streamflow scenarios in comparison with the methodology currently used by the electrical sector, known as Periodic Autoregressive with Annual Component – PAR(p)-A (Treistman et al., 2020a; Comissão Permanente para Análise de Metodologias e Programas Computacionais do Setor Elétrico, 2021), applying the NCRPS metric (Hersbach, 2000; Cassagnole et al., 2021) that assesses the overall quality of the generated distribution obtained by scenario generation models, rather than the traditional metrics that measure the accuracy of a predictive model. In addition to assessing the predictive distribution at each time step, moving averages for 2 to 12 months ahead will also be evaluated to analyze the ability of the models to represent the observed streamflow sequences.

This paper is structured into four sections. In section 1, we presented the context for the development of this work and its objectives. In Section 2, the five methodologies assessed by GT-CH and the current methodology used by the electrical sector are presented, including the case studies employed in the evaluation of these proposals, and the details of the metrics considered for the analysis. Section 3 presents the results obtained for each model, highlighting their strengths, and identifying areas for improvement. It is worth noting that all models were evaluated based on the same input datasets, ensuring a fair and consistent comparison. Finally, Section 4 presents the main conclusions from this evaluation, consolidating the results achieved by the GT-CH.

MATERIAL AND METHODS

Natural streamflow data

The historical dataset of natural monthly streamflow series from 146 hydropower plants (HPPs) in Brazil considered in this study is provided by ONS. It covers all major Brazilian basins from 1931 to 2021, and it can be obtained on ONS's website (Operador Nacional do Sistema Elétrico, 2023b).

Anthropogenic effects such as regulation and diversions, as well as reservoir evaporation, are removed from the observed streamflow record. In this work, all evaluations are done considering incremental natural streamflow, which is the difference between total natural streamflow from two subsequent HPPs.

Evaluated methodologies

A total of six methodologies representing the state-of-the-art scenario generation models, with different approaches, will be analyzed and compared with each other. The benchmark methodology is the PAR(p)-A, which is the model used in Brazilian energy operation planning (Maceira et al., 2018).

The second evaluated model is the Periodic Autoregressive with Exogenous Variable – PARX (Lima & Lall, 2010; Lappicy & Lima, 2023). This model uses large-scale climate indices, such as sea-surface temperature from specific Pacific and Atlantic oceans regions, and low zonal/southern winds, to improve monthly forecasts of natural flows. The use of climate variables in the regression equation of the PAR model tries to improve the forecast and scenario generation of streamflow, preserving the spatial correlation.

The third model is a hyper-multimodel (Souza Filho et al., 2023), which combines three families of models: conceptual rainfall-runoff model coupled with climate models; stochastic and machine learning models with endogenous variables; and stochastic and machine learning models with both endogenous and exogenous variables. This approach combines the best of each model, weighing them by the maximum likelihood.

The fourth methodology consists of a Markov-switching periodic autoregressive model with - MS-PAR(p) (Treistman et al., 2020b; Pessanha et al., 2023), which is the combination of the classic PAR model with the addition of climate variables as a state model by a Markov chain. This evaluation uses the El Niño Southern Oscillation as the represented climate variable, which is one of the most impacting climate variability in the Brazilian streamflow.

The fifth model is the LYNX-Series model, which is based on the Contemporaneous Autoregressive Moving Average - CARMA (Detzel et al., 2023). It represents a non-periodic and multivariate version of the Box & Jenkins or ARIMA family of stochastic models (Box et al., 2008). The contemporaneous portion of the model, in turn, is responsible for considering the spatial correlation between different locations. The framework also employs a sampling procedure in which a subset of synthetic scenarios is selected from the outputs of the CARMA model based on the recent historical streamflow regimes.

The last evaluated model is the Hybrid Generator of Synthetic Streamflow Scenarios - GHCen (Treistman et al., 2023) - which merges the conceptual rainfall-runoff modeling of the SMAP/ONS model (Operador Nacional do Sistema Elétrico, 2017) with a stochastic methodology for the simulation of synthetic daily precipitation scenarios. The stochastic methodology can reproduce the main characteristics of the precipitation historical data, while the conceptual modeling guarantees the correct physical behavior of the rainfall-runoff relationship.

Table 1 summarizes all evaluated methodologies and their main characteristics.

Evaluation metrics

The generation of synthetic scenarios of monthly streamflow is a complex problem to be represented due to the inherent uncertainties of this physical process, and the intricate interaction with the other variables that impact the hydrological cycle. It is not expected that a synthetic scenario generation model could make a perfect prediction many steps ahead, or that the observed cumulative inflows are always well distributed among the generated scenarios, especially during more critical periods of the historical record. Nevertheless, continuous efforts should be made to enhance the scenario generation models to make them more closely resemble the observed hydrological reality.

The purpose of this paper is to establish a framework for evaluating and comparing scenario generation models that are applied to monthly streamflows, which are one of the most important uncertainties in the Brazilian energy sector. The Continuous Ranked Probabilistic Score – CRPS - (Hersbach, 2000) in its normalized version (Cassagnole et al., 2021) is used as the main evaluation metric. The NCRPS (Equation 1) measures the difference between the cumulative distribution function of the simulated scenarios and the observation distribution (a Heaviside function on the observation), which is more suited to analyze scenario generation models.

$$NCRPS = \frac{\int_{-\infty}^{+\infty} \left(P_{cum_t}\left(q_t\right) - I_{o_t}\left(o_t\right) \right)^2 dq}{SD_o} \tag{1}$$

where q_t is the synthetic scenario of monthly inflow; o_t is the observed flow; $P_{cum_t}\left(q_t\right)$ is the cumulative distribution function of the synthetic scenarios generated for period t; SD_o is the standard deviation of the flow observed in the period; I_{o_t} is the Heaviside function.

For some point-specific analyses, MAPE (mean absolute percentage error) and NRMSE (normalized root mean square error) were also used. Both are described in Equations 2 and 3, respectively.

$$NRMSE = \frac{1}{T} \sum_{t=1}^{T} \sqrt{\frac{1}{N_{scen}} \sum_{iscen=1}^{N_{scen}} \frac{\left(o_t - q_t^{iscen}\right)^2}{SD_o}} \tag{2}$$

$$MAPE = \frac{1}{T} \sum_{t=1}^{T} \frac{1}{N_{scen}} \sum_{iscen=1}^{N_{scen}} \left| \frac{o_t - q_t^{iscen}}{o_t} \right| \tag{3}$$

where N_{scen} is the number of generated synthetic scenarios; q_t^{iscen} is the synthetic scenario of the $iscen$ index in period t; T is the total number of simulated cases.

Table 1. Summary of evaluated methodologies and their main characteristics.

Model	Type	Linear?	Climate variables?	References
PAR(p)-A	Statistical	Yes	No	Treistman et al. (2020a)
PARX	Statistical	Yes	Yes	Lima & Lall (2010); Lappicy & Lima (2023)
Hyper-multimodel	Hybrid (Statistical + Machine Learning + Conceptual)	No	Yes	Souza Filho et al. (2023)
CARMA	Statistical	Yes	No	Detzel et al. (2023)
MS-PAR(p)	Statistical	Yes	Yes	Treistman et al. (2020b); Pessanha et al. (2023)
GHCen	Hybrid (Statistical + Conceptual)	No	Yes	Treistman et al. (2023)

Case study

The case study presented in this paper includes the following characteristics:

1. 132 simulations, one for each month from January 2011 to December 2021.
2. The scenario generation horizon is one year ahead.
3. Monthly time steps that can be broken down into daily time steps and later grouped every month.
4. The simulation must be performed for all 146 HPPs in the National Interconnected System (NIS), and the scenarios must be spatially correlated.
5. Models must always be estimated with historical data available up to the year before the simulation. If it uses an exogenous variable, only the observed/predicted values up to the simulation date should be used.

The location of all considered HPPs is shown in the schematic diagram in Figure 1 (Operador Nacional do Sistema Elétrico, 2023a).

Comparative evaluation framework

To compare the scenario generation models an evaluation framework is proposed. The NCRPS is calculated for the predictive distribution of monthly incremental streamflow from t+1 to t+12 for all HPPs. Additionally, the NCRPS was calculated for the moving average from 2 months to 12 months, providing an analysis of the temporal evolution of the observed streamflow about the sequences of the synthetic scenarios. This analysis is crucial for energy planning models since the accumulation of resources over time heavily influences the results.

The NCRPS is a metric used to evaluate the predictive distribution, making it more suitable for assessing synthetic scenario generation. It has the unique characteristic of penalizing any observation that falls outside of the generated scenarios more severely. If we use this predictive distribution, the energy operation models will construct an energy policy that may not consider the observed natural resources, leading to a waste of resources.

To gain a better understanding of the analysis, we will begin by presenting a summary table. This table will display the percentage of HPPs, as defined in Equation 4, where the proposed model outperformed PAR(p)-A (with a lower NCRPS) for each horizon evaluated and for each of the proposed methodologies.

Furthermore, we calculated the percentage contribution of each HPP to the Natural Energy Inflow (NEI) of the NIS according to Equation 5, based on inflows from 1931 to 2021. This allows for greater weight to be given to HPPs that have a greater contribution to the system. Ultimately, this leads to a percentage improvement being calculated for each proposed model in terms of NEI.

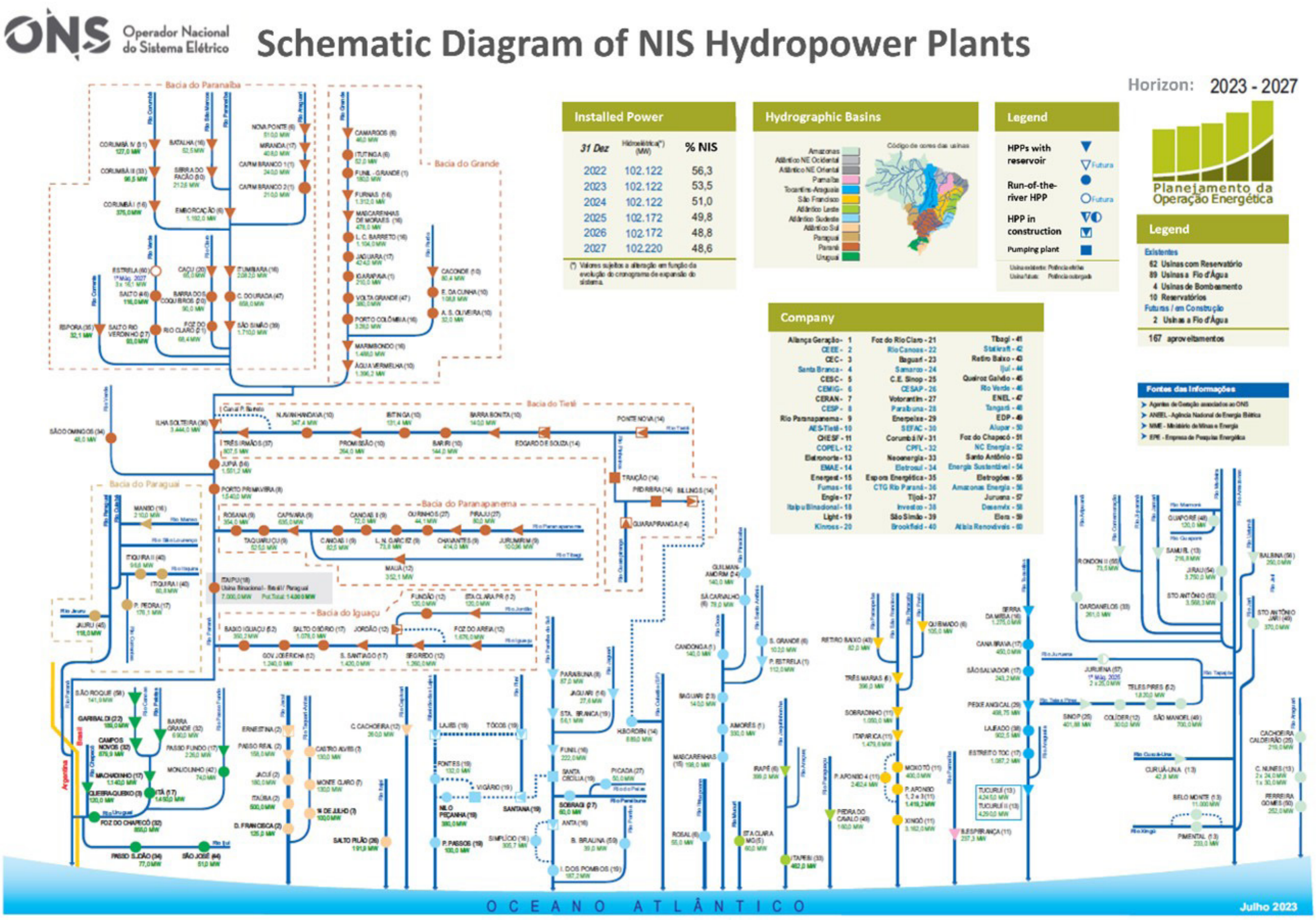


Figure 1. Schematic diagram of the HPPs in NIS.

$$\%HPP = \frac{N_{HPP}^{improve}}{N_{HPP}^{total}} \tag{4}$$

$$\%NEI = \sum_{iHPP=1}^{N_{HPP}^{improve}} \frac{NEI_{LTM}^{iHPP}}{NEI_{LTM}^{NIS}} \tag{5}$$

where $N_{HPP}^{improve}$ is the number of HPPs in which the proposed model was better than the current one, N_{HPP}^{total} is the total number of HPPs considered in the case study; NEI_{LTM}^{iHPP} is the average NEI between 1931 and 2021 referring to HPP with $iHPP$ index; and NEI_{LTM}^{NIS} is the NEI long-term mean of NIS for the same period.

The flowchart depicted in Figure 2 provides a concise overview of the devised framework for assessing the proposed models.

RESULTS

The results found in the case study are presented below. It is noteworthy that the objective of this first analysis is to provide an overview of the overall performance of each model. In the following sections, the presented results will be better explored individually.

First, Tables 2 and 3 bring, for each time step, the percentage of improvement of each proposed model about the current model, both in terms of the percentage of HPPs and in the percentage of representativeness of NEI. Tables 4 and 5 present the same evaluation, however, in moving average windows of two to twelve months. In these evaluations, the NCRPS was used as an evaluation metric. Values highlighted in red indicate the proposed model with the best performance among the five proposals.

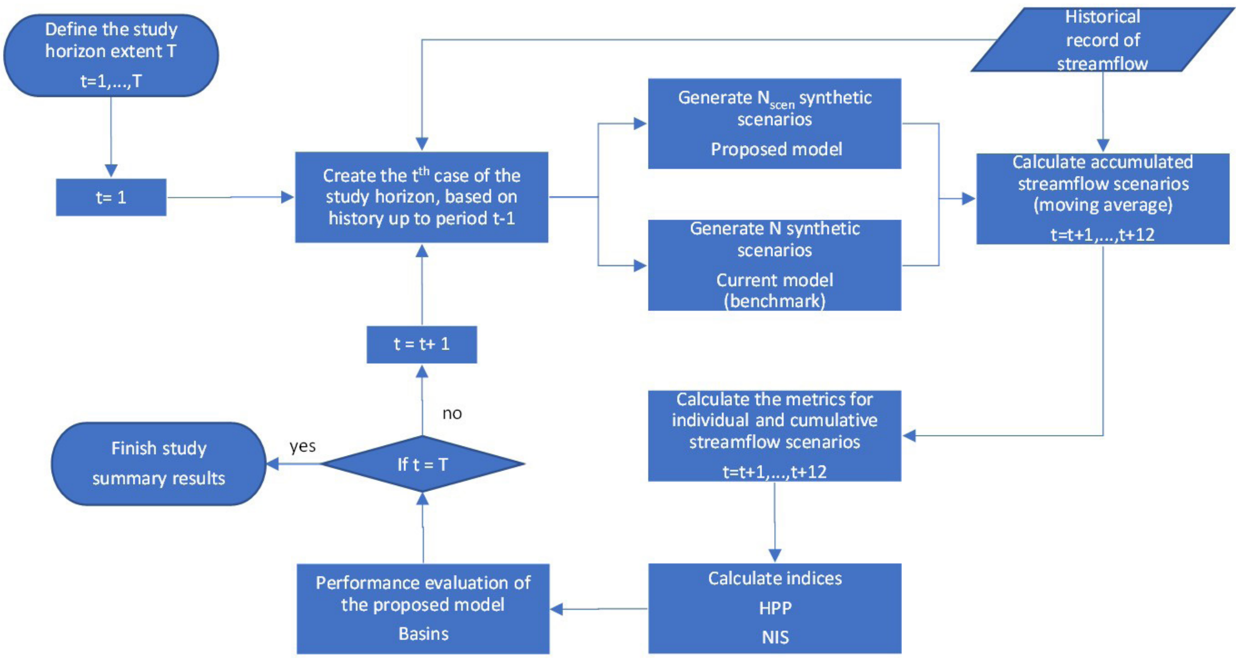


Figure 2. Flowchart of comparative evaluation framework.

Table 2. HPP's improvement percentage in comparison to the PAR(p)-A model, for each time step, using NCRPS as the metric. The highest-performing proposed model among the five is highlighted in red.

Model	t+1	t+2	t+3	t+4	t+5	t+6	t+7	t+8	t+9	t+10	t+11	t+12
CARMA	0.7%	6.8%	8.9%	6.2%	9.6%	10.3%	11.0%	11.0%	11.0%	13.0%	14.4%	10.3%
MS-PAR(p)	3.4%	3.4%	6.8%	8.2%	13.7%	13.7%	17.1%	12.3%	20.5%	22.6%	13.7%	23.3%
PARX	41.1%	**41.1%**	37.7%	34.2%	32.9%	34.9%	37.0%	28.8%	40.4%	41.1%	29.5%	30.8%
Hyper-multimodel	10.3%	15.1%	13.7%	13.7%	18.5%	17.8%	19.9%	16.4%	26.0%	26.0%	18.5%	24.7%
GHCen	**74.7%**	39.0%	**42.5%**	**42.5%**	**48.6%**	**50.0%**	**51.4%**	**54.1%**	**52.7%**	**52.7%**	**52.7%**	**54.8%**

Table 3. Improvement percentage for NEI in comparison to the PAR(p)-A model, for each time step, using NCRPS as the metric.

Model	t + 1	t + 2	t + 3	t + 4	t + 5	t + 6	t + 7	t + 8	t + 9	t + 10	t + 11	t + 12
CARMA	0.4%	4.1%	5.4%	2.0%	6.1%	16.2%	14.9%	18.4%	19.8%	19.5%	20.4%	17.0%
MS-PAR(p)	12.3%	6.9%	3.2%	6.5%	9.2%	6.7%	13.9%	9.4%	19.4%	19.8%	10.4%	27.1%
PARX	61.8%	**46.4%**	37.5%	35.7%	34.8%	33.6%	35.8%	30.6%	50.3%	52.7%	29.8%	31.9%
Hyper-multimodel	11.2%	7.4%	3.1%	12.3%	14.9%	16.5%	17.0%	16.3%	30.7%	29.2%	23.6%	28.5%
GHCen	**79.8%**	37.2%	**42.3%**	**36.1%**	**43.0%**	**53.4%**	**53.7%**	**55.5%**	**55.4%**	**55.2%**	**54.9%**	**56.0%**

Table 4. Improvement percentage for HPP in comparison to the PAR(p)-A model, for moving average, using NCRPS as the metric.

Model	MA 1	MA 2	MA 3	MA 4	MA 5	MA 6	MA 7	MA 8	MA 9	MA 10	MA 11	MA 12
CARMA	0.7%	1.4%	2.7%	2.1%	2.1%	2.1%	2.1%	2.1%	2.1%	2.7%	2.7%	2.1%
MS-PAR(p)	3.4%	6.2%	9.6%	12.3%	13.7%	14.4%	15.8%	14.4%	16.4%	17.1%	18.5%	17.8%
PARX	41.1%	45.2%	40.4%	33.6%	30.1%	28.1%	26.0%	22.6%	25.3%	24.0%	23.3%	21.9%
Hyper-multimodel	10.3%	15.1%	17.8%	18.5%	21.2%	22.6%	24.0%	22.6%	24.0%	26.0%	26.0%	27.4%
GHCen	**74.7%**	**51.4%**	**50.0%**	**52.7%**	**51.4%**	**50.7%**	**50.0%**	**50.0%**	**49.3%**	**51.4%**	**51.4%**	**50.7%**

Table 5. Improvement percentage for NEI in comparison to the PAR(p)-A model, for moving average, using NCRPS as the metric.

Model	MA 1	MA 2	MA 3	MA 4	MA 5	MA 6	MA 7	MA 8	MA 9	MA 10	MA 11	MA 12
CARMA	0.4%	2.1%	2.8%	2.4%	0.8%	0.8%	0.8%	0.8%	0.8%	3.9%	3.9%	3.8%
MS-PAR(p)	12.3%	12.7%	7.6%	11.6%	12.7%	11.9%	13.6%	8.1%	8.9%	9.4%	17.6%	17.6%
PARX	61.8%	**51.1%**	43.0%	33.4%	23.4%	20.5%	18.0%	16.2%	20.2%	19.6%	19.0%	18.9%
Hyper-multimodel	11.2%	18.3%	19.6%	17.0%	19.2%	20.8%	21.0%	20.6%	22.3%	22.7%	22.8%	24.1%
GHCen	**79.8%**	50.5%	**53.0%**	**64.7%**	**56.2%**	**54.8%**	**49.6%**	**50.5%**	**50.6%**	**51.1%**	**51.8%**	**51.1%**

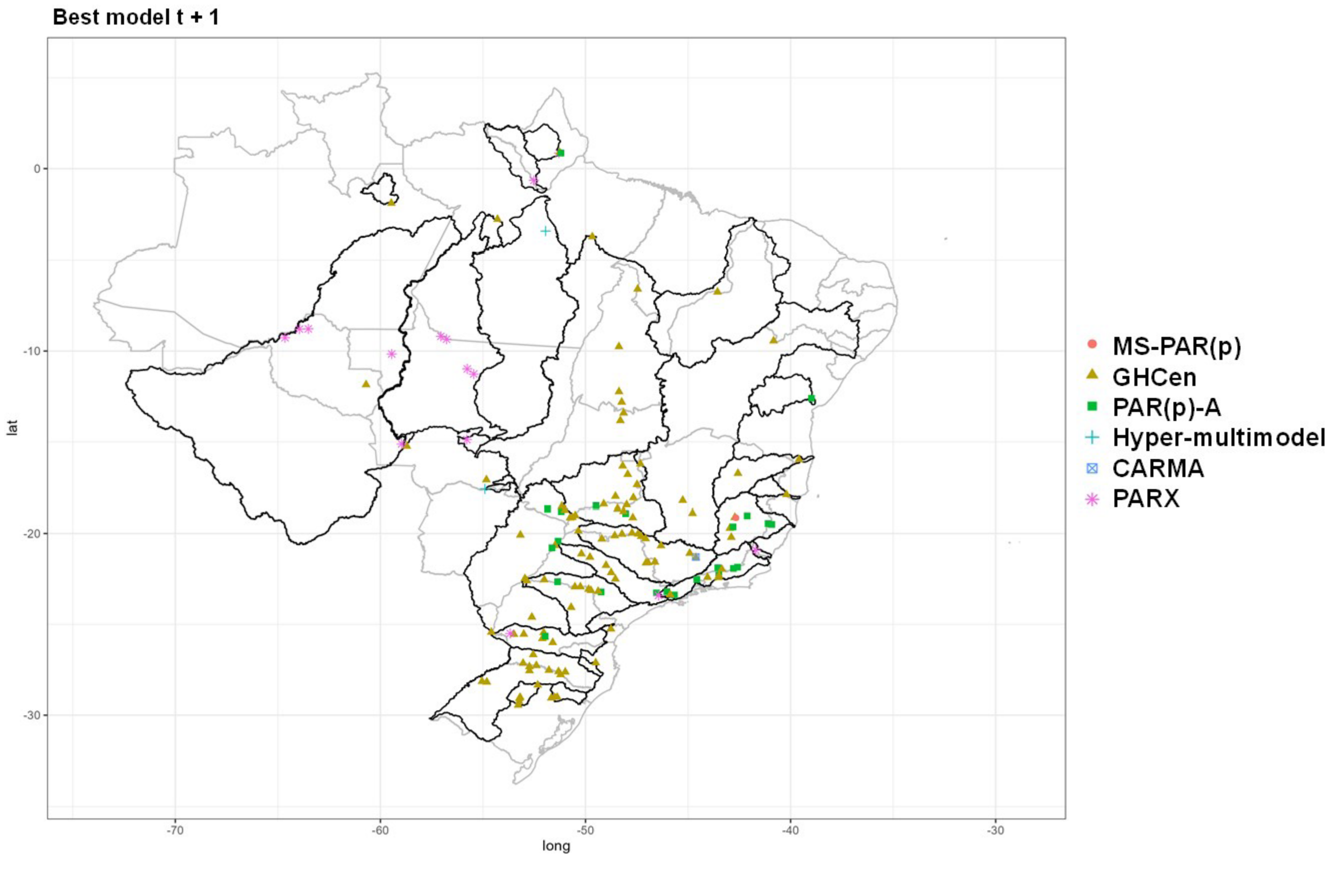


Figure 3. Lowest NCRPS model, in t + 1, for each HPP of NIS.

GHCen has, on average, the best performance both in the evaluation of the individual predictive distributions at each time step, and in the evaluation of moving averages. At t+1 for approximately 75% of the HPPs (80% of NEI) the GHCen model outperformed the current methodology. PARX has the second-highest percentage of improvement. Although the improvement about PAR(p)-A does not reach 50% of the HPPs in all horizons, for t+1 the PARX leads to an improvement of 61.8% of the NEI, indicating a better generation of synthetic inflow scenarios in HPPs with large representativeness in terms of NEI.

Figures 3 and 4 show the models with the lowest NCRPS in t+1 and the 12-month moving average, respectively, for each HPP of the NIS. The GHCen model represents most HPPs better for t+1, except for Teles Pires, Madeira (better represented by the PARX model), and Paraíba do Sul, Doce, and some parts of the incremental Paraná (better represented by the PAR(p)-A model). For the 12-month moving average, there is a more balanced representation between the different methodologies, with the current methodology and the GHCen being the most prevalent.

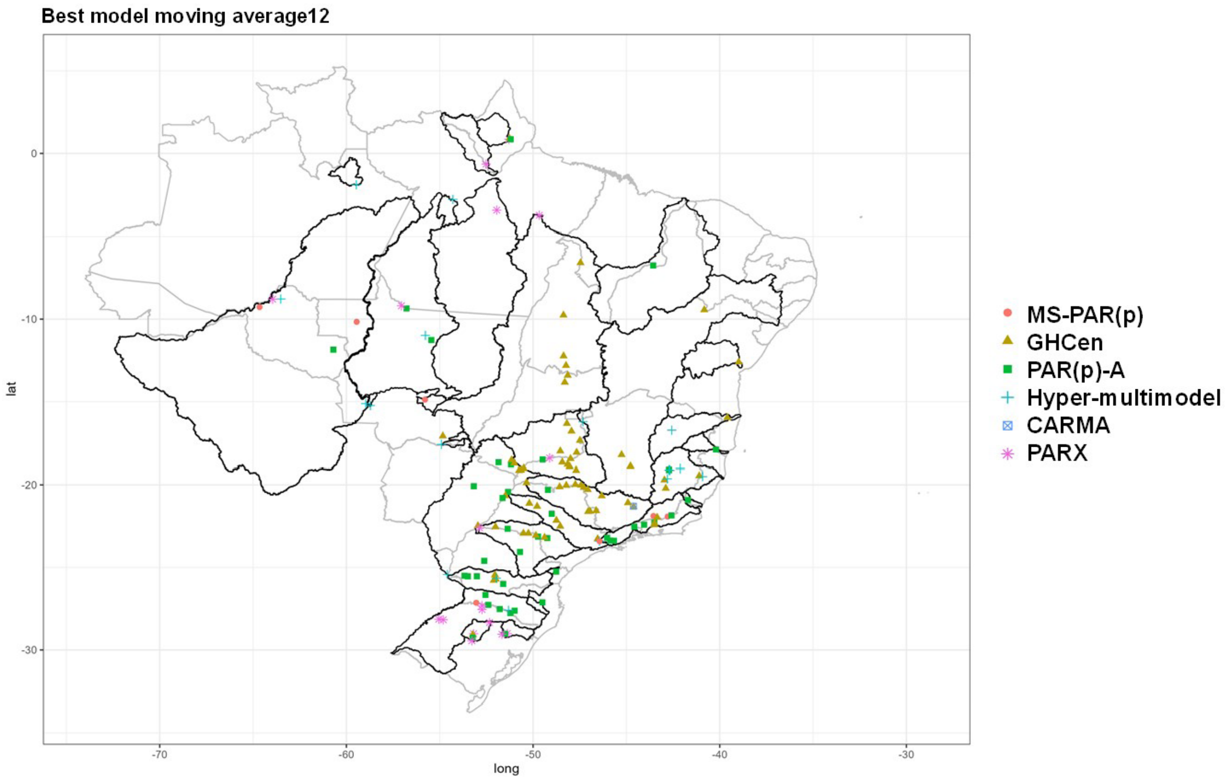


Figure 4. Lowest NCRPS model, for a moving average of twelve months, for each HPP of NIS.

CARMA results

From the results presented in Tables 2 to 5, it can be noted that the CARMA model has a lower performance than the current model and the other proposed models. In general, for the generation of scenarios from t+6, the model starts to present better performance compared to the previous steps. The CARMA model presents errors close to the current model, mainly in the South region, with a slight advantage mainly in the Uruguay River basin. Figure 5 shows the NCRPS for the predictive distributions at t+9 for the CARMA model (red bars) and the PAR(p)-A (blue bars), where, in general, similar values obtained by both models are observed.

However, in most of the analyses, the CARMA model is outperformed by PAR(p)-A. In the evaluation of moving averages, it is noted that the CARMA model has a lower performance for almost all HPPs. This indicates a low representativeness of the observed streamflow sequences, which may have occurred due to a small hydrological memory. Figure 6 shows the Boxplots of the predictive distributions in t+1 performed with the CARMA (red) and PAR(p)-A (blue) models, for the Furnas HPP over the period from 2014 to 2018, for a moving average of 12 months. It should be noted that PAR(p)-A generates predictive distributions that are the closest to the verified flows, represented by the black line.

MS-PAR(p) results

The results presented in Tables 2 to 5 indicate that the MS-PAR(p) model has, for most HPPs, a lower performance than the PAR(p)-A model when evaluated by the NCRPS, both in terms of predictive distributions at each step time and in terms of moving averages. In general, it is observed that the difference between the two models is greater in the short-term horizon, with a gradual improvement in the performance of the MS-PAR(p) model in longer time horizons of scenario generation.

Using the RMSE as an evaluation metric, despite not being the most suitable for the evaluation of predictive distributions, the MS-PAR(p) outperforms PAR(p)-A in a higher percentage of HPPs. Figure 7 shows the results of this analysis for each time step (a) and each moving average (b). Given that the RMSE is a metric that prioritizes errors at higher flows, this result indicates that the model has a better ability to capture flood events, especially in regions most affected by ENSO. Figure 8 presents the NRMSE of predictive distributions at t+1 in the northern region. It is possible to observe that for most HPPs the MS-PAR(p) model has smaller errors in this region when compared to the PAR(p)-A.

The analysis of the moving averages suggests that a better performance can also be observed in the North region. Although the southern region is notably a region that suffers from the influence of ENSO, even when evaluated through the NRMSE, the MS-PAR(p) model has lower results than the PAR(p)-A. Figures 9 and 10 show the predictive distributions of the 6-month moving averages performed with the MS-PAR(p) (red) and PAR(p)-A (blue) models, for the HPPs Itá and Santo Antônio Jari, corroborating previous analyses.

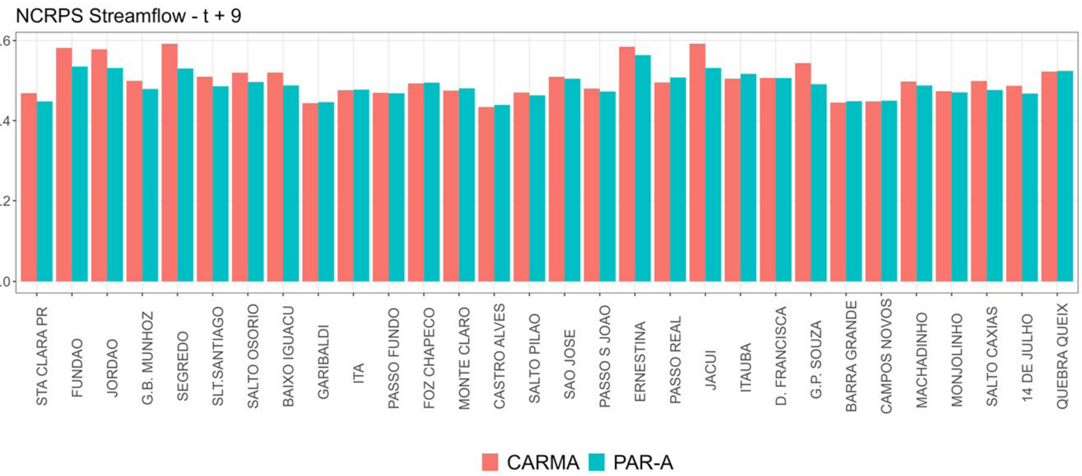


Figure 5. NCRPS of south region HPPs, for the predictive distributions at t + 9.

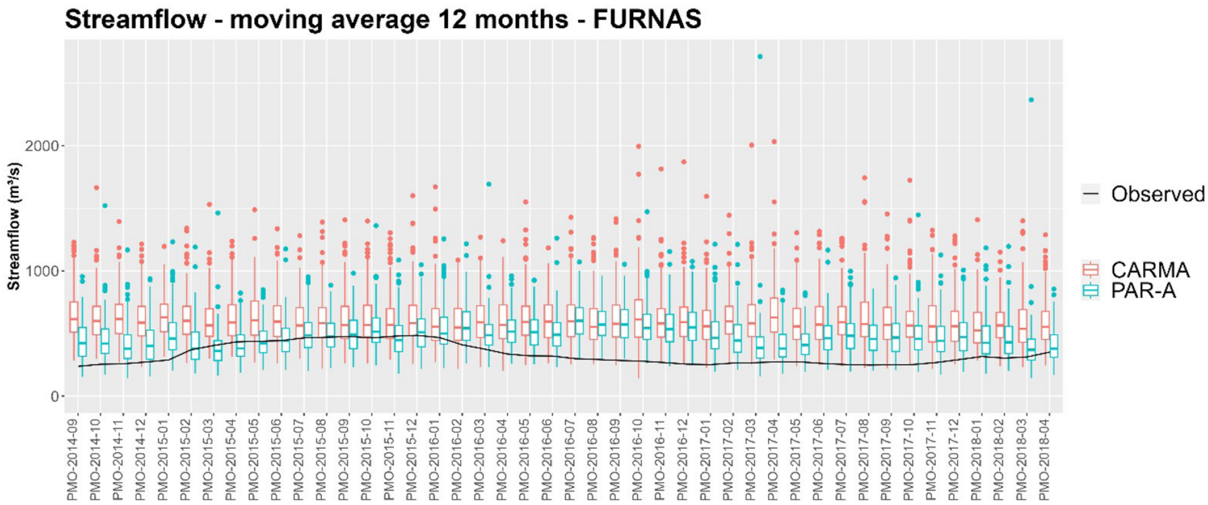
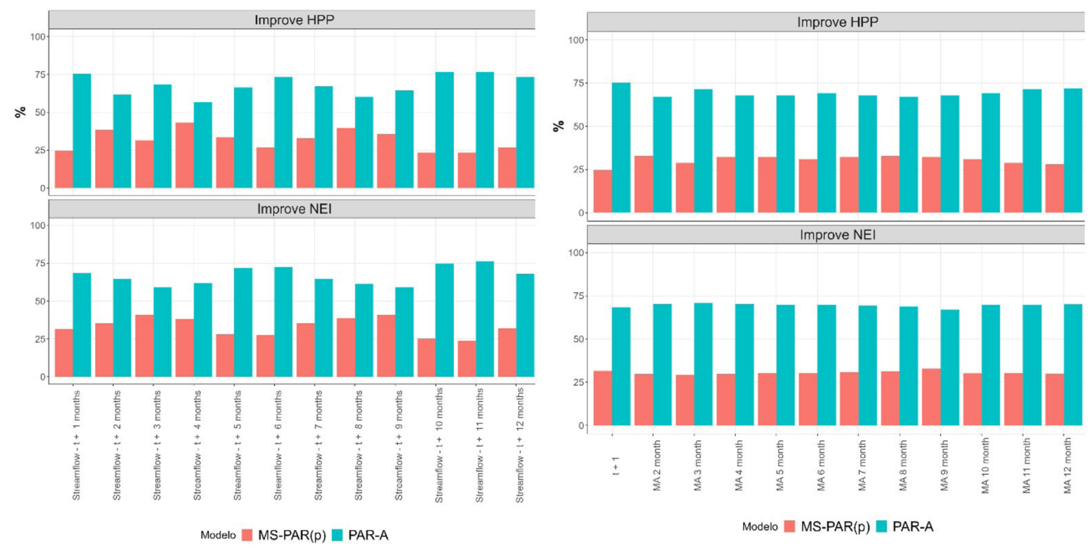


Figure 6. Boxplots of predictive distributions of twelve-month moving averages performed with CARMA (red) and PAR(p)-A (blue), for Furnas HPP.



(a) (b)

Figure 7. Percentage of improvement in HPP and NEI in terms of predictive distributions at each time step (a) and for moving averages (b), using NRMSE as a metric.

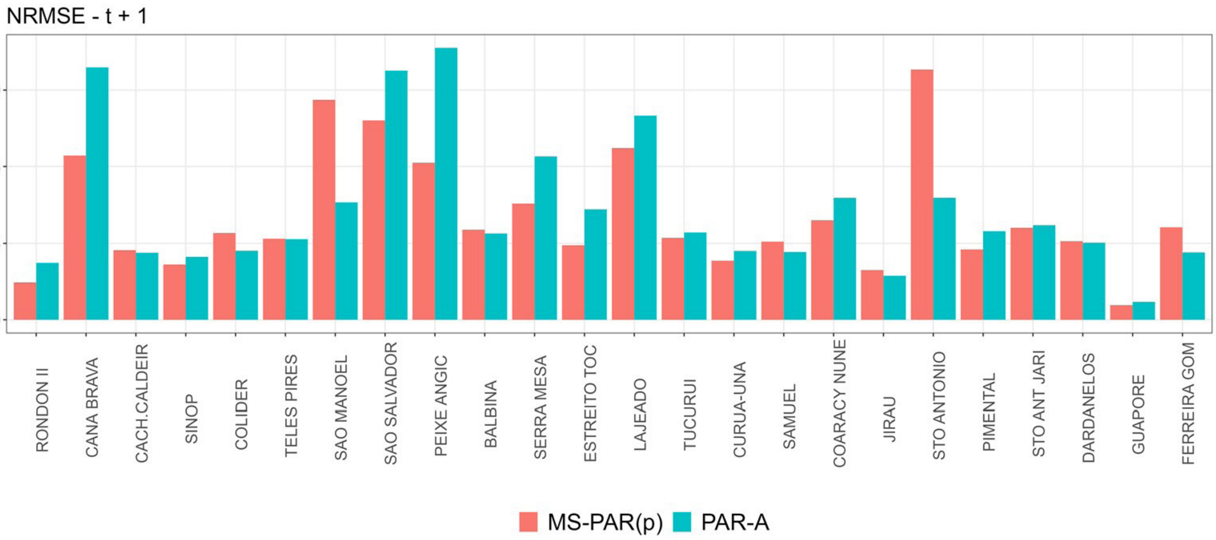


Figure 8. NRMSE of north region HPPs, for the predictive distributions at t + 1.

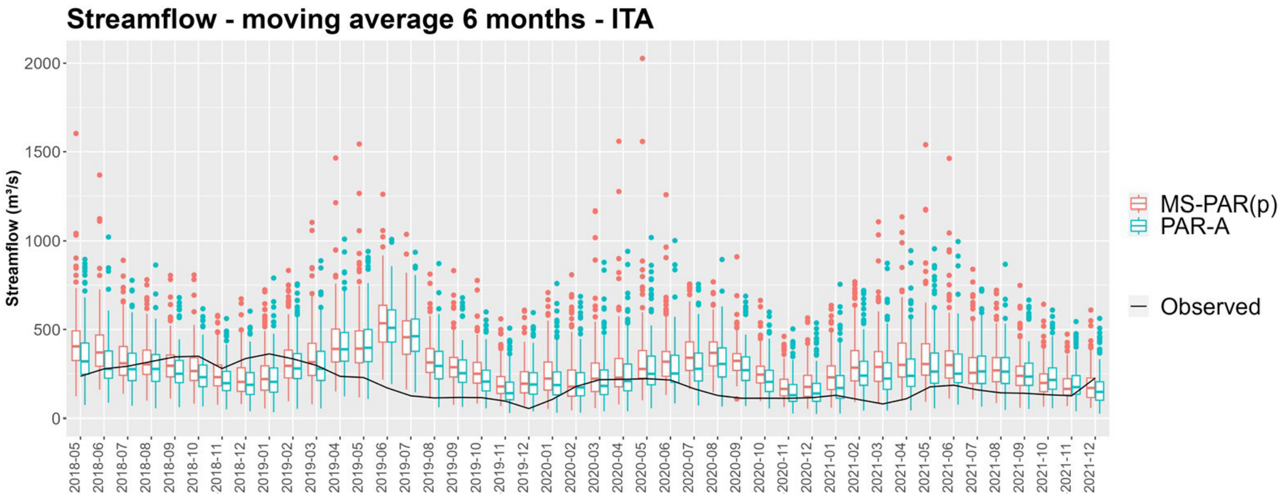


Figure 9. Boxplots of predictive distributions of six-month moving average performed with the MS-PAR(p) (red) and PAR(p)-A (blue), for the Itá HPP.

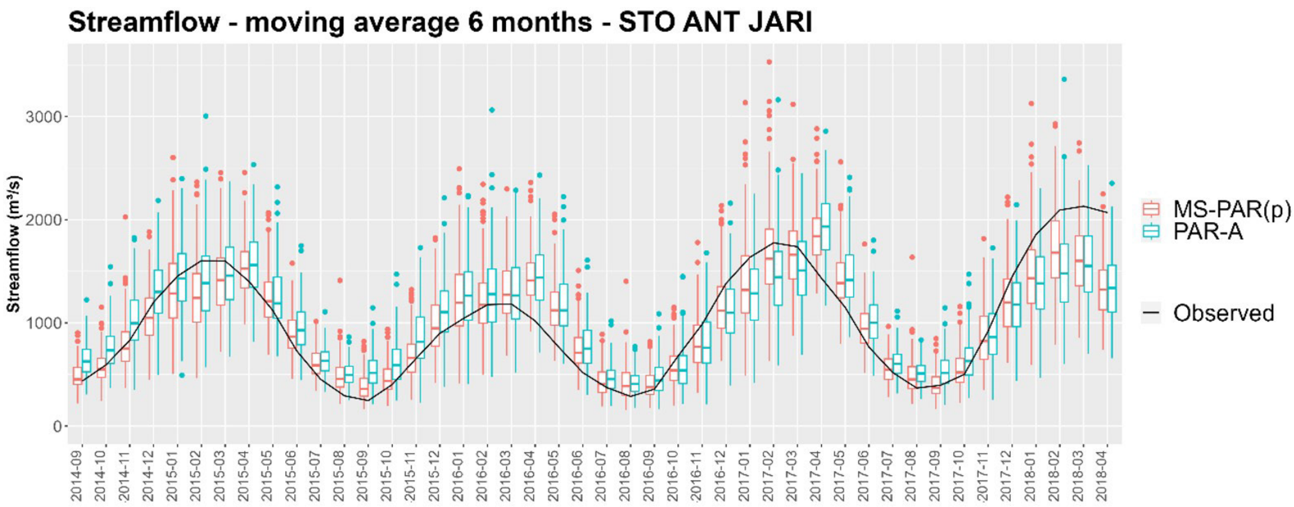


Figure 10. Boxplots of predictive distributions of six-month moving average performed with the MS-PAR(p) (red) and PAR(p)-A (blue), for the Santo Antônio Jari HPP.

Hyper-multimodel results

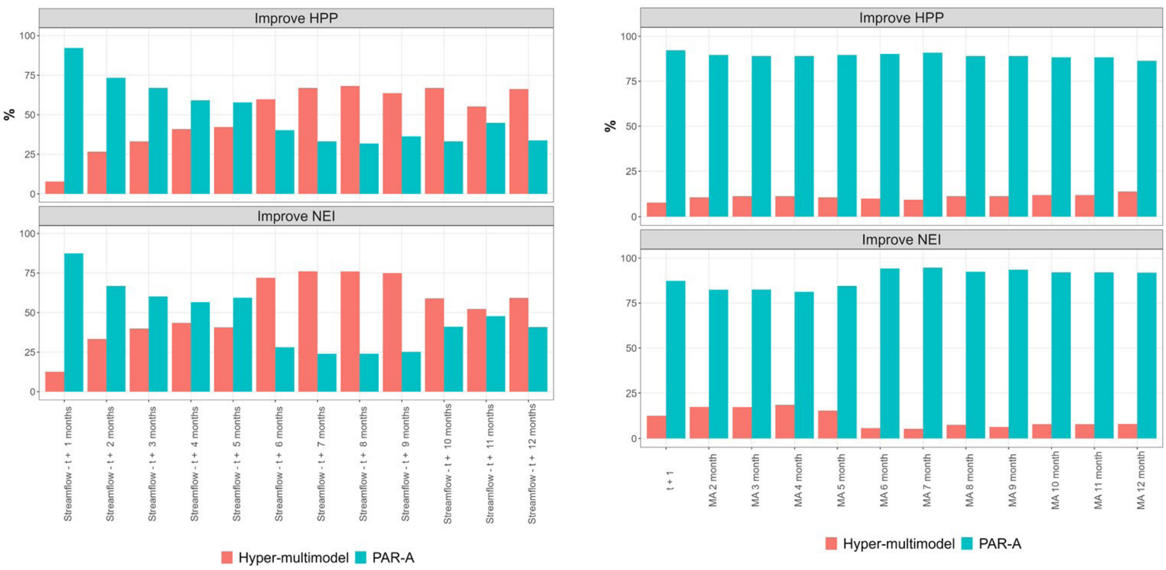
Analyzing the results presented in Tables 2 to 5, it is noted that the hyper-multimodel has a lower performance than the current model, for all evaluation horizons. The difference is generally greater in the short-term horizon, with a gradual increase in hyper-multimodel performance at each time step.

Based on the evaluations conducted, it was noticed that a considerable number of simulations produced scenarios that were either zero or close to zero. Generating scenarios with such low values is not desirable for a synthetic scenario creation model, and it might have adversely affected the outcomes.

Given this characteristic, MAPE was also used as an evaluation metric to deepen the analysis of the results a little more, even though this is not a metric indicated for the evaluation of predictive distributions. Figure 11 shows the percentage of improvement in UHE and ENA in terms of predictive distributions at each time step (a) and for moving averages (b), using MAPE as a metric. Two main points can be highlighted:

1. The hyper-multimodel shows an increasing performance at each time step, even being superior to the PAR(p) from t+6;
2. Despite having performed well for individual predictive distributions over the medium term, analysis of moving averages continues to underperform across the horizon.

To explain this poor performance in generating affluence sequences that can reproduce the observations, an evaluation of the synthetic temporal autocorrelations generated by the hyper-multimodel was carried out. Figure 12 presents for the Pimental



(a) (b)

Figure 11. Percentage of improvement in HPP and NEI in terms of predictive distributions at each time step (a) and for moving averages (b), using MAPE as a metric.

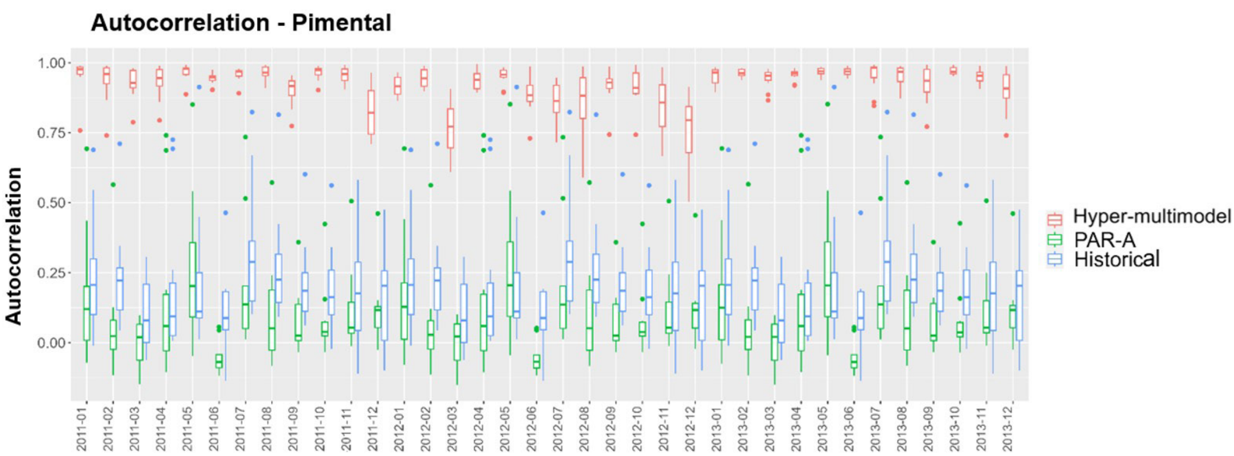


Figure 12. Boxplots of the synthetic temporal autocorrelations of hyper-multimodel (red), PAR(p)-A (blue), and for historical values (green), fixing the first month of each of the simulations, and varying the second vector of t +2 until t+12, for Pimental HPP.

HPP the boxplots of the temporal autocorrelations for the hyper-multimodel (red), for the PAR(p)-A (blue), and the historical values (green), setting the first month of each of the simulations and varying the second vector from t+2 to t+12. Therefore, each boxplot represents a set of 11 autocorrelations. Conditional scenario generation is not expected to reproduce historical values exactly. However, the temporal autocorrelations generated from the proposed model remain at very high values, regardless of the month of the simulation and the HPP considered. This behavior was observed in all HPPs.

Considering longer horizons, Figure 13 shows the simulations for the HPPs for t+6. It is noticeable that the predictive distributions of the Hyper-multimodel model have medians close to the observations. This fact occurs mainly in other HPPs with well-defined seasonality and for the months of the dry period when the distributions present a smaller dispersion. Even so, a large number of zeroed scenarios can be observed in most HPPs, negatively impacting the NCRPS.

In the evaluation of the moving averages, it is observed that, for some HPPs, there is a greater dispersion of the scenarios generated by the proposed model. However, the increase in the dispersion of the predictive distributions does not translate into a better performance of the model, as verified by the applied evaluation metrics. This behavior is illustrated in Figure 14, where it is notable that the hyper-multimodel has a lower performance than the current one. Finally, as stated in Souza Filho et al. (2023), it is important to notice that the results submitted in the GT-CH activity had operational errors in the hyper-multimodel scenario generation. The errors especially affected its spatial and temporal correlation, and may also impacted the performance of the proposed methodology. Further information about the errors can be found in the referred paper.

PARX results

As pointed out in Table 2 to Table 5, the PARX has one of the best performances among the evaluated models, mainly at t+1. Despite having an improvement in t+1 of 41% about the total number of HPPs, when weighted in terms of NEI the

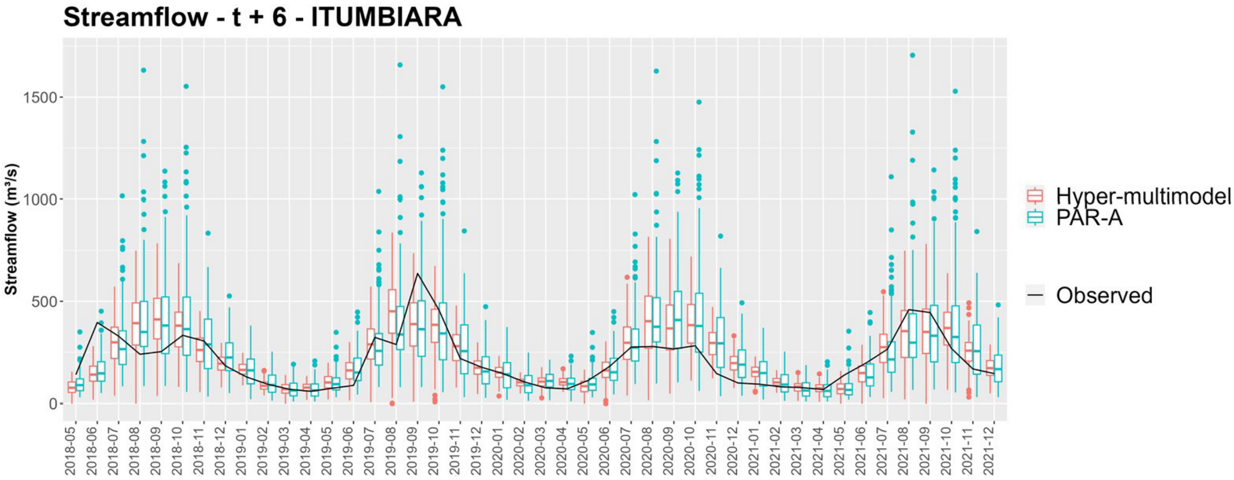


Figure 13. Boxplots of predictive distributions in t+6 performed with the Hyper-multimodel (red) and PAR(p)-A (blue), for the Itumbiara HPP.

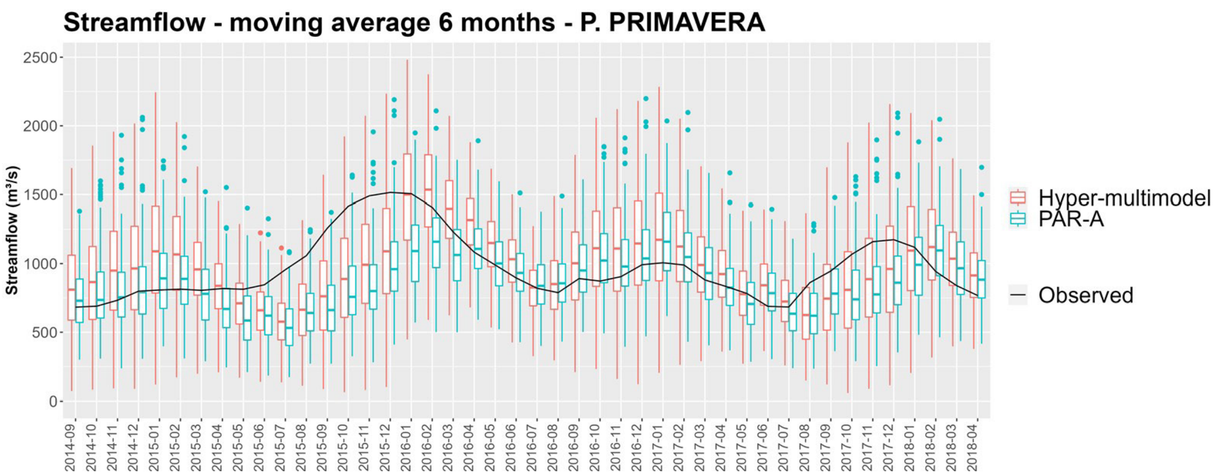


Figure 14. Boxplots of predictive distributions of six months moving average performed with the Hyper-multimodel (red) and PAR(p)-A (blue), for the Porto Primavera HPP.

improvement reaches 61.8%. Therefore, the model manages to improve HPPs that contribute proportionally more to the NEI. In general, it is observed that the model performs better in the short term, which gradually decreases with each time step.

By including exogenous variables related to weather events such as ENSO, the proposed model performs better mainly in the north and south regions of the country. Figure 15 presents the NCRPS of the HPPs present in the north region for the predictive distributions in t+1, generated with the PARX model (red) and PAR(p)-A (blue). It is noticeable that most HPPs in these regions have lower NCRPS values with the PARX model in comparison to PAR(p)-A.

It is interesting to note that at t+1, where the best PARX performance occurs, there are periods in which the generated predictive distribution has a notably different pattern than that observed in PAR(p)-A, due to the greater influence of climatic variables. To exemplify this characteristic of the proposed model, Figures 16 and 17 show the predictive distributions in t+1 performed with PARX (red) and PAR(p)-A (blue), for the Pimental and Baixo Iguaçu HPPs, respectively, between 2014 and early 2016. It is observed that for UHE Pimental, the proposed model generated drier distributions in the years 2014 and 2015. As for Baixo Iguaçu HPP, the predictive distributions generated have scenarios with higher flows. These results were likely influenced by the presence of the El Niño phenomenon that occurred during those years.

According to the analyses, the proposed model at each time step has a decrease in its performance in terms of moving average, from 45% in the two-month moving average to 21% in the twelve-month moving average. This suggests that the model has difficulty replicating the streamflow sequences. To understand the reason for this behavior, the reproduction of synthetic temporal autocorrelations generated by PARX was evaluated. Figure 18 shows for Furnas HPP the boxplots of synthetic temporal autocorrelations generated with PARX (red), with PAR(p)-A (blue) and historical

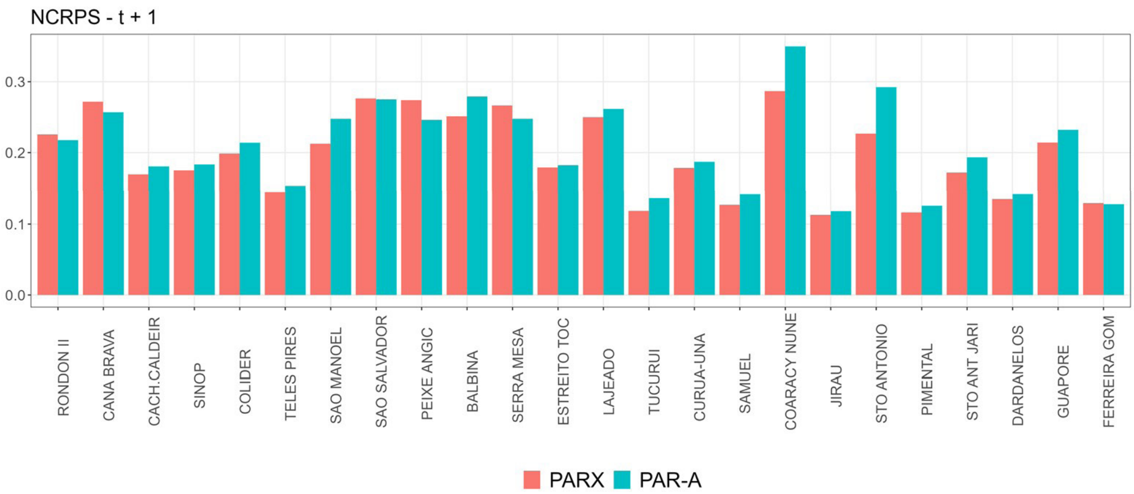


Figure 15. NCRPS of north region HPPs, for the predictive distributions at t + 1.

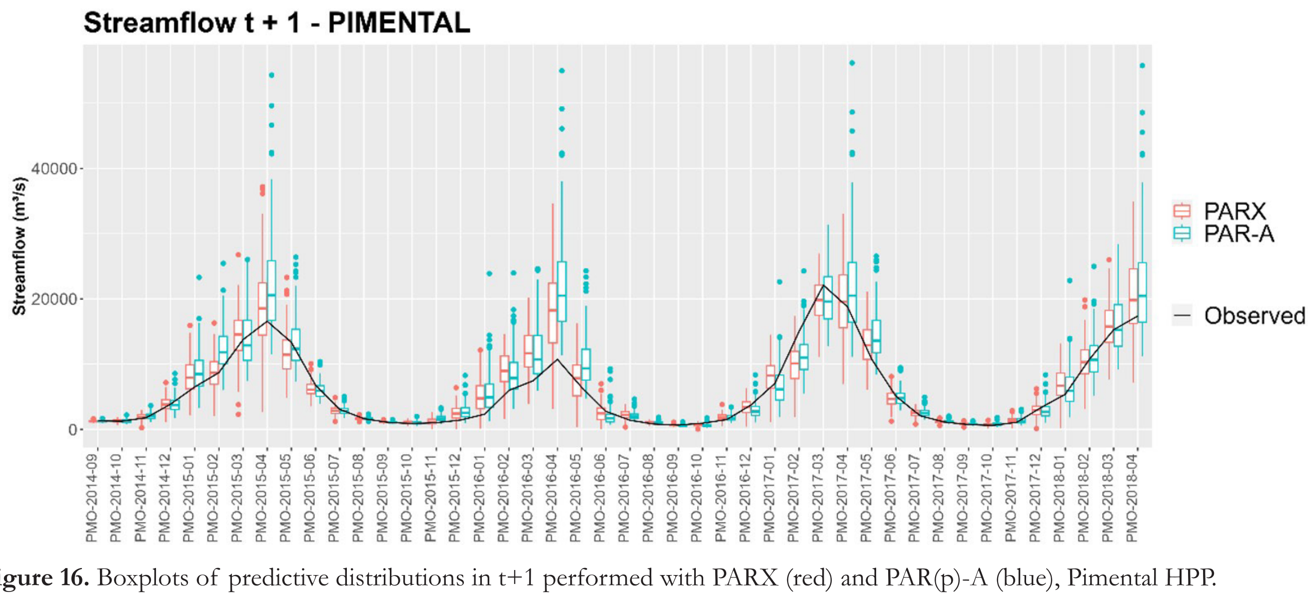


Figure 16. Boxplots of predictive distributions in t+1 performed with PARX (red) and PAR(p)-A (blue), Pimental HPP.

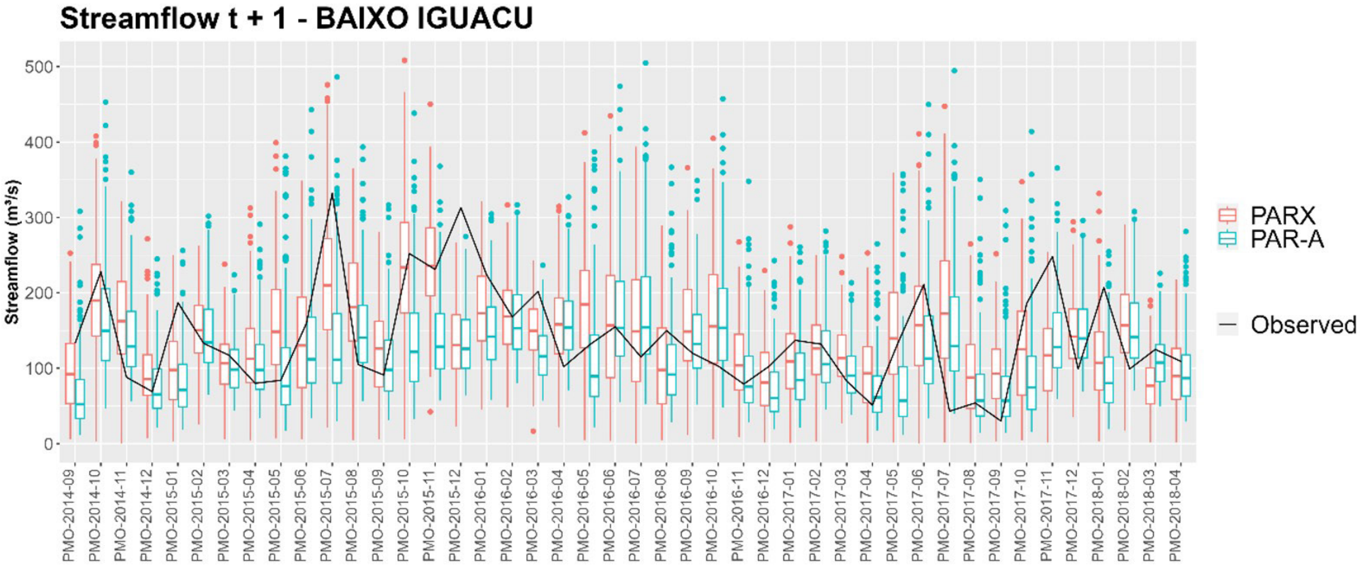


Figure 17. Boxplots of predictive distributions in t+1 performed with PARX (red) and PAR(p)-A (blue), Baixo Iguaçu HPP.

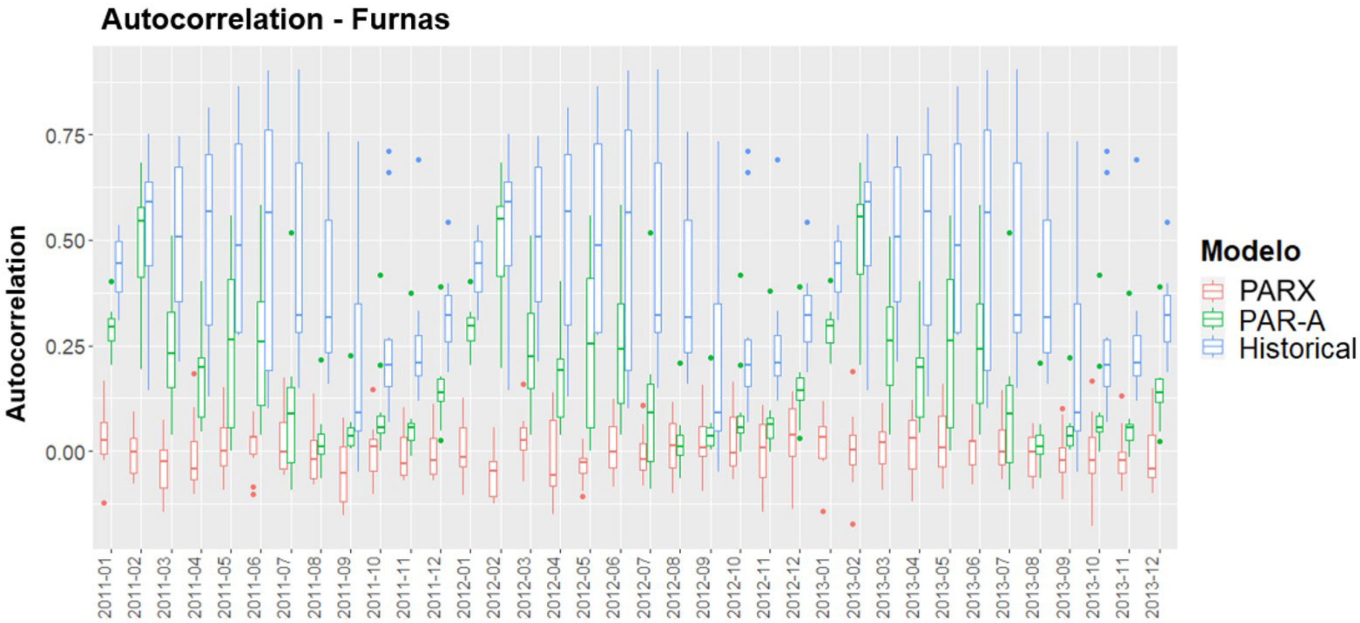


Figure 18. Boxplots of the synthetic temporal autocorrelations of the model proposed by the PARX (red), PAR(p)-A (blue), and for the historical values (green), fixing the first month of each of the simulations, and varying the second vector from t+2 to t+12, for Furnas HPP.

temporal autocorrelations (green), fixing the first month of each of the simulations, and varying the second vector from t+2 to t+12. The temporal autocorrelations generated from the proposed model remain at values close to zero.

This indicates that there is no relationship between the predictive distributions at each time step, which explains the lower performance when analyzing the moving averages. A justification for this behavior may lie in the fact that the models are adjusted for each time step, taking into account only the observed values. In other words, the synthetic scenarios generated in t+1 do not enter the equation for the generation of scenarios in t+2.

GHCen results

The evaluation of the GHCen model indicates that it performs better than the current model in most of the evaluated horizons, according to Tables 2 and 5. For horizon t+1, it is observed that in almost 80% of the NEI of the NIS, there is an improvement in the current model. As for horizons t+2 to t+5, this superiority is reversed, with more than 50% improvement in longer horizons. When analyzing the predictive distributions generated in terms of moving averages, the GHCen model maintains a superior performance in all evaluated horizons. This indicates that the trajectories of the synthetic scenarios generated by GHCen are closer to the observed flow sequences when compared to the current model.

In general, for horizon t+1, the proposed model presents superior results in most of the HPPs. Figure 19 presents the NCRPS of the HPPs present in the Rio Grande and Paranaíba basins for the predictive distributions in t+1 generated with the GHCen model (red) and PAR(p)-A (blue).

The exception occurs mainly in the North region, where the GHCen has its worst performance. In this region, the observed series of daily precipitation has low quality when compared to the rest of the SIN. This fact is mainly due to the low availability of rainfall stations and the poor performance of satellite rainfall estimates in this region. This issue ends up impacting both the calibration of the SMAP/ONS model, the data assimilation process produced by the model, and the generation of synthetic precipitation scenarios.

By using a conceptual rainfall-runoff model (SMAP/ONS), which provides estimates of surface flow, base flow, and percentage of soil moisture, together with a stochastic model that generates synthetic scenarios capable of reproducing the main characteristics of the historical series of precipitation, GHCen manages to generate more adherent predictive distributions mainly in the short term. Additionally, as it is a daily model, it can identify whether flows are rising or falling, which is an additional advantage over models with monthly discretization.

To exemplify the performance in t+1, Figure 20 presents the boxplots of the predictive distributions in t+1 carried out with the GHCen (red) and PAR(p)-A (blue), for the Furnas HPP. Notably, the GHCen model generates predictive distributions that are closer to the observations, and in some cases, with a lower variability than that generated by PAR(p)-A, mainly in recessions.

In addition to the quality of the precipitation data, a second factor was identified that has a direct influence on the quality of

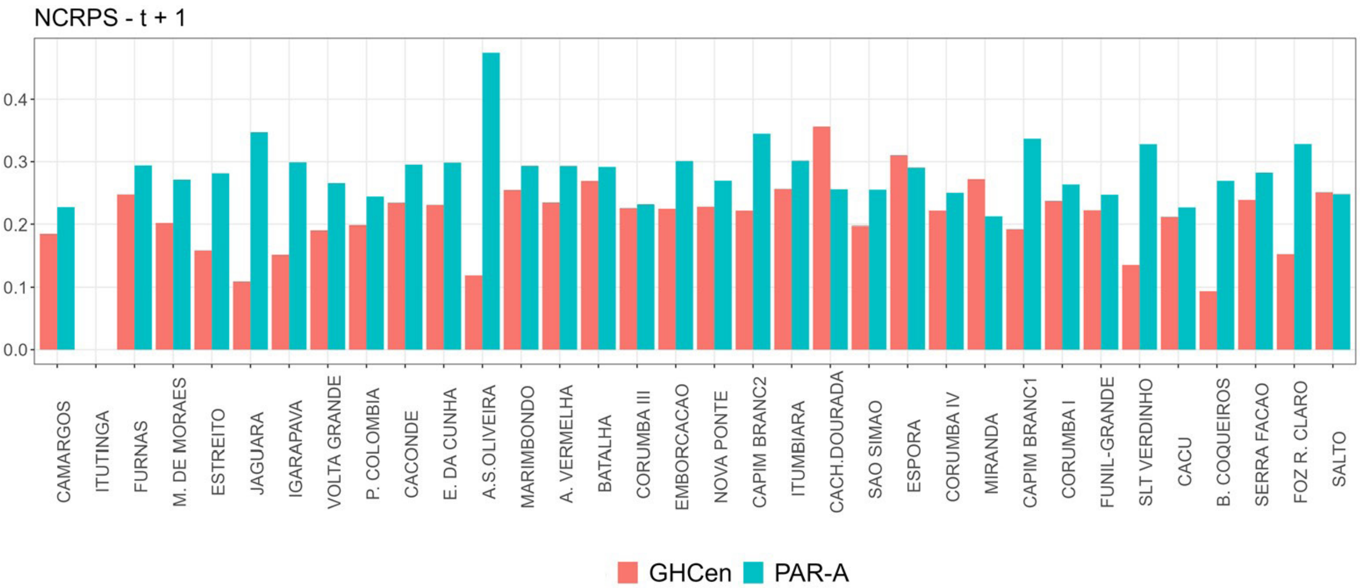


Figure 19. NCRPS Grande and Paranaíba basin HPPs, for the predictive distributions at t + 1.

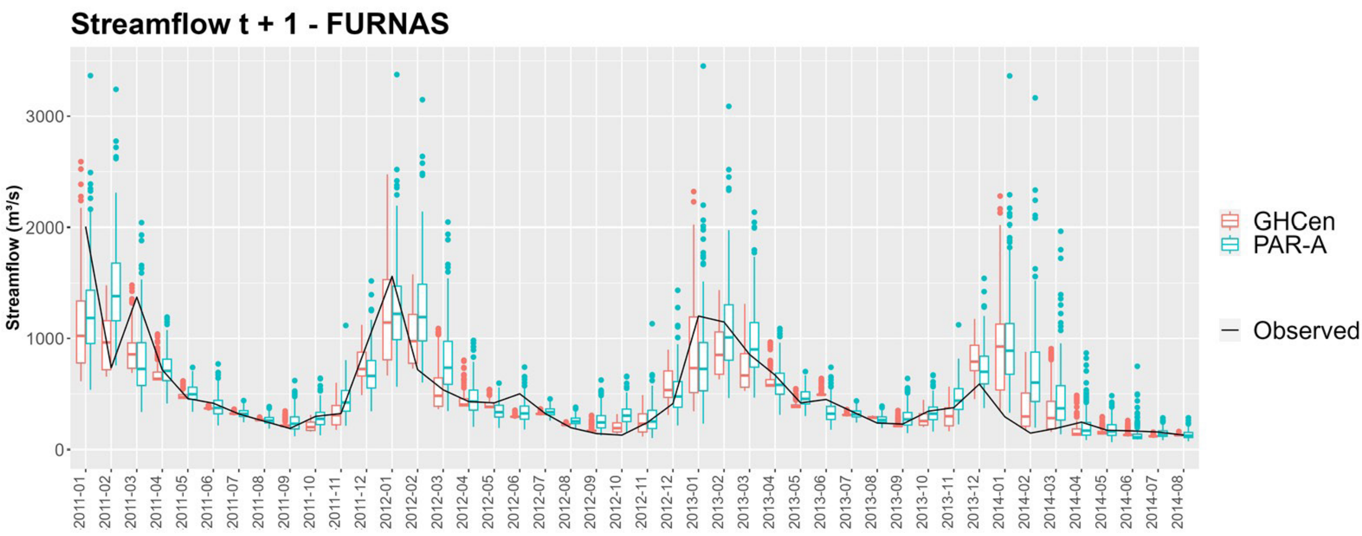


Figure 20. Boxplots of predictive distributions in t+1 performed with GHCen (red) and PAR(p)-A (blue), Furnas HPP.

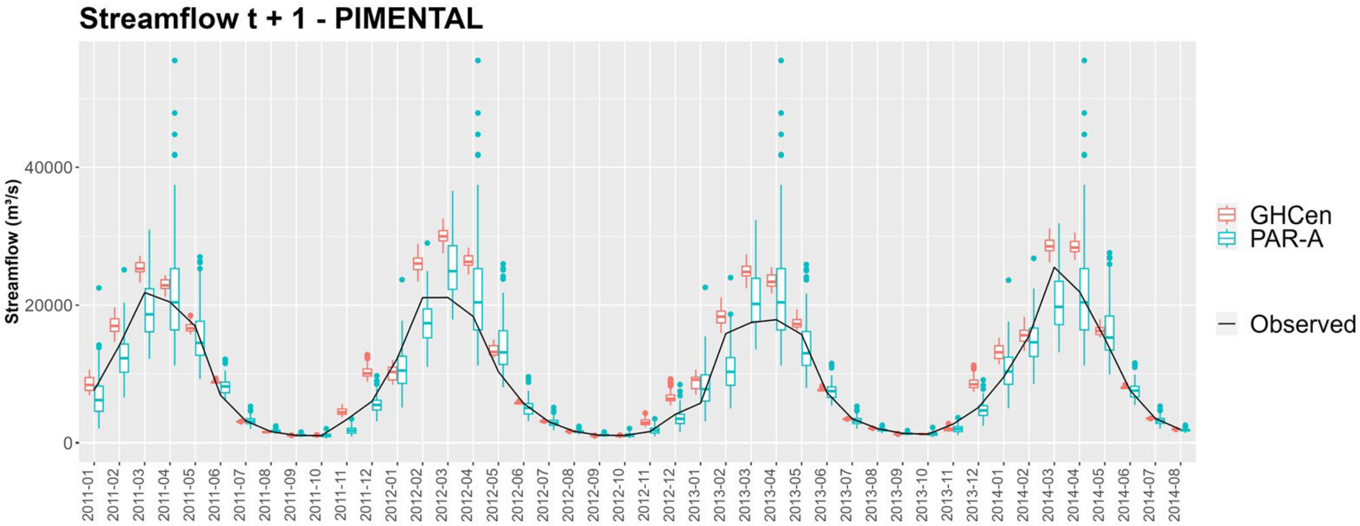


Figure 21. Boxplots of predictive distributions in t+1 performed with GHCen (red) and PAR(p)-A (blue), Pimental HPP.

the predictive distributions generated in t+1. For HPPs that are influenced by rainfall from many days ago, as in the case of Belo Monte with kt=60 days (calibration parameter of the SMAP/ONS model), considering only the rainfall uncertainty seems not to be enough to generate adequate predictive distributions in this horizon. This influence of precipitation many days ago usually occurs in basins with a large drainage area. For longer horizons, the conditioning effect of data assimilation is lost, resulting in distributions with greater variability.

Figure 21 presents the Boxplots of the predictive distributions generated with the GHCen (red) and PAR(p)-A (blue), for the Pimental HPPs in t+1. It is observable that the synthetic scenarios generated by GHCen have a much lower variability than the PAR(p)-A model. Additionally, the generated distributions fail to capture the observed values, especially during wet periods. These problems are not observed in longer horizons.

CONCLUSIONS

Streamflow is one of the main uncertainties in hydrothermal scheduling problems. Therefore, countries with those characteristics have a significant need to improve the accuracy of scenario generation models. Although there is plenty of literature on ways to compare forecast models, the same cannot be stated for scenario generation models, which have different characteristics and purposes.

This paper proposes a framework to evaluate and compare scenario generation models, based mainly on the NCRPS metric. In addition to assessing the predictive distribution each time step, the NCRPS was calculated for moving averages, giving insights into which model can better represent the observed streamflow sequences. The framework is applied in a case study for the NIS, comparing six streamflow scenario generation models, with distinct approaches, that represent the state-of-the-art in this field.

From the analysis, it was observed that GHCen was the model with the best results for both the t+1 step and the moving average. PARX model was able to better represent HPPs from the north and south regions, mainly in the t+1 step, but the representation of the streamflow sequences had a lower performance. Although the PAR(p)-A does not include any climate variable, it was, after the GHCen, the model with the best performance in the moving average analysis.

REFERENCES

Box, G. E., Jenkins, G. M., & Reinsel, G. C. (2008). *Time series analysis: forecasting and control.* Hoboken: John Wiley & Sons.

Cassagnole, M., Ramos, M.-H., Zalachori, I., Thirel, G., Garçon, R., Gailhard, J., & Ouillon, T. (2021). Impact of the quality of hydrological forecasts on the management and revenue of hydroelectric reservoirs: a conceptual approach. *Hydrology and Earth System Sciences*, *25*(2), 1033-1052. http://dx.doi.org/10.5194/hess-25-1033-2021.

Comissão Permanente para Análise de Metodologias e Programas Computacionais do Setor Elétrico – CPAMP. (2021). *Modelo PAR(p)-A de representação hidrológica e avaliação da parametrização do CVaR - Ciclo 2021/2022.* In Portuguese. Retrieved in 2023, October 14, from https://antigo.mme.gov.br/c/document_library/get_file?uuid=b88a8d0f-0f0f-13e8-6c85-5a1e84b6a258&groupId=36070

Detzel, D. H. M., Bessa, M. R., Ávila, L. R., Cantão, M. P., & de Geus, K. (2023). Generation of synthetic flow scenarios by means of multivariate sampling of contemporaneous ARMA model outputs. *Revista Brasileira de Recursos Hídricos*, *28*. In press. http://dx.doi.org/10.1590/2318-0331.282320230117.

Hersbach, H. (2000). Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting*, *15*(5), 559-570. http://dx.doi.org/10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2.

Lappicy, T., & Lima, C. H. (2023). Enhancing monthly streamflow forecasting for Brazilian hydropower plants through climate index integration with stochastic methods. *Revista Brasileira de Recursos Hídricos*, *28*. In press. http://dx.doi.org/10.1590/2318-0331.282320230118.

Lima, C. H., & Lall, U. (2010). Climate informed monthly streamflow forecasts for the Brazilian hydropower network using a periodic ridge regression model. *Journal of Hydrology*, *383*(3-4), 307-318. http://dx.doi.org/10.1016/j.jhydrol.2009.12.045.

Maceira, M. E. P., Penna, D. D. J., Diniz, A. L., Pinto, R. J., Melo, A. C. G., Vasconcellos, C. V., & Cruz, C. B. (2018, June 11-15). Twenty years of application of stochastic dual dynamic programming in official and agent studies in Brazil-main features and improvements on the NEWAVE model. In: Institute of Electrical and Eletronical Engineers, editor. *2018 Power Systems Computation Conference* (pp. 1-7). New York: IEEE.

Operador Nacional do Sistema Elétrico – ONS. (2017). *Manual de metodologia - aplicativo SMAP/ONS*. Rio de Janeiro: Operador Nacional do Sistema Elétrico.

Operador Nacional do Sistema Elétrico – ONS. (2023a). Retrieved in 2023, September 1, from https://www.ons.org.br/paginas/sobre-o-sin/mapas

Operador Nacional do Sistema Elétrico – ONS. (2023b). Retrieved in 2023, March 21, from https://sintegre.ons.org.br/sites/9/13/84//paginas/servicos/produtos-pasta.aspx?RootFolder=%2Fsites%2F9%2F13%2F84%2FProdutos%2F427%2F30-11-2022_115600&produto=S%C3%A9ries%20Hist%C3%B3ricas%20de%20Vaz%C3%B5es%20Naturais&descricaoProduto=Atualiza%C3%A7%C3%A3o%20de%20S%C3%A9ries%20Hist%C3%B3ricas%20de%20Vaz%C3%B5es%20-%201931%20a%202021

Pessanha, J. F. M., Almeida, V. A., & Chan, P. S. (2023). MS-PAR(p): generation of synthetic flow scenarios using a markov-switching periodic auto-regressive model. Hyper-multimodel. *Revista Brasileira de Recursos Hídricos*, *28*. In press. http://dx.doi.org/10.1590/2318-0331.282320230114.

Souza Filho, F. A., Rocha, R. V., Estacio, A. B. S., Rolim, L. R. Z., Pontes Filho, J. D. A., & Porto, V. C. (2023). Enhancing streamflow forecasting for the Brazilian electricity sector: a strategy based on a hyper-multimodel. *Revista Brasileira de Recursos Hídricos*, *28*. In press. http://dx.doi.org/10.1590/2318-0331.282320230120.

Treistman, F., Maceira, M. E. P., Damázio, J. M., & Cruz, C. B. (2020a, August 18-21). Periodic time series model with annual component applied to operation planning of hydrothermal systems. In: Institute of Electrical and Eletronical Engineers. PMAPS International Society, editors. *2020 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)* (pp. 1-6). New York: IEEE. http://dx.doi.org/10.1109/PMAPS47429.2020.9183472.

Treistman, F., Maceira, M. E. P., Penna, D. D. J., Damázio, J. M., & Rotunno Filho, O. C. (2020b). Synthetic scenario generation of monthly streamflows conditioned to the El Niño–Southern Oscillation: application to operation planning of hydrothermal systems. *Stochastic Environmental Research and Risk Assessment*, *34*(2), 331-353. http://dx.doi.org/10.1007/s00477-019-01763-2.

Treistman, F., Penna, D. D. J., & Khenayfis, L. S. (2023). GHCen: a stochastic-conceptual approach for generating synthetic streamflow scenarios. Hyper-multimodel. *Revista Brasileira de Recursos Hídricos*, *28*. In press. http://dx.doi.org/10.1590/2318-0331.282320230116.

Authors contributions

Felipe Treistman: Comparative framework conceptualization and discussions, activities coordinator, text writing and revision, GHCen results generation and analysis.

Débora Dias Jardim Penna: Comparative framework discussions, text writing and revision, results analysis.

Lucas de Souza Khenayfis: Comparative framework discussions, text revision, results analysis.

Nelson Bernardo Rodrigues Cavalcante: Comparative framework discussions, activities coordinator, text revision.

Francisco de Assis de Souza Filho: Comparative framework discussions, activities coordinator from Hyper-multimodel, text revision.

Renan Vieira Rocha: Comparative framework discussions, text revision, assistant coordinator of the activities from hyper-multimodel, generation of hyper-multimodel results.

Ályson Brayner Estácio: Comparative framework discussions, assistant coordinator of the activities from hyper-multimodel, generation of hyper-multimodel results.

Larissa Zaira Rafael Rolim: Generation of hyper-multimodel results.

João Dehon de Araújo Pontes Filho: Generation of hyper-multimodel results.

Victor Costa Porto: Generation of hyper-multimodel results.

Sullyandro Oliveira Guimarães: Generation of hyper-multimodel results.

José Francisco Moreira Pessanha: Comparative framework discussions and assistant coordinator of the activities from MS-PAR(p).

Victor Andrade de Almeida: Comparative framework discussions, coordinator of the activities from MS-PAR(p), generation and consolidation of the results from MS-PAR(p) and text revision.

Priscilla Dafne Shu Chan: Generation of the results from MS-PAR(p).

Thiago Lappicy: Comparative framework discussions, generation of PARX results, text revision.

Carlos Henrique Ribeiro Lima: Comparative framework discussions, text revision.

Daniel Henrique Marco Detzel: Comparative framework discussions, generation of CARMA results, text revision.

Marcelo Rodrigues Bessa: Comparative framework discussions, generation of CARMA results, text revision.

Editor-in-Chief: Adilson Pinheiro

Associated Editor: Adilson Pinheiro