MS-PAR(p): generation of synthetic flow scenarios using a Markov-switching periodic auto-regressive model

MS-PAR(p): geração de cenários sintéticos de vazões utilizando um modelo autoregressivo periódico com chaveamento markoviano

José Francisco Moreira Pessanha[1] ⓘ**, Victor Andrade de Almeida[1]** ⓘ **& Priscilla Dafne Shu Chan[1]** ⓘ

[1]Centro de Pesquisas de Energia Elétrica, Rio de Janeiro, RJ, Brasil
E-mails: francisc@cepel.br (JFMP), andrade@cepel.br (VAA), pchan@cepel.br (PDSC)

ABSTRACT

The operation planning of the National Interconnected System (NIS) is based on optimization models that use synthetic inflow scenarios to represent the periodic behavior observed in historical data. Currently, the PAR(p)-A model (Periodic Autoregressive with Annual Component) is officially employed in computational models by the responsible organizations for short and medium-term operation planning. This paper has the aim of presenting an experiment using an alternative model that takes into consideration information regarding climatic variables, which can influence the hydrological regime of river basins and therefore the entire energy planning. The evaluated model employs the ONI index as a measure of the El Niño-Southern Oscillation (ENSO) phenomenon, in addition to a Markovian switching process. The results of the experiment demonstrate that the methodology is able to capture the influence of this phenomenon on inflows and generate scenarios closer to observed flow values.

Keywords: Synthetic streamflow scenario generation; Markov chain; Autoregressive models; Operation planning; ENSO.

RESUMO

O planejamento da operação do Sistema Interligado Nacional (NIS) é baseado em modelos de otimização que utilizam cenários SINtéticos de afluências que buscam representar o comportamento periódico observado no histórico dos dados. Atualmente, o modelo PAR(p)-A (Autoregressivo Periódico com Componente Anual) é empregado de forma oficial nos modelos computacionais pelos órgãos responsáveis no planejamento da operação no curto e médio prazo. O objetivo deste trabalho é apresentar um experimento utilizando um modelo alternativo que considera alguma informação sobre variáveis climáticas, que podem influenciar no regime hidrológico das bacias e, por consequência, em todo o planejamento energético. O modelo utilizado emprega o índice ONI como medida do fenômeno El Niño – Oscilação Sul, além de um processo de chaveamento markoviano. Os resultados do experimento mostram que a metodologia é capaz a influência deste fenômeno nas afluências e gerar cenários mais próximos dos valores de vazão observados.

Palavras-chave: Geração de cenários sintéticos de afluências; Cadeia de Markov; Modelos autorregressivos; Planejamento da operação; ENOS.

INTRODUCTION

The Brazilian electrical energy matrix has an installed capacity of 210GW, with approximately 110GW coming from hydroelectric sources, which corresponds to about 52% of the total (Operador Nacional do Sistema Elétrico, 2023a). In addition, hydroelectric generation was responsible for approximately 72% of the power generated in 2022 (Câmara de Comercialização de Energia Elétrica, 2023). Due to this high dependence on the hydrological factor, it can be stated that the success of the operation planning is directly linked to the knowledge of the expected behavior of streamflows during the study horizon.

The operation planning of the National Interconnected System (NIS) officially uses, through the Ministry of Mines and Energy (MME), the National Operator of the Electric System (ONS), the Chamber of Electric Energy Commercialization (CCEE), and the Brazilian Energy Research Company (EPE), the NEWAVE (Maceira et al., 2018) and DECOMP (Diniz et al., 2018) models, which use as one of their inputs scenarios of future streamflows generated by the GEVAZP model (Jardim et al., 2001; Maceira & Damázio, 2006). All these models are part of the computational modeling chain of the Center for Electric Energy Research (CEPEL), which develops and enhances computational models that assist in both energetic and electrical planning and operation of the NIS.

The GEVAZP represents the behavior of historical streamflow series through periodic autoregressive models, in which the probabilistic properties of historical values (such as mean, variance, skewness, and correlation structure) exhibit a periodic pattern. This type of model is referred to as PAR(p) (Maceira & Damázio, 2006), where p is the order of the autoregressive model. In general, it involves linear regression of past inflow values with the addition of a random term called noise. There is also an extension of the PAR(p) model that considers the influence of an annual component on the behavior of streamflows, known as the PAR(p)-A model (Treistman et al., 2020a), which is the model officially used for NIS planning studies.

With the increasing occurrence of events possibly associated with climate change, efforts are needed to incorporate such information into models for generating future streamflow series, as these events highly influence streamflows in some regions of the planet. One of the climate phenomena most associated with changes in hydrological regimes in Brazilian basins is the El Niño-Southern Oscillation, which influences precipitation and, therefore, streamflows, thereby affecting the entire planning of NIS.

The objective of this work is to present an alternative to the models used in generating streamflow series and in official studies for the energy operation planning of the NIS. The model, known as MS-PARP (Treistman et al., 2020b), adds a Markovian switching to the traditional PAR(p) model. Once the presence of the El Niño phenomenon in the historical data is determined, it estimates the autoregressive model parameters separately.

This paper is structured into 4 sections. In Section 1, we provide the context in which the work is situated, its relevance to the electric sector, and the objectives of the paper. In Section 2, the El Niño phenomenon is discussed in further details, in addition to the index used in the study to quantify the intensity of the phenomenon, and the associated prediction models. Additionally,

in Section 2, we describe the employed methodology, the MS-PAR(p) model. Section 3 presents the case studies employed to evaluate the models and their results. Finally, Section 4 offers the conclusions of this paper.

MATERIAL AND METHODS

El Niño – Southern Oscillation – Data and predictions

El Niño-Southern Oscillation is a natural climatic phenomenon that plays a significant role in the dynamics of global atmospheric circulation and climate variations, including precipitation (Treistman et al., 2020b). It is characterized by changes in sea surface temperature (SST) and atmospheric pressure in different parts of the Pacific Ocean (Trenberth, 1997). These changes can be classified into three different states: La Niña (LN), Neutral (N), and El Niño (EN), according to the Oceanic Niño Index (ONI) (Huang et al., 2017), which has been available since 1950 and monthly measures sea surface temperature in the region known as NINO 3.4, as ilustred in Figure 1, located in the Pacific Ocean. It takes into account the three-month moving average of anomaly values observed. Thus, we have:

1. If the ONI index remains equal to or less than -0.5 °C for 5 consecutive periods, the period's state is considered a La Niña event;

2. If the ONI index remains equal to or greater than +0.5 °C for 5 consecutive periods, the period's state is considered an El Niño event;

3. If the ONI index falls within the range of -0.5 °C to +0.5 °C, the period's state is considered a Neutral event.

Another index frequently used to detect the ENSO phenomenon is the Southern Oscillation Index (SOI), which is based on the difference in pressure between the regions of Tahiti and Darwin and has been available since 1976. The SOI is calculated as the difference in pressure between these locations, normalized by the standard deviation of this difference. In
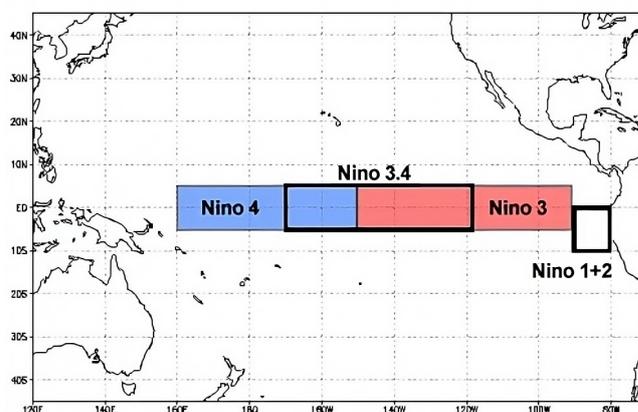
Figure 1. Regions of sea surface temperature (SST) measurement in the Pacific Ocean. Source: National Oceanic and Atmospheric Administration (2018).

this work, the ONI index will be used as an indicator of the ENSO phenomenon, mainly due to the greater availability of data and forecasts.

Since the presented model is autoregressive in nature, meaning it uses past values to represent future behavior, it is necessary to provide the model with input values that have not yet been observed over the forecast horizon. Therefore, access to forecasts of ONI values for the upcoming periods is required to estimate future ENSO states.

The International Research Institute for Climate and Society (IRI) provides forecasts made by various research institutions through its website (Columbia University, 2023).

Examples of the forecasts provided by the IRI can be seen in Figure 2 and Figure 3.

Streamflow data

Historical streamflow data can be found on the ONS website through the official Monthly Operation Planning (PMO) decks (Operador Nacional do Sistema Elétrico, 2023b). These decks contain, as part of their input data for the DECOMP, GEVAZP, and NEWAVE models, streamflows data from approximately 150 HPPs dating back to the beginning of their historical series in 1931.
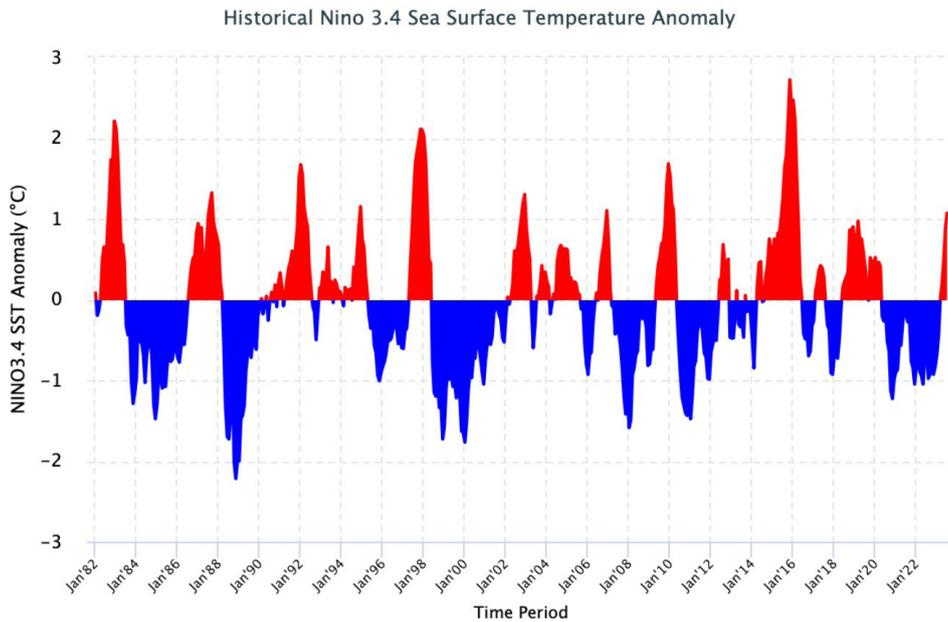

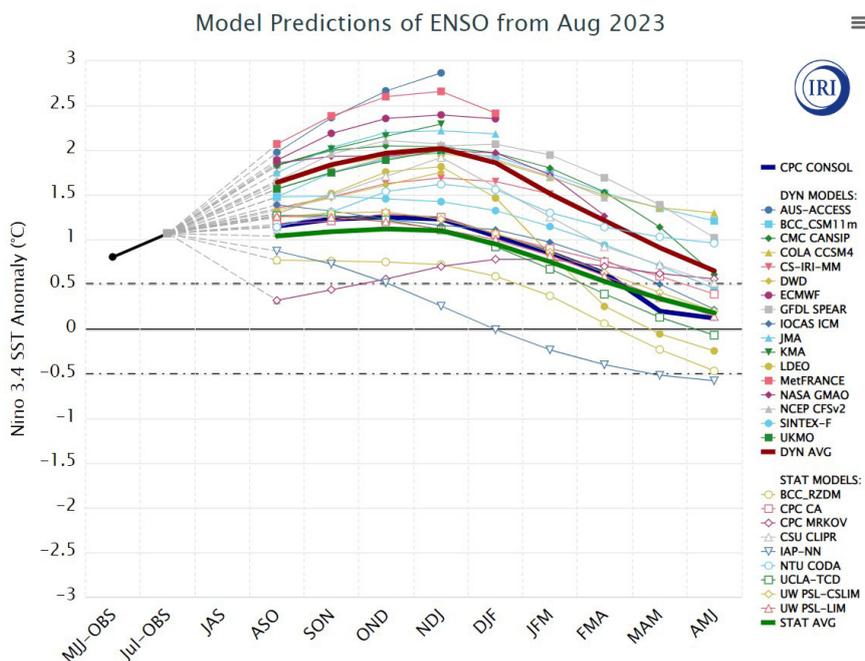**Figure 2.** Historical SST anomaly. Source: Columbia University (2023).

Figure 3. Prediction plume of SST anomaly in NINO 3.4 obtained in September. Source: Columbia University (2023).

Periodic autoregressive models

PAR(p)

The GEVAZP model generates monthly scenarios of future streamflows for the NEWAVE and DECOMP models, being essential to planning the NIS operation. In general, hydrological time series with intervals shorter than a year, such as monthly series, exhibit a periodic behavior in their probabilistic properties, such as mean, variance, skewness, and autocorrelation structure (Penna et al., 2018). The analysis of such series can be carried out using autoregressive formulations whose parameters exhibit a periodic behavior. This class of models is commonly referred to as periodic autoregressive models Maceira & Damázio, 2006). These models are denoted as PAR(p), where p is the order of the model, meaning the number of autoregressive terms in the model. In general, p is a vector, p = (p1, p2, ..., p12), where each element specifies the order for each period. The PAR(p1, p2, ..., p12) model can be mathematically described in Equation 1:

$$\left(\frac{z_t - \mu_m}{\sigma_m}\right) = \phi_1^m \cdot \left(\frac{z_{t-1} - \mu_{m-1}}{\sigma_{m-1}}\right) + \ldots + \phi_{p_m}^m \cdot \left(\frac{z_{t-p_m} - \mu_{m-p_m}}{\sigma_{m-p_m}}\right) + a_t \qquad (1)$$

where: z_t is a seasonal series with a period s, s is the number of periods (s = 12 for monthly series), N is the number of years, t is the time index, t = 1, 2, ..., sN, a function of year T (T = 1, 2, ..., N) and period m (m = 1, 2, ..., s), μ_m is the seasonal mean of period s, is the seasonal standard deviation of period s, $\phi^m(B)$ is the autoregressive operator of order p_m, $\phi^m(B) = (1 - \phi_1^m \cdot B - \phi_2^m \cdot B^2 - \ldots - \phi_{p_m}^m \cdot B^{p_m})$, B^i applied to Z_t results in Z_{t-i} ($B^i Z_t = Z_{t-i}$), p_m is the order of the autoregressive operator for period m, a_t is a series of independent noise with zero mean and variance $\sigma_a^{2(m)}$.

The generated scenarios can be understood as a linear combination of past observations, added to a random term called noise. In order to obtain scenarios of future inflows, one simply needs to draw multiple instances of this noise at with each noise representing a possible future realization of the stochastic process modeled by the model. The selected noises then go through an aggregation step, which aims to reduce the complexity of the optimization problem while maintaining the quality of generation (Jardim et al., 2001).

Before generating scenarios, it is necessary to obtain the linear coefficients that weigh the influence of each monthly inflow value from the past on the inflow occurring at time t. Box and Jenkins (Box & Jenkins, 1970) proposed a highly elaborate methodology for fitting stochastic models from the ARIMA family to time series data, which can be extended to models from the PAR(p) family. The first step, referred to by Box and Jenkins as model identification, involves choosing the order of the model based on estimates of the autocorrelation functions (ACF) and partial autocorrelation functions (PACF) obtained from the sample series (Penna et al., 2018). In periodic autoregressive modeling, this involves selecting the vector p. The selection is made using a confidence interval (e.g., 95%), testing from the highest order to the lowest order to see if the PACFs are significant. The second step is model estimation, which means estimating its parameters,

with maximum likelihood estimators or their approximations typically recommended. For this work, the method of moments was used. The third step involves model checking, that is, verifying through statistical tests whether the assumptions made during the previous steps are met. If the assumptions are not met, one must return to the first step until satisfactory results are achieved.

An important characteristic of the NIS is the complementarity of hydrological regimes in different regions of the country. Therefore, it is crucial that the model used in energy planning can consider and replicate this characteristic. To achieve this, spatial correlations seen in historical inflow values are calculated beforehand and then imposed on the drawn and aggregated noises. As a result, the model can generate multivariate inflow scenarios. The process of transforming initially uncorrelated noises (bt) into noises that preserve the spatial correlation of historical data (Wt) is done as described in Equation 2:

$$W_t = D b_t \qquad (2)$$

where D is a square matrix with dimensions equal to the number of HPPs used in the study.

The matrix D can be estimated as described in Equations 3 and 4:

$$\hat{U} = \mathbf{A} V \mathbf{A}' \qquad (3)$$

$$D = \mathbf{A} \sqrt{V} \mathbf{A}' \qquad (4)$$

where $\hat{U}$ can be the estimate of the covariance matrix of the residuals. The Equation 3 corresponds to the technique known as spectral decomposition, where matrix A is the diagonal matrix of eigenvalues, and V is the symmetric matrix of eigenvectors. Spectral decomposition is obtained using the Jacobi Method (Press et al., 1992).

Periodic autoregressive model with Markov switching - MS-PAR(p)

The model evaluated in this paper is an extension of the PAR(p) model, considering information about the ENSO climatic phenomenon through a Markovian switching model. This model can be understood as a combination of the PAR(p) modeling, in which its parameters are switched based on some states following a Markov chain (Wilks, 2011). The idea is that the autoregressive coefficients of the model, in addition to being estimated according to the period, will also be adjusted to vary according to the states of LN, N, and EN. Therefore, the MS-PAR(p) model follows the following mathematical formulation as defined in Equation 5:

$$\left(\frac{z_t - \mu_m^{\varepsilon_t}}{\sigma_m^{\varepsilon_t}}\right) = \phi_1^m \cdot \left(\frac{z_{t-1} - \mu_{m-1}^{\varepsilon_{t-1}}}{\sigma_{m-1}^{\varepsilon_{t-1}}}\right) + \ldots + \phi_{p_m}^m \cdot \left(\frac{z_{t-p_m} - \mu_{m-p_m}^{\varepsilon_t - p_m}}{\sigma_{m-p_m}^{\varepsilon_t - p_m}}\right) + a_t \qquad (5)$$

where: ε_t is the seasonal time series with a period s, with state space r = 1, 2, ..., d (d = 3 for representing ENSO), which follows a homogeneous Markov chain; $\mu_m^{\varepsilon_t}$ is the mean of period m for state ε_t; $\sigma_m^{\varepsilon_t}$ is the standard deviation of period m for state ε_t.

To calculate the sample mean and sample standard deviation, computed for each month and state, as well as the sample autocorrelation, which varies on a monthly basis only, we consider the Equations 6, 7 and 8 as follows.

$$\mu_m^r = \sum_{t=1}^{N} \mathbf{1}_{(\varepsilon_t = r)} \frac{Z_t}{N_m^s} \tag{6}$$

$$\sigma_m^r = \sqrt{\sum_{t=1}^{N} \mathbf{1}_{(\varepsilon_t = r)} \frac{\left(Z_t - \mu_m^r\right)^2}{N_m^r}} \tag{7}$$

$$\rho_m(k) = \frac{1}{N} \sum_{t=1}^{N} \left(\frac{Z_t - \mu_m^{\varepsilon_t}}{\sigma_m^{\varepsilon_t}} \right) \left(\frac{Z_{t-k} - \mu_{m-k}^{\varepsilon_{t-k}}}{\sigma_{m-k}^{\varepsilon_{t-k}}} \right) \tag{8}$$

where $\mathbf{1}_{(\varepsilon_t = r)}$ denotes the indicator function, assuming a unit value when $\varepsilon_t = r$ and zero for any other value and N_m^r is the number of occurrences of state r in month m.

The process of parameter selection and estimation, as well as the transformation of spatially correlated noises, follows the same steps as the PAR(p) model presented earlier.

It can be observed through Equation 5 that the generation of future inflow scenarios depends on the states of the ENSO events. In other words, to generate inflow scenarios, it is also necessary to generate scenarios for ENSO states, also on a monthly discretization basis.

The time series of ENSO phenomenon states can be understood as a discrete variable ε_t with a state space r related to its three possible states (LN, N e EN). The most used model class to represent the time series of a discrete variable is known as a Markov chain. The behavior of a Markov chain is driven by a set of transition probabilities between its states, P_m, with the simplest form being a first-order Markov chain (or a first-order autoregressive model). In other words, the next state depends only on the most recent state, independent of the sequence of previous states, as presented in Equation 9:

$$P(\varepsilon_t = r \mid \varepsilon_{(t-1)}, \varepsilon_{(t-2)}, \ldots, \varepsilon_1) = P(\varepsilon_t = r \mid \varepsilon_{(t-1)}) \tag{9}$$

The transition probabilities are conditioned probabilities with respect to the most recent state at time t-1. Since the ENSO phenomenon will be divided into its three possible states, the transition matrix P_m will have dimensions of 3x3, as described in Equation 10.

$$P_m = \begin{bmatrix} p_m^{1,1} & p_m^{1,2} & p_m^{1,3} \\ p_m^{2,1} & p_m^{2,2} & p_m^{2,3} \\ p_m^{3,1} & p_m^{3,2} & p_m^{3,3} \end{bmatrix} \tag{10}$$

Since there is no direct transition between EN and LN states, as well as for the conditions, the elements $p_m^{1,3}$ and $p_m^{3,1}$ will always be zero.

For each month "m," the transition matrix will be estimated for ENSO states and for ENSO conditions, both using the following mathematical expression in Equation 11:

$$p_m^{i,j} = P\left(\varepsilon_t = j \, \varepsilon_{\{(t-1)\}} = i \right)_m = \frac{N_m^{i,j}}{N_{\{m-1\}}^i} \tag{11}$$

where $p_m^{\{i,j\}}$ is the probability conditioned to transition to state j in month m, given that the previous month m-1 was in state; $N_m^{i,j}$ is the number of times the transition from state i to state j occurred in month m in the historical data; e $N_{\{m-1\}}^i$ " is the number of occurrences in the historical data of state i in month m-1. The transition matrix P_m can be estimated according to Equation 11 by varying the indices of the states.

Given the transition matrix between ENSO states obtained from historical data, the next step is using this data to produce forecasted transition matrices that adhere to the probability distribution predicted by the IRI. The calculation of monthly forecasted transition matrices between ENSO conditions can be understood as an optimization problem whose goal is to minimize the differences. Δ_i between the historical and forecasted matrices, as described in Equation 12:

$$\min \sum_{i=1}^{14} \Delta_i \tag{12}$$

s.t.,

$$P_c(LN|LN)_m^{prev} = P_c(LN|LN)_m^{hist} + \Delta_1^+ + \Delta_2^-$$

$$P_c(LN|N)_m^{prev} = P_c(LN|N)_m^{hist} + \Delta_3^+ + \Delta_4^-$$

$$P_c(N|LN)_m^{prev} = P_c(N|LN)_m^{hist} + \Delta_5^+ + \Delta_6^-$$

$$P_c(N|N)_m^{prev} = P_c(N|N)_m^{hist} + \Delta_7^+ + \Delta_8^-$$

$$P_c(N|EN)_m^{prev} = P_c(N|EN)_m^{hist} + \Delta_9^+ + \Delta_{10}^-$$

$$P_c(EN|N)_m^{prev} = P_c(EN|N)_m^{hist} + \Delta_{11}^+ + \Delta_{12}^-$$

$$P_c(EN|EN)_m^{prev} = P_c(EN|EN)_m^{hist} + \Delta_{13}^+ + \Delta_{14}^-$$

Constraints on the sum of conditional probabilities:

$$P_c(LN|LN)_m^{prev} + P_c(N|LN)_m^{prev} = 1$$

$$P_c(LN|N)_m^{prev} + P_c(N|N)_m^{prev} + P_c(EN|N)_m^{prev} = 1$$

$$P_c(N|EN)_m^{prev} + P_c(EN|EN)_m^{prev} = 1$$

where $P_c(LN)_m^{prev}$, $P_c(N)_m^{prev}$ and $P_c(EN)_m^{prev}$ are the forecasts provided for each of the conditions (LN, N and EL).

Constraints on the limits of conditional probabilities:

$$0 \le P_c(LN|LN)_m^{prev} \le 1$$

$$0 \le P_c(N|LN)_m^{prev} \le 1$$

$$0 \le P_c(LN|N)_m^{prev} \le 1$$

$$0 \leq P_c(N|N)_m^{prev} \leq 1$$

$$0 \leq P_c(EN|N)_m^{prev} \leq 1$$

$$0 \leq P_c(N|EN)_m^{prev} \leq 1$$

$$0 \leq P_c(EN|EN)_m^{prev} \leq 1$$

Constraints on the reproduction of probabilistic forecasts:

$$P_c(LN)_m^{prev} = P_c(LN)_{m-1}^{prev} \times P_c(LN|LN)_m^{prev} + P_c(N)_{m-1}^{prev} \times P_c(LN|N)_m^{prev}$$

$$P_c(N)_m^{prev} = P_c(LN)_{m-1}^{prev} \times P_c(N|LN)_m^{prev} + P_c(N)_{m-1}^{prev} \times P_c(N|N)_m^{prev} + P_c(EN)_{m-1}^{prev} \times P_c(N|EN)_m^{prev}$$

$$P_c(EN)_m^{prev} = P_c(N)_{m-1}^{prev} \times P_c(EN|N)_m^{prev} + P_c(EN)_{m-1}^{prev} \times P_c(EN|EN)_m^{prev}$$

The optimization problem above has a linear objective function as well as all its constraints, making it a linear programming problem. This problem can be solved using a simplex algorithm, for example.

Once the optimization problem above is solved for all months with ENSO condition forecasts, the forecasted transition matrix between ENSO states can be estimated. The calculation of the ENSO state transition matrix follows the same criteria used for historical classification, as described below:

1. If the previous state is classified as LN:
1.1. The probability of remaining as LN will be equal to the probability of persisting in the LN condition.
1.2. The probability of transitioning to the N state will be equal to the probability of transitioning to the N condition given that it was in an LN state.
2. If the previous state is classified as N:
2.1. The probability of transitioning to an LN state will be equal to the product of transitioning to the LN condition in month m and persisting in the LN condition for four more months.
2.2. The probability of transitioning to an LN state will be equal to the product of transitioning to the LN condition in month m and persisting in the LN condition for four more months.
2.3. The probability of persisting as neutral is given by the complement of the two previous options.
3. If the previous state is classified as EN:
3.1. The probability of remaining as EN will be equal to the probability of persisting in the EN condition.
3.2. The probability of transitioning to the N state will be equal to the probability of transitioning to the N condition given that it was in an EN state.

Mathematically, the above criteria can be summarized by the following expressions in Equations 13 to 19:

$$P(LN|LN)_m^{prev} = P_c(LN|LN)_m^{prev} \tag{13}$$

$$P(N|LN)_m^{prev} = P_c(N|LN)_m^{prev} \tag{14}$$

$$P(LN|N)_m^{prev} = P_c(LN|N)_m^{prev} \times \prod_{i=1}^{4} P_c(LN|LN)_{m+i}^{prev} \tag{15}$$

$$P(EN|N)_m^{prev} = P_c(EN|N)_m^{prev} \times \prod_{i=1}^{4} P_c(EN|EN)_{m+i}^{prev} \tag{16}$$

$$P(N|N)_m^{prev} = 1 - P(EN|N)_m^{prev} - P(LN|N)_m^{prev} \tag{17}$$

$$P(EN|EN)_m^{prev} = P_c(EN|EN)_m^{prev} \tag{18}$$

$$P(N|EN)_m^{prev} = P_c(N|EN)_m^{prev} \tag{19}$$

While forecasts are available, the model will calculate the matrices P_m^{prev}. For the calculation of $P(LN|N)_m^{prev}$ and $P(EN|N)_m^{prev}$, when there are no forecasts for the terms of months m+1, m+2, m+3, and m+4, these will be replaced by historical values. With this implementation, the aim is to make the most of the available probabilistic forecast.

RESULTS AND DISCUSSIONS

Case of study

In this section, we will present the results obtained from the generation of 200 streamflow scenarios for the period from 2011 to 2021 for the HPPs listed in the Table 1 will be presented, which were selected due to its localization in different river basins and regions of Brazil.

Once the parameters of the autoregressive model in Equation 5 are adjusted, each of the scenarios can be interpreted as a possible realization of the modeled stochastic process, with the randomly generated noise being responsible for the multidimensional aspect of the, while preserving the special correlation among the hydroelectric power plants.

At the beginning of the forecast horizon, the model terms related to past flows are obtained from historical data (hydrological trend). As the generation progresses through the periods, the generated flows are used as past values for the subsequent periods, as illustrated in the Figure 4. This type of generation is also known as conditioned generation.

Evaluation metrics

The streamflow scenarios generated by the MS-PAR(p) model are compared with the PAR(p) model using error a metric traditionally employed in assessing the average performance of forecasting models, calculated in relation to the values of streamflows originally observed in the historical data. The mean absolute percentage error (MAPE) is presented in Equation 20:

Table 1. List of Hydroelectric Power Plants used in the experiment.

HPP	Hydrographic Basin	Installed Capacity (MW)
Itá	Uruguai	1.450
Salto Caxias	Iguaçu	1.240
Itaipu	Paraná	14.000
Capivara	Paranapanema	619
Porto Primavera	Paraná	1.540
Furnas	Grande	1.312
Sobradinho	São Francisco	1.050
Santo Antônio Jari	Amazonas	390
Estreito Tocantins	Tocantins	1.087
Teles Pires	Teles Pires	1.820
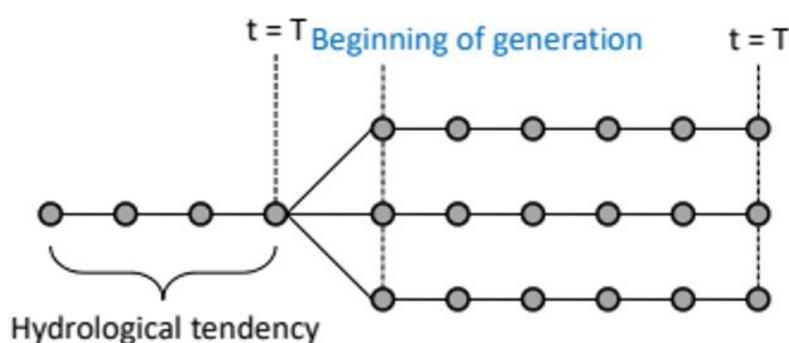
Figure 4. Sythentic scenarios generation conditioned to the recent past. Source: Treistman et al. (2020b).

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Z_t - Cen_t^i}{Z_t}\right| \quad (20)$$

Results

The Figure 5, Figure 6 and Figure 7 illustrates the scenarios generated one step ahead for the Sobradinho, Furnas, and Estreito Tocantins HPPs. For better visualization, only a portion of the generation horizon from 2011 to 2015 was chosen, with similar behavior in the other study years. In the figures, it is also possible to observe the quantiles of 50%, 75%, 90%, 95%, and 99%, where darker shades of blue represent the lower quantiles, and lighter shades represent the higher quantiles.

It can be observed that, in general, the used model is able to capture the periodic behavior of the inflow regime, generating scenarios close to the observed values, although in some cases, there is a deviation of the realized values from the range of generated scenarios.

The Table 2 displays the average performance indices calculated for the PAR(p) and MS-PAR(p) models, obtained in relation to historical values, for the entire horizon. The columns represent the performance of each model in each HPP, followed by a column representing the difference between the MS-PAR(p) model and the PAR(p) model. In other words, negative values in this column indicate an improvement of the used model compared to the original model. It is evident that the MS-PAR(p) model

achieved a superior average performance in the majority (7 out of 10) of the observed HPPs. The PAR(p) model exhibited better performance only in the Furnas and Santo Antonio Jari HPPs, while for the Salto Caxias HPP, both models performed very similarly. In general, better results are observed for the MS-PAR(p) model in the North, South, and Northeast regions (although only one HPP was analyzed in this region), while the PAR(p) model achieved better results in the Southeast region.

Due to the periodic behavior of inflow series, it is also interesting to observe the monthly performance of scenario generation. The average performance of the models for each HPP and for each month is shown in Tables 3 and 4 for the PAR(p) model and in Tables 5 and 6 for the MS-PAR(p) model. The difference between the average monthly performance of MS-PAR(p) and PAR(p) can be found in Tables 7 and 8. Once again, negative values indicate better performance of the evaluated model compared to the original model. For better visualization, negative values in Table 7 and Table 8 are represented in blue. The tables are separated according to dry periods (officially from May to November) and wet periods (officially from December to April), although this classification may vary from region to region.

In general, the performance of the MS-PAR(p) model is superior from May to November, with a lower MAPE index in 62% of the months, whereas in the wet season, the evaluated model was superior in 44% of the cases. It is worth noting the good performance in most of the months for the HPPs in the southern region, specifically Salto Caxias and Itá.
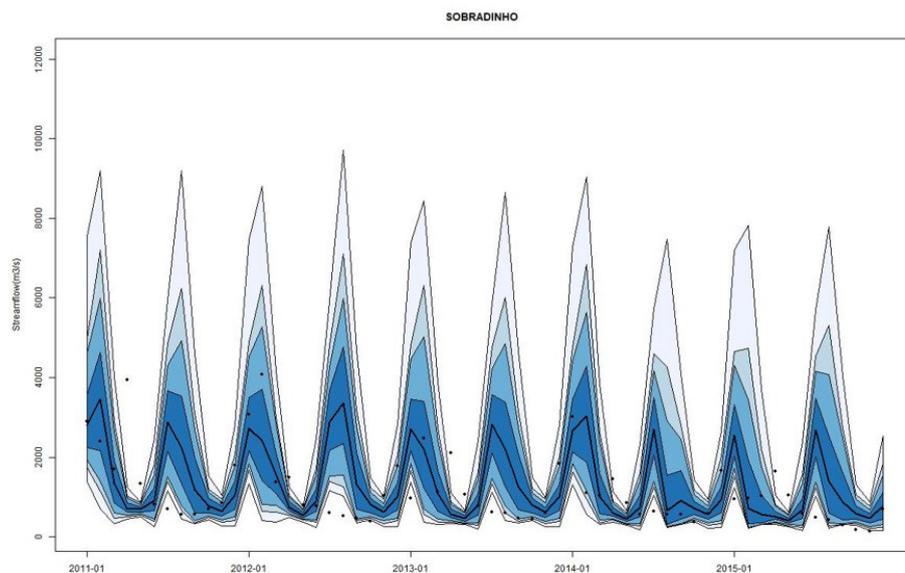
Figure 5. Sythentic MS-PAR(p) – Sobradinho.
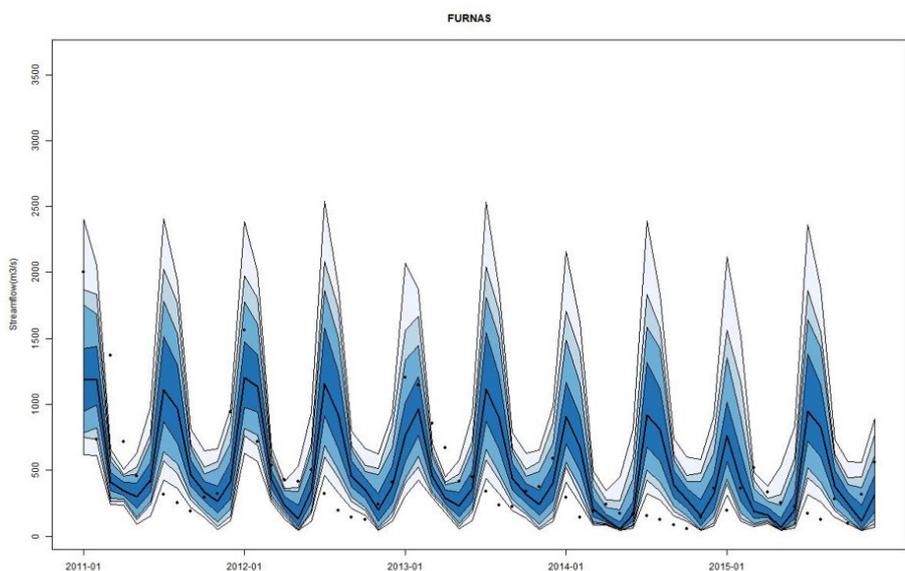
Figure 6. Sythentic MS-PAR(p) – Furnas.

Table 2. Average Performance of MS-PAR(p) and PAR(p) Models.

HPP	REGION	PAR(P)	MS-PARP(P)	Difference
TELES PIRES	North	234.6667	232.4525	-2.21
ESTREITO TOC	North	135.3333	132.5916667	-2.74
STO ANT JARI	North	110.4642	111.9066667	1.45
SOBRADINHO	Northeast	140.9525	136.2683333	-4.68
ITA	South	222.785	221.325	-1.46
SALTO CAXIAS	South	163.1167	155.9833333	-7.13
ITAIPU	South	31.76583	31.39833333	-0.37
FURNAS	Southeast	134.9825	141.5383333	6.56
CAPIVARA	Southeast	46.27417	44.19583333	-2.08
P. PRIMAVERA	Southeast	34.36833	34.395	0.03
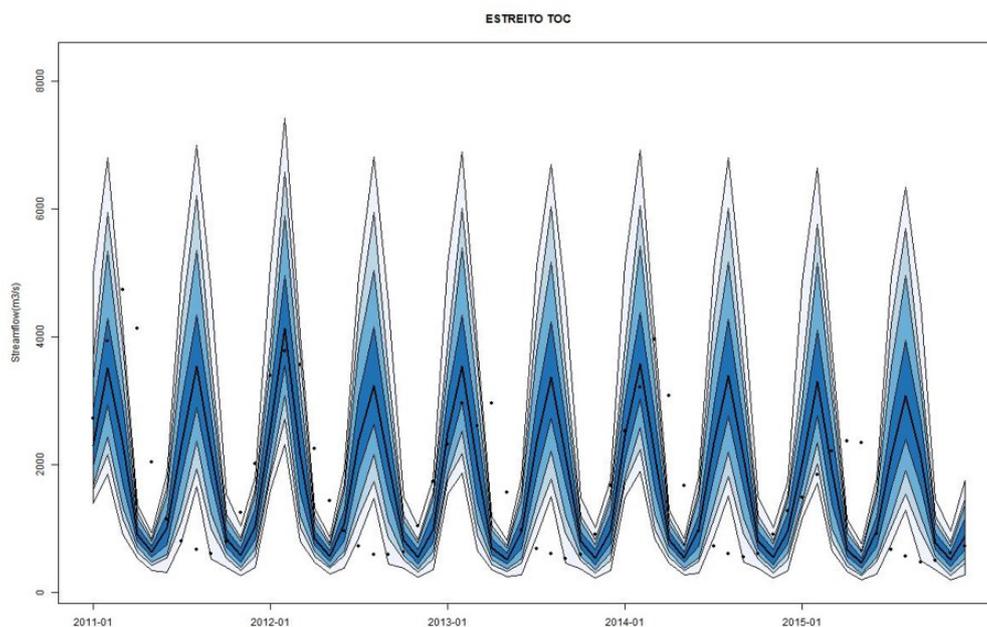
Figure 7. Sythentic MS-PAR(p) – Estreito Tocantis.

Table 3. Average monthly performance of the PAR(p) models during the wet season.

PAR(p)	DEC	JAN	FEB	MAR	APR
TELES PIRES	49.5	22.5	28.1	56.4	83.2
ESTREITO TOC	28.3	29.9	37.0	33.6	70.6
STO ANT JARI	46.3	40.9	38.5	67.3	26.5
SOBRADINHO	34.7	95.4	69.5	36.9	47.0
ITA	297.1	50.0	107.2	178.0	497.0
SALTO CAXIAS	185.1	52.4	78.1	120.0	165.3
ITAIPU	37.9	29.1	40.4	34.4	29.0
FURNAS	27.2	78.0	79.3	44.5	33.4
CAPIVARA	32.9	41.9	40.5	32.7	35.5
P. PRIMAVERA	52.6	26.0	42.1	20.1	29.0

Table 4. Average monthly performance of the PAR(p) models during the dry season.

PAR(p)	MAY	JUNE	JULY	AUG	SEPT	OCT	NOV
TELES PIRES	82.1	19.9	417.2	1119.6	815.9	64.8	56.9
ESTREITO TOC	63.6	15.6	290.8	594.4	375.4	51.0	33.8
STO ANT JARI	75.9	86.9	52.6	45.3	271.4	361.7	212.2
SOBRADINHO	33.1	80.1	447.8	407.3	192.6	170.1	77.0
ITA	352.4	62.6	61.8	112.3	224.8	131.7	598.7
SALTO CAXIAS	201.0	142.8	147.2	350.1	223.3	153.0	139.3
ITAIPU	20.9	33.9	28.1	25.3	34.9	41.0	26.3
FURNAS	38.0	31.7	442.4	467.5	230.6	115.5	31.6
CAPIVARA	36.3	46.0	99.9	61.8	44.2	47.4	36.4
P. PRIMAVERA	30.5	28.3	22.8	65.9	47.6	26.0	21.5

Table 5. Average monthly performance of the MS-PAR(p) models during the wet season.

MS-PAR(p)	DEC	JAN	FEB	MAR	APR
TELES PIRES	51.2	21.2	24.1	57.1	83.3
ESTREITO TOC	24.9	33.6	35.9	34.1	71.3
STO ANT JARI	49.1	43.8	40.5	71.7	26.5
SOBRADINHO	37.3	104.8	59.1	30.5	56.6

Table 5. Continued...

MS-PAR(p)	DEC	JAN	FEB	MAR	APR
ITA	291.7	50.2	82.0	195.1	531.5
SALTO CAXIAS	185.1	51.0	69.4	111.8	152.1
ITAIPU	37.1	29.8	40.3	26.5	21.6
FURNAS	26.9	85.3	82.1	46.7	35.6
CAPIVARA	34.6	28.2	41.7	34.1	37.1
P. PRIMAVERA	51.6	26.2	39.3	20.1	29.2

Table 6. Average monthly performance of the MS-PAR(p) models during the dry season.

MS-PAR(p)	MAY	JUNE	JULY	AUG	SEPT	OCT	NOV
TELES PIRES	82.0	16.9	449.6	1108.2	772.4	66.0	57.7
ESTREITO TOC	66.7	10.9	282.1	575.6	374.8	45.3	36.0
STO ANT JARI	75.9	86.9	51.4	47.1	281.0	356.0	213.1
SOBRADINHO	44.1	42.4	470.6	363.7	237.1	135.2	54.0
ITA	338.6	53.8	60.7	96.1	214.9	155.7	585.7
SALTO CAXIAS	197.4	132.6	139.7	320.8	225.1	150.3	136.5
ITAIPU	19.9	36.9	31.1	29.5	39.7	40.2	24.2
FURNAS	40.7	27.9	472.2	500.4	228.8	122.8	29.0
CAPIVARA	35.5	47.4	94.4	59.9	37.2	45.4	34.9
P. PRIMAVERA	30.5	26.7	23.5	64.6	50.8	29.0	21.3

Table 7. Average monthly performance difference between the MS-PAR(p) and PAR(p) models in the wet season.

Difference	DEC	JAN	FEB	MAR	APR
TELES PIRES	1.68	-1.31	-4.07	0.70	0.06
ESTREITO TOC	-3.48	3.68	-1.07	0.53	0.74
STO ANT JARI	2.78	2.83	2.05	4.38	0.04
SOBRADINHO	2.55	9.38	-10.41	-6.40	9.54
ITA	-5.43	0.19	-25.22	17.13	34.56
SALTO CAXIAS	-0.03	-1.34	-8.67	-8.21	-13.22
ITAIPU	-0.74	0.64	-0.09	-7.94	-7.44
FURNAS	-0.31	7.32	2.77	2.13	2.21
CAPIVARA	1.71	-13.67	1.28	1.41	1.53
P. PRIMAVERA	-1.01	0.15	-2.80	-0.01	0.25

Table 8. Average monthly performance difference between the MS-PAR(p) and PAR(p) models in the dry season.

Difference	MAY	JUNE	JULY	AUG	SEPT	OCT	NOV
TELES PIRES	-0.12	-3.02	32.38	-11.38	-43.50	1.17	0.84
ESTREITO TOC	3.12	-4.65	-8.76	-18.86	-0.62	-5.73	2.20
STO ANT JARI	-0.05	0.01	-1.21	1.76	9.56	-5.73	0.89
SOBRADINHO	11.00	-37.67	22.82	-43.55	44.47	-34.89	-23.05
ITA	-13.83	-8.80	-1.06	-16.18	-9.90	24.06	-13.04
SALTO CAXIAS	-3.55	-10.16	-7.50	-29.24	1.80	-2.64	-2.84
ITAIPU	-1.06	3.02	3.09	4.23	4.81	-0.77	-2.16
FURNAS	2.72	-3.75	29.86	32.92	-1.86	7.32	-2.66
CAPIVARA	-0.78	1.45	-5.45	-1.91	-7.04	-1.95	-1.52
P. PRIMAVERA	0.04	-1.66	0.66	-1.29	3.21	2.94	-0.16

It's important to highlight that in a potential official use of the MS-PAR(p) model, periodic assessments could be conducted to evaluate the recent performance of the models for each of the hydroelectric power plants included in the power system planning studies. This assessment could inform the decision of whether to consider the ENSO phenomenon on an individual basis.

CONCLUSIONS

The Brazilian Electrical System relies heavily on hydroelectric sources, which currently account for approximately 70% of the electricity generated in the country. This dependence also means that the inflows that generate this energy represent the greatest uncertainty in the planning of the NIS. To address this uncertainty,

the energy optimization models NEWAVE and DECOMP use inflow scenarios generated by the GEVAZP model, which employs the methodology of periodic autoregressive models PAR(p). This study aimed to present an additional study using a methodology originally proposed by Treistman et al. (2020b), in which climatic information related to the El Niño-Southern Oscillation (ENSO) phenomenon is added to the original model. The El Niño-Southern Oscillation is a natural climatic phenomenon that plays a significant role in the dynamics of global atmospheric circulation and climate variations, including precipitation and, consequently, inflows.

In the paper, the MS-PAR(p) methodology was described, which uses parameters that switch according to states following a Markov chain and are estimated segmentally for each of these states. In other words, in the original PAR(p) model, the statistical parameters mean and standard deviation vary by time period. In the MS-PAR(p) model, the mentioned parameters also vary with the state assigned to the period. The states in the MS-PAR(p) model correspond to the states of the ENSO phenomenon (La Niña - LN, Neutral - N, and El Niño - EN).

This paper also explained how to include ENSO forecasts, provided by the ONI index and released by the IRI, to generate ENSO scenarios by constructing predicted transition matrices that are similar to historical transition matrices using linear programming techniques.

For model evaluation, this work proposed an experiment involving the generation of inflow scenarios for 10 HPPs located in different river basins and regions of Brazil from 2011 to 2021. The performance metric used was the MAPE (Mean Absolute Percentage Error) of the generated scenarios compared to historical inflows, comparing the results obtained from the MS-PAR(p) and PAR(p) models.

The proposed experiment demonstrated that the MS-PAR(p) model was capable of generating inflow scenarios that were closer to observed values in most of the analyzed cases. Overall, the described model outperformed the original model in 70% of the HPPs when considering the average performance over the entire horizon. The paper also presented the average monthly performance of each model, with the MS-PAR(p) model outperforming the PAR(p) model in the majority (62%) of months during the Brazilian dry season, which spans from May to November. However, during the wet season, from December to April, the PAR(p) model achieved better overall performance, with a lower MAPE index in 56% of the months.

Although the MS-PAR(p) model showed slightly lower average performance in some months, the experiment demonstrated that the model can appropriately capture and incorporate the influence of ENSO, resulting in inflow scenarios closer to observed values. Therefore, the model is an interesting tool for official use in the models employed by the Brazilian Electrical Sector in the planning of the NIS. To this end, it is recommended to conduct periodic studies to assess the influence of the El Niño phenomenon and the model's performance for each of the HPPs. The MS-PAR(p) model can be used individually for HPPs where it exhibits better performance.

REFERENCES

Box, G. E., & Jenkins, G. M. (1970). *Time series analysis: forecasting and control.* San Francisco: Holden-Day.

Câmara de Comercialização de Energia Elétrica – CCEE. (2023, 10 de setembro). Geração de energia renovável bateu recorde em 2022, aponta CCEE. *CCEE*, São Paulo. Retrieved in 2023, September 10, from https://www.ccee.org.br/pt/web/guest/-/geracao-de-energia-renovavel-bateu-recorde-em-2022-aponta-ccee

Columbia University. International Research Institute for Climate and Society. (2023). *ENSO forecast. Retrieved in 2023, October 6*, from http://iri.columbia.edu/our-expertise/climate/forecasts/enso/current

Diniz, A., Santos, T., Cabral, R., Santos, L., Maceira, M. E., & Costa, F. (2018). Short/mid-term hydrothermal dispatch and spot pricing for largescale systems: the case of Brazil. In *Proceedings of the 20th Power Systems Computation Conference (PSCC)*, Dublin, Ireland.

Huang, B., Thorne, P. W., Banzon, V. F., Boyer, T., Chepurin, G., Lawrimore, J. H., Menne, M. J., Smith, T. M., Vose, R. S., & Zhang, H. M. (2017). Extended reconstructed sea surface temperature version 5 (ERSSTv5): upgrades, validations, and intercomparisons. *Journal of Climate*, *30*(20), 8179-8205. http://dx.doi.org/10.1175/JCLI-D-16-0836.1.

Jardim, D., Maceira, M., & Falcao, D. (2001). Stochastic streamflow model for hydroelectric systems using clustering techniques. In *Power Tech Proceedings* (Vol. 3), Porto. New York: IEEE. http://dx.doi.org/10.1109/PTC.2001.964916.

Maceira, M. E. P., & Damázio, J. M. (2006). Use of PAR(p) model in the stochastic dual dynamic programming optimization scheme used in the operation plannin of the Brazilian hydropower system. *Probability in the Engineering and Informational Sciences*, *20*(1), 143-156. http://dx.doi.org/10.1017/S0269964806060098.

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). 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 *Proceedings of the 2018 Power Systems Computation Conference (PSCC)* (pp. 1-7). New York: IEEE.

National Oceanic and Atmospheric Administration – NOAA, 2018. *Equatorial Pacific Sea Surface Temperatures (SST). Retrieved in 2023, October 9*, from https://www.ncei.noaa.gov/access/monitoring/enso/sst

Operador Nacional do Sistema Elétrico – ONS. (2023a, 1 de setembro). *Evolução da capacidade instalada no SIN - outubro 2023/dezembro 2027*. Retrieved in 2023, September 1, from https://www.ons.org.br/paginas/sobre-o-SIN/o-sistema-em-numeros

Operador Nacional do Sistema Elétrico – ONS. (2023b, 1 de setembro). *Programa Mensal da Operação (PMO)*. Retrieved in 2023, September 1, from https://www.ons.org.br/paginas/energia-no-futuro/programacao-da-operacao

Penna, D. D. J., Maceira, M. E. P., Damázio, J. M., Treistman, F., & Araujo, H. S. (2018). *Modelo de geração de séries sintéticas de energias e vazões: manual de referência*. Rio de Janeiro: CEPEL.

Press, W. H., Flannery, B. P., Teukolosjy, S. A., & Vetterling, W. T. (1992). *Numerical recipes in FORTRAN 77: the art of scientific computing* (2nd ed.). Cambridge: University of Cambridge.

Treistman, F., Maceira, M. E. P., Damázio, J. M., & Cruz, C. B. (2020a). Periodic time series model with annual component applied to operation planning of hydrothermal systems. In *Proceedings of the 2020 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)* (pp. 1-6). New York: IEEE.

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.

Trenberth, K. E. (1997). The Definition of El Niño. *Bulletin of the American Meteorological Society*, *78*(12), 2771-2778. http://dx.doi.org/10.1175/1520-0477(1997)078<2771:TDOENO>2.0.CO;2.

Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences* (Vol. 100). New York: Academic Press.

Authors contributions

José Francisco Moreira Pessanha: Methodology discussions and assistant coordinator of the activities.

Victor Andrade de Almeida: Methodology discussions, coordinator of the activities, generation and consolidation of the results and text writing and revision.

Priscilla Dafne Shu Chan: Methodology discussions, data organization and generation of the results.

Editor-in-Chief: Adilson Pinheiro

Associated Editor: Adilson Pinheiro