RBRH

Revista Brasileira de Recursos Hídricos
Brazilian Journal of Water Resources
Versão On-line ISSN 2318-0331
RBRH, Porto Alegre, v. 28, e46, 2023
Special Session

https://doi.org/10.1590/2318-0331.282320230117

Generation of synthetic flow scenarios by means of multivariate sampling of contemporaneous ARMA model outputs

Geração de séries sintéticas de vazão a partir de amostragem multivariada de saídas de modelo ARMA contemporâneo

Daniel Henrique Marco Detzel[1], Marcelo Rodrigues Bessa[1], Leandro Ávila[2], Mauricio Pereira Cantão[1] & Klaus de Geus[1]

[1]Universidade Federal do Paraná, Curitiba, PR, Brasil
[2]Institut für Bio- und Geowissenschaften, Forschungszentrum Jülich GmbH, Jülich, Germany
E-mails: detzel@ufpr.br (DHMD), marcelo.bessa@ufpr.br (MRB), leandroavilarangel@gmail.com (LÁ), mpcantao@gmail.com (MPC), klaus@degeus.com.br (KG)

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

ABSTRACT

This work constitutes one of the outcomes of the "Evaluation of hydrological scenario generation models" activity initiated by the Hydrological Scenario Representation Working Group (GT CH) and coordinated by ONS and CCEE. We introduce the LYNX-Series model, a contemporaneous non-periodic and multivariate variation of the autoregressive moving average model (CARMA) for generating synthetic time series of average inflow discharges to reservoirs in the Brazilian National Interconnected System (SIN). Notably, LYNX-Series couples the synthetic series generator with a multivariate sampling process to select a group of synthetic hydrological scenarios based on a similarity criterion with recent historical data. In addition to reducing the computational burden of the hydrothermal dispatch optimization process, the solution aims to enhance the representativeness of synthetic hydrological scenarios. The paper expounds on the theoretical aspects of the model and presents numerical simulations that validate its ability to replicate hydrological behaviors in various Brazilian basins.

Keywords: Synthetic flow series; Stochastic models; Multivariate sampling; National Interconnected System.

RESUMO

Este trabalho é um dos resultados da atividade "Avaliação de modelos de geração de cenários de afluências", proposta pelo Grupo de Trabalho de Representação de Cenários Hidrológicos (GT CH) e coordenada pelo ONS e CCEE. Apresenta-se o modelo LYNX-Series, uma formulação contemporânea autorregressiva de médias móveis (CARMA) não-periódica e multivariada para a geração de séries sintéticas de vazões médias afluentes aos reservatórios do Sistema Interligado Nacional (SIN). Como destaque, o LYNX-Series acopla o gerador de séries sintéticas a um processo de amostragem multivariada para a seleção de um grupo de cenários hidrológicos sintéticos com base em um critério de similaridade em relação ao passado recente. Além de reduzir o peso computacional do processo de otimização do despacho hidrotérmico, a solução tem por objetivo melhorar a representatividade dos cenários hidrológicos sintéticos. O artigo mostra os aspectos teóricos do modelo, bem como simulações numéricas que comprovam a sua capacidade na reprodução dos comportamentos hidrológicos das diversas bacias brasileiras.

Palavras-chave: Séries sintéticas de vazão; Modelos estocásticos; Amostragem multivariada; Sistema Interligado Nacional.

INTRODUCTION

The generation of synthetic flow series is a technique traditionally employed in hydrology to overcome limitations arising from analyses solely reliant on historical records (Medda & Bhar, 2019). Furthermore, observed time series are insufficient for analyses related to uncertainty and risk (Jardim et al., 2001), making the generation of synthetic series a particularly attractive solution for these purposes.

Since the early techniques proposed by Thomas & Fiering (1962) and Matalas (1967), numerous methods have been suggested. Among them, linear stochastic models of the Box & Jenkins type (Box et al., 2008) are among the most commonly employed for generating synthetic flow series. Such methods seek to adequately reproduce the persistence structure of time series, which justifies their success in modeling hydrological data. Hipel & McLeod (1977), Hipel et al. (1977), Stedinger et al. (1985), and Haltiner & Salas (1988) are examples of studies that brought significant advancements in synthetic series modeling using Box & Jenkins formulations at the time. Since then, the successful application of these methods has ensured their widespread use in hydrological studies over the years, including recent works (Singh & Ray, 2021; Tukiman & Harun, 2021; Medda & Bhar, 2019; Bayesteh & Azari, 2019; Pereira & Veiga, 2018).

An example of applying Box & Jenkins models for the generation of synthetic scenarios can be found in the Brazilian National Interconnected System (SIN). It is known that SIN is a large-scale hydro-thermal-wind system, with a predominant share of hydropower plants (approximately 53% of installed capacity – Agência Nacional de Energia Elétrica, 2023). The planning of SIN's operation and expansion is currently based on synthetic time series of natural energy inflows generated by a periodic autoregressive model (PAR-p, Maceira & Damázio, 2006). PAR-p is employed to obtain synthetic scenarios of natural energy inflows to equivalent reservoirs (Larroyd et al., 2017), a technique adopted to reduce the computational burden of hydrothermal dispatch optimization models (Jardim et al., 2001). More recently, Treistman et al. (2020) proposed PAR(p)-A, an enhancement to the original formulation that includes a component giving more weight to observations from the last 12 months in generating synthetic scenarios. This is the official version used in the planning and operation models of SIN at present.

This paper addresses the activity proposed by the Hydrological Scenario Representation Working Group (GT CH) of the PMO/PLD Technical Committee, coordinated by the National Operator of the Electrical System (ONS) and the Chamber of Electric Energy Commercialization (CCEE), titled "Evaluation of hydrological scenario generation models." Here, we present the LYNX-Series model, a formulation based on the contemporaneous autoregressive moving average (CARMA) model. The model was initially developed as part of Line 5 of the ANEEL Strategic Call 001/2008, where the Research and Development project PHOENIX – Optimization of Hydrothermal Dispatch through Hybrid Algorithms with High-Performance Computing – was conceived. The initial version of the model was further enhanced within the LYNX project – Large-Scale Optimization Applied to the Brazilian Hydrothermal Dispatch: Hierarchical Models for Medium and Short-Term Operation and Planning with Energy and Power Integration, earning the name LYNX-Series.

In contrast to the solution based on equivalent scenarios of natural energy inflow, the proposed approach applies the CARMA model to generate synthetic time series of mean natural inflow discharges for all operational power plants within the SIN. The CARMA model represents a non-periodic and multivariate version of the Box & Jenkins family of stochastic models. It is a parsimonious formulation, which allows a reduction in the number of parameters to be estimated compared to a conventional multivariate ARMA model. The choice of this formulation was made due to the large scale of the Brazilian hydroelectric system, as previously mentioned.

In addition to the development of LYNX-Series for generating 3000 synthetic time series, each with a duration of 60 months, for all power plants (HPPs) in the SIN, the modeling presented highlights the following aspects: (i) application of the Interior Point method for estimating CARMA model parameters through maximum likelihood estimation, considering non-linearity and constraints in the optimization process; (ii) implementation of a clustering-based sampling methodology using Mahalanobis distance as a similarity criterion; and (iii) identification and selection of synthetic scenarios with hydrological characteristics similar to those prevalent in recent years, aimed at improving the energy optimization process.

This paper presents the theoretical aspects of the LYNX-Series model, including a detailed explanation of the proposed sampling process. It also showcases results from numerical experiments, demonstrating the model's capabilities in replicating the hydrological characteristics of the analyzed time series. A final conclusion section ends the article.

LYNX-SERIES MODEL: THEORETICAL BACKGROUND

The LYNX-Series model is based on the CARMA formulation, which stands for Contemporaneous Autoregressive Moving Average model (Camacho et al., 1985). It is a multivariate approach derived from the linear stochastic models of the Box & Jenkins type, or ARIMA (Box et al., 2008). In general, this family of models operates with the autocorrelation structure of time series data, which justifies its significant applicability to flow series. The contemporaneous component of the model, on the other hand, is responsible for considering spatial correlation among different locations.

Synthetic series generation model

Let the vectors of l normally distributed series be denoted as $z_t = \left[z_{t,1}, z_{t,2}, \ldots, z_{t,l} \right]'$, each corresponding to a location u ($u = 1, 2, \ldots, l$). Also, let the vectors of l normally distributed residuals be denoted as $\varepsilon_t = \left[\varepsilon_{t,1}, \varepsilon_{t,2}, \ldots, \varepsilon_{t,l} \right]'$, independent in time but spatially correlated. The CARMA model is defined by Equation 1:

$$\phi_u(B) z_{t,u} = \theta_u(B) \varepsilon_{t,u}, \quad t = 1, 2, \ldots, n \tag{1}$$

Where $\phi_u(B)$ is the u-th autoregressive operator of order $p = \max(p_1, p_2, \ldots, p_l)$:

$$\phi_u(B) = 1 - \phi_{u1}B - \phi_{u2}B^2 - \ldots - \phi_{up}B^p$$

In turn, $\theta_u(B)$ is the u-th moving average operator of order $q = \max(q_1, q_2, \ldots, q_l)$:

$$\theta_u(B) = 1 - \theta_{u1}B - \theta_{u2}B^2 - \ldots - \theta_{uq}B^q$$

Finally, B is the lag operator, such that $z_{t-k} = B^k z_t$, for any lag k. The form of the CARMA model is identical to that of the multivariate ARMA model, except for the parameter matrices $\phi_u(B)$ e $\theta_u(B)$. In the CARMA formulation, both matrices are considered diagonal and take the form:

$$\phi_u = \begin{bmatrix} \phi_{11} & 0 & \ldots & 0 \\ 0 & \phi_{22} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \phi_{ll} \end{bmatrix} \qquad \theta_u = \begin{bmatrix} \theta_{11} & 0 & \ldots & 0 \\ 0 & \theta_{22} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \theta_{ll} \end{bmatrix}$$

In practical terms, considering diagonal matrices allows the model to replicate the individual statistics of each location and the instantaneous spatial correlation (lag zero) between them. Furthermore, the model is parsimonious as it does not require the estimation of an excessively large number of parameters. Haltiner & Salas (1988) and Stedinger et al. (1985) demonstrated the efficiency of this formulation when compared to the conventional multivariate ARMA model. Additionally, the CARMA model has the advantage of enabling its parameter matrices to be constructed with univariate estimates from each location (Hipel & McLeod, 1994, p. 784).

The process of fitting the CARMA model to flow series follows the methodological procedure of identification-estimation-validation originally proposed by Box et al. (2008). In the identification phase, the Bayesian Information Criterion (BIC) (Schwartz, 1978) is applied, defined as:

$$BIC = -2\ln L(r) + \ln nr \tag{2}$$

Where $L(r)$ is the likelihood function (defined below), r is the number of model parameters, and n is the size of the series. The criterion is applied to five variations of the model, CARMA(1,0), CARMA(2,0), CARMA(1,1), CARMA(2,1), and CARMA(2,2), with the chosen formulation being the one that minimizes Equation 2. Models of maximum order two were selected because they are sufficient for the satisfactory fitting of linear formulations to stationary series (Box et al., 2008).

The choice to use the BIC criterion over the traditional analysis of autocorrelation functions (ACF) and partial autocorrelation functions (PACF) is primarily due to the scale of the hydroelectric system under analysis. Working with a system that includes approximately 200 HPPs makes it impractical to visually analyze the ACF and PACF individually for each series, especially considering that the model needs to be re-estimated with each update to the historical data. Additionally, the BIC is an information criterion that penalizes models with an excessive number of parameters, prioritizing the parsimony of the formulation.

The estimation of parameters $\hat{\phi}$ e $\hat{\theta}$ is carried out using the maximum likelihood method, also individually for each location, by minimizing the log-likelihood function given by Equation 3:

$$\ln(r) = \ln L\left(\hat{\phi}, \hat{\theta}, \ \hat{\sigma}_a \mid z_t\right) = -n \cdot \ln \hat{\sigma}_a - \frac{SSQ\left(\hat{\phi}, \hat{\theta}\right)}{2\hat{\sigma}_a} \tag{3}$$

Where $\hat{\sigma}_a$ is the estimate of the standard deviation of the residual series $\boldsymbol{a}$ $(t = 1, 2, \ldots, n)$ and $SSQ\left(\hat{\phi}, \hat{\theta}\right)$ represents the sum of squares of residuals defined by Equation 4:

$$SSQ\left(\hat{\phi}, \hat{\theta}\right) = \sum_{t=1}^{n} \boldsymbol{a}_t^2 \tag{4}$$

It's important to emphasize that the minimization process, in addition to being non-linear, has constraints arising from the stationarity and invertibility properties of the model (Box et al., 2008). To address this, the Interior Point optimization method (Byrd et al., 2000) is used, with sample estimates serving as initial values for $\hat{\phi}$ e $\hat{\theta}$.

The spatially correlated field is estimated from the Equation 5:

$$\varepsilon_{t,u} = \Lambda a_t \tag{5}$$

Where $\boldsymbol{a}_t = \left[a_{t,1}, a_{t,2}, \ldots, a_{t,l}\right]'$ is the vector of l normally distributed and independent residuals (in both time and space), and Λ is a parameter matrix with dimensions $l \times l$, whose estimation is obtained from the solution of Equation 6:

$$\Lambda\Lambda' = \gamma_\varepsilon \tag{6}$$

Where γ_ε represents the variance-covariance matrix of the residuals $\varepsilon_{t,u}$. The procedure used is to apply the Cholesky decomposition to γ_ε to obtain the lower triangular matrix Λ. At this point, it's important to note that this solution only works for positive-definite γ_ε matrices. Therefore, for application to the flow series of HPPs in SIN, it is necessary to exclude duplicated series, as occurs in some cases (e.g., Camargos and Itutinga HPPs, Paulo Afonso-Moxotó Complex, etc.). The reason for this is that the presence of perfect correlations in the γ_ε matrix makes it not positive-definite, which renders the estimation procedure using the adopted method unfeasible.

As the final step in the CARMA model fitting process, there is theoretical validation. A suitable model is one that produces residuals that are independent, homoscedastic, and approximately normally distributed. These three checks are performed through statistical inferences of the Portmanteau type (Li & McLeod, 1981), Levene (Brown & Forsythe, 1974), and Shapiro-Wilk (Shapiro & Wilk, 1965), respectively. In all cases, the tests are applied at a significance level of 5%.

Sampling procedure

The choice of 3000 scenarios per HPP was established after a specific study aimed precisely at determining the appropriate quantity of synthetic series for energy-related applications (Detzel & Mine, 2017). However, this is an excessively large number to consider in the hydrothermal dispatch optimization process for which the LYNX-Series model provides the synthetic series. Therefore, the implementation of the sampling process was motivated to reduce the dimensions of the problem.

As a relevant issue, it should be noted that the synthetic series generation model is multivariate, and therefore, sampling should be performed in a way that preserves the entire spatial correlation structure among the power plants. The solution was to apply a sampling method based on clustering of synthetic series through the calculation of Mahalanobis distances between scenarios and historical data. In other words, the proposed method seeks the closest similarities between the generated synthetic series and the observed historical series. The Mahalanobis distance is defined according to Equation 7 (Maesschalck et al., 2000):

$$d(z,x) = \sqrt{(\bar{z}-\bar{x})\gamma_{zx}^{-1}(\bar{z}-\bar{x})} \tag{7}$$

Where $\bar{z}$ e $\bar{x}$ are the mean vectors of historical and synthetic series, respectively, and γ_{zx}^{-1} is the inverse of the joint variance-covariance matrix of historical and synthetic flows. Once the 3000 distances are obtained, 200 series are selected through stratified sampling based on ten classes. Implementation details of the entire sampling process can be found in the initial publication of the method by Detzel et al. (2013).

However, in the development of LYNX-Series, the sampling of synthetic scenarios has been revised to include a second objective. Now, in addition to reducing the dimensionality of the problem, the sampling aims to represent the recent state of inflows to the power plants in SIN. In other words, the process seeks to identify the prevailing hydrological characteristics in the recent past. This information is incorporated into the synthetic flow series to be sampled, aiming to bring greater refinement to the process.

To achieve this, the entire Mahalanobis distance-based sampling and stratification process was retained. The difference lies in the historical data used to calculate the distances. In this work, the decision was made to use the most recent 12 months of available historical data. Therefore, the length of the generated synthetic series was increased from n' to $n'+12$ months, with the first twelve months used only in the sampling process. Afterward, the first twelve months of each of the 200 scenarios are disregarded, leaving the remaining n' months as the synthetic output series from the model. Figure 1 graphically illustrates the procedure, assuming the goal is to obtain synthetic series with a final length of 60 months.

It's important to note that the model allows for the choice of any historical period for calculating distances and determining similarities. Additionally, it's also possible to choose samples with periods different from the proposed 12 months. These alternatives were tested during the model implementation phase, and for the series considered, no significant differences were detected. Therefore, the use of the last 12 months was established as a premise.

Considerations regarding seasonality, normality, and stationarity

Similar to the Box & Jenkins model family, CARMA is a formulation that requires normally distributed series. This condition is not met in many Brazilian rivers, so the choice is made to apply a logarithmic transformation to the series before subjecting them to the stochastic model itself. The logarithmic transformation is known to approximate the distribution of hydrological data to a normal shape (Helsel et al., 2020, p. 129).

Another numerical transformation applied aims to address the seasonality of the series. As mentioned in the introduction, CARMA is a non-periodic model and, therefore, does not have parameters specifically designed for seasonality. This choice was made during the initial conception of the method, aiming to maintain the parsimony of the hydrothermal dispatch model as a whole. Therefore, it was chosen to work with deseasonalized series, which are obtained through the standardization defined by the Equation 8:

$$z_{t,m} = \frac{q_{t,m} - \bar{q}_{t,m}}{\sigma_{t,m}}, \quad m = 1,2,\ldots,12 \tag{8}$$

Where $q_{t,m}$ represents the flow series with mean $\bar{q}_{t,m}$ and standard deviation $\sigma_{t,m}$, all calculated for a specific month m. Standardization is carried out individually for each location.

Finally, the LYNX-Series model was designed to deal with the non-stationarity of inflow series. Previous studies have shown evidence that non-stationarity affects a significant portion of the inflow to Brazilian reservoirs (Melchior, 2022; Silva & Detzel, 2021; Silva et al., 2019; Chagas & Chaffe, 2018; Detzel et al., 2011). In a more recent study, Detzel et al. (2023) showed that 48% of the flow series to hydroelectric power plants in the National Interconnected System exhibited some form of trend. Of these, 31% were identified as decreasing trends, and 69% as increasing

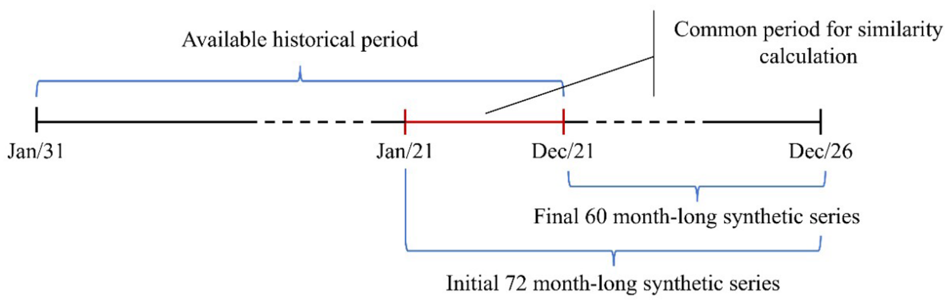

Figure 1. Representation of the sampling process. For this example, it is assumed that there is historical data available from Jan/31 to Dec/21, and the goal is to obtain synthetic series with a length of 60 months.

trends. Since CARMA is a stationary model, the occurrence of trends in historical records is undesirable.

As a solution, LYNX-Series implements the Mann-Kendall test (Hamed, 2009) for trends and the Pettitt test (Pettitt, 1979) to identify breakpoints in the series. Once trends are detected, historical data correction is carried out based on the angular coefficients of the cumulative flow series before and after the identified breakpoint (Detzel et al., 2011). This solution, based on historical data correction, is an alternative to using the ARIMA model, which is a non-stationary formulation of linear stochastic models. ARIMA models work by differencing the historical series into stationary segments, which are then subjected to the modeling process. However, the synthetic series obtained are equivalent to the differenced historical data, and returning to the original scale would require an individual integration constant for each series, which is not known. For this reason, models from this family are not commonly used for generating synthetic series (Salas et al., 1980, p. 279).

That being said, it should be emphasized that for the purposes of the activity proposed by the Hydrological Scenario Representation Working Group (GT-CH) within the Technical Committee of PMO/PLD (CT PMO/PLD), the module for identifying and correcting non-stationarity has been disabled because it alters historical data. If it were kept active, the model would lose a common basis for comparing results as shown in Treistman et al. (2023).

NUMERICAL EXPERIMENTS

Prior to submitting the synthetic flow scenarios for the activity proposed by the GT-CH, numerical experiments were conducted to validate the synthetic series generated by the LYNX-Series model. For this purpose, the static configuration of the Brazilian hydroelectric system as of January 2022 was selected, with a flow history ranging from January 1931 to December 2021. The data pertains to naturalized flows, excluding the effects of damming, non-consumptive uses, and reservoir evaporation (Braga et al., 2009). The data were collected from the SINtegre portal (https://sintegre.ons.org.br), which is maintained by ONS.

The flow series for the 146 plants in the four subsystems that make up the SIN, namely: Southeast/Central-West, South, Northeast, and North, were modeled. For all of them, initially, 3000 series were generated, from which 200 synthetic scenarios with a duration of 60 months were sampled. It is important to note that intensive investigations into the performance of the CARMA model have been conducted previously (Detzel et al., 2014, 2016). Furthermore, the last article in this special edition presents comparisons with other models using specific statistical metrics. Therefore, this section presents only general results of the model.

RESULTS AND DISCUSSIONS

Figure 2 displays the distribution of the selected model orders for the hydrological series studied. The BIC criterion indicated the ARMA(2,1) model for the majority of the series (67%). The simplest model, AR(1), was selected for 16% of the series, while AR(2) and ARMA(1,1) models were indicated for 12% and 5% of the series, respectively. The ARMA(2,2) model was not selected in any case. In a more in-depth analysis, it was found that purely autoregressive models were indicated for the series of HPPs in the South subsystem. Models with the moving average portion were selected for the series of HPPs in the other subsystems.

Figure 3 displays the HPPs considered in the study, distinguished by the type of reservoir (regulation or run-of-river). Examples of synthetic series generated for six projects are also shown: Samuel (Jamari River), Sobradinho (São Francisco River), Ilha Solteira (Paraná River), Barra Bonita (Tietê River), Gov. Bento Munhoz (or Foz do Areia, Iguaçu River), and Barra Grande (Pelotas River).

Through Figure 3, it is initially possible to observe the different hydrological regimes present in the six selected basins. The rivers in the Southern subsystem (Figures 3e and 3f) do not have a defined seasonality. However, as latitude decreases, the seasonal patterns of

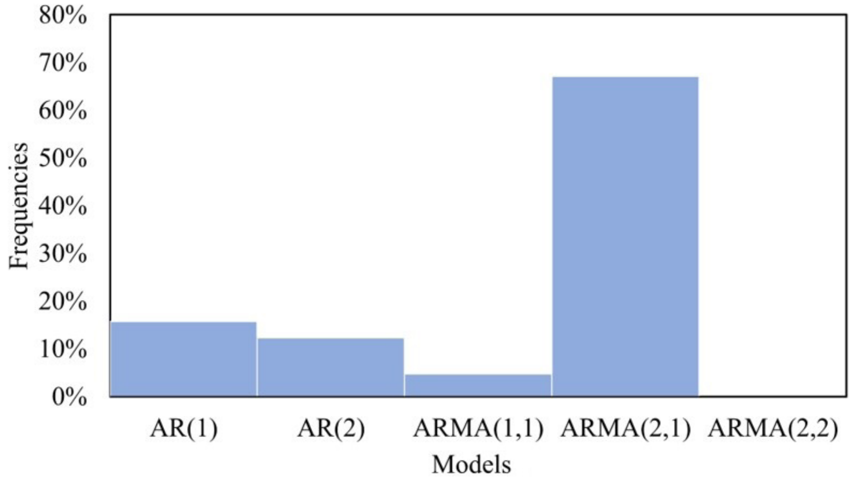

Figure 2. Frequency histogram for the orders of the selected models for the hydrological series in the study.

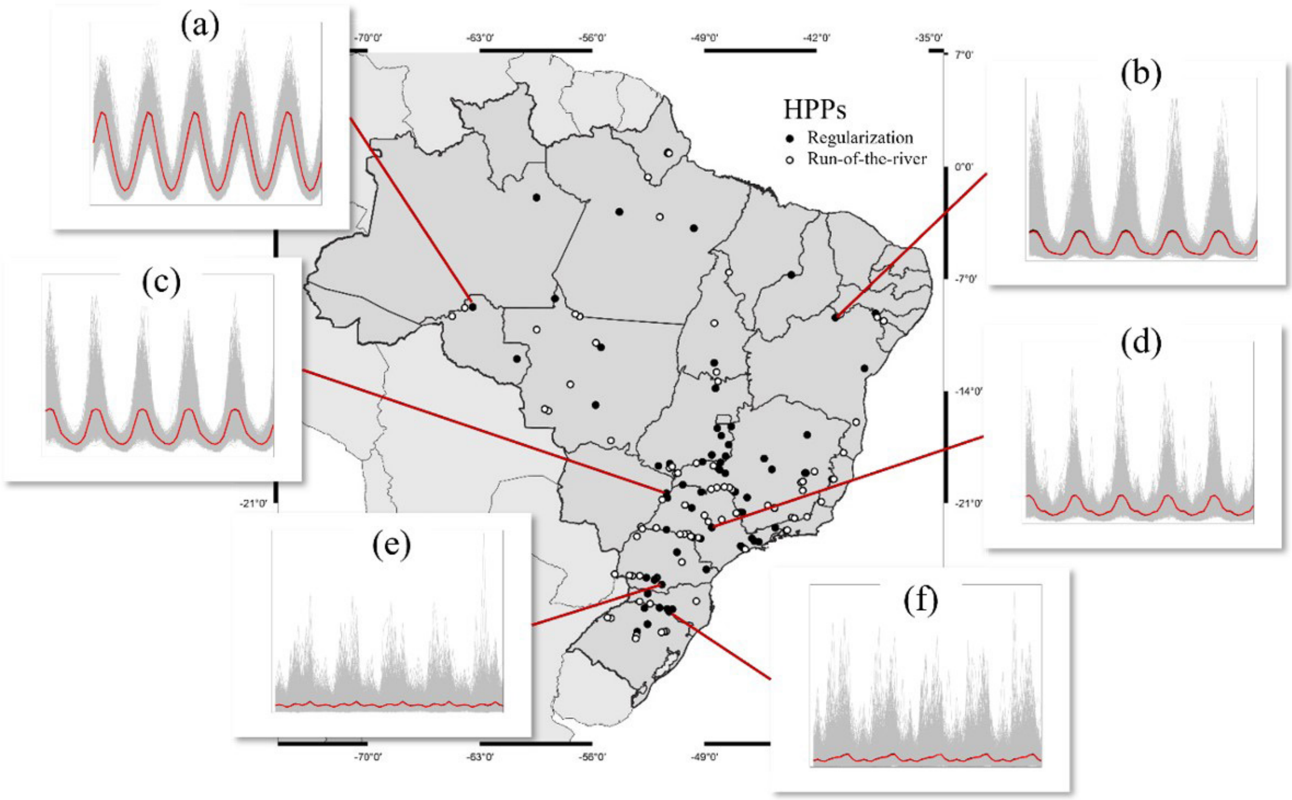

Figure 3. HPPs considered in the study, classified according to the type of reservoir. The synthetic flow series shown refer to the plants (a) Samuel, (b) Sobradinho, (c) Ilha Solteira, (d) Barra Bonita, (e) Gov. Bento Munhoz (Foz do Areia), (f) Barra Grande. The red lines display historical monthly average flows, the black lines represent synthetic monthly average flows, and the gray lines show the synthetic series.

the series become more evident. For these plants, the synthetic series are able to reproduce the seasonal behavior very well, as evidenced by the overlap of the lines of historical and synthetic monthly average flows. The only graph in which a distinction between the lines is identified is in Figure 3b, in the first few months.

Another interesting characteristic is the behavior of synthetic series around the average flows. Particularly in the Samuel HPP (Figure 3a), the synthetic series vary little and relatively uniformly for both high-flow and low-flow periods. On the other hand, the synthetic series for the Sobradinho HPP (Figure 3b) and Ilha Solteira HPP (Figure 3c) suggest greater uncertainty during high-flow periods, which is reflected in the wide range of scenarios generated during those times. In the case of the Gov. Bento Munhoz HPP (Figure 3e) and Barra Grande HPP (Figure 3f), the synthetic series also suggest high uncertainty around the mean values but without characterizing a specific period. This is a consequence of the greater annual hydrological regime regularity, as mentioned in the previous paragraph.

Figure 4 displays the comparisons between autocorrelation functions for the same power plants shown in Figure 3. These functions are used to assess the model's ability to reproduce the persistence structure of historical series in each case. What can be observed is that the CARMA model was able to adequately reproduce the autocorrelations of the first lags. In particular, for the series of the Samuel HPP (Figure 4a), Sobradinho HPP (Figure 4b), Ilha Solteira HPP (Figure 4c), and Barra Bonita HPP (Figure 4d), the autocorrelation is relatively well reproduced up to a lag of 24 months. On the other hand, for the series of the Gov. Bento Munhoz HPP (Figure 4e) and Barra Grande HPP (Figure 4f), the autocorrelation is satisfactorily reproduced only for the first few months.

These results are explained by the orders of the selected models in each case. As mentioned earlier, more complex models were indicated for power plants that operate in the Southeast/Central-West, Northeast, and North subsystems, and therefore, the persistence structure is well reproduced for longer lags. In the case of power plants in the South, purely autoregressive and lower-order models explain the limited reproduction of autocorrelations.

The last numerical results to be discussed pertain to the spatial correlation structure among the hydrological series modeled. The CARMA model was particularly successful in reproducing this characteristic, as evidenced by Figure 5. It is interesting to note that there is a group of power plants whose series have a negative correlation with the others. These power plants are located in the Southern subsystem and exhibit a hydrological behavior that complements the other subsystems. This behavior was also adequately reproduced by the model.

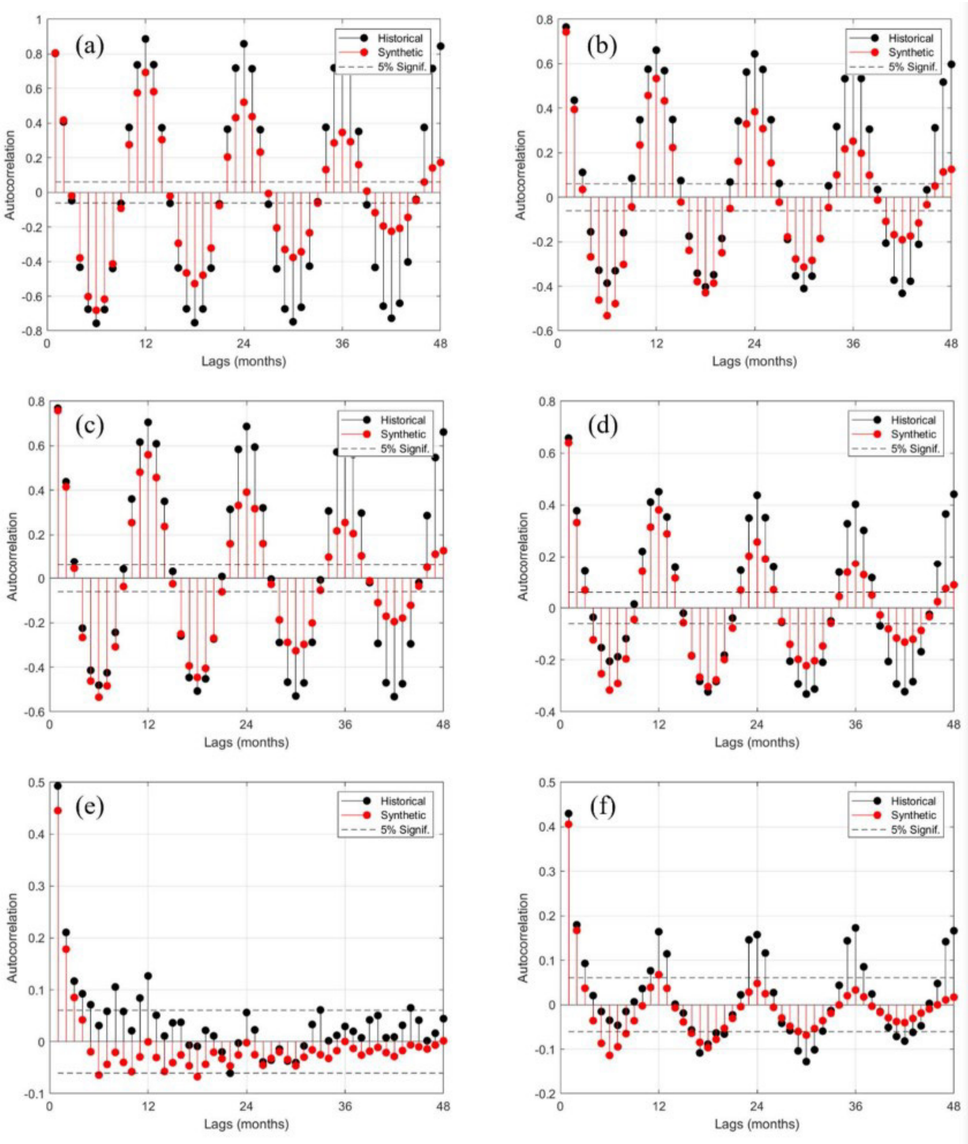

Figure 4. Comparison of autocorrelation functions for the HPPs (a) Samuel, (b) Sobradinho, (c) Ilha Solteira, (d) Barra Bonita, (e) Gov. Bento Munhoz (Foz do Areia), (f) Barra Grande. The red lines display historical monthly mean flows, the black lines show synthetic monthly mean flows, and the gray lines represent synthetic series.

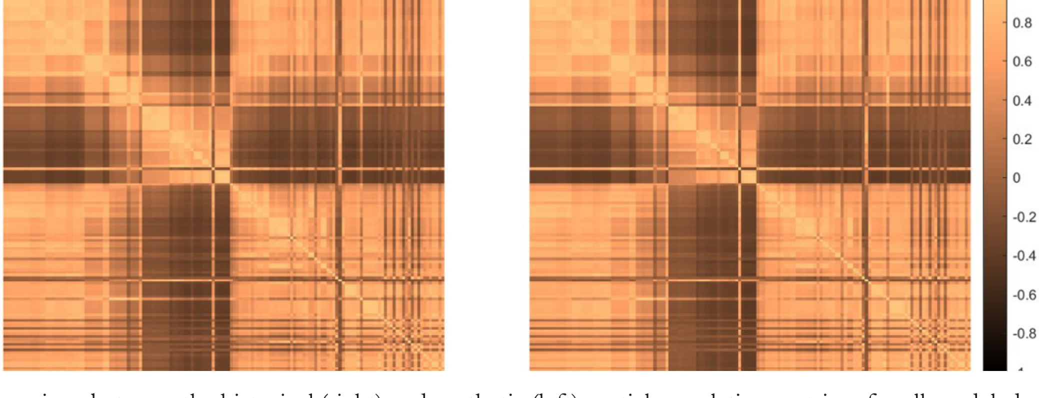

Figure 5. Comparison between the historical (right) and synthetic (left) spatial correlation matrices for all modeled power plants. The diagrams were constructed following the encoding sequence of the plants used by ONS.

CONCLUSION

This article aimed to provide detailed insight into the formulation of the LYNX-Series model, which was originally developed for applications related to the optimization of the Brazilian hydrothermal dispatch. The model's primary feature is the coupling between synthetic scenario generation and the method for sampling series that exhibit similarities with recent historical data. As demonstrated, the model is capable of adequately reproducing the hydrological behavior of rivers located in different regions of Brazil. In particular, LYNX-Series exhibits high skill in reproducing the spatial correlation structure among series, a crucial characteristic in the context of planning and operating a large-scale hydroelectric system like the Brazilian one.

An important note regarding the results presented here and their relation to the GT-CH's activity proposal: LYNX-Series was originally designed for generating natural inflow series to reservoirs. However, the activity's proposal was to assess the scenarios in terms of incremental flows between power plants. Therefore, an adaptation to the modeling was necessary to calculate and provide these synthetic incremental scenarios for comparison with other models. Consequently, the results to be presented in Treistman et al. (2023) pertain to incremental flows, and their interpretation should consider this characteristic.

REFERENCES

Agência Nacional de Energia Elétrica – ANEEL. (2023). *Sistema de informações de geração da ANEEL*. Retrieved in 2023, September 18, from https://tinyurl.com/2zp36k74

Bayesteh, M., & Azari, A. (2019). Comparison of the performance of stochastic models in the generation of synthetic monthly flows data: a case study on Marun river. *Journal of Applied Research in Water and Wastewater*, *6*(2), 117-125.

Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2008). *Time series analysis forecasting and control* (4th ed.). New Jersey: John Wiley & Sons.

Braga, R. S., Rocha, V. F., & Gontijo, E. A. (2009). Revisão das séries de vazões naturais nas principais bacias hidrográficas do Sistema Interligado Nacional. In: *Anais do XVIII Simpósio Brasileiro de Recursos Hídricos* (pp. 1-20). Porto Alegre: ABRHidro.

Brown, M. B., & Forsythe, A. B. (1974). Robust tests for the equality of variances. *Journal of the American Statistical Association*, *69*(346), 364-367. http://dx.doi.org/10.1080/01621459.1974.10482955.

Byrd, R. H., Gilbert, J. C., & Nocedal, J. (2000). A trust region method based on interior point techniques for nonlinear programming. *Mathematical Programming*, *89*(1), 149-185. http://dx.doi.org/10.1007/PL00011391.

Camacho, F., McLeod, A. I., & Hipel, K. W. (1985). Contemporaneous Autoregressive-Moving Average (CARMA) modeling in water resources. *Journal of the American Water Resources Association*, *21*(4), 709-720. http://dx.doi.org/10.1111/j.1752-1688.1985.tb05384.x.

Chagas, V. B. P., & Chaffe, P. L. B. (2018). The role of land cover in the propagation of rainfall into streamflow trends. *Water Resources Research*, *54*(9), 5986-6004. http://dx.doi.org/10.1029/2018WR022947.

Detzel, D. H. M., & Mine, M. R. M. (2017). Comparison between deseasonalized models for monthly streamflow generation in a Hurst-Kolmogorov process framework. *Journal of Hydrologic Engineering*, *22*(4), 05016040-05016042. http://dx.doi.org/10.1061/(ASCE)HE.1943-5584.0001488.

Detzel, D. H. M., Bessa, M. R., & Mine, M. R. M. (2013). Amostragem de séries sintéticas hidrológicas. *RBRH*, *18*(4), 173-184. http://dx.doi.org/10.21168/rbrh.v18n4.p173-184.

Detzel, D. H. M., Bessa, M. R., Vallejos, C. A. V., Santos, A. B., Thomsen, L. S., Mine, M. R. M., Bloot, M. L., & Estrócio, J. P. (2011). Estacionariedade das afluências às usinas hidrelétricas brasileiras. *RBRH*, *16*(3), 95-111. http://dx.doi.org/10.21168/rbrh.v16n3.p95-111.

Detzel, D. H. M., Medeiros, L., Oening, A. P., Marcilio, D. C., & Toshioka, F. (2016). Acerca da quantidade de simulações estocásticas de vazão no contexto do planejamento energético. *Revista Brasileira de Energia*, *22*, 21-32.

Detzel, D. H. M., Mine, M. R. M., Bessa, M. R., & Bloot, M. (2014). Cenários sintéticos de vazões para grandes sistemas hídricos através de modelos contemporâneos e amostragem. *Revista Brasileira de Recursos Hídricos*, *18*, 173-184. http://dx.doi.org/10.21168/rbrh.v18n4.p173-184.

Detzel, D. H. M., Monteiro, V. G., Bessa, M., Cantão, M. P., & Alberti, E. L. (2023). Dez anos de análises na estacionariedade das vazões afluentes às usinas do SIN: o que mudou? *Revista Brasileira de Energia*, *28*(4), 159-167. http://dx.doi.org/10.47168/rbe.v28i4.

Haltiner, J. P., & Salas, J. D. (1988). Development and testing of a multivariate, seasonal ARMA(1,1) model. *Journal of Hydrology*, *104*(1-4), 247-272. http://dx.doi.org/10.1016/0022-1694(88)90168-0.

Hamed, K. H. (2009). Enhancing the effectiveness of prewhitening in trend analysis pf hydrologic data. *Journal of Hydrology*, *368*(1-4), 143-155. http://dx.doi.org/10.1016/j.jhydrol.2009.01.040.

Helsel, D. R., Hirsch, R. M., Ryberg, K. R., Archfield, S. A., & Gilroy, E. J. (2020). Statistical methods in water resources. In: United States Geological Survey, editor. *Hydrologic analysis and interpretation* (pp. 1-458). Reston: USGS. .

Hipel, K. W., & McLeod, A. I. (1977). *Applied Box-Jenkins modeling for water resources engineers*. New York: Elsevier.

Hipel, K. W., & McLeod, A. I. (1994). *Time series modelling of water resources and environmental systems* (1012 p.). Amsterdam: Elsevier.

Hipel, K. W., McLeod, A. I., & McBean, E. A. (1977). Stochastic modelling of the effects of reservoir operation. *Journal of Hydrology*, *32*(1-2), 97-113. http://dx.doi.org/10.1016/0022-1694(77)90121-4.

Jardim, D. L. D. D., Maceira, M. E. P., & Falcão, D. M. (2001). Stochastic streamflow model for hydroelectric systems using clustering techniques. In: Saraiva JT, Matos MA, editors. *2001 IEEE Porto Power Tech Proceedings (Cat. No.01EX502)* (Vol. 3, p. 421-430). New York: IEEE. http://dx.doi.org/10.1109/PTC.2001.964916.

Larroyd, P. V., Matos, V. L., & Finardi, E. C. (2017). Assessment of risk-averse policies for the long-term hydrothermal scheduling problem. *Energy Systems*, *8*(1), 103-125. http://dx.doi.org/10.1007/s12667-016-0191-y.

Li, W. K., & McLeod, A. I. (1981). Distribution of the residual autocorrelations in multivariate ARMA time series models. *Journal of the Royal Statistical Society. Series B. Methodological*, *43*(2), 231-239. http://dx.doi.org/10.1111/j.2517-6161.1981.tb01175.x.

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 planning of the Brazilian hydropower system. *Probability in the Engineering and Informational Sciences*, *20*(1), 143-156. http://dx.doi.org/10.1017/S0269964806060098.

Maesschalck, R., Jouan-Rimbaud, D., & Massart, D. L. (2000). The Mahalanobis distance. *Chemometrics and Intelligent Laboratory Systems*, *50*(1), 1-18. http://dx.doi.org/10.1016/S0169-7439(99)00047-7.

Matalas, N. C. (1967). Mathematical assessment of synthetic hydrology. *Water Resources Research*, *3*(4), 937-945. http://dx.doi.org/10.1029/WR003i004p00937.

Medda, S., & Bhar, K. K. (2019). Comparison of single-site and multi-site stochastic models for streamflow generation. *Applied Water Science*, *9*(3), 67. http://dx.doi.org/10.1007/s13201-019-0947-3.

Melchior, J. (2022). *Vazões de projeto sob condições não estacionárias: uma aplicação para usinas hidrelétricas no Brasil* (Master thesis). Retrieved in 2023, September 12, from https://hdl.handle.net/1884/80554

Pereira, G., & Veiga, Á. (2018). PAR(p)-vine copula based model for stochastic streamflow scenario generation. *Stochastic Environmental Research and Risk Assessment*, *32*(3), 833-842. http://dx.doi.org/10.1007/s00477-017-1411-2.

Pettitt, A. N. (1979). A non-parametric approach to the change-point problem. *Applied Statistics*, *28*(2), 126-135. http://dx.doi.org/10.2307/2346729.

Salas, J. D., Delleur, J. W., Yevjevich, V., & Lane, W. L. (1980). *Applied modeling of hydrologic time series* (484 p). Littleton: Water Resources Publications.

Schwartz, G. (1978). Estimating the dimension of a model. *Annals of Mathematical Statistics*, *6*(2), 461-464. Retrieved in 2023, October 14, from https://www.jstor.org/stable/2958889

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52*(3-4), 591-611. http://dx.doi.org/10.1093/biomet/52.3-4.591.

Silva, N. R., & Detzel, D. H. M. (2021). Modificações do teste de Mann-Kendall: a importância das premissas de testes estatísticos. In: *Anais do XXIV Simpósio Brasileiro de Recursos Hídricos* (pp. 1-8). Belo Horizonte: ABRHidro.

Silva, W. L., Xavier, L. N. R., Maceira, M. E. P., & Rotunno, O. C. (2019). Climatological and hydrological patterns and verified trends in precipitation and streamflow in the basins of Brazilian hydroelectric plants. *Theoretical and Applied Climatology*, *137*(1-2), 353-371. http://dx.doi.org/10.1007/s00704-018-2600-8.

Singh, H., & Ray, M. R. (2021). Synthetic stream flow generation of River Gomti using ARIMA model. In: Gupta LM, Ray MR, Labhasetwar PK, editors. *Advances in civil engineering and infrastructural development* (Lecture Notes in Civil Engineering, No. 87, pp. 255-263). Singapore: Springer. http://dx.doi.org/10.1007/978-981-15-6463-5_24.

Stedinger, J. R., Lettenmaier, D. P., & Vogel, R. M. (1985). Multisite ARMA(1,1) and disaggregation models for annual streamflow generation. *Water Resources Research*, *21*(4), 497-509. http://dx.doi.org/10.1029/WR021i004p00497.

Thomas, H. A., & Fiering, M. B. (1962). Mathematical synthesis of streamflow sequences for the analysis of river basins by simulation. In: Maass A, editor. *Design of water resources systems* (pp. 459-493). Cambridge: Harvard University Press. http://dx.doi.org/10.4159/harvard.9780674421042.c15.

Treistman, F., Penna, D., Khenayfis, L., Cavalcante, N., Souza Filho, F., Rocha, R., Estácio, A., Rolim, L., Pontes Filho, J., Porto, V., Guimarães, S., Pessanha, J. F., Almeida, V., Chan, P. D., Lima, C., Detzel, D., & Bessa, M. (2023). A framework to evaluate and compare synthetic streamflow scenario generation models. *Revista Brasileira de Recursos Hídricos*, *28*, e43. http://dx.doi.org/10.1590/2318-0331.282320230115.

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

Tukiman, N. N. A., & Harun, S. (2021). For evaluation the performances of stochastic streamflow models for the multi reservoirs: synthetic streamflow generation for reservoir management. *International Journal of Integrated Engineering*, *13*(1), 303-310.

Authors contributions

Daniel Henrique Marco Detzel: Model design and implementation, interpretation of the results, paper writing and review.

Marcelo Rodrigues Bessa: Model design, interpretation of the results, paper writing and review.

Leandro Ávila: Model implementation, interpretation of the results, paper review.

Mauricio Pereira Cantão: Interpretation of the results, paper review.

Klaus de Geus: Model design, paper review.

Editor-in-Chief: Adilson Pinheiro

Associated Editor: Adilson Pinheiro