Renewable Energy Sources Spatio-Temporal Scenarios Simulation under Influence of Climatic Phenomena

Gustavo Melo[1], Tuany Barcellos[1], Rafaela Ribeiro[1], Rafael Couto[1], Bruno Gusmão[1]
Fernando Luiz Cyrino Oliveira[1], Paula Maçaira[1], Bruno Fanzeres[1], Reinaldo Castro Souza[1], Olavo Bet[2]
[1]Industrial Engineering Department, Pontifical Catholic University of Rio de Janeiro, PUC-Rio, Rio de Janeiro, Brazil
[2]China Three Gorges Corporation, CTG Brazil, São Paulo, Brazil
[1]gustavo.melo.rio@gmail.com

***Abstract*—This work proposes a methodology for modeling and simulating stochastic processes of Variable Renewable Energy (VRE) sources production. The main objective is to capture, in the scenarios, both the influence of climate phenomena and the complementarity between different sources based on the definition of joint states of VREs. Case studies were developed based on data from three renewable energies covering the Brazilian territory. Regarding the evaluation criteria, the results were satisfactory, and the proposed methodology performed better than the benchmark model applied in the Brazilian market.**

***Index Terms*—Renewable Energy; Simulation; Complementarity; Climatic Phenomena; ENSO.**

I. Introduction

Increasing the share of renewable energy sources in countries' electrical matrices is essential for energy decentralization, decarbonization and achieving several Sustainable Development Goals (SDGs) established by the United Nations [1]. However, according to Morales et al. [2], it brings crucial challenges to the power systems' planning and operation, mainly due to the intermittency of renewable production and the fact that they are not usually dispatchable. Pinson [3] highlights that one of the main challenges involves capturing the intermittent productions stochasticity, aiming to model and simulate possible scenarios of the stochastic processes of generation by renewable sources, constituting a powerful tool to support the decision-making process, both in the public and private sectors. In recent years, much research has been directed toward modeling, predicting or simulating Variable Renewable Energy sources (VREs). In this environment, it is highlighted that exploring the complementary effects of different renewable resources is a significant growing topic in the sector, given the possibility of obtaining a combined power with less variability and intermittency [4].

Also noteworthy is the relevance of studies that assess the influence of climate phenomena, such as the El Niño–Southern Oscillation (ENSO) and its different phases (El Niño, La Niña, and Neutral), on the behavior of primary resources and generation in different locations (see [5], [6]). The impacts of the ENSO phases on the VREs are not homogeneous across all countries and regions. In Brazil, for example, during El Niño, the South-Southeast regions experience an increase in precipitation and the North-Northeast regions become dry, while the opposite occurs in La Niña [7]. During dry periods, clear sky conditions prevail, resulting in increased availability of solar radiation for photovoltaic (PV) production, and vice versa. The phenomenon's influence on wind behavior varies significantly within the same region. However, in the Brazilian Northeast, for example, higher average wind speeds are generally expected during El Niño compared to other phases [8]. Furthermore, regarding the ENSO, since the 1960s, the Pacific temperature indices used to detect the occurrence of extreme phases (El Niño and La Niña) of the phenomenon have shown more extreme oscillations [5], mainly due to climate change, intensifying the impacts of ENSO on natural resources during the same period. This underscores the need to consider them in the VREs modeling.

Thus, capturing and representing both the complementary properties between different renewable resources and the influence of climatic phenomena, such as ENSO, in generating probabilistic scenarios are imperative. In this context, one of the possible classifications for the different existing methods to simulate or predict renewable resources is to divide them into parametric and non-parametric methods [9]. Most time series models, for example, are classified as parametric because they make statistical assumptions, such as the normality of the residuals [10]. However, one of the possible disadvantages is that, in most real cases concerning renewable resources, noise cannot be treated in this way, as it presents intrinsically

This work was partially supported by Rio Paraná Energia S.A. through RD project ANEEL PD-10381-0322/2022, CAPES Finance Code 001, CNPq (309064/2021-0, 422470/2021-0, 307084/2022-1, 311519/2022-9, 402971/2023-0) and FAPERJ (202.825/2019, 210.618/2019, 211.086/2019, 211.645/2021, 201.231/2022, 201.243/2022, 201.348/2022, 210.041/2023, 210.015/2024).

asymmetric tail behaviors [10].

Consequently, it is possible to perceive a high volume of recent publications in this area of research involving non-parametric methods in addition to hybrid or combined techniques. Machine learning models stand out among the non-parametric methods, especially variations of neural network techniques (see [11]–[13]). However, it is essential to underline that although they present advantages in representing the randomness of VREs concerning other methods, machine learning models tend to require great computational effort and are usually difficult to interpret and control.

Under these circumstances, many works model the stochastic processes of renewable resources using Markov Chains (see [10], [14]–[17]), which are easier to interpret than machine learning methods. Based on Markov state modeling, we identify several opportunities for scientific contribution and propose a novel stochastic simulation model of renewable sources in this work.

In short, the proposal can be summarized in two main stages: (i) renewable modeling and (ii) scenario simulation. In (i) renewable modeling, resource states are calculated to represent the renewable sources. From the defined states, transition probability matrices are calculated based on historical data, making it possible to obtain different generation scenarios through simulation techniques (ii). The main innovations of this work can be summarized as follows, and more details are provided in Section II: (i) in the modeling stage, the calculation of states is proposed to consider three aspects at the same time: (a) the different complementary effects (spatial and temporal) between resources, with states representing multiple renewable sources simultaneously; (b) the influence of the phases of the ENSO on primary energy sources. For this, the states of renewables are specifically calculated for each phase of the climatic phenomenon; (c) the seasonality of the data, so the states are specifically calculated for each period of the data; (ii) in the simulation stage, the scenarios are obtained from three main steps: (a) the first one draws the ENSO phase; (b) the second one draws the cluster of renewable resources based on the previously defined ENSO phase; (c) the third one draws primary resource values belonging to the cluster defined in the second stage.

To validate the proposed methodology, scenarios were generated for Brazilian data on hydro, wind, and solar resources. The synthetic series obtained were evaluated using statistical techniques to verify the generated scenarios' adherence to the historical series characteristics. Furthermore, the results of the proposed methodology were compared to the scenarios obtained by benchmark models applied in the context of the electricity sector.

The paper is structured as follows. Section II presents the proposed framework to simulate renewable scenarios and briefly explains the benchmark models. Then, in Section III, an application is performed to simulate scenarios for different regions of Brazil. The results are evaluated and compared to synthetic series obtained by benchmark models in the same section. Finally, Section IV brings conclusions and final remarks.

II. Renewable scenarios simulation framework

Fig. 1 summarizes the methodology and its two main stages, Renewable Modeling and Scenario Simulation, each composed of three steps, detailed in II-A and II-B.

A. Renewable Modeling

The first step of the renewable modeling stage consists of obtaining, analyzing and preprocessing data. Depending on the interest and data availability, its time series can comprise primary resource data (e.g., wind speed and solar radiation) or power generation. After defining the region of interest and the renewable sources to be modeled, it is necessary to identify and treat missing data and outliers, besides evaluating the application of data transformation methods, such as standardization or normalization. This step also involves data descriptive analysis and applying statistical tests, like the ADF [18] and Fisher tests [19], to assess, respectively, data stationarity and seasonality. It is also important to assess how ENSO influences VREs by applying statistical tests that compare data distributions in the different climatic phenomenon phases, such as the Kolmogorov-Smirnov test [20].

Furthermore, during the preprocessing step, the data must be divided into classes, named $c_{l,e,t}$. Each $c_{l,e,t}$ comprises the set, from a location l, of renewable resource data in ENSO phase e at time t. Considering a monthly database, each class is represented by $c_{l,e,t}$, where t is, in this case, a monthly index. To compose all the classes, divide the original data among all possible combinations of "ENSO phase - month". So, as a more specific example, $c_{l,neutral,January}$ from a local l are the observations of renewable resources (as many as considered) in the January months when ENSO is neutral.

The second step is discretizing the data and grouping them into finite states. In this context, each state represents the coincident availability of different types of resources at a particular time. Thus, a sequence of states comprises the discrete time series. In this step, clustering algorithms, such as the k-means [21], can be used due to their simplicity and accuracy [15]. Clustering algorithms aim to classify objects into several clusters so that objects within the same cluster are as similar as possible (high intraclass similarity) [21]. In contrast, objects from different clusters are as diverse as possible (low interclass similarity) [21].

So, the multivariate data (where each energy source constitutes a variable) within each class ($c_{l,e,t}$) is discretized into different clusters (states). Since each $c_{l,e,t}$ consists of different resources, the obtained states aim to capture the complementarity effects between the VREs. Then, each calculated cluster will have a number of dimensions

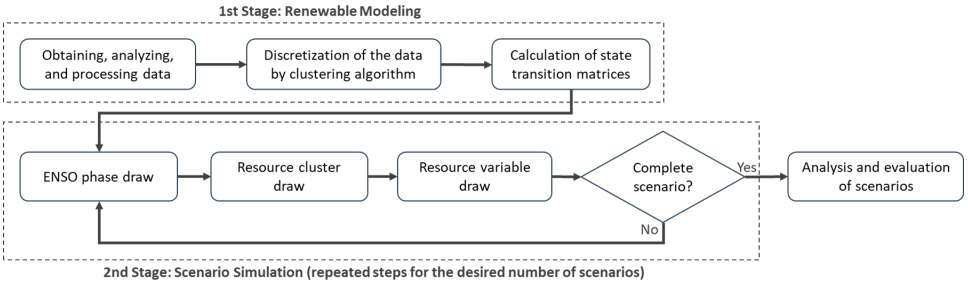

Fig. 1. Renewable scenarios simulation framework.

or attributes corresponding to the number of renewable sources. Additionally, clustering is carried out independently for each class, maintaining the characteristics of each location l of interest at time t for ENSO phase e. At the end of the discretization process, each vector of VREs at time t must belong to a particular cluster.

In the third step of the modeling, there are two types of state transition matrices: (i) between VREs and (ii) between ENSO phases.

Starting with the first ones, with the finite number of states defined in the previous step, the VREs matrices must be created at the same discretization periodicity for consistency. The stationary transition probability ($p_{a,b}$) from state a to b, for all indices $1 \leq (a,b) \leq k$, can be calculated according to: $p_{a,b} = (n_{a,b}/\sum_{i=1}^{k} n_{a,i})$, where $n_{a,b}$ is the number of transitions from a to b in the data history, and the denominator is the total transitions from a to all other possible states. After obtaining the transition probabilities for each state, it is possible to build the transition matrix according to (1), which represents a transition matrix of renewable resources for a specific location l from time t, considering that at time t the ENSO phase is e_i, to time $(t+1)$, considering that the ENSO phase will be e_j at $(t+1)$, with $(i,j) \in \{\text{Neutral, El Niño, La Niña}\}$.

$$P^{l}_{t_{e_i}\rightarrow(t+1)_{e_j}} = \begin{bmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,k} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ p_{w,1} & p_{w,2} & \cdots & p_{w,k} \end{bmatrix} \tag{1}$$

In the transition matrix, the rows correspond to the possible origin states (clusters) of the period t and the columns to the potential destination states of the period $(t+1)$. Detailing (1), w states are calculated for the period t in ENSO phase e_i. For period $(t+1)$ in ENSO phase e_j, there are a total of k states. Thus, $p_{1,1}$, for example, represents the transition probability from state 1 in period t in ENSO phase e_i to state 1 in period $(t+1)$ in ENSO phase e_j, which is another state compared to period t, given that, as detailed in the clustering step, the clusters are calculated independently for each combination of "ENSO phase - period". Furthermore, each state has one dimension for each VRE considered. Thus, for example, if two sources are being modeled (e.g., wind and PV), $p_{1,1}$ will be the transition probability from one bivariate state (wind-PV) to another bivariate state (wind-PV). By the end, the number of transition matrices for a location should correspond to the total number of possible combinations among the different time levels in the three ENSO phases.

From each matrix obtained by (1), the accumulated transition matrix is calculated according to (2):

$$P^{l,accum}_{t_{e_i}\rightarrow(t+1)_{e_j}} = \begin{bmatrix} p_{1,1} & p_{1,1}+p_{1,2} & \cdots & p_{1,1}+p_{1,2}+..+p_{1,k} \\ p_{2,1} & p_{2,1}+p_{2,2} & \cdots & p_{2,1}+p_{2,2}+..+p_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ p_{w,1} & p_{w,1}+p_{w,2} & \cdots & p_{w,1}+p_{w,2}+..+p_{w,k} \end{bmatrix} \tag{2}$$

To compute the accumulated matrix, the probabilities of each row in (1) are incrementally summed up until the last column. Consequently, all values in the last column of (2) must be 1, as each value in this column represents the sum of all possible transition probabilities starting from the cluster of the corresponding row. The accumulated matrix will be important in the simulation stage. The strategy to compute the transition matrices between ENSO is similar to the previous case, based on historical transitions. However, these represent the transition probabilities between the phenomenon phases for each period t to $(t+1)$. Thus, there is no need to apply clustering algorithms to identify the ENSO states. They should also have their accumulated version, named $P^{ENSO,accum}_{t\rightarrow t+1}$. Finally, they are independent of the VREs matrices and any location. Besides, these matrices are used in the first step of scenario simulation, detailed in the following subsection.

B. Scenario Simulation

The first step of the scenario simulation involves drawing the ENSO phase based on the accumulated transition matrix ($P^{ENSO,accum}_{t\rightarrow t+1}$) between the phases of the phenomenon calculated in the modeling. To draw the first value of the scenario, the last ENSO phase of the history is considered the state of origin. Then, a random number from a continuous Uniform distribution $[0,1]$ is compared with the elements of the row corresponding to the origin state in the accumulated matrix. Suppose the random number is greater than the cumulative probability of the previous state and less than or equal to the cumulative

probability of the subsequent state. In that case, this state is chosen to represent the next ENSO phase and will be used in the next stage of the simulation. The pseudo-code in Fig. 2 represents this procedure.

```
Begin
  Unif ~ U[0,1]
  Line = State(t)
  Column=1
  Repeat
    If (matrix(Line,Column)<Unif and Unif<=matrix(Line,Column+1)) then
      State(t+1) = Column+1
      Column = (No. columns matrix) + 1
    Else
      Column = Column + 1
  While (Column <= (No. columns matrix))
End
```

Fig. 2. Pseudo-code for the simulation based on state transition matrices.

The second step involves sampling the clusters from the accumulated transition matrix of VREs ($P^{l,accum}_{t_{e_i}\rightarrow(t+1)_{e_j}}$). The one is selected considering t as the origin time, state e_i (ENSO phase) at t, and e_j to the phase sampled in the first step for the time $(t+1)$. Thus, a new state is sampled for the period $(t+1)$, and the sampling mechanism is the same as shown in Fig. 2, now using the VREs matrix.

The third and final step involves randomly sampling an element of VREs belonging to the cluster drawn in the previous step. The possible elements of VREs were defined in the time series clustering step, and each has a number of dimensions corresponding to the number of renewable sources considered. Thus, we will finally have the simulated VREs for the period $(t+1)$.

The three simulation steps are repeated until the length for each scenario is reached. Finally, the proposed methodology must be evaluated to verify its performance and adherence to the historical features. To this end, we apply (i) a comparison between historical and simulated metrics, (ii) hypothesis tests to verify equality between historical and simulated distributions, and (iii) a comparison with the performance of benchmark models. Subsection II-C presents the benchmark models used in the context of the application developed in this work.

C. Benchmark Models

Observing the temporal granularity of the series and the simulation horizon is essential to select the benchmark models. In this sense, the proposed methodology was applied to the Brazilian monthly series to generate medium and long-term scenarios (Section III). From this, the benchmark models used for planning the Brazilian Electrical System (BES) for monthly simulations are the periodic autoregressive, or PAR (p) [22]–[29].

In general, monthly series of VREs are characterized by the periodic behavior of their probabilistic properties, such as mean, variance, asymmetry and autocorrelation structure [22]. Thus, the PAR (p) model fits an AR(p) model for each period of the series, where p is the order of the model, that is, the number of autoregressive terms. In general, p is a vector where each element gives the order of each period. For example, $p = [p_1,\ p_2,\ p_3, \ldots,\ p_{12}]$, in the case of monthly series. A Z series with S periods that repeat for N years can be represented by $Z = [z_{(1,1)}, z_{(1,2)}, \ldots, z_{(1,S)}, \ldots, z_{(N,S)}]$. The PAR model of this series Z in the period m is given by 3:

$$\left(\frac{z_{(t,m)} - \mu_m}{\sigma_m}\right) = \sum_{i=1}^{p_m} \varphi_i^{(m)} \left(\frac{z_{(t,m-i)} - \mu_{m-i}}{\sigma_{m-i}}\right) + u_{t,m} \qquad (3)$$

where μ_m is the mean over period m, σ_m is the standard deviation over period m, p_m is the autoregressive order over period m, $\varphi_i^{(m)}$ represents the ith autoregressive coefficient in the period m and $u_{t,m}$ is the series of independent residuals (mean zero and standard deviation = σ_m^u).

III. Case Studies

A. Data Base

To validate the proposed methodology, scenarios were generated for different regions of the Brazilian territory to obtain robust and generalized results, considering three renewable resources - hydro, wind, and PV - that account for more than 80% of Brazil's electrical matrix [30]. Data were obtained at the spatial level relative to the country's river basins. Seventeen basins cover the entire national territory. The data correspond to monthly time series of primary resources, that is, incremental flow, wind speed, and solar radiation, aiming to carry out medium-term simulations. It is noteworthy that the purpose of the case studies was not to transform primary resources into energy but to evaluate whether the proposed methodology captures the randomness of VREs and their complementary characteristics to generate representative scenarios.

The hydro data of the basins were obtained from NEWAVE [31], a tool used for planning the Brazilian Energy System Operation. The system receives hydrological flow data primarily through government agencies, such as the *Agência Nacional de Águas* (ANA) and the *Operador Nacional do Sistema Elétrico* (ONS), that collect inflow data at various hydrological stations distributed along the rivers and basins of the country. Two of the 17 basins are unavailable in NEWAVE, so these regions were not considered in this application.

The other variables, wind speed, extracted to a height of 100m, and surface solar radiation, were extracted from the Merra-2 reanalysis base [32], a widely used alternative by academia and practitioners when there is a lack of or limited measured data for renewable resources [4], [33]. Each basin covers a region with several points on the Merra-2 grid, so a monthly average of each wind and PV resource was obtained to obtain a single series per basin.

Although the incremental flow series are available from 1931 in NEWAVE, the others are only from January 1980 in Merra-2. As they all must have the same time window, they were obtained from 1980 onwards. The last

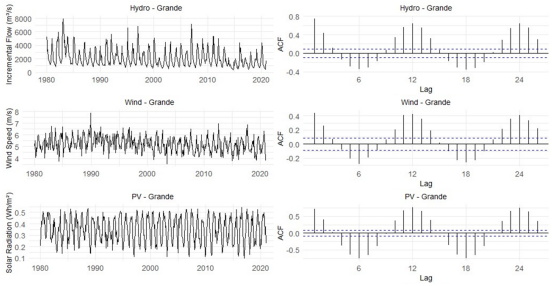

Fig. 3. Time series and ACFs - '*Grande*' Case.

observation corresponds to December 2022, totaling 42 years, or 504 monthly observations.

Fig. 3 shows an example of the three resources of the '*Grande*' basin and their respective autocorrelation functions (ACFs). The graphics of the other basins can be accessed in a GitHub repository through this link (https://github.com/gustavmelo/Paper-PSCC-2024), as well as the full results of the case studies developed in this work. Time series, in general, present a well-defined seasonal behavior. The ACF measures the degree of correlation of a variable at a given instant with itself at a later time instant. The ACF is, therefore, an essential tool for finding patterns of repetition/seasonality in time series. A relevant finding is related to the lags in which the autocorrelation values stand out compared to the others. In all cases, lags multiple of 12 stand out with positive autocorrelations, as the VREs tend to behave similarly at the same months due to the seasons. On the other hand, a lag multiple of 6 stands out with negative autocorrelations since the behaviors tend to be opposite in the inverted seasons. Furthermore, the VREs time series must be stationary so that modeling can be carried out using state transition matrices. To this end, the Augmented Dickey-Fuller (ADF) test [18] was applied, and, for a significance level of 5%, the null hypothesis of non-stationarity was rejected for all cases. Finally, the ONI index was used month by month to identify the ENSO phases, whose data were extracted from the National Oceanic and Atmospheric Administration (NOAA) [34].

B. Renewable Modeling

All stages of the methodology and analysis of the results were carried out in the R programming language [35]. Due to space limitations, we presented one example of each step. The other results are available in the same link to the GitHub repository presented in the previous section. Additionally, the data from the last two years of the dataset (2021 and 2022) were not utilized in the modeling process as an out-of-sample period for comparison with both, the benchmark and proposed scenarios.

Starting with the state transition matrices of the climate phenomenon calculated from the history of the ONI index, Table I shows the transition probabilities between the ENSO phases from February to March, as an example. There are 12 matrices of this type, corresponding to the total number of possible transitions between the months. Note that there are no transition probabilities between the El Niño and La Niña phases by definition of the phenomenon's index, which must always go through a neutral period between the two extreme phases.

TABLE I
ENSO Phase Transition Matrix from February to March - $P^{ENSO}_{February \rightarrow March}$.

State	Neutral	El Niño	La Niña
Neutral	0.941	0.059	0.000
El Niño	0.333	0.667	0.000
La Niña	0.083	0.000	0.917

To model the VREs, the clusters, and transition matrices were also calculated by month. All series were normalized before the clustering step to eliminate the influence of data dimensionality. For each basin, we apply the k-means algorithm 36 times, covering all combinations of 12 months of the year with the three phases of ENSO, and the cluster number for each of the 36 classes ranged from two to five. This number of clusters has enabled, in all cases, the estimation of the transition matrices from the available historical data. The centroid of each identified cluster represents a state for calculating the transition matrices. As for the monthly transition matrices per basin, we have 84 in total, 12 for each possible phase transition, i.e.: (i) between the months of the same ENSO phase: Neutral-Neutral, El Niño-El Niño, and La Niña-La Niña, totaling 36 matrices; (ii) between the months of different ENSO phases: Neutral-El Niño, Neutral-La Niña, El Niño-Neutral, and La Niña-Neutral, totaling 48 matrices. Table II shows the transition matrix between VREs from "January/Neutral Phase" to "February/Neutral Phase" for the '*Grande*' basin case.

TABLE II
Example of VREs Transition Matrix ($P^{Grande}_{January_{(Neutral)} \rightarrow February_{(Neutral)}}$) - '*Grande*' Case.

January - Neutral	**February - Neutral**			
	$1_{Feb,Neut}$	$2_{Feb,Neut}$	$3_{Feb,Neut}$	$4_{Feb,Neut}$
$1_{Jan,Neut}$	0.20	0.00	0.60	0.20
$2_{Jan,Neut}$	0.25	0.00	0.25	0.50
$3_{Jan,Neut}$	0.00	0.50	0.50	0.00
$4_{Jan,Neut}$	0.00	0.00	0.00	1.00

Note that although both "January-Neutral" and "February-Neutral" presented four VREs states each, it is noteworthy that they are not the same states since the clustering of each "month-phase" data set was carried out individually. Furthermore, each state is represented by the cluster's centroid to which it refers. So, each state simultaneously represents the three renewable sources since each centroid is a vector with three dimensions or attributes, one for each VRE, exploring the complementary effects between the renewable sources. As an example, state 1 of "January-Neutral" is represented by the following centroid: [6827 m^3/s, 5.01 m/s, 0.11 Wh/m^2], with the first

attribute referring to the incremental flow, the second, to the speed of wind and the third, to the solar radiation. State 1 of "February-Neutral" has the following centroid: [4723 m^3/s, 5.29 m/s, 0.22 Wh/m^2]. All transition matrices, whether between ENSO phases or between VREs, also have their version of accumulated probabilities, effectively used in the simulation steps, the results of which are detailed in the following subsection.

C. Scenario Simulation

For each basin, 2000 scenarios of 60 months, equivalent to five full years, were generated based on the simulation horizon of NEWAVE. To simulate the first state of each scenario, the last historical observation was used as the predecessor state. According to Fig. 1, an ENSO phase must be drawn in the first step. The last month in in-sample data was December 2020, whose ENSO phase was La Niña. Therefore, the scenarios start in January, and first, an ENSO phase is drawn from the accumulated matrix from December to January ($P^{ENSO,accum}_{Dec \rightarrow Jan}$). To do this, a uniform random number [0,1] is generated via Monte Carlo. With the random number, the next phase of the phenomenon is identified from the line corresponding to the predecessor state (La Niña), according to Fig. 2.

The second step is the VREs cluster draw. With the previous ENSO phase drawn, the VREs cluster for January is drawn using the correct VREs transition matrix. For each basin, this matrix must correspond to the transition of resources from "December-La Niña" to "January-Phase drawn firstly". The process in the matrix is similar to the first step, with the predecessor state corresponding to the last state of VREs identified by the clustering. In the third and final step, with the cluster of VREs for January defined, one of the resource vectors belonging to the cluster is randomly drawn. One could use the cluster's centroid to represent the final values of VREs, but to introduce higher variability to the scenarios, this third step was added, increasing the possibilities of VREs in the synthetic series. The three steps were repeated until the predefined number of scenarios was completed.

Following are two relevant observations: (i) the ENSO phases drawn in the first step are the same for all basins in all scenarios, keeping the correlation between the different regions concerning the phenomenon - the differences occur from the second step onwards with the VREs transition matrices and (ii) for the extreme phases of ENSO to occur, there must be at least five consecutive months of warming (El Niño) or cooling (La Niña) of the waters of the Pacific Ocean in the measurement region of the ONI index. Therefore, to represent this specificity of the phenomenon in the scenarios, the same phase is maintained when entering one of the two phases, interrupting the first stage of the simulation until completing five months. Then, after five months in the same phase, we return with the draw for the first step, enabling the continuation or not from the current ENSO state. If an exit and a subsequent return to one of the two phases (El Niño or La Niña) occurs, the five months must be fulfilled again.

D. Scenario Analysis

Comparing the ENSO phases' frequencies between the in-sample historical and the scenarios month by month, we see, in Fig. 4, that the phenomenon phases present seasonality, well reproduced by the simulations.

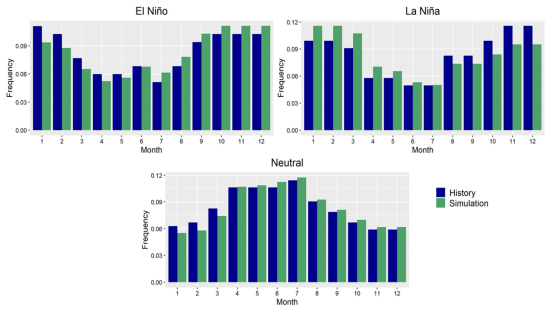

Fig. 4. ENSO phases histograms.

Proceeding to the VREs evaluation, analyzing them by ENSO phase, the Kolmogorov-Smirnov test [20] was applied to compare, with statistical robustness, whether the distributions of the simulated data follow the historical distributions by phase of the phenomenon. At a significance level of 10%, the results indicate that, for the three VREs, in the three phases of the phenomenon, considering all basins, the simulated scenarios present the same historical distribution in 98% of cases, as the null hypothesis of equality of distributions was not rejected.

Validating the results based on comparisons with the Brazilian market benchmark is essential. Thus, considering all 15 basins and 12 months of the year, 180 PAR(p) models were estimated. Each model had its order selected based on the partial autocorrelation function of the historical series, ranging from 1 to 6. The estimated models generated the same number of scenarios as the methodology proposed for each basin.

Using the '*Grande*' basin as an example, Table III presents several metrics of the historical in-sample data (*Hist*) and for the scenarios obtained from both the proposed methodology (*Prop*) and the benchmark (*Bench*). For the simulated metrics, it also presents the 1st and 99th percentiles of the scenarios, within parentheses. Finally, the table shows the discrepancies (*disc*) between the simulations and historical data. The discrepancy is defined as the absolute percentage difference between the value of the historical metric and the simulated mean for this metric. So, *disc* is calculated by $(|(Historical - Simulated)/Historical|) \times 100$. Analyzing all basins, in 80% of measurements, the proposed methodology's results are more accurate, i.e., they showed lower discrepancies (*Prop Disc*) compared to the benchmark (*Bench Disc*).

Analyzing Table III, we observe that the percentiles of the benchmark present more extreme values than the ones

TABLE III
METRICS COMPARISON – '*GRANDE*' CASE.

VREs	Data	Mean	Median	SD	Asymmetry	Kurtosis
Hydro	Hist	2098	1702	1298	1.36	4.79
	Bench	1971 (1407, 2606)	1607 (1134, 2193)	1232 (825, 1659)	1.46 (0.70, 2.41)	5.60 (2.39, 11.3)
	Prop	2105 (1824, 2482)	1698 (1416, 1981)	1300 (939, 1749)	1.36 (0.66, 2.04)	4.76 (2.30, 8.08)
	Bench Disc	6.1%	5.6%	5.1%	7.4%	16.9%
	Prop Disc	0.3%	0.3%	0.2%	0.4%	0.7%
Wind	Hist	5.22	5.18	0.64	0.21	3.01
	Bench	5.23 (5.01, 5.47)	5.22 (4.94, 5.49)	0.63 (0.51, 0.76)	0.12 (-0.50, 0.83)	2.76 (1.91, 4.29)
	Prop	5.21 (5.06, 5.36)	5.17 (4.96, 5.41)	0.64 (0.52, 0.79)	0.23 (-0.36, 1.11)	3.02 (1.96, 5.81)
	Bench Disc	0.3%	0.8%	0.5%	41.6%	8.0%
	Prop Disc	0.2%	0.1%	0.4%	13.8%	0.4%
PV	Hist	0.35	0.36	0.11	-0.23	1.91
	Bench	0.35 (0.32, 0.38)	0.36 (0.32, 0.40)	0.11 (0.10, 0.12)	-0.17 (-0.51, 0.16)	1.99 (1.60, 2.61)
	Prop	0.35 (0.33, 0.37)	0.36 (0.32, 0.39)	0.11 (0.10, 0.13)	-0.23 (-0.52, 0.05)	1.88 (1.56, 2.34)
	Bench Disc	0.5%	1.6%	0.7%	24.2%	4.2%
	Prop Disc	0.3%	0.1%	1.7%	0.7%	1.4%

of the proposed model in most cases. PAR(p) model is criticized for poorly representing the VREs asymmetry, while the proposal achieved better performance. When poorly estimated, it harms the representation of the tail of the distributions, thus potentially overestimating or underestimating extreme cases of VREs. This causes a misrepresentation of the risk of renewable energy assets.

According to the predefined criteria, the last two years of the database were set aside as an out-of-sample period. Fig. 5 presents, for each VRE of the '*Grande*' basin, the first two years of simulations, both from the benchmark and the proposal, along with the observed data in 2021 and 2022. Notably, although both methodologies reproduce the observed seasonal pattern, the PAR(p) presents more extreme scenarios, corroborating the in-sample analysis, where PAR(p) elongates the tails in the scenarios, often simulating unrealistic values. Our proposal generally keeps the scenarios within a narrower and more realistic range to cover the observed and feasible future. A similar pattern can be noted for all other basins. (GitHub link).

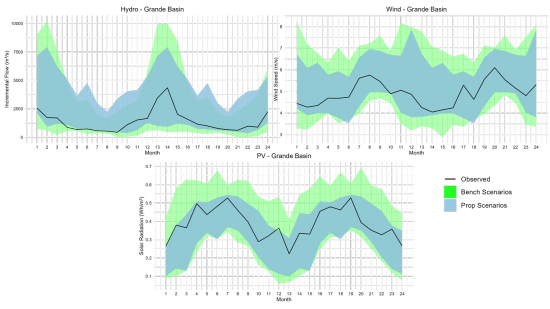

Fig. 5. Scenarios and observed data (2021-2022) - '*Grande*' Case.

Regarding the out-of-sample period, there was an observed prevalence of La Niña according to the ONI index - 83% of the months, while 17% were in the neutral phase. In the '*Grande*' basin region, La Niña tends to reduce rainfall volumes, increase solar incidence, and can bring wind patterns to lower levels. Thus, it is observed that compared to in-sample averages, 2021 and 2022 presented generally lower levels for incremental flow and wind speed, besides the opposite for solar radiation. At the same time, the two initial years of the scenarios of the proposed methodology presented 42% of the months in La Niña, 41% in the Neutral phase, and 17% in El Niño. Despite a significant portion of the months being in La Niña, its frequency was lower than observed, justifying the higher simulated average levels for the hydro and wind resources compared to the observed data, besides the opposite pattern for PV. Finally, we accurately simulated the frequency of ENSO phases compared to in-sample data (Fig. 4), but aiming to increase accuracy in out-of-sample periods in future work, we may use ENSO forecasts from NOAA institution [34]. However, we highlight the limitation of its forecast horizon, which is no longer than eight months.

IV. CONCLUSION

This work proposes a framework for simulating VREs stochastic processes, capturing the complementarity between different sources, besides the influence of climatic phenomena. To evaluate the proposal, monthly data from three sources covering the entire Brazilian territory were used. We emphasize that, for monthly series, relatively long historical records enable estimating states that better capture the characteristics of the resources in different ENSO phases. Using at least 20 years of monthly data is recommended, according to tests previously conducted.

It is worth noting that the monthly indices of Pacific temperature used to identify the ENSO phase have been at a more extreme level of oscillations in the last six decades. Therefore, as we have used data from this period, we consider this new behavior in the scenarios. Such patterns are unlikely to change for medium-term scenarios, as in the case of the study. However, if the interest lies in scenarios for many decades ahead, considering changes over time or simulations that encompass short-term changes, adaptations to the methodology would be necessary.

Finally, the proposal outperformed the official Brazilian benchmark model. As suggestions for future works: (i) applying the scenarios in energy planning models; (ii)

considering other climatic phenomena (e.g., sunspots); (iii) validating the methodology with sub-monthly data, such as hourly, assessing whether the scenarios reproduce the hour-to-hour ramping patterns; (iv) adapting the methodology to represent long-term climate trends or even very short-term oscillations in VREs states; (v) employing ENSO official forecasts in the simulation beginning.

References

[1] D. Çelik, M. E. Meral, and M. Waseem, "Investigation and analysis of effective approaches, opportunities, bottlenecks and future potential capabilities for digitalization of energy systems and sustainable development goals," *Electric Power Systems Research*, vol. 211, p. 108251, 2022.

[2] J. M. Morales, R. Minguez, and A. J. Conejo, "A methodology to generate statistically dependent wind speed scenarios," *Applied Energy*, vol. 87, no. 3, pp. 843–855, 2010.

[3] P. Pinson, "Wind energy: Forecasting challenges for its operational management," *Statistical Science*, vol. 28, no. 4, pp. 564–585, 2013.

[4] J. Jurasz, J. Mikulik, P. B. Dabek, M. Guezgouz, and B. Kaźmierczak, "Complementarity and 'resource droughts' of solar and wind energy in poland: an era5-based analysis," *Energies*, vol. 14, no. 4, p. 1118, 2021.

[5] NCEP, Available: https://origin.cpc.ncep.noaa.gov/products/precip/CWlink/MJO/enso.shtml, 2023.

[6] T. Iizumi, J.-J. Luo, A. J. Challinor, G. Sakurai, M. Yokozawa, H. Sakuma, M. E. Brown, and T. Yamagata, "Impacts of el niño southern oscillation on the global yields of major crops," *Nature communications*, vol. 5, no. 1, p. 3712, 2014.

[7] T. B. Murari, A. S. N. Filho, M. A. Moret, S. Pitombo, and A. A. Santos, "Self-affine analysis of enso in solar radiation," *Energies*, vol. 13, no. 18, p. 4816, 2020.

[8] A. T. S. dos Santos and C. M. Santos e Silva, "Seasonality, interannual variability, and linear tendency of wind speeds in the northeast brazil from 1986 to 2011," *The Scientific World Journal*, vol. 2013, 2013.

[9] H. Li, Z. Ren, M. Fan, W. Li, Y. Xu, Y. Jiang, and W. Xia, "A review of scenario analysis methods in planning and operation of modern power systems: Methodologies, applications, and challenges," *Electric Power Systems Research*, vol. 205, p. 107722, 2022.

[10] H. R. Baldioti and R. C. Souza, "Using a markov chain monte carlo technique to simulate synthetic natural inflow energy scenarios," in *2018 IEEE International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*. IEEE, 2018, pp. 1–6.

[11] S. Farah, N. Humaira, Z. Aneela, E. Steffen *et al.*, "Short-term multi-hour ahead country-wide wind power prediction for germany using gated recurrent unit deep learning," *Renewable and Sustainable Energy Reviews*, vol. 167, p. 112700, 2022.

[12] R. Yuan, B. Wang, Y. Sun, X. Song, and J. Watada, "Conditional style-based generative adversarial networks for renewable scenario generation," *IEEE Transactions on Power Systems*, vol. 38, no. 2, pp. 1281–1296, 2022.

[13] M. Alipour, J. Aghaei, M. Norouzi, T. Niknam, S. Hashemi, and M. Lehtonen, "A novel electrical net-load forecasting model based on deep neural networks and wavelet transform integration," *Energy*, vol. 205, p. 118106, 2020.

[14] G. Luo, D. Shi, J. Chen, and X. Wu, "A markov chain monte carlo method for simulation of wind and solar power time series," *Power System Technology*, vol. 38, no. 2, pp. 321–327, 2014.

[15] A. Almutairi, M. H. Ahmed, and M. Salama, "Use of mcmc to incorporate a wind power model for the evaluation of generating capacity adequacy," *Electric Power Systems Research*, vol. 133, pp. 63–70, 2016.

[16] S. Miao, K. Xie, H. Yang, H.-M. Tai, and B. Hu, "A markovian wind farm generation model and its application to adequacy assessment," *Renewable energy*, vol. 113, pp. 1447–1461, 2017.

[17] P. M. Maçaira, Y. M. Cyrillo, F. L. C. Oliveira, and R. C. Souza, "Including wind power generation in brazil's long-term optimization model for energy planning," *Energies*, vol. 12, no. 5, p. 826, 2019.

[18] D. A. Dickey and W. A. Fuller, "Distribution of the estimators for autoregressive time series with a unit root," *Journal of the American Statistical Association*, vol. 74, no. 366a, pp. 427–431, 1979.

[19] R. A. Fisher, *Statistical Methods for Research Workers*. Oliver and Boyd, 1925.

[20] F. J. Massey Jr, "The kolmogorov-smirnov test for goodness of fit," *Journal of the American statistical Association*, vol. 46, no. 253, pp. 68–78, 1951.

[21] J. MacQueen *et al.*, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, vol. 1, no. 14. Oakland, CA, USA, 1967, pp. 281–297.

[22] M. Maceira and J. Damázio, "Use of the 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*, vol. 20, no. 1, pp. 143–156, 2006.

[23] R. C. Souza, A. L. M. Marcato, B. H. Dias, and F. L. C. Oliveira, "Optimal operation of hydrothermal systems with hydrological scenario generation through bootstrap and periodic autoregressive models," *European Journal of Operational Research*, vol. 222, no. 3, pp. 606–615, 2012.

[24] P. G. C. Ferreira, F. L. C. Oliveira, and R. C. Souza, "The stochastic effects on the brazilian electrical sector," *Energy Economics*, vol. 49, pp. 328–335, 2015.

[25] I. F. Pereira, L. Hoffmann, L. d. O. Willer, I. da Silva Chaves, E. J. de Oliveira, T. P. Ramos, and A. L. M. Marcato, "Using stochastic dual dynamic programming and a periodic autoregressive model for wind-hydrothermal long-term planning," in *2015 IEEE Eindhoven PowerTech*. IEEE, 2015, pp. 1–6.

[26] C. M. B. de Castro, A. L. M. Marcato, R. C. Souza, I. C. S. Junior, F. L. C. Oliveira, and T. Pulinho, "The generation of synthetic inflows via bootstrap to increase the energy efficiency of long-term hydrothermal dispatches," *Electric Power Systems Research*, vol. 124, pp. 33–46, 2015.

[27] P. M. Maçaira, F. L. C. Oliveira, P. G. C. Ferreira, F. V. N. d. Almeida, and R. C. Souza, "Introducing a causal par (p) model to evaluate the influence of climate variables in reservoir inflows: a brazilian case," *Pesquisa Operacional*, vol. 37, pp. 107–128, 2017.

[28] X. Huang, P. M. Maçaira, H. Hassani, F. L. C. Oliveira, and G. Dhesi, "Hydrological natural inflow and climate variables: time and frequency causality analysis," *Physica A: Statistical Mechanics and its Applications*, vol. 516, pp. 480–495, 2019.

[29] M. L. de Menezes, R. C. Souza, and J. F. Pessanha, "Scenarios generation using bootstrap in the multichannel singular spectrum analysis approach and par (p) structures: Application to affluent natural energy," *International Journal of Computer Applications*, vol. 975, p. 8887, 2021.

[30] ONS, Available: https://www.ons.org.br/paginas/resultados-da-operacao/historico-da-operacao/dados-gerais, 2023.

[31] CCEE, Available: https://www.ccee.org.br/acervo-ccee?especie=44884&periodo=365, 2023.

[32] NASA, Available: https://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/, 2023.

[33] S. C. de Aquino Ferreira, F. L. C. Oliveira, and P. M. Maçaira, "Validation of the representativeness of wind speed time series obtained from reanalysis data for brazilian territory," *Energy*, vol. 258, p. 124746, 2022.

[34] NOAA, Available: https://www.ncei.noaa.gov/access/monitoring/enso/sst, 2023.

[35] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2021. [Online]. Available: https://www.R-project.org/