Energy Systems
https://doi.org/10.1007/s12667-023-00580-5

ORIGINAL PAPER

Combining monthly wind and inflow uncertainties in the stochastic dual dynamic programming

Application to the Brazilian interconnected system

Maria Elvira P. Maceira[1] · Albert C. G. Melo[1] · José Francisco M. Pessanha[1,2] · Cristiane B. Cruz[2] · Victor A. Almeida[2] · Thatiana C. Justino[2]

Received: 21 November 2022 / Accepted: 15 April 2023

Abstract

Following a global trend, intermittent sources, especially wind, have been experiencing accelerated growth in Brazil—in the last decade, wind power grew 13 times and became the second largest source in the electricity mix (12%), just behind hydropower (60%). Currently, although following regulatory guidelines, the representation of wind power in the long-term operation planning model is done in a simplified way, based on the monthly average of the last five years of aggregated generation, thus demanding improvements. The objective of this work is to describe an approach to be used by the Brazilian power industry to represent the uncertainties of monthly wind power production in the SDDP algorithm applied in the long-term operation planning model, keeping the large-scale stochastic problem still computationally viable, when applied to large interconnected systems, especially with hydroelectric predominance. The proposed methodology comprises statistical clustering of wind regimes and definition of equivalent wind farms; evaluation of monthly transfer functions between wind speed and power production; integrated generation of monthly multivariate synthetic scenarios of inflows and winds, considering associated cross-correlations; and representing monthly wind power in the SDDP algorithm. The application to real configurations of the Brazilian interconnected system, including case studies related to the monthly operation program and the calculation of the maximum amount of energy that can be traded in long-term power purchase agreements, points to its effectiveness and the relevance of modeling the wind uncertainties in the long-term operation planning of large hydro-dominated systems.

A. C. G. Melo and J. F. M. Pessanha contributed to the conception and design of the research; analysis of data and results; the creation of new software used in the work; and substantively revised the drafts of the paper. C. B. Cruz, V. A. Almeida and T. C. Justino contributed to the acquisition and analysis of data; and to the creation of new software used in the work.

Extended author information available on the last page of the article

Published online: 03 June 2023

Keywords Intermittent renewable sources · Wind power · Long-term operation planning · Interconnected systems · Stochastic optimization · Wind speed modelling correlated to inflows

1 Introduction

Intermittent sources, especially wind, have experienced accelerated growth in Brazil—in the last decade, wind power grew 13 times, reaching today 22.3 GW of installed capacity in around 830 wind farms and became the second largest source in the electricity mix (12%), just behind hydropower (60%). According to the Ten-Year Energy Expansion Plan 2022–2031 [1], in 2031 the wind power installed capacity will increase 1.5 times, reaching 39,336 MW (14% of the country's electricity mix). In turn, the current 6,230 MW of solar PV installed capacity will reach 10,383 MW in 2031, accounting for 4.7% of the country's electricity mix.

Despite the advantages, the intermittency of hourly wind generation, constitutes a challenge for its integration into the system. Thus, it is essential to develop and improve methodologies to represent the uncertainties of intermittent renewable sources in the long, medium and short-term operation planning models [2–4].

In many countries, expansion and operation planning in systems with hydroelectric power has been carried out by disaggregating the problems into specific horizons [5–7]. In the case of Brazil, this problem is divided into expansion planning (long-term), operation planning (medium and short term), and operation programming, being solved through a chain of computational models [6]. One of the key models is NEWAVE [8], based on the Stochastic Dual Dynamic Programming—SDDP [9], which since 1998 has been used in official studies and real decision making regarding the Brazilian Interconnected Power System (BIPS). The NEWAVE model is used in expansion planning and in medium-term operation planning, providing expected cost-to-go functions for the short-term operation planning model as well as for computing probabilistic performance indices of the system's operating conditions.

Currently, in accordance with the guidelines of the Electricity Regulatory Agency (ANEEL), the representation of wind power in the NEWAVE model is carried out in a simplified manner, based on the monthly average of the last five years of net generation of each individually wind farm (WF), aggregated by subsystem and load level, for the entire planning horizon.

To overcome this problem, a methodology to consider monthly wind energy uncertainties in the NEWAVE model has been developed since 2020 [10, 11], and improved in methodological terms and computational efficiency through its application in real BIPS configurations.

Due to the strategic uses of NEWAVE for the Brazilian power industry [8], its validation process starts at the Permanent Committee for the Analysis of Methodologies and Computational Programs for the Electricity Sector (CPAMP), established by the National Council for Energy Policy, and coordinated by the Brazilian Ministry of Mines and Energy. After successful initial tests regarding the

proposed methodology, CPAMP decided to continue the validation concurrently with the Validation Task Force on the NEWAVE model, jointly coordinated by the Brazilian National Electrical System Operator (ONS) and Electrical Energy Trading Chamber (CCEE), and under the supervision of ANEEL.

The objective of this work is to describe the main features of the approach that is being validated to be used by the Brazilian power industry to represent the uncertainties of monthly wind power in the SDDP algorithm applied in the long-term operation planning model, keeping the large-scale stochastic problem still computationally viable, when applied to large interconnected systems, especially with hydroelectric predominance, as is the case of the Brazilian system.

Case studies with the application of the proposed approach to actual configurations of BIPS are presented and discussed.

2 Long and medium-term operation planning model in Brazil

In the NEWAVE model [8], the operation planning problem is represented as a multistage stochastic linear programming problem. The objective is to minimize the expected operation cost during the planning period considering risk aversion mechanisms, given a known initial state of the system. Fuel costs and penalties for failure in load supply compose the operation cost. The solution of this problem results in an operation strategy.

The several hydropower reservoirs can be aggregated in energy equivalent reservoirs (EERs) [12–14] or represented by a hybrid modeling, allowing the NEWAVE model to represent the hydropower plants (HPPs) individually in entire or in part of its planning horizon [15]. In turn, the system state includes the energy storage level of the aggregated reservoirs and information about the "hydrologic trend", as the last p energy or water inflows in each aggregated or individual reservoir.

The representation of the inflows to hydropower reservoirs in the NEWAVE model is stochastic, through a scenario tree, where each path in the tree is called a hydrological scenario, and each node represents a possible realization of the inflow. These sequences follow a multivariate stochastic process, temporally and spatially correlated, with statistical properties similar to the historical record, which are preserved during the tree construction. To generate the energy/water inflows scenarios for the optimization problem, a periodic autoregressive model of order p, PAR(p) [16–18] is used, that is, the value obtained for the random variable in a given period is a function of the inflows of the previous periods.

The solution strategy in the SDDP algorithm, without traversing the complete tree of inflow scenarios, consists of traversing a sub-tree of inflow scenarios, which is chosen from the original distribution of the random variable, iteratively through two steps: (i) *forward simulation*, from $t=1$ to $t=T$ traversing the entire sub-tree, in order to generate new states for which the expected cost-to-go function (FCF) will be evaluated and new cuts of Benders constructed in the next backward recursion; (ii) *backward recursion*, from $t=T$ to $t=1$, the Benders cuts that represent the FCF are built for all nodes of the subtree resulting from the last forward simulation. The dual variables associated with these linear programming subproblems are used to

construct the Benders cuts. The set of future cost functions represents the system optimal operating policy.

Once the operation strategy is calculated, a simulation of the system operation is performed by using 2000 multivariate synthetic inflows scenarios, or by considering historical record sequences. Thus, statistics of several system performance indicators are provided, such as total operating cost, marginal operating cost, risk of deficit, energy deficit, hydro and thermal generation, spills, etc.

The compact formulation of the medium and long-term operation planning problem represented in the NEWAVE model, in its recursive form, is presented in (1), whereas a more detailed formulation is presented in Sect. 7.

In (1), c_t represents the system costs in time stage t, the decision variables x_t, from the feasible set X, are associated with the reservoir levels x^t_{y} and the allocation of water resources x^t_{gh} and thermal resources x^t_{gt}. The uncertainty of the inflows to the reservoirs is represented by the vector ξ_t. The set of constraints is denoted by g_t, which includes the system demand equation, water conservation equations in the reservoirs and operation constraints for the generation plants and interconnections. The recursive term φ_{t+1} is the recourse function for the subproblem of time step t, which can be obtained iteratively by applying nested Benders decomposition approaches to solve the problem [19].

$$\begin{aligned} \min_{x_1} \; & c_1 x_1 + \mathop{\mathrm{E}}_{\xi_2}\left[\varphi_2(x_1, \xi_2)\right] \\ s.t. \; & g_1(x_1) = b_1 \\ & x_1 \in X \end{aligned} \tag{1a}$$

$$\begin{aligned} \varphi_t(x_{t-1}, \xi_t) = \min_{x_t} & c_t x_t + \mathop{E}_{\xi_{t+1}|\xi_t,\ldots,\xi_{t+1-p}}\left[\varphi_{t+1}(x_t, \xi_{t+1})\right] \\ s.t. \; & g_t(x_t) = b_t(x_{t-1}, \xi_{t-j,j=1,\ldots,p}) \\ & x_t \in X;\; t = 2, \ldots, T \end{aligned} \tag{1b}$$

3 Overview of the proposed aproach to represent the monthly wind and inflow uncertainties in the SDDP algorithm

The number of wind farms currently installed in Brazil is already high (around 750) and it is expected that their growth will continue to accelerate over the next 10 years. Therefore, it is necessary to investigate how to represent the wind farms in the NEWAVE model, so that the number of state variables of the SDDP algorithm does not become too high. In this sense, and similarly to what already occurs with the representation of HPPs individually or by EERs, one possibility is to represent them individually (WF) or through equivalent wind farms (EWFs)—in the latter case, statistical grouping techniques are employed in the present work. To simplify the notation, henceforth the terms EER and EWF will be used interchangeably to represent hydropower plants and wind farms individually or in an equivalent way, respectively.

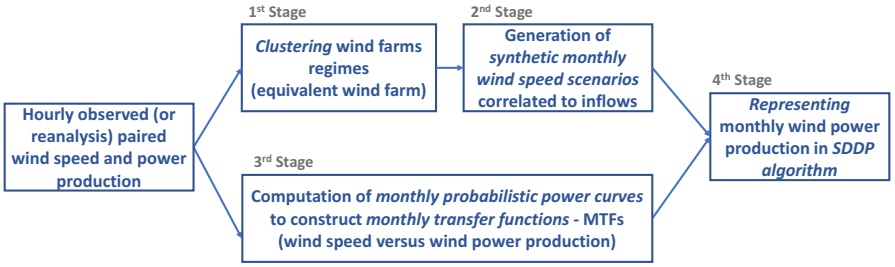

Fig. 1 Schematic diagram of the proposed approach

The existence of complementarity between the hydrological and wind speed regimes has been observed in Brazil, mainly in the Northeast region, where the greatest potential for wind generation in the country is concentrated [20]. Thus, when generating synthetic wind speed scenarios, it is interesting to consider the cross-correlation structure that may exist in the stochastic process of monthly average wind speeds and inflows to HPP reservoirs. In this way, this work proposes to extend the generation of inflows scenarios to make it an integrated model for the generation of monthly multivariate synthetic series of inflows and wind speeds, considering the correlations between wind speeds, between inflows and between wind speeds and inflows.

Once the synthetic scenarios of monthly wind speeds in the EWFs are obtained, it is necessary to estimate the associated wind power production to be considered in the SDDP algorithm's monthly dispatch problem. The proposed approach consists of constructing transfer functions (MTFs) between the monthly values of wind speeds and wind production, from EWF monthly probabilistic power curves. Then the MTFs are used in the operation dispatch problem of the SDDP algorithm.

As illustrated in Fig. 1, the proposed methodology consists of four main steps: (i) statistical clustering of wind regimes and definition of EWFs; (ii) an integrated model for the generation of monthly multivariate synthetic sequences of inflows and winds, considering the cross correlations between wind speeds, inflows, and wind speeds and inflows; (iii) evaluation of monthly transfer functions (MTFs) between wind speed and wind power production; and (iv) representing monthly wind power production in the SDDP algorithm.

Each step of the proposed approached is presented next, together with their application to the BIPS.

4 Aggregation of wind regimes into EWFs

The aggregation of wind regimes into EWFs is based on multivariate statistical methods and comprises two steps. Initially, the Exploratory Factor Analysis (EFA) [21] is applied to the covariance matrix between the wind speeds in the WFs in order to reduce the dimensionality of the data, from n WF to m ($m < n$) latent factors interpreted as wind regimes. Then, a cluster analysis algorithm (e.g., the K-Means or

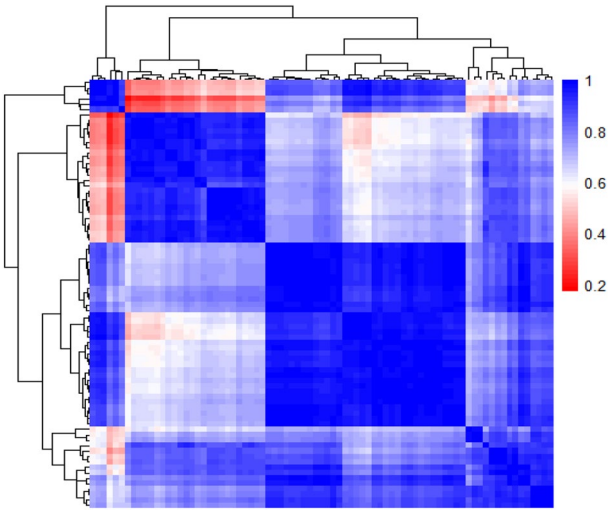

Fig. 2 Correlation matrix between monthly average wind speeds in 79 locations in Northeastern Brazil

Ward methods [21]) is applied to the coordinates of the WFs in m the latent factors in order to identify k groups of wind farms (EWFs).

For example, Fig. 2 presents a heatmap and associate dendrogram [21] of the correlation matrix between time series of monthly average wind speeds at 100 m height, from MERRA-2 (Modern-Era Retrospective analysis for Research and Applications from NASA) [22] for the period 2001–2017 at 79 municipalities in the Brazilian Northeast region. The 79 municipalities cover 498 wind farms with a total installed capacity of around 12,676 MW. The dendrogram on the heatmap of the correlation matrix in Fig. 2 indicates that wind regimes in the Northeast region can be grouped into 2, 3, 4 or 5 clusters. It is worth mentioning that in the past official expansion planning studies of BIPS considered 2 clusters for this region—coastal and inland [23]; more recently, [24] pointed out 3 clusters. Moreover, the diagonal blocks on the correlation matrix clearly show the existence of 3 clusters, being possible to identify even a fourth or a fifth cluster.

Let X be the data matrix, where each column stores the wind speed time series in a locality with wind farm. So, for the case of n sites with a time series with q hourly records of wind speed, the matrix X has dimensions q x n. From the matrix X one can obtain the matrix of covariances S between the wind speeds in the n locations. The matrix S has dimensions n x n and each element S_{ij} contains the covariance between wind speeds at locations i and j.

In EFA, it is assumed that the wind speed x_i in the WF at a site i (out of a total of n) is expressed by its average value μ_i plus the sum of the effects of m ($m < n$) wind regimes (latent factors $F_j \; \forall \; j = 1, \; m$) plus a residual term ε_i specific of the i-th site as shown in (2), where l_{ij} is the weight of the i-th site on the j-th latent factor F_j [21].

$$x_i = \mu_i + l_{i1}F_1 + l_{i2}F_2 + \cdots + l_{im}F_m + \varepsilon_i \forall i = 1, n \tag{2}$$

From the linear combination in (2) and the premise of independence between $F_1, F_2, \ldots, F_m$ and $\varepsilon_i \, \forall i = 1,n$ the following decomposition of the covariance matrix $S = LL^T + \Phi$ can be obtained, where L is a matrix with dimension of n x m, in which each row stores the weights of each location i in the m latent factors (wind regimes), i.e., each row is formed by the elements $l_{i1},\ldots,l_{im}$, $i = 1,n$. The first term (LL^T) is the commonality and correspond to the portion of the total variance of wind speeds at the n locations that is explained by the m wind regimes (latent factors). In turn, the term Φ is a diagonal matrix, whose elements capture the variability of wind speed in each location that is not explained by the m wind regimes.

The determination of the number of latent factors m and the formation of the matrix L consist in finding a value for m such that $S = LL^T$. By the Spectral Decomposition Theorem [21] the covariance matrix is expressed as a function of eigenvalues (λ) and respective eigenvectors (e):

$$S = \lambda_1 e_{i1} e_{i1}^T + \lambda_2 e_{i2} e_{i2}^T + \cdots + \lambda_n e_{in} e_{in}^T + \varepsilon_i \forall i = 1, n \tag{3}$$

Given that $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ the first eigenvalues concentrate the largest share of the total variance, then the first terms of the sum in (3) are the ones that most contribute to the formation of the matrix S. Then, a good approximation of the matrix S is achieved by the sum of the first m ($m < n$) terms in (3) such that the eigenvalues satisfy the following condition [21]:

$$100\%\left(\lambda_1 + \lambda_2 + \cdots + \lambda_m\right)/(\lambda_1 + \lambda_2 + \cdots + \lambda_n) \geq 80\% \tag{4}$$

Given the number of factors m, the matrix L can be generated based on the eigenvectors of S associated with the first m eigenvalues: $L = [\sqrt{\lambda_1} e_1 \ \sqrt{\lambda_2} e_2 \ \ldots \ \sqrt{\lambda_m} e_m]$. If the condition in (4) is satisfied with less than three factors ($m \leq 3$) it is possible to generate a visualization of the n locations with wind farms in a system of m factorial axes through of a map that allows the quick identification of groups of spatially correlated wind farms by a clustering algorithm, for example the K-Means and the Ward methods.

The application of EFA to the covariance matrix between the wind speeds in the 79 municipalities of Northeast Brazil revealed that 95% of the total variance is explained by the first 3 latent factors ($m = 3$), each one interpreted as a wind regime. The representation of the 79 municipalities in the space of the three latent factors is illustrated in Fig. 3a, where each point corresponds to one of the 79 sites. Figure 3a provides a good graphic representation of the correlation structure between the wind speeds in the 79 analyzed locations; the distances between the points in Fig. 3a reflect the correlations between the respective wind speeds, with close points indicating a greater correlation.

The dots (WFs) in Fig. 3a can be grouped into EWFs—visually or by using the K-Means method. In our case, 5 clusters captured about 90% of the data variability (measured by the ratio between the inertia between the clusters and the total inertia of the data). Figure 3b illustrates the classification of the 79 WFs into 5 EWFs, each

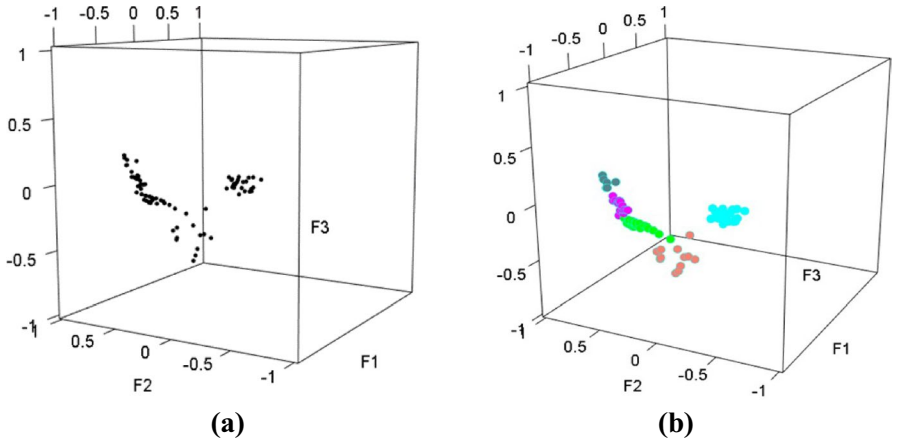

Fig. 3 Diagram of the 79 evaluated municipalities in the three latent factors (**a**) and groupings of wind farms (**b**)

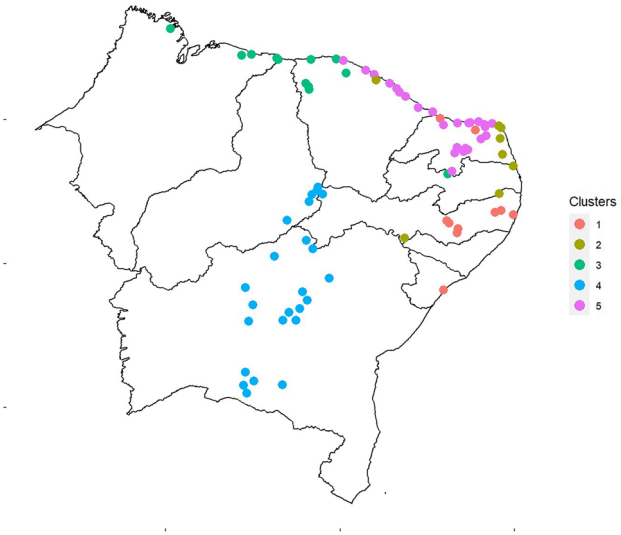

Fig. 4 Spatial representation of the wind farm clusters

indicated by a different color. In turn, the spatial representation of the 5 clusters is illustrated in Fig. 4, where it can observe the reduced overlap between the clusters.

Another possibility of WF aggregation could be to consider the substations that act as hubs to connect groups of wind farms (e.g., 37 in the Northeast and 12 in the South regions of Brazil). The definition of the final granularity of the EWF clusters is underway by the Standing Committee for Analysis of Methodologies

and Computational Programs of the Electric Sector—CPAMP, chaired by the Brazilian Ministry of Mines and Energy.

5 Generation of monthly multivariate wind speeds and inflows scenarios

In the current SDDP implementation of the NEWAVE model, a periodic autoregressive model of order p—PAR(p) is used to generate the energy/water inflows scenarios that are used in the forward and backward passes of the algorithm and in the simulation of the system operation with the calculated operation policy [25, 26]. This model can be written as:

$$\left(\frac{EI_{t,i}-\mu_{m,i}}{\sigma_{m,i}}\right)=\sum_{j=1}^{p_{m,i}}\phi_{t,j,i}\left(\frac{EI_{t-j,i}-\mu_{m-j,i}}{\sigma_{m-j,i}}\right)+a_{t,i} \tag{5}$$

where $EI_{t,i}$ is a random variable of a stochastic process with s seasonal periods and corresponds to the energy inflow of EER or HPP i at time t, which is a function of the year T and the seasonal period m: $t=(T—1)\ s+m$; p_m is the number of autoregressive terms in the model for the seasonal period m, $p_m<12$; $\mu_{m,i}$ and $\sigma_{m,i}$ are the mean and standard deviation of the stochastic process of the seasonal period m corresponding to stage t, respectively. The time uncorrelated series a_t is independent of EI_t, has zero mean and variance $\sigma_a^{2(m)}$, and can be written as a function of the $\rho^m(k)$ autocorrelations of EI_t and the periodic autoregressive coefficients ϕ [17].

The purpose of this work is to extend the synthetic inflow generation model to make it an integrated model for the generation of monthly multivariate synthetic sequences of inflows and wind speeds. In this sense, the random variable of the stochastic process with s seasonal periods that represents the monthly average wind speed in wind farm j at stage t is given by:

$$\left(\frac{V_{t,j}-\mu^v_{m,j}}{\sigma^v_{m,j}}\right)=explanatory\,component+a_{t,j} \tag{6}$$

Or, rewriting (6):

$$V_{t,j}=explanatory\,component+\sigma^v_m a_{t,j} \tag{7}$$

If the objective of the proposed methodology is not to extend the number of state variables of the SDDP algorithm (currently 84, in the case of representation by EERs and considering $p_m=6$), the time correlation structure, which may exist in the stochastic process of the monthly average wind speeds in any EWF, could not be explicitly represented in the synthetic series generation model. In this case, it would be represented by the spatial correlation between wind speeds and inflows to EERs or HPPs, which can be high in several months of the year, for different wind farms considered in the Monthly Operation Program carried out by ONS.

Hence, the explanatory component could be the average monthly wind speed for the seasonal period m corresponding to stage t, μ^v_m:

$$V_{t,j} = \mu^v_{m,j} + \sigma^v_{m,j} a_{t,j} \tag{8}$$

or could contain a portion related to inflows in a particular EER or HPP i from stage t, $EI_{t,i}$, or even from stage t-1, EI_{t-1}. The inclusion of this portion may contribute to the representation of the time correlation (lag 1) of $V_{t,j}$, if it exists. The process is then modeled by:

$$V_{t,j} = \mu^v_{m,j} + \theta_{t,j,i}\sigma^v_m \left(\frac{EI_{t,i} - \mu_{m,i}}{\sigma_{m,i}} \right) + \sigma^v_m a_{t,j} \tag{9}$$

where $\theta_{t,j,i}$ is the correlation coefficient between $V_{t,j}$, and $EI_{t,i}$, (or $EI_{t-1,i}$).

If the inclusion of new state variables is computationally viable, the process could be modeled as a PAR(1) in every month, where the representation of the temporal correlation (lag 1) of $V_{t,j}$, is explicitly considered:

$$V_{t,j} = \mu^v_{m,j} + \delta_{t,j}\sigma^v_m \left(\frac{V_{t-1,j} - \mu^v_{m-1,j}}{\sigma^v_{m-1,j}} \right) + \sigma^v_m a_{t,j} \tag{10}$$

where $\delta_{t,j}$ is the correlation coefficient between $V_{t,j}$ and $V_{t-1,j}$.

The developed scheme to generate monthly multivariate synthetic sequences of inflows and wind speeds comprises the following steps:

a. Obtain historical monthly EER incremental inflows;
b. Choose the order of the AR model for each seasonal period for each EER, by using the partial autocorrelation function [17];
c. Estimate the coefficients of the PAR(p) model using the Yule-Walker equation systems [17];
d. Generate a high cardinality sample (e.g., 100,000) of normal, time and spatially uncorrelated a_t residuals for both EERs and EWFs using simple random sampling, where they are considered to be equiprobable [25, 26];
e. Apply the K-Means method [21] to reduce the cardinality of the original sample; the resulting residuals then become non-equiprobable;
f. To generate multivariate monthly inflows and wind speeds, it is assumed that the standard normal residuals not spatially correlated, a_t, can be transformed into spatially correlated residual, e_t, through the following relationship:

$$e_t = D\, a_t \tag{11}$$

where the matrix D is obtained by decomposing the covariance matrix $\hat{U}$ between the residuals a_t [27]:

$$DD^T = \hat{U} \tag{12}$$

In practice, the behavior of the residuals does not follow the behavior of inflows and wind speeds: the residuals are not spatially correlated. However, in order to preserve the spatial dependencies of the stochastic processes of inflows and wind speeds, the spatial correlations between inflows to EERs, between wind speeds in EWFs and between inflows and wind speeds are employed, replacing the spatial correlations between the residuals.

g. A three-parameter Lognormal distribution is fitted to spatially correlated clustered residuals in order to better reproduce the skewness observed in this type of stochastic process [28]. However, unlike the inflow residuals, the monthly wind speed residuals can present positive as well as negative skewness in several months (see Fig. 6), which prevents, in the latter case, the use of the Lognormal distribution. In this case, one alternative is to use the Weibull distribution [29], which is quite flexible, allowing to deal with left or right skewness. In addition, the residuals have, by construction, negative values, which implies the need to use Weibull distributions with 3 parameters;
h. The synthetic monthly inflow scenarios are obtained by applying (5), while the monthly wind speed scenarios are obtained by (8), (9) or (10);
i. In each time period and scenario, the total inflows are calculated by the sum of the incremental inflows along the cascade of hydraulically coupled EERs.

Regarding step (g), several methods are available in the literature to estimate the shape (γ), scale (β) and location (α) parameters of tri-parametric Weibull distributions, most of them based on modifications of the method of moments (MoM) or maximum likelihood estimators (MLE) [30]. For example, the three basic statistics (central moments) of the Weibull distribution, i.e., expected value, standard-deviation and skewness, are respectively given by:

$$E(x) = \alpha + \beta\Gamma_1 \tag{13}$$

$$Var(x) = \beta^2(\Gamma_2 - \Gamma_1^2) \tag{14}$$

$$Sk(x) = \frac{\Gamma_3 - 3\Gamma_2\Gamma_1 + 2\Gamma_1^3}{\left(\Gamma_2 - \Gamma_1^2\right)^{3/2}} \tag{15}$$

where:

$$\Gamma_k(\gamma) = \Gamma(1 + k/\gamma) \tag{16}$$

and Γ(z) is the gamma function, defined as:

$$\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}dt \tag{17}$$

If the sample mean, standard-deviation and skewness are available, the 3 parameters can be estimated by, e.g., the MoM [30], solving sequentially *Sk(x)* for the shape γ (15), then *Var(x)* for the scale β (14) and, finally, *E(x)* for the location α (13).

However, as these estimations involves higher order statistics, both MoM and MLE may present difficulties and unsatisfactory results when considering the three Weibull parameters [31]. Indeed, we observed that the quality of estimates of these methods applied to the monthly average speeds of Brazilian wind farms varies greatly depending on the month of the year and the location of the wind farms, with the worst performances being associated with months with high negative asymmetries.

In this way, an approach for modeling residuals of monthly wind speeds through tri-parametric Weibull distributions was developed, seeking to preserve the mean, standard-deviation and skewness of monthly historical wind speeds, being especially suitable in situations of high asymmetries [32].

As described in [32], when the position parameter α is known, the estimates of the other two parameters can be calculated by MoM in a simpler way, since there is no need to use the skewness Eq. (14): the shape parameter γ can be estimated based on the coefficient of variation obtained from (13) and (14), however replacing the population mean and variance by the respective sample values; then the scale parameter β is obtained from (13). The approach starts from an initial value to estimate the position parameter, which can be obtained, e.g., through linear regressions; calculate estimates of other parameters using the method of moments; and, iteratively, updates the initial estimate in order to reduce the difference between the skewness of synthetic and sample (historical) monthly wind speeds. This proposal was applied to several EWFs, considering different months and skewness (positive and negative), presenting, in all cases, better performances than several methods available in the literature [32].

6 Transfer functions between monthly wind speed and power production

As the long and medium-term operation planning model adopts a monthly time step, the computed synthetic scenarios of monthly wind speeds for each EWF (stage 2 of Fig. 1) should be converted into monthly power production to be considered in the dispatch problem of the SDDP algorithm implemented in NEWAVE. This is achieved by obtaining mathematical functions, called monthly transfer functions (MTFs), that relate the monthly averages of wind speed with the monthly averages of energy production in each wind farm.

For this, it is necessary to use paired data of wind speed and wind power. However, due to the unavailability in Brazil of a public database of measured data, a procedure was proposed in [33] that uses the predicted values of wind speed and the respective wind power, on a half-hourly basis, made available since 2018 by ONS through the Sintegre system [34], for a set of hub substations. In addition to forecasts, this system provides 48 power curves, one for each half-hourly interval, used in converting wind speed forecasts into wind power. In order to expand the dataset,

it became necessary to consider hourly data from reanalysis, e.g., from MERRA-2 or ERA-5 (Reanalysis v5 from ECMWF) [35] for the geographic coordinates of the wind farms of each substation.

Reanalysis data is among the most commonly used datasets for studying weather and climate, and is produced by assimilating data, a technique for building long-term datasets that is widely used in climate studies, in a process known as retrospective analysis, or reanalysis. Reanalysis involves performing data assimilation for earlier periods using a current Numerical Weather Prediction model and data that is now available for those earlier periods. Thus, long and comprehensive sequences of atmospheric condition values are produced, forming a reanalysis dataset [36, 37].

For each EWF, the procedure consists of applying the power curves available in Sintegre to the hourly time series of wind speed reanalysis of each hub substation, to transform them into hourly estimates of wind power. Then, these hourly estimates are integrated, obtaining the time series of the monthly averages of wind speed and wind power, arranged in monthly probabilistic power curves. Finally, the MTFs are obtained through linear regression models—simple or piecewise [21]—adjusted to the monthly probabilistic power curves of each wind farm or hub substation.

7 Representation of wind power in the sddp algorithm of NEWAVE Model

As previously mentioned, currently, in accordance with ANEEL guidelines, the representation of wind, solar and biomass technologies in the NEWAVE model is carried out in a simplified way, based on the monthly average of the last five years of net generation of each individually plants, aggregated by subsystem and load level (*pqusi*—see (18.b)). In this way, *pqusi* is directly subtracted from the demand in the load supply equations. As in this work the monthly wind energy uncertainties in the SDDP algorithm will be explicitly represented, the term *pqusi* will no longer refer to wind energy, that is, it will only consider the amounts of solar and biomass power production.

Once the wind power synthetic productions of individualized or aggregated wind farms are obtained (see Sects. 5 and 6), they can be represented in the dispatch problem as an available generation source, but with null operating cost.

The formulation of the subproblem of each node of the subtree *(t,s)*, in each stage t, and forward scenario s and backward scenario ω, described in (1), is presented in detail, also modifying or adding constraints in which the wind power production must explicitly appear, according (18a)-(18d) plus the expected cost-to-go function (18e); for ease of viewing, modified/added terms are shown in bold. Table 1 describes the variables and parameters used in the mathematical formulation. For simplicity, the subscripts s and ω are omitted as well as the equations related to risk aversion mechanisms and dispatch of liquefied natural gas (*LNG*) thermal plant constraints. For details about risk aversion mechanisms and *LNG* constraints see [38–42].

The objective function (18a) is composed by fuel costs, penalties for failure in load supply and possible violations of operational constraints (minimum outflow,

Table 1 Variables and parameters used in the mathematical formulation in the SDDP algorithm of NEWAVE model

T	Time stage
m	Subsystem index
j	Thermal plant index
idef	Deficit level index
v	Operational constraints violation index
K, i	EER indices
c	Load level index
u	EWF index
l	Autoregressive term index
z	optimal value
cterm	Fuel costs
CDEF	Penalties for failure in load supply
DEF	Load supply deficit
NPDF	Number of deficit levels
NUT	Number of thermal plants in a subsystem
cv	Penalties for violations of operational constraints
viol	Operational constraints violation
nviol	Number of violations in each EER of a subsystem
β	Monthly discount rate
α	Expected cost-to-go function
NREE	Number of EERs
NPMC	Number of load levels
NSBM	Number of subsystems
GH	Controllable hydro generation
fpeng	Share of generation dispatch in a load level
GFIOL	Generation in run-of-river hydro plants
GT	Thermal generation
INT	Interchange energy between subsystems
EXC	Generation surplus
GW	Wind power production
NPEE	Number of EWFs
merc	Load to be supplied
submot	Generation of HPPs that have not yet reached their assured energy [43]
pqusi	Energy provided by non-dispatchable plants (constant power production), that is, solar and biomass technologies
gtmin	Minimum thermal generation
EA	Storage energy in the *EER* at the beginning of the stage
FDIN	When a new HPP start operating from stage *t* on, the storage capacity of the corresponding EER changes; the approach to reflect this on the storage energy at the beginning of the stage *t* is by applying a correction factor *FDIN* [44]
FC	Correction factor related to the initial state of EER as the energy inflows are previously calculated assuming that all HPPs of a EER are at 65% storage level
EC	Controllable energy inflow to HPP with regulation capability
EFIOL	Uncontrollable energy inflow to run-of-river HPP

Table 1 (continued)

EI	Energy inflow to an EER. The sum of EC and EFIOL results on EI in a specific EER
EVT	Spillage in an *EER*
EVP	Evaporation in an *EER*
EDV	Energy taken from an EER to meet multiple water use constraints
V	Average monthly wind speed of an EWF
a	Angular coefficient of the regression line that represents the associated MTF in an EWF
b	Linear coefficient of the regression line that represents the associated MTF in an EWF
$\overline{\pi}_{EA}$	Average partial derivative of the objective function related to state *EA* in an ERR at the beginning of the stage
$\overline{\pi}_{EAF}$	Average partial derivative of the objective function related to state *EI* in an ERR at the previous stages
p	Number of autoregressive terms in the PAR(p) model of an ERR
δ	Independent term of each Bender´s cut

water diversion, minimum hydraulic generation, etc.). The main constraints in each stage are the load balance equation in each load level and subsystem (18b), the controllable energy balance in each EER (18c) and the Benders cuts that represent the FCF (18e).

Objective Function

$$z_t = min \sum_{m=1}^{NSBM} \left(\sum_{c=1}^{NPMC} \left(\sum_{j=1}^{NUT_m} cterm_{t,m,j} \cdot GT_{t,m,j,c} \right) + \sum_{idef=1}^{NPDF} CDEF_{t,idef} \cdot DEF_{t,m,idef,c} \right) + \sum_{v=1}^{nviol} cv_{t,v}.viol_{t,m,k,c} + \frac{1}{1+\beta} \cdot \alpha_{t+1} \tag{18a}$$

Load supply equation in subsystem m in the load level c and stage t

$$\sum_{k \in NREE_m} (GH_{t,c,k} + fpeng_{t,c} GFIOL_{t,c,k}) + \sum_{j \in NUT_m} GT_{t,c,j} + \sum_{j=1, j \neq m}^{NSBM} \left(INT_{t,c}(i,k) - INT_{t,c}(k,i) \right) + \sum_{idef=1}^{NPDF} DEF_{t,m,idef,c} - EXC_{t,c,m} + \sum_{u=1}^{NPEE_m} \boldsymbol{GW_{t,u,c}} = merc_{t,m,c} - \left(submot_{t,m} + pqusi_{t,m} + \sum_{j \in NUT_m} gtmin_{t,m,j} \right).fpeng_{t,c} \tag{18b}$$

Controllable energy balance equation in EER k and stage t

$$EA_{t+1,k} = FDIN_{t.k} EA_{t,k} + FC_{t,k} EC_{t,k} - GH_{t,c,k} - EVT_{t,k} - EVP_{t,k} - EDVC_{t,k} \tag{18c}$$

Wind power production through the MTFs in WF/EWF u and period t

$$\sum_{c=1}^{NPMC} \boldsymbol{GW_{t,u,c}} \leq \boldsymbol{b_{t,u}^{W}} + \boldsymbol{a_{t,u}^{W}} \boldsymbol{V_{t,u}} \tag{18d}$$

Set of multivariate linear constraints (Bender´s cut) representing the cost-to-go function

$$\alpha_{t+1} - \sum_{k \in NREE} \overline{\pi}_{EA_{1,t+1,k}} EA_{t+1,k} + \sum_{l=1}^{p} \overline{\pi}_{EI_{1,j,t+1,k}} EI_{t-l+1,k} \geq \overline{\delta}_{1,t+1} \tag{18e}$$

Other constraints considered in the problem are: (i) *for each EER*—uncontrollable (run-of-river) energy balance equation; losses in uncontrollable (run-of-river) inflows; minimum and maximum hydropower generation per load level; minimum outflow; water deviation for other uses, such as irrigation and water supply; storage capacity; minimum operational storage; (ii) *for subsystems*—minimum and maximum energy interchange limits between subsystems per load level; limits in a group of energy interchanges between subsystems; energy interchange balance in subsystems with no load nor generation capacity; (iii) *for each thermal plant*—minimum and maximum generation; (iv) *for LNG power plants*—total anticipated thermal generation in each load level.

In this formulation, a new constraint must be added that provides the wind power production through the MTFs (18d). The left-hand side of the power balance equation in each subsystem *m*, for each load level *c*, by stage t receives a new term that represents the sum of the wind power of the EWFs belonging to subsystem *m*, as shown in (18b); as mentioned before, the term *pqusi* now considers only the amounts of solar and biomass power production.

In the integrated model proposed in (6), the explanatory component can be the average of the wind speed stochastic process of the seasonal period *m*, or contain a portion related to the inflows of stage *t*, $EI_{t,i}$, or of stage *t-1*, $EI_{t-1,i}$. or also the monthly wind speed process could be represented by a PAR(1) model. As a result, each of these modeling options has a distinct impact on the construction of the Benders cuts relative to the state variable inflow to the EER_i in period *t-1*:

- if the explanatory component is the mean itself, there is no change in the Benders cuts;
- in the case that *EI* is included, the calculation of the Benders cut coefficient associated with $EI_{t-1,i}$ should be reviewed. A portion given by the partial derivative of the objective function with respect to $EI_{t-1,i}$. in (18d), must be added;
- if the monthly wind speed process is represented by a PAR(1) model, a new state variable must be included in stage *t*, $V_{t-1,j}$,and the calculation of the associated Benders cuts coefficient is given by the partial derivative of the objective function in relation to V_{t-1} in (18d).

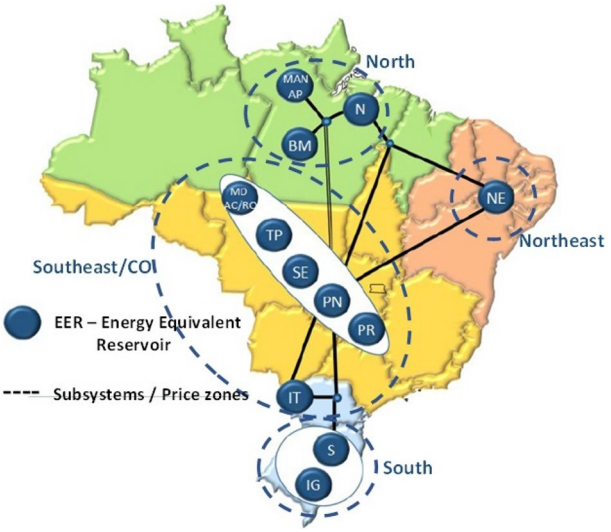

Fig. 5 Schematic representation of the Brazilian electric energy system

8 Application of the proposed methodology

The developed methodology was applied in real configurations of BIPS considering two key activities in the Brazilian power sector: (i) the operation planning, i.e., the Monthly Operation Program (MOP), carried out by ONS and CCEE; and (ii) auction for purchase new energy 4 years in advance (A-4), i.e., the calculation of the maximum amount of energy that can be traded in long-term Power Purchase Agreements (PPAs), carried by the Ministry of Mines and Energy (MME) together with the Energy Research Company (EPE) and the Electricity Regulatory Agency (ANEEL). In official studies, BIPS is divided in 4 interconnected subsystems/price zones and the hydropower configuration is represented by 12 EERs. Figure 5 shows a schematic representation of BIPS.

Initially, results from the integrated model for the generation of monthly multivariate synthetic sequences of inflows and winds as well as the evaluation of monthly transfer functions (MTFs) between wind speed and wind power production are presented and discussed. Then, case studies related to the impact of considering wind power uncertainties through the proposed methodology on the monthly operation program and on the calculation of the maximum amount of energy that can be traded in long-term PPAs are also presented and discussed.

8.1 Generation of monthly synthetic multivariate sequences of wind speeds and inflows

The approach described in Sect. 4 is illustrated by considering five EWFs (substations) located in five macro-regions (clusters) of wind regimes in Brazil, three of

them in the Northeast (*NE Interior*, *NE PE* and *NE Litoral*) and two in the South (*Sul Interior* and *Sul Litoral*). In this case, a sample of 37 years of monthly aggregated wind speeds measurements was considered together with normal correlated residuals with cardinality 2000.

Figure 6 shows the histograms of the historical and synthetic residuals of the average monthly wind speeds for two situations of high asymmetries, in two EWFs: (a) NE PE, in the month of August, with a skewness coefficient equal to 1.23; and (b) NE Interior, in June, with a skewness coefficient equal to -1.39. A successful fit of the Weibull distribution to the random residuals can be observed through the developed approach.

In turn, Fig. 7 compares, for the five EWFs, the monthly averages, standard deviations and skewness coefficients of the historical wind speeds with the ones produced by the synthetic series of wind speeds obtained with the proposed approach, using (8). Again, an excellent performance is observed, even for the skewness coefficients, thus confirming the successful fit of the Weibull distribution to the random residuals using the developed approach; this is mainly due to a special feature of the algorithm aiming to preserve in particular the skewness coefficient of the historical monthly wind speeds.

8.2 Monthly transfer functions

The procedure described in Section V was applied to each one of the 45 hub substations comprised in Sintegre system, obtaining hourly and monthly probabilistic power curves and estimating the associated MTFs. In this sense, time series of hourly wind speed reanalysis data from MERRA-2 were utilized, covering the period from 1980 to 2019 (40 years). Thus, the set of power curves available in Sintegre was applied for each hour h of each day d of the year 2019 (8760 curves) to the corresponding values of the wind speed reanalysis (i.e., at the same hour h and day d), in each year of the period 1980–2019.

Figure 8 shows the dispersion diagrams obtained with the Sintegre data (in red) and with the procedure developed (in blue) for the same five EWFs (substations) located in the Northeast (*NE Interior*, *NE PE* and *NE Litoral*) and South (*Sul Interior* and *Sul Litoral*) of Brazil. The Sintegre samples are the hourly wind power forecasts given by the system operator based on a set of numerical weather prediction providers which showed to be more disperse than those obtained from reanalysis data; furthermore, the developed procedure applies a power curve to the wind speeds from MERRA-2. As a consequence, the hourly probabilistic power curves obtained with the developed procedure are found within the Sintegre scatter diagrams, showing that the procedure using reanalysis data is reasonable.

The hourly values are then integrated to construct the monthly probabilistic power curves presented in Fig. 9. The scatter diagrams reveal high correlations (above 98.5%) between the monthly averages of wind speed and wind production in the analyzed EWFs—a typical behavior observed in other hub substations. This feature allows the construction of MTFs between monthly winds and power productions using linear regression models. The regression lines and corresponding

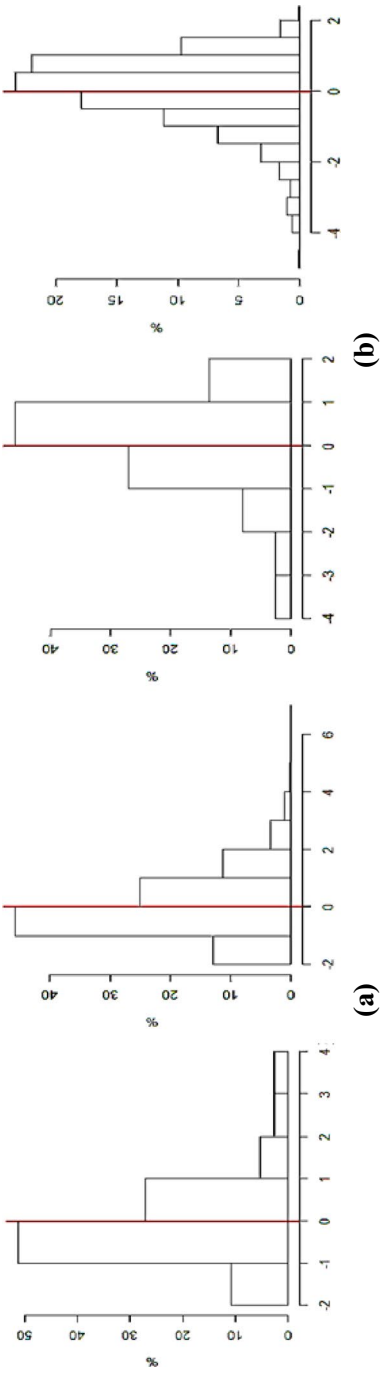

Fig. 6 A comparison of the historical and synthetic frequency distribution of random residuals for NE PE—August (**a**) and NE Interior—June (**b**)

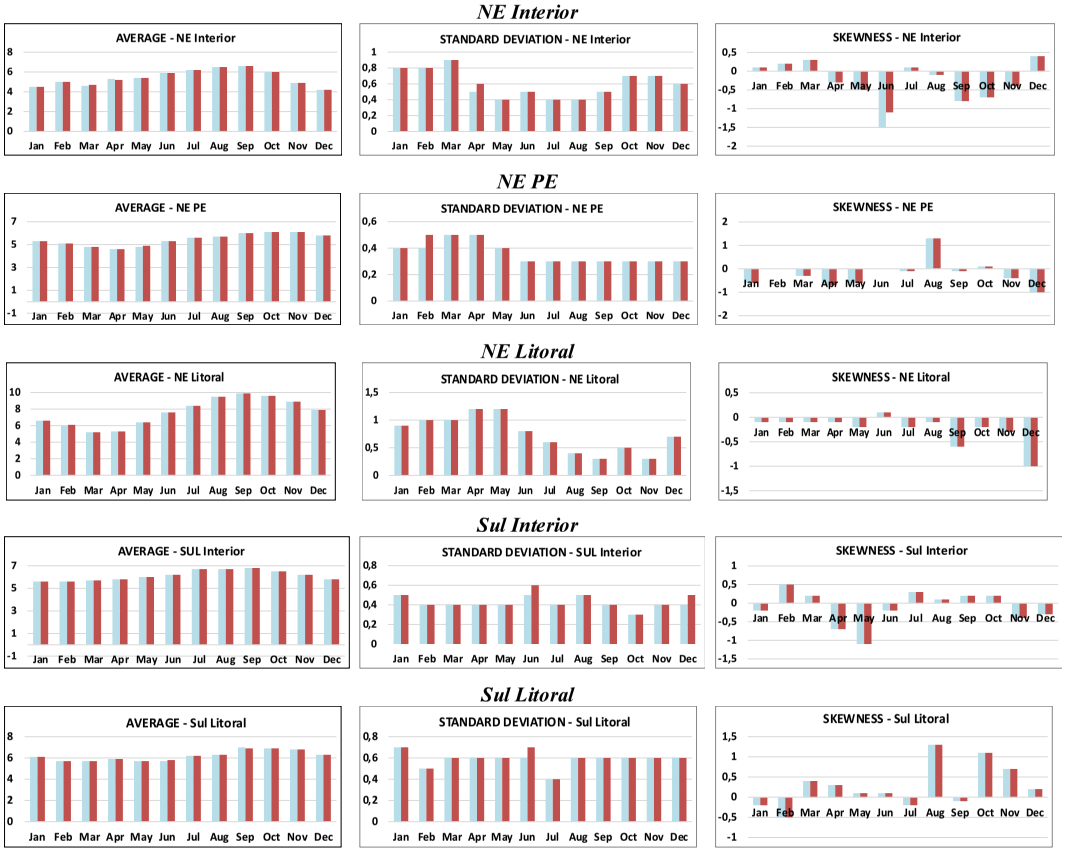

Fig. 7 Average, standard-deviation, skewness coeficient of monthly wind speed—historical (blue) and synthetic (red)

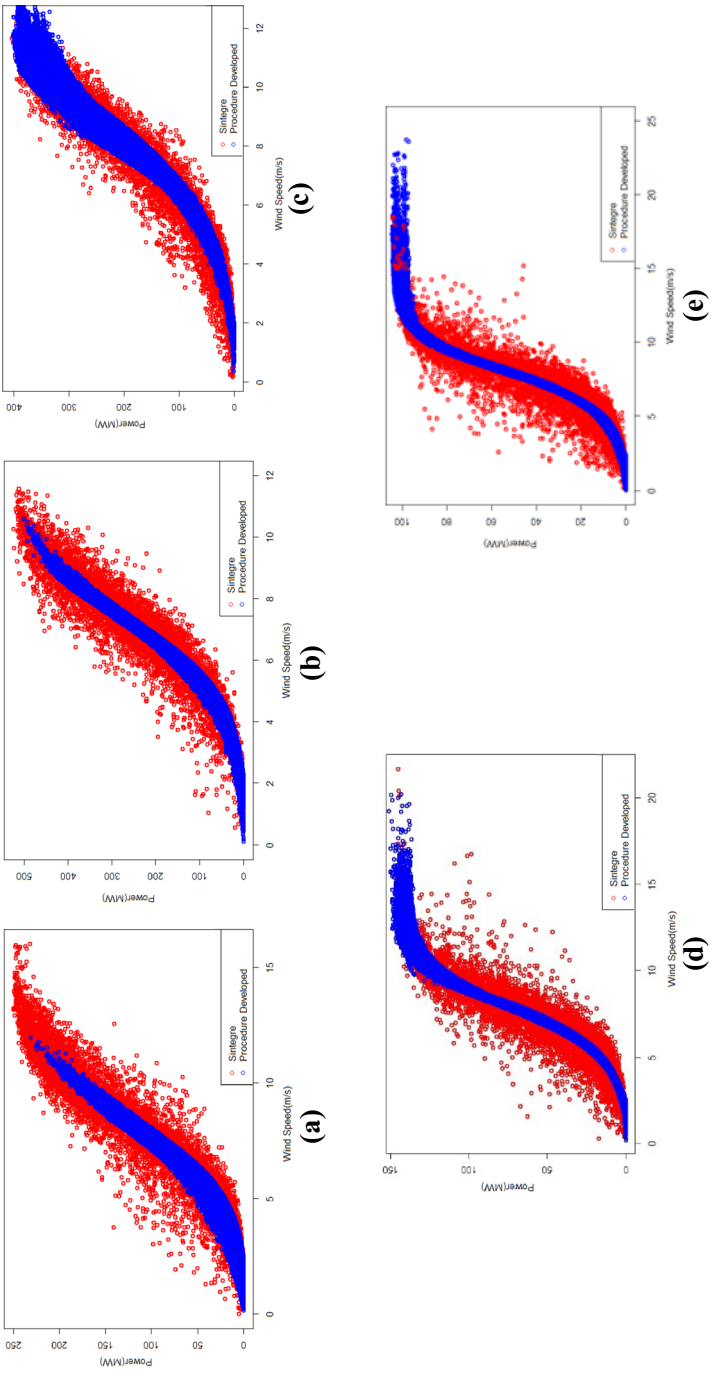

Fig. 8 Hourly probabilistic power curves with Sintegre data (red) and by the proposed procedure (blue) for five EWFs: Nordeste-Interior (**a**), Nordeste PE (**b**), Nordeste Litoral (**c**), Sul Interior(**d**) and and Sul Litoral (**e**)

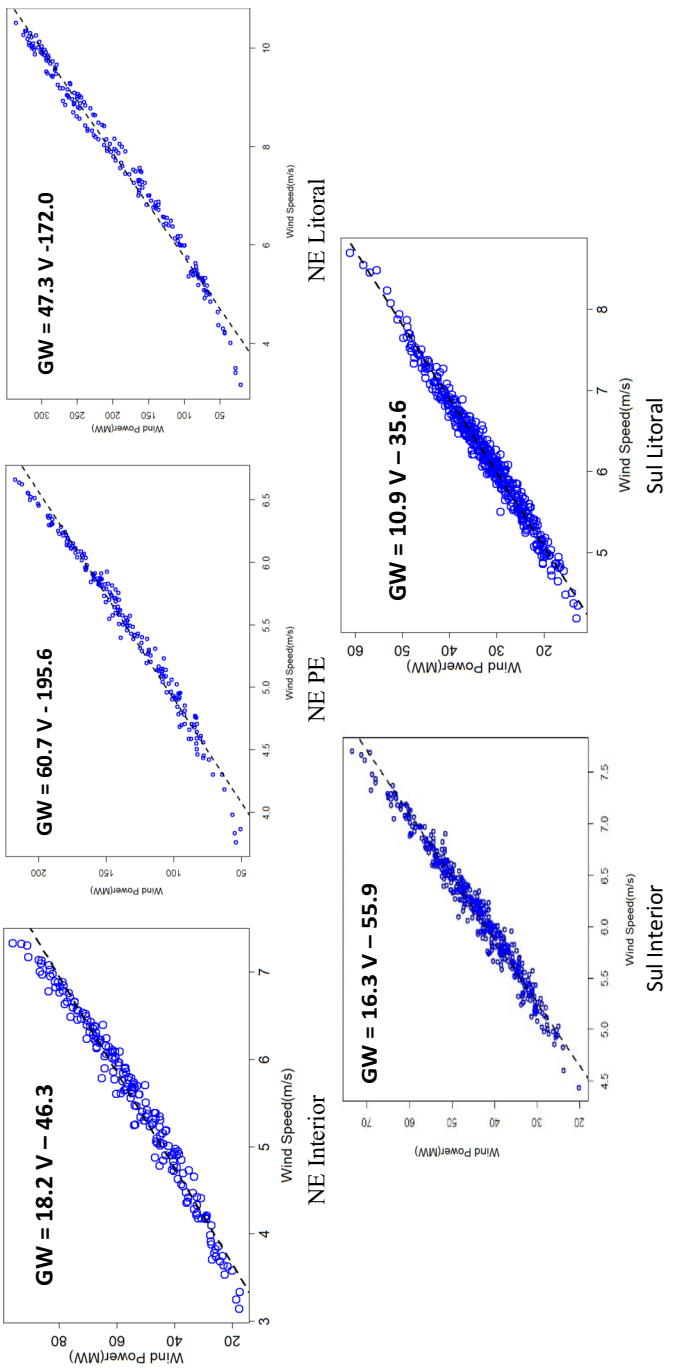

Fig. 9 Corresponding monthly probabilistic power curves and MTFs for 5 EWFs

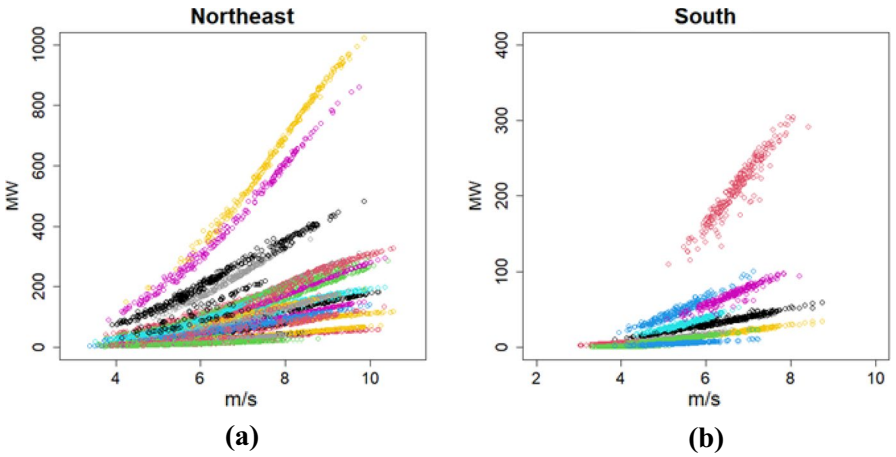

Fig. 10 Monthly probabilistic power curves for all 45 hub Sintegre hub substations (EWFs): 35 for the Northeast (**a**); 12 for the South subsystem (**b**)

equations are also shown in Fig. 9, where *GW* and *V* are the wind power production and the average monthly wind speed, for each EWF. It is important to check the seasonal behavior to see if there is a need to define, for each EWF, a single MTF valid for the whole year, or a set of MTFs, e.g., for each month of the year. For the set of EWFs presented, it was found that a single MTF would already be adequate.

By applying this procedure to all 45 Sintegre hub substations, we obtain the set of 35 and 12 monthly probabilistic power curves represented in Fig. 10 for the Northeast and South regions, respectively. Again, linear regressions showed to be adequate to construct the associated MTFs.

The definition of the final granularity of the EWF clusters and thus MTFs is underway by the Standing Committee for Analysis of Methodologies and Computational Programs of the Electric Sector—CPAMP.

In the case of adopting a smaller granularity, for example, considering the grouping of wind regimes in regions with large geographic coverage, it is also necessary to aggregate the MTFs of the EWFs belonging to each of these macro-regions. If the MTFs of each substation are represented by linear regressions, the aggregation of MTFs can be done by the sum of the angular and linear coefficients of the MTFs of each EWF.

8.3 Application to the operation planning

In this session, three cases based on the Monthly Operation Program (MOP) are studied, which will be described below. The MOP configuration comprises 162 hydropower plants disposed in 12 EERs, 121 thermal power plants distributed in 4 subsystems and price zones. Table 2. shows the installed hydro and thermal capacity for each subsystem. The operation planning horizon is 5 years and considers the evolution of the system configuration and demand along these years. MOP is carried

Table 2 Total installed capacity for each subsystem / price zone

	Hydropower (MW)	Thermal power (MW)
SE	60,842	10,503
South	15,144	2957
NE	10,831	7009
North	22,074	4117
Total	108,890	24,586

out by ONS and CCEE using NEWAVE every month, and weekly reviewed using a short-term operation planning model, followed by a daily operation programming.

The wind power is concentrated in Northeast and South subsystems. The temporal correlation structure that can be verified in the stochastic process of the monthly average wind speeds in any EWF will not be explicitly considered in the synthetic series wind generation model, and are represented indirectly, through of the spatial correlation verified between the stochastic processes of winds in EWFs and inflows in EERs, i.e., (8) is applied here. Consequently, no state variable will be added to the SDDP algorithm, so there is no addition of the FCF cardinality.

8.3.1 Comparison between cases considering and not considering wind uncertainty

The reference case (*without_uncertainty*) seeks to emulate the current procedure approved by the Regulator (ANEEL), where the average wind power is represented in NEWAVE as non-dispatchable plants, i.e., with constant power production. In this sense, the average monthly wind power productions of the Northeast (NE) and South (Sul) subsystems were determined by applying the MTFs of the 5 referred EWFs to their respective historical monthly wind speeds. The results are presented in Table 3.

In the case called *wind_uncertainty*, the information in Table 3 was suppressed. The wind power uncertainties were modeled by using the proposed approach. Initially, synthetic sequences of monthly average wind speeds in the EWFs are generated for the backward and forward passes of the SDDP algorithm, and for the

Table 3 Monthly wind power production average, obtained from the MTFs (MWmonth)

	Jan	Feb	Mar	Abr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
NE Interior	2086	2411	2183	2591	2706	3012	3232	3442	3456	3093	2344	1927
NE PE	278	258	225	215	233	276	300	312	335	347	343	314
NE Litoral	2146	1743	1223	1296	1967	2801	3411	4193	4443	4230	3773	3067
Subsystem NE	4509	4412	3631	4102	4907	6089	6943	7947	8234	7670	6460	5308
Sul Interior	94	94	95	98	101	107	116	116	119	112	106	98
Sul Litoral	611	533	537	565	536	544	621	639	747	738	713	641
Subsystem Sul	705	627	632	663	637	650	737	755	866	850	819	739

final simulation. Then they are transformed into synthetic sequences of wind power through the MTFs, which are also explicitly used in each operation dispatch problems, according to Sect. 7.

a) Expected total operation cost, annual deficit risk and annual EENS

For the *without_uncertainty* and *wind_uncertainty* cases, Table 4 shows the expected total operating costs, the annual deficit risks and expected energy not supplied (EENS) for the first 2 years of the planning horizon.

It is observed that there was a reduction in the expected total operation cost of 2.2% (i.e., R$ 556 million), when the uncertainty of wind speeds and associated wind power are explicitly modeled as proposed. On the other hand, the annual deficit risks were a little higher, but with lower expected values of EENS. The explicit consideration of the variability of the wind power source together with the complementarity between the hydro and wind power production allows the NEWAVE model to better optimize the operation strategy, thus reducing the operating cost and reflecting in the decrease in EENS. On the other hand, those sequences presenting very low wind power contribute to the increase of the annual risk of deficit, a relevant result for the system operator.

b) Frequency distributions of the synthetic wind power production

The aspects mentioned above are corroborated by Fig. 11. This figure shows, for the months of March and September (corresponding to wet and dry hydrological seasons, respectively), the frequency distributions of power production resulting from the synthetic sequences of wind speeds, which presents reasonable dispersion. Additionally, the single value of wind power considered in the current procedure (*without_uncertainty* case) is depicted.

c) Time evolution of the expected hydropower production

Figure 12 illustrates the evolution, over the planning horizon, of the expected hydropower generation of EERs NE and SE (the first located in the Northeast (NE) subsystem and the other in the Southeast (SE) subsystem). It is noted that the consideration of wind uncertainties impacts the optimization of the system, leading to differences in the expectation of hydroelectric generation, reaching values of up to 2,500 MW/month.

d) Frequency distributions of hydro and thermal power production

To better identify differences in the behavior of hydro and thermal power production, Fig. 13a, b, e, f present the cumulative frequency distributions of hydro and thermal power generation in EERs SE and NE on March and September for both case studies. For example, although there are differences in the histograms, they are not so high, which is in line with the difference in the expected value of the total

Table 4 Expected Total Operation Cost, Annual Deficit Risk, Annual EENS

Year	Southeast		South		Northeast		North	
	RISK (%)	EENS (MWmonth)	RISK (%)	EENS (MWmonth)	RISK (%)	EENS (MWmonth)	RISK (%)	EENS (MWmonth)
Without_Uncertainty Case								
Expected Total Operation Cost (R$ billion) = 25.790								
2021	0.2	1.4	0.1	0.2	0.0	0.0	0.0	0.0
2022	0.1	0.4	0.1	0.3	0.0	0.0	0.0	0.0
Uncertainty_Wind Case								
Expected Total Operation Cost (R$ billion) = 25.234								
2021	0.3	0.5	0.1	0.2	0.0	0.0	0.0	0.0
2022	0	0	0.1	0.1	0.0	0.0	0.0	0.0

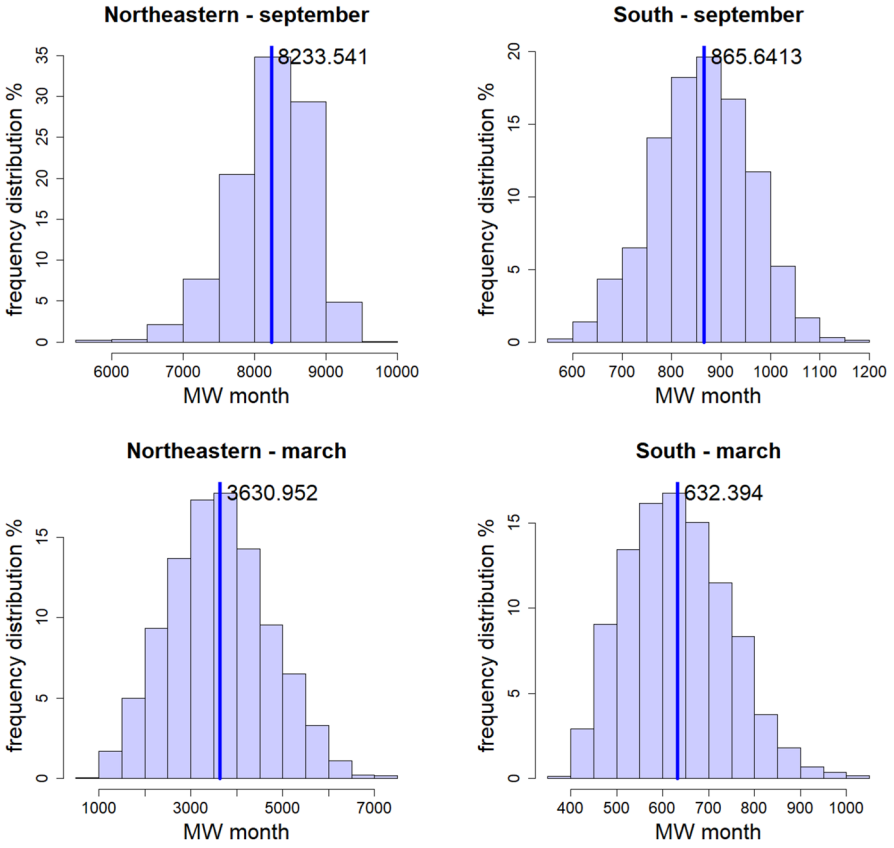

Fig. 11 Frequency distributions of the synthetic wind power production

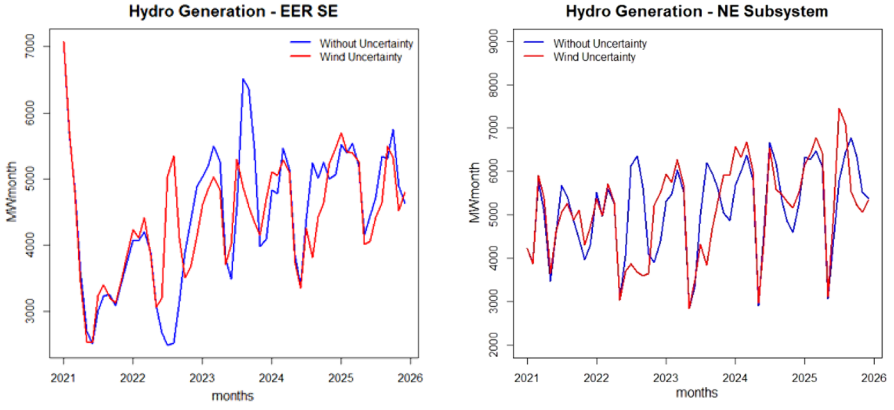

Fig. 12 Time evolution of EER-NE and EER-SE expected hydro generation

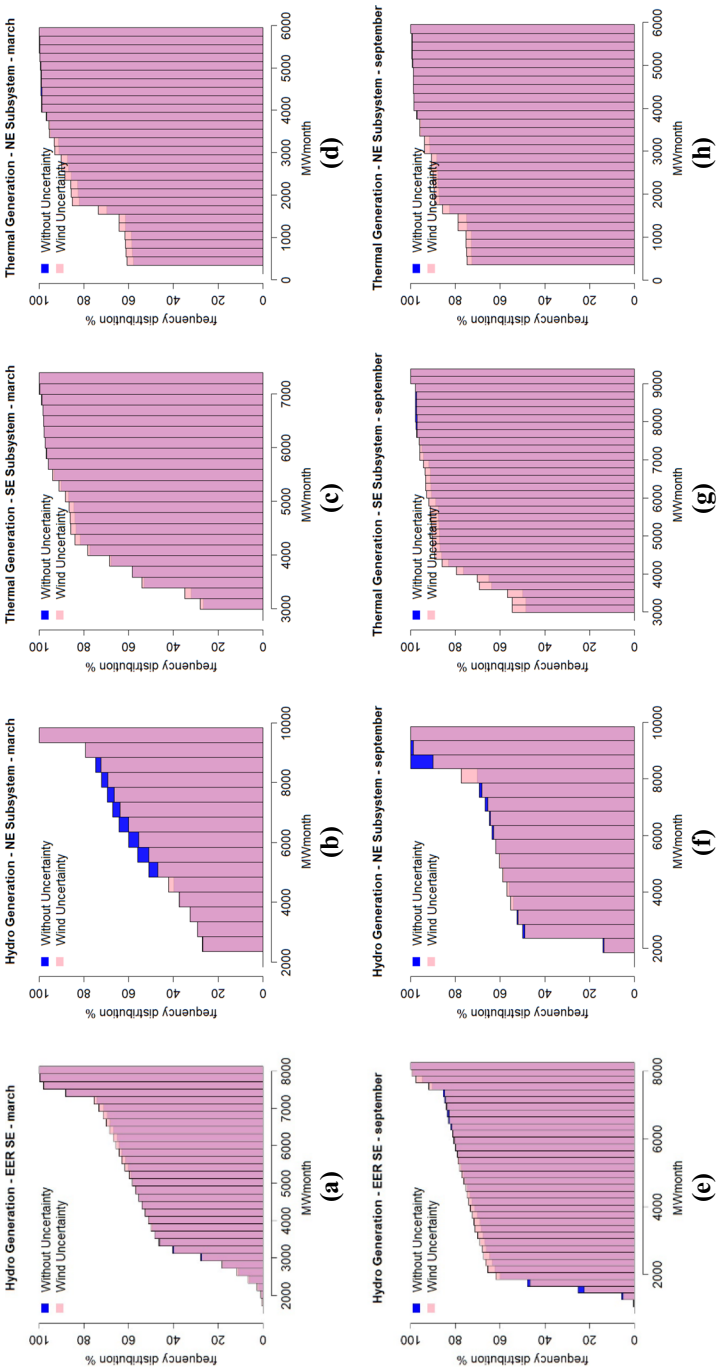

Fig. 13 Hydro generation for SE and NE EERs and thermal generation for SE and NE Subsystems—cumulative frequency distribution for without_uncertainty and wind_uncertainty cases

operation cost of 2%. In the SE subsystem, in March, there is a slight increase in the frequency of higher generation values; in September, there is a slight increase in the frequency of lower generation values, which implies reaching higher reservoir levels at the end of the dry season.

Regarding the thermal generation (Fig. 13c, d, g, h), in September, a month with low inflows in the Southeast and Northeast subsystems, a slightly higher frequencies of lower values of thermal generation are observed, when representing the uncertainty in the winds. This behavior justifies the decrease in the expected total operation cost. In March, a month with high inflows in these two subsystems, the same behavior is observed, but more attenuated. This variation is not so significant, since the thermal generation in BIPS is of the base load type, presenting little variation between the minimum and maximum values.

e) Operation marginal costs

Figure 14a, d shows the evolution, over the planning horizon, of the expected operation marginal costs (OMC) for the Southeast and Northeast subsystems; it can be observed that in general, the expected OMC is lower when wind speed uncertainties are taken into account, and this difference is more prominent in dry months. In turn, Fig. 14b, c present the OMC frequency distribution in March and September for the Southeast (SE) subsystem while Fig. 14e, f) show the same figures for the Northeast subsystem. As the month of September belongs to the dry hydrological season, the amplitude of OMC values is greater than that of the month of September (wet season) in both subsystems and in all case studies. However, for both subsystems and analyzed months, it can be seen that the frequency of lower OMC values increased in the *wind_uncertainty* case compared to the *without_uncertainty* case, proving the benefits of considering uncertainty in wind speeds and, therefore, in the wind power production.

8.3.2 Sensitivity analysis

To analyze the impact of a more accelerated penetration of wind energy, a sensitivity analysis was carried out, considering a third case study, where a 20% increase in the installed capacity of wind power was implemented, being denoted as *wind_uncertainty + 20% in capacity* case.

a) Expected total operation cost

The expected total operating costs in the *wind_uncertainty + 20% in capacity* case was R$ 23.555 billion, that is, 10.6% (R$ 2.735 billion) less than the *wind_uncertainty* case or 5 times greater than the reduction obtained in the *without_uncertainty* case. Thus, it is to be expected that the impacts on hydro and thermal generation, and also on OMCs, will be pronounced.

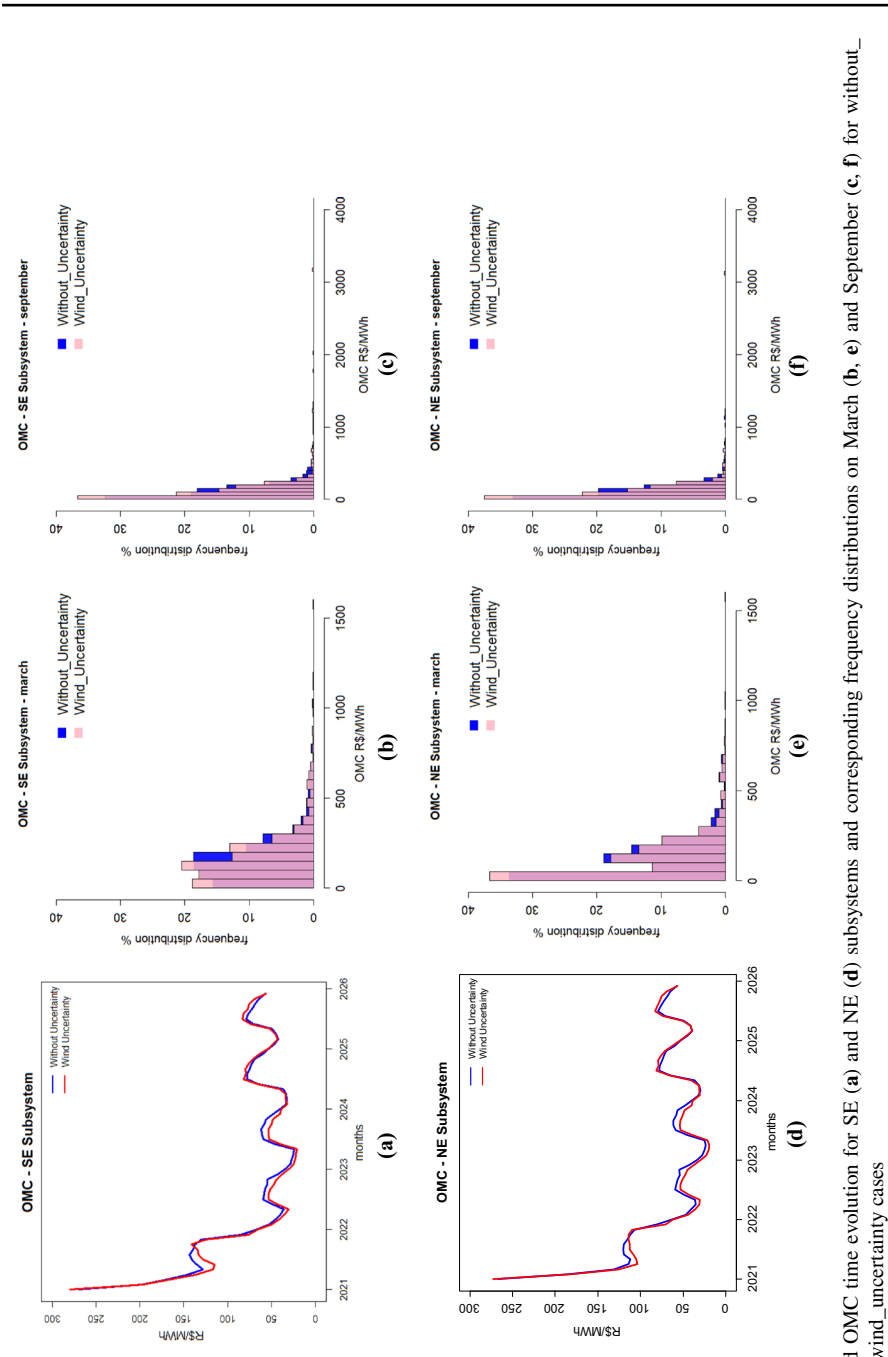

Fig. 14 Expected OMC time evolution for SE (**a**) and NE (**d**) subsystems and corresponding frequency distributions on March (**b**, **e**) and September (**c**, **f**) for without_uncertainty and wind_uncertainty cases

b) Frequency distributions of hydro and thermal power production

Figure 15 presents the cumulative frequency distributions of hydropower and thermal generation in EERs SE and NE on March and September for the *without_uncertainty* and *wind_uncertainty + 20% in capacity* cases. In general, the conclusions obtained for the *without_uncertainty* and *wind_uncertainty + 20% in capacity* case are the same as those for the *wind_uncertainty* case, but with much more pronounced differences in relation to the *without_uncertainty* case.

When analyzing Fig. 15a, b, e, f, one aspect worth highlighting is the increase in the frequency of high hydropower generation values for the NE subsystem in September (hydrologically dry season); this is probably due to the predominance of wind power capacity and the negative correlation between the hydrological regime and the wind speed in this region, meaning that when reaching lower reservoir levels at the end of the hydrological dry season (ie., storing energy in other subsystems) minimize the chances of spillage in the following season (wet), where wind speeds are lower.

Regarding the thermal generation (Fig. 15c, d, g, h), in September, a month with low inflows in the Southeast and Northeast subsystems, higher and meaningful frequencies of lower values of thermal generation are observed, when representing the uncertainty in the winds and considering a 20% increase in wind power installed capacity. This behavior justifies the greater decrease in the expected total operation cost.

c) Operation marginal costs

Figure 16a, d shows the evolution, over the planning horizon, of the expected operation marginal costs (OMC) for the Southeast and Northeast subsystems. It can be observed that, in general, the expected OMC in the wind_uncertainty + 20% capacity case is smaller than in the *without_uncertainty* case, and that this difference is much larger when compared to the *wind_uncertainty* case; again, the greatest differences occur in the dry months.

In turn, Fig. 16b, c present the OMC frequency distribution in March and September for the Southeast subsystem while Fig. 16e, f show the same figures for the Northeast subsystem. Again, for both subsystems and analyzed months, it can be seen that the frequency of lower OMC values increases when wind speed uncertainties are considered, and that this increase is much higher in case *wind_uncertainty + 20% in capacity*, compared to *wind_uncertainty* case.

These results highlight the importance of considering wind speed uncertainties in the long-term operation planning and points out that the representation of such uncertainties becomes more relevant with the more intense penetration of the wind power into the system.

8.4 Application to the long-term commercialization

The introduction of competition for the long-term market in the Brazilian 2004 Electrical Sector Reform was a milestone towards the creation of a more stable investment

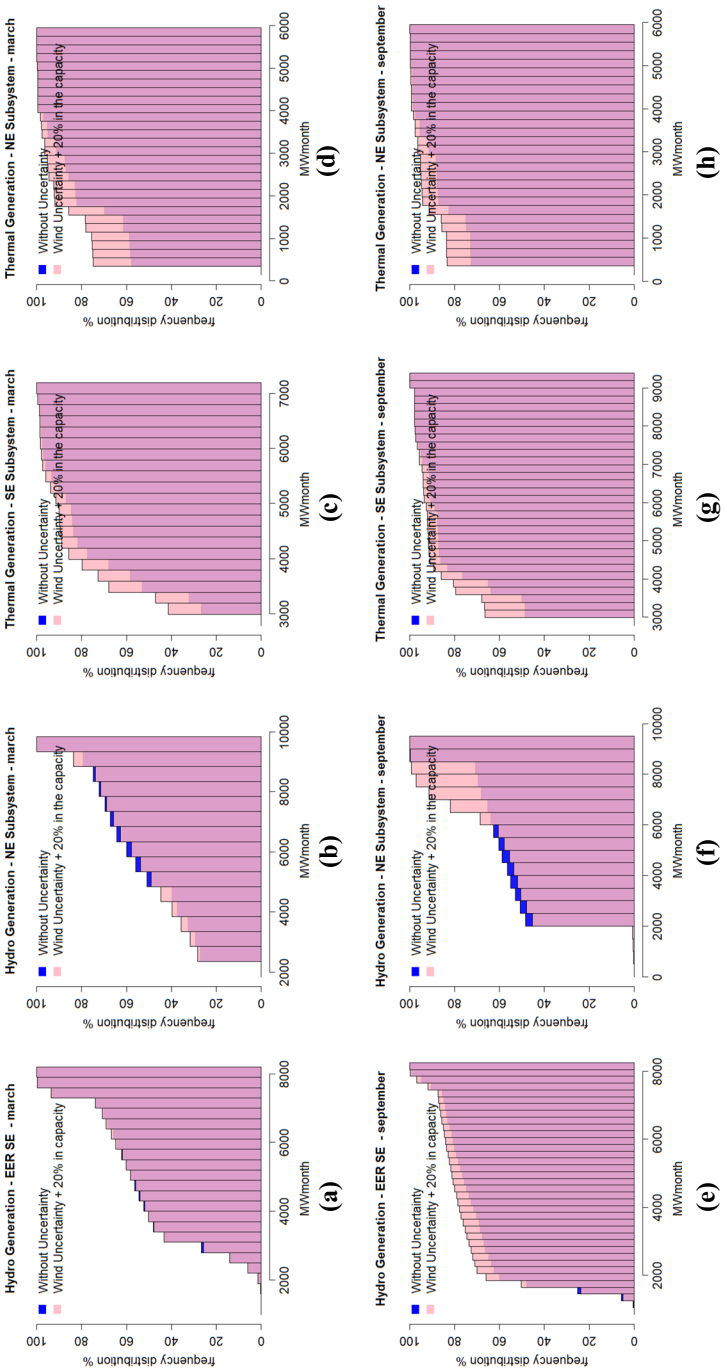

Fig. 15 Hydro generation for SE and NE EERs and thermal generation for SE and NE Subsystems—cumulative frequency distribution for without_uncertainty and wind_uncertainty + 20% in capacity cases

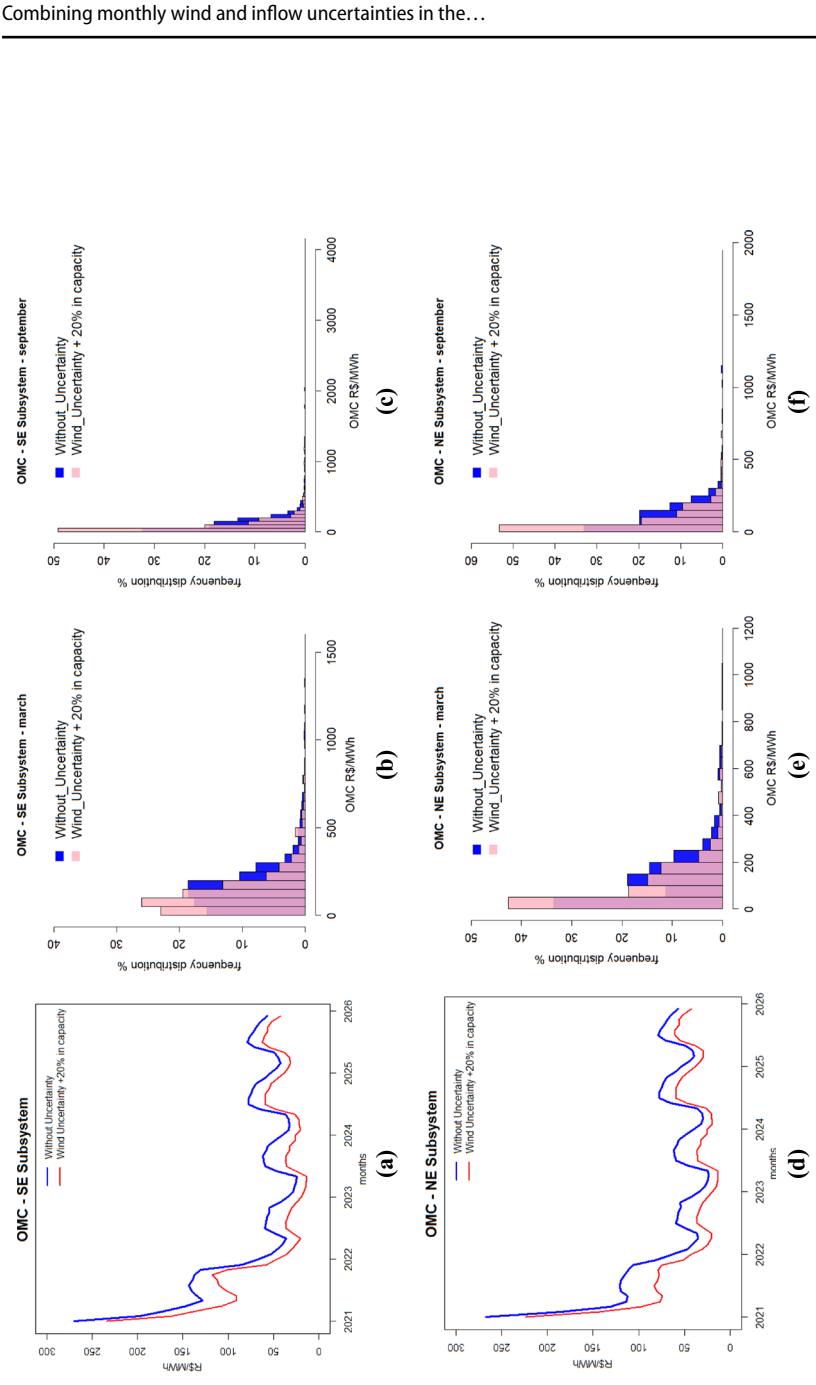

Fig. 16 Expected OMC evolution for SE (**a**) and NE (**d**) subsystems and corresponding frequency distributions on March (**b**, **e**) and September (**c**, **f**) for without_uncertainty and wind_uncertainty + 20% in capacity cases

environment for new generation capacity. Loads have to be 100% contracted and two environments for electricity trading were established: (i) a Free Contracting Environment, where free consumers can procure their energy needs as they wish, as long as they are 100% contracted; and (ii) a Regulated Contracting Environment, where generators must participate in centralized public auctions to be able to sign power purchase agreements (PPAs) with the regulated (captive) consumers supplied by the distribution companies, which must provide self-declaration of its forecasted loads for the next five years. In addition, the differences between the production and consumption of energy in relation to the contracts held are settled on the spot market by the spot prices (called Settlement Prices for Differences—PLDs) [45].

A question that arises is what is the maximum amount of energy that a power plant can trade in the long-run. In Brazil this is called *assured energy* and is calculated by a specific procedure taking into account the overall system optimization [43], summarized as follows. Initially, the *total assured energy* of BIPS (TAE), that corresponds to the maximum energy demand that the system could supply, is obtained through a simulation of the system operation provided by NEWAVE where the hydroelectric plants configuration is represented by EERs. The total system assured energy is the result of a procedure in which the energy demand is changed iteratively until the energy supply adequacy criteria are met. These criteria are defined by the Brazilian National Energy Policy Council (CNPE) and comprises the following requirements [46, 47]: (a) the annual expected value of the marginal operation cost (MOC) = the marginal expansion cost for each subsystem; (b) the conditional value at risk of the energy not supplied at 99% confidence level ($CVaR_{99\%}(ENS)) \leq 5\%$ annual energy demand for BIPS and subsystems; and (c) the conditional value at risk of the monthly marginal operation cost at 90% confidence level ($CvaR_{90\%}(MOC)) \leq$ R$ 800/MWh, for each subsystem.

At the end of the iterative process, the total assured energy is divided in two parts—a hydro block and a thermal block, based on the expected generation of the EERs and thermal plants, respectively. Then the hydro block is further allocated among individual HPPs using a specific approach [43].

In this section, the application is focused on the impact of considering wind speed uncertainties on the total assured energy calculation. To achieve this, two case studies were considered: a reference case (*TAE-without_uncertainty*) that seeks to emulate the current approved procedure, where the average wind power is represented in NEWAVE as non-dispatchabled plants; and the *TAE-wind_uncertainty* case, in which the wind power uncertainties were modeled by using the proposed approach.

The time horizon of this study is 5 years and a single future hydrothermal system configuration, considering new hydroelectric and thermal generators, is considered for the calculation of the total assured energy. The seasonality of demand is taken into account and a period of 10 years is added before the planning period so that the system loses memory of the initial stores in the EERs and the initial hydrological conditions. The marginal expansion cost for these case studies was R$ 90.38/MWh.

In the *TAE-without_uncertainty* case, the total assured energy to attain the CNPE criteria was 86,200 MWaverage/month. When considering the wind speed uncertainties, although criteria (b) e (c) were met, the expected MOC (R$ 81.17/MWh) was below the marginal expansion cost, meaning that there is room for BIPS meet a higher energy demand, i.e., for a higher total assured energy. As

a consequence, the iterative procedure is applied until the three criteria are achieved, resulting in a total assured energy for the *TAE-wind_uncertainty* case equals to 86,631 MWaverage/month, i.e., an increase of 0.5% (431 MWaverage/month) with respect to the *TAE-without_uncertainty* case. At first glance, this increase appears to be small. However, if we consider that the amount of assured energy can be traded in the long-term, with PPAs from 20 to 35 years, the economic benefit may not be negligible. For example, considering a 30-year contract priced at the marginal cost of expansion (R$ 90.38/MWh), 450 MWaverage/month represents the value of approximately US$ 24 billion. If the sale prices of the last auction for purchase electricity for wind power and hydropower are considered, the financial values become US$ 31 billion and US$ 47 billion respectively.

9 Conclusions

Following a world trend, Brazil is experiencing an accelerated growth of wind energy. The current representation of wind power production in the expansion and operation planning should be improved to consider the wind power uncertainties.

The objective of this work was to describe an approach to be used by the Brazilian power industry to represent the uncertainties of monthly wind power production in the SDDP algorithm applied in the medium and long-term operation planning model in Brazil. Due the dimensions of the Brazilian interconnected power system and to the hydropower predominance, attention is paid to keeping the large-scale stochastic problem computationally viable.

The proposed methodology comprises statistical clustering of wind regimes and definition of equivalent wind farms; evaluation of monthly transfer functions between wind speed and power production; integrated generation of monthly multivariate synthetic scenarios of inflows and winds, considering associated cross-correlations; and representing monthly wind power in the SDDP algorithm.

Each step of the proposed approach was applied in real configurations of the Brazilian interconnected power system including case studies related to the impact of considering wind power uncertainties on the monthly operation program and on the calculation of the maximum amount of energy that can be traded in long term power purchase agreements. The results obtained so far points to effectiveness of the proposed methodology and the relevance of modeling the wind uncertainties in the long-term operation planning of large hydro-dominated systems.

Further developments include the extension of the described approach to consider the uncertainties of photovoltaic solar energy production, which also has a high growth in the Brazilian system.

Acknowledgements The first two authors have led the development of the presented methodology when they were researchers at CEPEL, and the development has continued at UERJ. The authors are grateful for the support and discussions of Methodology WG of the Brazilian Standing Commission for the Analysis of Methodologies and Computational Programs in the Electricity Sector – CPAMP. This work was partially funded by CNPq/MCTI/FNDCT through Research Project 409715/2021-2.

Data availability Data available on request from the authors.

Declarations

Conflict of interest Conflict of interest the authors declare no conflict of interest regarding the publication of this work.

References

1. Ministry of Mines and Energy / Energy Research Company: Ten-Year Energy Expansion Plan 2022–2031 (2022)
2. Hamidpour, H., Aghaei, J., Pirouzi, S., Niknam, T., Nikoobakht, A., Lehtonen, M., Shafie-khah, M., Catalão, J.P.S.: Coordinated expansion planning problem considering wind farms, energy storage systems and demand response, Energy (ELSEVIER), Vol. 239 (2022)
3. Bessa, R.J., Miranda, V., Botterud, A., Zhou, Z., Wang, J.: Time adaptive quantile-copula for wind power forecasting. Renew. Energy **40**, 1–11 (2011)
4. Baptista, D., Carvalho, J.P., Morgado-Dias, F.: Comparing different solutions for forecasting the energy production of a wind farm. Neural Comput. Appl. **32**, 15825–15833 (2020)
5. Fosso, O.B., Gjelsvik, A., Haugstad, A., Mo, B., Wangensteen, I.: Generation scheduling in a deregulated system. The Norwegian case. IEEE Trans. Power Syst. **14**, 75–81 (1999)
6. Maceira, M.E.P., Terry, L.A., Costa, F.S., Damázio, J.M., Melo, A.C.G.: Chain of optimization models for setting the energy dispatch and spot price in the Brazilian System, 14th power systems computation conference—PSCC. Seville, Spain (2002)
7. Helseth, A., Melo, A.C.G.: Scheduling toolchains in hydro-dominated systems—evolution, current status and future challenges for Norway and Brazil, SINTEF Energy Research Technical Report, 2020–08–10 (2020)
8. Maceira, M.E.P, Penna, D.D.J., Diniz, A.L., Pinto, R.J., Melo, A.C.G., Vasconcellos, C.V., Cruz, C.B.: Twenty years of application of stochastic dual dynamic programming in official and agent studies in brazil—main features and improvements on the NEWAVE model. 20th Power systems computation conference—PSCC, Dublin, Ireland (2018)
9. Pereira, M.V.F., Pinto, L.: Multi-stage stochastic optimization applied to energy planning. Math. Programming **52**(1–3), 359–375 (1991)
10. Maceira, M.E.P., Melo, A.C.G., Pessanha, J.F.M., Cruz, C.B., Almeida, V.A., Justino, T.C.: Uma Abordagem para a Representação das Incertezas da Fonte de Geração Eólica no Modelo Newave, Cadernos do IME—Série Estatística, ISSN on-line 2317–4535 / ISSN impresso 1413–9022 - v. 48, p.1 – 36, 2020, https://doi.org/10.12957/cadest.2020.55395 (In Portuguese) (2020)
11. Maceira, M.E.P., Melo, A.C.G., Pessanha, J.F.M., Cruz, C.B., Almeida, V.A., Justino, T.C.: Wind uncertainty modeling in long-term operation planning of hydro-dominated systems. In: 17th IEEE International Conference on Probabilistic Methods Applied to Power Systems—PMAPS, Online (2022)
12. Arvantidis, N.V., Rosing, J.: Composite representation of multireservoir hydroelectric power system. IEEE Trans. Power Appar. Syst. **89**(2), 319–326 (1970)
13. Terry, L.A., Pereira, M.V.F., Araripe Neto, T.A., Silva, L.F.C.A., Sales, P.R.H.: Coordinating the Energy Generation of the Brazilian Hydrothermal Electrical Generating System. Interfaces, 16 (1986)

14. Maceira, M.E.P., Duarte, V.S., Penna, D.D.J., Tcheou, M.P.: An Approach to Consider Hydraulic Coupled Systems in the Construction of Equivalent Reservoir Model in Hydrothermal Operation Planning, 17th Power Systems Computation Conference—PSCC, Stockholm, Belgium (2011)
15. Maceira, M.E.P., Cruz, C.B., Penna, D.D.J., Diniz, A.L., Melo, A.C.G.: Combined Representation of Hydropower Plants and Inflow Scenarios Re-Sampling on Stochastic Dual Dynamic Programming—Application to the Brazilian System, 15th International Conference on Stochastic Programming, Trondheim, Norway (2019)
16. Salas, J.D., Delleur, J.W., Yevjevich, V., Lane, W.L.: Applied Modeling of Hydrologic Time Series, Water Resources Publications (1980)
17. Hippel, W.H., Mcleod, A.I.: Time series modeling of water resources and environmental system. ELSEVIER (1994)
18. Franses, P.H., Paap, R.: Periodic Time Series Models, Oxford University Press (2006)
19. Birge, J.R.: Decomposition and partitioning methods for multistage stochastic linear programs. Oper. Res. **33**(5), 989–1007 (1985)
20. Amarante, O.A.C., Schultz, D.J., Bittencourt, R.M., Rocha, N.A.: Wind/Hydro complementary seasonal regimes in Brazil. DEWI Mag **19**, 79–86 (2001)
21. Johnson, R.A., Wichern, D.W.: Applied multivariate analysis. Prentice Hall, New Jersey (1998)
22. Staffell, I., Pfenninger, S.: Using bias-corrected reanalysis to simulate current and future wind power output. Energy **114**, 1224–1239 (2016)
23. Energy Research Company—EPE: Acompanhamento de medições anemométricas—AMA: Caracterização do Recurso Eólico e Resultados Preliminares de sua Aplicação no Sistema Elétrico, Nota Técnica DEA 15/13, Rio de Janeiro (In Portuguese) (2013)
24. Pessanha, J.F.M., Almeida, V.A., Melo, A.C.G.: Fator de capacidade da geração eólica na região Nordeste: um estudo com dados de reanálises oriundos do MERRA-2 (global), VIII Simpósio Brasileiro de Sistemas Elétricos, Santo André, Brasil (In Portuguese) (2020)
25. Maceira, M.E.P., Bezerra, C.V.: Stochastic Streamflow Model for Hydroelectric Systems, 5th Probabilistic Methods Applied to Power Systems—PMAPS, Vancouver, Canada (1997)
26. Penna, D.D.J., Maceira, M.E.P., Damázio, J.M.: Selective sampling applied to long-term hydrothermal generation planning, 17th Power system computation conference—PSCC. Stockholm, Sweden (2011)
27. Kelman, J., Barth, F.T., Pompeu, C.T., Fill, H.D., Tucci, C.E.M., Braga JR., B.P.F.: Models for Water Resources Management. ABRH Publications (In Portuguese) (1987)
28. Charbeneau, R.J.: Comparison of the two and three parameter lognormal distributions used in streamflow synthesis. Water Resour. Res. **14**(1), 149–150 (1978)
29. Weibull, W.A, Statistical Theory of Strength of Materials: Ingeniors Vetenskaps Akademiens Handlingar (1939)
30. Cohen, A.C., Whitten, B.J.: Modified maximum likelihood and moment estimators for the three-parameter Weibull distribution. Communications in Statistics—Theory and Methods, 11 (1982)
31. Cousineau, D.: Fitting the three-parameter Weibull distribution: review and evaluation of existing and new methods. IEEE Trans. Dielectrics Electr. Insulation **16**, 281–288 (2009)
32. Melo, A.C.G., Maceira, M.E.P., Pessanha, J.F.M.: Adjustment of tri-parametric Weibull distributions with high skewness in SDDP: MSI UERJ Review—Statistical Series, v. 49 (In Portuguese) (2020)
33. Almeida, V.A., Pessanha, J.F.M., Melo, A.C.G., Maceira, M.E.P.: Modeling the monthly relationship between wind speed and wind power in Newave model, LIII Simpósio Brasileiro de Pesquisa Operacional—SBPO (In Portuguese) (2021)
34. Brazilian National Electric System Operator—ONS: Portal Sintegre, at: https://sintegre.ons.org.br/sites/6/27/48
35. Alessandrini, S., Sperati, S., Pinson, P.: The influence of the new ECMWF Ensemble Prediction System resolution on wind power forecast accuracy and uncertainty estimation. Adv. Sci. Res. **8**, 143–147 (2012)
36. Parker, W.S.: Reanalysis and observations: what´s the difference?, Bulletin of the American Meteorological Society, Vol. 97, No. 9 (2016)
37. Pfenninger, S., Staffell, I.: Long-term patterns of European PV output using 30 years of validated hourly reanalysis and satellite data. Energy - Elsevier **114**, 1251–1265 (2016)
38. Philpott, A.B., Matos, V.L.: Dynamic sampling algorithms for multi-stage stochastic programs with risk aversion. Eur. J. Oper. Res. **218**, 470–483 (2012)

39. Diniz, A.L., Tcheou, M.P., Maceira, M.E.P.: A direct approach to the consideration of the CVaR problem hydrothermal operation planning, XII Symposium of Specialists in Electric Operational and Expansion Planning - SEPOPE (In Portuguese) (2012)
40. Shapiro, A., Tekaya, W., Costa, J.P., Soares, M.P.: Risk neutral and risk averse Stochastic Dual Dynamic Programming method, European Journal of Operational Research, vol . 224 , n.2 (2013)
41. Maceira, M.E.P., Marzano, L.G.B., Penna, D.D.J., Diniz, A.L., Justino, T.C.: Application of CVaR risk aversion approach in the expansion and operation planning and for setting the spot price in the Brazilian hydrothermal system, 18th PSCC. Wroclaw, Poland (2014)
42. Diniz, A.L., Maceira, M.E.P.: Multi-lag Benders decomposition for power generation planning with nonanticipativity constraints on the dispatch of LNG thermal plants, Stochastic Programming—Applications in Finance, Energy, Planning and Logistics, 1ed (2013)
43. Maceira, M.E.P., Batista, F.R.S., Cerqueira, L.F., Melo, A.C.G., Marzano, L.G., Olasagasti, R.R.: A Probabilistic Approach to Define the Amount of Energy to be Traded in Hydro Dominated Interconnected Systems, 20th Power Systems Computation Conference—PSCC, Dublin, Ireland (2018)
44. Maceira, M.E.P., Penna, D.D.J., Melo, A.C.G., Moraes, L., Duarte, V.S.: Ten years of application of stochastic dual dynamic programming in official and agent studies in Brazil—description of the newave program. 16th Power System Computation Conference—PSCC, Glasgow, Scotland (2008)
45. Melo, A.C.G., Rodrigues, A.F., Batista, F.R.S., Marzano, L.G.B., Maceira, M.E.P.: Dominant contracting strategies for hydropower projects considering inflow uncertainties—application to the brazilian case, IEEE probabilistic methods applied do power systems, boise, pp. 24–28. Idaho, USA (2018)
46. Brazilian National Energy Policy Council (CNPE): Resolution CNPE Nº 29 (2019) (in Portuguese)
47. Ministry of Mines and Energy (MME): Ordinance Nº 59 (2020) (in Portuguese)

Authors and Affiliations

Maria Elvira P. Maceira[1] · Albert C. G. Melo[1] · José Francisco M. Pessanha[1,2] · Cristiane B. Cruz[2] · Victor A. Almeida[2] · Thatiana C. Justino[2]

✉ Maria Elvira P. Maceira
melvira@ime.uerj.br

Albert C. G. Melo
albert.melo@ime.uerj.br

José Francisco M. Pessanha
francisc@cepel.br

Cristiane B. Cruz
criscruz@cepel.br

Victor A. Almeida
andrade@cepel.br

Thatiana C. Justino
thatiana@cepel.br

[1] Mathematics and Statistics Institute, Rio de Janeiro State University — UERJ, Rio de Janeiro, Brazil

[2] Electric Energy Research Center — CEPEL, Rio de Janeiro, Brazil