Synthetic Time Series Generation Model for Analysis of Power System Operation and Expansion with High Renewable Energy Penetration

Rodrigo Palma-Behnke, *Senior Member, IEEE*, Jorge Vega-Herrera, *Member, IEEE*, Felipe Valencia, *Member, IEEE*, and Oscar Núñez-Mata, *Senior Member, IEEE*

***Abstract*—The increasing integration of renewable energy sources into current power systems has posed the challenge of adequately representing the statistical properties associated with their variable power generation. In this paper, a novel procedure is proposed to select a proper synthetic time series generation model for renewable energy sources to analyze power system problems. The procedure takes advantage of the objective of the specific analysis to be performed and the statistical characteristics of the available time series. The aim is to determine the suitable model to be used for generating synthetic time series of renewable energy sources. A set of indicators is proposed to verify that the statistical properties of synthetic time series fit the statistical properties of the original data. The proposal can be integrated into systematic tools available for data analysis without compromising the representation of the statistical properties of the original time series. The procedure is tested using real data from the New Zealand power system in a mid-term analysis on integrating wind power plants into the power system. The results show that the proposed procedure reduces the error obtained in analyzing power systems compared with reference models.**

***Index Terms*—Time series analysis, renewable energy source, solar energy, stochastic process, statistical analysis, wind energy.**

I. Introduction

The contribution of renewable energy sources (RESs) to power systems, e.g., wind- and solar-based energy, has undergone an accelerated expansion, and today, new wind farms and solar power plants are under construction or at the planning stage in different countries of the world [1]-[3]. For the wind power, before its integration into power systems, it is necessary to estimate their potential contribution to the system through a methodology on the wind speed and wind power modeling, such as the generation of synthetic series (SS) [4]. SS are series generated using mathematical models. SS aim to replicate the statistical properties of time series, often composed of historical datasets whose elements have causal relationships. Since time series have a stochastic component, the generation of each SS is the realization of the corresponding stochastic process [3], [5]-[7]. Thus, to represent the statistical properties of time series, a set of SS should be generated. The number of series to be generated and the model to be used are determined by the properties of the stochastic process.

With regard to its application to power systems, SS have other possible applications, e.g., ① the spinning-reserve required for their secure operation [8], [9]; ② the system conditions for reliability studies [10], [11]; ③ the commitment and economic dispatch of the generation units; ④ the size of the storage systems [12], [13]; ⑤ electrical consumption patterns [14]. These applications are fed by input parameters, which are usually subject to uncertainties. Different studies have attempted to solve the problems related to the operation and expansion of power systems in the presence of uncertainty and variability caused by RES. One way to overcome the uncertainty is the use of stochastic and robust optimization. These techniques have been used to solve the problems of unit commitment, generation planning, and transmission [5], [15]. The models above require the time series of the resources considered to model the variability and uncertainty adequately. This is a complex task, since in practice, it is often not possible to have time series of RES with the desired characteristics. Different models have been studied and developed to generate SS in response to this challenge. These models are used as input data in the tools for power system analysis [8]-[11], [16]. For instance, SS have been widely used to represent the uncertainty of RES in both generation and transmission planning [5], [17].

The selection of a model to generate SS is not a trivial task [8]-[11], [16], [17]. Few research works have been done on defining the criteria, indicators, or procedures to prioritize, organize, and systematize the selection of an appropriate model to generate SS. Based on a similar application in power systems, a model available in the literature is often

Manuscript received: November 10, 2020; accepted: May 19, 2021. Date of CrossCheck: May 19, 2021. Date of online publication: July 30, 2021.

This work was supported by FONDAP/ANID Solar Energy Research Centre SERC-Chile (No. 15110019), Fondecyt-ANID (No. 1211968), Fondecyt (No. 1181532), and the National Master Thesis (No. CONICYT/21161139).

R. Palma-Behnke (corresponding author) and F. Valencia are with the Energy Center, Department of Electrical Engineering, Faculty of Mathematical and Physical Sciences, University of Chile, Santiago, Chile (e-mail: rodpalma@cec.uchle.cl; fvalencia@centroenergia.cl).

J. Vega-Herrera is with the Department of Electrical Engineering, University of Antofagasta, Antofagasta, Chile (e-mail: jorge.vega@uantof.cl).

O. Núñez-Mata is with the School of Electrical Engineering, University of Costa Rica, San Pedro, Costa Rica (e-mail: oscar.nunezmata@ucr.ac.cr).

DOI: 10.35833/MPCE.2020.000747

tuned to fit the time evolution of the original time series. Several models for the applications in power systems have been reported in [8]-[11], [16]. Given the above, it is necessary to propose a methodology to systematize the selection of the model needed to generate SS depending on the problem that has to be solved with regard to power systems, e.g., for expansion and operation studies.

In the literature, Markov chains are widely used to generate SS for the analysis of power systems. These models can represent time series, whose statistical properties differ from the normal distribution [18], [19]. Nonetheless, their capabilities to represent the autocorrelation function (ACF) in series with seasonal variations have been questioned, and therefore, several modifications have been reported to achieve better representations of the ACF [20]. This is important because the ACF together with the seasonal trends determines the energy provided by the RES. Consequently, the use of Markov chains for the generation of SS is relegated to short-term analyses, where instant power becomes more important than the energy supplied by the RES [18], [19].

Furthermore, autoregressive integrated moving average (ARIMA) models are reported to generate long-term SS analyses, where the energy is more important than the short-term power produced by the RES [21]-[24]. These models provide a better representation of the ACF of time series compared with Markov chains [23]. However, the preprocessing of raw data is required so that their probability density function becomes normally distributed [25]-[27]. Numerical derivatives and integral transformation are often used for this purpose. The SS are obtained by applying the inverse of the preprocessing process to the model output, i.e., the numerical integration or the inverse integral transformation.

While Markov chains and ARIMA models can only represent temporal relationships in time series, vector autoregressive (VAR) models are used for the generation of SS in power system, where the spatial dependencies among the time series of RES are important (for instance, expansion planning analyses) [12], [16], [27]-[29]. Like ARIMA models, VAR models require the preprocessing of raw data to make their probability density functions normally distributed. As a result, the integral transformation is often used for this purpose [27]. VAR models adequately represent the statistical features of the datasets and their relationships, i.e., ACF, seasonal behavior, spatial dependencies, etc. Hence, they are suitable for generating SS in both long- and short-term analyses of power systems. However, the raw data have to be synchronized to turn the VAR model.

Significant efforts to develop new models for generating SS can be found in [16], [17]. However, little attention has been paid to formulating a procedure to select suitable models for generating SS depending on the study to be performed. This is highly needed considering the increase of RES and the importance of representing the uncertainty and variability to study the operation and expansion of power systems.

This paper proposes a novel procedure that selects the models needed to generate SS for a specific analysis of power systems. The procedure can be applied to study the operation and expansion of power grids with high penetration of renewable energy. The results indicate that the proposed procedure allows choosing, from several candidate models, the one that best fits the objective of the study and that better reconstructs the statistical characteristics of the original time series. The proposed procedure considers the following general steps.

1) The input of the specific analysis objective is considered.

2) With the information contained in the raw data, it determines the models that best fit the requirements of the analysis.

3) Finally, the model is selected by applying different statistical tests and computing a set of proposed indicators that account for the appropriateness of each model to represent the statistical properties of the original data.

Thus, the main contributions of this paper include: ① definitions and propositions of criteria needed for a proper selection of models to generate SS that can be used in different studies related to the operation and expansion of power systems; ② a systematic procedure to be followed and the statistical analyses necessary to select a suitable model for the generation of SS; ③ a procedure to systematize and define transformations by estimating the order and adjusting the models appropriately.

The remainder of this paper is organized as follows. Section II presents the proposed procedure. In Section III, the results are obtained by applying the proposed procedure. Finally, Section IV sets out the concluding remarks and future work.

II. Procedure for Selection of Models for Generation of SS

A. Overview

A suitable model for the generation of SS should produce an independent, identically distributed sample, each having the same fundamental properties as the original time series without replicating it exactly [15]. For an statistically viable SS, several stages must be followed. In [30], a methodology multi-step model-building strategy is developed with the following main steps: ① model specification, determining the order of model; ② model fitting, determining the parameter values of model; ③ model verification, checking whether the model analyzes the data correctly.

Firstly, in the model specification (also referred to as model identification), the different time series models that may be suitable for specified observed series are selected. In this step, the selected model is tentative and subject to revision later in the subsequent analysis. When choosing a model, in [30], the smallest number of parameters that will adequately represent the time series is suggested.

Secondly, the model fitting consists of finding the best possible estimates of these unknown parameters within a given model. For instance, the least-squares method can be considered.

Thirdly, the model verification is in charge of assessing the quality of the model that has been specified and estimated. In this step, two fundamental aspects are assessed. Initially, it is necessary to define whether the selected model fits the data. Next, it is necessary to assess whether the model assumptions are reasonably well satisfied. Thus, two situations may arise after the model verification, i.e., ① if there are no deficiencies found, it can be assumed that the modeling is complete and can be used; ② if deficiencies are found, another model is selected, i.e., we return to the model specification step. Finally, the three steps are repeated until, ideally, an acceptable model is found.

B. Proposed Procedure

The proposed procedure involves selecting and tuning the parameters of a model for generating SS for power system analysis. The procedure is based on an approach that classifies the most common models developed in the existing literature and simultaneously compares them using various indicators. Figure 1 shows the classification of the models reported in the existing literature about the generation of SS for solar- and wind-based RESs. Further, Fig. 1 shows the models that can represent the temporal and the spatial-temporal dependencies. The models comprise contemporaneous autoregressive moving average (CARMA), autoregressive moving average (ARMA), principal component analysis plus autoregressive moving average (PCA + ARMA) are identified in Fig. 1.

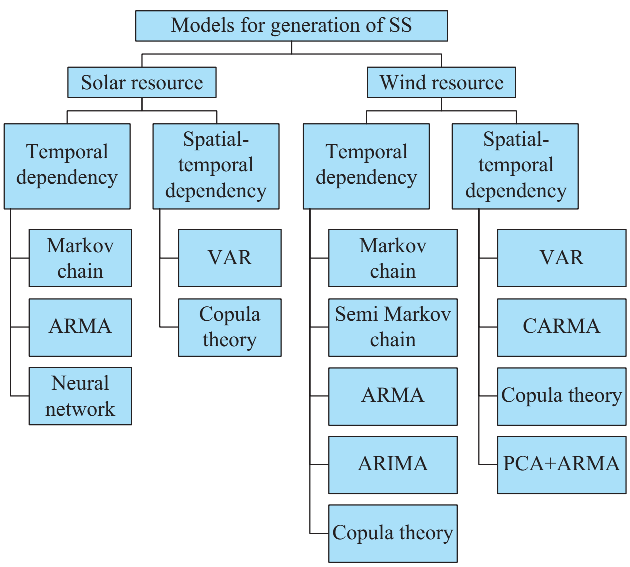

Fig. 1. Classification of models for generation of SS.

The proposed procedure focuses on modeling time series of wind speed, wind power, solar radiation, and solar power. This is because wind- and solar-based RESs are currently the most-widely employed ones across power systems.

The proposed procedure takes the analysis and available time series (dataset) of the RES as inputs. Figure 2 presents the flow chart of the proposed procedure including six steps. The first two steps involve the inputs of the procedure, namely, the analysis to be performed and the time series (raw data) available from the energy resource. The third step defines whether the information contained in the raw data is enough to carry out the analysis. The fourth step identifies the probability density function, the ACF, the partial autocorrelation function (PACF), seasonality (and their periods), and trends of the time series. The fifth step determines the set of candidate models that can be used for the generation of SS. The candidate models are included in Fig. 1. Finally, the sixth step selects the final model for the generation of SS. As can be observed, this procedure follows a sequence conceived to be later integrated into the available tools for data analysis and/or to be converted in a new model selection tool.

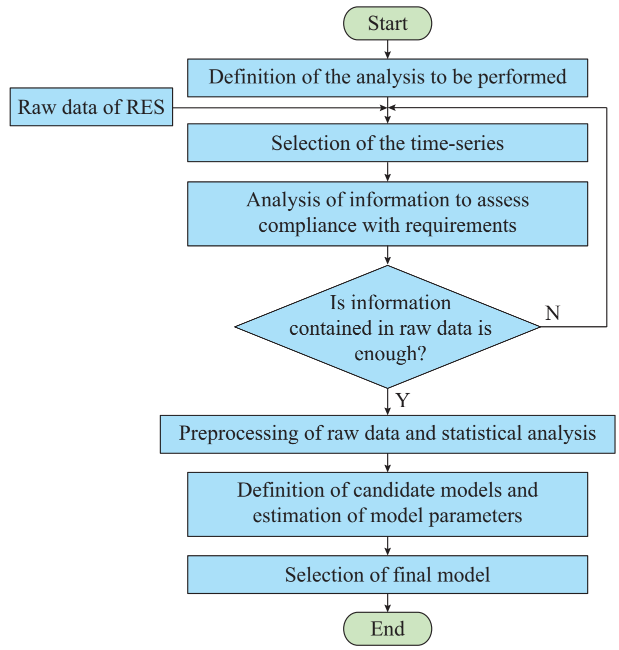

Fig. 2. Flow chart of model selection for generation of SS.

The steps shown in Fig. 2 are described in details as follows.

1) Definition of Analysis to Be Performed

This step defines the specific type of power system analysis to be performed and some modeling requirements for the time series of the energy resources. The following aspects should be clarified in this step: ① time horizon of the analysis; ② time frame; ③ simulation type; ④ the role of RES in the analysis; ⑤ power system model. There are two types of analysis in this paper: ① system expansion planning; ② system operation planning. Figure 3 shows the analysis carried out in this first step. In the analysis of system expansion planning, the power system model is simplified. The time horizon of the analysis ranges from 5 to 25 years, the RESs are considered as energy development centers, and in some cases, an hourly resolution is required for the time series. In the analysis of system operation planning, different models of the power system are used. The time horizon of the analysis ranges from a week (very short-term planning) to five years (long-term planning), only the RESs currently in operation are considered, and an hourly resolution for the time series is required. It is desirable to consider spatial-temporal dependencies among time series related to power plants based on RES.

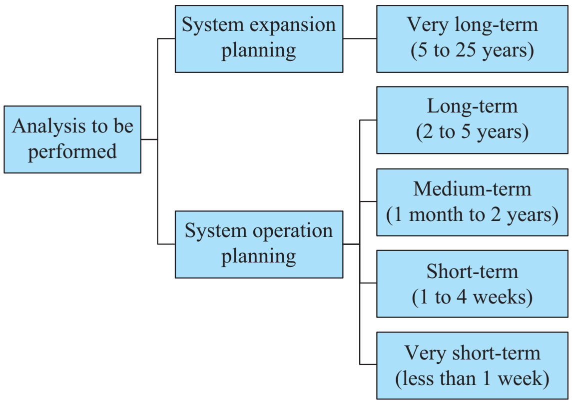

Fig. 3. Types of analysis to be performed in power system and corresponding time horizon.

2) Selection of Time Series

In this step, the time series (raw data) of the RES are selected. The following aspects must be verified in this step: ① location; ② measured variables; ③ data length; ④ sampling time. The time series available to perform the analyses of power systems are often not generated at the locations of interest. In these cases, atmospheric models are used to generate time series at the locations of interest [31], [32]. In this step, it is also suggested to check whether new RESs are added in the analysis (e.g., new RES projects become operative in the analysis). If so, the analysis must be conducted considering the time series of wind speed and/or solar radiation since this brings opportunity to reduce the uncertainty of raw data. Hence, the results are obtained [19]. The general scheme of this step is presented in Fig. 4.

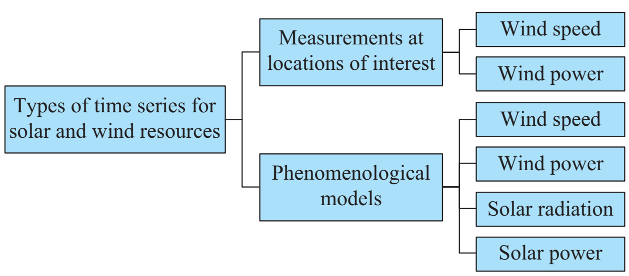

Fig. 4. General scheme for selection of time series.

In order to exemplify what is shown in Fig. 3, the following time series are selected: ① those measured at the energy development centers for system expansion planning; and ② those measured at the RES power stations currently in operation for system operation planning. In both cases, the sampling time must be at least one hour, while its duration should cover more than one year.

3) Analysis of Information to Assess Compliance with Requirements

In this step, the compliance with the requirements in the first step and the features of the available time series of the energy resources is assessed. If the time series meet all the requirements, there will be enough information to perform the desired analysis. Otherwise, the step of selection of time series is to be revisited for the missing information. Note that this is a checking step. It is verified that the following conditions are met, depending on the analysis to be performed.

1) As for the system expansion planning analysis, there are time series available for each energy development center; the time series are at least one year; and the sampling time of the time series is at least one hour.

2) As for the system operation planning, there are time series available for each RES currently in operation; the sampling time of the time series is one hour at the most; and the time series are synchronized (if the focus is in a very-short-time system operation planning).

If these conditions are not satisfied by the available time series, additional information about the RES centers should be gathered, or the focus of the analysis should be adapted to the information that the series provides. There is a dependency between the first and the second steps, which motivates the addition of a link between the first and the second steps in Fig. 2. Accordingly, this crosscheck prevents misleading results and/or issues arising during the data processing and the model selection steps (from the fourth step onwards).

4) Preprocessing of Raw Data and Statistical Analysis

In this step, a statistical characterization of the time series is carried out. This involves the preprocessing of raw data to convert the RES time series to meet the requirements for a proper application of SS models. Figure 5 shows the preprocessing and the transformation of the RES time series. The trends, seasonality, and temporal and spatial correlations are analyzed. This is important since wind speed/power and solar radiation/power series exhibit marked trends and seasonality that could be represented through deterministic models. Hence, less complicated models could be used to represent the behavior of the remaining stochastic process. In the proposed procedure, the following tests are used to identify trends, seasonality, temporal, and spatial correlations of the time series.

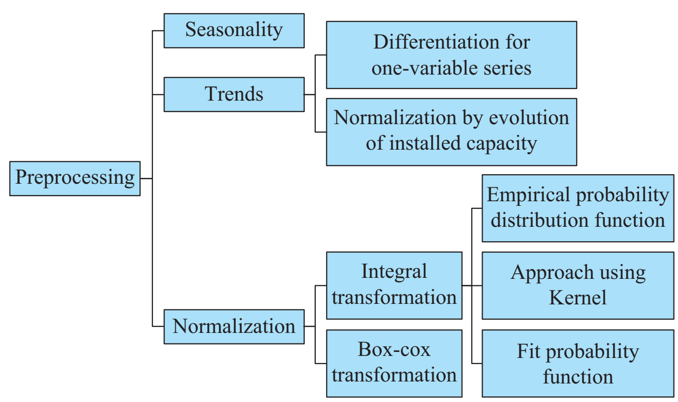

Fig. 5. General scheme for preprocessing of raw data and statistical analysis.

1) Trends and seasonality are identified through an augmented Dicker-Fuller test [33], and their results complement the results obtained through the box-plot technique and the Kruskal Wallis test [33].

2) Temporal and spatial correlations are identified using the ACF and the cross-correlation function (CCF), respective-

ly. When there is more than one power plant based on RES, the cross-correlation matrix (CCM) is used to identify the spatial correlation. Further, the single- and multi-variable versions of the Ljung-Box test are applied as complements [34].

Furthermore, since several models require that the time series have a normal probability density function, the procedure in [35] and the hypothesis tests reported in [36] are followed. They include the Anderson-Darling and the χ^2 tests.

Next, we describe how the tests and functions are used in this step. Firstly, the trends and seasonality are identified with hourly, daily, weekly, and monthly resolutions [32]. Secondly, the temporal dependency is verified by analyzing the ACF and the PACF of the individual time series. If these functions yield the coefficients significantly different from zero, it is concluded that these time lags are temporarily coupled. Thirdly, the spatial-temporal dependency is identified using the CCF. If there are lags significantly different from zero, then there exists a spatial correlation among the time series. Finally, as mentioned above, the hypothesis tests in [35], [36] are carried out to determine if additional processing of raw data is required to obtain the time series with a normal probability density function. It is important to remark that, although these analyses are performed with hourly, daily, weekly, and monthly resolutions (for illustrative purposes), just carrying out these assessments using a weekly and a monthly resolution provides enough information for the system expansion planning. Nonetheless, for the system operation planning, these tests should be conducted with hourly and even daily resolutions.

5) Definition of Candidate Models and Estimation of Model Parameters

In this step, a set of models suitable for the generation of SS is defined. The analysis requirements to be performed and the statistical features of the time series are considered for this purpose. Also, the assumptions/properties of the available models are checked. The set of models suitable for the generation of SS is formed by following a decision tree. In this step, the following three conditions are checked.

1) The need for spatial-temporal representation

Since just a few models satisfy this requirement (as shown in Fig. 1), those models determine a set of suitable models for themselves and define the first possible outcome of the decision tree, i. e., the remaining conditions are not checked, and the procedure moves towards the next step: the selection of the final model. However, if the analysis does not require the spatial-temporal representation, then the candidate models are those able to represent the temporal dependency, as shown in Fig. 1. Thus, the process moves towards the second condition.

2) Type of time series where the model is used

This includes solar radiation/power and/or wind speed/power in similar analyses. As a result, a set of available models is brought down to those previously used to represent the same energy source in similar power system analysis.

3) Compliance degree between statistical properties of time series and assumptions/requirements of models

The models that do not require the preprocessing of raw data are identified, thereby defining the set of suitable models, i.e., the second possible outcome of the decision tree. If all models require the preprocessing of the raw data, then they are classified into two categories: ① the models that only require a numerical derivative to comply with all their assumptions/requirements; ② the models that require the use of an integral transformation to comply with all their assumptions/requirements. The two sets of suitable models and their preprocessing methodologies constitute the third possible outcome of the decision tree.

6) Selection of Final Model

In this step, the final model is selected from the set of suitable models defined in the previous step. The structures and parameters are also identified. In order to select the suitable model, a five-step procedure is proposed as follows: ① analysis of data from the previous processing; ② determination of model parameters; ③ generation of SS; ④ calculation of assessment indicators on model benefits; ⑤ analysis of indicators and selection of the final model. To determine the model parameters, we follow the instruction given in the works where the models are presented.

The best model is selected considering the performance indicators of each model, which are based on the deviation of the statistics obtained from the SS and the original time series. The statistics considered are the mean, variance, standard deviation, and the quantiles 10%, 25%, 50%, 75%, and 90%. Besides, the deviations from the probability density function, ACF, PACF, and the CCM are also used as indicators. All these indicators are computed with the same resolution used in the statistical analysis, e.g., hourly, daily, weekly, and/or monthly, to prevent representation errors [12], [22], [33]. The error measurements proposed in this paper to assess the accuracy of each model are as follows.

1) Root mean square error (RMSE)

$$RMSE = \sqrt{\frac{1}{N}\sum_{k=1}^{N} e_k^2} \tag{1}$$

2) Root mean squared relative error (RMSRE)

$$RMSRE = \frac{RMSE}{\sqrt{\frac{1}{N}\sum_{k=1}^{N} x_k^2}} \tag{2}$$

3) Error measurement

$$\overline{RMSRE} = \frac{1}{2}\left(RMSRE_{\mathrm{Month}} + RMSRE_{\mathrm{CCM}}\right) \tag{3}$$

where e_k is the difference between the model and the time series for the k^{th} statistical feature; N is the number of statistical features considered; and x_k is the k^{th} statistical feature of the time series to be represented through the SS; $RMSRE_{\mathrm{Month}}$ is the RMSRE calculated using monthly statistics; $RMSRE_{\mathrm{CCM}}$ is the RMSRE of the CCM; and $\overline{RMSRE}$ is the average of $RMSRE_{\mathrm{Month}}$ and $RMSRE_{\mathrm{CCM}}$. Equation (3) is the error measurement used when the original time series have a seasonal behavior and spatial-temporal dependencies.

The model that gives the minimum error measurements is selected as the final model to generate SS for the power system analysis. If the error measurements are not significantly different from one to another, an additional analysis is carried out based on the objective of the analysis to be performed. For instance, in very short-term analyses that include chronological simulations, the spatial-temporal relationships are essential. Hence, the model with the minimum error measurements in the CCM is selected to generate SS. In contrast, for the medium, large, and very-large analyses, wherein energy simulations are considered, the energy contribution of the RES is highly important. Then, the model with the minimum error measurements in the statistical seasonality indicators is selected to generate SS. If chronological simulations of typical days are included, the SS must adequately represent the spatial-temporal relationships, the seasonality, and the energy contribution of the RES. Then, the model with the minimum error measurements in both statistical seasonality indicators and CCM is selected.

Note that the proposed procedure for the model selection considers that all assumptions are made in the formulation of the models (e.g., the requirements of co-variance stationarity and normal distribution of time series for ARMA models). The proposed procedure also attempts to prevent the models without previous statistical knowledge of the time series to be represented. Furthermore, by following the procedure, the model that better fits with the requirements of the analysis and the time series features is selected. This avoids the review and model selection procedures when only a model exploration search is performed. Finally, by following the proposed procedure, we diminish the error in the analysis that could appear due to the lack of representation in the statistical properties of the time series related to RES. In addition, the proposed procedure follows the logic of expert systems. Therefore, it could be integrated with other data analysis tools and/or used independently as an additional tool by system operators. This is beyond the scope of this paper, and therefore, such implementation is not included.

A case study is presented in Section III, where the proposed procedure is applied in system expansion planning.

III. Case Study and Results

In this section, the proposed procedure is applied to a mid-term operation planning analysis considering two wind power stations. The proposed procedure is implemented in the R and MATLAB software.

Step 1: the case study considers the system operation planning with a two-year horizon. The system is mainly thermal, and chronological simulations are included in the analysis. The model for the simulations considers the main transmission lines and the short-term technical constraints of the system. The addition of new power plants is not included in the analysis. Two wind power plants already in operation in the system are modeled.

Step 2: for this specific case, the time series for these power plants have an hourly resolution and correspond to the measurements made in New Zealand from 2004 to 2008 at STH1 and CKS1 locations (STH1 and CKS1 mean one location in wind sites in the Southland and Otago, and Cook Strait, respectively) [33], [37].

Step 3: the time series requirements for this case study are as follows: a time frame of two years, a resolution of at least an hour, and spatial-temporal correlation. The latter is required because a chronological simulation is considered. The available time series of wind speed meet the first two requirements. The spatial-temporal correlation is verified through the CCF of the series.

Step 4: Fig. 6(a) shows the ACF of time series at STH1, whereas Fig. 6(b) shows the CCF of time series between STH1 and CKS1 locations. These functions allow us to determine if there are spatial-temporal dependency. Figure 6 shows that several coefficients significantly differ from zero, which means that there is a spatial-temporal dependency between the two locations.

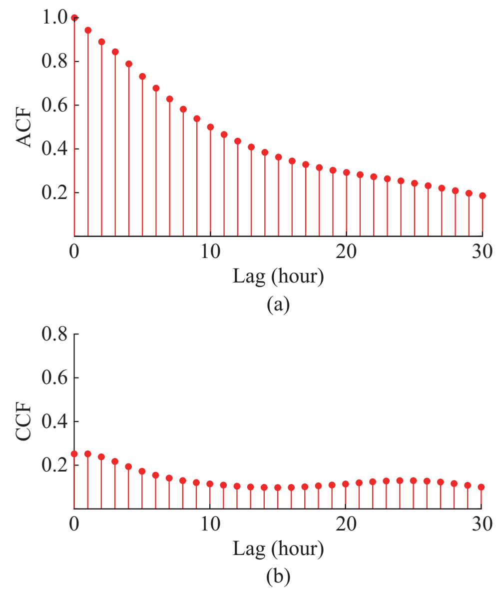

Fig. 6. Time series spatial-temporal correlation. (a) ACF of time series at STH1 location. (b) CCF of time series at STH1 and CKS1 locations.

The next step is to determine whether the time series are normally distributed. Figure 7 shows the histogram of wind speed time series at STH1 location. The histogram presents an evident asymmetry towards the left. This indicates that the time series are not normally distributed. To corroborate this statement, the χ^2 and Anderson-Darling tests are conducted, resulting in the rejection of the null hypothesis that the time series are normally distributed (the p-values of the tests are close to zero).

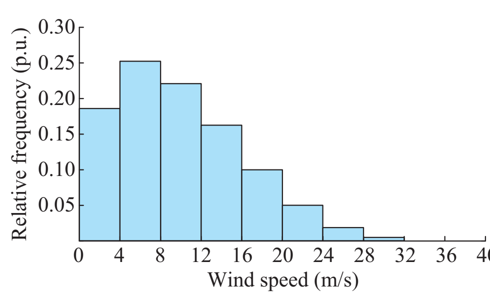

Fig. 7. Histogram of wind speed time series at STH1 location.

Then, the seasonality in the time series is assessed. The box plot of the times series is depicted in Fig. 8 considering both hourly and monthly resolutions to perform exploratory analysis, where the results obtained from STH1 location with an hourly resolution. The results show that the time series present a daily seasonality. In order to verify this fact, the Kruskal Wallis test is applied. As a result, the *p*-values tend to be zero for both hourly and monthly resolutions at STH1 and CKS1 locations. This fact implies that there are hourly and monthly seasonality in both time series. Since both time series have seasonality, it is also possible to conclude that they are not stationary in terms of co-variance. This is important for defining the suitable models and selecting the final model.

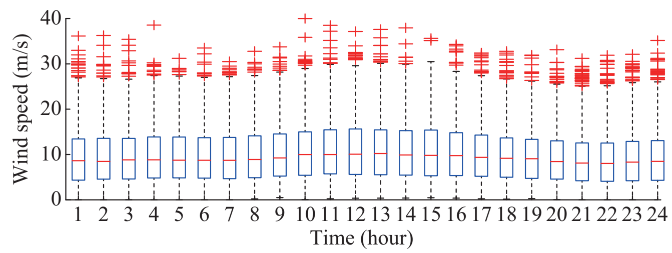

Fig. 8. Box plot for STH1 location with hourly resolution.

Step 5: the set of suitable models should be defined. Given the spatial-temporal representation constraint, only the models with this capability are considered. Since raw datasets are synchronized with hourly and monthly seasonalities, and are not normally distributed, VAR models with different preprocessing methods [16], [23], CARMA models [8], models based on Copula theory [12], and PCA + ARMA models [29] are finally selected. These models constitute a set of 17 suitable models as follows: ① 12 VAR models with different preprocessing methods; ② three different CARMA models; ③ a model based on Copula theory; ④ a PCA + ARMA model. For this set of models, the integral transformation is selected to preprocess the original data for VAR and PCA plus VAR models (the remaining models do not require a preprocessing of the raw data). Then, (4)-(8) define the mathematical procedure to transform the original data.

$$\boldsymbol{Y}_t = \boldsymbol{X}_t - \boldsymbol{X}_{at} - \boldsymbol{X}_{dt} \tag{4}$$

$$\boldsymbol{Y}_t = \frac{\boldsymbol{X}_t - \boldsymbol{\mu}_{tk}}{\boldsymbol{\sigma}_{tk}} \quad \forall k \in S \tag{5}$$

$$\boldsymbol{Y}_t' = \boldsymbol{F}_{inN}^{-1}(\boldsymbol{F}_k(\boldsymbol{Y}_{tk})) \quad \forall k \in S \tag{6}$$

$$\boldsymbol{Y}_t = \boldsymbol{X}_t - \boldsymbol{\mu}_{tk} \quad \forall k \in S \tag{7}$$

$$\boldsymbol{Y}_t' = \boldsymbol{F}_{inN}^{-1}(\boldsymbol{F}(\boldsymbol{Y}_t)) \tag{8}$$

where $\boldsymbol{X}_t$ is the vector of each time series; $\boldsymbol{X}_{at}$ is the vector of the means computed with a monthly resolution; $\boldsymbol{X}_{dt}$ is the vector of the means computed with an hourly resolution; $\boldsymbol{Y}_t$ is the vector of the transformed time series; $\boldsymbol{F}_k$ is the vector of probability density functions; $\boldsymbol{F}_{inN}^{-1}$ is the vector of inverse standard normal distributions; $\boldsymbol{Y}_t'$ is the resulting stationary normally distributed time series; $\boldsymbol{\mu}_{tk}$ is the vector of seasonality means; $\boldsymbol{\sigma}_{tk}$ is the vector of seasonality standard deviation; and k is the seasonality pattern, which belongs to the set S composed of hourly and monthly seasonality.

Once the raw data are preprocessed, the resulting time series $\boldsymbol{Y}_t'$ associated with each original series are used to identify the structure and parameters of each model. The identification procedure ends when the residuals of the model related to $\boldsymbol{Y}_t'$ behave as uncorrelated white noise. Furthermore, the SS generated using the model is normally distributed. The SS of the renewable resources are obtained by applying the inverse process described in (4)-(7).

Step 6: considering the objectives of the case study, the statistical indicators used to measure the accuracy of the models are computed for both hourly and monthly resolutions. Since the chronological simulation requires that the seasonality, the spatial-temporal correlations, and the energy contribution of the wind power plants are adequately represented, (3) is used to select the model to generate SS. This error measurement consists of a combination of RMSRE computed with a monthly resolution and RMSRE of the CCM. This error measurement is selected since the error metrics (1) and (3) are similar for all models and all statistical features presented in Section II. Furthermore, (3) is used as error measurement since high-order models tend to have similar RMSRE values when computed with an hourly resolution. Figure 9 presents the RMSRE obtained for the set of suitable models, considering the wind speed time series at STH1 location. In Fig. 9, models 1 to 12 are the VAR models; models 13, 16, and 17 are the CARMA models; model 15 is based on the Copula theory, and model 14 is the PCA + ARMA model. As can be observed, the model with the minimum RMSRE value is model 15. Figure 10 shows the RMSRE of CCM and monthly statistics. It can be observed that model 15 gives the highest value of RMSRE of CCM, i.e., 35%. This is an unacceptable error considering that the analysis to be performed involves a chronological simulation.

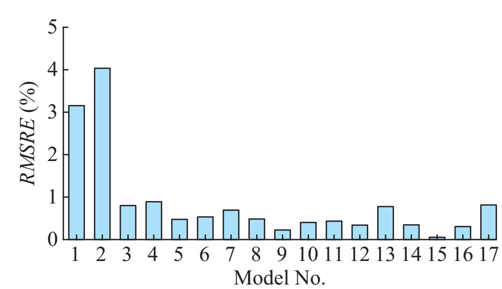

Fig. 9. RMSRE of each available model considered.

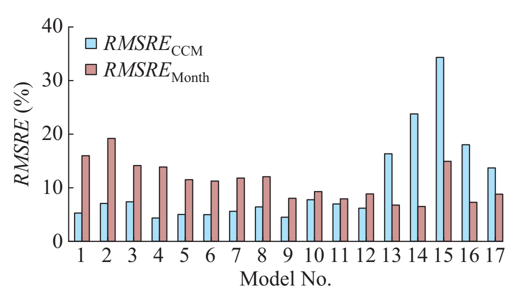

Fig. 10. RMSRE of CCM and monthly statistics for each available model considered.

Figure 11 shows the RMSRE values for each available model computed by solving (3). Contrary to the hourly resolution analysis suggested, (3) indicates that the best model to generate SS is the VAR model 9. This is because the RMSRE of the VAR model 9 is smaller than that of model 15. This result is expected due to the significant difference in the RMSRE of CCM shown in Fig. 10, where $RMSRE_{\mathrm{CCM}}$ of the VAR model 9 is lower than 10%, whereas that of model 15 is about 35%. This result shows a considerable difference between selecting any model available in the literature and following the proposed procedure for model selection.

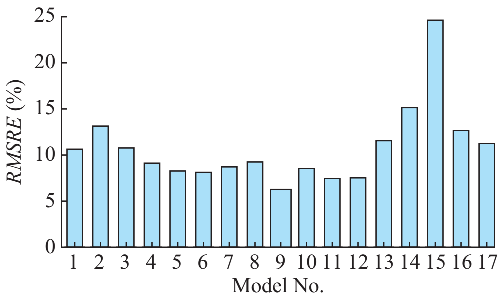

Fig. 11. RMSRE values for each available model computed by solving (3) considering wind speed time series at STH1 location.

Now, we assess the accuracy of the VAR model 9, which is the selected model. The histogram of the SS generated with this model, its ACF, its PACF, its CCF, and its box plot are compared with the original wind speed time series. In this paper, only the results of the STH1 location are presented.

Figure 12 shows the histograms of the original time series (OTS) and the SS generated by the VAR model 9. It is shown that the SS can represent the probability density function of OTS. This point is relevant, as the type of analysis to be performed requires that the energy contribution of the wind power plants is adequately represented, and the probability density function is an indicator of the contribution.

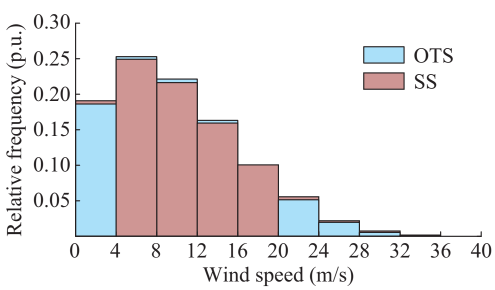

Fig. 12. Histograms of OTS and SS generated by VAR model 9.

Figure 13(a) and (b) shows the comparison of the ACF and PACF, respectively. Figure 13(a) shows that the ACF restructed through SS using VAR model 9 closely follows the ACF of OTS until lag 20. In this lag, the difference between the two ACFs achieves its maximum value and keeps constant until lag 50. This is an expected result since VAR model 9 is of order 16. Therefore, it cannot represent the influences of all lags in the current value of the time series. However, the difference observed is not significant. This is supported by the result presented in Fig. 13(b), where the PACF of OTS and PACF restructed through SS using VAR model 9 are practically the same. Since the PACF determines the effect of a given lag in the current value of the series, the VAR model 9 is suitable to represent the temporal dependency of the time series.

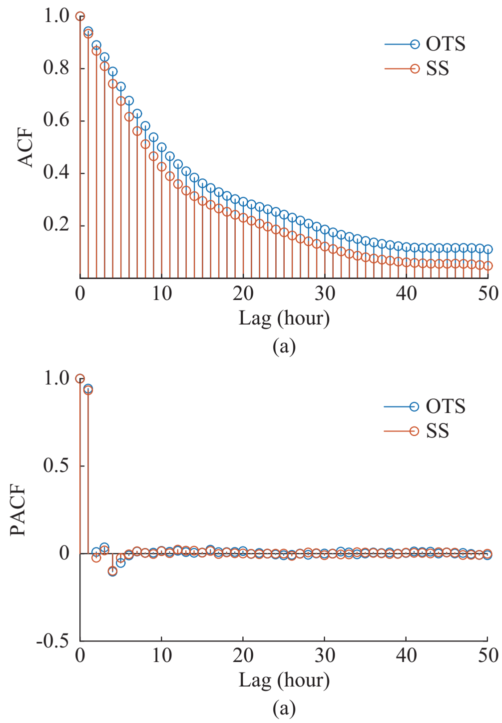

Fig. 13. ACF and PACF of OTS and SS reconstructed through SS using VAR model 9 at STH1 location. (a) ACF. (b) PACF.

Figure 14 shows the comparison of the CCF of OTS and the CCF reconstructed through SS using VAR model 9 at STH1 location. It is observed that the SS can represent the shape and values of the original CCF. However, the relationship between the past values of CKS1 location and the present values of STH1 location has a better representation than the influence of the future values of CKS1 location on the present values of STH1 location. This is an expected result since the VAR model used for the generation of SS is causal. Therefore, it only considers the influence of the cast lags in the current value of the time series. Despite this result, the difference observed for the future measurements at CKS1 location is not significant. Thus, it is possible to conclude that model 9 also adequately represents the spatial relationships between the wind speed time series at STH1 and CKS1 locations. This statement is further supported by the result in Fig. 10, where the RMSRE error for model 9 is about 4% in the representation of the cross-correlation matrix.

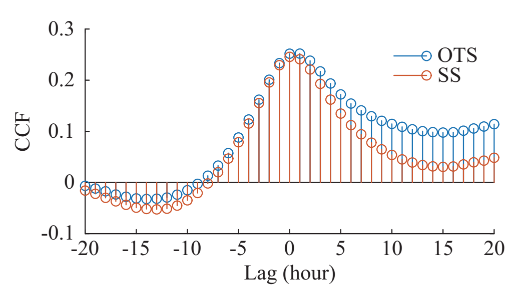

Fig. 14. Comparison of CCF of OTS and CCF reconstructed through SS using VAR model 9.

Finally, Fig. 15 shows a comparison between the box plots of the OTS and SS. Like the previous observations, the SS can represent the seasonal behavior on an hourly basis. As shown in Fig. 15, wind speed measurements at CKS1 location are used for comparison to show that the VAR model 9 does not allow only a representation of the statistical indicators of the measurements at STH1 location. This result confirms that the proposed procedure for selecting a suitable model to generate SS for power system analysis works appropriately. A model that achieves the requirement of power system analysis and the statistical properties of time series associated with RES can be obtained. Table I shows the statistical properties of the selected model.

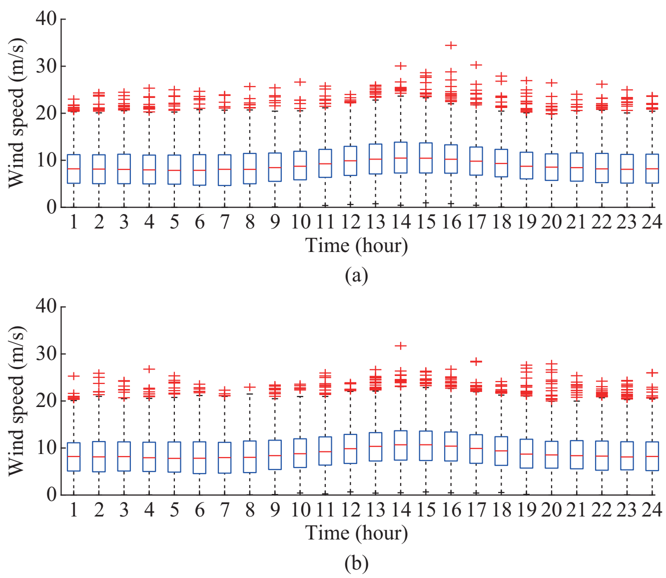

Fig. 15. Box plots of OTS and SS using VAR model 9. (a) OTS. (b) SS.

TABLE I
STATISTICAL PROPERTIES FOR CKS1 LOCATION

Statistics	Property	
	Original model	VAR model 9
Average	9.057	9.061
Standard deviation	4.411	4.375
Variance	19.453	19.140
Minimum	0.000	0.000
Maximum	34.438	32.184
Quantile 10%	3.470	3.470
Quantile 50%	8.803	8.851
Quantile 75%	11.986	11.986
Quantile 90%	14.908	14.834

IV. CONCLUSION

This paper studies the challenge of selecting the appropriate model for generating SS for operation, planning, and expansion studies of power systems considering RESs. In this sense, a methodological proposal to select a suitable model to generate SS is proposed.

It has been demonstrated that if an adequate analysis is not carried out and the available models are applied without verifying their assumptions and application conditions, seasonal energy contributions or dependency structures may not be adequately characterized. According to the obtained results, it is found that the proposed procedure for model selection allows choosing the model that best achieves the objective of this paper, and can represent the statistical characteristics of the OTS.

Furthermore, the proposed approach is independent of the type of RES modeled, the nature of the time series (i.e., solar power, solar radiation, wind power, or wind speed), and the statistical features of the time series. Additionally, the results show that, if the proposed procedure is followed, it is possible to reduce the error in the analysis of power systems compared with a traditional approach. Future research shall be focused on scenario reduction strategies and the simplification of different steps of the proposed procedure. Besides, incorporating other types of modeling (e.g., machine learning, physics, and stochastic models, etc.) in the proposed framework is considered as future work.

REFERENCES

[1] IRENA. (2019). Renewable Energy Statistics 2019. [Online]. Available: https://www.irena.org/publications/2019/Jul/Renewable-energy-statistics-2019

[2] I. Andres, "CVAR constrained planning of renewable generation with consideration of system inertial response, reserve services and demand participation," M. S. dissertation, Pontificia Universidad Católica de Chile, Santiago, Chile, 2014.

[3] P. Chen, "Generation, stochastic modeling and analysis of power system with renewable generation," M.S. dissertation, Aalborg University, Aalborg, Denmark, 2010.

[4] B. Brown, R. Katz, and A. Murphy, "Time series models to simulate and forecast wind speed and wind power," *Journal of Climate and Applied Meteorology*, vol. 23, no. 8, pp. 1184-1195, Aug. 1984.

[5] A. Soroudi and T. Amraee, "Decision making under uncertainty in energy systems: state of the art," *Renewable and Sustainable Energy Reviews*, vol. 28, pp. 376-384, Dec. 2013.

[6] F. Ezbakhe and A. Pérez-Foguet, "Decision analysis for sustainable development: the case of renewable energy planning under uncertainty," *European Journal of Operational Research*, vol. 292, pp. 1-13, Mar. 2020.

[7] V. Onishi, C. Antunes, E. Fraga *et al*., "Stochastic optimization of trigeneration systems for decision-making under long-term uncertainty in energy demands and prices," *Energy*, vol. 175, pp. 781-797, May 2019.

[8] A. Papavasiliou and S. Oren, "Multiarea stochastic unit commitment for high wind penetration in a transmission constrained network," *Operation Research*, vol. 61, no. 3, pp. 578-592, May-Jun. 2013.

[9] A. Papavasiliou, S. Oren, and R. O'Neill, "Reserve requirements for wind power integration: a scenario-based stochastic programming framework," *IEEE Transactions on Power Systems*, vol. 26, no. 4, pp. 2197-2206, Nov. 2011.

[10] R. Billinton, H. Chen, and R. Ghajar, "Time-series models for reliability evaluation of power systems including wind energy," *Microelectronics Reliability*, vol. 36, no. 9, pp. 1253-1261, Sept. 1996.

[11] R. Billinton and W. Wangdee, "Reliability-based transmission reinforcement planning associated with large-scale wind farms," *IEEE Transactions on Power Systems*, vol. 22, no. 1, pp. 34-41, Feb. 2007.

[12] H. Haghi and S. Lotfifard, "Spatiotemporal modeling of wind generation for optimal energy storage sizing," *IEEE Transactions on Sustainable Energy*, vol. 6, no. 1, pp. 113-121, Jan. 2015.

[13] L. Kotzur, P. Markewitz, M. Robinius *et al*., "Time series aggregation for energy system design: modeling seasonal storage," *Applied Energy*, vol. 213, pp. 123-135, Mar. 2018.

[14] X. Serrano-Guerrero, G. Escrivá-Escrivá, S. Luna-Romero *et al*., "A time-series treatment method to obtain electrical consumption patterns for anomalies detection improvement in electrical consumption profiles," *Energies*, vol. 13, no. 5, pp. 1-23, Feb. 2020.

[15] A. Conejo, M. Carrión, and J. Morales, *Decision Making Under Uncertainty in Electricity Markets*. New York: Springer, 2010.

[16] J. Ekström, M. Koivisto, I. Mellin *et al*., "Assessment of large scale

wind power generation with new generation locations without measurement data," *Renewable Energy*, vol. 83, pp. 362-374, Nov. 2015.

[17] M. Koivisto, J. Ekstrom, J. Seppanen *et al*., "A statistical model for comparing future wind power scenarios with varying geographical distribution of installed generation capacity," *Wind Energy*, vol. 19, no. 4, pp. 665-679, May 2016.

[18] H. Nfaoui, H. Essiarab, and A. Sayigh, "A stochastic Markov chain model for simulating wind speed time series at Tangiers, Morocco," *Renewable Energy*, vol. 29, no. 8, pp. 1407-1418, Jul. 2004.

[19] G. Papaefthymiou and B. Klockl, "MCMC for wind power simulation," *IEEE Transactions on Energy Conversion*, vol. 23, no. 1, pp. 234-240, Mar. 2008.

[20] S. Karatepe and K. Corscadden, "Wind speed estimation: incorporating seasonal data using Markov chain models," *ISRN Renewable Energy*, vol. 2013, pp. 1-9, Dec. 2013.

[21] S. Kennedy and P. Rogers, "A probabilistic model for simulating long-term wind-power output," *Wind Engineering*, vol. 27, no. 3, pp. 167-181, May 2003.

[22] R. Aguiar and M. Collares-Pereira, "TAG: a time-dependent, autoregressive, Gaussian model for generating synthetic hourly radiation," *Solar Energy*, vol. 49, no. 3, pp. 167-174, Sept. 1992.

[23] D. Hill, D. McMillan, K. Bell *et al*., "Application of auto-regressive models to U.K. wind speed data for power system impact studies," *IEEE Transactions on Sustainable Energy*, vol. 3, no. 1, pp. 134-141, Jan. 2012.

[24] Y. Ge, Y. Nan, and L. Bai, "A hybrid prediction model for solar radiation based on long short-term memory, empirical mode decomposition, and solar profiles for energy harvesting wireless sensor networks," *Energies*, vol. 12, no. 24, pp. 1-21, Dec. 2019.

[25] A. Lojowska, D. Kurowicka, G. Papaefthymiou *et al*., "Advantages of ARMA-GARCH wind speed time series modeling," in *Proceedings of IEEE 11th International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, Singapore, Singapore, Jun. 2010, pp. 83-88.

[26] V. Graham, K. Hollands, and T. Unny, "A time series model for K_t with application to global synthetic weather generation," *Solar Energy*, vol. 40, no. 2, pp. 83-92, 1988.

[27] B. Klockl and G. Papaefthymiou, "Multivariate time series models for studies on stochastic generators in power systems," *Power Systems Research*, vol. 80, no. 3, pp. 265-276, Mar. 2010.

[28] J. Morales, R. Mínguez, and A. Conejo, "A methodology to generate statistically dependent wind speed scenarios," *Applied Energy*, vol. 87, no. 3, pp. 843-855, Mar. 2010.

[29] D. Le, G. Gross and A. Berizzi., "Probabilistic modeling of multisite wind farm production for scenario-based applications," *IEEE Transactions on Sustainable Energy*, vol. 6, no. 3, pp. 748-758, Jul. 2015.

[30] J. Cryers and K.-S. Chan, *Time Series Analysis with Applications in R*. New York: Business Media LLC, 2008.

[31] A. Papavasiliou and S. Oren, "Stochastic modeling of multi-area wind power production," in *Proceedings of 48th Hawaii International Conference on System Sciences*, Hawaii, USA, Jan. 2015, pp. 1-10.

[32] J. Dowds, P. Hines, T. Ryan *et al*., "A review of large-scale wind integration studies," *Renewable and Sustainable Energy Reviews*, vol. 49, pp. 768-794, Sept. 2015.

[33] G. Pritchard. (2016, Dec.). Code and data on wind power correlation. [Online]. Available: https://www.stat.auckland.ac.nz/~geoff/

[34] R. Tsay, *Multivariate Time Series Analysis with R and Financial Applications*, New Jersey: John Wiley & Sons, 2013.

[35] W. Martinez, A. Martinez, and J. Solka, *Exploratory Data Analysis with MATLAB*, Abingdon: Taylor & Francis Group, 2017.

[36] G. Box, G. Jenkins, G. Reinsel *et al*., *Times Series Analysis Forecasting and Control*, New Jersey: John Wiley & Sons, 2016.

[37] K. Suomalainen, G. Pritchard, B. Sharp *et al*., "Correlation analysis on wind and hydro resources with electricity demand and prices in New Zealand," *Applied Energy*, vol. 137, pp. 445-462, Jan. 2015.

Rodrigo Palma-Behnke received the M.Sc. degrees in electrical engineering from the Pontificia Universidad Católica de Chile, Santiago, Chile, in 1994, and the Ph.D degree from the University of Dortmund, Dortmund, Germany, in 1999. He is a Full Professor at the Electrical Engineering Department at the Universidad de Chile, Santiago, Chile, since 1999. His research interests include the planning and operation of electrical systems in competitive power markets, renewable energy, smart grids, education, and microgrids.

Jorge Vega-Herrera received the B.S. degree in electrical engineering from the Universidad de Antofagasta, Antofagasta, Chile, in 2012, and the M.Sc. degree in electrical engineering from the Universidad de Chile, Santiago, Chile, in 2016. He is currently pursuing the Ph.D. degree in electrical engineering at the Universidad de Chile. He is a Full Professor at the Electrical Engineering Department at the Universidad de Antofagasta. His research interests include the modeling and stability of power systems, power systems transients, and integration of converter-interfaced generation in power systems.

Felipe Valencia received the Master's and Ph.D. degrees in control engineering from the Universidad Nacional de Colombia, in Manizales and Medellin, Colombia, in 2009 and 2012, respectively. He is a Control Engineer with the Universidad Nacional de Colombia, Medellín, Colombia. He is currently an Associate Researcher with the Solar Energy Research Center and a Full Researcher with the ATAMOSTEC technology program. His research interests include the design of distributed and hierarchical strategies for controlling large-scale systems, different fields such as power energy generation, transmission, and distribution, transportation, and smart grids.

Oscar Núñez-Mata received the B.S. degree in electrical engineering from the Universidad de Costa Rica, San Pedro, Costa Rica, in 1993, and the M.S. and Ph.D. degrees in electrical engineering from Universidad de Chile, Santiago, Chile, in 2014 and 2018, respectively. He is a Full Professor at the School of Electrical Engineering at the University of Costa Rica, San Pedro, Costa Rica. His research interests include the development of new protection systems, energy conversion systems, and the development of microgrids and renewable energy solutions.