A Synthesis Method to Generate Hourly Electricity Production Time-series of Wind Plants in Peru for Long-term Expansion Planning

Ruben Felix, *Member, IEEE* and Jaime E. Luyo, *Member, IEEE*

***Abstract*—Plenty of works have treated the system expansion planning problem in the presence of intermittent renewable energy resources like wind. However, most of those proposals have been approached from scenarios of plenty of data, which is not the rule in developing countries, where principal investment actors have recently switched their focus. In contrast to operation problems where existing literature can be successfully applied since it requires short-term historical time-series gathered from the same studied plants, proposals for planning problems are almost impossible to apply because of a lack of information and measurement about renewable resources in places where no renewable plants have been previously installed. In order to fill this information gap, this paper presents a novel methodology to synthesize wind production time-series on an hourly time scale, taking as inputs aggregate data such as monthly wind speed average values and Weibull annual parameters. The methodology comprises four steps, from data gathering to calculating electrical power produced by a wind farm. Three application tests are performed for different places in India, Chile, and Peru to validate the proposed methodology. The results show that the methodology successfully synthesizes time-series of output power, correctly achieves persistence characteristics, and slightly over or underestimates the produced wind energy, having a discrepancy of $\pm$6.2% in the yearly total.**

***Index Terms*—Expansion planning, power systems, renewable energy, synthesis methods, wind energy.**

I. Introduction

As non-conventional renewable generation technologies continue decreasing their investment cost and increasing their efficiencies [1]–[3], electrical systems all over the world will get their renewable energy resources (RER) penetration level growing up, as is forecasted by many global organizations like [1] who projects that RER production will account for a 86% of global power generation. At the same time, electricity will have 49% of the share in final consumption by 2050. This fact will bring several benefits to human development, but also new challenges that energy sector actors shall face to smooth the path of this energy transition.

On the one hand, both operation problems: real-time operation and dispatch scheduling, must deal with the intermittent behavior of wind resources. It is known that while penetration level increases, the system gets more vulnerable to RER generation fluctuation, as studied in [4]–[6]. Cited studies show that short-period fluctuations in wind speed change drastically load demands as seen by the independent system operator (ISO) creating a significant ramp up or down, which make it compulsory to have must-run machines or energy storage systems in the grid. Also, the unpredictable availability of the generation capacity of this kind of plant makes the scheduling process of day-ahead electricity markets difficult. Sometimes, complexity is such that the ISO assumes a deterministic generation time-series that plant owners prepare and send to them, avoiding the responsibility to accurately forecast this input, as evidenced in [7], [8] for the Peruvian case. This problem affects physical markets and the wholesale electricity market, as studied in [9]. Plenty of approaches have been studied in the literature [10]–[13] to overcome these problems.

Although uncertainties are present everywhere in the energy sector's value chain, it should be noticed that the farther the time horizon analysis is, the more challenging it becomes to resolve the problem. Thus, on the other hand, system expansion planning problems are at a higher level of complexity [14], [15].

Furthermore, successfully overcoming these problems is conditioned to information on resources measurements and existing plants' historical records. Unfortunately, developing countries that are recently increasing their investment expenditures in RER plants usually lack this information. Although private agents make their field studies, the gathered information is not socialized, and public and academic organisms are obliged to resort to other methods to develop their activities.

Indeed, this is of the points on which recent works on barriers to renewable energy in developing countries strongly coincide. Specifically on the lack of capabilities [16] and information barriers [17], since these countries do not have access to databases on the potential and production of renewable energies. Neither do they have the capabilities to develop proper mathematical models in order to study its phenomena.

In that sense, this paper addresses the question of what methodology should be used to estimate the behavior of renewable wind plants in any part of a country in the context of scarcity of information, which is an issue identified as one of the most critical challenges in developing countries.

The main contributions of the paper are:

1) To develop a novel methodology to generate hourly electricity production time-series for renewable wind plants, which:
   a) Only uses aggregate data as input, and
   b) Does not require historical time-series.
2) To achieve persistence characteristics and realistic stochastic behavior in generated time-series.

Ruben Felix and Jaime E. Luyo are with Facultad de Ingeniería Mecánica, Universidad Nacional de Ingeniería e-mail:rfelixr@uni.pe

3) To compare synthetic wind speed and energy against actual measurements and plants.

The remainder of the paper is organized as follows. Section II provides a comprehensive literature review of the approaches to synthesizing wind speed time-series. Section III presents and discusses the proposal to fill the central gap found in the previous section. Section IV performs a set of applications in different places worldwide, such as India, Chile, and Peru, and presents its comparison with historical values. Finally, concluding remarks and future work are highlighted in Section V.

II. Literature Review

Significant research effort has been made for modelling wind speed values. In that way, one of the first proposals was presented in [18], which compared six well-established approaches to generate synthetic wind time-series. According to their test results, authors concluded that methods based on independent and identically distributed values, one-step Markov models, two-step models, and Box-Jenkins models do not generate representative and accurate synthetic time-series. Therefore, recommend using the Shinozuka method or the embedded Markov chain model.

Better quality time-series were obtained in [19], where a Weibull distribution for each hour of a month's typical day is modeled to preserve the wind speed's daily pattern. The paper proves that the performance of an electrical system remains similar when using measured and synthesized time-series as evaluated by the loss of load probability (LLP). As a limitation, three years of data were required to model Weibull distributions.

It is essential to mention that a unique Weibull distribution cannot model an annual wind speed time series, as demonstrated in [20], where an overestimation of 40% in produced energy is obtained, besides producing a divergent probability density function (PDF).

Despite the findings presented in [18], plenty of works have been published later using supposedly not recommended methods. One of them is [21], where it is used a first-order Markov chain model to synthesize wind speed time-series obtaining good results with more than 90% of the agreement for statistical parameters between historical and synthetic series. However, to improve results, a higher-order Markov model must be used as recommended by the authors.

In [22], first- and second-order Markov chain models are employed to generate synthetic wind speed time-series for two localities in Malaysia. After comparing principal statistics such as mean, standard deviation, auto-correlation, Weibull distribution parameters, and spectral density of real and generated series, the authors concluded that both models have a good performance synthesizing wind speed time-series. Although the cited paper does not analyze similarity in wind speed profile, it must be said that neither the first- or second-order Markov model achieves reproducing actual behavior accurately. However, it is true that a higher-order or larger size transition matrix of Markov models better preserves statistical characteristics of historical data, as studied in [23].

An interesting analysis of the pitfalls of using Markov models to synthetically generate wind speed time-series to be later used in planning processes, specifically for energy storage planning, is presented in [24], where it is demonstrated that generated time-series lacks the persistence of actual data and would predict a radically different storage requirement.

Some works, as [25], stated that Markov models are good enough to generate hourly time-series since it preserves statistical parameters and that Weibull and Gaussian distributions should be used for intra-hourly values (minutes and seconds). However, results obtained demonstrate a lack of persistence in synthetic time-series.

It must be said that proposals using Markov models require historical measurements in order to calculate parameters.

A novel approach was presented by Negra *et al.* [26], who developed a model inspired by the foundation of Markov transition matrixes, obtaining very realistic wind speed time-series. However, for the calibration of the model, a set of 7 years of measurements was required.

The use of autoregressive moving-average (ARMA) models was also proposed to synthesize wind speed time-series, as in [27], where synthetic series for 15 New Zealand farms were generated, although methodology requires several years of wind registers over the country. Chen *et al.* [28] also produced a realistic wind speed time-series using Fourier series and ARMA models to characterize seasonal trends and auto-correlation in residue. Nevertheless, three years of historical data are employed to train the model.

An interesting approach was presented in [29], where a hybrid method that uses measured and synthetic values achieves creating wind speed time series for analyzing frequency (1 min) and flicker phenomena (10 sec). Starting from 10-min average wind speed values, the authors introduce zero-mean, high-frequency turbulence to create a high-resolution time series, which is then used to comprehensively analyze ramp rates and spinning reserve requirements, finding a good agreement with the actual results.

Similarly, Hagspiel *et al.* [30] proposed using a model based on Copula theory yielding good results that help perform steady-state analysis of power systems. A comparison between synthetic and natural energy generated by some existing wind farms in Europe was taken as a validation criterion. Regrettably, no graphs of real and generated time-series were provided in the publication, making it difficult to compare the reproduction accuracy of actual behavior. Another attempt using Copula theory was presented by Sarmiento *et al.* [31], who also addresses the synthesis of wind direction. The statistical validations against one-year measurements data obtained positive results; however, the wind speed profile does not reflect an actual wind speed time-series.

Meanwhile, Carapellucci *et al.* [32] presented a simple yet exciting methodology to produce synthetic wind speed time-series. Starting from the premise that wind speed comprises periodic deterministic and stochastic components, a five-step methodology is developed, supported by a genetic algorithm for tuning parameters. The paper proposal's main advantage is that the model only needs aggregate statistical parameters such as yearly mean, monthly mean, and monthly maximum wind speed. Generated time-series properly preserves actual wind behavior, although persistence is still not achieved.

Indeed, persistence is a characteristic that is not discussed in most of the mentioned papers, even though this characteristic can represent the difference between a completely random and a realistic wind speed time-series. An interesting approach to achieve this feature was presented by Naimo [33], who uses the optimization model of the Assignment Problem to re-sequence the generated values to accomplish persistence requirements. Nevertheless, an hourly measured wind speed data was needed to resolve this optimization problem.

A recent proposal was presented in [34] where a synthesis process with a special focus on preserving diurnal patterns is developed, i.e. persistence. The methodology initially generates random values that are then reordered to be consistent with the PDF and power spectral density function (PSD), and to accurately reproduce the average diurnal variation (ADV) of analyzed wind-speed datasets. Although the proposal generate different time-series in each run and does not depend on the time interval, the algorithm requires a wind speed dataset as input. Application case uses an 18-year long time-series recorded from a meteorological mast in the Netherlands.

One of the common challenges that almost all South America and other developing countries face is the scarcity of wind atlas or an official source of information about wind resources within its territory. Hence, obtaining measurement data at scales of 1-min, 5-min, or even 1-hr is almost impossible most of the time, making it very difficult to adopt some of the models proposed in developed countries, which make extensive use of historical data at the scale of 1-h or more frequently.

Therefore, the main objective of this paper, which is also its main contribution, is to develop a methodology to generate hourly electricity production time-series for wind renewable plants using aggregate data as input. This methodology can be helpful for project evaluation and system expansion planning purposes.

III. Methodology

The proposed methodology consists of four steps, as shown in Fig. 1.

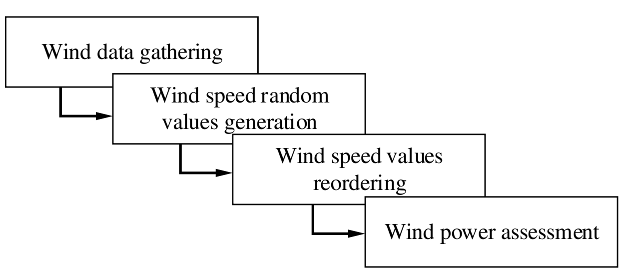

Fig. 1. Steps flow of synthesis methodology.

The first step gathers relevant wind data about a specific geographic coordinate. The second step shows the randomly generating process of wind speed values using the retrieved information. In step three, an optimization model is used to reorder the randomly generated values. Finally, the fourth step presents the process to obtain wind power generation values.

A. Wind Data Gathering

The Peruvian Ministry of Energy and Mines (MEM) released an interactive tool called Wind Atlas in 2016 [35]. For a given pair of latitude and longitude (ϕ and λ in °), this tool allows obtaining the yearly scale (β_y) and shape (α_y) parameters for the Weibull distribution approximation at 100 m height (h_{ref}) anywhere in the country.

The corresponding average wind speed and standard deviation ($\overline{v}_y$ and σ_y in m⁄s) can be obtained with the functions $NtWeibullMean(\alpha_y, \beta_y)$ and $NtWeibullStdev(\alpha_y, \beta_y)$, respectively, provided by the free Microsoft Excel add-in Ntrand [36].

Since the Peruvian Wind Atlas (PWA) does not provide monthly average wind speed values, the open-access POWER database provided by the American National Aeronautics and Space Administration (NASA) [37] will be used to acquire monthly averages ($\overline{v}^p_m$ in m⁄s). A parameter k_m is defined that relates these monthly values and their average, as indicated in Eq. 1. This parameter must be calculated for each month $m \in [1, M]$.

$$k_m = \frac{\overline{v}^p_m}{\frac{1}{M}\sum_{m=1}^{M} \overline{v}^p_m} \tag{1}$$

For other countries, the RE Data Explorer [38], a renewable energy resource geospatial explorer created by the National Renewable Energy Laboratory (NREL), could gather wind speed time-series at different heights. Although this explorer covers only a minor part of the world, it is a valuable tool to obtain free data.

If another source of data were used, e.g., in-site measurement, Weibull scale and shape parameters could be obtained:

1) using one of the methods presented in [39]–[41],
2) using the $NtWeibullParam(\mu, \sigma)$ function in Excel with the Ntrand add-in, where μ the mean and σ the standard deviation of speed values, or
3) using the $fitdistr(ws, densfun = "weibull", lower = 0)$ function in R [42] with package MASS [43], where ws is an array containing wind speed values

Monthly averages, as well, may be obtained from the same data source. Special care should be taken to use values of β_y, α_y, $\overline{v}^p_m$, and k_m referred to the same height h_{ref}.

Finally, the air density value (ρ in kg⁄m^3) must be calculated for the total hub height, which includes the wind tower height (h_{hub} in m) and the selected location's total altitude above sea level.

B. Wind Speed Random Values Generation

The inverse Weibull distribution will be employed to generate random values through Eq. (2):

$$v^u_{syn} = \vartheta \left(\ln \frac{1}{1-U} \right)^{1/\varphi} \tag{2}$$

Where v^u_{syn} represents the set of unordered random generated values and U is a uniform random number. Parameters ϑ and φ stand for the scale and shape, respectively.

In an Excel worksheet, a ($N \times M$) matrix must be constructed to obtain $N = 8760$ random values for every month ($M = 12$) of a typical year. The content of each matrix column should be the Ntrand matrix function $NtRandWeibull(N, \varphi, \vartheta, 0)$ considering $\varphi = \alpha_m$, and $\vartheta = \beta_m$ for the corresponding month column.

Since no values were recovered for these monthly parameters, they will be calculated using the matrix function $NtWeibullParam(\overline{v}_m, \sigma_m)$ having $\overline{v}_m = k_m\overline{v}_y$ and $\sigma_m = k_m\sigma_y$. These relationships represent a good approximation of actual parameter monthly variation as inferred from the data shown in [44]. Do not confuse the calculated parameter $\overline{v}_m$ with the gathered $\overline{v}^p_m$ value.

Two additional terms must be appended when calling the Ntrand Excel function to avoid the problem of producing identical random values for places with the same aggregate input parameters. The final formula would be $NtRandWeibull(\ldots, R_\phi, R_\lambda)$, where R_ϕ and R_λ are the last five digits of ϕ and λ starting from the right, which usually belongs to the decimal part when working with a precision of at least six digits. These constants are used in the function as random seeds.

Usually, h_{ref} is not the same as h_{hub}, so randomly generated values must be scaled using the power law presented in Eq. (3) suggested by [45].

$$\hat{v}^u_{syn} = v^u_{syn}\left(\frac{h_{hub}}{h_{ref}}\right)^{1/\gamma} \tag{3}$$

Where $\hat{v}^u_{syn}$ and v^u_{syn} are the wind speed values for heights h_{hub} and h_{ref}, respectively. Power law exponent γ, also known as Hellman's wind shear or friction exponent, could be approximated using one of the formulas presented in [46], [47], but since both NASA POWER and site measurements most frequently report wind speed values for more than one height, it is possible to clear the value of γ for each month using the monthly average speed values.

On the other side, since natural wind behavior has a sinusoidal-like form, the diurnal pattern formulation proposed by [48] will be adapted and applied to synthetic values.

$$v_{final,i} = v^o_{syn,i}\left(1 + \delta\cos\left[\frac{2\pi}{24}(i - \phi)\right]\right) \tag{4}$$

In Eq. (4), parameter $i \in [1, 24]$ represents the hours of the day, ϕ is the hour of the day at which peak wind speed used to occur, and δ is the diurnal pattern strength whose typical value goes from 0.0 to 0.4. v^o_{syn} is the corresponding synthetic value for the hour i, after being ordered using the algorithm presented next.

C. Wind Speed Values Reordering

Until the previous step, generated values represented a random set of numbers, as with almost all wind speed synthesis processes reviewed. However, it is possible to reorder the generated time-series to be realistic, attending to the persistence characteristics that wind speed has.

In that sense, an algorithm that uses the Assignment Problem optimization model is proposed in Algorithm 1, similar to [33].

Algorithm 1 Reordering algorithm

$a = 0$
$b = rand(20 : 32)$ ▷ Chooses a random size
while $a < N$ **do**
  $\hat{v}^{sub}_{syn} = \hat{v}^u_{syn}[a : b]$ ▷ Extracts a subset
  $c = rand(12 : 24)$ ▷ Chooses a random peak hour
  $L = length(\hat{v}^{sub}_{syn})$
  for j in $1 \ldots L$ **do**
    for i in $1 \ldots L$ **do**
      $disordered = \hat{v}^{sub}_{syn}[j]/\max\left(\hat{v}^{sub}_{syn}\right)$
      $pattern = 1 + \cos\left(\frac{2\pi}{L}(i - c)\right)$
      $distances[j, i] = \left|pattern^2 - disordered^2\right|$
    end for
  end for
  $\hat{v}^u_{syn}[a : b] = AssignmentProblem(\hat{v}^{sub}_{syn}, distances)$
  $a \mathrel{+}= b$
  if $a \geq N$ **then break** ▷ Finishes process
  $b = rand(20 : 32)$ ▷ Sets a new random size
  if $a + b > N$ **then** $b = N - a$
end while

This algorithm takes as input the values of $\hat{v}^u_{syn}$ and replaces them in chunks. In each iteration, a subset of the generated time-series is created. The size of this subset is chosen randomly between 20 and 32 hours, which tries to represent a natural "diary" cycle length. After that, this subset is compared with a sinusoidal pattern whose hour of peak speed is set by a random function between 12 and 24 since, typically, the wind blows stronger in the afternoon and night. The distance between the pattern and the normalized subset is calculated as suggested in [33].

With these values, the assignment problem is finally solved. As a result, it is obtained a reordered time-series which is used to replace the original correspondent values in $\hat{v}^u_{syn}$. This newly ordered time-series is represented by v^o_{syn}.

Parameters of the random functions can be revised to adapt to other locations' diary cycle patterns.

D. Wind Power Assessment

Although wind turbine manufacturers provide their products' power-wind speed (P-S) curves, no equation exists. In that sense, a piece-wise linearization function should be constructed to represent P-S curves [49].

$$P_W = \mathfrak{F}(v_{final})\frac{\rho}{\rho_0}\eta_W \tag{5}$$

Eq. (5) shows the power (P_W) calculation formula for a specific wind speed (v_{final}). In this equation, ρ is the air density defined in step one, ρ_0 is the standard air density at sea level, and $\mathfrak{F}$ is the piece-wise linearization function that models the P-S curve of the wind turbine.

Typical derating factor (η_W) values are in a range of 80-90% [44], [50]–[55] and reflect the losses caused due to wake effect, availability, electrical efficiency, turbine performance, environment, and curtailments.

TABLE I
DEVIATIONS OF MONTHLY MEAN WIND SPEED (M/S) VALUES AT h_{hub} FOR A POINT IN INDIA.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Actual	4.693	4.952	4.850	5.210	6.714	8.321	6.490	6.546	6.171	4.478	3.735	4.078
Synthetic	4.555	4.896	4.723	5.215	6.759	8.069	6.299	6.548	6.134	4.377	3.657	4.051
Error	-2.9%	-1.1%	-2.6%	0.1%	0.7%	-3.0%	-2.9%	0.0%	-0.6%	-2.2%	-2.1%	-0.7%

IV. VALIDATIONS

Three tests are performed for different places to validate the proposed methodology. Both tests compare the total energy produced when wind speed is converted by the simulation plant, which uses 42 wind turbines of the model Siemens SWT-2.3-108 [56], a turbine of 2.3 MW nominal power, with $h_{hub} = 80$ m and wind cut-in, cut-out, and rated speed values of 3, 25, and 11 m/s, respectively.

A. Test N° 1 – Comparison of Wind Energy for a Point in India

In this test, full-year measurement data of wind speed at different heights are gathered from the RE Explorer for a place in India located at $\phi = 25.44°$ and $\lambda = 78.57°$, which is 248 meters above sea level (MASL).

Parameters β_y and α_y are calculated using the measurement at 100 m (h_{ref}) as the average result of the R function indicated in the previous section and methods 1, 2, and 3 of [39]. Measurements at 40 and 80 m are used to clear the value of γ for each month. From the monthly mean wind speed at 40 m and cleared γ values, the corresponding $\overline{v}_m^p$ values at h_{ref} are calculated. Now, it is possible to estimate k_m.

Once model parameters are established, random values are generated, referred to h_{hub}, and ordered. A comparison between gathered (80 m) and ordered (synthetic) wind speed values are shown in Table I. It is seen that approximate values have a mean absolute percentage error (MAPE) of 1.6%, having a maximum discrepancy in June when variation reaches up to -3.0%, which is still acceptable. The corresponding normalized root-mean-square deviation (nRMSD) is 2.1%.

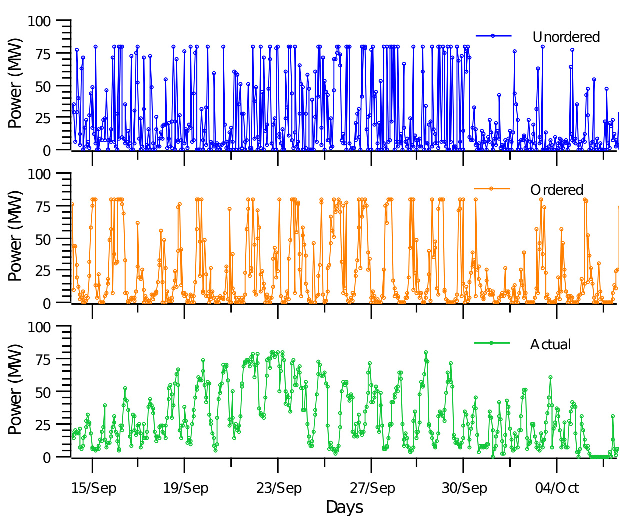

Fig. 2. Time-series of total wind power output (MW) for a point in India.

The power generation profiles produced by the simulation plant when exposed to the unordered, ordered (synthetic) and, actual wind speed time-series are presented in Fig. 2.

It is found that the time-series before performing the re-ordering process have completely erratic behavior. After this process, a more realistic time-series is successfully obtained whose diary cycle coincides acceptably with the actual series. However, since the proposed model only uses monthly aggregate parameters, it is found that synthetic values do not present weekly trends, as is evident in the actual one.

An interesting comparison of the daily wind speed profiles of synthetic and actual time-series is presented in Fig. 3. Although a similar daily pattern is revealed, it can be noticed that historical time-series have a smoother behavior and more dispersed values during the whole day, with a notable exception near the hour 19. In contrast, dispersion is minor and almost the same along the year in the synthetic time-series.

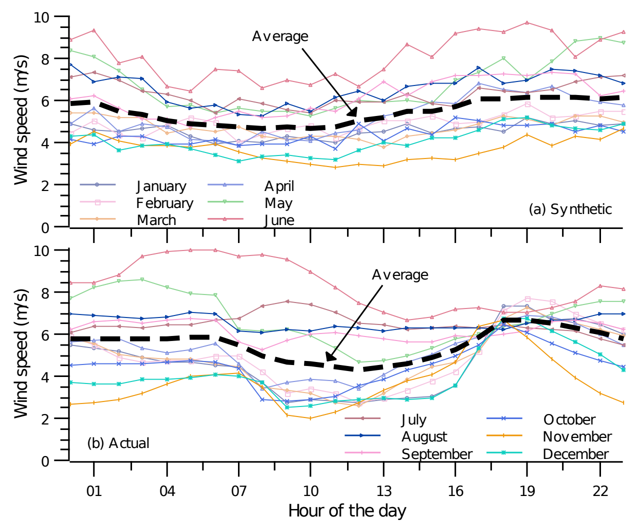

Fig. 3. Monthly and annual daily profile of wind speed (m/s) for a point in India.

In terms of total monthly wind energy produced, the proposed methodology achieves good results in general, as shown in Table II, where estimated annual production is -5.6% lower than the natural generation.

The corresponding MAPE is 8.3%, while the nRMSD is 13.7%. The maximum discrepancies occur in June and November, with -17.8% and 20.6%, respectively. However, it must be noticed that almost all deviations are within the root-mean-square error (RMSE), which is 2,117.01 MWh. For instance, the error in November is just 783.65 MWh.

The June's deviation of -6,307.08 MWh could be explained by two principal factors: 1) model parameters were derived from a unique measurement that does not represent year-to-

TABLE II
DEVIATION OF TOTAL MONTHLY WIND ENERGY OUTPUT (MWH) VALUES FOR A POINT IN INDIA.

	Actual	Synthetic	Error
January	10,445.60	9,298.15	-11.0%
February	11,372.26	9,729.97	-14.4%
March	11,382.02	10,400.56	-8.6%
April	13,069.84	13,220.91	1.2%
May	23,178.49	22,987.15	-0.8%
June	35,351.94	29,044.86	-17.8%
July	22,610.76	20,006.27	-11.5%
August	21,399.94	22,009.06	2.8%
September	17,773.97	18,839.08	6.0%
October	7,906.84	7,679.01	-2.9%
November	3,798.43	4,582.08	20.6%
December	6,819.88	6,919.43	1.5%
Total	185,109.99	174,716.53	-5.6%
Mean	15,425.83	14,559.71	

year resource variability, and 2) the occurrence of an atypical daily pattern as evidenced in Fig. 3-b.

It is essential to mention that $\overline{v}_m^p$ values were inferred from measurements at 40 m, even though data were available in RE Data Explorer in order to evaluate the methodology performance in a scenario where no wind speed time-series at h_{ref} exists.

This test uses a value of $\phi = 22$, $\delta = 0.2$, and $\eta_W = 85\%$.

B. Test N° 2 – Comparison of Wind Energy for a Point in Chile

A place in Chile located at $\phi = -24.65°$ and $\lambda = -70.24°$, with an altitude of 2,588 MASL, is analyzed in this second test. Data of annual measurements of wind speed at different heights are retrieved from [57].

Similar to the previous test, β_y and α_y are calculated using the time-series of wind speed at 50 m (h_{ref}). Then, γ is cleared each month using the monthly mean wind speeds at 10 and 40 m. Next, using this value and the monthly mean speed at 10 m, $\overline{v}_m^p$ values at h_{ref} are calculated.

A comparison of ordered (synthetic) and gathered values, both scaled at h_{hub}, is presented in Table III. Approximate values have a MAPE of 1.1% and an nRMSD of 1.5%. Maximum deviations occurred in February and November, reaching 2.3% and -4.7%, respectively. The latter deviation is most likely related to the fact that actual wind speed does not always follow the smooth vertical profile suggested by the Hellman law.

The hourly behavior of power generation when using the unordered, ordered, and actual wind speed time-series over the simulation plant are shown in Fig. 4. It can be seen that, again, the first output of the methodology corresponds to a completely random series. Nevertheless, after the reordering process, synthetic values achieve a realistic behavior, even obtaining attractive curves such as the one produced for August 11, which shows a pattern also appearing in the natural series. Grantham *et al.* [58] indicate that this kind of methodology aims not to produce patterns that have occurred in the recorded data but to generate patterns that are equally as likely to occur, which is being accomplished.

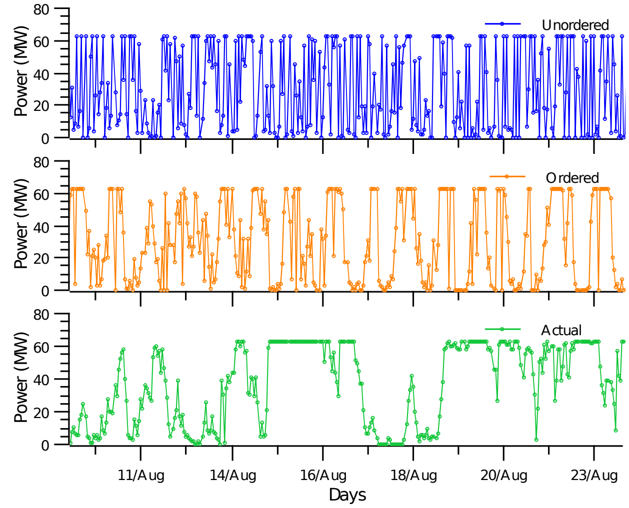

Fig. 4. Time-series of total wind power output (MW) for a point in Chile.

Although it is widely accepted that the daily profile of wind speed fits a sinusoidal-like form, it can be seen in Fig. 5 that this Chilean point hardly accomplishes this convention. In this figure, the spiky daily profile of the synthetic values is also evident, in contrast with the smooth curves of the natural series. In part (b) of the figure, it is notorious, as well, that July is far above the annual mean, although June and October also replicate this trend. Synthetic values mimic this characteristic, too, mainly between 6 and 18 hours.

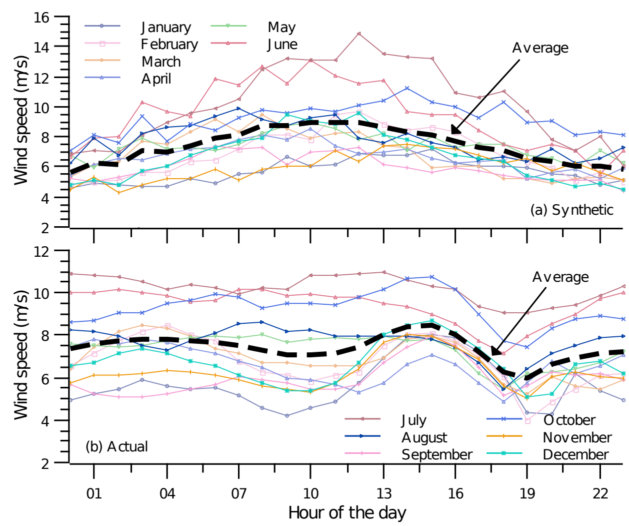

Fig. 5. Monthly and annual daily profile of wind speed (m/s) for a point in Chile.

Table IV shows the simulation plant's energy when exposed to the synthetic (ordered) and actual wind speed time-series. The monthly variations between both series produce a MAPE of 4.2%, with maximum divergences in January (5.9%) and July (-11.4%). Estimated annual production is 2.1% lower than the actual value.

July's error reaches -3,663.31 MWh, higher than the RMSE, which is 1,262.02 MWh. The corresponding nRMSD is 6.2%. Explaining factors are 1) month's speeds are higher than the

TABLE III
DEVIATIONS OF MONTHLY MEAN WIND SPEED (M/S) VALUES AT h_{hub} FOR A POINT IN CHILE.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Actual	5.688	6.756	6.894	6.587	7.337	9.394	10.212	7.702	5.963	9.265	6.265	6.681
Synthetic	5.684	6.911	6.876	6.671	7.391	9.445	10.239	7.829	5.954	9.144	5.972	6.688
Error	-0.1%	2.3%	-0.3%	1.3%	0.7%	0.5%	0.3%	1.7%	-0.1%	-1.3%	-4.7%	0.1%

TABLE IV
DEVIATION OF TOTAL MONTHLY WIND ENERGY OUTPUT (MWH) VALUES FOR A POINT IN CHILE.

	Actual	Synthetic	Error
January	12,547.62	13,287.62	5.9%
February	16,525.08	16,823.45	1.8%
March	19,579.51	18,759.00	-4.2%
April	16,451.77	17,333.11	5.4%
May	21,838.07	21,177.52	-3.0%
June	26,015.00	25,655.07	-1.4%
July	32,140.78	28,477.46	-11.4%
August	23,289.61	22,459.02	-3.6%
September	14,595.80	14,283.17	-2.1%
October	26,930.13	27,670.87	2.8%
November	15,543.86	14,243.40	-8.4%
December	18,047.32	18,158.87	0.6%
Total	243,504.55	238,328.55	-2.1%
Mean	20,292.05	19,860.71	

annual value, and 2) a flat daily profile, instead of sinusoidal, as analyzed previously.

This test uses a value of $\phi = 11$, $\delta = 0.2$, and $\eta_W = 85\%$.

C. Test N° 3 – Comparison of Wind Energy with an Existing Wind Farm in Peru

The third test simulates an existing wind farm in Peru called Tres Hermanas, which is located at $\phi = -16.39°$ and $\lambda = -75.08°$ and has an altitude of 217 MASL.

Weibull parameters β_y and α_y at 100 m (h_{ref}) are given by PWA, and POWER gives monthly mean wind speed values at 10 and 50 m. From these values, γ is cleared, and then $\overline{v}^p_m$ at h_{ref} is calculated starting from values at 10 m.

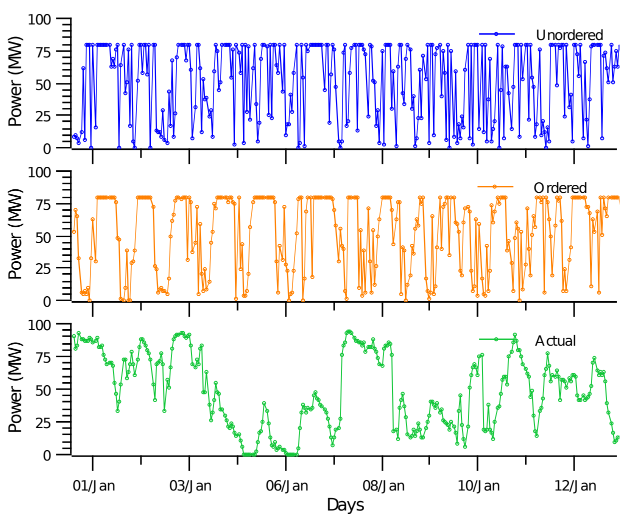

Fig. 6. Time-series of total wind power output (MW) for a point in Peru.

Fig. 6 depicts the approximate power generated using the wind speed time-series unordered and ordered. It also shows the values injected into the grid in 2017. It can be seen that peaks and valleys between the ordered and actual series coincide acceptably. However, sub-monthly trends are not captured by the methodology.

When daily profiles of power generation of these time-series are revised in Fig. 7, it is evidenced that synthetic time-series have a noisier behavior, although the annual mean is quite similar. Tuning this profile would require working with hourly parameters instead of monthly.

This fact is not necessarily a bad aspect of the proposed methodology whenever it produces a more exigent-to-integrate wind farm time-series. Indeed, developing planning studies considering this kind of synthetic time-series could guarantee that when actual implementation is done, the power system will not get into a more stressful situation than it was considered when simulated, which in essence, represents the worst case.

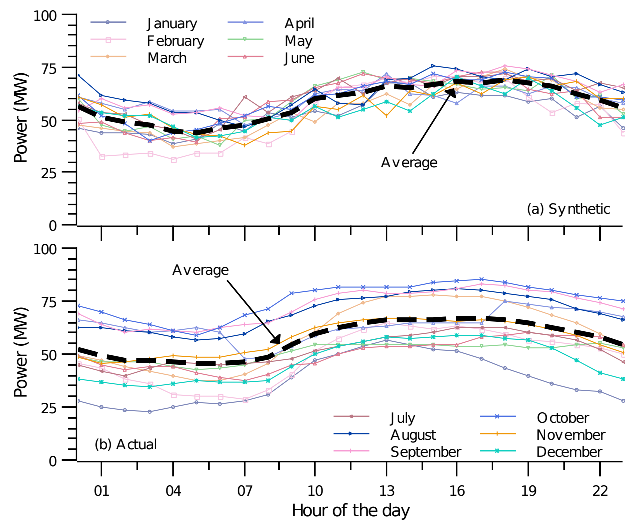

Fig. 7. Monthly and annual daily profile of wind power (MW) for a point in Peru.

Regarding energy, the methodology yields an annual amount higher by 6.2% than the average of measured values of 2016, 2017, and 2018, as shown in Table V. Month-to-month variations have a MAPE of 8.1%. Comparisons to each year independently produce measures between 11% and 14.5%. Regarding the nRMSD, values of 12.8%, 15.8%, and 15.3% are found for 2016, 2017, and 2018, respectively. However, compared to the average, this value is reduced to 9.8%, which is expected.

As shown in Fig. 8, actual monthly energy production fluctuates year after year, with some exceptions in July and

TABLE V
DEVIATION OF TOTAL MONTHLY WIND ENERGY OUTPUT (MWH) VALUES FOR A POINT IN PERU.

	Average	Synthetic	Error
January	32,119.13	39,340.20	22.5%
February	29,769.70	34,978.23	17.5%
March	40,884.02	40,548.28	-0.8%
April	41,615.48	42,975.77	3.3%
May	42,490.23	43,293.23	1.9%
June	36,113.21	41,967.99	16.2%
July	38,053.91	45,320.03	19.1%
August	45,491.20	46,922.11	3.1%
September	44,592.62	44,895.43	0.7%
October	46,599.45	44,286.17	-5.0%
November	39,327.17	40,588.14	3.2%
December	39,266.72	40,716.75	3.7%
Total	476,322.84	505,832.32	6.2%
Mean	39,693.57	42,152.69	

November, where production is quite the same. Synthetic injection for January obtains the maximum deviation in Table V; nevertheless, this value is similar to the one that occurred in 2018. Likewise, the situation that occurred in 2017 during August, September, and October, where actual production surpassed the values of the other years, makes the synthetic injections obtained for June and July conceivable.

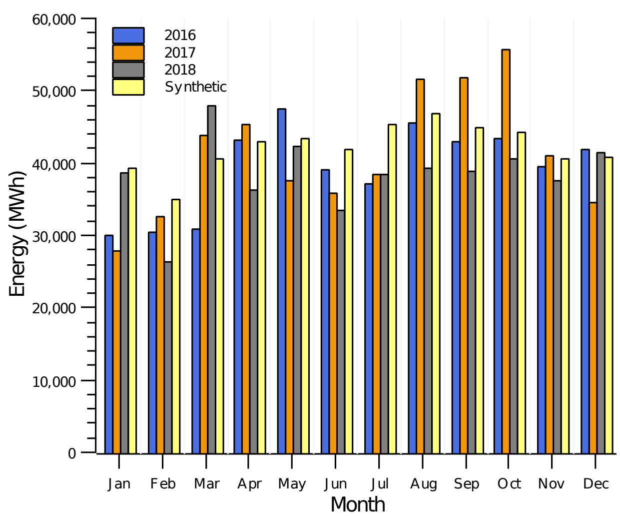

Fig. 8. Total monthly wind energy output (MWh) for a point in Peru.

Actual energy values originally correspond to an installed power of 97.15 MW [59], so they were scaled to the simulation plant size.

This test uses a value of $\phi = 17$, $\delta = 0.2$, and $\eta_W = 85\%$.

V. CONCLUSIONS

The proposed methodology presents a simple but effective approach for producing synthetic hourly production values for renewable wind plants.

According to the results presented, the randomness of the proposed methodology initially produces a time-series in which variability is not realistic. This unrealistic time-series can be used as a worst-case scenario because such variability makes it hard to integrate its corresponding wind plant into a non-flexible power system. Such a scenario could be a positive attribute for planning activities but, in contrast, could increase the expansion plan cost. This behavior is corrected by the reordering algorithm, which gives a more realistic profile.

The produced energy over or underestimates actual production within a variation range of $\pm 6.2\%$. This aspect should be considered when using these synthetic time-series to perform medium and long-term planning.

It should be remembered that the proposed methodology does not try to adjust natural wind curves but to generate realistic probable ones.

The presented methodology is novel because it uses aggregate parameters as input and does not require historical time-series, which is suitable for developing countries lacking RER information. Besides, this flexible and parametric methodology can generate multiple time-series scenarios modifying aggregate input parameters to achieve enough range of cases to incorporate uncertainties that may be used in future research work.

Future studies must be carried out to improve sub-monthly trends and estimated monthly injections.

ACKNOWLEDGMENT

We want to thank the Concytec-ProCiencia for their support to the research activities in the field of energy at the Mechanical Engineering School of the Universidad Nacional de Ingeniería (UNI).

REFERENCES

[1] IRENA, "Global energy transformation: A roadmap to 2050 (2019 edition)," tech. rep., IRENA, 2019.
[2] IRENA, "Renewable Power Generation Costs in 2021," tech. rep., 2022.
[3] IEA, "World Energy Investment 2022," tech. rep., 2022.
[4] M. R. R. Tabar, M. Anvari, G. Lohmann, D. Heinemann, M. Wächter, P. Milan, E. Lorenz, and J. Peinke, "Kolmogorov spectrum of renewable wind and solar power fluctuations," *The European Physical Journal Special Topics*, vol. 223, pp. 2637–2644, 10 2014.
[5] M. Anvari, G. Lohmann, M. Wächter, P. Milan, E. Lorenz, D. Heinemann, M. R. R. Tabar, and J. Peinke, "Short term fluctuations of wind and solar power systems," *New Journal of Physics*, vol. 18, p. 063027, 6 2016.
[6] D. K. Critz, S. Busche, and S. Connors, "Power systems balancing with high penetration renewables: The potential of demand response in Hawaii," *Energy Conversion and Management*, vol. 76, pp. 609–619, 12 2013.
[7] COES, "Technical Procedure PR-01 - Short-Term Ooperation programming," 2014.
[8] COES, "Technical Procedure PR-37 - Medium-term operation programming," 2016.
[9] F. A. Wolak, "Long-Term Resource Adequacy in Wholesale Electricity Markets with Significant Intermittent Renewables," tech. rep., National Bureau of Economic Research, Cambridge, MA, 7 2021.
[10] A. Heydari, D. Astiaso Garcia, F. Keynia, F. Bisegna, and L. De Santoli, "A novel composite neural network based method for wind and solar power forecasting in microgrids," *Applied Energy*, vol. 251, p. 113353, 10 2019.
[11] Ã. B. Filik and T. Filik, "Wind Speed Prediction Using Artificial Neural Networks Based on Multiple Local Measurements in Eskisehir," *Energy Procedia*, vol. 107, pp. 264–269, 2 2017.
[12] C. Fu, G.-Q. Li, K.-P. Lin, and H.-J. Zhang, "Short-Term Wind Power Prediction Based on Improved Chicken Algorithm Optimization Support Vector Machine," *Sustainability*, vol. 11, p. 512, 1 2019.
[13] T. Ouyang, X. Zha, L. Qin, Y. He, and Z. Tang, "Prediction of wind power ramp events based on residual correction," *Renewable Energy*, vol. 136, pp. 781–792, 6 2019.
[14] H. C. Byllling, S. Pineda, and T. K. Boomsma, "The impact of short-term variability and uncertainty on long-term power planning," *Annals of Operations Research*, vol. 284, no. 1, pp. 199–223, 2020.

[15] Y. Feng, *Scenario generation and reduction for long-term and short-term power system generation planning under uncertainties*. PhD thesis, Iowa State University, Ames, 2014.
[16] S. Tabrizian, "Technological innovation to achieve sustainable development - Renewable energy technologies diffusion in developing countries," *Sustainable Development*, vol. 27, pp. 537–544, 5 2019.
[17] S. Sen and S. Ganguly, "Opportunities, barriers and issues with renewable energy development - A discussion," *Renewable and Sustainable Energy Reviews*, vol. 69, pp. 1170–1181, 3 2017.
[18] F. C. Kaminsky, R. H. Kirchhoff, C. Y. Syu, and J. F. Manwell, "A Comparison of Alternative Approaches for the Synthetic Generation of a Wind Speed Time Series," *Journal of Solar Energy Engineering*, vol. 113, no. 4, p. 280, 1991.
[19] H. G. Beyer and K. Nottebaum, "Synthesis of long-term hourly wind speed time series on the basis of European wind atlas data," *Solar Energy*, vol. 54, pp. 351–355, 5 1995.
[20] R. S. R. Gorla, M. K. Pallikonda, and G. Walunj, "Use of Rayleigh Distribution Method for Assessment of Wind Energy Output in Cleveland-Ohio," *Renewable Energy Research and Applications*, vol. 1, no. 1, pp. 11–18, 2020.
[21] A. D. Sahin and Z. Sen, "First-order Markov chain approach to wind speed modelling," *Journal of Wind Engineering and Industrial Aerodynamics*, 2001.
[22] A. Shamshad, M. A. Bawadi, W. M. Wan Hussin, T. A. Majid, and S. A. Sanusi, "First and second order Markov chain models for synthetic generation of wind speed time series," *Energy*, 2005.
[23] F. O. Hocaoglu, Ã. N. Gerek, and M. Kurban, "The Effect of Markov Chain State Size for Synthetic Wind Speed Generation," *Probabilistic Methods Applied to Power Systems, 2008. PMAPS'08. Proceedings of the 10th International Conference*, 2008.
[24] K. Brokish and J. Kirtley, "Pitfalls of modeling wind power using Markov chains," in *2009 IEEE/PES Power Systems Conference and Exposition, PSCE 2009*, 2009.
[25] A. N. Legesse, A. K. Saha, and R. P. Carpanen, "Generating wind speed time series for time domain simulation of wind turbines," in *5th Southern African Universities Power Engineering Conference*, (Stellenbosch), 2017.
[26] N. B. Negra, O. Holmstrøm, B. Bak-Jensen, and P. Sørensen, "Model of a synthetic wind speed time series generator," *Wind Energy*, 2008.
[27] R. Turner, X. Zheng, N. Gordon, M. Uddstrom, G. Pearson, R. de Vos, and S. Moore, "Creating Synthetic Wind Speed Time Series for 15 New Zealand Wind Farms," *Journal of Applied Meteorology and Climatology*, vol. 50, pp. 2394–2409, 12 2011.
[28] J. Chen and C. Rabiti, "Synthetic wind speed scenarios generation for probabilistic analysis of hybrid energy systems," *Energy*, vol. 120, pp. 507–517, 2 2017.
[29] S. Rose and J. Apt, "Generating wind time series as a hybrid of measured and simulated data," *Wind Energy*, vol. 15, pp. 699–715, 7 2012.
[30] S. Hagspiel, A. Papaemannouil, M. Schmid, and G. Andersson, "Copula-based modeling of stochastic wind power in Europe and implications for the Swiss power grid," *Applied Energy*, 2012.
[31] C. Sarmiento, C. Valencia, and R. Akhavan-Tabatabaei, "Copula autoregressive methodology for the simulation of wind speed and direction time series," *Journal of Wind Engineering and Industrial Aerodynamics*, 2018.
[32] R. Carapellucci and L. Giordano, "A methodology for the synthetic generation of hourly wind speed time series based on some known aggregate input data," *Applied Energy*, 2013.
[33] A. Naimo, "A Novel Approach to Generate Synthetic Wind Data," *Procedia - Social and Behavioral Sciences*, 2014.
[34] D. D. Ambrosio, J. Schoukens, T. D. Troyer, M. Zivanovic, and M. C. Runacres, "Synthetic wind speed generation for the simulation of realistic diurnal cycles," *Journal of Physics: Conference Series*, vol. 1618, p. 062019, 9 2020.
[35] MEM Peru, "Peruvian Wind Atlas," 2016.
[36] Numerical Technologies, "Ntrand," 2022.
[37] NASA, "POWER," 2022.
[38] NREL, "RE Explorer," 2022.
[39] C. G. Justus, W. R. Hargraves, A. Mikhail, and D. Graber, "Methods for Estimating Wind Speed Frequency Distributions," *Journal of Applied Meteorology*, vol. 17, pp. 350–353, 3 1978.
[40] K. S. P. Kumar and S. Gaddada, "Statistical scrutiny of Weibull parameters for wind energy potential appraisal in the area of northern Ethiopia," *Renewables: Wind, Water, and Solar*, vol. 2, p. 14, 12 2015.
[41] N. Aghbalou, A. Charki, S. Elazzouzi, and K. Reklaoui, "A probabilistic assessment approach for wind turbine-site matching," *International Journal of Electrical Power & Energy Systems*, vol. 103, pp. 497–510, 12 2018.
[42] R Core Team, "R: A Language and Environment for Statistical Computing," 2017.
[43] W. N. Venables and B. D. Ripley, *Modern Applied Statistics with S*. New York: Springer, fourth ed., 2002.
[44] A. Colmenar-Santos, S. Campíez-Romero, L. Enríquez-Garcia, and C. Pérez-Molina, "Simplified Analysis of the Electric Power Losses for On-Shore Wind Farms Considering Weibull Distribution Parameters," *Energies*, vol. 7, pp. 6856–6885, 10 2014.
[45] IEC, "IEC 61400-12-1:2017 Wind energy generation systems - Power performance measurements of electricity producing wind turbines," tech. rep., 2017.
[46] J. F. Manwell, J. G. McGowan, and A. L. Rogers, *Wind Energy Explained: Theory, Design and Application, Second Edition*. 2010.
[47] D. Spera and T. Richards, "Modified Power Law Equations for Vertical Wind Profiles," *Wind Characteristics and Wind Energy Siting Conference*, 1979.
[48] HOMER, "HOMER Pro User Manual," 2022.
[49] M. Lydia, S. S. Kumar, A. I. Selvakumar, and G. E. Prem Kumar, "A comprehensive review on wind turbine power curve modeling techniques," *Renewable and Sustainable Energy Reviews*, vol. 30, pp. 452–460, 2 2014.
[50] E. Diaz-Dorado, C. Carrillo, J. Cidras, and E. Albo, "Estimation of energy losses in a Wind Park," in *2007 9th International Conference on Electrical Power Quality and Utilisation*, pp. 1–6, IEEE, 10 2007.
[51] P. K. Steimer and O. Apeldoorn, "Medium voltage power conversion technology for efficient windpark power collection grids," in *The 2nd International Symposium on Power Electronics for Distributed Generation Systems*, pp. 12–18, IEEE, 6 2010.
[52] N. Inaba, R. Takahashi, J. Tamura, M. Kimura, A. Komura, and K. Takeda, "A consideration on loss characteristics and annual capacity factor of offshore wind farm," in *2012 XXth International Conference on Electrical Machines*, pp. 2022–2027, IEEE, 9 2012.
[53] A. Madariaga, C. Martinez de Ilarduya, S. Ceballos, I. Martinez de Alegria, and J. Martin, "Electrical losses in multi-MW wind energy conversion systems," *Renewable Energy and Power Quality Journal*, pp. 322–327, 4 2012.
[54] G. S. Böhme, E. A. Fadigas, A. L. Gimenes, and C. E. Tassinari, "Wake effect measurement in complex terrain - A case study in Brazilian wind farms," *Energy*, vol. 161, pp. 277–283, 10 2018.
[55] EWEA, "Wind Farm Energy Loss Factors," tech. rep., European Wind Energy Association, Brussels, Belgium, 2019.
[56] Siemens, "Siemens Wind Turbine SWT-2.3-108," tech. rep., Erlangen, Germany, 2011.
[57] ME Chile, "Wind and Solar resource measurement campaign," 2019.
[58] A. P. Grantham, P. J. Pudney, and J. W. Boland, "Generating synthetic sequences of global horizontal irradiation," *Solar Energy*, vol. 162, no. November 2017, pp. 500–509, 2018.
[59] COES, "Technical Specification of Power Plants," 2022.

Ruben Felix received a B.Sc. in Mechanical Electrical Engineering in 2013 and an M.Sc. in Energetics in 2019 from the UNI, Peru. Specialized in Machine Learning by the Universidad Ricardo Palma (URP), Peru and in Solar Energy by the National Institute of Solar Energy (NISE), India.

He is currently a Ph.D. researcher at UNI with interest in renewable energies, energy planning and planning optimization.

Jaime E. Luyo holds a Ph.D. in Economics from the Universidad Nacional Mayor de San Marcos (UNMSM), Peru. Mechanical Electrical Engineer, graduated with "Unanimous Distinction" at UNI. Master in Electrical Engineering (Systems and Control) by the Rensselaer Polytechnic Institute, New York, USA.

Former Director of the Doctoral Program in Energetics at UNI and former Vice Minister of Electricity of Peru.