Journal of
Renewable Energy Research and Applications (RERA)

Vol. 4, No. 2, 2023, 171-181 *DOI: 10.22044/rera.2022.12305.1173*

A Method to Synthesize Hourly Time-series of Solar Power Plants in Peru for Long-Term Planning

R. Felix[1*] and L. Sayas[2]

1. Facultad de Ingeniería Mecánica, Universidad Nacional de Ingeniería, Rímac, Lima, Peru.
2. División de Supervisión de Electricidad, OSINERGMIN, Magdalena del Mar, Lima, Peru.

Received Date 30 September 2022; Revised Date 08 November 2022; Accepted Date 23 December
**Corresponding author: rfelixr@uni.pe (R. Felix)*

Abstract
Plenty of works have treated the system expansion planning problem in the presence of intermittent renewable energy resources like solar. However, most of those proposals have been approached from scenarios of plenty of data, which is not the rule in developing countries, where the principal investment actors have recently switched their focus. In contrast of operation problems where the 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 solar production time-series on an hourly time scale, taking as inputs aggregate data such as monthly average, maximum or minimum values of basic parameters like global horizontal insolation, air temperature, and surface albedo. The methodology comprises five steps, from data gathering to calculating electrical power produced by a solar photovoltaic system. Three application tests are performed for different places in Chile, Slovakia, and Peru to validate the proposed methodology. The results show that the methodology successfully synthesizes time-series of output power, correctly replicates typical solar resource behavior, and slightly underestimates the produced solar energy, having a discrepancy of 2.4% in the yearly total.

Keywords: *Expansion Planning, Power Systems, Renewable Energy, Solar Energy, Synthesis Methods.*

1. Introduction

The recent reports have shown that the investment cost of renewable generation technologies continues to decrease, in contrast to their efficiencies, which get higher year to year [1–3]. According to [1], renewable energy resources (RER) will account for 86% of the global electricity generation, while 49% of the share in final consumption will use electricity by 2050. Although several benefits to human development will come with this higher RER penetration level, energy agents shall face many challenges to smooth this energy transition.

As studied in [4–7], power systems get vulnerable to RER generation fluctuation when the penetration level increases. In this scenario, the independent system operator (ISO) experiences critical load demand ramp up or down created by short-period fluctuations in solar irradiance, which reveals the necessity of must-run machines or energy storage systems in the grid to absorb variability. Also ISO finds it very difficult to perform the day-ahead dispatch scheduling due to the unpredictable availability of this kind of plant's generation capacity. Although the ISO's responsibility is to forecast this time-series accurately, technical complexity sometimes obligates it to assume a deterministic generation prepared by plant owners, as evidenced in [8, 9] for the Peruvian case. This situation affects the physical and wholesale electricity markets, as studied in [10]. Several approaches have been proposed [11–18] to overcome these problems.

Although RER uncertainties affect operation activities to some relevant degree, as detailed before, system expansion planning problems are at a higher level of complexity since it involves a farther time horizon analysis [19, 20].

In that sense, the availability of information on resource measurements and existing plants' historical records is a key to successfully

overcoming these problems. Unfortunately, the developing countries that are recently experiencing more investment expenditures in RER plants usually lack this information. Even though private developers conduct field studies to profile their projects, the gathered information is not socialized with academia and public organisms, which must appeal to alternative sources to develop their studies.

Recent publications on the existing barriers to renewable energy in developing countries strongly coincide with these points: the lack of capabilities to develop proper mathematical models to study these problems [21] and information barriers to access to databases on the potential of these resources [22].

Therefore, this paper addresses the question of what methodology should be used to estimate the behavior of solar photovoltaic (PV) renewable plants in any part of a country in the context of information scarcity, 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 solar PV renewable plants that
   a. only use aggregate data as input, and
   b. Do not require historical time-series
2. To achieve realistic stochastic behavior in generated time-series
3. To compare synthetic solar irradiance and energy against actual measurements and plants

The remainder of the paper is organized as what follows. Section 2 provides a comprehensive literature review of the approaches presented to synthesize solar time-series. Section 3 presents and discusses in detail the proposal to fill the central gap found in the previous section. Section 4 performs a set of applications in different places worldwide such as Chile, Slovakia, and Peru; and presents its comparison with historical values. Finally, concluding remarks and future work are highlighted in Section 5.

2. Literature review

Significant research effort has been made for modeling solar hourly radiation values. In that sense, one of the first proposals was presented in [23], where hourly insolation values' daily and annual periodicity was removed using a Fourier analysis, and time-dependent frequency distribution (TDFD) was employed to synthesize insolation values. Likewise, Graham et al. [24] presented an autoregressive-moving average (ARMA) technique to generate synthetic values starting from monthly mean values of the clearness index (k_T).

A novel approach was described in [25], demonstrating that k_T was the variable that induces the randomness in the series, in contrast with previously published research that treated solar irradiance as a random variable. Then an ARMA model for k_T was used to generate synthetic irradiance values.

However, the previously mentioned works coincide in presenting complex statistical analyses and models that became too difficult to understand and implement in practical applications.

On the other side, Hontoria et al. [26] developed an artificial intelligence (AI) model to synthesize solar hourly radiation time-series. Although formulation does not require a deep statistical analysis of variables' interrelation, it is necessary to have historical time-series values to train the AI model. As reviewed in [27], AI models that do not need historical solar measurements require historical values for many other parameters.

Another reasonable attempt was presented by Celik [28], where an energy output model was used to generate synthetic hourly radiation values. Although energy output is coherent with historical measurements, the results showed that the power generation profile did not reflect the actual stochastic solar behavior.

An interesting approach was developed by Polo et al. [29], who, through a simple model based on a beta distribution, generates a 10-min solar irradiance time-series. Validation of the model showed that daily and monthly means between actual and synthetic series coincided and that the 10-min solar irradiance profile preserves the actual behavior acceptably. However, the proposed procedure needs a historical hourly solar irradiance time-series to start.

Laslett et al. [30] developed a simple algorithm to generate hourly solar radiation values implemented using a web page with the same aim. The algorithm needs as arguments distances instead of coordinates, and starting from that, hourly values are generated. Nonetheless, algorithm equations were explicitly defined for the southwest region of Western Australia. Thus applicability to other places is limited and must be carried out carefully since model calibration depends on comprehensive historical measurement data.

Another major group of works looks to generate synthetic time-series at high temporal resolutions. In that line, Ngoko et al. [31] proposed a model

that used a second-order Markov transition matrix (MTM) to synthesize a 1-min time resolution time-series. While final model only needs as input a daily k_T, it is needed data at the same time scale of 1-min to calibrate the model.

Also Grantham et al. [32] presented a simple approach to generate 5-min global horizontal irradiance (GHI) and direct normal irradiance (DNI) time-series. The method uses 1-min observed data (re-scaled to 5-min) to calibrate the model, which then is used to interpolate hourly mean solar irradiance values obtaining good results though validation is performed only in the Australian territory. The authors suggest that the method would be suitable for applying to Typical Meteorological Year (TMY) records.

However, the same authors later recommend using another proposed attempt that combined Fourier series, autoregressive models, and white noise terms. This new attempt synthesizes coherent daily and hourly time-series since it was demonstrated that the generated series included patterns that had not occurred in the recorded data but were equally as likely to occur, thus better suits evaluation and planning requirements [33]. The application of the described method needs to have historical data to force the generated series' distribution.

Recently, a novel approach to downscale DNI time-series from 1-h to 1-min that can be applied in any location without requiring local adaptation was presented in [34]. After training the model with 14 years of measured 1-min DNI data, it only needs an hourly DNI time-series for any location. Indeed, downscaling proposals have gain attention within developed countries in recent years [35, 36].

Wang et al. [37] proposed an interesting approach to predict long-term time-series combining a variety of decomposition and reconfiguration methods. The authors achieve generating time-series for 10 years in advance, using the same amount of registers. A related approach was presented in [38], where the empirical mode decomposition (EMD) was complemented with deep learning (DL) techniques to improve the performance of the method. A case study in Korea to guide the development of sustainable energy policies is developed, using registers of 4 years to synthesize 1 year of data.

The lack of a solar atlas or an official source of information about solar resources within the territory of almost all of South America and other developing countries makes it very difficult to obtain measurement data at scales of 1-min, 5-min, or even 1-hr. As a result, adopting the models proposed in developed countries, which make extensive use of historical data, is almost impossible most of the time.

For that reason, 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 solar PV renewable plants using aggregate data as input. This methodology can be helpful for project evaluation and system expansion planning purposes.

3. Methodology

The proposed methodology consists of five steps, as shown in figure 1.

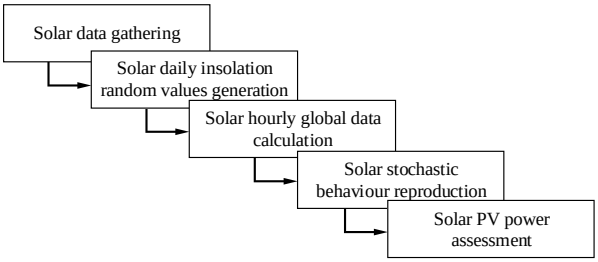

Figure 1. Steps flow of synthesis methodology.

The first step gathers relevant solar data for a specific coordinate. Then it is used in the second step to generate random daily insolation values. Subsequently, the third step calculates the hourly global radiation incident on the PV array starting from the previous values. The fourth step models the stochastic behavior of clouds. Finally, step five converts solar radiation into PV power values.

3.1. Solar data gathering

Available open-access information about solar irradiance and insolation in Peru is limited to very few sources. On the one hand, the Peruvian Ministry of Energy and Mines (MINEM) released a non-interactive solar map in 2003 [39] that showed average monthly solar insolation values for the whole country. However, the ability to extract precise values for any given coordinate is limited since it was delivered as static images instead of being implemented over a geographic information system (GIS) platform.

On the other hand, some international companies and organizations provide this kind of information. While many of them are private commercial services (e.g. Meteonorm or Solargis), there is a couple of open-access services provided by the American National Aeronautics and Space Administration (NASA) called POWER [40] and by the European Commission called PVGIS [41]. This paper

employs five parameters provided by POWER service and one by PVGIS.

For a given pair of latitude and longitude (ϕ and λ in $^{\circ}$), the monthly mean daily global horizontal insolation ($\overline{H}_m$ in kWh/m^2/day) is obtained from PVGIS (average month of all available years), while the maximum and minimum variations of $\overline{H}_m$ (H_m^{max} and H_m^{min} in %), the maximum and minimum air temperature at 2 m (T_m^{max} and T_m^{min} in ^{o}C), and the surface albedo ($\rho_{g,m}$) values are retrieved from POWER. These parameters must be obtained for each month $m \in [1, M]$.

3.2. Solar daily insolation random value generation

Random values for global horizontal insolation (H_{syn}) are generated using Ntrand [42], which is a free Microsoft Excel add-in that uses pseudo-uniform random numbers generated by Mersenne Twister algorithm which achieves a long period of 2^{19937}-1 and high order of equidistribution.

One of the distributions supported by Ntrand is the Truncated Normal Distribution shown in (1), which is proposed to be employed in this paper.

$$\mathbf{H_{syn} = \Phi^{-1}\left(\Phi(\vartheta_m) + U\left(\Phi(\zeta_m) - \Phi(\vartheta_m)\right)\right)\sigma_m + \mu_m} \qquad (1)$$

where Φ represents the cumulative distribution function of standard normal distribution, and U is a uniform random number. Parameters ϑ_m and ζ_m are the normalized values of the lower (g) and upper (t) bounds of truncated normal for a given mean (μ) and standard deviation (σ).

In an Excel worksheet, a $(N \times M)$ matrix must be constructed to obtain $N = 8760$ random values for every month ($M = 12$). The content of each matrix column should be the Ntrand matrix function $\mathrm{NtRandTruncnorm}(N, g, t, \mu, \sigma, 0)$ considering $\mu = \overline{H}_m$, $g = \overline{H}_m(1 + H_m^{max})$, and $t = \overline{H}_m(1 - H_m^{min})$ for the corresponding month column.

Since no value was recovered for the standard deviation of $\overline{H}_m$, it will be assumed equal to the multiplication of a constant k by the mean value ($\sigma = k\mu$). In this paper, the value of k is set to 0.5.

Therefore, random generation process results depend on the value of these four parameters. From these, g and t could be treated as quasi-fixed values since they come from historical measures. On the other side, μ and k values should be considered factors that introduce uncertainty in the process and could be used to generate distinct synthetic time-series to represent multiple scenarios for the stochastic behavior of the solar resource.

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 $\mathrm{NtRandTruncNorm}(\ldots, R_\phi, R_\lambda)$, 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.

Although different values for each hour of a typical year are obtained after this process, the generated values are still a daily insolation value, which must be converted into hourly values.

3.3. Solar hourly global radiation calculation

This third step explains the calculation flow required to convert the global daily horizontal radiation synthesized in the second step into hourly total radiation over a tilted solar panel. This flow consists of four primary relations, which must be applied following the order presented in figure 2.

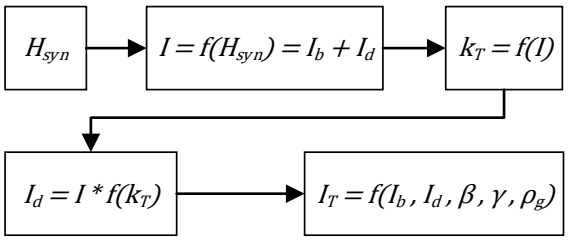

Figure 2. Calculation flow to convert Daily Horizontal Radiation (H_{syn}) into Hourly Total Radiation (I_T).

To start, it should be clear that the total radiation incident on any tilted surface corresponds to the sum of beam radiation (I_b), diffuse radiation (I_d), and albedo or ground reflected radiation.

However, when working with horizontal radiation values, the albedo component is neglected due to this physical disposition.

In that sense, the estimated hourly global horizontal radiation value (I) obtained through (2) from H_{syn} will account just for beam and diffuse components, i.e. $I = I_b + I_d$.

$$\mathbf{r_t = I / H_{syn}} \qquad (2)$$

The definition of r_t is explained in the Appendix, from (S1) to (S8).

After this first relation, the value of I is obtained. This value could be related to the extraterrestrial horizontal radiation (I_o) by the ratio k_T as indicated in (3). Special care should be taken to consider non-negative values for I, and therefore, for k_T.

$$\mathbf{k_T = I/I_o} \quad (3)$$

The equations needed for the calculus of I_o are given in the Appendix, from (S9) to (S10).
This second relation gives us the value of k_T for each hour. This index can then define the hourly diffuse fraction, which relates I_d and I, as in (4).

$$\mathbf{I_d/I = f(k_T)} \quad (4)$$

where $f(k_T)$ represents a piecewise function whose definition is presented in (S11) of the Appendix.
After applying this third relation, the value of I_d is obtained. With this value, it is possible to clear the value of the beam radiation since $I_b = I - I_d$.
As suggested by [43], for PV panels placed in the southern hemisphere, the HDKR anisotropic model is suitable. For the sake of simplicity, suggested model will be used to calculate the total incident radiation on a tilted surface with slope β (in rad) and surface azimuth angle γ (in rad), as defined in (5). For further studies, a selection process could be carried out to choose the best model [44].

$$\mathbf{I_T = (I_b + I_d A_i)R_b + I_d(1 - A_i)\left(\frac{1+\cos\beta}{2}\right)\left[1 + f\sin^3\left(\frac{\beta}{2}\right)\right] + I\rho_{g,m}\left(\frac{1-\cos\beta}{2}\right)} \quad (5)$$

Complementary equations for this fourth relation are shown in the Appendix, from (S12) to (S16).

3.4 Solar stochastic behavior reproduction

The effect of shadows over the panel is represented by a factor f_{CLD}. Although clouds interrupt, I_b, I_d continues to be present. Hence, only this component is affected by f_{CLD}, as indicated in (6).

$$\mathbf{I_b^* = f_{CLD} I_b} \quad (6)$$

f_{CLD} is generated randomly for each hour using the algorithm presented in table 1. Note that global radiation hour-to-hour variation is achieved using step 2, so additional noise is introduced in this step.

Table 1. Noise generator algorithm.

```
aux = 0
for h in 1..N do
    if aux ≥ rand_1^h then
        f_CLD^h = aux
        aux = 0
    Else
        f_CLD^h = 1
        aux+= rand_2^h
    end if
end for
```

$rand_1$ and $rand_2$ are random numbers between 0 and 1 generated for each hour h by the non-volatile Excel matrix function NTRAND(N).
Consequently, I_b^* should be used in replace of I_b for all subsequent equations, including (5).

3.5 Solar PV power assessment

Air temperature varies within a day; however, the available data from the first step only provides each month's minimum and maximum values. A relationship between global radiation and air temperature has to be used to approximate the air temperature value for each hour of the day.
Hence, proposal of [45] is adapted, as shown in (7).

$$\mathbf{\overline{H}_m/\overline{H}_{o,m} = c_m \cdot ((T_m^{max} - T_m^{min})/T_m^{min})^{0.5}} \quad (7)$$

According to [45], c_m is a function of the altitude and distance to the sea, and its value is determined using historical data. In this paper, c_m will be estimated using available data from step one.
The monthly mean daily extraterrestrial radiation ($\overline{H}_{o,m}$) can be evaluated using (8):

$$\mathbf{\overline{H}_{o,m} = \frac{24}{\pi} G_{on}(\cos\varphi\cos\delta\sin\omega_s + \omega_s\sin\varphi\sin\delta)} \quad (8)$$

Extraterrestrial radiation should be calculated for the recommended average day for each month indicated in [43]. Finally, since [45] define this parameter as a constant, the average value $c = \sum c_m/M$ will be used for the following calculations.
Prieto's proposed model [45] is then extended in this paper to calculate hourly values. Therefore, the air temperature for a specific hour h is determined by (9). It is assumed that T_m^{min} is the same for every day of the corresponding month and that the maximum temperature solved from the equation for each hour is the representative air temperature for that hour.

$$\mathbf{T_h = \left((I/I_o)\cdot(T_m^{min})^{0.5}\cdot 1/c\right)^2 + T_m^{min}} \quad (9)$$

Since the temperature has an inertial behavior, the temperature hour value and the previous hour value are weighted using a factor z. In that sense, the final ambient temperature value is defined, as shown in (10). In this paper, the value of z is set to 0.75.

$$\mathbf{T_h^a = zT_h + (1 - z)T_{h-1}^a} \quad (10)$$

Now, it is possible to calculate cell temperature for each hour (T_h^c) using (11), which is the inferred equation presented in [46].

$$T_h^c = \frac{T_h^a + (T_{NOCT}^c - T_{NOCT}^a)\left(\frac{G_T}{G_{T,NOCT}}\right)\left[1 - \frac{\eta_{STC}^{mp}(1-\alpha_p T_{STC}^c)}{\tau\alpha}\right]}{1 + (T_{NOCT}^c - T_{NOCT}^a)\left(\frac{G_T}{G_{T,NOCT}}\right)\left(\frac{\alpha_p \eta_{STC}^{mp}}{\tau\alpha}\right)} \quad (11)$$

Nominal operating cell temperature (NOCT) is defined for solar irradiance $G_{T,NOCT} = 0.8$ kW/m^2 and ambient temperature $T_{NOCT}^a = 20$ °C. Values for cell temperature under NOCT conditions (T_{NOCT}^c in °C), efficiency at maximum power point (η_{STC}^{mp} in %), temperature coefficient of power (α_p in %/°C) and cell temperature (T_{STC}^c in °C) under standard test conditions (STC) can be found in the solar PV module datasheet.

Although solar irradiance striking the PV panel (G_T) is an instantaneous value, the hourly total incident radiation (I_T) value obtained in step 3 will be used here, assuming that flat radiation occurs for the entire hour bin.

Finally, (12) is used to calculate the output power of the PV panel:

$$P_{PV} = P_{STC} f_{PV} \left(\frac{G_T}{G_{T,STC}}\right)\left[1 + \alpha_p (T_h^c - T_{STC}^c)\right] \quad (12)$$

Rated capacity of PV panel under STC (P_{STC} in W) is obtained from the module datasheet. Solar irradiance for STC ($G_{T,STC}$) is equal to 1 kW/m^2.

In this equation, a derating factor (f_{PV}) is employed to account for losses occasioned by soiling, mismatching, transformation, degradation, etcetera. Do notice that losses by irradiance level, temperature, and clouds are internalized in the proposed procedure, so f_{PV} should only account for the other factors. Its value strictly depends on the place conditions where the solar PV panel is mounted such as the level of dust and rain, among others. Typical values range between 75-95% [47–54].

4. Validations

Three application tests are performed for different places to validate the proposed methodology. The first test evaluates the accuracy of the proposed data sources for a point in Chile. Then second test compares the methodology output until the fourth step with a historical time-series of irradiance in Slovakia. Finally, in third test an existing PV power plant in Peru is simulated through the proposed methodology, and its results are contrasted with actual values.

4.1 Test N° 1 – comparison of solar data sources for a point in Chile

Both POWER and PVGIS databases employ complex models to approximate solar radiation for a wide range worldwide. Hence, these values are not historical in-site measurements but a mathematical approximation.

In order to quantify how accurate is the data provided by these platforms, figure 3 compares them with values obtained from the Solar Explorer published by the Ministry of Energy of Chile [55], which has a normalized root-mean-square deviation (nRMSD) of 5.7% for the northern part of Chile (between latitudes -17° and -30°).

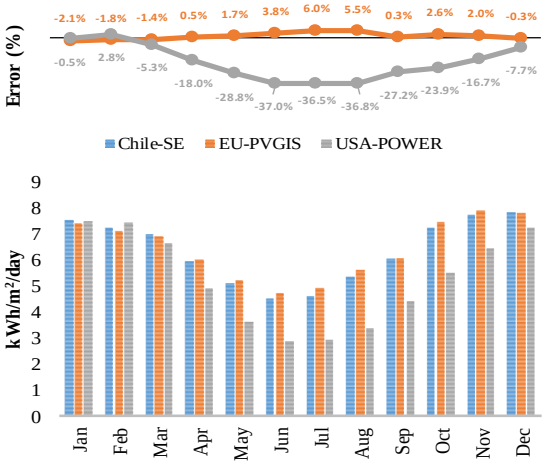

Figure 3. Comparison of $\overline{H}_m$ gathered from different solar data sources for a point in Chile.

Values for the comparison are gathered for a place located at $\phi = -18.37°$ and $\lambda = -70.14°$.

As shown in figure 3, values offered by PVGIS are significantly more accurate than those provided by POWER. Indeed, PVGIS dataset has a mean absolute percentage error (MAPE) of 2.3%, while POWER ranges 20.1%.

On the one hand, NASA's data display information lower by more than 10% than Chilean's values for 8 out of 12 months, having deviations above 23% for months from May to October, with a peak in June where the error reaches -37%.

On the other hand, PVGIS's data remains between an error of ±2.6% for 9 out of 12 months. Maximum discrepancy occurs on July with an error of 6%, followed by a deviation of 5.5% on August.

In that sense, it is validated that using PVGIS to obtain information about $\overline{H}_m$ should be preferred over POWER, as proposed in step one.

This test used the solar radiation database called PVGIS-NSRDB for years 2005-2015.

4.2 Test N° 2 – comparison of I_T for a point in Slovakia

The second test validates the quality of synthetic hourly values generated. Hence, historical hourly

measurement data is used to compute the total monthly, and annual global horizontal irradiance received, then is compared with the amounts produced with the synthesized I_T time-series.

The available data of Slovakia for the years 2014, 2017, and 2020 was downloaded from [56]. Geographic coordinates correspond to $\phi = 49.03°$ and $\lambda = 20.32°$.

Figure 4 represents the daily behavior of the measured and synthetic values for the recommended average day for each month indicated in [43], although both time-series correspond to a full year.

The plotted series have been ordered to allow a correct visualization.

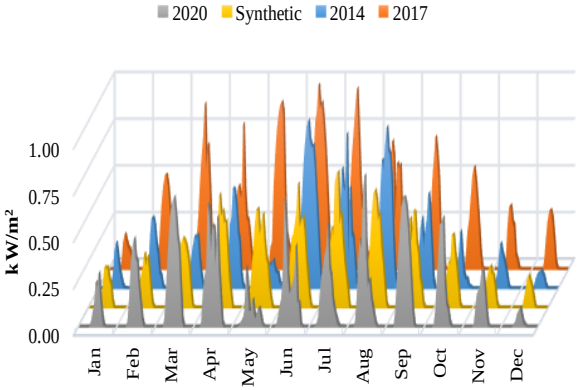

Figure 4. Time-series of I_T for a point in Slovakia.

As seen, there are some days like April, July or August where synthetic and historical daily time-series are very similar to each other. But also exists days like May, October or December, where not even historical values coincide between them. In fact, this figure evidences that solar resource variability is such that radiation for the same day differs from year to year.

This figure also demonstrates that the proposed methodology correctly replicates typical solar resource behavior, although synthetically generated values barely coincide with some historical days. However, as mentioned by Grantham et al. [33], the objective of this kind of methodologies is not to produce patterns that have occurred in the recorded data, but to generate patters that are equally as likely to occur.

Regarding the monthly insolation, figure 5 reflects that historical total radiation received each month fluctuates year after year. Despite this, synthetic amounts seem similar to actual values for some months like March, September or December.

The monthly radiation received calculated using the synthetic I_T time-series has a MAPE of 11.9%, 12.5%, and 11.7% when compared with actual values of years 2014, 2017, and 2020, respectively.

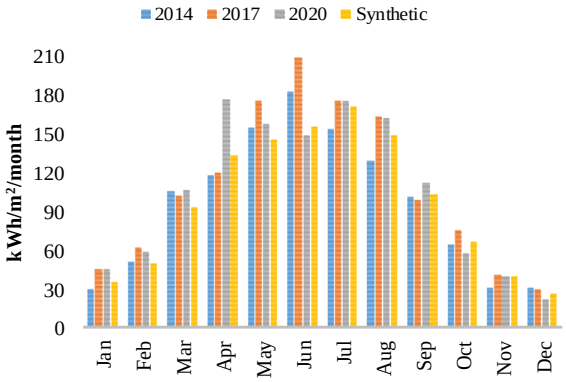

Figure 5. Comparison of total monthly insolation for a point in Slovakia.

Nevertheless, if a mean aggregates the three historical hourly time-series, resulting amounts are closer to the synthetic ones, as shown in table 2.

Table 2. Deviations of total monthly insolation values for a point in Slovakia.

	Average	Synthetic	Variation
January	39.48	34.72	-12.0%
February	56.49	49.02	-13.2%
March	104.50	92.86	-11.1%
April	137.86	132.86	-3.6%
May	162.31	145.40	-10.4%
June	179.64	155.21	-13.6%
July	167.94	170.96	1.8%
August	151.26	148.45	-1.9%
September	103.25	102.89	-0.3%
October	65.00	65.49	0.7%
November	36.53	39.02	6.8%
December	26.81	25.49	-4.9%
Total	1231.07	1162.38	-5.6%
Mean	102.59	96.86	

Specifically, deviation gets reduced to a MAPE of 6.7%.

Significant variations values are obtained for some months like January or February; however, these values respond to a numerical issue instead of a methodology instability, since average values are rather small. If the root-mean-square error (RMSE) is calculated, a value of 9.78 kWh/m^2 is obtained, which contains, for instance, the difference between synthetic and average value for January (4.75 kWh/m^2).

The nRMSD, which achieves 9.5% when comparing synthetic values to the average of historical measurements. Individual comparisons produce values of 13.6%, 18.3% and 14.7% for years 2014, 2017, and 2020, respectively.

These results validate the proposed methodology until step four and also suggest that to deal with uncertainties, synthetic series should be generated

considering a variation of ±9.5% the values of $\overline{H}_m$.
This test used the solar radiation database called PVGIS-SARAH2 for years 2005-2020.

4.3 Test N° 3 – comparison of solar energy with an existing solar park in Peru

The third test corroborates the results of all steps. In that sense, the historical power output of an existing PV plant in Peru is used to evaluate the quality of synthetic time-series generated for a similar plant.
The studied plant is Majes Solar Park, located at $\phi = -16.44°$ and $\lambda = -72.22°$.
This plant uses fixed PV panels sloped at 15° facing the north. Although the existing plant comprises modules of 350, 370, 390, and 410 W manufactured by TSolar, simulation is made using technical specifications of Hanwha Solar Q.Peak Duo L-G5.2 for a rated power of 380 W. A total of 57 894 panels are considered for the simulation, giving a total installed power of 22 MW, as it is in reality.
Synthetized solar resource for the plant's location is good and almost stable throughout the year, as shown in figure 6 (yellow bars).
The yearly average daily total insolation received over a PV panel, calculated using I_T time-series, is 6,365 Wh/m^2/day. The greatest deviations from the yearly mean are found in June, with a value lower by 11.3% (5,646 Wh/m^2/day), and in November, where radiation achieves 10.4% more than the average (7,027 Wh/m^2/day).
Figure 6 also plots the values of $\overline{H}_m$ retrieved in step one (blue line). It is seen that even though these values are taken as inputs, the proposed methodology achieves a significant variability (notice orange marks of minimum and maximum synthetic daily values in each month column) yielding monthly averages different from data gathered. Discrepancies between gathered and calculated averages fluctuates in a range of ±15%, with a corresponding MAPE of 8.4%.

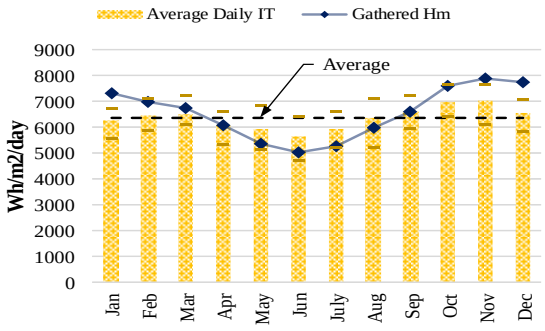

Figure 6. Monthly average daily insolation over a PV panel for a point in Peru.

As analyzed in the previous test, this behavior is expected and desired for the solar resource since historical registers usually differs year to year.
The solar park's total monthly energy produced, actual and synthetic, are reported in table 3. The variation between both series produces a MAPE of 4.9%, with maximum divergences in April (-9.3%) and July (9%).

Table 3. Deviation of total monthly solar energy output (MWh) values for a point in Peru.

	Actual	Synthetic	Variation
January	3630.41	3512.84	-3.2%
February	3345.25	3246.72	-2.9%
March	3474.82	3624.32	4.3%
April	3615.47	3280.31	-9.3%
May	3210.09	3319.55	3.4%
June	3164.76	3074.66	-2.8%
July	3061.85	3338.34	9.0%
August	3806.71	3551.88	-6.7%
September	3818.65	3555.02	-6.9%
October	4070.51	3888.83	-4.5%
November	3841.99	3774.42	-1.8%
December	3811.57	3667.79	-3.8%
Total	42852.08	41834.70	-2.4%
Mean	3571.01	3486.22	

However, when looking at the total annual energy injected by the solar plant into the grid, the difference between both values results 2.4%, being that the proposed methodology slightly underestimates this energy contribution. The nRMSD of the synthetic series is 5.4%.
Besides validating total energy output between actual and synthetic values, a comparison between output power profiles is also required. Figure 7 displays the power output time-series for the first and second weeks of the year.
It can be noticed that synthetic series achieves reproducing stochastic behavior of solar resources, without overemphasizing erratic behavior and being coherent with measurements.
This sample also evidences that significant erratic injections occur in existing solar PV plants, as seen in days 4, 6, 10 or 12, although other plotted days present this phenomenon as well.
For the simulation, a value of 86% was considered for f_{PV}.

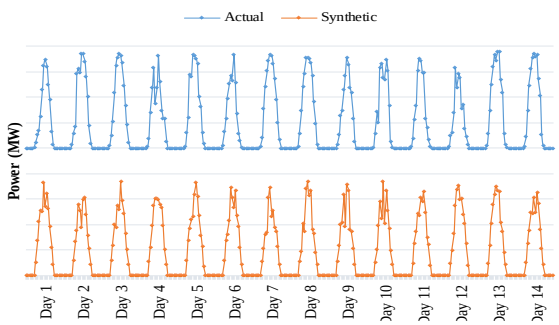

Figure 7. Time-series of total solar power output (MW) for a point in Peru.

This test used the solar radiation database called PVGIS-NSRDB for years 2005-2015.

5. Conclusions

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

According to the results presented, the methodology successfully synthesizes the time-series of output power, which correctly replicates typical behavior of solar resource.

The results show that the proposed methodology slightly underestimates the calculus of produced solar energy, having a discrepancy of 2.4% in the yearly totals.

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

Comparison with the existing solar plants could help estimate the expected error for future uses of synthetic time-series.

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 find alternatives ways to estimate c_m presented in (7), since it will fail when analyzing places where historical minimum temperatures were negative.

This methodology can help to implement a software to improve the access to RER information in order to accelerate the adoption of distributed solar PV generation in developing countries, and to enable further research in the field of smart grids, as stated in [57].

6. Appendix

See Supplement for supporting content.

7. 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 UNI.

8. References

[1] IRENA (2019). Global energy transformation: A roadmap to 2050 (2019 edition).

[2] IRENA (2022). Renewable Power Generation Costs in 2021.

[3] IEA (2022). World Energy Investment 2022.

[4] Tabar, M.R.R. et al. (2014). Kolmogorov spectrum of renewable wind and solar power fluctuations. The European Physical Journal Special Topics. Vol. 223, No. 12, pp. 2637–2644.

[5] Mossoba, J. et al. (2012). Analysis of solar irradiance intermittency mitigation using constant DC voltage PV and EV battery storage. 2012 IEEE Transportation Electrification Conference and Expo (ITEC), 2012.

[6] Anvari, M. et al. (2016). Short term fluctuations of wind and solar power systems. New Journal of Physics. Vol. 18, No. 6, p. 063027.

[7] Traube, J. et al. (2013). Mitigation of Solar Irradiance Intermittency in Photovoltaic Power Systems With Integrated Electric-Vehicle Charging Functionality. IEEE Transactions on Power Electronics. vol. 28, no. 6, pp. 3058–3067.

[8] COES (2014). Technical Procedure PR-01-Short-Term Ooperation programming.

[9] COES (2016). Technical Procedure PR-37-Medium-term operation programming.

[10] Wolak, F.A. (2021). Long-Term Resource Adequacy in Wholesale Electricity Markets with Significant Intermittent Renewables. Cambridge, MA.

[11] Dyson, J. et al. (2017). Utility scale solar short term generation forecasting for improved dispatch and system security. 16th Wind Integration Forum, Berlin, 2017.

[12] Dong, J. et al. (2020). Novel stochastic methods to predict short-term solar radiation and photovoltaic power. Renewable Energy. Vol. 145, pp. 333–346.

[13] Heydari, A. et al. (2019). A novel composite neural network based method for wind and solar power forecasting in microgrids. Applied Energy. Vol. 251, p. 113353.

[14] Paulescu, M. and Paulescu, E. (2019). Short-term forecasting of solar irradiance. Renewable Energy. Vol. 143, pp. 985–994.

[15] Caldas, M. and Alonso-Suárez, R. (2019). Very short-term solar irradiance forecast using all-sky imaging and real-time irradiance measurements. Renewable Energy. vol. 143, pp. 1643–1658.

[16] Brancucci Martinez-Anido, C., et al. (2016). The value of day-ahead solar power forecasting improvement. Solar Energy. Vol. 129, pp. 192–203.

[17] Behera, M.K., Majumder, I., and Nayak, N. (2018). Solar photovoltaic power forecasting using optimized modified extreme learning machine technique. Engineering Science and Technology, an International Journal. Vol. 21, No. 3, pp. 428–438.

[18] Fentis, A. et al. (2019). Short-term nonlinear autoregressive photovoltaic power forecasting using statistical learning approaches and in-situ observations.

International Journal of Energy and Environmental Engineering. Vol. 10, No. 2, pp. 189–206.

[19] Bylling, H.C., Pineda, S., and Boomsma, T.K. (2020). The impact of short-term variability and uncertainty on long-term power planning. Annals of Operations Research. Vol. 284, No. 1, pp. 199–223.

[20] Feng, Y. (2014). Scenario generation and reduction for long-term and short-term power system generation planning under uncertainties. Iowa State University.

[21] Tabrizian, S. (2019). Technological innovation to achieve sustainable development - Renewable energy technologies diffusion in developing countries. Sustainable Development. Vol. 27, No. 3, pp. 537–544.

[22] Sen, S. and Ganguly, S. (2017). Opportunities, barriers and issues with renewable energy development – A discussion. Renewable and Sustainable Energy Reviews. Vol. 69, pp. 1170–1181.

[23] Balouktsis, A. and Tsalides, P. (1986). Stochastic simulation model of hourly total solar radiation. Solar Energy. Vol. 37, No. 2, pp. 119–126.

[24] Graham, V.A., Hollands, K.G.T., and Unny, T.E. (1988). A time series model for Kt with application to global synthetic weather generation. Solar Energy.

[25] Graham, V.A. and Hollands, K.G.T. (1990). A method to generate synthetic hourly solar radiation globally. Solar Energy.

[26] Hontoria, L., Aguilera, J., and Zufiria, P. (2002). Generation of hourly irradiation synthetic series using the neural network multilayer perceptron. Solar Energy.

[27] Mellit, A. (2008). Artificial Intelligence technique for modelling and forecasting of solar radiation data: a review. International Journal of Artificial Intelligence and Soft Computing.

[28] Celik, A.N. (2003). Long-term energy output estimation for photovoltaic energy systems using synthetic solar irradiation data. Energy.

[29] Polo, J. et al. (2011). A simple approach to the synthetic generation of solar irradiance time series with high temporal resolution. Solar Energy.

[30] Laslett, D., Creagh, C., and Jennings, P. (2014). A method for generating synthetic hourly solar radiation data for any location in the south west of Western Australia, in a world wide web page. Renewable Energy.

[31] Ngoko, B.O., Sugihara, H., and Funaki, T. (2014). Synthetic generation of high temporal resolution solar radiation data using Markov models. Solar Energy. Vol. 103, pp. 160–170.

[32] Grantham, A.P. et al. (2017). Generating synthetic five-minute solar irradiance values from hourly observations. Solar Energy. Vol. 147, pp. 209–221.

[33] Grantham, A.P., Pudney, P.J., and Boland, J.W. (2018). Generating synthetic sequences of global horizontal irradiation. Solar Energy. Vol. 162, No. November 2017, pp. 500–509.

[34] Larrañeta, M. et al. (2018). Methodology to synthetically downscale DNI time series from 1-h to 1-min temporal resolution with geographic flexibility. Solar Energy. Vol. 162, No. October 2017, pp. 573–584.

[35] Zhang, W. et al. (2018). A stochastic downscaling approach for generating high-frequency solar irradiance scenarios. Solar Energy. Vol. 176, pp. 370–379.

[36] Frimane, Â. et al. (2019). Non-parametric Bayesian-based recognition of solar irradiance conditions: Application to the generation of high temporal resolution synthetic solar irradiance data. Solar Energy. Vol. 182, pp. 462–479.

[37] Wang, S.-Y., Qiu, J., and Li, F.-F. (2018). Hybrid Decomposition-Reconfiguration Models for Long-Term Solar Radiation Prediction Only using Historical Radiation Records. Energies. Vol. 11, No. 6, p. 1376.

[38] Nam, K., Hwangbo, S., and Yoo, C. (2020). A deep learning-based forecasting model for renewable energy scenarios to guide sustainable energy policy: A case study of Korea. Renewable and Sustainable Energy Reviews. Vol. 122, p. 109725.

[39] MEM Peru (2003). Peruvian Solar Energy Atlas, Available: http://dger.minem.gob.pe/atlassolar/.

[40] NASA (2022). NASA POWER - Prediction of Worldwide Energy Resources, Available: https://power.larc.nasa.gov/.

[41] JRC (2022). JRC Photovoltaic Geographical Information System (PVGIS)-European Commission, Available: https://re.jrc.ec.europa.eu/pvg_tools/en/.

[42] Numerical Technologies (2022). Ntrand, Available: http://www.ntrand.com/.

[43] Duffie, J.A. and Beckman, W.A. (2013). Solar Engineering of Thermal Processes: Fourth Edition.

[44] Shukla, K.N., Rangnekar, S., and Sudhakar, K. (2015). Comparative study of isotropic and anisotropic sky models to estimate solar radiation incident on tilted surface: A case study for Bhopal, India. Energy Reports.

[45] Prieto, J.I., Martínez-García, J.C., and García, D. (2009). Correlation between global solar irradiation and air temperature in Asturias, Spain. Solar Energy. Vol. 83, No. 7, pp. 1076–1085.

[46] HOMER (2022). HOMER Pro User Manual, Available: https://homerenergy.com/products/pro/docs/latest/.

[47] Shiva Kumar, B. and Sudhakar, K. (2015). Performance evaluation of 10 MW grid connected

solar photovoltaic power plant in India. Energy Reports. Vol. 1, pp. 184–192.

[48] Gostein, M., Caron, J.R., and Littmann, B. (2014). Measuring soiling losses at utility-scale PV power plants. 2014 IEEE 40th Photovoltaic Specialist Conference (PVSC), 2014.

[49] Zorrilla-Casanova, J. et al. (2011). Analysis of dust losses in photovoltaic modules.

[50] Micheli, L., Deceglie, M.G., and Muller, M. (2018). Mapping Photovoltaic Soiling Using Spatial Interpolation Techniques. IEEE Journal of Photovoltaics. pp. 1–6.

[51] Deceglie, M.G., Micheli, L., and Muller, M. (2018). Quantifying Soiling Loss Directly From PV Yield. IEEE Journal of Photovoltaics. Vol. 8, No. 2, pp. 547–551.

[52] Caron, J.R. and Littmann, B. (2012). Direct monitoring of energy lost due to soiling on first solar modules in California. 2012 IEEE 38th Photovoltaic Specialists Conference (PVSC) PART 2, 2012.

[53] IFC (2015). Utility-Scale Solar Photovoltaic Power Plants: A project Developer's Guide. Washington, D.C.

[54] Dobos, A.P. (2014). PVWatts Version 5 Manual.

[55] ME Chile (2017). Chilean Solar Explorer, Available: https://solar.minenergia.cl/exploracion.

[56] WMO (2022). World Radiation Data Centre, Available: http://wrdc.mgo.rssi.ru/wrdc_en.htm.

[57] Zhang, C. et al. (2018). Generative Adversarial Network for Synthetic Time Series Data Generation in Smart Grids. 2018 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm), 2018.