



Evaluation of Univariate Time-Series Models for Short-Term Solar Energy Forecasting



Luis F. Martínez-Soto , Omar Rodríguez-Zalapa ,
José Alberto López-Fernández, José Joaquín Castellanos-Galindo,
and José Horacio Tovar-Hernández

1 Introduction

1.1 Background

The energy transition that most countries are currently experiencing toward renewable energies is part of the strategies to mitigate the effects of climate change and comply with the Paris Agreement and the Sustainable Development Goals [1], intending to limit the global temperature rise in this century below 2 °C. Among the technologies that favor this transition are those that operate on the side of the electric power user, such as distributed generation, microgrids, and demand management. These allow increasing the efficiency of electrical systems by avoiding energy losses in the transmission and distribution stages. Besides, the correct operation of these new technologies requires energy control strategies that allow the electrical systems to continue to operate reliably and safely in the face of this new paradigm, in which energy management systems are increasing their relevance [2].

Energy management systems make it possible to cope with the intermittency and natural variability of renewable energies related to local climatic conditions, the constant increase in demand, and the complexity of the electricity consumption

L. F. Martínez-Soto (✉)

Centro de Ingeniería y Desarrollo Industrial (CIDESI), Querétaro, México

TecNM-Campus Querétaro, Querétaro, México

e-mail: lmartinez@cidesi.edu.mx

O. Rodríguez-Zalapa · J. A. López-Fernández · J. J. Castellanos-Galindo

TecNM-Campus Querétaro, Querétaro, México

J. H. Tovar-Hernández

TecNM-Campus Morelia, Morelia, México

habits of users. However, to achieve proper energy management, it is essential to adequately know the current and future energy production and consumption behavior and even estimate the batteries' state of charge in storage systems. This is possible by applying forecasting techniques, which can be done in various ways, whether from environmental forecast data, sky images, or data obtained with on-site measurements. The technique used depends on the objective of the forecast to be made, which in turn determines the required forecast horizon, that is, the time in advance in which the predictions are made. In the case of user-side applications involving distributed renewable generation sources, such as photovoltaic (PV) systems, the appropriate forecast horizon is usually short-term or one day in advance [3].

On the other hand, it is also necessary to use well-established and robust baseline forecasting techniques to evaluate the performance of new forecasting methodologies and algorithms. However, there are no generalized conclusions in the literature on which models are suitable in different circumstances [4], so it is necessary to previously analyze and implement reference models of forecasting to have a common reference in the evaluation of new methodologies or algorithms.

1.2 Forecast of Solar Energy in Microgrids

One of the biggest problems in forecasting PV power is its variability. For example, partly cloudy conditions can reduce solar radiation by up to 80% in one second, which, depending on the degree of penetration, represents a great challenge for power system operators. It is common to use classical or statistical forecasting techniques based on time-series to predict solar energy in microgrid applications. Still, the application of supervised machine learning algorithms is becoming increasingly relevant. There are deep learning models based on recurrent neural networks (RNN), such as gated recurrent units (GRU) [5, 6], long short-term memory (LSTM) [7, 8], and combined models or ensemble [9, 10]. Some combine multiple algorithms to achieve better prediction results than those obtained with individual algorithms [11].

The machine learning and statistical forecasting models can work adequately in intra-hour temporal resolutions (less than one hour) and intra-day (up to 72 h) and with spatial resolutions mainly at the microscale level, that is, up to 1 km [12]. This spatial resolution mainly favors the application of these forecasting techniques in microgrids and photovoltaic solar plants in areas of up to 100 ha.

There are also forecasting applications in the state of the art that use models composed of sub-processes for different time scales, known as multi-horizon [13, 14] (e.g., 5 min to 1 h, 1 to 6 h, 6 to 48 h, and greater). Each subprocess can use different datasets as inputs and even different granularity in the data for each time scale. This implies working with different data storage intervals for each prediction subprocess. Consequently, the initial and most crucial step in data analytical process applications is preparing or preprocessing the data [15]. It is also possible to forecast photovoltaic generation using fuzzy prediction interval (FPI)

models [16, 17], as was done in projects developed in the experimental microgrid of Huatocondo, Chile (24 kW), and the Goldwind microgrid in China (two PV systems: 200 and 250 kW). In the first case, better results were obtained with FPI than those obtained with linear regression algorithms for horizons of 15 min and 24 h in advance. In the second case, in combination with a particle swarm optimization (PSO) algorithm, superior performance was obtained, compared to the application of other optimization algorithms, for horizons of 10 min and 24 h in advance.

Besides, combinations of artificial intelligence techniques with PSO-type optimization algorithms have also been used. For example, in [18], the PSO is combined with an adaptive neuro-fuzzy inference system (ANFIS) to forecast the photovoltaic power of 3 photovoltaic generation units of 100 kW each. The algorithm results were evaluated in terms of RMSE, mean absolute error (MAE), and nMAE (normalized MAE), obtaining superior results for a forecast horizon of one hour in advance. Results are based on algorithms based on backpropagation artificial neural networks (ANN) and the persistence method. As part of the design of a centralized controller for energy management of a microgrid, in [19], two forecasting algorithms of the photovoltaic generation are applied: autoregressive moving average (ARMA) using the final prediction error (FPE) and a model based on multilayer perceptron nonlinear autoregressive (NAR). The neural network could predict sudden changes better, while the ARMA algorithm could only follow trends.

The potential for improving the prediction models of photovoltaic power when applying rolling forecasting is shown in [20]. This technique makes it possible to enhance the values predicted by the forecasting model, simultaneously extending and correcting the time-series model using real-time measurements. The results of this strategy were combined with a hybrid forecasting model based on support vector regression (SVR) sub-elements or sub-SVRs and an ARIMA algorithm to forecast instantaneous PV power. The precision of the model was evaluated with the RMSE and mean absolute percentage error (MAPE), and the results were better for sunny days than those obtained in the presence of clouds.

Combining two ANFIS algorithms has proven to be effective in applying Sugeno-type inference systems [21] to perform forecasting in microgrids in a 10 kW PV system. The analysis was conducted via simulation, using monitoring datasets from a PV system in Targoviste, Romania. The results were evaluated using MAE and the RMSE, observing that the error increases as the forecast horizon increases. Improvements in predictions were obtained by increasing the number of membership functions for each variable (increment from 2 to 3).

On the other hand, a method of forecasting PV power applicable to community microgrids has been proposed, based on deep learning [22], using deep recurrent neural networks (DRNN) with an LSTM forecasting algorithm with a short-term horizon. The algorithm was tested on a residential microgrid in Austin, Texas, consisting of a 100 kW PV system. The results were evaluated in terms of RMSE, MAE, MAPE, and Pearson's correlation coefficient and were compared with support vector machine (SVM)-based prediction algorithms and a multilayer perceptron ANN. The forecast results showed the best performance for the DRNN algorithm,

followed by the multilayer perceptron (MLP) and finally the SVM. The DRNN model practically outperformed the SVM algorithm by 2 to 1.

Statistical linear models are the time-series prediction methods most traditionally used as an alternative to numerical weather prediction (NWP) techniques. They are based on a physical approach where atmospheric models describe the variability of meteorological processes. Descriptions are given at the mesoscale level or global datasets of meteorological measurements [23]. On the other hand, time-series statistical models are based on historical data, as a sequence of observations, measured at certain moments in time. They are ordered chronologically and evenly spaced, which makes the data dependent on each other in most cases to develop a descriptive relationship between various magnitudes to obtain an estimate of the future of a certain magnitude [24, 25]. In such a model, the dependent or output variable depends only on its past values [26].

There are also forecasting applications for solar energy and other energy fields using time-series, such as SARIMA models, to forecast hydroelectric production in Ecuador [27, 28]. It was shown that the proposed models are suitable for forecasting time-series with seasonal components, which can also be improved with the use of exogenous variables, such as precipitation, to predict monthly hydroelectric production in forecasting horizons up to a year in advance. On the other hand, triple exponential smoothing or Holt-Winters models have also been used to forecast solar irradiance in PV-distributed generation facilities for intra-hour prediction horizons in a PV system located on the roof of the University of Utrecht, Netherlands [29]. The implemented model was suitable for forecasting the time-series, including its seasonal component, with a high degree of accuracy, in terms of the MAE, compared to other forecasting methods, such as the persistence and the average methods.

To evaluate forecasting models based on univariate time-series, which can later be applied as forecasting benchmarks to assess the performance of other more advanced multivariate forecasting techniques, a comparison between the Box-Jenkins and exponential smoothing approaches has been conducted. A general methodology proposal for forecasting in a univariate time-series model, applied and validated in this study, is also revealed.

2 Methods

2.1 *Time-Series Decomposition*

A decomposition of the PV power time-series was applied, which consists of a statistical analysis that allows separating the time-series into different components or sub-series of time, each representing one of the characteristics of the original time-series: trend, seasonality, and residuals [30]. This makes it possible to analyze each of the components separately, allowing to complement the determination of stationarity of the time-series, that is, if its mean and variance are constant over

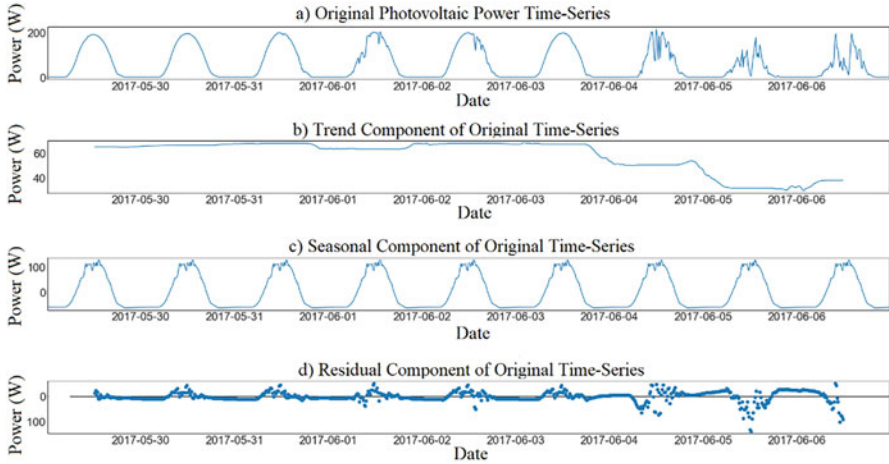


Fig. 1 Time-series decomposition of photovoltaic power

time. In addition, it will enable us to analyze how dominant the seasonal component of the time-series is, determining whether there is periodicity in the seasonal variation of the time-series, which can be annual, monthly, daily, or on another time scale [31]. Therefore, determine if the forecasting model should contain components to predict those dominant characteristics.

In general, there are two methods to perform time-series decomposition, the multiplicative and the additive [32]. In this case, the additive method has been used using the Python *statsmodels* package. The result is shown in Fig. 1, where it is observed that the time-series has a small trend component, with a value of up to 60 W, when there are variations in photovoltaic power, unrelated to the daily cyclical behavior of irradiance (e.g., shadows). In the case of seasonality, it is the dominant component of the time-series, reflecting the daily cycles of photovoltaic power. Finally, the residual component is shown as the weakest, with maximum values coinciding with the reductions in photovoltaic power caused by shadows.

2.2 Exponential Smoothing Forecasting Model

The exponential smoothing forecasting techniques, or error, trend, and seasonality (ETS), are state-space models widely used to forecast time-series in other areas of knowledge [33, 34], despite being proposed between 1957 and 1960 [35]. In issues related to electrical energy forecasting, they are applied, either individually or in combination with other techniques, in applications such as wind power forecasting [36] and photovoltaic solar power [37] in forecast horizons from short-term to long-term. This forecasting method uses weighted combinations of past observations from the time-series, giving the most recent observations more weight than the oldest

ones. The weight given to old observations exponentially decreases as they get older, so this method is called exponential smoothing. These forecasting techniques are divided into three: simple exponential smoothing, used to predict time-series that do not have seasonality or trend; Holt's linear smoothing, applicable to trend time-series; and finally, Holt-Winters smoothing method, useful for forecasting time-series that include seasonality and trend. Eqs. (1, 2, 3 and 4) describe a Holt-Winters forecasting model [38]:

$$S_t = \alpha \left(\frac{X_t}{I_{t-s}} \right) + (1 - \alpha)(S_{t-1} + T_{t-1}) \quad (1)$$

$$T_t = \gamma(S_t - S_{t-1}) + (1 - \gamma)T_{t-1} \quad (2)$$

$$I_t = \delta \left(\frac{X_t}{S_t} \right) + (1 - \delta)I_{t-s} \quad (3)$$

$$\widehat{X}_{t+h} = (S_t + hT_t)I_{t-s+h} \quad (4)$$

where S_t and T_t are the equations of the level of the time-series (average value) and the additive trend, with smoothing parameters α and γ . I_t is the index of the seasonal length of the time-series, with smoothing parameters δ . X_t represents the observed data, while \widehat{X}_{t+h} is the forecasting equation for h instants of time in advance.

2.3 Box-Jenkins Methodology

On the other hand, the Box-Jenkins methodology of ARIMA models has dominated for almost 50 years in the field of time-series forecasting [39]. It is widely used to forecast photovoltaic production and solar irradiance, either as a reference model for validating new forecasting techniques [40] or as a functional forecasting technique [41]. This approach indicates that a non-seasonal time-series can be modeled as a combination of the values and past errors, which is denoted as ARIMA (p, q, d), which can be expressed according to Eq. (5):

$$x_t = \sum_{i=1}^p \varphi_i x_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i} + k + \varepsilon_t \quad (5)$$

where x_t is the predicted parameter at each instant t , φ represents the autoregression (AR) coefficients, θ represents the moving average (MA) coefficients, k is a constant, and ε_t models the white noise. For their part, p and q are the orders of the AR and MA models, respectively, while d represents the order of differentiation of the model. Similarly, a time-series that contains seasonality can be expressed with a seasonal ARIMA model. It is also known as SARIMA or ARIMA (p, d, q) (P, D, Q), where $P, D,$ and Q represent the seasonal order of autoregression, differentiation, and moving average, respectively.

2.4 Forecasting Errors

To evaluate the performance of forecasting techniques, several indices allow measuring the performance of prediction algorithms by comparing the predicted values against actual or observed values. The most common indices are the MAE, the MAPE, the mean square error (MSE), and RMSE. Each of these indices has its particular characteristics, so selecting the most suitable index depends on the approach of the case of application [42]. However, it has been observed in the state of the art that, for the operation of microgrids and short-term forecasting, the most widely used indices are the RMSE, the MAE, and the MSE, which is why they are the ones used in this work. These indices are calculated according to Eqs. (6, 7 and 8):

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2} \quad (6)$$

$$MSE = \frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2 \quad (7)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i| \quad (8)$$

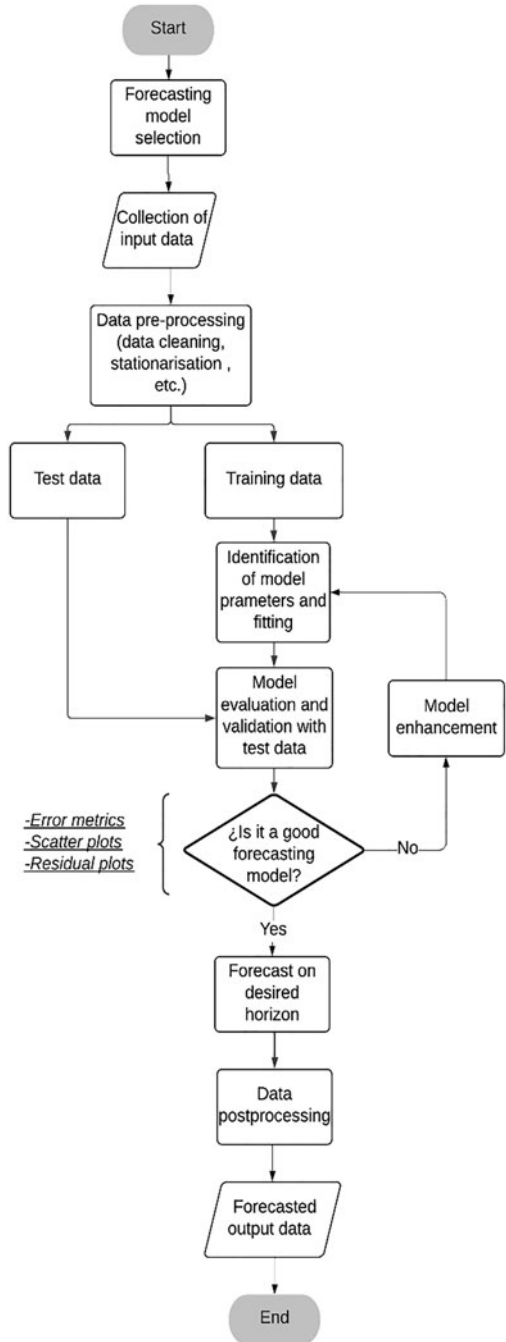
where y_i is the real value or observed at the time i , \hat{y}_i is the predicted value at the time i , and N represents the number of data samples used to make the forecast and therefore represents the forecast horizon. Additionally, it is possible to evaluate the forecast results using scatter plots and histograms of the frequency of errors, among other indicators [43, 44]. The prediction model results are evaluated by comparing the predicted data with the test or actual data in all cases.

2.5 Proposed Forecasting Approach

By comparing the process of identification of parameters, training, and testing of the prediction algorithms used for one-variable or univariate time-series, a methodology has been identified, proposed, and implemented for its general application in these cases, which has been reflected in the flow diagram of Fig. 2. As an initial stage, three substages are included: (1) collection of input data, (2) selection of forecasting model, and (3) data preprocessing. The forecasting model selection stage allows the methodology to be generalized and applicable to any univariate time-series forecasting technique applicable to PV power or solar irradiance.

On the other hand, data collection can be done in various ways, either with data obtained from measurements, or using state-of-the-art datasets, as was done in this work. The preprocessing data stage is of great relevance, and its characteristics will depend on the forecasting technique to be used. For example, if ARMA models are used, the stationarity of the data should be evaluated to determine if a prediction

Fig. 2 Proposed forecasting methodology for univariate time-series models



model is adequate. Data resolution change and cleaning can also be performed at this stage. The methodology considers error indicators and dispersion and residual graphs to compare the predicted values against the real values of the PV power or solar irradiance.

This methodology has been implemented using Python programming language and its Auto ARIMA function and the statsmodels package for the ARIMA and exponential smoothing models, respectively, to determine the coefficients and adjust the prediction models. The forecast horizon used is short-term, 24 h in advance, applicable to demand-side installations, such as microgrids.

2.6 *Experimental Design and Characteristics of Datasets*

Like other research on forecasting applications of other energy fields [45], an empirical comparative evaluation of forecasting techniques has been carried out. This was done with two methods based on time-series, ARIMA and exponential smoothing. The research was conducted with historical data to predict solar energy on a short-term forecasting horizon of 24 h ahead and to evaluate both forecasting methods as possible benchmarks for future stages of this research.

The validation of both forecasting methods was done with two different datasets. The first predicts solar irradiance using a dataset collected over one and a half years and 1 s time resolution. It was obtained from a 29.25 kW PV system installed in Queretaro, Mexico, where the project is developed. This PV system consists of 90 polycrystalline photovoltaic modules of 325 W each one, an inclination angle of 20°, and an azimuth orientation of 18°. Currently, the measurement system does not include data on PV power. Therefore, only the irradiance forecast is considered in this case. Another purpose with this dataset was to evaluate the forecast accuracy changing the data resolution from 1 min to 15 min.

Irradiance and PV power data are relevant to test the forecasting algorithms as indicative variables of the available solar energy due to the high Pearson correlation. It varies between 0.89 and 0.99, depending on the season of the year and the presence or absence of clouds [46–48]. Therefore, the study is performed with a dataset available on IEEE DataPort [49] to forecast PV power tests. This dataset includes six magnitudes corresponding to 2017, with a time resolution of 1 min, obtained from a PV installation located at the Polytechnic of Milan. The dataset contains information on photovoltaic power (W), ambient temperature (°C), global horizontal irradiance (GHI, W/m²), the plane of photovoltaic array irradiance (POA, W/m²), wind speed (m/s), and wind direction (°).

The PV installation of Polytechnic of Milan consists of a PV module with a peak power of 285 W, made of monocrystalline silicon, and installed with an inclination angle of 30° and an azimuth orientation 6° 30'. This dataset has been already referenced in other PV power forecasting research work based on ANN [50], which will allow in the future to compare the obtained forecast results with those already reported. A relevant characteristic of this dataset is the presence of snowfall

days, a meteorological condition never available locally in México, where the project is currently being carried out. The main reason and motivation to use this dataset is the possibility of assessing the proposed forecast methods with that kind of meteorological condition. It ensures a good generality of application for different places and weather conditions.

Two general conditions were considered with the second dataset: sunny days in winter and cloudy days in spring. As a subcase for sunny days, it was considered a variation in the number of training data days, with variations in PV power caused by clouds, to assess its impact on the accuracy of forecasting models.

3 Results

3.1 Solar Irradiance Forecast

Initially, a Holt-Winters model with additive error component (A), additive trend component (A), and additive seasonal component (A) or (A, A, A), to forecast a solar irradiance time-series of 7 days, with sampling per minute is performed. The model adjusted to the time-series and made an adequate forecast, as shown in Fig. 3. The curves reflect minor variations, which can be observed when using a sampling per minute. Despite the variations, the Holt-Winters model (blue curve) manages to

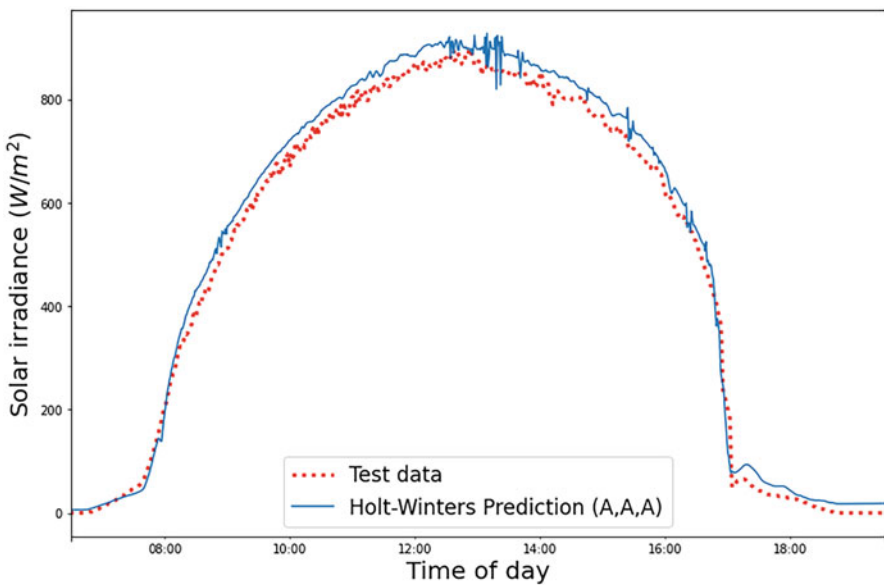


Fig. 3 Irradiance data and predicted data, using a Holt-Winters model, with data sampling per minute

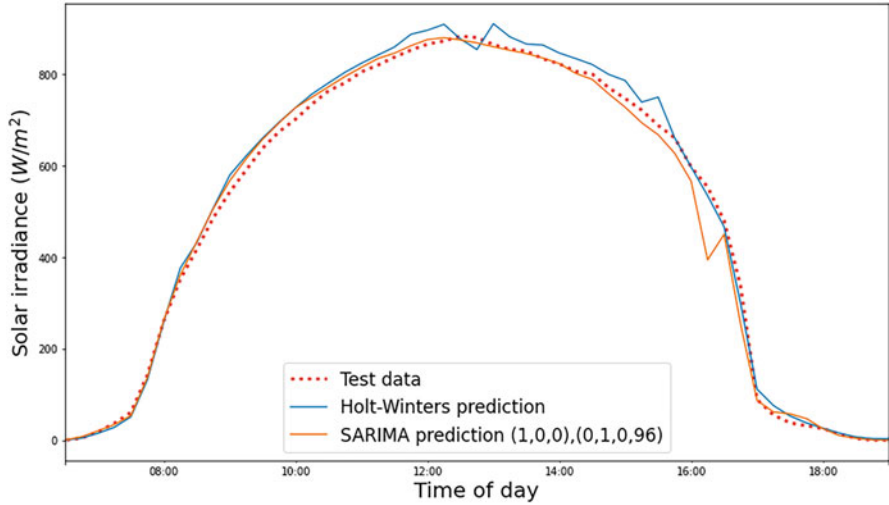


Fig. 4 Irradiance data and predicted data, using a Holt-Winters model, with a sampling interval of 15 min

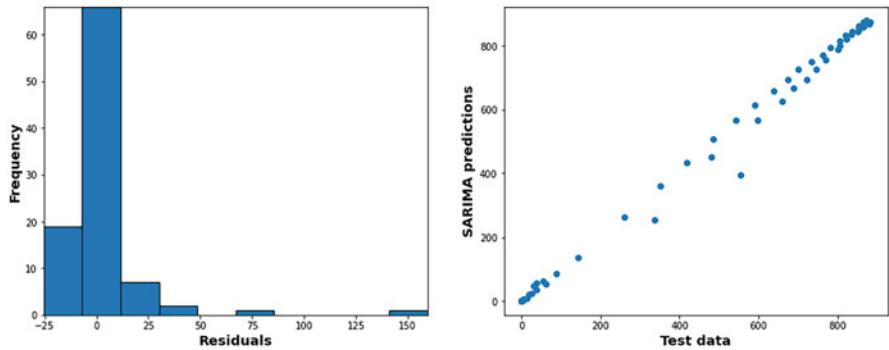


Fig. 5 Residual and dispersion plots of SARIMA model forecasting results for irradiance, with a sampling interval of 15 min

follow the general trajectory of the measured irradiance curve (red color). The RMSE obtained was $23.95 W / m^2$ (Fig. 4).

Subsequently, the same model was applied to the same time-series with an adjusted sampling to 15 min to analyze the impact of data preprocessing on forecasting results. In this case, the RMSE was lower, with a value of $17.15 W / m^2$. A second model was identified and trained as a SARIMA $(1, 0, 0) \times (0, 1, 0)$ [96], showing a similar graphical performance but with a bigger RMSE of $21.13 W / m^2$. These results are confirmed with residuals and scatter plots of Figs. 5 and 6, showing most of the residuals or differences between the actual and predicted values are close to zero with the Holt-Winters model. In contrast, its scatter plot resembles a linear regression.

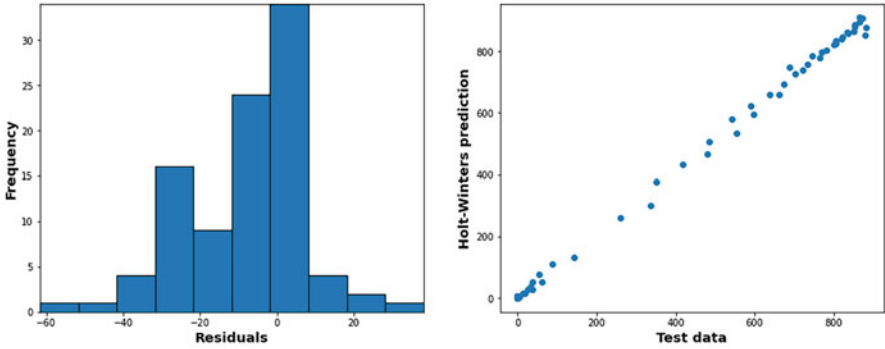


Fig. 6 Residual and dispersion plots of Holt-Winters model forecasting results for irradiance, with

Table 1 Summary of irradiance forecast results

Conditions	Dataset resolution	Number of training data days/ test data days	Model	RMSE (W/m ²)
Sunny	1 min	6/1	Holt-winters (A, A, A)	1.89
Sunny	15 min	6/1	Holt-winters (A, A, A)	37.19
Sunny	15 min	6/1	SARIMA (1,0,0) × (0,1,0) [96]	21.13

a sampling interval of 15 min

A summary of obtained irradiance forecasts results is shown in Table 1, indicating the general weather conditions in each case, the data resolution, the number of training and test data days used, the characteristics of the fitted model, and the RMSE obtained.

4 Photovoltaic Power Forecast

4.1 Forecast for the Winter Season: First Case

In all cases of PV power forecasting, the dataset has been preprocessed to a resolution of 15 min.

The first analysis case of PV power forecasting considers a time-series of 9 days, from January 13 to 21, considering eight days as training data and the ninth as test data. Two forecasting models were identified and trained, the first SARIMA (5, 0, 4) × (0, 1, 0) [96] and the second from Holt-Winters. The SARIMA model yielded an RMSE of 1.89 W, while the Holt-Winters model had an RMSE of 37.19 W. The

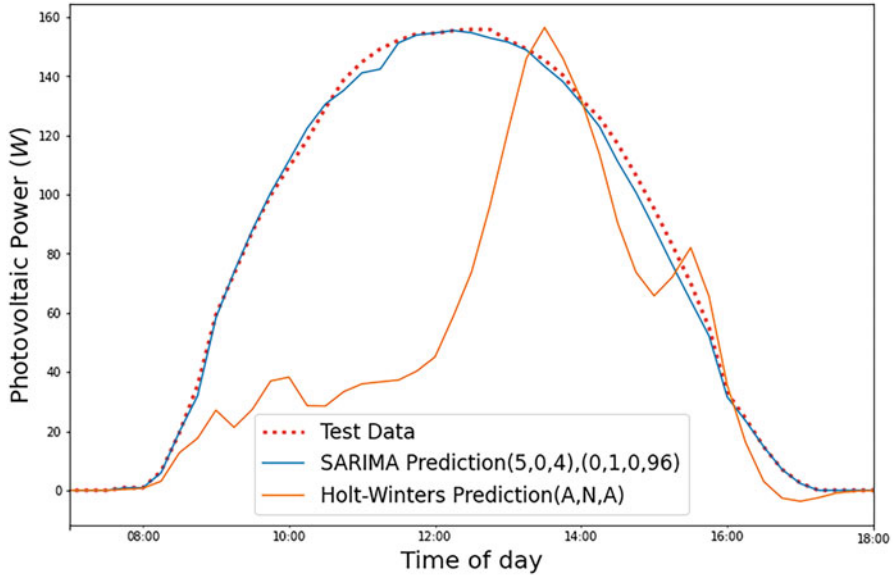


Fig. 7 PV power test data and predicted data, using both models, case 1

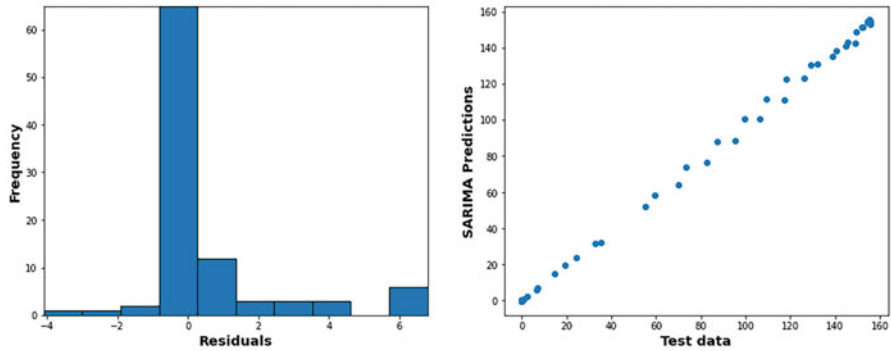


Fig. 8 Residual and dispersion plots of SARIMA model forecasting results, case 1

comparison of the forecast results is shown in Fig. 7 with a significant similarity between the test data and the SARIMA prediction. It is confirmed in the results of the residuals and scatter plots (Fig. 8) since most of the residuals or differences between the actual and predicted values are close to zero. In contrast, the scatter plot strongly resembles a linear regression. On the other hand, the Holt-Winters prediction model shown in Fig. 7 with orange color could not accurately predict the PV power. It was confirmed in the results of residuals and dispersion (Fig. 9), with residuals far from zero and a marked dispersion between the test or actual data and the predictions.

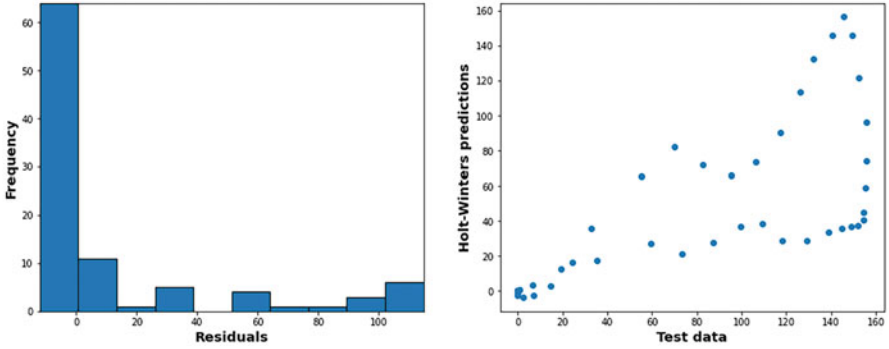


Fig. 9 Residual and scatter plots of Holt-Winters model forecasting results, case 1

4.2 Forecast for the Winter Season: Second Case

The second case also considers the PV power forecast of January 22. In this case, a 7-day time-series is used, eliminating the two first days (January 13 and 14) with the presence of clouds, to evaluate the potential improvement of the Holt-Winters model. In this way, the training data is reduced from 8 to 6 days. The test data is the same (January 21). The models identified and fitted to the training data are a SARIMA $(4, 0, 0) \times (0, 1, 0)$ [96] model and the second Holt-Winters model. The SARIMA model yielded an RMSE of 1.93 W, very similar to the previous case, while the Holt-Winters model obtained an RMSE of 10.31 W, 20 W lower than the previous case. The comparison of the test data with the results of both forecasting models is shown in Fig. 10. The similarity between the SARIMA curve model and test data is confirmed, unlike with the Holt-Winters model. However, the behavior of the latter is improved in comparison to the previous case.

4.3 Forecast for the Spring Season

The third case evaluated is a time-series of nine days of photovoltaic power, from May 29 to June 6, 2017, using eight days as training data and the ninth day as test data. In this case, the primary purpose was to evaluate the behavior of the models to forecast cloudy days. Two forecasting models were identified and trained, the first SARIMA $(5, 0, 1) \times (0, 1, 0)$ [96] and the second from Holt-Winters. The SARIMA model yielded an RMSE of 37.74 W, while the Holt-Winters model had an RMSE of 54.45 W. In both cases, the errors are considerably high, considering that the maximum value of the photovoltaic power during that period is around 200 W. A

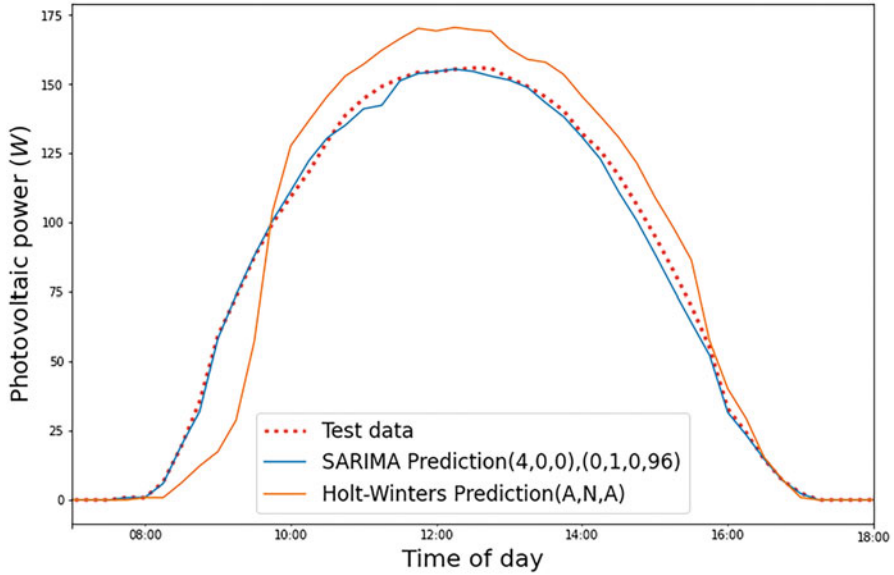


Fig. 10 PV power test data and predicted data using both models, case 2

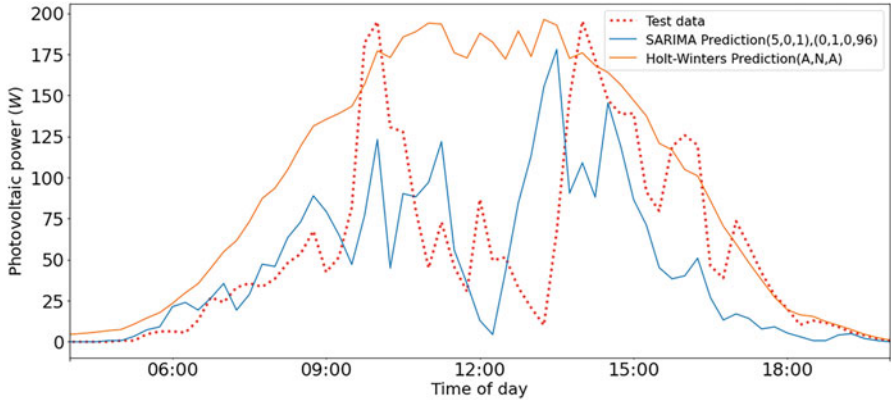


Fig. 11 PV power test data and predicted data using both models, case 3

graphical comparison of the real PV power curve concerning the prediction result of both models is shown in Fig. 11. The SARIMA model presents typical day variations with a high incidence of clouds, although it does not follow the general trajectory of the curve of real values. On the other hand, the Holt-Winters model shows only some slight variations close to the peak of the photovoltaic production curve.

Table 2 Summary of PV power forecast results

Conditions	Number of training data days/test data days	Model	RMSE (W/m ²)
Winter, sunny	8/1	SARIMA (5,0,4) × (0,1,0) [96]	1.89
Winter, sunny	8/1	Holt-Winters (A, N, A)	37.19
Winter, sunny	6/1	SARIMA (4,0,0) × (0,1,0) [96]	1.93
Winter, sunny	6/1	Holt-Winters (A, N, A)	10.31
Spring, cloudy	8/1	SARIMA (5,0,1) × (0,1,0) [96]	37.74
Spring, cloudy	8/1	Holt-Winters (A, N, A)	54.45

Finally, a summary of obtained PV power forecasts is shown in Table 2. General weather conditions are presented in each case, such as sunny or cloudy days and the year's season. The number of training and test data days used, the characteristics of the fitted model, and the RMSE obtained are also shown. In all cases, the dataset had a resolution of 15 min.

5 Discussion

Both scenarios of irradiance forecasting during sunny days show a similar predicted irradiance profile compared to the measured values, as can be seen in Figs. 3 and 4. Nevertheless, when reducing the data resolution from 1 min to 15 min, the irradiance profile shows fewer variations in the general trend due to data resampling. When forecasting with 1 min data resolution, the Holt-Winters method lets us predict the fast variation of the general trend of irradiance. This could be useful to forecast irradiance in PV systems with high accuracy, Class A, monitoring systems, suggested by the standard IEC 61724-1:2017 [51], for applications of fault location, electric network interaction assessment, PV technology assessment, and precise PV system degradation measurement.

In the first case analyzed of PV power forecasting, the Holt-Winters prediction model (Fig. 7), in orange color, presents a significant distortion, apparently caused by the effect of cloudy days existing in the training data. This was confirmed when two cloudy days were removed in training data, as seen in Fig. 10 from the second case of PV power forecasting. Therefore, for nonhomogeneous datasets, where the training data include variations due to cloud cover, the SARIMA model is more suitable for predicting photovoltaic power.

The second case of irradiance forecasting shown in Fig. 4 and the first case of PV power forecasting during the winter season depicted in Fig. 10, showed that, when there are fewer fast variations in the magnitude of solar energy caused by clouds, the SARIMA model can eliminate its moving average (MA) component. Again, this can be considered a result of the presence of prevailing homogeneous datasets.

As a relevant finding in the implementation and analysis of these two univariate forecasting techniques of solar energy, it was found that, unlike cases reported in the state of the art [19], the ARIMA forecasting models are not enough to forecast solar energy accurately if they do not include the seasonal component, as the SARIMA model does. For the same reason, the Holt-Winters model predicts solar energy with acceptable results by including a third equation to handle the time-series seasonality. The application of the proposed forecasting methodology has allowed validating the applicability of the exponential smoothing and Box-Jenkins forecasting methods, allowing us to identify and verify which weather conditions are adequate to predict solar irradiance and photovoltaic power. In addition, the use of this general methodology to evaluate forecasting techniques in the field of solar energy for microgrids was validated.

6 Conclusions

Based on the obtained results, the SARIMA model yields better results in all the cases analyzed in terms of the RMSE index and the dispersion of the results. Specifically, it could adequately represent the characteristics of trend, seasonality, and average values of the time-series of photovoltaic power or solar irradiance. However, in cases with minimum or no cloud incidence, expressly with predominantly homogeneous data, the Holt-Winters exponential smoothing model is also an alternative to fit the time-series training data to represent the dominant component of the seasonality of this type of time-series. As a general conclusion, it was confirmed that they are not the most suitable for representing variations in highly nonlinear time-series, as is the case of rapid variations in the magnitude of the photovoltaic power caused by the incidence of clouds. Moreover, it was possible to identify, propose, and implement a forecasting methodology applicable to forecast univariate time-series of solar energy in microgrids, which can be generalized and applied to state-of-the-art forecasting algorithms. Therefore, as future work, forecasting models of solar energy in microgrids, capable of predicting nonlinearities and sudden changes in the time-series, will be implemented, with a multivariable approach, such as machine learning and deep learning techniques.

References

1. Dzebo, A., Janetschek, H., Brandi, C., & Iacobuta, G. (2019) *Connections between the Paris agreement and the 2030 agenda. The case for policy coherence*, p. 38.
2. Van Der Meer, D., Mouli, G. R. C., Mouli, G. M. E., Elizondo, L. R., & Bauer, P. (2018). *Energy management system with PV power forecast to optimally charge EVs at the workplace*. IEEE Transactions on Industrial Informatics.
3. A. Ahmad, N. Javaid, A. Mateen, M. Awais, and Z. A. Khan, "Short-term load forecasting in smart grids: An intelligent modular approach," *Energies*, vol. 12, no. 1, p. 164, Jan. 2019.
4. Salamanis, A. I., et al. (2020). Benchmark comparison of analytical, data-based and hybrid models for multi-step short-term photovoltaic power generation forecasting. *Energies*, 13(22). <https://doi.org/10.3390/en13225978>
5. Wang, Y., Liao, W., & Chang, Y. (2018). Gated recurrent unit network-based short-term photovoltaic forecasting. *Energies*, 11, 2163.
6. Gao, B., Huang, X., Shi, J., Tai, Y., & Xiao, R. (2019). Predicting day-ahead solar irradiance through gated recurrent unit using weather forecasting data. *Journal of Renewable and Sustainable Energy*, 11.
7. Harrou, F., Kadri, F., & Sun, Y. (2020). Forecasting of photovoltaic solar power production using LSTM approach. In *Advanced statistical modeling, forecasting, and fault detection in renewable energy systems*. IntechOpen.
8. Ge, Y., Nan, Y., & Bai, L. (2019). A hybrid prediction model for solar radiation based on long short-term memory, empirical mode decomposition, and solar profiles for energy harvesting wireless sensor networks. *Energies*, 12, 4762.
9. Liu, D., & Sun, K. (2019). Random forest solar power forecast based on classification optimization. *Energy*, 187, 115940.
10. M. W. Ahmad, M. Mourshed, and Y. Rezgui, "Tree-based ensemble methods for predicting PV power generation and their comparison with support vector regression," *Energy*, vol. 164, pp. 465–474, Dec. 2018.
11. Martínez, C. G. (2018). *Árboles de decisión y métodos de ensemble*. RPubS.
12. Ahmad, T., Zhang, H., & Yan, B. (2020). *A review on renewable energy and electricity requirement forecasting models for smart grid and buildings* (Sustainable cities and society) (Vol. 55, p. 102052). Elsevier Ltd.
13. Mishra, S., & Palanisamy, P. (2018). Multi-time-horizon solar forecasting using recurrent neural network. In *2018 IEEE energy conversion congress and exposition*. ECCE.
14. Fan, C. et al. (2019). Multi-horizon time-series forecasting with temporal attention learning. In *Proceedings of the ACM SIGKDD international conference on knowledge discovery and data mining*.
15. Huang, Q. (2018). *Application of machine learning in power systems: Part I – An overview*. IEEE Smart Grid.
16. Sáez, D., Ávila, F., Olivares, D., Cañizares, C., & Marín, L. (2015). Fuzzy prediction interval models for forecasting renewable resources and loads in microgrids. *IEEE Transactions on Smart Grid*, 6(2), 548–556.
17. Rafique, F., Jianhua, Z., Rafique, R., Guo, J., & Jamil, I. (2018). Renewable generation (wind/solar) and load modeling through modified fuzzy prediction interval. *International Journal of Photoenergy*, 2018, 1–14.
18. Semero, Y. K., Zheng, D., & Zhang, J. (2018). A PSO-ANFIS based hybrid approach for short term PV power prediction in microgrids. *Electric Power Components and Systems*, 46.
19. Hernández-Hernández, C., Rodríguez, F., Moreno, J. C., Da Costa Mendes, P. R., Normey-Rico, J. E., & Guzmán, J. L. (2017). The comparison study of short-term prediction methods to enhance the model predictive controller applied to microgrid energy management. *Energies*, 10(7), 35.
20. Mei, F., Pan, Y., Zhu, K., & Zheng, J. (2018). A hybrid online forecasting model for ultrashort-term photovoltaic power generation. *Sustainability*, 10(3), 1–17.

21. Dragomir, O. E., Dragomir, F., Stefan, V., & Minca, E. (2015). Adaptive neuro-fuzzy inference systems as a strategy for predicting and controlling the energy produced from renewable sources. *Energies*, 8(11), 13047–13061.
22. Wen, L., Zhou, K., Yang, S., & Lu, X. (2019). Optimal load dispatch of community microgrid with deep learning based solar power and load forecasting. *Energy*, 171, 1053–1065.
23. Ma, J., & Ma, X. (2018). A review of forecasting algorithms and energy management strategies for microgrids. *Systems Science & Control Engineering*, 6(1), 237–248.
24. Villavicencio, P. J. (2010). Introducción a Series de Tiempo. *Man. Metodol. Ser. tiempo*, 33.
25. Ssekulima, E. B., Anwar, M. B., Al Hinai, A., & El Moursi, M. S. (2016). Wind speed and solar irradiance forecasting techniques for enhanced renewable energy integration with the grid: A review. *IET Renewable Power Generation*, 10(7), 885–898.
26. Atique, S., Noreen, S., Roy, V., Subburaj, V., Bayne, S., & MacFie, J. (2019). Forecasting of total daily solar energy generation using ARIMA: A case study. In *2019 IEEE 9th Annual Computing and Communication Workshop and Conference* (pp. 114–119). CCWC.
27. Mite-León, M., & Barzola-Monteses, J. (2018). Statistical model for the forecast of hydropower production in Ecuador. *International Journal of Renewable Energy Research*, 10(2), 1130–1137. 1309-0127.
28. Barzola-Monteses, J., Mite-León, M., Espinoza-Andaluz, M., Gómez-Romero, J., & Fajardo, W. (2019). Time series analysis for predicting hydroelectric power production : The Ecuador case. *Sustainability*, 11(6539), 1–19. <https://doi.org/10.3390/su11236539>
29. Dev, S., Alskaf, T., Hossari, M., Godina, R., Louwen, A., & Van Sark, W. (2018). Solar irradiance forecasting using triple exponential smoothing. In *International conference on smart energy systems and technology*. SEST. <https://doi.org/10.1109/SEST.2018.8495816>
30. Yang, D., Sharma, V., Ye, Z., Lim, L. I., Zhao, L., & Aryaputera, A. W. (2015). Forecasting of global horizontal irradiance by exponential smoothing, using decompositions. *Energy*, 81.
31. Prema, V., & Rao, K. U. (2015). Time-series decomposition model for accurate wind speed forecast. *Renewables: Wind, Water, and Solar*, 2.
32. Tran Anh, D., Duc Dang, T., & Pham Van, S. (2019). Improved rainfall prediction using combined pre-processing methods and feed-forward neural networks. *J*, 2.
33. Farias, R. L., Puig, V., Rangel, H. R., & Flores, J. J. (2018). Multi-model prediction for demand forecast in water distribution networks. *Energies*, 11.
34. Liu, L., & Wu, L. (2020). Predicting housing prices in China based on modified Holt's exponential smoothing incorporating whale optimization algorithm. *Socio-Economic Planning Sciences*, 72.
35. Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice*.
36. Ferreira, M., Santos, A., & Lucio, P. (2019). Short-term forecast of wind speed through mathematical models. *Energy Reports*, 5.
37. Sánchez-Durán, R., Barbancho, J., & Luque, J. (2019). Solar energy production for a decarbonization Scenario in Spain. *Sustainability*, 11(24).
38. Trull, Ó., García-Díaz, J. C., & Troncoso, A. (2020). Stability of multiple seasonal holt-winters models applied to hourly electricity demand in Spain. *Applied Sciences*.
39. Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2013). *Time-series analysis: Forecasting and control* (4th ed.). Wiley.
40. Suresh, V., Janik, P., Rezmer, J., & Leonowicz, Z. (2020). Forecasting solar PV output using convolutional neural networks with a sliding window algorithm. *Energies*, 13, 723.
41. David, M., Ramahatana, F., Trombe, P. J., & Lauret, P. (2016). Probabilistic forecasting of the solar irradiance with recursive ARMA and GARCH models. *Solar Energy*.
42. Xu, Z., Hu, Z., Zhao, J., Song, Y., Lin, J., & Wan, C. (2016). Photovoltaic and solar power forecasting for smart grid energy management. *CSEE Journal Power Energy System*, 1(4), 38–46.
43. Sharafi, M., Ghaem, H., Tabatabaee, H. R., & Faramarzi, H. (2017). Forecasting the number of zoonotic cutaneous leishmaniasis cases in south of Fars province, Iran using seasonal ARIMA time-series method. *Asian Pacific Journal of Tropical Medicine*, 10(1), 79–86.

44. Sabir, E. C., & Batuk, E. (2013). Demand forecasting withof using time-series models in textile dyeing-finishing mills. *Tekst. ve Konfeksiyon*, 23(2), 143–151.
45. Divina, F., Torres, M. G., Vela, F. A. G., & Noguera, J. L. V. (2019). A comparative study of time series forecasting methods for short term electric energy consumption prediction in smart buildings. *Energies*, 12(10). <https://doi.org/10.3390/en12101934>
46. Yadav, H. K., Pal, Y., & Tripathi, M. M. (2020). Parameter optimization using PSO for neural network-based short-term PV power forecasting in Indian electricity market. *Lecture Notes in Electrical Engineering*, 597, 331–348.
47. Zhu, H., Li, X., Sun, Q., Nie, L., Yao, J., & Zhao, G. (2015). A power prediction method for photovoltaic power plant based on wavelet decomposition and artificial neural networks. *Energies*, 9(1), 11.
48. Li, L.-L., Cheng, P., Lin, H.-C., & Dong, H. (2017). Short-term output power forecasting of photovoltaic systems based on the deep belief net. *Advances in Mechanical Engineering*, 9(9), 168781401771598.
49. Leva, E., Sonia, P., & Silvia, O. (2020). *Photovoltaic power and weather parameters* | *IEEE DataPort*. IEEE DataPort.
50. Nespoli, A., et al. (2019). Day-ahead photovoltaic forecasting: A comparison of the most effective techniques. *Energies*, 12. <https://doi.org/10.3390/en12091621>
51. IEC. (2017). *IEC 61724-1:2017 photovoltaic system performance - part 1: Monitoring*. IEC.