UNIVERSIDAD DISTRITAL
FRANCISCO JOSÉ DE CALDAS

Tecnura

https://revistas.udistrital.edu.co/index.php/Tecnura/issue/view/1328

DOI: https://doi.org/10.14483/22487638.23971

INVESTIGACIÓN

Characterization and prediction of solar radiation in the municipality of Mompox-Bolívar using time series models

Caracterización y predicción de la radiación solar en el municipio de Mompox-Bolívar usando modelos de series de tiempo

Gabriel Elías Chanchí-Golondrino [1], Manuel Alejandro Ospina-Alarcón [2], Álvaro Realpe-Jiménez [3],

Fecha de Recepción: 30 de julio de 2025 **Fecha de Aceptación**: 31 de julio de 2026

Cómo citar: G. E. Chanchí-Golondrino, M. A. Ospina-Alarcón and A. Realpe-Jiménez, «Characterization and prediction of solar radiation in the municipality of Mompox-Bolívar using time series models», *Tecnura*, vol. 30, n.° 89, jul. 2026. 4–28. https://doi.org/10.14483/22487638.23971

Abstract

Objective: Considering the growing transition towards renewable energies and specifically the solar potential of Colombia and the Caribbean region with respect to photovoltaic solar energy, as well as the need to develop solar potential characterization studies, in this article, the adjustment and evaluation of four stationary and seasonal time series models (ARIMA, SARIMA, SARIMAX, and Exponential Smoothing) on solar radiation data from the municipality of Mompox-Bolívar are proposed as a contribution.
Methodology: Methodologically, within this research, the CRISP-DM methodology was adapted into 4 phases, namely: F1. Business and data understanding, F2. Data preparation, F3. Modeling, F4. Model evaluation and deployment.
Results: As a result, it was obtained that at the level of stationary models, the ARIMA (3,0,3) model was the one that achieved the best fit, with a moderate coefficient of determination for the training and test sets. Likewise, with respect to models with a seasonal component, it was determined that the SARIMA model based on ARIMA (3,0,3) and with a seasonal component (1,0,1,8) had the best performance, with a good fit on the training set (R2 of 0.8571) and test set (R2 of 0.8152).
Conclusions: It is concluded that the incorporation of the exogenous variable into the SARIMA model did not improve the SARIMA model's performance, such that the SARIMAX model obtained good performance on the training set and moderate performance on the test set. Likewise, the models obtained in this research are of great utility for the characterization of solar radiation potential, as well as for efficient planning in the deployment of photovoltaic solar energy systems.
Keywords: ARIMA, solar energy, solar radiation, SARIMA, time series.

1 Ph.D en Ingeniería Telemática. Facultad de Ingeniería de la Universidad de Cartagena. ROR Email: gchanchig@unicartagena.edu.co

2 Ph.D en Ingeniería – Ciencia y Tecnología de Materiales. Facultad de Ingeniería de la Universidad de Cartagena. ROR Email: mospinaa@unicartagena.edu.co

3 Ph.D en Ingeniería Química. Facultad de Ingeniería de la Universidad de Cartagena. ROR Email: arealpe@unicartagena.edu.co

Resumen

Objetivo: Teniendo en cuenta la transición creciente hacia las energías renovables y específicamente el potencial solar de Colombia y de la región caribe con respecto a la energía solar fotovoltaica, así como la necesidad de desarrollar estudios de caracterización del potencial solar, en este artículo se propone como contribución el ajuste y evaluación de cuatro modelos de series de tiempo estacionarios y estacionales (ARIMA, SARIMA, SARIMAX y Exponential Smoothing) sobre los datos de radiación solar del municipio de Mompox-Bolívar.
Metodología: A nivel metodológico, dentro de esta investigación fue adaptada la metodología CRISP-DM a 4 fases a saber: F1. Entendimiento del negocio y de los datos, F2. Preparación de los datos, F3. Modelado, F4 Evaluación y Despliegue del modelo.
Resultados: Como resultado, se obtuvo que a nivel de los modelos estacionarios, el modelo ARIMA (3,0,3) fue el que obtuvo un mejor ajuste, con un coeficiente de determinación moderado para el conjunto de entrenamiento y prueba. Así mismo, con respecto a los modelos con componente estacional, se determinó que el modelo SARIMA basado en ARIMA(3,0,3) y con componente estacional (1,0,1,8) fue el que tuvo el mejor rendimiento, con un buen ajuste en el conjunto de entrenamiento (R2 de 0.8571) y prueba (R2 de 0.8152).
Conclusiones: Se concluye que la incorporación de la variable exógena al modelo SARIMA no mejoró el rendimiento del modelo SARIMA, de tal modo que el modelo SARIMAX obtuvo un buen rendimiento en el conjunto de entrenamiento y un rendimiento moderado en el conjunto de prueba. Así mismo, los modelos obtenidos en esta investigación, son de gran utilidad para la caracterización del potencial de la radiación solar, así como para la planificación eficiente en el despliegue de sistemas de energía solar fotovoltaica.

Palabras clave: ARIMA, energía solar, radiación solar, SARIMA, series de tiempo.

Introduction

Solar radiation is the radiant energy emitted by the Sun through visible light as a result of the nuclear fusion process, manifesting as electromagnetic waves with wavelengths between 0.30 μm and 3 μm, including photons in the visible spectrum [1]- [3]. Upon reaching the Earth's atmosphere, some of this radiation is attenuated and divided into direct and diffuse components, both of which are usable for electricity generation [4]. This component categorization is essential for understanding the availability of solar energy under different meteorological and geographical conditions [5]. Solar radiation is fundamental to life, climate, biological processes, and nuclear fusion processes on Earth, spanning wavelengths from ultraviolet to infrared and including chemicals, as well as regulating ecosystems [6], [7]. However, the continued use of fossil fuels generates CO_2 emissions that accelerate climate change, worsening global environmental challenges [8], [9]. Conventional oil and coal reserves could be depleted within the next century, making early adoption of renewable energy sources critical [9], [10].

As the foundation of all renewable energy technologies, solar radiation drives wind generation, water evaporation, and biomass growth [7], [11]. Photovoltaic panels and concentrated solar power (CSP) systems reliably convert sunlight into electricity or heat [7], [12]. Therefore, the use of solar radiation not only reduces dependence on fossil fuels but also significantly reduces greenhouse gas emissions [2],

[11]. In this regard, solar energy's technical potential far exceeds the global electricity consumption, making it a viable and scalable option for the future energy landscape [7].

Solar energy stands out as one of the most promising renewable sources compared to other alternatives. Its main advantage is its virtually inexhaustible nature and its worldwide availability, making it a viable option even in remote regions or those with limited access to other energy sources [13]-[15]. Off-grid photovoltaic systems represent the most cost-effective solution for basic electricity supply in regions without grid connection [16], [17]. In this regard, unlike fossil fuels, solar energy also has the advantage of not producing polluting emissions or greenhouse gases during its operation, helping to combat climate change and environmental pollution [18], [19]. Likewise, solar energy reduces dependence on imported energy resources, such as fossil fuels (oil, natural gas, coal), while decreasing vulnerability to price fluctuations and geopolitical risks, simultaneously promoting economic development and local job creation in manufacturing, installation, and maintenance of solar systems [13], [14].

The study and characterization of solar radiation is fundamental because it is an essential resource that enables the generation of electricity in PV systems. In this regard, an increase in the intensity of solar radiation significantly enhances the power and energy generated by PV panels. For example, when going from 1000 to 3000 W/m^2, electrical power can increase almost three times, and the total energy generated can increase by up to 179% [20], [21]. Thus, the variation in the duration and intensity of solar radiation according to the time of day affects both daily and annual energy production [22], [23]. According to IDEAM data, Mompox-Bolívar records average irradiance levels exceeding 5 kWh/m^2·day, positioning it as a high solar potential [24]. Therefore, regions with high radiation levels can generate more electricity with the same infrastructure, thereby optimizing investment and reducing costs [23], [25]. Based on the above, evaluating and harnessing the maximum potential of solar radiation allows the optimization of the location, orientation, and operation of PV panels, thereby increasing the overall efficiency of the system [26].

Various state-of-the-art studies have been conducted on solar radiation and its characterization at different latitudes worldwide. Thus, in [27], [28], several studies were conducted in the Turkish provinces of Kırklareli, Tokat, Nevşehir, and Karaman, using algorithms such as artificial neural networks (ANN), support vector machines (SVM), k-NN, and deep learning to predict daily solar radiation with high accuracy. Similarly, in [29], empirical models and ML algorithms were compared to predict diffuse solar radiation at 17 meteorological stations in China, covering different climates and atmospheric conditions. Similarly, in [30], recurrent neural network (RNN) and ANN models were applied using meteorological data from Alabama, United States, with the result that RNN models provide higher prediction accuracy. Although these studies report high prediction accuracy, most of them rely on AI or ML approaches that require multiple meteorological variables, large training datasets, and high computational resources.

Consequently, their implementation may be difficult in regions where long-term historical information or multiple environmental variables are not readily available.

In Colombia, a study was conducted in [5] based on the descriptive statistics of radiation data from the city of Cúcuta, taking into account a 9-year history, aiming at the quantification and prediction of available solar resources. Similarly, in [31], the contribution of the so-called Solar Atlas of Colombia, an online tool that allows visualization of meteorological data and estimation of photovoltaic generation potential across the entire Colombian territory, was proposed, using satellite data validated with local measurements and regional climate models. Furthermore, in [32]-[34], algorithms such as ANFIS and regression models were used to predict solar radiation in specific areas, such as Cajicá and the Atlantic coast, achieving low prediction errors (below 1% in some cases). Similarly, in [35], statistical models, such as Markov chains and WRF, were used for characterization and short-term and hourly forecasting, especially in urban areas like Medellín, showing that these methods can be computationally efficient alternatives. In the same vein, in [36], a study was conducted in La Guajira, Colombia, using solar radiation meteorological data from IDEAM for more than five years, evaluated with the Bird and Hulstrom model, with the objective of quantifying and identifying areas with greater potential for solar technology development, considering geometric, geographic, astronomical, physical, and meteorological variables. Similarly, in [37], a statistical methodology was developed in Bogotá, Colombia, to characterize solar irradiance and ambient temperature, using real measurements from 6:00 a.m. to 6:00 p.m., with probability distribution function models and Monte Carlo simulations, to assess the performance of solar systems, such as photovoltaic and thermal systems. Although previous studies have compared statistical and ML approaches for solar radiation forecasting, their results strongly depend on the geographical location, temporal resolution, climatic conditions, input variables, and forecasting horizon [38], [39]. In the Colombian Caribbean, previous research has evaluated statistical and machine-learning models for solar radiation forecasting [35], [40]; however, the comparative performance of ARIMA, SARIMA, SARIMAX, and Exponential Smoothing models under a common evaluation framework has not been specifically examined for the municipality of Mompox-Bolívar. In addition, the effect of incorporating wind speed as an exogenous variable into statistical time-series models requires a local evaluation because the contribution of meteorological predictors may vary according to regional climatic conditions [39]. Therefore, this study addresses this location-specific methodological gap by comparing classical time-series models for solar radiation forecasting in Mompox-Bolívar.

To address the aforementioned research gaps, this study proposes the implementation, adjustment, and evaluation of various time series models for the characterization, analysis, and prediction of solar radiation in the municipality of Mompox-Bolívar, with the aim of identifying the feasibility and solar potential of this municipality for the deployment of renewable energy projects, specifically photovoltaic solar energy. Thus, using historical solar radiation data from the municipality of Mompox-Bolívar for the years 2021 and 2022, the adjustment and evaluation of the fitting capacity of four time series models were performed: ARIMA, SARIMA (without exogenous variable), SARIMAX with the exogenous variable of wind speed, and finally the Exponential Smoothing model. The parameters providing the best fit were adjusted for each of the considered models, and the performance of these models was compared using error metrics (mean absolute error (MAE), root mean squared error (RMSE)) and the coefficient of determination (R^2). The main scientific contributions of this research are as follows: (i) a systematic comparative evaluation of four classical statistical time-series models for solar radiation forecasting, (ii) identification of the most suitable forecasting model for solar radiation in Mompox-Bolívar, (iii) evaluation of the contribution of wind speed as an exogenous variable in SARIMAX-based forecasting, and (iv) proposal of a reproducible forecasting methodology based on CRISP-DM and free and open-source software that can be readily adapted to other geographical regions.

The results obtained in this research are highly useful for photovoltaic energy providers and government entities because they allow for identifying the available energy potential in different months of the year and, crucially, for anticipating the availability of this resource in advance. Thus, the potential of the forecasts generated by the time series models lies in optimizing the operational and strategic planning of solar plants, improving grid management, minimizing risks associated with the intermittency of the source, and facilitating informed decision-making for investment and efficient deployment of photovoltaic infrastructure in the municipality of Mompox, providing a clear view of the viability of solar energy projects in the region and promoting sustainable development. Furthermore, given the use of free and open-source software (FOSS) technologies, universities and research centers can use this work to extrapolate the considered models in the characterization and prediction of solar radiation data in different latitudes.

The rest of the article is organized as follows: Section 2 outlines the various phases of the CRISP-DM methodology used in carrying out this research. Section 3 presents the results obtained in this study, which includes the identification of the parameters for the 4 models considered, associated with stationarity and seasonality, as well as the adjustment and evaluation of the different models, aimed at determining the parameters that provide the best fit in each case, considering the error metrics and the coefficient of determination (R^2). Finally, this phase includes forecast generation for each model. Lastly, Section 4 presents the conclusions and future work derived from this research.

Methodology

For the development of this research, an adaptation of the CRISP-DM methodology was made into four phases: P1. Business and data understanding, P2. Data preparation, P3. Modeling, P4. Evaluation and Deployment. This methodology was adapted considering that it constitutes a standardized process model, widely used in the development of data mining and machine learning projects, characterized by its independence from the industrial sector and the type of technology used, making it versatile and applicable to various contexts [41]-[44].

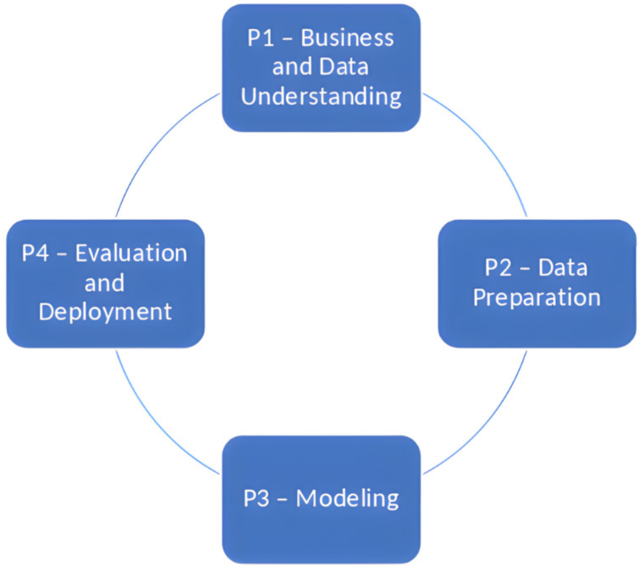


Figure 1.*Adaptation of CRISP-DM*

Source: Authors.

In phase 1 of the methodology, solar radiation data for the municipality of Mompox for the years 2021 and 2022 were obtained from the Institute of Hydrology, Meteorology, and Environmental Studies (IDEAM) portal, with data captured at an hourly frequency, resulting in a total dataset of 17,529 data points. Likewise, in this phase, the existence of possible missing values was identified, and a visual analysis of the weekly and monthly radiation for the Mompox municipality was conducted. Thus, Figure 2 presents a sample of the solar radiation time series, where one data point is shown for every 30 points to make the series less dense and facilitate the analysis of its general behavior. Based on the data in Figure 2, it is possible to identify that the time series exhibits variability, suggesting some stationarity around a constant mean over time. Additionally, a clear periodicity is observed, indicating a possible seasonal behavior, with peaks and valleys repeating in regular cycles, which will be verified by identifying the model parameters.

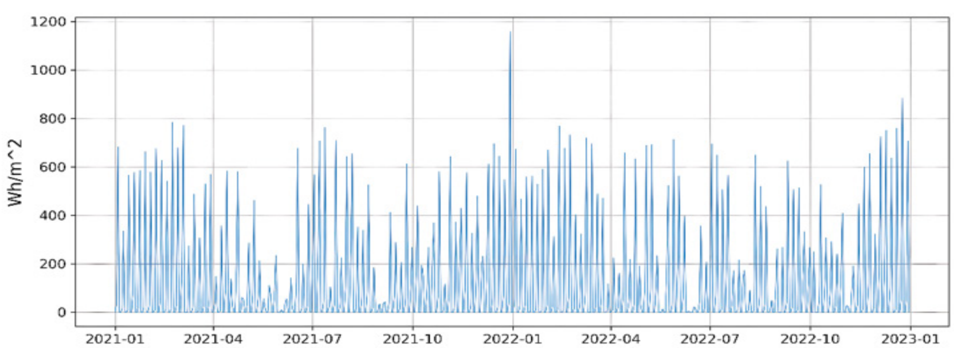


Figure 2. *Solar radiation in the municipality of Mompox during the years 2021 and 2022*

Source: Authors.

In line with Figure 2, Figure 3 presents the weekly average of solar radiation for the municipality of Mompox during 2021 and 2022.

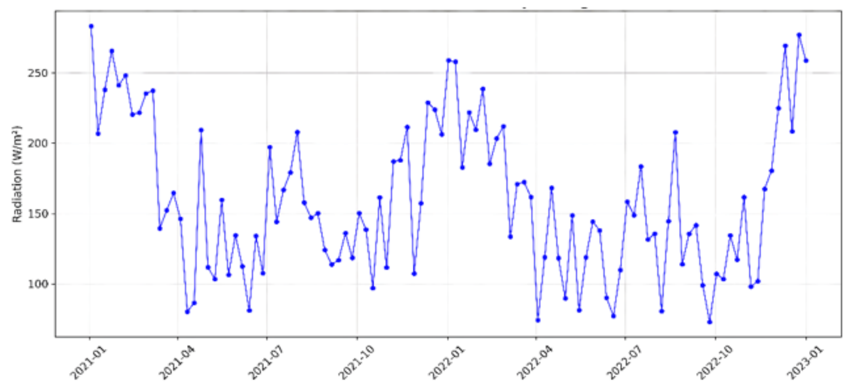


Figure 3. *Weekly Average Solar Radiation for the Municipality of Mompox, Bolívar*

Source: Authors.

From Figure 3, a pattern of seasonality characterized by recurrent peaks and pronounced valleys can be clearly identified. These higher radiation peaks are consistently observed toward the end and beginning of each year, specifically during the periods of December-January-February and, to a lesser extent, around June-July. Similarly, the minimum radiation values tend to occur in periods such as April-May and October-November, coinciding with climate transition periods. Despite these seasonal fluctuations, the series exhibits variability, suggesting some stationarity in its overall behavior, as the values seem to oscillate around a constant mean without any evident long-term upward or downward trend throughout the two years analyzed. This will be further analyzed in the subsequent phases of the methodology.

In phase 2 of the methodology, the radiation data were captured at a frequency of 1 h, meaning 24 samples per day; therefore, an initial resampling of the time series data was performed at a frequency of 3 h. This process was specifically implemented to achieve greater series smoothing by averaging the radiation every 3 h using the *resample()* function from the pandas library. The goal was to attenuate high-frequency fluctuations, smooth the data trend, and consequently improve the fitted models' efficiency and predictive performance. Likewise, once the series resampling and smoothing were completed, the training and test sets were obtained with a respective ratio of 80% and 20%. Thus, all the radiation data from 2021 were assigned to the training set (80%), while the remaining 20% were derived from the corresponding proportion of the 2022 data. This strategic decision to include a full annual cycle of 2021 in the training set is crucial for the model to thoroughly capture the annual trend and, importantly, the series' inherent seasonality, which is ideal for fitting and improving the performance of models such as ESS, SARIMA, and SARIMAX. Thus, Figure 4 presents the separation of the time series data into the training and test sets, with one data point shown for every 10 points to improve visualization.

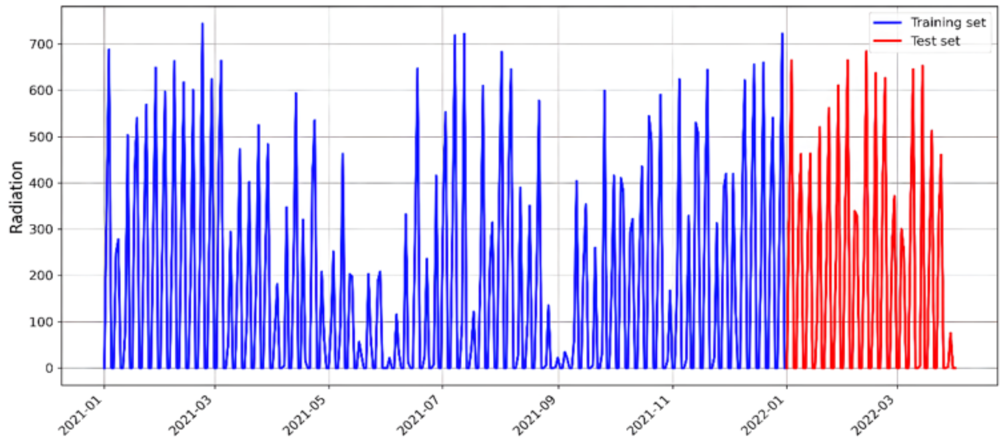


Figure 4. *Time series training and test sets.*

Source: Authors.

Figure 4 shows the training set of the series in blue (80%), which corresponds to the complete radiation data for 2021 (2920 readings representing the average radiation every 3 h). Likewise, in red, the test set is shown, which corresponds to the portion of the 2022 data, representing one-fifth of the total data or one-fourth of the total training set (730 readings of the average radiation every 3 h).

Once the training and test sets of the time series with the solar radiation data from the municipality of Mompox were obtained, phase 3 of the methodology proceeded with the fitting of the ARIMA, SARIMA, exponential smoothing, and SARIMAX models using the data from the training set of the series. For the ARIMA model, prior to fitting with the training set, the potential model parameters *d*, *p*, and *q* were identified, respectively: the order of differencing required to make the series stationary (*d*), the order of the autoregressive part (AR) (*p*), and the order of the moving average part (MA) (*q*). The estimation of the parameter *d* was performed using the Dickey-Fuller stationarity test implemented in the Python statsmodel library, while a visual inspection was carried out on the partial autocorrelation and simple autocorrelation plots, respectively, for the potential values of the parameters p and q, to identify the relevant lags, using the same library. Based on this set of potential parameters, iterations were carried out with several ARIMA models using the previously identified combinations of p, q, and d parameters to determine the ARIMA models whose components are significant and, from these, identify the model with the best fit according to the AIC and BIC metrics. In this same regard, the same non-seasonal parameters d, p, and q from the best-fitting ARIMA model were used as reference for the SARIMA and SARIMAX models (the latter including wind speed as an exogenous variable). Then, the seasonal parameters P, D, Q, and s were identified, and systematic iterations were performed over several configurations of these parameters to identify those that provided the best seasonal fit to the observed periodicity in the time series, using the coefficient of determination (R^2). Similarly, the seasonal component was obtained for the exponential smoothing model, which is associated with the parameter of seasonal periods, and iterations were carried out with different values of this parameter to identify the model with the best seasonal fit according to the R^2 metric. The ARIMA model can be defined according to Equation (1). This formulation is the standard in time series analysis and aligns with [45], who established the theoretical foundation for ARIMA models [45]. Further validation of its application to environmental data can be found in [46].

$$(1-\phi_1 B-\phi_2 B^2-\cdots-\phi_p B^p)(1-B)^d y_t=(1+\theta_1 B+\theta_2 B^2+\cdots+\theta_q B^q)\varepsilon_t \tag{1}$$

Equation (1), which mathematically represents the ARIMA model, y_t corresponds to the time series analyzed at different time periods *t*; B represents the lag operator; $(1-B)^d$ represents the differencing component of order *d*; $\phi_1,\phi_2,....,\phi_p$ are the coefficients of the autoregressive (AR) component of order *p*; $\theta_1,\theta_2,....,\theta_q$ are the coefficients of the moving average (MA) component of order *q*. For ARIMA models, it is necessary to identify the parameters *p*, *q*, and *d*. The SARIMA model can be mathematically expressed using equation (2). SARIMA's efficacy for seasonal solar radiation data is supported by research from [47], who demonstrated its superiority in capturing periodic patterns in renewable energy datasets [47].

$$\phi(B^s)\phi(B)(1-B)^d(1-B^s)^D y_t=\Theta(B^s)\theta(B)\varepsilon_t \tag{2}$$

In equation (2), which mathematically represents the SARIMA model, y_t corresponds to the time series analyzed at different time periods *t*; B is the lag operator; $\phi(B)$ is the autoregressive (AR) polynomial of order *p*; $\theta(B)$ is the polynomial representing the moving average (MA) of order *q*; $(1-B)^d$ is the differencing operator of order *d*; *s* is the seasonality period parameter; $(1-B^s)^D$ corresponds to the seasonal differencing of order *D*; $\phi(B^s)$ is the seasonal autoregressive polynomial of order *P*; and $\Theta(B^s)$ represents the seasonal moving average polynomial of order *Q*. Thus, the SARIMA model, in addition to the parameters *p*, *q*, and *d*, includes parameters related to seasonality *P*, *D*, and *Q*. In the same regard, equation (3) mathematically represents the SARIMAX model, which is the SARIMA model with an exogenous variable. The integration of exogenous variables follows the framework proposed by [48] for structural time series models [48].

$$\phi(B^s)\phi(B)(1-B)^d(1-B^s)^d y_t=\Theta(B^s)\theta(B)\varepsilon_t+\beta X_t \tag{3}$$

Equation (3), which mathematically represents the SARIMAX model, includes, in addition to the same components as in equation (2), X_t, also known as the exogenous variable vector over time, and β, which represents the coefficient vector that measures the influence of X_t on y_t. Finally, equation (4) mathematically represents the exponential smoothing model. This formulation is derived from Holt-Winters' method, as detailed in [49], and is widely used for forecasting seasonal time series in energy applications [50].

$$\hat{y}_{t+h|t} = l_t + hb_t + s_{t+h-m(k+1)} \quad (4)$$

In equation (4), which represents the Exponential Smoothing model, $\hat{y}_{t+h|t}$ represents the forecast for time $t+h$, made at time t; l_t is the level of the series at time t; b_t is the estimated trend at time t; $s_{t+h-m(k+1)}$ is the seasonal component corresponding to $t+h$; h is the number of periods ahead; m is the length of the seasonal cycle, and k is the number of complete cycles that have passed since t.

Finally, in phase 4, the reports for each of the models included the MAE metric, RMSE metric and the coefficient of determination (R^2), both for the training set and the test set, comparing the latter with each model's forecasts. The choice of these metrics was crucial, as RMSE quantifies the average magnitude of the prediction errors, while R^2 allows for evaluating the proportion of variance in the dependent variable that can be predicted from the independent variables, thus determining the best model among the four evaluated. These metrics were also contrasted with a visual inspection, where the training and test sets were compared with the time series model curves.

Results

An initial analysis was performed on the monthly average of the solar radiation values from the time series for the years 2021 and 2022, which are presented in Figure 5. This was done to identify the series' monthly trends, as well as its seasonality and stationarity.

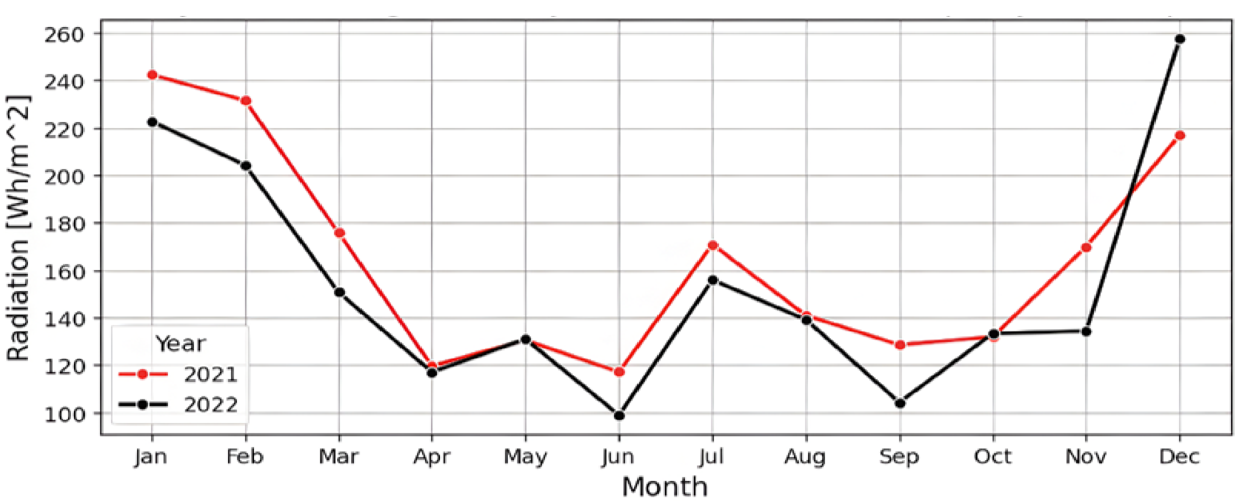


Figure 5. *Average monthly radiation in the municipality of Mompox, Bolívar.*

Source: Authors.

Based on the results from Figure 5, a clear seasonal trend can be observed throughout 2021 and 2022, such that consistently high radiation values are observed in December in both years, with a secondary peak in January for 2021 and a sharp increase toward December in 2022. Similarly, the minimum radiation values generally occurred between May and June in both years. This seasonal variability, characterized by higher radiation during the dry season (December-March) and lower values coinciding with the rainy season (April-November), is consistent with the Colombian Caribbean region's typical climatic behavior. Thus, December 2022 exhibited the highest radiation peak, approximately 258 Wh/m^2, while June 2022 recorded the lowest value, approximately 98 Wh/m^2. The seasonal trend of the series suggests that it can be better modeled using seasonal models than those based purely on stationarity.

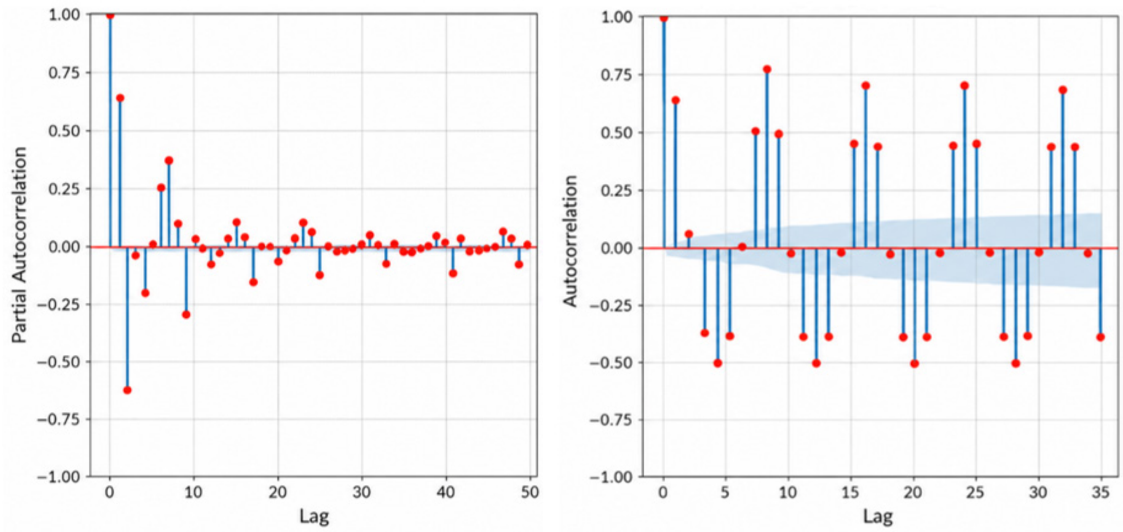
Now, regarding the ARIMA model and considering that this model requires the identification of the parameters p, d, and q, the first step is to perform the Dickey-Fuller stationarity test on the initial series, as well as on the first and second differences, to identify the possible values of the parameter d, which refers to the number of times the series needs to be differenced to achieve stationarity. Table 1 presents the results of the Dickey-Fuller test applied to the training set of the series.

Table 1. *Results of the Dickey-Fuller test application.*

DIFFERENCING OF THE SERIES	**DICKEY-FULLER STATISTIC**
ORIGINAL SERIES	ADF STATISTIC: -7.509485682578794 P-VALUE: 4.0510952708071326E-11
FIRST DIFFERENCING	ADF STATISTIC: -15.689051499568787 P-VALUE: 1.4413817315620087E-28
SECOND DIFFERENCING	ADF STATISTIC: -22.488760186276057 P-VALUE: 0.0

Source: Authors.

According to the results shown in Table 1, considering the p-value obtained for the original series, as well as for the first and second differencing, in all three cases, the series rejects the null hypothesis of non-stationarity, meaning that the original series, along with its first two differences, are stationary. However, for the original series, the p-value is significantly lower than conventional significance levels, leading to the decisive rejection of the nonstationarity hypothesis.



(a) (b)

Figure 6. *Partial (a) and simple autocorrelation (b) of the original series.*

Source: Authors.

Likewise, given that the highly negative ADF statistic for the original series strongly indicates that the series does not contain a unit root and is therefore stationary with an integration order of d = 0. Thus, although d = 1 and d = 2 are feasible, excessive differencing may introduce spurious correlations and reduce the model's interpretability. Regarding the identification of the parameters p and q of the ARIMA model, the process proceeded with the visual analysis of the partial and simple autocorrelation of the original series using a lag number of 35, which corresponds to approximately 10% of the training data.

From Figure 6, a significant peak is observed at the first lag (lag 1) in the partial autocorrelation (PACF) plot (Figure 6a), suggesting a strong direct correlation. Likewise, the PACF plot shows other notable peaks at lags 2 and 3, as well as at 7, 8, 11, 12, 13, and 14, among others, which cross the significance threshold, indicating that the current radiation is influenced by values up to three months prior and also by more complex seasonal or sub-seasonal patterns, offering potential values for the autoregressive order (p) that could include 1, 2, 3. On the other hand, the ACF plot (Figure 6b) shows a gradual decay, which is consistent with a stationary series. Therefore, significant peaks are identified at lags 1, 2, and 3, along with patterns that repeat annually (lag 12, 24, etc.) and subannually (lag 6, 7), crossing the significance threshold. This suggests the presence of moving average components (q) that could be 1, 2, 3, reflecting past errors' dependencies. The persistence of autocorrelation at multiples of 12 also suggests potential seasonality in the moving average. Therefore, it is important to evaluate the significance of the components in the equation for ARIMA models, considering the possible ranges of the previously identified parameters p, q, and d.

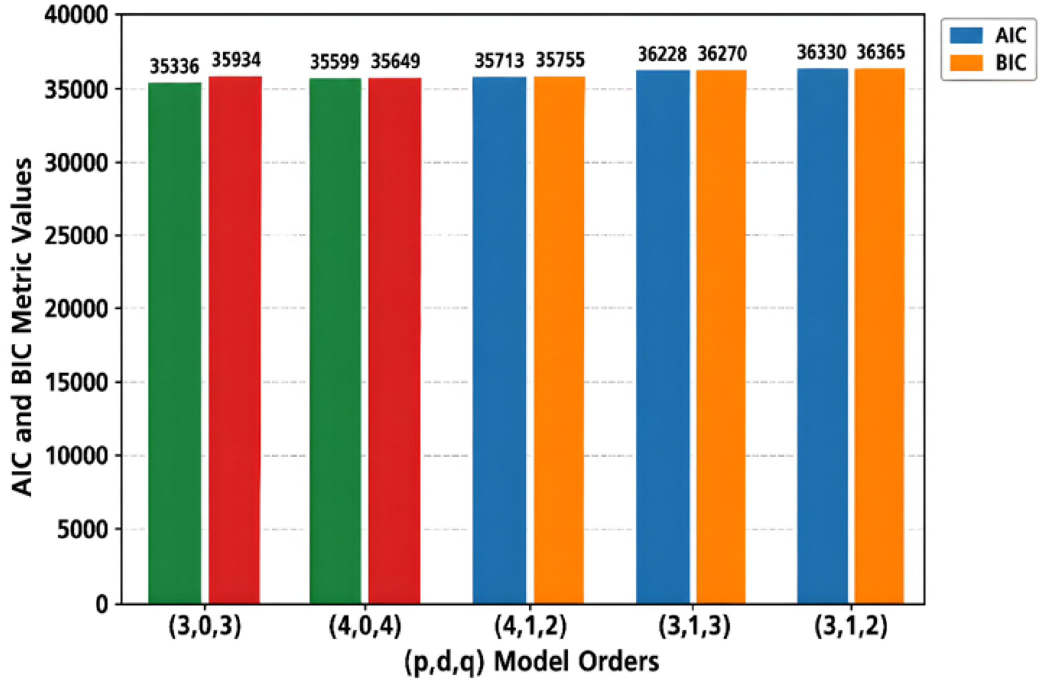


Figure 7. *Comparison of the 5 ARIMA models with the best AIC and BIC metrics.*

Source: Authors.

Once the evaluation of the different ARIMA models was carried out, it was found that the 10 models whose equation components are significant (p-value < 0.05) and also provide a better fit according to the AIC and BIC fit and goodness metrics are: ARIMA(2,0,1), ARIMA(2,1,1), ARIMA(3,1,1), ARIMA(4,0,0), ARIMA(2,1,2), ARIMA(3,1,2), ARIMA(3,1,3), ARIMA(4,1,2), ARIMA(4,0,4), and ARIMA(3,0,3). In this regard, Figure 7 presents a graph that allows comparing the top 5 models based on the AIC and BIC metrics from the 10 models mentioned earlier, which help identify the models with the best optimal trade-off between fit and complexity, indicating that they allow for selecting the most parsimonious model without sacrificing the goodness of fit.

Figure 7 shows that the ARIMA (3,0,3) model achieves the lowest values in the AIC and BIC metrics, indicating that it is the model with the best fit for the evaluated solar radiation time series, capturing temporal dependencies and underlying patterns in the data more accurately. Figure 8 shows the comparison of the fit of the ARIMA (3,0,3) model against the training and test data of the time series.

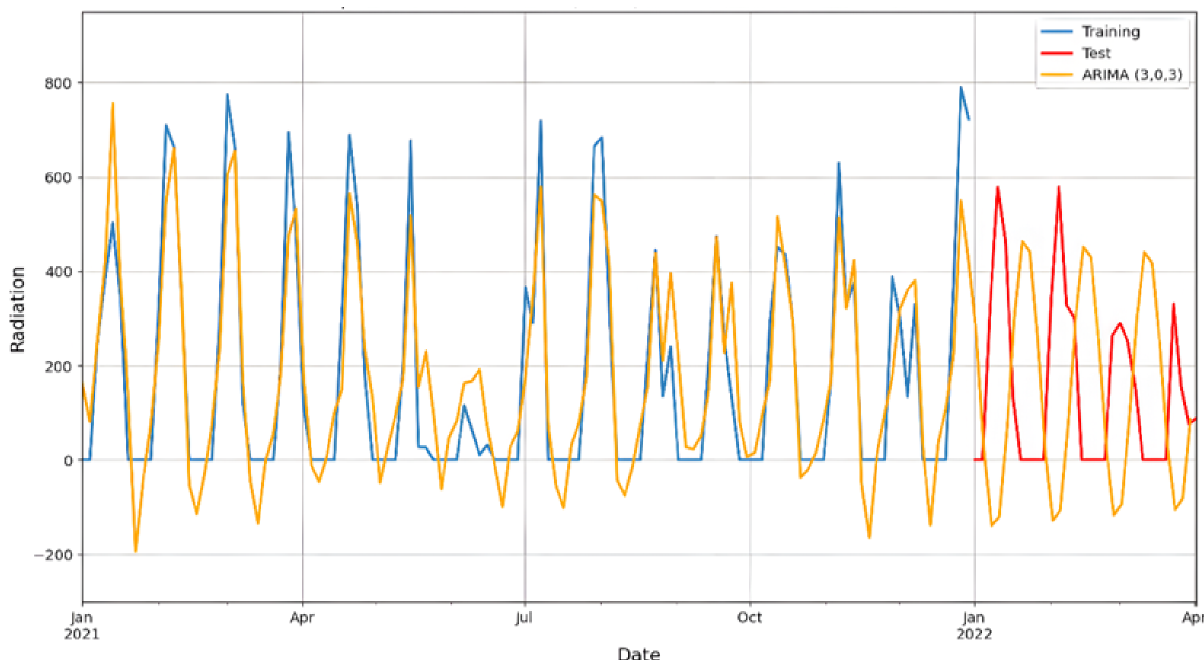


Figure 8. *Comparison of the ARIMA (3,0,3) model with the time series data.*

Source: Authors.

In Figure 8, it is possible to observe how the ARIMA(3,0,3) model attempts to closely follow the peaks and valleys presented in the training set, while the model's predictions show lower precision and less adherence to the fluctuations. The metrics obtained in the coefficient of determination (R^2), which is 0.7751 for the training set and 0.7031 for the test set, more clearly demonstrate this. That is, while the fit in the training set is good, with 77.51% of variability explained, the fit in the test set is moderate, explaining only 70.31% of the variability, reflecting a loss in the model's ability to generalize the fluctuations in the test set.

Regarding the SARIMA time series model, it was found that starting with the p, d, and q parameters of the ARIMA (3,0,3) model, the 5 models that are significant in their components and provide the best fit when iterating over the seasonal parameters P, D, Q, and s are presented in Figure 9.

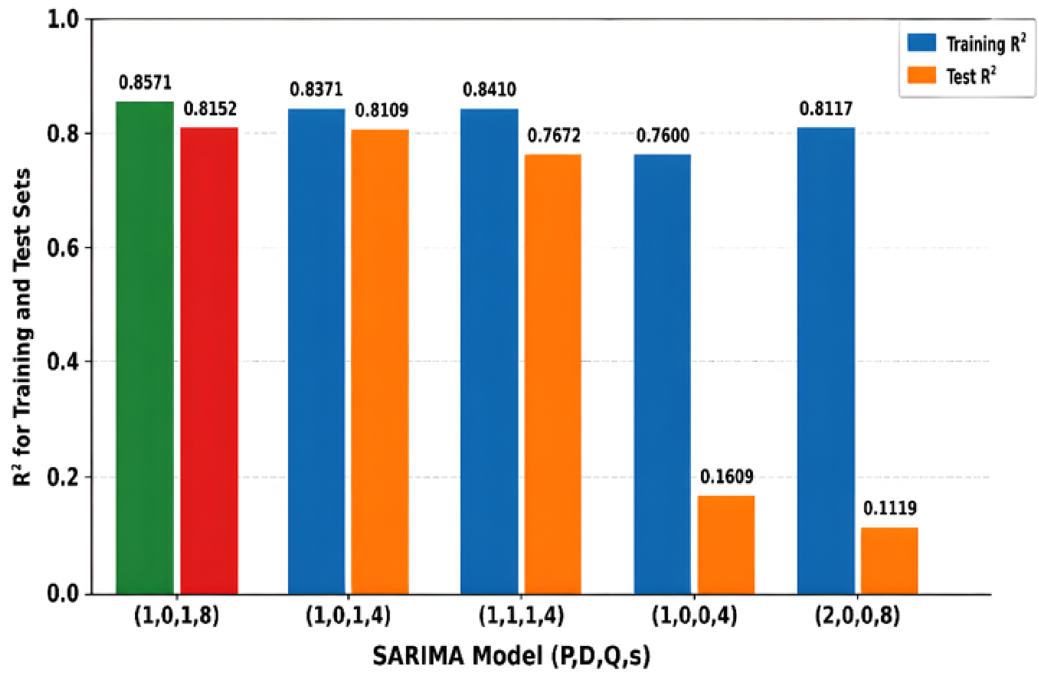


Figure 9. *SARIMA models with the best fit in the training and testing datasets.*

Source: Authors.

According to the results presented in Figure 9, the SARIMA model with the best fit is the one with the parameters P, D, Q, and S: 1, 0, 1, 8, which indicates that the seasonal period fits best every 24 hours because the data were captured every 3 hours. This seasonality can be explained by the fact that the dataset corresponds to solar radiation data from the municipality of Mompox, which follows a daily cycle characterized by variations in radiation throughout the day, making the SARIMA model with these parameters well-suited to capture the solar fluctuations that repeat over a 24-hour cycle. Figure 10 shows the comparison of the fit of the SARIMA model with the parameters P, D, Q, and s (1,0,1,8) against the training and test data of the time series.

Figure 10 shows how the SARIMA model with the best fit accurately follows both the training and test sets of the solar radiation time series from the municipality of Mompox. This good fit is clearly reflected in the coefficients of determination (R^2), which are 0.8571 for the training set and 0.8152 for the test set, indicating that the model explains 85.71% and 81.52% of the variability in the training and test data, respectively, demonstrating that the model maintains a high level of explainability and good performance in both datasets, suggesting an excellent ability to generalize the fluctuations in solar radiation.

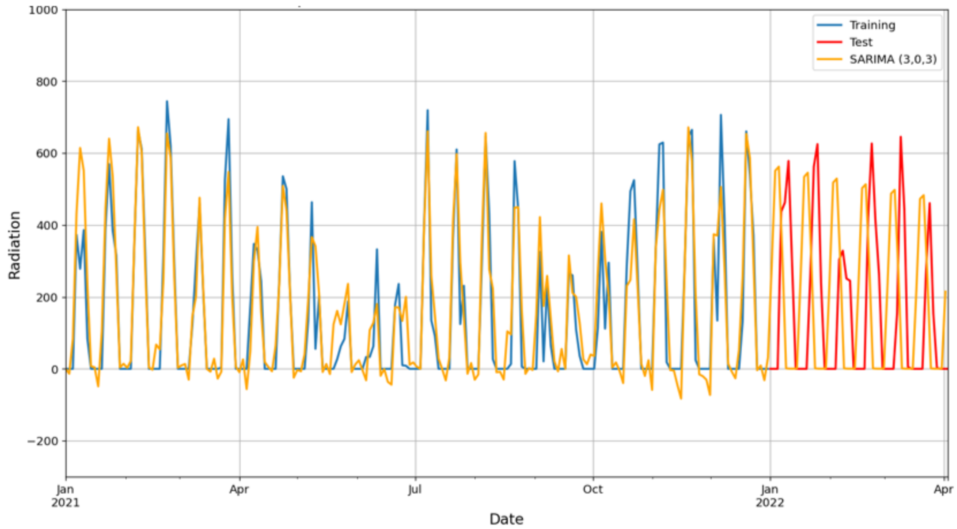


Figure 10. *Comparison of the SARIMA model and time series data.*

Source: Authors.

Wind speed was taken as the exogenous variable in the SARIMAX model, which was also partitioned into the training and test sets and included in the SARIMA model with the best fit (seasonal order (1,0,1,8)). Thus, Figure 11 presents a graph in which the obtained SARIMAX model is compared against the training and test sets of the solar radiation time series.

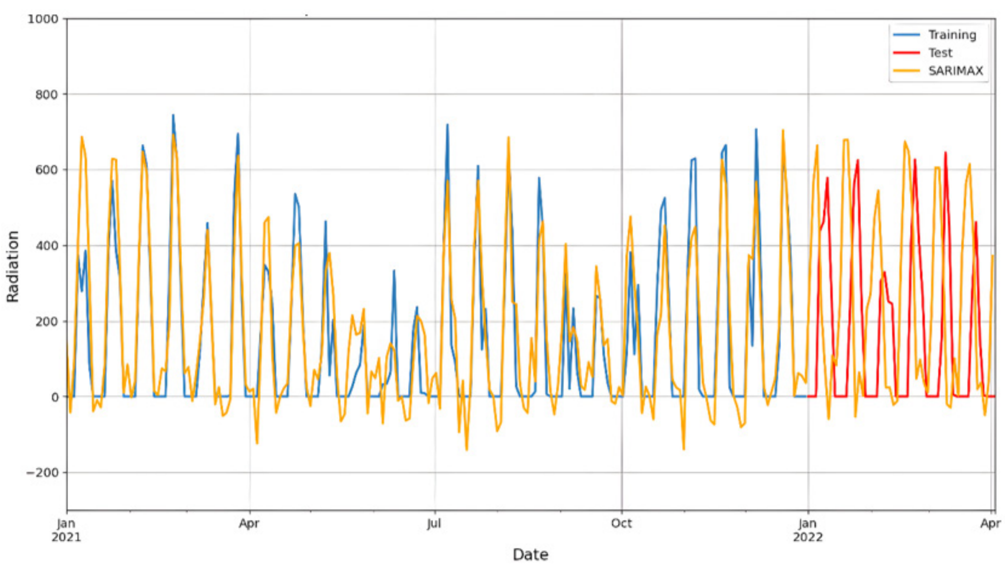


Figure 11. *Comparison of the SARIMAX model and time series data.*

Source: Authors.

Figure 11 shows how the SARIMAX time series model, which incorporates the exogenous variable wind speed, provides a good fit in the training set, which is comparable to the performance of the SARIMA model. However, the fit or performance of the model in the test set is lower than that of the SARIMA model. This is clearly reflected in the determination coefficients obtained by the SARIMAX model, which are 0.8203 for the training set and 0.7344 for the test set. The results show that the SARIMAX model explains 82.03% of the variability in the training set but only 73.44% in the test set, suggesting that its generalization ability is lower than that of the SARIMA model, especially when facing data not observed during training.

Finally, the graph shown in Figure 12 is obtained when adjusting and evaluating the exponential smoothing model against the training and test sets of the series. In this case, the model achieved the highest performance with a seasonal parameter s of 272.

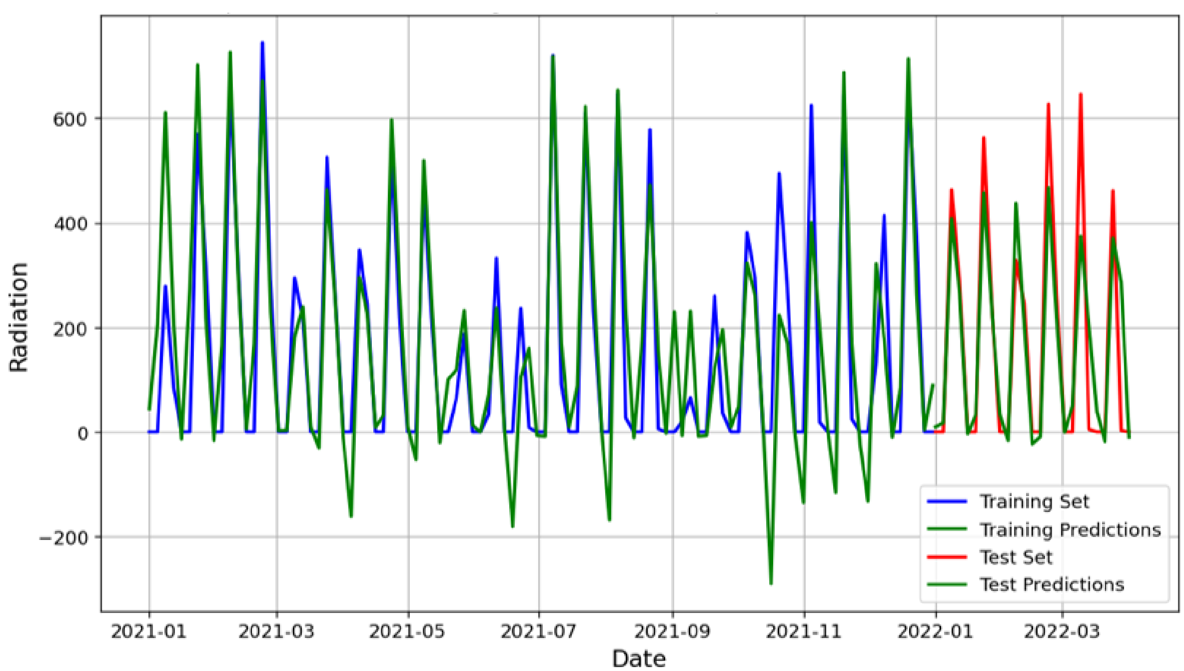


Figure 12. *Comparison of the exponential smoothing model with respect to the time series data.*

Source: Authors.

In Figure 12, it is possible to observe how the exponential smoothing model presents valleys below zero in the curve, which are more pronounced compared to the ARIMA and SARIMA models, although it attempts to closely follow the fluctuations of the time series. According to the above, the coefficient of determination obtained for the training set is 0.7863, whereas that for the test set is 0.7561, indicating that the model explains 78.63% and 75.61% of the variability in the training and test sets, respectively.

this shows a moderate performance of the exponential smoothing model, which is superior only to the ARIMA(3,0,3) model but slightly inferior to the performance of the SARIMA models, which achieve a more robust fit, especially in the test set.

Finally, in order to compare the results obtained by the four evaluated models, Table 2 presents a comparative table with the MAE, RMSE, and R^2 metrics obtained for each model.

Table 2. *Error and performance metrics obtained from each model.*

DIFFERENCE OF THE SERIES	TRAINING METRICS	TEST METRICS
ARIMA (3,0,3)	MAE: 84.3953 RMSE: 106.0545 R^2: 0.7751	MAE: 106.0854 RMSE: 130.1029 R^2: 0.7031
SARIMA (3,0,3,1,0,1,8)	MAE: 58.1052 RMSE: 84.5519 R^2: 0.8571	MAE: 57.2445 RMSE: 102.6592 R^2: 0.8152
SARIMAX	MAE: 69.8946 RMSE: 94.8052 R^2: 0.8203	MAE: 86.7577 RMSE: 123.0639 R^2: 0.7344
EXPONENTIAL SMOOTHING	MAE: 71.0793 RMSE: 103.3768 R^2: 0.7863	MAE: 77.4377 RMSE: 118.1749 R^2: 0.7561

Source: Authors.

According to the results in Table 2, the SARIMA model provides the best fit and lowest error level among the four models, both in the training and test sets, demonstrating consistent performance across both datasets. This is supported by its higher R^2 (0.8571 in training and 0.8152 in testing) and the lowest MAE (58.1052 and 57.2445, respectively) and RMSE (84.5519 and 102.6592, respectively) values. In contrast, the ARIMA (3,0,3) model exhibited the lowest performance in both the training and test sets, with the highest error levels in the MAE and RMSE metrics, particularly with an R^2 of 0.7031 in the test set and an MAE of 106.0854. Based on these findings, it can be concluded that the SARIMA model, built on the ARIMA (3,0,3) model with seasonal parameters (1,0,1,8), is the most effective in characterizing and fitting the time series solar radiation data from the municipality of Mompox, Bolívar. It demonstrates robust predictive capabilities and superior fit in both the training and test data, making it the optimal choice for future estimations.

As a discussion, it is important to mention that this work proposed the characterization of the solar radiation data from the municipality of Mompox-Bolívar as its contribution, as well as the fitting and evaluation of these data using four time series models (ARIMA, SARIMA, SARIMAX, and Exponential Smoothing). For this, the parameters that provide the best fit were determined in each case, taking into account the fit and goodness metrics (AIC and BIC), as well as the error and performance metrics (MAE, RMSE, and R^2). In this regard, it was first determined that the time series, by default and according to the Dickey-Fuller test, is stationary, justifying the use of time series models for its study and characterization. The stationarity of the series is a key feature that makes time series models, such as ARIMA, SARIMA, or exponential smoothing, particularly suitable for modeling and forecasting solar radiation, as these models are capable of capturing and adjusting the inherent temporal fluctuations and patterns in such data. This finding complements previous research, such as those conducted in [5] and [32]-[34], where descriptive statistics and regression models have primarily addressed characterization, fitting, and prediction. However, these approaches may not be effective in capturing the temporal and seasonal variations in the data given the stationary nature of solar radiation, making time series models more accurate and appropriate for this type of analysis.

These findings agree with previous studies reporting that statistical time-series models constitute robust alternatives for solar radiation forecasting when the analyzed series exhibit stationary and seasonal characteristics. In particular, recent comparative studies have shown that SARIMA models generally outperform conventional ARIMA models when strong periodic patterns are present, since the explicit incorporation of seasonal components allows a more accurate representation of recurrent solar radiation cycles [38], [39]. Similarly, the results obtained in this study indicate that the solar radiation time series in Mompox exhibits a strong seasonal component in addition to stationarity, as evidenced by the superior performance of SARIMA, SARIMAX, and exponential smoothing compared with the best-performing ARIMA model.

Although machine learning models have demonstrated excellent predictive performance in previous studies [29], the results obtained in this work indicate that seasonal statistical models constitute a suitable alternative for the analyzed dataset, offering good predictive performance while maintaining low computational complexity and high interpretability. Likewise, the results obtained for the SARIMAX model indicate that incorporating wind speed as an exogenous variable did not improve the predictive performance achieved by the SARIMA model. Similar observations have been reported in previous studies, where the contribution of exogenous meteorological variables depends strongly on local climatic conditions, data quality, and the degree of correlation between the auxiliary variables and solar radiation. Consequently, the usefulness of incorporating exogenous variables should be evaluated on a case-by-case basis rather than assumed a priori [39], [40].

Unlike many ML approaches, statistical time-series models explicitly preserve the chronological structure of observations and account for temporal dependence. Although recent machine learning and deep learning models have demonstrated excellent predictive capabilities, they generally require larger datasets, multiple meteorological variables, and higher computational resources than classical statistical models [29]. In contrast, using only historical solar radiation observations, the proposed SARIMA model achieved a good balance between forecasting accuracy, computational simplicity, and model interpretability. This characteristic makes it particularly suitable for regions such as many municipalities in developing countries where long historical records of multiple meteorological variables are not available [38], [39].

Finally, one limitation of stationary and seasonal time-series models is that they require periodic recalibration to adapt to long-term changes in climatic conditions. Climate change may alter solar radiation historical patterns, affecting both seasonal behavior and long-term trends. Consequently, the predictive performance of these models may deteriorate over time if they are not periodically updated. Therefore, future research should investigate adaptive and hybrid forecasting approaches capable of maintaining predictive accuracy under changing climatic conditions. In addition, future studies may evaluate the incorporation of additional meteorological variables and compare the proposed statistical models with state-of-the-art ML and DL approaches using the same dataset and evaluation framework.

Conclusions

This study proposed the characterization, fitting, and comparative evaluation of four classical time-series models (ARIMA, SARIMA, SARIMAX, and Exponential Smoothing) using solar radiation data from the municipality of Mompox-Bolívar. The Dickey-Fuller test confirmed that the analyzed time series is stationary, supporting the application of statistical time-series models for solar radiation forecasting. Furthermore, the superior performance of the seasonal models demonstrated that the solar radiation series exhibits a marked seasonal behavior that can be effectively represented through seasonal statistical models. Beyond the comparative evaluation of forecasting techniques, the main scientific contribution of this work lies in providing a systematic characterization of solar radiation through a unified evaluation framework that integrates stationarity analysis, model comparison, and the assessment of exogenous variables. This distinguishes the proposed methodology from previous studies that have mainly focused on descriptive analyses, regression models, or the application of individual forecasting techniques. In addition, the obtained results provide useful information for estimating the solar potential of Mompox and supporting the planning of photovoltaic energy systems.

At the level of stationary models, it was determined through the Dickey-Fuller stationarity test that the original series, as well as its first and second differences, reject the null hypothesis of non-stationarity. Similarly, through partial and simple autocorrelation, the feasible values for the p and q parameters were determined to be 1, 2, and 3. Thus, by iterating over the different feasible parameters and comparing the fit and goodness metrics, the significant model with the best fit is ARIMA(3,0,3), which achieved a coefficient of determination of 0.7751 for the training set and 0.7031 for the test set, allowing for a moderate adjustment of the model to the data's characteristics.

The SARIMA model based on the ARIMA(3,0,3) model at the stationary level and with a seasonal component (1,0,1,8) had the best fit compared to the SARIMAX (exogenous variable wind speed) and exponential smoothing models. In the training set, it achieved a coefficient of determination of 0.8571, while in the test set, it obtained a coefficient of determination of 0.8152, making it the best-performing model evaluated in this study. In contrast, although the SARIMAX model showed a good fit in the training set, its fit in the test set was moderate. Similarly, regarding the exponential smoothing model, although it closely follows the peaks and fluctuations of the series visually, the pronounced negative valleys affected its fitting capability, resulting in moderate performance for both the training and test sets, but still outperforming the ARIMA model.

Because this research was implemented entirely using free and open-source Python libraries, the proposed methodology can be readily reproduced and adapted for both academic research and practical applications, particularly for solar resource assessment in the Colombian Caribbean region. Specifically, the pandas library was used for temporal indexing and resampling of the data, statsmodels for model identification, fitting, and evaluation, and matplotlib for visualization of the time series and diagnostic plots.

Future research should evaluate forecasting models, such as Meta's Prophet and LSTM neural networks, under the same evaluation framework adopted in this study. Incorporating other meteorological variables, such as cloud cover, temperature, or relative humidity, into SARIMAX and hybrid forecasting models may provide further insight into the contribution of exogenous information to solar radiation prediction. Finally, extending the proposed methodology to other regions of the Colombian Caribbean would allow the evaluation of its robustness under different climatic conditions.

Acknowledgments

This article was developed within the framework of the project: "DESARROLLO SOSTENIBLE MEDIANTE LA IMPLEMENTACIÓN DE ENERGÍA SOLAR FOTOVOLTAICA EN ZONAS RURALES NO CONECTADAS A LA RED ELÉCTRICA NACIONAL EN MOMPOX, COLOMBIA (022-2024)", which is part of an internal research initiative at the University of Cartagena.

References

[1] G. N. Tiwari and L. Sahota, "Solar Radiation and Heat Transfer," 2017, pp. 63–119. [Online]. Available: https://doi.org/10.1007/978-981-10-4672-8_2.

[2] H. Hissou, S. Benkirane, A. Guezzaz, M. Azrour, and A. Beni-Hssane, "A novel machine learning approach for solar radiation estimation," *Sustainability*, vol. 15, no. 13, p. 10609, Jul. 2023. [Online]. Available: https://doi.org/10.3390/su151310609.

[3] J. Polo, L. Martín-Pomares, C. A. Gueymard, J. L. Balenzategui, F. Fabero, and J. P. Silva, "Fundamentals: Quantities, Definitions, and Units," 2019, pp. 1–14. [Online]. Available: https://doi.org/10.1007/978-3-319-97484-2_1.

[4] A. Madhlopa, "Solar Radiation Resource," 2018, pp. 51–64. [Online]. Available: https://doi.org/10.1007/978-3-319-68388-1_3

[5] W. Contreras Sepúlveda, M. G. Galban Pineda, and S. B. Sepúlveda Mora, "Análisis estadístico de la radiación solar en la ciudad de Cúcuta," *Entre Ciencia e Ingeniería*, vol. 12, no. 23, pp. 16–22, Jun. 2018. [Online]. Available: https://doi.org/10.31908/19098367.3698.

[6] A. E. Gürel, Ü. Ağbulut, H. Bakır, A. Ergün, and G. Yıldız, "A state of art review on estimation of solar radiation with various models," *Heliyon*, vol. 9, no. 2, p. e13167, Feb. 2023. [Online]. Available: https://doi.org/10.1016/j.heliyon.2023.e13167.

[7] J. Tovar-Pescador, "Modelling the statistical properties of solar radiation and proposal of a technique based on Boltzmann statistics," in *Modeling Solar Radiation at the Earth's Surface*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 55–91. [Online]. Available: https://doi.org/10.1007/978-3-540-77455-6_3.

[8] IEA, "World Energy Outlook 2023," 2023. [Online]. Available: https://www.iea.org/news/the-energy-world-is-set-to-change-significantly-by-2030-based-on-today-s-policy-settings-alone.

[9] E. Alagoz and Y. Alghawi, "The future of fossil fuels: Challenges and opportunities in a low-carbon," Int. J. Earth Sci. Knowl. Appl., vol. 5, no. 3, pp. 381–388, 2023.

[10] Á. Realpe, J. Garzón, and D. A. Montes, "Mathematical modelling of electric energy potential of Piedra Bolivar Campus of Cartagena University, obtained from solar radiation," *Int. J. ChemTech Res.*, vol. 10, no. 1, pp. 209–214, 2017.

[11] E. W. Likta, J. K. Gadzama, and D. I. Malgwi, "Evaluation of solar energy potentials for power energy generation in Maiduguri, Borno State," *FUDMA J. Sci.*, vol. 7, no. 3, pp. 223–231, Jul. 2023. [Online]. Available: https://doi.org/10.33003/fjs-2023-0703-1863.

[12] S. A. Shehzad, T. Hayat, A. Alsaedi, and B. Chen, "A useful model for solar radiation," *Energy, Ecol. Environ.*, vol. 1, no. 1, pp. 30–38, Feb. 2016. [Online]. Available: https://doi.org/10.1007/s40974-016-0014-y

[13] N. Kannan and D. Vakeesan, "Solar energy for future world: A review," *Renew. Sustain. Energy Rev.*, vol. 62, pp. 1092–1105, Sep. 2016. [Online]. Available: https://doi.org/10.1016/j.rser.2016.05.022

[14] P. Nikolaidis, "Solar energy harnessing technologies towards de-carbonization: A systematic review of processes and systems," *Energies*, vol. 16, no. 17, p. 6153, Aug. 2023. [Online]. Available: https://doi.org/10.3390/en16176153

[15] S. Dixit, "Solar technologies and their implementations: A review," *Mater. Today Proc.*, vol. 28, pp. 2137–2148, 2020. [Online]. Available: https://doi.org/10.1016/j.matpr.2020.04.134

[16] N. K. Sharma, P. K. Tiwari, and Y. R. Sood, "Solar energy in India: Strategies, policies, perspectives and future potential," *Renew. Sustain. Energy Rev.*, vol. 16, no. 1, pp. 933–941, Jan. 2012. [Online]. Available: https://doi.org/10.1016/j.rser.2011.09.014

[17] H. S. K. Nathan, "Solar energy for rural electricity in India: A misplaced emphasis," *Econ. Polit. Wkly.*, vol. 49, no. 50, pp. 60–67, 2014

[18] F. M. Guangul and G. T. Chala, "Solar energy as renewable energy source: SWOT analysis," in *2019 4th MEC International Conference on Big Data and Smart City (ICBDSC)*, IEEE, Jan. 2019, pp. 1–5. [Online]. Available: https://doi.org/10.1109/ICBDSC.2019.8645580

[19] R. V. Moharir and S. S. Dodani, "Solar energy as an energy source: An approach," *IOSR J. Mech. Civ. Eng.*, vol. 14, no. 3, pp. 87–92, Jun. 2017. [Online]. Available: https://doi.org/10.9790/1684-1403058792.

[20] R. Nasrin, M. Hasanuzzaman, and N. A. Rahim, "Effect of high irradiation on photovoltaic power and energy," *Int. J. Energy Res.*, vol. 42, no. 3, pp. 1115–1131, Mar. 2018. [Online]. Available: https://doi.org/10.1002/er.3907

[21] N. U. Putri, E. Pranita, J. P. Sembiring, A. Jayadi, A. Aryanto, and F. K. Muzamil Suat, "The effect of solar radiation on solar panels in aeroponic plant systems," in *2023 International Conference on Networking, Electrical Engineering, Computer Science, and Technology (IConNECT)*, IEEE, Aug. 2023, pp. 121–126. [Online]. Available: https://doi.org/10.1109/IConNECT56593.2023.10327307.

[22] G. Senger, B. Chtirkova, D. Folini, J. Wohland, and M. Wild, "Persistent extreme surface solar radiation and its implications on solar photovoltaics," *Earth's Future*, vol. 12, no. 8, Aug. 2024. [Online]. Available: https://doi.org/10.1029/2023EF004266

[23] A. Manowska, A. Dylong, B. Tkaczyk, and J. Manowski, "Analysis and monitoring of maximum solar potential for energy production optimization using photovoltaic panels," *Energies*, vol. 17, no. 1, p. 72, Dec. 2023. [Online]. Available: https://doi.org/10.3390/en17010072

[24] IPSE, "La energía no es el fin, es el motor de la transformación," Instituto de Planificación y Promoción de Soluciones Energéticas para Zonas No Interconectadas. [Online]. Available: https://ipse.gov.co/blog/2023/10/05/la-energia-no-es-el-fin-es-el-motor-de-la-transformacion-ipse/

[25] R. Prăvălie, C. Patriche, and G. Bandoc, "Spatial assessment of solar energy potential at global scale: A geographical approach," *J. Clean. Prod.*, vol. 209, pp. 692–721, Feb. 2019. [Online]. Available: https://doi.org/10.1016/j.jclepro.2018.10.239

[26] H. Khodayar Sahebi, S. Hoseinzadeh, H. Ghadamian, M. H. Ghasemi, F. Esmaeilion, and D. A. Garcia, "Techno-economic analysis and new design of a photovoltaic power plant by a direct radiation amplification system," *Sustainability*, vol. 13, no. 20, p. 11493, Oct. 2021. [Online]. Available: https://doi.org/10.3390/su132011493

[27] Ü. Ağbulut, A. E. Gürel, and Y. Biçen, "Prediction of daily global solar radiation using different machine learning algorithms: Evaluation and comparison," *Renew. Sustain. Energy Rev.*, vol. 135, p. 110114, Jan. 2021. [Online]. Available: https://doi.org/10.1016/j.rser.2020.110114.

[28] V. Demir and H. Citakoglu, "Forecasting of solar radiation using different machine learning approaches," *Neural Comput. Appl.*, vol. 35, no. 1, pp. 887–906, Jan. 2023. [Online]. Available: https://doi.org/10.1007/s00521-022-07841-x.

[29] L. Wang *et al*., "Prediction of diffuse solar radiation based on multiple variables in China," *Renew. Sustain. Energy Rev.*, vol. 103, pp. 151–216, Apr. 2019. [Online]. Available: https://doi.org/10.1016/j.rser.2018.12.029

[30] Z. Pang, F. Niu, and Z. O'Neill, "Solar radiation prediction using recurrent neural network and artificial neural network: A case study with comparisons," *Renew. Energy*, vol. 156, pp. 279–289, Aug. 2020. [Online]. Available: https://doi.org/10.1016/j.renene.2020.04.042

[31] G. Narvaez *et al*., "An interactive tool for visualization and prediction of solar radiation and photovoltaic generation in Colombia," *Big Earth Data*, vol. 7, no. 3, pp. 904–929, Jul. 2023. [Online]. Available: https://doi.org/10.1080/20964471.2023.2185920

[32] E. Miranda, J. F. G. Fierro, G. Narváez, L. F. Giraldo, and M. Bressan, "Prediction of site-specific solar diffuse horizontal irradiance from two input variables in Colombia," *Heliyon*, vol. 7, no. 12, p. e08602, Dec. 2021. [Online]. Available: https://doi.org/10.1016/j.heliyon.2021.e08602

[33] G. Valencia, J. Duarte, and L. Obregon, "Global solar radiation prediction in Colombia using a backpropagation neural network architecture," *Int. Rev. Autom. Control*, vol. 12, no. 6, p. 293, Nov. 2019. [Online]. Available: https://doi.org/10.15866/ireaco.v12i6.18388

[34] L. C. Ruiz, D. Amaya, and O. L. Ramos, "Prediction of solar radiation through the ANFIS algorithm," *Appl. Mech. Mater.*, vol. 823, pp. 389–395, Jan. 2016. [Online]. Available: https://doi.org/10.4028/www.scientific.net/AMM.823.389

[35] J. Urrego-Ortiz, J. A. Martínez, P. A. Arias, and Á. Jaramillo-Duque, "Assessment and day-ahead forecasting of hourly solar radiation in Medellín, Colombia," *Energies*, vol. 12, no. 22, p. 4402, Nov. 2019. [Online]. Available: https://doi.org/10.3390/en12224402

[36] M. Vanegas Chamorro, E. Villicaña Ortíz, and L. Arrieta Viana, "Cuantificación y caracterización de la radiación solar en el departamento de La Guajira-Colombia mediante el cálculo de transmisibilidad atmosférica," *Prospectiva*, vol. 13, no. 2, p. 54, Dec. 2015. [Online]. Available: https://doi.org/10.15665/rp.v13i2.487

[37] J. A. Hernández Mora, C. L. Trujillo Rodríguez, and W. A. Vallejo Lozada, "Modelamiento de la irradiancia y la temperatura ambiente utilizando funciones de probabilidad," *Rev. Tecnura*, vol. 18, no. 39, p. 128, Dec. 2013. [Online]. Available: https://doi.org/10.14483/udistrital.jour.tecnura.2014.1.a09

[38] E. Chodakowska, J. Nazarko, Ł. Nazarko, and H. S. Rabayah, "Solar radiation forecasting: A systematic meta-review of current methods and emerging trends," *Energies*, vol. 17, no. 13, p. 3156, Jun. 2024. [Online]. Available: https://doi.org/10.3390/en17133156

[39] E. Chodakowska, J. Nazarko, Ł. Nazarko, H. S. Rabayah, R. M. Abendeh, and R. Alawneh, "ARIMA models in solar radiation forecasting in different geographic locations," *Energies*, vol. 16, no. 13, p. 5029, Jun. 2023. [Online]. Available: https://doi.org/10.3390/en16135029

[40] G. Carrascal, J. Barrios, F. Ferreira, and J. Plaza, "Solar radiation forecasting: A study case in the Colombian Caribbean region," *Springer Proc. Math. Stat.*, vol. 469, pp. 199–210, 2025. [Online]. Available: https://doi.org/10.1007/978-3-031-68949-9_15.

[41] J. Nava and P. Hernández, "Optimization of a hybrid methodology (CRISP-DM)," in *Logistics Management and Optimization through Hybrid Artificial Intelligence Systems*, IGI Global, 2012, pp. 356–379. [Online]. Available: https://doi.org/10.4018/978-1-4666-0297-7.ch014

[42] C. Schröer, F. Kruse, and J. M. Gómez, "A systematic literature review on applying CRISP-DM process model," *Procedia Comput. Sci.*, vol. 181, pp. 526–534, 2021. [Online]. Available: https://doi.org/10.1016/j.procs.2021.01.199.

[43] S. Maataoui, G. Bencheikh, and G. Bencheikh, "Predictive maintenance in the industrial sector: A CRISP-DM approach for developing accurate machine failure prediction models," in *2023 Fifth International Conference on Advances in Computational Tools for Engineering Applications (ACTEA)*, IEEE, Jul. 2023, pp. 223–227. [Online]. Available: https://doi.org/10.1109/ACTEA58025.2023.10193983

[44] F. Martinez-Plumed *et al.*, "CRISP-DM twenty years later: From data mining processes to data science trajectories," *IEEE Trans. Knowl. Data Eng.*, vol. 33, no. 8, pp. 3048–3061, Aug. 2021. [Online]. Available: https://doi.org/10.1109/TKDE.2019.2962680

[45] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, Time Series Analysis: Forecasting and Control, 5th ed. Wiley, 2015

[46] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*. OTexts, 2013.

[47] Y. Patil, T. Shruti, and R. K., "Time-series forecasting using ARIMA and SARIMA models for solar NASA POWER data," in *2025 3rd International Conference on Intelligent Systems, Advanced Computing and Communication (ISACC)*, IEEE, Feb. 2025, pp. 946–952. [Online]. Available: https://doi.org/10.1109/ISACC65211.2025.10969243

[48] D. K. Wijaya, R. Widjaya, A. Kurniawan, and Diana, "Using time series analysis to predict earthquake," in *2024 International Conference on Artificial Intelligence, Blockchain, Cloud Computing, and Data Analytics (ICoABCD)*, IEEE, Aug. 2024, pp. 25–30. [Online]. Available: https://doi.org/10.1109/ICoABCD63526.2024.10704296

[49] E. S. Gardner, "Exponential smoothing: The state of the art," *J. Forecast.*, vol. 4, no. 1, pp. 1–28, Jan. 1985. [Online]. Available: https://doi.org/10.1002/for.3980040103.

[50] S. Dev, T. AlSkaif, M. Hossari, R. Godina, A. Louwen, and W. van Sark, "Solar irradiance forecasting using triple exponential smoothing," in *2018 International Conference on Smart Energy Systems and Technologies (SEST)*, IEEE, Sep. 2018, pp. 1–6. [Online]. Available: https://doi.org/10.1109/SEST.2018.8495816