Probabilistic Irradiation Forecast for Photovoltaic Cells

A project highlighting the application of ARIMA Model in the forecasting of solar irradiance of the photovoltaic cells for the MA-202 course

Done By:

Part	Name	Roll No.
1-2	Jinious Sheth	21110087
3	Yash Bothra	21110242
4	Hirva Patel	21110154
5	Khushal Ramani	21110104
6	Kumar Harsh	21110109

Index

Introduction

A photovoltaic cell, mostly known as a solar cell, is a device that converts light energy into electricity. It consists semiconductor materials such as silicon, which absorbs photons and releases electrons, resulting in an electric current flow. This current flow causes the generation of electricity which is stored. It has been in more use in modern times when we need different and sustainable sources of energy.

As light strikes a photovoltaic cell, electrons in the semiconductor material are excited to a higher energy level, allowing them to move freely. The electric field created by the cell's multiple components separates positively charged holes from negatively charged electrons. Electrons are then sent through an external circuit, creating an electrical current. Solar panels with photovoltaic cells are widely used to produce electricity for residential and commercial usage. They are also found in tiny devices such as calculators and watches, as well as bigger facilities such as power plants. Photovoltaic cells are gaining popularity as a sustainable and clean source of energy as a way to reduce dependency on fossil fuels and combat climate change.

It is important for us to understand in order to predict the external factors affecting the power produced by the photovoltaic cells in solar panel. One of the most important factors is Solar Irradiation. It is important to analyse the amount of irradiance in a particular area when setting up the solar panels. For IITGN campus, we have analysed the irradiance data of the solar panels fitted along the solar carport. Solar irradiance is the quantity of solar energy radiated from a source per unit area, which is commonly measured in watts per square metre (W/m2). Solar irradiance is a measure of the intensity of solar radiation, which is the transfer of thermal energy in the form of electromagnetic waves.

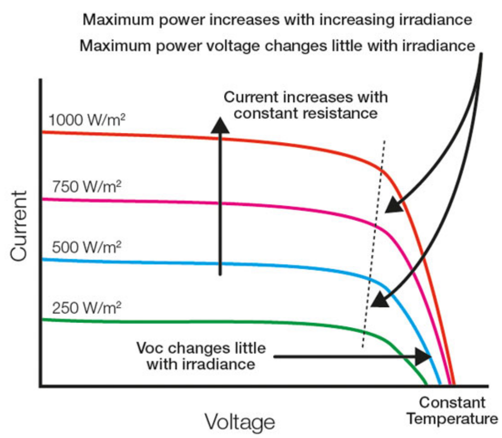
Solar irradiance can affect the efficiency and performance of a solar panel. As a solar panel becomes too hot, the resistance of the materials rises, resulting in a decrease in efficiency and power production. Additionally, extreme heat can destroy solar cells and limit their lifespan.

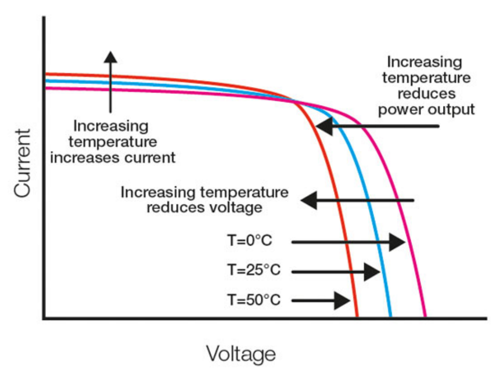
To further understand the statistics, we used Arima Model. Arima Model is a statistical method for modelling time series data that employs three components: Autoregression, Integrated and Moving Average. Time series data is a collection of observations over time, and Arima Model method allows us to incorporate previous knowledge or beliefs about the data into our analysis. We decided to use the data collected from one month(May) of year 2021 of a time gap of 15 minutes to for calculations and plotting of graphs.

Solar Irradiance

To maximise the efficiency of a solar panel, evaluate the amount of solar irradiance it receives and design the panel and its mounting mechanism to dissipate surplus heat. This can be accomplished by strategies like as ventilation, insulation, and the use of reflective materials to limit the amount of heat absorbed by the panel. It is also critical to set the panel in a particular way so that it receives the maximum direct sunlight while receiving the least amount of heat. When PV cells are exposed to high temperatures for extended periods of time, they might suffer irreparable damage, lowering the cell's lifespan. This is why, in order to ensure optimal performance and longevity, PV cells should be kept at a reasonable temperature.

The open circuit voltage of a PV module varies with cell temperature. The open circuit voltage (Voc) drops as temperature rises owing to external changes or heat generated by internal power dissipation during energy production. This, in turn, reduces the power output. The PV module temperature coefficient must be considered when designing a solar PV system, comparing the projected average cell temperature in its operational environment to the STC data used to calculate module output. Similarly, irradiance will affect module performance, with a decrease in sunshine largely resulting in a decrease in current and, as a result, a decrease in power production[1].





Forecasting Time Series

Probability theories are fundamental in modelling the many forms of time series in time series forecasting. Time series data are modelled and analysed by using probability concepts like that of random variables, probability distributions. Probabilistic forecasts that take uncertainty into account are then created.

A progressive change in the mean or amount of the data through time characterises one of the most common forms of time series. One standard approach for modelling trend time series is to use a linear regression model where the time variable is used as a forecast of the response variable. In this case, the trend is modelled as a linear variation in time, and unrelated residuals are taken for granted.[2]

A key component of time-series analysis is seasonality. Seasonal changes might affect time-series since they are indexed forward in time. For instance, we anticipate that ice cream sales will be higher in the summer and lower in the winter.

Seasonality can occur throughout a range of time frames, including days, weeks, and months. Understanding how seasonality impacts our series is crucial for time-series analysis since it enables us to make better future projections. Cyclical time series show patterns that may not have a definite length or regularity and occur over a longer time period than seasonal trends. A cyclical time series is represented in Fourier analysis as a collection of sine and cosine waves with various frequencies. The series can be modelled and predictions about future values can be produced by varying the amplitudes and frequency of these waves.

Random or unpredictable variations that show no obvious pattern or trend characterise irregular time series. One popular method for modelling irregular time series is to employ a stochastic process model, in which the time series is represented as a string of random variables produced by a probabilistic mechanism. The autoregressive integrated moving average (ARIMA) model, a linear time series model that includes both autoregressive and moving average components, is the stochastic process model that is most frequently used for irregular time series.

Some time series models used for forecasting include:

1. Autoregressive Integrated Moving Average model: By taking into account the data's autocorrelation and partial autocorrelation, this model is used to forecast stationary time series data.

2. State space models: An adaptable framework known as a state space model is used to represent time series data as a collection of unobserved states that change over time.

3. Seasonal Autoregressive Integrated Moving Average model: The seasonality of the data is taken into account in this model, which is an extension of the ARIMA model.

4. Vector Autoregression model: This model forecasts numerous time series variables by taking into account their interactions and interdependencies.

5. Dynamic regression model: A dynamic regression model is a linear regression model that uses other independent variables as predictors and lagged values of the dependent variable as its input.

6. Bayesian model: Using Bayesian methods to estimate the model's parameters, Bayesian time series forecasting is an alternate way for time series modelling.

7. Long Short-Term Memory model: By taking into account long-term dependencies and non-linear correlations between the past and future values, this model, a subtype of recurrent neural network, forecasts time series data.

These models can help analysts and decision-makers make informed decisions and plan for the future by predicting the behavior of a variable over time.[3]

Autoregressive Integrated Moving Average (ARIMA) model

The time series model known as ARIMA (AutoRegressive Integrated Moving Average) is frequently used for forecasting. It combines the moving average (MA) model with the autoregressive (AR) model, two less complex models. The ARIMA model is used to record historical forecasting errors as well as linear correlations between a time series' present value and its previous values. P stands for the order of the autoregressive component in the ARIMA model, while d stands for the amount of differencing necessary to make the series stationary. q denotes the order of the moving-average component in the ARIMA model. Time series data can be fitted with ARIMA models utilising statistical tools.

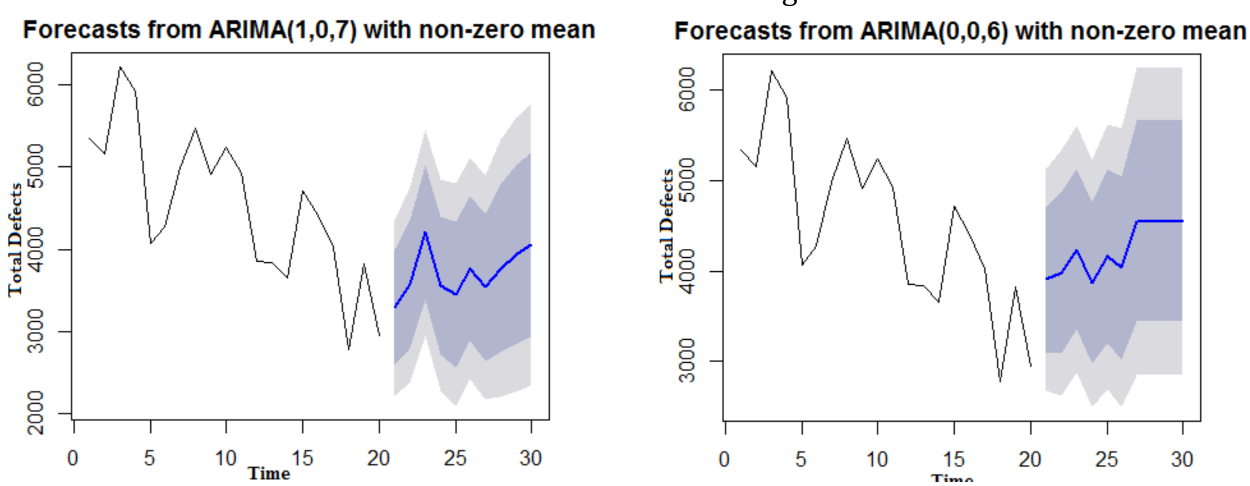


Figure 1[4]

Seasonal Autoregressive Integrated Moving Average (SARIMA) model

SARIMA model is advancement of ARIMA model where the value of the data set depends on the value of it at the same time around last year. Suppose, the price of a particular type of fish depends on the season and availability. Thus, it is safe to assume that the price of the fish will be same as the same month from last year.

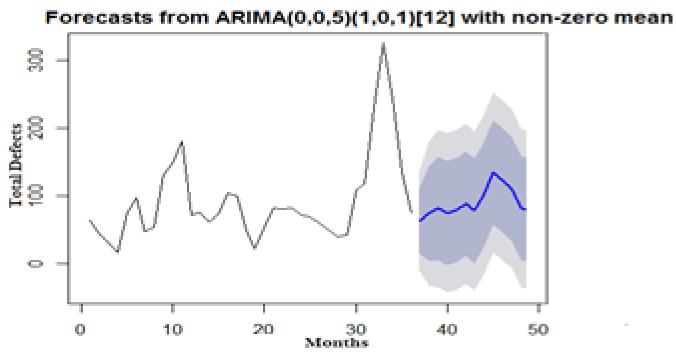


Figure 2[4]

Bayesian Time Series model

Bayesian time series forecasting used the Bayes theorem to predict the future value. It is said to be a stronger model than classical models because of its flexibility. Probabilistic predictions of future values of the time series are made using the posterior distribution of the parameters.

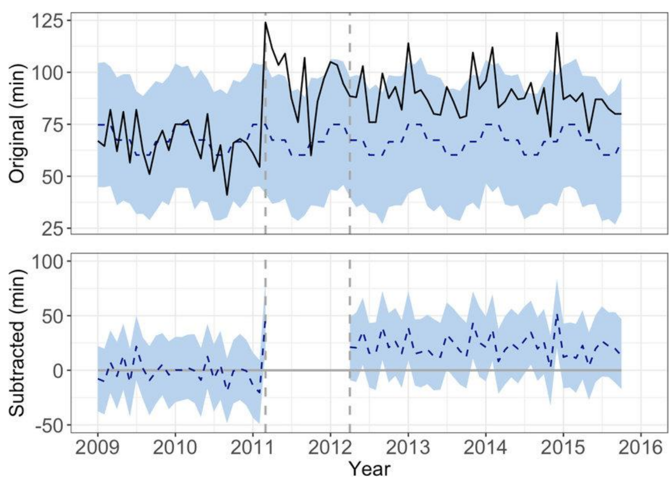


Figure 3[5]

Vector Autoregression (VAR) model

To study the relationship between several time series variables, a statistical model known as vector autoregression (VAR) is used. It is an upgrade of the univariate autoregression (AR) model, which considers takes into account a variable's dependence on its previous values.

Dynamic regression model

A subset of time series models called dynamic regression models permits the forecast to take into account extraneous factors. These models are helpful when variables outside of the time series themselves have an impact on how the time series behaves. The link between the external variables and the time series is measured by a set of regression coefficients in the model. To account for autocorrelation in the residuals, ARIMA components can be added to dynamic regression models.

State space model

State space models is another type of time series model that allows the inclusion of both observable and unobservable time series components. The model consists of two parts: an observation equation that links the unobservable and observable data and a state equation describes how the unobservable data change over time. State space models make it easier to model complex time series with multiple elements in account.

Thus, there are variety of time series forecasting methods, each with their own advantages and disadvantages. The features, objectives, and the resources available all influence the method that is chosen. The many methods for time series forecasting include ARIMA, Bayesian, dynamic regression, exponential smoothing, state space models, etc.

Why ARIMA Model?

The AR (autoregressive model) forecasts the future or current value based on the past values of the data. It takes the PACF (Partial Auto Correlation Function) of each of the previous data sets and looks at the significance of each on the current data set. The current data set is predicted based on some of the previous lags only if it has significant amount of effect on the current data set.

The AR (1) model looks like:

$$\boldsymbol{X}_t = \phi_1 \boldsymbol{X}_{t-1} + \varepsilon_t$$

where $\boldsymbol{X}_t$ is the current data set that we want to predict, ϕ_1 is a constant that depends on the significance of the previous data set on the current, $\boldsymbol{X}_{t-1}$ is the previous data set and ε_t is the error in prediction. This model concentrates on how the past data can influence the current data, and thus, has constant factors related to each past data that can cause a significant change in the current.

MA (Moving Average) model is conceptually a linear regression of the current value of the series against current and previous (observed) white noise error terms or random shocks. The random shocks at each point are assumed to be mutually independent and to come from the same distribution, typically a normal distribution, with location at zero and constant scale.[6]
The actual value of the current data will be in an MA(1) model will look like:

$$\boldsymbol{X}_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

and the estimated value for current data will be:

$$\boldsymbol{X}'_t = \mu + \theta_1 \varepsilon_{t-1}$$

where $\boldsymbol{X}_t$ is the actual current data set, θ_1 is a constant, ε_t is the error in current prediction, and ε_{t-1} is the error in previous prediction. This model concentrates on the error committed in calculating each term and correcting it the next time.

ARMA (AutoRegressive Moving Average) model applies the methods of both AR and MA model. However, it is assumed that the time-series is stationary, i.e. it has a constant mean, and it fluctuates uniformly. This model will predict on the basis of both: on the previous data sets collected and also the errors committed for each calculation. Thus, this proves to be a stronger model than AR and MA implemented individually.

The value of the current data will look like:

$$X_t = \varepsilon_t + \sum_{i=1}^{p} \varphi_i X_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}.$$

Considering ARMA (1,1) model we get:

$$\boldsymbol{X}_t = \varepsilon_t + \phi_1 \boldsymbol{X}_{t-1} + \theta_1 \varepsilon_{t-1}$$

Given a time series of data X, the ARMA model is a tool for predicting future values in a time series. The AR part involves regressing the variable on its own lags. The MA part involves including the error term as a linear combination of error terms occurring previously.[7]

AR and MA are classical models of forecasting time series, but there was certainly a need for another. We cannot always assume that the time series is stationary. Thus, ARIMA model was developed that takes the difference of two consecutive data points, and applies ARMA model on it. This new data set is termed $\boldsymbol{Z}_t$.

$$\boldsymbol{Z}_t = \boldsymbol{X}_t - \boldsymbol{X}_{t-1}$$

We can adjust the number of differences that we take depending on the kind of data we are dealing with. For example, we can apply ARMA model on the difference of differences, which will make it 2^{nd} order of differences. Thus the ARMA model has three variables: p, d and q, where p and q come from AR and MA models respectively and d is the number of differences taken.
Thus an ARIMA (1,1,1) model will look like:

$$\boldsymbol{Z}_t = \varepsilon_t + \phi_1 \boldsymbol{Z}_{t-1} + \theta_1 \varepsilon_{t-1}$$

where $\boldsymbol{Z}_t$ is the first order difference.

This adds an extra benefit to the forecasting by accounting for the change in mean in upcoming intervals of time. For example, in stock market, where the share prices increase in long period of time, assuming a constant mean won't give us accurate predicitons.

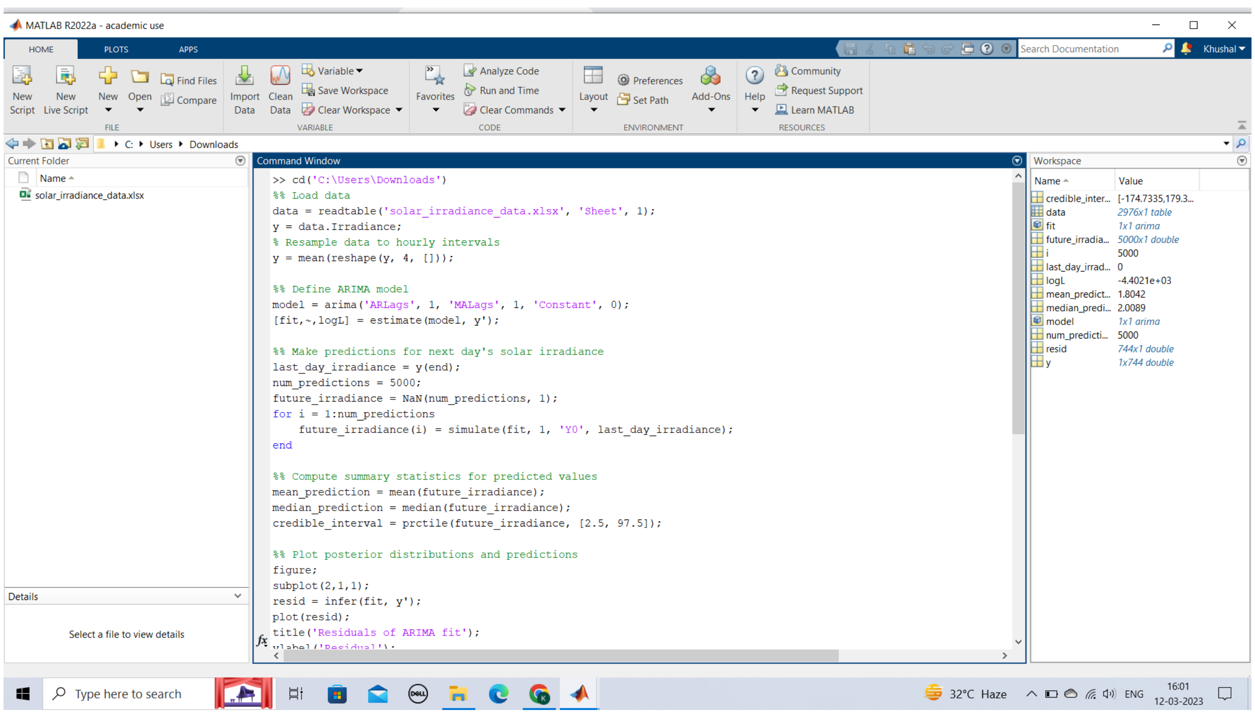
However, as our data points is irradiance of the sun, it remains more or less same around same time each day. We can see similar patterns in irradiance on yearly basis, i.e. the value of irradiance is almost same as last year at that same time. This means that our data is stationary. Therefore, we have used the function of ARIMA model in MATLAB considering the constant = 0. That means, our data has constant mean and the value of mean does not change in time.

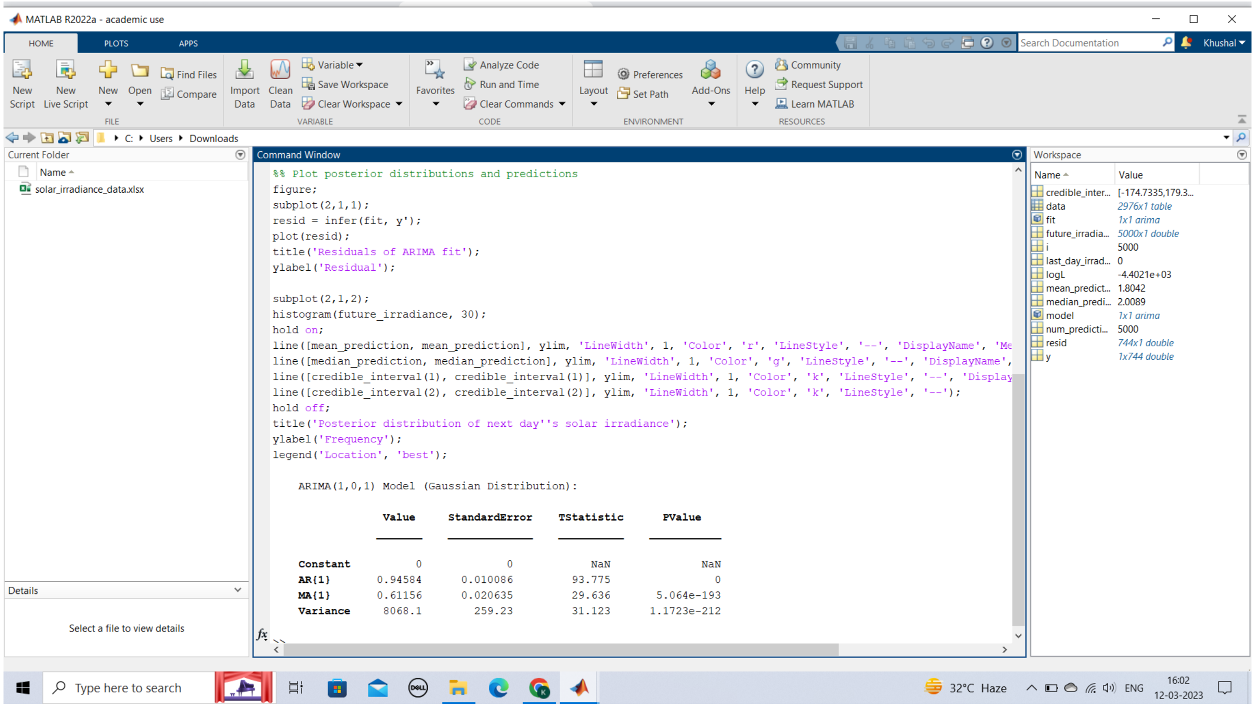
Calculation of Solar Irradiance And The

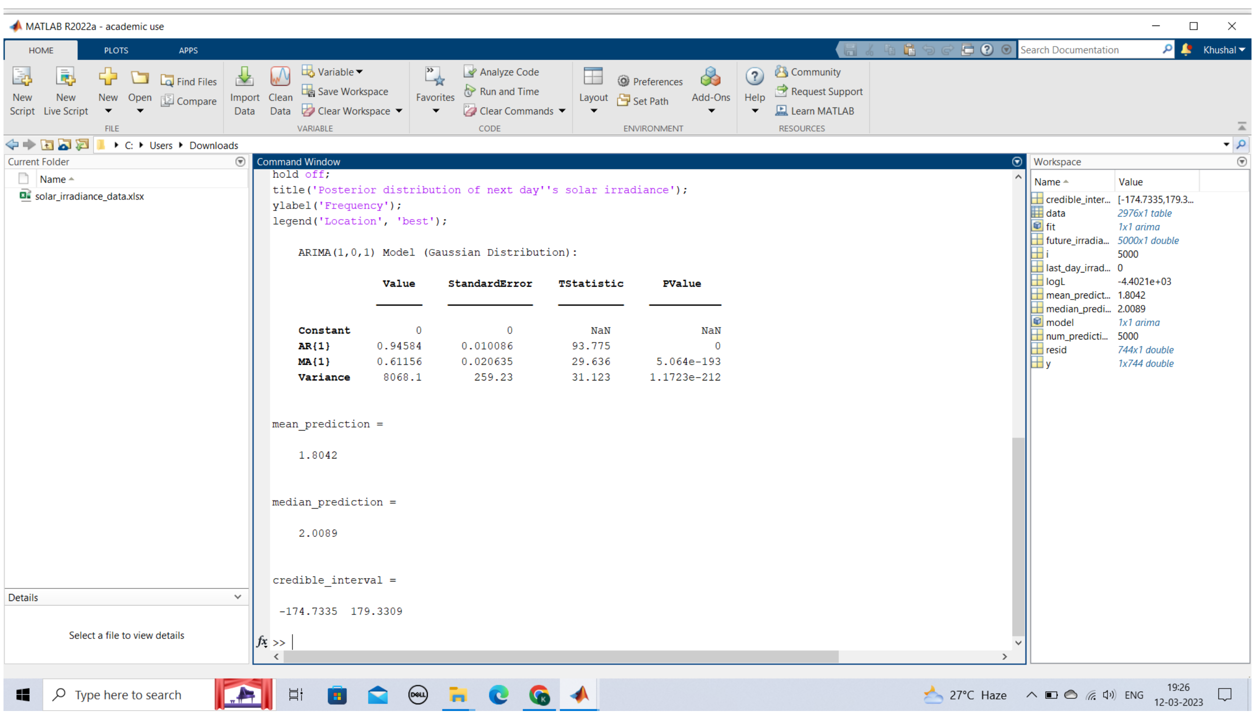
Parameters Involved

Here is a MATLAB code for calculating the solar irradiance on the next day of the next month. We have taken the solar irradiance values of IITGN's solar carport during May, 2021 into account. The data points are in a 15 minutes intervals. The data in this sheet is taken from the sheet that was provided by Prof. Naran Pindoriya. Here is the link to the excel sheet whose data is taken into account.

solar_irradiance_data

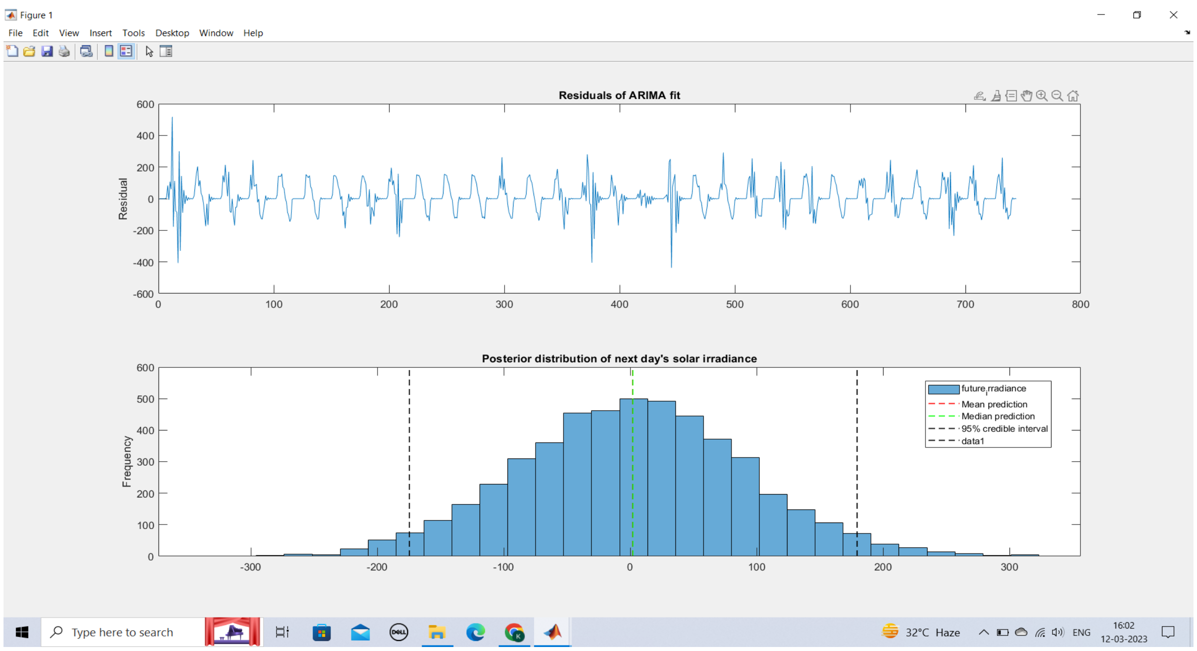






This code in MATLAB reads the solar irradiance data of IITGN's solar carports from a column named "Irradiance" of an Excel file named"solar_irradiance_data.xlsx". The solar irradiation for the following day is then forecasted using the ARIMA model. Then we generated two plots. One is for the ARIMA fit's residuals which implies that the model is fit for the given data.

The other one shows the mean, median, and 95% confidence interval. The distribution of blue bars shows the range of future values of solar irradiance and the confidence interval tells where the predicted values are likely to fall.



Applications

For PV facilities to effectively participate in the energy market and organize their resources, accurate solar energy forecasting is crucial. The research has documented several techniques for predicting PV energy. On the other hand, we'll discuss statistical methods for time series forecasting using observed past data. (ARIMA).

The ARIMA (Autoregressive Integrated Moving Average) model establishes an outstanding short-term sun radiation projection technique. The predictive analysis tool used to create the forecast, IBM SPSS, offers various statistical techniques, including geospatial analysis, Monte Carlo simulations, and linear regressions.

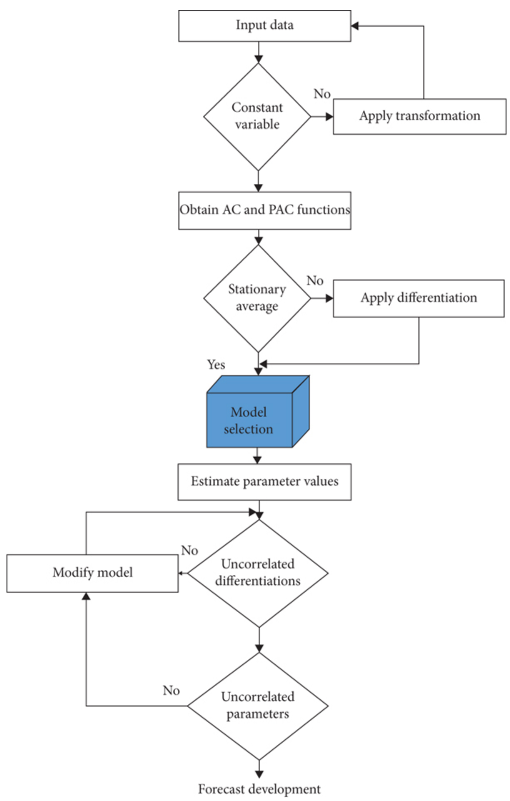


Fig. 1

Fig1. Operation of the ARIMA model for the energy production forecast of a photovoltaic system; AC and PAC autocorrelation functions are used.

The ARIMA (1,0,0) or (1,1,0) type model was selected for solar radiation prediction in this part of application. The Akaike Informational Criterion led to this decision. (AIC). The AIC is used to assess the model depending on the inputted data. The goal is to ascertain the quality of the statistical model on a dataset and determine whether it is more effective than other models. The information used to identify the processes that might produce the data is estimated by the ARIMA model.

The Akaike statistical criterion for the ARIMA model can be calculated based on the following relation:

$$\mathrm{AIC} = 2k - 2\ln(L)$$

L represents the maximum value given by an estimation function (MLE (Maximum Likelihood Estimation)) and the number of estimated parameters. The minimum value of the AIC result defines that the corresponding model is the most efficient.

The SPSS program automatically calculates the AIC, a model-fitting parameter.

If the forecast process contains seasonal fluctuations, as in this case, the process becomes SARIMA (p,d,q) (P, D, Q), where p is the order of the AR process, d represents the differentiation term, q is the order of the moving average, P represents the order of AR seasonal processes, Q represents the MA order, D is the order of seasonal differentiation, and s represents the length of the seasonal period.[9]

Conclusion

We learned many things during this project and concluded that the ARIMA model could provide valuable insights into forecasting photovoltaic (PV) cell solar irradiance and help optimize the performance and maintenance of the PV system.

We analyzed the historical time series data and identified the trend and seasonality of the solar irradiance. We analyzed the data of IIT Gandhinagar itself and came to know about various things about it. On analyzing and calculating the probability, we learned how accurate the forecasting is, and there is a potential impact of many external factors like weather patterns and maintenance of the PV system may affect the accuracy of the forecast.

References

[1].*How does temperature and irradiance affect I-V curves?* (n.d.). Seaward. Retrieved March 11, 2023, from https://www.seaward.com/gb/support/solar/faqs/00797-how-does-temperature-and-irradiance-affect-i-v-curves/

[2]"Time Series Forecasting Methods, Techniques & Models | InfluxData," *InfluxData*, Jan. 03, 2023. https://www.influxdata.com/time-series-forecasting-methods/

[3]"Time Series Forecasting: Definition, Applications, and Examples," *Tableau*. https://www.tableau.com/learn/articles/time-series-forecasting

[4]"Time Series-based Forecasting using ARIMA Models," *Time Series-based Forecasting using ARIMA Models | HCLTech*, Apr. 01, 2015. https://www.hcltech.com/blogs/time-series-based-forecasting-using-arima-models

[5]https://www.researchgate.net/publication/330991913/figure/fig2/AS:724564287975424@1549760996180/A-Bayesian-time-series-analysis-using-monthly-median-total-emergency-medical-services.jpg

[6]"Moving-average model - Wikipedia," *Moving-average model - Wikipedia*, Feb. 27, 2023. https://en.wikipedia.org/wiki/Moving-average_model#:~:text=In%20time%20series%20analysis%2C%20the,identical%20to%20itself%20random%2Dvariable

[7]"Autoregressive–moving-average model - Wikipedia," *Autoregressive–moving-average model - Wikipedia*, Aug. 01, 2010. https://en.wikipedia.org/wiki/Autoregressive%E2%80%93moving-average_model

[8] 1-month data of forecasting of solar irradiance of the photovoltaic cells provided by Prof. Naran Pindoriya.
15 Minute.xlsx

[9] https://www.hindawi.com/journals/ijp/2021/6777488/