Print ISSN : 0022-2755
Journal of Mines, Metals and Fuels
Contents available at: www.informaticsjournals.com/index.php/jmmf

Statistical Analysis of PV Cell Power Generation and Influence of Weather on Power Generation

Vinay Mavuthanahalli[1], Channabasavegowda[1] and Hamritha Shankar[2*]

[1]Student, M.Tech, Department of Industrial Engineering and Management M S Ramaiah Institute of Technology, Bangalore 560054, India.
*[2]Assistant Professor, Department of Industrial Engineering and Management M S Ramaiah Institute of Technology, Bangalore 560054, India. *E-mail: hamritha.ap@gmail.com*

Abstract

The need for renewable energy is growing as a result of climate change, global warming and the subsequent increase in natural disasters brought on by the use of carbon-emitting resources. An earlier examination of potential future energy pathways demonstrates that it is theoretically feasible to simultaneously improve energy security, air quality, and access while preventing disastrous climate change. Land, energy, and water are some of our most precious resources, but how and how much we use them also affects the climate. At the national level, renewable energy prospects would be very diverse. The rise in global economic activities results in higher demand in electricity, as its one of the main sources of energy. This leads to search for renewable energy sources. Solar cells are one of the technological innovations that directly convert light energy into electricity through the photovoltaic effect, creating electrical charges that are free to travel through semiconductors. But due to uncertainty in the weather and the influence of weather on power generation makes the integration of PV cell into existing grid a difficult part. To operate the electricity distribution system efficiently one should know both the demand and supply of distribution center. To solve this issue, a statistical model that forecasts the power generation at the PV cell plant and aids in grid operation, as well as information about the impact of weather on power generation, is required. In addition to using traditional time series models fundamental machine learning methods like Linear Regression, Random Forest Model, XG Boost, and Support Vector Regression (SVR) models are also trained and applied for prediction.

Keywords : *Time Series, ML, XGBoost, Optimization, Blackout.*

1.0 Introduction

Climate change is changed in temperatures and weather patterns. It can be natural due to solar cycle, but from 1800's, human activities are driver of climate change, due to burning of fossil fuels and other modernization activities. Air pollution accumulates sunlight and solar radiation that have bounced off the earth's surface. Normally this needs to go into space, but pollution traps the heat and makes the planet hotter. These heat-trapping pollutants are known as greenhouse gases, and their impact is called the greenhouse effect.

The International Energy Agency (IEA) estimates that "there will be over 1700 GW of solar power capacity installed globally by 2030 [2] and solar power should be the world's largest generator of electricity to achieve Net Zero by 2050". From this information it can be inferred that the whole world is looking towards the renewable energy sources.

One of such renewable energy is solar power. It's easily installable and maintainable in all the possible renewable energy sources. It easily merges with existing infrastructure

*Corresponding Author

with the bare minimum alteration in the power system. From Fig.1. it's observed that in India there is an exponential growth in the installation of PV power cell and the use of solar energy, it has also been in the peak in recent years.

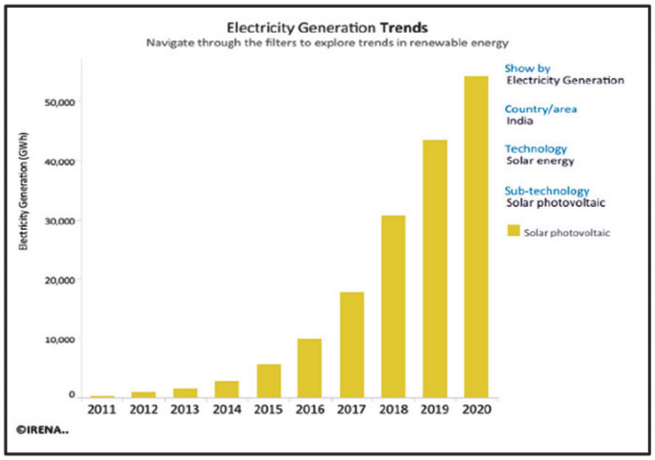

Figure 1: PV Power Output from 2011 to 2020 in India (Source: IRENA)

2.0 Literature Review

Worldwide, it is acknowledged that energy is the driving force behind economic growth. There are three major categories of energy resources available worldwide: fossil, nuclear and renewable. The literature on solar panels and renewable energy seems to be growing and expanding annually. The renewable energy generation and utilization is greatly affected by the environmental factors. The generated energy must be stored and utilized in the best possible ways. Gopi et al.[1] demonstrated that performance suffers during rainy seasons, which are frequent in the humid tropical region. In contrast, Kim et al[5] discovered that weather-related variables have a significant impact on the power generation by a solar power plant in Samcheonpo, Korea. They created many regression models and discovered that the intensity of insolation during daylight hours and daylight time had the strongest link with the solar electricity generated by the plant. They also discovered that the amount of evaporation affects solar power generation. Kumar et al. investigated the performance of PV plants built over water bodies and observed that it is slightly lower than that of land-based plants[6,7]. Maitanova et al. studied that by increasing the size of the training set, one can lower the seasonal mean absolute scaled error of prediction[8].

From above literature review it is inferred that with the big dataset one can predict the value accurately as well as reliably. Basic time series models give good result with dataset with less noise. Weather parameters have high influence on the solar power generation, to have better control on generation of solar power, the distribution centre must know the weather parameters that influences the power generation.

3.0 Research Problem

When the electrical power network supply to an end user is lost for number of reasons, it is referred as power blackout. Blackouts caused by or arising from energy station tripping out are particularly challenging to recover quickly. The results of global warming and the exponential growth of the economy have been the primary causes of blackouts. The increase in power consumption in some urban areas brought on by an increase in crypto mining is one recent phenomenon that stands out. The distribution center must have accurate demand and supply forecasts in order to meet the aforementioned objectives. Due to the influence of parameters that affect power production, there is considerable output variability in solar power generation.

4.0 Research Objectives

Aim of the project is,

1. Determining the relationship between meteorological variables and how they affect the production of electricity.
2. To develop the best model possible for predicting power generation, evaluating itagainst the available options, and facilitating efficient power distribution.
3. Project the next period's electricity production.

5.0 Methodology

The data on electricity generation show both trend and seasonality. When there is a trend or seasonality in the data that has to be analysed, Time Series Models are employed. For time series data, the following models are used:

(a) Auto-regressive (AR)
(b) Moving Average (MA)
(c) Auto-regressive Moving Average (ARMA)
(d) Auto-regressive Integrated Moving Average (ARIMA)
(e) Seasonal Auto-regressive Integrated Moving-Average (SARIMA)

According to the literature review, machine learning models perform quite well most of the time because they have a feedback loop also can create good prediction models. Every model is unique and performs in unique way.

The following basic machine learning algorithms have been tested:

(a) Random Forest
(b) Linear Regression

(c) XGBoost
(d) Support Vector Machine (SVM)

Python 3 has been used as a tool to make the models and version is 3.9.7 of Python 3. Jupyter Note Book is used to write the code.

5.1 Procedure Used

Step 1: Obtain the solar power generation and weather Dataset for pre-processing.
Step 2: Data Preparation and Visualization.
Step 3: Selecting TimeSeries Models and Machine Learning Models.
Step 4: Create the Classifier Model.
Step 5: Put the Classifier Model to the Test.
Step 6: Evaluate the Experimental Results of Each Classifier's Performance.
Step 7: Recommending the Best Prediction Algorithm.

Documentation is the eighth step.

5.2 Dataset and Parameters

Two Kaggle datasets are chosen for the investigation. One dataset includes the solar modules total solar power output, which has maximum capacity of 5kWh, andvariables are solar power produced in cumulation in kWh, consumed electricity in kWh and gas in m^3 volume withresolution of one day.

Weather information was obtained from the Time and Date website from 2012 to 2019 as shown in Table 1. This dataset contains weather status characteristics such as wind, humidity, air pressure, and temperature in Antwerp, Belgium. [9,10]

6.0 Results and Discussion

6.1 Dickey-Fuller Test

David Dickey and Wayne Fuller devised the Dickey-Fuller test, a statistical test that is frequently used to examine the stationarity of time series data. They made their proposal in 1979[3,4]. It claims that the mean and standard deviation are unaffected by changes in the datapoints[11,12]. The hypothesis tests state the below assumptions:

Null hypothesis (H_0): Data is not stationary.

Alternative hypothesis (H_1): Data is stationarity. Dickey Fuller's test yielded a "stationary" result and rejects the null hypothesis. Therefore, tests can be carried on because statistical parameters do not change over the passage of time.

6.2 Influence of Weather on Power Generation

From the correlation matrix as shown in Fig.2, parameters like wind and pressure are less influential in the power produced by the PV cell because they have weak correlations with the values of -0.22 and 0.18, respectively.

Temperature have positive correlation of 0.63 on the power generation, as the temperature increases there is increase in the solar power output, they are directly proportional to each other. Power generation and humidity are

Figure 2: Correlation Matrix

Table 1: Dataset preview (Source: Kaggle)

Temp	12.33	7.79	8.21	7.35
Weather	Overcast	Scattered clouds	Light rain. Fog	Broken clouds
Wind	21.33	17.21	32.35	26.38
Humidity	89.33	81.43	78.23	71.04
Barometer	1007.87	1009.17	1007.92	1012.69
Date 2012	01-Jan	02-Jan	03-Jan	04-Jan
Day power	0.8	2.9	0.8	2.7
Current/day	16.5	19.6	18.4	20.1

Table 2: Electricity (%) Generated on Seasons and Output Productivity

Season	summer	fall	winter	Spring	Total
Days	496	488	1208	728	2920
% Day	17	17	41	25	
Power Produced kW	8836.9	4671	5572	13047	32127
% Power	28	15	17	41	
O/p	1.62	0.87	0.42	1.63	

inversely related, with a negative correlation of 0.75 as the humidity increases there is decrease in the solar power output.

Seasons like summer and spring have shown to be more productive at the solar power plant, with 1.62 and 1.63 output productivity per day respectively. Winter has the lowest output productivity at 0.42 as shown in Table 2.

6.3 Statistical Models for Power Generation Forecasting

6.3.1 Time Series Models for Power Generation Forecasting

A time series is a sequential set of data points, which are of successive time's period. It is defined as a set x(t), t = 0, 1, 2... Where 't' is time elapsed. The variable) x(t) is random variable function.

A time series in generally have three components they are trend, seasonality, error (residual) (Fig.3).

Models Performance Evaluation done using:

(a) Coefficient of Determination (R^2)

(b) Mean Squared Error (MSE)

(c) Root Mean Square Error (RMSE)

As seen from Table 2, all the time series models R^2 prediction values are in negative range, and near to zero. From this it's evident that this model's performance in the prediction is not satisfactory.

Table 3: Time Series PG Models Performance Evaluation

Prediction model	MSE	RMSE	R^2
AR model	58.25	7.63	-0.03
MA model	57.67	7.59	-0.02
ARMA model	64.55	8.03	-0.14
ARiMA model	64.55	8.03	-0.14
SARiMA model	64.55	8.03	-0.14

The results obtained from Table 3; it can be concluded that the time series model falls short of expectations although power generation data is confirmed to be periodic. The moving average has a tough time detecting minor spikes and drops between periods. The time series models consequently perform poorly as seen in result. This leads to the further study on the performance of machine learning models.

6.3.2 Machine Learning Models for Power Generation Forecasting

Machine Learning (ML) is a branch of computer science and a subfield of artificial intelligence techniques. Machine learning uses historical data and elements to establish a relationship between input and output, even if the relationship is complex. As a result, it is advised to use appropriate information or prepare them effectively. The models chosen are:

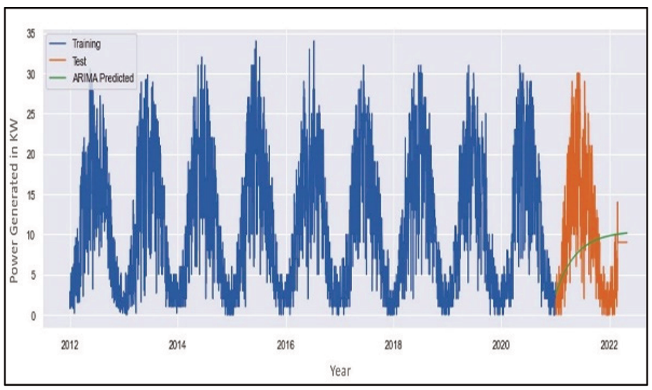

Figure 3: ARIMA Model PG Prediction Plot

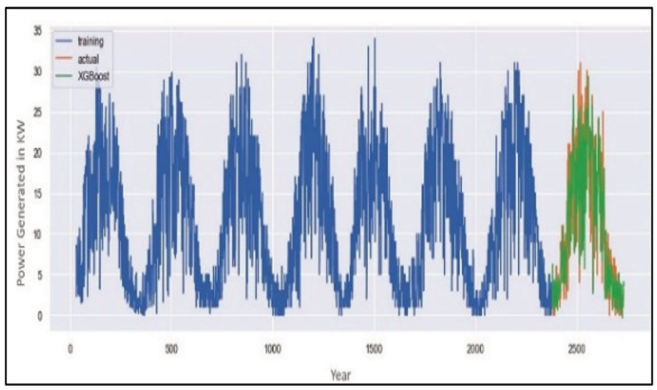

Figure 4: XGBoost Model PG Prediction Plot

1. Linear Regression (LR)
2. Random Forest (RF)
3. XG Boost (XG Boost)
4. Support Vector Machine (SVM).

As seen in Table 3, both models SVM and XG boost performed very well. Even LR model and RF models are also nearer to the best results. From the literature review it can be said that the XG Boost Fig.4 should be chosen, due to fact that it will perform better even with the large datasets (Table 4).

Table 4: ML Models Performance Evaluation

Prediction Model	MSE	RMSE	R^2
Linear Regression Model	15.875	3.984	0.765
Random Forest Model	13.632	3.692	0.798
SVM Model	13.215	3.635	0.805
XG Boost Model	13.179	3.63	0.805

Table 5: Power Generation Prediction Sheet

Lag	Actual Power	Predicted Power	Diff.
2380	1	6.27	5.27
2381	3	2	-1
2382	1	1.9	0.9
2383	0	1.22	1.22
2384	0	2.52	2.52
2385	1	1.66	0.66
2386	1	1.15	0.15
2387	4	3.81	-0.19
2388	2	4.33	2.33
2389	2	0.69	-1.31
2390	0	1.16	1.16
2391	2	0.56	-1.44
2392	2	4.04	2.04

7.0 Conclusions

The results of this research reveal the following understandings:

A. Wind and pressure to be less influential variables in the power generated by the PV cell and temperature to be highly influential parameter. Power generation and humidity have a negative correlation and are inversely related.

B. Summer and spring seasons are productive seasons at the solar power plant, with 1.62 and 1.63 output productivity per day, respectively. Winter has the lowest output productivity of 0.42.

C. Time series model falls short even when power generation data is verified to be periodic, due to a brief increase and a period of decline.

D. Machine learning models like SVM and XG Boost performed exceptionally well, with prediction accuracy of 80.5 %. XGBoost has been determined to be the most accurate prediction model for power generation as studies can be carried in this model even when the data set is large.

8.0 Limitation and Scope for Future Work

The works listed below can be used for the research's next stage, as it is observed from the outcomes and conclusions. More research in the area of hyper parametric optimization techniques, such as Manual Search, Randomized Search, Halving Grid Search, Halving Randomized Search, HyperOptSklearn, and Bayes Search, can improve the outcomes of machine learning models.

To enhance the effectiveness of machine learning models, work can be carried on Deep Learning techniques such as Classic Neural Networks, Convolutional Neural Networks, Recurrent Neural Networks (RNNs), Generative Adversarial Networks, Self-Organizing Maps, Boltzmann Machines, Deep Reinforcement Learning, Autoencoders, Backpropagation, and Gradient Descent. This technique can be applied to the optimization and dynamic pricing of power distribution in the distribution facility.

8.0 References

[1] Gopi, A., Sudhakar, K., Keng, N. W., Krishnan, A. R. and Priya, S. S. (2021): Performance modeling of the weather impact on a utility-scale PV power plant in a tropical region. *International Journal of Photoenergy, 2021*.

[2] Akhter, M. N., Mekhilef, S., Mokhlis, H., & Mohamed Shah, N. (2019): L Review on forecasting of photovoltaic power generation based on machine learning and metaheuristic techniques. *IET Renewable Power Generation*, 13(7), 1009-1023.

[3] Dickey, D. A. and Fuller, W. A. (1979): Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, 74(366a), 427-431.

[4] Gupta, P. and Singh, R. (2021): PV power forecasting based on data-driven models: a review. *International Journal of Sustainable Engineering,* 14(6), 1733-1755.

[5] Kim, Y. S., Joo, H. Y., Kim, J. W., Jeong, S. Y. and Moon, J. H. (2021): Use of a big data analysis in regression of solar power generation on meteorological variables for a Korean solar power plant. *Applied Sciences*, 11(4), 1776.

[6] Gupta, P. and Singh, R. (2021, June): Univariate model for hour ahead multi-step solar irradiance forecasting. In 2021 IEEE 48th Photovoltaic Specialists Conference (PVSC) (pp. 0494-0501). IEEE.

[7] Kumar, M., Chandel, S. S. and Kumar, A. (2020): Performance analysis of a 10 MWp utility scale grid-connected canal-top photovoltaic power plant under Indian climatic conditions. *Energy*, 204, 117903.

[8] Maitanova, N., Telle, J. S., Hanke, B., Grottke, M., Schmidt, T., Maydell, K. V. and Agert, C. (2020): A machine learning approach to low-cost photovoltaic power prediction based on publicly available weather reports. *Energies*, 13(3), 735.

[9] Power Generation Data: https://www.kaggle.com/datasets/fvcoppen/solarpanelspower

[10] Weather Data: https://www.kaggle.com/datasets/ramima/weather-dataset-in-antwerp-belgium?select=weather_in_Antwerp_future2.csv

[11] Idman, Ebru, EmrahIdman, and Osman Yildirim. "Estimating Solar Power Plant Data Using Time Series Analysis Methods." In 2020 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), pp. 1-6. IEEE, 2020.

[12] Innani, Abhishek, and Shweta Dhatwalia. "Solar Power Predictor Using Ensemble Learning."