An Exploratory Analysis of Time Series Photovoltaic Stationary Data

Executive Summary

This report finds that the photovoltaic (PV) output is stable and predictably linked to solar radiation, showing no significant trends or fluctuations over time. While cloudy days reduce output, high UV index days compensate, ensuring reliable performance. The use of seasonal decomposition and Granger causality tests confirms solar radiation's predictive influence on PV output. Based on these findings, the report recommends expanding the PV array.

Univariate Analysis

We conducted univariate analysis on the PV output to understand the solar panel array's current energy production level.

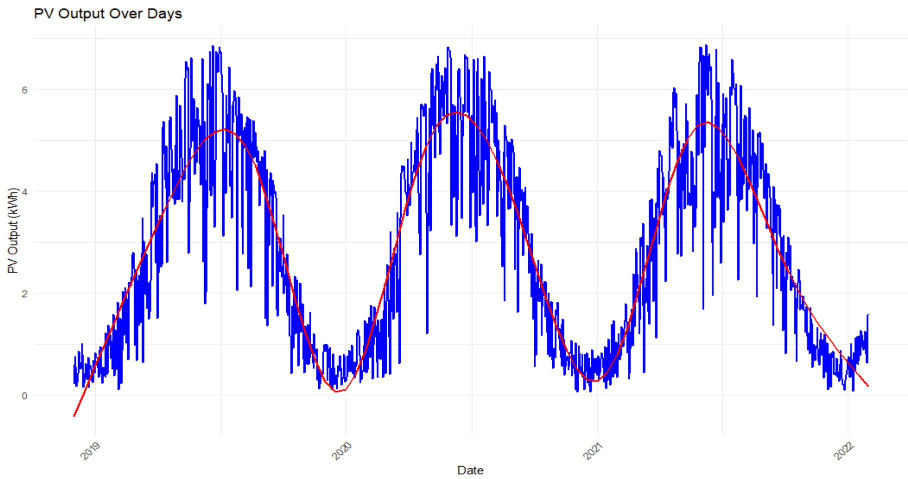

Figure 1 Plot of Seasonal Pattern in Time Series PV Output

The plot of daily PV output over time shows visible seasonality, indicating higher outputs during certain periods of the year and lower outputs in others. The red smoothing line filters out the noise of daily fluctuations to better highlight the seasonal pattern. The contrasting colours, customized axis labels and titles also enhance the interpretation of the data.

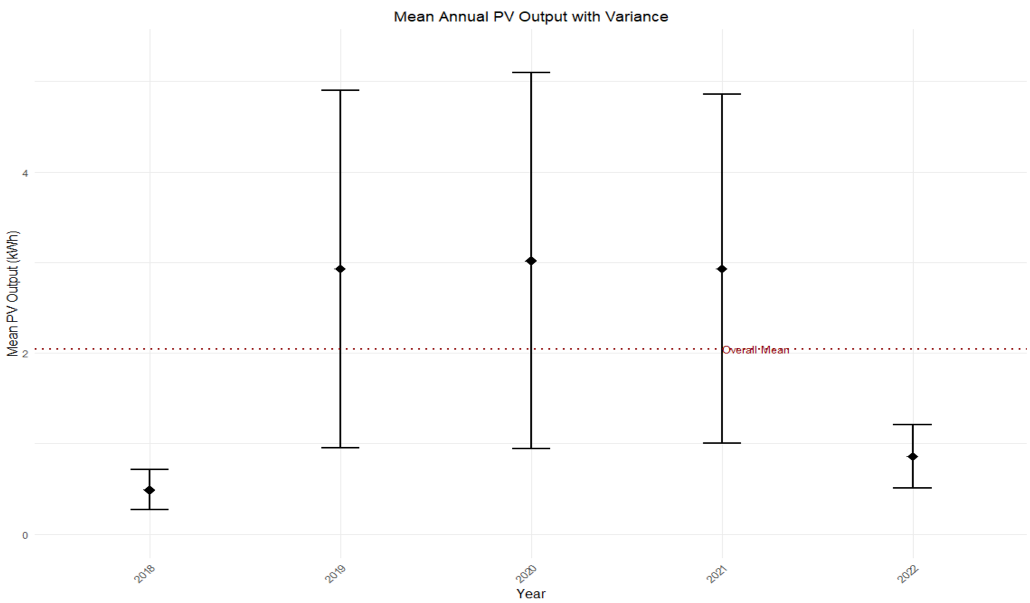

Figure 2 Box Plot Showing the Mean Annual PV Output with Variance

The last four years of data shows that on average we can expect 2.045 kWh per year. The boxplot underscores the variability in PV output, with particularly high yearly differences in both mean and variance during the period of 2019-2021.

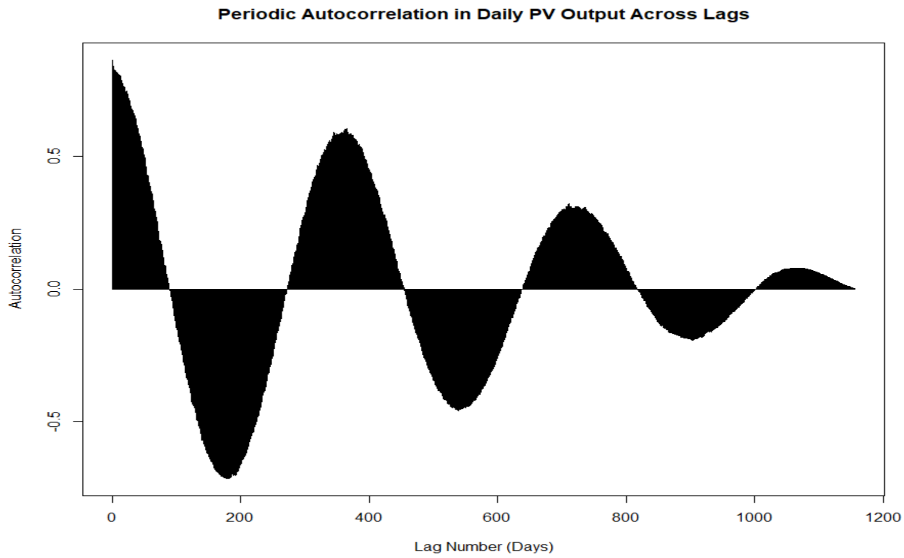

Figure 3 Periodic Autocorrelation in Daily PV Output Across Lags

The Periodic Autocorrelation Function (ACF) plot shows a sinusoidal pattern indicating a strong annual cycle, with diminishing correlation over time and fluctuating peak sizes. This weakening correlation may stem from natural attenuation and increasing data variability, while peak size variation might result from changing seasonal intensity, nonlinear effects, intervention events, or yearly anomalies.

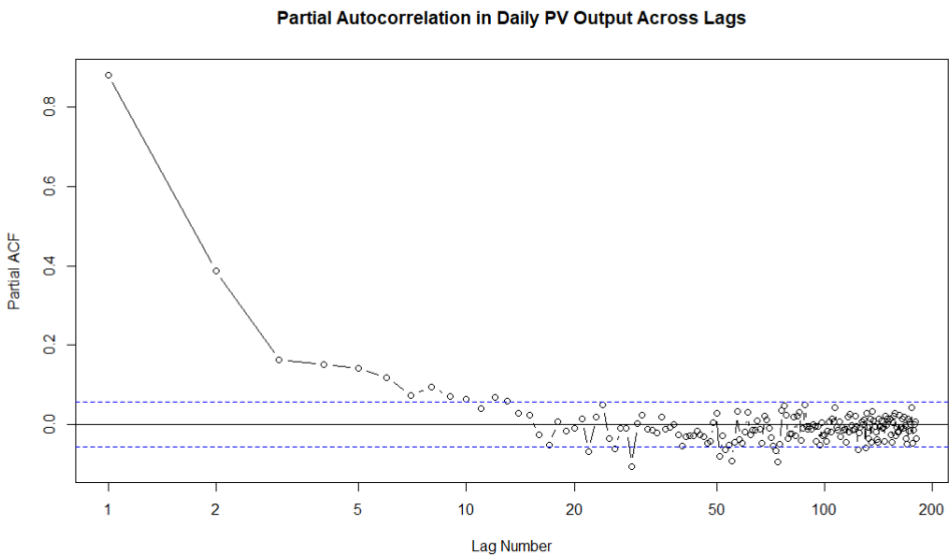

Figure 4 Partial Autocorrelation in Daily PV Output Across Lags (measured in days)

The Partial Autocorrelation Function (PACF) plot emphasizes the non-persistent nature of autocorrelation in daily PV output. There is an 80% correlation between today and tomorrow's weather, with a rapid decline in correlation strength at subsequent lags. To focus on meaningful correlations and data, we truncated the lag values to be between 1 and 180 and applied a logarithmic transformation to the x-axis.

Given the clear seasonality in the data and it's non-stationary, which we concluded after conducting the Augmented Dickey-Fuller (ADF) test[1] and analysing the ACF[2], we transformed the data into the frequency domain using Fast Fourier Transform. Using Nyquist's theorem and Welch's method, we generated a spectral density plot for PV output, revealing a peak that indicates a significant annual cycle in solar energy generation. This plot employs a logarithmic scale on the x-axis to improve visibility into relevant frequencies.

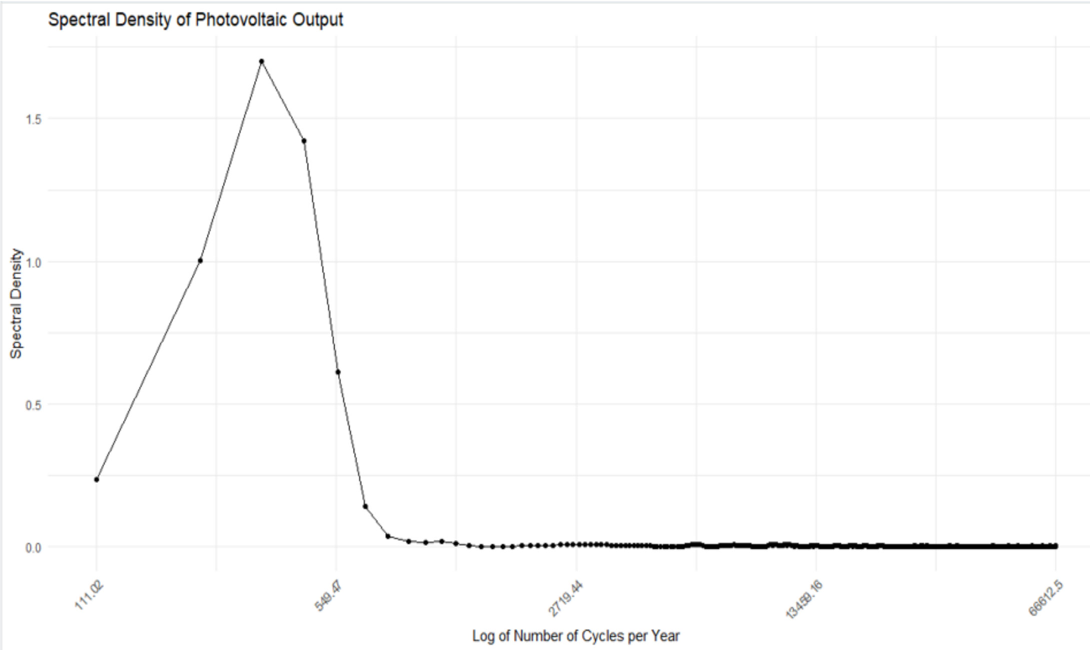

Figure 5 Spectral Density Analysis of Solar Energy Production: Highlighting Seasonal Variations

Multivariate Analysis

The correlation matrix (Figure 6) highlights that PV output is strongly linked to solar radiation (0.93) and UV index (0.89), indicating a direct impact of sunlight on solar energy production. A subsequent regression analysis of solar radiation and PV output supported the interpretation of their strong linear relationship. The heatmap also shows a positive relationship with maximum temperature (0.74), suggesting increased efficiency or sunlight hours at higher temperatures, and a significant negative correlation with humidity (-0.77), implying that higher humidity levels, typically associated with clouds and rain, negatively affect solar power. Additionally, the nearly perfect correlation between solar radiation and UV index (0.95) underscores their shared reliance on sunlight intensity. The matrix utilizes a diverging colour palette to clearly differentiate between positive and negative correlations.

The Granger causality test results indicate that both solar radiation and cloud cover possess a statistically significant ability to predict future values of PV output in the time series model. There is substantial evidence of instantaneous causality, which suggests that solar radiation and cloud cover, at any given time point, have a significant concurrent association with PV output. The past values of solar radiation demonstrate a statistically significant impact on the forecasting of future PV output values, as do the past values of cloud cover. This denotes that both solar radiation and cloud cover carry distinct predictive signals relevant to the PV output, confirming the existence of Granger causality. The statistical significance of these relationships is underscored by extremely low p-values,

[1] The ADF test yields a p-value of 0.5442 for the PV output data, suggesting non-stationarity as it exceeds the threshold of 0.05.

[2] The ACF plot reveals substantial autocorrelation at higher lags and seasonal patterns, indicating time dependency and seasonality in the data.

underscoring the strong dynamic linkage among these specific meteorological variables in the context of solar energy generation.

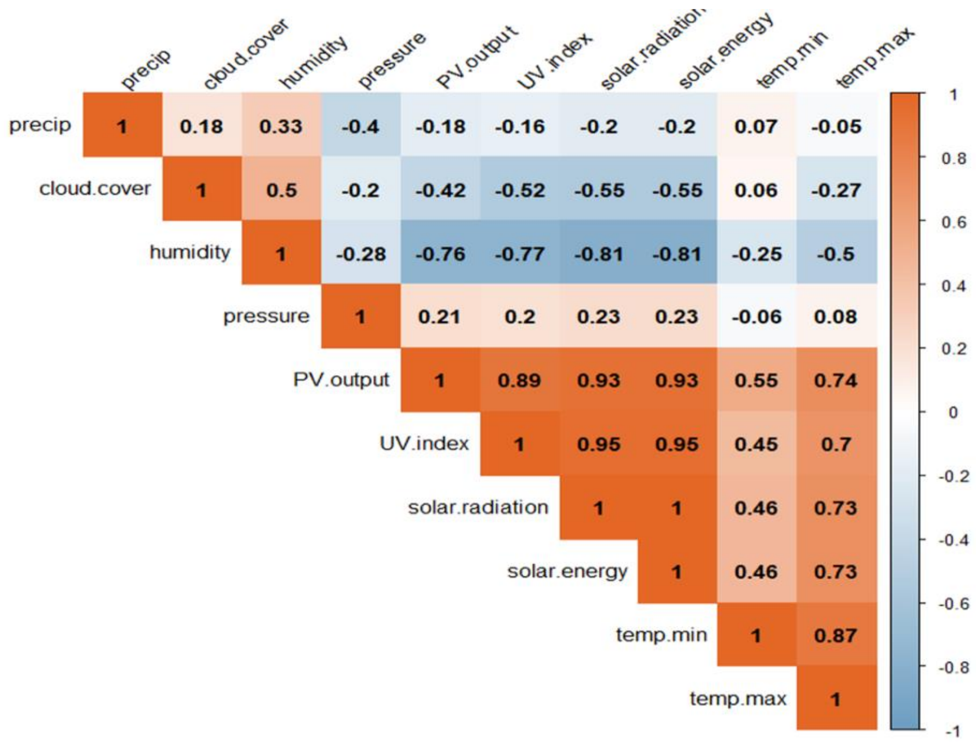

Figure 6 Correlation Matrix Heatmap of Features

The scatter plot suggests a stronger negative correlation between cloud cover and PV output with increasing UV index levels, indicating that cloudiness more significantly reduces PV output when the UV index is higher. Lower UV index categories show a flatter trend, implying other factors may influence PV output. Variability in the data points around the trend lines suggests additional variables may affect PV efficiency. The 'NA' category shows a negative trend as well, but conclusions from this subset are less reliable due to missing UV data.

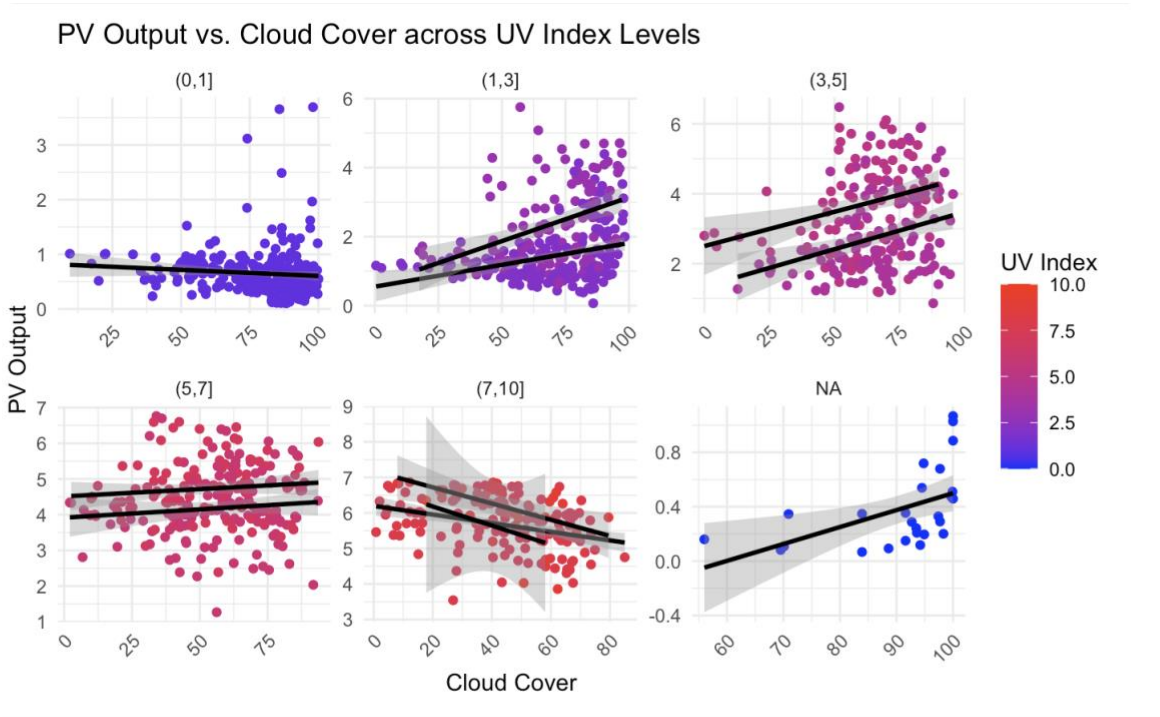

Figure 7 PV Output vs Cloud Cover Across UV Index Levels.