TITLE: ANALYSIS OF SOLAR POWER GENERATION & ENVIRONMENTAL FACTORS REPORT BY TRACY PREST

INTRODUCTION

The global transition to renewable energy has placed photovoltaic (PV) systems at the forefront of sustainable power generation. Solar energy, in particular, offers a clean and scalable alternative to fossil fuels. However, the performance of PV systems is influenced by several factors, including environmental conditions and system parameters.

Solar power play a critical role in addressing the world's growing energy demands while reducing greenhouse gas emissions. Photovoltaic (PV) systems convert sunlight into electricity, offering a sustainable, clean, and increasingly affordable solution.

This analysis focuses on an Electric Power Generation Dataset collected between October 1, 2022, and December 25, 2022, at five-minute intervals. The dataset includes key parameters such as voltage (Ua), current (Ia), power output (P), temperature (TEMP), and humidity (H). By leveraging **Excel** for data cleaning, **Python** for exploratory data analysis, and **Power BI** for visualization, this analysis aims to uncover trends, patterns, and recommend optimization strategies that can help improve system performance and decision-making.

PROBLEM STATEMENT

While PV systems are widely adopted, their performance fluctuates with varying environmental conditions. Without proper monitoring and data analysis, these fluctuations can lead to reduced energy output, operational inefficiencies, and revenue loss. The goal of this analysis is to understand how environmental and system parameters influence PV output and provide actionable insights for improving system performance.

Detailed analysis will make it possible to:

- Identify how temperature and humidity influence power generation.
- Detect patterns in voltage, current, and power output over time.
- Develop a clear, visual performance monitoring tool for quick decision-making.

Lack of insight may lead to suboptimal operation, reduced energy generation, and missed opportunities for system optimization.

This analysis seeks to address these challenges by cleaning and exploring the dataset, creating visualizations, and providing actionable insights through an interactive dashboard.

AIMS AND OBJECTIVES

Aim:

The aim of this analysis is to analyze photovoltaic power generation data to understand performance patterns, assess the influence of environmental factors, and recommend improvements for operational efficiency.

Objectives:

1. Clean and prepare the dataset for analysis using Excel.
2. Conduct exploratory data analysis (EDA) using Python to uncover trends and relationships.
3. Create subplots to visualize each variable against time.
4. Build an interactive Power BI dashboard to filter and explore data dynamically.
5. Identify key factors influencing power generation.
6. Recommend strategies for optimizing PV system performance.

METHODOLOGY

Data Collection:

Dataset covering October 1, 2022 – December 25, 2022, with five-minute intervals was gotten from Kaggle.

Data Cleaning:

The dataset was cleaned in Microsoft Excel to handle missing values, format timestamps, and standardize variable names.

Python Analysis:

Using Python's matplotlib and pandas, subplots were created for each variable (voltage, current, temperature, humidity) against time, revealing patterns and anomalies.

Visualization in Power BI:

A fully interactive dashboard was developed, enabling users to:

- View daily, weekly, and monthly trends
- Filter by time of day (morning/evening)
- Compare environmental conditions against power output

Deliverables

Clean dataset (Excel format).

Python scripts for exploratory analysis.

Power BI dashboard file.

POSSIBLE QUESTIONS

1. What time of the day produces high power output?
2. Does high temperature increase or decrease panel efficiency?
3. How does temperature affect solar output?

INSIGHTS

From the analysis, the following insights were gotten:

1. The Key Performance Insights (KPI) were Average Power Generated was **86.53kW,** Energy Generated kWh was **2.14 K,** Maximum Temperature was **29.10** °C and Average humidity was **61.78%**.

2. **Datetime** allows precise tracking of time from beginning of October to Christmas of 2022.

3. **Slicers** helps the monitoring process by the navigation through date and time for one to get the information they are looking for.

4. Power output (kW) of the PV system is a crucial parameter for assessing energy generation efficiency. At **8:30:00 AM** for the 3 months span, the highest power output was a total of **9710.84kW.** This means a lot of power was outputted by this time.

5. The peak Power output in the time series was **343.86kW** at **8:30:00AM**. This occurred in December 2022 within December 18 and 21st.

6. From the scatter plot, the total Power output at **8:30:00 AM** in December 2022 was **5420.36kW**. It was the highest of all the months.

7. **11:30:00 AM** was when the power output was minimum in November 2022 at **12.70kW**.

8. Voltage is the voltage measured in the PV System, capturing fluctuations in electrical potential. The average voltage when power was at its peak was **235.99V** and average current was **332.55A**. We can see that the current is more than the voltage.

9. Current is the current generated by the PV System, reflecting flow of electric charge. The **Bar chart** shows that:

In October 2022, the Average Current produced was **68.09A** and the Average Voltage **238.53V**.

In November 2022, the Average current was **86.78A** while the Average Voltage was **238.44V**.

In December 2022, the Average Current was **295.06A** and Average Voltage **236.59V**.

10. Average humidity of the peak power was **54.40%** and Average temperature at the time was **7.80** °C.

11. At minimum power output, the average humidity was **84.30 %** and the average temperature was **26.20** °C. The average voltage and current were **238.77V** and **69.04A**.

12. From the Scatter plot, there is a rise in power output in the Evening compared to Afternoons and Mornings.

13. A bell curve is recognized to show power normally distributed from **7:50:00 AM** to **12:50:00 PM**. From **8:30:00 AM** till **11:55:00 PM** there is no drop in power production. There is a possibility of stored energy even when the sun is not shining.

14. We can see from the Pie chart, the power output for the months are:

    October 2022, **17.17%**

    November 2022, **23.02%**

    December 2022, **59.8%**

From the analysis, the following key observations were made:

Daily Power Generation Patterns

Power output follows a bell-shaped curve, peaking during hours when solar irradiance is highest.

Power generation occurs during nighttime but stored energy was used.

Impact of Temperature

As temperature rises beyond a certain point, a slight decline in power efficiency is observed aligning with the known negative temperature coefficient of PV cells.

Effect of Humidity

High humidity days showed reduced average power output, possibly due to atmospheric scattering and absorption of sunlight.

Voltage–Current Relationship

Voltage and current are directly related to power output, with noticeable variations during cloudy periods or sudden weather changes.

ANALYSIS

Temporal Trends:

Using Python subplots, clear visual patterns emerged showing seasonal shifts, with slightly lower average power outputs in early October compared to late December which were higher. The subplots created with Python displays patterns in the figure below.

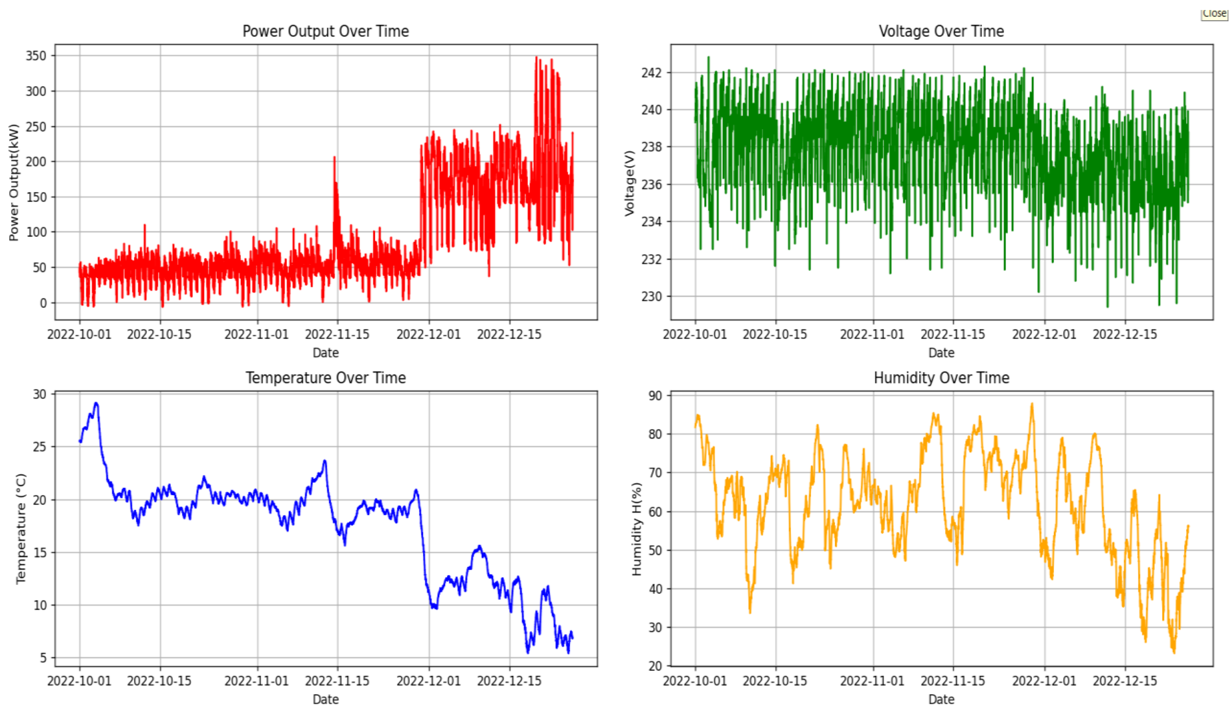


Figure 1: Subplots from Python.

Environmental Influence:

Both temperature and humidity demonstrated measurable impacts on energy yield. The relationship was not linear; extreme conditions negatively influenced output. Low temperature resulted in increased power output while high temperature reduces efficiency. High humidity lowers power output while reduced humidity increases solar radiation reaching the panels.

Dashboard Findings:

The Power BI dashboard made it easy to filter by date, identify performance dips, and cross-reference them with environmental factors.

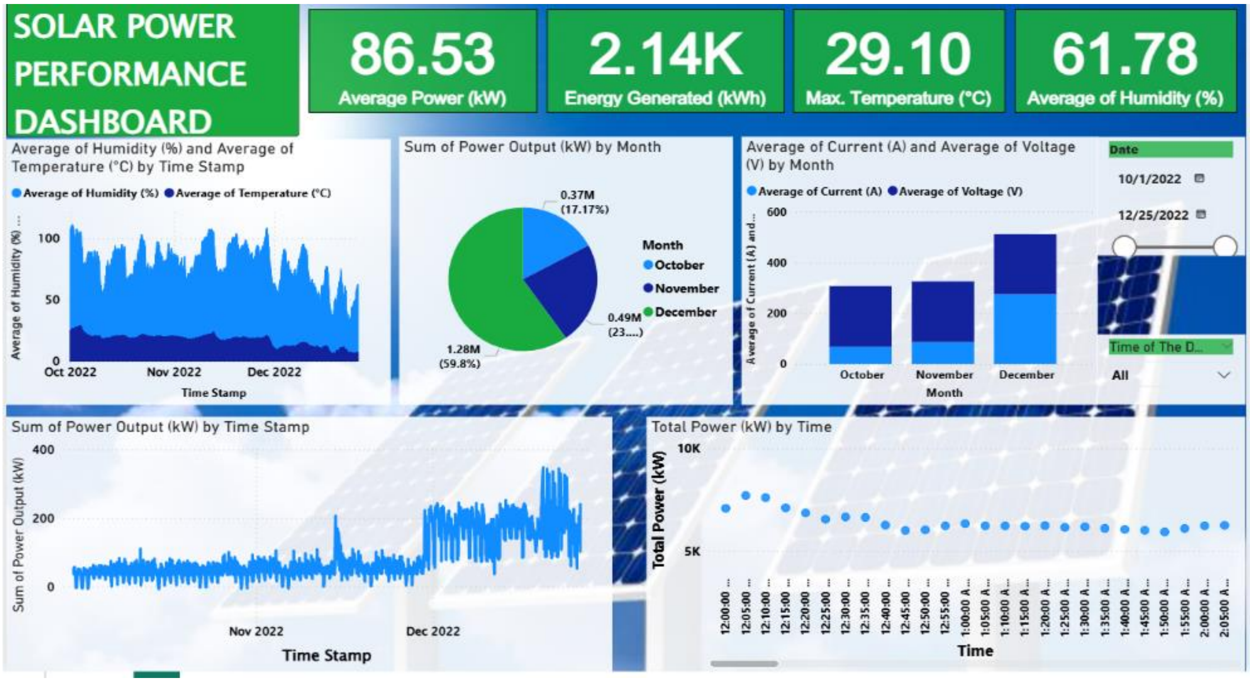


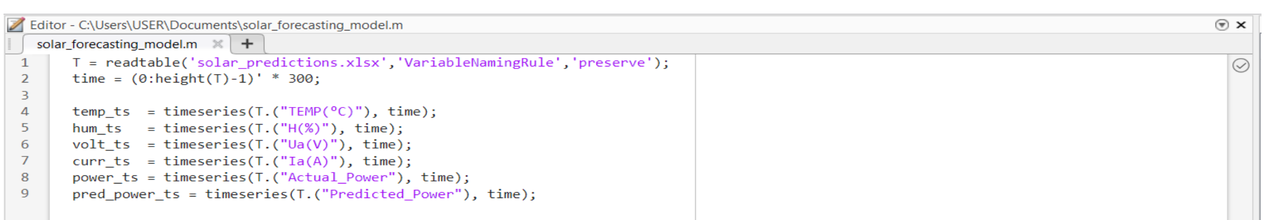
Figure 2: Power BI Dashboard for Visualization

To interact with the dashboard and identify various performance in each timestamp, view the Power BI file on the GitHub Repository.

MATLAB and Simulink Modeling

The dataset was prepared for dynamic system modeling using MATLAB, which transformed important variables into Simulink-compatible time-series objects. The original column headers were retained when the cleaned dataset was imported into MATLAB after being exported from Python. A relative time vector was created to show how the system changed over time because the data was recorded at consistent intervals of five minutes.

Electrical variables (voltage, current, and power output) and environmental elements (temperature and humidity) were transformed into separate timeseries objects. This stage made it possible to enter data into Simulink in an organized and consistent manner for additional analysis and visualization.


Editor - C:\Users\USER\Documents\solar_forecasting_model.m
solar_forecasting_model.m

```
T = readtable('solar_predictions.xlsx','VariableNamingRule','preserve');
time = (0:height(T)-1)' * 300;

temp_ts  = timeseries(T.("TEMP(°C)"), time);
hum_ts   = timeseries(T.("H(%)"), time);
volt_ts  = timeseries(T.("Ua(V)"), time);
curr_ts  = timeseries(T.("Ia(A)"), time);
power_ts = timeseries(T.("Actual_Power"), time);
pred_power_ts = timeseries(T.("Predicted_Power"), time);
```

Figure 3: A MATLAB script that creates time-series signals from electrical and environmental variables.

System Performance Model in Simulink

Using measured operational data, a Simulink model was created to visualize and study the behavior of a photovoltaic power system. Rather than focusing on intricate photovoltaic cell physics, the model emphasizes system-level performance monitoring.

From Workspace blocks were used to import time-series inputs from MATLAB into the Simulink environment. Environmental conditions, electrical characteristics, and system power output were monitored over time using different scopes. This structure makes it possible to visually represent how electrical features and environmental factors affect the overall performance of the system.

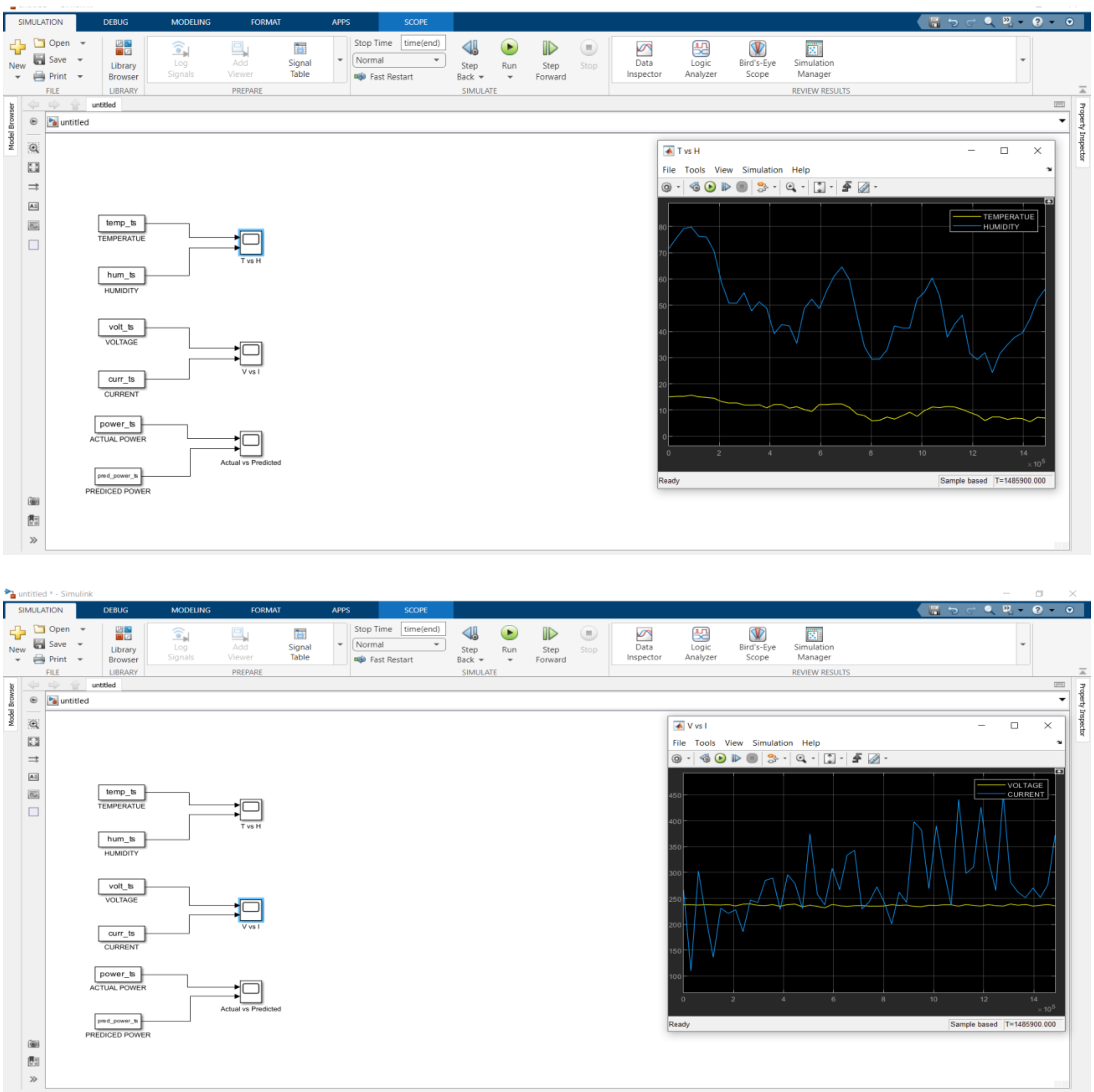


Figure 4: Simulink models displaying system environmental inputs.

Comparing Actual and Predicted Power

Both measured (actual) and machine-learning-predicted power outputs were imported into Simulink in order to assess model performance. As a result, system behavior and forecast accuracy over time could be directly compared visually. The comparison reveals times when the anticipated and measured outputs deviate more from one another and sheds light on the model's dependability.

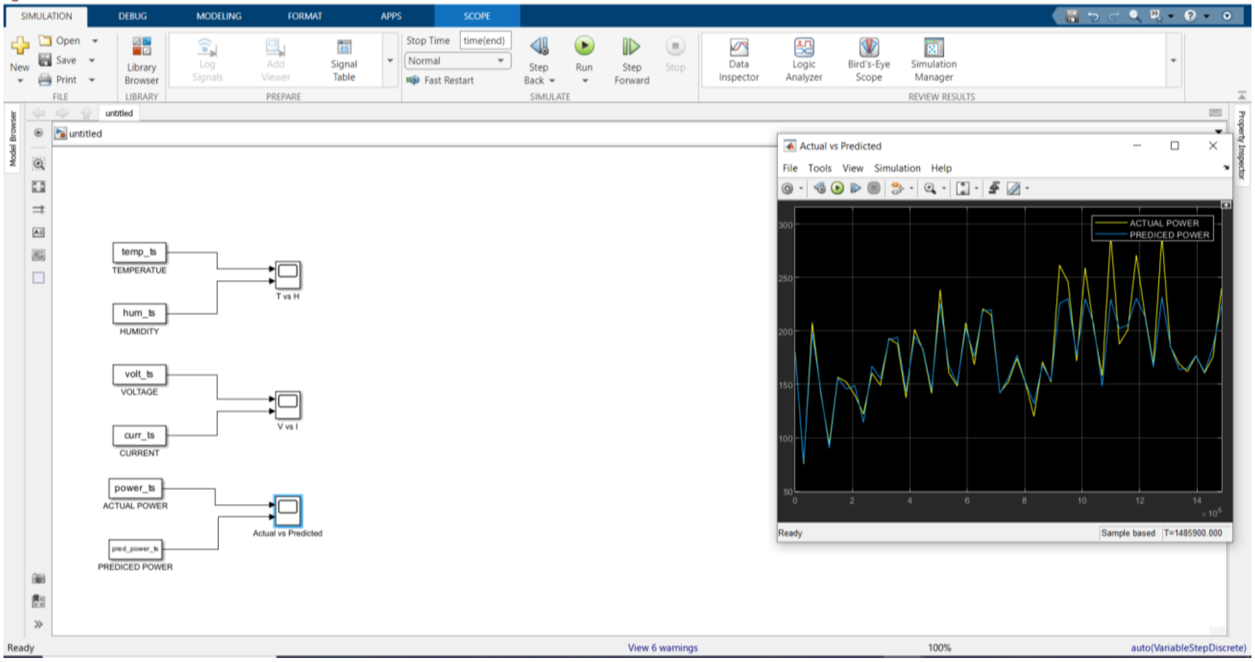


Figure 5: A comparison between expected and actual PV power output using Simulink.

CONCLUSION AND RECOMMENDATIONS

Conclusion

This analysis highlights the importance of detailed monitoring and visualization in maximizing PV system efficiency. By understanding the influence of temperature, humidity, and seasonal variation, operators can take informed actions to optimize performance.

The integration of Excel for cleaning, Python for in-depth analysis, and Power BI for visualization proved highly effective in translating raw data into actionable insights.

This MATLAB–Simulink integration, combines statistical modeling with dynamic system visualization and exemplifies a data-driven method for analyzing renewable energy systems. The process offers a useful framework for solar system performance monitoring, validation, and optimization by bridging data analytics and system-level engineering.

Recommendations

From the analysis carried out, the following recommendation should be followed:

1. **Performance Optimization:**
   Stake holders and users should consider installing a cooling mechanism or selecting PV panels with a better temperature coefficient for hot climates.

2. **Weather-Adaptive Scheduling:**
   Integrating real-time weather forecasting into system operations will help anticipate and mitigate efficiency drops.

3. **Preventive Maintenance:**
   Inspections should be scheduled after prolonged high-humidity periods to prevent moisture-related degradation.

4. **Dashboard Utilization:**
   Continuously monitor the dashboard for anomalies and act promptly on deviations from normal patterns.

Full code, dataset, and interactive dashboard are available on GitHub for further exploration.