Analyzing the Impact of Weather Patterns on Renewable Energy Production

Yilin Yang
Georgetown University DSAN

Problem Statement

In the context of escalating global demand for renewable energy and the imperative to reduce carbon emissions, China stands as a leader in renewable energy production. However, the integration of inherently variable solar and wind energy into the power grid presents significant challenges. The volatility and intermittency of these energy sources necessitate advanced forecasting models to ensure grid stability and operational efficiency.

This project aims to develop a predictive model that leverages extensive datasets from six wind farms and eight solar stations across China, analyzing key weather-related variables such as wind speed, wind direction, solar irradiance, air temperature, atmospheric pressure, and humidity. The goal is to accurately forecast energy output measured in megawatts (MW) using a combination of sophisticated machine learning techniques. These techniques include Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM) Networks, Gradient Boosting Machines (GBM), Random Forests and Gated Recurrent Unit (GRU). These models are chosen to address the complex dynamics between weather conditions and power generation, enhancing the accuracy of renewable energy forecasts.

Furthermore, the project commits to assessing the environmental impact of its computational modeling processes. Utilizing the CodeCarbon library, it will estimate and document the carbon emissions associated with the model training, aligning the project with sustainable practices and environmental stewardship. The insights generated by this research will not only enhance renewable energy forecasting but also support informed decision-making in energy management and policy development, contributing to more effective integration of renewable sources into the national grid.

Data

The dataset comprises over two years (2019–2020) of power generation and weather-related data collected at 15-minute intervals from six wind farms and eight solar stations in China. The data includes key parameters such as wind speed, wind direction, total solar irradiance, air temperature, atmospheric pressure, and relative humidity. These stations are strategically located across various climate zones and terrains, including deserts, mountains, and plains, to ensure diverse and representative data.

In addition to the on-site data collected from six wind farms and eight solar stations in China, this project will also utilize the CarbonMonitor-Power dataset to provide a broader context of renewable energy generation in China. The CarbonMonitor-Power dataset offers near-real-time global power generation data, including daily time scales for renewable sources such as solar and wind power. With over two million records, this dataset provides valuable insights into the energy mix of China and other countries, highlighting the contributions of renewable energy to the overall power generation.

Methodology

Data Processing: The initial step involved consolidating the datasets, followed by addressing missing data through methods such as interpolation and imputation, as well as identifying and handling outliers using statistical approaches like the interquartile range (IQR). Feature selection was guided by correlation analysis, which helped pinpoint relevant seasonal and temporal patterns that could influence the model's performance.

Model Development: We employed an ensemble of sophisticated machine learning techniques to capture the complex dynamics of energy production:

Recurrent Neural Networks (RNN): To model sequential dependencies within time-series data.

Long Short-Term Memory (LSTM) Networks: To harness temporal sequence information for enhanced prediction accuracy.

Gradient Boosting Machines (GBM): For their efficacy in capturing complex, nonlinear relationships within data.

Random Forests: For their robustness and ability to handle diverse datasets.

Gated Recurrent Unit (GRU) Networks: Specifically chosen for their efficiency in learning from large time series data, making them particularly suitable for modeling the volatile patterns in renewable energy generation.

Model Validation: To ensure the robustness of the model, we utilized cross-validation techniques to minimize overfitting and boost generalizability. The model's performance was gauged using a suite of metrics, including Mean Absolute Error (MAE), Mean Squared Error (MSE), and the Coefficient of Determination (R^2), ensuring a comprehensive assessment of accuracy and reliability.

Environmental Impact Assessment: The environmental cost of the modeling procedures was scrutinized through the CodeCarbon tool, which calculates the carbon emissions attributed to computational activities based on factors like energy use, geographical location of computation, and the duration of the model training. This assessment is integral to align the project with sustainable practices and environmental stewardship.

Our methodical and integrative approach guarantees the development of a potent predictive model that can significantly improve the forecasting of renewable energy outputs, thereby aiding strategic planning and policy formulation in the renewable energy sector.

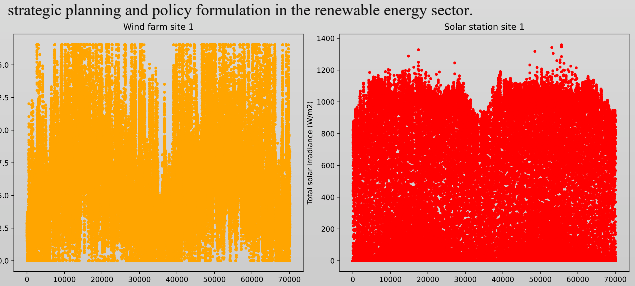

Figure 1: Seasonal and Daily Variability of Wind and Solar Energy

Evaluation metrics and Results

Correlation Analysis:

Wind Farms (Figure2):

- The wind speed measurements at **different heights (10m, 30m, 50m)** and the **central wind speed (WS_cen)** are strongly positively correlated with each other, which is expected because they are likely measuring the same physical phenomenon at slightly different locations or conditions.
- Wind direction at different heights also shows a high degree of correlation, though slightly less than wind speed, indicating that the direction is relatively consistent at different measurement heights.
- There is a moderate to strong positive correlation between wind speed and the **power output (Power(MW))**, which suggests that higher wind speeds lead to higher power generation, as would be expected for wind turbines.
- Other factors like **air temperature (Air_T), air pressure (Air_P),** and **humidity (Air_H)** have weaker correlations with power output.

Solar Stations (Figure 3):

- **Total solar irradiance (TSI), direct normal irradiance (DNI),** and **global horizontal irradiance (GHI)** are all strongly positively correlated with each other and with power output. This is expected since solar power generation directly depends on the intensity of sunlight received.
- Similar to the wind farms, the air temperature, air pressure, and humidity show varying degrees of correlation with power output, suggesting that these factors might influence the efficiency of solar panels to some extent but are not as directly impactful as the solar irradiance variables.
- There is a strong positive correlation between solar irradiance and **power output (Power(MW))**, indicating effective conversion of solar energy into electrical power.

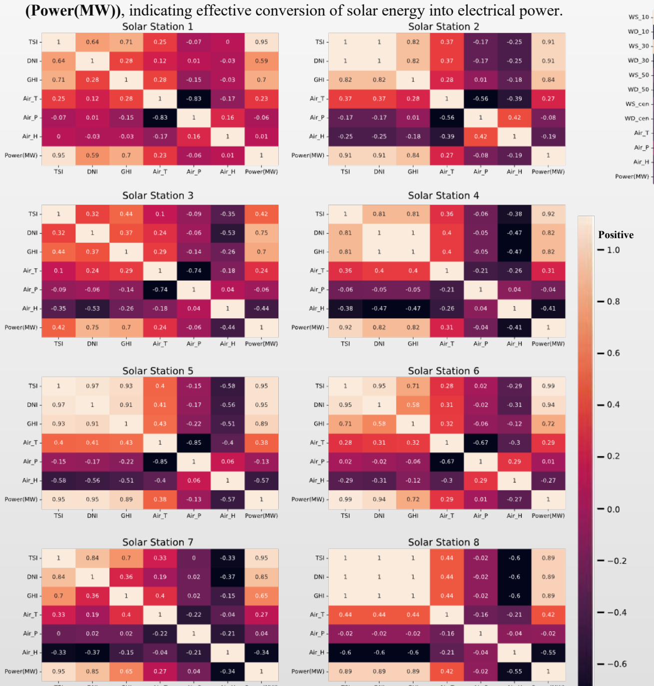

Figure 3: Correlation Analysis for Solar Station Sites

Solar Station Sites	MSE	MAE	R^2
Random Forest	16.2410	1.4084	0.9413
LSTM	0.0167	0.0645	-2.2759
Gradient Boosting Machines	16.5204	1.5440	0.9403
RNN	0.0045	0.0364	0.8610
GRU	**0.0022**	**0.0192**	**0.9327**

Table 2: Performance Metrics of Predictive Models for Solar Station Sites

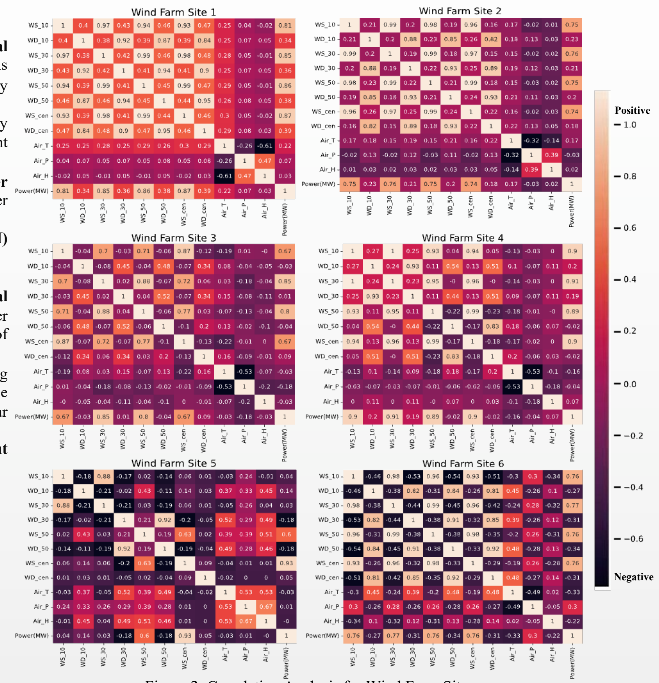

Figure 2: Correlation Analysis for Wind Farm Sites

Model Comparison for Wind Farm Sites:

- For wind farm sites, the GRU model demonstrates superior performance with the lowest MSE at 0.0025 and the lowest MAE at 0.0276, suggesting highly accurate predictions. Its R^2 value of 0.8677 indicates it can explain a significant portion of the variance in the data, which is a robust indicator of its predictive capacity. Although the Random Forest and GBM models boast higher R^2 scores, their much larger MSE and MAE figures point to less accurate individual predictions. In comparison, the LSTM and RNN models have lower error metrics than Random Forest and GBM, yet they don't match the GRU's performance. Overall, for precise forecasting in wind energy, the GRU model's blend of low error rates and high explanatory power makes it the preferred choice.

Wind Farm Sites	MSE	MAE	R^2
Random Forest	128.3873	6.0228	0.9051
LSTM	0.0041	0.0475	0.6081
Gradient Boosting Machines	132.1294	6.2613	0.9018
RNN	0.0037	0.0367	0.8063
GRU	**0.0025**	**0.0276**	**0.8677**

Table 1: Performance Metrics of Predictive Models for Wind Farm Sites

Model Comparison for Solar Station Sites

- For solar station sites, the GRU model emerges as the clear leader, achieving an impressive balance between error minimization and predictive power, with the lowest MSE at 0.0022 and the smallest MAE at 0.0192. Its R^2 score of 0.9327 reflects a high correlation between the model's predictions and actual values, indicating its capacity to capture the majority of the variance within the dataset.
- While the Random Forest and GBM models demonstrate a strong ability to fit the data as evidenced by their R^2 scores, their higher MSE and MAE values suggest that their predictions are not as close to the true values as the GRU model, potentially compromising their precision.
- The LSTM model lags significantly behind, with a negative R^2 value indicating that its predictions are not well aligned with the actual data, which could be attributed to overfitting, underfitting, or model misspecification.
- In sum, the GRU model's superior performance metrics underscore its reliability and precision, making it the preferable model for solar energy output forecasting and for stakeholders seeking dependable predictions in the renewable energy sector.

Reference

- Liu, H., Zhang, Z., Wang, Z., Chen, M., Ding, J., & Cai, Z. (2021). Solar and wind power data from the Chinese State Grid Renewable Energy Generation Forecasting Competition [Data set]. Figshare. https://doi.org/10.6084/m9.figshare.17304221.v4
- Zhu, B. et al. CM-Power near-real-time monitoring of global power generation on hourly to daily scales. Figshare https://doi.org/10.6084/m9.figshare.21938102.v2 (2022).

Discussion&Limitations

Technical Discussion

The observed data from Figure 3 underscores the inherent variability and seasonal trends in solar and wind power generation, which are crucial for the development of our predictive models. The dominance of wind power and the pronounced variability in solar output highlight the complex dynamics these models must handle. This validates our choice of sophisticated machine learning techniques including GRU, LSTM, and RNN, which are capable of capturing these temporal and sequential dependencies effectively.

The granularity of data collected at 15-minute intervals provides a detailed view of the fluctuations within days and across seasons, reinforcing the need for models that can dynamically adapt to sudden changes in weather conditions and daylight. Our methodology, which integrates comprehensive data preprocessing and robust machine learning algorithms, is designed to handle these complexities, ensuring accurate forecasts and aiding in efficient grid management.

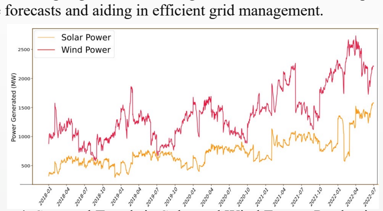

Figure 4: Seasonal Trends in Solar and Wind Energy Production in China

Limitations

While the project employs advanced modeling techniques and extensive datasets, there are several limitations:

1. **Data Quality and Completeness**: Discontinuities and anomalies in the data, as observed in the solar power outputs, pose challenges. These gaps could affect model training and predictions, necessitating sophisticated data imputation and anomaly detection techniques.
2. **Model Generalizability**: The models are trained on specific datasets from China, which may limit their applicability to regions with different climatic and geographical characteristics. Further testing is required to validate these models across diverse environments.
3. **Environmental Impact of Computational Processes**: Although the CodeCarbon tool provides estimates of the carbon footprint of our computational models, the accuracy of these estimates and the scalability of such assessments remain areas for improvement. This is crucial for ensuring that our research aligns with sustainable practices.

Future Directions

To address these limitations and further enhance the utility of our research, future efforts will focus on:

- **Enhancing Data Integrity**: Implementing real-time monitoring and error correction mechanisms during data collection to improve the quality and reliability of the datasets.
- **Expanding Model Testing**: Conducting cross-regional studies to assess and enhance the generalizability of the models, adapting them as needed to cater to varying environmental conditions.
- **Reducing Computational Environmental Impact**: Refining the methods for estimating and mitigating the carbon emissions associated with model training, aiming to develop more energy-efficient algorithms.

Carbon Cost

The observations emphasize the balance between computational efficiency and environmental impact within machine learning applications for renewable energy forecasting. Both RNN models demonstrate strong performance metrics, which is crucial for accurate forecasting in renewable energy applications.

For the wind farm dataset, the GRU model achieved commendable predictive accuracy, as reflected by its evaluation metrics. However, it consumed 0.083587 kWh of electricity and emitted about 0.030855 kg of CO_2 during training. These figures underline the environmental costs associated with deploying machine learning models but also highlight the relatively low impact compared to traditional energy-intensive processes.

Similarly, the solar station GRU model exhibited high accuracy with slightly higher energy consumption of 0.220511 kWh and CO_2 emissions of 0.081398 kg. The increased energy demand and emissions are balanced by the model's superior predictive performance, which can enhance the operational efficiency of solar energy production.

Model	Energy Consumed (kWh)	CO2 Emissions (kg)
Best GRU for Wind Farm	0.083587	0.030855
Best GRU for Solar Station	0.220511	0.081398

Table 3: Environmental Impact and Performance Metrics of RNN Models for Renewable Energy Forecasting