Multivariate Wind Speed Forecasting based on Vector AutoRegressive models

Group 9 Project Members

	Name	Student ID
1	Jodie Marquez	300566408
2	Riya Rana	300567404
Individual	Nguyen Quoc Dinh	300550781

I. Executive Summary

Research Goal:

This project endeavors to forecast wind speeds in Christchurch, New Zealand, incorporating natural factors such as rainfall and temperature to examine their impact on predictive accuracy. The primary objective is to discern the extent to which these factors contribute meaningfully to the daily wind speed forecasting performance.

Methods:

Utilizing historical daily weather data sourced from the weather25 website, the dataset encompasses key natural factors - average daily wind speed, rainfall, minimum temperature, and maximum temperature.

Employing the Vector Autoregression (VAR) statistical model on this time series data facilitates wind speed prediction based on its historical values, along with the concurrent influence of rainfall and temperature. The VAR model relates current observations of wind speed with past observations of itself and past observations of other factors in this dataset.

Findings:

The analysis reveals relative associations among wind speed, rainfall, and temperature across time. However, it also indicates that the inclusion of rainfall and temperature does not significantly enhance the predictive accuracy of wind speed compared to utilizing solely wind speed data. Consequently, these findings prompt exploration into other natural factors that could potentially bolster the predictive performance of wind speed. Emphasizing the necessity of identifying additional influential factors, this study opens avenues for refining accuracy in wind speed forecasting.

Table of Contents

II. Background

According to the International Renewable Energy Agency (IRENA), wind power, coupled with solar energy, would lead the way for the transformation of the global electricity sector and it could cover more than one-third of global power needs (35%), becoming the world's foremost generation source by 2050. Wind is a reliable, sustainable, clean, and commercially viable energy. Wind energy is now successfully competing across the globe, building new industries, creating hundreds of thousands of jobs, and leading the way towards a clean energy future.

New Zealand, with its exceptional wind resources, stands as an ideal ground for the development of the wind power industry. Anticipated data from the New Zealand Wind Energy Association shows that New Zealand will have 19 operational on-shore wind farms with a total capacity of 1.45 MW by the end of 2023, constituting around 10% of the nation's installed generation capacity, and catering to over 450,000 Kiwi homes annually.

Christchurch, the largest city in the South Island, has a few wind turbines around, making it a promising location for wind research. This study aims to explore potentially influential natural factors that could enhance the predictive performance of wind speed in this region. The imperative to understand and predict wind patterns drives the motivation behind this project.

While New Zealand's national meteorological service, MetService, employs in-house high-resolution weather regional forecast models based on the Weather Research and Forecasting (WRF) models, the complexity and cost associated with collecting a long list of input variables for these models in the initial stages of comprehensive wind speed research led to the adoption of a simpler approach. Utilizing the statistical technique known as Vector Autoregression (VAR), this study presents an economical yet effective model with a limited set of input variables suitable for the preliminary research phase.

III. Data Description

Our primary data source for this study is a comprehensive weather dataset crawled from Weather25. After being crawled, merged, and cleaned, the dataset encompasses a span from January 1, 2019, to September 29, 2023, with a daily frequency, resulting in a total of 1733 entries. The data is organized with a DatetimeIndex and comprises four key variables crucial to our analysis.

Variable	Data Type	Description
rain_mm	Float	Rainfall in Millimeters
temp_max_c	Float	Maximum Temperature in Celsius
temp_min_c	Float	Minimum Temperature in Celsius
wind_speed_km	Float	Wind Speed in Kilometers per Hour

The time indices exhibiting missing values display a discernible pattern that spans the entire time range. Notably, most of the missing values occur consistently and simultaneously across all four variables, manifesting as a singular missing value per month consistently falling on the month's end day. Several plausible explanations for this pattern include considerations related to the data collection processes, data reporting protocols, or idiosyncrasies in the data processing workflow.

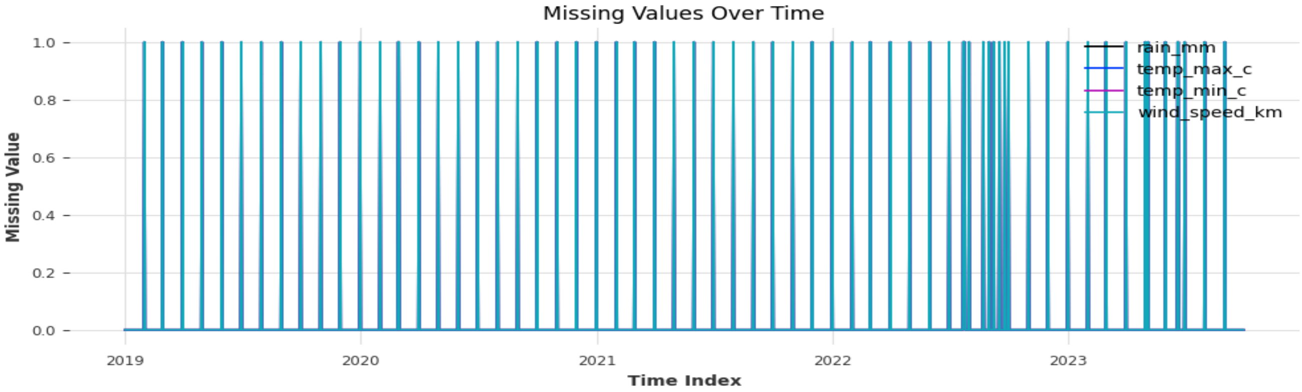
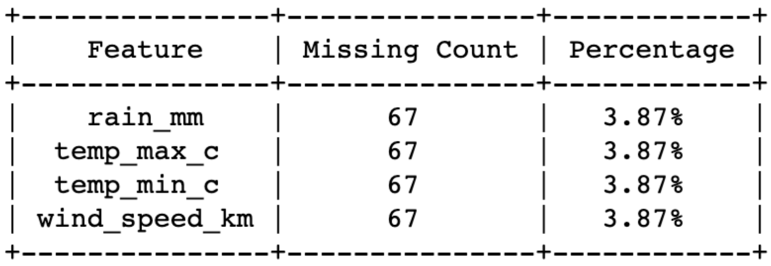


Figure 1: Missing Values occurring on specific dates for all 4 variables

Missing values, constituting about 3.87% of the total dataset, equate to 67 instances per feature. Remarkably, these gaps manifest uniformly across all four variables, occurring simultaneously. Given their minority status, a straightforward imputation approach involves filling them with the preceding or succeeding values within the respective features.



```
+---------------+---------------+------------+
|    Feature    | Missing Count | Percentage |
+---------------+---------------+------------+
|    rain_mm    |      67       |   3.87%    |
|   temp_max_c  |      67       |   3.87%    |
|   temp_min_c  |      67       |   3.87%    |
| wind_speed_km |      67       |   3.87%    |
+---------------+---------------+------------+
```

Table 1: Total of missing values and their percentage for 4 variables

IV. Ethics, Privacy, and Security

Ethical Considerations

Our commitment to ethical conduct in research involves a thorough examination of potential implications and consequences. In meteorological data analysis, where information is often derived from public records, ethical concerns primarily revolve around responsible and unbiased interpretation. Fortunately, there is no need to aggregate and anonymize data in this project.

Privacy Concerns

Since meteorological data is typically drawn from publicly available sources, our privacy concerns are relatively minimal. However, we recognize the importance of safeguarding individual identities and sensitive information. Throughout the analysis, our approach involves aggregating and anonymizing data to mitigate any potential privacy risks associated with specific data points. Our commitment is to uphold privacy standards while harnessing the power of data to derive meaningful insights.

Security Measures

Securing both our data and results is fundamental to maintaining the credibility and integrity of our research. While our dataset may not contain highly sensitive information, ensuring its confidentiality and preventing unauthorized access is paramount. To achieve this, we've implemented secure data storage practices, including password protection and restricted access limited to authorized personnel. Encryption methods, where applicable, add an extra layer of security.

Our commitment to secure practices extends to the sharing of results. Intermediate and final outcomes will be disseminated judiciously, aligning with academic and ethical standards. Communication channels within our collaborative project are secured, and sensitive information is shared through approved and secure platforms.

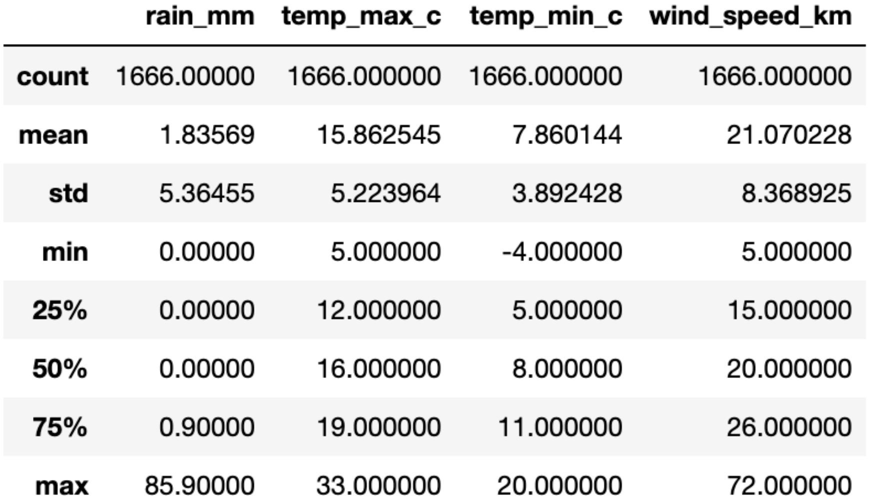
By proactively addressing ethical, privacy, and security considerations, our research is a testament to our commitment to upholding the highest standards of integrity and responsibility. This not only enhances the robustness and credibility of our findings but also ensures that our data-driven insights are derived and shared responsibly.

V. Exploratory Data Analysis

In this Exploratory Data Analysis (EDA), the primary objective is to gain a comprehensive understanding of the data related to the VAR model. This includes wind speed, rainfall, minimal temperature, and maximum temperature. The analysis is accomplished by employing both visual and statistical techniques.

1. Summary Statistics

The table below presents summary statistics for the input features:



	rain_mm	temp_max_c	temp_min_c	wind_speed_km
count	1666.00000	1666.000000	1666.000000	1666.000000
mean	1.83569	15.862545	7.860144	21.070228
std	5.36455	5.223964	3.892428	8.368925
min	0.00000	5.000000	-4.000000	5.000000
25%	0.00000	12.000000	5.000000	15.000000
50%	0.00000	16.000000	8.000000	20.000000
75%	0.90000	19.000000	11.000000	26.000000
max	85.90000	33.000000	20.000000	72.000000

Table 1: Summary Statistics for all features used as part of VAR modeling

- Rainfall Distribution: The average rainfall is 1.84 mm, with a notable presence of zero values (25th and 50th percentiles at 0 mm). However, the maximum rainfall recorded is quite high at 85.9 mm, indicating the potential for extreme weather events such as heavy rainfall and the risk of flooding.

- Temperature Variation: The temperature data shows a range from a minimum of -4°C to a maximum of 33°C. This reflects a diverse climate, with temperatures spanning from cold to relatively hot. The mean maximum temperature is 15.86°C, and the mean minimum temperature is 7.86°C.

- Wind Speed Range: Wind speed varies from a minimum of 5 km/h to a maximum of 72 km/h. The mean wind speed is 21.07 km/h, indicating a moderate average wind condition. The data suggests a broad spectrum of wind speeds, potentially influencing factors like local weather patterns and climate.

- Rainfall Impact on Averages: The presence of high rainfall values, particularly the maximum of 85.9 mm, contributes to an elevated mean rainfall value. It's worth noting that this outlier may significantly impact the overall average, as evidenced by the mean being higher than both the 75th percentile (0.9 mm) and the median (0 mm).

2. Time Series Overview

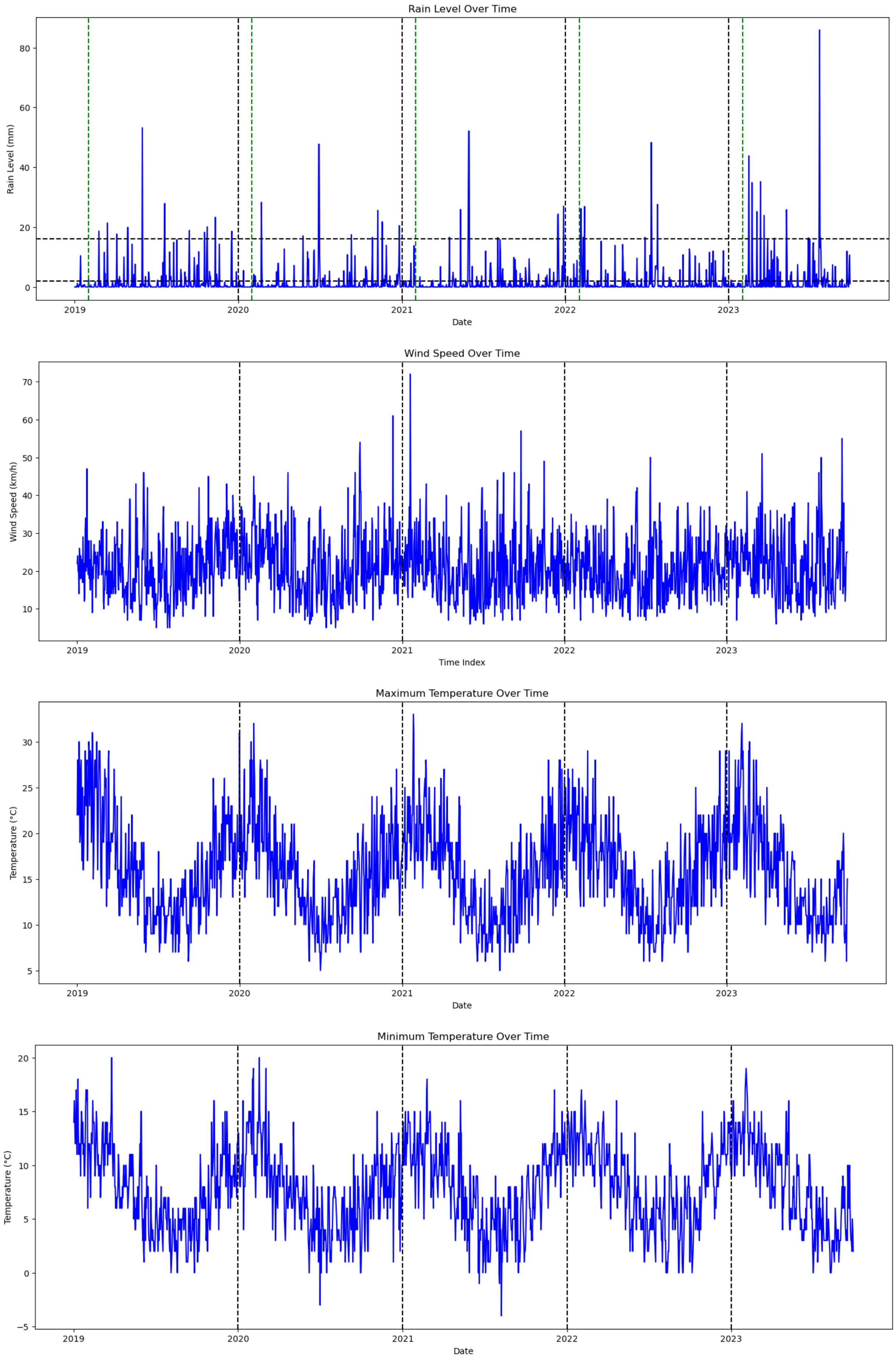


Figure 2: Rain Level, Wind Speed, Minimum Temperature, and Maximum Temperature (January 1, 2019 - September 29, 2023*)*

3. Correlation

By utilizing a correlation matrix, the linear relationship between wind speed and other variables is evaluated at the same point in time.

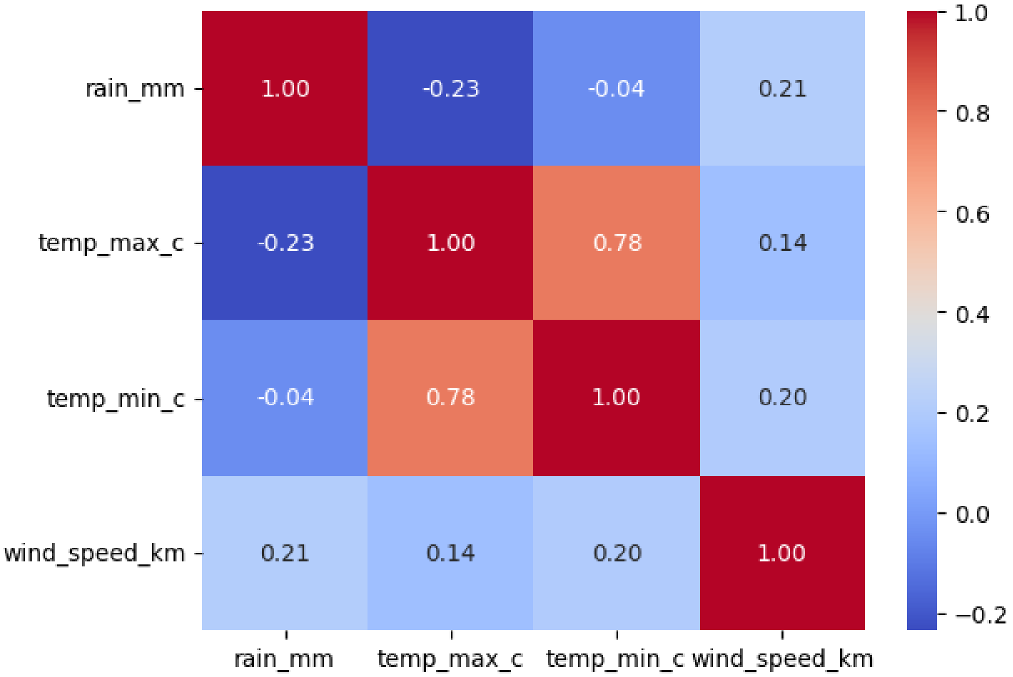
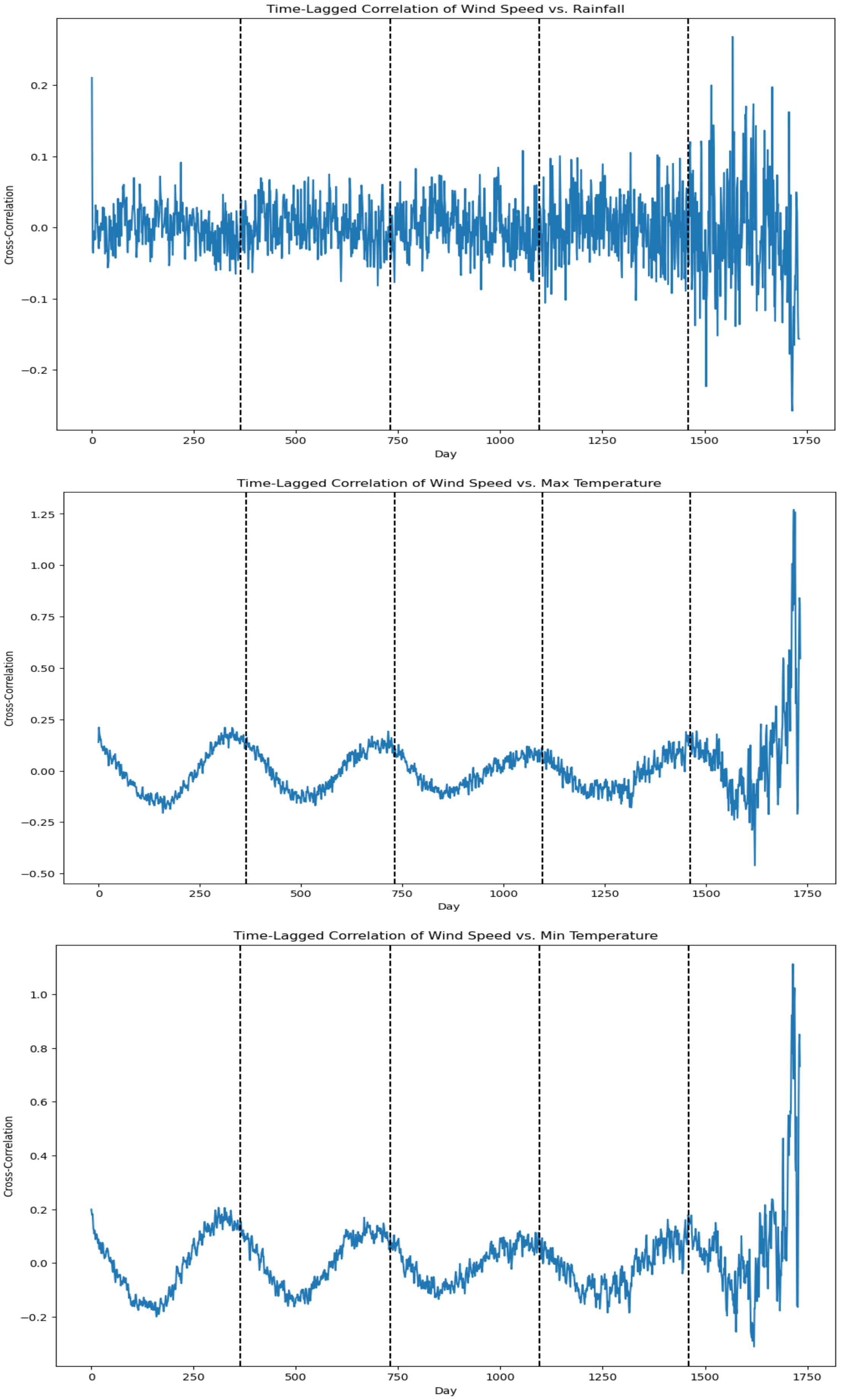


Figure 4: Correlation Matrix

- Rainfall and Wind Speed: A positive correlation of around 0.21 indicates a modest association between rainfall and wind speed, hinting that heightened rainfall may coincide with slightly elevated wind speeds.
- Temperature and Wind Speed: The correlation between maximum temperature and wind speed registers at a modest 0.14, implying a weak positive connection. In contrast, the correlation between minimum temperature and wind speed strengthens slightly to 0.2. Both findings suggest a modest positive relationship between temperature and wind speed, implying that higher temperatures may be marginally linked to increased wind speed.
- Temperature Interplay: A robust correlation of 0.78 between maximum and minimum temperatures signifies a pronounced positive relationship, aligning with the expected pattern of these temperatures moving in sync. This temperature interplay should be taken into account when examining variations in wind speed, as the coupling of maximum and minimum temperatures may influence wind patterns.

While the correlation matrix provides us with a snapshot of the contemporaneous relationships among variables, the time-lagged cross-correlation function can reveal more nuanced patterns by considering the influence of past observations. The time-lagged correlation is when the first metric increases or decreases in sync with the second but with a lag between the first and second metric. The lag time here is a number of days delayed between the two metrics.

- Significant cross-correlation values, both negative and positive, appear at lags from day 1500 onward, with peaks reaching +0.2 and -0.2. A positive correlation may indicate that high rainfall precedes significantly increased wind speeds 1500 days later, while a negative correlation suggests decreased wind speeds after a certain lag following heightened rainfall. These findings highlight intricate temporal dynamics between rainfall and wind speed, revealing prolonged influence and potential delayed relationships.
- Similar behavior is observed in the case of wind speed and temperature, with plots displaying significant variance at lags from day 1500. Additionally, a distinct annual seasonality pattern is evident for the four years preceding 2023, starting from day 1500 onward. Correlations gradually decrease toward the middle of the year and increase towards the year-end, indicating a seasonal influence on the relationship between temperature and wind speed.



Time-Lagged Correlation of Wind Speed vs. Rainfall

Time-Lagged Correlation of Wind Speed vs. Max Temperature

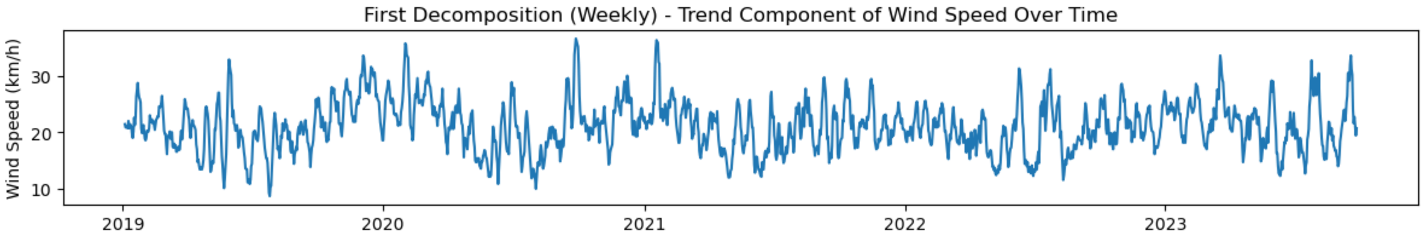
Time-Lagged Correlation of Wind Speed vs. Min Temperature

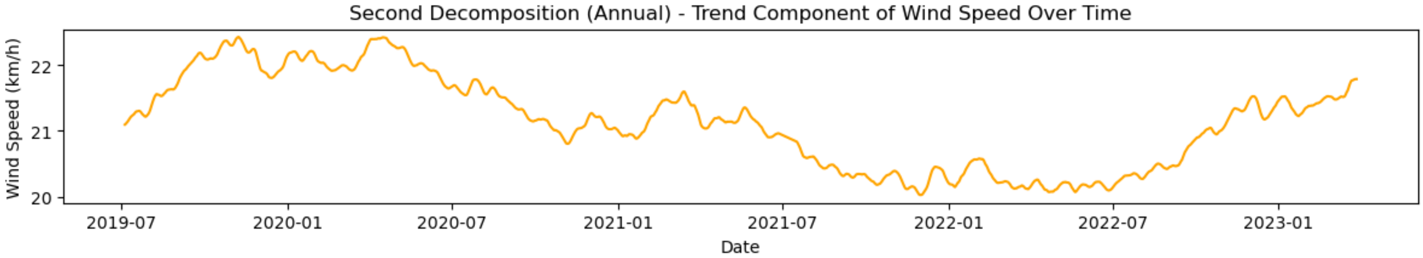
Figure 5: Time-lagged correlation of
Wind Speed vs. Rainfall, Wind Speed vs. Max Temperature, and Wind Speed vs. Max Temperature

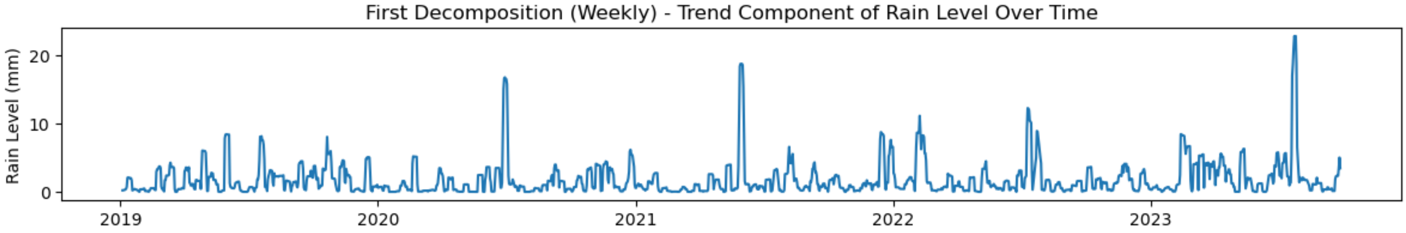
4. Trend, Seasonal, and Remainder Detection by Time Series Decomposition

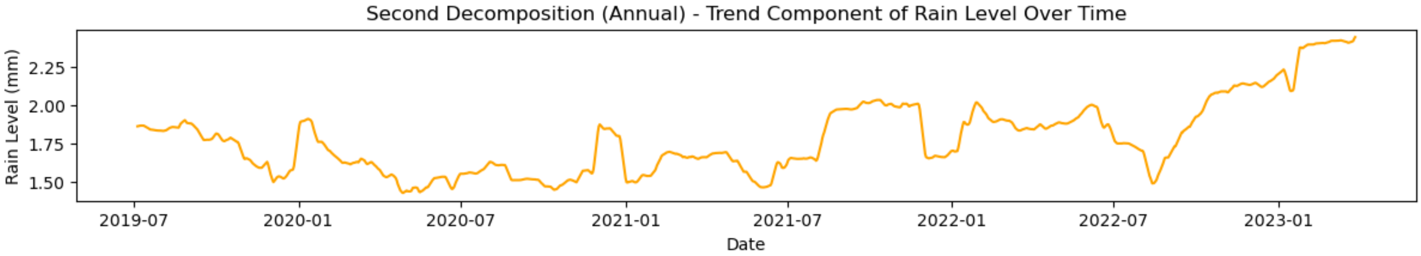
Time series decomposition is a statistical technique used to break down a time series into its constituent components, including trend, seasonal, and residual (or remainder). Time series decomposition is done to help improve understanding of the time series, but it can also be used to improve forecast accuracy.
There are two rounds of time series decomposition: the first decomposition from daily to weekly, and the second decomposition from weekly to yearly. The first try helps to identify patterns that repeat on a weekly basis. Focusing specifically on the trend and seasonal components obtained from the first try, this second try at a yearly level allows to investigate longer-term seasonality trends.

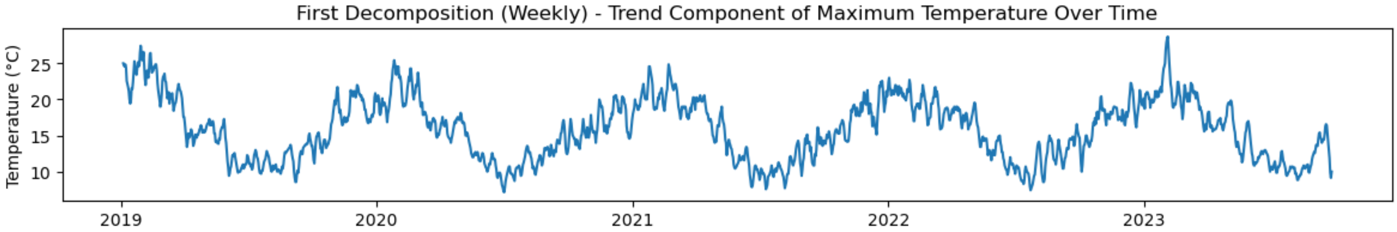
Trend Detection

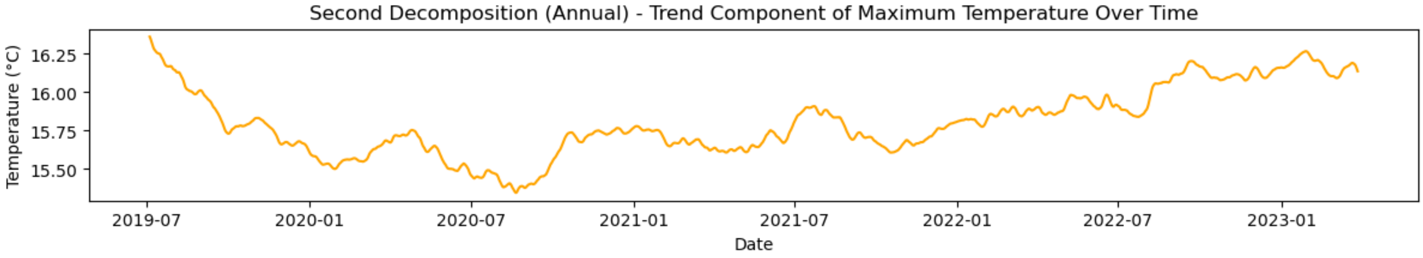


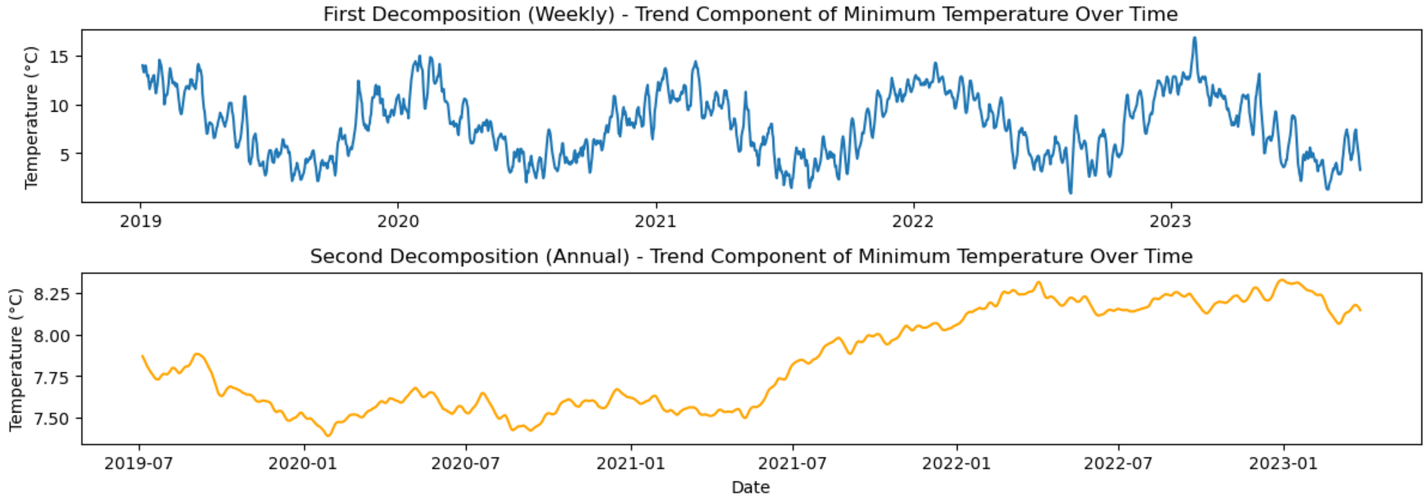












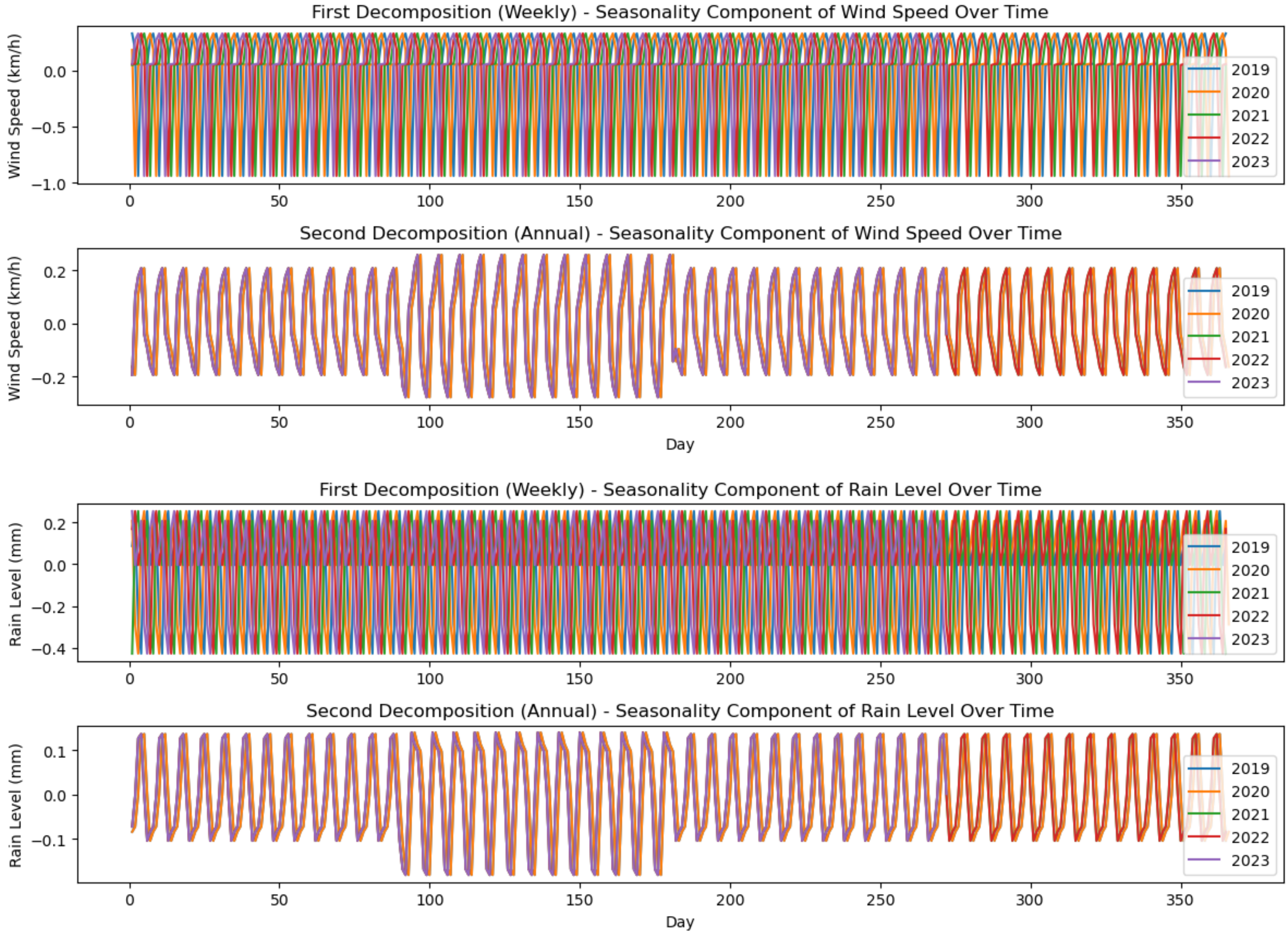
First Decomposition (Weekly) - Trend Component of Minimum Temperature Over Time

Second Decomposition (Annual) - Trend Component of Minimum Temperature Over Time

Figure 6: Trend Component of Wind Speed, Rainfall, Min Temperature and Max Temperature after the First and Second Decompostions

While the weekly decomposition of all four variables doesn't reveal an apparent trend, the annual decomposition paints a different picture. It highlights an upward trend in rainfall, suggesting a gradual increase over the years. In contrast, the wind speed and temperature show no discernible trend at the annual level.

Seasonal Detection



First Decomposition (Weekly) - Seasonality Component of Wind Speed Over Time

Second Decomposition (Annual) - Seasonality Component of Wind Speed Over Time

First Decomposition (Weekly) - Seasonality Component of Rain Level Over Time

Second Decomposition (Annual) - Seasonality Component of Rain Level Over Time

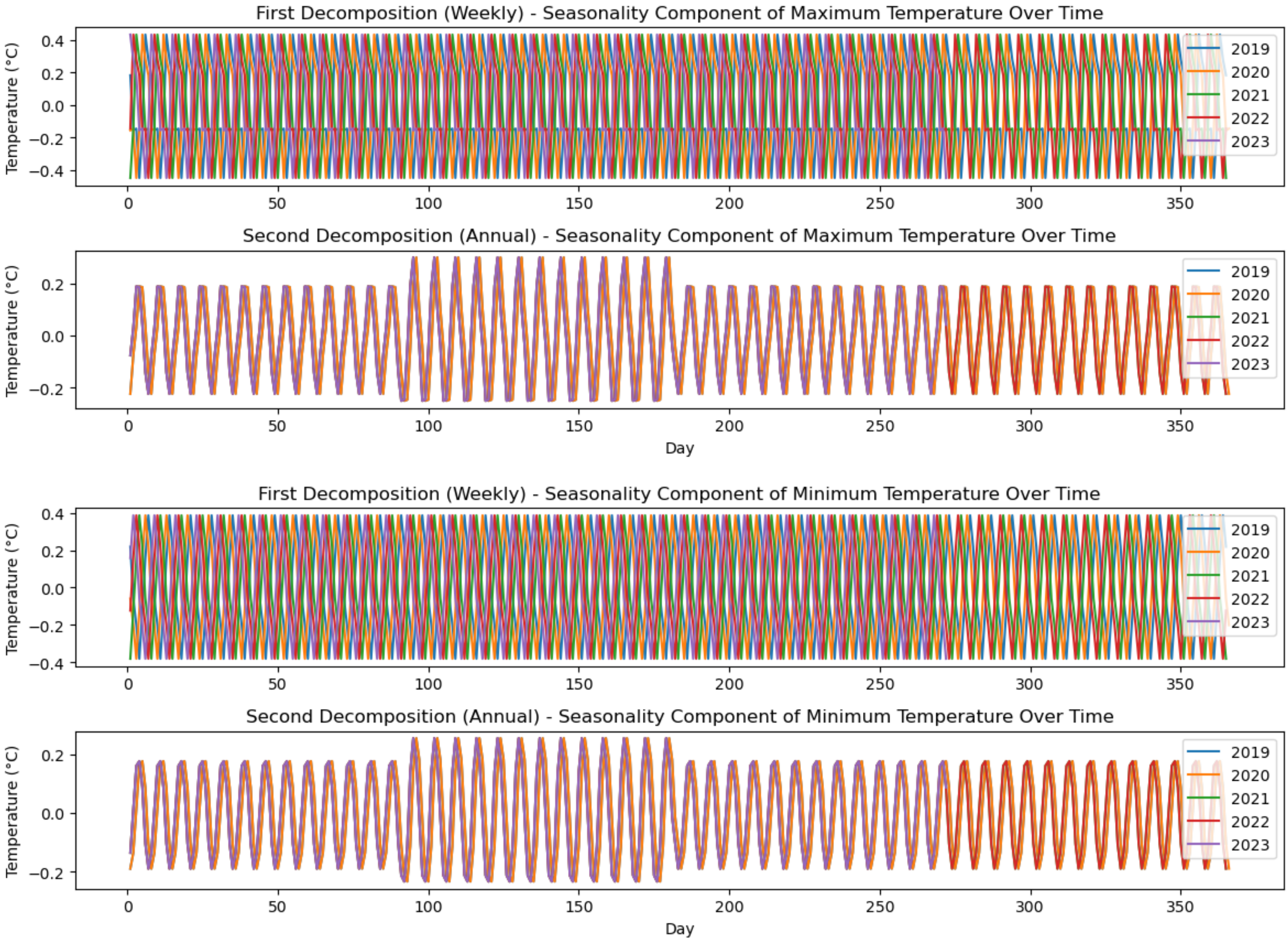
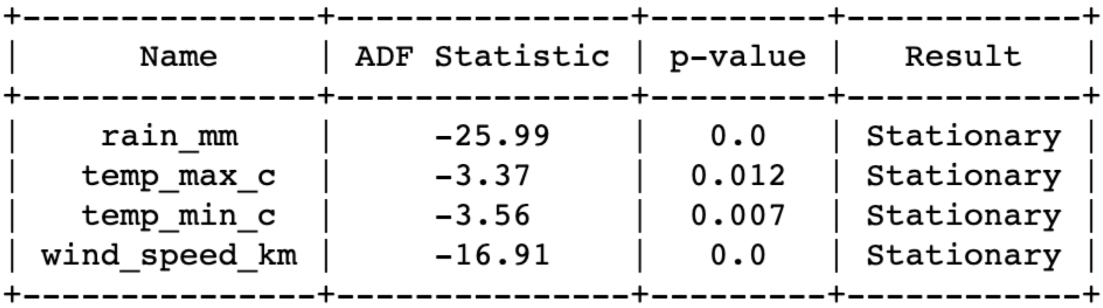


Figure 7: Seasonal Component of Wind Speed, Rainfall, Min Temperature and Max Temperature through 365 days over 5 years after the First and Second Decompositions

The weekly decomposition reveals a consistent pattern in the seasonal component across all four variables, repeating on a weekly basis. However, the annual decomposition provides a more detailed perspective, highlighting distinct patterns throughout the year. Specifically, the data exhibits consistent seasonal patterns in the first, third, and last quarters, suggesting similar shapes and trend during these periods. In contrast, the second quarter displays a comparable pattern but with higher volatility, introducing additional variability. This consistent beehavior across all four variables emphasizes the recurring nature of these patterns throughout the year, with a notable spike in volatility during the second quarter.

5. Stationarity

Stationarity means that the statistical properties of the process generating the data remain constant over time. This stability is vital for certain models and techniques, as non-stationarity can introduce bias into estimates and compromise the reliability of forecasts. The augmented Dickey-Fuller (ADF) test is used for assessing stationarity in time series. By testing the null hypothesis that a unit root is present (indicating non-stationarity), the ADF test helps determine whether a time series requires differencing to achieve stationarity. The alternative hypothesis suggests that the time series is already stationary.



```
+--------------+--------------+---------+------------+
|     Name     | ADF Statistic | p-value |   Result   |
+--------------+--------------+---------+------------+
|   rain_mm    |    -25.99    |   0.0   | Stationary |
|  temp_max_c  |    -3.37     |  0.012  | Stationary |
|  temp_min_c  |    -3.56     |  0.007  | Stationary |
| wind_speed_km |    -16.91    |   0.0   | Stationary |
+--------------+--------------+---------+------------+
```

Table 2: Results from the ADF tests for all four variables

The ADF tests are conducted on the original time series. The p-value of features are all below a significance level of 0.05 and that would lead to the rejection of the null hypothesis of a unit root, suggesting stationarity. This stationary behavior at the original level is beneficial for modeling and forecasting, as it simplifies the analysis and enhances the reliability of statistical inferences.

6. AR and MA terms of Wind Speed

While understanding the data, a preliminary examination of autoregressive (AR) and moving average (MA) terms provides insights into potential modeling strategies, guiding decisions in later phases. AR models predict time series data with a dependence on past values, suitable for noticeable trends. MA models predict data considering past error terms, beneficial for short-term fluctuations. Combining both and the seasonal component into a Seasonal Autoregressive Integrated Moving Average (SARIMA) model, with differencing for stationarity (if necessary) before applying the AR and MA components, ensures a comprehensive approach to capturing complex temporal patterns.

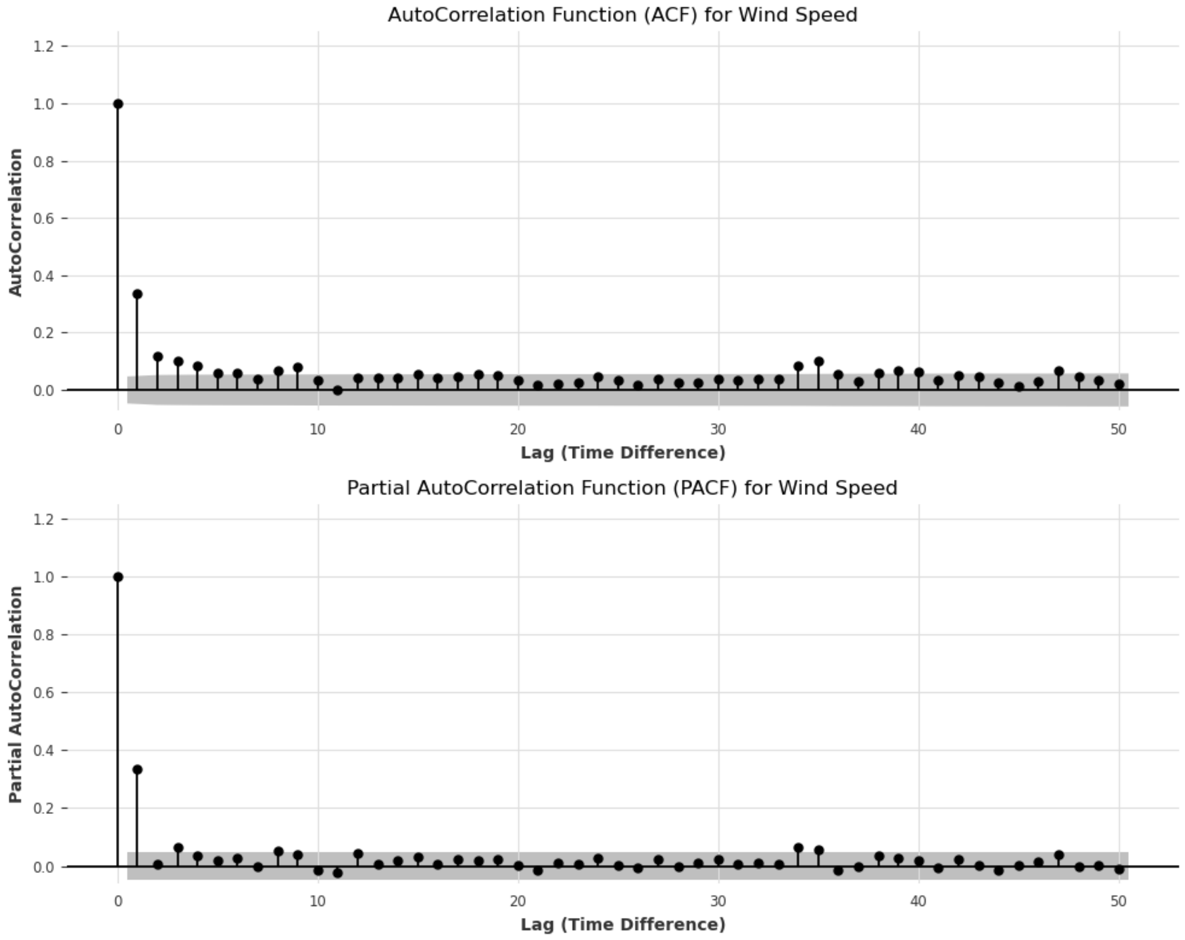


Figure 8: Autocorrelation and Partial Autocorrelation for Wind Speed

In this research, forecasting wind speed is the primary focus. Based on the analysis of the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots, it recommends using an AR model with p=1, reflecting the influence of the immediate past value, and aa MA model with q=1 or q=2, accounting for one or two past error terms.

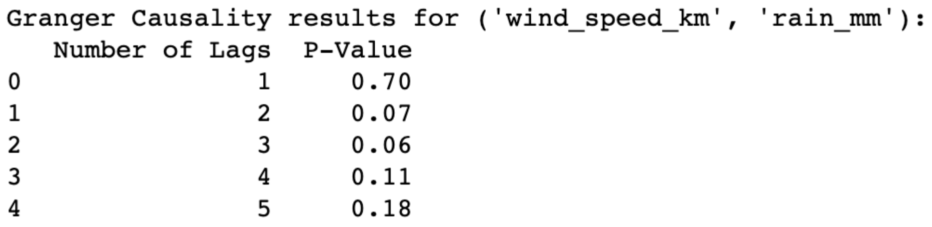
VI. Vector Autoregression (VAR) model

The data of 1733 days has been split into the train set and the test set with the last 60 days has been used to be the test set.

1. Granger Causality Test

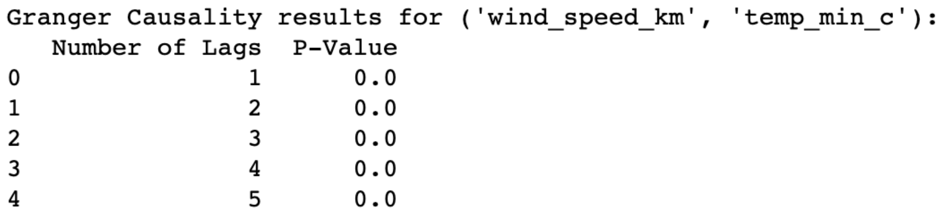
Granger Causality tests will be employed to examine whether rainfall has an association with wind speed and whether minimum temperature and maximum temperature have associations with wind speed as well. Grange causality means that past values of rainfall, and temperature have a statistically significant effect on the current value of wind speed, taking past values of wind speed into account as regressors.
Up to 5 lags will be investigated with the following hypotheses:

- Null hypothesis: The time series in another variable does not Granger cause the time series in wind speed. In other words, the coefficients corresponding to past values of the other variable are zero.
- Alternate hypothesis: The time series in another variable Granger causes the time series in wind speed



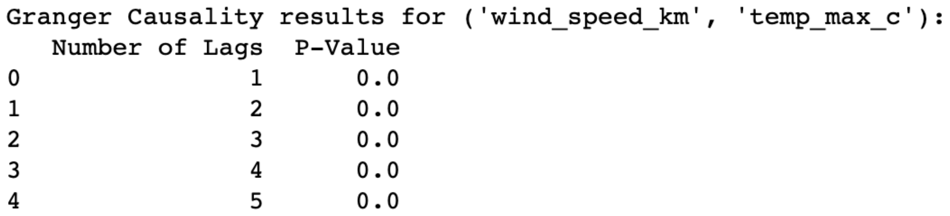
```
Granger Causality results for ('wind_speed_km', 'rain_mm'):
   Number of Lags  P-Value
0               1     0.70
1               2     0.07
2               3     0.06
3               4     0.11
4               5     0.18
```

Table 3: Granger Causality results between Wind Speed and Rainfall



```
Granger Causality results for ('wind_speed_km', 'temp_min_c'):
   Number of Lags  P-Value
0               1      0.0
1               2      0.0
2               3      0.0
3               4      0.0
4               5      0.0
```

Table 4: Granger Causality results between Wind Speed and Minimum Temperature



```
Granger Causality results for ('wind_speed_km', 'temp_max_c'):
   Number of Lags  P-Value
0               1      0.0
1               2      0.0
2               3      0.0
3               4      0.0
4               5      0.0
```

Table 5: Granger Causality results between Wind Speed and Maximum Temperature

The p-values for all lag orders of temperature are below the 0.05 significance level, providing sufficient evidence to reject the null hypothesis. This indicates the Granger causality from temperature to wind speed, suggesting that past minimum and maximum temperatures influence wind speeds. Conversely, for rainfall, the null hypothesis cannot be rejected, with p-values above 0.05, indicating the absence of Granger causality from rainfall to wind speed.
For accurate forecasting of wind speed using the VAR model, temperature should be included to prevent biased parameter estimates. While Granger causality is not established for rainfall, including it in the VAR model is still beneficial, as it maintains a relationship with wind speed, even though past values of rainfall lack predictive information for wind speed.

2. Grid Search for Lag Order

Having the multivariate time series being a stationary form and establishing a relationship between temperature, rainfall to wind speed, we now aim to select the most suitable VAR model.
To determine the optimal number of lags for the model, utilizing four scoring metrics: AIC (Akaike Information Criterion), BIC (Bayesian Information Criterion), FPE (Final Prediction Error), and HQIC (Hannan-Quinn Information Criterion). The optimal model will be characterized by the lowest scores across all these metrics. The investigation will encompass models ranging from 1 to 11 lags.

```
Lag Order     AIC     BIC         FPE   HQIC
        1   11.07  11.14*    64468.72  11.1*
        2   11.06   11.17    63595.58   11.1
        3   11.04   11.21     62529.9   11.1
        4   11.02   11.24    61201.29   11.1
        5    11.0   11.27    60154.95   11.1
        6    11.0   11.31    59744.82  11.11
        7    11.0   11.37    59905.87  11.14
        8    11.0   11.42    59911.03  11.15
        9    11.0   11.47    59730.88  11.17
       10  10.99*   11.51   59146.61*  11.18
       11   10.99   11.56    59339.56   11.2
```

Table 6: AIC, BIC, FPE, HQIC score for VAR(1) to VAR(11) Lag Order Selection (highlights the minimums)*

The analysis reveals that the optimal lag order varies based on different criteria and the candidates are 1 and 10. A lag order of 10 demonstrates the lowest AIC and FPE values, indicating a favorable balance between model fit and complexity according to these criteria. However, it’s important to note that the FPE criterion tends to be more sensitive to overfitting, leading to a lack of a clear minimum. On the other hand, the lowest BIC and HQIC values are observed for a lag order of 1. This value provides a favorable balance between model fit and complexity.

3. Modeling Vector Autoregression

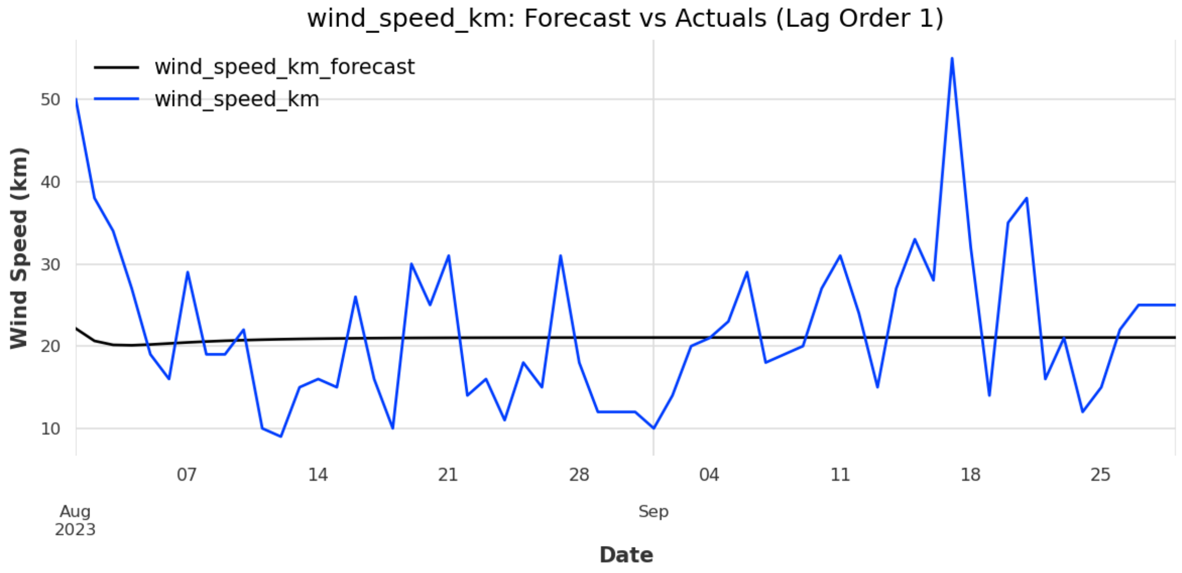


Figure 9: Predicted vs. Original values of Wind Speed for test data using Lag Order 1 over a 60-day Period

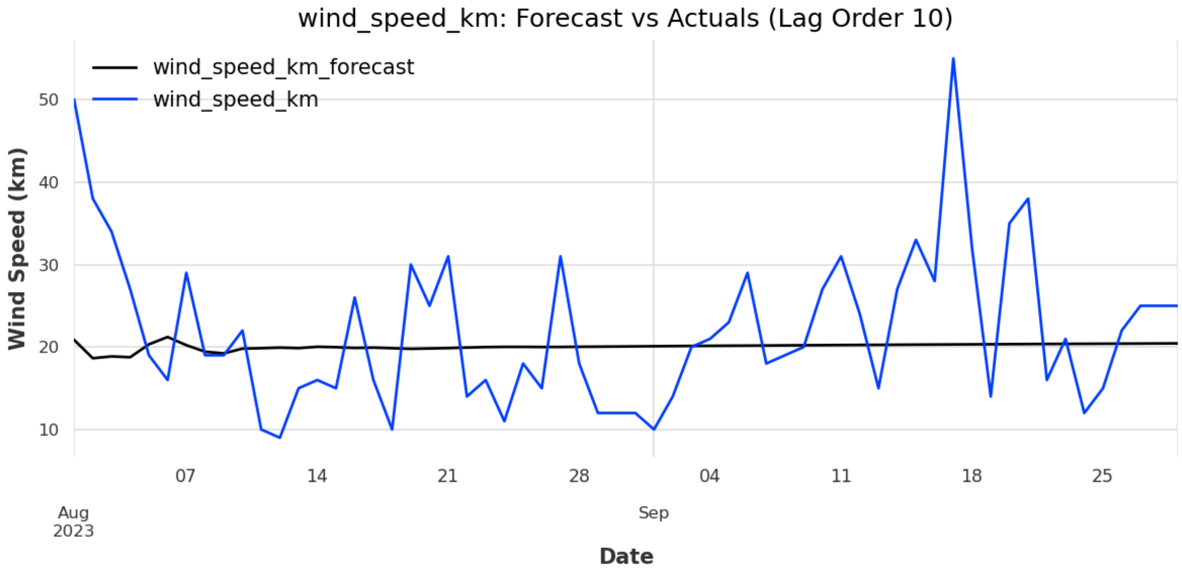


Figure 10: Predicted vs. Original values of Wind Speed for test data using Lag Order 10 over a 60-day Period

Both lag orders exhibit consistent patterns across two plots. While the original wind speed values vary significantly from 10 to 50, the predicted values remain relatively stable, hovering around 20 km/hour.

4. Evaluating the VAR(1) and VAR(10) models

In evaluating the performance of VAR(1) and VAR(10) models, we use specific metrics to assess the model's accuracy and effectiveness in capturing the underlying patterns and trends in the data.

- MAE (Mean Absolute Error): it gauges the average magnitude of errors between predicted and actual values. With all error-related metrics, a lower value indicates better accuracy.
- RMSE (Root Mean Square Error): it measures the average difference between predicted and actual values, taking into account the squared differences to emphasize larger errors.
- MAPE (Mean Absolute Percentage Error): it calculate the average percentage difference between the predicted values and the actual values.
- MASE (Mean Absolute Scaled Error): it measures the relative forecast accuracy compared to the naive or benchmark model, which refers to simply using the last observed value of the time series as the forecast for all future points.
- Forecast Bias: it measures the systematic deviation of forecast values from the actual values. A positive bias indicates overestimation, while a negative bias suggests underestimation. Smaller is better.
- Prediction Direction Accuracy (PDA): it measures the percentage of correct directional predictions made by the model. Direction refers to the movement or trend of future values, indicating whether they are predicted to increase, decrease, or remain unchanged. Larger is better.

```
+-----------+------+------+-------+------+---------------+-----------------------------------+
| Lag Order | MAE  | RMSE | MAPE  | MASE | Forecast Bias | Prediction Direction Accuracy (%) |
+-----------+------+------+-------+------+---------------+-----------------------------------+
|  Lag 10   | 7.33 | 9.65 | 35.04 | 1.0  |      2.07     |               50.85               |
|   Lag 1   | 7.33 | 9.46 | 36.67 | 1.0  |       1.2     |               52.54               |
+-----------+------+------+-------+------+---------------+-----------------------------------+
```

Table 7: Results of Performance Metrics for Lag Order 1 and 10

The results show that the VAR(1) model would be the preferred model. Although the VAR(10) gives a bit less error on prediction in terms of MAPE metric, the VAR(1) shows that it gives less error in terms of RMSE metric, less bias in prediction, and predicts the direction with more accuracy.

VII. Conclusions and Recommendations

Conclusions

The analysis aims to assess the effectiveness of VAR models in forecasting wind speed using its own historical data alongside variables like rainfall, minimum and maximum temperatures. The evaluation metrics applied to VAR(1) and VAR(10) models reveal that VAR(1) outperforms in both magnitude and direction of predictions.

However, overall VAR model performance appears suboptimal, attributed to challenges posed by the high volatility in wind speed data. Additionally, low correlation values between wind speed and other variables suggest a limited contribution to forecasting accuracy, with correlations falling below the 25% threshold.

Recommendations

Expand the dataset by collecting additional relevant time series data, especially from neighboring areas. Incorporating data on wind power generation from adjacent regions can provide valuable insights into the interconnected dynamics of wind patterns. This expanded dataset can be instrumental in training more robust multivariate models.

Collaborate with meteorological experts to gain domain-specific insights. Engage professionals with expertise in atmospheric sciences to better understand the intricacies of wind patterns and identify additional variables that could enhance the predictive capabilities of the multivariate models.

Investigate the application of advanced multivariate time series forecasting algorithms beyond Vector Autoregression (VAR). Algorithms such as Long Short-Term Memory (LSTM) networks or Gated Recurrent Units (GRU) in the realm of recurrent neural networks (RNNs) may capture intricate dependencies within the multivariate data more effectively.

Limitations

Obtaining the most recent and comprehensive historical data poses a challenge, especially when popular weather websites do not offer free access to up-to-date information. The lack of real-time and high-quality data may impact the accuracy and relevance of the forecasting models.

The analysis assumes that the observed patterns are solely a result of the inherent characteristics of the variables. External factors, such as weather anomalies or changes in environmental conditions, may introduce additional complexities not accounted for in the current study.

VIII. References

- Lütkepohl, H. (2013). Vector autoregressive models. Handbook of research methods and applications in empirical macroeconomics, 30.
- Liu, Y., Roberts, M. C., & Sioshansi, R. (2018). A vector autoregression weather model for electricity supply and demand modeling. Journal of Modern Power Systems and Clean Energy, 6(4), 763-776.
- Andrés, D. (2023, July 6). Performance Metrics for Time Series Forecasting - ML Pills. Machine Learning Pills. Retrieved December 3, 2023, from https://mlpills.dev/time-series/performance-metrics-for-time-series-forecasting/
- Andrés, D. (2023, June 22). Error Metrics for Time Series Forecasting - ML Pills. Machine Learning Pills. Retrieved December 3, 2023, from https://mlpills.dev/time-series/error-metrics-for-time-series-forecasting/
- Soman, S. S., Zareipour, H., Malik, O., & Mandal, P. (2010, September). A review of wind power and wind speed forecasting methods with different time horizons. In North American power symposium 2010 (pp. 1-8). IEEE.
- Lei, M., Shiyan, L., Chuanwen, J., Hongling, L., & Yan, Z. (2009). A review on the forecasting of wind speed and generated power. Renewable and sustainable energy reviews, 13(4), 915-920.
- Shang, Z., He, Z., Chen, Y., Chen, Y., & Xu, M. (2022). Short-term wind speed forecasting system based on multivariate time series and multi-objective optimization. Energy, 238, 122024.