COMP6235 Further Assessment coursework

Analysis of Climate Characteristics in London's Temperature Data

Yuhua Liu
32870639
University of Southampton
yl13u22@soton.ac.uk

ABSTRACT

The climate is becoming unstable and global warming is affecting people's daily lives. This article starts with the most intuitive temperature data, utilizing MongoDB to store a large amount of feature data. Based on analytical requirements, it filters and visualizes the changing trends in the data, and assesses the impact of extreme weather conditions. Utilizing historical meteorological data, analyze the temperature variations and precipitation patterns in different seasons in London. Explore potential climate trends. Such data analysis provides a tangible representation of urban climate data. It can be effectively integrated into city planning, future weather forecasting, the frequency of extreme weather events, and even further into areas such as energy usage and public health and safety. Assessing the potential impact of meteorological events on London's infrastructure, including transportation, water supply, and energy, enables the formulation of corresponding response strategies.

1 Introduction

Climate change and long-term temperature dependence are pressing issues that require global attention. A significant body of research indicates that the Earth's temperature is gradually rising.
The UK has broken high-temperature records, and Australia, as well as parts of Europe, has experienced devastating wildfires due to dry weather. According to predictions in the report, global temperatures may rise by 4 ° C or more by the end of this century[1]. In the UK, reports based on climate change predictions jointly released by the Tyndall Centre for Climate Change Research and the Hadley Climate Prediction Centre suggest that climate change could lead to phenomena such as flooding, droughts, and slope instability[2]. London, as an international metropolis, is naturally a part of global climate change. By observing temperature changes and other meteorological features in the London area, there can be a trend-setting indication for the weather in the UK and even the broader European context. This information allows scholars and the general public to gain a better understanding of the direction of global climate change.

2 Literature Review

As early as 2008, the UK Parliament passed the world's first Climate Change Act[3]. Since then, there has been a gradual increase in attention to climate change, widely recognized across various sectors. This marked a historic step in addressing climate-related issues.

Temperature change, may have interconnected effects and variations with many other meteorological features. Geert Lenderink's research indicates that contemporary climate is influenced by changes in daily average temperature and precipitation, and there is a close relationship between large-scale atmospheric circulation and temperature changes[4].

Climate change may also have a certain impact on energy production. The fluctuation in energy demand is, to some extent, driven by the seasonal variations in temperature[5]. Similarly, there are seasonal variations in wind energy. According to research and predictions, wind speeds in the UK follow a strong seasonal pattern. As a crucial component of renewable energy, wind power is significantly influenced by climate changes[6].

Predicting temperature changes is indeed a worthwhile area of study. Prior to the onset of an abnormal heatwave, the UK Meteorological Department issued warnings, and the UK Health Security Agency also released a Level 4 high-temperature health alert[7]. Jia, Xiaoyan, et al., utilized LSTM neural networks to predict temperatures in the East China Sea, observing regional temperature changes and seasonal trends[8]. Temperature forecasting has become an effective tool and method for monitoring climate change.

3 Methodology

3.1 data collection

The data selected for this article all comes from the Kaggle website and is entirely open source. Relevant data links are provided in the appendix. Three datasets were chosen, encompassing multiple feature dimensions. The primary city's historical meteorological data table contains **4,238,670** records over the past decade. The London historical climate data table consists of **15,341** records. The UK energy data table comprises **31** records. The London Historical Climate Data Sheet was used in the analysis of temperature, the Historical Meteorological Data Sheet for the last ten years for major cities was used in the analysis focusing on more climatic characterization and comparisons between cities, and the UK Energy Data Sheet was used in the analysis focusing on energy correlations.

3.2 database

During the experiment, it was discovered that the selected data volume was too large, and CSV text files were not suitable for displaying the data completely. It was not possible to directly perform data filtering and subsequent operations. The speed of querying and filtering data using Python libraries was not sufficient. Therefore, considering inserting the data into a database for more efficient querying and filtering.

The database selected for this article is MongoDB. MongoDB is a NoSQL database that falls under the non-relational database category, suitable for storing large-scale data. It also supports horizontal scaling to handle data of varying sizes. Research indicates that MongoDB has faster insertion speed and query efficiency compared to relational databases like MySQL. However, the complexity of MongoDB in the later stages is a point that could be improved in future research[9]. Data in MongoDB is stored in JSON format, making it compatible with most commonly used programming languages and providing a simpler, more efficient, and intuitive method for data interaction.

3.3 data pre-processing

Read the data using the *read_csv* tool in pandas, check for missing values, and fill the missing values with 0. In MongoDB, data is stored in a key-value pair dictionary format. Therefore, before executing the insertion statement, it is necessary to format the CSV file using the *to_dict()* method.

3.4 data analyse, relativity and prediction

Depending on the needs and changes in different features visualize trends in temperature as well as other information. Observe the distribution characteristics of the data from the macro level. Observe the indicators of the data from the microscopic point of view, set the threshold of extreme weather according to the percentile, and judge the number of days when extreme weather occurs and the corresponding characteristics.

There are outliers between two continuous variables in the data. A better method for assessing the linear correlation between the two variables is to use the Spearman correlation coefficient[10]. The Spearman correlation coefficient is less sensitive to specific numerical values and focuses more on the relative order of the data. The meteorological data selected in this article has continuity in terms of dates and is ordered. The Spearman correlation coefficient is better suited to uncover potential non-linear relationships in the data.

Another feature dimension of temperature data is time, an ordered and continuous feature. The prediction of temperature can provide an effective observation of future climate change and a more adequate solution to deal with the emergence of extreme weather. In practice, the use of LSTM deep learning model for prediction can effectively meet the demand. The forget gate unique to Long Short-Term Memory (LSTM) models can adaptively select crucial information. A prominent feature of temperature data is its strong cyclical variation, exhibiting not only a single trend but fluctuations around a time cycle. LSTMs introduce the concept of gates to determine the significance of information, providing excellent predictive capabilities for time series data[11].

3.5 data visualization

This article employed various visualization charts to present the data. Bar charts provided an intuitive comparison of numerical values for different categories. Histograms observed the distribution of data across different seasons, aiding in determining central tendencies. Radar charts illustrated comparisons between multiple dimensions in climate data. Scatter plots depicted the correlation and trends between two feature dimensions. Pie charts showcased the proportion of highlighted features within the overall dataset. Line charts delineated the changing trends and magnitudes of the data.

4 Discussion and Analyse

4.1 London Temperature Analysis

London is an internationally diverse metropolis with distinctive climate features and abundant meteorological resources. Over the past decade, the annual average temperature has shown an overall upward trend. In 2022, the average daily maximum temperature reached 17.28 degrees, marking a 16.36% increase compared to the 14.85 degrees recorded a decade ago in 2013. Furthermore, there was a notable 13.61% increase from 2021 to 2022, achieving the highest year-over-year growth in nearly a decade. When extending the timeline to 40 years, spanning from 1981 to 2020, it is observed that London's average temperature has increased by 20.74%. Additionally, within each decade, there is a consistent trend of higher average temperatures. This warming trend has significant implications for human life, as high temperatures and humidity are recognized factors contributing to mortality and are also associated with triggering violent events[12].

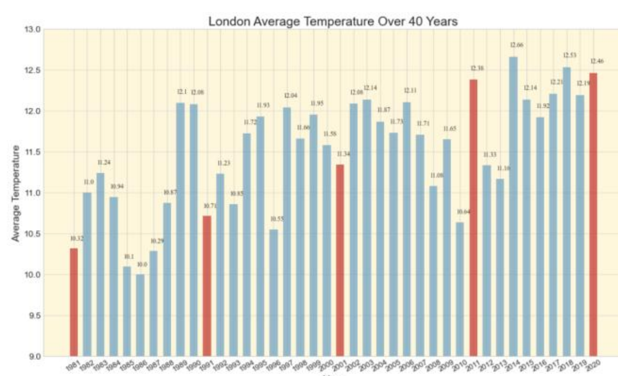

Figure 4.1.1: London Average Temperature 40 Years

Temperature data often exhibits strong seasonality, with temperatures following periodic patterns. Therefore, the data was aggregated and classified in the database based on the 'season' dimension. Upon observation, it was found that the temperatures in all three seasons roughly follow a normal distribution. However, summer exhibits a less pronounced skewness in its distribution. By comparing the line chart depicting the average temperature changes in different seasons over the past decade, it was observed that the temperature trends in summer do not show significant differences from other seasons, and in fact, it has a lower rate of increase. Over the past ten years, the average temperature in London during the summer increased by 7.36%, while spring, autumn, and winter experienced increases of 39.89%, 10.00%, and 15.57%, respectively. Therefore, there is consideration as to whether there are certain extreme weather events that have caused a shift in temperature distribution.

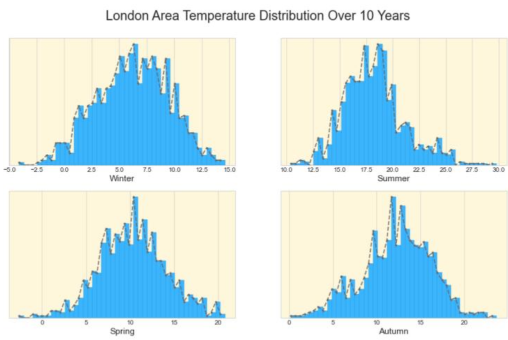

Figure 4.1.2: London Temperature Distribution by Seasons

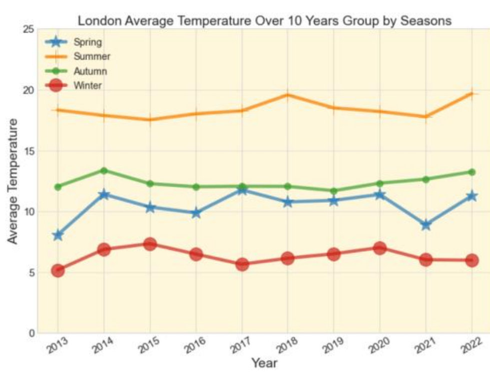

Figure 4.1.3: London Average Temperature by Seasons

This article employs a method based on percentile thresholds to determine extreme temperatures. In MongoDB, the percentile values for summer and winter temperatures are calculated as thresholds for extreme high and extreme low temperatures. The database is then queried to filter the number of days where temperatures meet the specified threshold conditions.

In practical applications, it is common to set the 95th percentile as the threshold for extreme hot weather and the 5th percentile as the threshold for extreme cold weather[13]. This approach takes into account the overall distribution of the data, providing a better understanding of the dynamic changes in temperature. A pie chart is used to visualize the proportions of extreme high and extreme low-temperature days throughout the year, allowing for an analysis of the changes in extreme weather over the past decade.

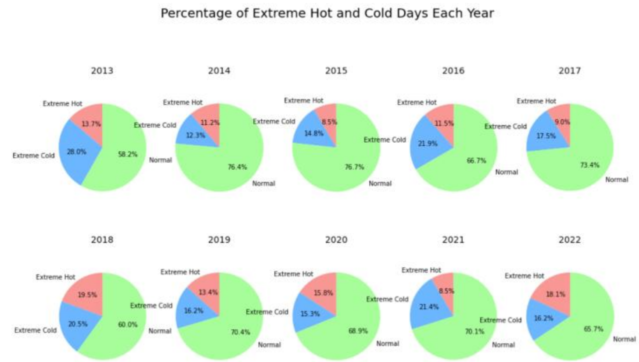

Figure 4.1.4: London Extreme Weather Percentage

In the pie chart, it is observed that the proportion of extreme weather in the last five years has significantly increased compared to the previous five years. In the past year, 34.3% of the time required enduring the impact of extreme weather, with extreme high-temperature days accounting for 18.1%, the second-highest proportion in the past decade. Besides extreme high temperatures, the phenomenon of extreme low temperatures is also noteworthy. It can be observed that the proportion of data below the 5th percentile in the dataset, representing extreme cold weather, often surpasses the proportion of extreme high-temperature days.

Examining the change in the number of extreme weather days between adjacent years, while using the annual average temperature as a reference, provides an intuitive sense of whether there is a correlation.

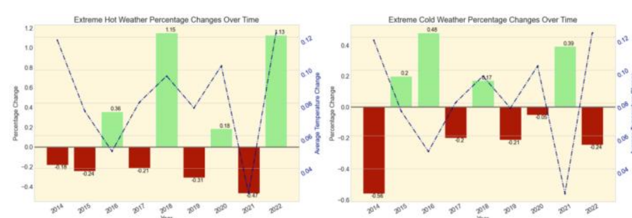

Figure 4.1.5: The Rate of Change in Extreme Weather

Intuitive data analysis reveals that, over the past decade, in nine out of ten years, the trend in extreme high-temperature weather aligns with the overall trend in average annual temperature. However, the trend in extreme low-temperature weather aligns with the average temperature trend in only two of the ten years. Extreme low temperatures and extreme high temperatures share the same growth trend in five out of the ten years. When predicting extreme high-temperature weather in the future, it may be beneficial to consider the changes in average temperature as a reference. However, numerical analysis indicates that extreme high-temperature weather exhibits significant fluctuations, and the weather's instability may pose challenges to temperature predictions and extreme weather contingency plans. In the case of predicting extreme low-temperature weather, additional auxiliary features may be necessary for more accurate forecasts.

4.2 London Temperature Prediction

This article chooses to use the Long Short-Term Memory (LSTM) network model for temperature prediction. Historical average temperature data for London over the past forty years is extracted from the database, with a time step set to 10. The data is split, with 80% used as the training set and 20% as the testing set. The model's prediction accuracy and other metrics are evaluated on the testing set. In assessing the test results, the mean_absolute_error and mean_squared_error metrics from the sklearn library are employed. Additionally, Root Mean Squared Error (RMSE) is calculated, representing the square root of MSE, which reduces the impact of larger errors on the parameters. Accuracy is also calculated to provide an intuitive representation of the model's performance on the test set. MAE is less sensitive to outliers, MSE focuses more on the impact of larger errors, and RMSE moderates the influence of larger errors on the parameters. Accuracy directly showcases the model's performance in predicting the test set.

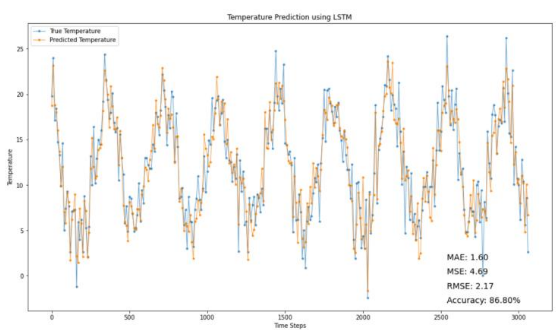

Figure 4.2.1: LSTM Prediction

The prediction results indicate that, given a substantial amount of historical data, temperature prediction is a feasible method for assessing temperature trends. It exhibits good capturing ability for the cyclical changes and correlations in temperature. In practical applications, it can provide value for temperature warnings and environmental solutions related to temperature.

Based on the previous analysis, extreme weather is also an important dimension in observing temperature data. Predicting the number of extreme weather days can provide valuable insights into future temperature conditions. The UK Met Office, for instance, issues nationwide heatwave alerts when daytime temperatures exceed 30°C for two consecutive days, with nighttime temperatures exceeding 15°C, as observed during the monitoring period from June to September 2022[14]. Therefore, this article selects 30°C as the threshold for extreme high-temperature prediction and calculates the number of hot days for each year over the past forty years to make predictions.

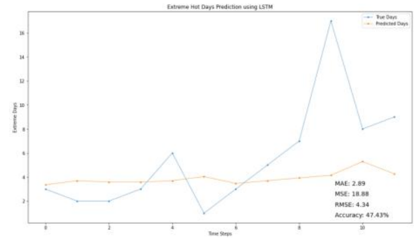

Figure 4.2.2: LSTM Prediction on Extreme Hot Weather

In the face of small-sample data, the model's prediction accuracy and various metrics show mediocre performance, compounded by the inherent randomness in the occurrence of extreme weather. The observed fluctuation in the number of days over a decade is considerable. In future work, expanding the training dataset and adjusting model parameters can be considered to achieve more accurate predictions.

4.3 The Correlation of Climate Indicators

The global climate is a complex system, and while temperature is a crucial meteorological indicator, other metrics such as wind speed and rainfall are also important components that directly or indirectly influence the world's climate. For instance, the temperature difference in air masses leads to pressure differences, forming winds. Changes in temperature gradients can result in deviations in wind speed from normal values[15]. Therefore, this article selects temperature data, wind speed data, and rainfall data from the past decade in London from the database for analysis. The correlation among these three features is assessed using the Spearman correlation coefficient.

Based on the calculation results of the Spearman correlation coefficient, shown in Table4-3-1, no strong individual correlations were found among the three variables. However, during periods of drastic climate change, the interactive effects may become more pronounced. Therefore, considering the perspective of extreme temperatures, we assess whether there is a stronger correlation. To do this, extreme temperature data is filtered from the database, and scatter plots are observed to determine if there is a stronger correlation among meteorological variables during extreme temperatures.

Table 4-3-1 Spearman Correlation Result

Spearman Correlation	Temperature_c	Precipitation_mm	Wind_speed_kmh
Temperature_c	1.000	-0.1511	-0.0173
Precipitation_mm	-0.1511	1.000	0.2294
Wind_speed_kmh	-0.0173	0.2294	1.000

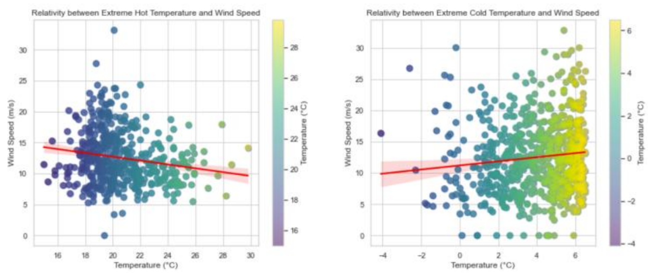

Figure 4.3.1: Correlation between Temperature and Wind

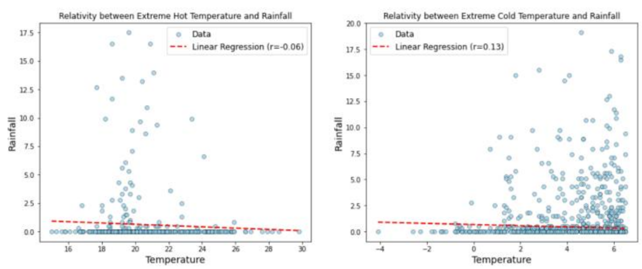

Figure 4.3.2: Correlation between Temperature and Rainfall

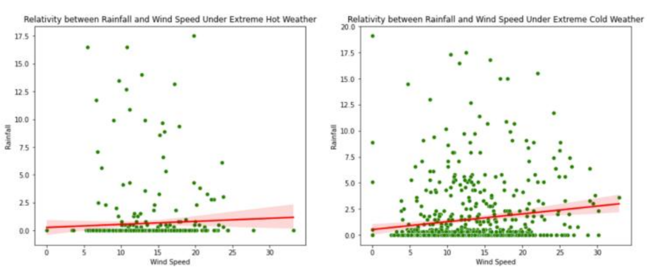

Figure 4.3.3: Correlation between Rainfall and Wind

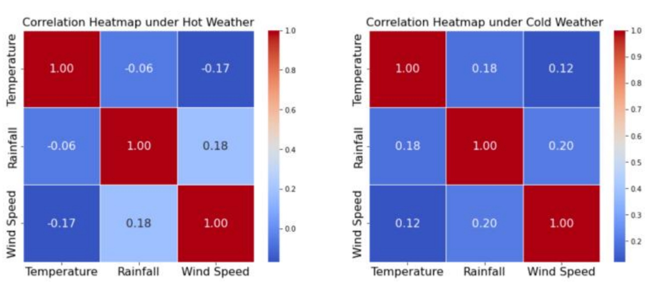

Figure 4.3.4: Correlation Heatmap under Extreme Weather

Upon observing the scatter plots, the most distinctive contrast is found between temperature data and wind speed data. Although the correlation is not high, under high and low-temperature conditions, they exhibit completely opposite correlations. This trend allows researchers to estimate the rise and fall of wind speed while observing temperature.

The trend of climate features not only directly influences the weather but also has an indirect impact on the usage and consumption of energy. It is precisely the global climate change that has led to a shift in the direction of energy development. The proportion of high-energy-consumption and high-pollution energy development is decreasing, while the research and utilization of renewable energy are on the rise.

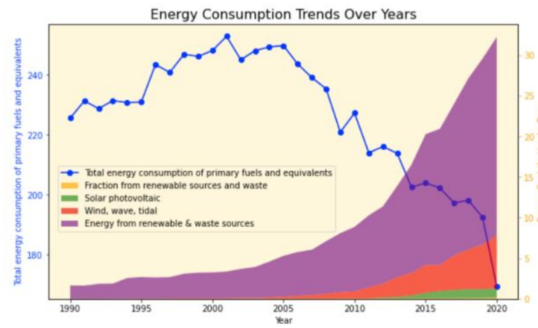

Figure 4.3.5: Energy Consumption Trends

From the charts, it can be observed that the usage proportion of renewable energy has been steadily increasing. Over the decade from 2011 to 2020, the usage of renewable energy has grown by 276.64%, which is closely related to the overall upward trend in global temperatures over the past decade. The increase in temperature leads to changes in atmospheric pressure, affecting wind speed and resulting in an increase in the utilization of wind energy resources. Similarly, the increase in sunlight hours contributes to the rise in solar energy resource utilization. What is particularly encouraging is the rapid decline in the consumption of primary fuels and equivalents, reaching 20.81%. This is a positive signal, indicating improved energy efficiency with the progress of renewable energy development, leading to a decrease in the use of primary combustion materials and rapid advancement in low-carbon technologies.

4.4 Comparison

Comparing the three major cities in the UK, it can be observed that the overall temperature in London is higher than the other two cities, while the wind speed is the lowest among the three. Analyzing the geographical features, London has a latitude of 51.51 degrees N, Edinburgh is at 55.95 degrees N, and Belfast is at 54.60 degrees N. It can be inferred that the average temperature is correlated with latitude.

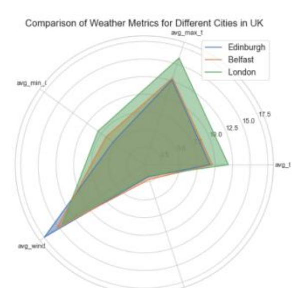

Figure 4.4.1: Comparison in Three Main Cities in UK

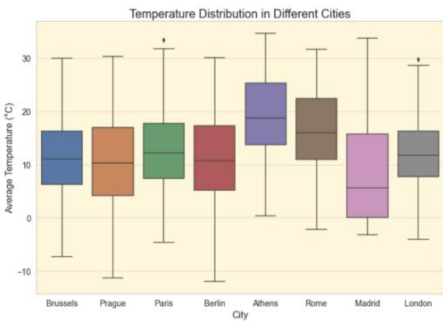

Figure 4.4.2: Comparison of Several Major Cities in EU

Observing the box plots of several major cities in Europe over the past decade, it is noticed that London's box plot is shorter, indicating a more evenly distributed temperature with less variation. The average maximum temperature in London is also among the lowest compared to other major cities. However, both London and Paris have outliers, indicating the need to be cautious about the occurrence of extreme high temperatures. The median of London's average temperature is similar to Brussels and Paris, remaining above 10°C. In contrast to Athens and Paris with higher temperatures, the UK does not have to endure prolonged periods of high temperatures during the summer.

5 Future Work

The project has some limitations. In the predictive functionality, increasing the amount of data and optimizing the complexity of parameters and models would be beneficial, especially in the section predicting extreme temperatures. The model did not capture trends well when faced with drastic fluctuations in the data, and adjustments are needed in future work. In the comparative analysis with other cities, additional dimensions such as rainfall, distance to the sea, atmospheric pressure, and more could be included for a more comprehensive comparison. In terms of visualization, using more interactive and advanced charts for dynamic displays could be considered in future developments.

6 Conclusion

This study analyzes the climate change trends in the London and global regions over different time periods by acquiring temperature and other meteorological feature data. The findings indicate a rapid increase in temperatures over the past decade, with a 20.74% rise in London's average temperature over the forty-year period. Temperature exhibits seasonal cyclic patterns, and the analysis observes temperature changes and distribution across different seasons. The frequency of extreme weather events has increased in the last five years, highlighting the need for measures to address such extremes. LSTM models are employed for temperature and extreme weather predictions, demonstrating good accuracy in temperature forecasts but leaving room for improvement in predicting extreme weather occurrences. The study also explores correlations between various meteorological indicators, considering their associations during extreme weather conditions. Additionally, it examines the correlation between temperature trends and the use of renewable energy sources, revealing a positive trend in clean energy utilization. Future work could further explore data correlations, monitor global temperature trends, and encourage the adoption of clean energy sources to address climate change concerns.

REFERENCES

[1] "Effects of Climate Change." Met Office, 2022, https://www.metoffice.gov.uk/weather/climate-change/effects-of-climate-change.

[2] Hulme, Michael. Climate change scenarios for the United Kingdom: the UKCIP02 scientific report. Tyndall Centre for Climate Mental Sciences University, 2002.

[3] Lockwood, Matthew. "The political sustainability of climate policy: The case of the UK Climate Change Act." Global Environmental Change 23.5 (2013): 1339-1348.

[4] Lenderink, Geert, and Erik Van Meijgaard. "Increase in hourly precipitation extremes beyond expectations from temperature changes." Nature Geoscience 1.8 (2008): 511-514.

[5] Hunt, Lester C., Guy Judge, and Yasushi Ninomiya. "Underlying trends and seasonality in UK energy demand: a sectoral analysis." Energy Economics 25.1 (2003): 93-118.

[6] Cradden, Lucy C., Gareth P. Harrison, and John P. Chick. "Will climate change impact on wind power development in the UK?." Climatic change 115 (2012): 837-852.

[7] Zachariah, Mariam, et al. "Without human-caused climate change temperatures of 40 C in the UK would have been extremely unlikely." (2022).

[8] Jia, Xiaoyan, et al. "Prediction of sea surface temperature in the East China Sea based on LSTM neural network." Remote Sensing 14.14 (2022): 3300.

[9] Wei-Ping, Zhu, L. I. Ming-Xin, and Chen Huan. "Using MongoDB to implement textbook management system instead of MySQL." 2011 IEEE 3rd International Conference on Communication Software and Networks. IEEE, 2011.

[10] Khamis, Harry. "Measures of association: How to choose?." Journal of Diagnostic Medical Sonography 24.3 (2008): 155-162.

[11] Wang, Jian Qi, Yu Du, and Jing Wang. "LSTM based long-term energy consumption prediction with periodicity." Energy 197 (2020): 117197.

[12] Cohn, Ellen G. "Weather and crime." The British Journal of Criminology 30.1 (1990): 51-64.

[13] Kephart, Josiah L., et al. "City-level impact of extreme temperatures and mortality in Latin America." Nature medicine 28.8 (2022): 1700-1705.

[14] "London Cool Spaces Summer 2022," Mayor of London 2022, https://www.london.gov.uk/what-we-do/environment/climate-change/climate-adaptation/cool-spaces.

[15] How Does Cold Weather Affect Wind Speed? January 2019. https://www.windlogger.com/blogs/news/how-does-cold-weather-affect-wind-speed

GitHub Repository:

https://github.com/YuhuaLiu929/COMP6235-Further-Assessment