Evaluating the Impact of Ultra Low Emission Zone Policies on Air Quality and Public Transportation Choices in Central London

Xiaoyi Chen

CASA0010: MSc Dissertation

Supervisor: Simon Doyle

GitHub Repository

Word Count: 11,805

This dissertation is submitted in part requirement for the MSc in the Centre for Advanced Spatial Analysis, Bartlett Faculty of the Build Environment, UCL.

28th of August 2024

Abstract

This study evaluates the impact of the Ultra-Low Emission Zone (ULEZ) policy on air quality and public transportation in Central London from 2019 to 2024. Utilizing data on Nitric Oxide (NO), Nitrogen Dioxide (NO2), PM2.5, and PM10 levels, along with traffic flow and public transportation statistics, the analysis employs both spatial interpolation and direct observation methods, specifically Kriging, to assess changes in air quality. Kriging is used to estimate weighted average air quality by leveraging the geographical coordinates of sensors and corresponding natural climate conditions, such as wind direction and speed, at specific times. The research explores changes in air quality across multiple stages of ULEZ implementation, highlighting the policy's effectiveness in reducing key pollutants and influencing public transport patterns. Results reveal significant improvements in air quality, particularly in the later stages of ULEZ expansion, with notable reductions in NO and NO2 levels, especially in regions such as the City of London, Camden, and Kensington and Chelsea. This study provides critical insights into the effectiveness of targeted environmental policies in mitigating urban air pollution and promoting public health, offering valuable guidance for future urban planning and policy-making.

Acknowledge

I would like to express my sincere gratitude to my supervisor, Simon Doyle, for his invaluable guidance and support throughout the development of my dissertation. His insightful suggestions on the thesis topic, structural framework, data collection methods, and the content of the results and conclusions have been instrumental in shaping my research.

I would like to deeply thankful to my personal tutor, Claire Dooley, for her expert advice on time management and planning. Her guidance has been crucial in helping me stay organized and meet my deadlines.

Additionally, I acknowledge the use of ChatGPT 4.0 (OpenAI, https://chat.openai.com) for summarizing my initial notes, proofreading my final draft, as well as Copilot (GitHub, https://github.com/features/copilot) for helping to organize and implement repetitive coding methods.

Declaration of Authorship

I, Xiaoyi Chen, hereby declare that this dissertation is all my own original work and that all sources have been acknowledged. It is 11,805 words in length.

Xiaoyi Chen

Xiaoyi Chen, 28th of August 2024

Contents

1 Introduction 7
2 Literature Review 7
2.1 Air Quality and Urban Pollution 7
2.1.1 Urban Air Quality Influencing Factors 7
2.1.2 ULEZ and Similar Policies 8
2.2 Data Aggregation and Spatial Interpolation Methods 9
2.2.1 IDW and Kriging 9
2.2.2 Meteorological Effect on Air Spread 10
2.3 Research Question 10
2.3.1 Time Period 10
2.3.2 Study Area 11
2.3.3 Research Question 12
3 Methodology 13
3.1 Data Sources 13
3.1.1 Data Collection 13
3.1.2 Data Feasibility Analysis 16
3.2 Data Processing 19
3.2.1 Missing value interpolation processing 19
3.2.2 Kriging Weights Calculation 20
4 Result 24
4.1 Exploratory Data Analysis(EDA) 24
4.1.1 Air pollutants EDA 24
4.1.2 Road traffic EDA 28
4.2 Time Series Analysis 28
4.2.1 Air Quality Trend Analysis on ULEZ changes 28
4.2.2 Difference Between Periods 36

4.3 Relationship among Air Pollution and Traffic Types 41
4.3.1 Correlation Analysis 41
4.3.2 Regression Analysis 44
5 Discussion 45
5.1 Critical Evaluation of Methodology 45
5.2 Limitations of the Study 46
5.2.1 Acknowledgment of Constraints 46
5.2.2 Potential Biases 46
5.2.3 Considering Additional Influencing Factors 46
5.4 Future Work 47
5.4.1 Further analysis 47
5.4.2 Further Investigation 48
6 Conclusion 49
Reference 51
Appendix 56
Appendix A: Tables 56
Appendix B: Graghs 59
Appendix C: Important Code 63
C.1 Interpolation Method 63
C.2 Kriging weighed air quality calculation 65
Appendix D: Research Log 68

1 Introduction

Urban air quality is a critical environmental and public health issue, particularly in densely populated cities like London. The adverse effects of air pollution, primarily from vehicular emissions, include respiratory and cardiovascular diseases, premature deaths, and overall diminished quality of life. To address these concerns, Great London has implemented various policies, including the Ultra-Low Emission Zone (ULEZ)(Greater London Authority, 2023, Transport for London, 2019), aimed at reducing pollution levels and encouraging the use of cleaner transport options.

This study seeks to evaluate the impact of the ULEZ policy(Mayer of London, 2023) on air quality in central London, focusing on key pollutants such as PM2.5, PM10, and NO2. By analyzing historical air quality data and changes in public transport usage, this research aims to provide insights into the effectiveness of ULEZ and its implications for urban policy and planning. The findings will contribute to a better understanding of how targeted environmental policies can mitigate air pollution and promote sustainable urban living.

2 Literature Review

2.1 Air Quality and Urban Pollution

Urban air pollution is a critical issue affecting public health and environmental quality worldwide. Numerous studies have highlighted the primary sources of urban air pollution, including industrial emissions, residential heating, and, significantly, vehicular traffic. According to the World Health Organization, fine particulate matter (PM2.5), nitrogen dioxide (NO2), and ozone (O3) are the most concerning pollutants due to their adverse health impacts, including respiratory and cardiovascular diseases(World Health Organization, 2021b).

The EEA (European Environment Agency, 2013) reports that road transport is one of the major contributors to urban air pollution, particularly in terms of NO2 and particulate matter emissions. In dense urban environments, the contribution of vehicles to air pollution can be substantial, often exceeding 50% of total emissions.

Several studies have confirmed the direct impact of vehicular emissions on urban air quality. Research conducted by the International Council on Clean Transportation (ICCT) (Susan Anenberg, 2019) shows that traffic-related emissions are a significant source of NO2 and PM2.5 in cities. This pollution is exacerbated by high traffic volumes and congestion, leading to increased exposure levels in urban populations.

2.1.1 Urban Air Quality Influencing Factors

Urban air quality is influenced by a myriad of factors, ranging from vehicular traffic to industrial emissions and natural elements like wind patterns. Traffic-related air pollution (TRAP) is particularly significant in urban settings, where the concentration of vehicles contributes heavily to pollutants such as PM2.5 and NO2. Khreis et al. (2020) highlight that TRAP is a primary contributor to respiratory and cardiovascular diseases, underscoring the critical need for policies like ULEZ. Moreover, the complex interplay between vehicular emissions, urban infrastructure, and environmental conditions makes it challenging to model and predict pollution levels accurately.

In the study conducted by Nathvani et al. (2023), a permutation importance analysis was performed on data from Accra, Ghana, to assess the factors contributing to noise and PM2.5 levels. The results, depicted in the accompanying figure, show that the most significant contributors to environmental noise and PM2.5 concentrations are traffic-related factors, including 'Cars,' 'Taxis,' 'Trucks,' and 'People,' as well as 'Tro-tros' (a form of mini buses used for public transportation in Ghana). These elements consistently rank high in their contribution to pollution, underscoring the impact of vehicular activity on air quality in urban environments.

Similarly, in the study by Danesh Yazdi et al. (2020), a gradient boosting machine (GBM) was employed to determine variable importance based on SHAP (Shapley Additive Explanations) values. The analysis, focused on London, not only identified 'Daily traffic counts,' 'Number of bus stops,' and 'Distance to nearest bus stops' as significant traffic-related factors affecting air quality but also highlighted the importance of meteorological conditions. Specifically, variables such as 'Average wind speed,' 'Wind direction,' and 'Light at night' were found to have the highest importance, indicating that weather and environmental factors play a more dominant role in influencing PM2.5 levels in London compared to purely traffic-related factors.

These findings from both studies highlight the critical interplay between traffic and meteorological factors in determining urban air quality, with the specific influences varying by region.

2.1.2 ULEZ and Similar Policies

The Ultra Low Emission Zone in London, implemented in stages starting in April 2019, is designed to reduce vehicular emissions by imposing strict standards on vehicles entering the zone. Similar low emission zones (LEZs)(Urban Access Regulations in Europe, 2012) have been established in various cities across Europe, including Berlin, Milan, and Stockholm, each showing varying degrees of success in improving air quality.

A study by Holman, Harrison, and Querol (Holman et al., 2015) compared the effectiveness of different LEZs across Europe, concluding that such zones generally lead to reductions in NO2 and PM concentrations, though the extent of these improvements varies based on local conditions and enforcement mechanisms. In Berlin, for example, the implementation of LEZs led to a 9% reduction in PM10 and a 5% reduction in NO2 levels within the first year.

In London, the introduction of the ULEZ has shown promising results(Tom Edwards, 2019, Victoria Jones, 2019). Early assessments indicated a 29% reduction in roadside NO2 concentrations within the central zone shortly after its implementation in April 2019. These findings support the hypothesis that stringent vehicular emission controls can significantly improve urban air quality.

2.2 Data Aggregation and Spatial Interpolation Methods

2.2.1 IDW and Kriging

Accurate assessment of air quality relies on robust data collection and analysis techniques. Spatial interpolation methods were revealed by Wong et al. (2004), such as Inverse Distance Weighting (IDW) and Kriging, are commonly used to estimate air quality levels a10/15/2024cross different regions based on data from monitoring stations.

IDW is a straightforward method where the influence of a data point decreases with distance from the point of interest. It has been widely used in environmental studies to create continuous surface maps of pollutant concentrations. For example, Chen and Lin (2022) propose an enhanced variant called Clustering-based Inverse Distance Weighting (CIDW) to improve PM2.5 exposure estimates using dense networks of low-cost microsensors in Taiwan. However, the feasibility of applying CIDW in London is uncertain due to differences in sensor density, spatial heterogeneity, and environmental conditions. London's current sensor network may not be as dense as Taiwan's, potentially limiting CIDW's effectiveness. Moreover, London's unique urban landscape and regulatory context may require modifications to the method to account for local air quality challenges and data integration issues.

Kriging(Chilès and Desassis, 2018, Papritz and Stein, 2002, Stein, 1999), a more advanced geostatistical method, not only considers the distance between data points but also the spatial correlation between them. It provides more accurate estimates by modeling the spatial structure of the data. A study by Chung et al. (2019) explored the use of Kriging and conditional simulation to supplement missing groundwater-level data, finding Kriging effective for capturing spatial correlations even with significant data gaps. Further supporting the efficacy of Kriging, Shukla et al. (2020) applied Kriging and Inverse Distance Weighting (IDW) techniques to predict PM2.5 concentrations at unmonitored locations in Delhi, aiming to improve air pollution mapping. The study involves two phases, predicting concentrations at specific sites and during winter months, with an average error of 22% for Kriging and 24% for IDW. Similarly, Bayraktar and Turalioglu (2005) utilized this method to determine suitable sampling sites for air quality monitoring in metropolitan areas, highlighting its ability to reduce sampling and analysis costs while providing reliable city-wide air quality estimates.

Kriging's robustness in various environmental contexts, including air quality assessment and groundwater monitoring, is well-documented. In urban areas like Central London, Kriging excels due to its ability to handle spatial heterogeneity. In London, where pollution levels vary significantly over short distances because of factors like traffic patterns, building density, and green spaces, Kriging's spatial interpolation capabilities allow for more accurate estimations by incorporating surrounding data points. The dense network of sensors in London provides a rich dataset, enhancing the reliability of the interpolated surfaces. Studies by Beauchamp et al. (2018) and Van Zoest et al. (2020) have demonstrated Kriging's superior performance in urban air quality assessments compared to simpler methods like inverse distance weighting (IDW).

Kriging's ability to incorporate both spatial and temporal variability, along with adjustments for local environmental factors such as wind speed and direction, ensures precise air quality estimates. Incorporating wind data into Kriging allows for adjustments based on pollutant dispersal patterns, leading to more accurate representations at unsampled locations. This flexibility makes Kriging a powerful tool for policymakers, as shown in studies by Contreras and Ferri (2016) , which highlight its improved accuracy in urban air quality modeling and its applicability in policy-making.

2.2.2 Meteorological Effect on Air Spread

To accurately estimate air quality in urban environments, considering climatic conditions such as wind direction and speed is crucial. These factors influence the dispersion and concentration of pollutants. Existing studies have utilized meteorological data to enhance the accuracy of spatial interpolation methods for air quality estimation, helping in understanding how pollutants spread across different areas and under varying weather conditions.

Incorporating wind effects can be effectively achieved through weighted models that adjust pollutant concentrations based on wind direction and speed. Such models often use dispersion algorithms to simulate the movement and dilution of pollutants in the atmosphere(Snoun et al., 2023). For example, Gaussian dispersion models in study by Bruzzone and Nocera (2021) have been widely adopted for estimating the impact of traffic emissions on surrounding areas, adjusting concentration levels according to wind vectors. These models are particularly useful in urban settings, where the complex interplay of buildings and infrastructure can significantly influence air flow and pollutant distribution.

In the context of this study, which focuses on central London, incorporating climatic conditions - especially wind speed and direction - into the calculation of weighted average air quality offers several benefits. It allows for more accurate and realistic assessments of pollutant exposure, reflecting the dynamic nature of urban air quality. By accounting for these factors, the study can better understand and predict areas of higher pollution concentration and potential exposure risks, contributing to more informed urban planning and public health strategies.

2.3 Research Question

2.3.1 Time Period

The ULEZ was first implemented in the Central Zone on April 8, 2019(Transport for London, 2019), during the pre-pandemic period. The next major change was the Inner Expansion, which took effect on October 25, 2021. This expansion occurred during the period when England was beginning to ease lockdown restrictions. On Monday, February 22, 2021, the UK Government published its 4-step plan to ease lockdown restrictions in England (Institute For Government, 2022). The most recent change was the Outer Expansion (Transport for London, 2023), which was implemented on August 29, 2023, in the post-pandemic period.

Wherever possible, the study will symmetrically frame the pre- and post-implementation periods around each ULEZ measure. This approach ensures that the data samples are balanced, providing a more accurate comparison of air quality and traffic conditions before and after each implementation phase. Therefore, the research time period division is as follows:

Table 2.4.1 Research Time Period Division Table: ULEZ vs. London Pandemic Lockdown Timelines

Research Time Period	ULEZ Policies	COVID-19 Lockdown
01 July. 2018 – 07 Apr. 2019 08 Apr. 2019 – 31 Dec. 2019	2019 Central Zone Implementation (April 8, 2019) *ULEZ was officially implemented in the central London zone.*	Pre-Pandemic Period
22 Feb. 2021 – 24 Oct. 2021 25 Oct. 2021 –31 Mar. 2022	2021 Inner Expansion (October 25, 2021) *ULEZ was expanded to include all vehicle types within Inner London.*	During Pandemic Period *On Monday 22 February 2021, the UK Government published its 4-step plan to ease lockdown restrictions in England.*
1 Aug. 2022 – 28 Aug. 2023 29 Aug. 2023 – present	2023 Outer Expansion (August 29, 2023) *ULEZ was further expanded to cover all 32 London boroughs, incorporating an additional five million people into the zone.*	Post-Pandemic Period

2.3.2 Study Area

The research will focus on central London, which includes the following boroughs: Camden, City of London, Islington, Kensington and Chelsea, Lambeth, Southwark, Westminster.

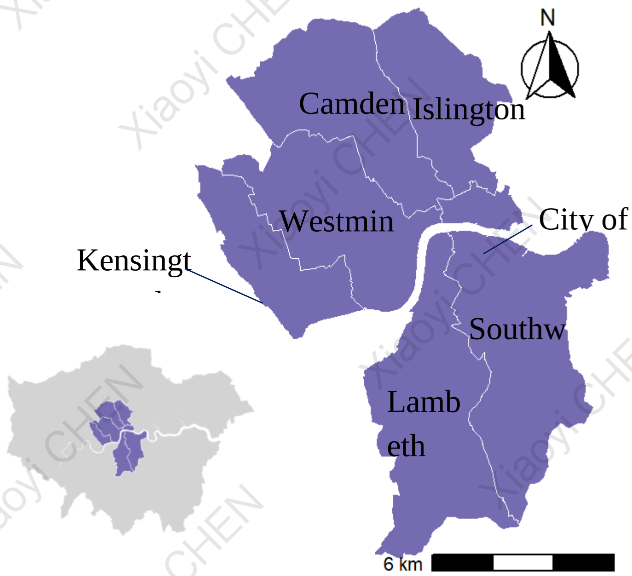


Figure 2.4 Study area of central London

These Boroughs are selected due to their central location and the availability of comprehensive air quality and public transport data. The study specifically targets air quality monitoring sites within these areas, analyzing their geographic information and air quality data. Additionally, it examines the geographic information and passenger flow data of subway stations within these boroughs.

The **Table A.1** in Appendix A below lists the underground stations within each study area. This comprehensive list provides the specific locations where passenger flow data and geographic information have been collected and analyzed.

2.3.3 Research Question

By dividing the analysis into periods before and after the ULEZ implementation and considering the effects of the COVID-19 pandemic, we aim to isolate the impact of ULEZ on air quality and public transport usage. Therefore, our main research questions are as follows:

1. How has the air quality in central London changed before and after the implementation of the ULEZ policy on these three time periods April 8, 2019, October 25, 2021, and August 29, 2023?

2. Has the introduction and expansion of ULEZ policies affected public transport usage in central London, specifically around the key dates of April 8, 2019, October 25, 2021, and August 29, 2023?

To further explore the impact and mechanisms underlying these changes, the following supplementary questions are considered:

3. What are the spatial patterns of air quality changes across different boroughs of central London before and after ULEZ implementation?

4. How do variations in traffic volumes, specifically of different vehicle types, correlate with changes in air quality in the study areas?

5. How did the COVID-19 pandemic influence the patterns of air quality and public transport usage, and how can these effects be separated from the impacts of ULEZ?

6. Are there differences in the effectiveness of ULEZ policies on air quality and public transport usage between major and minor roads within the study areas?

7. How do socio-economic factors and urban characteristics of different boroughs influence the effectiveness of ULEZ policies?

3 Methodology

3.1 Data Sources

3.1.1 Data Collection

The data sources utilized in this study are diverse and comprehensive, covering various aspects of:

1. Air quality data source (Imperial College London, 2018, Imperial College London, 2024, Ricardo Energy & Environment, 2020)
2. Public transport data source (Department for Transport, 2018, Transport for London, 2017)
3. Meteorology data source (NOAA & Deutscher Wetterdienst & Environment Canada, 2016)
4. and Geographic boundaries within central London (Greater London Authority, 2003, London, 2018).

The following table provides a detailed description of each dataset, including its source, format, and application in this research.

Table 3.1 Data Collection, description and application

Dataset Source		Description	Application/Process
Air Quality Data	Air Quality Data London Datastore	database consolidates information from two sources: 1. the London Air Quality Network, which measures NO2, PM10, O3, and PM2.5 at regulatory air quality monitoring sites. 2. Breathe London, which measures NO2 and PM2.5 using sensors. This study primarily utilizes air quality data from specific monitoring sites within defined areas on the London Air Quality Map, covering specific time ranges from both data sources. Format: .csv .xls	As a collection of Air Quality data sources, this data is mainly used to arrive at London Air Quality Network Statistics Maps and Breathe London air quality data in this analysis.
	London Air Quality Network Statistics Maps & Data Search Engine	The London Air Quality Network, managed by Imperial College London, shares data through the Air Quality Data London Datastore. It provides information on various air quality objectives at different monitoring sites for specific time periods, including: Annual mean NO2 objective, Hourly mean NO2 objective (number of hours), Annual mean PM10 objective, Daily mean PM10 objective (number of days), 15-minute mean SO2 objective, Hourly mean SO2 objective, Daily mean SO2 objective, Rolling 8-hour mean ozone objective, Rolling 8-hour mean CO objective, Annual mean PM2.5 objective Format: .csv	In addition to obtaining daily and hourly NO2 and PM2.5 data across different date ranges before and after the three ULEZ implementations, I also need the precise geographic information (latitude and longitude) for each monitoring site. Using the Kriging method, this study will calculate the average air quality for the target area.
	Breathe London	Breathe London also shares its data through the Air Quality Data London Datastore. The monitoring sites for Breathe London are distinct from those of the London Air Quality Network. Breathe London measures hourly NO2 and PM2.5 across different date ranges. It provides comparisons with the UK annual objective, WHO annual guideline, and WHO daily guideline. Format: data version(.csv .xls), Image version(.jpeg .png .svg .pdf)	As a supplement to the London Air database's record of missing NO2 and PM2.5 after 2021.
	London Air Quality Network, Defra's Air Quality Monitoring	The London Air Quality Network and Defra's Air Quality Monitoring provide specific geographic information (latitude and longitude) for each monitoring site, along with some climatic information (e.g., temperature, wind direction, and wind speed). They offer hourly measurements of various pollutants for specific date ranges, including: Ozone (O3), Nitrogen Dioxide (NO2), Sulphur Dioxide (SO2), Particulate Matter (PM2.5), Particulate Matter (PM10)	This study will obtain hourly measurements of various pollutants for specific date ranges, calculate their daily mean values, and then use the Kriging method to determine the average air quality for the target area.

		Format: .csv	
Public Transport Data	The public Transport for London (TfL) data	The public Transport for London (TfL) data includes Annual Station Counts and Network Demand. It represents the travel demand on a typical autumn weekday (Monday-Thursday), Friday, Saturday, and Sunday at all stations and lines of the London Underground, London Overground, Docklands Light Railway, TfL Rail / Elizabeth Line, and London Trams. Format: .csv	The study focuses on the Network Demand data, particularly the *Entry Tap Count* and *Exit Tap Count* at various tube stations, as well as *Tube Journey Count* and *Bus Journey Count*. Data from the years 2019, 2021, and 2023 will be examined to compare passenger volumes and transport demand before and after the ULEZ implementation.
	Road traffic statistics - London region (dft.gov.uk)	The Road Traffic Statistics for the London region, provided by the Department for Transport (dft.gov.uk), include data on various vehicle types such as Cars, Taxis, and Private Hire Vehicles, Cycles, Vans and Lorries, and Buses and Coaches. The statistics cover different types of roads, including Motorways, A-roads, B and C-roads, and Minor roads. These files are utilized as vehicle counts recorded at specific count points in London regions, with data spanning from 1993 to 2023, though not for every year continuously. Available Data: Count points; Regional traffic: by vehicle type; Regional traffic; Average annual daily flow; Average annual daily flow by direction; Raw counts. Format: .json .csv	This study intends to download the "Raw counts" files for the following central London boroughs: Camden, City of London, Islington, Kensington and Chelsea, Lambeth, Southwark, Westminster The main focus will be on the traffic volumes of different modes of transportation, including Pedal cycles, Two-wheeled motor vehicles, Cars and taxis, Buses and coaches The objective is to analyze the time series changes in traffic volumes from 2018 to 2023. This data will be compared with changes in underground passenger flows to explore the relationship between London's air quality changes and these factors, including their spatial patterns and spatial autocorrelation.
Meteorological Data	Weather History & Climate | The Weather's Record Keeper | Meteostat	The meteorological data provides essential weather parameters for different areas of London, covering both historical dates and specific date ranges. The key meteorological parameters included are: Air Temperature, Average Temperature, Minimum Temperature, Maximum Temperature, Dew Point, Total Precipitation, Wind (From) Direction, Average Wind Speed, Wind Peak Gust, Relative Humidity, etc. Format: JSON API, .ods .html .csv .xslx	This data will be used to apply the Kriging method to process air quality data from specific monitoring points within a defined area. By incorporating wind speed (in km/h) and wind direction (0~359°, 0° indicates due south, increasing degrees clockwise), we will calculate a weighted average air quality for the region, providing a more accurate representation of air quality that accounts for meteorological factors.
Geographic file	London Boroughs and Wards	statistical-gis-boundaries-london.zip - Statistical GIS Boundary Files for London, containing Borough, Ward, MSOA, LSOA and OA boundaries as at 2011 Format: .zip .shp	By integrating these boundary files with air quality and transport datasets, we can accurately map and analyze the spatial distribution of air quality measurements and transport flows. This integration will facilitate the examination of the relationship between air quality changes and public transport usage across different administrative regions, enhancing the spatial

			autocorrelation and spatial pattern analysis in this research.
	[FOI request detail - Transport for London (tfl.gov.uk)]	Coordinates of London Underground stations (including Elizabeth Line stations), updated on 24 August 2018. Format: .csv	As Entry tap count and Exit tap count data for all metro stations within the study area, correlation and regression analysis with air quality or other modes of transportation are performed.

3.1.2 Data Feasibility Analysis

3.1.2.1 Air Quality Data

In this research, several challenges with the air quality data have been identified. These issues must be addressed to ensure the feasibility and accuracy of the study, as detailed in **Table A.2** in Appendix A.

1. Data Gaps

A significant challenge encountered is the presence of data gaps within the air quality datasets. Data gaps can lead to incomplete analyses and potential biases, particularly if the missing periods coincide with significant pollution events or trends. This issue can be mitigated by using available data from the affected sensors while supplementing it with data from other sensors in the study area. Cross-referencing with other databases that cover the same region can help fill these gaps, ensuring a more comprehensive dataset.

Specifically, for the study area of Islington, PM2.5 data can only be collected from the three major Air Quality data sources for the period of September 8, 2023, to July 2024 (as shown in **Figure B.1** in Appendix B). Therefore, when comparing PM2.5 indicators between regions, Islington cannot be directly referenced in the analysis results. This limitation also prevents us from conducting correlation and regression analyses between PM2.5 time series records and public transport flow data for Islington.

However, this limitation does not significantly impact the overall research objectives. Aside from PM2.5, the data for the other three pollutants (NO, NO2, and PM10) are comprehensive across the seven study areas in central London. Thus, the analysis of these pollutants in relation to public transport usage will not be significantly affected. The broader study of air pollution and public transport in central London remains robust and informative.

This approach ensures that the integrity of the analysis is maintained while acknowledging and addressing the limitations posed by data gaps. By clearly documenting these gaps and their potential impact, we can provide a transparent and reliable analysis.

2. Time Scale

The study requires data on a daily scale. However, the "London Air Quality Network" database provides data on an hourly basis. Direct use of hourly data without aggregation results in an unmanageable dataset and complicates the analysis. Hourly data must be preprocessed to derive daily statistics such as the maximum, mean, and minimum values. For subsequent Kriging method calculations to determine air quality weighted values, the daily mean values are used as the observed values for computing the weights. This ensures that the spatial interpolation of air quality data is based on a consistent and manageable time scale, enhancing the reliability and interpretability of the results.

By aggregating hourly data into daily statistics, the study achieves a balance between data granularity and analytical feasibility, enabling robust and meaningful insights into the impact of ULEZ policies on air quality and public transport usage in central London.

3. Air Quality Categories

Not all sensors record a comprehensive set of air quality species. This study focuses on the most commonly recorded species: PM2.5, PM10, NO, and NO2 as supplementary species. The selective approach may omit the analysis of other potentially important pollutants. The solution is concentrating on PM2.5, PM10, NO, and NO2, which are consistently recorded across most sensors, providing a robust baseline for analysis. Including additional species where available enhances the depth of the study without compromising the consistency of the core dataset.

4. Sensor Coordinates

Access to precise geographic coordinates of sensors via API has proven difficult due to maintenance or lack of updates. Inaccurate or missing coordinates can significantly affect spatial analysis and interpolation accuracy. To address this, coordinates will be manually obtained from maps by identifying the closest landmark or building. This approach ensures accurate spatial referencing necessary for effective spatial analysis.

3.1.2.2 Public Transport Data

In the dataset obtained from Road Traffic Statistics - London Region, there are instances of non-continuous traffic flow data. The primary challenge is that non-continuous data can lead to inaccuracies in analyzing trends and making inferences about the traffic flow over specific periods. Gaps in data can obscure the true impact of interventions like the ULEZ policy on traffic and air quality.

The approach to mitigate data gaps is to employ interpolation techniques (Shumway and Stoffer, 2011), such as linear interpolation or more sophisticated methods like spline interpolation to estimate missing data points based on available observations. Spline interpolation is particularly suitable for handling large datasets, as it uses low-order polynomials (e.g., cubic splines) between each pair of data points, avoiding the oscillation problems associated with high-order polynomial(Press and Teukolsky, 2007). Furthermore, addressing the lack of specific data periods, particularly before the ULEZ Central Zone Implementation, April 8, 2019, we propose using data from July 1, 2018, to April 8, 2019. Zhang and Batterman (2013) assumes that traffic conditions from early 2018 to early 2019 are relatively stable and representative of the period before the ULEZ implementation. Therefore, it provides a broader view of traffic trends that is essential for this analysis.

Additionally, it is important to note that in the subsequent Exploratory Data Analysis focused on comparing and analyzing different regions and transportation modes, there is a significant data gap in the `Minor Roads` category for the City of London study area. This absence of data poses a challenge in fully understanding the traffic dynamics in the City of London compared to other boroughs. Consequently, the analysis of the City of London's traffic flow will primarily rely on the data collected from `Major Roads`. However, this limitation must be acknowledged, as it could potentially bias the results when comparing traffic volumes across different regions. The lack of data for minor roads might obscure some of the more granular details of traffic behavior, particularly in less congested areas, and could affect the overall ranking of traffic volumes among the study areas. Therefore, while the data from Major roads offers valuable information, the results is interpreted with caution, considering the possible underrepresentation of certain traffic flows within the City of London.

3.1.2.3 Meteorological Data

As Kousis et al. (2021) has noted, central urban city shows relatively homogeneous microclimatic conditions due to the urban heat island effect, leading to consistent higher temperatures and wind patterns across the urban area. Boroughs like Westminster, Camden, and the City of London have high building density, extensive road networks, and significant human activity, resulting in similar air movement and temperature profiles(Mills, 2008). Thus, the London Weather Centre is a valid representative source for these areas.

3.2 Data Processing

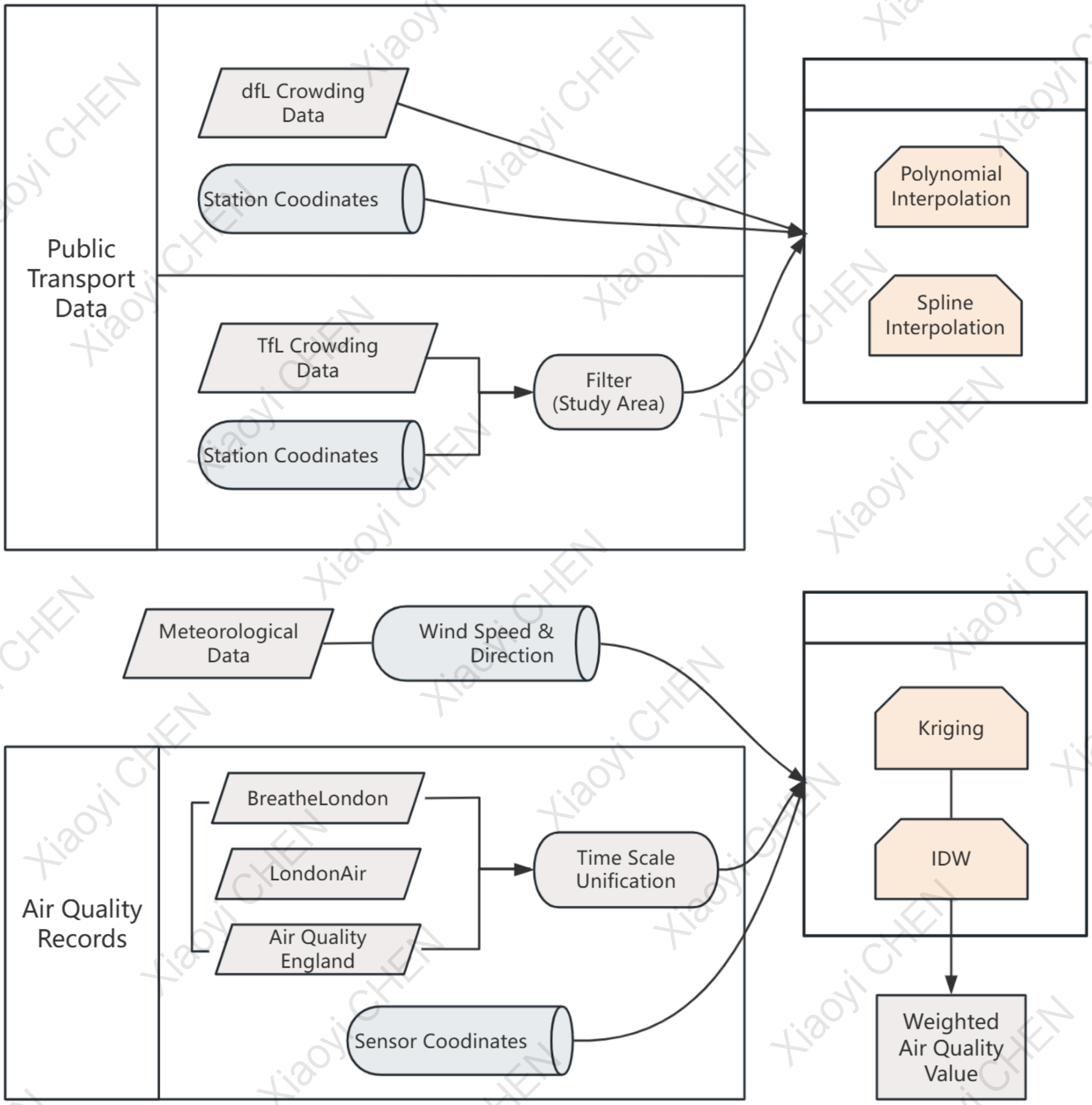


Figure 3 Data processing flowchart

3.2.1 Missing value interpolation processing

3.2.1.1 Spline interpolation process

Spline interpolation fits a smooth curve through the known data points, providing more accurate estimates for larger gaps. It is useful for capturing underlying trends in the data.

For the Road Traffic Statistics - London Region dataset, the raw counts are provided in CSV files. These files are named `dft_rawcount_local_authority_ID.csv`, where the ID corresponds to specific regions (id_145: Camden; id_174: City of London; id_96: Islington; id_110:

Kensington and Chelsea; id_107: Lambeth; id_103: Southwark; id_109: Westminster.). Each file contains vehicle counts recorded at specific count points from major roads and minor roads within these areas, with the complete data can be found in **Table A.3** in Appendix A.

Spline interpolation method ensures that the continuity and trends in the dataset are maintained (Hemelrij.J, 1965):

Firstly, the entire range of the dataset is divided into intervals between the known data points. For each interval, a cubic polynomial $S_i(x) = a_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3$ is constructed, where x_i is the known data point in the interval. The cubic polynomials are fitted such that they are continuous in the first and second derivatives at each of the data points, and this system of equations is solved to obtain the coefficients a_i, b_i, c_i, and d_i.

To illustrate the spline interpolation process, we provide a simplified demo in the appendix. The code snippet (see **Appendix C.1.1**) demonstrates the steps involved in loading the data, performing the spline interpolation, and saving the interpolated values.

3.2.1.2 Polynomial interpolation process

Polynomial interpolation is used for the air quality data, spline interpolation. This method is particularly suitable for large datasets and avoids the oscillation problems associated with high-order polynomials. Polynomial interpolation uses low-order polynomials between each pair of data points, ensuring smooth and accurate interpolation. To address the data gaps in the metastatic datasets, spline interpolation is used as it is well-suited for filling small gaps in the data.

Polynomial interpolation involves finding a single polynomial that passes through all the given data points. This method is often used for smaller datasets or when the data is expected to follow a polynomial trend (Press and Teukolsky, 2007):

For a set of n known data points $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, a polynomial of degree $n - 1$ is formulated as $P(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_{n-1} x^{n-1}$. To solve for the coefficients $a_0, a_1, \dots, a_{n-1}$, a Vandermonde matrix is constructed using the known data points. The system of linear equations $V \cdot a = y$, where V is the Vandermonde matrix, a is the vector of coefficients, and y is the vector of known values, is solved to determine the polynomial coefficients. Once the polynomial $P(x)$ is determined, it is used to interpolate and estimate the missing values within the range of the known data points.

The detailed implementation of this method can be found in **Appendix C.1.2**.

3.2.2 Kriging Weights Calculation

Spatial interpolation is a crucial technique in geographical analysis, allowing us to estimate unknown values at specific locations based on known values from surrounding locations. To calculate the weighted air quality in a specific region, especially in urban areas like central London, we employ the Kriging method to interpolate air quality data, specifically focusing on air quality categories(such as $PM_{2.5}$, NO_2) concentration levels across a defined study area with the influence of climatic conditions such as wind speed and direction. The primary goal of using Kriging is to generate a spatially continuous surface of air quality metrics that can be used to assess pollution levels at unsampled locations within the study area.

3.2.2.1 Variogram Calculation

The first step in Kriging is to compute the experimental variogram, which quantifies spatial dependence by measuring the dissimilarity between data points as a function of distance. The semi-variance $\gamma(h)$ is calculated as follows:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_i + h)]^2 \tag{3.1}$$

where $\gamma(h)$ is the semi-variance at distance h, defined as a function of the separation distances among observations(Matheron, 1971). $N(h)$ is the number of point pairs separated by distance h, and $Z(x_i)$ is the measured value at location x_i. This approach is similar to the methodology used by Beers and Kleijnen (2004) and Shad et al. (2009), where the mean squared error (MSE) is employed.

This involves computing the distances between all pairs of sensor points using the Haversine formula(Ashari and Purwono, 2023), which accounts for the spherical nature of the Earth, providing more accurate distance measurements than simple Euclidean distances:

$$Haversine = 2R \cdot arcsin\left(\sqrt{sin^2\left(\frac{\Delta\phi}{2}\right) + cos(\phi_1)\, cos(\phi_2)\, sin^2\left(\frac{\Delta\lambda}{2}\right)}\right) \tag{3.2}$$

Where R is the Earth's radius, ϕ and λ are the latitude and longitude in radians.

After obtaining the variogram, we construct the Kriging system of equations to estimate the weights λ for each known data point(Szidarovszky et al., 1987):

$$\begin{bmatrix} \gamma(0) & \gamma(h_{1,2}) & \dots & \gamma(h_{1,n}) & 1 \\ \gamma(h_{2,1}) & \gamma(0) & & \gamma(h_{2,n}) & 1 \\ \vdots & & \ddots & \vdots & \\ \gamma(h_{n,1}) & \gamma(h_{n,2}) & \dots & \gamma(0) & 1 \\ 1 & 1 & & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_1 \\ \vdots \\ \lambda_1 \\ \mu \end{bmatrix} = \begin{bmatrix} \gamma(h_{0,1}) \\ \gamma(h_{0,2}) \\ \vdots \\ \gamma(h_{0,n}) \\ 1 \end{bmatrix} \tag{3.3}$$

Here, $\gamma(h_{i,j})$ represents the semi-variance between points i and j, and $\gamma(h_{0,i})$ represents the semi-variance between the interpolation point and each known point.

3.2.2.2 Weight Adjustment for Interpolation

To account for wind effects, which influence the dispersion of pollutants, the Kriging weights are adjusted by considering wind direction and speed:

$$i_{\prime}^{\lambda} = \lambda_i \left(1 + \frac{v \cdot cos(\theta - \theta_i)}{v_{max}}\right) \tag{3.4}$$

where v is the wind speed, θ is the wind direction, θ_i is the direction from the interpolation point to sensor i, and v_{max} is the maximum wind speed for normalization. The adjusted weights are then normalized(Hartkamp et al., 1999):

$$i_{\prime\prime}^{\lambda} = \frac{i_{\prime}^{\lambda}}{\sum_{i=1}^{n} i_{\prime}^{\lambda}} \tag{3.5}$$

Incorporating wind effects in this manner allows for more realistic air quality estimates.

Finally, based on the weights $i_{\prime\prime}^{\lambda}$ obtained by the features of the inverse distance weighting method(Ozelkan et al., 2016), the air species concentration at the interpolation point is estimated using the normalized weights:

$$Z^{*}(x_0) = \sum_{i=1}^{n} i_{\prime\prime}^{\lambda} Z(x_i) \tag{3.6}$$

where $Z(x_i)$ are the corresponding air quality species values at the known locations.

3.2.2.3 Application

In this section, the Kriging interpolation method is applied to the air quality data for various regions in central London. The goal is to estimate a spatially continuous surface of air quality metrics, specifically focusing on the weighted concentration levels of pollutants such as NO, NO_2 , PM10, and PM2.5. The detailed code implementation for this process can be found in **Appendix C.2**.

The process begins by iterating through each defined region in the study. For each region, the air quality data is preprocessed, which includes filtering and cleaning the data to ensure consistency. As shown in **Equation (3.1)**, the semi-variogram is calculated using `calculate_semivariogram` function for each daily subset. This function computes the semi-variance between all pairs of sensor points based on their spatial locations (latitude and longitude) and the difference in their measured pollutant values. The semi-variance values are then averaged for unique distances to produce the experimental variogram, which quantifies the spatial dependence of air quality values across the study area. Next, the distances between all sensor points are calculated using Euclidean metrics as **Equation (3.2)**. Any `NaN` or infinite values within the distance matrix are identified and replaced with zeros to prevent computational issues during the Kriging weight calculation. The Kriging weights are then computed using the `calculate_kriging_weights` function, which solves the Kriging system of **Equation (3.3)** to determine the influence of each sensor point on the interpolation.

To account for the influence of meteorological conditions, particularly wind speed and direction, the Kriging weights are adjusted using the `adjust_weights` function in **Equation (3.4)(3.5)**. This adjustment incorporates wind effects by modifying the Kriging weights based on the wind speed, direction, and the relative orientation of each sensor point with respect to the interpolation point. The adjusted weights are then normalized to ensure they sum to one, which is necessary for the subsequent interpolation step.

The normalized weights, **Equation (3.6),** are used to estimate the weighted air quality value at the interpolation point. The interpolation is performed by the `interpolate` function, which calculates a weighted average of the pollutant values, ensuring that the resulting value is non-negative.

Finally, the weighted values for each date are compiled into a data frame and saved to a csv file for each region. This output represents the spatially interpolated air quality data, which can be used for further analysis, such as trend analysis, change-point detection, or evaluating the impact of environmental policies like the ULEZ.

This application of Kriging, with the incorporation of wind effects, provides a robust method for estimating air quality across a heterogeneous urban environment, enabling more accurate assessments of pollution levels and their potential impact on public health.

4 Result

4.1 Exploratory Data Analysis(EDA)

4.1.1 Air pollutants EDA

4.1.1.1 Analysis of weighted NO and NO2 Distributions

The histograms for weighted NO and NO2 across central London reveal notable differences in air quality distribution among the study areas. NO concentrations show a skew towards lower values, with Camden, City of London, Kensington and Chelsea, Islington, and Southwark displaying lower average concentrations and occasional high-pollution events. In contrast, Westminster and Lambeth exhibit broader distributions with consistently higher NO concentrations, indicating more frequent and sustained pollution episodes.

For NO2, the histograms demonstrate elevated levels across all boroughs, but with broader and more varied distributions compared to NO. Westminster stands out with significantly higher NO2 concentrations, often peaking around 50 $\mu g/m^3$, suggesting a persistently high pollution level, likely driven by dense traffic and vehicular activity. Camden and Southwark also show substantial NO2 pollution, though their peaks are lower, around 30 $\mu g/m^3$. Other boroughs generally have peaks below 20 $\mu g/m^3$, indicating relatively lower NO2 pollution.

Boxplots further illustrate these patterns, highlighting higher median values and significant outliers in Westminster, Lambeth, and the City of London. This suggests frequent episodes of high NO and NO2 pollution in these boroughs, with the City of London showing a slightly narrower range for NO2, possibly due to stricter regulations or more effective traffic management.

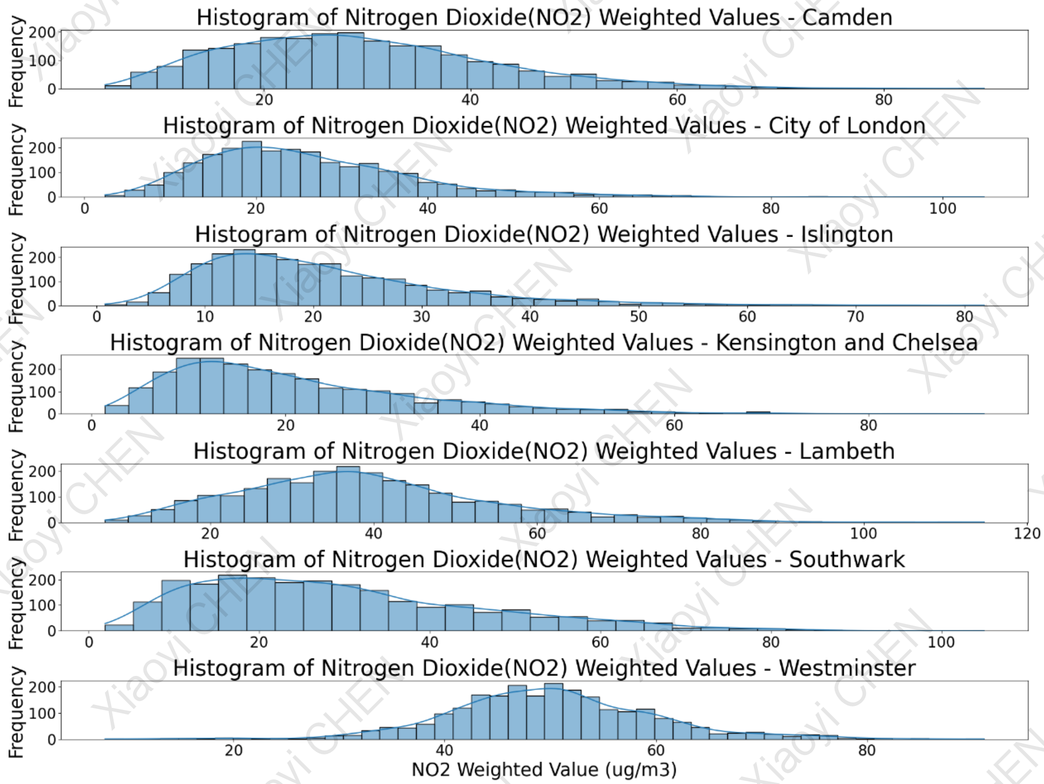
4.1.1.2 Analysis of weighted PM2.5 and PM10 Distributions

The distribution of particulate matter (PM2.5 and PM10) varies significantly across the study areas. For PM2.5, Westminster and the City of London have the highest frequency peaks around 10 $\mu g/m^3$, indicating consistent pollution levels, likely due to high traffic density and urban activity. In other boroughs, PM2.5 levels generally peak below 10 $\mu g/m^3$, suggesting lower pollution levels.

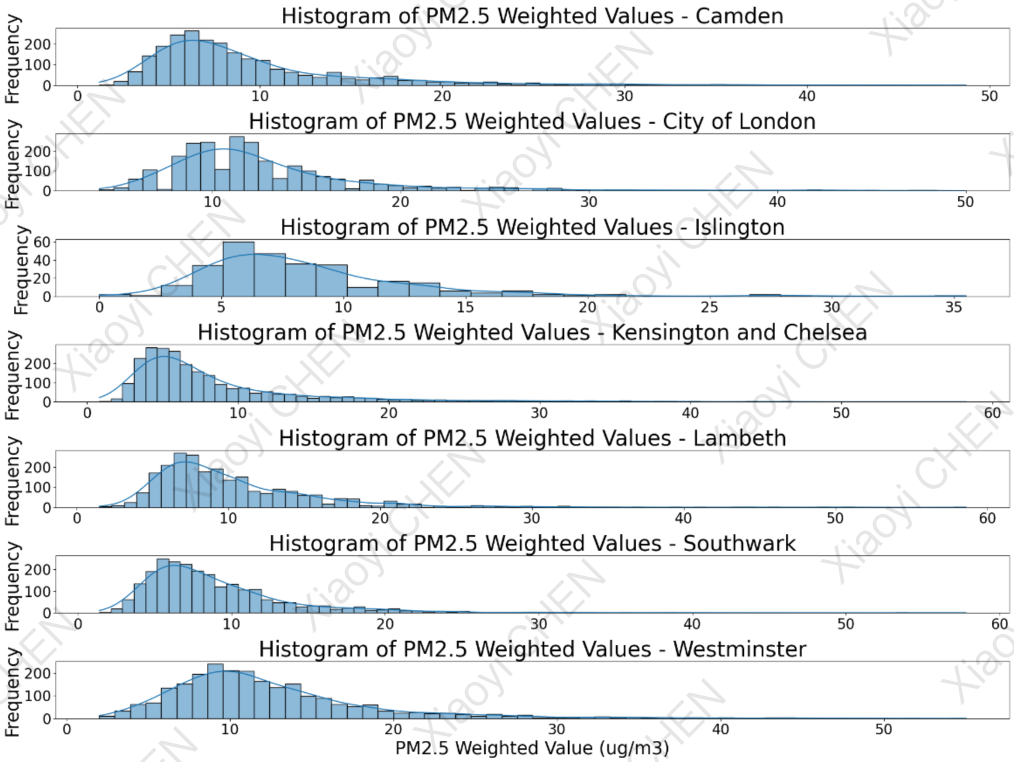
PM10 distributions tell a different story, with Lambeth showing the broadest distribution and the highest concentration peaks around 35 $\mu g/m^3$. This is notably higher than in other boroughs, where PM10 peaks generally range between 15-20 $\mu g/m^3$. The elevated PM10 levels in Lambeth suggest localized pollution sources, such as industrial activities or construction, warranting further investigation.

Boxplots for PM2.5 and PM10 align with these observations, with Lambeth showing greater variability and higher median values, particularly for PM10. Southwark, adjacent to Lambeth, exhibits similar patterns, though to a lesser extent, suggesting that industrial emissions may significantly impact air quality in the southern part of central London.

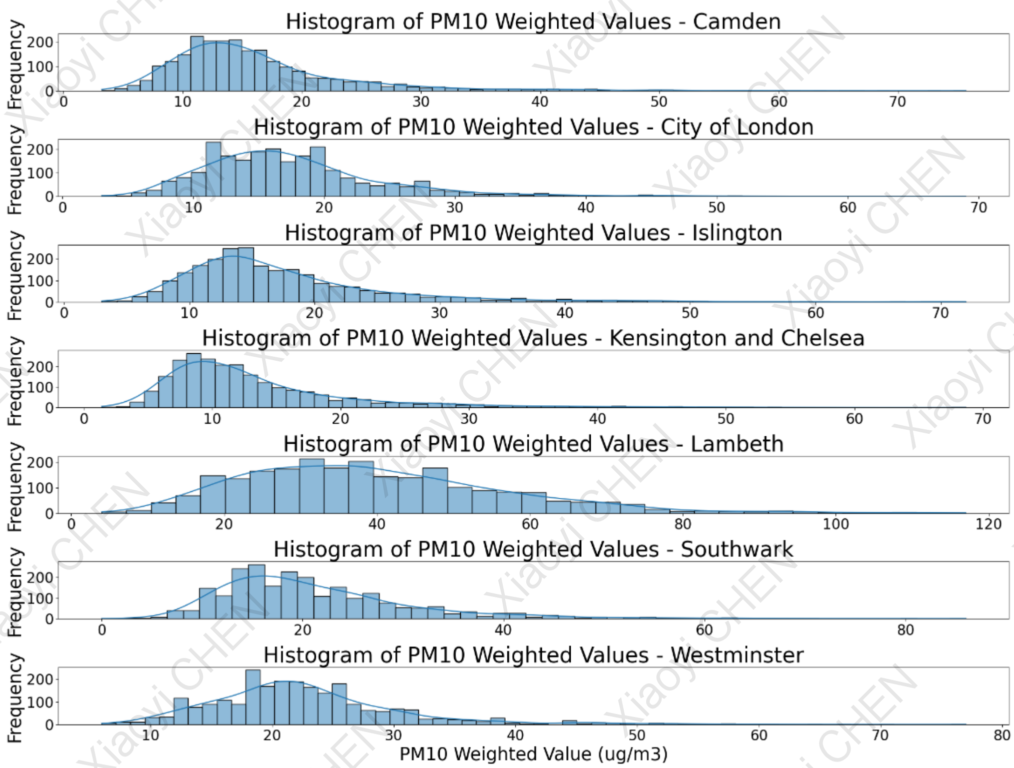
(a)



(b)



(c)



(d)

Figure 4.1.1 Histogram of weighted (a)NO, (b)NO2, (c)PM2.5 and (d)PM10 (ug/m3) on central London's regions

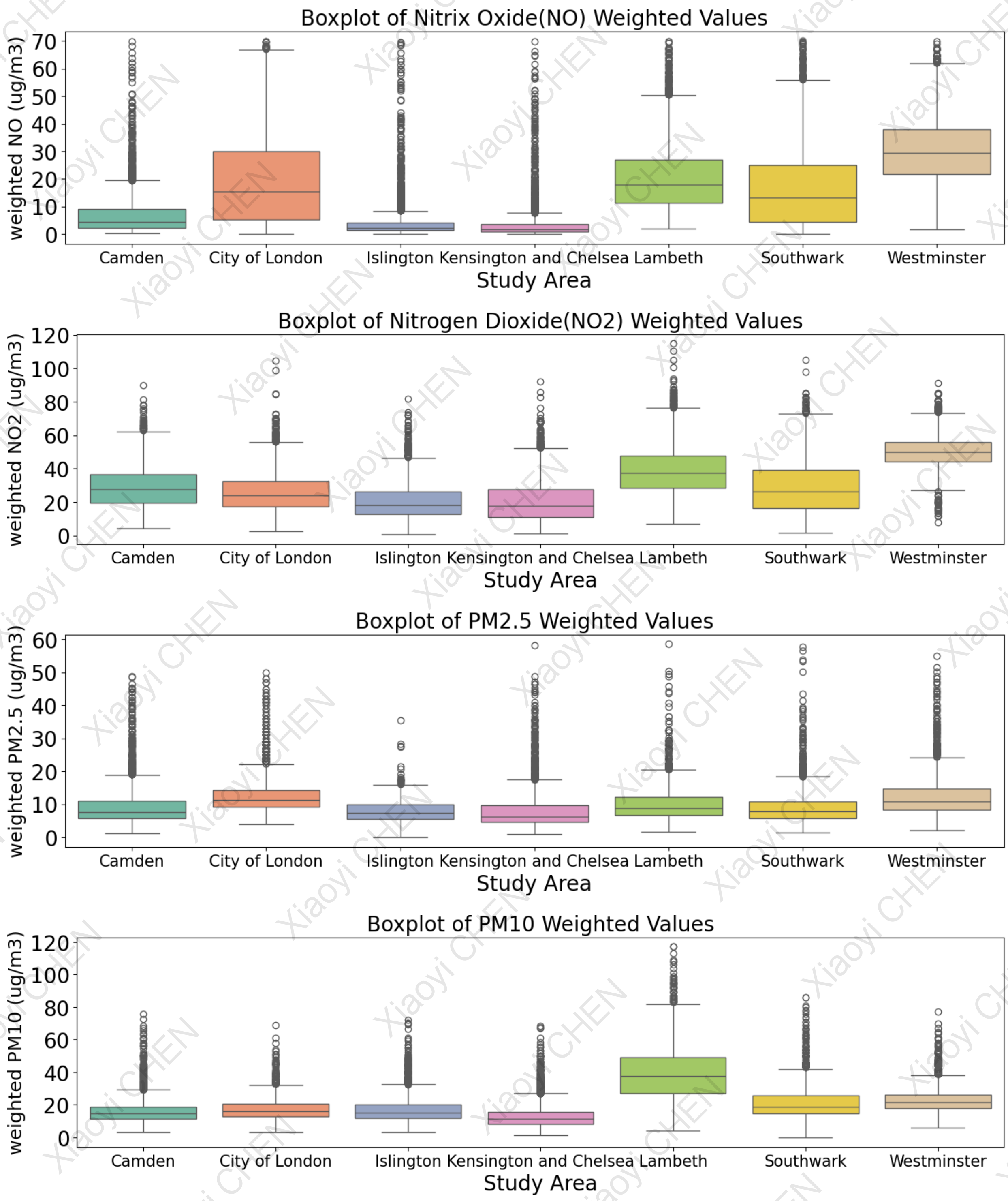


Figure 4.1.2 Boxplot of weighted NO, NO2, PM2.5 and PM10 (ug/m3) on central London's regions

Overall, the analysis underscores the multifaceted nature of air quality issues in central London, with different pollutants being more prevalent in specific boroughs due to a combination of traffic, industrial activities, and localized sources. Westminster, the City of London, and Lambeth are of particular concern, with Westminster showing high levels of NO2 and PM2.5, while Lambeth is notably affected by PM10 pollution. These findings highlight the need for targeted air quality management strategies that consider the specific pollution sources and their distribution across different areas.

In addition, we also have statistical charts for Time Series of Air pollutants in each study area, as shown **Figure B.1** in Appendix B. Three contrasting lines have also been added to the figure from (World Health Organization, 2021a) (details in **Table A.2** in Appendix A)

4.1.2 Road traffic EDA

The vehicle types analyzed include pedal cycles, two-wheeled motor vehicles, cars and taxis, buses and coaches, light goods vehicles (LGVs), and heavy goods vehicles (HGVs). Details of vehicle types is on **Table A.3** in Appendix A. The exploratory data analysis (EDA) for road traffic reveals key patterns in the distribution of different vehicle types across major and minor roads within the seven study areas on **Figure B.2** of Appendix B:

- Central Boroughs (Westminster, Lambeth, Southwark): High vehicle counts, especially on major roads, underscore their roles as commercial and residential centers.
- Peripheral Boroughs (Kensington and Chelsea, Camden): A more balanced traffic distribution with a notable emphasis on cycling, particularly on minor roads.
- Geographic Influence: The proximity of central roads to commercial centers in Westminster, Lambeth, and Southwark increases traffic flow, especially on major roads, while Kensington and Chelsea's residential nature leads to different road usage patterns.

The EDA underscores the distinct functionalities of major and minor roads in different boroughs. Urban traffic management should consider these differences when developing policies, as each borough's unique characteristics heavily influence road usage.

4.2 Time Series Analysis

4.2.1 Air Quality Trend Analysis on ULEZ changes

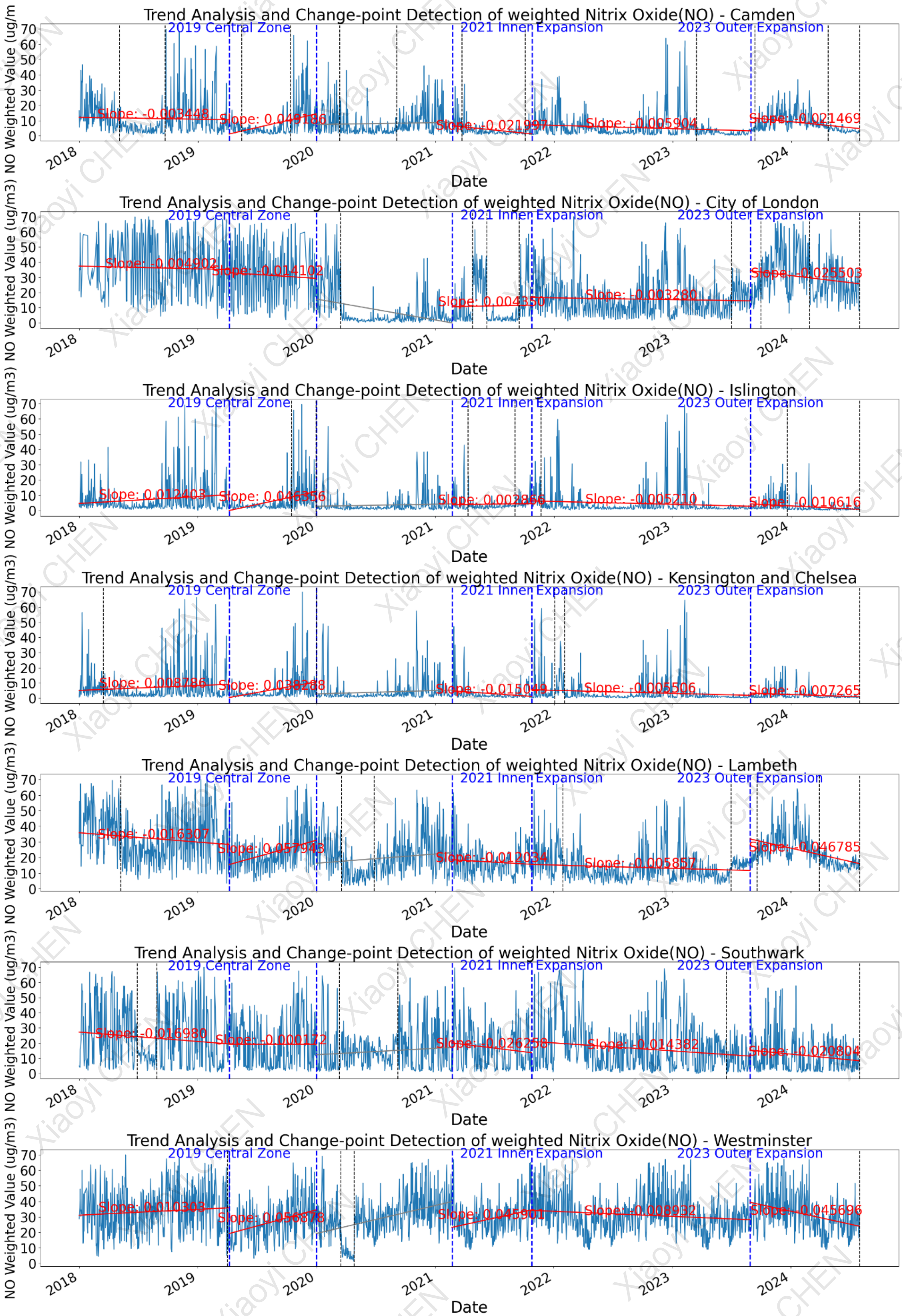
The time series figures presented in this study are designed to illustrate the trends and changes in air pollutant concentrations over time, with a focus on the periods surrounding the implementation of the ULEZ policies. In **Figures 4.2.1**:

- The vertical blue dotted line marks the beginning of each ULEZ policy implementation phase, distinguishing different study periods.

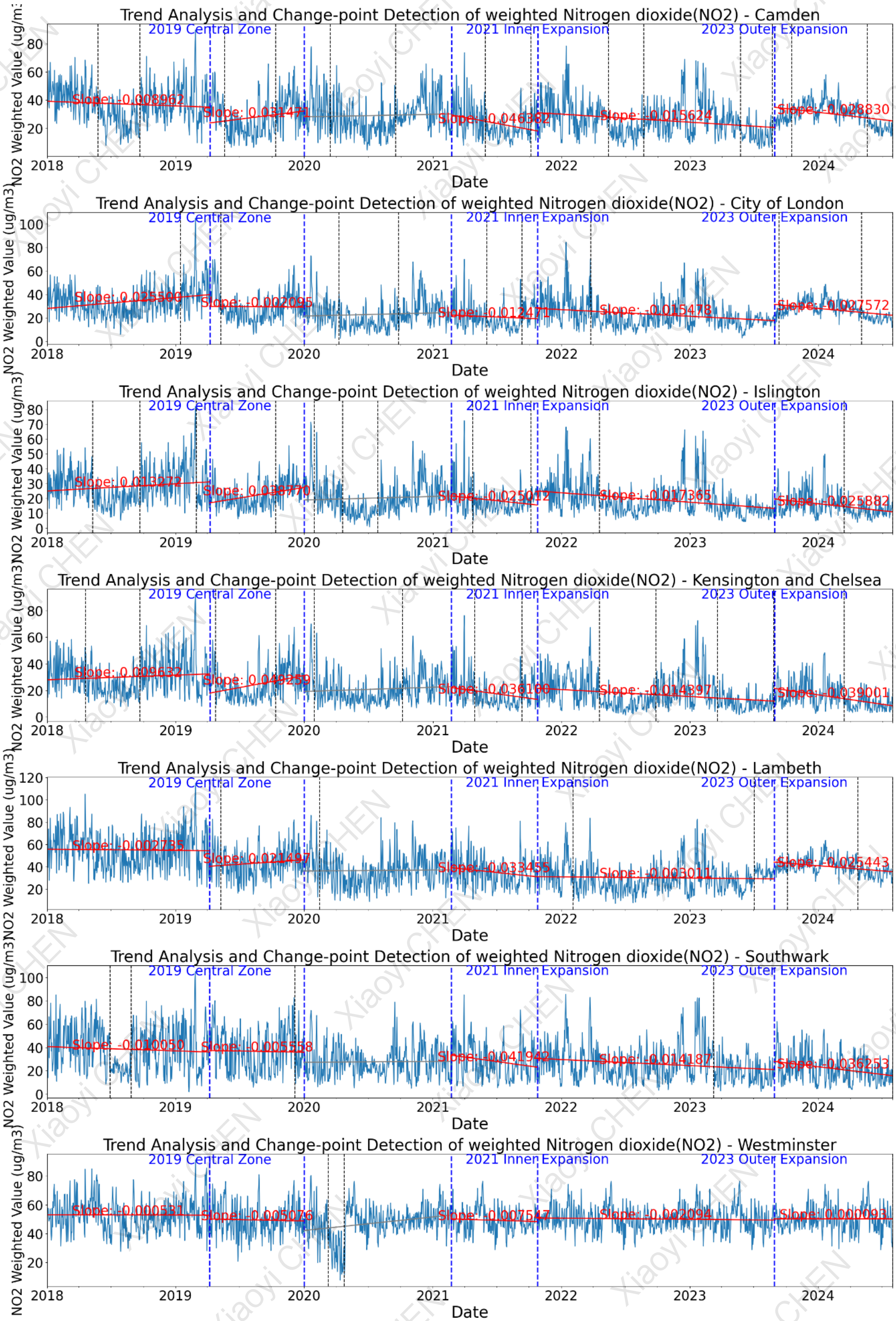
- The vertical black dotted line represents the location of the detected change point within the time series, indicating when significant shifts in the data occurred.
- The red line depicts the trend of the air pollutant time series during different periods, both before and after each ULEZ policy implementation, calculated using the linear regression method.
- The gray line represents the trend of the air pollutant time series during the COVID-19 pandemic, specifically from January 1, 2020, to February 22, 2021, when lockdown restrictions were in place in London.

It's important to note that the gray line during the COVID-19 pandemic is not included in the assessment of the ULEZ policy's impact on air quality. The pandemic period significantly affected travel behavior and transportation choices, leading to an abnormal reduction in traffic volumes that do not accurately reflect the impact of the ULEZ. Therefore, this period is excluded from the analysis of ULEZ policy effects, and the trend during this time is shown separately in gray.

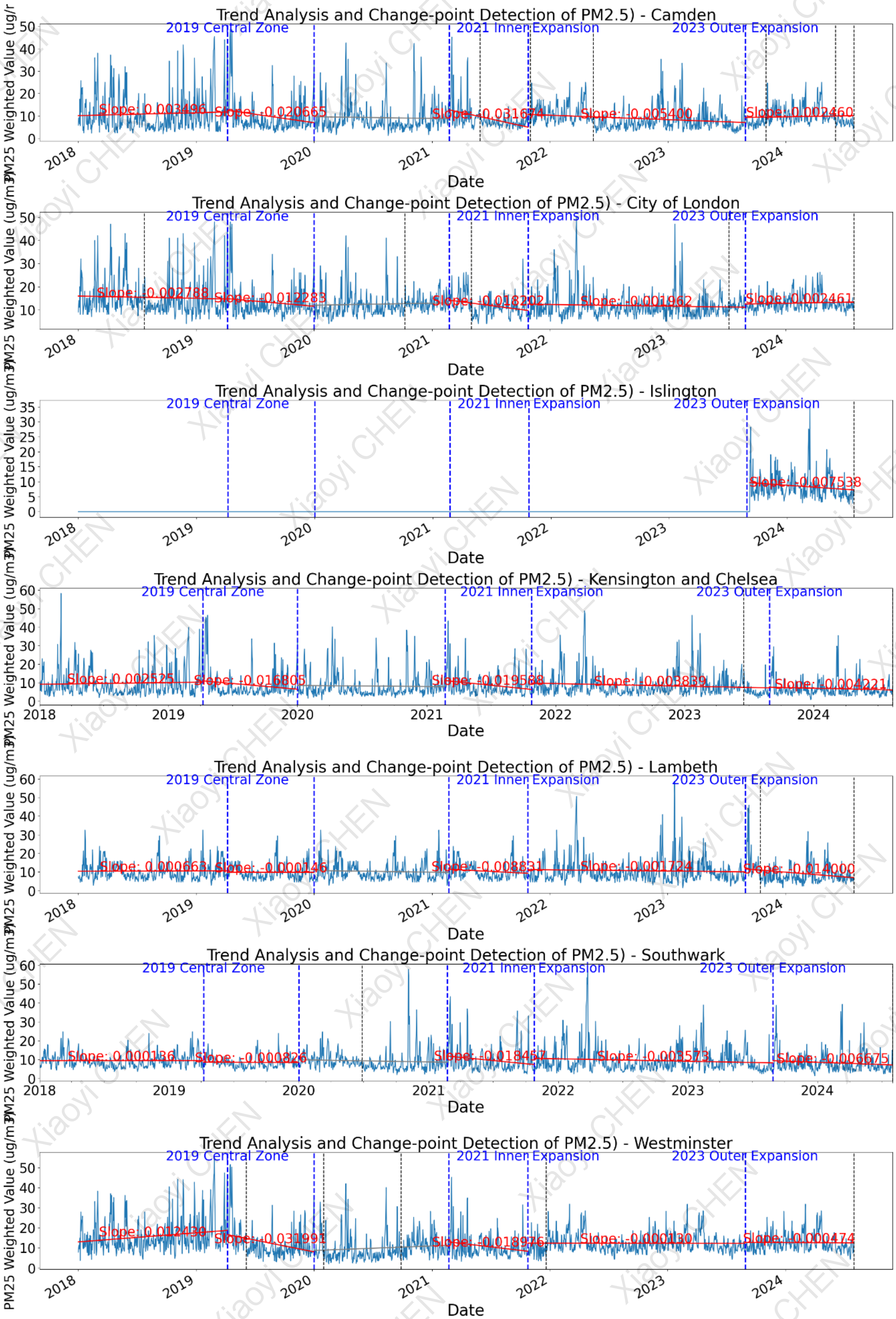
The results, as illustrated in **Figure 4.2**, show the trends and changes in NO, NO2, PM2.5, and PM10 concentrations across different boroughs during the study periods defined by ULEZ policy implementations and COVID-19 lockdown phases.


Trend Analysis and Change-point Detection of weighted Nitrix Oxide(NO) - Camden
Trend Analysis and Change-point Detection of weighted Nitrix Oxide(NO) - City of London
Trend Analysis and Change-point Detection of weighted Nitrix Oxide(NO) - Islington
Trend Analysis and Change-point Detection of weighted Nitrix Oxide(NO) - Kensington and Chelsea
Trend Analysis and Change-point Detection of weighted Nitrix Oxide(NO) - Lambeth
Trend Analysis and Change-point Detection of weighted Nitrix Oxide(NO) - Southwark
Trend Analysis and Change-point Detection of weighted Nitrix Oxide(NO) - Westminster
2019 Central Zone
2021 Inner Expansion
2023 Outer Expansion
NO Weighted Value (ug/m3)
Date

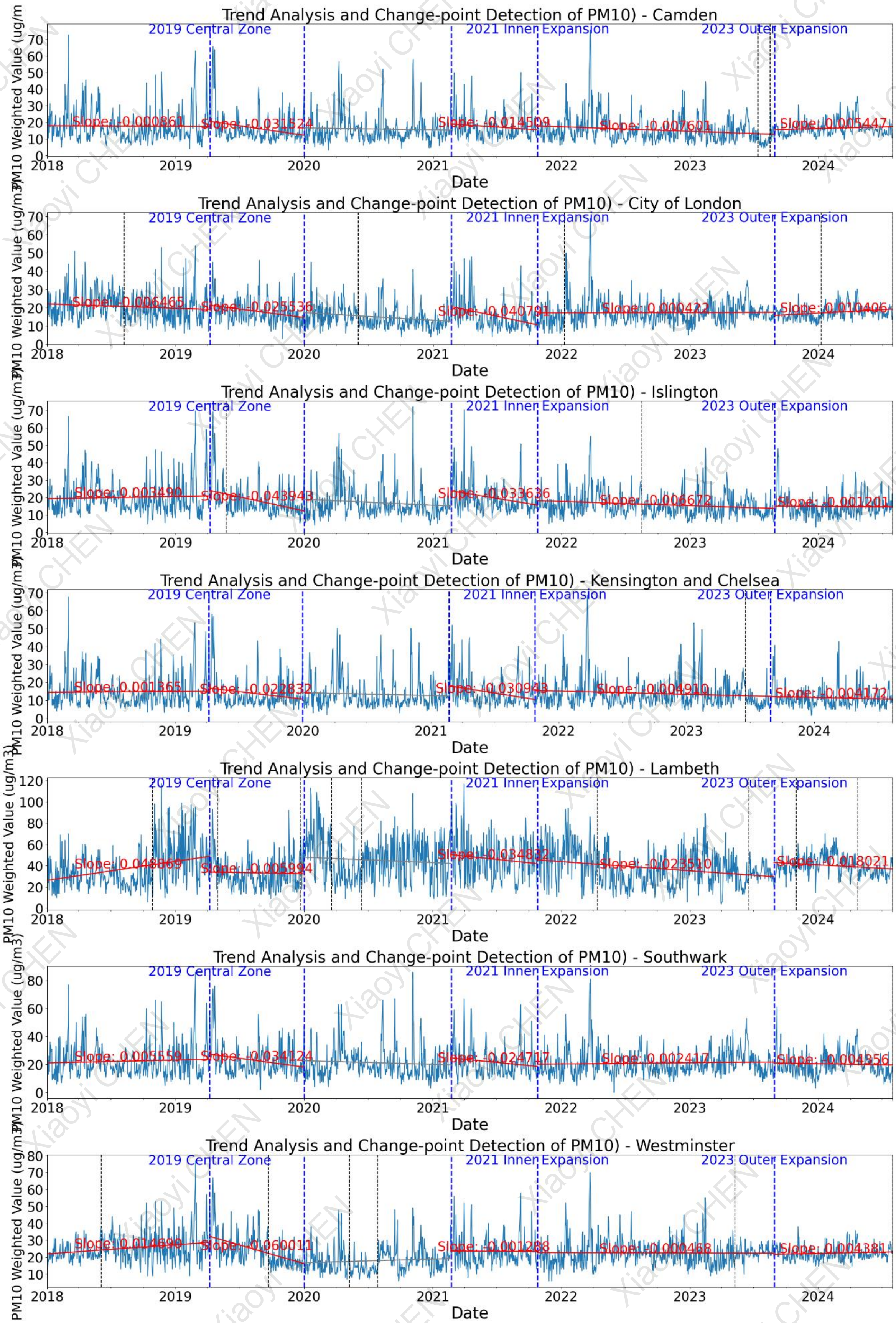
(a)


Trend Analysis and Change-point Detection of weighted Nitrogen dioxide(NO2) - Camden
Trend Analysis and Change-point Detection of weighted Nitrogen dioxide(NO2) - City of London
Trend Analysis and Change-point Detection of weighted Nitrogen dioxide(NO2) - Islington
Trend Analysis and Change-point Detection of weighted Nitrogen dioxide(NO2) - Kensington and Chelsea
Trend Analysis and Change-point Detection of weighted Nitrogen dioxide(NO2) - Lambeth
Trend Analysis and Change-point Detection of weighted Nitrogen dioxide(NO2) - Southwark
Trend Analysis and Change-point Detection of weighted Nitrogen dioxide(NO2) - Westminster
2019 Central Zone
2021 Inner Expansion
2023 Outer Expansion
NO2 Weighted Value (ug/m3)
Date
2018
2019
2020
2021
2022
2023
2024

(b)


Trend Analysis and Change-point Detection of PM2.5) - Camden
2019 Central Zone
2021 Inner Expansion
2023 Outer Expansion
Slope: 0.003496
Slope: -0.020665
Slope: -0.031674
Slope: -0.005400
Slope: 0.002460
PM25 Weighted Value (ug/m3)
Date
Trend Analysis and Change-point Detection of PM2.5) - City of London
Slope: -0.002788
Slope: -0.012283
Slope: -0.018202
Slope: -0.001962
Slope: 0.002461
Trend Analysis and Change-point Detection of PM2.5) - Islington
Slope: -0.007538
Trend Analysis and Change-point Detection of PM2.5) - Kensington and Chelsea
Slope: 0.002525
Slope: -0.016805
Slope: -0.019588
Slope: -0.003839
Slope: -0.004221
Trend Analysis and Change-point Detection of PM2.5) - Lambeth
Slope: 0.000663
Slope: -0.000146
Slope: 0.008831
Slope: -0.001724
Slope: -0.014000
Trend Analysis and Change-point Detection of PM2.5) - Southwark
Slope: 0.000136
Slope: -0.000826
Slope: -0.018457
Slope: -0.003573
Slope: -0.006675
Trend Analysis and Change-point Detection of PM2.5) - Westminster
Slope: 0.012430
Slope: -0.031991
Slope: -0.018976
Slope: -0.000130
Slope: 0.000474
2018
2019
2020
2021
2022
2023
2024

(c)


Trend Analysis and Change-point Detection of PM10) - Camden
Trend Analysis and Change-point Detection of PM10) - City of London
Trend Analysis and Change-point Detection of PM10) - Islington
Trend Analysis and Change-point Detection of PM10) - Kensington and Chelsea
Trend Analysis and Change-point Detection of PM10) - Lambeth
Trend Analysis and Change-point Detection of PM10) - Southwark
Trend Analysis and Change-point Detection of PM10) - Westminster
2019 Central Zone
2021 Inner Expansion
2023 Outer Expansion
PM10 Weighted Value (ug/m3)
Date
2018
2019
2020
2021
2022
2023
2024

(d)

Figure 4.2.1 Trend Analysis and Change-Point Detection of Air Quality [(a)NO, (b)NO2, (c)PM2.5 and (d)PM10] on central London. The vertical blue dotted line is the distinguishing precedent in this study period. The vertical black dotted line represents the timing location of the detected changed point. The red line is the trend of the air pollutant time series in different periods before and after the ULEZ by the linear regression method, and there is a gray line that is the trend of the air pollutant time series during the Covid-19 pandemic.

4.2.1.1 The extent of the impact on air quality in each study area

The trend of NO showed different variations in each region. Before and after the implementation of ULEZ in the Central Region in 2019, several regions such as Lambeth and Southwark showed a downward trend. With the implementation of the expansion in 2021, the NO concentration in all regions showed a slight downward trend, indicating that the policy was effective in controlling NO pollution. After the 2023 expansion, the decrease in NO concentrations was more significant in some regions (e.g., Westminster and Lambeth).

In terms of the change of NO2 concentration, most regions showed a downward trend after each policy stage. After the implementation of the Chuo District in 2019, NO2 concentrations began to decline in most areas. In particular, after the expansion in 2021, the downward trend in NO2 concentrations was particularly pronounced in Camden, Kensington and Chelsea, and Southwark. After the expansion in 2023, NO2 concentrations decreased to varying degrees in all regions.

In terms of PM2.5 concentrations, most regions showed a downward trend after each stage of policy implementation. In particular, after the implementation of Chuo Ward in 2019 and the expansion in 2021, the decrease in PM2.5 concentration was even more pronounced. However, after the outward expansion in 2023, the downward trend of PM2.5 has slowed, and some areas (such as Camden and the City of London) have even shown a slight upward trend.

PM10 concentrations showed a downward trend in most regions, especially after the implementation in the Central District in 2019 and the expansion in 2021. However, after the outward expansion in 2023, the PM10 trend has rebounded in some areas, notably in the City of London and Camden, showing a slight upward trend.

Table 4.2 Trends of Air Categories (NO, NO2, PM2.5, PM10) on ULEZ Policy Implications

	NO	NO2	PM2.5	PM10
2019 Central District Implementation	Most areas showed an upward trend in NO concentrations during this phase, particularly Islington, Westminster and Kensington and Chelsea, possibly due to a temporary increase in traffic flow.	NO2 concentrations declined before and after the implementation of the policy in some regions (e.g., City of London, Kensington and Chelsea), suggesting that the early ULEZ policy was starting to play a role in controlling NO2 emissions.	After this phase, PM2.5 concentrations began to decline in most areas (e.g. City of London and Kensington and Chelsea), indicating that the effect of the policy on PM2.5 control is beginning to appear.	PM10 concentrations showed significant declines in most areas (e.g., Lambeth and Westminster) during this phase, indicating that the early implementation of the policy was effective in controlling PM10.

2021 Intra Expansion	During this period, NO concentrations in most regions tended to stabilize or decreased slightly, indicating that policy control of NO emissions began to emerge.	NO2 concentrations decreased significantly in most areas during this period, especially in Southwark and Kensington and Chelsea, indicating that the expanded implementation of the policy was more effective in controlling NO2.	After this phase, PM2.5 concentrations decreased in all regions, especially in Camden and Lambeth, indicating that the internal expansion had a significant effect on PM2.5 control.	The downward trend of PM10 concentrations during this period became more pronounced in several regions (e.g., Southwark and Islington), indicating that the effect of internal expansion on PM10 control was enhanced.
2023 Expansion	After this phase, NO concentrations decreased significantly in most regions, especially in Westminster and Lambeth, indicating that the policy had a more significant effect on NO control after the expanded implementation.	NO2 concentrations decreased in all regions after this phase, especially in Camden and Lambeth, further indicating the effective control of NO2 after the policy expansion.	The downward trend in PM2.5 concentrations has slowed after this phase, suggesting that the policy expansion may not have as much of an impact on PM2.5 as the previous two phases.	After this phase, PM10 concentrations have shown a slight upward trend in some areas (e.g., Camden and City of London), suggesting that the impact of policy expansion on PM10 may be more complex and may require more time to observe its long-term effects.

4.2.1.2 Summary of ULEZ Policy Impacts Time Series

Effectiveness Across ULEZ Policies

The effectiveness of ULEZ policies varied across different implementation phases and study areas, reflecting a complex interaction between policy measures and local environmental conditions. The 2019 Central Zone implementation had a modest impact on reducing air pollutants like NO2 and PM2.5 in most areas. However, the reductions were not uniform, with some areas, such as the City of London, experiencing stability or slight increases in pollutants. This suggests that the initial phase of ULEZ may not have been sufficiently comprehensive to trigger substantial improvements across all regions.

The 2021 Inner Expansion generally reinforced the trends set by the 2019 policy, contributing to further reductions, particularly in NO2 levels. Despite this, the overall impact appeared moderate, with only minor changes observed in some regions. The limited scope of the 2021 expansion may have resulted in less dramatic shifts, indicating that the benefits of ULEZ policies might be incremental rather than immediate.

The most pronounced effects were observed following the 2023 Outer Expansion, which led to significant reductions in NO and NO2 levels across most areas. However, the impact on PM2.5 and PM10 was less consistent, with some regions experiencing stabilization or even slight increases in these pollutants. This variability suggests that while ULEZ is effective in targeting nitrogen oxides, particulate matter may require additional or alternative strategies for substantial reduction.

Top Three Study Areas with Significant Impact

Westminster, Lambeth, and Southwark emerged as the top three areas where ULEZ policies had the most significant impact. Westminster showed notable reductions in NO levels post-2023, though NO2 trends were less consistent, likely due to its dense downtown area and high traffic volumes. Lambeth exhibited significant downward trends in NO, NO2, and PM2.5 concentrations across all ULEZ stages, with the most marked changes observed post-2023, indicating a high sensitivity to the policy. Southwark consistently demonstrated decreases in NO2 and PM2.5, particularly after the 2023 expansion, highlighting the effectiveness of ULEZ in this region.

Areas with relatively low impact

Other areas including City of London and Camden exhibited less consistent trends, with some increases. Due to City of London high base traffic flow, although the concentration of NO and NO2 has decreased after the implementation of the policy, the overall trend is relatively stable and the change is small. Although the policy phases have some impact on PM10 and NO2 concentrations on Camden, the change is small, and PM2.5 concentrations even show a slight increase in some phases.

Overall, the ULEZ policy has a significant effect on air quality, but the effect varies by region and pollutant type. The expansion in 2021 had the greatest impact on NO and NO2 concentrations, especially in areas such as Lambeth, Southwark and Westminster. In 2019, the implementation of the initial control of PM2.5 and PM10 in the central district has played a positive role, but its effect needs to be further consolidated in the subsequent outward expansion. Although the expansion in 2023 will further cover a larger area, the effect on some pollutants (such as PM2.5 and PM10) remains to be seen. Overall, Southwark, Lambeth and Westminster were the three regions most affected by the ULEZ policy, while Camden and the City of London had a relatively small impact, possibly because they already had high baseline environmental standards or other factors before the policy was implemented.

4.2.2 Difference Between Periods

Unlike the approach in section **4.2.1 Air Quality Trend Analysis on ULEZ changes**, where the Kriging method is used to calculate weighted air quality values, this section analyzes the air quality data directly from actual detectors at specific geographic locations. By observing the changes across the three ULEZ periods, we aim to gain insights into the effectiveness of the ULEZ implementations.

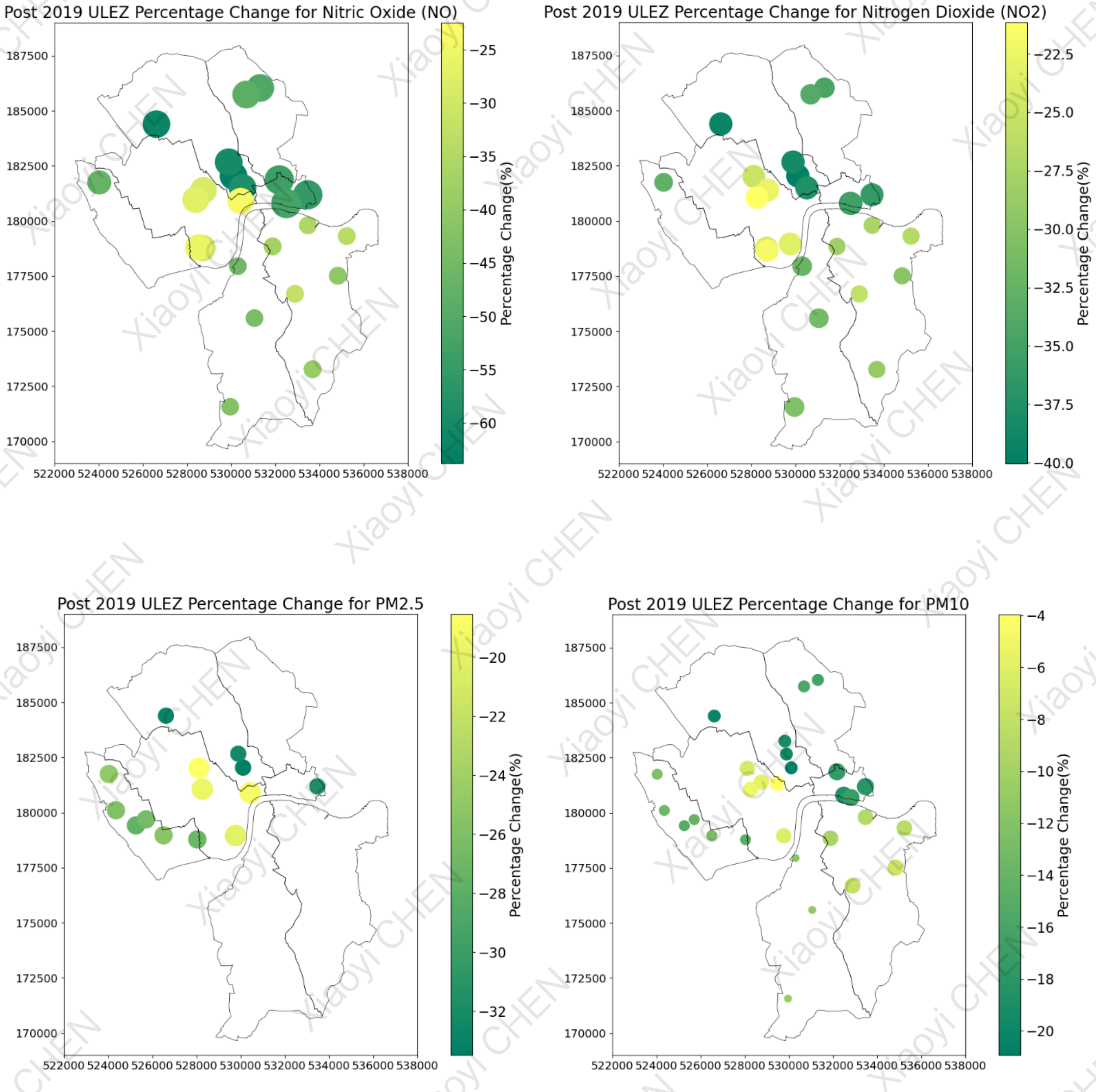
To observe changes in air quality, we calculate the difference in air quality values between successive periods. This subtraction method directly highlights areas where air quality has either improved or worsened due to ULEZ. We define the percentage change between consecutive periods using the formula:

$$Percentage\ Change = \frac{Mean\ Value_{Post-ULEZ} - Mean\ Value_{Pre-ULEZ}}{Mean\ Value_{Pre-ULEZ}} \times 100 \qquad (4.2)$$

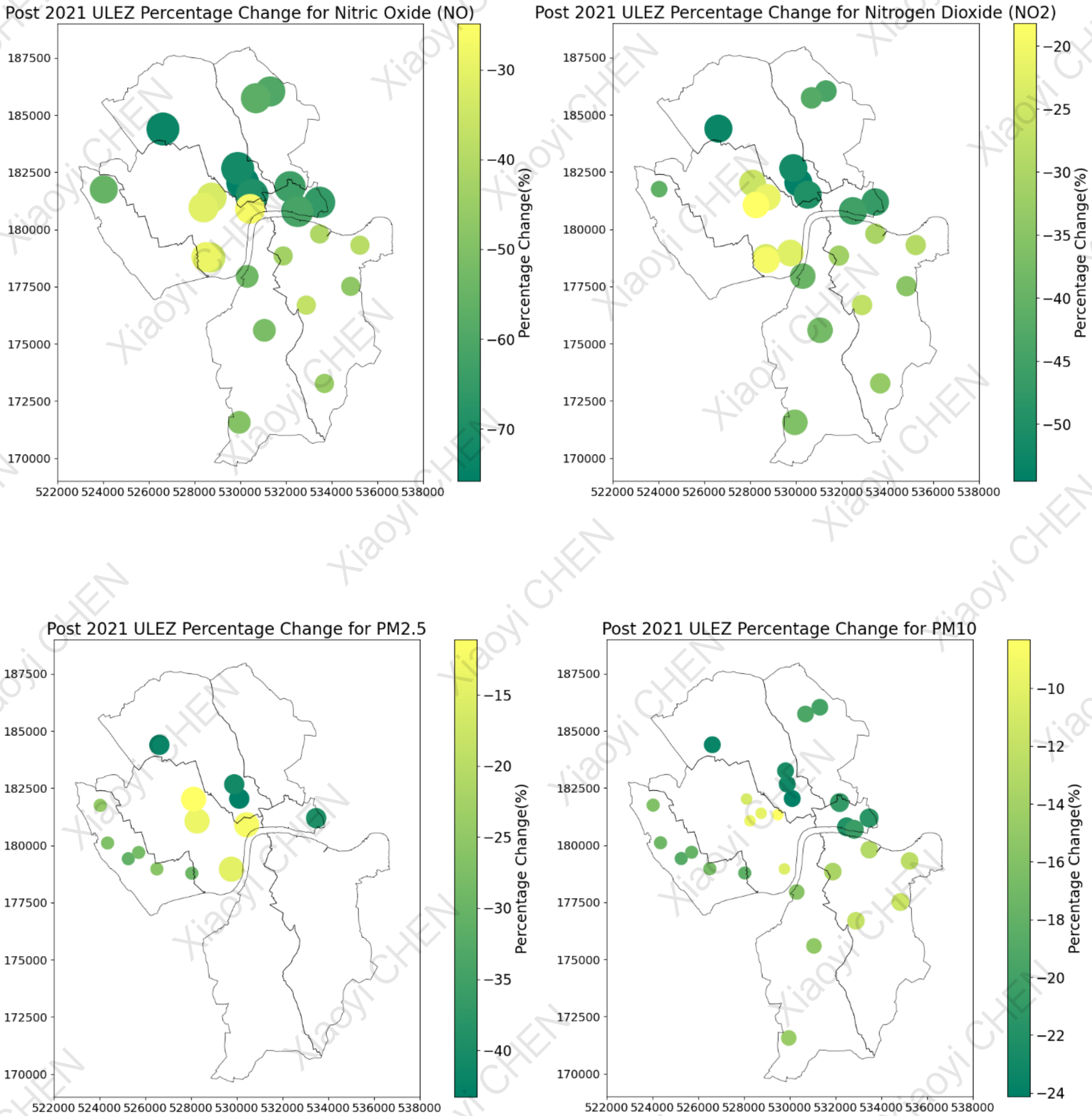
This method is particularly useful for identifying significant changes or anomalies, such as drastic reductions in pollutant levels, which can then be visually represented on difference maps.

Using direct air quality data from detectors instead of the Kriging method is more straightforward and efficient, as it avoids the complexities and potential inaccuracies of spatial interpolation. This method allows for a more accurate and immediate reflection of changes in air quality, ensuring that the analysis is based on observed data rather than model estimates, thereby providing clearer insights into the impact of ULEZ policies.

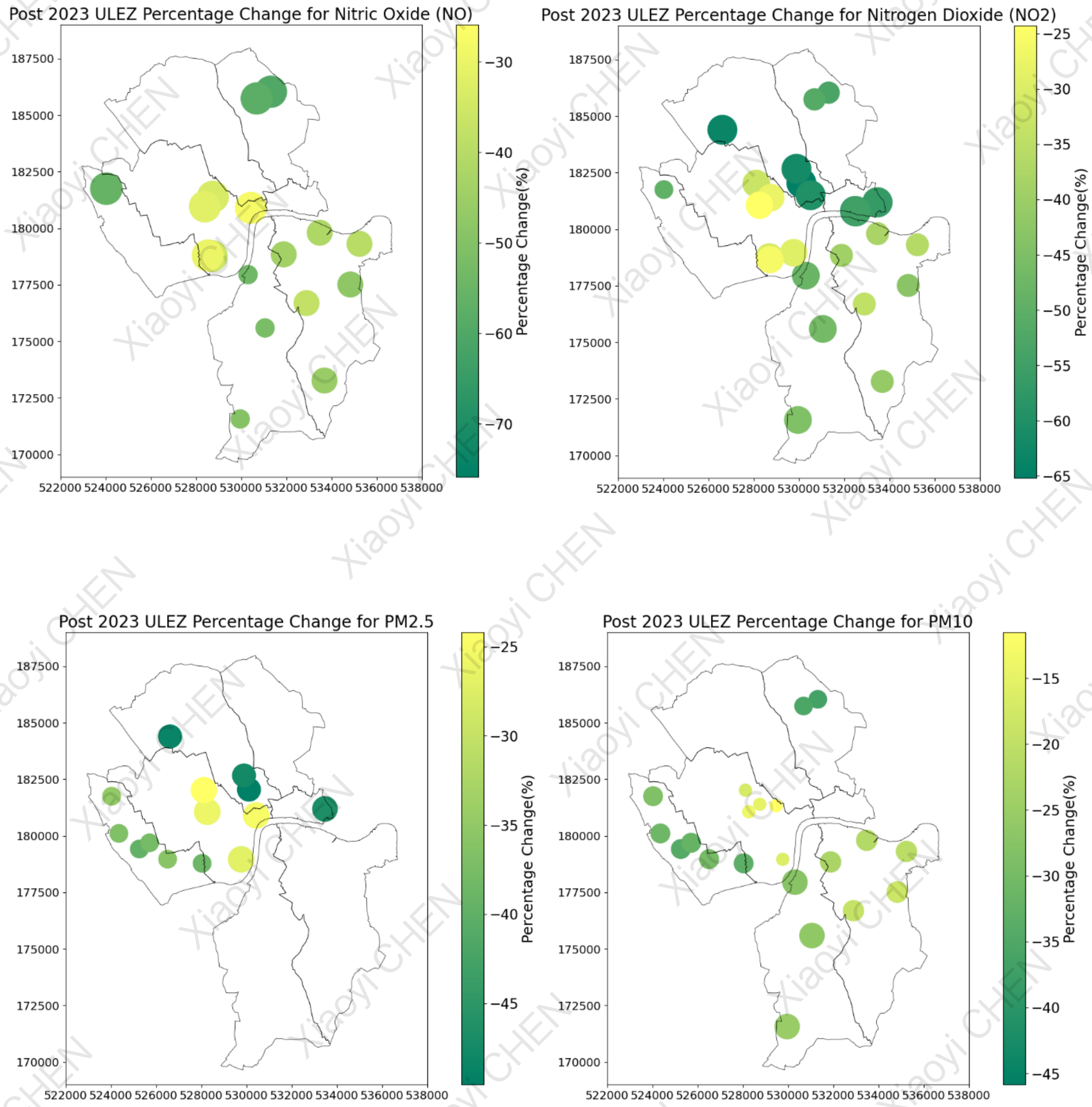
With the calculation of percentage changes and the identification of significant shifts in air quality, we can now move on to the interpretation of the difference maps on **Figure 4.2.2**:



(a) Post 2019 ULEZ on Central Zone



(b) Post 2021 ULEZ of Inner Expansion



(c) Post 2023 ULEZ on Outer Expansion

Figure 4.2.2 Difference change percentage of detectors' Air quality mean value in central London

Each ULEZ implementation has led to a reduction in air pollutants across the regions. The most significant changes are observed post the 2023 ULEZ expansion, indicating that as the ULEZ coverage area expanded, the impact on air quality became more pronounced.

As for the magnitude of changes over time, initially, the 2019 ULEZ implementation resulted in relatively modest changes in air quality. However, subsequent expansions, particularly the 2023

ULEZ, showed much larger percentage changes. This trend suggests that as the ULEZ boundary expanded from inner zones to outer areas, the improvements in air quality became more substantial.

Then observing the areas with significant changes, City of London, Camden, and Kensington and Chelsea experienced the most significant air quality improvements. Despite the gradual changes observed in the time series analysis for these areas, the overall difference in air quality before and after ULEZ implementation is quite pronounced. Nevertherless, Westminster, Lambeth, and Southwark, on the other hand, showed less dramatic changes in air quality difference. Even though these areas showed notable trends in the time series analysis, their overall change percentage is lower. This could be due to the higher baseline levels of pollution, making percentage changes appear less significant despite actual reductions.

Finally, we can know the pollutants most affected by ULEZ is Nitric Oxide (NO) and Nitrogen Dioxide (NO2), which saw the most significant reductions, followed by PM2.5 and PM10. This pattern is consistent with the findings from the Air Quality Trend Analysis, highlighting the particular sensitivity of these pollutants to ULEZ policies.

4.3 Relationship among Air Pollution and Traffic Types

4.3.1 Correlation Analysis

In examining the relationship between air pollution and various traffic types, it is crucial to understand how different modes of transportation contribute to the levels of pollutants such as NO, NO2, PM2.5, and PM10 in urban environments. Correlation analysis provides valuable insights into these dynamics by quantifying the strength and direction of associations between traffic volumes and air quality parameters. By identifying these correlations, we can better comprehend the impact of specific vehicle types on pollution levels, thereby informing targeted policy measures to improve air quality. The following sections delve into the correlation analysis between air pollution and traffic types over the central London.

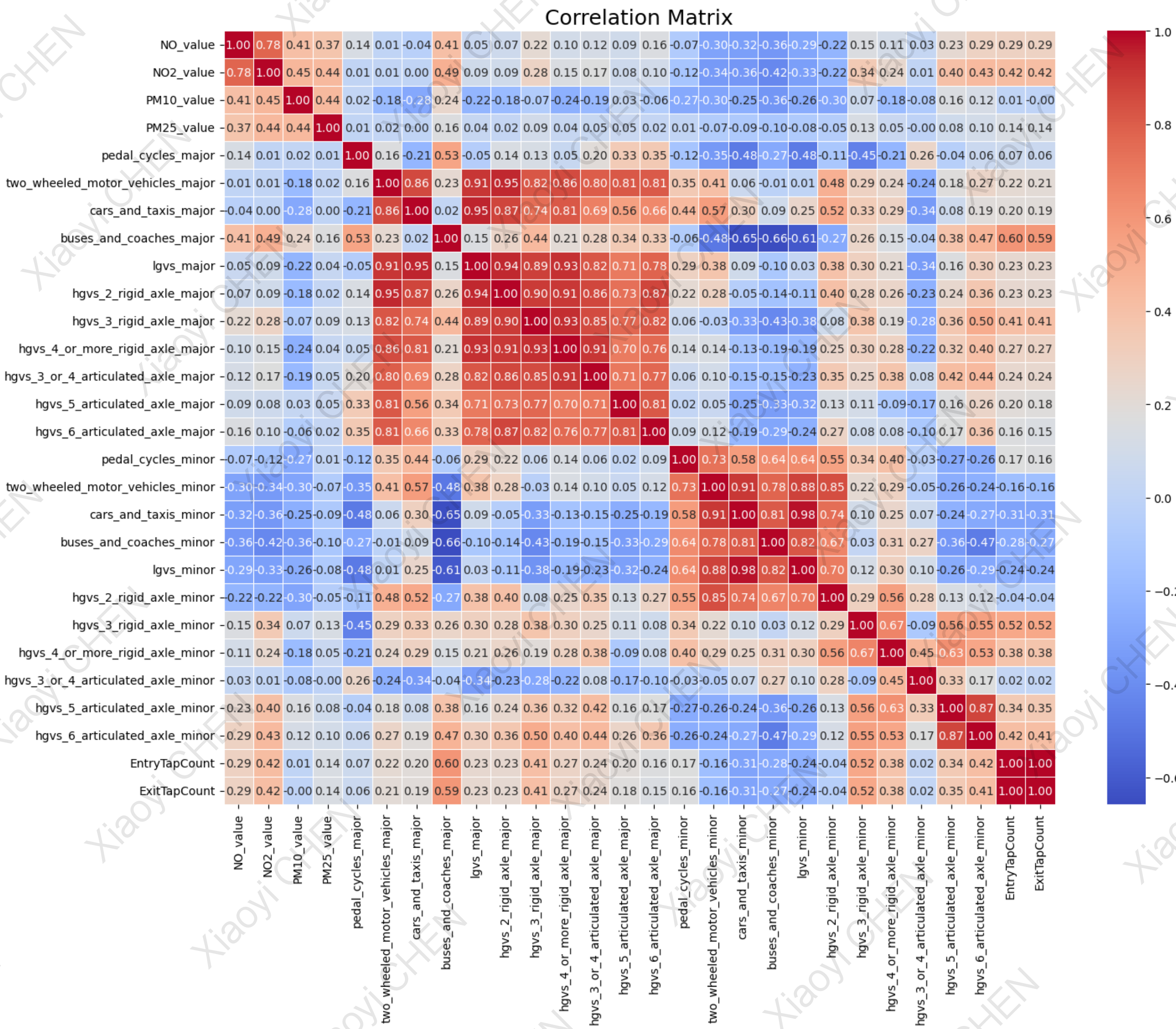


Figure 4.3 Correlation matrices between air pollutants and traffic types. `EntryTapCount` and `ExitTapCount` represent the entry and exit flow of London metro subways station.

Based on the correlation table provided, here are the key relationships between traffic factors and air quality indicators (NO, NO2, PM10, PM2.5). These correlations highlight the factors that have a significant positive or negative impact on air quality:

Table 4.3 Correlation Analysis of Air Categories(NO, NO2, PM2.5, PM10)

	NO_value	NO2_value	PM2.5_value	PM10_value
Positive Correlations (r > 0.2)	**buses_and_coaches_major (0.41)**: This indicates that as the number of buses and coaches on major roads increases, the NO (Nitric Oxide) levels tend to rise. This suggests that emissions from these vehicles are a significant contributor to NO pollution.	**buses_and_coaches_major (0.49)**: A strong positive correlation suggests that more buses and coaches on major roads are associated with higher NO2 levels. This further supports the idea that diesel-powered buses and coaches are key contributors to nitrogen dioxide pollution. **EntryTapCount (0.42)**: This correlation suggests that higher public transport usage, as indicated by entry tap counts at stations, is associated with increased NO2 levels. This may be due to higher traffic and associated emissions in areas with heavy public transport usage.	**buses_and_coaches_major (0.16)**: While this correlation is slightly below 0.2, it still shows a positive trend, indicating that buses and coaches on major roads may contribute to PM2.5 levels.	**buses_and_coaches_major (0.24)**: A positive correlation indicates that an increase in buses and coaches on major roads is linked with higher PM10 levels (particulate matter less than 10 micrometers). This suggests that these vehicles contribute to particulate pollution.
Negative Correlations (r < -0.2)	**buses_and_coaches_minor (-0.36)**: This suggests that on minor roads, an increase in buses and coaches is associated with lower NO levels. This might be due to less congestion and more efficient fuel combustion on these roads. **cars_and_taxis_minor (-0.32)**: Similarly, more cars and taxis on minor roads are associated with lower NO levels, possibly due to reduced congestion and idling.	**buses_and_coaches_minor (-0.42)**: A strong negative correlation suggests that more buses and coaches on minor roads are linked with lower NO2 levels, likely due to similar reasons as NO.	**buses_and_coaches_minor (-0.10)**: This is a weak negative correlation, indicating a slight trend where more buses and coaches on minor roads are associated with lower PM2.5 levels.	**buses_and_coaches_minor (-0.36)**: This suggests that an increase in buses and coaches on minor roads is associated with lower PM10 levels, possibly due to better traffic flow and less idling.

The analysis indicates that buses and coaches on major roads are key contributors to elevated levels of NO, NO2, and PM10, likely due to their diesel engines and higher emissions rates. In contrast, on minor roads, an increase in vehicles—especially buses, coaches, and cars—is paradoxically linked to lower pollution levels. This may be attributed to reduced congestion, which allows for more efficient fuel combustion and thus lower emissions.

From **Figure 4.3** Correlation matrices and **Table 4.3** Correlation Analysis of different air pollutants, we can recommend that targeted emission reduction policies, particularly focusing on buses and coaches on major roads, could significantly reduce NO and NO2 pollution levels. Additionally, the unexpected negative correlations observed on minor roads warrant further investigation to better understand the underlying mechanisms. This could provide valuable insights for shaping traffic management strategies and improving air quality in urban areas.

4.3.2 Regression Analysis

Building on the correlation analysis, we now delve into the regression analysis, which quantifies the relationship between various traffic variables and air pollutant levels (NO, NO2, PM10, and PM2.5).

$$Value_{NO} =$$
$$4.8144 + (0.5234 * Buses\&Coaches_{MajorRoad}$$
$$+(-1.3989 * Buses\&Coaches_{MinorRoad})$$
$$+(0.0042 * Cars\&Taxis_{MinorRoad})$$
$$+\left(0.0001 * Metro_{Entry}\right) + (-0.0001 * Metro_{Exit})$$

$$Value_{NO2} =$$
$$20.1872 + \left(0.4291 * Buses\&Coaches_{MajorRoad}\right)$$
$$+(-1.9009 * Buses\&Coaches_{MinorRoad})$$
$$+(0.0363 * Cars\&Taxis_{MinorRoad})$$
$$+\left(0.0000 * Metro_{Entry}\right) + (0.0000 * Metro_{Exit})$$

$$Value_{PM2.5} =$$
$$5.8488 + \left(0.1021 * Buses\&Coaches_{MajorRoad}\right)$$
$$+(-0.1041 * Buses\&Coaches_{MinorRoad})$$
$$+(0.0082 * Cars\&Taxis_{MinorRoad})$$
$$+\left(0.0000 * Metro_{Entry}\right) + (-0.0000 * Metro_{Exit})$$

$$Value_{PM10} =$$
$$24.5807 + \left(0.2308 * Buses\&Coaches_{MajorRoad}\right)$$
$$+(-2.3836 * Buses\&Coaches_{MinorRoad})$$
$$+(0.0446 * Cars\&Taxis_{MinorRoad})$$
$$+ (0.0001 * Metro_{Entry}) + (-0.0001 * Metro_{Exit})$$

The regression formulas derived offer insights into how specific factors contribute to the concentration of these pollutants in the study areas. Below are the regression formulas for each pollutant, which will be used to further understand the impact of traffic dynamics on air quality.

5 Discussion

5.1 Critical Evaluation of Methodology

Data Gaps

One of the primary challenges encountered during this study was the issue of non-continuous traffic flow data, particularly the lack of data for the City of London in minor roads. This limitation restricted the scope of analysis and potentially affected the robustness of the findings. The absence of data in these areas meant that some traffic patterns could not be fully analyzed, which may have influenced the conclusions drawn about the impact of the ULEZ on air quality in specific regions.

Kriging Method for Weighted Air Quality Calculation

The initial motivation for using the Kriging method was to obtain a daily weighted air quality value for specific regions. This approach accounted for the geographic dispersion of air quality detectors within an area and incorporated wind speed and direction as key meteorological factors influencing air quality distribution. However, Kriging has its limitations. One significant drawback is that while Kriging excels in spatial interpolation, it may not always capture the complex dynamics of air pollution, especially in highly urbanized areas where pollution sources are varied and not evenly distributed. Alternative air quality dispersion models as mention in Holmes and Morawska (2006) book, such as Gaussian plume models or computational fluid dynamics (CFD) models, may offer more accurate simulations of pollutant dispersion in urban environments. These models are particularly effective in considering the intricacies of building layouts, traffic flow, and other urban infrastructure, which can significantly affect air pollution distribution.

Variability in Air Quality Trends

In this study, air quality time series analysis was conducted using Linear Regression, specifically Ordinary Least Squares (OLS) and Linear Least Squares methods. While these methods are straightforward and widely used, they have limitations in accurately reflecting real-world air quality trends. As observed in **Figure 4.2**, each air quality category exhibits clear annual cyclical patterns that do not align with the ULEZ implementation periods. Linear regression, being a simplistic approach, only provides a linear trend, which can be overly reductive in the presence of such cyclical behavior. To better capture the true trends in air quality data, Nonlinear Regression methods could be employed. For instance, Seasonal Decomposition of Time Series (STL), Fourier series models, or Generalized Additive Models (GAMs) (Contreras and Ferri, 2016, Hyndman, 2018, Szidarovszky et al., 1987, Terzi and Cengiz, 2009, Wen et al., 2020) could be more suitable. These methods allow for the incorporation of periodic components and can better fit the annual cycles observed in air quality data. Such models offer a more nuanced understanding of how ULEZ and other factors influence air quality over time, particularly in complex urban environments like Central London.

5.2 Limitations of the Study

5.2.1 Acknowledgment of Constraints

One of the major constraints of this study is the uneven geographic distribution of air quality detectors across Central London. The concentration of detectors in specific regions means that air quality measurements are more reflective of conditions in those regions, potentially overlooking pollution levels in less-monitored areas. This limitation could introduce bias into the analysis, as areas with fewer detectors may not be accurately represented in the final air quality assessment.

However, emerging technologies in environmental monitoring, particularly the development of low-cost micro-scale sensors, present a potential solution to this issue. As highlighted by Kumar et al. (2015), the conventional approach to air quality monitoring—relying on static and sparse measurement stations—can be prohibitively expensive and may fail to capture the full spatial and temporal variability of urban pollution. This limitation underscores the need for more comprehensive and dynamic monitoring solutions that can provide real-time, fine-grained data across a wider area.

In the future, incorporating these low-cost sensors into the existing monitoring network could help mitigate the geographical limitations observed in this study. By providing more evenly distributed and real-time data, these sensors could enhance the accuracy of air quality assessments, particularly in less-monitored regions. This would allow for a more comprehensive understanding of air pollution patterns and better inform urban policy decisions aimed at reducing exposure and improving public health.

5.2.2 Potential Biases

Potential biases introduced by the data collection process, such as the selection of study areas and the time periods chosen for analysis, also warrant discussion. The data gaps identified earlier in Chapter ***3.1.2 Data Feasibility Analysis*** could lead to biased conclusions, particularly if the missing data coincided with periods of significant changes in traffic or air quality. The solutions to these gaps, discussed in Chapter ***3.2.1 Missing Value Interpolation Processing***, included the use of spline interpolation for larger gaps in the Road traffic statistics dataset and polynomial interpolation for smaller gaps in air quality data. The solutions to these gaps, discussed in Chapter ***3.2.1 Missing Value Interpolation Processing***, included the use of spline interpolation for larger gaps in the road traffic statistics dataset and polynomial interpolation for smaller gaps in air quality data.

5.2.3 Considering Additional Influencing Factors

When discussing the impact of the ULEZ on central London's air quality, it's crucial to recognize that while the ULEZ targets vehicle emissions, this focus might be too narrow. The policy adjusts vehicle-related charges and restricts certain types of traffic, aiming to reduce pollutants such as NO, NO_2 , PM2.5, and PM10. However, by concentrating on these specific pollutants, the study potentially overlooks other significant factors that can influence air quality in an urban environment.

Nathvani et al. (2023) illustrate that people and marketplace activities are also critical in determining air quality. This suggests that public behavior, including pedestrian density and commercial activities, can contribute to pollution levels in ways not directly related to traffic. The absence of these variables in our analysis is a key limitation, as it means this study may not fully account for all contributors to air quality in central London.

Furthermore, while our study focused on pollutants directly linked to vehicle emissions and meteorological conditions, it did not include an analysis of object feature importance, which could provide deeper insights into the various factors impacting air quality. Danesh Yazdi et al. (2020) conducted a comprehensive examination of object feature importance, considering a broader range of variables, such as proximity to bus stops, building heights, and population density, among others. Their findings underscore the complexity of air quality determinants and highlight that limiting the analysis to traffic-related factors and basic meteorological data might oversimplify the issue.

This omission suggests that our study's conclusions about the ULEZ's effectiveness might be incomplete. Future research should consider a more holistic approach, incorporating a wider array of factors that can influence air quality. By doing so, it would provide a more comprehensive understanding of the impacts of ULEZ and similar policies, leading to more informed policy decisions and interventions.

5.4 Future Work

The analysis conducted in this dissertation provides valuable insights into the impact of ULEZ policies on air quality in Central London. However, several areas warrant further investigation and methodological improvements to deepen our understanding and enhance the effectiveness of future research.

5.4.1 Further analysis

Spatial Cluster Analysis

To further understand the spatial distribution of air pollution in Central London, cluster analysis was employed to identify regions with similar air quality and traffic characteristics. The primary objective of this analysis was to pinpoint areas with consistently high or low pollution levels. K-means clustering, a widely used method in spatial analysis, was applied to group the study areas based on their air quality and traffic data. By clustering these regions, the analysis revealed significant patterns, such as areas where high traffic volumes correlate with elevated pollution levels. These clusters serve as critical zones for targeted policy interventions, enabling more efficient allocation of resources and the implementation of measures to reduce pollution in the most affected areas.

Logistic Regression Analysis

The regression analysis aimed to quantify the relationship between the implementation of ULEZ policies and improvements in air quality. Specifically, logistic regression was used to model the probability of air quality improvement as a function of several variables, including traffic volume, public transport usage, and the stages of ULEZ implementation. This approach provided a statistical framework to assess the effectiveness of the ULEZ policies over time. The results indicated a significant correlation between ULEZ expansions and the likelihood of reduced pollution levels, offering strong evidence of the policy's impact. By modeling these relationships, the analysis not only confirmed the benefits of the ULEZ but also highlighted the potential areas where further improvements could be made, thereby guiding future policy decisions.

5.4.2 Further Investigation

Long-term Monitoring

While this study provided a snapshot of air quality changes pre- and post-ULEZ, long-term monitoring is essential to assess the sustained impact of these policies. Specifically, future research should focus on Camden and the City of London, where the ULEZ's impact on air quality was relatively small. A more extended observation period might reveal more subtle trends and longer-term improvements that were not immediately apparent in the current analysis.

Moreover, Westminster, despite showing significant improvements due to ULEZ implementation, continues to face air quality challenges. A more granular study focusing on Westminster could identify persistent sources of pollution and the reasons behind the area's ongoing struggles with air quality. Such research could inform targeted interventions to address these specific issues.

Granular Studies on Pollutants and Vehicles

Building on the correlation and multiple linear regression analyses, future work could focus on the individual pollutants or specific types of vehicles that have the most significant impact on air quality. This approach would enable a more detailed understanding of the relationship between vehicle emissions and specific pollutants, potentially leading to more effective policy measures. For instance, if certain types of vehicles, such as heavy goods vehicles (HGVs) or buses, are found to be major contributors to NO2 or PM10 levels, stricter regulations or incentives for cleaner technologies could be implemented.

Exploring ULEZ Benefits Beyond Air Quality

Finally, future research could explore the broader benefits of ULEZ beyond air quality improvements. This could include examining the policy's impact on public health, traffic congestion, and economic factors such as property values and business activity in Central London. Understanding these additional benefits could provide a more comprehensive evaluation of ULEZ's overall effectiveness and inform the design of similar policies in other urban areas.

6 Conclusion

This study comprehensively explored the effects of the Ultra-Low Emission Zone (ULEZ) policies on air quality across different boroughs of Central London from 2019 to 2023. By employing a range of analytical techniques, including exploratory data analysis, time series analysis, and regression modeling, the research has highlighted both the successes and challenges associated with ULEZ implementation.

Significant Impact of ULEZ on Air Quality

The findings underscore the effectiveness of ULEZ policies in reducing concentrations of Nitric Oxide (NO) and Nitrogen Dioxide (NO2), particularly following the 2021 and 2023 expansions. These later phases of the ULEZ, which extended coverage to broader areas, demonstrated marked improvements in air quality, especially in outer zones that were newly included. The 2023 expansion, in particular, led to a substantial drop in NO and NO2 levels, affirming the importance of broad coverage and stringent enforcement in achieving air quality objectives.

Regional Variations in ULEZ Effectiveness

The impact of ULEZ policies, however, was not uniform across all boroughs or pollutants. While Westminster and Lambeth showed consistently high levels of NO and NO2, reflecting the persistent challenges posed by dense traffic and urban activity, boroughs such as Camden, City of London, and Kensington and Chelsea exhibited notable reductions in these pollutants over time. This suggests that while some areas responded rapidly to the ULEZ interventions, others, particularly those with heavier traffic, may require additional measures or more time to experience similar benefits.

In contrast, particulate matter (PM2.5 and PM10) showed more varied responses, with some boroughs like Lambeth experiencing increases in PM10 levels post-2023 ULEZ, indicating the potential influence of non-traffic-related sources such as construction activities. This anomaly suggests that while ULEZ policies are effective in addressing vehicle emissions, they may need to be supplemented by other regulations targeting different sources of particulate pollution.

Traffic Patterns and Pollution Dynamics

The regression analysis provided further insights into the relationship between traffic and air quality, particularly highlighting the role of buses and coaches on major roads as significant contributors to NO, NO2, and PM10 pollution. Interestingly, on minor roads, an increase in vehicle numbers was associated with lower pollution levels, possibly due to reduced congestion and more efficient traffic flow. This finding suggests that urban traffic management strategies that reduce congestion could have a beneficial impact on air quality, even in high-traffic areas.

Implications for Future Policy

The results of this study indicate that while ULEZ has been successful in improving air quality, especially concerning nitrogen oxides, the varying impacts across pollutants and regions point to the need for a more nuanced approach. Future policies should consider regional differences, targeting areas with persistent pollution challenges and addressing a wider range of pollutants. Moreover, ongoing monitoring and adaptation of ULEZ and similar policies will be essential to ensure their continued effectiveness in an evolving urban environment.

Overall, the ULEZ policy has been effective in improving air quality in Central London, particularly in reducing harmful nitrogen oxides. However, the varying impact across different pollutants and regions suggests that a one-size-fits-all approach may not be sufficient. Future policies should consider these nuances, perhaps by incorporating stricter controls on specific vehicle types or by addressing other sources of particulate matter. Additionally, the study highlights the importance of continued monitoring and analysis to assess the long-term effectiveness of ULEZ and similar initiatives, ensuring that they adapt to changing urban dynamics and continue to protect public health.

Reference

Ashari, I. A. and Purwono, I. A. (2023) 'Monitoring the pH Levels of Well Water in the Home Industri Sarung Goyor Village, Pemalang, Using IoT Technology and Inverse Distance Weight Method', *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI)*, 9(4), pp. 1119-1130.

Bayraktar, H. and Turalioglu, F. S. (2005) 'A Kriging-based approach for locating a sampling site—in the assessment of air quality', *Stochastic Environmental Research and Risk Assessment*, 19(4), pp. 301-305. DOI: 10.1007/s00477-005-0234-8.

Beauchamp, M., Malherbe, L., de Fouquet, C., Létinois, L. and Tognet, F. (2018) 'A polynomial approximation of the traffic contributions for kriging-based interpolation of urban air quality model', *Environmental Modelling & Software*, 105, pp. 132-152.

Beers, W. C. M. v. and Kleijnen, J. P. C. 'Kriging interpolation in simulation: a survey'. *Proceedings of the 2004 Winter Simulation Conference, 2004.*, 5-8 Dec. 2004, 1-121.

Bruzzone, F. and Nocera, S. 'Issues in Modelling Traffic-Related Air Pollution: Discussion on the State-Of-The-Art'. *Computational Science and Its Applications – ICCSA 2021*, Cham, 2021//: Springer International Publishing, 337-349.

Chen, P.-C. and Lin, Y.-T. (2022) 'Exposure assessment of PM2.5 using smart spatial interpolation on regulatory air quality stations with clustering of densely-deployed microsensors', *Environmental Pollution*, 292, pp. 118401. DOI: https://doi.org/10.1016/j.envpol.2021.118401.

Chilès, J.-P. and Desassis, N. (2018) 'Fifty Years of Kriging', in Daya Sagar, B.S., Cheng, Q. & Agterberg, F. (eds.) *Handbook of Mathematical Geosciences: Fifty Years of IAMG*. Cham: Springer International Publishing, pp. 589-612.

Chung, S. Y., Venkatramanan, S., Elzain, H. E., Selvam, S. and Prasanna, M. V. (2019) 'Chapter 4 - Supplement of Missing Data in Groundwater-Level Variations of Peak Type Using Geostatistical Methods', in Venkatramanan, S., Prasanna, M.V. & Chung, S.Y. (eds.) *GIS and Geostatistical Techniques for Groundwater Science*: Elsevier, pp. 33-41.

Contreras, L. and Ferri, C. (2016) 'Wind-sensitive interpolation of urban air pollution forecasts', *Procedia Computer Science*, 80, pp. 313-323.

Danesh Yazdi, M., Kuang, Z., Dimakopoulou, K., Barratt, B., Suel, E., Amini, H., Lyapustin, A., Katsouyanni, K. and Schwartz, J. (2020) 'Predicting Fine Particulate Matter (PM2.5) in the Greater London Area: An Ensemble Approach using Machine Learning Methods', *Remote Sensing*, 12(6), pp. 914. Available at: https://www.mdpi.com/2072-4292/12/6/914.

Department for Transport, M. o. L. (2018) 'London Road Traffic Statistics'. Available at: https://roadtraffic.dft.gov.uk/regions/6 (Accessed: July 15, 2024).

European Environment Agency (2013) *Every breath we take - Improving air quality in Europe* (978-92-9213-363-4. Available at: https://www.eea.europa.eu/publications/eea-signals-2013 (Accessed: June 24, 2024).

Greater London Authority (2003) 'Statistical GIS Boundary Files for London'. Available at: https://data.london.gov.uk/dataset/statistical-gis-boundary-files-london (Accessed: June 15, 2024).

Greater London Authority (2023) 'London Atmospheric Emissions Inventory (LAEI) 2019'. Available at: https://data.london.gov.uk/dataset/london-atmospheric-emissions-inventory--laei--2019 (Accessed 30 July, 2024).

Hartkamp, A. D., De Beurs, K., Stein, A. and White, J. W. (1999) *Interpolation techniques for climate variables.* Cimmyt.

Hemelrij.J (1965) *HANDBOOK OF MATHEMATICAL FUNCTIONS WITH FORMULAS GRAPHS AND MATHEMATICAL TABLES. Revue De L Institut International De Statistique-Review of the International Statistical Institute*.

Holman, C., Harrison, R. and Querol, X. (2015) 'Review of the efficacy of low emission zones to improve urban air quality in European cities', *Atmospheric Environment*, 111, pp. 161-169. DOI: https://doi.org/10.1016/j.atmosenv.2015.04.009.

Holmes, N. S. and Morawska, L. (2006) 'A review of dispersion modelling and its application to the dispersion of particles: An overview of different dispersion models available', *Atmospheric Environment*, 40(30), pp. 5902-5928. DOI: 10.1016/j.atmosenv.2006.06.003.

Hyndman, R. (2018) *Forecasting: principles and practice.* OTexts.

Imperial College London (2018) 'Air Quality Data for London - Local Air Quality Management Monitoring Statistics'. Available at: https://www.londonair.org.uk/london/asp/publicstats.asp? (Accessed: June 22, 2024).

Imperial College London (2024) 'Breathe London air quality data'. Available at: https://www.breathelondon.org/ (Accessed: June 24, 2024).

Institute For Government (2022) *Timeline of UK government coronavirus lockdowns and restrictions - Key lockdowns and measure introduced during the Covid pandemic between March 2020 and December 2021*. Available at: https://www.instituteforgovernment.org.uk/data-visualisation/timeline-coronavirus-lockdowns (Accessed: June 15, 2024).

Khreis, H., Nieuwenhuijsen, M. J., Zietsman, J. and Ramani, T. (2020) 'Chapter 1 - Traffic-related air pollution: Emissions, human exposures, and health: An introduction', in Khreis, H., Nieuwenhuijsen, M., Zietsman, J. & Ramani, T. (eds.) *Traffic-Related Air Pollution*: Elsevier, pp. 1-21.

Kousis, I., Pigliautile, I. and Pisello, A. L. (2021) 'Intra-urban microclimate investigation in urban heat island through a novel mobile monitoring system', *Scientific Reports*, 11(1), pp. 9732. DOI: 10.1038/s41598-021-88344-y.

Kumar, P., Morawska, L., Martani, C., Biskos, G., Neophytou, M., Di Sabatino, S., Bell, M., Norford, L. and Britter, R. (2015) 'The rise of low-cost sensing for managing air pollution in cities', *Environment International*, 75, pp. 199-205. DOI: 10.1016/j.envint.2014.11.019.

London, T. f. (2018) 'Coordinates of London Underground stations (including Elizabeth Line stations)'. Available at: https://tfl.gov.uk/corporate/transparency/freedom-of-information/foi-request-detail?referenceId=FOI-1451-1819 (Accessed: July 20, 2024).

Matheron, G. (1971) *The Theory of Regionalized Variables and Its Applications.* École national supérieure des mines, p. 211.

Mayer of London (2023) *Inner London Ultra Low Emission Zone Expansion One Year Report* [General]: Mayer of London. Available at: https://www.london.gov.uk/programmes-strategies/environment-and-climate-change/environment-and-climate-change-publications/inner-london-ultra-low-emission-zone-expansion-one-year-report (Accessed: June 28, 2024).

Mills, G. (2008) 'Luke Howard and The Climate of London', *Weather*, 63(6), pp. 153-157. DOI: https://doi.org/10.1002/wea.195.

Nathvani, R., D, V., Clark, S. N., Alli, A. S., Muller, E., Coste, H., Bennett, J. E., Nimo, J., Moses, J. B., Baah, S., Hughes, A., Suel, E., Metzler, A. B., Rashid, T., Brauer, M., Baumgartner, J., Owusu, G., Agyei-Mensah, S., Arku, R. E. and Ezzati, M. (2023) 'Beyond here and now: Evaluating pollution estimation across space and time from street view images with deep learning', *Science of The Total Environment*, 903, pp. 166168. DOI: https://doi.org/10.1016/j.scitotenv.2023.166168.

NOAA & Deutscher Wetterdienst & Environment Canada (2016) 'London Weather Centre Data'. Available at: https://meteostat.net/en/place/gb/holborn?s=03779&t=2022-01-01/2022-12-31 (Accessed: June 22, 2024).

Ozelkan, E., Chen, G. and Ustundag, B. B. (2016) 'Spatial estimation of wind speed: a new integrative model using inverse distance weighting and power law', *International Journal of Digital Earth*, 9(8), pp. 733-747. DOI: 10.1080/17538947.2015.1127437.

Papritz, A. and Stein, A. (2002) 'Spatial prediction by linear kriging', in Stein, A., Van der Meer, F. & Gorte, B. (eds.) *Spatial Statistics for Remote Sensing*. Dordrecht: Springer Netherlands, pp. 83-113.

Press, W. H. and Teukolsky, S. A. (2007) *Numerical recipes with source code CD-ROM 3rd edition: the art of scientific computing.* Cambridge university press.

Ricardo Energy & Environment (2020) 'Greater London Monitoring Data'. Available at: https://www.airqualityengland.co.uk/local-authority/?la_id=999 (Accessed: July 15, 2024).

Shad, R., Mesgari, M. S., abkar, A. and Shad, A. (2009) 'Predicting air pollution using fuzzy genetic linear membership kriging in GIS', *Computers, Environment and Urban Systems*, 33(6), pp. 472-481. DOI: https://doi.org/10.1016/j.compenvurbsys.2009.10.004.

Shukla, K., Kumar, P., Mann, G. S. and Khare, M. (2020) 'Mapping spatial distribution of particulate matter using Kriging and Inverse Distance Weighting at supersites of megacity Delhi', *Sustainable Cities and Society*, 54, pp. 101997. DOI: https://doi.org/10.1016/j.scs.2019.101997.

Shumway, R. and Stoffer, D. (2011) *Time Series Analysis and Its Applications With R Examples.*

Snoun, H., Krichen, M. and Chérif, H. (2023) 'A comprehensive review of Gaussian atmospheric dispersion models: current usage and future perspectives', *Euro-Mediterranean Journal for Environmental Integration*, 8(1), pp. 219-242. DOI: 10.1007/s41207-023-00354-6.

Stein, M. L. (1999) *Interpolation of Spatial Data: Some Theory for Kriging*. Springer Series in Statistics.

Susan Anenberg, J. M., Daven Henze, Ray Minjares (2019) *A global snapshot of the air pollution-related health impacts of transportation sector emissions in 2010 and 2015*: International Council on Clean Transportation. Available at: https://theicct.org/publication/a-global-snapshot-of-the-air-pollution-related-health-impacts-of-transportation-sector-emissions-in-2010-and-2015/ (Accessed: June 24, 2024).

Szidarovszky, F., Baafi, E. Y. and Kim, Y. C. (1987) 'Kriging without negative weights', *Mathematical Geology*, 19(6), pp. 549-559. DOI: 10.1007/BF00896920.

Terzi, Y. and Cengiz, M. A. (2009) 'Using of generalized additive model for model selection in multiple poisson regression for air pollution data', *Scientific Research and Essays*, 4(9), pp. 867-871. Available at: <Go to ISI>://WOS:000270306000005.

Tom Edwards (2019) 'ULEZ: The politics of London's air pollution', *BBC News*. Available at: https://www.bbc.co.uk/news/uk-england-london-47814416 (Accessed: 30 July 2024).

Transport for London (2019) *Ultra Low Emission Zone*. Available at: https://tfl.gov.uk/modes/driving/ultra-low-emission-zone (Accessed: July 30, 2024).

Transport for London (2023) *ULEZ Expansion 2023*. Available at: https://tfl.gov.uk/modes/driving/ultra-low-emission-zone/ulez-expansion-2023 (Accessed: June 28, 2024).

Transport for London, M. o. L. (2017) 'Crowding data of the public Transport For London data'. Available at: http://crowding.data.tfl.gov.uk/ (Accessed: July 15, 2024).

Urban Access Regulations in Europe (2012) *Impact of Low Emission Zones*. Available at: https://urbanaccessregulations.eu/low-emission-zones-main/impact-of-low-emission-zones (Accessed: June 28).

Van Zoest, V., Osei, F. B., Hoek, G. and Stein, A. (2020) 'Spatio-temporal regression kriging for modelling urban NO2 concentrations', *International journal of geographical information science*, 34(5), pp. 851-865.

Victoria Jones (2019) 'Ultra-low emission zone comes into force in central London', *ITVX*. Available at: https://www.itv.com/news/2019-04-08/ultra-low-emission-zone-comes-into-force-in-central-london (Accessed: 30 July 2024).

Wen, Q. S., Zhang, Z., Li, Y., Sun, L. and Assoc Comp, M. (2020) 'Fast RobustSTL: Efficient and Robust Seasonal-Trend Decomposition for Time Series with Complex Patterns'. *26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, Electr Network, Aug 23-27, 2203-2213.

Wong, D. W., Yuan, L. and Perlin, S. A. (2004) 'Comparison of spatial interpolation methods for the estimation of air quality data', *Journal of Exposure Science & Environmental Epidemiology*, 14(5), pp. 404-415. DOI: 10.1038/sj.jea.7500338.

World Health Organization (2021a) *WHO global air quality guidelines: particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide*. Available at: https://www.who.int/publications/i/item/9789240034228 (Accessed: August 23, 2024).

World Health Organization (2021b) *WHO global air quality guidelines: particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide* (978-92-4-003422-8. Available at: https://www.who.int/publications/i/item/9789240034228 (Accessed: June 22, 2024).

Zhang, K. and Batterman, S. (2013) 'Air pollution and health risks due to vehicle traffic', *Sci Total Environ*, 450-451, pp. 307-16. DOI: 10.1016/j.scitotenv.2013.01.074.

Appendix

Appendix A: Tables

Table A.1 Underground stations in study areas(*Return to the text*.)

Study Area	Underground Station
City of London	Aldgate, Bank, Barbican, Blackfriars, Cannon Street, Liverpool Street, Mansion House, Monument, Moorgate, St Paul's, Chancery Lane
Westminster	Baker Street, Bayswater, Bond Street, Charing Cross, Covent Garden, Edgware Road, Embankment, Great Portland Street, Green Park, Hyde Park Corner, Lancaster Gate, Leicester Square, Maida Vale, Marble Arch, Marylebone, Oxford Circus, Paddington, Piccadilly Circus, Pimlico, Queensway, Regent's Park, Royal Oak, St James's Park, St John's Wood, Temple, Victoria, Warwick Avenue, Westbourne Park, Westminster
Camden	Belsize Park, Camden Town, Chalk Farm, Euston, Euston Square, Finchley Road, Goodge Street, Hampstead, Holborn, Kentish Town, King's Cross St Pancras, Mornington Crescent, Russell Square, Swiss Cottage, Tottenham Court Road, Warren Street, West Hampstead
Islington	Angel, Archway, Arsenal, Caledonian Road, Farringdon, Finsbury Park, Highbury & Islington, Holloway Road, Tufnell Park, Old Street
Kensington and Chelsea	Earl's Court, Gloucester Road, High Street Kensington, Holland Park, Kensington (Olympia), Knightsbridge, Ladbroke Grove, Latimer Road, Notting Hill Gate, Sloane Square, South Kensington, West Brompton
Lambeth	Brixton, Clapham Common, Clapham North, Lambeth North, Oval, Stockwell, Vauxhall, Waterloo
Southwark	Bermondsey, Borough, Canada Water, Elephant & Castle, Kennington, London Bridge, Southwark

Table A.2 Data limitations and potential solutions(*Return to the text*.)

Dataset Source		Explanation of Limitations	Potential Solutions and Critical Thinking
Air Quality Data	Air Quality Data London Datastore London Air Quality Network Statistics Maps & Data Search Engine Breathe London London Air Quality	**1. Data Gaps** For example, the sensor at Westminster - Elizabeth Bridge lacks PM2.5 data before 2020/04/29. Similarly, the sensor at City of London - Senator House was decommissioned in 2013, resulting in no recent data. **2. Inconsistent Time Scale** Unlike other dataset, the "London Air Quality Network" only provides data in an hourly format. 3. **Fragmentary Air Quality Species** Not all sensors record the full spectrum of air quality species. Some sensors record species NO_2 and $PM_{2.5}$ but lack of other air species' records.	**1. Sensors supplement** Continue to be utilized while supplementing the dataset with information from other sensors within the study area. Cross-referencing with additional databases covering the same region could fill these gaps **2. Unified time scale** Preprocessing the hourly data to derive daily statistics such as maximum, mean, and minimum values for each day is necessary **3. Main air species selection** The focus will be on the most commonly recorded species: PM2.5, PM10, NO, and NO2. **4. Manual record**

	Network, Defra's Air Quality Monitoring	**4. Inaccessible Sensor Coordinates** The precise geographic coordinates of sensors via the API have been problematic due to maintenance issues or lack of updates. For example, the "Breathe London" database does not provide latitude and longitude information for its sensors.	For the "London Air Quality Network" and "Defra's Air Quality Monitoring" coordinates can be manually extracted from each sensor's introduction page. For "Breathe London" sensors, coordinates will be manually obtained from a map by identifying the closest landmark or building within a 10-meter radius of the sensor.
Public Transport Data	The public Transport for London (TfL) data	——	——
	Road traffic statistics - London region (dft.gov.uk)	**1. Non-continuous Traffic Flow Data:** All of traffic flow data 'dft_rawcount_local_authority.csv' exhibits significant gaps, such as the period between 2019/5/10 and 2019/7/16. **2. Missing Data for ULEZ Implementation Periods:** There is no data for the City of London from 2019/1/1 to 2019/4/8, making it challenging to analyze the immediate effects of the ULEZ Central Zone implementation on April 8, 2019.	**1. Data Interpolation Methods:** employs spatial interpolation methods as spline interpolation. **2. Selection of Data Period:** - expand count date from 2018/1/1 to represent the pre-ULEZ period. - for the pre-2019/04/08 ULEZ Central Zone Implementation, we will Focus on the period from 2018/01/01 to 2019/4/7 - for the post-2021/10/25 ULEZ Inner Expansion period, we will use 2022 data.
Meteorological Data	Weather History & Climate | The Weather's Record Keeper | Meteostat	The Meteostat database categorizes climate data for the London boroughs into seven major areas. This broad classification can result in less specific climate data for smaller, localized studies, particularly in urban areas where microclimates can vary significantly. The data from the London Weather Centre, chosen to represent central London, may not fully capture the climatic nuances of all specific areas within central London. This can introduce inaccuracies in studies that require precise, location-specific climate data.	**Urban Climatic Homogeneity:** By selecting the London Weather Centre as the primary source for central London, we aim to use the most geographically relevant climate data available. This approach ensures that the data is as representative as possible within the given constraints.
Geographic file	London Boroughs and Wards	——	——

Table A.3 Vehicle Category in Traffic Dataset(*Return to the text*.)

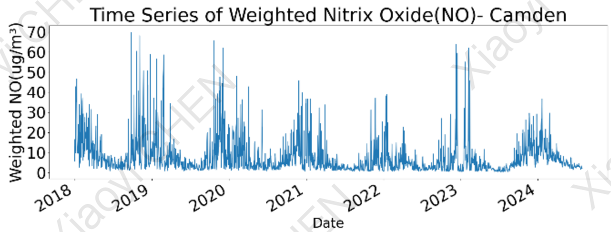
Category	Description
Pedal Cycles	Ordinary bicycles
Two-Wheeled Motor Vehicles	Motorcycles and electric bicycles
Cars and Taxis	Private cars and taxis

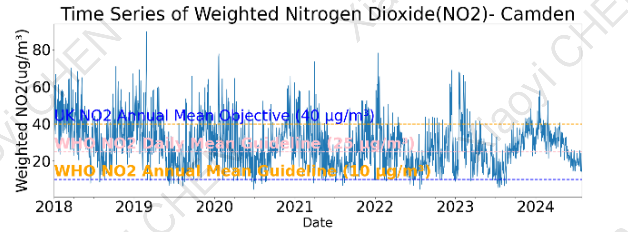
Buses and Coaches	Public buses and long-distance coaches
LGVs	Light Goods Vehicles, typically with a maximum gross weight not exceeding 3.5 tonnes
HGVs	Heavy Goods Vehicles, categorized as:
	- HGVs 2 Rigid Axle: Heavy goods vehicles with two rigid axles
	- HGVs 3 Rigid Axle: Heavy goods vehicles with three rigid axles
	- HGVs 4 or More Rigid Axle: Heavy goods vehicles with four or more rigid axles
	- HGVs 3 or 4 Articulated Axle: Heavy goods vehicles with three or four articulated axles
	- HGVs 5 Articulated Axle: Heavy goods vehicles with five articulated axles
	- HGVs 6 Articulated Axle: Heavy goods vehicles with six articulated axles
All HGVs	Includes all types of heavy goods vehicles, both rigid and articulated axles

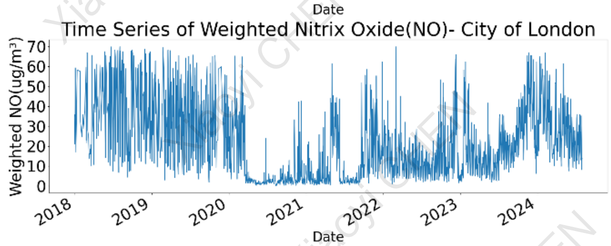
Table A.4 National Air Quality Objectives (*Return to the text.*)

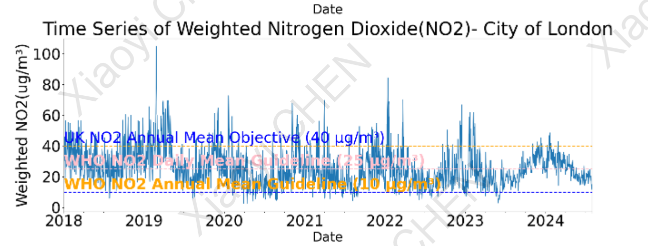
Pollutant	**Standard**	**Value ($\mu g/m^3$)**
NO2	UK Annual Mean Objective	40
	WHO Daily Mean Objective	25
	WHO Annual Mean Objective	10
PM2.5	UK Annual Mean Objective	20
	UK Daily Mean Objective (not to be exceeded more than 35 times a year)	15
	WHO Annual Mean Objective	5
PM10	UK Annual Mean Objective	40
	UK Daily Mean Objective	45
	WHO Annual Mean Objective	15

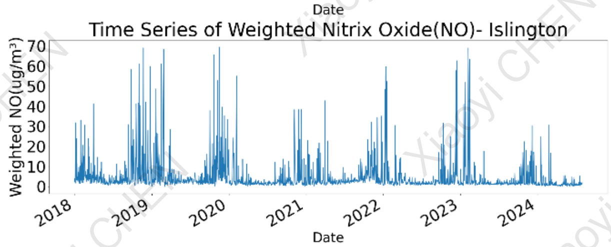
Appendix B: Graghs

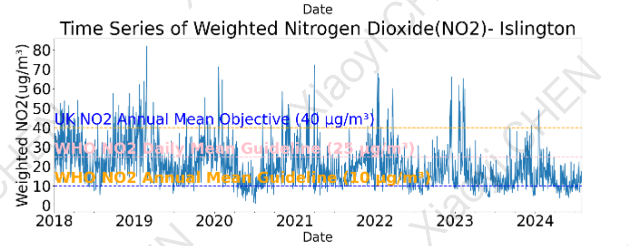


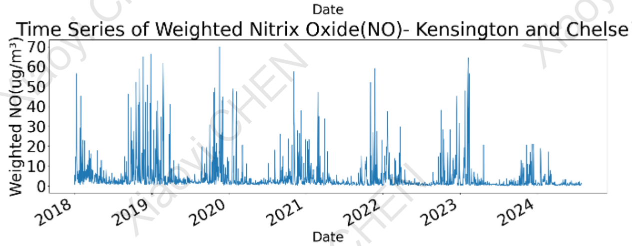


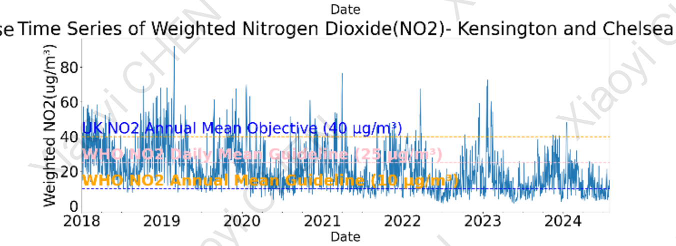


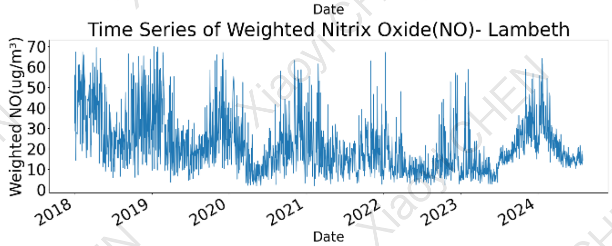


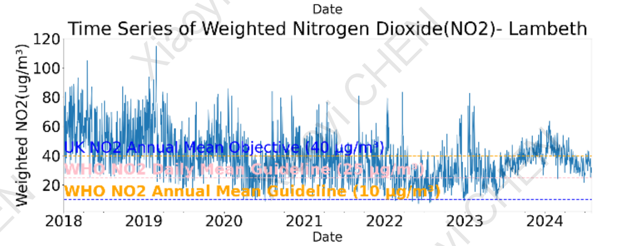


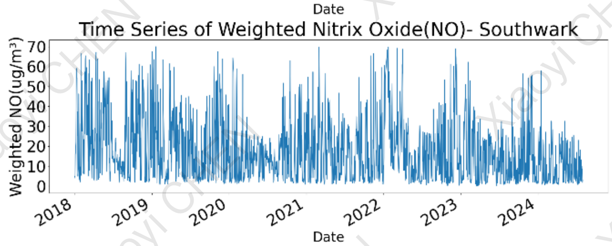


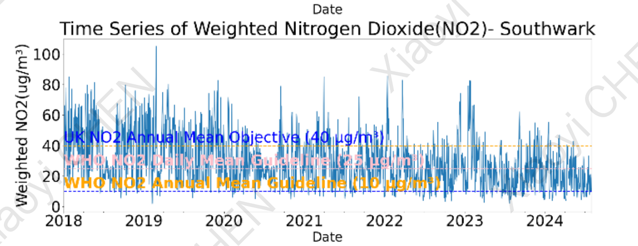


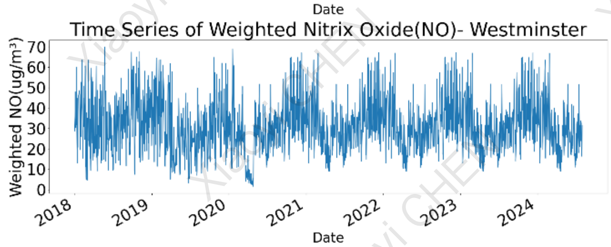


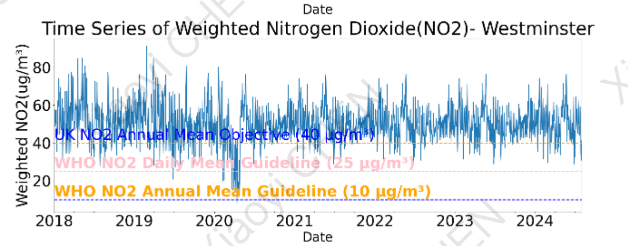












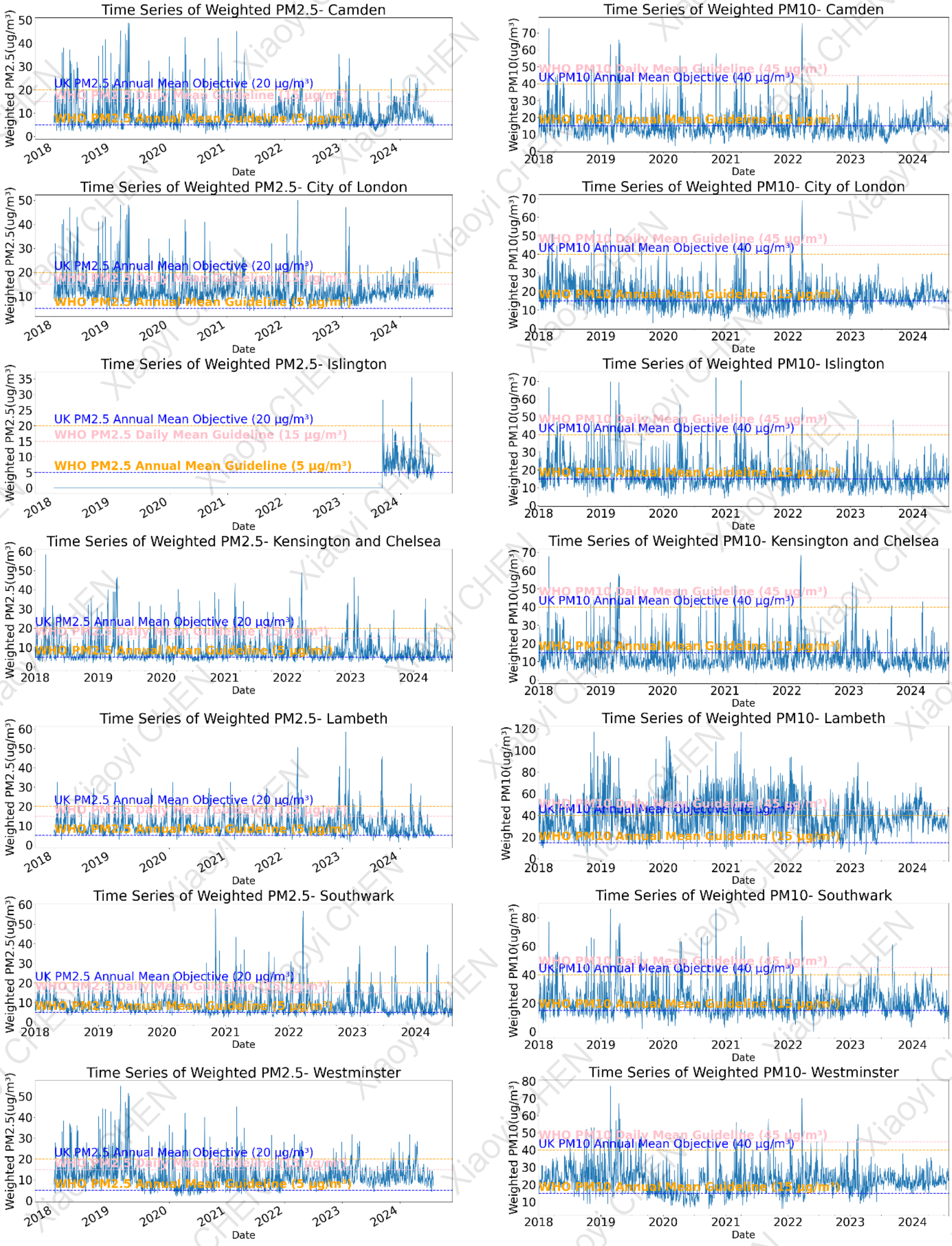


Figure B.1 Time Series on central London, including (a)NO, (b)NO2, (c)PM2.5 and (d)PM10. In each time series figure, there are guidelines showing National Air Quality Objectives (**Table A.4**), which orange dashed line, pink dashed line and deep blue dashed line respectively representing UK Annual Mean Objective, WHO Daily Mean Objective, and WHO Annual Mean Objective. (*Return to the text*.)

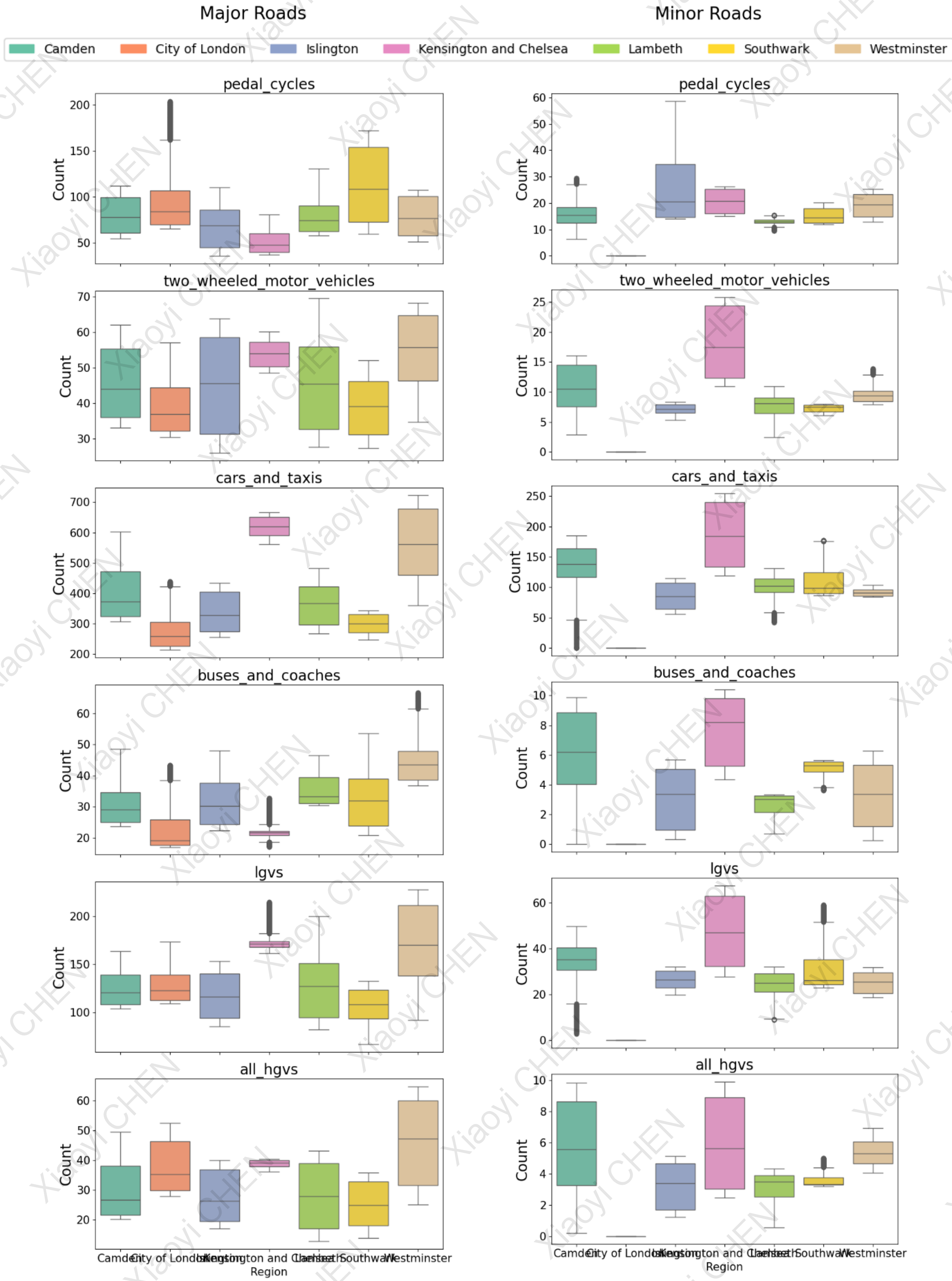


Figure B.2 Boxplots for Vehicle Types on Central London. Vehicle types explanation is in **Table A.3**. Minor roads' vehicle records are insufficient in City of London. (*Return to the text*.)

Appendix C: Important Code

C.1 Interpolation Method

C.1.1 Traffic data spline interpolation

(*Return to the text*.)

```
# Function to perform linear interpolation
def linear_interpolation(df, date_col, value_cols):
  df[date_col] = pd.to_datetime(df[date_col])
  min_date = df[date_col].min()
  max_date = df[date_col].max()

  # Create a complete date range
  date_range = pd.date_range(start=min_date, end=max_date, freq='D')

  # Create a new DataFrame with the complete date range
  full_df = pd.DataFrame(date_range, columns=[date_col])

  for col in value_cols:
    # Perform linear interpolation
    df.set_index(date_col, inplace=True)
    df = df.resample('D').mean()  # Resample to daily frequency and calculate mean for existing data
    df[col] = df[col].interpolate(method='linear')

    # Ensure interpolated values do not fall below the minimum original values
    min_value = df[col].min()
    df[col] = np.where(df[col] < min_value, min_value, df[col])
    full_df = full_df.merge(df[col], left_on=date_col, right_on=date_col, how='left')

  return full_df

# Specify the columns to interpolate
value_columns = ['pedal_cycles', 'two_wheeled_motor_vehicles', 'cars_and_taxis', 'buses_and_coaches',
        'lgvs', 'hgvs_2_rigid_axle', 'hgvs_3_rigid_axle', 'hgvs_4_or_more_rigid_axle',
        'hgvs_3_or_4_articulated_axle', 'hgvs_5_articulated_axle', 'hgvs_6_articulated_axle',
        'all_hgvs']

for file_path in file_paths_filtered:
  # Read the CSV file
  df = pd.read_csv(file_path)

  # Interpolate separately for each road type
  for road_type in df['road_type'].unique():
    road_df = df[df['road_type'] == road_type]
    interpolated_df = linear_interpolation(road_df, 'count_date', value_columns)
```

```
    # Calculate 'all_motor_vehicles' as the sum of specific columns
    interpolated_df['all_motor_vehicles'] = interpolated_df[value_columns].sum(axis=1)

    # Merge interpolated data back into the main DataFrame
    interpolated_df['road_type'] = road_type
    if road_type == df['road_type'].unique()[0]:
      full_interpolated_df = interpolated_df
    else:
      full_interpolated_df = pd.concat([full_interpolated_df, interpolated_df])

  # Construct new file path
  processed_file_path = file_path.replace('dft_rawcount_local_authority_id_', '')
  processed_file_path = processed_file_path.replace('145', 'Camden').replace('174', 'City of London').replace('96', 'Islington').replace('110', 'Kensington and Chelsea').replace('107', 'Lambeth').replace('103', 'Southwark').replace('109', 'Westminster')
```

C.1.2 Meteostat data polynomial interpolation

(*Return to the text*.)

```
# Function to perform polynomial interpolation
def polynomial_interpolation(df, date_col, value_cols, max_degree=3):
  # Create a complete date range
  date_range = pd.date_range(start=min_date, end=max_date, freq='D')

  # Create a new DataFrame with the complete date range
  full_df = pd.DataFrame(date_range, columns=[date_col])

  for col in value_cols:
    # Adjust polynomial degree based on the number of data points
    degree = min(max_degree, len(x) - 1)
    if degree < 1:
      full_df[col] = np.nan
      continue
    try:
      polynomial = np.poly1d(np.polyfit(x, y, degree))
    except np.linalg.LinAlgError:
      print(f"Failed to fit polynomial for (Chilès and Desassis) due to insufficient data.")
      full_df[col] = np.nan
      continue

    # Apply the polynomial to the full date range
    full_df[col] = polynomial((full_df[date_col] - min_date).dt.days.values)

    # Ensure interpolated values do not fall below the minimum original values
    min_value = non_null_data[col].min()
```

```
        full_df[col] = np.where(full_df[col] < min_value, min_value, full_df[col])

    # Merge the interpolated values back into the original dataframe
    df = pd.merge(df, full_df, on=date_col, how='outer', suffixes=('', '_interpolated'))
    for col in value_cols:
        df[col] = df[col].combine_first(df[f'(Chilès and Desassis)_interpolated'])
        df.drop(columns=[f'(Chilès and Desassis)_interpolated'], inplace=True)

    return df

# Specify the columns to interpolate
value_columns = ['tavg', 'tmin', 'tmax', 'wdir', 'wspd', 'wpgt', 'pres']

# Perform polynomial interpolation on the weather data
interpolated_weather_df = polynomial_interpolation(weather_df, 'date', value_columns, max_degree=3)
```

C.2 Kriging weighed air quality calculation

This section contains the Python code used for performing Kriging interpolation on air quality data. (*Return to the text*.)

C.2.1 Definition

```
# calculate_semivariogram
def calculate_semivariogram(data):
    num_points = len(data)
    semivariances = []

    for i in range(num_points):
        for j in range(i + 1, num_points):
            dist = np.linalg.norm([data['Latitude'].iloc[i] - data['Latitude'].iloc[j],
                                   data['Longitude'].iloc[i] - data['Longitude'].iloc[j]])
            squared_diff = (data['Value'].iloc[i] - data['Value'].iloc[j]) ** 2
            semivariances.append((dist, squared_diff))

    unique_distances = sorted(set([item[0] for item in semivariances]))
    avg_semivariances = []
    for dist in unique_distances:
        squared_diffs = [item[1] for item in semivariances if item[0] == dist]
        avg_semivariances.append((dist, np.mean(squared_diffs) / 2.0))

    return np.array(avg_semivariances)

# calculate_kriging_weights
def calculate_kriging_weights(semivariogram, distances, n, nugget=1e-10):
    A = np.zeros((n + 1, n + 1))
```

```
  for i in range(n):
    for j in range(n):
      if i == j:
        A[i, j] = semivariogram[0][1] + nugget
      else:
        dist = int(distances[0, j])
        A[i, j] = semivariogram[dist][1] if dist < len(semivariogram) else semivariogram[-1][1]

  A[-1, :-1] = 1
  A[:-1, -1] = 1

  b = np.zeros(n + 1)
  for i in range(n):
    dist = int(distances[0, i])
    b[i] = semivariogram[dist][1] if dist < len(semivariogram) else semivariogram[-1][1]

  weights = solve(A, b)
  return weights[:-1]

def adjust_weights(weights, wind_speed, wind_dir, sensor_directions, max_wind_speed):
  adjustments = 1 + (wind_speed * np.cos(np.radians(wind_dir - sensor_directions))) / max_wind_speed
  adjusted_weights = np.clip(weights * adjustments, 0, None)  # Ensure weights are non-negative
  return adjusted_weights

def normalize_weights(weights):
  total_weight = np.sum(weights)
  if total_weight == 0:
    return np.zeros_like(weights)
  return weights / total_weigh
```

C.2.2 Kriging application

```
# Go through each area
for region in regions:
  print(f"Processing region: {region}")
  no_df = preprocess_data(base_folder, region)

  # Kriging calculation
  weighted_values = []
  dates = no_df['Date'].unique()

  for date in dates:
    daily_data = no_df[no_df['Date'] == date]
    if len(daily_data) > 1:
      semivariogram = calculate_semivariogram(daily_data)
      distances = cdist(daily_data[['Latitude', 'Longitude']], daily_data[['Latitude', 'Longitude']], metric='euclidean')
```

```
            # Examine and process NaN and inf values
            if np.isnan(distances).any() or np.isinf(distances).any():
                print(f"NaN or inf values found in distances for {region} on {date}")
                distances = np.nan_to_num(distances, nan=0.0, posinf=0.0, neginf=0.0)

            kriging_weights = calculate_kriging_weights(semivariogram, distances, len(daily_data))


            wind_speed, wind_dir = get_wind_data(date, date.year)
            sensor_directions = np.arctan2(daily_data['Longitude'] - daily_data['Longitude'].mean(), daily_data['Latitude'] - daily_data['Latitude'].mean()) * 180 / np.pi
            meteostat_path = f'{meteostat_folder}\\meteostat{date.year}.csv'
            meteostat_df = pd.read_csv(meteostat_path)
            max_wind_speed = meteostat_df['wspd'].max()
            adjusted_weights = adjust_weights(kriging_weights, wind_speed, wind_dir, sensor_directions, max_wind_speed)
            normalized_weights = normalize_weights(adjusted_weights)
            weighted_value = interpolate(daily_data, normalized_weights)
            weighted_value = max(0, weighted_value)  # Make sure there are no negative values
            weighted_values.append({'Date': date, 'weighted_value(ug m-3)': weighted_value})
        else:
            weighted_values.append({'Date': date, 'weighted_value(ug m-3)': daily_data['Value'].values[0]})

    # Save the weighted values
    weighted_df = pd.DataFrame(weighted_values)
    output_path = os.path.join(weighted_output_folder, f'{region}- weighted.csv')
    weighted_df.to_csv(output_path, index=False)
    print(f"Weighted data saved to {output_path}")
```

Appendix D: Research Log

Date	Task/Meeting	Summary of Discussion/Notes
05/13/2024	Review of air quality data	Discussed issues with air quality monitoring points and their correlation with subway station locations. Consideration of other factors and databases affecting air quality.
05/21/2024	Literature review and data collection	Continued reading papers and collecting relevant data for analysis.
06/10/2024	Experimental analysis and draft framework	Discussed air pollution dispersion models (IDW + Kriging) and their limitations.
06/11/2024	Feedback on dissertation outline	Received feedback on dissertation outline.
07/05/2024	Feedback on June draft	Addressed data gaps: suggested using alternative data for missing time periods, explaining impacts on accuracy, and exploring other interpolation methods.
07/22/2024	Meeting with Simon Doyle	Discussed improving citations and discussion, incorporating critical reflection, and enhancing data and analysis visualizations.
07/23/2024	Meeting with Claire Dooley	Discussed further refinements and critical thinking on data gaps.
07/26/2024	Report to Simon Doyle	Reported challenges and methods used. Simon agreed with the approach but emphasized the importance of critical discussion and updating the draft.
08/02/2024	Email update to Simon	Provided updates on citations, data acquisition, and technical challenges. Mentioned pending technical flowcharts and requested guidance on Extenuating Circumstances (EC).
08/12/2024	Follow-up email to Simon	Updated on progress with Kriging techniques and encountered technical difficulties. Requested advice on EC application and indicated concern about meeting the deadline.
08/16/2024	Email to Simon regarding feedback availability during leave	Expressed concerns about Simon's leave affecting feedback. Mentioned pressure from resit exams and progress on data analysis. Requested possible email communication during Simon's leave.
08/21/2024	Simon's feedback on draft	Simon reviewed the draft and provided comments. Emphasized completing all sections and focusing on meeting the submission deadline.