Imperial College London

School of Public Health

Assessing the ULEZ Expansion using Boosted Regression Trees, Difference-in-Differences Modelling and Change Point Analysis

Author: Megan Pete
CID: 02292827
Degree: MSc Health Data Analytics and Machine learning
Word count: 8765

Submitted in partial fulfillment of the requirements for the MSc degree in Health Data Analytics and Machine Learning of Imperial College London

29 August 2023

Abstract

Background: Air pollution is one of the biggest threats to public health, and has been linked to increased mortality by numerous studies. The UK has made significant progress in reducing air pollution in recent years, however, levels remain well above WHO recommendations. The Ultra-Low Emission Zone (ULEZ) was introduced in London in 2019 to encourage vehicle compliance with Euro emissions standards, in a bid to lower emissions, since vehicles are one of the leading causes of $PM_{2.5}$ and NO_2 pollution. It was then expanded in 2021.
Aims: We aimed to evaluate the success of the 2021 ULEZ expansion by firstly estimating the effect of ULEZ compliance on $PM_{2.5}$ and NO_2 levels, and then investigating whether the 2021 ULEZ expansion caused an increase in compliance.
Methods: To estimate the effect of ULEZ compliance on pollution levels, we used a combination of meteorological normalisation using boosted regression trees, and linear regression. To estimate the effect of the ULEZ expansion on compliance levels, we used difference-in-differences modelling and change point analysis.
Results: We found that each unit increase in ULEZ compliance was associated with a decrease of 0.072 $\mu g/m^3$ of $PM_{2.5}$ and 0.14 $\mu g/m^3$ of NO_2. Moreover, the October 2021 ULEZ expansion caused an increase in compliance of 9.30 (per 100 vehicles) in the expansion zone.
Conclusions: We concluded that the 2021 ULEZ expansion was successful, since it caused a surge in compliance, bringing the total compliance to over 92.79%.

Contents

1 Background and literature review

Air pollution is a pressing global issue, negatively impacting both the environment and public health. In order to combat this problem, many countries have implemented policies aiming to reduce pollution. London's Ultra Low Emission Zone (ULEZ) is considered to be one of the most stringent emissions zones in the world, and has been a topic of much discussion in recent years. In this paper, we assess the efficacy of the ULEZ on lowering emission levels using a multi-faceted approach of meteorological normalisation, difference-in-differences modelling and change point analysis, both on compliance data and on pollution data.

1.1 Air pollution

Air pollution, as defined by the World Health Organization (WHO), arises from contamination of the environment by any chemical, physical or biological agent that modifies natural characteristics of the atmosphere. Common sources of air pollution are crude household combustion devices, motor vehicles, industrial facilities and forest fires. Alarmingly, WHO data reveals that 99% of the global population breathes air that contains high levels of pollutants, with low-income and middle-income countries having the highest exposures (1). Consequently, there were 4.2 million premature deaths in 2016 alone resulting from air pollution (13). The main air pollutants, as categorised by the US Environmental Protection Agency, are Carbon Monoxide (CO), Lead (PB), Nitrogen Dioxide (NO_2), Ozone (O_3), Particulate Matter (PM) and Sulphur Dioxide (NO_2) (17).

1.2 Health effects

The detrimental health impacts of air pollution are well-documented. Many studies have shown that inhalation of air pollution increases mortality (5). In fact, the European Air Quality Report (2013) stated that PM, O_3, and reactive nitrogen substances (mainly NO_2) were the pollutants posing the greatest risk to health in Europe (11). Epidemiological studies have suggested that PM with aerodynamic diameters less than 2.5 μm ($PM_{2.5}$) has larger negative effects on cardio and respiratory disorders than larger particles (22). The largest health burdens arising from $PM_{2.5}$ and NO_2 are Chronic obstructive pulmonary disease (COPD) and Type 2 diabetes respectively (12), with current evidence showing a strong association between $PM_{2.5}$ and all-cause mortality (5). Other conditions arising from inhaling these pollutants include asthma, lung cancer and respiratory conditions (4).

1.3 UK policies

Over the last 50 years, the United Kingdom has made significant progress reducing air pollution risks through policy interventions and technological advancements. Between 1970 and 2010, the attributable mortality due to exposure to $PM_{2.5}$ and NO_2 declined by 56% and 44% respectively (6). However, it remains important to monitor and reduce the concentrations of these pollutants due to the regulations not being stringent enough, and the current regulations not being adhered to. The UK's current regulations are that the annual average concentration of $PM_{2.5}$ may not exceed 20 μg/m^3, with a goal to lower this to 10 μg/m^3 by 2040 (2). However, there is evidence to suggest that there is no lower threshold for the relationship between $PM_{2.5}$ and mortality, meaning that there is, in reality, no safe level of $PM_{2.5}$ (3). Therefore, a far greater reduction is needed. The current rules for NO_2 concentration in the UK state that the annual mean concentration of NO_2 must not exceed 40 $\mu g/m^3$ (18). However, the Annual Air Quality Assessment (2021) found that the UK was not compliant with this limit (19). Therefore, it is clear that NO_2 and $PM_{2.5}$ levels in the UK still pose a public health threat, and should both be reduced.

1.4 Ultra Low Emission Zone

Motor vehicle traffic has remained one of the leading causes of $PM_{2.5}$ and NO_2 pollution in London. As per the London Atmospheric Emissions Inventory report (2019), road vehicles were responsible for over 25% of NO_x (nitrogen oxides), and almost 40% of $PM_{2.5}$ emissions (20). To tackle this problem, London introduced the world's most stringent emissions zone called the 'Ultra-Low Emission Zone' (ULEZ) in 2019 (9). The scheme works by introducing a daily charge for owners who drive cars in the zone that do not meet specific emissions standards, derived from Euro vehicle standards. The ULEZ was subsequently expanded on October 25, 2021, to be 18 times larger than before, measuring 380km^2 (16), with further plans to expand on August 29, 2023. Our goal was to evaluate the efficacy of the October 2021 ULEZ expansion.

1.5 Models and current research

Causal inference methods commonly used to study the ULEZ and other emissions zones are a Regression Discontinuity Design (RDD), Difference in Difference (DID) framework and state-space modelling. Each of these methods have their strengths and limitations, which has led to varying results being published on the 2019 implementation of the ULEZ. The most optimistic estimate came from the Greater London Authority, who estimated that the ULEZ caused a 29% reduction in roadside NO_2 concentrations July to September 2019 and a 37% reduction from January to February 2020 (23). Hajmohammadi et al (2022) used state space modelling to show the ULEZ implementation caused a 20% reduction of NO_x in the ULEZ area (24). Zhai et al (2021) and Prieto-Rodriguez et al (2022) used DID frameworks and found that the ULEZ reduced PM_{10} by 5.5% (25) and NO_2 by 19% (26). However, Liang Ma et al. (2021) found, using an RDD model, that only a small portion of air quality improvements in London were attributable to the ULEZ, with an average reduction of less than 3% for NO_2 and no significant effects on $PM_{2.5}$ (9). Considering the huge variation in results, it is vital to carefuly consider the underlying assumptions of each method in order to choose one that is suitable to assess the 2021 ULEZ expansion.

A host of confounding factors make air quality data particularly difficult to analyse. These include changing traffic severity, types of vehicles on the road, smaller government policies, and changes in traffic flow near monitoring stations (10). Arguably, the most notable confounders are meteorological conditions and the COVID-19 pandemic, with research showing that both $PM_{2.5}$ and NO_2 levels in London decreased significantly during the COVID-19 lockdowns in 2020 (9). Meteorological normalisation can be used in order to mitigate the effects of weather patterns and seasonality. Methods developed in recent years use boosted regression trees and random forest models. Machine learning methods perform better for meteorological normalisation than other techniques as they can handle a mix of variables, missing data, non-linear effects and outliers (27).

1.6 Breathe London network

The Breathe London network is a collection of over 400 air quality monitoring sensors around London which measure hourly $PM_{2.5}$ and NO_2 concentrations. The hardware was produced by Clarity Movement Co, and they are managed, corrected and used for analysis by the Environmental Research Group at Imperial College London. They are different to traditional air quality sensors, such as those in the London Air Quality Network (LAQN), as they are significantly smaller and cheaper. This means they are able to be placed in various locations around the city, often close to traffic sources, which provides data that is highly specific to each neighbourhood.

1.7 Objective

Although the 2019 implementation of the ULEZ has been studied, there is little existing research on the October 2021 expansion, which affected significantly more vehicles than the implementation. We therefore aimed to provide novel research on this topic using data from the Breathe London network, in order to assess whether further expansions are necessary, especially in light of the August 2023 expansion. In our evaluation of the ULEZ expansion, we looked at two key aspects of the scheme: compliance and emissions. We defined the scheme to have been successful if it caused an increase in compliance ULEZ standards, and if this, in turn, caused a decrease in air pollution. In order to account for meteorological factors, and to remove noise from the data, we applied meteorological normalisation using boosted regression trees, using the `deweather` package in R (28). Then, to assess the efficacy of the ULEZ expansion, we used difference-in-differences (DID) modelling and change point analysis to answer the questions: (R1) Is an increase in ULEZ compliance correlated with a decrease in NO_2 or $PM_{2.5}$ concentrations? (R2) Did the 2021 ULEZ expansion cause an increase in ULEZ vehicle compliance?

2 Methods

2.1 Data collection and pre-processing

Data was collected from Breathe London, the Mayor of London, and London City airport for our analyses.

ULEZ compliance data

ULEZ compliance data was collected from the Expanded Ultra Low Emission Zone Six Month Report (37). ULEZ emissions standards are defined in terms of Euro emissions standards, depending on the vehicle and emission type, as shown in table 1. Compliance data was used from the ULEZ Expansion 6 Month Report captured through Automated Number Plate Recognition (ANPR) data derived from anonymised daily camera detections, and then cross-referenced with DVLA records (39). This means that one vehicle is able to contribute to both the expansion and outer zones' compliance rates if it drives in both zones. The ULEZ zones are shown in figure 1, where the yellow area represents the 2021 expansion zone and the red area represents the existing 2019 ULEZ zone.

Pollution data

Pollution data was collected from the Breathe London network, which contains sensors at various locations around London. The sensors measured hourly concentrations of NO_2 and $PM_{2.5}$, and the sensor locations were recorded. Data was used from between 1 March 2021 and 31 May 2022. The sensors were regularly calibrated, and after concentrations were measured, the data was scaled using using a number of "colocated" nodes (i.e. nodes for which an LAQN reference node exists at the same location), to reduce uncertainty. Outlier concentrations were then removed, defined to be smaller than $Q_1 - 1.5 \times IQR$ and larger than $Q_3 + 1.5 \times IQR$. Finally, the hourly concentrations were aggregated by day to remove noise. The sensors in the Breathe London network are classed as either "roadside", "kerbside" or "background", depending on their distance from a main road. Our analyses were performed on roadside sensors due to the fact that these would be the most

Table 1: Euro vehicle emissions standards required to be followed within the ULEZ to avoid the penalty (38) (39).

Vehicle type	Euro standard	Emission limits
Motorcycles, mopeds, motorised tricycles and quadricycles (L category)	Euro 3	Nitrogen oxides for diesel (0.5 g/km) and petrol (0.15 mg/km) engines
Petrol cars, vans, minibuses and other specialist vehicles	Euro 4	Nitrogen oxides (0.08g/km)
Diesel cars, vans and minibuses and other specialist vehicles	Euro 6	Nitrogen oxides (0.08g/km), Particulate matter (0.005g/km)

sensitive to detecting changes. The sensors were split into two groups: "expansion" and "outer", depending on whether or not they were in the area affected by the expansion. Figure 2 shows the roadside sensors for both pollutants in relation to the different ULEZ zones. The inner red zone in this figure represents the original ULEZ zone, introduced in 2019, but it was excluded due to the fact that this zone was already required to follow emissions standards and so no major change would be expected.

Meteorological data

Meteorological data was collected from London City Airport using the `importNOAA()` function from the `worldmet` package in R. In particular, hourly air temperature, wind speed and wind direction were recorded. Furthermore, geographical information about London was collected from Google Maps using the `get_map`() function from the `ggmap` library.

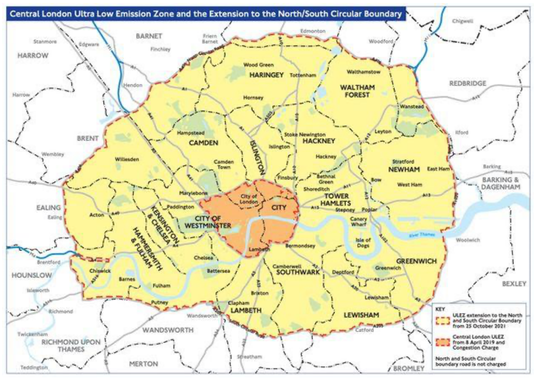


Figure 1: Ulez expansion map (35).

2.2 Meteorological Normalisation

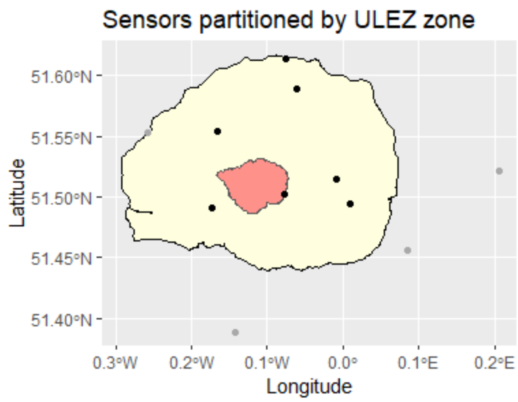
In order to answer our first research question (R1) "Is an increase in ULEZ compliance correlated with a decrease in NO_2 or $PM_{2.5}$ concentrations?", it was necessary to first perform meteorological normalisation on the $PM_{2.5}$ and NO_2 time series to remove the natural variation in pollution levels caused by meteorological factors, which would have confounded our analysis. The `deweather` package in R was used to do this, which was developed by Carslaw et al (28). The package works by fitting boosted regression trees to estimate the concentrations of $PM_{2.5}$ and NO_2 from meteorological predictor variables, and then performing a meteorological averaging procedure to account for variability in weather conditions.

Gradient boosted regression tree algorithm

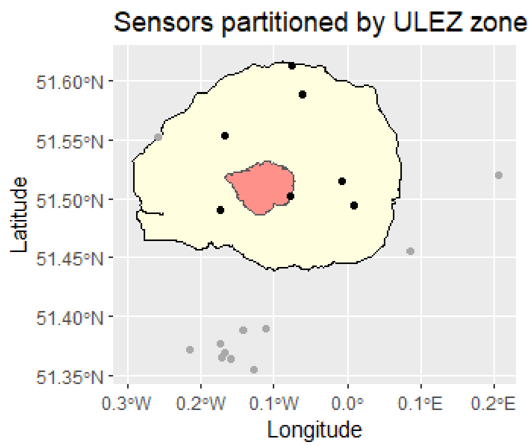
Gradient boosted regression trees (GBRT) is a prediction algorithm combining regression decision trees and gradient boosting.

Decision trees are hierarchical machine learning models that can be used for regression or classification. We used regression trees since our outcome variable, pollution concentration, was continuous. They facilitate the decision making process by recursively partitioning the data based on feature values. In our case, we partitioned our dataset consisting of pollution concentrations, by meteorological features such as the air temperature, wind speed and wind direction at the time the concentration was measured. The main idea for building a decision tree is to divide the predictor space of observations (X_1, X_2, ..., X_N) into K distinct regions, (R_1, R_2, ..., R_K). The prediction for each region R_j becomes the most prevalent outcome in that region for classification, or the mean of the observations in that region for regression (32), i.e.

$$\text{Prediction}(R_j) = \text{mean}(X_i), \quad \forall X_i \in R_j.$$



(a) $PM_{2.5}$



(b) NO_2

Figure 2: Breathe London sensors in the expansion and outer zones.

Ideally, the dataset should be split in such a way that minimises the loss function, Ψ, quantifying the prediction error overall. However, this would mean calculating and measuring the loss function of every possible splitting of the nodes, which would be too computationally expensive. Therefore, an impurity metric is used as a proxy to estimate the potential improvement of the loss function that would be made with a particular split. Recursive binary splitting is used whereby the data is split into two recursively on the feature that minimises the impurity at that given point (i.e. a greedy algorithm). Common impurity metrics for regression and classification tasks are given in the appendix in tables 9 and 10. In our case, the `gbm()` function used the Mean Squared Error (MSE). For regression trees, since there are no obvious ways to split each feature, various threshold values are tested for each feature split and the optimal one is chosen. The process continues until stopping criteria, such as maximum tree depth or minimum samples per leaf, are met. Figure 12 in the appendix illustrates an example of a decision tree. Decision trees are flexible and interpretable, however they are prone to overfitting, meaning their prediction may be poor on test data. Therefore, strategies such as pruning, random forests and gradient boosting can be used in order to improve their generalisation.

Gradient boosting is an iterative algorithm which combines multiple decision trees, introduced by Friedman in 2001 (29). Each decision tree is called a "weak learner" which has low accuracy due to the fact that it is prone to overfitting. However, gradient boosting is a method to build and combine multiple decision trees, and usually outperforms larger single decision trees (34). Gradient boosting regression can be modelled as follows, as explained by Li and Bai (33). Assume we have a regression problem with a training set $D = (\mathbf{x}, \mathbf{y})$ for $\mathbf{x} \in R^{n \times d}$, $\mathbf{y} \in R^n$, where y_i is the observation corresponding to the vector of features $x_i \in R^d$.

Regression trees aim to find the function f^* mapping $\mathbf{x}$ to $\mathbf{y}$ which minimises the expected value of the loss function Ψ over the joint distribution of $\mathbf{x}$ and $\mathbf{y}$,

$$\mathrm{f}^* = \arg\min_{\mathrm{f}} \{E_{\mathbf{y},\mathbf{x}} \Psi(\mathbf{y}, \mathrm{f}(\mathbf{x}))\}. \tag{1}$$

Gradient boosting is a method to approximate f^* using a linear combination of regression trees $h(\mathbf{x})$,

$$\mathrm{f}(\mathbf{x}) = \sum_{m=0}^{M} \beta_m h(\mathbf{x}), \tag{2}$$

where β_m are the expansion coefficients.

The approximation $\mathrm{f}(\mathbf{x})$ is defined recursively, starting with a simple constant function. A good initial guess is choosing a constant that is as "close" as possible to the outcome vector i.e. a constant that minimises the sum of the losses between it and each outcome. More formally, this can be written as

$$\mathrm{f}_0(\mathbf{x}) = \arg\min_{\rho} \sum_{i=1}^{n} \Psi(y_i, \rho). \tag{3}$$

Subsequently, a new weak learner $h_m(\mathbf{x})$ is added. This new weak learner $h(\mathbf{x})$ is a decision tree which is built and trained to predict the residual errors of the previous iteration. Hence, each iteration improves the prediction by accounting for part of the error. In other words, the model learns how to correct itself. A learning rate parameter is also added to each weak learner to control how quickly the function $f_m(x)$ converges, and it is a trade-off between convergence speed and stability of the model.

In order to train the new weak learner $h(\mathbf{x})$ to predict the residual errors of the previous iteration (i.e. "pseudo-residuals" of the previous iteration), a loss function such as the mean squared error (MSE) is used to quantify the accuracy of the previous prediction. The weak learner is then trained to predict the function h and learning rate ρ which minimises the loss between the prediction of the previous iteration and the pseudo-residuals of the previous iteration. In mathematical terms we write this as

$$h_m = \arg_{h,\rho} \min \sum_{i=1}^{n} [\tilde{y}_{im} - \rho h(\mathbf{x}_i)]^2, \tag{4}$$

where for the mth iteration,

- h_m is the best weak learner,
- $h(\mathbf{x}_i)$ is the prediction of the weak learner for the ith sample,
- ρ is the learning rate associated with the weak learner $h(\mathbf{x}_i)$,
- $\tilde{y}_{im}$ is the pseudo-residual for the ith sample, defined as the negative gradient of the loss function Ψ with respect to the current prediction $f(\mathbf{x}_i)$

$$\tilde{y}_{im} = -[\frac{\partial \Psi(y_i, \mathrm{f}(\mathbf{x}_i))}{\partial \mathrm{f}(\mathbf{x}_i)}]_{\mathrm{f}(\mathbf{x})=\mathrm{f}_{m-1}(\mathbf{x})} \tag{5}$$

The approximation f is then updated after each iteration as follows

$$\mathrm{f}_m = \mathrm{f}_{m-1} + \beta_m h_m(\mathbf{x}), \quad m = 1, 2, \ldots M, \tag{6}$$

where the optimal weight β_m for the weak learner $f(\mathbf{x}_i)$ is defined to be the constant that minimizes the loss function Ψ when we combine the current and the previous model's predictions,

$$\text{i.e. } \beta_m = \arg_\beta \min \Psi(y_i, \mathrm{f}_{m-1}(\mathbf{x}_i) + \beta h(\mathbf{x}_i)). \tag{7}$$

Weak learners are added until a pre-defined stopping criterion is reached, such as a maximum number of iterations or a minimum level of accuracy.

Model parameters

We used the following parameters in our GBRT model:

- To estimate the optimal value for the maximum number of trees, 5-fold cross-validation was used. This number provides a balance between model accuracy and computational complexity.
- The learning rate was set at 0.1 to control the step size at which the boosting algorithm adapted to errors.
- The maximum depth of individual trees in the ensemble was set to 5 to prevent overfitting.
- The `bag.fraction` parameter was set to 0.5. This parameter controls the fraction of the training data that is randomly selected (with replacement) to train each individual tree during the boosting process, in order to introduce additional randomness and diversity into the ensemble of trees.
- The minimum number of observations required to create a terminal node was set to 10.

Meteorological normalisation pipeline

Using the `deweather` package in R, we performed meteorological normalisation as follows.

1. We employed the `testMod()` function from the `deweather` package to find the optimal number of trees, and evaluate the performance of the GBRT model in predicting pollutant concentrations while considering relevant meteorological variables – in our case week, month, wind speed, wind direction and air temperature. By testing the model against withheld data and comparing its predictions with actual measurements, we determined the suitability of the GBM model for meteorological normalisation.
2. We then used the `buildMod()` function to fit the optimal GBRT model, found in the previous step, to our data.
3. After training a GBRT to predict the outcome concentration from the meteorological features, the `metSim()` function was used to apply a meteorological averaging procedure to both the expansion and the outer data separately. It aims to disentangle the effects of meteorological conditions from the variable of interest, allowing for a more accurate assessment of underlying trends and changes. The function works as follows
    - Initially the GBRT model built in the previous step is trained using a dataset that includes both the variable of interest and relevant meteorological variables.
    - The function then simulates scenarios where meteorological conditions are randomly varied while keeping other non-meteorological variables constant. This generates a range of hypothetical instances of the variable of interest, representing how it might have behaved under different meteorological circumstances.
    - For each simulation, the GBM model predicts the variable's values based on the altered meteorological conditions. The difference between the predicted values and the actual values represents the impact of meteorological conditions, or the "residuals."

- The residuals from multiple simulations are aggregated by calculating their mean or other statistical summaries. This average impact of meteorological factors is considered as the correction to be applied to the actual variable values.
- The calculated correction (average residual) is subtracted from the actual variable values. This results in normalised data, where the meteorological influence has been effectively removed. The normalised data is now clearer of meteorological noise, providing a more accurate representation of the variable's behavior.

2.3 Effect of compliance on air pollution (R1)

Once meteorological normalisation was performed, we proceeded to estimate the correlation between ULEZ compliance rates and air pollution levels. Separate linear regression models were built for $PM_{2.5}$ and NO_2, which estimated concentration level as a function of compliance. Since meteorological normalisation was already performed, there was no need to include these covariates in the models, as they had already been accounted for. Therefore, simple linear regressions were used, given by

$$\mathrm{Conc}_t = \alpha + (\beta \times \mathrm{C}_t) + \epsilon_t$$

where:

- Conc_t represents the air pollution concentration (of $PM_{2.5}$ or NO_2) at time t;
- α is the intercept term;
- β is the change in concentration per unit change in compliance rate;
- ϵ_t denotes the error term.

2.4 Effect of ULEZ on compliance (R2)

Once we had quantified the correlation between ULEZ vehicle compliance and pollution levels, we proceeded to answer R2. Our goal was to understand whether the new legislation in October 2021, also known as the ULEZ expansion, increased ULEZ compliance levels. We employed both difference-in-differences (DID) modelling and change point analysis to quantify the manner in which the ULEZ expansion affected compliance levels.

Difference-in-differences model

We used a difference-in-differences (DID) model to assess the impact of the ULEZ expansion on the the ULEZ compliance rates, both in the expansion (intervention) and outer (control) zones. A DID model is a statistical method used for causal inference, which compares the outcome variable before and after the intervention, in both the intervention and the control groups. It does so under the assumption that these confounding variables affect both the control and intervention groups in a similar way. By comparing the two groups, it accounts for a host of confounders which are constant between the two groups, and aims to isolate the effect of the intervention. Using a DID model we were able to explain whether the ULEZ expansion affected the expansion (intervention) and outer (control) zones differently. The DID method hinges on the "parallel trends" assumption, which states that, in the absence of the intervention, both the ULEZ expansion (intervention) and outer (control) groups would have experienced similar trends in outcomes over time. This assumption isolates the causal effect of the intervention from other confounding factors.

Model specification

Baseline trends were established in both the expansion (intervention) and outer (control) groups using compliance rates before the ULEZ expansion. By comparing their outcomes in this pre-ULEZ period, we established a baseline trend for each group. The intervention period was defined to be from the introduction of the ULEZ period, i.e. October 25, 2021, until the end of the study period. A linear regression model was set up with dummy variables representing the post-ULEZ time period and the expansion zone. The model was specified as

$$\mathrm{C}_t = \alpha + (\beta \times \text{Expansion zone}) + (\gamma \times \text{Post}_t) + (\delta \times \text{Expansion zone} \times \text{Post}_t) + \epsilon_t$$

where:

- C_t represents the compliance rate at time t;
- α is the intercept term;
- β captures the average difference in outcomes between the expansion and outer zones before the ULEZ expansion;
- γ captures the average change in outcomes over time in the outer zone;
- δ is the DID estimate, which quantifies the average intervention effect by comparing the change in outcomes between the expansion and outer groups during the intervention period;
- ϵ_t denotes the error term.

Hypothesis tests were conducted to evaluate the statistical significance of the model coefficients. The coefficient of the interaction term, δ, is the DID estimate, and quantifies how differently the expansion (intervention) and outer (control) zones were affected by the ULEZ expansion. This can also be computed by calculating the increase in compliance in each zone from pre-ULEZ to post-ULEZ, and then taking the differences between them.

Change point analysis

DID modelling was used to model the mean effects before and after the ULEZ expansion. However, this does not shed light on whether any sharp changes took place around the time of the intervention, which why we used change point analysis. Change point analysis is a technique used to identify points in a dataset where the underlying statistical properties have changed. Therefore, this method should pick up any sudden, significant changes within the compliance time series, potentially caused by ULEZ. It was performed using the `changepoint` package in R.

Parameters

In order to perform change point analysis, a number of parameters needed to be specified.

Algorithm

The binary segmentation method was used to detect change points. This method involves iteratively partitioning the data into segments, and then testing for change points within each segment. The algorithm works as follows

- First, the entire dataset is treated as one segment.
- A statistical test (determined by the distribution of the data) is performed to assess whether or not a significant change point lies within the segment. A hypothesis test is performed using the test statistic under the null hypothesis, H_0, which is that there are no change points in the segment. If the test statistic exceeds the critical value chosen, it suggests that a change point is likely to exist within the segment. If the test concludes a change point is likely to exist, the segment is split into two subsegments at the location of the detected change point.
- This process continues iteratively until no more change points are detected in any of the segments.

Test-statistic

A test statistic needed to be chosen, and therefore, we needed to determine the distribution of the compliance data. Using density and Q-Q plots, we hypothesised that our data was normally distributed. To confirm this hypothesis, we used the Shapiro-Wilk normality test, which resulted in a test statistic of $W = 0.86$ (p-value $= 2.2 \times 10^{-16}$) for the expansion and $W = 0.87$ (p-value $= 2.2 \times 10^{-16}$) for the outer zone. The fact that these test statistics are close to 1 indicates that the data is likely normally distributed.

By assuming the data is normally distributed, we were then able to choose the likelihood-ratio test statistic for our change point analysis (40),

$$LR = -2 \cdot \ln \left(\frac{L(\boldsymbol{\theta}_{\text{with CP}})}{L(\boldsymbol{\theta}_{\text{without CP}})} \right),$$

Where

- $L(\boldsymbol{\theta}_{\text{with CP}})$ represents the likelihood of the data assuming the presence of a change point within the segment.
- $L(\boldsymbol{\theta}_{\text{without CP}})$ represents the likelihood of the data assuming no change point within the segment.

The LRT compares the likelihood of the data under the assumption that there is a change point within the segment with the likelihood under the assumption that there is no change point. This evaluates whether the inclusion of a change point significantly improves the fit of the model to the data.

Other parameters

The other parameters that we specified were

1. We focused on detecting changes in the mean concentration value.
2. We tried detecting only one change point to see whether the algorithm could pick up the ULEZ expansion date.
3. The MBIC penalty was applied to avoid picking up on very small fluctuations in the data.

3 Results

3.1 Exploratory analysis

In order to understand the underlying distributions of our data used in both of our research questions, we performed an exploratory analysis, for ULEZ compliance rates as well as air pollution concentrations.

Air pollution concentrations

We aimed to gather information about London's air pollution levels both in space and in time. To gain an overview of the spatial distribution of $PM_{2.5}$ and NO_2 concentrations, we created a map of London with scatter plot nodes colored by their concentration levels, in figures 3 and 4. These visualisations allow us to identify potential clusters and hot spots of higher concentrations. We see a relatively high concentration of $PM_{2.5}$ in the south-west of London, and a lower concentration in central London, which may allude to the ULEZ expansion zone being lower. However, the trend for NO_2 is more difficult to spot. Table 2 and 3 show summary statistics of the pollution concentrations within the ULEZ expansion zone and the outer (control) zone. Table 2 suggests that mean $PM_{2.5}$ concentrations in the expansion zone are lower than in the outer zone, but table

Table 2: Summary statistics of mean hourly $PM_{2.5}$ concentrations in the ULEZ expansion and outer (control) zones.

Measure for $PM_{2.5}$	Expansion (intervention)	Outer (control)
Mean	9.17	9.70
Median	8.43	9.10
Min	-2.77	-4.17
Max	23.54	23.52
SD	4.17	4.18

3 suggests that NO_2 concentrations in the outer zone are lower than the expansion zone, however not by much as the median values are extremely close together in both cases.

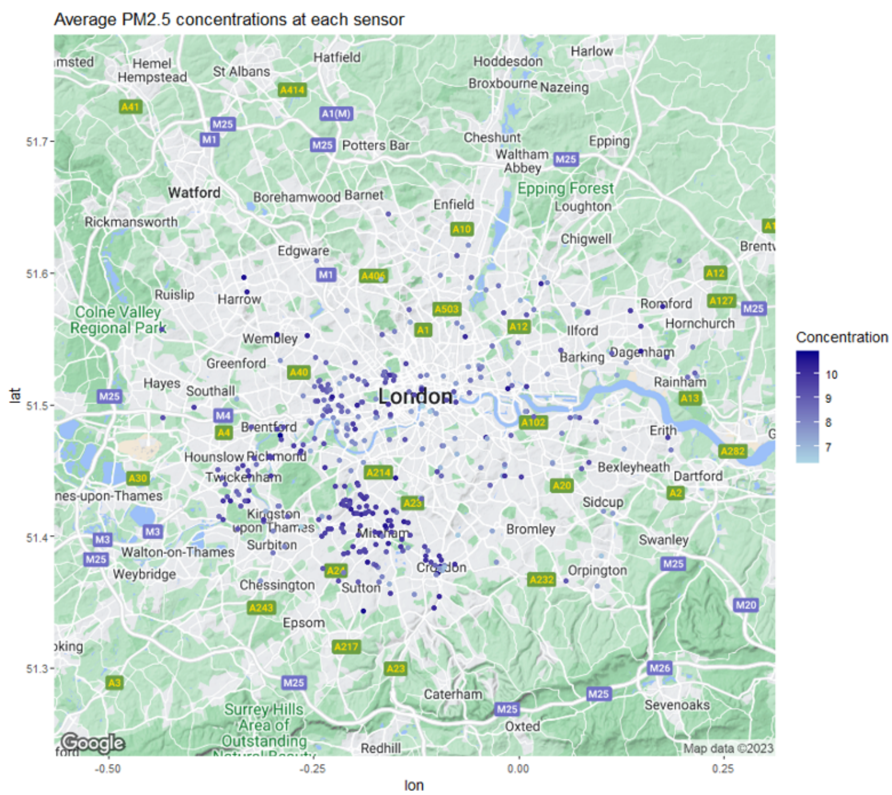


Figure 3: Average $PM_{2.5}$ concentration at each sensor.

To gain a more thorough understanding of annual trends, we examined the average concentration patterns across all data points. We tracked the pollution concentrations from the beginning of 2021 to the start of 2023, as illustrated in figure 5. Additionally, we integrated data on Covid-19 restrictions into the graph. Here, red segments denote periods of stricter restrictions, while blue segments represent times with no restrictions. The dotted line shows the ULEZ expansion. However, uncovering distinct temporal patterns within the data proved to be challenging due to its considerable noise. This noise is likely attributed to fluctuating weather conditions that introduce unpredictable variability. By computing the correlations, we found that both $PM_{2.5}$ and NO_2 concentrations are correlated with wind speed, wind direction and air temperature, with wind speed being the most correlated. Figure 13 in the appendix displays heat maps illustrating the correlation between meteorological variables and pollution concentrations. While we could potentially incorporate these meteorological factors into our models as covariates to account for

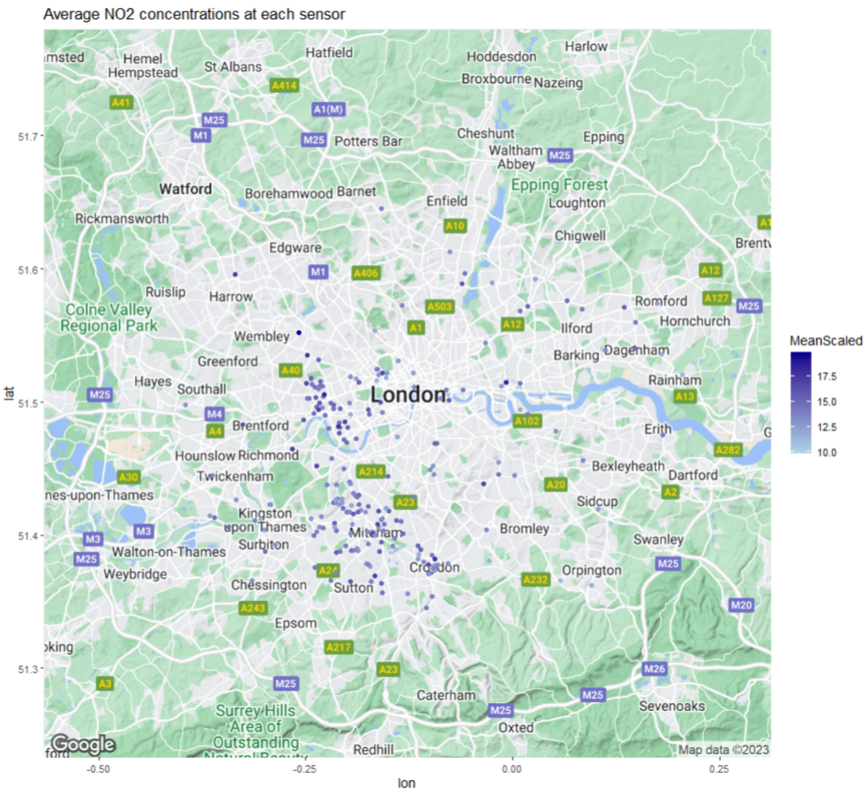


Figure 4: Average NO_2 concentration at each sensor.

their influence, the noise in the data makes meaningful interpretation and conclusive findings difficult. This highlights the importance of implementing meteorological normalisation.

ULEZ compliance

The graph of ULEZ compliance rates over time is shown in figure 6, split by zone. The ULEZ zone (orange) is the expansion zone, and the outer (green) is the control zone. The graph shows that all zones experienced an increase in compliance between June 2020 and May 2022. From the graph we can see that all zones had a compliance rate of over 70 per 100 vehicles in June of 2020, long before the expansion came into effect. In May 2022, the ULEZ (orange) and outer (green) zones had compliance of approximately 94 and 84 per 100 vehicles respectively. The two zones followed relatively similar trends up until October 2021, after which the ULEZ zone experienced a more drastic increase in compliance. This suggests that the ULEZ expansion encouraged a higher rate in compliance, which we rigorously investigated in section 3.4.

Table 3: Summary statistics of mean hourly NO_2 concentrations in the ULEZ expansion and outer (control) zones.

Measure for NO_2	Expansion (intervention)	Outer (control)
Mean	14.70	14.51
Median	13.81	13.66
Min	2.78	0.51
Max	35.52	35.49
SD	5.95	6.12

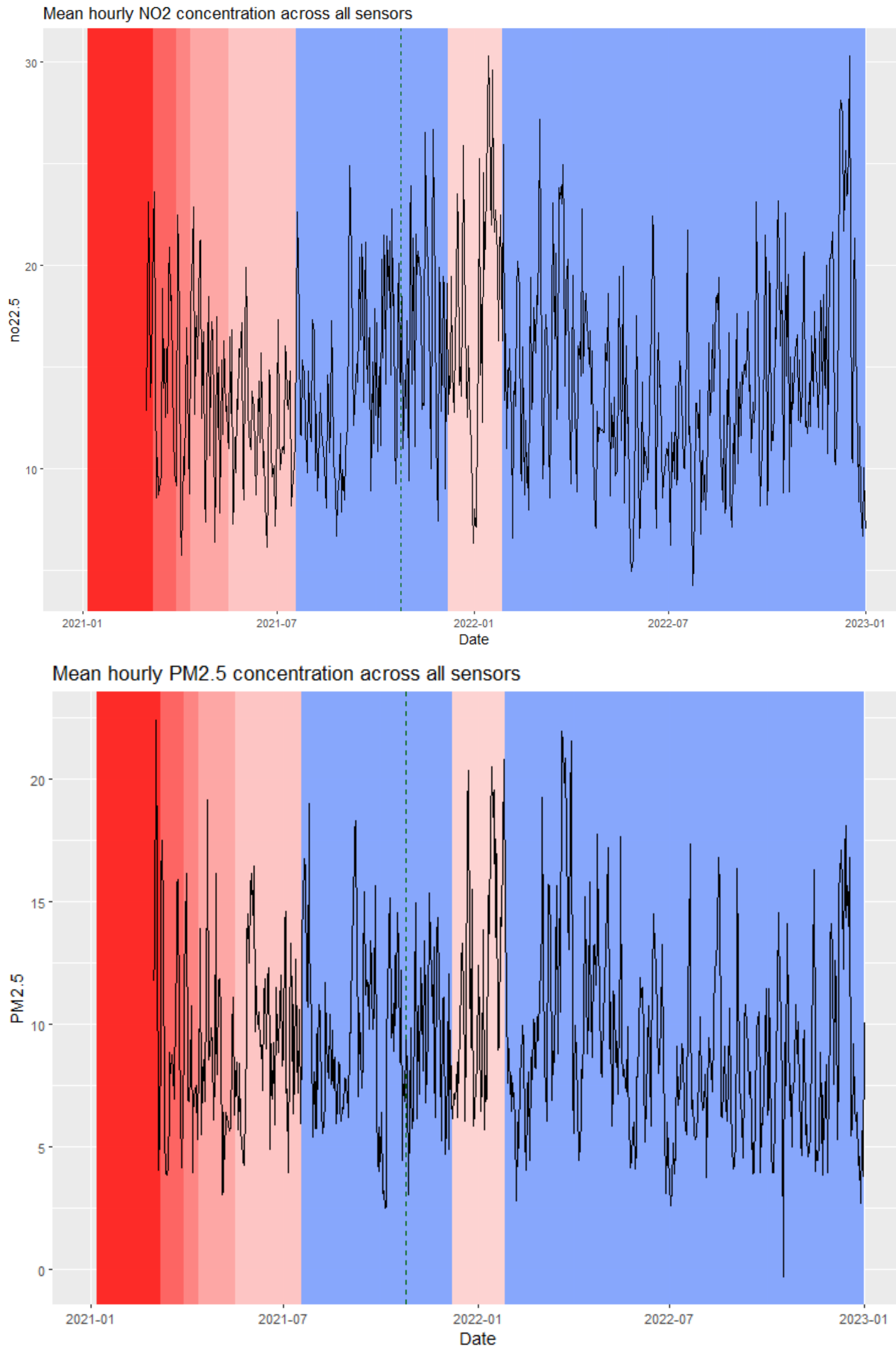


Figure 5: Average hourly $PM_{2.5}$ and NO_2 concentration across all sensors for the study period before meteorological normalisation. **Key:** Red represents Covid-19 restrictions in place and blue represents time periods were all restrictions were lifted. The dotted line represents the introduction of the ULEZ.

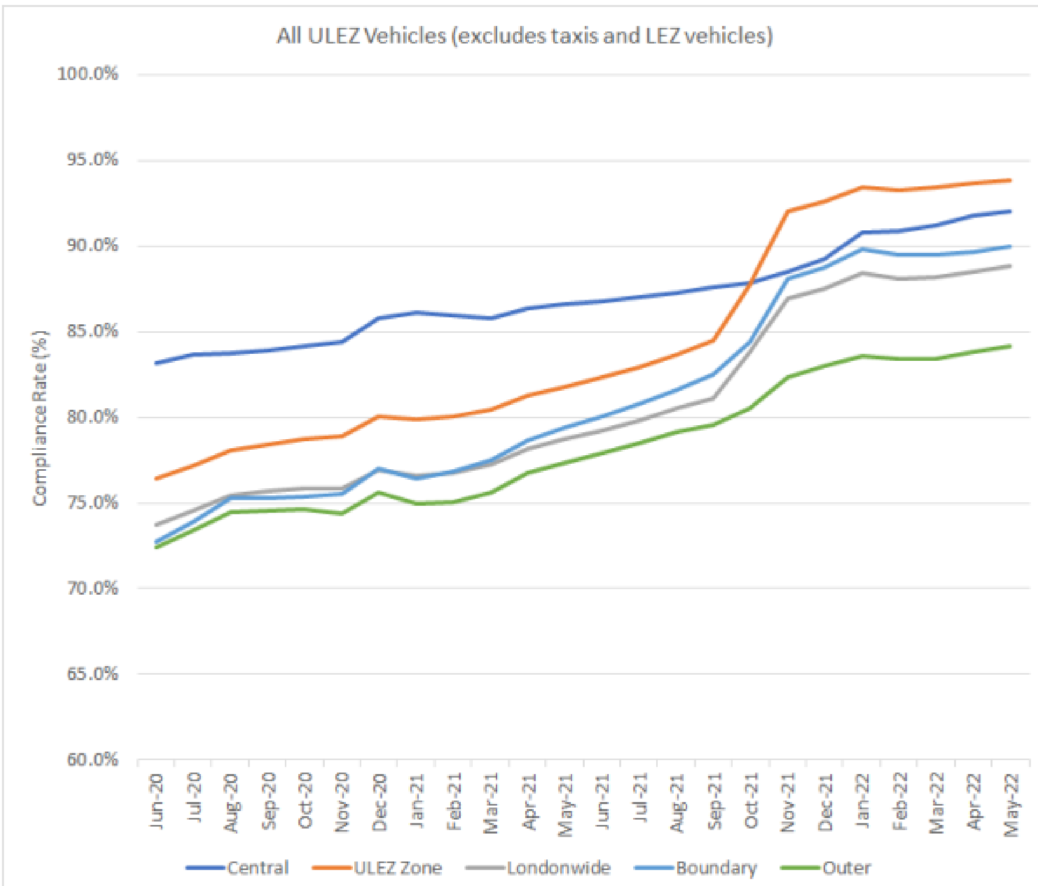


Figure 6: ULEZ vehicle compliance with Euro emissions standards, split by zone (37).

3.2 Meteorological normalisation

In preparation for answering our first research question (R1), quantifying the correlation between compliance and pollution levels, we performed meteorological normalisation to reduce noise in the $PM_{2.5}$ and NO_2 time series and make the data suitable for analysis. The GBRT models were built using the full study period's data, with a random 80% vs 20% split for the training and test set. The features in the model were wind speed (ws), wind direction (wd), air temperature, day of the week, month of the year, and a trend term. 5-Fold cross-validation determined that the optimal number of trees were 3986 for the $PM_{2.5}$ model and 4773 for the NO_2 model.

Model evaluation

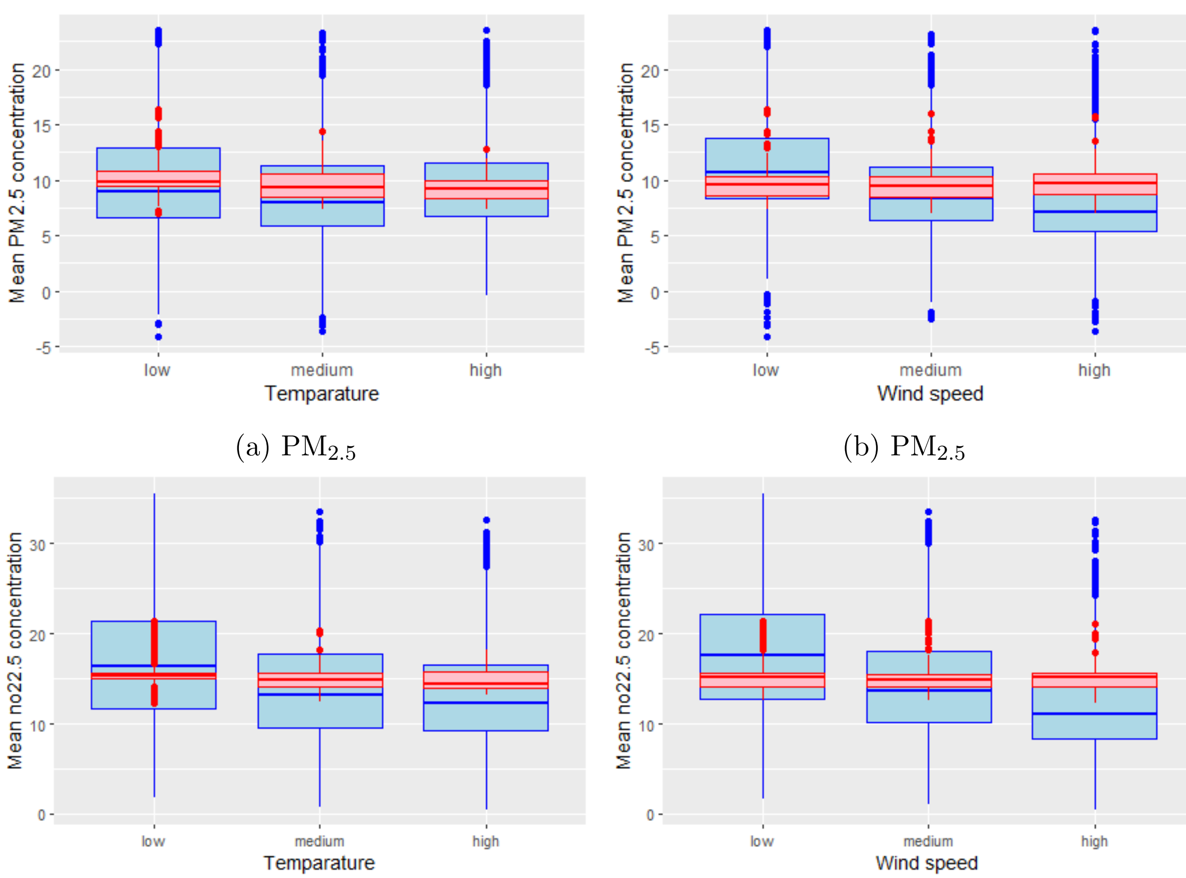
The model showed a good fit, with measured versus predicted values approximating $y = x$, as shown in figure 14 in the appendix. Moreover, a number of useful metrics were calculated (table 14) in order to evaluate the models. The results can be summarised as follows:

- Both models had exceptional performance as indicated by low FAC2 (Fractional Gross Error), MB (Mean Bias), NMB (Normalized Mean Bias), and NMGE (Normalized Mean Gross Error) values.
- Strong correlation coefficients showed how well the models captured the trends in the data.
- Although the models displayed root mean square errors (RMSE) of 2.02 and 4.86 for $PM_{2.5}$ and NO_2, respectively, it is crucial to emphasize that meteorological normalisation aims not at perfect prediction, but rather at the substantial reduction of weather-related biases and noise in the data. Moreover, the RMSE values, albeit non-negligible, represent only 7% and 11% of the total concentration value range for $PM_{2.5}$ and NO_2, respectively. This modest deviation from the observed data range has minimal impact on the overall analysis

The consistently favourable results across multiple metrics highlighted both models' generalisability and accuracy in making predictions on unseen data. To test the robustness of the model, different parameters were used, however the model's performance didn't change much. Figure 7 illustrates visually how the effects of air temperature and wind speed on $PM_{2.5}$ and NO_2 were significantly reduced after applying meteorological normalisation. This is seen by the fact that the box plots after normalisation have relatively constant medians and smaller ranges. Three temperature and wind speed bands were created by splitting the features into 3 equal parts, labelled as "low", "medium" and "high". Overall, these metrics suggested that the meteorological normalisation achieved its goal, so we were able to apply the meteorological averaging procedure and normalise the concentrations.

Table 4: Metrics for model evaluation of $PM_{2.5}$ and NO_2 GBRT models.

Statistic	Pollutant	Training	Testing
$PM_{2.5}$		5680	1420
NO_2	n	8823	2206
$PM_{2.5}$		0.98	0.98
NO_2	FAC2	0.97	0.96
$PM_{2.5}$		0.00	0.04
NO_2	MB	0.00	-0.14
$PM_{2.5}$		0.98	1.33
NO_2	MGE	2.36	2.92
$PM_{2.5}$		0.00	0.00
NO_2	NMB	0.00	-0.01
$PM_{2.5}$		0.10	0.14
NO_2	NMGE	0.16	0.20
$PM_{2.5}$		1.34	2.02
NO_2	RMSE	3.09	3.86
$PM_{2.5}$		0.98	0.88
NO_2	r	0.86	0.76



(a) $PM_{2.5}$ (b) $PM_{2.5}$

(c) NO_2 (d) NO_2

Figure 7: Box plots of normalised and non-normalised pollution concentrations for temperature and wind speed bands.

Model interpretation

Our meteorological models provide us not only with a way to remove the effects of meteorological variables from our pollution time series, but also help us understand how each of these covariates affect the pollution concentrations. Our approach allows us to isolate the individual impacts of each covariate, which is particularly noteworthy, because in many cases, disentangling these effects from one another is challenging.

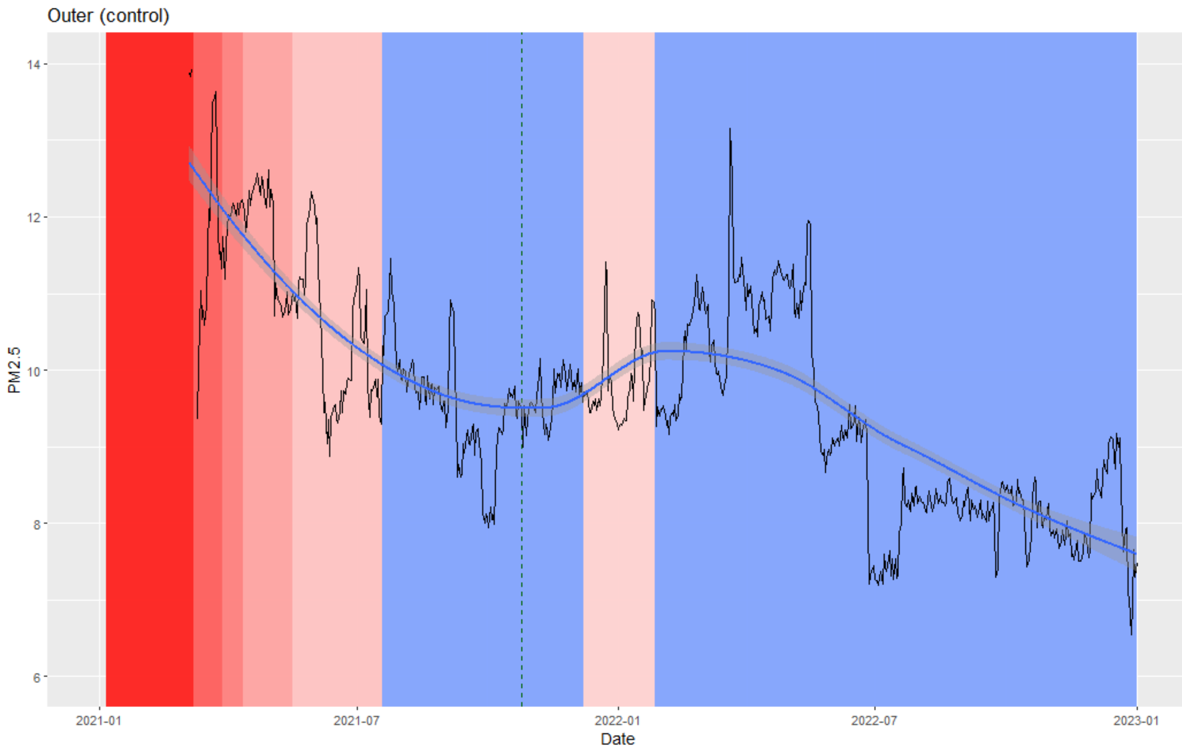
In order to understand how the pollutant concentrations are affected by meteorological conditions, we examined the partial dependencies, shown in figure 15 in the appendix. The partial dependencies give the relationship between the pollutant of interest and the covariates used in the model while holding the value of other covariates at their mean level. We can infer that:

- Higher air temperature corresponds with higher pollution concentrations.
- Lower wind speeds correspond with higher pollution levels.
- Pollution concentrations seem to stay relatively constant with wind direction.
- Pollution concentrations are the highest on Tuesday through Thursday, and the lowest on Sunday and Monday.
- Pollution levels are the lower in summer months and higher in winter months. Note that this trend is not affected by the temperature at the time since it is calculated by keeping the temperature constant.

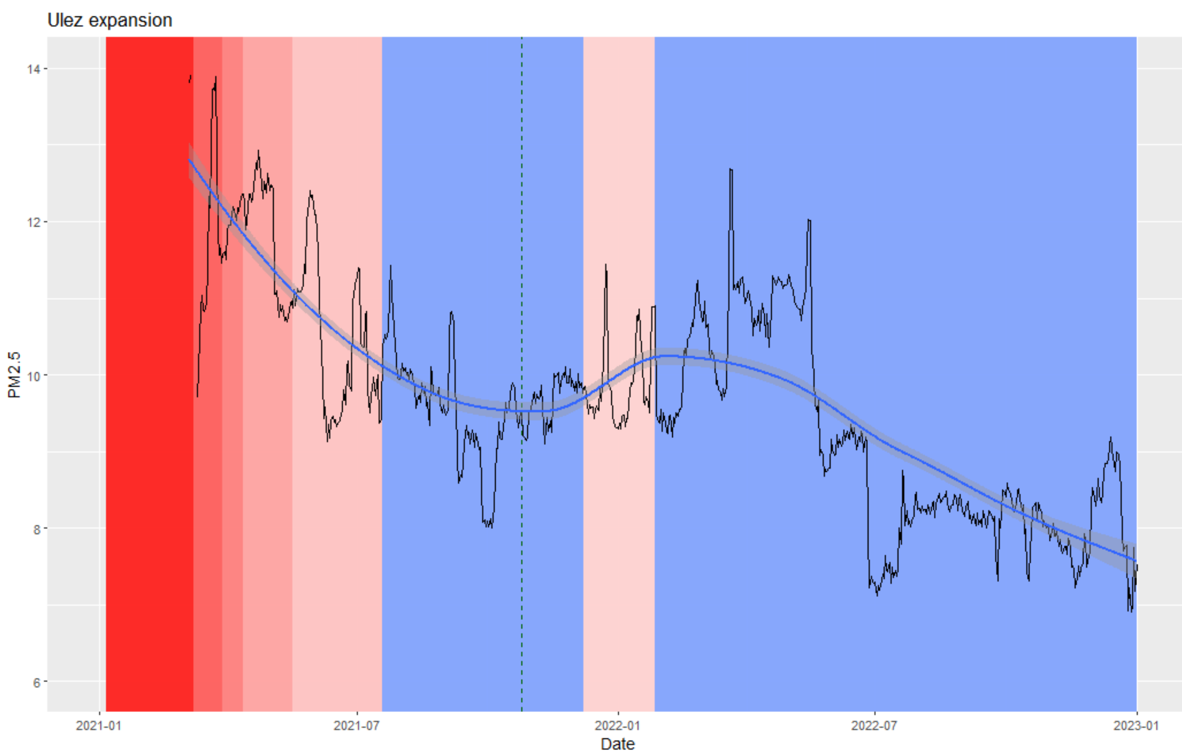
Next, we were able to check whether the interaction between different meteorological variables affected the pollution concentrations by looking at the two way dependencies. Most notably, we saw that a low air temperature combined with high wind speed resulted in low pollution concentrations, shown in figure 16 in the appendix. Then, to understand how much each meteorological variable influenced our model, we checked the the relative influences, shown in figure 17 in the appendix. Wind speed had the greatest influence and weekday had the smallest influence for both $PM_{2.5}$ and NO_2. Month of the year had a larger influence on both models than air temperature.

Normalised concentration graphs

The result of our meteorological normalisation is shown in the normalised time series concentration graphs in figures 9 ($PM_{2.5}$) and 8 (NO_2) for both the expansion (intervention) and outer (control) zones. The smoothed line was calculated using Locally Weighted Scatterplot Smoothing. The red areas of the graphs represent the severity of the Covid-19 lockdown at the time, and the dashed line represents the introduction of the ULEZ expansion. The trend for both the expansion and non-expansion areas is that the mean pollution concentrations are generally decreasing over time. However, it is difficult draw meaningful conclusions simply comparing these graphs, since they graphs do not take into account ULEZ compliance rates, nor do they give insight into whether any results are statistically significant. Therefore, we proceed to apply change point detection and a difference-in-differences model to rigorously investigate whether the ULEZ expansion caused an increase in Euro vehicle standard compliance, and whether this in turn caused emissions to fall.

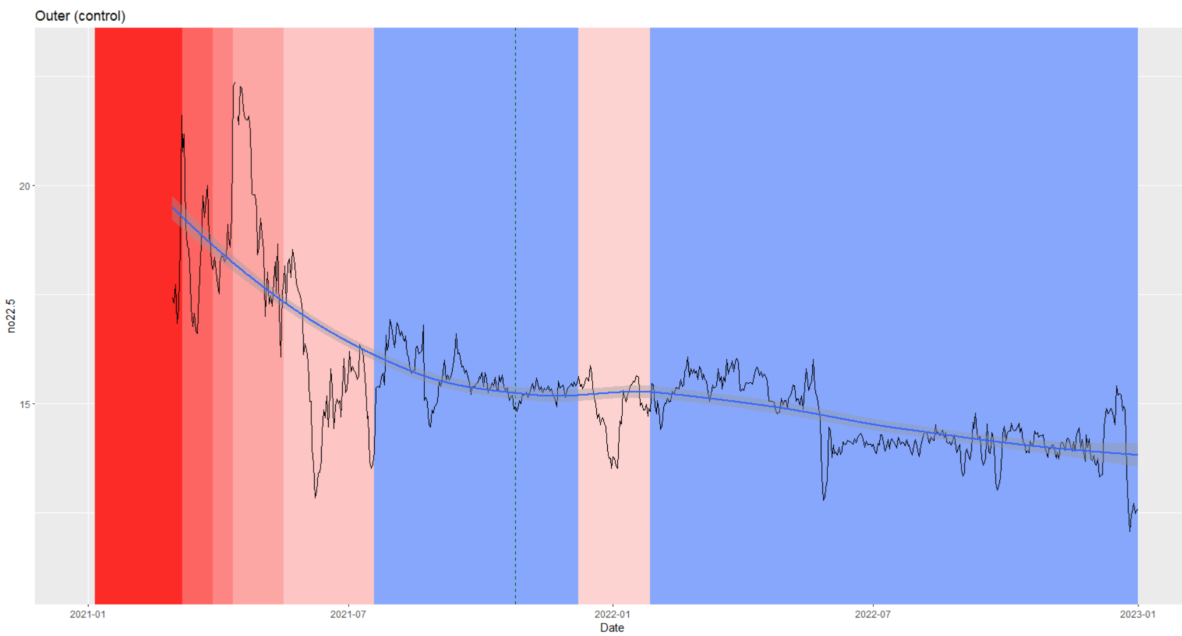


(a) Outer zone

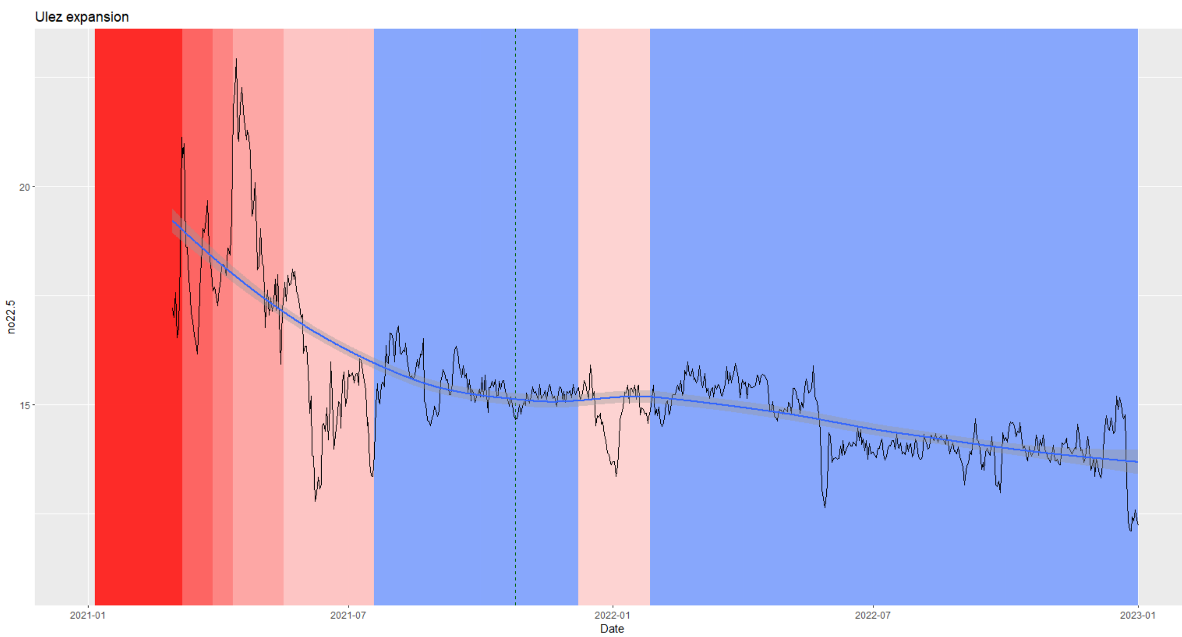


(b) Expansion zone

Figure 8: $PM_{2.5}$ time series concentration graph in the expansion versus the outer (control) region.



(a) Outer zone



(b) Expansion zone

Figure 9: NO_2 time series concentration graph in the expansion versus the outer (control) region.

3.3 Effect of compliance on pollution levels (R1)

After the meteorological normalisation had been performed, we moved on to our main research aim, which was evaluating the success of the ULEZ expansion. As described in section 2, we split this into two research questions, our first being "Is an increase in ULEZ compliance correlated with a decrease in pollution levels?". To answer this question, we built linear regression models for $PM_{2.5}$ and NO_2, modelling air pollution concentration levels as a function of ULEZ compliance. Covariates were not included in the model due to the fact that meteorological variables had already been accounted for. The results of the regressions are shown in tables 5 and 6.

For $PM_{2.5}$, the baseline (i.e no ULEZ compliance) concentration estimate was 16.4 $\mu g/m^3$, with each unit increase in compliance correlated with a decrease of 0.072 $\mu g/m^3$. This means that in theory, if all vehicles were ULEZ compliant, we would expect a decrease of 7.2 $\mu g/m^3$ of $PM_{2.5}$, (i.e. a 43.81% decrease) than if no vehicles were compliant. The NO_2 regression gave a baseline of 27.5 $\mu g/m^3$, with each unit increase in compliance decreasing this by 0.14 $\mu g/m^3$. Therefore, at full compliance, we would expect a decrease of 14 $\mu g/m^3$, or 49.76% compared to baseline. There results show the strong inverse correlation between ULEZ compliance and air pollution levels, and suggest that increasing ULEZ compliance is an effective way to lower air pollution levels.

Table 5: Estimating the effect of ULEZ compliance on $PM_{2.5}$ concentration.

	$PM_{2.5}$ concentration estimate ($\mu g/m^3$)	Std. Error	t value	Pr(>|t|)
(Intercept)	16.41	0.60	27.17	0.00 ***
compliance	-0.072	0.0072	-10.01	0.00 ***
$^{*}p < 0.05$,	$^{**}p < 0.01$,	$^{***}p < 0.001$		

Table 6: Estimating the effect of ULEZ compliance on NO_2 concentration.

	NO_2 concentration estimate ($\mu g/m^3$)	Std. Error	t value	Pr(>|t|)
(Intercept)	27.47	0.80	34.17	0.00***
compliance	-0.14	0.0096	-14.30	0.00***

3.4 Effect of ULEZ on vehicle compliance (R2)

Once we had quantified the correlation between ULEZ compliance levels and air pollution concentrations, we moved on to understanding whether the October 2021 ULEZ expansion caused a change in ULEZ compliance levels. Answering this question was more difficult, as there are multiple ways in which an intervention may affect an outcome. Therefore we used a multi-faceted approach to quantify this effect, including difference-in-differences modelling and change point analysis.

Difference-in-differences model

We began by applying difference-in-difference modelling in order to estimate how the expansion and outer zones changed in relation to one another during the ULEZ expansion. We used the expansion and outer groups, along with the intervention date of October 25, 2021. A linear regression was performed, as described in section 2.4, with compliance (per 100 vehicles) estimated as a function of the time period and zone. The results are shown in table 7. 'Intercept' is the baseline compliance estimate, which is an estimate for mean compliance in the outer zone, during the pre-expansion time period. 'Expansion zone' and 'Post expansion' are indicator functions (dummy variables).

Table 7: Results of difference-in-differences model for estimating the effect of the ULEZ expansion on ULEZ vehicle compliance rates.

	Compliance estimate (per 100 vehicles)	Std. Error	t value	Pr(>|t|)
(Intercept)	78.37	0.11	741.67	0.00***
Expansion zone	5.016	0.15	33.57	0.00***
Post expansion	4.75	0.16	29.83	0.00***
Expansion zone × post expansion	4.55	0.23	20.21	0.00***
$^{*}p < 0.05$,	$^{**}p < 0.01$,	$^{***}p < 0.001$		

Our model gave statistically significant results, showing that the average compliance for vehicles driving pre-expansion and in the outer zone (which we assumed to be the baseline scenario), was 78.47 per 100 vehicles. This increased by 5.02 (to 83.49 per 100) for drivers in the expansion zone, and by 4.75 (to 83.22 per 100) for those driving post-expansion. If both of these conditions were met, there was a further increase of 4.55, reaching a compliance rate of 92.79 per 100 vehicles. This means that the compliance rate post-expansion was 14.32 per 100, or 18.2% higher than the baseline scenario. For ease of reading, the coefficients from the model are calculated for each possible scenario (pre/post-intervention and expansion/outer zone), and shown in table 8.

The DID estimate (interaction term coefficient) of 4.55 (per 100 vehicles), indicates that the outer and expansion zones both experienced an increase in compliance after the ULEZ expansion, but the expansion zone experienced an increase that was 4.55 (per 100 vehicles) higher than the outer zone. The fact that the expansion and outer zones do not experience a very large difference between them could be due to various factors, including the possibility that the ULEZ expansion was not actually responsible for a large increase in compliance, or the fact that the ULEZ expansion was responsible for the increase in compliance in both the outer zone, as well as the expansion zone. To investigate the second possibility, change point analysis was applied to check whether the outer zone experienced a sharp increase in compliance at the time of the ULEZ expansion.

Change point analysis

In the DID section, we found a difference of 4.55 (per 100 vehicles) between the expansion zone's compliance increase and the outer zone's compliance increase. However, we suspected that this was due to the fact that the outer zone was also affected by the ULEZ expansion, and not that only 4.55 was attributable to the ULEZ expansion. To investigate this, we applied change point analysis to assess whether there were any sharp, sudden changes in the mean compliance level at the point of the ULEZ introduction in either the outer or expansion zones. October 2021 was excluded from the analysis to account for the fact that people likely upgraded their cars or changed their driving

Table 8: Difference-in-differences model coefficients for predicting ULEZ compliance (per 100 vehicles) as a function of zone and time period.

	Outer zone	Expansion zone
Pre-expansion	78.47	83.49
Post-expansion	83.22	92.79
Difference	4.75 (6.05% increase)	9.30 (11.14% increase)
Difference-in-differences	**4.55**	

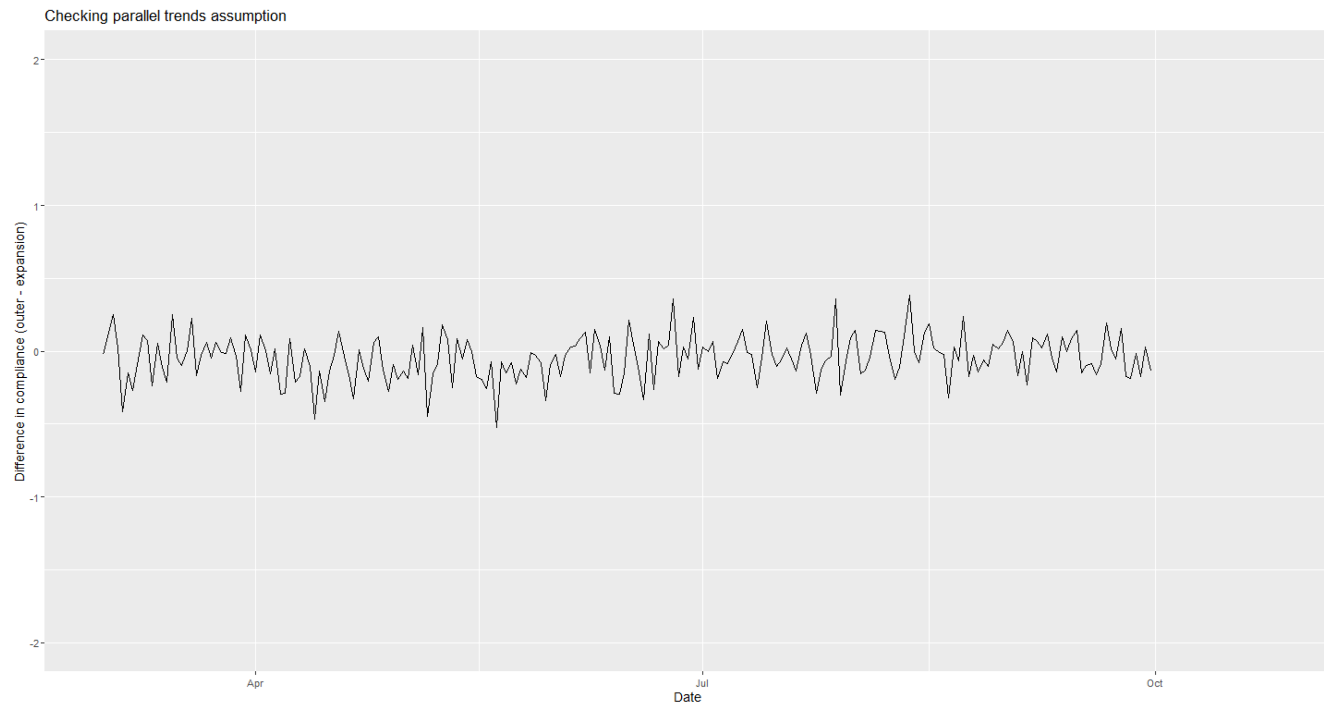


Figure 10: Checking the parallel trends assumption: outer zone compliance minus expansion zone compliance pre-ULEZ expansion.

habits prior to the expansion date on October 25. The parallel trends assumption was checked visually by looking at the difference between compliance rates in the two zones, as shown in figure 10, where we can see that the differences in compliance are very small. Binary segmentation was applied along with the MBIC penalty. Figures 11a and 11b show the results of this change point analysis for the expansion and outer zones respectively. In both zones, there is a change point exactly at the time of the ULEZ expansion, indicating there was an increase in vehicle compliance both in the ULEZ expansion and the outer zones. This provides evidence to suggest that the outer zone was also affected by the ULEZ expansion, and therefore not a perfect control, meaning that the DID estimate of 4.55 was very likely an underestimate. It also provides new evidence to suggest the expansion was responsible for the increase in both zones, as the significant changes in the model are all recorded around the time of the ULEZ expansion. Therefore, we conclude that the ULEZ was responsible for a 9.30 and 4.75 (per 100 vehicles) increase in compliance. for the expansion and outer zones. A change point analysis was performed on the difference in compliance between the zones. This result is shown in figure 11c, indicating a change point shortly before, and at the ULEZ expansion. This agrees with our DID estimate, that the expansion zone experienced an increase in compliance that was 4.55 (per 100 vehicles) higher than the outer zone's increase.

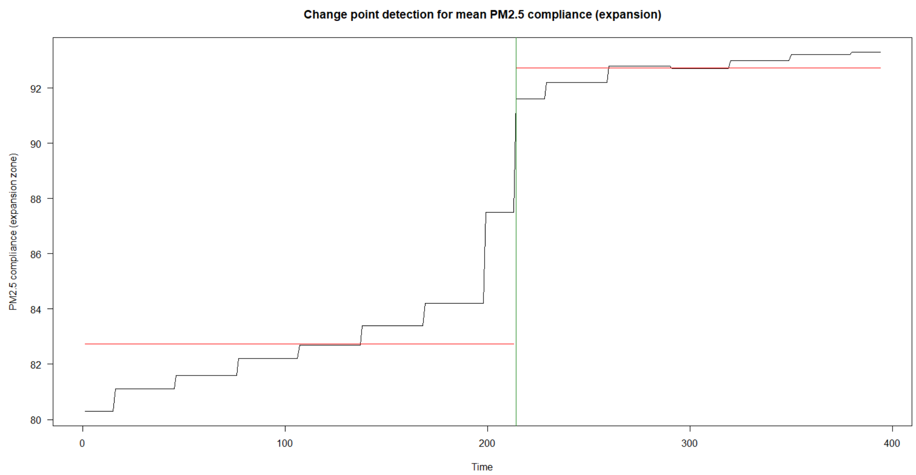
3.5 Summary of results

In order to evaluate the success of the ULEZ expansion, we looked at two main aspects of the scheme: Firstly, whether an increased ULEZ compliance is correlated with lower air pollution levels, and secondly, whether the October 2021 ULEZ expansion caused an increase in compliance levels. To answer the former, we performed meteorological normalisation to account for confounding variables and remove noise from the data, and then performed two separate linear regressions, modelling $PM_{2.5}$ and NO_2 concentration levels as functions of ULEZ compliance rates. The results showed strong inverse correlations between ULEZ compliance levels and both pollutants, with a 43.81% decrease for $PM_{2.5}$ and a 49.76% decrease for NO_2, for full compliance compared to no compliance.

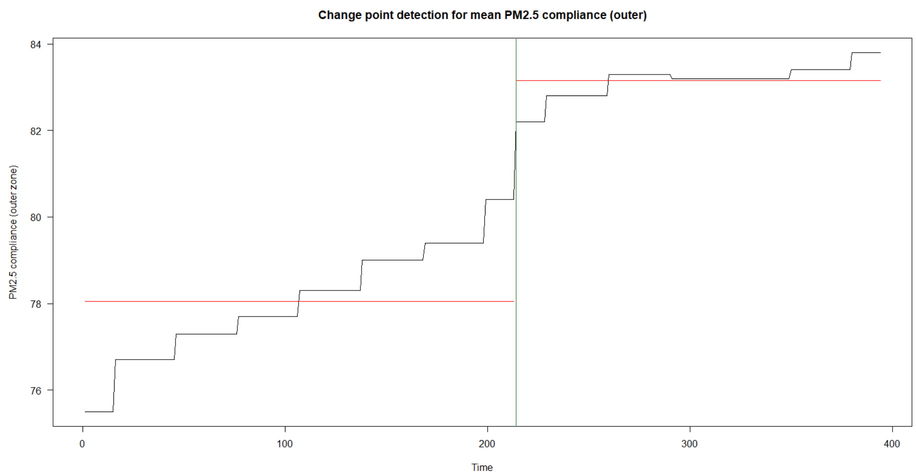
Next, to quantify the effect of the ULEZ expansion on compliance levels, Difference-in-differences

(DID) modelling was performed, which showed an 18.4% increase in compliance post-expansion, in the expansion zone, from the baseline scenario (pre-expansion, in the outer zone). Both the outer and expansion zones experienced an increase in compliance, at 4.75 and 9.30 (per 100 vehicles) respectively. Therefore, our DID estimate, calculated as the difference between those increases, was only 4.55 (per 100 vehicles). This either meant that the ULEZ expansion was only responsible for a 4.55 increase in compliance in the expansion zone, or it meant that the outer group was not in fact a perfect control, and was also affected by the ULEZ expansion, resulting in an underestimate of the effect. To investigate the second possibility, we employed change point analysis to check whether there was a sharp increase in compliance in the outer zone at the time of the expansion. We did, indeed, find a sudden, sharp increase in compliance in both zones at the intervention date, indicating that the ULEZ expansion was likely responsible for the increase in compliance in both zones, but caused a larger increase in the expansion zone.

Our analysis indicates the importance of ULEZ vehicle compliance in reducing air pollution levels, and helps us to understand how an intervention such as the ULEZ expansion changes vehicle compliance rates. In light of our results, we will evaluate the ULEZ expansion policy in detail in section 4.



(a) Expansion zone (penalty of 9773)



(b) Outer zone (penalty of 2537)

(c) Expansion minus outer zone (penalties of 2350 and 82)

Figure 11: Change point analysis results for ULEZ compliance level for expansion and outer zones. The green line represents the ULEZ expansion date.

4 Discussion

4.1 Overview

The core objective of this study was to assess the success of the 2021 ULEZ expansion in London. In order to evaluate the success of the scheme, we looked at two main aspects: compliance and air pollution trends. We asked: (R1) "Is an increase in ULEZ compliance associated with a decrease in NO_2 or $PM_{2.5}$ concentrations?" and (R2) "Did the 2021 ULEZ expansion cause an increase in ULEZ vehicle compliance?".

Effect of ULEZ compliance on air pollution concentration (R1)

In order to answer R1, we first needed to remove the meteorological noise from our pollution time series. We employed a gradient boosted regression tree (GBRT) model to effectively mitigate the influence of meteorological variables, including air temperature, wind speed, wind direction, day of the week, and month of the year, from our concentration time series data. Separate models were trained for $PM_{2.5}$ and NO_2, and subsequently applied to both the outer (control) and expansion zones, due to their shared weather conditions. Notably, both models demonstrated commendable performance on unseen test data, with the $PM_{2.5}$ model exhibiting superior predictive accuracy. The meteorological averaging procedure was then performed by running the model 200 times on different subsets of the data, in order to come up with an "average" scenario. Finally, the pollution concentrations were then normalised by removing the residual concentrations from the time series. The new normalised time series for both pollutants showed general downward trends in both the expansion and outer zones.

R1 was then investigated by building two separate linear regression models, using compliance rate as the independent variable and pollution concentration (of $PM_{2.5}$ or NO_2) as the dependent variable. Meteorological covariates were not included due to the fact that they had already been accounted for in the normalisation. These models indicated a strong correlation between compliance and pollution concentration. For every 1/100 cars compliant with ULEZ standards, $PM_{2.5}$ levels decreased by 0.072 $\mu g/m^3$ and NO_2 by 0.14 $\mu g/m^3$. In other words, if compliance was at 100/100 cars, concentrations of $PM_{2.5}$ and NO_2 would be 43.8% and 49.8% lower compared to if there was no compliance.

Effect of the ULEZ expansion on vehicle compliance (R2)

To answer R2, we used compliance rates from the ULEZ expansion 6 month report (37), which were captured through ANPR data and cross-checked with DVLA records. This means that compliance was not defined using people normally residing in an area, but rather as a collection of all the vehicles driving through the zone, meaning that one vehicle could contribute to compliance rates in both the expansion and outer zones.

We aimed to quantify the effect of the ULEZ expansion on compliance levels using a multi-faceted approach. Difference-in-differences modelling was performed, which indicated that the outer zone experienced an increase in compliance of 4.75 (per 100 vehicles) and the expansion zone experienced an increase in compliance of 9.30 (per 100 vehicles). This corresponds to relative increases of 6.05% and 11.14% respectively. The DID estimate, calculated as the difference between those increases, was only 4.55 (per 100 vehicles). This either meant that the ULEZ expansion was only responsible for the 4.55 increase in compliance in the expansion zone, or that the outer group was not actually a perfect control, and was also affected by the ULEZ expansion. We used change point analysis to investigate whether the ULEZ expansion caused an increase in the outer zone. We did, indeed, find a sudden, sharp increase in compliance in both zones at the intervention date, indicating that the ULEZ expansion was likely responsible for the increase in compliance in both zones, but caused a larger increase in the expansion zone.

4.2 Our unique approach

What sets our study apart from many others assessing the success of ULEZ implementations is our inclusion of compliance rates. While numerous studies directly investigate the impact of ULEZ on pollution concentrations, we believe that focusing solely on pollution levels can obscure critical insights. The rationale behind framing our research question in terms of compliance rates stems from our perspective on how to define the success of the intervention. In the context of our research question, the study started with a high compliance rate. Therefore, the possible increase in compliance was limited, and so the goal of the ULEZ expansion was not to cause a drastic increase in compliance, but rather to make compliance as close to 100% as possible. Therefore, evaluating the effect of compliance on air pollution and the effect of the expansion on compliance separately, allows us to fairly evaluate both whether ULEZ compliance is helping reduce air pollution, as well as how well government interventions succeeded in encouraging compliance. To make our perspective more clear, allow us to illustrate it with an anaology: if a vaccine is developed to combat a rare, yet deadly disease, and every susceptible individual is effectively protected, one would not observe a significant reduction in disease prevalence. However, this does not imply that the vaccine failed. In fact, it achieved a 100% success rate. This is similar to our situation in the sense that the compliance being high to begin with does not take away the success of the ULEZ expansion.

4.3 Interpretation of results

Our results for R1 show that 100% ULEZ vehicle compliance is associated with a 43.8% and 49.8% decrease in $PM_{2.5}$ and NO_2, compared to 0% vehicle compliance. It is important to note that our analysis did not use causal inference, but only estimated correlation. Therefore, it is a possibility that another influence is responsible for the decrease in air pollution levels. However, the fact that the main meteorological covariates were accounted for in our normalisation process rules them out as a possible confounders, and makes an inverse causal relationship between ULEZ compliance and air pollution concentration a likely reality. There is also a host of existing research concluding that Euro vehicle standards (which ULEZ standards are derived from) cause air pollution to decrease (42; 43; 44). We therefore have high confidence that our identified correlation is in fact causal, and so an increase in ULEZ compliance does cause $PM_{2.5}$ and NO_2 to decrease.

After concluding that an increase in ULEZ compliance likely causes air pollution levels to fall, our second research question (R2) aimed to investigate whether the October 2021 ULEZ expansion caused an increase in ULEZ vehicle compliance. This question was more difficult to answer due to the fact that ULEZ compliance levels are only recorded for London, making it difficult to find a suitable control. We began by using outer London as a control area, but change point analysis indicated a sharp increase in compliance in both the expansion and outer zones at the expansion date, likely due to the fact that many people drove in both zones. The fact that both sharp changes in compliance were detected at the time of the expansion is evidence to suggest that these changes were a result of the expansion. This corresponded to an increase of 9.30 vehicles (per 100 vehicles) in the expansion and 4.74 in the outer zone. Although this may not seem large at first glance, this brings the final expansion compliance up to 92.79 (per 100 vehicles), which is a very high rate of compliance. Furthermore, it is also plausible that people began upgrading their cars before the intervention date, in anticipation for the scheme, or due to driving in the original 2019 ULEZ zone. The ULEZ expansion was announced back in 2018, meaning effects from the expansion may go as far back as then which have not been taken into account. The changes in compliance we recorded are therefore likely an underestimate. In light of our results, we view the expansion as a success in raising the compliance levels to a high level, but acknowledge that there is room for improvement of the scheme in order to get to full compliance.

4.4 Limitations

It is important to acknowledge the limitations of our study. Our main limitation for R1 was the fact that we only tested correlation between ULEZ compliance and air pollution concentration, and used literature to confirm that this relationship was causal. Moreover, we did not include traffic data in our models which may have been a confounder. For R2, we did not have a perfect control area when estimating the effect of the ULEZ expansion on compliance levels, since ULEZ compliance is currently only recorded in London. To get around this, we used change point analysis to infer causality, as the likelihood of a compliance increase at the ULEZ expansion date being caused by another variable is very low. Other studies do not have this problem since they do not structure their research questions using compliance rates. However, they miss out a host of valuable insights into how the ULEZ intervention affected people's driving behaviours. There were also likely spillover effects from people upgrading their cars or changing their driving habits before the introduction of the ULEZ, leading to an underestimate of the effect size. Despite the limitations in our study, our results were robust, and offered valuable insight into the efficacy of the ULEZ expansion.

4.5 Existing results and future research directions

It is difficult to compare our results to existing research due to the fact that almost all current research is about the 2019 ULEZ implementation, and does not use compliance estimates as an integral part of their analysis. Moreover, most existing studies use DID modelling and incorporate meteorological variables into their models, but the advantage of us having used meteorological normalisation instead is that it produces a concentration time series which is much more interpretable. This allows us to visually analyse the yearly trends in air pollution, which is hugely beneficial when the trend is difficult to quantify. As far as future research goes, it is clear that more climate interventions need to be implemented in order to mitigate the public health risk from air pollution. We found that ULEZ compliance is currently very high (over 90%) in the expansion zone, and our research suggests that if the outer zone experiences a similar trend with the new 2023 expansion, it will increase to a similar level. However, even in the best case scenario of 100% compliance, our models still predicted NO_2 concentration of 13.8 $\mu g/m^3$ and a $PM_{2.5}$ concentration of 9.22 $\mu g/m^3$. These air pollution levels are still both higher than the WHO recommended guidelines of 10 and 5 $\mu g/m^3$ for NO_2 and $PM_{2.5}$ (41). Moreover, there is no safe lower level of $PM_{2.5}$, with every 10 $\mu g/m^3$ of $PM_{2.5}$ associated with a relative increase of 1.05% in daily mortality rate (37). Therefore, it is vital to research new interventions to further decrease London's air pollution levels. Other areas to research include energy production and construction. We also suggest that more studies are carried out on compliance rates of air quality interventions, as this would be hugely beneficial in guiding climate policies. We believe that investigating compliance levels is as important as pollution trends, since an intervention can only work if the public compiles with it. Therefore, we urge researchers to use this multi-faceted approach in assessing climate policy.

Conclusion

In conclusion, our study sheds new light on the efficacy of the 2021 ULEZ expansion, using compliance rates as an integral part of our research. Firstly, we found a strong inverse correlation between ULEZ compliance and both $PM_{2.5}$ and NO_2 pollution concentrations, with every unit increase in ULEZ compliance responsible for a decrease of 0.072 $\mu g/m^3$ of $PM_{2.5}$ and 0.14 $\mu g/m^3$ of NO_2. Moreover, our results showed that the 2021 ULEZ expansion caused an increase of 9.04 (per 100 vehicles) in the expansion zone, which is high considering the fact that the mean compliance before the expansion was 83.49 (per 100 vehicles). We therefore concluded that the 2021 ULEZ expansion was successful. The implications of these findings serve not only to direct policy on future ULEZ expansions, but also as a test case for low emissions zones globally.

Acknowledgements

I would like to extend my sincere thanks toward the Environmental Research Group at Imperial College London. In particular, I am appreciative of my supervisor, Dr Iq Mead, as well as Dr Ian Gang Chen, for providing me with expert guidance during our weekly meetings, and giving me feedback on two of my drafts; to Laura Buchanan, for collecting and helping me get acquainted with the data, and to Timothy Baker, for providing useful suggestions at the beginning of the project. Moreover, I am eternally grateful to Professor Stephen Peté, Dr Marí Peté, Reilly Tilbury and Puzzle Tilbury-Peté for their unwavering support, belief and encouragement throughout my academic journey, as this would not have been possible without them.

References

[1] World Health Organization (2023). *Air Pollution*. [online] Who.int. Available at: https://www.who.int/health-topics/air-pollution#tab=tab_1 [Accessed 23 May 2023].

[2] GOV.UK. (n.d.). Particulate matter (PM10/PM2.5). [online] Available at: https://www.gov.uk/government/statistics/air-quality-statistics/concentrations-of-particulate-matter-pm10-and-pm25#:~:text=The%20Environmental%20Targets%20(Fine%20Particulate. [Accessed 23 May 2023].

[3] Di, Q., Dai, L., Wang, Y., Zanobetti, A., Choirat, C., Schwartz, J.D. and Dominici, F. (2017). Association of Short-term Exposure to Air Pollution With Mortality in Older Adults. *JAMA*, 318(24), p.2446. doi: https://doi.org/10.1001/jama.2017.17923.

[4] Kurt, O.K., Zhang, J. and Pinkerton, K.E. (2016). Pulmonary health effects of air pollution. *Current Opinion in Pulmonary Medicine*, 22(2), pp.138–143. doi: https://doi.org/10.1097/mcp.0000000000000248.

[5] Al-Kindi, S.G., Brook, R.D., Biswal, S. and Rajagopalan, S. (2020). Environmental determinants of cardiovascular disease: lessons learned from air pollution. *Nature Reviews Cardiology*, 17. doi: https://doi.org/10.1038/s41569-020-0371-2.

[6] Carnell, E., Vieno, M., Vardoulakis, S., Beck, R., Heaviside, C., Tomlinson, S., Dragosits, U., Heal, M. and Reis, S. (2019). Modelling public health improvements as a result of air pollution control policies in the UK over four decades—1970 to 2010. *Environmental Research Letters*, 14(7), p.074001. doi: https://doi.org/10.1088/1748-9326/ab1542.

[7] Greater London Authority (2022). *AIR QUALITY IN CITY OF LONDON: A GUIDE FOR PUBLIC HEALTH PROFESSIONALS*. [online] London Assembly . Greater London Authority. Available at: https://www.london.gov.uk/sites/default/files/col_air_quality_for_public_health_professionals.pdf [Accessed 23 May 2023].

[8] Vega, E., Namdeo, A., Bramwell, L., Miquelajauregui, Y., Resendiz-Martinez, C.G., Jaimes-Palomera, M., Luna-Falfan, F., Terrazas-Ahumada, A., Maji, K.J., Entwistle, J., Enríquez, J.C.N., Mejia, J.M., Portas, A., Hayes, L. and McNally, R. (2021). Changes in air quality in Mexico City, London and Delhi in response to various stages and levels of lockdowns and easing of restrictions during COVID-19 pandemic. *Environmental Pollution*, 285, p.117664. doi: https://doi.org/10.1016/j.envpol.2021.117664.

[9] Ma, L., Graham, D.J. and Stettler, M.E.J. (2021). Has the ultra low emission zone in London improved air quality? *Environmental Research Letters*, 16(12), p.124001. doi: https://doi.org/10.1088/1748-9326/ac30c1.

[10] 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.

[11] Costa, S., Ferreira, J., Silveira, C., Costa, C., Lopes, D., Relvas, H., Borrego, C., Roebeling, P., Miranda, A.I. and Paulo Teixeira, J. (2014). Integrating Health on Air Quality Assessment—Review Report on Health Risks of Two Major European Outdoor Air Pollutants: PM and NO2. *Journal of Toxicology and Environmental Health, Part B*, 17(6), pp.307–340. doi: https://doi.org/10.1080/10937404.2014.946164.

[12] www.eea.europa.eu. (n.d.). *Health impacts of air pollution in Europe, 2022 — European Environment Agency*. [online] Available at: https://www.eea.europa.eu/publications/air-quality-in-europe-2022/health-impacts-of-air-pollution.

[13] Barnes, J.H., Chatterton, T.J. and Longhurst, J.W.S. (2019). Emissions vs exposure: Increasing injustice from road traffic-related air pollution in the United Kingdom. *Transportation Research Part D: Transport and Environment*, 73, pp.56–66. doi: https://doi.org/10.1016/j.trd.2019.05.012.

[14] Wakeling, D., Passant, N., Murrells, T., Pang, Y., Thistlethwaite, G., Walker, C., Garcia, J. and Webb, J. (2016). UK Informative Inventory Report (1990 to 2014). [online] Available at: https://uk-air.defra.gov.uk/assets/documents/reports/cat07/1603150959_GB_IIR_2016_Final.pdf.

[15] More mileage in reducing urban air pollution from road traffic. (2021). *Environment International*, 149, p.106329. doi: https://doi.org/10.1016/j.envint.2020.106329.

[16] Greater London Authority (2021). *EXPANDED ULTRA LOW EMISSION ZONE – FIRST MONTH REPORT*. [online] London, p.2. Available at: https://www.london.gov.uk/sites/default/files/ulez_first_month_report_december_2021.pdf [Accessed 25 May 2023].

[17] US EPA (2018). *NAAQS Table — US EPA*. [online] US EPA. Available at: https://www.epa.gov/criteria-air-pollutants/naaqs-table.

[18] GOV.UK (2021). *Concentrations of nitrogen dioxide*. [online] GOV.UK. Available at: https://www.gov.uk/government/statistics/air-quality-statistics/ntrogen-dioxide.

[19] GOV.UK (2022). *Emissions of air pollutants in the UK – Nitrogen oxides (NOx)*. [online] GOV.UK. Available at: https://www.gov.uk/government/statistics/emissions-of-air-pollutants/emissions-of-air-pollutants-in-the-uk-nitrogen-oxides-nox.

[20] London Datastore. (2023). *London Atmospheric Emissions Inventory (LAEI) 2019*. [online] Available at: https://data.london.gov.uk/dataset/london-atmospheric-emissions-inventory--laei--2019 [Accessed 26 May 2023].

[21] Polichetti, G., Cocco, S., Spinali, A., Trimarco, V. and Nunziata, A. (2009). Effects of particulate matter (PM10, PM2.5 and PM1) on the cardiovascular system. *Toxicology*, 261(1-2), pp.1–8. doi: https://doi.org/10.1016/j.tox.2009.04.035.

[22] Pope, C.A. and Dockery, D.W. (2006). Health Effects of Fine Particulate Air Pollution: Lines that Connect. *Journal of the Air & Waste Management Association*, 56(6), pp.709–742. doi: https://doi.org/10.1080/10473289.2006.10464485.

[23] Greater London Authority (2019). *CENTRAL LONDON ULTRA LOW EMISSION ZONE – SIX MONTH REPORT*. [online] Available at: https://www.london.gov.uk/sites/default/files/ulez_six_month_evaluation_report_oct19.pdf [Accessed 13 Jun. 2023].

[24] Hajmohammadi, H. and Heydecker, B. (2022). Evaluation of air quality effects of the London ultra-low emission zone by state-space modelling. *Atmospheric Pollution Research*, 13(8), p.101514. doi: https://doi.org/10.1016/j.apr.2022.101514.

[25] Zhai, M. and Wolff, H. (2021). Air pollution and urban road transport: evidence from the world's largest low-emission zone in London. *Environmental Economics and Policy Studies*, 23. doi: https://doi.org/10.1007/s10018-021-00307-9.

[26] Prieto-Rodriguez, J., Perez-Villadoniga, M.J., Salas, R. and Russo, A. (2022). Impact of London Toxicity Charge and Ultra Low Emission Zone on NO2. *Transport Policy*. doi: https://doi.org/10.1016/j.tranpol.2022.10.010.

[27] Ceballos-Santos, S., González-Pardo, J., Carslaw, D.C., Santurtún, A., Santibáñez, M. and Fernández-Olmo, I. (2021). Meteorological Normalisation Using Boosted Regression Trees to Estimate the Impact of COVID-19 Restrictions on Air Quality Levels. International Journal of Environmental Research and Public Health, 18(24), p.13347. doi: https://doi.org/10.3390/ijerph182413347

[28] Carslaw D.C. Deweather—An R Package to Remove Meteorological Variation from Air Quality Data. 2021. Available at: https://github.com/davidcarslaw/deweather [Accessed on 20 July 2023].

[29] Friedman, J.H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. The Annals of Statistics, [online] 29(5), pp.1189–1232. Available at: JSTOR [Accessed on 29 July 2023].

[30] Elith, J., et al. "A Working Guide to Boosted Regression Trees." Journal of Animal Ecology, vol. 77, no. 4, July 2008, pp. 802–813, https://doi.org/10.1111/j.1365-2656.2008.01390.x.

[31] "How to Interpret Decision Trees with 1 Simple Example - inside Learning Machines." Inside Learning Machines, 3 Feb. 2023, https://insidelearningmachines.com/interpret_decision_trees/ [Accessed on 1 Aug. 2023].

[32] Gareth Michael James, et al. An Introduction to Statistical Learning : With Applications in R. New York, Springer, 2013.

[33] X. Li and R. Bai, "Freight Vehicle Travel Time Prediction Using Gradient Boosting Regression Tree," 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA), Anaheim, CA, USA, 2016, pp. 1010-1015, doi: https://doi.org/10.1109/ICMLA.2016.0182.

[34] Huang, Yifei, et al. "GBRTVis: Online Analysis of Gradient Boosting Regression Tree." Journal of Visualization, vol. 22, no. 1, 24 Sept. 2018, pp. 125–140, https://doi.org/10.1007/s12650-018-0514-2.

[35] "ULEZ 2022 Map - the Ultra Low Emission Zone Map in London." ULEZ, 8 Aug. 2020, https://ulez.co.uk/ultra-low-emission-zone-2021-map/. [Accessed on 25 July 2023].

[36] Carslaw D (2023). worldmet: Import Surface Meteorological Data from NOAA Integrated Surface Database (ISD). https://davidcarslaw.github.io/worldmet/index.html, http://davidcarslaw.github.io/worldmet/ [Accessed on 3 August 2023].

[37] EXPANDED ULTRA LOW EMISSION ZONE – SIX MONTH REPORT. (n.d.). [online] London Assembly. Mayor of London. Available at: https://www.london.gov.uk/sites/default/files/expanded_ultra_low_emission_zone_six_month_report.pdf [Accessed 22 August 2023].

[38] AIR Index. (n.d.). What are the Euro Emission Standards: Check Your Car. [online] Available at: https://airindex.com/emissions-ratings/euro-emission-standards/ [Accessed 12 August 2023].

[39] Matters, T. for L. — E.J. (n.d.). ULEZ standards. [online] Transport for London. Available at: https://tfl.gov.uk/modes/driving/ultra-low-emission-zone/ways-to-meet-the-standard [Accessed 24 August 2023].

[40] Likelihood Ratio Test Statistic - an overview (n.d.). ScienceDirect Topics. [online] Available at: https://www.sciencedirect.com/topics/mathematics/likelihood-ratio-test-statistic#:~:text=The%20LRT%20statistic%20for%20testing [Accessed 25 August 2023].

[41] World Health Organization (2021). What are the WHO Air quality guidelines? [online] www.who.int. Available at: https://www.who.int/news-room/feature-stories/detail/what-are-the-who-air-quality-guidelines [Accessed 27 August 2023].

[42] Harrison, R.M. and Beddows, D.C. (2017). Efficacy of Recent Emissions Controls on Road Vehicles in Europe and Implications for Public Health. Scientific Reports, 7(1). doi: https://doi.org/10.1038/s41598-017-01135-2.

[43] Mehlig, D., Woodward, H., Oxley, T., Holland, M. and ApSimon, H. (2021). Electrification of Road Transport and the Impacts on Air Quality and Health in the UK. Atmosphere, 12(11), p.1491. doi: https://doi.org/10.3390/atmos12111491.

[44] A comparison of real driving emissions from Euro 6 diesel passenger cars with zero emission vehicles and their impact on urban air quality compliance Urban air quality study: extension I. (n.d.). Available at: https://www.concawe.eu/wp-content/uploads/2018/04/Rpt_18_8.pdf[Accessed29Aug.2023] [Accessed 27 August 2023].

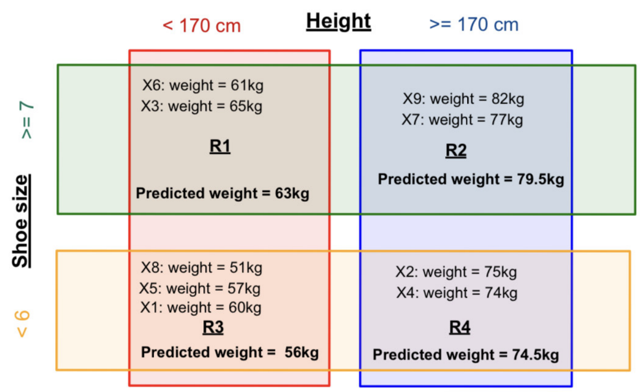
A Appendix

Impurity Measure	**Formula**	**Description**
Mean Squared Error (MSE)	$\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$	Average squared difference.
Mean Absolute Error (MAE)	$\frac{1}{N}\sum_{i=1}^{N}\lvert y_i - \hat{y}_i\rvert$	Average absolute difference.
Huber Loss	$\frac{1}{N}\sum_{i=1}^{N}L_\delta(y_i - \hat{y}_i)$	Combination of MSE and MAE.
Poisson Deviance	$2\sum_{i=1}^{N}(y_i \log(y_i)$ $-y_i \log(\hat{y}_i))$	For Poisson regression tasks.
Exponential Deviance	$2\sum_{i=1}^{N}(y_i - \hat{y}_i)$ $-y_i \log(y_i - \hat{y}_i)$	For exponential regression tasks.
Variance Reduction (Anova)	$\mathrm{Var}(D) - \mathrm{Var}(D\vert A)$	Reduction in variance achieved by a split.

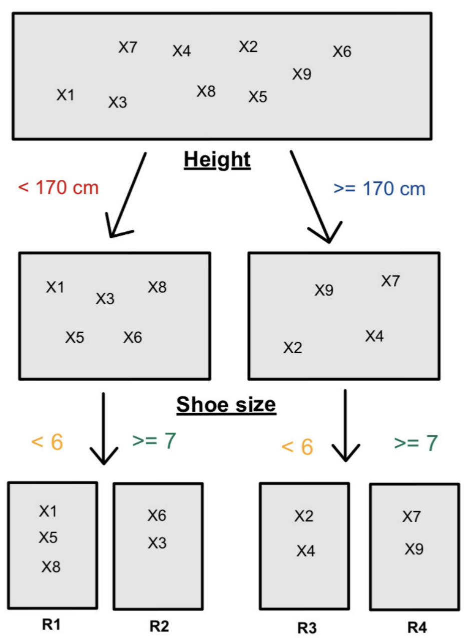
Table 9: Impurity measures for regression tasks.

Impurity Measure	**Formula**	**Description**
Gini Impurity	$\sum_k p_k(1-p_k)$	Probability of misclassification.
Entropy	$-\sum_k p_k \log(p_k)$	Level of uncertainty in class distribution.
Misclassification Error	$1 - \max(p_k)$	Misclassification rate at a node.
Cross-Entropy (Log Loss)	$-\sum_k y_k \log(p_k)$	Difference between true and probabilities. predicted probabilities.
Information Gain	$H(D) - H(D\vert A)$	Reduction in entropy achieved by a split.
Brier Score	$\frac{1}{N}\sum_{i=1}^{N}(p_i - y_i)^2$	Accuracy of probabilistic predictions.

Table 10: Impurity measures for classification tasks.



(a) Space partition



(b) Corresponding decision tree

Figure 12: Regression decision tree example.

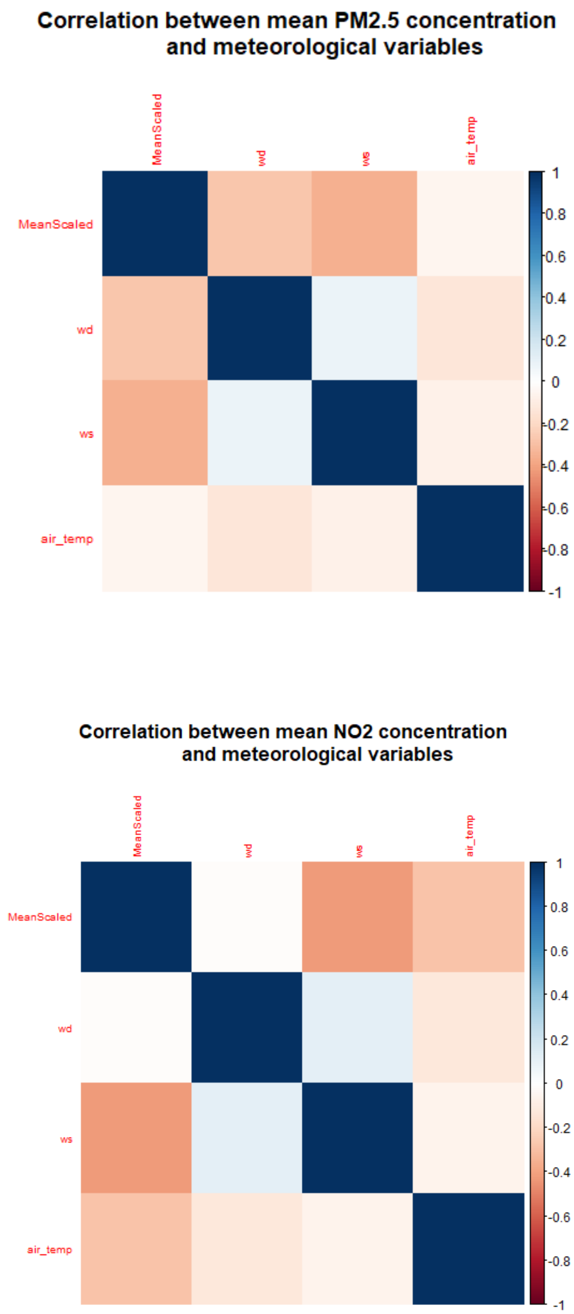
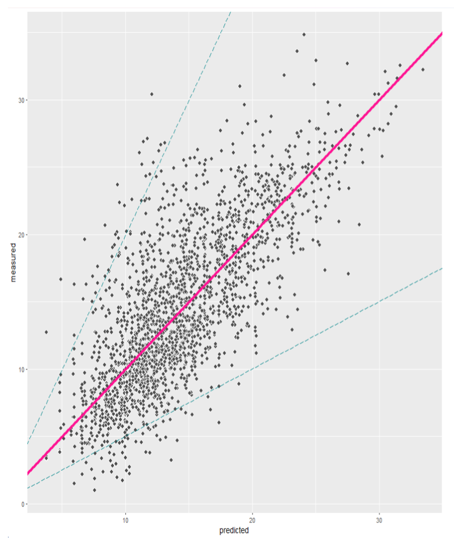
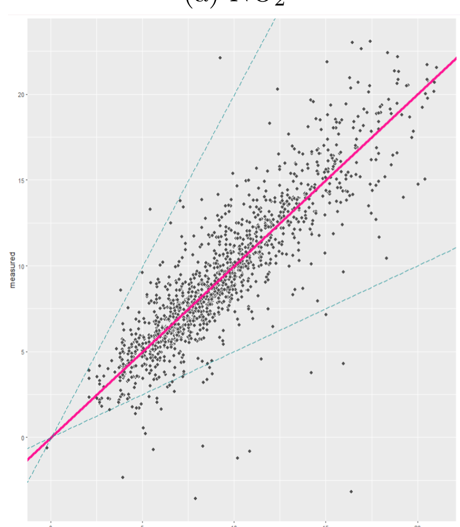


Figure 13: Correlation heat maps between non-normalised NO_2 and $PM_{2.5}$ concentrations with meteorological variables.

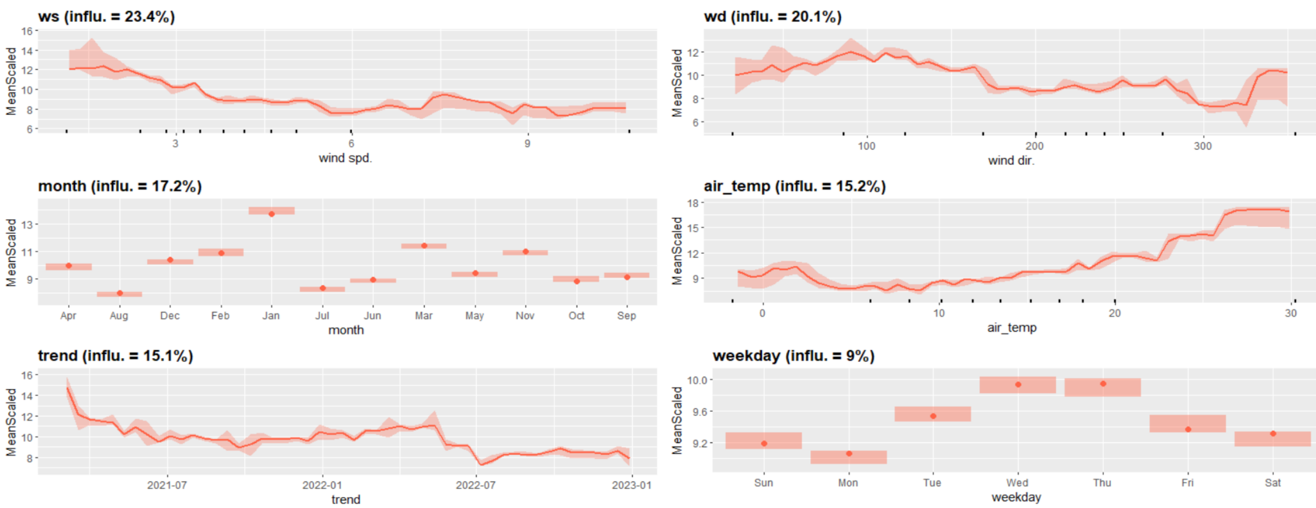


(a) NO_2

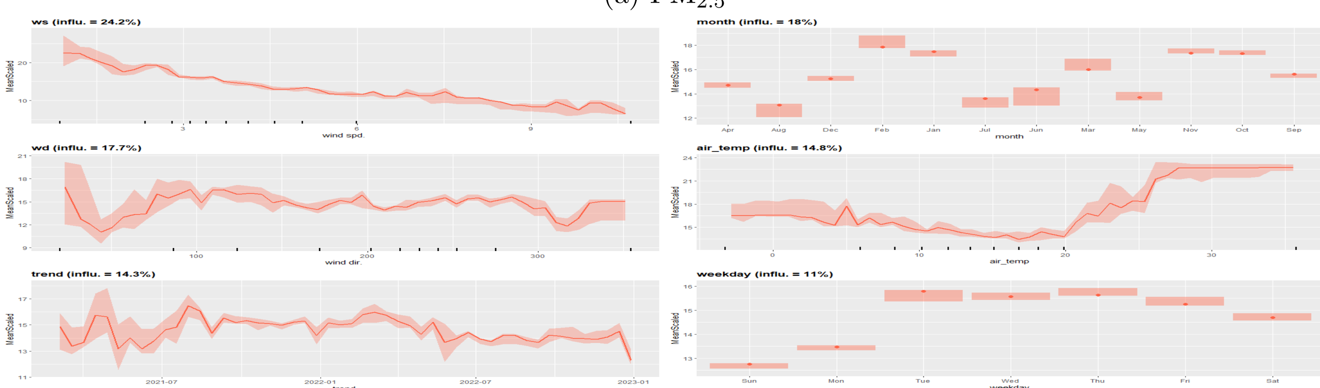


(b) $PM_{2.5}$

Figure 14: GBRT measured versus predicted pollution concentrations.

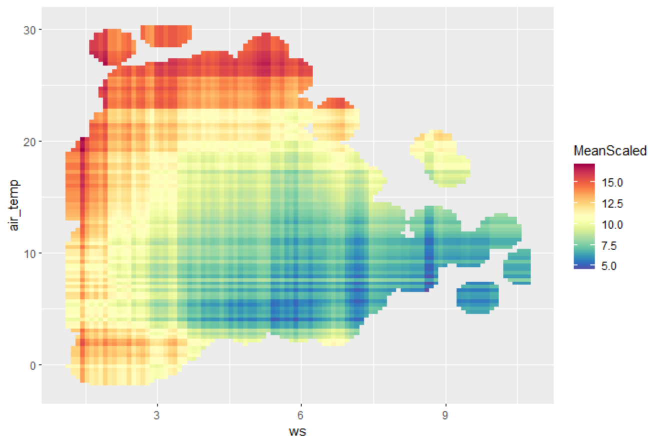


(a) $PM_{2.5}$

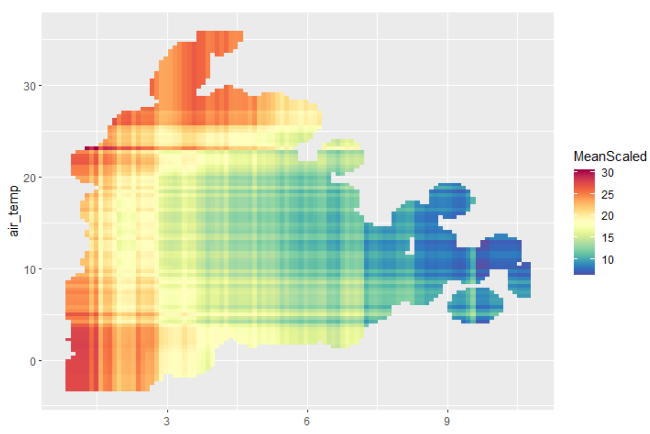


(b) NO_2

Figure 15: Partial dependencies of each meteorological covariate with pollution concentrations.



(a) $PM_{2.5}$



(b) NO_2

Figure 16: Two-way dependencies for pollution concentration versus air temperature and wind speed.

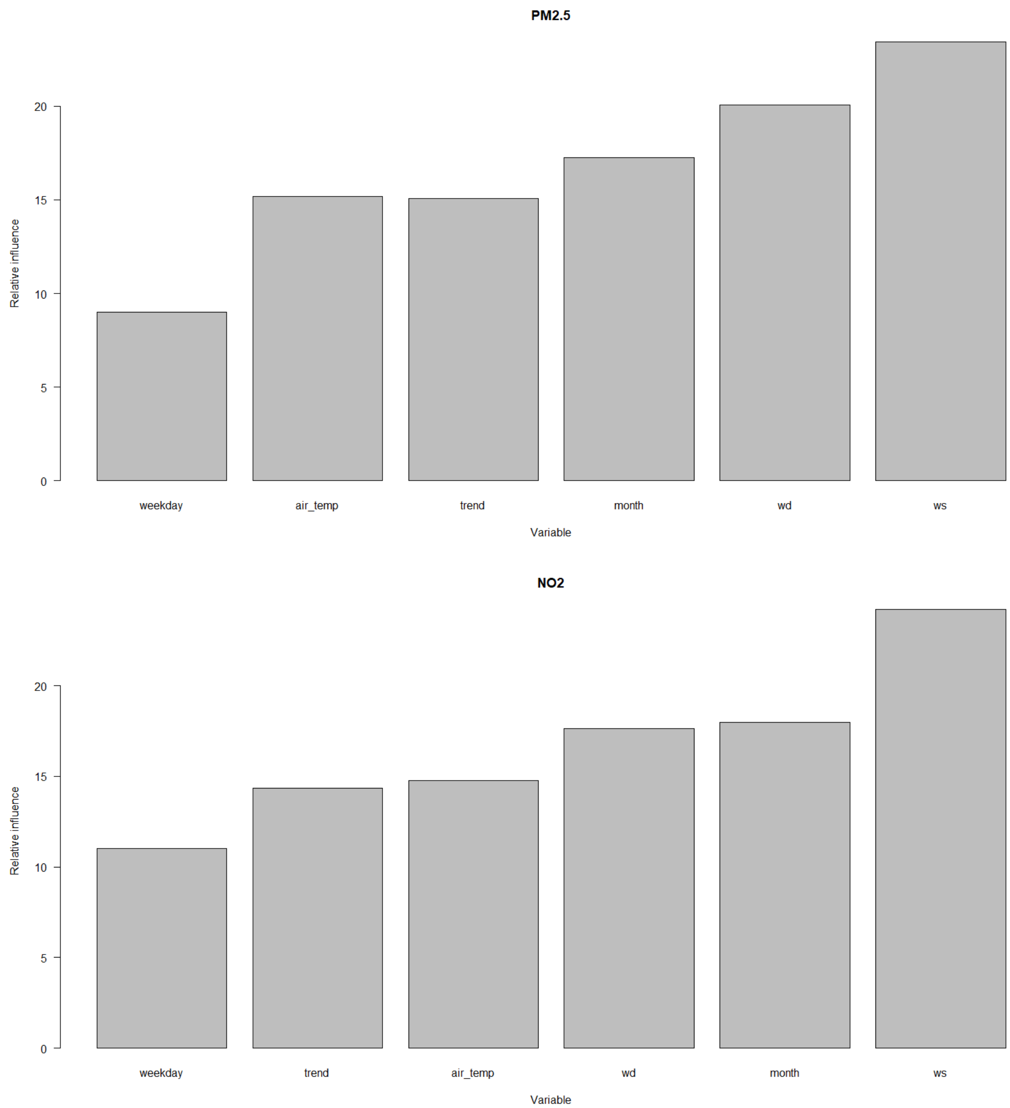


Figure 17: Relative influences of each meteorological variable in the GBRT models.