

Examining the effect and its temporal variations of near-road greenspaces for different types of roads on regional PM_{2.5} levels – a case study in London



(The Independent, 2014)

Yulun Lin (19040824)

2022

This dissertation is submitted as an Independent Geographical Study as a part of a BSc degree in Geography at King's College London.

KING'S COLLEGE LONDON

UNIVERSITY OF LONDON

DEPARTMENT OF GEOGRAPHY

INDEPENDENT GEOGRAPHICAL STUDY

I, Yulun Lin, hereby declare (a) that this dissertation is my own original work and that all source material used is acknowledged therein; (b) that it has been specially prepared for a degree of King's College London; and (c) that it does not contain any material that has been or will be submitted to the Examiners of this or any other university, or any material that has been or will be submitted for any other examination.

This Dissertation is 8134 words.

Signed: Yulun Lin

Date: 7th April 2022

ACKNOWLEDGEMENTS

I would like to thank my supervisor Dr. James Millington, without whom this Independent Geographical Study would have never become possible.

I would also like to thank my family and friends for the support and company during the completion of this IGS as well as throughout the three years of my undergraduate course.

ABSTRACT

The effect of near-road green space on regional PM_{2.5} levels was examined using data from 21 London air quality monitoring sites. A linear regression model was used to investigate the relationship between the annual mean regional PM_{2.5} concentrations and the conditions of green spaces near different classes of roads. The temporal variations in the relationship were also explored by modelling the regional average PM_{2.5} concentrations over different periods as functions of the conditions of green spaces near different types of roads. The results show that near-road green spaces for Unclassified and Other roads have a negative effect on regional PM_{2.5} levels, while the other types have a positive effect, and there are large temporal variations detected. The findings contribute to the planning and application of near-road green spaces in terms of PM pollution mitigation.

Key words: fine particulate matter; near-road greenspace; multivariate linear regression model

Table of Contents

LIST OF FIGURES.....	6
LIST OF TABLES	7
1. Introduction	8
2. Methods	12
2.1 Background	12
PM _{2.5} pollution in London	12
Air Quality Monitoring in London.....	12
Road classification	13
Greenspace in London	13
2.2 Data sources	14
PM _{2.5} data.....	14
Greenspace	15
Road	15
2.3 Data pre-processing and EDA.....	16
Data cleaning for PM _{2.5} data	16
Spatial geometry manipulation	19
Generating explanatory variables.....	20
2.4 Multivariate linear regression models	21
3. Results.....	23

3.1 Modelling annual mean PM _{2.5} concentration	23
3.2 Modelling annual mean PM _{2.5} concentrations for each hour.....	24
3.3 Modelling monthly mean PM _{2.5} concentrations	27
4. Discussion	31
4.1 The indicators of the greenspaces near different types of roads	31
4.2 Overall effects of near-road greenspaces in changing regional PM _{2.5} levels	31
4.3 Temporal changes in the effect of near-road greenspaces	34
5. Conclusion.....	36
Reference List	37
Appendix.....	45

LIST OF FIGURES

<i>Figure 1: Locations of the 21 air quality monitoring sites in London.</i>	14
<i>Figure 2: daily mean concentrations of $PM_{2.5}$ in 2019 compared to WHO guideline and monthly mean.</i>	18
<i>Figure 3: Diurnal change in the $PM_{2.5}$ concentration in London</i>	18
<i>Figure 4: site buffer of CD1 (Camden - Swiss Cottage) as an example.</i>	19
<i>Figure 5: Performance of the 24 models in terms of r-squared value and LOO cross-validation r-squared value. The two dashed lines indicate the average r-squared and LOOCV r-squared values of the 24 models.</i>	25
<i>Figure 6: Feature importance estimations for the 24 models. The five bars from left to right represent: Unclassified road, A road, B road, Classified unnumbered road, Other road. The error bar is presented in black line, which is generated from the calculated standard deviation of each feature importance estimation.</i>	26
<i>Figure 7: Intercepts and coefficients of all features for the 24 models. The dashed lines represent the average of the coefficients from the 24 models.</i>	27
<i>Figure 8: Performance of the 12 models. The dashed lines indicate the average performance of the 12 models.</i>	28
<i>Figure 9: Feature importance estimations for the 12 models. The five bars from left to right represent: Unclassified road, A road, B road, Classified unnumbered road, Other road. The error bar is presented in black line, which is generated from the calculated standard deviation of each feature importance estimation.</i>	29
<i>Figure 10: Intercepts and coefficients of all features for the 12 models. The dashed lines represent the average of the coefficients from the 12 models.</i>	30

LIST OF TABLES

<i>Table 1: Annual mean $PM_{2.5}$ concentration for each site and the annual mean for London in 2019.</i>	<i>17</i>
<i>Table 2: Summary statistics of the explanatory variables and the $PM_{2.5}$ data.</i>	<i>21</i>
<i>Table 3: Multivariate linear regression model for the annual mean $PM_{2.5}$ concentration as a function of the indicators for greenspaces near five types of roads. The sample size of the model is 21.</i>	<i>24</i>

1. Introduction

Particulate matter (PM) refers to small solid and liquid matter suspended in the air, and is one of the most serious threats to human health among all the ambient air pollution. High exposure to PM can cause damage to the human body including the lung (Löndahl *et al.*, 2006), the heart (Sun *et al.*, 2010; Brook *et al.*, 2010) and the airway (González-Flecha, 2004), depending on the size of the particle. PM is mainly classified into two categories according to their aerodynamic diameter (Kim *et al.*, 2015), namely fine particulate matter (PM_{2.5}), which has a diameter smaller than 2.5 µm, and coarse particulate matter (PM₁₀), which has a diameter between 2.5 to 10 µm. The sizes of PMs decide their transport abilities in the atmosphere as well as in the human body. PM_{2.5} tends to travel longer in the atmosphere and penetrate deeper into the human body than PM₁₀. As a result, major health problems related to PM_{2.5} are associated with the lungs, Bronchi branches and Bronchioli (Löndahl *et al.*, 2006), while PM₁₀ mainly causes damage to respiratory systems (airway) (González-Flecha, 2004). It is estimated that more than two million deaths worldwide each year are directly related to diseases caused by air pollution, most of which by fine particulate matter (Shah *et al.*, 2013). PM_{2.5} is the primary contributor to human health issues relating to ambient air pollution.

The hazard mainly comes from exposure to a high concentration of PM_{2.5}, and the lower the concentration, the less the danger it exposes to human health. The WHO guideline value for PM_{2.5} is 15 µg/m³ daily mean or 5 µg/m³ annual mean (WHO, 2021). This guideline represents the highest possible concentration to which the effect of PM_{2.5} on human health is acceptable, but does not guarantee no damage to health. However, most regions around the world, especially regions in developing countries, have PM_{2.5} levels higher than the WHO guidelines (World Bank, 2017).

The fine particulate matter in the atmosphere comes from both anthropogenic and natural sources. The former include combustion of fossil fuels, industrial and agricultural activities, and erosion of pavement by road traffic

(Srimuruganandam and Nagendra, 2012). The natural sources include volcanoes, wildfires, dust storms and sea spray (Anderson *et al.*, 2012). Natural sources contribute only 18% to global PM_{2.5} pollution, with the rest from anthropogenic sources, among which traffic section takes up the highest percentage (Karagulian *et al.*, 2015). Hence, finding a way to mitigate the PM_{2.5} pollution from road transport emissions can greatly reduce the PM_{2.5} level in urban areas.

One proposed approach to this is developing near-road greenspaces. Many researchers have examined the effect of green spaces in reducing regional PM levels (Kończak *et al.*, 2021; Song *et al.*, 2015; Nowak *et al.*, 2006; Lei *et al.*, 2018; Irga *et al.*, 2015; Beckett *et al.*, 2000; Hofman *et al.*, 2016), which is mainly through two mechanisms - mass removal and transmission block. On one hand, vegetation in green spaces can help directly remove the PM from the air by capturing and storing them on the leaf surface as well as in the wax layer (Kończak *et al.*, 2021). On the other hand, green spaces can act as a 'windbreak' that interrupts the dispersion of particulate matter (Morakinyo and Lam, 2016) as well as alter other local meteorological environments including temperature, barometric pressure, relative humidity, etc. which also affect PM level (Hofman *et al.*, 2016).

Based on these theoretical and empirical foundations, near-road greenspaces are believed to have a positive effect on lowering PM concentrations. Indeed, there have been studies finding near-road air quality is significantly improved by vegetation (Morakinyo and Lam, 2016), especially on busy roadsides in open areas (Baldauf *et al.*, 2011). However, Vos *et al.* (2013) found that in some cases, instead of reducing PM_{2.5} concentration, roadside vegetation can actually enhance PM pollution nearby by hindering the wind flow and resulting in an accumulation of particulate matter in the area (Abhijith *et al.*, 2017). Such a finding brings uncertainty to the effect of near-road greenspaces on lowering PM_{2.5} levels in urban areas, and further investigation is needed.

Previous researches on examining the effect of green spaces, especially near-road greenspaces, in reducing PM concentration in urban areas can be divided into

two streams. The first stream primarily focused on assessing the abilities in capturing particles in the air of greenspaces. Liu *et al.* (2015) found that canopy density, leaf area, mean diameter at breast height, average tree height and grass coverage and height in forests could greatly alter PM_{2.5} concentration. Jeanjean *et al.* (2017) and Steffens *et al.* (2012) suggested that vegetation has an overall higher ability in reducing PM pollution during summer because of higher leaf area density. Different vegetation species also have different levels of impact on PM concentrations. For example, cypress trees reduce PM levels more than pine trees (Ji and Zhao, 2014). Variations in the location of vegetation in relation to wind direction can also lead to changes in its ability to reduce PM concentrations (Al-Dabbous and Kumar, 2014). Greenspaces are most effective in reducing PM concentrations when the wind blows from areas of high PM levels (e.g. roads) towards them. Lei *et al.* (2018) found that patterns of greenspaces can also influence their ability to reduce PM pollution. Increasing the differences between areas of greenspace patches as well as their edge complexities can significantly lower PM concentrations. A series of meteorological factors including wind (Przybysz *et al.*, 2018; Wang *et al.*, 2015; Przybysz *et al.*, 2014; He *et al.*, 2020), precipitation (Xu *et al.*, 2017; Wang *et al.*, 2015) and solar radiation (temperature) (Wang *et al.*, 2015) can also change the effect of green space on reducing PM_{2.5} levels. These research findings contribute extensively to the academic understanding and policy-making of the urban greenspaces in tackling PM pollution. However, most of them failed to consider temporal changes. PM_{2.5} concentrations in different seasons can vary greatly, and even within one day, the concentrations have highs and lows. In these cases, the influences of greenspaces on PM concentrations could also be changing. Moreover, the studies that did consider temporal changes all had an approach that was through field measurements, which, while delivering valuable first-hand data and solid mechanism-level understandings, were not convincing enough if were to be applied to a larger scale.

The other stream that includes this subject used land-use regression (LUR) extensively to examine how land use types affect spatial-temporal changes in

PM_{2.5} levels. Wu *et al.* (2017) utilized a LUR model with PM_{2.5} concentrations and monthly NDVI (Normalized Difference Vegetation Index) data in Taipei, and found a strong negative correlation between them. Xu *et al.* (2019) also found a relation between forest land type and PM_{2.5} level through LUR. However, the problem with the LUR technique is that it always suffers from multicollinearity (Ross *et al.*, 2007), which makes the model output less reliable. To overcome this, Kim (2020) developed a partial least-squares regression model, which minimizes the influence of multicollinearity of the variables, to study the effects of land use on PM levels in different seasons in Seoul, South Korea, and found that the percentage of green space area is negatively related to regional PM concentrations. Yet, none of them was able to evaluate the relation between near-road greenspaces and regional PM_{2.5} concentrations.

Therefore, although urban greenspaces have been proven to have a significant effect on PM reduction, the influence of near-road greenspaces and the temporal changes in the influence are still not clear. Given the fact that road traffic is the largest contributor to PM_{2.5} pollution in most parts of the world (Karagulian *et al.*, 2015), it is important to determine whether near-road greenspaces have a positive or negative effect on reducing PM concentrations. Hence, this study aims to examine the role of near-road greenspaces with regard to PM_{2.5} concentration, taking London as a case study city. To be more specific, PM_{2.5} data from 21 selected air quality monitoring sites across London were used to examine the relationships between different types of near-road greenspaces and PM_{2.5} concentration as well as the temporal changes in the relationships. It is recognised that greenspaces that are near different types of roads will have different effects on reducing PM concentrations, so the near-roads greenspaces were classified into several categories according to their road types. The result of this study can enrich the understanding of near-road greenspaces' effect on lowering regional PM_{2.5} concentrations and its temporal change, and inform local urban green space planning.

2. Methods

2.1 Background

London has a population of approximately 9 million and covers a land area of around 1500 km². It is the largest city in the UK and has one of the busiest road traffic in the country. It is characterised by a temperate oceanic climate, with warm to hot summer and cool winter, and high precipitation all year.

PM_{2.5} pollution in London

The annual average PM_{2.5} level in London was reported as 13.3 µg/m³ in 2016 (Mayor of London, 2019), which was above the WHO guideline for annual mean concentration (5 µg/m³). It is estimated that apart from transboundary sources, the largest proportion (30%) of the PM_{2.5} pollution comes from the road transport section (Mayor of London, 2019). In areas with intensive traffic flow (e.g. central London), the PM_{2.5} may be much higher than the annual average level.

Air Quality Monitoring in London

London has one of the largest air quality monitoring networks in the world, with participation from all kinds of organisations and departments. The LAQN (London Air Quality Network) is one of them and is operated by the Environmental Research Group at Imperial College London, in cooperation with TfL (Transport for London), Defra (Department for Environment, Food and Rural Affairs) and local authorities where the monitoring sites are located (London Air, 2022a). It provides the public with open air quality data collected from its monitoring sites all across London. Apart from LAQN, the AURN (Automatic Rural and Urban Network) is another network that provides nationwide hourly air quality data to the public (Defra, 2022), with several sites in London.

The richness of the air quality data is a very important reason for choosing London as the case study city. The spatial change of PM_{2.5} concentrations across London is a very important dimension of this study, hence it is crucial to gather data from different monitoring sites.

Road classification

The roads in London (and the UK in general) are classified into four categories (GOV.UK, 2012):

1. A roads - major roads aiming to provide transport links within or between areas. This type of road should have the highest volume of traffic of the four.
2. B roads - a lower class of roads, often with a poorer physical standard. Intended to feed traffic between A roads and smaller roads
3. Classified unnumbered roads - smaller roads connecting A, B roads with unclassified roads. Also known as C roads
4. Unclassified roads - local roads supporting local traffic. Most roads in the UK fall into this category. This class of roads have the lowest volume of traffic.

Except for those four categories, the motorway is another category of road that provides high-speed long-distance transportation. The number of motorways is much lower than the other four types of roads.

Greenspace in London

London is a green city, with roughly 40% of its area being greenspaces. However, the greenspaces are not evenly distributed across the whole city, with a much larger portion in Outer London and a smaller portion in Inner London. The uneven spatial distribution pattern gives an opportunity to study its relationship with regional air quality, and in this case, with regional PM_{2.5} levels.

There are currently two schemes to protect the urban greenspaces in London, with one focusing on protecting undeveloped land around the city called Green Belt and the other aiming to protect greenspaces within the city called Metropolitan Open Land (MOL). The two designations helped develop and maintain extensive urban green areas in London. 22% of London is specified as Green Belt and another 10% is specified as MOL (GiGL, 2018).

2.2 Data sources

PM_{2.5} data

Hourly mean PM_{2.5} concentration data from 21 air quality monitoring sites across London were downloaded from the London Air website (London Air, 2022b), which is the website of the LAQN. Most of the sites are in the LAQN and the others are in the AURN. The 21 selected sites are located mainly in Inner London, with 2 of them in Outer London. **Figure 1** shows their locations. Their location information was downloaded from London Datastore (London Datastore, 2019).

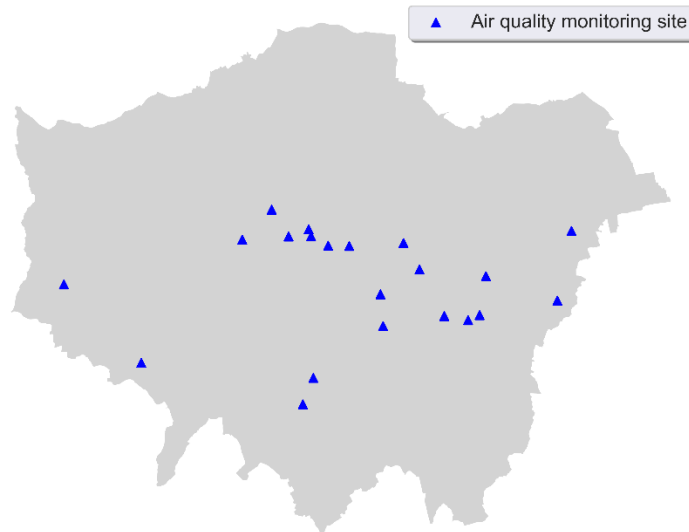


Figure 1: Locations of the 21 air quality monitoring sites in London.

The hourly mean PM_{2.5} concentration data is for the year 2019. This is due to the concern about the impact of the COVID-19 pandemic and lockdown since March 2020. During the lockdown, most PM_{2.5} sources (road traffic in particular) were significantly reduced (Wang and Li, 2021), and therefore the PM_{2.5} level was much lower than normal (pre-COVID level). To minimize the interference, the study period was determined to be the most recent year prior to the pandemic.

It is noteworthy that the PM_{2.5} data used are all provisional (not ratified), so the result of this study should be evaluated and used with caution.

Greenspace

The greenspace information was generated from the OS MasterMap Greenspace Layer (OS, 2021a) provided by the Ordnance Survey, which is the UK's national mapping agency. The MasterMap Greenspace Layer contains all accessible (public parks, sports facilities, etc.) and non-accessible (private garden) urban green spaces in the UK. The map is in the form of vector data divided into 5km x 5km grids. 26 grids were downloaded from the EDINA Digimap Ordnance Survey Collection (Digimap, 2021), which is a collection of OS data owned by EDINA at the University of Edinburgh. The MasterMap Greenspace Layer is updated twice a year, in April and October respectively. The dataset used in this study is from October 2019 in order to synchronise with the PM_{2.5} data.

Road

The road information was generated from the OS Open Roads (OS, 2021b) which is also provided by the Ordnance Survey. This dataset contains not only the spatial geometry of every road in the UK, but also their information such as classification, name, function, etc. The road dataset is also in the form of vector data with a grid size of 100km x 100km, and is also updated twice a year in April and November. The data used in this study is from November 2019.

2.3 Data pre-processing and EDA

For the investigation of near-road greenspaces' effect on regional PM_{2.5} concentrations, this study focus on examining the relationship between the PM_{2.5} data from each air quality monitoring site and the near-road greenspace conditions in the 1km surrounding area around each site. The 1km buffer was decided based on some previous studies (Lei *et al.*, 2018; Chen *et al.*, 2019; Cai *et al.*, 2020) that investigated the effect of urban greenspaces on PM_{2.5}. Before the analysis, some data pre-processing procedures were performed, and explanatory data analysis was then conducted on both the dependent and independent variables.

Data cleaning for PM_{2.5} data

Before the analysis, the PM_{2.5} data was first cleaned. This includes (the concrete process in the Appendix):

1. removing unusual values - some of the values that were very abnormal (e.g. negative PM_{2.5} readings) were removed (set to be null).
2. filling missing values - then all the null values were filled using a technique called the mean-before-after method (Norazian *et al.*, 2008), which is to replace a missing value with the mean of the data points before and after it. In cases where there were several continuous missing values, the closest non-null data points before and after the missing value were used to generate the replacement using linear regression. If the number of continuous missing values exceeded 12 (i.e. half a day), the 12 missing values were then replaced with the values from the same period of the previous day. This is a method commonly used in dealing with missing values within environmental datasets (Chen and Xiao, 2018).

After removing all unusual values and filling all missing values, an initial observation of the PM_{2.5} data found an annual mean concentration of 11.8 µg/m³ with all 21 sites exceeded the WHO guideline of 5 µg/m³, as shown in **table 1**. However, compared to the reported 13.3 µg/m³ annual mean in 2016, most of the

sites had a lower annual mean, which proved that London's past efforts on reducing PM_{2.5} pollution have been working, although the pollution level is still significantly harmful to human.

<i>Siteid</i>	<i>Sitename</i>	<i>Annual mean PM_{2.5} concentration (µg/m³)</i>
<i>BX9</i>	Bexley - Slade Green FDMS	11.2
<i>BL0</i>	Camden - Bloomsbury	10.9
<i>CD9</i>	Camden - Euston Road	13.7
<i>CD1</i>	Camden - Swiss Cottage	11.1
<i>CT2</i>	City of London - Farringdon Street	13.9
<i>CT3</i>	City of London - Sir John Cass School	12.1
<i>CR8</i>	Croydon - Norbury Manor	10.1
<i>GR4</i>	Greenwich - Eltham	10.9
<i>GB0</i>	Greenwich - Falconwood FDMS	12.6
<i>GN6</i>	Greenwich - John Harrison Way	11.0
<i>GN3</i>	Greenwich - Plumstead High Street	13.4
<i>GR9</i>	Greenwich - Westhorne Avenue	10.5
<i>HV1</i>	Havering - Rainham	11.4
<i>LH0</i>	Hillingdon - Harlington	9.4
<i>KC1</i>	Kensington and Chelsea - North Ken	9.6
<i>HP1</i>	Lewisham - Honor Oak Park	9.9
<i>LW2</i>	Lewisham - New Cross	15.4
<i>TD5</i>	London Teddington Bushy Park	11.7
<i>ST5</i>	Sutton - Beddington Lane north	11.7
<i>TH4</i>	Tower Hamlets - Blackwall	12.6
<i>MY7</i>	Westminster - Marylebone Road FDMS	14.2
-	Annual mean for the Whole London	11.8

Table 1: Annual mean PM_{2.5} concentration for each site and the annual mean for London in 2019.

In terms of daily mean, London's PM_{2.5} concentration exceeded the WHO guideline of 15 µg/m³ on 74 out of 365 days, as illustrated in **Figure 2**. Most of these days were in winter (November to January) and spring (February to April), which revealed a fluctuation in the annual trend with more days of high concentrations and higher monthly means during November to April and fewer days and lower monthly means during May to October. As other studies have shown (Lei *et al.*, 2018; Kim, 2020; Liu *et al.*, 2014; Li *et al.*, 2015), there is a very strong seasonal difference in the PM_{2.5} level, and the presence of such difference could have an impact on the relationships between near-road greenspace and PM_{2.5} concentration.

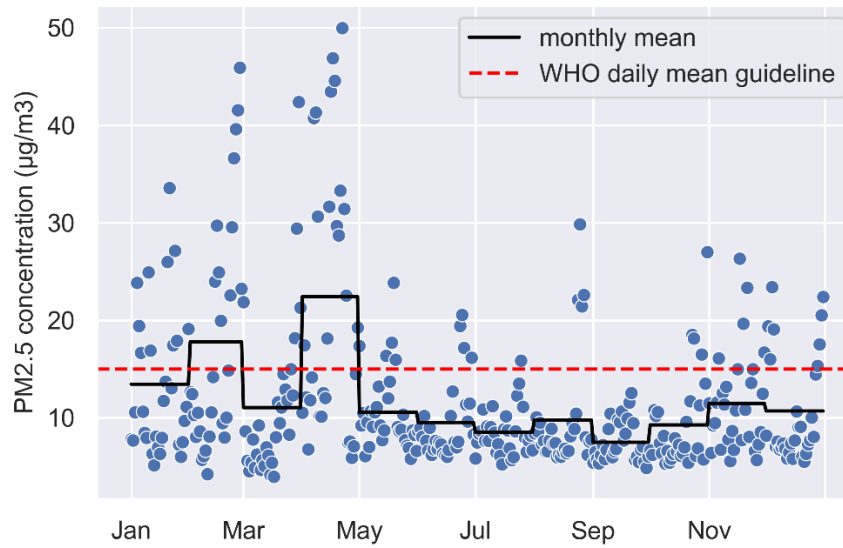


Figure 2: daily mean concentrations of PM_{2.5} in 2019 compared to WHO guideline and monthly mean.

Similarly, the fluctuation in the daily trends of London's PM_{2.5} level is also notable. As **Figure 3** shows, there are two peaks in PM_{2.5} concentration throughout the day - one between 7-9 am with a concentration of around 12.5 µg/m³, and the other around midnight with a concentration over 13 µg/m³. The lowest concentration is typically reached between 2-3 pm with an average concentration below 10 µg/m³. This daily pattern also adds uncertainties as well as possibilities to the effect of near-road greenspace on reducing PM_{2.5} pollution.

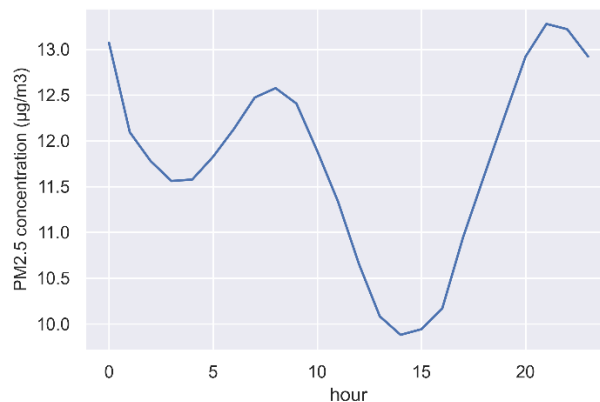


Figure 3: Diurnal change in the PM_{2.5} concentration in London

The summary statistics for the dependent variable are shown down in table 2 together with the explanatory variables.

Spatial geometry manipulation

The 1km-radius buffers around all sites were generated based on site locations using Python package geopandas. Then all greenspaces and different types of roads in each site buffer were found using package shapely. The classification of roads contains the five types (including motorways) as mentioned above, as well as a sixth class 'Other' which represents all roads that are not assigned a road classification at national or local level (labelled 'Not Classified') or do not have the classification information (labelled 'Unknown') (OS, 2017). **Figure 4** shows an example of the site buffer as well as the roads and green spaces that are in it.

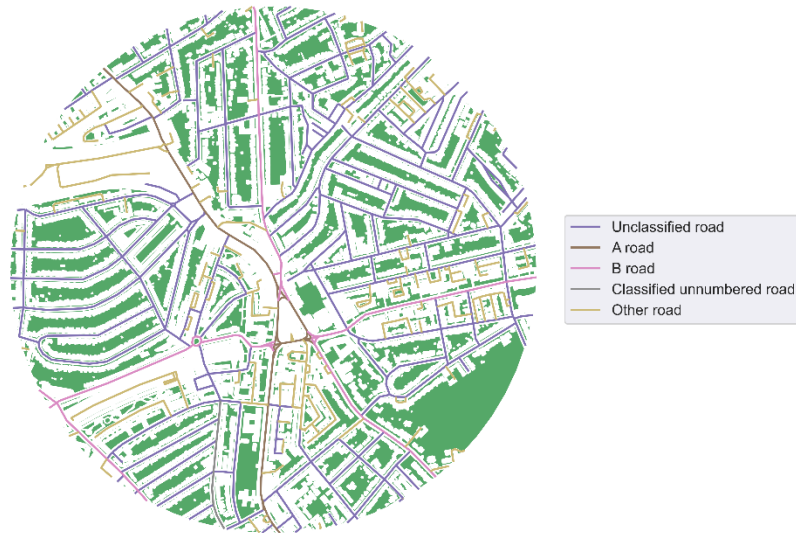


Figure 4: site buffer of CD1 (Camden - Swiss Cottage) as an example.

It is worth mentioning that there were two pairs of sites that were very close to each other (less than 1.5km). However, after performing a Student t-test (Kim, 2015) on their PM_{2.5} data, it was found that they were significantly different (p-value < 0.05), hence they were kept in the study.

A 50m buffer zone was then generated for each road and used to find all near-road greenspaces. Several studies (Kassomenos *et al.*, 2014; Eeftens *et al.*, 2012; Holguin *et al.*, 2007) on road traffic and PM_{2.5} pollution backed up the choice of a 50m buffer. These near-road greenspaces were also classified according to their nearby roads. For an area of green space that was close to more than one type of road, it was counted multiple times as near-road greenspace. This means that for each road class there is a set of marked near-road green spaces, and an area of green space can be marked as several different types of near-road greenspaces at the same time.

Generating explanatory variables

In order to investigate the effects of different types of near-road greenspace on reducing PM_{2.5} concentration, it is important to determine the proper variables to represent the near-road greenspace conditions in each site buffer. The simplest possible choice would be to use the percentage of near-road greenspace, which is the area of near-road greenspaces divided by the area of all green spaces in a site buffer. The problem with using the area percentage as an indication is that for places where there are only a small number of roads the percentage will be very small while for places with many roads the percentage will be very large. As a result, instead of being an indication of near-road greenspace condition, it is actually an indication of the number of roads.

One approach to mitigate the influence of the number of roads is to divide the area of near-road greenspaces by the total length of roads. In this way, the division result becomes near-road greenspace area per road length, which only reflects the conditions of near-road sections of the greenspace regardless of the number of roads in each site buffer.

Therefore, in this study, a greenspace-area-per-road-length number was calculated for each type of road as an indicator for the near-road greenspace conditions in each site buffer. This was accomplished based on the road and greenspace geometries prepared in the previous section. All site buffers with no

specific type of roads were assigned zero for their indicator for the specific road type.

Variable Name	Description	mean	std	min	25%	50%	75%	max
UnC_area_per_len	Unclassified road	28.52	14.02	6.46	13.89	32.96	38.75	47.13
A_area_per_len	A road	26.25	16.16	7.08	12.6	21.63	35.94	60.89
B_area_per_len	B road	25.79	21.17	0	5.63	27.49	44.13	56.16
CUn_area_per_len	Classified unnumbered road	34.01	22.54	0	17.72	33.96	50.73	64.51
Other_area_per_len	Other road	42.5	21.03	16.42	26.38	37.12	55.77	84.76
Value	PM _{2.5} concentration	11.78	10.66	0.1	5.5	8.6	14	592.8

Table 2: Summary statistics of the explanatory variables and the PM_{2.5} data.

The summary statistics for the explanatory variables are shown in **table 2**. It is obvious that there are large variations within the same indicator of near-road greenspaces, which suggests a great difference between the conditions of near-road greenspaces in different areas. Therefore it becomes more practically relevant to study the relationship between the near-road greenspace conditions and regional PM_{2.5} concentrations.

2.4 Multivariate linear regression models

A preliminary analysis of annual mean concentrations for all 21 sites was first performed to evaluate their overall relationship with the near-road greenspace. The global Moran's I of the annual means indicated that there was no obvious spatial auto-correlation in the dependent variable, so a non-spatial multivariate linear regression was performed.

The model performance was evaluated through a LOOCV (LeaveOneOut cross-validation) which is a type of cross-validation method that works well with small sample size data (Scikit-learn, 2013a). A typical cross-validation (k-fold) splits a dataset into k subsets and uses each subset once as the testing set to evaluate the performance of the model trained by all the other subsets (Scikit-learn, 2013b). The result of cross-validation is the average performance of the k models. When the sample size is small, the k-fold cross-validation result can have a large variance because how the samples are split will greatly alter the result. A LOOCV,

on the other hand, split the samples into training and testing sets N times, where N is the sample size, with only one sample as the testing set and all the other $N-1$ samples as the training set. Each time the testing sample is used to evaluate the performance of the model fitted with the training set, and the cross-validation result is the average performance of the N models. The advantage of LOOCV is that the estimation is deterministic, meaning that there is no variance in the estimated performance of the model because every sample is used once to evaluate the model, and the process can be repeated (Wei *et al.*, 2019). The downside of LOOCV is its high computational cost (Syed, 2011; Wei *et al.*, 2019), although for a small sample size it is neglectable.

The effect of each type of near-road classification was determined according to their corresponding feature importance and model coefficient. The feature importance was computed using the permutation feature importance technique from the Python package *sklearn*, which calculates the decrease in the model performance when the specific feature (independent variable) is shuffled (Breiman, 2001). A common method is to repeat the shuffle procedure several times (in this case 50 times), and calculate the mean and the standard deviation from all the repeated samples for each feature. The feature importance reflects how much a model depends on a feature, and in the case of this study, how much effect each type of near-road greenspace has on the regional $PM_{2.5}$ level.

After the initial investigation of the relationship between near-road greenspace and regional $PM_{2.5}$ level, the temporal changes in it were explored in further depth. This was accomplished using a series of multivariate linear regression models. The temporal changes were analysed along with two time series: 12 months throughout a year and 24 hours throughout a day. The hourly $PM_{2.5}$ concentrations were first used to generate monthly mean concentrations as well as average concentrations at each hour during the year. The two sets of concentrations at different time intervals were then analysed in groups separated by each unit time interval. In other words, 12 groups of monthly mean concentrations and 24 groups of hourly mean concentrations on an average day during the year were analysed independently. Each analysis included an

identification of spatial auto-correlation, a fit to a multivariate linear regression model, a LOOCV for model performance, permutation feature importance and a check for residual's normality. None of the variables was transformed or scaled, because the coefficients of the model were to be used as an indication of the effect of the near-road greenspace on PM_{2.5} concentration.

3. Results

3.1 Modelling annual mean PM_{2.5} concentration

The preliminary analysis of the annual mean PM_{2.5} concentrations found a global Moran's I of 0.096 using a Gaussian kernel weights matrix, and the multivariate linear regression model as a function of the near-road greenspace conditions had an r-squared value of 0.365 and a LOOCV r-squared of 0.081. The residuals of the model were normally distributed and not spatially auto-correlated. **Table 3** shows the coefficient and feature importance mean and error estimations for the explanatory variables. The indicator for greenspace near Unclassified road had the highest estimated feature importance, which was even higher than the r-squared value of the model. This means that the performance depends heavily on the variable, and shuffling it would alter the r-squared value to negative. The effect of near-Unclassified-road greenspace, therefore, was the strongest among the five types, and the estimated coefficient (-0.108) indicated that the higher the indicator (near-Unclassified-road greenspace area per road length), the lower the PM_{2.5} level. The B road and Classified unnumbered road indicators had median levels of estimated importance among the five, with similar coefficient estimations (0.038 for the former and 0.039 for the latter), which represented the positive effects of these two types of greenspaces on PM_{2.5} concentrations. The feature importance estimations of the rest two indicators (A road and Other road) were both comparatively low, but their estimated effects were opposite. A road indicator had a positive effect (0.030) while Other road indicator had a negative effect (-0.025).

Road type	Coefficient	Feature importance	Std of feature importance
Unclassified road	-0.108	1.636	0.552
A road	0.030	0.195	0.107
B road	0.038	0.490	0.228
Classified unnumbered road	0.039	0.614	0.244
Other road	-0.025	0.200	0.135
R-squared		0.365	
LOOCV R-squared		0.081	

Table 3: Multivariate linear regression model for the annual mean $PM_{2.5}$ concentration as a function of the indicators for greenspaces near five types of roads. The sample size of the model is 21.

However, the low LOOCV r-squared value of the model and the high variations (high standard deviation) in the feature importance estimations made the result less convincing. There might be some spatial-temporal changes in the relationship that were altering the annual mean model performance, and since the potential influence from spatial auto-correlation had already been excluded, it was necessary to check the temporal changes.

3.2 Modelling annual mean $PM_{2.5}$ concentrations for each hour

The 24 groups of annual mean $PM_{2.5}$ concentrations for each hour were separately tested for global Moran's I and none of them was found spatial auto-correlated. They were then used to fit a multivariate linear regression model as a function of their near-road greenspace indicators, and their residuals were normally distributed with no spatial auto-correlation spotted. Their performances are shown in **Figure 5**. The models of the annual means for hours between 0 am and 3 am and between 9 am and 12 pm had better performance

than the others, with an r -squared value higher than 0.4 and LOOCV r -squared value higher than 0.1. These hourly intervals were exactly the intervals at which the annual mean $PM_{2.5}$ concentrations were falling (Figure 3). The model with the highest performance was for 9 am where the r -squared (0.481) and the LOOCV r -squared value (0.227) were both the highest out of the 24 models. Conversely, the hourly intervals when the corresponding model performance was low coincided with the time periods when the concentrations were rising. The lowest-performance model was for 5 pm with an r -squared value of 0.203 and a LOOCV r -squared value of 0.005. The average r -squared value of the 24 models was 0.337 and the average LOOCV r -squared value was 0.065.

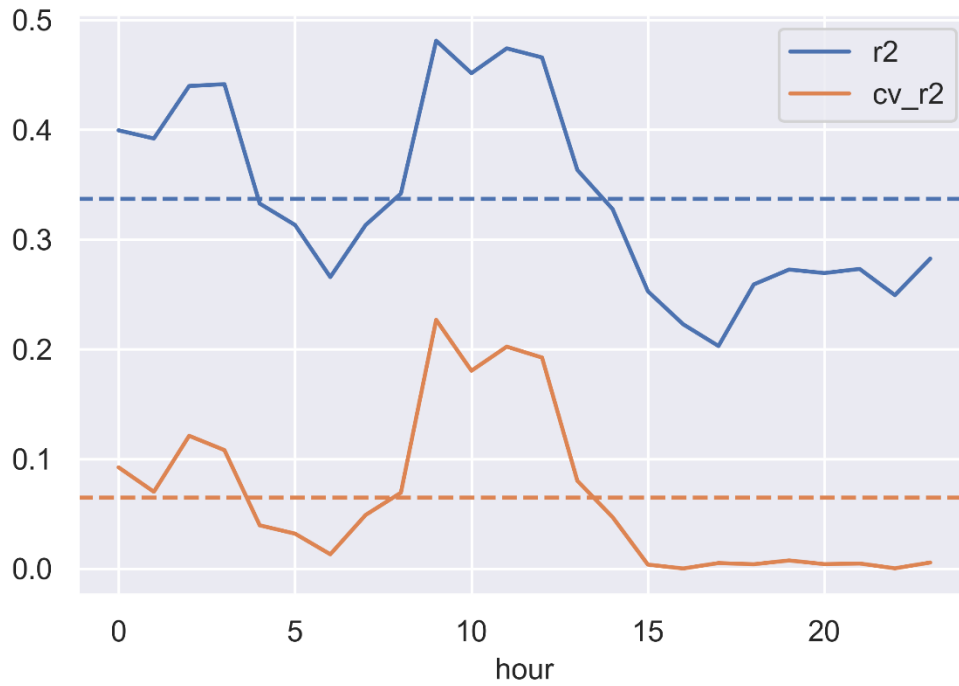


Figure 5: Performance of the 24 models in terms of r -squared value and LOO cross-validation r -squared value. The two dashed lines indicate the average r -squared and LOOCV r -squared values of the 24 models.

The feature importance estimations for models of all hourly intervals are presented in **Figure 6**. In almost all models the indicator for greenspaces near Unclassified roads had the highest feature importance, especially in the high-performance models where its importance was significantly higher than the other

four features. In contrast, A road and Other road greenspace indicators had the two lowest estimated feature importance (between 0.2 and 0.3) in most models except for 0 am-4 am when the estimation for A road was relatively higher and 5 am-8 am when that for Other road was around 0.5. The indicators for B road and Classified unnumbered road had comparatively stable feature importance with a value between 0.3 and 0.6 in most models. The variations in the estimations were still relatively large, with a higher average deviation (0.41) for the Unclassified road indicator and lower for the rest.

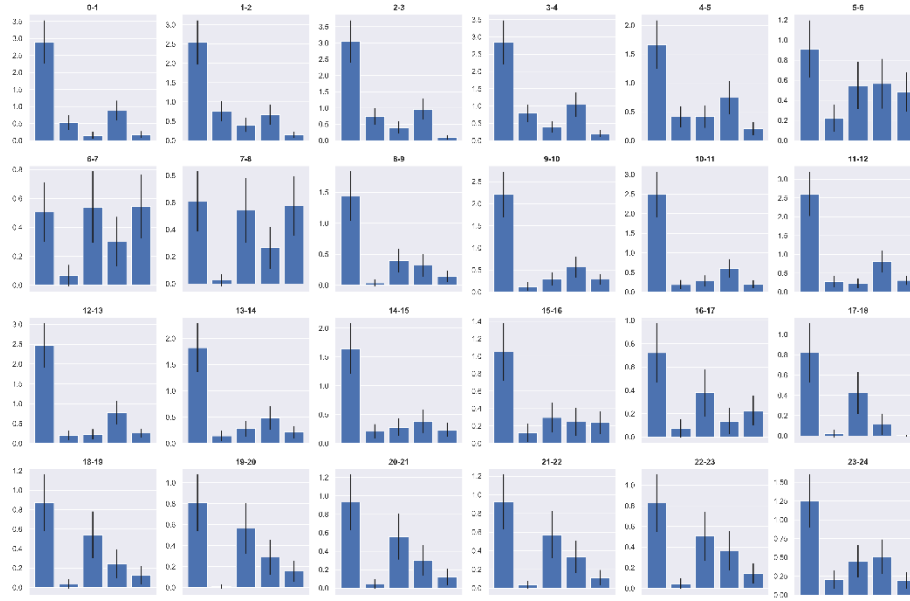


Figure 6: Feature importance estimations for the 24 models. The five bars from left to right represent: Unclassified road, A road, B road, Classified unnumbered road, Other road. The error bar is presented in black line, which is generated from the calculated standard deviation of each feature importance estimation.

The model coefficients were used to determine the effect of different indicators, and hence different types of near-road greenspace, on $PM_{2.5}$ levels. **Figure 7** shows the coefficients of all indicators for every model. In all 24 models, the effects of greenspaces near Unclassified and Other roads on $PM_{2.5}$ levels were negative while the other three types of greenspaces had positive effects. The changing patterns in the effect sizes of the five features as well as the intercept

were relatively similar, which was a declining-rising-declining-rising-declining trend.

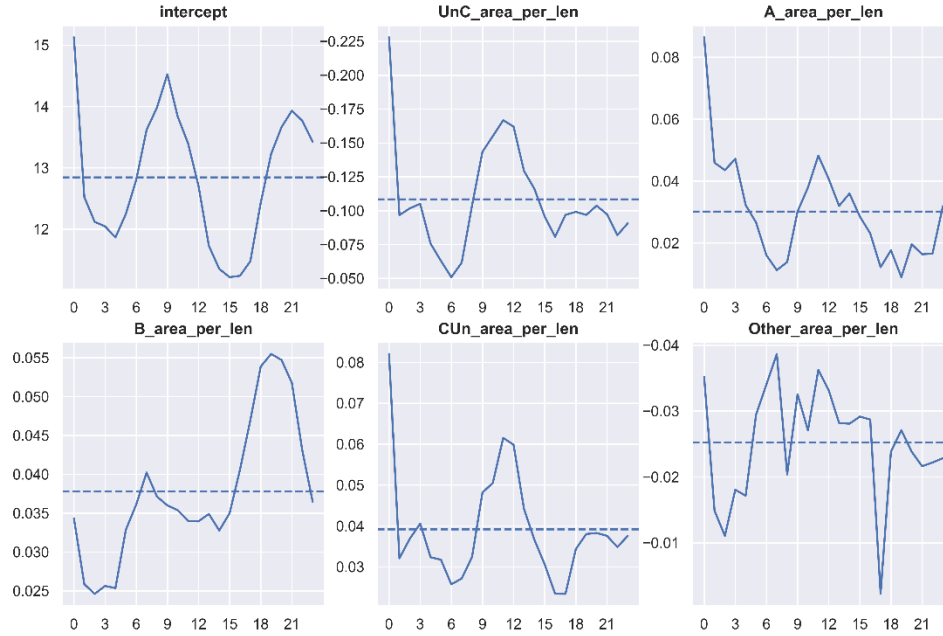


Figure 7: Intercepts and coefficients of all features for the 24 models. The dashed lines represent the average of the coefficients from the 24 models.

3.3 Modelling monthly mean PM_{2.5} concentrations

The same routine of analyses was performed on the 12 groups of monthly mean PM_{2.5} concentrations. Since no spatial auto-correlation was spotted in all groups of monthly means, multivariate linear regression models as a function of near-road greenspace conditions were fitted and tested. The residuals for all 12 models were approximately normally distributed, and no spatial auto-correlation was found in the residuals. The model evaluation results are shown in **Figure 8**. The models for the monthly means during the summer months (June to August) overall performed better than the others. The June model had the highest r-squared value (0.545) as well as the highest LOOCV r-squared value (0.280). The second-highest performing model was the August one with an r-squared value of 0.499 and a LOOCV r-squared value of 0.189. On the other hand, winter and

spring month models had a lower performance. The model for April had the lowest r-squared value of 0.079 with a LOOCV r-squared value of 0.102. The high- and low-performance seasons coincided with the high and low seasons of $PM_{2.5}$ as shown in Figure 2. The r-squared mean for the 12 months was 0.323 and the LOOCV r-squared mean was 0.079.

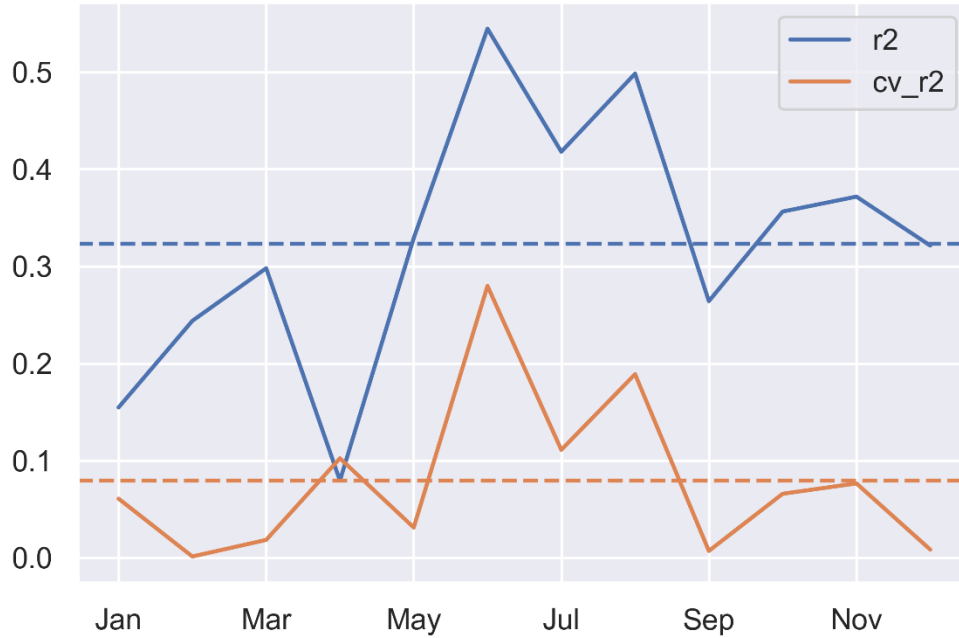


Figure 8: Performance of the 12 models. The dashed lines indicate the average performance of the 12 models.

As for the feature importance estimation, The Unclassified road indicator had the highest average importance, as shown in **Figure 9**. It was significantly higher than other features from February to July, but was lower during September to November, and almost to the same level as B road in August and during December to January. A road had low feature importance estimations in most months, except for June when it was estimated to be the second most important feature with an estimation of 1.932. Similarly, Other road had relatively low estimations from January to July and November, with high estimations from August to October and December. B road and Classified unnumbered road both had low feature importance estimations between February and June, while

estimations for July, September and December were median. In the models for August, October and November, estimations for B road were relatively high while estimations for Classified unnumbered road were low.

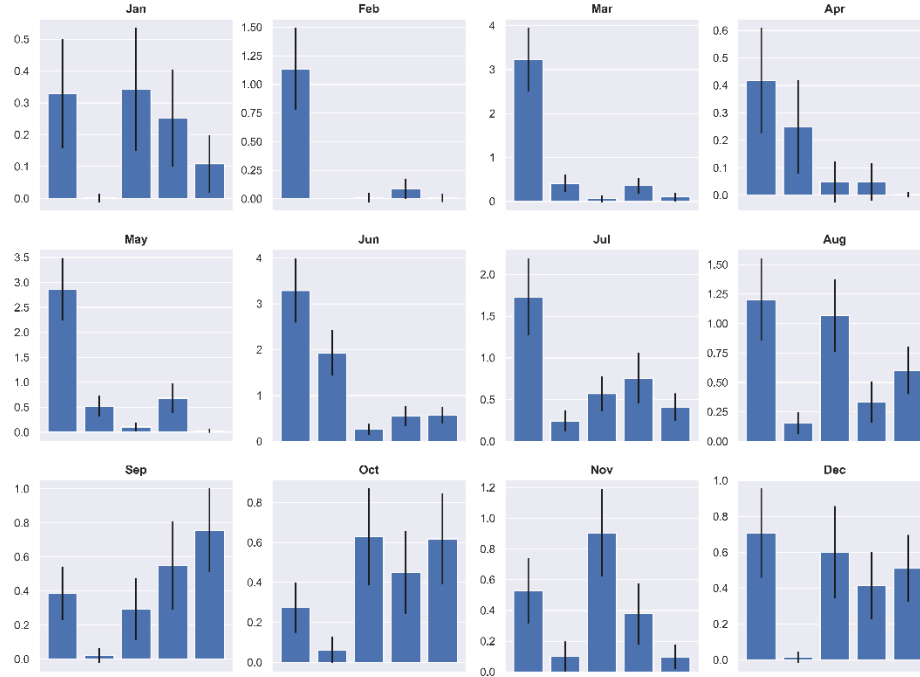


Figure 9: Feature importance estimations for the 12 models. The five bars from left to right represent: Unclassified road, A road, B road, Classified unnumbered road, Other road. The error bar is presented in black line, which is generated from the calculated standard deviation of each feature importance estimation.

The variations in the estimated feature importance were overall lower than that in the 24 groups of annual means for each hour. The average standard deviation of the estimations for Unclassified road indicator was the highest (0.361), with the other four indicators having about half the average standard deviation.

The patterns of the coefficients are presented in **Figure 10**. The changes in the intercepts of the 12 models followed a similar pattern to the monthly $PM_{2.5}$ concentrations as shown in Figure 2, with peaked values in February and April. The coefficients of the five features were much more chaotic. On one hand, the

effect of Unclassified road indicator remained negative for the entire year. Its size was higher from February to August and peaked (-0.205) in June, while lower during autumn (September to November) and early winter (December and January). On the other hand, the effects of B road and Classified unnumbered road indicators remained positive, with the effect size of the former peaking in August and the latter from May to July. The rest two features, being the indicators for A road and Other road, had relatively higher variations in their model coefficients. The effect of A road was positive in most months, with the size of the effect reaching its maximum in June at 0.137 . However, its effect in October and November was negative, although the effect size was not large. Conversely, the effect of Other road was negative except for in February and March, with the highest negative effect size in June at -0.054 and the highest positive effect size in March at 0.020 . Overall, the average effect size of the Unclassified road was the highest (-0.109) and that of the Other road was the lowest (-0.025).

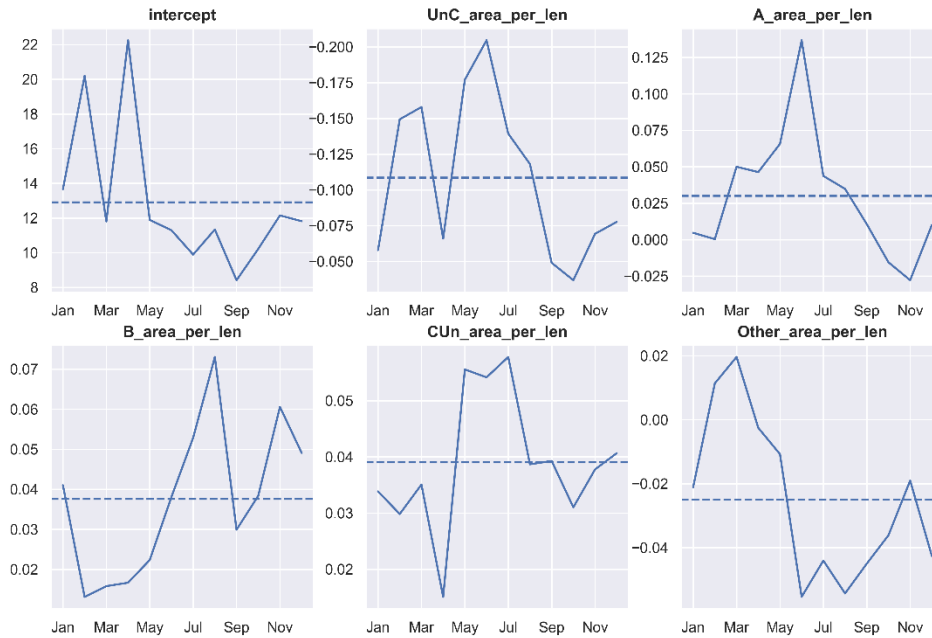


Figure 10: Intercepts and coefficients of all features for the 12 models. The dashed lines represent the average of the coefficients from the 12 models.

4. Discussion

4.1 The indicators of the greenspaces near different types of roads

The explanatory variables of the models used in this study were the total greenspace area near each type of road divided by the total length of that type of road (for each site buffer). This is an indication of the condition of the greenspaces for the specific type of road, and a higher value means a higher coverage of near-road greenspaces. However, such an indicator does not represent any information about the area of near-road greenspaces, hence a higher value of the indicator does not mean more near-road greenspaces for a specific type of road in a site buffer. Nor does it represent a relationship between the near-road greenspaces between two roads; for example, a higher indicator of near-road greenspace for one road type than that for the other type within the same buffer can only represent a higher greenspace coverage around the former road than the latter within that buffer, and not any information beyond that.

As a result, the detected relationships between those explanatory variables and regional $PM_{2.5}$ concentrations represent how the greenspace coverage near a specific type of road affects the local $PM_{2.5}$ levels (i.e. higher coverage for a specific road type will result in a higher or lower $PM_{2.5}$ level). Since the levels of coverage of the different near-road greenspaces can have different effects on the local $PM_{2.5}$ concentrations, this can also be referred to as the effect of the near-road greenspace for a specific road type on regional $PM_{2.5}$ levels.

4.2 Overall effects of near-road greenspaces in changing regional $PM_{2.5}$ levels

The results from the model for the annual mean $PM_{2.5}$ concentration highlight that there is a strong negative relationship between the indicator for greenspaces near Unclassified roads and the regional $PM_{2.5}$ levels, with the size of the effect being the highest among the five types of roads. Its feature importance is also

estimated to be the highest, indicating the highest impact on the PM_{2.5} levels. On the other hand, the results suggest a positive effect of the coverage of greenspaces near B roads and Classified unnumbered roads on regional PM_{2.5} levels, which means higher coverage of near-road greenspaces is related to a higher level of PM_{2.5}. This indicates that instead of reducing PM_{2.5} pollution, these green spaces are actually intensifying it. The moderate feature importance estimations for the two indicators imply that the near-road greenspaces for the two types of road have a median impact on regional PM_{2.5} levels. As for A road, the predicted effect is also positive, and the estimated feature importance suggests that the impact of greenspaces near A roads is even lower than that of B road and Classified unnumbered road. The coverage of greenspaces near Other roads is estimated to have a weak negative effect on regional PM_{2.5} levels with low estimated feature importance.

The higher impact (feature importance) of the greenspaces near Unclassified roads is likely due to the difference in the average traffic volume of different types of roads. The Unclassified roads generally have a lower traffic volume than A roads, B roads and Classified unnumbered roads (Roads.org.uk, 2017; GOV.UK, 2012). Therefore, the greenspaces near Unclassified roads are normally exposed to lower PM_{2.5} particle intensities than those near the other types of roads. This typically results in a better function of the leaf in storing PM_{2.5} particles over a longer period of time. When air containing particulate matter passes through a green space, part of the particulate matter will be removed from the air by the leaf surface and the wax layer on it (Kończak *et al.*, 2021). If the air contains a large amount of particulate matter, the leaf will soon reach its maximum in storing them (Liu *et al.*, 2013) and maintains a low PM removal efficiency before its recovery through the wind (Schaubroeck *et al.*, 2014) and rainfall wash-off (Weerakkody *et al.*, 2018; Xu *et al.*, 2020; Schaubroeck *et al.*, 2014; Xu *et al.*, 2017). Conversely, if the air only contains a small amount of PM, the leaf can capture particles and remove them from its surface at the same time. In this case, the green space will have an overall higher efficiency in capturing PM. The coverage of greenspaces near Unclassified roads that are exposed to a lower

amount of PM_{2.5} hence have a higher effect on the regional PM_{2.5} levels. Similarly, B road and Classified unnumbered road have lower traffic volumes than A road so they have higher impacts on regional PM_{2.5} levels.

In terms of the division between the positive and negative effects of coverages of greenspaces near different types of roads on the PM_{2.5} concentrations, the determinants are more complicated. On one hand, greenspaces reduce regional PM_{2.5} pollution by removing particles from the air (Beckett *et al.*, 2000; Nowak *et al.*, 2006; Kończak *et al.*, 2021) as well as blocking its transmission (Hofman *et al.*, 2016; Morakinyo and Lam, 2016). On the other hand, the block of transmission also results in a higher regional concentration (Morakinyo and Lam, 2016). As Vos *et al.* (2013) suggested, at least locally, greenspaces near roads have negative effects on reducing pollutants, because the existence of the vegetation reduces the ventilation, and therefore the pollutants accumulate in the area. This aerodynamic especially adds to pollution in street canyons where there are built-ups on both sides of the road. The trees slow the wind speed in a street canyon and reduce the exchange between the air within the canyon and above the roof, which results in an accumulation of pollution (Abhujith *et al.*, 2017; Jeanjean *et al.*, 2017). As a result, the effect of green space on regional PM_{2.5} levels depends on the combination of the removal and the increase of the particulate matter. The results from the model suggest that the overall removal effect of PM exceeds the intensification effect for greenspaces near Unclassified roads and Other roads, and the intensification effect surpasses the removal effect for the other three types of roads. The specific reasons behind such a pattern need some further investigation. One proposed conjecture is that this is due to morphological differences. A roads, B roads and Classified unnumbered roads are those higher-class roads that form the main network, and are therefore connected to more nearby built-ups. Hence the street canyon pattern is more common on those roads than on Unclassified roads and Other roads, so the near-road greenspaces on those roads tend to increase the regional PM_{2.5} concentration.

4.3 Temporal changes in the effect of near-road greenspaces

The results from the two groups of models that examined the temporal changes in the relationships between the coverage of near-road greenspaces for different types of roads and regional PM_{2.5} levels highlight the significant role of greenspaces near Unclassified roads in reducing PM_{2.5} pollution. In most scenarios greenspaces near Unclassified roads have the highest influence on regional PM_{2.5} levels. This is especially true from February to July on a monthly basis, and before 5 am and after 8 am on a daily basis, when its estimated impact is significantly higher than the greenspaces near the other four types of roads.

The estimated effect of greenspaces near Unclassified roads is negative in all periods. In terms of monthly variations, the relatively high estimated effect from May to July aligns with results from other studies (Wang *et al.*, 2015; Jeanjean *et al.*, 2017; Xu *et al.*, 2017; Steffens *et al.*, 2012) that found an overall higher reducing effect of vegetation on PM levels during summer. This is partly due to the fact that the leaf area density is higher in summer (Jeanjean *et al.*, 2017; Steffens *et al.*, 2012) so more PMs are deposited on the leaf surface. Other factors include a higher frequency of rainfall (Xu *et al.*, 2017; Wang *et al.*, 2015), higher wind speed and more vertical air exchange due to higher solar radiation (Wang *et al.*, 2015). In contrast, the lowest estimated effect is from September to January, which also coincides with a series of other studies (Przybysz *et al.*, 2018; Przybysz *et al.*, 2014; He *et al.*, 2020) that observed a lower reducing effect of vegetation on PM pollution during winter mainly due to lower wind speed, cooler temperature and less precipitation.

As for the diurnal variations, the highest effect of greenspaces near Unclassified roads on PM_{2.5} concentration was observed at midnight and a relatively high effect during 9-12 am, and the lowest effect was during 5-7 am. There are very few studies on the diurnal variations in the PM removal effect of vegetation (Deng *et al.*, 2019; Brantley *et al.*, 2014), and neither of them found a significant change in the effect throughout an average day. The possible cause of such detected

variation may be the change in temperature and solar radiation, and the resulting wind speed change. When the sun rises, the increasing temperature creates more convections, which generate stronger wind that washes off the leaf surface. However, such speculation can not explain the peak in the effect at midnight, and therefore the diurnal variation may need further investigations.

The effect of near-road greenspaces for Other roads is similar to that for Unclassified roads, except for in February and March when it is positive, and overall much lower influence on $PM_{2.5}$ levels. Nevertheless, the effect of greenspaces near A roads, B roads and Classified unnumbered roads is, as estimated for the overall effect, positive in virtually all periods. This suggests that the intensification effect of these green spaces on PM pollution surpasses the removal effect all the time. The positive effect of all three peaks in summer and troughs in winter or spring, with the effect of A road greenspaces, shortly being negative during October to November.

The peaked positive effect in summer is likely due to the higher temperature and the resulting low pressure around the road that prevents airborne PM from leaving by creating inward airflow (Al-Dabbous and Kumar, 2014). The presence of near-road greenspaces strengthens such an effect by lowering air circulation at the surface level whose consequence is much more impactful than the removing capacity of the vegetation leaves, especially in street canyons (Jeanjean *et al.*, 2017). Similarly, the low positive (and even negative) effect in winter and spring can be explained by the high-pressure climate that promotes the outward flow of near-surface air in street canyons, and therefore reduce the intensification effect of green spaces on regional $PM_{2.5}$ levels (Al-Dabbous and Kumar, 2014). This theory can also be implied to the diurnal variations where the high effect occurs during 9 am-2 pm and the low effect during nighttime. The exceptions are during 6-9 pm for B road and at midnight for A road and Classified unnumbered road when the effect is very high. The former can be explained by higher traffic flows during the period, but the cause of the latter still needs some further investigation.

5. Conclusion

This paper examined the effect of near-road greenspaces for different types of roads on regional $PM_{2.5}$ levels and the temporal changes in the effect through modelling the relationships between coverages of near-road greenspaces and local $PM_{2.5}$ concentrations. This effect is determined by the combination of removal ability and intensification ability. Several conclusions were drawn from this study:

1. The greenspaces near Unclassified roads have the highest overall reducing effect on regional $PM_{2.5}$ levels, with a relatively high effect in summer and from 9-12 am on an average day. The potential pushes are most possibly the higher leaf area density and warmer temperatures.
2. The A road, B road and Classified unnumbered road greenspaces have an opposing effect to reduce $PM_{2.5}$ levels. The most likely reason is that street canyon morphology, which tends to create intensified PM pollution, is much more common around these roads. The positive effect generally peaks seasonally in summer and diurnally during 9-12 am.
3. The temporal change in the effect of Other road greenspaces is most varied. The overall effect on regional $PM_{2.5}$ levels is negative, but in spring the effect is positive.
4. Except for B road, greenspaces near all the other four types of roads have a high effect (either positive or negative) on $PM_{2.5}$ level at midnight.

The aforementioned points derived from this study should lead to some practical policy implications in urban greenspace planning. It proves that greenspace, especially near-road greenspace, is not guaranteed to reduce PM pollution in all circumstances. However, due to some constraints of this study, including the small sample size and failure in explaining some of the temporal changes, some further explorations should be made for a better understanding of the function of green spaces, and for more cost-effective implementations of urban greenspaces.

Reference List

- Abhijith, K. V., Kumar, P., Gallagher, J., McNabola, A., Baldauf, R., Pilla, F., Broderick, B., Sabatino, S. D. and Pulvirenti, B. (2017) Air pollution abatement performances of green infrastructure in open road and built-up street canyon environments—A review. *Atmospheric Environment*, 162, 71-86.
- Al-Dabbous, A. N. and Kumar, P. (2014) The influence of roadside vegetation barriers on airborne nanoparticles and pedestrians exposure under varying wind conditions. *Atmospheric Environment*, 90, 113-124.
- Anderson, J. O., Thundiyil, J. G. and Stolbach, A. (2012) Clearing the air: a review of the effects of particulate matter air pollution on human health. *Journal of medical toxicology*, 8(2), 166-175.
- Baldauf R., Jackson L., Hagler G., Isakov V., McPherson G., Cahill T.A., Zhang K.M., Cook J.R., Bailey C. and Wood P. (2011) The role of vegetation in mitigating air quality impacts from traffic emissions. *EM. Jan*: 1-3.
- Beckett, K. P., Freer-Smith, P. H. and Taylor, G. (2000) Particulate pollution capture by urban trees: effect of species and windspeed. *Global Change Biology*, 6(8), 995–1003.
- Brantley, H. L., Hagler, G. S. W., J. Deshmukh, P. and Baldauf, R. W. (2014) Field assessment of the effects of roadside vegetation on near-road black carbon and particulate matter. *Science of the Total Environment*, 468-469, 120–129.
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.
- Brook, R.D., Rajagopalan, S., Pope III, C.A., Brook, J.R., Bhatnagar, A., Diez-Roux, A.V., Holguin, F., Hong, Y., Luepker, R.V., Mittleman, M.A. and Peters, A. (2010) Particulate matter air pollution and cardiovascular disease: an update to the scientific statement from the American Heart Association. *Circulation*, 121(21), 2331-2378.

- Cai, L., Zhuang, M. and Ren, Y. (2020) Spatiotemporal characteristics of NO₂, PM_{2.5} and O₃ in a coastal region of southeastern China and their removal by green spaces. *International Journal of Environmental Health Research*, 1-17.
- Chen, X. and Xiao, Y. (2018) A Novel Method for Air Quality Data Imputation by Nuclear Norm Minimization. *Journal of Sensors*, 2018, 1–11.
- Chen, M., Dai, F., Yang, B. and Zhu, S. (2019) Effects of neighborhood green space on PM_{2.5} mitigation: Evidence from five megacities in China. *Building and Environment*, 156, 33–45.
- Defra. (2022) Automatic Urban and Rural Network (AURN). [Online]. Available from: <https://uk-air.defra.gov.uk/networks/network-info?view=aurn> [Accessed 06 Feb 2022]
- Deng, S., Ma, J., Zhang, L., Jia, Z. and Ma, L. (2019) Microclimate simulation and model optimization of the effect of roadway green space on atmospheric particulate matter. *Environmental Pollution*, 246, 932–944.
- Digimap. (2021) Ordnance Survey. [Online]. Available from: <https://digimap.edina.ac.uk/os> [Accessed 18 Dec 2021]
- Eeftens, M., Beelen, R., De Hoogh, K., Bellander, T., Cesaroni, G., Cirach, M., Declercq, C., Dedele, A., Dons, E., De Nazelle, A. and Dimakopoulou, K. (2012) Development of land use regression models for PM_{2.5}, PM_{2.5} absorbance, PM₁₀ and PM_{coarse} in 20 European study areas; results of the ESCAPE project. *Environmental science & technology*, 46(20), 11195–11205.
- González-Flecha, B. (2004) Oxidant mechanisms in response to ambient air particles. *Molecular aspects of medicine*, 25(1-2), 169-182.
- GOV.UK. (2012) Guidance on road classification and the primary route network. [Online]. Available from: <https://www.gov.uk/government/publications/guidance-on-road-classification-and-the-primary-route-network/guidance-on-road-classification-and-the-primary-route-network> [Accessed 08 Feb 2022]

- Greenspace Information for Greater London (GiGL). (2018) Mapping London's Green Belt and Metropolitan Open Land. [Online]. Available from: <https://www.gigl.org.uk/mapping-londons-green-belt-and-mol/> [Accessed 14 Feb 2022]
- He, C., Qiu, K., Alahmad, A. and Pott, R. (2020) Particulate matter capturing capacity of roadside evergreen vegetation during the winter season. *Urban Forestry & Urban Greening*, 48, 126510.
- Hofman, J., Bartholomeus, H., Janssen, S., Calders, K., Wuyts, K., Van Wittenberghe, S. and Samson, R. (2016) Influence of tree crown characteristics on the local PM₁₀ distribution inside an urban street canyon in Antwerp (Belgium): A model and experimental approach. *Urban forestry & urban greening*, 20, 265-276.
- Holguin, F., Flores, S., Ross, Z., Cortez, M., Molina, M., Molina, L., Rincon, C., Jerrett, M., Berhane, K., Granados, A. and Romieu, I. (2007) Traffic-related Exposures, Airway Function, Inflammation, and Respiratory Symptoms in Children. *American Journal of Respiratory and Critical Care Medicine*, 176(12), 1236–1242.
- Irga, P. J., Burchett, M. D. and Torpy, F. R. (2015) Does urban forestry have a quantitative effect on ambient air quality in an urban environment?. *Atmospheric Environment*, 120, 173-181.
- Jeanjean, A. P. R., Buccolieri, R., Eddy, J., Monks, P. S. and Leigh, R. J. (2017) Air quality affected by trees in real street canyons: The case of Marylebone neighbourhood in central London. *Urban Forestry & Urban Greening*, 22, 41–53.
- Ji, W. and Zhao, B. (2014) Numerical study of the effects of trees on outdoor particle concentration distributions. In *Building Simulation*. 7(4), 417-427. Tsinghua University Press.
- Karagulian, F., Belis, C. A., Dora, C. F. C., Prüss-Ustün, A. M., Bonjour, S., Adair-Rohani, H. and Amann, M. (2015) Contributions to cities' ambient particulate matter (PM): A systematic review of local source contributions at global level. *Atmospheric environment*, 120, 475-483.

- Kassomenos, P. A., Vardoulakis, S., Chaloulakou, A., Paschalidou, A. K., Grivas, G., Borge, R. and Lumbreras, J. (2014) Study of PM₁₀ and PM_{2.5} levels in three European cities: Analysis of intra and inter urban variations. *Atmospheric Environment*, 87, 153–163.
- Kim, T.K. (2015) T test as a parametric statistic. *Korean Journal of Anesthesiology*, 68, 540 - 546.
- Kim, H. (2020) Land use impacts on particulate matter levels in Seoul, South Korea: comparing high and low seasons. *Land*, 9(5), 142.
- Kim, K.-H., Kabir, E. and Kabir, S. (2015) A review on the human health impact of airborne particulate matter. *Environment International*, 74, 136–143.
- Kończak, B., Cempa, M. and Deska, M. (2021) Assessment of the ability of roadside vegetation to remove particulate matter from the urban air. *Environmental Pollution*, 268, 115465.
- Lei, Y., Duan, Y., He, D., Zhang, X., Chen, L., Li, Y., Gao, Y.G., Tian, G. and Zheng, J. (2018) Effects of urban greenspace patterns on particulate matter pollution in metropolitan Zhengzhou in Henan, China. *Atmosphere*, 9(5), 199.
- Li, R., Li, Z., Gao, W., Ding, W., Xu, Q. and Song, X. (2015) Diurnal, seasonal, and spatial variation of PM 2.5 in Beijing. *Science Bulletin*, 60(3), 387–395.
- Liu, Z., Hu, B., Wang, L., Wu, F., Gao, W. and Wang, Y. (2014) Seasonal and diurnal variation in particulate matter (PM₁₀ and PM_{2.5}) at an urban site of Beijing: analyses from a 9-year study. *Environmental Science and Pollution Research*, 22(1), 627–642.
- Liu, X., Yu, X. and Zhang, Z. (2015) PM_{2.5} concentration differences between various forest types and its correlation with forest structure. *Atmosphere*, 6(11), 1801-1815.
- Liu, L., Guan, D., Peart, M. R., Wang, G., Zhang, H. and Li, Z. (2013) The dust retention capacities of urban vegetation—a case study of Guangzhou, South China. *Environmental Science and Pollution Research*, 20(9), 6601-6610.

- Löndahl, J., Pagels, J., Swietlicki, E., Zhou, J., Ketzel, M., Massling, A. and Bohgard, M. (2006) A set-up for field studies of respiratory tract deposition of fine and ultrafine particles in humans. *Journal of Aerosol Science*, 37(9), 1152-1163.
- London Air. (2022) About Londonair. [Online]. Available from: <https://www.londonair.org.uk/LondonAir/General/about.aspx> [Accessed 06 Feb 2022]
- London Air. (2022b) Data Downloads. [Data]. Available from: <https://www.londonair.org.uk/london/asp/datadownload.asp> [Accessed 03 Jan 2022]
- London Datastore. (2019) Air Quality Monitoring Sites. [Data]. Available from: https://data.london.gov.uk/download/air_quality_monitoring_sites/b22020b1-c42a-4653-afb0-7f52fd658635/Air_Quality_Monitoring_Sites.gpkg [Accessed 12 Dec 2021]
- Mayor of London. (2019) PM_{2.5} in London: Roadmap to meeting WHO guidelines by 2030. [Report]. Available from: https://www.london.gov.uk/sites/default/files/pm2.5_in_london_october19.pdf [Accessed 06 Feb 2022]
- Morakinyo, T. E. and Lam, Y. F. (2016) Simulation study of dispersion and removal of particulate matter from traffic by road-side vegetation barrier. *Environmental Science and Pollution Research*, 23(7), 6709-6722.
- Norazian, M. N., Shukri, Y. A., Azam, R. N. and Al Bakri, A. M. M. (2008) Estimation of missing values in air pollution data using single imputation techniques. *ScienceAsia*, 34(3), 341-345.
- Nowak, D. J., Crane, D. E. and Stevens, J. C. (2006) Air pollution removal by urban trees and shrubs in the United States. *Urban forestry & urban greening*, 4(3-4), 115-123.
- Ordnance Survey (OS). (2017) OS Open Roads User guide and technical specification. [Report]. Available from:

- <https://www.ordnancesurvey.co.uk/documents/os-open-roads-user-guide.pdf> [Accessed 02 Feb 2022]
- Ordnance Survey (OS). (2021a) OS MasterMap Greenspace - Layer. [Online]. Available from: <https://www.ordnancesurvey.co.uk/business-government/products/mastermap-greenspace> [Accessed 17 Dec 2021]
 - Ordnance Survey (OS). (2021b) OS Open Roads. [Online]. Available from: <https://www.ordnancesurvey.co.uk/business-government/products/open-map-roads> [Accessed 17 Dec 2021]
 - Przybysz, A., Sæbø, A., Hanslin, H. M. and Gawroński, S. W. (2014) Accumulation of particulate matter and trace elements on vegetation as affected by pollution level, rainfall and the passage of time. *Science of the Total Environment*, 481, 360–369.
 - Przybysz, A., Nersisyan, G. and Gawroński, S. W. (2018) Removal of particulate matter and trace elements from ambient air by urban greenery in the winter season. *Environmental Science and Pollution Research*, 26(1), 473–482.
 - [Roads.org.uk](https://www.roads.org.uk). (2017). Road Numbers: Numbers for A and B-roads. [Online]. Available from: <https://www.roads.org.uk/articles/road-numbers/numbers-and-b-roads> [Accessed 12 Feb 2022]
 - Ross, Z., Jerrett, M., Ito, K., Tempalski, B. and Thurston, G. (2007) A land use regression for predicting fine particulate matter concentrations in the New York City region. *Atmospheric Environment*, 41(11), 2255–2269.
 - Schaubroeck, T., Deckmyn, G., Neiryneck, J., Staelens, J., Adriaenssens, S., Dewulf, J., Muys, B. and Verheyen, K., (2014) Multilayered modeling of particulate matter removal by a growing forest over time, from plant surface deposition to washoff via rainfall. *Environmental science & technology*, 48(18), 10785-10794.
 - Scikit-learn. (2013a) 3.1.2.1.3. Leave One Out (LOO). [Online]. Available from: https://scikit-learn.org/stable/modules/cross_validation.html#leave-one-out-loo [Accessed 09 Feb 2022]

- Scikit-learn. (2013b) 3.1.2.1.1. K-Fold. [Online]. Available from: https://scikit-learn.org/stable/modules/cross_validation.html#k-fold [Accessed 09 Feb 2022]
- Shah, A. S., Langrish, J. P., Nair, H., McAllister, D. A., Hunter, A. L., Donaldson, K., Newby, D. E. and Mills, N. L. (2013) Global association of air pollution and heart failure: a systematic review and meta-analysis. *The Lancet*, 382(9897), 1039–1048.
- Song, Y., Maher, B. A., Li, F., Wang, X., Sun, X. and Zhang, H. (2015) Particulate matter deposited on leaf of five evergreen species in Beijing, China: Source identification and size distribution. *Atmospheric environment*, 105, 53-60.
- Steffens, J. T., Wang, Y. J. and Zhang, K. M. (2012) Exploration of effects of a vegetation barrier on particle size distributions in a near-road environment. *Atmospheric Environment*, 50, 120–128.
- Sun, Q., Hong, X. and Wold, L. E. (2010) Cardiovascular effects of ambient particulate air pollution exposure. *Circulation*, 121(25), 2755-2765.
- Srimuruganandam, B. and Nagendra, S. S. (2012) Source characterization of PM₁₀ and PM_{2.5} mass using a chemical mass balance model at urban roadside. *Science of the total environment*, 433, 8-19.
- Syed, A. R. (2011) A review of cross validation and adaptive model selection.
- The Independent. (2014) 47 per cent of London is green space: Is it time for our capital to become a national park? [Online] Available from: <https://www.independent.co.uk/climate-change/news/47-per-cent-of-london-is-green-space-is-it-time-for-our-capital-to-become-a-national-park-9756470.html> [Accessed 26 Feb 2022]
- Vos, P. E., Maiheu, B., Vankerkom, J. and Janssen, S. (2013) Improving local air quality in cities: to tree or not to tree?. *Environmental pollution*, 183, 113-122.
- Wang, Q. and Li, S. (2021) Nonlinear impact of COVID-19 on pollutions – Evidence from Wuhan, New York, Milan, Madrid, Bandra, London, Tokyo and Mexico City. *Sustainable Cities and Society*, 65, 102629.

- Wang, H., Shi, H. and Wang, Y. (2015) Effects of Weather, Time, and Pollution Level on the Amount of Particulate Matter Deposited on Leaves of *Ligustrum lucidum*. *The Scientific World Journal*, 2015, 1–8.
- Wei, J., Chu, X., Sun, X., Xu, K., Deng, H., Chen, J., Wei, Z. and Lei, M. (2019) Machine learning in materials science. *InfoMat*, 1(3), 338–358.
- Weerakkody, U., Dover, J. W., Mitchell, P. and Reiling, K. (2018) The impact of rainfall in remobilising particulate matter accumulated on leaves of four evergreen species grown on a green screen and a living wall. *Urban Forestry & Urban Greening*, 35, 21–31.
- World Bank. (2017) PM_{2.5} air pollution, mean annual exposure (micrograms per cubic meter). [Online]. Available from: <https://data.worldbank.org/indicator/EN.ATM.PM25.MC.M3?end=2017&start=2017&view=map> [Accessed 03 Feb 2022]
- World Health Organization (WHO). (2021) Ambient (outdoor) air pollution. [Online]. Available from: [https://www.who.int/news-room/fact-sheets/detail/ambient-\(outdoor\)-air-quality-and-health](https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health) [Accessed 03 Feb 2022].
- Wu, C. D., Chen, Y. C., Pan, W. C., Zeng, Y. T., Chen, M. J., Guo, Y. L. and Lung, S. C. C. (2017) Land-use regression with long-term satellite-based greenness index and culture-specific sources to model PM_{2.5} spatial-temporal variability. *Environmental Pollution*, 224, 148–157.
- Xu, X., Zhang, Z., Bao, L., Mo, L., Yu, X., Fan, D. and Lun, X. (2017) Influence of rainfall duration and intensity on particulate matter removal from plant leaves. *Science of the Total Environment*, 609, 11–16.
- Xu, M., Sbihi, H., Pan, X. and Brauer, M. (2019) Local variation of PM_{2.5} and NO₂ concentrations within metropolitan Beijing. *Atmospheric Environment*, 200, 254–263.
- Xu, X., Xia, J., Gao, Y. and Zheng, W. (2020) Additional focus on particulate matter wash-off events from leaves is required: A review of studies of urban plants used to reduce airborne particulate matter pollution. *Urban Forestry & Urban Greening*, 48, 126559.

Appendix