Road transport and emissions modelling in England and Wales: A machine learning modelling approach using spatial data

Alexandros Sfyridis

Bartlett School of Energy, Environment and Resources

Institute for Sustainable Resources

University College London

Supervisors:

Prof. Paolo Agnolucci

Prof. Paul Ekins

A thesis submitted for the degree of

Doctor of Philosophy

Declaration

"I, Alexandros Sfyridis confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis."

Acknowledgements

I would first like to thank my supervisors, Prof Paolo Agnolucci and Prof Paul Ekins, for their continued support, inspiration, guidance and supervision during my studies. Paolo's active engagement and experience have been significant factors to improve my academic thinking and writing and while being critical it helped me to get the most out of my PhD. Paul has always been a source of inspiration and through our interaction I have been motivated to continue my endeavour in academia. I would also like to acknowledge the financial support from the Natural Environmental Research Council (NE/M019799/1) and the UK Energy Research Centre (Grant Number: EP/L024756/1). Equally, I would like to thank my examiners, Prof Helena Titheridge and Dr Astley Hastings for taking the time to examine and discuss my work during the viva.

I would also like to thank the research staff and PhD students at the ISR for all the interesting discussions on various scientific topics that have been intellectually stimulating, and for all the fun times in and out of the office. Specifically, I would like to thank Vincenzo De Lipsis, Chrysanthi Rapti, Arkaitz Usubiaga, Simon Damkjaer, Tony Carr, Wan-Ting Hsu and Theodoros Arvanitopoulos. I will cherish the moments we shared throughout those years. Without them, the office would have been less pleasant.

I would like to thank Dr Zack Protogeros for the insightful conversations over technical issues on machine learning and data mining, showing me ways to enrich my knowledge and think out of the box. Moreover, I would like to thank Argyris Konstantinidis and Margarita Mitsika for their sincere emotional support and all those who supported me and passed away during these years.

Finally, I would like to express my sincere gratitude to my sister Maria, my father Yiorgos and my mother Tania for their tireless encouragement and for always being truly supportive on my decisions.

My PhD research was driven by curiosity and seek for knowledge. I hope that the end of this journey will be the beginning of another one full of knowledge and experiences.

Abstract

An expanding street network coupled with an increasing number of vehicles testifies to the significance and reliance on road transportation of modern economies. Unfortunately, the use of road transport comes with drawbacks such as its contribution to greenhouse gases (GHG) and air pollutant emissions, therefore becoming an obstacle to countries' objectives to improve air quality and a barrier to the ambitious targets to reduce Greenhouse Gas emissions.

Unsurprisingly, traffic forecasting, its environmental impacts and potential future configurations of road transport are some of the topics which have received a great deal of attention in the literature. However, traffic forecasting and the assessment of its determinants have been commonly restricted to specific, normally urban, areas while road transport emission studies do not take into account a large part of the road network, as they usually focus on major roads.

This research aimed to contribute to the field of road transportation, by firstly developing a model to accurately estimate traffic across England and Wales at a granular (i.e., street segment) level, secondly by identifying the role of factors associated with road transportation and finally, by estimating CO_2 and air pollutant emissions, known to be responsible for climate change as well as negative impacts on human health and ecosystems. The thesis identifies potential emissions abatement from the adoption of novel road vehicles technologies and policy measures. This is achieved by analysing transport scenarios to assess future impacts on air quality and CO_2 emissions. The thesis concludes with a comparison of my estimates for road emissions with those from DfT modelling to assess the methodological robustness of machine learning algorithms applied in this research.

The traffic modelling outputs reveal traffic patterns across urban and rural areas, while traffic estimation is achieved with high accuracy for all road classes. In addition, specific socioeconomic and roadway characteristics associated with traffic across all vehicle types and road classes are identified. Finally, CO_2 and air pollution hot spots as well as the impact of open spaces on pollutants emissions and air quality are explored. Potential emission reduction with the employment of new vehicle

technologies and policy implementation is also assessed, so as the results can support urban planning and inform policies related to transport congestion and environmental impacts mitigation. Considering the disaggregated approach, the methodology can be used to facilitate policy making for both local and national aggregated levels.

Impact Statement

The thesis contributes significantly both from an academic and policy point of view to the challenges faced by transport planners, environmental scientists, and decision makers, related to GHG and air pollution mitigation in line with the targets set by the UK government.

My contribution to the academic literature is reflected by the work presented in chapter 3, which develops a novel methodological approach to model traffic on a disaggregated level achieving a high estimation accuracy. The chapter has been published in the Journal of Transport Geography – one of the journals with the highest impact factor in transport studies, where it has already a number of citations. Moreover, the work presented in chapter 5, which presents a novel hybrid methodology to estimate emissions and assesses current and future policy implementations has been published in the special issue "Air Quality in the UK" in the Atmosphere Journal. The work presented in chapter 4, where the determinants of road traffic are assessed, is also under review at the Transport Research Record Journal, completing the publications of the thesis in academic journals. In addition, this thesis contributes to the academic discussion on road transport and associated emission modelling, highlighting the importance of utilizing multiple and diverse data as well as the employment of machine learning algorithms in transport studies. Finally, the thesis contributes to the academic discussion by identifying and raising important challenges in road transport studies that can be addressed in future academic research, to improve road transport modelling, better understand traffic determinants and facilitate the process of mitigating emissions from road transport.

From a policy perspective, the thesis has informed urban, transport and environmental planners, as well as policy makers on the effects specific policies and infrastructure can have on GHGs and air pollution, so that land use, infrastructure construction, and policies development can be improved and implemented. By applying my knowledge on transport modelling and applied machine learning, I have contributed to understanding the complexity of road transport systems and its interrelation with urban and rural infrastructure, socioeconomic indicators, and the environment.

Table of contents

List of Publications

Peer-reviewed publications based on this thesis:

Delafield, G., Donnison, C., Roddis, P., Arvanitopoulos, T., **Sfyridis, A.**, Dunnett, S., Ball, T., Logan, K.G., 2021. Conceptual framework for balancing society and nature in net-zero energy transitions. Environ. Sci. Policy 125, 189–201.

https://doi.org/10.1016/j.envsci.2021.08.021

Sfyridis A., Agnolucci, P., 2021. Road Emissions in London: Insights from Geographically Detailed Classification and Regression Modelling. Atmosphere (Basel). 12, 188.

https://doi.org/10.3390/atmos12020188

Sfyridis A., Agnolucci, P., 2020. Annual average daily traffic estimation in England and Wales: An application of clustering and regression modelling. J. Transp. Geogr. 83, 102658

https://doi.org/10.1016/j.jtrangeo.2020.102658

In progress:

"Factors affecting Road Traffic: Identifying drivers of Annual Average Daily Traffic (AADT) using LASSO regression", with Paolo Agnolucci – under review at Transportation Research Record

List of Abbreviations

AADT	Annual Average Daily Traffic
ABM	Activity Based Model
ALBATROSS	A Learning-based Transportation Oriented Simulation System
ANN	Artificial Neural Network
AON	All-or-nothing
ARTEMIS	Assessment and Reliability of Transport Emissions Models and Inventory Systems
ATC	Automatic Traffic Counters
BEIS	Department for Business, Energy and Industrial Strategy
BRUTAL	Background, Road and Urban Transport modelling of Air quality Limit
CAZ	Clean Air Zone
CCC	Committee on Climate Change
CEMDAP	Comprehensive Econometric Micro-simulator for Daily Activity-Travel Patterns
CO	Carbon Monoxide
CO_2	Carbon Dioxide
COPERT	COmputer Program to calculate Emissions from Road Transport
DDM	Direct Demand Model
DEFRA	Department for Environment, Food & Rural Affairs
DfT	Department for Transport
FAMOS	Florida Activity Mobility Simulator
FSM	Four Step Model
GBM	Gradient Boosting Machine
GHG	Greenhouse Gases
GIS	Geographic Information System
GLMM	Generalised Linear Mixture Model
GWR	Geographically Weighted Regression
HBEFA	Handbook of Emission Factors for Road Transport
HC	Hydrocarbons
HGV	Heavy Goods Vehicle
IPCC	Intergovernmental Panel on Climate Change
ITN	Integrated Transport Network
ITNUP	Integrated Transport Network Urban Paths
KNN	K-Nearest Neighbour
LGV	Light Goods Vehicle
LSOA	Lower Super Output Areas
MAPE	Mean Absolute Percentage Error
ML	Machine Learning
MNL	Multinomial Logit
MOVES	Motor Vehicle Emission Simulator
MSE	Mean Square Error
NAEI	National Atmospheric Emissions Inventory
NAPTAN	National Public Transport Access Nodes
NEMO	Network Emission Model
NL	Nested Logit
NO_x	Nitrogen Oxides
NTM	National Transport Model

OLEV	Office for Low Emission Vehicles
OLS	Ordinary Least Squares
ONS	Office for National Statistics
OS	Ordnance Survey
PHEM	Passenger car and Heavy duty vehicle Emission Model
PM	Particulate Matter
RF	Random Forest
RMSE	Root Mean Square Error
SMOTE	Synthetic Minority Over-sampling Technique
SO	Sulfur Monoxide
SO_2	Sulfur Dioxide
SUE	Stochastic User Equilibrium
SVR	Support Vector Regression
TAZ	Traffic Analysis Zone
TREMOD	Transport Emission Model
UE	User Equilibrium
UKTCM	UK Transport Carbon Model
UKTM	UK Times Model
UKTM	UK Times Model
ULEZ	Ultra Low Emission Zone
VKT	Vehicle Kilometres Travelled
VOA	Valuation Office Agency
VOC	Volatile Organic Compounds
WHO	World Health Organisation

List of Tables

List of Figures

1. Introduction

1.1. Background

Anthropogenic Greenhouse gases (GHGs) have significantly increased since the pre-industrial era and are considered as main contributors of climate change and daily temperature extremes that have been observed around the globe (IPCC, 2022). In particular, among the GHGs, Carbon Dioxide (CO_2) has been identified to hold the largest share, estimated to contribute between 72% and 76% of the total GHG emissions globally (Olivier and Peters, 2019; US Environmental Protection Agency, 2017). The continuous growth of CO_2 emissions – which according to recent data peaked in the previous decade – may have irreversible impacts on the Earth's ecosystems and human health and can damage the economy (Chaabouni and Saidi, 2017; Freitas et al., 2017).

However, the increase in GHGs emissions on a global scale is not the only disturbance related to the environment and human health. Other gases such as Nitrogen Oxides, Carbon Monoxide and Particulate Matter, (normally referred as Air Pollutants) originating from the combustion of fossil fuels (Aoki, 2017) for the needs of electricity, industry and transportation have been proved to affect the economy (Taghizadeh-Hesary and Taghizadeh-Hesary, 2020; Xie et al., 2019), and also have significant impacts on human health (Hamanaka and Mutlu, 2018), such as cardiovascular and respiratory diseases, as well as mortality (Bergstra et al., 2018). The World Health Organisation (WHO) identifies six major air pollutants, namely Particulate Matter (PM), Carbon Monoxide (CO), Nitrogen Oxides (NO_x), Ozone (O_3), Sulfur Oxides (SO_2) and Lead (Pb) (Manisalidis et al., 2020), estimated to cause approximately 4.2 million deaths every year around the globe (WHO, 2016). Nonetheless, air pollutants do not only relate to impacts on human health. In fact, air pollution and climate change are interrelated, since several air pollutants are also significant contributors to global warming (Guerreiro et al., 2016), while air quality can also affect climate and vice versa, with both having direct or indirect effects on health (Orru et al., 2017).

Consequently, to tackle the threats of climate change and air pollution, countries have introduced several measures aiming to reduce their emissions, such as the ratification of the Paris Agreement. Signed by 195 countries around the world committing to reduce GHG emissions, the agreement functions as a tool to keep global temperature rise below 2 ^{0}C compared to pre-industrial levels (U.N, 2015).

To mitigate the impacts of GHGs the UK, as an independent country, has introduced the Climate Change Act 2008, committing to reduce its GHG emissions to 80% by 2050, compared to 1990 figures (HM Government, 2011). More recently, the UK has announced a more ambitious plan to reduce carbon emissions by 68% by 2030 compared to the 1990 levels (UK Government, 2020) supported by investing in green energy across various sectors (HM Government, 2020) and also introduced the 'Net Zero' strategy, aiming to eliminate GHG emissions from all sectors (including transport) by the middle of this century (HM Government, 2021).

In addition, the UK has adopted the Clean Air Strategy where specific targets and associated actions are set to reduce the emissions of air pollutants from various sectors. The desired emission reductions are set for 2030 using 2005 as the base year and differ depending on the pollutant (DEFRA, 2019). Moreover, some local authorities in the UK have introduced Clean Air Zones[1] (CAZs), where local measures are applied to achieve immediate improvement in air quality and health, but also focus on the transition towards low emission economies (DEFRA, 2020). At the time of writing, London is the first city in the UK that has introduced a CAZ – namely Ultra Low Emission Zone (ULEZ) – in April 2019 and is about to be expanded in 2021 (Greater London Authority, 2019), while the cities of Bath and Birmingham have recently joined the scheme.

[1] CAZs are essentially predefined geographic areas – normally urban – focusing on the improvement of air quality, encouraging the operation of low emission vehicles. Encouragement can be supported by the introduction of restrictions – in the form of charges – to enter the CAZ (DEFRA, 2020).

1.2. Road Transport: Significance, complexity, and the impact on air quality

Road transportation is a complex and broad field that has been explored from different perspectives and disciplines, at aggregate and disaggregate levels for purposes such as traffic flow forecasting (e.g., Abadi et al., 2015; Yu and Cho, 2008), planning urban infrastructure (e.g., Batty et al., 2003) and assessing environmental impacts (e.g. Chapman, 2007; Ellison et al., 2013). The significance of and dependence on road transport can be appreciated based on both the historically increasing number of vehicles and an expanding street network. In particular, there has been an increase of over five million vehicles in the UK during the last decade, with an increase of 53% since the records began in 1994 (Department for Transport, 2020a), the majority of those being cars. In addition, the last decade, 2,000 more miles of roads have been constructed, when for the last twenty years the figures are more than double, exceeding 5,000 miles (Department for Transport, 2020b).

However, the increasing use of road transport comes with its drawbacks, such as its contribution to noise levels, GHGs and air pollutant emissions. Recent data show that despite a decrease in total UK's GHG emissions by 32% since the 1990s, emissions from road transport have increased by 6% during the same period (Office for National Statistics, 2019). In fact, road transport alone makes up approximately 20% of the UK's total GHG emissions (Office for National Statistics, 2019) and 92% of emissions originating from transport – and in particular CO_2 (Latake and Pawar, 2015) – having global impacts and contributing to climate change (IPCC, 2022). In addition, road transport is a significant source of air pollutants – such as Nitrogen Oxides (NO_x), Particulate Matters (PM) and Carbon Monoxide (CO) – contributing up to 80% of total transport pollutant emissions (Department for Transport, 2018a). These pollutants are responsible for negative impacts on human health and ecosystems (DEFRA, 2018) and even though there has been a significant reduction in air pollutant emissions, damage to human health can occur even at low levels (European Environmental Agency, 2014; Ricardo Energy & Environment, 2019). Consequently, there can be no argument that for the UK to meet its targets, focus should be placed on the road transport sector.

1.3. Literature gap and motivation

Considering the significance and contribution of road transport to GHG emissions and air pollution, to date, a number of studies has focused on the estimation of emissions from road transport to further understand environmental and health implications and facilitate policy making. However, numerous limitations and diverse results are observed, usually depending on the selected modelling approach and data availability, subject to area/country of application. For example, emissions are usually estimated at an aggregated level such as national, regional or city-wide (e.g., Borge et al., 2012; Ong et al., 2011; Sookun et al., 2014). Therefore, conclusions about local impacts cannot be drawn. In addition, some countries – such as the UK – estimate emissions on roads of minor importance based on average regional flows (Pang et al., 2016; Tsagatakis et al., 2017), due to lack of traffic measurements on these roads. Considering that minor roads are less crowded, but make up 87% of total road length in the UK (Department for Transport, 2019), the aforementioned methodological approach implies incomplete and unreliable emission estimation across the full extent of the road network.

Consequently, it becomes obvious that availability of traffic information across the full extent of the road network, is critical to address this issue. However, traffic data collection to this extent is challenging and costly, and therefore transport departments rely on automatic traffic counters (ATCs). Still, ATCs are normally not integrated throughout the road network. In the UK – as in most countries – ATCs are only installed at selected locations on major roads covering only a fraction of the network, while data for minor roads are usually collected manually, seasonally and at selected locations.

Lack of traffic count measurements across the road network, underlines the need for a method to estimate these values as accurately as possible at all possible locations on the road network. To date, a number of attempts has been made to estimate traffic for roads where data is not available. In particular, estimation of Annual Average Daily Traffic (AADT) – a measure of traffic volume[2] defined as the average number of vehicles at a given location on an average day over a year[3] (McCord et al., 2003; Roess et al., 2011) – has been investigated in studies on motorized (e.g., Lowry, 2014), and non-motorized transport (e.g., Hankey et al., 2017; Lu et al., 2017). AADT data that can be collected by ATCs where passing vehicles are monitored on a 24 hour basis, are used for a number of applications in road transport studies, such as accident prediction (e.g., Çodur & Tortum, 2015), GHG emission estimation (e.g., Puliafito et al., 2015), noise exposure estimation (e.g. Morley and Gulliver, 2016; Shu et al., 2014) and economic evaluations of safety projects (e.g. Wang et al., 2013) among others. AADT values are also fundamental for road construction, planning, maintenance and pavement design studies (Leduc, 2008).

However, although research on AADT estimation has been improved with the incorporation of novel modelling approaches, significant limitations can still be observed. Firstly, studies are usually limited within the boundaries of urban areas (e.g. Doustmohammadi & Anderson 2016, Kim et al. 2016), or preoccupied with AADT estimation on particular road classes, such as major roads (e.g. Caceres et al. 2012). This implies that estimations for rural or suburban areas are neglected, while only a few studies incorporate minor roads (e.g. Apronti et al. 2016, Morley & Gulliver 2016). Secondly, estimations are mainly conducted on total AADT while traffic volumes for different vehicle types is largely unexplored. The latter is considered fundamental to estimate emissions originating from road transport, since GHG and air pollutant emissions are expected to differ depending on the vehicle type, while it can also facilitate policy making as well as urban and environmental planning. Thirdly, the majority of studies focuses on estimation accuracy (e.g. Fu et al., 2017; Shojaeshafiei et al., 2017), without considering the impact that several characteristics – such as land use – have on road transport. Furthermore, the

---

[2] I am aware that the terms "traffic flow" and "traffic volume" are used interchangeably in the literature. In my thesis, I follow Zhao and Park, (2004) and use the term "traffic volume" for AADT. Further information on these terms can be found in Appendix A.

[3] AADT is given by (Leduc, 2008): $AADTi = \sum_{j=1}^{365} \frac{TC_{i,j}^{24}}{365}$, where $TC_{i,j}^{24}$ is the 24 hour traffic count on road link i at day j.

explanatory variables[4] used in several models are limited and consequently many potential affecting factors are not taken into account. This is also fundamental not only to facilitate traffic volume estimation, but also to examine the complexity of road transport and how it interrelates with urban – and where possible rural – infrastructure and demographics. The latter is again vital for decision making in the transport field, but also across a wide range of interconnected sectors such as urban and environmental planning and of course the economy.

1.4. Aims and objectives

The aim of my thesis is two-fold. First, is to identify the degree of influence specific factors have on traffic volume (i.e., AADT) variations across the road network. Second, is to assess the quantity of CO_2 and three air pollutants (PM, $NO_{x,}$ and CO) originating from road transport and identify potential emissions abatement through technological developments of road vehicles and policies development.

The thesis intends to contribute to the road transport literature by addressing and potentially overcoming the identified limitations of the modelling implemented so far, as discussed in section 1.3. That is, CO_2 and the three air pollutants emission estimation – and potential abatement – is conducted at link (i.e., street segment) level for all segments, while a comprehensive set of factors that affect traffic volumes for different vehicle types is examined. England and Wales are used as an empirical study to examine the influence of factors on traffic as well as estimate emissions and emission abatement. In addition, while emissions will be estimated for England and Wales and robustness checks will be performed to assess the methodological effectiveness (i.e., evaluation), an additional – experimental – case study is conducted for the Greater London area. The case study aims to perform a policy assessment, where the effects of the recently introduced ULEZ and its potential extension on the three air pollutants will be examined.

[4] Explanatory variables indicate the characteristics considered to "explain" the deviations of AADT on the road network. These are essentially variables considered to affect AADT – further discussed in chapters 2, 3 and 4.

To meet the aim of the thesis, the following objectives have been drawn:

- Objective 1 – To understand the road transport and emission modelling approaches currently used in the literature.
- Objective 2 – To model traffic volume and evaluate the model performance.
- Objective 3 – To model and estimate present (i.e., base-year) emissions and emission projections

Addressing the first objective, enables the identification of the structure and underlying processes of road transport and emission models. Provided that to estimate emissions from road transport traffic information is required, this will allow to conclude on the most suitable road transport and corresponding emission models – i.e., a competent conjoint approach. After addressing the first objective, the modelling approach to the case of estimating AADT to the full extent of the road network in the study area can be applied (objective 2). Considering that the selected approach can model traffic volumes as a function of affecting characteristics, objective 2 allows to address one of the aims of the thesis – to identify and assess the impact specific factors have on traffic volume. Moreover, by evaluating the model's performance will allow to proceed with confidence to estimate emissions. As a word of clarification, it should be noted, that AADT estimation can generally be divided into current-year and future-year estimations (Castro-Neto et al., 2009), with the former using data from existing traffic counters to develop models capable of estimating AADT at locations where counts are not available when new data are used (Selby & Kockelman, 2013) and the latter incorporating historical traffic data, aiming to estimate short or long term future AADTs at the same locations. This work focuses on the former, where the model is developed and applied on data from existing traffic counters, so as to test its accuracy and potential application on street segments where counters are not available. Finally, modelling present and future emissions corresponding to the third objective, builds on the findings of the second objective and facilitates to address the second aim of the thesis – to assess the quantity of CO_2 and the three air pollutants and identify potential emission abatements.

Taking into account – current and future – air quality policies as well as considering present and projected vehicle fleet composition and traffic projections figures, will allow to estimate both existing and succeeding emissions and thus conclude on potential abatement and evaluate policy efficiency. Moreover, by comparing the emission modelling outcomes with other established modelling approaches, will allow the assessment of the emission model, and also to double-validate the transport model and evaluate the overall process.

As a final note, it should be highlighted, that the scope of the thesis is to assess the impact of technological developments and implementation of policies on CO_2 and air pollutants' emissions in England and Wales. The thesis develops and assesses a modelling approach experimenting with novel methodologies, where traffic and emissions can be estimated on a granular level. Taking into account that policies are implemented for the UK as a whole and that road transport forms only a part of the total emissions where targets are set, the model and its associated outcomes can be used to inform policy making and also provide some evidence for the potential to achieve the emission goals.

1.5. Thesis structure

The thesis is organised in six chapters and five appendices, the latter containing additional information and results for corresponding chapters, that have not been included in the main body of the thesis. Following the introduction, chapter 2 focuses on the literature review, related to the concepts of transport and emission modelling. Specifically, in section 2.1 the chapter is briefly introduced. Section 2.2 presents the traffic data collection processes and discusses potential uncertainties in the derived data, while in section 2.6 the factors that have been found to affect traffic volumes are investigated, with the factors grouped and discussed individually. Section 2.4 presents the three major road transport modelling approaches that have been widely used in the literature, followed by corresponding applications in research. In section 2.5 focus is placed on studies where AADT estimation is conducted, while in section 2.6 the emission factor development methods are presented. In section 2.7 the road emission modelling approaches and their applications in research are discussed and in section 2.8 the concept of transport scenarios is briefly discussed, where examples from three different families of scenarios are presented. Finally, in section 2.9 the chapter is concluded.

Chapter 3 focuses on the development of a road transport model. Section 3.1 briefly introduces the chapter contents and section 3.2 presents the data that has been collected as well as the data manipulation process. In section 3.3 the methodology is presented, where firstly the selected road transport modelling approach is justified and then the steps to develop the model are described. Section 3.4 presents the modelling results and section 3.5 discusses the findings. In section 3.6 the chapter is summarised.

Chapter 4 identifies the road transport determinants, so as to address the first aim of the thesis. Again, section 4.1 introduces the chapter and section 4.2 describes the utilized dataset. In section 4.3 the method to identify and quantify the effect of the characteristics on traffic volumes is presented and the data is analysed. The analysis results are presented in section 4.4 and discussed in detail in section 4.5. Finally, section 4.6 summarises the chapter.

Chapter 5 focuses on the estimation of emissions for the two case studies – England and Wales, and Greater London. Section 5.1 provides the overview of the chapter. In section 5.2 additional datasets are presented and merged with data from preceding chapters. Section 5.3 firstly discusses and justifies the selected road emission modelling approach. Secondly, it presents the emission model and finally, the ULEZ and scenario analysis is undertaken. In section 5.4 the results for both case studies are demonstrated, while extensive discussion of the findings is presented in section 5.5. Again, section 5.6 summarises and concludes the chapter.

Finally, chapter 6 concludes the thesis. This chapter summarises the findings and identifies and discusses existing limitations and offers suggestions for potential future research.

2. Transport and emission concepts

2.1. Chapter overview

This chapter intends to cover the major topics related to the aims and objectives of this research that can be found in the literature, so as to understand the progress that has been made in the field of road transport studies and facilitate the process of traffic volume and emission modelling. In particular, section 2.2 starts with investigating the traffic data collection methods, where potential uncertainties in data quality are also identified. In section 2.3 the literature on studies investigating the factors affecting traffic and their relative effect on traffic volumes is reviewed. This will enable strengthening the knowledge on the effect these factors have on traffic volumes and also to build the list of potential factors to be explored. Moreover, based on the reviewed literature – and domain knowledge – specific factors not taken into account can be identified, so that they can be considered and assessed in this research. Furthermore, considering the outcomes of the reviewed studies, the variation of effects of these factors across different areas, regions, and countries, can be identified and compared with the findings in the UK. That is, the effect of the same factors can differ significantly depending on the study area, since homogeneity across countries and/or cities cannot be assumed.

Considering that estimating emissions from road transport traffic data extracted from transport models is required, in section 2.4 road transport modelling approaches are explored and then focus is placed on AADT estimation models (section 2.5). The goal is to understand these processes and conclude on the most suitable approach for this research. In section 2.6 a review of the emission factor development methods is taking place and in section 2.7 emission modelling approaches are explored. Coupled with the transport modelling literature, this is necessary to conclude on the most suitable model to estimate emissions from road transport, based on available data and the selected transport model. Section 2.8 investigates transport scenarios developed by different institutions. Based on the scenarios' reliability, relevance, and data availability, the scenarios to explore can be concluded, so as to produce emission projections. Finally, in section 2.9 the chapter is summarised and a brief discussion on how the topics covered will be used in the thesis takes place.

2.2. Traffic data collection

To extract information and monitor the road network for better management the Department for Transport (DfT) utilizes raw data extracted from Automatic Traffic Counters (ATCs), which are installed at selected locations across the network (Morley and Gulliver, 2016), normally placed on roads of major importance exhibiting higher traffic levels such as motorways and Trunk roads. However, considering the costs to integrate ATCs to the full extent of the road network the DfT also collects raw data from manual counts that are undertaken seasonally for different road classes (Department for Transport, 2016a), with each road being classified as either major or minor. The major road network includes Motorways and 'A' roads indicating main arteries with heavy traffic flows and often many lanes, used for long distance travel, with 'A' roads further subclassified in Trunk and Principal roads (Department for Transport, 2018b). The minor road network includes 'B', 'C' and Unclassified ('U') roads, which are of lesser significance, carry lower traffic and are normally maintained by local authorities (Department for Transport, 2018b). By utilizing the collected raw data from both the ATCs and manual counts, the DfT performs a series of calculations to extrapolate AADT values, at the locations where traffic counts are undertaken. The methodology applied to extrapolate these AADT values, depends on the road class, with the concluding AADT providing information about the total counts as well as traffic counts by different vehicle types. In this section the methodology to collect traffic count data for the different road classes is introduced. Moreover, the methodology to extrapolate AADT values from the raw data is presented and associated discrepancies are discussed.

2.2.1. Manual Traffic Counts

Manual counts comprise the majority of traffic counts on the road network, with approximately 10,000 counts conducted every year in both major and minor roads, represented as links with a unique ID (Department for Transport, 2013). Counts are undertaken by trained enumerators over a 12-hour period – 7am to 7pm – (Department for Transport, 2016a) during 'neutral days' – indicating days where traffic is expected to behave similarly. There are 110 neutral days in a year, which are normally weekdays between March and October with public holidays excluded (Department for Transport, 2013). Considering the large amount of road links in the country, not all links are counted every year. In the case of major roads, different links are manually counted, with counts undertaken annually or on a cycle of 2, 4 or 8 years, while for minor roads, a representative sample is selected every year, and the growth between two consecutive years is applied to estimate traffic counts for the latest year (Department for Transport, 2013).

2.2.2. Automatic Traffic Counters (ATCs)

ATCs are permanent installations normally embedded on the road surface using inductive loops and piezoelectric sensors – or a combination of both – to record information such as vehicle length and wheelbase, used to distinguish between vehicle types (Department for Transport, 2017). Inductive loop detectors essentially consist of wires embedded under the road surface and are widely utilized sensors in traffic management systems (Leduc, 2008). These sensors function as an electrical circuit, where the metal parts of a passing vehicle interact with the loop, acting as a magnetic field and therefore creating electric current (Oluwatobi et al., 2021). A roadside device can then record these signals. Piezoelectric sensors, offer easy installation and low maintenance cost (Han et al., 2013; Yonar, 2019), and operate in a similar principle converting kinetic to electric energy (Ayaz et al., 2022). These sensors are also embedded under the road surface, although the system utilizes the pressure applied by the vehicle on the sensor to create an electric signal, proportional to the degree of pressure (Jinturkar and Pawar, 2016; Rajab et al., 2011). The signal can again be recorded by a device, placed on the side of the road. However, other technologies are occasionally utilized, such as video counts, radar, rubber tubes and Automatic Number Plate Recognition cameras (ANPR), while the DfT has also plans to enrich the dataset in collaboration with Highways England and local authorities (Department for Transport, 2018c).

ATC data are collected continuously from approximately 180 devices spread across the major road network in Great Britain (GB). The devices record both the number of passing vehicles as well as their physical properties, producing a 24-hour figure as a combination of two 12-figures (7am-7pm and 7pm-7am) for each day of the year (Department for Transport, 2013). In addition to traffic counts, ATCs produce supplementary information – the expansion factors and growth factors – that is used to extrapolate AADT for the major and minor road network (Nosal et al., 2014). To calculate both expansion and growth factors, all roads with ATCs are firstly classified in 22 categories (shown in Table B-1 in the Appendix). Notice that ATCs within the Greater London area are handled differently, where only four categories are used. Expansion factors are calculated by firstly dividing the 24-hour counts over 365 days by the 12-hour day figure (7am to 7pm) for each neutral day, and then calculating the median of each ATC expansion factor for each of the categories (Department for Transport, 2010). This results into over 16,000 factors, corresponding to neutral days, expansion factor categories and vehicle types (Department for Transport, 2013). The growth factors indicate traffic growth at national level and are used to calibrate manual count points when counts have not been undertaken in the reference year.

2.2.3. Annual Average Daily Traffic (AADT) extrapolation

To extract AADT values for major and minor road, the expansion and growth factors derived from ATCs are utilized in a two-step process. Firstly, the relevant expansion factors are multiplied by the 12-hour total, depending on the vehicle type, count date and expansion factor category (Table B-1) for all manual counts and then, the growth factors are applied to the previous year's AADT for calibration, for the remaining roads (Department for Transport, 2013).

However, the reliability of the final AADT is questionable, considering the procedure to conclude on these values. Uncertainty in the data is mainly related to the data collection process, although drawbacks can be identified in the calculation of AADT with the utilisation of expansion and growth factors. For example, in the case of ATCs, although the devices are overall considered reliable, like any device they may suffer from malfunctions (Bickel et al., 2007; Chen et al., 2018; Medeiros et al., 2010). This can consequently result into short- or long-term data gaps, with missing traffic count values being propagated through the process and consequently providing erroneous AADT values (El Esawey et al., 2015). Moreover, studies have identified discrepancies in the extracted traffic counts from ATCs when evaluated with ground truth data. For example, Chauke, (2015) collected manual traffic count data at the location where ATCs are already installed. This research concluded that discrepancies between the recorded traffic counts and the ground truth data normally range between 2% to 5%, although in some cases, discrepancies can exceed 5%. In addition, variations in traffic count accuracy have been observed between urban and rural areas with traffic counts in urban areas being more accurate (Gadda et al., 2007). Issues related to ATCs traffic counts can also be observed in the case of vehicles' classification, where vehicles may be misclassified (Bharadwaj et al., 2016), with LGVs and cars being confused in most cases (Yu et al., 2010), due to comparable size and characteristics. This confusion occurs mainly due to low sensitivity of the systems used to monitor traffic, an issue that can be resolved with systems using multiple technologies combined (e.g., inductive loops with laser scanners), considered to provide more accurate data (Bellucci and Cipriani, 2010). However, counting and classification errors are not explicit for ATCs, with similar issues being observed in the case of manual counts. Specifically, Zheng and Mike, (2012) in their study in the UK, have concluded that manual counting errors are normally less than 1%, although classification errors are more significant, averaging between 4% and 5%. Finally, considering that AADT from manual counts are essentially estimated from a sample, it is fair to conclude that the reliability of the dataset is questionable.

2.3. Drivers of road traffic

Traffic volume on each road is dependent on various factors that can be attributed to specific characteristics of the road and its surrounding environment, as well as its adjacent roads and neighbouring areas. Moreover, different factors may have a different degree of influence on traffic volume, while some may specifically affect particular types of vehicles. Thus, to identify these factors and the magnitude at which they affect traffic volume is important, not only to understand road transport and traffic deviations across the road network, but also to provide further insights for interrelated sectors, such as urban planning, policy making and the economy.

The factors and their effect on road transport are usually identified with the employment of statistical modelling, or through behavioural studies. In behavioural studies, the effects are explored based on qualitative analysis normally relying on data collected from surveys and interviews, where statistical modelling can also be involved. The effects of the characteristics (i.e., factors) in cases where statistical modelling is present, can be explored by examining the derived coefficients. However, depending on the method used, the effect of factors on traffic cannot always be directly assessed – e.g., in the case where 'black box'[5] or simulation models are employed (Burns et al., 2020).

In the following subsections (2.3.1–2.3.5), a synopsis of studies exploring several characteristics thought to affect traffic is provided. It should be noted that an attempt to capture all the potential factors identified in the literature has been conducted. However, due to the different nature of the reviewed studies as discussed above, the impacts of factors on traffic cannot necessarily be assessed. Thus, focus on the effects of these factors on traffic is presented and discussed when available.

The identified attributes can generally be classified in five major categories: roadway characteristics, socioeconomic factors, land use, public transport and parking facilities.

2.3.1. Roadway attributes

Roadway attributes are related to various characteristics of the road segment at the location where the traffic count point is placed. For example, Xia et al., (1999) investigated the effect of 12 variables (i.e., attributes) on AADT, to conclude that the number of lanes, road environment and functional classification (i.e., the road being arterial or collector) all exhibit high positive coefficients, with the

[5] 'Black-box' models are considered the models where there is lack of transparency on the predicted outcomes and consequently the predictions are not explained in a way that people can understand (Rudin, 2019).

number of lanes and functional classification explaining 50% and 26% of traffic volume variation respectively. Out of the four regression models tested by Zhao and Chung, (2001) in Florida, the best performing model – validated by Mean Square Error (MSE) – indicates that functional class and number of lanes are the most significant predictors with high positive coefficients. Number of lanes and road class have since been used in many following studies. Specifically, Zhao & Park (2004), Yang et al., (2011), Doustmohammadi & Anderson (2016) and Shojaeshafiei et al. (2017) have all incorporated these attributes in various US states, with Zhao and Park, (2004) and Doustmohammadi and Anderson, (2016) again concluding that that number of lanes are statistically significant with high positive coefficients. Road width, has been used as an variable among others by Chen et al., (2019).

Selby & Kockelman (2013) used speed limits for street segments to conclude that they are important to capture traffic deviations – coupled with number of lanes and road class – in their study in Texas, although the impact is minor in urban areas. Speed limits have also been used by Fu et al. (2017), while Apronti et al. (2016) incorporate type of road surface to distinguish between paved and unpaved roads, considering the large number of unpaved roads in the study area (Wyoming, US). The same study, also incorporates a 'highway accessibility' variable applied to low volume roads which is found in a number of previous studies as well (e.g. Mohamad et al. 1998; Selby & Kockelman 2013; Xia et al. 1999; Zhao & Park 2004), although applied to capture connectivity of higher class roads with motorways. In addition, Sarlas & Axhausen (2014) take into account road density in the vicinity of traffic count points, found to be statistically significant, although having low correlation with traffic. Other studies have also introduced topological roadway characteristics for traffic analysis, such as the degree of connectivity[6] (e.g. Jiang and Liu, 2009; Pun et al., 2019) and several centrality measures (e.g. Gao et al., 2013; Jayasinghe et al., 2015; Zhao et al., 2017) . Finally, Chen et al., (2019) used a Generalised Linear Mixture Model (GLMM) and the Synthetic Minority Over-sampling Technique (SMOTE) with twenty variables in Seattle. The findings show that 17 and 15 variables respectively are

[6] From graph theory, the degree of connectivity can be defined as the number of connecting links of a node – i.e. the number of adjacent nodes connected to a certain node (Lieberthal and Gardner, 2021).

statistically significant with the "Spatial weighted volume"[7] having the highest positive coefficient. Roadway characteristics also indicate high positive signs, while five variables for GLMM and seven for SMOTE have negative coefficients with local streets and one-way roads exhibiting the highest.

2.3.2. Socioeconomic characteristics

Socioeconomic characteristics are the most common attributes used in the literature and are taken into account in almost all studies. Specifically, population of local settlements exhibits a positive coefficient in most studies (e.g. Apronti et al., 2016; Mohamad et al., 1998; Seaver et al., 2000; Zhao and Chung, 2001; Zhao and Park, 2004) indicating that the higher the population the higher the traffic values will be observed. Population has also been considered by Selby & Kockelman (2013), Fu et al. (2017), Raja et al., (2018) and Zhang and Chen, (2020) drawing similar conclusions, with these studies conducted in three US states (Texas, Alabama, Kentucky) and the Republic of Ireland (Fu et al., 2017). However, Doustmohammadi and Anderson, (2016) find that population has negative correlation with traffic volume as opposed to findings from other studies in their study in Alabama, US.

The number of households and household income is used by Eom et al. (2006), to find that household income has negligible association with traffic volumes, based on two different models, while Apronti et al. (2016) used employment and per capita income. Other socioeconomic attributes considered as drivers of traffic used in applied studies include age of population, gender balance in the population and car ownership (e.g. Cervero & Kockelman 1997; Stead 2001; Zhao & Chung 2001; Zhang 2007; Aditjandra et al. 2012). Findings from these – behavioural – studies indicate that age has negligible association with driving – and consequently higher traffic volumes – and that car ownership is related with increased driving. Similarly, Jahanshahi & Jin (2016) state that car ownership indicates higher traffic volumes, although correlation varies across areas. However, one has to bear in mind that car ownership is strongly connected to household income (Silva et al., 2012).

[7] Defined as the "the sum of other roads' weighted AADT divided by the reciprocal of squared Euclidean distance, in counts/sq. feet" (Chen et al., 2019)

2.3.3. Land use

Land use variables indicate the surrounding environment at the location where the count point is placed, with the majority of studies mainly distinguishing the traffic count points at either being on urban or rural areas (e.g. Eom et al. 2006, Fu et al. 2017, Zhao & Chung 2001, Zhao & Park 2004), with urban areas usually demonstrating higher traffic values. However, other studies have introduced more detailed land use classification. For example, Xia et al. (1999) classified land use by introducing business, residential and fringe areas while Kim et al. (2016) classified land use as commercial, residential, industrial and miscellaneous, to find that commercial areas are the ones that are highly correlated with traffic among the rest of the land use types.

In Seaver et al., (2000) among the 45 characteristics considered, the attribute indicating the number of farms is statistically significant with a positive sign. However, one has to bear in mind that the study is conducted on local rural roads, indicating that traffic is likely to be affected by different factors as opposed to roads in urban areas. Apronti et al. (2016) refined this approach by introducing a more detailed classification, considering several types of agricultural land use, forest and recreational sites among others, while finally, Chen et al., (2019) considered the number of industrial and commercial properties within 150ft of the road segment, together with a land use mixture variable[8].

2.3.4. Public transport

Public transport variables are almost entirely absent from AADT estimation studies. As an example, Sarlas & Axhausen (2014) incorporated density of public transport stops in the vicinity of traffic count points. On the other hand, behavioural studies have investigated the impact of public transport supply on mode choice and road traffic. Cervero (1994) finds that residents near rail stations are more likely to use public transport, which is associated with lower use of private road vehicles. Stead (2001) discovered that bus frequencies are associated with travelled distances by individuals and mode

[8] Defined as the entropy of five different land use types within 150ft of road segment.

choice, albeit findings differ across geographic areas. Aditjandra et al. (2012) also conclude that public transport accessibility is associated with lower individual driving.

2.3.5. Parking availability

The influence of parking availability and costs in AADT is also absent from the majority of AADT estimation studies that have been reviewed in the literature. The only identified study is Chen et al., (2019), where the average parking cost in the vicinity of each road segment is considered. On the other hand, parking correlation with mode choice and therefore road traffic, is an established area of research in behavioural studies, where Hess (2001) and Zhang (2007) conclude that availability of parking encourages individual car use. In addition, there is a considerable literature about the occurrence of parking as a pulling factor for traffic, especially in those cases where free or low-cost parking is available. Studies conclude that traffic can be generated to the areas where parking is available, but can also contribute to the increase of traffic at the surrounding areas (e.g. Arnott and Inci, 2006; Shoup, 2006; Kelly and Clinch, 2009; Arnott and Williams, 2017; Inci et al., 2017).

2.4. Road transport modelling approaches and applications

Road transport has been studied through the lenses of several disciplines with numerous attempts to capture its complexity. To date, the available road transport modelling approaches make use of simulation and statistical models, usually also incorporated within larger interactive models. In this section, the three most common transport modelling approaches are presented and briefly discussed: (i) The Four Step Model (FSM), (ii) the Activity Based Model (ABM) and (iii) the Direct Demand Model (DDM).

2.4.1. Four Step Model (FSM)

2.4.1.1. Model description

The FSM is a procedure during which the study area is split into Traffic Analysis Zones (TAZs)[9] where the modelling process takes place based on trips travelled between TAZs. Modelling is conducted on four steps, where different models are applied on each of the steps, namely (i) trip generation, (ii) trip distribution, (iii) mode choice and (iv) traffic assignment.

- Step 1 – Trip Generation

The trip generation step, estimates the total number of trips that either originate or terminate in each TAZ based on the amount of activity (Brustlin et al., 2012) and it is the most important step due to the fact that if errors are introduced they can be propagated to the next steps (Ortuzar and Willumsen, 2011). Trips are generated by trip generation models using a growth factor[10] and regression modelling, and are defined as a movement from a point of origin to a point of destination (Ortuzar and Willumsen, 2011).

- Step 2 – Trip Distribution

The trip distribution step links the trips for each pair of TAZs so that travel patterns are represented through an Origin–Destination (OD) matrix (Martin and McGuckin, 1998). In this step, distribution models are estimating the values (i.e., the number of trips) in each OD matrix cell on the basis of any available information. The most common model for trip distribution is the Gravity Model which takes the trips that are produced at one zone and distributes them to other zones, on the basis of impedance which is assumed to capture underlying travel behaviour (McNally, 2007).

---

[9] TAZs are subject to constraints and guidelines which are difficult to consider and implement in a single TAZ design process (Martínez et al., 2009). Consequently, TAZs number is based on the study's purpose and data availability (McNally, 2007).

[10] This growth factor is related to variables such as population, income and car ownership (Ortuzar and Willumsen, 2011).

- Step 3 – Mode Choice

Mode choice estimates and allocates OD trips to the available transportation modes based on trip maker, journey and transport facility characteristics (Ortuzar and Willumsen, 2011). The modelling is based on discrete choice analysis (Ben-Akiva and Bierlaire, 1999), with Multinomial Logit (MNL) and its extension, the Nested Logit (NL) (Ben-Akiva, 1973) models being the most common approaches.

- Step 4 – Trip Assignment

Finally, the trip assignment step, distributes traffic on the network according to a certain route choice principle (Lam and Lo, 2004). The most widely used models for trip assignment are:

i. All-or-nothing assignment (AON)

This is the simplest route choice and assignment method, which assumes that there are no congestion or capacity effects, and the trips are assigned based on the minimum cost path (i.e., travel time) between origin i and destination j.

ii. User equilibrium assignment (UE)

First defined by Wardrop, (1952) and later formalised by Ortuzar and Willumsen, (2011), "Under equilibrium conditions traffic arranges itself in congested networks in such a way that no individual trip maker can reduce the path costs by switching routes".

iii. Stochastic user equilibrium assignment (SUE)

The Stochastic User Equilibrium (SUE) models include both pure stochastic as well as user optimised equilibrium conditions, and a more realistic perspective is examined, where a traveller chooses the path according to the minimum perceived travel cost rather than the actual one (Lam and Lo, 2004).

2.4.1.2. Model applications

FSMs have been explored and applied to estimate travel demand in several studies. In this subsection, two representative applications of FSMs are briefly described. First, the National Transport Model (NTM) for Great Britain and second, the European-level Trans–Tools (TT) model.

NTM, is a passenger transport model combining cross sectional with time series data extracted from surveys, census as well as forecast datasets (Department for Transport, 2012). In particular, the National Trip End Model (NTEM) produces the total number of trips so that the Demand Model, which is the core module of the system, estimates trips between OD pairs and mode types based on generalised costs of possible alternatives. The National Rail Model and the Great Britain Freight Model also provide information to the Demand Model for rail and freight, respectively.

The TT model is built in a Geographic Information System (GIS) framework and it incorporates an economic model, a freight trade and freight choice model as well as a passenger demand model (Rich et al., 2009). The passenger demand model creates matrices for car demand and air passengers estimated and adjusted based on traffic counts (Rich et al., 2009). The road traffic assignment model calculates averaged daily traffic, split into particular time periods using Stochastic User Equilibrium.

2.4.2. Activity Based Model (ABM)

2.4.2.1. Model description

The Activity Based Model (ABM) shares similarities with the FSM in the sense that activities are generated, destinations are identified, modes are determined and routes are predicted and like FSMs, ABMs involve the utilization of different models (Castiglione et al., 2014). ABMs, which have been developed as an advanced alternative to FSMs, seek to represent travel choices made by individuals (Chiu et al., 2011) and model a sequence of trips, defined as tours and classified by purpose (Ortuzar and Willumsen, 2011). ABMs are part of a wider model system where interaction between individual models occurs with the major component being the 'Population Synthesis' providing sociodemographic information to the system (Brustlin et al., 2012). ABMs can be classified in three different modelling approaches namely (i) constraints based models, (ii) utility maximising models and (iii) computational process models (Rasouli and Timmermans, 2014).

i. Constraint based models

These are the first generation of ABMs and their objective is to examine whether an individual activity agenda is feasible within a space time context (Rasouli and Timmermans, 2014).

ii. Utility maximising models

Utility maximising is the most widely ABM approach in use. These models are based on the concept that individuals maximise their utility when scheduling daily activities (Ben-Akiva and Bowman, 1998) and consist of a series of utility maximisation based discrete choice models, such as Nested Logit (NL) and Multinomial Logit (MNL) (Rasouli and Timmermans, 2014).

iii. Computational Process models

The computational process is a new experimental rule-based method, that aims to go a step further and imitate the way individuals think and act when building schedules (Pinjari and Bhat, 2011), hence in a more realistic way as opposed to the unrealistic utility maximising assumption of econometric models.

2.4.2.2. Model applications

ABMs have also been incorporated in several applications. As examples the Comprehensive Econometric Micro-simulator for Daily Activity-Travel Patterns (CEMDAP) model, developed at the University of Texas, Austin by Bhat et al., (2004) is a system of econometric models representing decision making behaviour of individuals (Pinjari et al., 2008). The two systems comprised in the model are the generation–allocation module which models the decisions of household adults to undertake various travelling activities, and the scheduling module that uses these decisions as inputs to model the complete activity patterns considering constraints such as work or school and determine choices, such as number of tours and stops. The Florida Activity Mobility Simulator (FAMOS) model, is a two-module microsimulation model system (Pendyala et al., 2005), with one module generating the household attributes and another the activity travel patterns for individuals (Kitamura and Fujii, 1998). Finally, the ALBATROSS model[11] developed at the Eindhoven University by Arentze and Timmermans, (2004), is considered the most comprehensive computational process ABM (Rasouli and Timmermans, 2014). The model takes as input an activity schedule, a list of constraints, household and individual characteristics, zone data and transport system characteristics. The scheduling module is the system's core part which controls the scheduling processes in a sequence of steps, with each step modelled using decision trees to indicate the probability of choice, given personal characteristics and previous choice. ALBATROSS has been used, among others, by Beckx et al., (2009) combined with an emission model to assess emissions generated from passenger cars in the Netherlands. In addition, the Forecasting Evolutionary Activity-Travel of Households and their Environmental Repercussions (FEATHERS) model (Bao et al., 2018) which is the extended Flemish version of ALBATROSS (Balac and Axhausen, 2016; Zhuge et al., 2017) has been applied by Lee et al., (2012) and validated with smart card data by Cho et al., (2015) in Seoul.

2.4.3. Direct Demand Model (DDM)

2.4.3.1. Model description

Direct Demand Models (DDMs) are statistical – empirical approaches developed as an alternative to the standard travel demand models (Hankey et al., 2017) such as FSMs and ABMs. They can be seen as aggregated models aiming to explain traffic volume as a function of relevant factors thought to influence traffic. DDMs, which have been described as "the empirical equivalent of the conventional demand function of economic theory" (Wardman et al., 1994), aim to incorporate trip generation, distribution and mode choice into one single equation (Ortuzar and Willumsen, 2011). DDMs are

[11]ALBATROSS stands for: A Learning-based Transportation Oriented Simulation System (Arentze and Timmermans, 2004).

classified into purely direct models, relating demand to mode, as well as trip and traveller characteristics with the use of a single estimated equation, and partially direct models that "exploit attributes of differences between modal choice and total demand for travel" (Kepaptsoglou et al., 2017). The latter, is essentially a two-stage process where the first step explains the total demand across all modes and the second distributes the demand with a mode choice model (Gaudry et al., 2000).

2.4.3.2. Model applications

DDMs have been extensively used in several transport fields ranging from air transport to rail and metro ridership, road transport and bike and pedestrian studies. For example, Gelhausen et al., (2018) developed a DDM to forecast passenger and flight volumes at German airports. Regression analysis of passenger and flight volume time series data is applied, while taking into account economic variables as well as major demand shocks, such as the German Unity and 9/11 attacks. Fagnant and Kockelman, (2016) developed a DDM to estimate peak hour cyclist counts at locations where counts are not available. Number of lanes, parking presence and socioeconomic variables are used, among others, as inputs to a Poisson regression and two negative binomial models. Similarly, Hankey et al., (2017) implement DDMs in a small town to estimate bicycle and pedestrian traffic flow, using land use and socioeconomic variables within a buffer distance around the count stations, by applying stepwise linear regression and using R^2 as a goodness of fit measure. DDMs have also been extensively applied in rail and metro ridership. Gutiérrez et al., (2011) developed multiple regression DDM models to estimate the number of passengers boarding each metro station as a function of the station characteristics, socioeconomic factors, and environment dimensions. Focus is placed on the use of a distance decay weighting function to capture particular characteristics within the station's catchment area. Cardozo et al., (2012), conduct a study on metro ridership in Madrid and compare a Geographically Weighted Regression (GWR) with Ordinary Least Squares (OLS) using station, socioeconomic and land use characteristics. Zhao et al., (2014) use DDM to identify associations between station ridership and influencing factors in Nanjing metro in China through a multiple regression model. They employ 11 variables associated with ridership at station level, such as land use, intermodal and station characteristics. Kepaptsoglou et al., (2017) provide an example of partial DDM where a traffic assignment model is used to obtain travel times and costs between TAZ centroids, fed into a linear regression model to estimate traffic demand as a function of travel times and costs.

DDMs have also been extensively used in road transport studies. For example, Anderson et al., (2006) applies a multiple linear regression model to predict average daily traffic and uses road characteristics and socioeconomic variables within a buffer as predictors. DDMs have also been used to predict other transport-related variables. As an example, Sarlas and Axhausen, (2014), use spatial regression models to estimate average morning peak hour speed of each road link at an extended part of Switzerland's major road network. Finally, DDMs are also often used to estimate AADT, as discussed in more detail in the next section. As an example, Lowry, (2014) estimates AADTs in a small community based on modified versions of stress centrality (OD Centrality) (Shimbel, 1953) as the only predictors. OD Centrality is calculated using multipliers for parcel land use data, similar to but, different from the trip production and attraction steps of FSMs.

2.5. Annual Average Daily Traffic (AADT) estimation models

AADT estimation in particular, is not a novel concept with analyses conducted for over 30 years now (e.g. Neveu 1983, Fricker & Saha 1987). To date, a number of approaches has been applied and tested using known traffic volumes extracted from traffic count points to estimate AADT values at locations where traffic counters are not available. These are usually statistical approaches (i.e., DDMs) where regression models are applied to estimate AADT with the utilization of explanatory variables. However, it should be noted that statistical modelling can also be used to identify and assess the effects of the explanatory variables on traffic volumes as discussed in section 2.3, although in this section, focus is placed elsewhere. Specifically, in this section I precent and discuss the three principal – statistical – approaches normally employed for AADT estimation that I have identified in the literature, namely Linear Regressions, Spatial Statistic models and several Machine Learning (ML) techniques.

2.5.1. Linear regression models

Linear regression models are the most popular models in the literature, with applications identified in early AADT estimation studies. For example, Mohamad et al. (1998) applied linear regression in county roads[12] across forty counties in Indiana. Out of the 11 predictors tested, four have been found to be statistically significant – county population, location of count point (i.e., urban/rural), access to motorway and total arterial mileage of the county. The model has been validated by collecting new AADT data in eight randomly selected counties in the State. Xia et al. (1999) used a linear regression model from a sample of 450 count stations to estimate AADT for non-state roads in Florida. The model used six independent variables and is validated by using 10% of the sample. A multiple regression model has also been used by Seaver et al., (2000) to estimate AADT in rural roads in Georgia, US. Zhao & Chung (2001) extended the work of Xia et al., (1999) by using a larger dataset, incorporating land use and accessibility variables. The model is again validated using 10% of the sample and examining the R^2 and Mean Square Error (MSE) values. Zhao and Park, (2004) used a similar set of variables and applied OLS this time to compare with other modelling approaches, to conclude that linear regression provides the lowest estimation accuracy. Yang et al., (2011), tested linear regression models in North Carolina and more recently, Apronti et al., (2016) used linear regression to estimate AADT values on low volume roads in the state of Wyoming, US. Similarly, Doustmohammadi & Anderson (2016) applied linear regression with land use data in two small and medium sized cities in Alabama. For validation, different but similar sized cities were selected to apply the developed model and a T-test is used to test for model accuracy. Raja et al., (2018) tested three different linear regression models using 150 traffic count points in low volume roads in the same US state. The models are validated

[12] County roads are defined as roads under the responsibility of county highway departments, rather than state departments of transportation (Mohamad et al., 1998).

using 55 points from the main sample and are compared based on R^2. Pun et al. (2019) applied a multiple linear regression model in Hong Kong using a set of 850 traffic counters in the area. The model is again assessed using the R^2 metric, while the Root Mean Square Error (RMSE) is used to assess its accuracy. Finally, OLS has still been used recently among other models by Zhang and Chen, (2020) for AADT estimation in Kentucky, US and by Pulugurtha and Mathew, (2021) to estimate AADT on local roads in North Carolina.

2.5.2. Spatial statistical models

Evolution in the field of spatial statistics and development and availability of spatial datasets has led to the application of spatial methods for AADT estimation. In these models, spatial location and correlation are taken into account so that data points are weighted according to their distance from the location where the dependent variable is to be estimated (Loyd, 2007). Among several approaches, Kriging interpolation and Geographic Weighted regression (GWR) are the most popular spatial models used in road transport studies. However, although these models are available and can be easily applied within various GIS platforms[13] – such as GeoDa, ArcGIS and GWR4 – and other data analytics software[14], their application is fairly limited in this field. For example, Wang & Kockelman (2009) applied Kriging interpolation with Euclidean distances among traffic count stations in Texas, using 20% of the data to validate the model and finding that AADT values are overestimated by 33%. Kriging with additional – explanatory – covariables (i.e., CoKriging) has been applied by Eom et al. (2006) and Shamo et al., (2015) in North Carolina, and by Selby & Kockelman (2013) in the US state of Washington, all using different types of semivariograms[15] and part of the dataset to validate the models. Kim et al. (2016) also used CoKriging[16] in South Korea, while more recently, Kriging for AADT

[13] GeoDa and GWR4 have been developed in academia by Anselin et al., (2010) and Nakaya, (2014) respectively, while ArcGIS is developed by the Environmental Systems Research Institute (ESRI). GWR4 can be used for GWR analysis only, while GeoDa and ArcGIS also incorporate Kriging analysis.

[14] Open-source data analytics software such as R and Python can conduct spatial analysis.

[15] The semivariogram is a tool that is used to measure spatial autocorrelation – essentially a measure of variance (Hohn, 1999; Olea, 1999)

[16] CoKriging is essentially similar to the Kriging technique, where additional variables are used to predict the dependent variable (Stein and Corsten, 1991).

estimation has been used in comparison with GWR by Selby & Kockelman (2013) in Texas and by Mathew and Pulugurtha, (2021) again in North Carolina. In both of these studies, Kriging outperforms GWR, although the models' validation – measured with Mean Absolute Percentage Error (MAPE) – shows that estimation errors still remain high, between 59% and 63% in Selby and Kockelman, (2013)[17] and at 84% in Mathew and Pulugurtha, (2021). GWR to estimate AADT has also been used by Zhao & Chung (2001) and Zhao & Park (2004) in Florida.

2.5.3. Machine Learning and Data Mining models

More recently, rise on the applications of Machine Learning (ML) and Data Mining algorithms in various disciplines, has shown that these methods can provide higher estimation accuracy in regression problems (Brathwaite et al., 2017; Paredes et al., 2017; Sekhar et al., 2016) and consequently these approaches have reached AADT estimation studies. However, to my knowledge at the time of writing, applications are mainly focused on production of future AADT predictions based on historical data, while ML use has been scarce for AADT estimation at unmeasured locations. Studies where ML is used to estimate AADT are presented below[18].

ML applications can be found in Shojaeshafiei et al. (2017), where the K-STAR (K*) and Random Forest (RF) algorithms are applied to estimate AADT in Alabama. The models are validated using the R^2 and the Nash-Sutcliffe (N-S) statistic. By comparing the models, this study found that RF performs better compared to K*. Wu and Xu, (2019) also used RF to estimate AADT at selected roads in Washington state and compared it with a linear regression model, to find that the models exhibit similar performance. RF has also been tested by Das and Tsapakis (2019) for AADT estimation in Vermont.

[17] The range of MAPE values refers to different types of road classes modelled in this study.

[18] Considering the large and constantly expanding number of ML algorithms developed by computer scientists, one has to take into account that identifying all algorithms in the literature is a very challenging task. I acknowledge that numerous new models – possibly applied on AADT estimation and more broadly in road transport studies – can potentially be found in the literature. However, this is beyond the purposes of the thesis. Instead, some well-established algorithms in the literature are identified, presented, and also applied in the following chapters.

The model has again been compared with linear regression using RMSE. As opposed to Wu and Xu, (2019), in this study the outcomes show that RF provides better estimates.

Together with RF, Pun et al., (2019) applied Support Vector Regression (SVR) and multiple linear regression using AADT from 850 counters in Hong Kong. The models are evaluated using R^2 and RMSE to find that multiple linear regression performs – slightly – better compared to the other models, with higher R^2 and lower RMSE values. SVR has also been used by Sun and Das, (2019) in low volume roads of 8 parishes in Luisiana. SVR is compared with Poisson and Negative Binomial models to conclude that SVR delivers higher estimation accuracy compared to the other models.

SVR has also been compared with Artificial Neural Networks (ANNs). For example, Khan et al., (2018) applied SVR and ANN to estimate AADT in South Carolina, using a sample of 164 traffic count points located on different road classes in the State. Model training is conducted using two thirds of the data and validation is taking place with the rest of the sample and by calculating RMSE and MAPE for each model. The results show that SVR provides more accurate estimations.

Fu et al. (2017) have also used ANNs to estimate AADTs in the Republic of Ireland. MAPE values are compared with OLS estimations based on 96 traffic counters that have been used to train and test the models (i.e., MAPE is calculated 'in-sample'). Validation reveals that ANNs perform better than OLS with MAPE being 23% and 67% for each model respectively. This study also appears to be the first attempting to extend the study area to country level.

Several modifications of ANNs can also be found in the literature. For example, Tawfeek and El-Basyouny, (2019) used a Deep Neural Network (DNN) to estimate AADT in Alberta, Canada, using a sample of 1,350 traffic counters. The model is compared with OLS using R^2, to conclude that DNN can improve the R^2 values by around 35%.

Finally, cluster analyses can also be found in traffic volume estimation studies. Specifically, Gecchele et al. (2011) tested different clustering methods (i.e., K-means, K-medoids, PAM[19]) to group traffic counters using temporal traffic volume patterns, indicating that groups of roads with similar traffic patterns can also be identified. These groups where used to create seasonal factors that can be later used for AADT calculation. However, although the study makes use of clustering algorithms, it focuses on temporal pattern identification and seasonal traffic volume estimation and does not take into account other characteristics that can affect traffic volumes. A similar approach has been used by Caceres et al. (2018), where hourly traffic patterns and AADT values from 455 intercity road traffic counters, have been used. The hourly patterns have been utilised to create groups (i.e., clusters) and three different linear regression models have been applied within each group to estimate AADT. The models are validated using the initial (i.e., measured) AADT values from the traffic counters, with 75% of the sample been used for training and 25% for validation. MAPE values range from 15% to 46% depending on the model.

2.6. Emission factors

Emission factors are coefficients that relate the amount of the pollutant released with the corresponding activity that causes the emission (Abdallah et al., 2020) and are considered vital to estimate the amount of pollutant released from a specific source (Nghiem et al., 2019). These factors have been developed for numerous pollutants in various sectors – such as industry, housing and transport – and are essential for the computation of emissions. For road transport in particular, the emission factors indicate the amount of the emitted pollutant per unit mass of fuel burned, energy consumed or more commonly per distance driven (Facanha and Horvath, 2007). The accurate development of emission factors is critical, considering that imprecise factors and their utilization in corresponding emission models (further discussed in section 2.7) can result in significant discrepancies in emission estimations (Shen et al., 2014). In this section, the various methods that are used to

[19] PAM stands for: Partitioning Around Medoids (Kaufman and Rousseeuw, 1990).

measure road vehicle emissions and develop the emission factors are presented. Based on this, several advantages and limitations related to each of the emission measurement methods are discussed.

2.6.1. Emission measurement methods

Emission factors are normally developed by utilizing emission data collected during vehicle emission monitoring experiments. The factors are derived for several vehicle categories – or even single vehicles as well as entire vehicle fleets – using different techniques. The extrapolated factors from the experiments depend on various parameters such as vehicles' type and age, corresponding emission technology, fuel type and quality as well as maintenance, operating and ambient air conditions (Franco et al., 2013; Nghiem et al., 2019). This indicates that emission factors can change over time, considering increased vehicle mileage, improvement in engine emission technologies and change in fuel specifications (Brimblecombe et al., 2015; Carslaw and Rhys-Tyler, 2013; Dallmann et al., 2012). The measurement technique, coupled with the operating conditions and the types of vehicles selected for each experiment, all indicating significant determinants related to the quality of the derived emission factors. Measurement of emissions to derive emission factors are normally conducted using two different methods, namely (i) controlled conditions and (ii) real-world conditions (Jamriska and Morawska, 2001; Smit et al., 2010).

2.6.1.1. Controlled conditions

The controlled conditions method is the most widely used (Raparthi et al., 2021; Seo et al., 2021) and refers to experiments conducted in lab environments where several parameters introducing variability such as driver behaviour, environmental and traffic conditions are controlled (Ježek et al., 2015). The controlled condition experiments are conducted with the use of either chassis or engine dynamometer devices under different predefined settings – known as test cycles – aiming to simulate real driving conditions, such as free flow or urban driving (André et al., 2006).

Test cycles are essential for all chassis and engine dynamometer experiments, and consequently, the capacity to which real-world driving conditions are in fact represented is vital for the quality of the experimental outcomes (Joumard et al., 2006). The test cycles are classified in two types, the steady-state and transient cycles. Steady-state cycles involve continuous operation of the engine using constant speed and load, testing on different modes for a sufficient amount of time, until the emissions are stabilized (Franco et al., 2013). Then, when the predefined number of modes is tested, the emission measurements are normally joined using a weighted average method, with weights being specified for each mode (Artelt et al., 1999). On the contrary, variations in the vehicle's operating conditions are introduced during the transient test cycle experiments (Franco et al., 2013). The variations may include continuous change in vehicle's speed and load among others, aiming to represent real-world operating conditions and account for driving situations such as idling deceleration and acceleration (Barlow et al., 2009; Franco et al., 2013).

The chassis dynamometer test cycles are mainly transient (Yanowitz et al., 2000), where the – chassis – dynamometer imitates the resistance force that is exercised on the wheels of the vehicle for each cycle (Yang et al., 2018). During the testing, the vehicle is tied down to remain still and it is operated by a driver who follows a gear change pattern to comply with a predefined time-speed profile and stay within the required speed threshold (Nine et al., 1999). The tailpipe emissions are finally fed into a gas analyser to measure emissions (Yang et al., 2018). On the other hand, the engine dynamometer test cycles are mainly steady-state (Franco et al., 2013), although tests can also be undertaken on a transient cycle basis (e.g., Jiang et al., 2009; Jin et al., 2017). During these tests, the engine and exhaust are removed (Bunker et al., 1997; Jiang et al., 2009) and the dynamometer is connected to the engine shaft, so that the resistance imposed to the engine is directly simulated (Kęder et al., 2014; McCormick et al., 1998).

2.6.1.2. Real-world conditions

Measuring emissions under real-world conditions involves the utilization of several techniques such as tunnel measurements, remote sensing and on-road or on-board measurements, with all being used extensively in numerous studies.

Tunnel measuring methods involve the collection of emissions data from both ends of the tunnel, split by direction. The overall flow of pollutants from all vehicles passing through the tunnel is measured for a predefined timeframe (Jamriska et al., 2004), where the pollutant is related to the traffic flow (Hueglin et al., 2006). The difference in pollutant concentrations between inlet and outlet is measured and the total pollution is then calculated by multiplying it with the estimated airflow through the tunnel, with wind speeds also taken into consideration (Franco et al., 2013). Tunnel measurement emissions are very common in the literature and examples can be found in Grieshop et al., (2006) who conducted emission measurements in the state of Pennsylvania, US, Abdallah et al., (2020) in Beirut, Lebanon and Raparthi et al., (2021) in India among others.

Measuring emissions via remote sensing is a method that has been firstly introduced by Stedman and Bishop, (1990) and involves instantaneous measurement of the concentration ratios of pollutants in the exhaust plume of the passing vehicle (Huang et al., 2018). The system for measuring emissions is based on the concept of absorption spectroscopy (Burgard et al., 2006) indicating that gases absorb light at particular wavelengths and normally consists of an infrared (IR) and ultraviolet (UV) beam sources as well as speed and acceleration sensors. When the vehicle crosses the sensor, the system is triggered and it measures the speed and acceleration of the passing vehicle, while the IR and UV sensors identify the absorption of light at specific wavelengths when the beam passes through the exhaust plume (Smit et al., 2021). Remote sensing is a quick, efficient and effective method to monitor tailpipe emissions under real-world driving conditions (Chan et al., 2004) and consequently has become popular in many studies over the years, with examples demonstrated in Jimenez et al., (2000) who tested remote sensors to measure NO and NO_2 for Heavy Goods Vehicles (HGVs) and Davison et al., (2020) measuring CO_2 and NO_x.

On-road – or chase – emission measurements are undertaken with the use of two vehicles following each other (Rubino et al., 2008). Essentially, a mobile laboratory equipped with emission measurement devices “chases” (i.e., follows) individual vehicles and collects sample of the gases extracted from its tailpipe (Milenković et al., 2020; Shorter et al., 2005). Similar to the remote sensing method, the equipment consists of IR and UV analyzers (Krecl et al., 2021) with additional instruments such as heated flame ionization detectors (HFIDs), non-dispersive ultraviolet (NDUV) analyzers (e.g., Chen et al., 2007; Liu et al., 2011), video recording equipment as well as meteorological and positioning systems (Rubino et al., 2008; Shorter et al., 2005) being used among others. The method can thus provide real-world emission data under different operating conditions and has been applied in numerous studies around the world. For example Westerdahl et al., (2009) measured CO and Black Carbon (BC) in Beijing, China, and Wen et al., (2019) who measured NOx, CO, CO_2 and BC also in Chengdu, China. In other countries, Herndon et al., (2005) used the chase method to measure SO_2 and CH_4 from buses in New York City, while multiple experiments can be identified in Finland. Specifically, Järvinen et al., (2019) again measured emissions from city buses in Helsinki while Karjalainen et al., (2014) in Alastaro and Wihersaari et al., (2020) experimented on PM emission measurements from petrol and diesel cars respectively.

Finally, on-board (or Portable Emission Measurement Systems – PEMS) are arrangements of emission measurement devices such as gas analyzers and data recorders (Kihara et al., 2000) that are carried on board the vehicle and collecting data under real traffic conditions, while GPS, weather stations and accelerometers are also usually integrated on the vehicle (Oprešnik et al., 2012). In PEMS the emissions are transferred directly from the tailpipe to the on-board unit through pipes, where the analyzer estimates the real-time emissions (Ma et al., 2012). Emission measurements are normally undertaken for CO, CO_2, HC and NO_x (e.g., Boughedaoui et al., 2008; Unal et al., 2004) and although its applications have not been very common before the 2000s (Kihara et al., 2000), relatively recent improvements and developments that made it commercially available, has increased its use (Mamakos et al., 2011). For example, Cheng et al., (2019) and Zhang et al., (2020) used an on-board system to measure NO_x emissions from heavy-duty vehicles in China, while heavy- and light-duty vehicle's emissions where also measured by Weller et al., (2019) in Austria. PEMS for two wheeled vehicles is also demonstrated in Vojtisek-Lom et al., (2020) who used a mini-PEMS to measure CO,

CO_2, HC and NO_x and similarly, by Murena et al., (2019) for CO, CO_2 and NO_x emissions. PEMS applications on diesel and petrol cars can be found in many studies in recent years (e.g., Fitz et al., 2021, 2020; Mahesh et al., 2019, 2018; Pouresmaeili et al., 2018; Seo et al., 2021).

2.6.2. Emission factor development

The development of emission factors is dependent on the emission measurement methods with several uncertainties associated with each of the methods. The development of emission factors under controlled conditions is normally undertaken by plotting the aggregated results from various test cycles and then fitting a regression line on the data (Franco et al., 2013). Hence, it becomes obvious that this method cannot sufficiently capture the exact emission impact of each different driving cycle, while variabilities in the emission factors from different vehicles and pollutants have been observed in numerous studies (Choi and Frey, 2010). Although emission factors developed using chassis or engine dynamometer testing can model emissions and fuel consumption for multiple vehicle configurations and driving patterns (Kousoulidou et al., 2012), they cannot be representative of real on-road driving conditions, while there is normally a limited number of vehicles tested (Jamriska and Morawska, 2001; Morawska et al., 2005). Specifically, absence of real world driving conditions such as ambient temperatures and road gradients are simulated under favourable conditions, resulting in lower fuel consumption and corresponding emissions rates (Mellios et al., 2011). Moreover, engine dynamometer testing in particular, is even less useful due to the fact that results are provided in quantity of pollutant per unite of engine energy output (e.g., KWh^{-1}) and therefore it is again not directly pertinent to real driving conditions (Franco et al., 2013).

On the other hand, measuring emissions in real-world conditions offers a more accurate and realistic approach (Unal et al., 2004), although several limitations and uncertainties still apply, depending on the method. For example, emission factors developed from tunnel measurements are calculated by considering the pollutant mass concentrations in each side of the tunnel, its cross sectional area in m^2 and the distance between the two monitoring stations as well as the wind speed, number of sampled vehicles and sampling duration (Franco et al., 2013; Pierson et al., 1996; Pierson and Brachaczek, 1982). However, these capture emissions originating not only from tailpipes but also from brake and tyre wear, while the emissions are captured at aggregated levels and cannot be distinguished by vehicle type (Geller et al., 2005). Moreover, the emissions measurements are affected by wind, normally generated by larger vehicles and thus having impact on the resistance imposed particularly to smaller vehicles, therefore affecting emissions too (Corsmeier et al., 2005).

In remote sensing emission measurements, emission factors are normally fuel-based, where the tailpipe concentration of the pollutant is taken into consideration and divided by the estimated fuel consumption of the vehicle (Singer and Harley, 1996). Conversion from fuel-based to distance-based emission factors requires an estimation of instantaneous fuel economy (Franco et al., 2013). In contrast with tunnel measurements, remote sensing techniques can monitor a large number of vehicles at disaggregated level – i.e., by vehicle type (Franco et al., 2013) although these methods are also associated with numerous limitations. For example Unal et al., (2004) state that there is a significant difficulty dealing with multiple lanes on a given road as well as vehicles in close vicinity. Moreover, the vehicle load which affects the generated emissions cannot be evaluated precisely (Boughedaoui et al., 2008), while the emissions are measured for only part of the vehicle's journey and are not representative for emissions at different driving conditions (Franco et al., 2013; Unal et al., 2004). The latter can be overcome with the utilization of on-road (i.e., chase) emission measurement methods, where emission factors are developed in a similar way to remote sensing, although a representative sample of vehicles across numerous driving conditions can be monitored. However, a major disadvantage of this method is that the mobile laboratory should be placed to a minimum of ten meters distance from the vehicle being chased (Morawska et al., 2007) and therefore introducing uncertainty with regard to the prevision of emission measurement, while a maximum chase speed of 120km/h is imposed (Franco et al., 2013).

Finally, emission factor development from PEMS – i.e., on-board systems – is analogous to the chassis dynamometer process, where bins of large emission datasets are collected and mass emissions are plotted against the corresponding mean speed from the bin (Franco et al., 2013), where a regression line can be fitted. PEMS are simple and relatively inexpensive to install on a wide variety of vehicles and have therefore been extensively used in recent years. PEMS, overall overcome several of the limitations identified for both controlled and real-world conditions, considering that the equipment is carried on-board, while measurements are taking place during several driving conditions. However, the latter also introduces a degree of uncertainty, since ambient temperature may vary, while the impact of human factor (e.g., the driver's behaviour) can also affect the generated emissions (Matzer et al., 2017). Hence, one of the main disadvantages of PEMS is the low reproducibility of the measurements (Lozhkina and Lozhkin, 2016; Weiss et al., 2011). Moreover, the equipment needed can add significant weight on the vehicle – approximately 30kg-80kg – (Leatherman, 2018; Weiss et al., 2011), and therefore can introduce bias in the measurements, specifically in the case of low-weight vehicles. Finally, the number of pollutants that can be modelled is limited (Franco et al., 2013; Hausberger et al., 2022), while the measurement of hot and pulsating tailpipe gas flow during real-world driving conditions has been shown to exhibit a 23% uncertainty with PEMS (Hausberger et al., 2022).

Overall, it has been seen that emission factors can be derived, using numerous different methods to measure emissions with several uncertainties associated with each of these methods. This can result into significant discrepancies in the extrapolated emission factors depending on the selected method, also affecting the outputs of emission estimation models (discussed in section 2.7) that make use of these factors. However, the factors do not only depend on the method selected, but on numerous other characteristics, such as the criteria for vehicle selection for the experiments, the ambient conditions and the impact of human factor (e.g., driving conditions) that has been found to significantly affect the tailpipe emissions (Wihersaari et al., 2020). Moreover, the fuel characteristics are also significant contributors in emissions, with fuel type, volatility and composition all being important in the emission factor development process (Unal et al., 2004). Again, it should be noted that emission factors can change over time due to various factors such as vehicle deterioration due to increased mileage (Brimblecombe et al., 2015), technological improvement in fuel specifications as well as emission control technologies (Carslaw and Rhys-Tyler, 2013; Dallmann et al., 2012). Considering the uncertainties associated with the development of emission factors, disparities between official reported emission figures and real-world emissions from vehicles are likely to be identified, as it has also been reported in numerous studies (e.g., Fontaras et al., 2017; Tietge et al., 2015)

2.7. Emission modelling approaches and applications

As with transport models, estimation of GHG and air pollutant emissions from road transport can be conducted with the use of various emission models classified depending on geographic scale of application, methodological approach and generic model type (Boulter et al., 2007). In particular, road transport emission models can be classified into static and dynamic, with each type exhibiting advantages and disadvantages, mainly related to data availability, required computer processing and the scale of application. Static and dynamic models are further classified into (i) traffic situation, (ii) instantaneous, (iii) average speed and (iv) aggregate emission factor models (Elkafoury et al., 2013). The first three classes are also introduced in De Blasiis et al., (2013) and are generally accepted and widely used in the literature, although some studies use different classifications. For example, Boulter et al., (2007) introduce variations of average speed and instantaneous models, Esteves - Booth et al., (2002) classify the models based on the type of emissions, and Fallahshorshani et al., (2012) based on input data, study scale and type of pollutants being modelled. The major requirement for all emissions models is activity (i.e., traffic) data, extracted from transport models. For the sake of clarity and to avoid confusion among the aforementioned approaches; the traffic situation, instantaneous and average speed terminology is henceforth used. In this section, emission modelling approaches and their applications in research studies are presented. Moreover, other alternative emission modelling approaches that have been identified, are also discussed.

2.7.1. Traffic situation models

Traffic situation models estimate emissions related to particular traffic patterns, using emission factors for each situation (Smit et al., 2009). The situations are defined in terms of road type, area type, speed limit and congestion level where a specific traffic patterns occurs – i.e., “stop-and-go” driving, “free-flow”, “heavy” and “saturated” (Smit et al., 2009). These models require information on Vehicle Kilometres Travelled (VKT) and on the traffic situation applied to each road link (Baškovic and Knez, 2013).

Incorporation of traffic situation models can be found in a number of cases, such as the Handbook of Emission Factors for Road Transport (HBEFA) developed on behalf of and used by several European countries, such as Germany, Austria, Switzerland and Sweden (iCET, 2015). HBEFA, which is essentially an emission factor database, provides emission factors based on defined traffic situations (Wyatt, 2017) and has been used in several studies. For example, Borge et al., (2012) used HBEFA to estimate Nitrogen Oxides (NO_x) in Madrid and Fontaras et al., (2014) to estimate Carbon Monoxide (CO), Carbon Dioxide (CO_2) and NO_x emissions, with both studies concluding that the model overestimates emissions. On the contrary, Elkafoury et al., (2015) tested HBEFA on CO to conclude that the model underestimates emissions. The Assessment and Reliability of Transport Emissions Models and Inventory Systems (ARTEMIS) is another traffic situation model (Boulter and McCrae, 2007; Wang and McGlinchy, 2009) built to improve the European tools for emission modelling from all transport modes at national, international and regional levels (iCET, 2015). It consists of a collection of sub-models (Joumard et al., 2008), where emission estimations are based on the classification of vehicles – e.g., heavy duty, motorcycles, etc. (Andre et al., 2008). The model has been used by Martinet et al., (2017) to measure compound emissions for diesel and petrol vehicles and by Iodice and Senatore, (2015) for

CO, Hydrocarbons (HC) and NO_x for two wheeled vehicles. Liu et al., (2016) also used ARTEMIS emission factors to estimate CO_2 in Sweden.

2.7.2. Instantaneous emission models

Instantaneous emission models relate emission rates to vehicle operational modes (Baškovic and Knez, 2013), so that a traffic simulation module provides vehicle operation data and the emission module assigns an emission factor to each combination of instantaneous speed and acceleration rates (Elkafoury et al., 2013) for each interval.

The Passenger car and Heavy duty vehicle Emission Model (PHEM) developed at the Graz University of Technology (Hauseberger and Rexeis, 2017) is the most significant example of an instantaneous model (Fallahshorshani et al., 2012). PHEM is based on parameters from real driving conditions considering factors such as road gradients and vehicle loading (Hausberger et al., 2009). It is essentially a system of different modules that simulates engine power and speed, where emissions are interpolated based on "engine maps" obtained from emissions measurements and empirical engine test (Rexeis et al., 2013). PHEM is incorporated in HBEFA model, providing evaporation emission factors for air pollutants and CO_2 emissions (Wyatt, 2017).

2.7.3. Average speed emission models

Average speed models use average rather than instantaneous emission factors varying according to the average speed of a vehicle (Boulter et al., 2007) and applied to a street segment or an entire journey (Smit et al., 2009).

The COmputer Program to calculate Emissions from Road Transport (COPERT) is the most widely used average speed tool for air pollutants and GHGs (Ntziachristos and Samaras, 2000), where emission factors are expressed as a function of the average speed over a complete driving cycle (Ntziachristos et al., 2009) and can also be used to provide distance-based emission factors (Ren et al., 2016). COPERT has been used and integrated in numerous studies and models, such as the National Atmospheric

Emissions Inventory (NAEI) in the UK, providing emission maps based on spatial datasets and traffic count data (Tsagatakis et al., 2017). In other studies COPERT has been used by Vanhulsel et al.,(2014) to estimate CO_2, NO_x and $PM_{2.5}$ on major roads in Belgium and by Ong et al., (2011) to assess the impact of a shift from private cars and motorcycles to public transport, as well as the impact of a shift from conventional fuel use to natural gas on GHG and air pollutant emissions in Malaysia. In China, Wang et al., (2011) used the model to estimate emissions from passenger cars for three future scenarios, while in the UK Mascia et al., (2017) used COPERT to predict impacts of CO_2, NO_x and Black Carbon based on different traffic management measures in Glasgow.

2.7.4. Other models

There are many other – less common – models that can be found in the literature for estimating emissions from road transport, where the extent of application for each model varies across counties. For example, in the US the most popular being the Motor Vehicle Emission Simulator (MOVES) that has replaced MOBILE (Zhou et al., 2015), while in Europe the Transport Emission Model (TREMOD) and the Network Emission Model (NEMO) are in use, although their application is fairly limited compared to the models described. In the UK, the UK Transport Carbon Model (UKTCM) and the Background, Road and Urban Transport modelling of Air quality Limit values (BRUTAL) model have also been used. UKTCM is a system of sub-models and developed to provide annual projections for all passenger and freight transport supply and demand as well as estimate CO_2, CO, NO_x, Sulfur Dioxides (SO_2), Total Hydrocarbon (THC) and PM emissions (Brand, 2010). BRUTAL is based on the previous ASAM (ApSimon et al., 1994) and UKIAM (Oxley et al., 2003) models, using GIS and incorporating datasets extracted from NAEI and COPERT. A different approach based on dispersion kernels is incorporated in the SHERPA-city application developed by Degraeuwe et al., (2021) and applied in Madrid to estimate NO_x and NO_2 concentrations.

However, as an alternative and sometimes combined with the set models discussed, a different methodological approach has been applied. In particular, GHG and air pollutant emissions can be estimated by multiplying emission factors with activity data (e.g., annual vehicle kilometres travelled – VKT), similar to the methodology used in NAEI to estimate hot exhaust emissions (Brown et al., 2018). Vehicle Kilometres Travelled (VKT) can be calculated by multiplying AADT with the length of each link (Zheng and Weng, 2016), which is an essential indicator for accurate VKT calculation (Leduc, 2008). This approach has been used in Sookun et al., (2014) to estimate GHG, NO_x, CO and SO_2 in Mauritius where traffic counts (i.e. AADT) for all road classes are split by fuel type, and fuel consumption is calculated by vehicle type and road class. Similarly, Setyawan et al., (2015) estimate NO, CO_2, SO, PM and CO for a particular road in Indonesia and Jung et al., (2017) calculated VKT from traffic counts to estimate CO, NO_x, $PM_{2.5}$ and VOC emitted from trucks on major roads in a Korean metropolitan area. Labib et al., (2018) also utilised traffic counts to calculate VKT and emissions factors for each fuel and vehicle type to estimate CO_2 in Dhaka, Bangladesh and Patarasuk et al., (2016) estimated road transport emissions from VKT based on AADT, to distribute emissions along the road network in Salt Lake City, Utah. Finally, Puliafito et al., (2015) used AADT values to calculate VKT and estimate CO_2 on a grid level in Argentina and Fu et al., (2017) calculated VKT from estimated AADT values to finally estimate $PM_{2.5}$, NO_x and HC emissions for each road link in the Republic of Ireland, again being the only identified study to extend estimation at a very granular level.

2.8. Transport scenarios

Transport scenarios are developed at national and international levels by different institutions, public bodies, private companies, and the academia. The scenarios aim to project the future of transport based on several factors such as demographic characteristics, technological developments and policies implementation. Depending on the scenario, transport is considered either as part of the total energy system or as an individual sector, while road transport usually forms a separate sector of total transport figures. Individual scenarios for road transport are normally developed by transport departments across countries. To date, numerous scenarios have been developed. Investigation of these scenarios will allow to conclude on the ones to be assessed to facilitate emission projections. In this section, examples from three families of these scenarios are described.

In the UK the Department for Transport, (2018) develop five different scenarios based on projections from the National Transport Model (NTM). NTM uses the national trip end model (NTEM) dataset, considering 2015 as the base year and providing scenarios up to 2050 based on a number of factors affecting travel demand, such as car ownership, population, and income. Projections are conducted for miles travelled by vehicle type (i.e., Cars, Light Good Vehicles, Heavy Goods Vehicles and Public Service Vehicles), road class (i.e., motorways, trunk, principal and minor roads) and area type (urban and rural) for the nine English regions[20] and Wales. Average speeds as well as CO_2, NO_x and PM_{10} emissions are also projected for the same areas. All scenarios project increased total traffic growth (in vehicle miles travelled), for all vehicle types, although changes in road traffic vary depending on the area, while under some scenarios, distance travelled is reduced for certain vehicle types and/or regions. For example, scenario 5 projects a decrease of travelled miles for Heavy Goods Vehicles (HGVs) in all areas except London. Emissions decrease in all areas in all five scenarios mainly due to assumptions of increasing fuel efficiency of the vehicle fleet and increasing use of biofuels (counted as zero emissions). The NO_x emission reductions also rely on the effectiveness of European standards to control actual driving conditions.

Road transport projections have also been considered using the UK Times Model (UKTM) – an energy system model developed at the regional and national levels (Daly and Fais, 2014). UKTM also includes CO_2, Methane (CH_4), Nitrous Oxide (N_2O) and Hydrofluorocarbons (HFCs) emission projections and is divided into three supply and five demand sectors, where transport is considered as one of the demand sectors. Projections for road transport are made up to 2050, considering 2010 as the base year. Among the scenarios used in the ADVENT[21] research project, a scenario for total fuel demand in

[20] The nine English regions are: North West, North East, Yorkshire and the Humber, West Midlands, East Midlands, East of England, Greater London, South East, and South West.
[21] ADVENT stands for: Addressing the Valuation of Energy & Nature Together

road transport implies biofuels remaining at the same levels as the base year, oil fuels decreasing and electricity and hydrogen increasing. Moreover, transport demand (i.e., VKT) by type of vehicle is also collapsed by engine type, where hydrogen cars and vans show a significant increase, while diesel and petrol cars and vans are eliminated by 2050. HGVs remain constant, while the total road transport demand is increasing. UKTM is also used by the Committee on Climate Change (CCC) in the 5th Carbon Budget Report (Daly and Fais, 2014), where scenarios for several energy systems are incorporated. However, transport scenarios are based on the NTM for baseline projections of VKT and emissions (Committee on Climate Change, 2015). These scenarios combine expectations without additional effort to reduce emissions, with assessments of cost-effective requirements to achieve 2050 targets. The scenarios are focused on GHG reductions at national level that can further be disaggregated. However, road transport is incorporated in the wider "surface transport" category that also includes rail (freight and passenger).

Four different scenarios have been developed in collaboration between the universities of Sussex and Leeds (Watson et al., 2004) for the use of hydrogen in the UK, considering 2000 as the base year and projecting to 2050. The scenarios are defined around two dimensions based on governance (regionalisation and globalisation) and values (consumerism or community). Hydrogen use in transport varies in all scenarios from 5% at the low investment in transport technology scenario to 80%-100% at the Global Sustainability scenario where high investment on new low energy and low emission vehicles occurs and moderate car ownership growth applies. Based on these scenarios, Page et al., (2004) focus on the use of hydrogen in road transport, to identify its impact on total transport energy consumption, using a transport model developed for the purposes of the research. Total transport energy consumption is increased in all scenarios where the "Global Sustainability" is the only one where approximately 50% of the consumption is covered by hydrogen.

2.9. Chapter summary

In this chapter the progress around major topics on road transport to date has been reviewed. In accordance with the aims of the thesis, focus has been placed on the topics of traffic data collection and emission factor development, road transport and emission modelling, the identification of the drivers of road traffic volume and the development of scenarios, so as the future of road transport can be assessed.

It has been seen that road transport can be modelled via simulation (i.e., FSM and ABM) or statistical approaches (i.e., DDM) and that emission models require traffic information to produce estimations. Consequently, the selection of transport modelling approach is vital, since depending on the transport model a corresponding emission model should be applied. However, the selection of road transport modelling approach is highly dependent on the availability of the corresponding activity (i.e., traffic) and the associated (i.e., factors) data that have also been identified as the drivers of road traffic. In addition, it has also been seen that the correlation of these factors with traffic may vary, subject to the modelling approach as well as the geographic area of application (i.e., country, state, city, etc.). This implies that transport model selection is also dependent on the availability of data and vice versa.

Following the above, the next chapter acts upon and addresses two of the topics discussed in this chapter: transport related datasets and transport modelling; that will also form the foundation for the rest of the thesis. Specifically, data availability is explored, and related datasets are collected, introduced, and manipulated. Moreover, the approach to model traffic volumes is determined and applied. Finally, the modelling outcomes are critically discussed and assessed.

3. Road transport modelling

3.1. Chapter overview

The task of road transport modelling involves the utilization of multiple information corresponding to various characteristics affecting traffic, and of course the employment of the analogous model. Specifically, traffic volume modelling and AADT estimation is normally approached as a regression problem and conducted with the application of statistical models, such as the ones presented in section 2.5. However, as it has been discussed (see section 1.3) there are numerous limitations associated both with the models as well as the data.

This chapter focuses on the development of a methodology to estimate AADT at locations where traffic measures have not been conducted, while also attempting to address the identified limitations of the modelling implemented so far. In particular, this approach aims to address three key points: (i) the incorporation of a comprehensive set of driving factors the majority of which are not taken into account in the current literature, (ii) the consideration of all road classes (i.e., major and minor) and regions in the area of interest (i.e., England and Wales) and (iii) the improvement in estimation accuracy. In order to do so, information is extracted from a number of spatial and non-spatial datasets from different sources, with the datasets being manipulated in a GIS environment and fed into a hybrid model, based on ML techniques. The methodological output reveals traffic patterns across urban and rural areas and is demonstrated to produce accurate results for the road classes examined. The method can be used to provide outputs at different geographical scales, so it can be used both for macro and micro analyses.

The chapter is presented in six sections. Section 3.2 describes the datasets that will be used and the corresponding sources. In section 3.3 the methodology to estimate AADT at unmeasured locations is presented, the selected modelling approach is justified and then the process to estimate AADT and validate the model is presented. Section 3.4 presents the derived modelling results and in section 3.5 I discuss the findings. Finally, section 3.6 summarises the chapter.

3.2. Data

A number of spatial and non-spatial datasets have been extracted from various sources and manipulated within a GIS environment to design features (i.e., variables/factors) – discussed in subsection 3.3.2. Selection of specific datasets is based on the identified factors in section 2.3, although not all factors identified and used in the literature are available for England and Wales. For example, at the time of data collection, the number of lanes and speed limits were not available for the road network. However, I make use of many datasets to create features potentially affecting AADT, additional to those considered in the literature so far, since incorporating more data has the potential to improve model performance (Domingos, 2012). In this section a detailed description of the datasets is presented. A summary of all used datasets is shown in Table C-1 in the Appendix. More specifically, the datasets used are:

i. Traffic count points

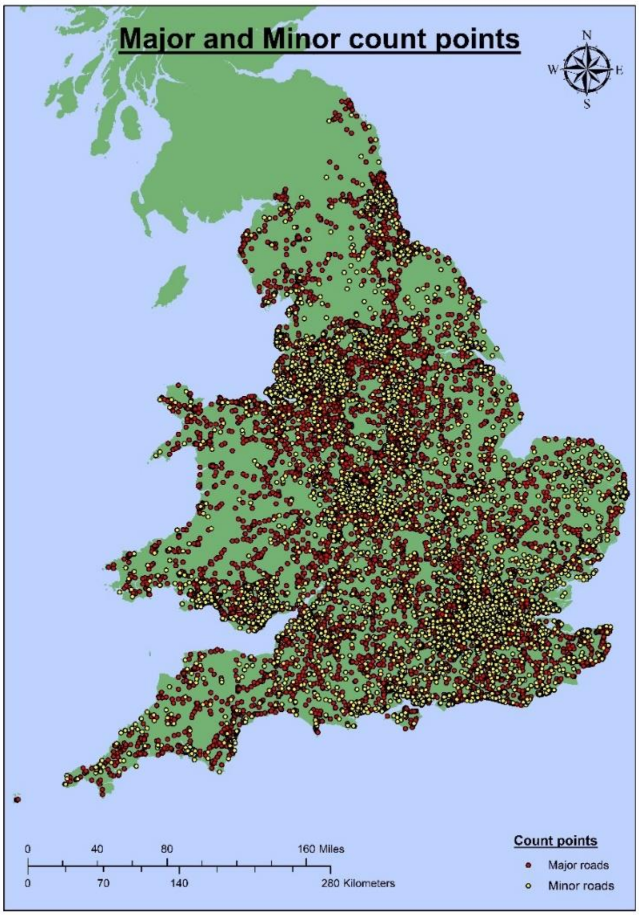
Traffic count points were derived from the UK's Department for Transport (DfT) and consist of approximately 19,000 geocoded count points in England and Wales for 2015, and are classified as Major (Motorways[22] and 'A' roads) and Minor ('B', 'C' and 'U' roads). The count points provide information about the number of vehicles (i.e., AADT) driving at that particular point. It is important to mention that the counts further distinguish among five vehicle types (Two-wheeled motor vehicles,

[22] Motorways have not been utilised at this stage of my research considering that traffic on these roads is not directly affected from its surrounding characteristics (Eom et al. 2006, Zhao & Chung 2001). Moreover, traffic volume is available for all motorways in England and Wales and consequently there is no need to estimate AADT for these roads. Motorways have been used at a later stage of the thesis to facilitate emission estimation (see chapter 5).

Cars and Taxis, Buses and Coaches, Light and Heavy Goods Vehicles). However, in this chapter, focus is placed on the estimation of total AADT – i.e., not distinguishing between vehicle types[23].

For this dataset, I further check for potential missing information and exclude faulty counters where identified. Using the locational information this dataset has been converted into a spatial form and mapped as shown in Figure 3-1.

Figure 3-1: Major and minor traffic count points locations in England and Wales



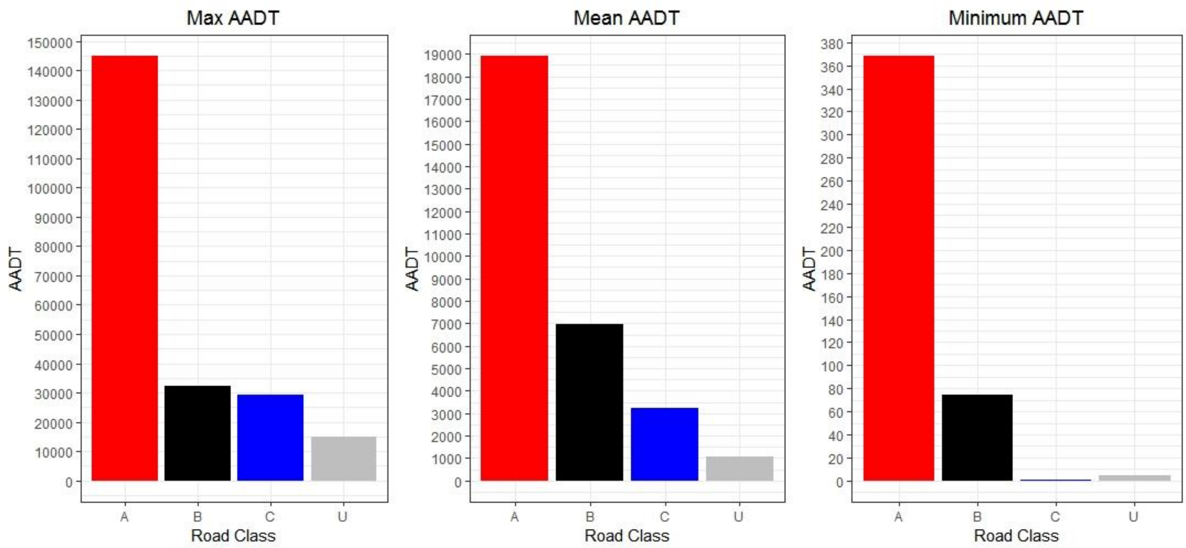
In Table 3-1 the number of traffic count points and the corresponding traffic volume (i.e., AADT) for the four road classes modelled in this chapter (i.e., 'A', 'B', 'C' and 'U' roads) are presented while in Figure 3-2, the average and the range of AADT (i.e., total number of motorized vehicles) for the same four classes is shown.

[23] Utilisation of vehicle information is further explored in chapter 4 where the impact of variables on each vehicle type is explored and in chapter 5 where emissions are estimated.

Table 3-1: Number of traffic counters and volume for each road class

Road Class	Total number of points	Total traffic volume	Traffic volume per point
A	14,670	276,364,386	18,839
B	1,032	7,151,442	6,930
C	1,058	3,592,136	3,395
U	2,041	2,182,208	1,069

Figure 3-2: Maximum (left), average (centre) and minimum (right) AADT values for all road types



ii. Road Network

The Integrated Transport Network (ITN) and ITN Urban Paths (ITNUP) spatial datasets have been extracted from Ordnance Survey (OS) and consist of the entire road network in Great Britain (GB) as of 2015. The ITN dataset contains information such as length of each segment in the network, road class and locations of junctions for all roads, while the ITNUP dataset contains man-made footpaths, subways, steps, and footbridges as well as cycle paths in all urban areas of Great Britain over 5km^2 (Ordnance Survey, 2018) as shown in Figure 3-3.

Figure 3-3: Road network spatial datasets (ITN & ITNUP)

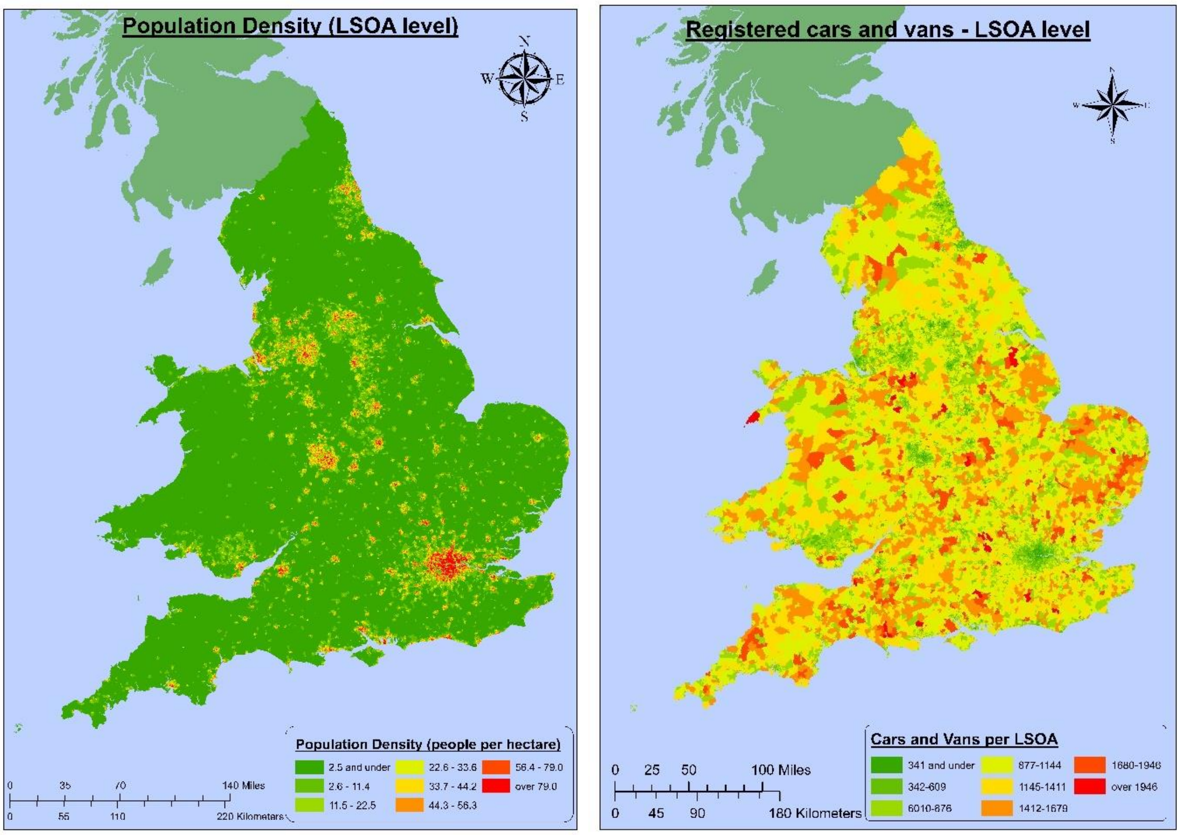
iii. Socioeconomic characteristics

Socioeconomic characteristics are derived from the Office for National Statistics (ONS), and include information about population, population density, workplace population, workplace density, number of households and median income based on data collected in 2011, and are available at the Lower Super Output Areas (LSOAs)[24] level. LSOAs are spatial datasets derived from OS for the same year (i.e., 2011) where the socioeconomic characteristics from ONS and the number of registered cars and vans – derived from the Office for Low Emission Vehicles (OLEV) for 2011 – are matched. Essentially, there are three datasets used to create a spatial dataset incorporating the characteristics needed for this research. The two non-spatial datasets (i.e., socioeconomic characteristics and registered number of vehicles) are matched with the spatial LSOA dataset. Matching is conducted in GIS using LSOA names

[24] Lower Super Output Areas are approximately 35,000 areas designed by the Office for National Statistics (ONS) for England and Wales, with population minimum of 1000.

and corresponding unique coded IDs, that have been available for all three datasets. A sample is shown in Figure 3-4.

Figure 3-4: Population density (left) and registered number of vehicles (right) by LSOA in England and Wales



iv. Public transport

Geolocated bus stops and bus stations as well as Train and Light Rail stations[25] have been derived from the National Public Transport Access Nodes (NaPTAN) database in 2016. Again, using the location information, the datasets have been mapped as shown in Figure 3-5 and Figure 3-6.

[25]This dataset includes all National Rail as well as all local metro, tram, and light rail system stations. These include the Tyne and Wear Metro (Newcastle), Merseyrail (Liverpool), Manchester Metrolink (Manchester), Nottingham Express Transit – NET (Nottingham), Supertram (Sheffield), West Midlands Metro (Birmingham and Wolverhampton), West Yorkshire Metro (Leeds) and Blackpool Tramway as well as all London rail-based transport modes – i.e., London Underground, London Overground, Docklands Light Railway (DLR) and London Tramlink.

Figure 3-5: Bus stops locations in Central London

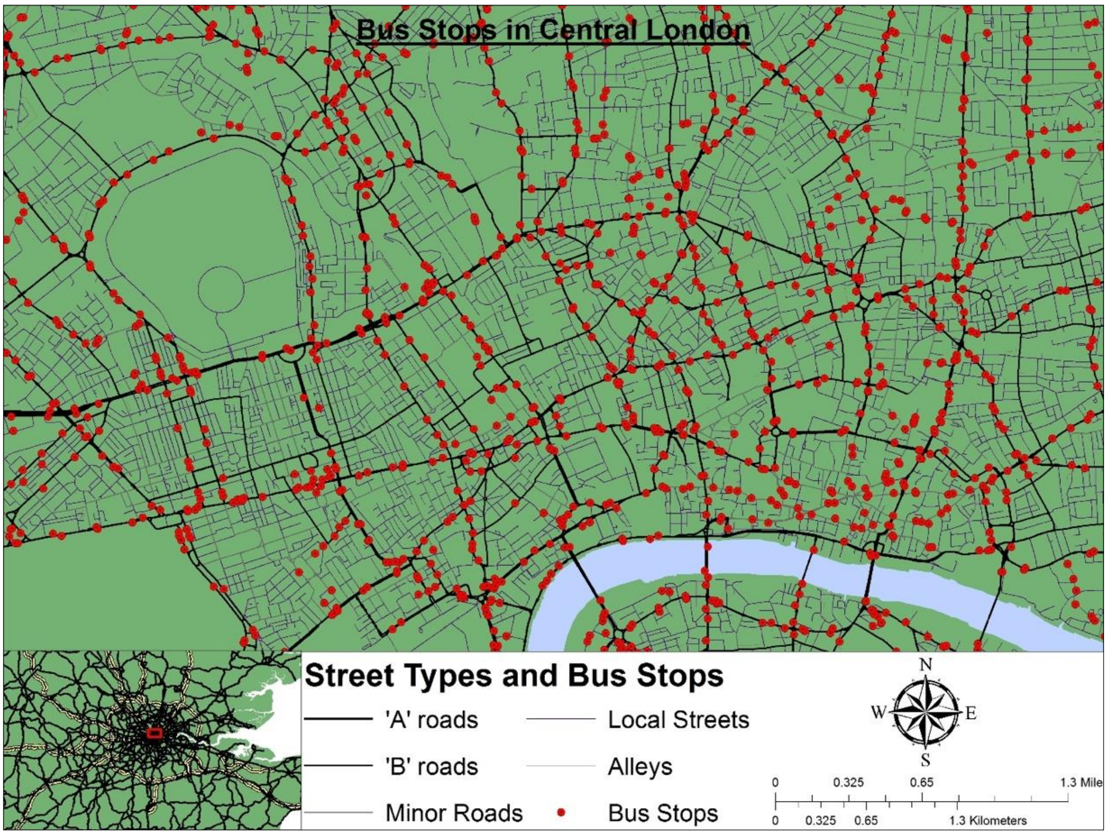
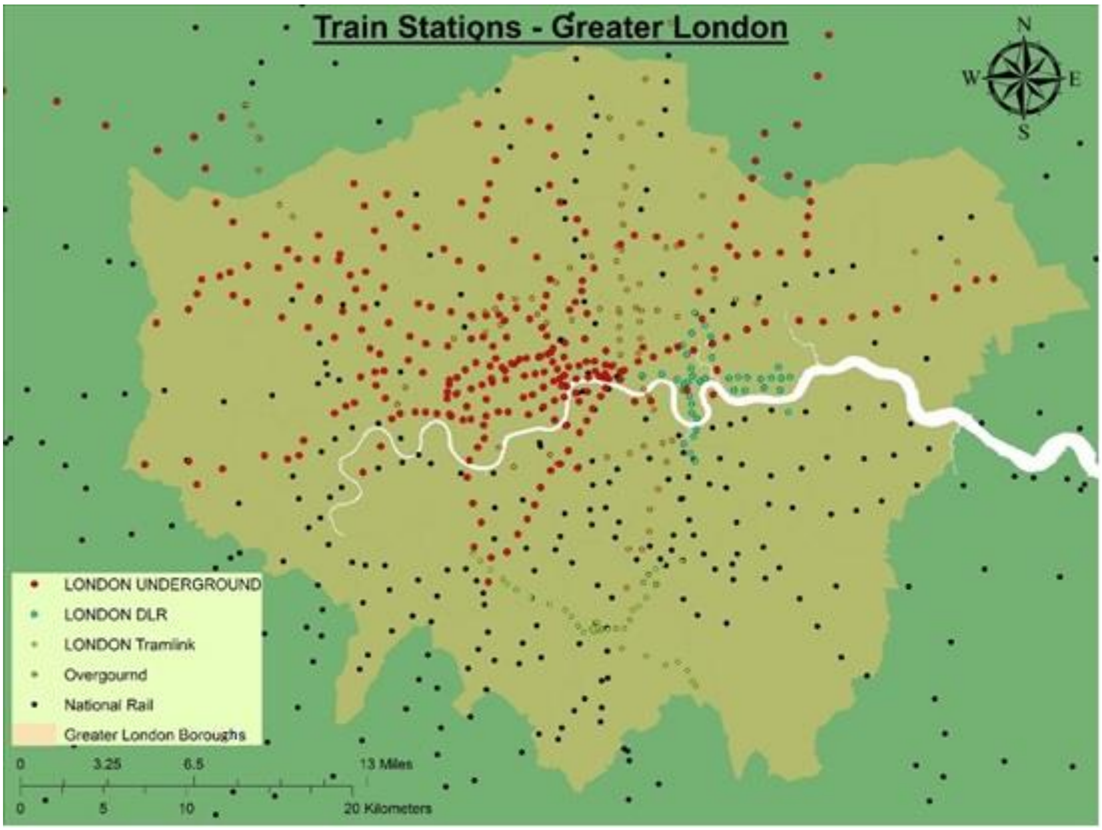


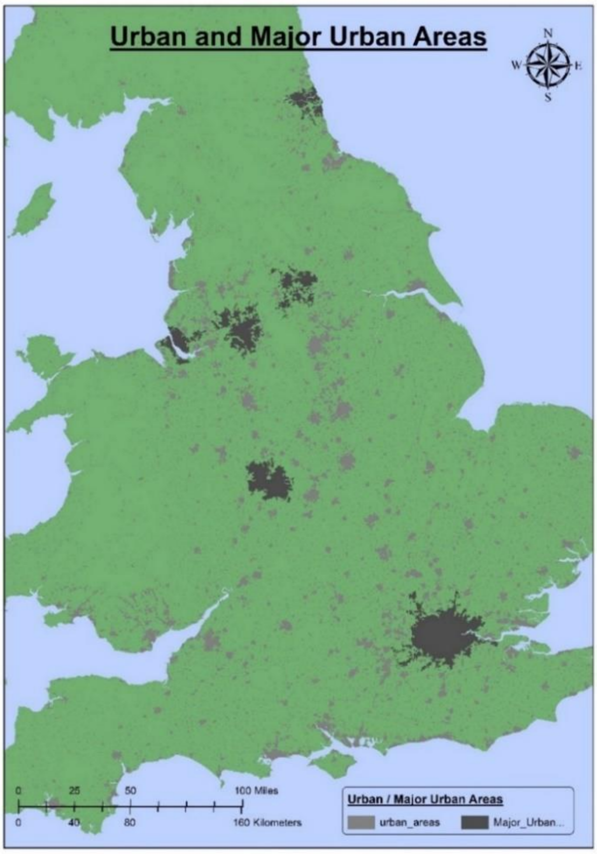
Figure 3-6: Train and light rail stations in the Greater London Area



v. Urban areas

Urban area polygons are spatial datasets derived from OS that designate the urban areas' boundaries in 2016, as these areas are defined by the Ministry of Housing Communities and Local Government, and Defra report (Bibby & Brindley, 2014). In Figure 3-7 all urban areas in England and Wales are shown. In addition, the six largest (i.e., major) urban areas in England and Wales are highlighted as defined by Pointer, (2005). These areas are: The Greater London, West Midlands (Birmingham, Wolverhampton, and Coventry), Greater Manchester, West Yorkshire (Leeds and Bradford), Tyneside (Newcastle and Sunderland) and Liverpool (also including Wirral and Knowsley) Urban Areas. These polygons are used to indicate whether a point is located in an urban or rural environment.

Figure 3-7: Urban and major urban areas in England and Wales

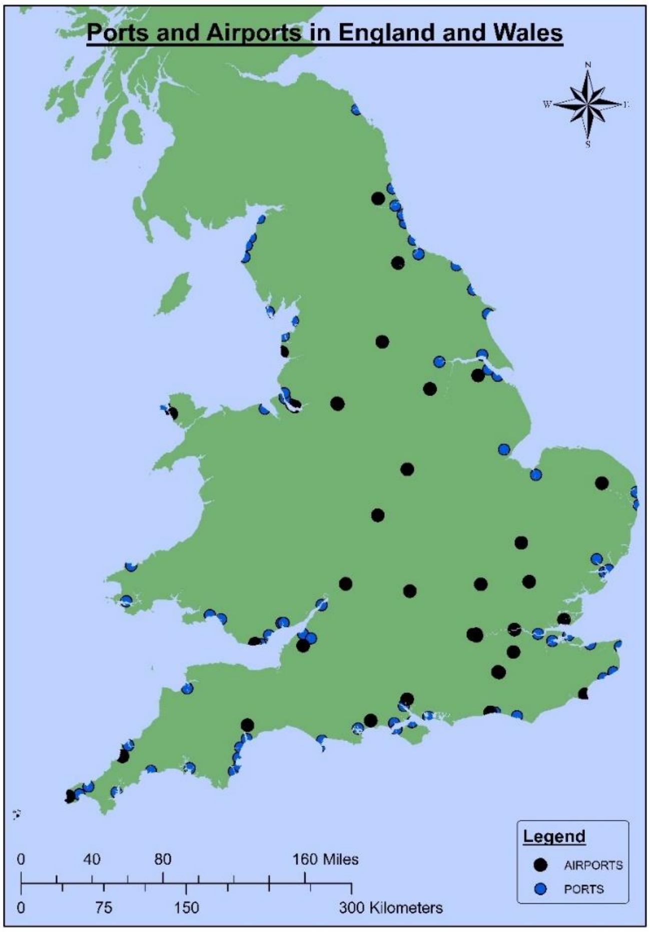


vi. Land use

Finally, land use data have been extracted from various sources. First, a list of rateable values for non-domestic properties in England and Wales as of 2017 is provided by the Valuation Office Agency (VOA).

The VOA dataset contains approximately 2.5 million records classified in over 100 classes based on a coding system, while addresses and postcodes for the properties are also included. This dataset had to be geocoded and existing categories were reclassified to 17 new classes, to reduce complexity. The new classes are shown in Table C-2 in the Appendix. Moreover, considering that ports and airports have an impact on the transport network of their surrounding area (Hesse, 2013), their locations as of 2015, are derived from the British Port Association and the Civil Aviation Authority respectively. Using the locational information ports and airports can be converted to spatial datasets and mapped as shown in Figure 3-8. Finally, electric vehicle charging point locations in 2016 are taken from OLEV. Considering location availability, the land use datasets can again be mapped. A summary of all used datasets is shown in Table C-1 in the Appendix.

Figure 3-8: Ports' and airports' locations in England and Wales



3.3. Methodology

The methodology to estimate AADT comprises of two major stages followed by a series of consecutive steps. Firstly, in subsection 3.3.1 the three major modelling approaches presented in section 2.4 are revised and the major issues and challenges for these approaches are discussed, so as to conclude on the most suitable approach for the research and justify the model selection. Then, in subsection 3.3.2.2 a detailed description of each step followed to develop a model for AADT estimation is provided. All the analysis in this and the following chapters has been conducted using ArcGIS (ArcMap and ArcCatalog) – version 10.4, QGIS Desktop – version 2.18.7 with GRASS 7.2.0 and RStudio – version 1.2.1335.

3.3.1. Modelling approach

From section 2.4, where the three main approaches for transport modelling have been presented, one can see that FSMs and ABMs are mainly simulation models attempting to explain and replicate transportation behaviour, either at aggregated (FSMs) or disaggregated levels (ABMs). These models almost exclusively utilise data extracted from surveys (Wang et al., 2016). However, big data for transport studies – such as smart card data – have been used in recent studies (e.g. Aslam and Cheng, 2018) and have provided researchers with additional sources and methods to study behaviour and estimate mode choice at zonal or individual level, offering larger samples at long observation periods at negligible cost (Anda et al., 2017). In addition, novel methodological approaches, such as sophisticated statistical models and machine learning algorithms, are nowadays applied for mode choice analysis (e.g. Bolbol et al. 2012, Sekhar et al. 2016) which are credited with reducing prediction errors compared to the traditional discrete choice models (Brathwaite et al., 2017; Paredes et al., 2017; Sekhar et al., 2016). In fact, FSM and ABM studies based on either travel surveys or even newer datasets have their own drawbacks. Firstly, studies based on sample datasets to draw inferences for the total population, may be susceptible to sampling bias[26] so that it is likely that errors generated during the first step of an FSM or the population synthesis of an ABM and propagated across the process. This is called by Castiglione et al. (2014) “aggregation bias”, related to the fact that individuals of the same group with similar characteristics are assumed to behave similarly, although this mainly applies to FSMs. Secondly, data requirements and computational costs to simulate travelling behaviour for a large area, such as a country, at a granular level are difficult to accommodate. Thirdly, both FSMs and ABMs include several assumptions about transport behaviours, such as household’s income, access to auto mode, and availability of alternative travel modes, which may fail to account for particular population groups and their daily travel needs (Nostikasari, 2015). ABMs also require input assumptions about sociodemographic and economic characteristics, multimodal transportation networks, and other key factors influencing travel behaviour (Castiglione et al., 2014). Fourthly, the use of utility maximising assumption incorporated in some models can also be unrealistic (Pinjari and Bhat, 2011). Finally, validation for FSMs and ABMs is an iterative process that would be best conducted after each step of FSMs and each output of ABMs, although data limitations sometimes prevent this with the implication that the accuracy of the data validation may suffer. In any case, this makes model validation complex and time consuming.

On the other hand, DDMs can address some of the limitations related to FSMs and ABMs. First, the DDMs’ aggregated nature and ability to model demand in a single equation imply lower information and data requirements (Cardozo et al., 2012). This approach also implies that data samples are not required, as estimation can be implemented on the whole population, so that one does not need to make any assumptions and extrapolation on the impact in the population based on observed samples. In addition, as the DDMs are statistical empirical approaches (Hankey et al., 2017) not aiming to simulate travel behaviours, model outputs can easily be validated against observed data. In fact,

[26] Sampling (or sample selection) bias is defined as the bias that can be evident if the individuals participating/sampled from the population are systematically different from those who do not participate (i.e., those who have not been included in the sample) (Cuddeback et al., 2004).

Horowitz (2006) states that aggregated DDMs are particularly useful when trying to predict a single quantity such as Vehicle Kilometres Travelled (VKT) or emissions in large geographic area such as a country, an argument that can partially be based on the lower data requirements and less complexity of DDMs compared to FSMs and ABMs. Finally, DDMs can also provide a very granular output, which allows for in depth analysis between the dependent and independent variables, so that one can capture dynamics missed by other models or more complicate effects such as self–selection[27] (Cervero, 2006) or provide a more direct measure of the impact of independent variables on the dependent (Cardozo et al., 2012).

However, DDMs have also been sometimes criticised due to the fact that they do not explicitly incorporate mode choice (Hancock, 2008). Nevertheless, for the aims of this research, DDMs are considered more suitable. Due to their statistical perspective and ability to provide a sensible relationship between traffic and its determinants, DDMs allow drawing inferences for the most significant drivers of traffic volumes (discussed in section 2.3), and associated emissions, in specific geographical areas. Consequently, by using relevant data and variables there is no need to make assumptions for travel behaviours and related characteristics; thus, providing a more realistic perspective on traffic volumes and produced pollutants. Finally, the performance of these statistical models can easily be assessed, by using validation metrics based on observed values (i.e., measured traffic volumes), allowing for precise estimation. Consequently, a statistical approach – rather than simulation – to model AADT is considered more suitable and is used in this chapter.

3.3.2. AADT modelling

To estimate AADT, three major steps are considered. First, the data described in section 3.2 are used to design the variables to be used as model inputs. All variables are designed using GIS. Second, the variables are fed into the selected algorithms and validation metrics to assess model's performance

[27] Self-selection is defined by Mokhtarian and Cao, (2008) as "the tendency of people to choose locations based on their travel abilities, needs and preferences" (Van Wee, 2009).

are used. Finally, the weighted average errors for each road type and across the road network are calculated.

3.3.2.1. Feature design

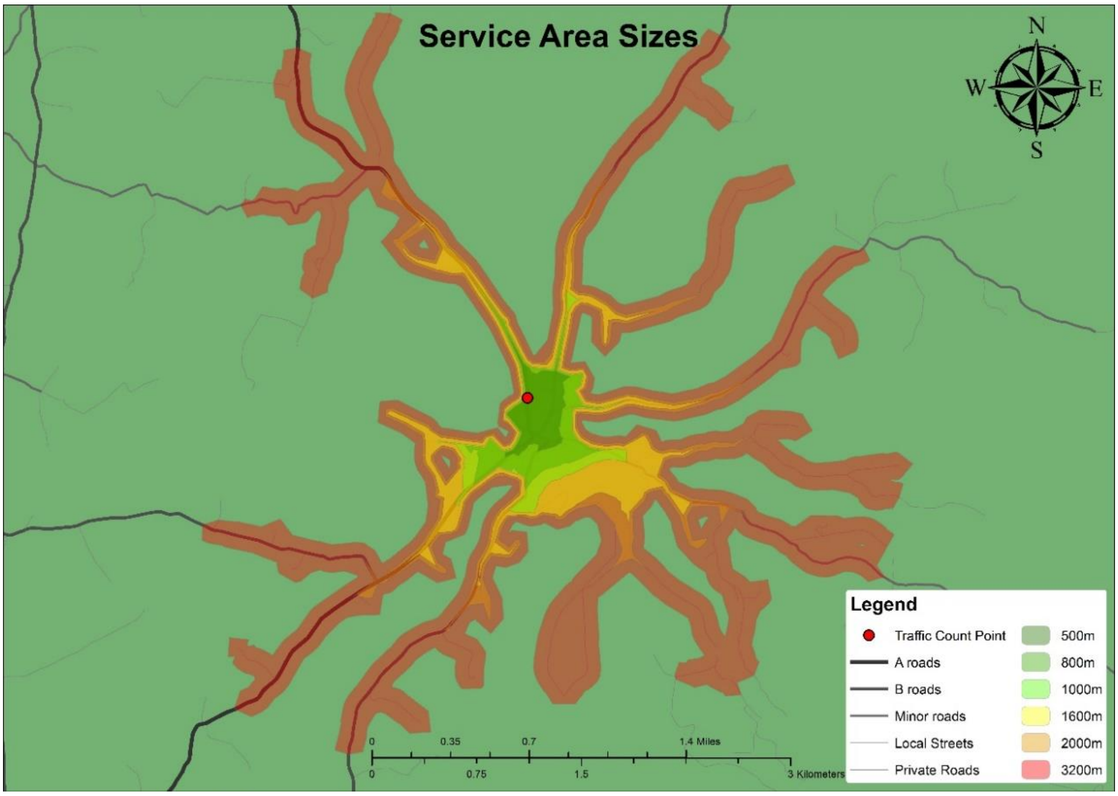
The initial step in the approach is to consider each point's spatial position and incorporate characteristics of the point's environment and location. In order to do this, the fact that urban areas generate and attract more activity and that the larger the area the more transport is generated (Caceres et al., 2018) has to be taken into account. However, one has to bear in mind that the urban areas dataset contains all build up areas whether they are large urban centres or small towns, likely to exhibit different traffic. Moreover, points marginally contained within or marginally excluded from the urban area polygons (Figure 3-7) have to also be taken into account. To address these three issues, it is firstly determined whether each point is located at either urban or rural environment and also four distance measures are calculated: (i) distance from urban area (ii) distance from major urban area (iii) distance from urban area centroid and (iv) distance from major urban area centroid. A centroid is defined as the geometric centre of each urban/major urban area polygon. This variable is designed to capture the distance of each point to the relative urban/major urban area centre (i.e., city/town centre) – likely to affect traffic. Distances to urban areas are calculated as straight lines (i.e., Euclidean distances) from each point to the nearest edge of the nearest urban/major urban area polygon, while for centroids, distances are calculated as straight lines from each point to the centroids.

In terms of roadway characteristics, two indicators for toll roads[28] and ring roads are introduced and also the "road nature" related to each count point is taken into account, which demonstrates whether a point is located on a single carriageway, dual carriageway, slip road or roundabout as indicated by OS, and either Trunk[29] or Principal road as indicated by the Department for Transport (2014).

[28] The "toll road" feature also includes the London Congestion Charge Zone, where all count points within the zone are considered to be toll roads.
[29] Trunk roads indicate long distance roads, usually connecting cities and having heavy traffic flows (Department for Transport, 2014)

In terms of variables reflecting the characteristics of an area, rather than a single point, the work of Koperski et al. (1998) based on the concept of service areas which are created around each point is followed. This is an improvement on the work of Zhao and Chung, (2001), Zhao and Park, (2004), Sarlas and Axhausen, (2014) and Doustmohammadi and Anderson, (2016) who use buffers of different radii around traffic count points. Service areas construct buffers by taking into account the street network instead of Euclidean distances. This measure is considered to be more suitable for this case study, since it can capture the actual predefined distance, a vehicle has to cover from/to the traffic count point. The service areas are of six different sizes (500m, 800m, 1000m, 1600m, 2000m and 3200m) for all road types as shown in Figure 3-9. The concept of service area is used in the case of land use, accessibility to motorways and some of the public transport characteristics. Service areas are overlaid with the VOA and charging points datasets as well as with the ports and airports datasets, to assess land use within each area. Accessibility to motorways which is associated with higher traffic volumes (Apronti et al., 2016; Zhao & Park, 2004), is also assessed by overlaying service areas with motorway junctions. Bus stops and bus stations are treated the same way.

Figure 3-9: Count point service areas



Finally, the socioeconomic characteristics – already available at LSOA level as described in section 3.2 – are taken into account as are train and light rail stations. For the latter the ITNUP[30] dataset is firstly utilized and 800 metres service areas around each train station are created[31]. When ITNUP is not available (e.g., if a train station is located at any rural area), the ITN is used instead. Then, the proportion of each LSOA covered by station service areas is calculated, so as a station accessibility attribute is also available at LSOA level. Lastly, the count point service areas with LSOAs are overlayed so as to consider all intersecting LSOAs with each service area and introduce socioeconomic and station accessibility characteristics for each count point. Specifically, the mean values for station accessibility, population density, workplace density, income and workplace plus population density are incorporated, the last variable being used in Fu et al. (2017). In addition, the summed values of population, workplace population, number of households and registered vehicles are also calculated. The feature design process generates 41 independent (33 numerical - 8 categorical) and the dependent variable (AADT). The variables are summarised in Table 3-2.

Table 3-2: Independent variables

Variable	Description	Type
1. Urban/Rural	A count point's surrounding environment	Categorical
2. Distance to Urban Area	The straight Euclidean distance from a count point to an urban area polygon edge	Numerical
3. Distance to Major Urban Area	The straight Euclidean distance from a count point to a Major urban area polygon edge	Numerical
4. Distance to Urban Area Centroid	The straight Euclidean distance from a count point to the geometrical centre of an urban area polygon	Numerical
5. Distance to Major Urban Area Centroid	The straight Euclidean distance from a count point to the geometrical centre of a major urban area polygon	Numerical
6. Toll Road	Whether or not the count point is located at a toll road	Categorical
7. Ring Road	Whether or not the count point is located on a ring road	Categorical
8. Road Nature	Whether the count point lies on a single or dual carriageway, slip road or roundabout	Categorical
9. Road Category	Whether the count point lies on a Primary or Trunk Road	Categorical
10. Junction Accessibility	Whether the road where the count point is located has access to a motorway based on the specified service area	Categorical

[30] Notice that the use of ITNUP indicates that access to train stations can be by foot as well, using the footpaths, subways, steps, footbridges, and cycle paths, although this data is available for urban areas only.
[31] The 800 metres threshold is considered as the standard distance one would consider walking to reach a station in most research (e.g. Cardozo et al. 2012, Gutiérrez et al. 2011).

11. Charging Points	The number of charging points within each service area	Numerical
12. Ports	Whether the road where the count point is located has access to a port based on the specified service area	Categorical
13. Airports	Whether the road where the count point is located has access to an airport based on the specified service area	Categorical
14. Bus Stops	The number of bus stops within each service area	Numerical
15. Bus Stations	The number of bus stations within each service area	Numerical
16. Train Accessibility	The adjacent LSOAs' average train station coverage	Numerical
17. Population	The total population of a count point's intersecting LSOAs	Numerical
18. Population Density	The average population density of a count point's intersecting LSOAs	Numerical
19. Workplace Population	The total workplace population of a count point's intersecting LSOAs	Numerical
20. Workplace Population Density	The average workplace population density of a count point's intersecting LSOAs	Numerical
21. Workplace plus Population Density	The average workplace plus population density of a count point's intersecting LSOAs	Numerical
22. Income	The average median income of a count point's intersecting LSOAs	Numerical
23. Households	The total number of households of a count point's intersecting LSOAs	Numerical
24. Registered Vehicles	The total number of registered cars and vans of a count point's intersecting LSOAs	Numerical
25-41. VOA (17 features – Table C-2)	The total number of VOA elements in the predefined count point's service area	Numerical

3.3.2.2. Traffic volume (AADT) estimation

Considering the large geographic extent and mixed characteristics, it is expected that AADT values and other variables to exhibit large variations across the study area. For example large differences are expected between urban and rural areas (Morley & Gulliver, 2016). For this reason, a clustering algorithm to take into account (dis)similarities among count points and their surroundings and group points with similar characteristics is firstly applied. Then, three models, namely standard multivariate linear regression, Random Forests (RF) and Support Vector Regression (SVR) within each cluster are applied and each model's accuracy is assessed by using validation metrics.

In order for the models to be comparable based on the selected validation metrics, all the designed features are fed to all models without undertaking further statistical tests (e.g., checking for collinearity or feature importance). That is, if one model is able to automatically handle complexities within the dataset, it is considered as an asset of the particular model. The process is applied for each road class ('A', 'B', 'C' and 'U')[32] and each service area size individually. Finally, for each road class the service area where the algorithm resulted into the lowest errors is selected and the selected points are merged to construct the full dataset, so as the optimal service area size for each road class and point location can be detected. That allows to identify the optimal distance where the traffic on a particular road is influenced by its surroundings.

3.3.2.2.1. Clustering

For the clustering stage, the K-prototypes (Huang, 1998) algorithm, suggested by He (2006) is used, which integrates the K-means and K-modes processes for numeric and categorical data respectively (Huang, 1997a) to cluster mixed type data[33] – see list of variables in Table 3-2. K-prototypes, instead of taking samples from the dataset, uses the whole dataset and thus it does not suffer from sampling bias, and it is less computationally intensive compared to K-medoids or various Hierarchical Clustering algorithms that can handle mixed variable types. For numerical variables, K-prototypes uses squared Euclidean distances as in K-means, while for categorical variables, the dissimilarity measure is defined by the total mismatches of the attribute categories of two objects (Huang, 1998) so that the overall distance metric is equal to the squared Euclidean distance to measure (dis)similarity for numerical variables and the matching (dis)similarity for the categorical variables,

$$d(X,Y) = \sum_{j=1}^{m_r} \left(x_j - y_j\right)^2 + \gamma \sum_{j=1}^{m_c} \delta\left(x_j, y_j\right) \tag{1}$$

[32] Again, motorways are excluded at this stage of the analysis considering that traffic on these roads is not directly affected from its surrounding characteristics (Eom et al., 2006; Sun and Das, 2019; Zhao and Chung, 2001).

[33] This choice has been dictated by the fact that most clustering algorithms, for example the K-means, do not take into account categorical data, as based on the Euclidean distance. Alternatives such as the chi-square (Greenacre & Primicerio, 2015) have been found to perform poorly (Faith et al., 1987; McCune & Grace, 2002). Kaufman & Rousseeuw (1990) advocates the use of the K-medoids algorithm incorporating the Gower's similarity coefficient (Gower, 1971) although the computational cost when using this type of similarity metric increases significantly and it is therefore unsuitable for large datasets (Huang, 1998).

where X and Y are the two mixed-type objects for each point j, m_r and m_c are the numbers of numeric and categorical attributes respectively and γ is a weight to avoid favouring either type of attribute (Huang, 1997b). δ indicates the dissimilarity (mismatches) for the categorical variables, where:

$$\delta(x_j, y_j) = \begin{cases} 0 \ (x_j = y_j) \\ 1 \ (x_j \neq y_j) \end{cases} \tag{2}$$

Moreover, the data are also transformed to address the problem of different measurement units and ranges. Data transformations make features dimensionless to overcome the problems resulting from the dependence on different measurement units and the deviations among variable variances that affect cluster quality and formations (Rokach & Maimon, 2015; Zhang et al., 2019) so that each variable can play an equal role in the analysis (Greenacre & Primicerio, 2015; Han et al., 2012; Mohamad & Usman, 2013). Large variable range tend to have large effect on the resulting clustering structure (Kaufman and Rousseeuw, 1990; Mohamad and Usman, 2013). As variable measurement units and their respective ranges play a significant role in the cluster formations, methodological guidance on the use of transformation is very clear-cut in the literature, as applying data transformation is considered essential for most practical applications to enhance performance (Bishop, 2013). In particular, numerical variables should be transformed to scale their effect on the results (Larose, 2005) and conventional distance measures (e.g. Euclidean) should not be used without applying transformations on the data (Mohamad and Usman, 2013).

In terms of the specific transformation, the most common form of normalisation – the min-max normalisation – is applied, which sets all variables within the range of 0 to 1 based on:

$$x' = \frac{x_i - \min{(x)}}{\max(x) - \min(x)} \tag{3}$$

where i indicates the data instance and $\min{(x)}$ and $\max{(x)}$ are the minimum and maximum values of each variable, respectively. It is also taken into account that the parameters thought to be more relevant in separating the groups should be assigned a higher influence factor (Hastie et al., 2009), –

i.e. weight[34] – to raise their importance of certain variables which are considered more critical in cluster formation (Gebotys & Elmasry, 1989; Hummel et al., 2017). Weights can be assigned by multiplying the variables with a constant (Akhanli & Hennig, 2017; Hammah & Curran, 1999). In this case, the scope is to form clusters where AADT values are similar and independent variables are relatively correlated with the dependent variable (i.e., AADT) within the same cluster, to achieve accurate predictions. The upper goal is for the dependent variable to have a high enough weight (range) to influence the formation of the cluster, although without dominating it. Considering that 41 independent and 1 dependent features of different types and ranges are available, applying the K-prototypes algorithm without transforming the data, results into clusters dominated by the independent variables only, while the same output is observed when all variables have equal influence. On the other hand, transforming only the predictors, results into clusters dominated by AADT values, since the ranges are extremely different. Hence, the work of Bacher et al. (2004) who apply random lower weights to variables separating the clusters to achieve equal influence and similarly, Opsahl & Panzarasa (2009) who also assign random weights between 1 and 10 to links (edges) on their work on clustering networks is followed. That is, the variable ranges are changed and weights of 1 and 10 to the independent and dependent variables are assigned respectively[35]. This is achieved by implementing a generalised version of the min-max standardisation above which can be used to transform a range of values into another $[\alpha, \beta]$, i.e.

$$x' = (\beta - \alpha)\frac{x - \min(x)}{\max(x) - \min(x)} + a \qquad (4)^{36}$$

[34] The weights can be unequal among the variables to define their influence (Friedman & Meulman, 2004) and can also be zero if they do not possess any important information (Hammah & Curran, 1999).

[35] I acknowledge that some variables do not directly affect all vehicles; hence their contribution to AADT may be questionable – e.g., charging points are only useful for electric vehicles. However, in this stage I do not examine the contribution of each variable to different types of vehicles, but to AADT for all motorised vehicles. Moreover, in this chapter I focus on AADT estimation and model comparison, thus I have chosen to give all independent variables the same weight, so as to be able to draw rational inferences when comparing models.

[36] As it can be seen, the required ranges are set by applying data transformations and consequently weighting is achieved without multiplying by a constant.

with $\alpha = 1$ and $\beta = 10$ the required range can be obtained. In the case of the AADT, values within the range of 1 to 10 are set, while the value of 0 for the dependent variable is not considered, since there are no observations with zero value. Regarding the choice of weights on the variables, more information is provided in Appendix C.

As the K-prototypes algorithm requires defining the number of clusters (K) before clustering is implemented, the "elbow" method which is considered as the optimal since it is the only one considering mixed data types[37] is employed. The elbow method examines the percentage of variance as a function of the number of clusters (Bholowalia & Kumar, 2014), the idea being that starting with K=2 and increasing the number of clusters, at some point the marginal gain drops dramatically and gives an angle in the graph (Kodinariya & Makwana, 2013) indicating the optimal K. When testing 20 clustering processes (i.e., $K = 2, 3, \ldots, 20$), for each of the 4 road classes and 6 service area sizes, the optimal number for K ranged between 4 and 6 depending on the case examined each time. For simplicity, five clusters for all cases are selected, e.g., Figure C-1 in the Appendix.

3.3.2.2.2. AADT estimation

To estimate AADT in each cluster, the dataset is firstly randomly split into two groups, 80% of the observations for training and 20% for testing. The training dataset is used to implement three different models, namely (i) standard multivariate linear regression (OLS), (ii) Random Forest (RF) and (iii) Support Vector Regression (SVR).

The multivariate linear regression model is as follows:

$$AADT_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_j x_{ij} + \varepsilon_i \tag{5}$$

where: $AADT_i$ is the dependent variable at the ith observation, $i = 1, \ldots, n$, x_{ij} is the value of the jth independent variable in the ith observation, $j = 1, \ldots, m$, β_0 is a constant term, β_j is the regression

[37] Other methods include the "Silhouette" method (Rousseeuw, 1987), the Calinsky – Herabasz Criterion (Calinski and Harabasz, 1974), Bayesian Information Criterion – BIC (Schwarz, 1978) and Akaike Information Criterion – AIC (Akaike, 1974) among others.

coefficient for the jth independent variable, ε_i is the error term and m is the number of independent variables.

Random Forest (RF) is a machine learning technique, used both for classification and regression modelling (Strecht et al., 2015), introduced by Breiman (2001). RF is a collection of decision trees, an example of so-called ensemble methods, based on bootstrapping (Efron, 1979) and bootstrap aggregation (Breiman, 1996). The RF regression prediction is given by:

$$\widehat{f_{rf}^{B}} = \frac{1}{B}\sum_{b=1}^{B} T_b(x) \tag{6}$$

where: B is the number of trees and $T_b(x)$ is the b^{th} random forest tree grown from b bootstrapped data. Here, 500 trees and 5 variables for the forest to sample at each split are used.

Finally, Support Vector Regression (SVR) is the extension of Support Vector Machine (SVM) classifier (Cortes & Vapnik, 1995) proposed by Drucker et al. (1997). SVR aims to find a function $f(x)$ where predicted values are at most ε from the observed ones. The general SVR equation for non-linear predictions is given by Basak et al., (2007):

$$f(x) = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*)\, k\langle x_i, x\rangle + b \tag{7}$$

where: α_i, α_i^* are the Lagrange multipliers for each data instance i, $k\langle x_i, x\rangle$ is the kernel[38] and b is the bias. Here the radial basis Kernel is used and by replacing in (7) the equation becomes:

$$y = \sum_{i=1}^{N} (a_i - a_i^*) * \exp\left(-\gamma \left\| x_i - x_j \right\|^2\right) \tag{8}$$

where: $\gamma = \frac{1}{2\sigma^2}$ and set to 0.1, and σ is the standard deviation.

[38] The kernel refers to a function that maps data from one space to another higher dimensional feature space (Hastie et al., 2009)

3.3.2.2.3. Validation

Prediction accuracy is validated using the test set comprising 20% of the dataset. Two validation measures are used, the Mean Absolute Percentage Error (MAPE):

$$MAPE = \frac{100}{n}\sum_{t=1}^{n}\left|\frac{A_i - F_i}{A_i}\right| \tag{9}$$

where: A_i is the observed value, F_i is the predicted value and n is the number of observations, and the Root Mean Square Error:

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(F_i - A_i)^2}{n}} \tag{10}$$

3.3.2.2.4. Weighted average

Weighted average is calculated on the lowest identified MAPEs for each model j across clusters i for each road class c,

$$WMAPE_{j,c} = \sum_{i=1}^{K}\left(\frac{AADT_{i,c}}{\sum AADT_{i,c}}\right) * MAPE_{i,j,c} \tag{11}$$

where: $WMAPE_{j,c}$ is the weighted average MAPE for model j in road class c, $AADT_{i,c}$ is the total traffic volume for cluster i at road class c, $MAPE_{i,j,c}$ is the MAPE for cluster i, model j and road class c and K is the number of clusters. Then, the overall weighted average MAPE across road types for the whole road network for each model j, is similarly calculated:

$$OWMAPE_j = \sum\left(\frac{AADT_c}{\sum AADT_c}\right) * WMAPE_{j,c} \tag{12}$$

where: $AADT_c$ is the total traffic counted for road class c. Similarly, the weighted values for the corresponding RMSEs are also calculated.

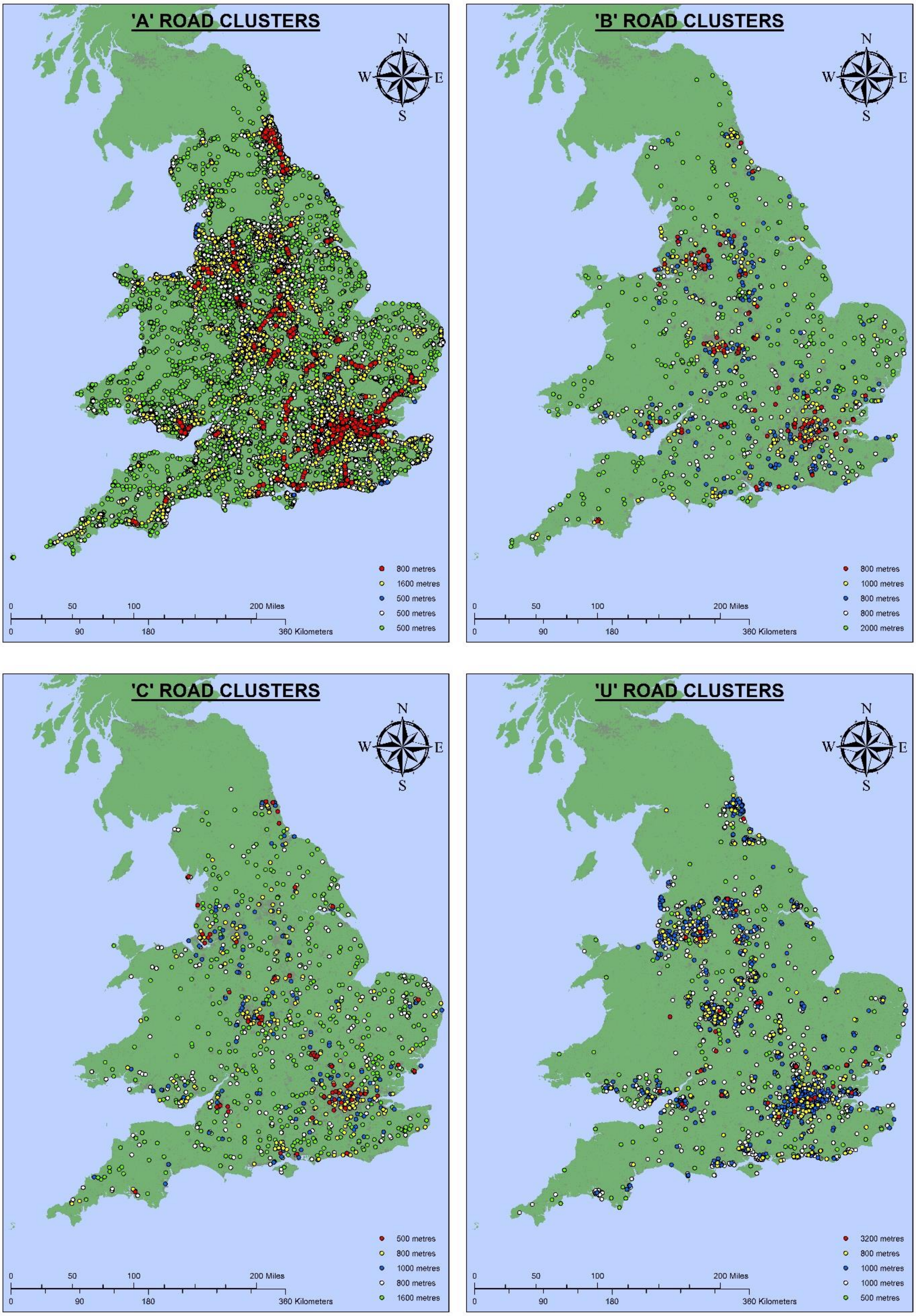
3.4. Results

Based on the evaluation metrics (i.e., MAPE and RMSE) used to assess the performance of each regression model (i.e., RF, SVR and OLS), in this section the results for the clusters and corresponding service areas where higher estimation accuracy is achieved are presented.

As the first result from this chapter, it is interesting to comment on the estimated clusters, as they exhibit similar patterns across road types. This is shown in Figure 3-10, where the clusters and related optimal service area sizes are colour coded. In particular, for each of the four road types, i.e. 'A', 'B', 'C' and 'U':

- **Cluster 1 (red)** contains points located on roads where traffic counts tend to be higher, such as ring and trunk roads in the case of 'A' road class and evenly split between urban and rural areas. For 'B', 'C' and 'U' roads points are placed at locations of higher transport significance, almost exclusively located in urban areas.
- **Cluster 2 (yellow)** includes relatively high traffic values with points in 'A' roads located both in urban and rural environments, while for other road types, this cluster is mainly formed by points within urban locations.
- **Cluster 3 (blue)** consists of medium AADT values with points for all road types located within urban areas, mainly concentrated in city centres. In particular, 'A' road points are observed within designated major urban areas as well as the city centres of some medium and small urban centres.
- **Cluster 4 (white)** also contains medium AADT, although usually with smaller values than those in cluster 3. These points are mainly located in suburban areas of large urban centres as well, but also in the centre of smaller settlements. Some of the points are also observed in rural areas, especially in the case of lower-class roads.
- Finally, **cluster 5 (green)** contains the lowest AADT values which are normally located in rural areas and the outskirts of urban centres.

Figure 3-10: Clusters for 'A' (top left), 'B' (top right), 'C' (bottom left) and 'U' (bottom right) roads



Moreover, this work casts light on the performance of different methods across the five clusters formed for each of the four road types. Figure 3-11 to Figure 3-14 display both the MAPEs and RMSEs for the 120 combinations of clusters and road types for each of the three methods implemented in this chapter.

Figure 3-11: MAPE (top) and RMSE (bottom) for 'A' roads

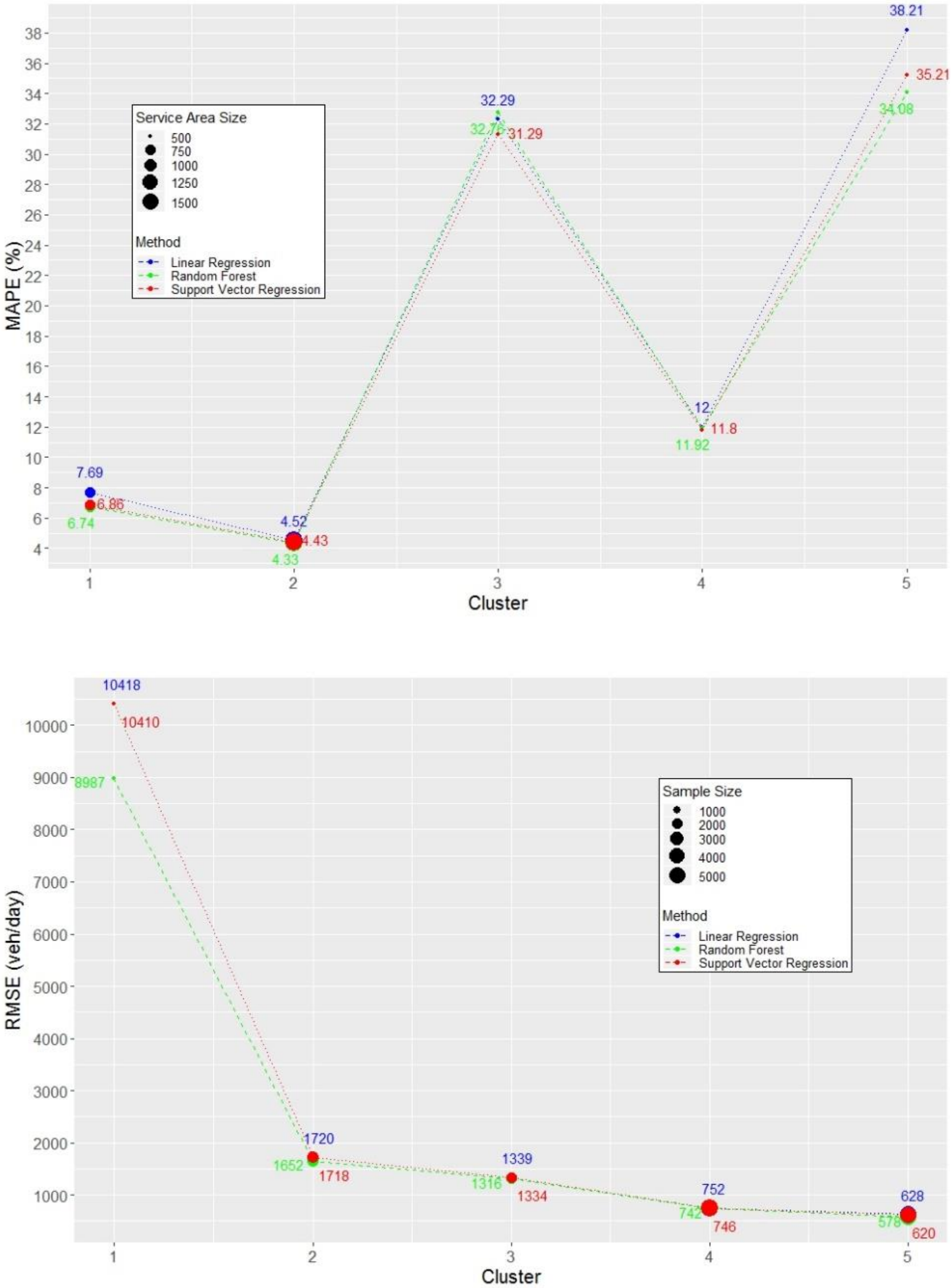


Figure 3-12: MAPE (top) and RMSE (bottom) for 'B' roads

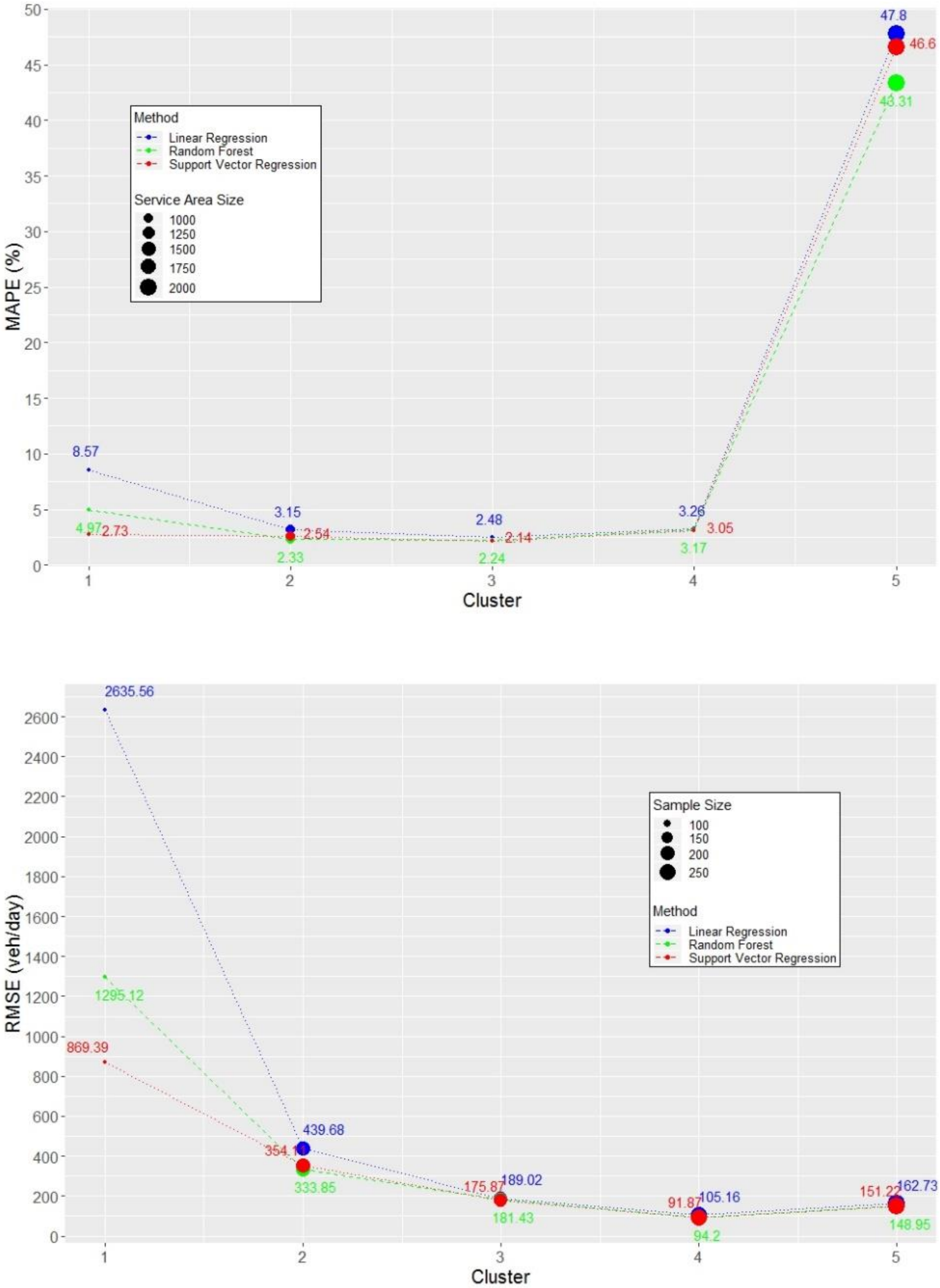


Figure 3-13: MAPE (top) and RMSE (bottom) for 'C' roads

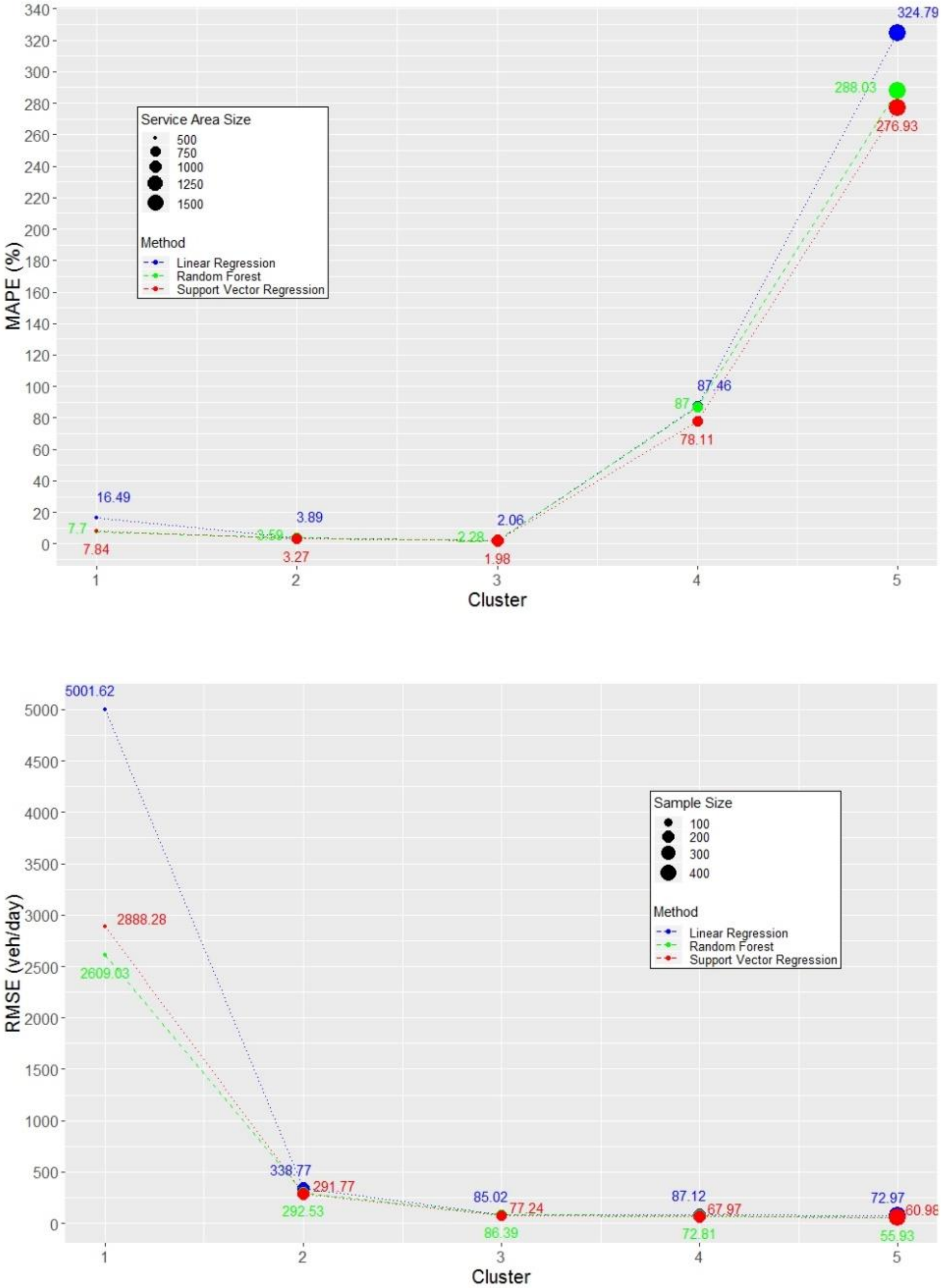
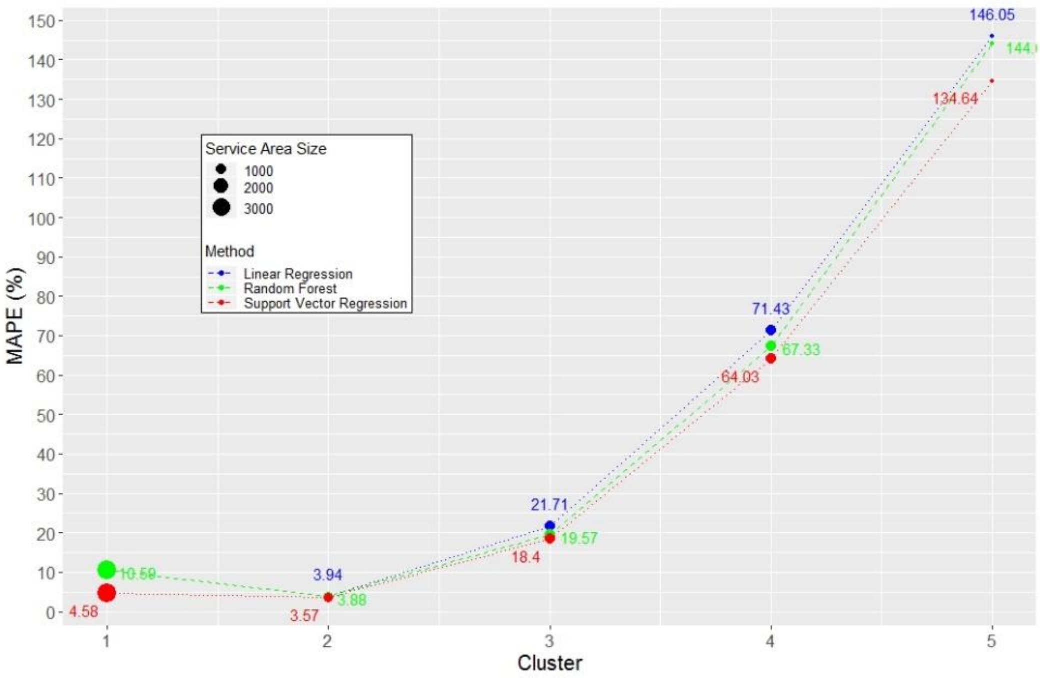
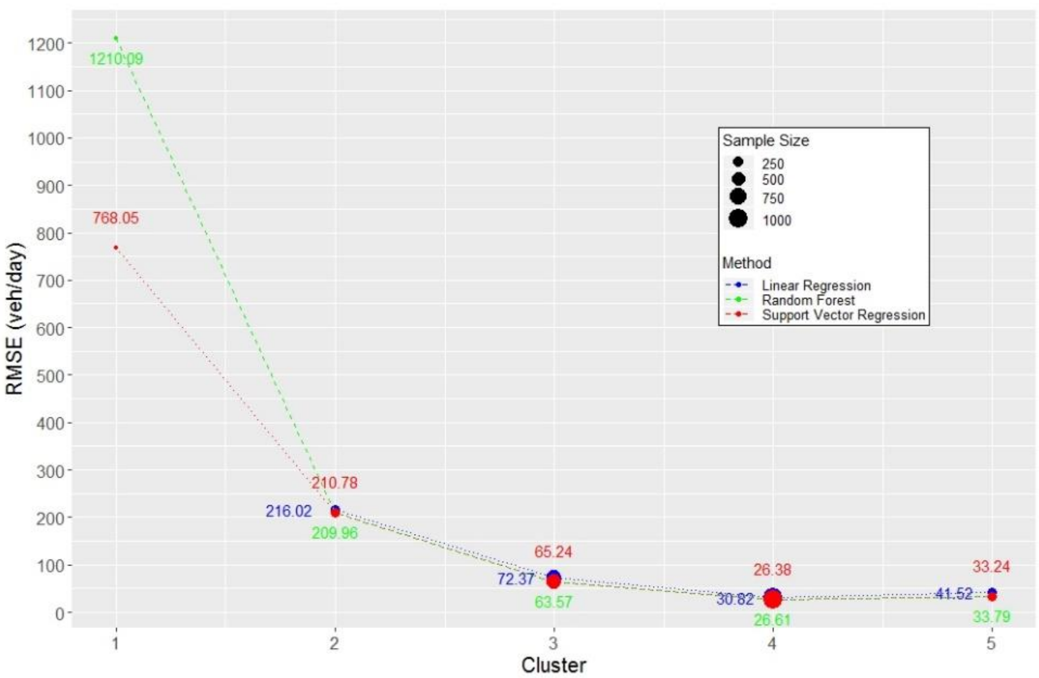


Figure 3-14: MAPE (top) and RMSE (bottom) for 'U' roads





As one can see in Figure 3-11 to Figure 3-14 and Table 3-3, the two Machine Learning methods are fairly equivalent and outperform the regression method. In the case of the SVR, the MAPE ranges between 2% (cluster 3 in 'C' roads) and 276.9% (cluster 5 in C roads) while the MAPEs achieved by RF range between 2.2% (cluster 3 in 'B' roads) and 288% (cluster 5 in 'C' roads). Among the three methodologies implemented, linear Regression exhibits the highest MAPEs in almost all cases, with values falling between 2.1% (cluster 3 in ‘C’ roads) and 324.8% (cluster 5 in ‘C’ roads). Linear Regression also produce a very big error in cluster 1 of ‘U’ roads, probably due to the very small number of observations in this cluster (31 points – 25 for training and 6 for testing). Considering this result unreliable, it is excluded from Figure 3-14, although it is interesting to see that the SVR performs well also in this case despite the very small sample. Similar conclusion emerges when assessing the performance of the methodologies implemented in this study based on the RMSEs, therefore adding robustness to the conclusion that the two Machine Learning methods are fairly equivalent and outperform the regression method (Table 3-3). It is worth mentioning however that RF produces lower RMSE than SVR in the case of cluster 1 – ‘A’ roads which is by far the combination with the highest level of traffic (Table C-3). Linear Regression continues to produce the largest errors and again results into very large error in cluster 1 at ‘U’ roads (not shown in Figure 3-14).

One can also appreciate from Figure 3-11 to Figure 3-14 that the predicting patterns, as measured by the MAPE and RMSE are similar for all models, with higher MAPEs usually observed in cluster 5 and higher RMSEs in cluster 1 across road types. This is however a simple reflection of the fact that cluster 5 comprises observations with relatively low traffic which translates in higher MAPE, while cluster 1 contains cases with high level of traffic so that the RMSE (which tends to be influenced by the level of the observations) is corresponding high. The range of the RMSE values across clusters make comparison difficult – as an example it goes from about 5,000 in the case of cluster 1 to as a low as 55 in the case of cluster 5 in C roads. The relatively small values for most combinations of clusters and roads in Figure 3-11 to Figure 3-14 shows that the 3 methods produce similar results when measured in terms of vehicles per day, in a way that they may all satisfy users’ needs unless they focus on specific types of roads and traffic volumes for which a specific method can work better than another.

This is confirmed by Table 3-3, which presents MAPEs and RMSEs, first averaged across clusters and eventually across road types to obtain an overall MAPE and an overall RMSE for each model. Traffic volumes – presented in Table C-3 – are used throughout as weights in the averaging process. One can see that MAPE is highest for ‘C’ and ‘U’ roads and smaller for ‘A’ and ‘B’ roads with the lowest ones observed at ‘B’ roads. Moreover, one can see more clearly that SVR is the best performing model when measured based on the weighted MAPE, while regression has the highest MAPE for all road types. SVR again outperforms RF in ‘B’, ‘C’ and ‘U’ roads, with the MAPE of SVR being 0.2% lower than RF in ‘B’ roads and increasing at ‘C’ and ‘U’ roads respectively, e.g., 27% versus 29.3% in the case of ‘U’ roads. For ‘A’ roads, however, the performance of SVR and RF is essentially identical and the gap of these two methods with the linear regression shrinks to 0.8 percent, as is the overall MAPE with SVR performing slightly better than RF but only by 0.01%.

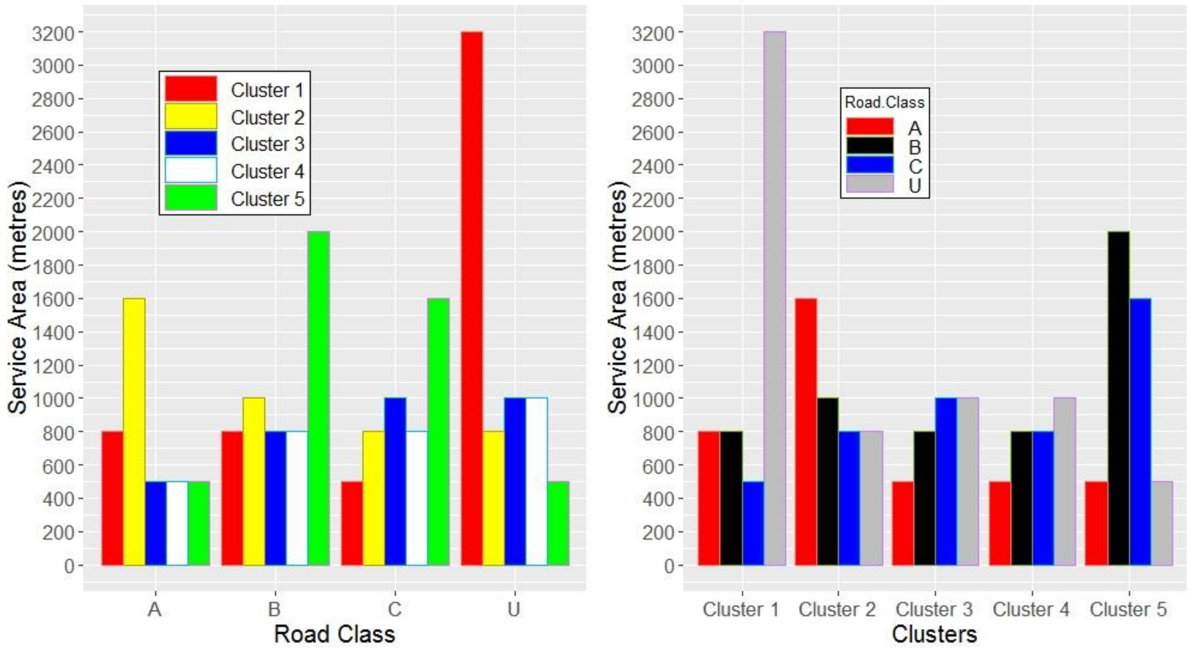
In terms of RMSEs, it can be seen that the errors are higher for higher class roads and decrease for the lower-class roads as expected. This same expected pattern is also observed within the clusters of each road class for the unweighted errors as shown in Figure 3-11 to Figure 3-14 and Table 3-3. However, RMSE values are lower in ‘B’ compared to ‘C’ roads where AADT values are usually lower as shown in Figure 3-2. The averaged RMSEs show that errors are again higher for Linear Regression and are also balanced between SVR and RF, with RF resulting to lower errors half of the time. However, observed differences are small and the mean difference between RF and SVR is 217 vehicles across all road types in favour of RF.

Table 3-3: Original and Weighted MAPEs and RMSEs

Road Class	Cluster	Service Area	MAPE (%)			RMSE (vehicles per day)			Number of points
			OLS	RF	SVR	OLS	RF	SVR	
A Roads	1 – red	800	7.7%	6.7%	6.9%	10,418	8,987	10,410	521
	2 – yellow	1600	4.5%	4.3%	4.4%	1,720	1,652	1,718	2,170
	3 – blue	500	32.3%	32.8%	31.3%	1,339	1,316	1,334	1,672
	4 – white	500	12.0%	11.9%	11.8%	752	742	746	5,627
	5 – green	500	38.2%	34.1%	35.2%	628	578	620	4,680
	Weighted		**15.2%**	**14.4%**	**14.4%**	**2,429**	**2,193**	**2,423**	14,670
B Roads	1 – red	800	8.6%	5.0%	2.7%	2,636	1,295	869	86
	2 – yellow	1000	3.2%	2.3%	2.5%	440	334	376	216
	3 – blue	800	2.5%	2.2%	2.1%	207	166	177	184
	4 – white	800	3.3%	3.2%	3.1%	105	94	92	252
	5 – green	2000	47.8%	43.3%	46.6%	163	149	151	284
	Weighted		**7.4%**	**5.9%**	**5. 7%**	**815**	**471**	**381**	1,022
C Roads	1 – red	500	16.5%	7.7%	7.8%	5,002	2,609	2,888	59
	2 – yellow	800	3.9%	3.6%	3.3%	339	293	292	207
	3 – blue	1000	2.1%	2.3%	2.0%	85	86	77	147
	4 – white	800	87.5%	87.0%	78.1%	87	73	68	218
	5 – green	1600	324.8%	288.0%	276.9%	73	56	61	427
	Weighted		**37.1%**	**31.8%**	**30.3%**	**1,408**	**796**	**863**	1,058
U Roads	1 – red	3200	368.7%	10.6%	4.6%	58,650	1,210	768	31
	2 – yellow	800	3.9%	3.9%	3.6%	216	210	211	187
	3 – blue	1000	21.7%	19.6%	18.4%	72	64	65	557
	4 – white	1000	71.4%	67.3%	64.0%	31	27	26	1,070
	5 – green	500	146.1%	144.0%	134.6%	42	34	33	196
	Weighted		**75.2%**	**29.3%**	**27.0%**	**7,327**	**235**	**181**	2,041
ALL ROADS	**Overall Weighted**		**15.69%**	**14.48%**	**14.47%**	**2,413**	**2,119**	**2,336**	**18,791**

Finally, the pattern of the optimal service areas, i.e., the area producing the lowest MAPE, across clusters and road types, as shown in Figure 3-15 can be elaborated. Here, a clear pattern is evident for road classes 'B' and 'C', where the service areas are small and similar for clusters 1 to 4 and increase at cluster 5. On the contrary, the optimal service areas for 'U' roads follow the opposite pattern starting at large service area for cluster 1 and gradually decreasing to reach the minimum (500 metres) for cluster 5. Service areas for 'A' roads are of medium size and also minimise for clusters 3, 4 and 5. Moreover, it can be observed that small to medium service areas dominate the figures, with only two large service areas observed at 'U' roads cluster 1 (3200 metres) and 'B' roads cluster 5 (2000 metres).

Figure 3-15: Optimal Service Areas by road class (left) and cluster (right)



In addition, one can observe that clusters 3 and 4 fluctuate around small to medium service area sizes and tend to increase as the road class decreases in significance (from 'A' to 'U'), in contrast with cluster 2 where the service area decreases together with the respective road class. Cluster 1 exhibits an increasing pattern and cluster 5 – representing the "rural" areas – is optimised at small service areas for road classes 'A' and 'U' and at higher ones for 'B' and 'C' roads.

3.5. Discussion

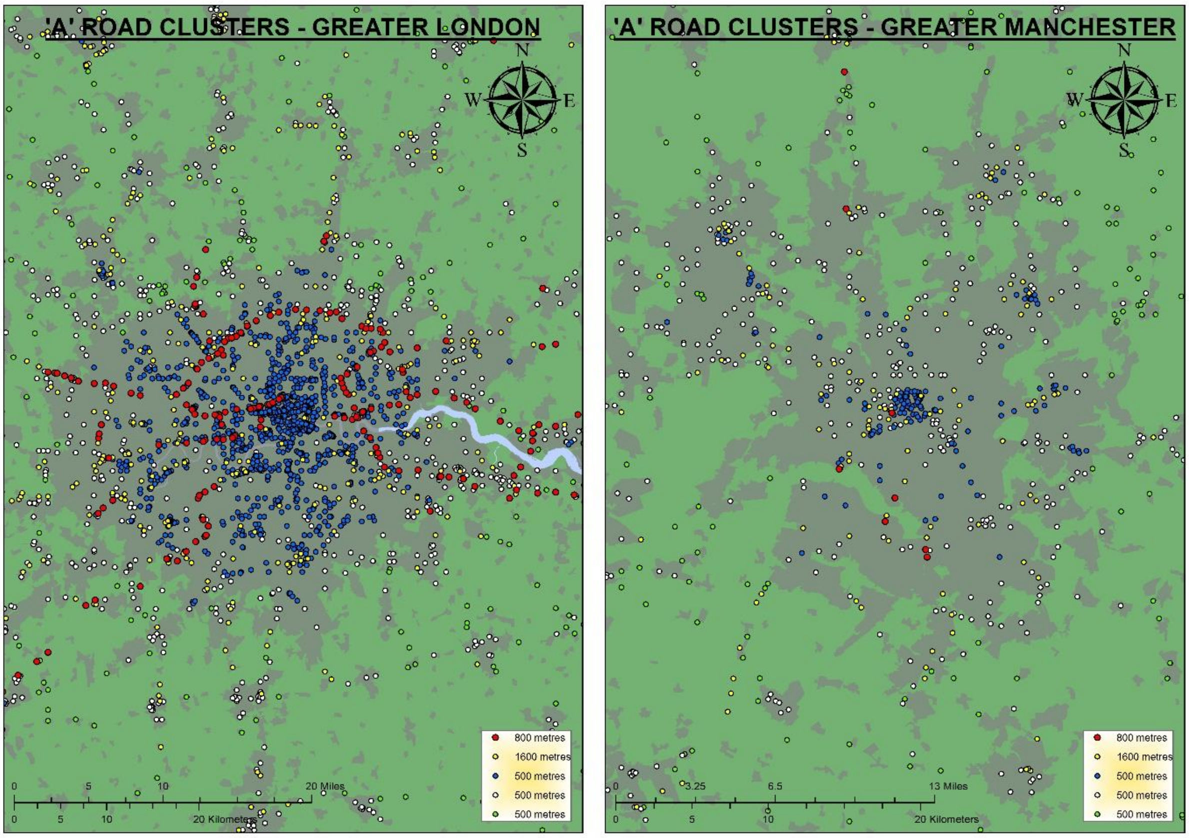
This chapter has focused on the development of a methodology to estimate AADT at locations where traffic counters are not available. The procedure has been applied to AADT figures collected from DfT for all available road classes in England and Wales from this dataset, therefore providing a rigorous and comprehensive test of the process outlined in this chapter. Several variables postulated to affect traffic flows were firstly included, based on results from previous studies in the literature, as inputs into predictive models. These variables portray a detailed representation of roadway, land use, socioeconomic and public transport characteristics. Specifically, utilisation and manipulation of spatial data within GIS, facilitated feature design and the analysis, so as to incorporate related socioeconomic, land use and roadway attributes – used as AADT predictors – which are directly associated with the count points' spatial locations.

The output from the models has been assessed using statistical validation metrics normally employed in the literature, in particular MAPE and RMSE. As the focus of this approach in this chapter is on estimation, the metrics above were computed on the test dataset, i.e., 20% of the sample, that were available. The fact that the choices conform to the standard in the literature both in terms of inputs and metrics to assess the output make the results compelling, as high accuracy of the AADT predictions obtaining out-of-sample MAPEs as low as 2% can be delivered. This contrast with the majority of results arising from the applications in the literature, where in some cases lowest errors are 50% for similar road types (e.g. Selby & Kockelman 2013) or ranging between 39% and 400% in others (e.g., Wang et al., 2013).

The significant improvement in accuracy can be attributed to two interrelated aspects of this approach: data transformation and clustering. First of all, the clustering algorithm revealed groups where data exhibit similar characteristics while the application of data transformations allowed the clustering algorithm to create groups where both similar AADT values and related characteristics have been taken into account. This can be concluded both from section 3.4 where the clusters are presented and even more so from Figure 3-16 where points in city centres are clustered together, indicating areas with similar characteristics (e.g., a large number of shops and businesses) and picking up underlying roads[39].

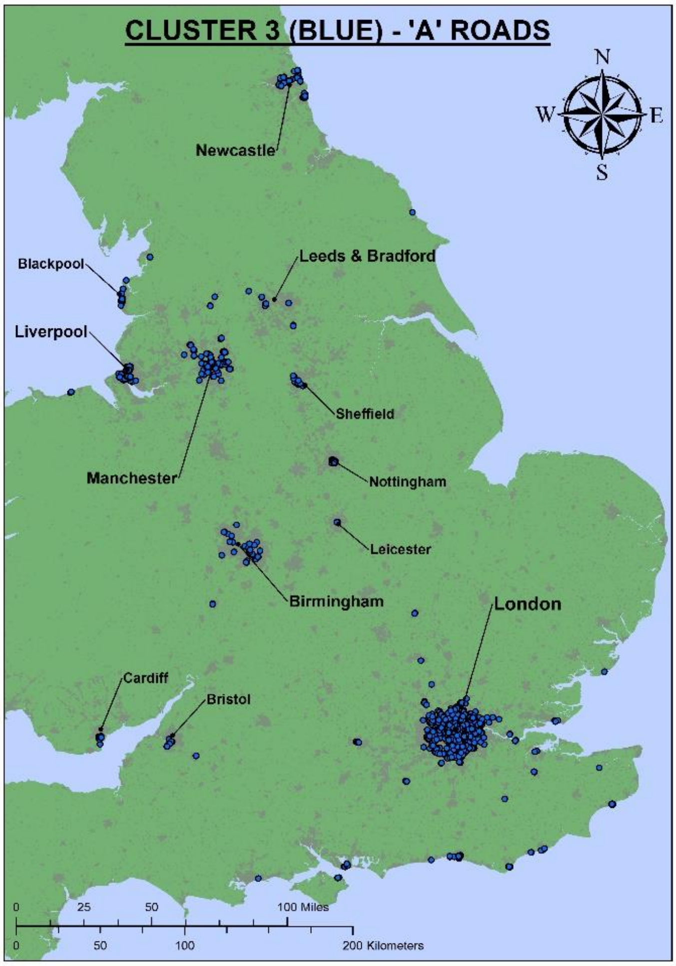
[39] In the case of roads, the north circular (the ring road in north London) is clearly visible among the red dots.

Figure 3-16: A road clusters for Greater London (left) and Greater Manchester (right)



However, error deviations among the models, clusters and road types presented in Table 3-3, show that the models' performance – in terms of MAPE – is dependent on two conditions. First is the value of the dependent variable (i.e., the amount of traffic per traffic count point) within each cluster, where high MAPEs are observed for clusters with low AADT values – usually clusters 4 and 5 – in most cases. Nonetheless, this expected outcome is due to the fact that the estimated variable can have values very close to zero (Caceres et al., 2018) and consequently even slight deviations would exaggerate the error. For example, a misprediction of 10 vehicles would have a different impact on MAPE for an observed value of 100 compared to an observed value of 100,000. However, the exception of the unexpectedly high MAPE in cluster 3 (blue) for 'A' roads, can be due to the characteristics of the areas where the points are located. As it is shown in more detail in Figure 3-16 and Figure 3-17 the points tend to be at city centres of large and major urban areas, usually associated with diverse land use and complexity. Thus, this cluster could be further disaggregated so that patterns not currently identified could be revealed, having the potential to improve accuracy (Greenacre & Primicerio, 2015).

Figure 3-17: Cluster 3 - 'A' roads



Second condition to affect models' performance is the number of data instances (i.e., sample) within each cluster, particularly in the case of Linear Regression. Specifically, Linear Regression results into very high MAPE for the smallest sample across the data set (31 points at cluster 1 – 'U' roads) and also is over 9% higher compared to RF and SVR for the second smallest sample (59 points at cluster 1 – 'C' roads) and 3.5% - 6% higher for the third smallest sample (86 points at cluster 1 – 'B' roads). Sample effect is also noticeable at the RMSEs, where Linear regression again produces very large error in cluster 1 at 'U' roads, while all models also result into high RMSEs at cluster 1 at 'C' roads even compared with cluster 1 – 'B' roads where traffic counts are higher, as shown in Table 3-3.

It is important to mention that sample size affects overall model performance. For example, models perform similarly – and potentially more accurately – in the case of 'A' roads, including most of the traffic count points comprised in the sample (i.e., approximately 15,000 out of 19,000 points). As ‘A’ roads account for over 95% of the total traffic in the database – see Table C-3 – it turns out that the overall MAPE is fairly similar to the MAPE for ‘A’ roads, to the great benefit of linear regression in terms of comparison across methods. That is, if one is to take into account the traffic values estimated by DfT (Table C-4) it appears that ‘A’ roads account for 57%, while ‘C’ and ‘U’ roads combined – where errors are higher – account for 34% of the total traffic. Consequently, MAPEs weighted based on Table C-4 results in only 8% higher error for OLS compared to RF (27.2% versus 19.2%) and 8.6% higher error compared to SVR (27.2% versus 18.6%). This leads to the conclusion that again SVR performs better than RF and Linear Regression performance is overstated by the sample.

As a final remark, I look into cluster 1 at ‘A’ roads. From Figure 3-18 it can be observed that the points clustered here are mainly associated with ring roads (e.g., north circular in London – also in Figure 3-16), motorway extensions (e.g., part of A23 from Crawley to Brighton – Figure 3-19) as well as roads connecting urban centres – usually trunk roads – such as the A19 (Figure 3-19). Moreover, 96% of these points in this cluster are dual carriageways, 15% ring roads and over 10% have access to motorways within 800 metres. In addition, from Table C-3 it can be seen that points in this cluster have an average of approximately 75,000 vehicles per count point, while for motorways (not included in this stage of the analysis) there are approximately 74,000 vehicles per point. This leads to the conclusion that count points included in this cluster can potentially have strong similarities with motorways, indicating that traffic on these roads is not necessarily related with the road’s immediate surrounding environment.

Figure 3-18: Cluster 1 - 'A' Roads

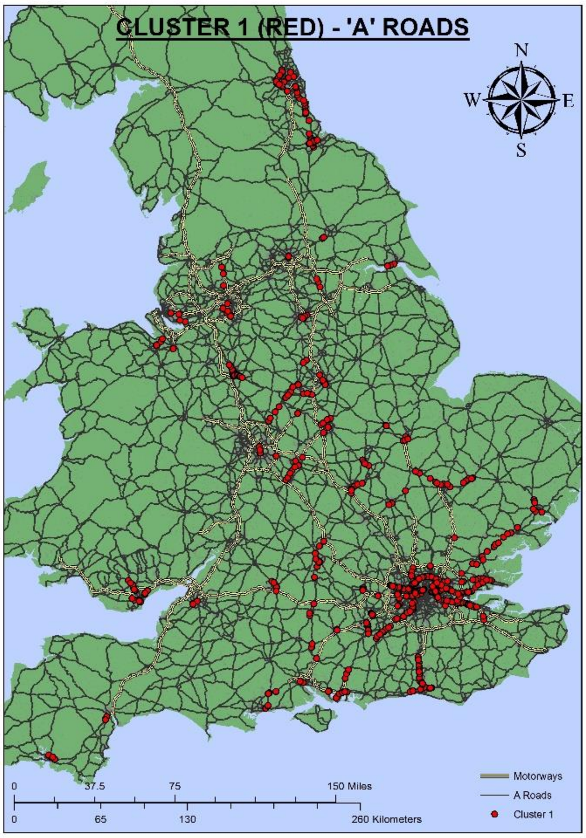
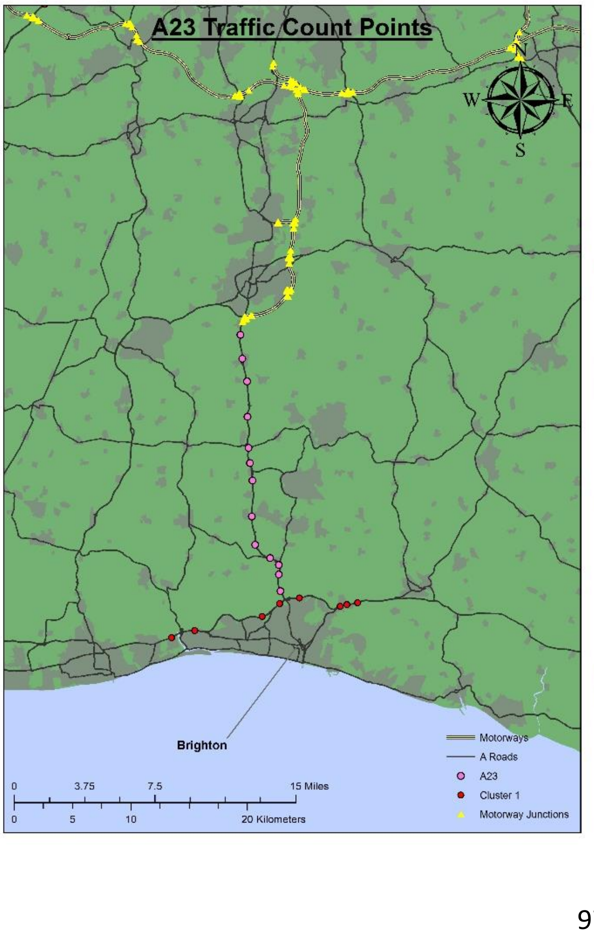
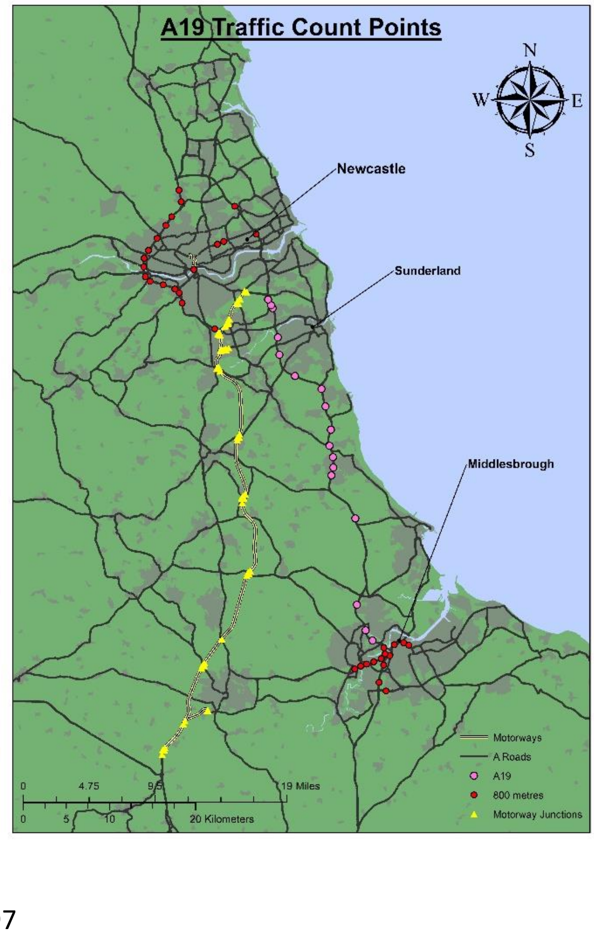


Figure 3-19: A23 (left) and A19 (right) roads in cluster 1





3.6. Chapter summary

In this chapter two main topics that form the foundation of this thesis have been covered, and by doing addressing some of the road transport modelling limitations identified in the literature. First, a comprehensive set of drivers of road traffic volumes based on the factors discussed in the reviewed literature has been created. Second, a hybrid – clustering-regression – model to estimate AADT for all road classes has been developed and validated.

Based on the aims of this research, the developed model can be used to inform the selection of the emission modelling approach. Moreover, considering the AADT estimation accuracy achieved, it is safe to assume that emissions can also be accurately estimated. On the other hand, the proposed methodology exhibits several limitations that need to be addressed before modelling emissions. These are identified and discussed at a later stage of the thesis (chapter 5).

As a further limitation, the methodology presented in this chapter does not assess or discuss the impacts of the identified factors on road transport, since focus has been placed on estimation accuracy. However, the dataset created in this chapter and the model outcomes can be utilised to investigate the effects of the variables on AADT, as discussed in the following chapter.

4. Determinants of road traffic volume

4.1. Chapter overview

AADT modelling presented in chapter 3 has focused on the accurate estimation of traffic volumes for all vehicles across different road classes. The model presented, although it can be useful for numerous applications in road transport studies – such as accident prediction, noise exposure and emissions estimation – it is not informative in terms of the precise effect the identified drivers (Table 3-2) have on AADT. This is fundamental not only to facilitate traffic volume estimation, but also to examine the complexity of road transport and how it interrelates with urban – and where possible rural – infrastructure and demographics. The latter is vital for decision making in the transport field, but also across wide range of interconnected sectors such as urban and environmental planning and of course the economy.

In this chapter, the association of driving factors on traffic volumes (i.e., AADT) is investigated, aiming to understand the complexity of road transport and address the first aim of the thesis – to identify the degree of influence specific factors have on traffic flow variations across the road network. The analysis is again focused on the four different road classes as classified by Department for Transport, (2014) and are examined individually ('A', 'B', 'C' and 'U' roads)[40]. The analysis is conducted for five different vehicle types[41] where a statistical model is applied, and the most statistically significant variables are identified so that their impact on AADT can be assessed.

The chapter is presented in six sections. Section 4.2 presents the dataset used and section 4.3 describes the methodology applied to understand the impact of various factors on AADT. In section 4.4, the results for each road class and vehicle type are presented, and in section 4.5 the findings for all road classes are analysed and discussed. Finally, section 4.6 summarises the chapter.

[40] As a reminder, motorways are not examined due to the fact that traffic on these roads is not affected from the surrounding characteristics (Eom et al., 2006; Sun and Das, 2019; Zhao and Chung, 2001).
[41] Cars, buses, Light Good Vehicles, Heavy Good Vehicles, and two-wheeled vehicles.

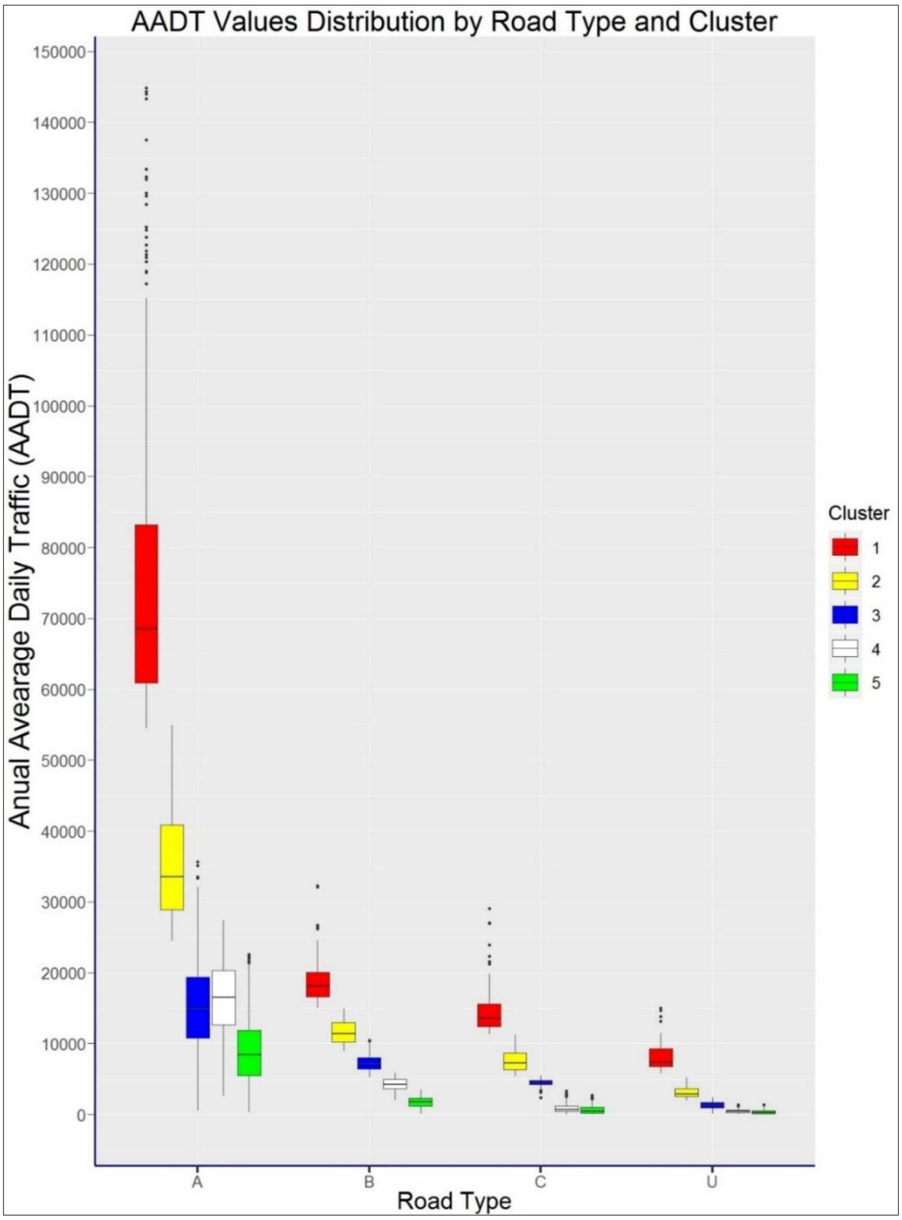
4.2. Data

In this chapter, the dataset constructed in chapter 3 is used, as presented in subsection 3.3.2.1. However, for each road type and count point, instead of looking at total AADT values, the traffic volumes for each of the five different vehicle types – i.e., Cars, Buses, Light Good Vehicles (LGVs), Heavy Good Vehicles (HGVs) and two-wheeled – as classified by DfT are taken into account. One has to recall that the count points are partitioned in five different clusters and therefore, are characterised by the variables where in some cases exhibit considerable variation. Consequently, the distribution of the variables is not similar across the clusters and some variables have zero values in some clusters. However, a generalised pattern related to traffic volumes emerged – as shown in section 3.4 – with AADT decrease starting from cluster 1 towards cluster 5 for all road types (Figure 4-1). Moreover, the clusters appear to be significantly affected by the location of traffic counters, being placed in urban, suburban or rural areas (Table 4-1). The independent variables are presented in Table D-1 in the Appendix.

Table 4-1: Traffic counters' locations and service areas for each cluster

Group Number	Colour Code	Road Class				Location
		'A'	'B'	'C'	'U'	
		Service Areas Size (in metres) / Number of Points				
1	Red	800	800	500	3200	'A' roads – evenly split between urban and rural areas 'B','C' and 'U' roads – mainly in urban areas
		521	86	59	31	
2	Yellow	1600	1000	800	800	'A' roads – evenly split between urban and rural areas 'B','C' and 'U' roads – predominantly in urban areas
		2,170	216	207	187	
3	Blue	500	800	1000	1000	'A' roads – predominantly in major urban areas and centres of smaller urban 'B' roads – evenly split between urban and rural 'C' and 'U' roads – predominantly urban
		1,672	184	147	557	
4	White	500	800	800	1000	'A' and 'U' roads – predominantly in urban areas 'B' and 'C' roads – split between urban and rural (many in the centres of smaller settlements as well as outskirts and suburbs of large urban centres)
		5,627	252	218	1,070	
5	Green	500	2000	1600	500	Almost exclusively rural. Some points for 'U' roads are located at smaller settlements.
		4,680	284	427	196	
Total Number of Points		14,670	1,022	1,058	2,041	18,791
Proportion of Points in each road class		78.07%	5.44%	5.63%	10.86%	

Figure 4-1: Total AADT Values by Road Class and Cluster



4.3. Methodology

As discussed in section 2.3 to understand the impacts of the identified factors on AADT the employment of statistical or behavioural models is required, so as quantitative or qualitative analysis can be conducted. In chapter 3, a statistical approach to model AADT based on numerous variables (i.e., features) has been applied, where it is found that ML models perform better compared to the traditional OLS statistical approach. However, although ML algorithms have proved to provide more accurate results, these models may lack strong theoretical basis and are sometimes considered 'blackboxes' (Adler et al., 2018; Krause et al., 2018; Steindl and Pfeiffer, 2017), which means there is a lack of knowledge in the internal process (Burns et al., 2020; Lakkaraju et al., 2017) making them unsuitable for interpreting the outcomes (Brathwaite et al., 2017). This is reflected both in AADT estimation studies (e.g. Das and Tsapakis, 2019) and other disciplines (e.g. Churpek et al., 2017; Zhang et al., 2018), where ML models provide more accurate results compared to other models, but the exact impact of predictors on the response variable cannot be assessed and interpreted.

Hence, to understand the impact of the identified factors on AADT, the coefficients obtained from statistical modelling are examined. This is done by first transforming the data in a form enabling comparison of coefficients across variables and then, by applying a regression model and extract the statistically significant coefficients. From Table D-1 in the Appendix, one can see that numerical variables are mainly count variables although a number of continuous variables measured in different units are also included – e.g., distances (in meters) and income (in British pounds – GBP). To allow comparison of the coefficients, all variables are standardised, using what is sometimes called the z-transformation. Any count or continuous variable, x, regardless of the distribution, can be transformed into a variable with mean of 0 and a standard deviation of 1, given that they have finite mean μ and standard deviation (SD) σ :

$$x' = \frac{x - \mu}{\sigma} \tag{13}$$

This transformation will allow to compare and assess the impact of each – independent – variable on the dependent – i.e., AADT (Diez et al., 2012).

Moreover, logarithmic transformation on the dependent variable is also applied to address the fact that the distribution of the dependent variable is skewed to the right for all roads and subgroups (Zoico et al., 2010). Therefore, the coefficients of predictors on the dependent variable as percentage impacts (Troy et al., 2012), rather than absolute units, can be interpreted.

To model the relationship between the independent and dependent variables, the Least Absolute Shrinkage and Selection Operator (LASSO) method (Tibshirani, 1996) is used. Lasso is both a regression and variable selection method, that aims to produce a set of statistically significant predictors to minimise the estimation error (Musoro et al., 2014). This is done by imposing a penalty term on the model parameters so that some regression coefficients shrink to zero (Ogutu et al., 2012) and consequently these variables are excluded. The so-called L1 penalty is applied by using the parameter λ controlling the shrinkage level, so that the set of coefficients estimated by Lasso minimises the expression in the curly braces below (Hastie et al., 2009):

$$\beta^{lasso} = argmin\left\{\frac{1}{2}\sum_{i=1}^{N}\left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j\right)^2 + \lambda\sum_{j=1}^{p}\left|\beta_j\right|\right\} \tag{14}$$

where y_i and x_i are the outcome and predictor variables respectively and β represents the coefficients.

The parameter λ is key, as changing its values can affect the number of chosen variables and estimated coefficients. Specifically, the larger the value of λ, the greater the level of shrinkage and the fewer variables retained by Lasso. Therefore, for each λ, the k-fold cross validation method (e.g. Melkumova and Shatskikh, 2017; Yang and Zou, 2015) is applied to select the optimal value. During this process, the sample is randomly split into k partitions. The k-1 subsets are used to train the model and the remaining subset is used to test how well the model fit the data by computing the cross-validation error. The process is repeated with a different subsample and the error is computed for each iteration. The value providing the lowest error is selected (Moreno-Torres et al., 2012; Wong, 2015).

However, to account for randomness in the subset selection, a repeated k-fold cross-validation is applied. This process replicates k-fold cross-validation multiple times, with the data being rearranged for each iteration (Yadav and Shukla, 2016). Here k=10 is used and tested on 100 repetitions. The average value for optimal lambda across the repetitions is calculated and used as the parameter in Lasso.

4.4. Results

In this section, results are presented for each road class and vehicle type separately. The results from Lasso include only a subset of variables, as some of the variables have zero coefficient. In some cases, this happens for a great share of the variables, while in other cases most variables are retained, and in those instances, only those having a high impact on AADT are discussed. Moreover, variables with coefficients exhibiting similar patterns across all groups are identified and presented, in addition to those with high positive or negative coefficients. It is worth pointing out that the meaning of coefficients is different across variable types. In the case of continuous variables, the estimated coefficients represent percentage changes in traffic volumes arising from an increase of 1 standard deviation (SD) in the variable. In the case of categorical variables, the coefficients represent percentage change when switching from the base category to the other category as seen in Table D-2 in the Appendix, showing the results for all road classes. In this section, the Lasso regression outcomes for all roads are presented. The complete set of results can be seen in Table D-3 to Table D-10. Discussion and interpretation of all the findings are presented in section 4.5.

4.4.1. 'A' roads

In this road type it can be observed that in the case of roads with high traffic volumes (e.g., cluster 1 and cluster 2), Lasso returned a set with relatively few variables. Similarly, a relatively small number of variables is selected for high volumes vehicle types (e.g., cars) compared to vehicles with lower volume. Moreover, the categorical variables have a significant impact across all groups and vehicle types (Table D-3 and Table D-4).

4.4.1.1. Cars

In the case of cars, the effect of some variables varies significantly across clusters. For example, the number of households and the number of registered vehicles exhibit high positive coefficients in cluster 3 and high negative in cluster 5. On the other hand, some **categorical** variables have similar major impact on car volumes across almost all clusters. Dual carriageways are correlated with high volumes in most clusters, compared to single carriageways, indicating an increase ranging from 17% in cluster 3 to 49% in cluster 5. Ring roads have a higher level of traffic, up to 20% in clusters 3 and 4, while Road category is also statistically significant, particularly in the case of low volume clusters such as 4 and 5. Compared to primary rural roads, primary urban roads have lower traffic volumes up to 30%. Trunk roads, on the other hand, are related to higher traffic volumes, up to 12%. Several categorical variables have a strong effect only in some clusters, while in other clusters they are dropped. This occurs in the case of access to motorways, associated with high traffic volumes only in cluster 5 by 30% compared to roads with no access. Occurrence of ports is significantly negatively correlated with car volumes in clusters 3 and 5, by 49% and 36%, respectively. Similarly, traffic volumes are lower in urban areas in the case of cluster 4 (by 14%) and 5 (by 26%). In some cases, correlation of a certain factor with traffic volumes is high but does not manifest itself in the same direction across clusters. For example, traffic volumes are 27% lower in the case of toll roads in cluster 3 but 33% higher in cluster 5.

The **distances** from urban and major urban areas are statistically significant on most clusters, having high signs in low volume clusters. It can be observed that the further a count point is from an urban or major urban area the lower the car traffic volumes with values as high as 42% in cluster 5, considering all other factors being equal. Distances to these areas' centres have the opposite effect, indicating a potential increase in traffic volumes with values reaching 66% in cluster 5. Two **land use** variables – petrol stations and sporting facilities – have positive coefficients in all clusters indicating a potential increase on traffic volumes as high as 5%. The number of shops is correlated with lower car volumes by 11% in cluster 3. In terms of **socioeconomic** variables, it can be observed that population exhibits high positive coefficients in clusters 3 and 5 (17% and 9%), while population and workplace population densities usually have a positive sign ranging from 1% to 5% for most clusters. Income also has a positive effect across all clusters reaching 13% in cluster 5. For **public transport**, train station accessibility has a minor negative sign in clusters 1, 2 and 5 and positive for clusters 3 and 4 up to 6%, while the number of bus stops is associated with lower car volumes across all clusters although there is a positive sign of 5% on cluster 5.

4.4.1.2. Buses

In the case of buses, among the **categorical** variables, existence of dual carriageway correlates with higher bus volumes from 6% to 32% in all clusters except cluster 2, while fewer buses are expected in ring roads, where the there is a negative correlation by -11% in cluster 1 to as much as -30% in cluster 4. Road category tends to have a strong positive correlation in clusters with high traffic volume while it tends to have a negative coefficient in those with low volumes. Specifically, a very high positive coefficient (64%) is observed in the case primary urban roads in cluster 1, while the same category exhibits a negative (-20%) in cluster 5 where counters are located mainly in rural areas. Among the categorical variables which have an effect only in some clusters, access to motorways is associated with higher traffic by 11% and the road being in urban areas with lower traffic volumes by 33% in cluster 5.

The set of **distances** from urban and major urban areas and the respective centres have mixed signs on bus traffic. In the case of distance to major urban areas a potential increase by 44% on traffic volumes is indicated, although the further apart from major urban areas centres there is a negative sing indicating a potential decrease up to 25%. In terms of **land use** variables, negative signs are observed for variables supporting the use of private vehicles, such as vehicle facilities and parking variables, each being related to a 10% reduction in cluster 3. There are however instances where the impact of these variables is positive, e.g., vehicle facilities in cluster 1. Among the **socioeconomic** variables, workplace population and population densities have positive signs in almost all cases up to 17% in cluster 3. Population has a positive sign ranging from 16% to 60% while number of registered vehicles has a negative up to 21% in cluster 4. Among the factors related to **public transport**, the number of bus stops is associated with higher bus traffic between 8% in cluster 1 and 33% in cluster

5. Train station accessibility has a positive coefficient only in the case of cluster 3, indicating higher bus traffic by 13%.

4.4.1.3. Light Good Vehicles (LGVs) and Heavy Good Vehicles (HGVs)

LGVs and HGVs are overall affected by similar variables, with estimated coefficients having the same sign, most likely because these vehicles are used for similar purposes, i.e., goods transport. In the case of **categorical** variables, ring roads exhibit high positive signs ranging from 6% to 37% and are normally higher for HGVs. Also, trunk roads are correlated to more LGVs and HGVs, up to 22% for LGVs and 112% for HGVs compared to primary rural roads. Dual carriageways are associated with high traffic volumes as opposed to single carriageways, up to 44% for LGVs and 76% for HGVs, while access to motorways is related to 27% and 50% increase respectively, particularly in low volume clusters. Presence of ports has high positive impact on HGVs traffic, up to 55%, although small negative effects are observed for LGVs. **Distance** to urban areas has negative signs in most cases, ranging from -0.13% up to -47% for LGVs and -53% for HGVs, although distance to urban city centres has positive signs up to 86% and 95% for LGVs and HGVs respectively. In terms of **socioeconomic** variables, workplace population is related to higher LGVs and HGVs volumes in almost all clusters up to 5% for LGVs and 21% for HGVs. For the **public transport** variables, bus stops have negative signs on both vehicles in most clusters with the coefficients ranging from -2% to -6% for LGVs and up to -17% for HGVs.

4.4.1.4. Two wheeled vehicles

Similar to other vehicles, the **categorical** variable related to dual carriageways correlates with significantly increased two-wheeled vehicle volumes up to 32%, with the highest increase occurring in cluster 3. Among the **distance** variables, high positive coefficients (up to 83%) are observed in relation to distance to city centres, while negative signs up to -45% can be seen for the distances to urban area boundaries, particularly in the case of major cities. For the **socioeconomic** characteristics, income has high positive coefficients across all clusters ranging from 11% in group 5 to 81% in cluster 3. Similarly, the number of households is associated with increased volumes from 4% in cluster 5, to 24% in cluster

2. Population density is also associated with higher volumes – about 35% for the mainly 'urban' cluster and the high-traffic cluster. High negative signs are evident for the registered vehicles variable reaching -21% in cluster 3.

4.4.2. 'B' roads

For 'B' roads, there are fewer instances of coefficients taking high values regardless of the sign across clusters of roads, compared to what was discussed in the case of 'A 'roads, indicating that different characteristics correlate to traffic volumes across different road clusters. Only a few variables are observed to be highly correlated with traffic volumes in the case of buses and two-wheeled vehicles, where patterns similar to 'A' roads are observed (Table D-5 and Table D-6).

4.4.2.1. Cars

In the case of cars, the association between traffic volumes and the majority of statistically significant variables is negligible, while only a small number of variables exhibit a relatively high positive or high negative correlation. In particular, in the case of **categorical** variables, one can observe that the location of points in urban areas (compared to the rural benchmark) is significant only in cluster 3, with a negative coefficient of -7% for counters in urban areas – indicating that urban 'B' roads are related with less car volumes compared to rural. In the case of **distances** to urban areas and urban centres negative coefficients are observed at -3% and -2% for clusters 1 and 3 respectively. However, a few **socioeconomic** variables are statistically significant and are sometimes highly correlated with car traffic volume. As an example, income has a positive coefficient of 8% in cluster 5, but only 2% cluster 1 and a negative coefficient of -1% in cluster 3. Workplace population has a positive coefficient of 5% on cluster 1, while workplace density has a negative of -7% for cluster 4. For **public transport** variables, it is observed that accessibility to train/light rail stations is correlated with lower car traffic by 2% for cluster 4 and by 3% for cluster 3.

4.4.2.2. Buses

Also, in the case of buses, relatively few variables are retained by the lasso estimator. **Distance** to urban area centres has a high coefficient (12% and 33%) for clusters 2 and 1 respectively, where points in cluster 1 are concentrated around major urban areas. Among the **socioeconomic** variables, population and population density also have high positive signs, up to 29% in cluster 2. With regards to **public transport** variables, bus stops have a high positive coefficient, ranging between 22% and 24% in three out of five clusters, as one would expect.

4.4.2.3. Light Good Vehicles (LGVs) and Heavy Good Vehicles (HGVs)

For LGVs and HGVs the **categorical** variable indicating road category has high negative coefficient, particularly in the case of HGVs, being associated with lower HGV traffic volumes of 26% and 21% in clusters 2 and 3 respectively. In the case of LGVs, road category indicates lower traffic volumes by 6% in cluster 3 being the highest sign estimated across the dataset for this particular road class. **Distance** to urban areas has the highest positive coefficient for HGVs, up to 24% in cluster 4 and distance to urban centres a negative sign indicating lower HGVs traffic volume by 13% in the same cluster. For LGVs and HGVs two statistically significant **land use** variables have positive signs for both vehicle types across road groups. Warehouses are correlated with higher LGVs and HGVs volumes ranging from 1% to 7% for LGVs and from 11% to 22% for HGVs. Similarly, the coefficients for factories range from 2% to 4% for LGVs and from 1% to 5% for HGVs. In the case of **public transport** variables, bus stops exhibit a negative coefficient for both vehicle types, although it tends to be significantly higher for HGVs, reaching -20% in cluster 4. Finally, among the **socioeconomic** variables, the number of households has a negative sign in most clusters, up to -15% in group 3 for HGVs, indicating that household occurrence is associated with lower HGVs volumes.

4.4.2.4. Two wheeled vehicles

Similar to the results for ‘A’ roads presented earlier, **distance-**related variables exhibit high positive signs, with values up to 15% in cluster 2, while **socioeconomic** characteristics are again those mostly associated with two-wheeled vehicle traffic. In particular, income has positive sign across all groups, ranging from 3% in cluster 3 to 25% in cluster 5. Workplace and population density are also highly correlated with the two-wheeled vehicles volumes in most clusters, with coefficient signs ranging from 11% up to 38%. On the contrary, negative coefficients are evident for the registered vehicles variable (i.e., “car_van” – Table D-5), ranging between -3% and -14% indicating the lower traffic volume levels of two-wheeled vehicles in those areas where a higher number of cars are registered. Finally, among the **public transport** variables, the occurrence of bus stops indicates lower volumes up to 14% in cluster 3.

4.4.3. ‘C’ roads

Also, in the case of ‘C’ roads, only a small number of variables have been selected by Lasso. Specifically, a small number of variables is selected in cases where traffic volumes are low, such as in clusters 4 and 5, or for those vehicle types where traffic is relatively low across the clusters such as buses (Table D-7 and Table D-8).

4.4.3.1. Cars

In the case of cars in ‘C’ roads, like with other road classes, the highest positive coefficients are observed in relation to the **categorical** variable indicating dual carriageway roads, which is related to car volumes being higher by 2% in cluster 2 and 19% in cluster 1 compared to single carriageways. The highest negative coefficients are observed in the case of **distance-**related variables up to 24% in cluster 4 for distances to urban areas, and up to 13% in cluster 5 in the case distance to urban areas’ centres.

The **land use** variables representing vehicle related facilities (see Table D-1 in the Appendix and Table 3-2) and education related buildings and facilities (i.e., “Research” – e.g., schools and universities – in

Table D-7) have positive signs for clusters 2 and 4 ranging from 2% to 5%. Sporting facilities are also related to higher traffic, as much as 9%, while parking infrastructure has a positive sign up to 5%. On the contrary, occurrence of factories is associated with higher car volumes in some clusters (up to 8%) but lower in others, up to -3%. The presence of offices also has a negative coefficient in clusters 2 and 3 indicating lower traffic volumes of 2% and 7% respectively. Public services buildings – indicated by the variable "public" in Table D-1 – has a negative coefficient of 7% for cluster 2.

Socioeconomic variables are particularly important in cluster 4 and 5, where count points are primarily located in rural areas. Specifically, population is related to higher car traffic volumes by 12% and 21%, respectively. The number of households has a positive sign of 9% in cluster 5, while the coefficient of income is 13% in cluster 4. In the case of the other clusters, these variables tend to have a small signs, if any.

4.4.3.2. Buses

In the case of buses, the **distance** to urban areas variable has high negative coefficients across most clusters, ranging from -7% to -22%. The signs of distance to urban city centres on bus volumes, however, is less uniform, going from a negative of 13% in cluster 4 to a positive of 23% in cluster 2. In the case of **land use** variables, healthcare and education facilities have both positive signs ranging from 5% to 16%, while the leisure facilities are related to a positive sign of 12% in cluster 3. Large differences across clusters are observed for other variables related to land use, such as parking facilities, exhibiting positive coefficients in cluster 1 and 4 at 5%, but a high negative coefficient for cluster 3 (-20%). Similarly, the variable "Public" – indicating public services buildings (see Table D-1 in the Appendix) – has a negative sign of -10% for cluster 2, but positive for clusters 1 and 3, at 2% and 12% respectively. Vehicle related infrastructure – indicated by the variable "Vehicle" in Table D-1 – tend to be related with lower bus volumes as high as 22% in cluster 3.

Socioeconomic variables play a significant role for buses in this road class. In particular, high signs can be observed in the case of population density, with coefficients ranging from -7% in cluster 5 to 39%

in cluster 2. The number of households also exhibits negative signs in clusters 2 and 5, being 28% and 2%, respectively.

Finally, **public transport** variables have a positive coefficient across all clusters where they are retained. The signs of the number of bus stops are particularly strong in clusters 2 and 3, at 32% and 37% respectively. Bus stations also have a positive sign of 12% on cluster 3 while accessibility to train stations is correlated with higher bus volumes by 10% in cluster 4.

4.4.3.3. Light Good Vehicles (LGVs) and Heavy Good Vehicles (HGVs)

For LGVs and HGVs, **distance**-related variables have mixed coefficients signs. Distance to urban areas has a low positive coefficient for both vehicle types in cluster 3 (4% for LGVs and 2% for HGVs). The coefficients signs are higher but negative in other clusters – e.g., -11% for LGVs in cluster 4 and -7% for HGVs in cluster 5. Similarly, distance to major urban areas has a positive coefficient for LGVs in cluster 1 (4%) but negative in cluster 4 (-8%). For HGVs the sign is also negative in cluster 3 (-2%). Distance to urban centres has also a negative sign on LGVs volumes at -5% in cluster 4 and -6% in cluster 5.

The **land use** variables of warehouses and factories have significant correlation with both vehicle types across almost all clusters. In particular, the presence of warehouses indicates higher LGVs volumes up to 10% and HGVs up to 26%, and the presence of factories indicates higher LGV volumes up to 13% in cluster 4 and HGV volumes up to 15% in cluster 3. Sporting facilities, healthcare infrastructure, and farming and agriculture facilities are all associated with higher LGVs volumes by 9%, 5% and 4.5% respectively in low traffic clusters. On the contrary, the number of shops is associated with lower LGV volumes by -4% and -3% in clusters 1 and 3 respectively.

In terms of **socioeconomic** characteristics, population is related to higher LGVs volumes by 6% in cluster 4 and by 12% in cluster 5, while workplace population is correlated with higher HGVs volumes in cluster 5 by 13%. On the contrary, the number of registered vehicles – indicated by the variable “car_van” in Table D-1 – indicates fewer HGVs by -7% in cluster 3 and by -18% in cluster 2. Among the **public transport** variables, number of bus stations is associated with lower LGV and HGV volumes in clusters 2 and 3 up to -6%, although train/light rail accessibility has a positive coefficient (8%) for HGVs in cluster 2.

4.4.3.4. Two wheeled vehicles

Again, similar to road classes ‘A’ and ‘B’, the **categorical** variable “road nature” is statistically significant indicating that dual carriageways carry 22% more two-wheeled vehicles in cluster 2 compared to single carriageways. The **distance**-related variables exhibit positive signs in most of the clusters, indicating higher two-wheeled traffic volume ranging from 4% in cluster 3, up to 18% in cluster 2. The highest negative coefficient is observed for the **land use** variable indicating the number of shops at -17% in cluster 2. Similar to ‘A’ and ‘B’ road classes, the **socioeconomic** characteristics play a significant role for two-wheeled vehicles, where the high coefficients are observed. In particular, population density is related to higher traffic volumes up to 20%, while workplace and population density also having a positive sign in clusters 2 and 3 at 14% and 29% respectively. In the same clusters, income is related to higher traffic volumes by 10% and 5%. The number of registered vehicles has a negative sign in cluster 2 at 9%. Finally, the **public transport** variable of bus stops also exhibit a negative coefficient of -3% in cluster 2 and by -16% in cluster 3.

4.4.4. ‘U’ roads

Again, in the case of ‘U’ roads, few variables have been selected by Lasso. However, in contrast with the other road classes, there is no specific pattern of variable coefficients across clusters or vehicle types (Table D-9 and Table D-10).

4.4.4.1. Cars

For cars, a larger number of variables has been selected in clusters with lower traffic volumes, but none of the variables is observed to have similar either positive or negative sign across clusters. In fact, more variables are selected in cluster 5 where the highest and lowest coefficients are observed. More specifically, **distance** to urban areas is associated with lower car volumes in cluster 5 by -16%. In terms of **land use** characteristics, four variables have a relative significant coefficient. In particular, the number of public buildings and infrastructure – indicated by the variable “Public” in Table D-1 – has a positive sign of 6% in cluster 5. In the same cluster, education related facilities – indicated by the variable “Research” in Table D-1 – are related to more cars by 8%, while the number of offices has a significant positive sign of 22%. Charging points have a positive coefficient of 7% in cluster 5, although there is a negative sign of -7% in cluster 4.

The highest positive coefficient is observed for the **socioeconomic** variable of registered vehicles at 66% in cluster 5, although the coefficient is 4% in cluster 4. Population density also has a positive sign, indicating higher car volumes in cluster 5 by 13%, although it indicates lower traffic by 10% and 6% in clusters 3 and 4, respectively. Workplace and population density is correlated with higher car traffic by 26% in the low traffic volume cluster as well. The highest negative coefficient is observed for the population variable in cluster 5, indicating lower traffic volume by -36%. Finally, income has a positive coefficient of 4% in cluster 4.

In terms of **public transport** characteristics, it is observed that the presence of bus stations correlates with lower car volumes up to -9% (cluster 1). Train station accessibility also indicates lower car volume by -12% in cluster 3, although it correlates with higher volumes by 13% in cluster 5. The number of bus stops is related with higher car volumes by 33% in cluster 5 only.

4.4.4.2. Buses

For buses, **distance** to urban areas has a negative sign across all clusters with values ranging from -9% in cluster 1 to -20% in cluster 2.

The **land use** variable of warehouses has a positive coefficient in cluster 1 at 70% although the coefficient signs for clusters 2 and 3 are negative, indicating lower bus traffic by -14% and -9%

respectively. Similarly, the presence of superstores is associated with higher bus volumes at cluster 1 by 36% and lower in cluster 4 by -2%. The presence of factories indicates lower bus traffic by -22% at cluster 2. In terms of **socioeconomic** variables, the number of registered vehicles correlates with lower bus volumes at clusters 3 and 4 by -22% and -3% respectively. Income also has negative coefficients in three clusters. In particular, signs are negative in clusters 2, 3 and 4 indicating lower bus traffic volumes by -10%, -5% and -6% respectively.

Similar to other road classes, among the **public transport** variables, the number of bus stops is significantly correlated with higher bus volumes in all clusters where it is selected by Lasso, with coefficient signs ranging from 2% in cluster 4 to 86% in cluster 2.

4.4.4.3. Light Good Vehicles (LGVs) and Heavy Good Vehicles (HGVs)

For LGVs and HGVs **distance**-related variables have both negative and positive coefficient signs ranging from -8% to 4% for LGVs and from -10% to 32% for HGVs.

The **land use** variable of warehouses has positive coefficients for both vehicle types and across all clusters, ranging from 1% to 11% for LGVs and from 6% to 35% for HGVs. However, in contrast with other road classes the occurrence of factories has mixed signs on LGV volumes, with a negative coefficient of -5% for LGVs in cluster 3 and a positive of 2% in cluster 4. In the same cluster, factories are correlated with higher HGVs volume by 12%. In terms of **socioeconomic** characteristics, income has a high positive coefficient of 45% for HGVs in cluster 1 and a negative of -3% for LGVs in cluster 3. The number of registered vehicles is associated with lower HGV traffic in clusters 1, 2 and 3 by 30%, 17% and 22% respectively. Finally, population and population density also indicate lower traffic volumes for both vehicle types up to -26% for LGVs in cluster 3 and up to -5% for HGVs in cluster 4.

4.4.4.4. Two wheeled vehicles

For this vehicle type the majority of **distance**-related variables have positive signs. In particular, the coefficients for the distance to urban and major urban centres variables range from 2% in cluster 5, to

16% in cluster 3. However, the distance to urban areas has mixed signs on two-wheeled vehicles, indicating higher traffic volumes by 2% in cluster 2, but lower in cluster 3 by 4%. The highest negative coefficients among the **land use** variables are observed for healthcare facilities at -8% and factories at -7%, both observed in cluster 3. Moreover, petrol stations are related with lower two-wheeled vehicle traffic by -3% in cluster 2 and by -10% in cluster 3 respectively.

Socioeconomic variables again have significant coefficient signs for this vehicle type, similar to other road classes. Population density, workplace density and income, are all associated with higher traffic volume up to 26%. Moreover, the number of households is associated with increased traffic volumes by 10% in cluster 2 and by 66% in cluster 3. On the contrary, the highest negative coefficient is observed for the population variable in cluster 3, indicating lower traffic volume by -24%.

Finally, the effects of **public transport** characteristics vary across groups. Train station accessibility is related with decreased traffic by 5% in cluster 4, although the variable exhibits a positive sign of 19% in cluster 2. Similarly, the number of bus stops is related with decreased volumes by 5% in cluster 3 and by 1% in cluster 4, although it exhibits a positive sign in cluster 5, increasing traffic volume by 13%.

4.5. Discussion

This chapter focused on the identification of several characteristics that are associated with traffic volumes of the five different vehicle types, normally used by DfT. The coefficients were analysed by road class and vehicle type, further classified in five subgroups (i.e., clusters). In this section, the results presented in section 4.4 are discussed and the causes behind the statistically significant variables produced by Lasso are identified. Due to the large number of variables, focus is placed on those exhibiting similar patterns across road types and clusters and those with the highest impact on each vehicle's traffic volume as indicated by the coefficients.

As a general observation it can be seen that most coefficients take the expected value across road classes, vehicle types and clusters. However, one has to bear in mind the imbalanced number of traffic counters for each road type (Table 4-1) and that for smaller road classes traffic counts are taken seasonally and manually, and eventually adjusted to estimate AADT (Department for Transport, 2013, 2014), therefore adding additional uncertainty to the results. Moreover, the number of vehicles driving through roads with lower traffic volumes is small, so that zero counts are not uncommon for some vehicle types, and most count points are in urban areas.

The significant effect of some variables across all road classes and clusters is also observed. Specifically, carriageway type is significant for 'A', B' and 'C' roads, with dual carriageways being correlated with higher traffic volumes for most vehicle types. More vehicles are expected to drive through dual carriageways – as opposed to single – due to safety and increased speed limits (Butcher, 2017; Pitaksringkarn et al., 2018) and lower design standards for single carriageways (Gitelman et al., 2017). The effect of this variable is not significant for 'U' roads, due to the fact that there are not many dual carriageways in these roads. Moreover, it can be observed that the set of distance variables affects all vehicles in all road classes. The effects of the variables selected by Lasso for each vehicle type are discussed for all the investigated road classes as follows:

4.5.1. Cars

In the case of cars, income is statistically significant in all road classes and most clusters with positive sign, apart from cluster 3 in 'B' roads. The positive signs are explained by the fact that the higher the income the more likely one is to own a car (Silva et al., 2012). For 'A' roads, income has a positive coefficient on all clusters with higher sign in cluster 5 – the one with the lowest public transport coverage. For 'B', 'C' and 'U' roads the variable is significant with smaller signs in clusters 4 and 5 where counters are located mainly in rural and/or suburban areas, confirming results for 'A' roads. The positive coefficient of income, particularly in rural areas confirms findings in Oakil et al., (2018), where commuting is more likely to occur with private vehicles, due to lack of public transport (Diao and Ferreira, 2014) and long waiting times (Becker and Axhausen, 2017).

Distance to urban and major urban areas is also statistically significant across all road classes and most clusters. The pattern observed indicates that the further from urban areas the less cars are on the streets. The coefficients tend to be higher for low volume clusters across all roads, a logical outcome, since most counters are within urban areas where volumes are higher. Similarly, distance to urban centres is significant at all road classes, although the signs are usually positive and higher for 'A' roads. Moreover, the high positive coefficients for distance to major urban centres observed for 'A' roads, are explained by the fact that urban centres are usually better served by public transport. In the case of major urban centres, the results are significantly affected by the Greater London area, where there is a high public transport coverage and congestion charges apply in the city centre.

Studies have shown that the effect of sporting facilities on traffic volumes depend on the scale of facilities and activities taking place (Humphreys and Pyun, 2018), with large scale events taking place in large sport facilities significantly increasing traffic in the surrounding area (Ghosh et al., 2019; Kim et al., 2017). Considering that many different types of facilities are included in the "sport" variable (Table D-1), it can be assumed that higher coefficients relate to larger venues, where people usually drive to attend events (Ghosh et al., 2019). Smaller ones relate to smaller sport infrastructure such as sporting centres or golf courses, attracting fewer people.

Considering that the number of shops and offices attracts more people, car traffic would be expected to follow (Choudhary and Gokhale, 2019). This is observed in cluster 5 (mainly rural), providing one more evidence about the reliance to private vehicles in rural areas (Jain et al., 2018; Zahir and Haron, 2019). However, high negative coefficients are observed in clusters 3 and 4 for 'A' roads and clusters 2 and 3 for 'C' roads. Here, traffic counters are located in urban centres (Table 4-1) and the coefficients can be related to higher public transport coverage (Figure 4-2 and Figure 4-3).

The occurrence of public transport where one would expect to be associated with lower car volumes (Cervero, 1994; Handy, 1996) is evident across all road classes and most clusters with some exceptions. Train accessibility is associated with lower car volumes in clusters 3 and 4 in 'B' and 'U' roads (i.e.

predominantly urban); an expected outcome, also validated by the majority of similar studies (e.g. Aditjandra et al., 2012; Cervero, 1994). However, in ‘U’ roads, public transport has mainly positive coefficients in cluster 5. This is due to the fact that cluster 5 represents rural areas (Table 4-1), indicating that people access stations with private vehicles, considering that these areas are more likely to exhibit an absence of buses serving the stations. For ‘A’ roads, where train station accessibility has negative signs in cluster 5 and positive in clusters 3 and 4, the coefficient is related to competition between car and public transport modes, due to the fact that train stations are usually located near heavy traffic roads (Kwon et al., 2016).

Figure 4-2: Bus Stops Distribution for ‘A’ Road Groups 3 (left), 4 (middle) and 5 (right)

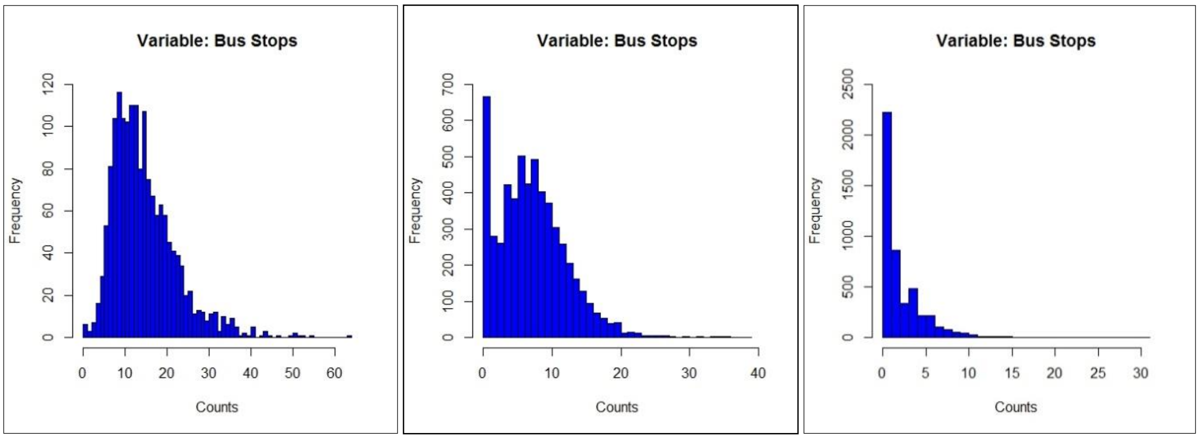
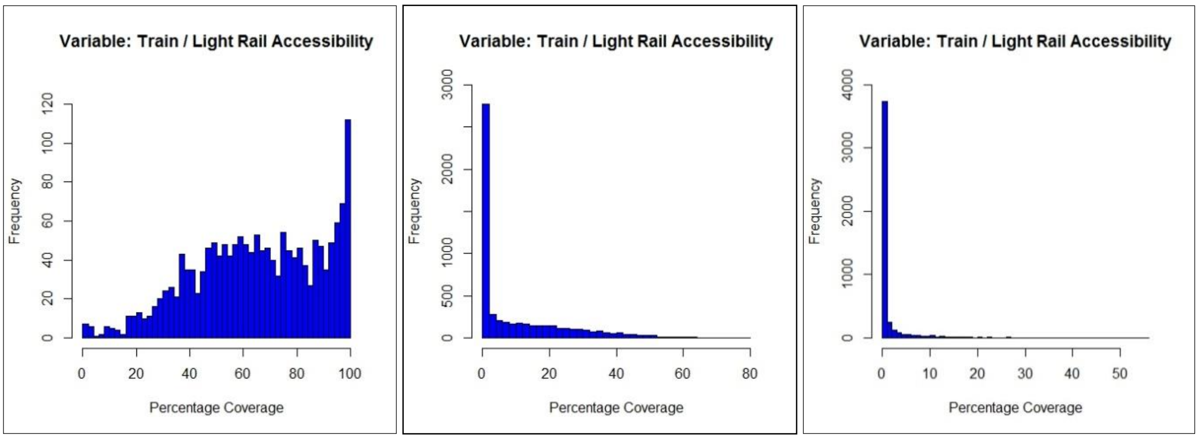


Figure 4-3: Train/Light Rail Accessibility Distribution in ‘A’ Road Groups 3 (left), 4 (middle) and 5 (right)



The coefficients from socioeconomic variables associated to population, workplace population and related densities indicate that the higher the population, the higher the number of cars on the streets, a finding that complies with outcomes from other models (e.g. Mohamad et al., 1998; Zhao and Chung, 2001; Zhao and Park, 2004). On the contrary, density indicators are usually associated with lower car volumes confirming findings of Zhang, (2007), as compact areas discourage car usage. However, densities are usually associated with higher traffic volumes in cluster 5, again indicating that these – predominantly rural – areas depend more on private vehicles.

Finally, a few variables are correlated with car volumes on specific road classes and clusters, particularly in the case of 'A' roads. Specifically, one would expect high positive coefficients on ring roads, since they are usually large roads designed to carry and distribute large traffic volumes radially across the network (Jianqin et al., 2015). Access to motorways with positive coefficient in cluster 5 is also expected, since roads service as entry points to motorways carry more vehicles compared to roads of the same class without access and these are often located in rural areas. A number of studies also confirms my findings (e.g. Dombalyan et al., 2017) related to the negative coefficient for toll roads in cluster 3, with lower traffic volume compared to toll-free roads of similar size[42]. Toll roads tend to obstruct traffic volumes due to toll stations (Abdelwahab, 2017) and people tend to avoid roads where a fee is needed. The negative signs for ports in clusters 3 and 5 are expected, as freight vehicles are likely to generate the highest share of traffic, although one should notice that only very few counters are located close to ports for both clusters. Finally, the presence of petrol stations associated with higher levels of car traffic is related to additional traffic generated by the station itself as cars drive in those roads to refuel but also because petrol stations are placed at locations with already high traffic.

4.5.2. Buses

Measurements for this vehicle type also incorporate coaches (Department for Transport, 2014), which may cause uncertainty in relation to the estimated coefficient since buses are used for commuting within urban areas, while coaches are used for intercity travels.

[42] The positive coefficient in cluster 5 (see Table D-3), is probably explained by the fact that there is only one traffic counter indicated as toll road, hence we can question this coefficient's reliability.

Among the significant variables, the sign of estimated coefficients confirms expectations with few exceptions. Bus stops correlate to bus volumes across all road classes and clusters where the variable is selected. Distance to urban areas has negative coefficient across all clusters, confirming that more buses are expected in urban rather than rural areas (Carli et al., 2015). On the contrary, distance to major urban areas is statistically significant for ‘A’ roads only in clusters 1 and 4 and it does not reflect the expected outcome, possibly due to the fact that coaches are also incorporated in this vehicle type.

Similarly, distances to urban area centres indicate an increase in bus volumes for ‘A’, ‘B’ and ‘C’ roads. For ‘A’ roads, the coefficient is positive probably due to the fact that more than half of the counters are placed on trunk roads and ring roads where buses are less likely to drive through. For ‘B’ and ‘C’ road classes the coefficients can be explained by the fact that points in the clusters where the variable is significant are located within urban areas, but the majority is located around city centres (e.g., suburbs). However, for ‘A’ roads, distance to major urban area centres indicates that bus volumes reduce the further away from city centres, although signs are opposite for clusters 4 and 5 in ‘B’ roads. This is an expected outcome, since most of buses drive through ‘A’ roads and one would expect major city centres to be better served by these vehicles, while for ‘B’ roads again the coefficients could possibly be affected by coaches.

Similar to the case of cars, population, workplace population and related densities are associated with increased bus volumes across ‘A’, ‘B’ and ‘C’ roads and most clusters indicating that bus routes are placed and used in heavily populated urban areas, where public transport has a significant effect (Bento et al., 2005; Friedman et al., 1994). The few negative coefficients normally observed in cluster 5 indicate reliance on private vehicles in rural areas. These variables are not statistically significant in the case of ‘U’ roads – usually small roads in residential areas – where less buses are likely to drive through.

The negative coefficients for income for ‘B’, ‘C’ and ‘U’ roads indicate that higher income results into lower bus use, since households with higher incomes are more likely to own private vehicles (Cervero and Kockelman, 1997; Papagiannakis et al., 2018; Silva et al., 2012) and an increased number of private vehicles is related to lower bus traffic, since transport needs are met by cars (Aditjandra et al., 2012; Diao and Ferreira, 2014). These results confirm expectations and are correlated with coefficients related to private vehicles, such as the registered vehicles (i.e., "Car_van" – Table D-1), parking facilities and vehicle related infrastructure variables, more clearly indicated in the case of 'A' roads. Here, in areas with parking availability and places where vehicle related infrastructure is evident – e.g., vehicle repairs workshops and garages – bus volumes are lower.

Again, as with cars, several variables are correlated with bus volumes on specific road classes and clusters. For 'A' roads, the positive coefficients for primary and trunk roads in cluster 1 are expected, although the coefficient for primary urban roads is negative for cluster 5, most likely because cluster 5 predominantly consists of rural roads. Moreover, the occurrence of factories, reduces bus traffic in 'A' roads as expected, since these facilities are usually located in industrial areas where people usually commute to by private vehicles. Train accessibility in 'B' roads, takes positive sign in cluster 2 (mainly urban areas), due to buses connecting to train stations (Hong et al., 2016; Kang et al., 2019; Seriani and Fernández, 2015), while it has a negative sign in cluster 3 (rural and smaller urban areas) where coaches and trains might be substitute (Webb et al., 2016). For 'U' roads, the high coefficient (36%) for "Superstores" is explained by the fact that 29 out of the 31 counters in this group have at least one superstore in the vicinity, while 26 of them are in urban and/or major urban areas, with the majority being near the respective centres. Therefore, it is likely to have buses servicing these facilities and it can be safely assumed that the coefficient is reasonable. Overall it can be concluded that bus volumes are negatively correlated with variables related to high car volumes, as it is also confirmed by numerous other studies (e.g., Aditjandra et al., 2012; Bento et al., 2005; Cervero, 1994; Stead, 2001).

4.5.3. Light Good Vehicles (LGVs) and Heavy Good Vehicles (HGVs)

In the case of LGVs and HGVs, the most significant effect is related to the occurrence of warehouses and factories, indicating an increase in AADT for both vehicle types across all roads and almost all clusters. This is a logical outcome since these vehicles access these facilities to deliver or pick up goods (Nugmanova et al., 2019). For 'B' roads, the coefficients are remarkably high for the warehouse variable, particularly in cluster 4, with most points being located in rural and suburban areas where HGVs are more likely to drive (Chhorn et al., 2018a).

The set of distance variables is correlated LGV and HGV volumes across all road classes. For 'A' roads, the higher the distance from urban and major urban areas, the lower the level of traffic; a reflection that higher economic activity – usually observed in large urban areas – drives demand for LGVs and HGVs. On the contrary, high coefficients for HGVs in 'B' roads, indicate that large vehicles are difficult to operate within dense urban environments on secondary roads, therefore are usually observed far from urban areas, an expected outcome. For smaller road classes (i.e., 'C' and 'U') in high volume clusters (i.e., clusters 1, 2 and 3) the signs are similar to 'B' roads. However, the negative coefficients observed in low volume rural clusters 4 and 5 particularly for HGVs, indicate that operation of these vehicles is related with urban areas of higher economic activity.

Distance to urban and major urban centres coefficients for 'A' roads, indicate that these vehicles operate far from city centres, with convenient access required particularly by HGVs, but also due to restrictions usually applied, leading to extensive use of ring or peripheral roads (Tzouras and Lázaro, 2018). 'B' and 'C' roads exhibit similar patterns, with negative coefficients for HGVs indicating reduced traffic volumes close to urban centres due to operating restrictions (Browne et al., 2010), although the number of LGVs increases slightly due to the fact that these vehicles are used for goods transportation in inner city roads (Chhorn et al., 2018b). For 'U' roads the high positive coefficient (32%) for distance to urban centres for HGVs in cluster 1 is explained by the fact that the majority of counters are located within urban/major urban areas. Similarly, the negative coefficient of -10% for distance to major urban centres in cluster 1, is possibly affected by the number of points within the Greater London area, where specific permits are needed for HGVs to drive there (Transport for

London, 2019). In both cases, it can also be observed that the number of HGVs is affected by the location of warehouses and factories, which are usually placed far from urban areas and their centres.

The conclusions drawn for the distance variables are also conveyed by socioeconomic variables such as the number of households, population and workplace population densities, which have negative coefficients across all road classes and most clusters. The pattern is slightly different for 'U' roads, where areas with high population densities exhibit lower GV traffic, while workplace population and density both positive impacting traffic volumes – probably indicating industrial areas where these vehicles are likely to operate. Overall, it can be concluded that HGVs are unlikely to drive through lower class roads (i.e., 'C' and 'U') or extensively residential or urban areas, since the roads are smaller, and transportation is more convenient with smaller vehicles. This can also be confirmed by the Road Category variable coefficients in 'B' roads, where lower traffic volume is observed in urban areas as opposed to rural, and also by the coefficients in 'A' roads, where traffic is lower in primary urban roads, although trunk roads, carry higher volume of GVs both within urban and rural areas.

Finally, a few variables are statistically significant in 'A' roads only. For example, the presence of ports is correlated with higher HGVs traffic volumes and lower LGVs volumes across clusters apart from cluster 3. In this cluster, only a few ports can be found in the vicinity of the counters and the variable is not statistically significant. Transportation of goods to ports usually occurs at a large scale, more suited to HGVs, therefore explaining the different coefficients of ports on HGVs and LGVs. One would also expect to see a large number of LGVs and HGVs in ring roads and trunk roads as expressed by the positive coefficients on these variables and the variable indicating access to motorways (i.e., "Junction" – Table D-1). Specifically, HGVs use ring roads to access facilities before transportation switches to smaller vehicles in inner city roads (Chhorn et al., 2018b).

4.5.4. Two wheeled vehicles

Traffic volumes for these vehicles is mostly associated with socioeconomic characteristics. It can be observed that higher income is correlated with higher two-wheeled vehicle volumes (Department for

Transport, 2016b) possibly due to the fact that in the UK, motorcycles are a means of pleasure rather than transport as opposed to other European countries (Delhaye and Marot, 2015).

Similarly, socioeconomic characteristics such as the number of households, population and workplace population densities, are associated with higher numbers of two-wheeled vehicles. This is also confirmed by findings in RAND, (2004) and Wong (2013), indicating that these vehicles are concentrated in dense urban areas. The latter is also confirmed by the set of distance variables, particularly in the case of ‘A’ roads. The coefficients indicate that the further apart from urban and major urban areas the fewer motorbikes are on the streets, while the further apart from city centres (but within urban areas) the more of these vehicles will be. It is worth noting that population density is related to increased motorcycle use as discussed above, yet this is not necessarily true for population totals. Two-wheeled vehicles are likely to be seen in urban areas with higher population and population density, but densities may vary across different parts of urban areas, particularly in major cities.

Although two-wheeled vehicles allow for cheaper and more efficient transport in dense urban areas (Law et al., 2015), the Department for Transport (2016) shows that motorcycle owners are likely to own at least one car, which explains the fact that lower two-wheeled traffic levels are correlated with large numbers of registered cars. This is also indicated by the negative sign for bus stops in ‘B’ and ‘C’ roads.

For ‘A’ roads, the positive coefficient for toll roads in cluster 3 can be explained by the fact that counters in this cluster are mainly withing urban areas including London, where all points within London’s congestion charge zone (covering Central London) are indicated as toll roads. As two-wheeled vehicles are common in dense urban areas, this outcome indicates higher volumes in Central London, an outcome facilitated by the fact that motorbikes are exempt from charges (Prud’homme and Bocarejo, 2005). For cluster 4 the result is considered unreliable, due to the fact that only two points are marked as toll roads. These indicate river crossings where more traffic than other counters in the group is expected.

4.6. Chapter summary

In this chapter, the full set of variables created in chapter 3 has been examined, to understand their relation with traffic volumes of five different vehicle types in England and Wales. The analysis of traffic volumes has been undertaken for four different road classes where traffic counters have been subdivided into five groups (i.e., clusters) based on specific land use, socioeconomic, public transport, and roadway characteristics in the vicinity of each counter. The results produced by Lasso reveal patterns for specific explanatory variables across vehicle types and road classes. In some cases, heterogeneous results across estimated models have been reconciled by looking at the characteristics of the counters and areas in each model.

This chapter has to a great extent fulfilled the first aim of this thesis - *to identify the degree of influence specific factors have on traffic volume variations across the road network.* Certainly, the outputs of the modelling conducted in this chapter can be perplexing, since many different road classes, vehicle types, clusters and numerous variables are examined and discussed. This process can result into some of the variables producing unexpected or challenging to explain coefficients. However, considering the complexity of transport systems, this is considered to be a contribution to the field of road transport studies, since the effect of diverse characteristics on road transport has been captured. Moreover, as the transport literature is constantly enriched one would expect an even larger number of factors being considered and novel ideas to investigate their effects using statistical – and potentially other types of – models. Nevertheless, the outputs from the model applied in this chapter can be used to inform policy making in road transport and interrelated fields, such as urban and regional planning, economics and of course the road transport contribution in GHG and air pollution emissions; the latter being examined in the following chapter.

5. Emissions modelling

5.1. Chapter overview

In chapter 2 it has been discussed that GHG and air pollutants from the road transport sector can be estimated with the utilisation of emission models. These models use activity (i.e., traffic) data related to speed or distance travelled coupled with emission factors – that attempt to relate the activity with the quantity of the pollutant. Activity data are normally inferred from transport models, while the emission factors are usually extracted from related databases.

In this chapter, the contribution of road transport in CO_2 and specific air pollutant ($PM_{2.5}$, NO_x, CO) emissions for England and Wales is investigated, intending to address the second aim of the thesis – *to assess the quantity of CO_2 and three air pollutants (PM, NOx, and CO) originating from road transport and identify potential emissions abatement through technological developments of road vehicles and policies development*. To meet the aim, a methodological approach is presented, where emissions are estimated for the current (i.e., base) year on a street segment level for all roads – major and minor – in the study area. Emissions are also estimated for 2035 considering the base year estimations and projected trends in road transport. Trends in road transport are considered both in terms of traffic volumes as well as vehicle technological developments (e.g., electric vehicles), using governmental resources (e.g., Department for Transport, 2018b). By doing so the identification of potential abatement – or rise – in emissions originating from road transport can be achieved and hence insights on the future contribution of road transport to air pollution and CO_2 can be provided. The estimations are compared with the outputs from other modelling approaches, to assess the methodological approach and also draw conclusions on the performance of different models to estimate emissions[43]. Furthermore, in a separate case study, air pollutant emissions in the Greater

[43] It should be noted that emissions from vehicles are mainly generated from tailpipes, although other parts of the vehicle also contribute, such as tyre and brake wear (Jeong et al., 2019). This chapter focuses exclusively on emissions generated from vehicles' exhausts.

London area[44] are estimated, to assess the effects of recently introduced policies to mitigate air pollution from road transport as well as the potential effects of policies to be introduced in the near future. Specifically, air pollutant emissions in Greater London are estimated to assess the effects of ULEZ and also to identify potential emissions abatement from the ULEZ extension[45].

The aim of this chapter is attained by implementing a two-step methodological approach based on classification and regression models, where emissions are estimated for all road classes and all available vehicle types. Considering the outputs from the model are produced at detailed local levels, it is considered that the model can be used to identify both hot spots of air pollution exposure and GHG emissions, as well as to estimate total (i.e., aggregated) emissions.

The chapter is presented in six sections. Section 5.2 presents the datasets used and the corresponding sources while Section 5.3 presents the methodology to estimate emissions at a granular level, applied to both case studies. Firstly, the selected emission modelling approach is justified and then the methodological steps followed to estimate emissions and associated projections are presented. Section 5.4 presents the outcomes from the modelling process and in section 5.5 the results are discussed. Finally, section 5.6 summarises the chapter.

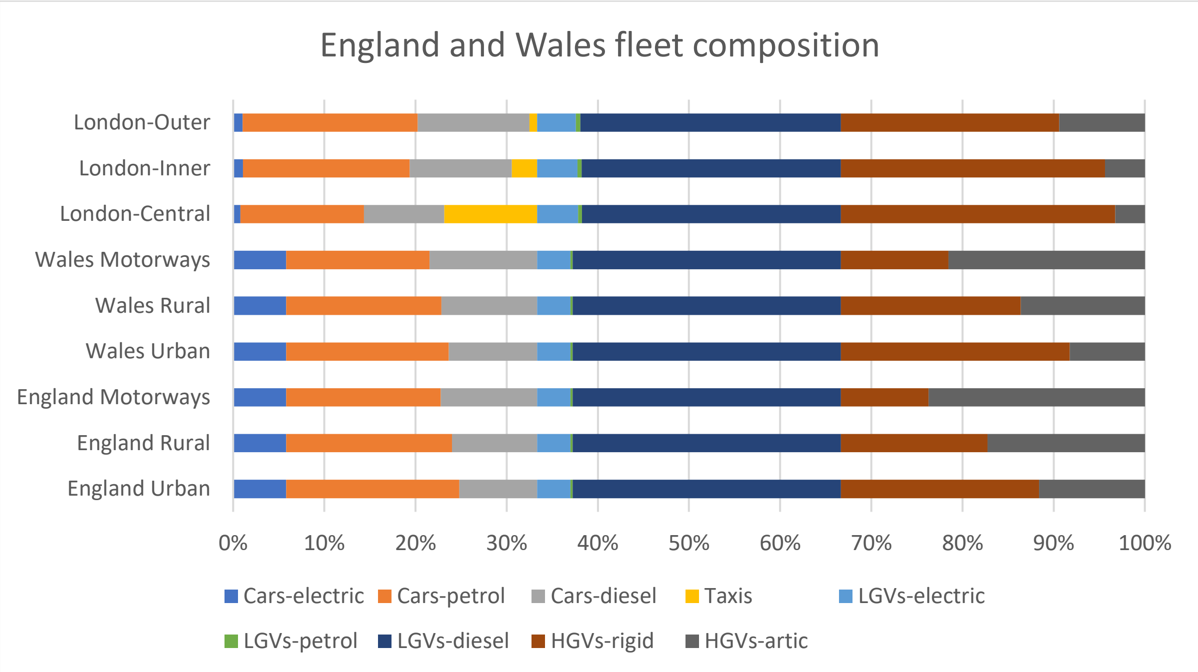
5.2. Data

To estimate emissions for England and Wales, four different datasets are utilized. Firstly, for activity data the dataset presented in chapter 3 is used, and – instead of total AADT – the five different vehicle types (i.e., Cars and Taxis, Buses, Light Good Vehicles (LGVs), Heavy Good Vehicles (HGVs) and Two-wheeled vehicles) are considered as also discussed in chapter 4. In addition to the four road types (i.e., ‘A’, ‘B’, ‘C’ and ‘U’) examined in the previous chapters, and to account for all roads within the study

[44] Emissions for Greater London are estimated for the air pollutants only, since Clean Air Zones focus on the reduction of air pollutants rather than GHGs (Bernard et al., 2020).
[45] At the time of writing ULEZ extension is planned to come into effect in October 2021.

area traffic count points for motorways[46] from DfT are also extracted, where again AADT values for the same five vehicle categories are provided. Secondly, the vehicle fleet composition for England and Wales from NAEI is extracted, where the vehicle types are further segregated. Specifically, cars and LGVs/vans are classified by fuel type (i.e., indicating whether a vehicle uses petrol, diesel or being electric), while HGVs are subdivided to rigid and artic vehicles as is also shown in Figure 5-1.

Figure 5-1: Vehicle fleet composition in England and Wales



Moreover, emission factors for the air pollutants ($PM_{2.5}$, NO_x, CO) are also extracted from NAEI, while CO_2 emission factors are extracted from the Department for Business, Energy and Industrial Strategy (BEIS) as shown in Table 5-1. Finally, traffic projections (in Vehicle Miles Travelled – VMT) are extracted from two scenarios developed by DfT with the National Transport Model (NTM)[47]: (i) the "Reference"[48] scenario – as the standard baseline scenario and (ii) the "High GDP, Low Fuel"[49] scenario, where its

[46] As a reminder, motorways have been excluded from chapters 3 and 4, because traffic on these roads is not directly affected from its surrounding characteristics (Eom et al., 2006; Zhao and Chung, 2001).
[47] The NTM is a FSM simulation-based model, taking into account population growth, economic factors, and demand for goods and freight among others (Department for Transport, 2018d), as it has also been discussed in chapter 2 – sections 2.4 and 2.8.
[48] Assumes updated central forecasts for Gross Domestic Product (GDP), BEIS Central Forecasts for fuel, Central projection for Population and 25% of car and LGV mileage powered by zero emission technologies by 2050 (Department for Transport, 2018d).
[49] Assumes High GDP Growth and Low Fuel Cost Projection (Department for Transport, 2018d).

previous years' traffic projections are more accurate compared to other scenarios. Notice that DfT provides travelled distance in miles, while distance values used for emission estimation and model comparison later in the chapter are all in kilometres (i.e., Vehicle Kilometres Travelled – VKT). Where values are given in miles, the unit has been converted considering that 1 mile = 1.6 kilometres.

Table 5-1: Emission factors (g/km) by vehicle type and road location

Vehicle Type	Road location	NOx	$PM_{2.5}$	CO	CO_2
Car - petrol	Urban	0.074	0.001	0.393	0.18717
	Rural	0.061	0.001	0.439	
	Motorway	0.058	0.001	0.622	
Car - diesel	Urban	0.582	0.009	0.071	0.16308
	Rural	0.473	0.007	0.032	
	Motorway	0.539	0.006	0.030	
Taxi	Urban	0.582	0.009	0.071	0.20638
	Rural	0.473	0.007	0.032	
	Motorway	0.539	0.006	0.030	
LGV - petrol	Urban	0.068	0.001	0.999	0.19748
	Rural	0.068	0.002	0.736	
	Motorway	0.076	0.003	1.495	
LGV - diesel	Urban	0.915	0.009	0.078	0.18129
	Rural	1.017	0.011	0.077	
	Motorway	1.362	0.011	0.074	
HGV - rigid	Urban	1.712	0.018	0.483	0.75418
	Rural	0.999	0.016	0.367	
	Motorway	0.780	0.014	0.337	
HGV - artic	Urban	1.140	0.014	0.437	0.63295
	Rural	0.492	0.010	0.298	
	Motorway	0.410	0.009	0.278	
Bus	Urban	2.954	0.029	0.890	0.26760
	Rural	1.167	0.017	0.421	
	Motorway	1.121	0.018	0.447	
Two-wheeled	Urban	0.081	0.007	1.761	0.09826
	Rural	0.104	0.006	1.851	
	Motorway	0.145	0.006	2.298	

The datasets for these two scenarios provide information about growth and/or decline in VMT (Table 5-2[50]) and associated emissions (NO_x, PM, and CO_2), for three different road classes[51] distributed across ten regions in England and Wales[52]. However, it should be mentioned that the selected scenarios do not reflect the current transport policy aiming to reduce the emissions from road transport as also

[50] All PSVs are considered to be buses.
[51] Scenarios distinguish roads classes into Motorways, 'A' roads and minor roads. Minor roads include 'B', 'C' and 'U' road classes.
[52] These include the nine English regions (i.e., South West, South East, Greater London, West Midlands, East Midlands, East of England, North East, North West and Yorkshire and the Humber) and Wales as shown in Table 5-2.

discussed in section 1.1. Scenarios as such, have not been produced by DfT to date, probably due to the fact that the emission reduction targets have been recently updated. The selected scenarios are essentially based on the standard road transport growth projections and the associated characteristics considered, such as GDP and population growth, indicating a business-as-usual case.

Table 5-2: VKT growth/decline for each scenario by vehicle type and region

Region	**Cars[53]**		**LGVs**		**HGVs**		**PSVs[5050]**	
	Scenario1	Scenario2	Scenario1	Scenario2	Scenario1	Scenario2	Scenario1	Scenario2
East Midlands	22.6%	26.4%	31.3%	52.1%	0.4%	2.7%	-10.0%	
Eastern England	21.6%	24.8%	29.4%	49.8%	8.4%	10.7%	-10.0%	
London	24.0%	26.3%	35.0%	55.9%	0.3%	0.0%	1.8%	
North East	19.0%	21.7%	31.7%	52.4%	-0.9%	0.8%	-11.1%	
North West	22.2%	25.7%	29.3%	49.6%	0.3%	2.3%	-11.4%	
South East	22.3%	25.8%	31.2%	52.0%	9.1%	11.8%	-10.0%	
South West	23.9%	27.8%	28.3%	48.6%	-0.8%	0.2%	-10.0%	
West Midlands	21.5%	24.9%	33.5%	54.5%	1.5%	3.3%	-11.2%	
Yorks & Humber	21.5%	25.0%	31.0%	51.7%	1.4%	3.2%	-11.5%	
Wales	21.7%	25.0%	30.2%	50.8%	0.2%	2.8%	-10.0%	
All Regions	**22.2%**	**25.6%**	**30.8%**	**51.5%**	**5.0%**	**4.9%**	**-8.9%**	
	Scenario 1				**Scenario 2**			
All vehicles	**22.20%**				**28.10%**			

For the London-ULEZ case study, six different datasets are utilized. Four datasets are similar to England and Wales – activity (i.e., traffic information), vehicle fleet composition and two different emission factor datasets. For activity data, the dataset presented in chapter 3 is again used, enriched with the traffic counts for motorways from DfT. For vehicle fleet composition, a different, more detailed dataset is again extracted from NAEI where the fleet for the current ULEZ as well as the inner and outer London zones is further split based on Euro standards for each of the five vehicle types (i.e., cars, LGVs, HGVS, buses and two-wheeled) as shown in Figure 5-2 to Figure 5-4.

[53] Note that this category also includes taxis.

Figure 5-2: Fleet composition in Inner London

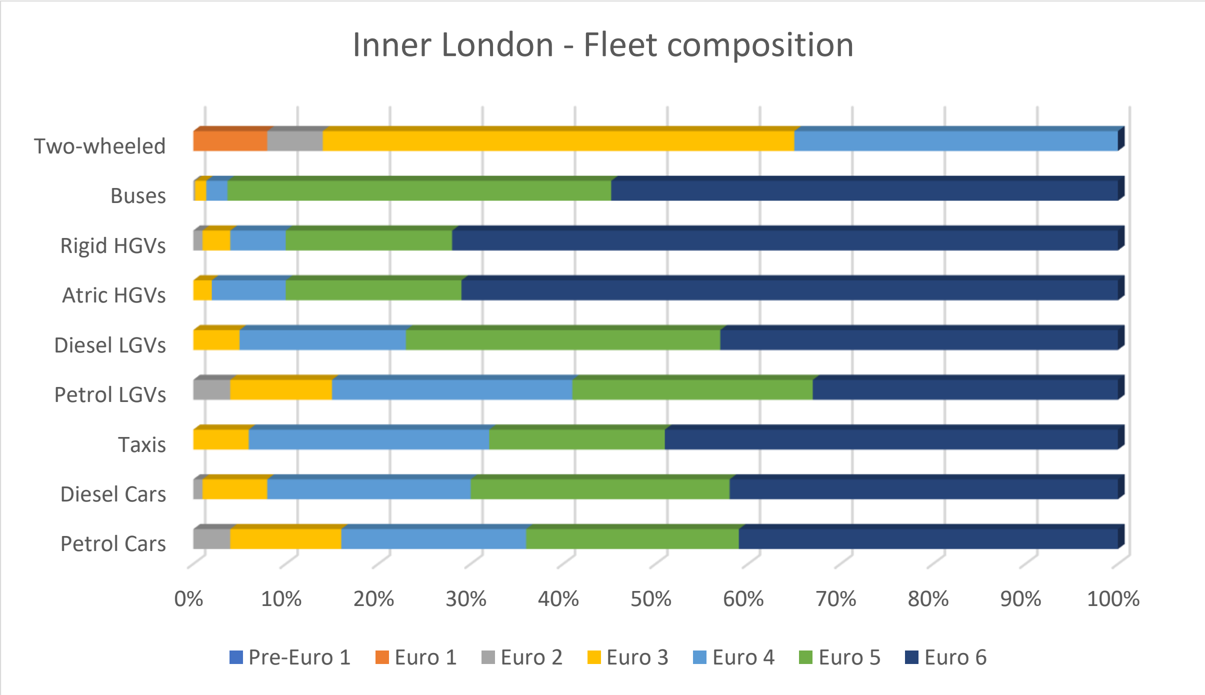


Figure 5-3: Fleet composition in Outer London

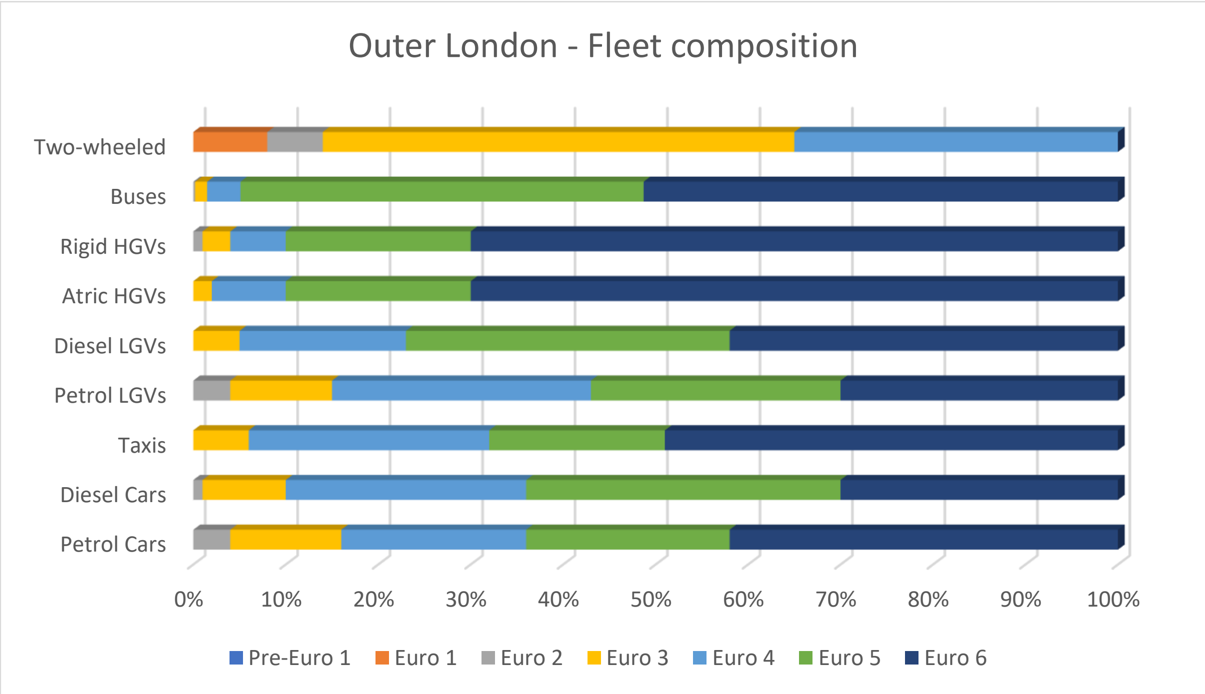
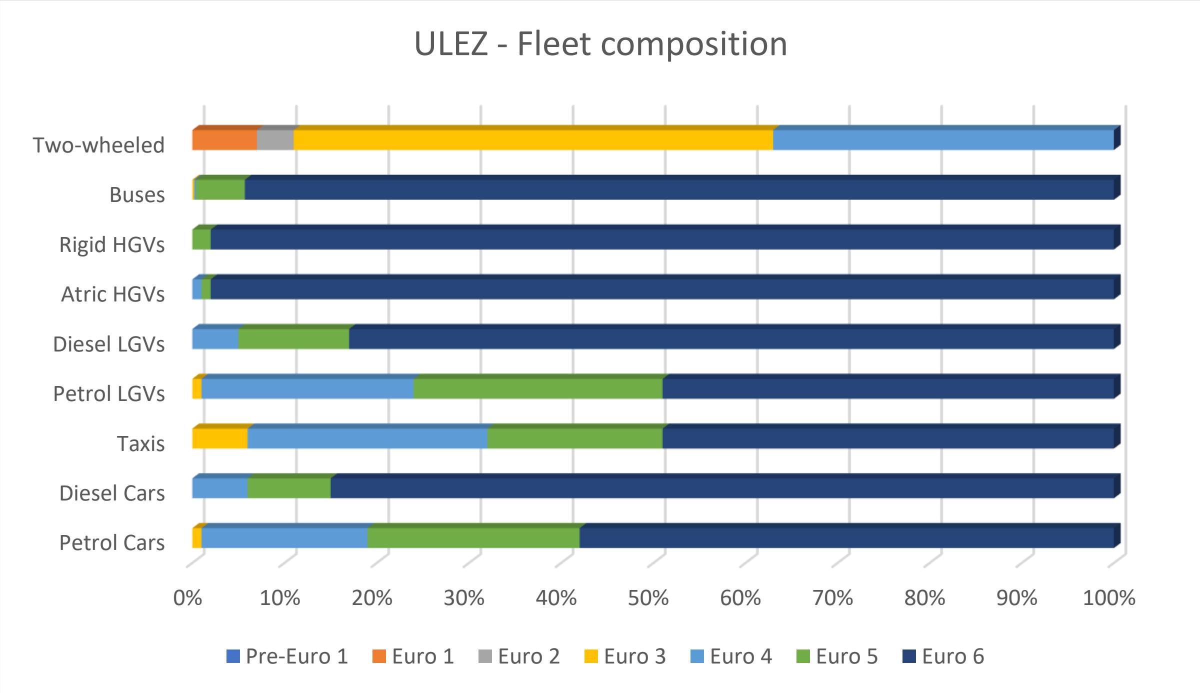


Figure 5-4: Fleet composition in London ULEZ zone



For the emission factors related to the three air pollutants ($PM_{2.5}$, NO_x, CO), two datasets have been used. The emission factors from the first dataset are the same as those used for England and Wales and extracted from NAEI (Table 5-1). The second dataset consists of detailed emission factors for the Euro standards for each vehicle (Table 5-3) so as to correspond to the more detailed vehicle composition.

Table 5-3: Emission factors (g/km) by Euro vehicle type

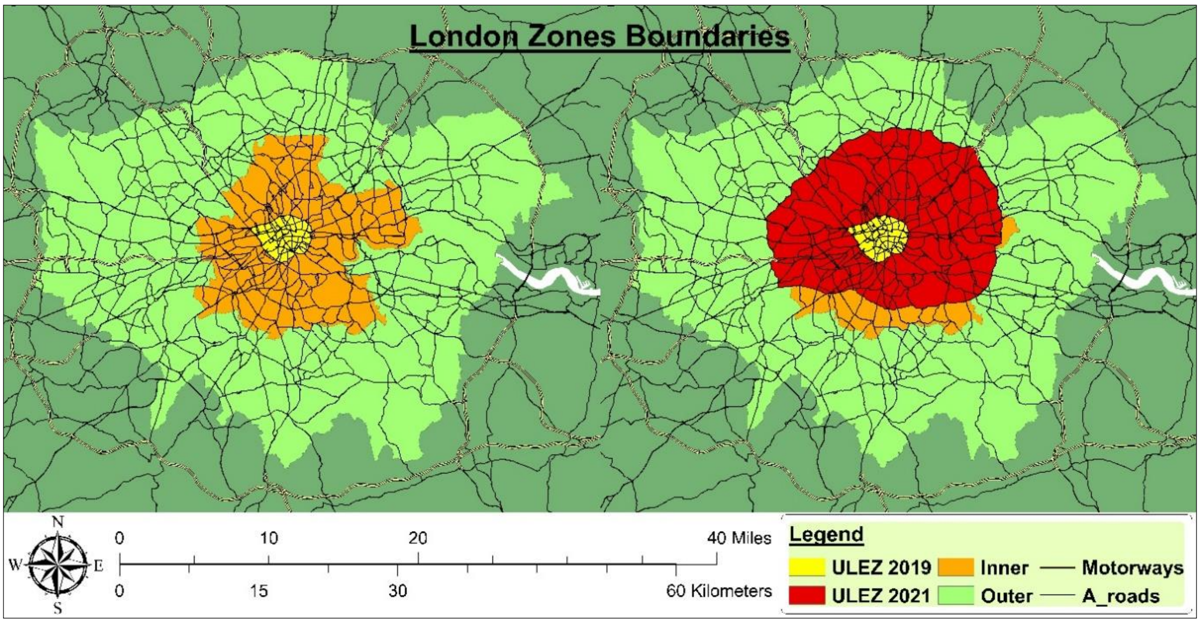
Car type	**Pollutant**			**Car type**	**Pollutant**		
	NOx	$PM_{2.5}$	CO		NOx	$PM_{2.5}$	CO
Cars - petrol				HGVs (Artic & Rigid)			
Euro 1	0.97	0.14	2.72	Euro 1	6.56	0.486	4.5
Euro 2	0.5	0.01	2.2	Euro 2	5.15	0.246	2.75
Euro 3	0.15	0.01	1.82	Euro 3	3.18	0.138	1.28
Euro 4	0.08	0.01	1	Euro 4	2.75	0.025	1.13
Euro 5	0.06	0.005	0.62	Euro 5	1.81	0.019	0.086
Euro 6	0.06	0.0045	0.62	Euro 6	0.35	0.009	0.086
Cars - diesel				Buses			
Euro 1	0.97	0.14	2.72	Euro 1	8	0.486	4.32
Euro 2	0.7	0.09	1	Euro 2	7	0.2	2.75
Euro 3	0.56	0.05	0.092	Euro 3	3.5	0.075	2.25
Euro 4	0.25	0.025	0.089	Euro 4	2.75	0.0408	1.86
Euro 5	0.18	0.005	0.049	Euro 5	2	0.025	0.1865
Euro 6	0.08	0.0045	0.04	Euro 6	0.43	0.01	0.1865

LGVs - petrol			
Euro 1	0.563	0.0023	4.93
Euro 2	0.23	0.0023	3.73
Euro 3	0.18	0.0011	3.89
Euro 4	0.096	0.0011	2.01
Euro 5	0.072	0.0014	1.69
Euro 6	0.064	0.0012	1.3
LGVs - diesel			
Euro 1	1.22	0.19	4.93
Euro 2	1.22	0.14	1.25
Euro 3	1.03	0.07	0.79
Euro 4	0.831	0.04	0.375
Euro 5	0.64	0.005	0.075
Euro 6	0.248	0.005	0.075

Two-wheeled			
Euro 1	0.38	0.047625	6.05
Euro 2	0.35	0.033	5
Euro 3	0.15	0.029	1.3
Euro 4	0.11	0.08	1.07
Euro 5	0.06	0.0045	1
Euro 6	-	-	-

Finally, instead of the traffic projections – used for England and Wales – the latest traffic data from DfT, where traffic (in VKT) for Greater London is provided for each of the five vehicle types for each year are utilized. Throughout the Greater London case study, the boundaries of ULEZ (introduced in 2019) and ULEZ extension (to be introduced in 2021)[54] are extracted from the London Datastore[55] as shown in Figure 5-5.

Figure 5-5: Inner, Outer, ULEZ (left) and ULEZ extension (right) zones in Greater London



[54] For clarity, ULEZ and ULEZ extension are henceforth referred as ULEZ2019 and ULEZ2021 respectively.
[55] The London Datastore is created and maintained by Greater London Authority (GLA) and provides free data and statistics about London.

5.3. Methodology

The methodology to estimate emissions comprises of two major stages. Firstly, in subsection 5.3.1, the major modelling approaches presented in section 2.7 are revised and the major issues and challenges for these approaches are discussed, to conclude on the most suitable approach for this research and justify the model selection. Then, in subsection 5.3.2, a detailed description of each step followed to estimate emissions for both case studies (i.e., England & Wales and Greater London) is provided.

5.3.1. Modelling approach

To estimate GHG and air pollutant emissions, one of the identified approaches discussed in section 2.7 has to be applied. However, from section 2.7 it can be concluded that traffic situation and instantaneous emissions models cannot be utilised. Traffic situation models require detailed statistics for speed and determination of each traffic situation for each road link (Baškovic and Knez, 2013) making these models unsuitable for extended application (Elkafoury et al., 2013). Also instantaneous models are very rare (Boulter et al., 2007) due to precision issues that have been identified when measuring emissions, which are normally related to the vehicles' operating conditions while taking the measurements (Ajtay and Weilenmann, 2004), as it is also discussed in section 2.6. In addition, the need to include traffic simulation models, requiring a wide range of data which are difficult to obtain, calibrate and process (Zhou et al., 2015), limits the use of instantaneous models. On the contrary, data required as inputs for average speed models is usually available and the models are comprehensive in terms of number of pollutants that can be modelled, vehicle types as well as influencing factors (Smit et al., 2009). However, they do not take into account different driving behaviours and operational modes (Sturm et al., 1998) usually resulting into different emissions and fuel consumption factors for the same speed (Elkafoury et al., 2013). Therefore, considering data availability emissions are estimated following the methodologies applied by Sookun et al., (2014) and Fu et al., (2017) among others, presented in subsection 2.7.4. That is, to estimate emissions for all street segments in the study area, traffic volumes – i.e., AADT – at locations where counts have not been measured are firstly estimated. Then VKT is calculated and finally, emission factors on the estimated VKT are applied. For England and Wales projections, the VKT growth and/or decline on the estimated traffic values is applied and then emissions are estimated by again applying the same emission factors on the updated VKT. A similar – although slightly different – approach is applied for Greater London.

5.3.2. Emissions estimation

As described, to estimate emissions three steps are undertaken. Firstly, the model presented in chapter 3 to estimate AADT at locations where traffic counters are not available is used. Secondly, VKT for these locations is calculated using the estimated AADT values. For both steps, data processing is conducted in GIS. Finally, emissions are estimated utilizing both estimated and measured traffic counts – and VKT – and the corresponding emission factors.

5.3.2.1. AADT estimation at unmeasured locations

To estimate traffic counts at unmeasured locations, "artificial" traffic counters at unmeasured street segments are created, by placing a point in the middle of each unmeasured segment using GIS. Specifically, the traffic count points are firstly matched with the spatial road dataset described in section 3.2 and the unmatched segments are considered to be unmeasured. Slip roads and roundabouts are considered as individual street segments. For each new point, the same methodology presented in section 3.3 is followed and service areas of different sizes are created, to consider the land use, socioeconomic, roadway and public transport characteristics, based on service areas selected for each road class. For example, for count points on 'A' roads only service areas of 500, 800 and 1600 metres are considered as shown in Table 5-4. The process results into traffic counters with the same ID occurring multiple times in the dataset, with variable values different for each service area.

Table 5-4: Service area by cluster for each road class

Road Class \ Cluster	1	2	3	4	5
	Service Area Size (m)				
A	800	1600	500	500	500
B	800	1000	800	800	2000
C	500	800	1000	800	2000
U	3200	800	1000	1000	2000

Then, each new point is allocated to the cluster with the most similar characteristics, and the most suitable service area for each point is identified, based on the classification process and associated problems discussed below. As the dataset created in chapter 3 incorporates values for AADT – i.e., the value to be estimated – the dataset is firstly randomly split – 80% for training and 20% for testing – and the dependent variable (i.e., AADT) is excluded. Then, the training set is used to train three different classification algorithms (i.e., Random Forest (RF), Gradient Boosting Machine (GBM), and K-nearest Neighbour) and the classification accuracy is evaluated by utilizing the testing set using a confusion matrix[56]. The process is repeated for each road class individually. Among the three algorithms, GBM provided the highest classification accuracy in both case studies (Table E-1 and Table E-2 in the Appendix) and it is therefore selected to classify the new points. Finally, to account for points with repeated IDs a GBM probabilistic classification (Friedman, 2001) is applied for all new points and service areas, so as to identify the probability of each point – and respective service area – belonging to each of the clusters. For example, for 'A' roads if a service area of 800 metres is selected for a point, then it should be assigned to cluster 1, but if a service area of 500 metres is selected, the point can belong to any of clusters 3, 4 or 5 as shown in Table 5-4. Therefore, a probabilistic approach will allow identifying the cluster where each point should be classified at.

GBM is essentially an ensemble of "weak" prediction models – usually decision trees – aimed to improve accuracy by minimising the average value of the loss function $L(y_i, F(x))$ on the training set

[56] For the case of London, the data points are firstly clipped to the London spatial boundaries, to select the points falling within the Greater London area. Then the same methodology is applied.

(Brown and Mues, 2012), for a number of M iterations (Rawi et al., 2017), where y_i is the observed value and $F(x)$ is the corresponding function. The probability of a point belonging to a class is given by:

$$\mathbb{P}(y|x) = \frac{e^{\log(odds)}}{1 + e^{\log(odds)}} \tag{15}$$

where $\log(odds)$ represents the odds of a point belonging to a class.

The algorithm, starts with a constant function:

$$F_0(x) = argmin \sum_{i=1}^{n} L(y_i, \gamma) \tag{16}$$

where γ is the value for log(odds).

For each iteration (i.e., $m = 1\ to\ M$), *pseudo-residuals* (i.e., the difference between the observed and predicted values) are computed for each data instance i:

$$r_i^m = -\left[\frac{\partial L\big(y_i, F(x_i)\big)}{\partial F(x_i)}\right]_{F(x)=F(m-1)(x)}, \forall_i = 1,2, \dots, n \tag{17}$$

and a weak learner $h_m(x)$ is fitted to the residual values. Moreover, the parameter γ_m is calculated:

$$\gamma_m = argmin \sum_{i=1}^{n} L\big(y_i, F_{(m-1)}(x_i) + \gamma h_m(x)\big) \tag{18}$$

and the model is updated as $F_m(x) = F_{(m-1)}(x) + \gamma_m h_m(x)$ where h_m is the weak learner for iteration *m* and γ_m is the corresponding value extracted from equation 18. Finally, the algorithm concludes into the final $F_M(x)$ output after M iterations.

After the new points are classified, to estimate AADT for each vehicle type, the Random Forest (RF) regression within each cluster is applied, using the existing points to train the algorithm and all the available independent variables[57]. RF is applied for regression as the algorithm with the highest

[57] Note, that clusters are now composed by the original points (i.e., traffic counters) as formed in chapter 3, as well as the new "artificial" counters created in this chapter to estimate emissions.

estimation accuracy for this model and dataset, based on the results from chapter 3. As it has been shown in Table 3-3, RF results into identical MAPE with SVR, while it has lower RMSE. Thus, it is considered the best performing algorithm that can provide more accurate estimations. Again, RF is a collection of decision trees based on bootstrapping and bootstrap aggregation (Breiman, 2001, 1996) and the regression is performed as:

$$\widehat{f_{rf}^{B}} = \frac{1}{B}\sum_{b=1}^{B} T_b(x) \tag{19}$$

where: B is the number of trees and $T_b(x)$ is the b^{th} tree grown from b bootstrapped data.

5.3.2.2. Vehicle Kilometres Travelled (VKT) and emission calculation

To calculate VKT the work of Kim et al., (2016) is followed and the AADT values are multiplied with the length of each street segment for all vehicle types.

$$VKT_{ij} = AADT_{ij} * \ length_j \tag{20}$$

where i and j represent vehicle type and traffic counter location respectively. For the new points, the length of each link is extracted from GIS.

To estimate GHG and air pollutant emissions, for the base year, the observed (i.e., known) and the estimated traffic count points are firstly merged. Then, the points are grouped based on their location as shown in Figure 5-1 – i.e., on motorways, urban/rural areas or in central, inner, and outer London – to account for the different vehicle fleet composition in these areas. Similarly, for the Greater London case study the points lying in Central, Inner and Outer London zones as shown in Figure 5-5[58] are distinguished. Then the vehicle fleet composition data extracted from NAEI are utilized and the proportion of vehicle types in each of these regions/zones is calculated. For example, in English urban

[58] Notice, that at this stage emissions are calculated for the pre-ULEZ base year (i.e., 2015). Consequently, the fleet composition data used for London are those presented in Figure 5-1 and not those shown in Figure 5-2 to Figure 5-4. Confusion may occur since ULEZ boundaries as shown in Figure 5-5 are essentially identical to those defining Central London.

areas, petrol cars account for 58% of the total fleet of cars. Hence, the corresponding VKT for petrol cars in these areas is set to be the total car VKT multiplied by 0.58. Finally, to estimate emissions for each vehicle, the estimated VKT is utilized and multiplied with the corresponding emission factor as follows[59]:

$$E_{if} = A_{if} F_{if} \tag{21}$$

where: E is the emissions expressed, A is the activity (i.e., VKT), F is the pollutant emission factor (in grams per km travelled), i indicates vehicle type and f indicates fuel type.

5.3.3. Scenario analysis and Ultra Low Emission Zone (ULEZ) – Emission projections

To estimate emission projections for the two scenarios in England and Wales, the VKT growth/decline from Table 5-2 for each vehicle type and region is firstly applied to the estimated VKT values for all points:

$$VKT'_{i,r} = VKTestimated_{i,r} * VKTgrowth_{i,r} \tag{22}$$

where i is the vehicle type and r is the region. For example, the projected VKT for cars the in East Midlands region is the estimated VKT times 1.226 for scenario 1 and the estimated VKT times 1.264 for scenario 2 (see Table 5-2). Then for each vehicle type i, the projected vehicle fleet composition is used to estimate VKT for each fuel type as described in 5.3.2.2. Finally, emissions are estimated using equation 21.

In the case of London, emission estimation for the ULEZ2019 case is conducted in three steps. Firstly, for ULEZ2019 the DfT road traffic data are used to identify the change (i.e., growth/decline) in traffic compared to the base year for each vehicle type. Changes in traffic are shown in Table E-3 in the Appendix. Then, the change is applied on the estimated VKT as follows:

[59] Emission factor for electric vehicles is considered to be zero.

$$VKT_{ULEZ2019_i} = VKT_{estimated_i} * \left(\frac{VKT_{2019_i}}{VKT_{2015_i}}\right) \tag{23}$$

where again i represents the vehicle type.

Secondly, the points are grouped based on their location as shown in Figure 5-5 – i.e., Inner/Outer London and ULEZ – and again the vehicle fleet composition data is utilized to calculate the proportion of vehicle type in each zone, based on Euro standards as shown in Figure 5-2 to Figure 5-4. Finally, emissions are then estimated using the Euro standard emission factors Table 5-3 and the new VKT values in equation 21.

For ULEZ2021 emissions estimation two assumptions are made. Firstly, it is assumed that traffic levels remain the same as in 2019[60], indicating that VKT is identical to 2019 levels. Secondly, it is assumed that the fleet within the extended ULEZ zone (i.e., ULEZ2021) is identical to the one within the current ULEZ (i.e., ULEZ2019), due to lack of further figures (i.e., data) that would inform an alternative decision. This means that the change in the total fleet composition within Greater London occurs only on a spatial context – the ULEZ zone extension – as shown in Figure 5-5. Then, the traffic count points based on their location[61] (Figure 5-5) are grouped to apply the fleet composition proportions on the VKT. Emissions are again calculated using equation 21 and the corresponding emission factors for Euro standards.

5.4. Results

In this section the estimated AADT, VKT, CO_2 and air pollutant emissions are presented for both case studies. Firstly, focus is placed on base year estimations and then on the scenarios and ULEZ projections. Results for base year estimations and projections are presented separately.

[60] This is done due to the fact that in 2020 traffic has been significantly decreased across Britain compared to 2019 (Department for Transport, 2020c), probably due to COVID-19 restrictions. Hence, it is assumed that in 2021, due to the relaxing of these restrictions, traffic levels will return – at least to a certain extent – to 2019 levels.

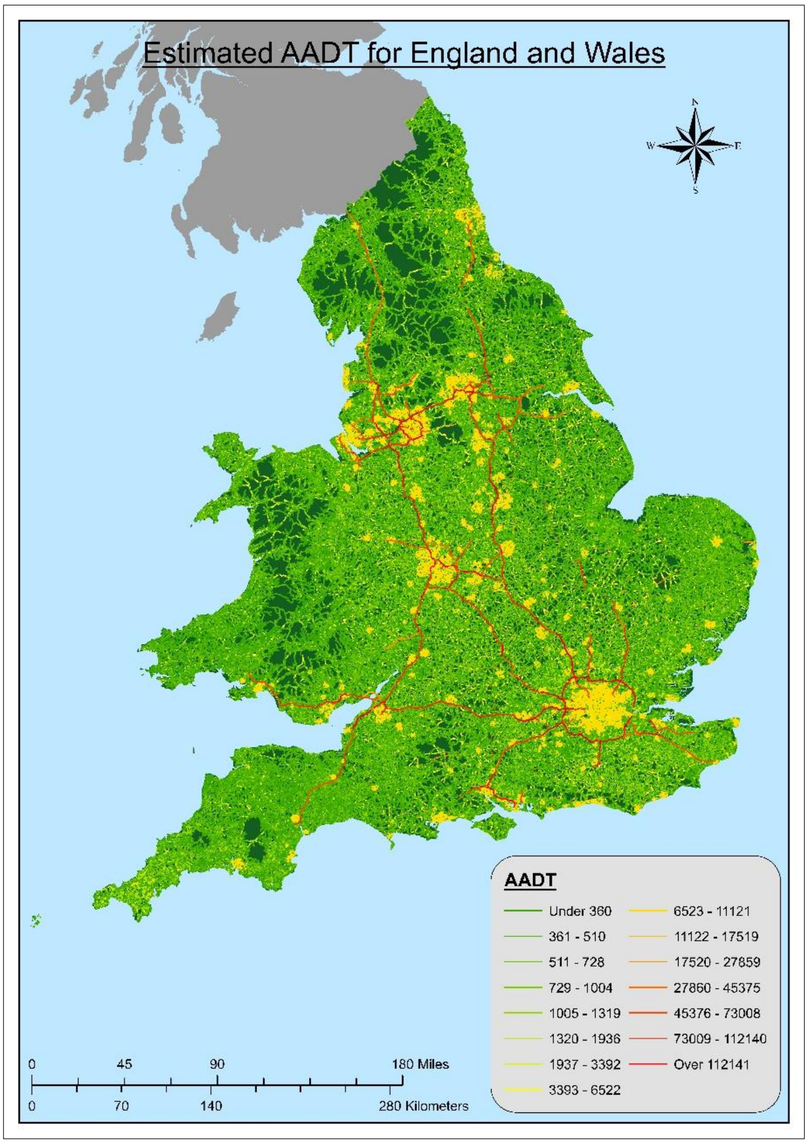
[61] From Figure 5-5 one can clearly notice that ULEZ extension (i.e., ULEZ2021) covers all central London and the majority of Inner London as well.

5.4.1. Base year AADT and emission estimations

5.4.1.1. England and Wales

In Figure 5-6 the estimated AADT for all street segments in England and Wales is presented, where roads with significantly higher traffic volumes can be distinguished. These are usually motorways and 'A' roads. Secondary (i.e., minor) roads in urban areas also appear to carry heavy traffic loads, while secondary roads in rural areas have lower AADT values.

Figure 5-6: AADT by street segment in England and Wales



The aggregated VKT estimations for all vehicle types and road classes in England and Wales are presented in Table 5-5. One can observe that streets are mainly dominated by cars. Interesting also appears the high volume of HGVs in Motorways as opposed to other road classes. On the contrary, estimated bus traffic volumes is lowest in motorways. LGVs also appear to be significantly present across all road types composing 14% of the total distance travelled by all vehicles. Two-wheeled vehicles have a small share – although larger than buses – and mainly operate in lower class (i.e., minor) roads.

Table 5-5: Aggregate VKT proportions by Vehicle Type and Road Class in England and Wales

Vehicle Type	**Road Class**					
	Motorways	**A**	**B**	**C**	**U**	**All roads**
Cars	73.39%	78.43%	81.09%	82.24%	79.72%	78.24%
Buses[62]	0.30%	0.83%	1.94%	1.42%	1.46%	0.96%
LGVs	14.62%	14.72%	13.35%	13.21%	12.41%	14.16%
HGVs	11.35%	5.14%	2.35%	2.14%	1.63%	5.52%
Two-wheeled	0.34%	0.87%	1.28%	1.00%	4.78%	1.13%
Total	22.08%	45.40%	10.31%	14.21%	8.00%	100%

Total average and annual[63] estimated emissions for each vehicle type are presented in Table 5-6. Again, it is observed that highest emissions levels are originating from cars. Moreover, CO_2 emissions are similar for LGVs and HGVs although the number of HGVs and associated VKT is much lower compared to LGVs.

Table 5-6: Daily average and total annual emissions by vehicle type (thousand tonnes) in England and Wales

Vehicle Type	**NOx**	**$PM_{2.5}$**	**CO**	**CO_2**
Cars	345.58	8.49	582.77	170,460
Buses[6262]	48.79	0.56	16.21	4,866
LGVs	219.35	3.05	70.71	43,969
HGVs	96.58	1.32	33.30	52,025
Two-wheeled	2.17	0.12	77.70	1,572
Total – Average Daily	712.47	13.54	780.69	272,892
Total – Annual	260,051.55	4,942.1	284,951.85	99,606,710

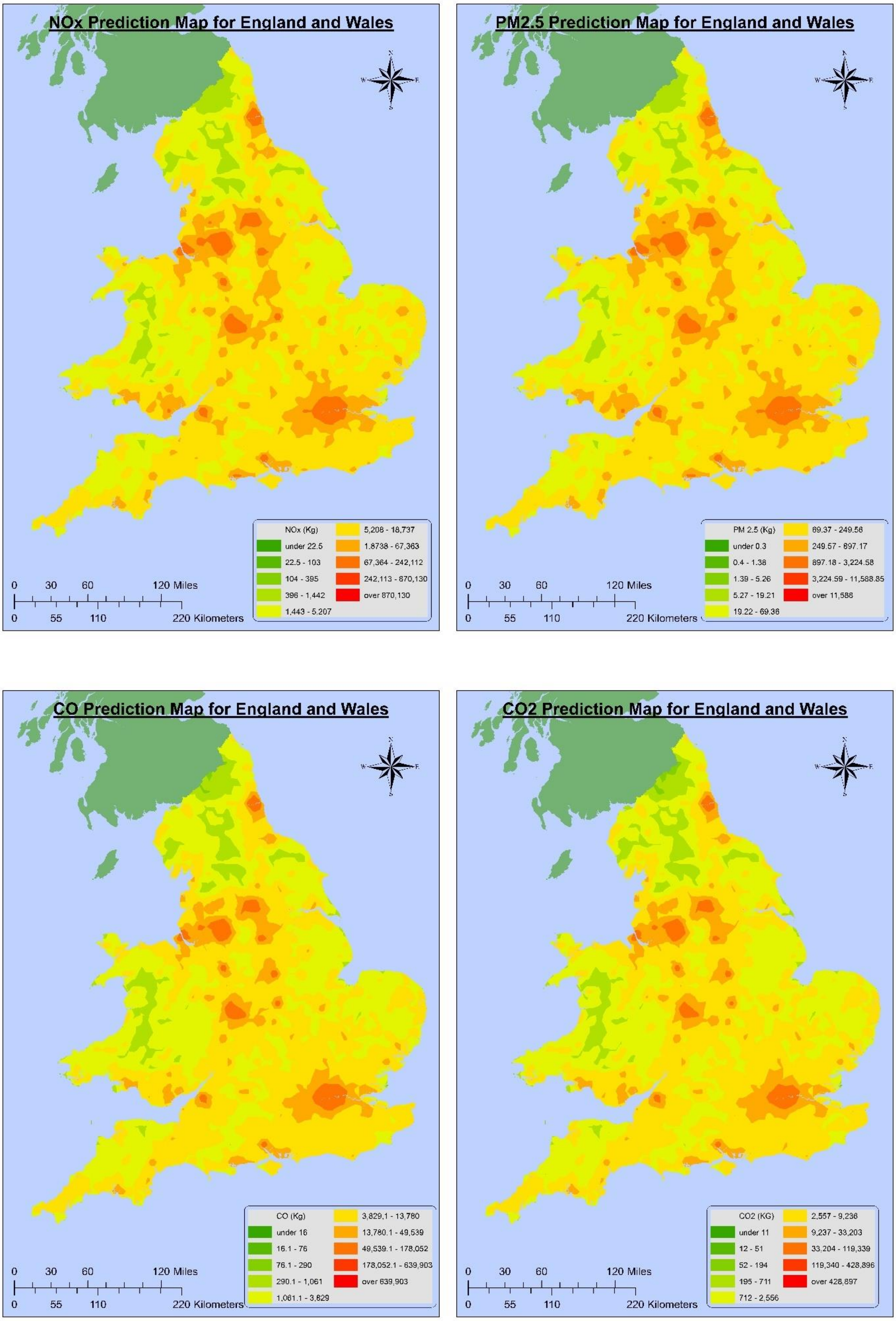
[62] Note that this vehicle type includes buses and coaches (Department for Transport, 2014).
[63] Annual emissions are estimated by multiplying Average Daily Emissions by 365.

In Figure 5-7 the estimated emissions are shown as heat maps for all air pollutants and CO_2, where similar patterns can be observed due to the fact that emissions are driven by the same factors – i.e., the number and type of vehicles and the corresponding length of the streets. In fact, at first sight, one can consider the maps being identical, although this is imprecise. The figures do reveal similar patterns, but at a closer look the differences can be observed specifically when focusing on less populated areas, such as small urban centres. Similarities in the figures occur firstly due to the same driving factors, but also due to the visualisation techniques used to create the figures. For visualisation and computational processing purposes emissions are mapped based on a 1km x 1km grid that has been created within a GIS environment.

Considering that more vehicles and more streets can be found in urban areas the associated emissions in these areas are also expected to be higher. This is clearly seen in Figure 5-7 where higher pollution levels can be observed in and around urban and major urban areas and conurbations, such as London, Birmingham, West Yorkshire as well as the Manchester – Liverpool conurbation, also complying with the traffic volume estimation shown in Figure 5-6.

On the contrary less populated areas, where less vehicles operate exhibit lower emissions. In particular significantly lower levels of pollution can be observed in rural and mountainous regions – such as the Cambrian mountains in Wales – and national parks (e.g., the Lake District in the North West and Northumberland national park in the North East of England), reflecting the lower levels of traffic as well as the lack of roads in these areas as again shown in Figure 5-6.

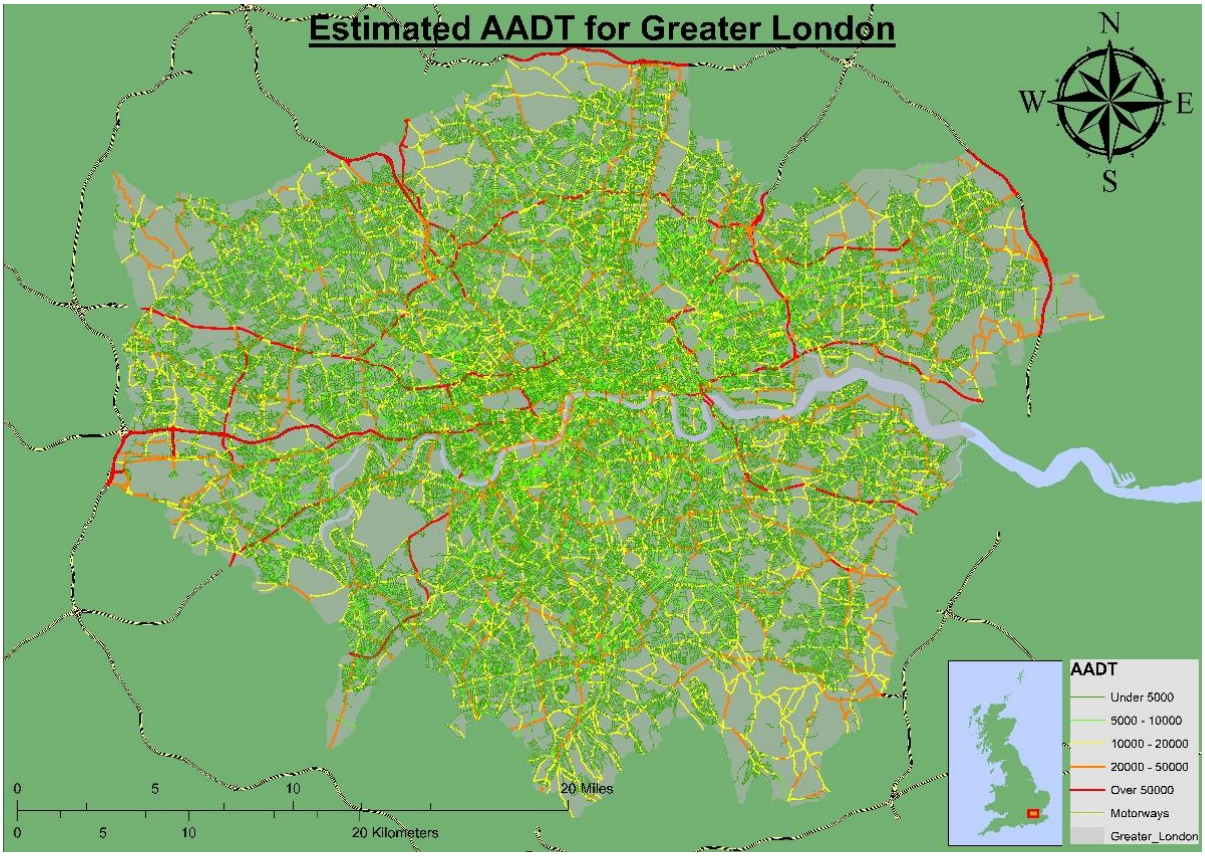
Figure 5-7: NOx (top left), $PM_{2.5}$ (top right), CO (bottom left) and CO_2 (bottom right) emissions distribution in England and Wales



5.4.1.2. Greater London

In Figure 5-8 the estimated AADT for all street segments in the Greater London area are presented, where again roads with significantly higher traffic volumes can be distinguished. Similar to the whole of England and Wales, these are usually 'A' roads and parts of motorways, although there is just a short length of motorway roads within Greater London. Roads around Heathrow airport, river crossings and major arteries also appear to carry heavy traffic loads, while secondary roads and streets in residential areas have lower AADT values.

Figure 5-8: AADT by street segment in Greater London



The aggregated VKT estimations for all vehicle types and road classes are presented in Table 5-7. Again, one can observe that street traffic is mainly dominated by cars, although there is a low number of registered cars as opposed to other areas in the country (Figure 3-4). High volume of HGVs is observed in Motorways as opposed to other road classes, while LGVs comprise 13% of the total distance travelled. Bus and two-wheeled vehicles' volumes are similar at approximately 2% each, where one can also observe that the figures for these two vehicle types are nearly double compared to the figures for England and Wales (Table 5-5).

Table 5-7: Aggregate VKT proportions by Vehicle Type and Road Class in Greater London

Vehicle Type	Road Class					
	Motorways	A	B	C	U	All roads
Cars	72.79%	77.37%	76.46%	81.65%	80.60%	78.62%
Buses[62]	0.43%	2.46%	5.06%	2.19%	1.63%	2.28%
LGVs	15.91%	13.70%	13.11%	12.07%	12.98%	13.29%
HGVs	10.18%	3.78%	2.10%	2.84%	2.29%	3.61%
Two-wheeled	0.69%	2.69%	3.26%	1.26%	2.49%	2.21%
Total	6.76%	44.60%	6.4%	24.32%	17.91%	100%

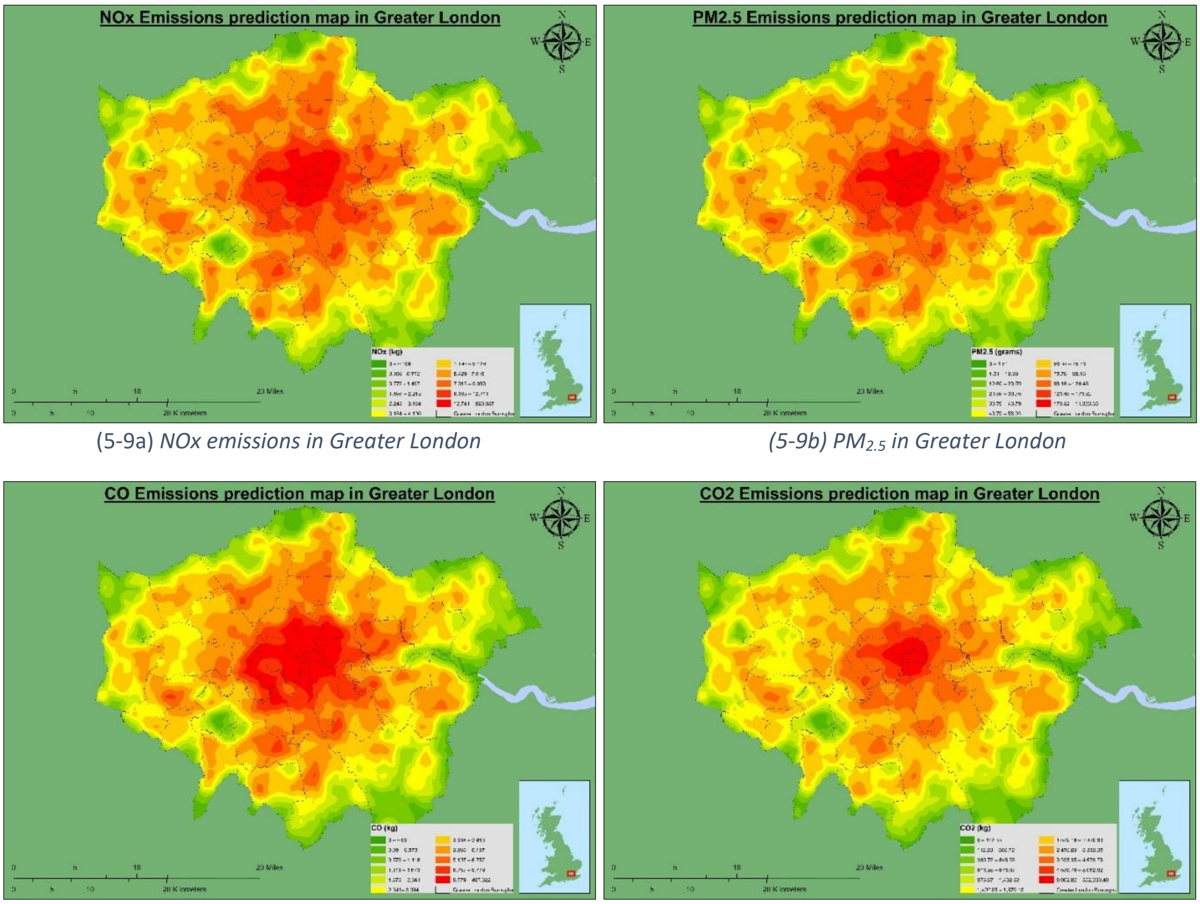
Total average and annual estimated emissions for each vehicle type are presented in Table 5-8. Again, it is observed that highest emissions levels are originating from cars. Moreover, CO_2 emissions are similar for LGVs and HGVs although the number of HGVs and associated VKT is much lower compared to LGVs.

Table 5-8: Daily average and total annual emissions by vehicle type (tonnes) in Greater London

Vehicle Type	NOx	$PM_{2.5}$	CO	CO_2
Cars	23.93	0.38	24.62	14,810
Buses[62]	9.50	0.10	2.70	0,181
LGVs	13.77	0.18	1.88	3,446
HGVs	7.66	0.09	2.35	3,038
Two-wheeled	0.28	0.02	8.74	0,250
Total – Average Daily	55.14	0.77	40.30	21,725
Total – Annual	20,126.10	281.05	14,709.50	7,929,625

In Figure 5-9 emissions for the estimated pollutants and CO_2 are shown with London boroughs overlaid with heat maps for each pollutant and CO_2. Higher pollution can be observed in central and inner London as opposed to boroughs located far from the city centre. South London boroughs have lower-level emissions compared to the north part of the city, reflecting the lower levels of traffic shown in Figure 5-8.

Figure 5-9: Air Pollutants and CO_2 estimation heat maps in Greater London



(5-9a) *NOx emissions in Greater London*

(5-9b) $PM_{2.5}$ in Greater London

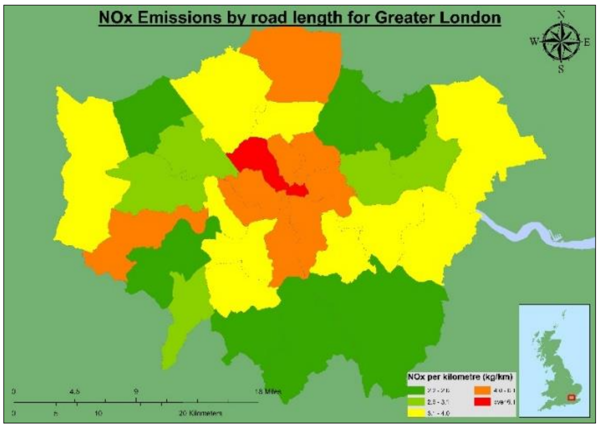
(5-9c) CO emissions in Greater London

(5-9d) CO_2 emissions in Greater London

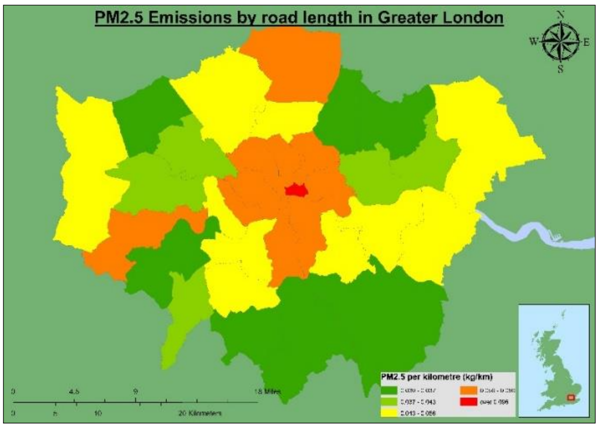
Similar patterns are observed in Figure 5-10 where emissions for each borough are mapped by the length of streets within the borough, indicating kilograms of pollutant per kilometre of street length (kg/km). One can observe that some boroughs in the north part of the city have higher values of emissions, particularly in the case of NOx and $PM_{2.5}$.

Interestingly, one can identify the impacts of large open spaces such as Richmond Park indicated by green patches on the maps as well as the – negative – impacts of town centres and busy streets, indicated by smaller red patches around the city centre (Figure 5-11).

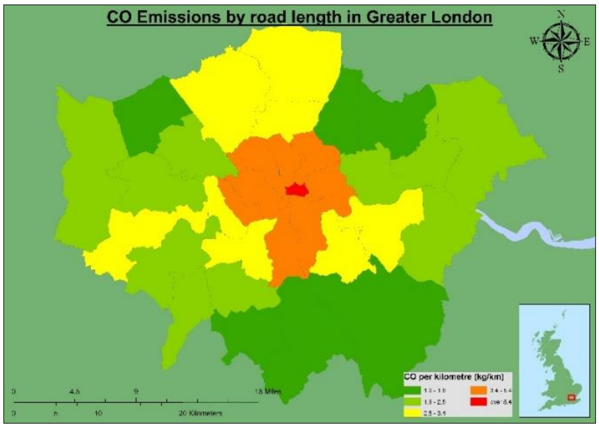
Figure 5-10: Air Pollutants and CO_2 emissions in kilograms per street kilometre (kg/km) for London boroughs



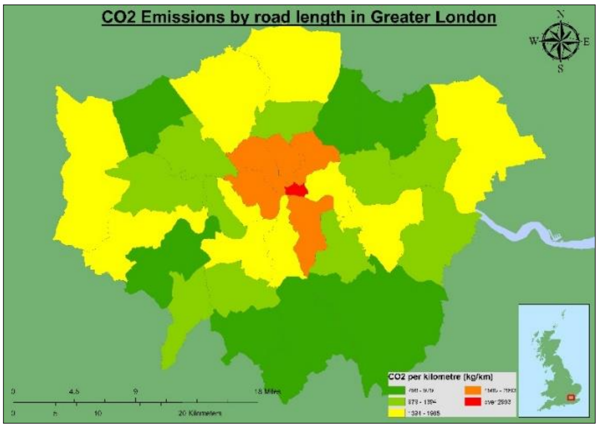
(5-10a) NOx by road length in Greater London boroughs



(5-10b) $PM_{2.5}$ by road length in Greater London boroughs

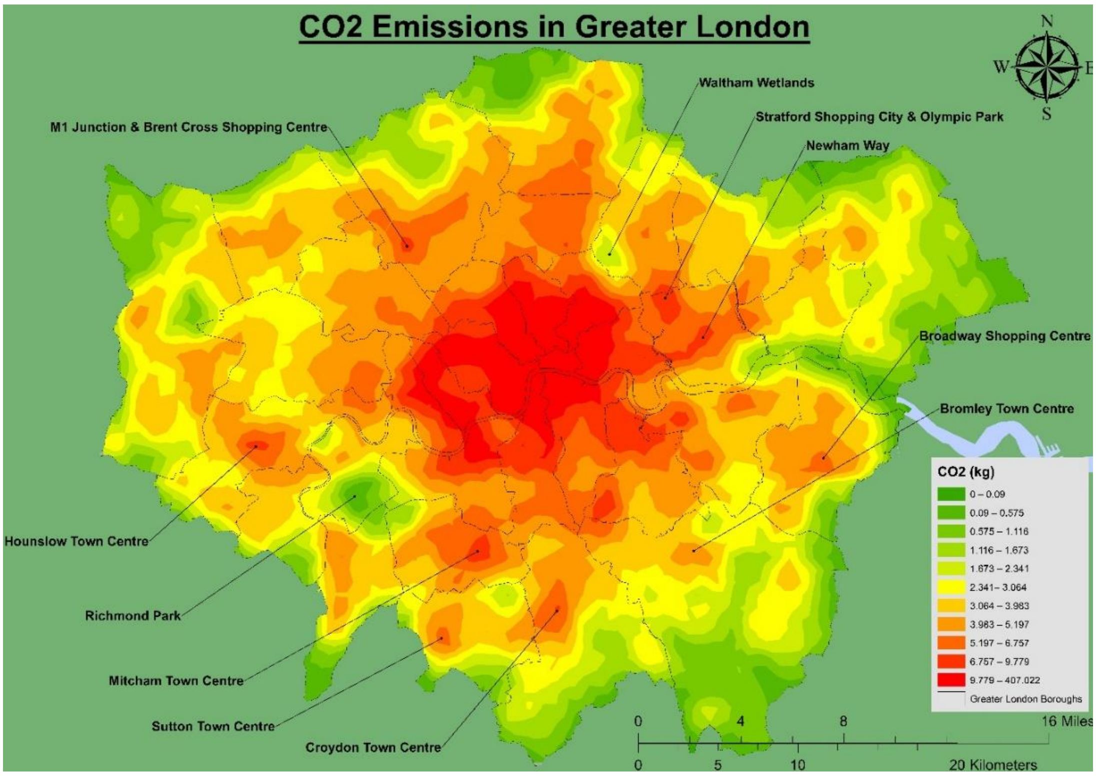


(5-10c) CO_2 by road length in Greater London boroughs



(5-10d) CO by road length in Greater London boroughs

Figure 5-11: Open spaces and town centre's impact on CO_2 emissions in Greater London



5.4.2. Scenarios and ULEZ emission estimations

5.4.2.1. England and Wales

In Table 5-9 the estimated emissions from the two selected scenarios are summarised. One can observe that for both scenarios there is a decrease in all air pollutants and CO_2 emissions compared to the base year. In particular, significant reductions for all air pollutant emissions can be observed for both scenarios, with NOx emissions reduced by 62% in the case of the "reference" scenario (i.e., scenario 1) and by 54% for the "High GDP, Low Fuel" scenario (i.e., scenario 2). $PM_{2.5}$ is reduced by 80% for both scenarios and CO emissions are reduced by 47% in the case of scenario 1 and by 44% in scenario 2 respectively. CO_2 emissions are also reduced based on the estimations for both scenarios, although reductions are lower compared to the air pollutants. Specifically, CO_2 is reduced by 15% for scenario 1 and by 12% for scenario 2.

Table 5-9: Projected emission estimations (thousand tonnes) for England and Wales

Pollutant	Base year	Scenario 1	Scenario 1 – Reductions (%)	Scenario 2	Scenario 2 – Reductions (%)
NOx	260.05	98.43	62.15%	118.55	54.41%
$PM_{2.5}$	4.90	0.94	80.82%	0.98	80.00%
CO	284.95	149.77	47.44%	158.79	44.27%
CO_2	99,606.71	84,917.64	14.75%	88,040.45	11.61%

In Table 5-10 to Table 5-13 estimated emissions for each air pollutant and CO_2, by vehicle type are shown. In the case of NO_x (Table 5-10), emission reduction for cars is approximately 68 thousand tonnes for both scenarios, while emissions from LGVs are also significantly reduced by 51 and 32 thousand tonnes for each scenario respectively. However, the relative reductions appear to be higher for buses and HGVs, where NO_x emissions from buses are reduced by 86% for both scenarios. HGV emissions are also reduced significantly by 76% and 74% for scenario 1 and scenario 2 respectively.

Table 5-10: NOx emission projections (thousand tonnes) by vehicle type in England and Wales

Vehicle Type	Base Year	Scenario 1	Scenario 1 – Reductions (%)	Scenario 2	Scenario 2 – Reductions (%)
Cars	126.14	57.53	54.40%	58.11	53.93%
Buses[62]	17.81	2.56	85.63%	2.56	85.63%
LGVs	80.06	29.35	63.34%	48.09	39.93%
HGVs	35.25	8.56	75.72%	9.1	74.18%
Two-wheeled	0.79	0.44	44.30%	0.69	12.66%
Total	260.05	98.43	62.15%	118.55	54.41%

In the case of $PM_{2.5}$, cars account for the largest proportion of total $PM_{2.5}$ emissions. In Table 5-11 it can be observed that emissions from cars are reduced by 2.4 thousand tonnes in both scenarios, contributing more than 50% on the overall emission reductions. Again, high relative reductions can be observed for HGVs at 87.5% as well as LGVs where $PM_{2.5}$ emissions are significantly reduced by 89% for scenario 1 and 86% at scenario 2.

Table 5-11: $PM_{2.5}$ emission projections (thousand tonnes) by vehicle type in England and Wales

Vehicle Type	Base Year	Scenario 1	Scenario 1 – Reductions (%)	Scenario 2	Scenario 2 – Reductions (%)
Cars	3.10	0.66	78.71%	0.67	78.39%
Buses[62]	0.20	0.10	50.00%	0.10	50.00%
LGVs	1.11	0.12	89.19%	0.15	86.49%
HGVs	0.48	0.06	87.50%	0.06	87.50%
Two-wheeled	0.004	0.002	50.00%	0.003	25.00%
Total	4.90	0.94	80.82%	0.98	80.00%

In Table 5-12 CO emissions for the base year and the reductions for each scenario are shown. Car emissions are reduced by approximately 90 thousand tonnes for both scenarios, while significant reductions are observed in the case of LGVs by 16.75 thousand tonnes (65%) for scenario 1 and 15.28 tonnes (59%) for scenario 2. Interestingly, very large reductions in CO emissions can be seen for two-wheeled vehicles, where in scenario 1 emissions are reduced by 89% – approximately 25 thousand tonnes – and by 68% (19 thousand tonnes) in scenario 2. LGVs' emissions are also reduced significantly, although emissions from HGVs are only slightly reduced as opposed to other pollutants and vehicle types.

Table 5-12: CO emission projections (thousand tonnes) by vehicle type in England and Wales

Vehicle Type	Base Year	Scenario 1	Scenario 1 – Reductions (%)	Scenario 2	Scenario 2 – Reductions (%)
Cars	212.71	121.59	43.84%	122.91	42.22%
Buses[62]	5.92	4.12	30.41%	4.12	30.41%
LGVs	25.81	9.06	64.90%	10.53	59.20%
HGVs	12.16	11.86	2.47%	12.11	0.41%
Two-wheeled	28.36	3.14	88.93%	9.11	67.88%
Total	284.95	149.77	47.44%	158.79	44.27%

CO_2 emissions computations in Table 5-13 show that significant gains can be achieved in both scenarios for cars. In particular, emissions are reduced by approximately 21% in both cases, saving about 13,000 thousand tonnes of annual CO_2 production by private cars and taxis. CO_2 abatement is also high for buses and two-wheeled vehicles. In the case of LGVs and HGVs that contribute about 35% to the total emissions – in base year – smaller gains are observed, where for scenario 2 a small increase in annual emissions is also estimated.

Table 5-13: CO_2 emission projections (thousand tonnes) by vehicle type in England and Wales

Vehicle Type	Base Year	Scenario 1	Scenario 1 – Reductions (%)	Scenario 2	Scenario 2 – Reductions (%)
Cars	62,218.13	48,994.97	21.25%	49,409.26	20.59%
Buses[62]	1,776.35	1,542.28	13.17%	1,542.28	13.17%
LGVs	16,048.87	15,380.15	4.17%	16,485.93	+2.72%
HGVs	18,989.29	18,810.28	0.94%	20,147.43	+6.10%
Two-wheeled	574.06	239.95	58.20%	455.55	20.64%
Total	99,606.71	84,917.64	14.75%	88,040.45	11.61%

5.4.2.2. Greater London

In Table 5-14 to Table 5-16 the computed air pollutant emissions resulting from ULEZ2019 and ULEZ2021 are shown by vehicle type. Similar to the abatements observed for England and Wales the largest emission reductions are related to cars, since these vehicles have the largest share on the streets in London (Table 5-7). NO_x emissions for cars (Table 5-14) are estimated to be reduced by 14.95 tonnes on a daily basis, saving approximately 5,500 tonnes annually with the current ULEZ, while the zone extension is estimated to save approximately 200 tonnes additionally per year. One can also observe the significant reductions for buses by 83%, equal to 7.9 tonnes of NO_x daily – 2,884 annually – and 74% abatement in the case of HGVs, reducing emissions by 2,000 tonnes per year, in the case of ULEZ2019. An extension of the zone will result into total savings of nearly 3,000 tonnes annually for buses and 2,120 for HGVs, compared to the base year. Two-wheeled vehicles contribute only slightly to emissions savings, by approximately 4% resulting in 3.65 tonnes less NO_x emissions from this vehicle type per year, in the case of ULEZ2021. However, an increase in emissions from LGVs is observed in both cases, although an expansion of the zone will result into a decrease in NO_x emissions from LGVs by 22%. Nevertheless, the total emission reduction from all vehicle types is remarkable for both cases at 31% for ULEZ2019 and 43% for ULEZ2021 compared to the base year, which can be translated to 6,200 and 8,600 less tonnes of NO_x per year respectively.

Table 5-14: NOx emission projections (tonnes) by vehicle type in Greater London

Vehicle Type	Base Year	ULEZ2019	ULEZ2019 change (%)	ULEZ2021	ULEZ2021 change (%)	2019 – 2021 change (%)
Cars	23.93	8.98	-62.47%	8.45	-64.69%	-5.90%
Buses[62]	9.50	1.60	-83.16%	1.33	-86.00%	-16.88%
LGVs	13.77	25.27	+83.51%	19.66	+42.77%	-22.20%
HGVs	7.66	2.03	-73.50%	1.85	-75.85%	-8.87%
Two-wheeled	0.28	0.28	0.00%	0.27	-3.57%	-3.57%
Total – Average Daily	55.14	38.16	-30.79%	31.56	-42.77%	-17.30%
Total – Annual	20,126.10	13,928.40	-30.79%	11,519.40	-42.77%	-17.30%

Similar to the NO_x emissions, the impact of ULEZ and the extension is evident on $PM_{2.5}$, although to a lesser extent, as it can be seen in Table 5-15. Specifically, in the case of ULEZ2019 emissions are reduced for all vehicle types, except of cars that contribute the most due to the large number of these vehicles on the streets. The estimations show that even with the current ULEZ regulations, $PM_{2.5}$

emissions increase compared to the base year. An extension of the zone is estimated to reduce the emissions only by 1%. Although emissions from cars are still projected to rise compared to the base year, the ULEZ extension is estimated to reduce $PM_{2.5}$ from cars by 23% (approximately 60 tonnes annually) compared to the current conditions. Total emission reductions during the expansion are estimated to reduce emission by 1,300 tonnes per year compared to the base year.

Table 5-15: $PM_{2.5}$ emission projections (tonnes) by vehicle type in Greater London

Vehicle Type	Base Year	ULEZ2019	ULEZ2019 change (%)	ULEZ2021	ULEZ2021 change (%)	2019 – 2021 change (%)
Cars	0.38	0.71	+86.84%	0.55	+44.74%	-22.54%
Buses[62]	0.10	0.03	-70.00%	0.02	-80.00%	-33.33%
LGVs	0.18	0.16	-11.11%	0.14	-22.22%	-12.50%
HGVs	0.09	0.04	-55.56%	0.04	-55.56%	0.00%
Two-wheeled	0.02	0.018	-10.00%	0.017	-15.00%	-5.56%
Total – Average Daily	0.77	0.95	+23.38%	0.76	-1.30%	-20.00%
Total – Annual	281.05	346.75	+23.38%	277.40	-1.30%	-20.00%

Again, as with the other pollutants, CO emissions are significantly reduced in both cases compared to the base year. From Table 5-16 one can observe that total emissions are reduced by 45%, equivalent to 6,600 tonnes per year in the case of ULEZ2019, while the extension of the zone is expected to reduce CO by 12% and an additional 1,000 tonnes annually. Cars contribute the most in emission reductions by 9.78 tonnes daily, resulting to a total abatement of 3,500 tonnes annually – essentially more than 50% of the total abatement. The estimated savings from cars for ULEZ2021 are 2.52 tonnes daily compared to the current ULEZ estimations, resulting into a reduction of more than 900 tonnes per year and over 90% of the total reductions. Significant reductions are also observed for LGVs by 48% and two-wheeled vehicles by 70% considering ULEZ2019, contributing to emission savings by 330 tonnes and 2,250 tonnes of CO on an annual basis respectively. An extension of the zone is estimated to reduce CO emissions from LGVs by further 5%, although further emission reduction from two-wheeled vehicles is minimal. Finally, emissions from buses and HGVs are also reduced, although to a lesser extent, resulting into 120 and 157 tonnes per year respectively for ULEZ2019 and additional savings of approximately 18 tonnes combined in the case of a zone extension.

Table 5-16: CO emission projections (tonnes) by vehicle type in Greater London

Vehicle Type	Base Year	ULEZ2019	ULEZ2019 change (%)	ULEZ2021	ULEZ2021 change (%)	2019 – 2021 change (%)
Cars	24.62	14.84	-39.72%	12.32	-49.96%	-16.98%
Buses[62]	2.70	1.98	-26.67%	1.97	-27.04%	-0.51%
LGVs	1.88	0.97	-48.40%	0.92	-51.06%	-5.15%
HGVs	2.35	1.84	-21.70%	1.80	-23.40%	-2.17%
Two-wheeled	8.74	2.56	-70.71%	2.54	-70.94%	-0.78%
Total – Average Daily	40.30	22.19	-44.94%	19.55	-51.49%	-11.90%
Total – Annual	14,709.50	8,099.35	-44.94%	7,135.75	-51.49%	-11.90%

5.5. Discussion

This chapter focused on tailpipe emission estimation of NOx, $PM_{2.5}$, CO and CO_2 originating from road transport in England and Wales and separately for the Greater London area. Moreover, emissions have been projected for 2035 for England and Wales, considering two scenarios developed by DfT and potential emission abatement for the three air pollutants has been estimated considering the current ULEZ and potential ULEZ extension in the Greater London area. All estimations have been conducted for five road classes and five different vehicle types used by DfT in the UK, by combining data from numerous sources and applying classification and regression modelling. In this section, the results are discussed. Moreover, an attempt to assess the findings, using base year and projected estimations from other models used by DfT and NAEI has been conducted.

In the case of England and Wales, the higher levels of pollution are concentrated around urban and in particular major urban areas as expected and has also been confirmed by similar studies in other areas (e.g. Alotaibi et al., 2019; Schmitz et al., 2018). This can be explained by the fact that higher traffic volumes are usually observed in these areas (Caselli et al., 2010; Rahman et al., 2021) as it can also be seen by the AADT estimations in Figure 5-6 and also discussed in chapter 4.

Base year estimations from the modelling approach presented in this chapter are also comparable with the published DfT[64] data as shown in Table 5-17. One can observe that VKT estimations from my model are very similar to the DfT modelling approach using the NTM. Emission estimations are also similar although, deviations are slightly larger.

Table 5-17: Base year estimations comparison for England and Wales (in thousand tonnes)

	VKT[65]	NOx	$PM_{2.5}$	CO	CO_2
DfT	457.80	273.27	5.50	266.00	96,167.00
Author's Method	452.97	259.26	4.90	256.59	99,032.65
Deviation (%)	1.06%	5.13%	10.99%	3.54%	2.98%

Similarly, the VKT, air pollutants and CO_2 estimations are again comparable with the DfT scenario outputs, where similar patterns can be observed (Table 5-18 and Table 5-19). In particular, VKT estimations are almost identical with the figures provided by DfT, while NO_x and CO_2 exhibit low deviations. Again, $PM_{2.5}$ appears to be the most challenging pollutant to be modelled.

Table 5-18: Scenario 1 estimations comparison for England and Wales (in thousand tonnes)

	VKT[65]	NOx	$PM_{2.5}$	CO	CO_2
DfT	559.40	103.70	0.72	N/A	77,148.00
Author's Method	564.00	97.99	0.94	146.63	84,677.69
Deviation (%)	0.82%	5.51%	30.5%	N/A	9.76%

Table 5-19: Scenario 2 estimations comparison for England and Wales (in thousand tonnes)

	VKT[65]	NOx	$PM_{2.5}$	CO	CO_2
DfT	586.30	110.50	0.76	N/A	81,084.00
Author's Method	584.39	117.86	0.98	149.68	87,584.90
Deviation (%)	0.33%	6.66%	28.95%	N/A	8.02%

[64] The Department for Transport (DfT) excludes two-wheeled vehicles from the estimations. Hence, the estimations from my method shown in Table 5-17, also exclude these vehicles, so as a reliable comparison can take place.
[65] In billion Vehicle Kilometres Travelled (VKT).

The low deviations in VKT values can be straightforwardly explained. Considering that the base year estimations are comparable, and traffic growth factors from the selected scenarios are applied on the base year estimations, one would expect small deviations at the projected figures. However, the – larger – deviations on the estimated emissions can be attributed to several factors. Firstly, linear assumptions about the use of vehicles have been considered. For example, it is assumed that all different vehicle types – as presented in the vehicle fleet composition in Figure 5-1 – are used with the same frequency, while – for instance – petrol cars may be operational for longer time frames as opposed to electric vehicles. Secondly, the emission projections provided by DfT are estimated using a combination of speed flow and speed emission curves (Department for Transport, 2018d) derived from NTM, which is a simulation model. Although a simulation approach would be expected to better capture behavioural aspects of road transport (Munigety and Mathew, 2016), and hence associated emissions, the modelling approach used by DfT is based on average curves (Department for Transport, 2018d) and consequently detailed – local – figures may be questionable.

In terms of specific pollutants, it is worth commenting on the high contribution of HGVs on CO_2. Specifically, it has been seen that the distance travelled by HGVs is only one third compared to LGVs (Table 5-5), although HGVs contribute more to CO_2 emissions (Table 5-6 and Table 5-13). This outcome agrees with findings from similar studies (e.g., Aditjandra et al., 2016; Jacyna-Gołda et al., 2017) and can be attributed to the high emission factors for these vehicles, placing them in the top rank of CO_2 emissions by km driven among the five vehicle types. However, this does not apply to CO emissions, where HGVs have lower contribution as opposed to other vehicles (Table 5-6), also explaining the low reduction in CO emissions in the case of scenarios (Table 5-12).

Emissions from all vehicles are also overall higher for scenario 2 compared to scenario 1, reflecting the higher increase in traffic for scenario 2 as it is shown in Table 5-2. This also reflects the projections for LGVs and HGVs, where in the case of scenario 2 it is estimated that CO_2 emissions from these two vehicle types will be higher compared to the base year. One should notice that for scenario 2, VKT for

LGVs is expected to grow by 50% and for HGVs by 4.9% (Table 5-2). Although VKT growth for HGVs is slightly higher in scenario 1 (5%), the projected CO_2 emissions are lower. This can be attributed to the different levels of traffic across regions. For example, regions with higher traffic may be given higher traffic growth values and therefore affecting the total VKT – and associated emissions – for HGVs, which can also be inferred by the aggregated VKT, and emission figures shown in Table 5-18 and Table 5-19.

In the case of Greater London, the results show higher levels of pollution in the city centre as expected, as it can also be confirmed by similar studies in other urban areas (e.g., Fu et al., 2017; Munir et al., 2020). For London in particular, the results concur with the ones presented in Fecht et al., (2016) where $PM_{2.5}$ distribution is studied and higher levels of pollution are concentrated in central and inner London, as well as major road arteries. Estimated emissions from the method presented in this chapter, are also comparable in most cases – again with the exception of $PM_{2.5}$ – with estimations from the London Atmospheric Emissions Inventory (LAEI) forming fraction of the NAEI for London (LAEI, 2013). The aggregated emission estimations are provided in Table 5-20. The large deviation in estimations observed for $PM_{2.5}$ can be attributed to the methodology and factors taken into consideration when estimating emissions by LAEI. In the case of $PM_{2.5}$ emissions from breaks and tyres are also considered on top of the exhaust emissions.

Table 5-20: Estimation comparison for Greater London- tonnes/year

	NOx	**$PM_{2.5}$**	**CO**	**CO_2**
LAEI	23,852.5	1,253.4	N/A	6,651,511
Authors' Method	20,126.10	281.05	14,709.50	7,929,625

Unfortunately, at the time of writing the thesis, the impacts of ULEZ and the corresponding ULEZ extension have not been investigated from an environmental perspective so far, probably due to the fact that results from ongoing studies are not yet publicly available due to the recent implementation of the scheme. Consequently, the estimates cannot be directly assessed or compared with other similar studies. However, some, preliminary results presented in reports suggest that ULEZ has reduced NO_x emissions by 30% (Laurent, 2021; Osei et al., 2021), while there is also the argument that NO_x and $PM_{2.5}$ emissions have dropped by 31% and 15% respectively (Bernard et al., 2020) within the current ULEZ boundaries – covering central London. Assuming reliability of these reports, the computation for NO_x estimations presented in Table 5-14 conform with these findings, showing a 30% reduction for this pollutant, while my computation for $PM_{2.5}$ indicates an increase, again showing the challenges in estimating $PM_{2.5}$ emissions. Another noteworthy feature in Table 5-14 is the increase in NO_x emissions from LGVs for ULEZ2019 that can be attributed to the 26% traffic increase compared to base year (Table E-3), indicating that current policy implementations are inadequate to reduce the air pollution impact from these vehicles.

Significant VKT growth by 29% can also be noticed for two-wheeled vehicles (Table E-3), that are well known to contribute substantially to CO emissions (Iodice and Senatore, 2015), where in some cases it has been found that they emit three times more CO compared to conventional cars (Vasic and Weilenmann, 2006). This is also reflected in the high levels of CO produced by these vehicles as shown in Table 5-16, where two-wheeled vehicles account for over 10% in all cases in London. However, the implementation of clean air policies such as ULEZ has the potential to significantly reduce CO emissions both for two-wheeled and other vehicle types, as shown in my results presented in Table 5-12 and Table 5-16.

Nevertheless, it is interesting to comment on the overall modelling results and deviations on the emission estimations, observed on the base year and projections for both case studies. Deviations in estimations from statistical approaches compared to standard emissions models are subject to various factors also reported in similar studies in Greater London and the UK as well as other countries, while significant deviations in estimations can be observed among different emission models. For example Fontaras et al., (2014) compared COPERT and HBEFA to identify deviations in the estimations and that both over-predicted PM emissions, with HBEFA results deviating more from actual measurements. In the case of NAEI which is widely used in the UK estimations have been usually found to deviate from other modelling approaches. For example, Vaughan et al., (2016) indicate that NAEI tends to overestimate NO_x emissions, although it seems to underestimate Volatile Organic Compounds (Valach et al., 2015) – the latter pollutant not examined in this study. On the contrary Chatterton et al., (2015) compare NOx, PM_{10} and CO_2 estimation with NAEI to find strong correlations, although in this study annual distance travelled is used and there is no differentiation between vehicle types.

However, over and under estimations cannot be solely attributed to each modelling approach per se, but can be related to limitations in methodology and available data (Borge et al., 2012). In the approach presented in this thesis, the modelling limitations should also be considered, such as the classification and regression outcomes, the length of roads within the study area and the study area per se. First, from Table E-1 and Table E-2 it can be observed that classification accuracy is significantly higher for 'A' roads, as opposed to other road classes, although accuracy for 'U' roads is relatively high for both case studies as well – 71% for England and Wales, and 70% for London. Thus, an uncertainty in the classification of points and consequently in the estimation of AADT and VKT can result into biased emission estimations. However, from Table 5-5 and Table 5-7 one can see that VKT on 'A' roads and Motorways account for over 50% of total VKT. Including 'U' roads, the percentage increases to 75% for England and Wales, and to 68% for London indicating that for most modelled street segments, emissions are accurately estimated. Moreover, taking into account classification accuracy for 'A' roads as well as the fact that there is a large sample to train the algorithm for this road class and that AADT on Motorways is directly counted – and not estimated with a model – it can also be concluded that the results for these two road types are more reliable as opposed to other road classes.

Second, the deviations observed in the projected emissions can also be attributed to the linear assumptions made – and applied – to the model, which are related to vehicle fleet composition. This assumption implies that equal use of all vehicles is considered, while – for instance – petrol cars could travel more compared to electric vehicles, thus producing more emissions. The later also suggests that a simulation model, could potentially capture this behaviour and provide more accurate outcomes, although this would have not been possible due to data requirements and high computation costs, particularly in the case of England and Wales. Moreover, linear assumptions have been made in the case of emission factors, indicating that even improved fuel technologies would still emit pollutants at the same level. This can specifically be argued in the scenario emission estimations for England and Wales, where fuel types (e.g., petrol, diesel, electric) rather than Euro standards have been used as opposed to the estimations for Greater London, due to lack of data. Consequently, the same emission factors are used for all vehicles of the same fuel type, rather than improved – and to some extent lower – emission factors that are normally applied for Euro standard vehicle technologies, explaining the fact that the presented model usually results into higher emissions compared to DfT outputs.

Finally, in the case of London, roads have been spatially clipped so as not to extend further than the Greater London boroughs and associated boundaries. That is, initial street segment lengths could be longer affecting estimation of VKT and related emissions. This can significantly affect estimations for Motorways and attached Outer London boroughs, where one can observe that even though these road types usually carry heavy traffic, VKT accounts for a smaller fraction of total distance travelled in London (Table 5-7), due to the short length of motorways within the study area. Finally, occurrence of satellite cities[66] and roads attached to but outside Greater London can significantly affect emissions on the edges of the study area, an effect that cannot be captured when the focus is placed on this case study.

5.6. Chapter summary

This chapter has focused on the estimation of tailpipe emissions from road transport vehicles using classification and regression modelling, facilitated by, and addressing the limitations of the transport volume modelling presented in chapter 3. Emissions have been estimated at a disaggregated level, for

[66] Satellite cities are smaller settlements around larger cities, separated from the metropolitan core by belts of rural territories (Bontje, 2019; Sorensen, 2001).

each road class and vehicle type, for the base year, while projections have been made for two case studies, considering technological development of vehicles, projected traffic for all regions in the study area as well as current and future policies.

This chapter has fulfilled the second aim of the thesis – *to assess the quantity of CO_2 and three air pollutants (PM, NO_x, and CO) originating from road transport and identify potential emissions abatement through technological developments of road vehicles and policies development.* Specifically, the findings that have been validated for both the base year and projections (in the case of England and Wales) indicate that the disaggregated nature of the presented approach can provide detailed results at granular level, and consequently sound inferences related to road usage and associated emissions can be drawn at different geographic scales. The revealed patterns on the contribution of specific vehicles on emissions by road class can be particularly useful for urban and environmental planners and can facilitate in the design, implementation, and assessment of relevant policies.

With this chapter the analysis of this thesis is complete. Certainly, considering the amount of data derived from multiple sources, the employment of various models and the attempt to complete numerous tasks throughout the modelling process, this study, like any other of this extent, exhibits limitations and has the potential to be improved, enriched, and extended. The following chapter draws the conclusions of the thesis and also attempts to address these issues by discussing its limitations and potentials for future research.

6. Conclusions

6.1. Thesis summary

The introduction in chapter 1 and literature review in chapter 2 demonstrated the need to mitigate CO_2 and air pollutant emissions and highlighted the contribution of road transport in these figures. Moreover, the significant challenges and literature gaps in road transport modelling, the estimation of associated emissions and the identification of factors affecting traffic have been presented. Difficulties in the identification of factors and emission estimation occur due to the complexity of road transport systems that become challenging to capture in transport models. This issue is usually reflected in the frequent lack of data in transport models as well as the occurrence of multiple emission modelling approaches.

In chapter 3, a comprehensive – spatial – dataset of drivers of traffic flows has been created, attempting to capture the maximum number of traffic determinants, by extracting and combining information from different sources, and a hybrid (i.e., clustering-regression) method to estimate AADT has been presented. The presented model revealed groups (i.e., clusters), where traffic patterns have been uncovered, allowing for further analysis to be conducted. Moreover, the modelling outcomes have shown high level of accuracy based on usual evaluation metrics. The dataset and model outputs from this chapter formed the foundation of the analysis to meet the aims of the thesis.

Chapter 4 provides the response to the first aim of the thesis – to identify the degree of influence specific factors have on traffic volumes. Thorough analysis of the drivers of traffic volumes has been conducted with the application of the Lasso regression and variable selection method, facilitated by the utilization of the full dataset and the modelling outputs from chapter 3. The analysis that has been conducted by road class on five different vehicle types, revealed that the effects of specific factors on traffic vary depending on vehicle type and road class, although some factors appear to significantly affect traffic volumes in all roads and vehicle types, such as the carriageway type and the set of variables indicating distances to urban areas.

Finally, tailpipe emissions from road transport have been estimated in chapter 5 as a response to the second aim of the thesis – to assess the quantity of CO_2 and three air pollutants (PM, NO_x, and CO) originating from road transport and identify potential emissions abatement through technological developments of road vehicles and policies development. It should be clarified that the analysis is conducted by making use of two business-as-usual scenarios from DfT, therefore not necessarily representing the current – or future – policy plans and implementation. The analysis has been conducted by applying a hybrid classification-regression model, based on the dataset and outputs extracted from chapter 3. Focus has been placed on base year emission estimations for England and Wales and separately for Greater London, with the two different case studies further examined for two different purposes. Firstly, emissions have been projected for England and Wales to identify potential abatement, given trends in traffic growth and vehicles' technological development. Secondly, emission estimations have been conducted for Greater London to assess current policies and future policy implementations. The modelling approach has been evaluated against models used by DfT, while the results show that significant emission reductions can be achieved with the introduction of new vehicle technologies and the implementation of clean air/low emission policies.

6.2. Concluding discussion

Severe impacts on ecosystems, human health and the economy caused by the high levels of CO_2 and air pollutant emissions has led the UK to take measures towards mitigating the effects. The contribution of road transport in these emissions is of high significance, since it has been found to hold a significant portion of the total GHG emissions in the UK – approximately 20% (Office for National Statistics, 2019) – and contributing up to 80% of total transport pollutant emissions (Department for Transport, 2018a), thus forming a barrier for the UK to meet its targets.

To date several studies have attempted to estimate emissions from road transport facilitated by the application of transport models, that are essential tools to provide traffic information to emission models. Emission modelling is key to quantify emissions from road transport and assess the potential effects of introduced and future policies. However, transport and emission modelling approaches to date exhibit several limitations. It has been observed that traffic estimations are usually restricted within city boundaries, with the exclusion of minor roads also being evident in the majority of traffic estimation studies. As a result, traffic in roads of minor importance is estimated based on average regional flows, and emissions are usually estimated on regional aggregated levels, implying incomplete and questionable estimations, where insights about the impacts of road transport on local levels cannot be drawn. Moreover, it has been seen that different modelling approaches can provide significant deviations in emission estimations values for the same areas (Fontaras et al., 2014), where in some cases the models tend to overestimate or underestimate emissions (e.g., Valach et al., 2015; Vaughan et al., 2016). In addition, several vital characteristics that are thought to affect traffic volume are usually not taken into account in transport models, probably also affecting the low estimation accuracies achieved in traffic estimation studies. The latter, also affects the accuracy of emission estimation, while the exclusion of traffic determinants does not allow the creation of links between the occurrence of these characteristics and the emissions originating from road transport.

Considering the gaps above, this thesis aimed to contribute to the transport literature, by addressing the identified limitations. Specifically, the thesis aimed to identify the influence of several characteristics on traffic volume, as well as to assess the quantity of CO_2, PM, $NO_{x,}$ and CO originating from road transport and identify potential emissions abatement through policy developments given the latest data available. This is done by proposing and applying an alternative methodology implemented in three stages, based on the investigation of novel statistical methods where traffic volumes and the corresponding emissions have been estimated, and the relation of several characteristics on traffic volume have been identified. In this section the major steps undertaken to meet the aims of the thesis and the corresponding outcomes as discussed in sections 3.5, 4.5 and 5.5 are concluded.

6.2.1. Traffic volume modelling

From the AADT modelling results presented in section 3.4, it can be seen that the formation of clusters has significantly contributed to the high estimation accuracy. Preliminary analysis on cluster formation has revealed that there are several variables taking a distinct set of values in each group and road type. Moreover, the analysis has shown that the RMSEs produced by the models[67] are similar to the AADT standard deviation (SD) values within each cluster. This implies that the model replicates the dispersion of the dependent variable (Meyer, 2012) and consequently, it is safe to assume that lower errors would be difficult to achieve.

Moreover, the analysis of the transport modelling results (sections 3.4 and 3.5) shows that 'A' roads in cluster 1 have strong similarities with motorways – which have not been modelled – in terms of traffic volumes (Table C-3). With regards to this point, road classes can be confused specifically when data from various sources such as DfT and OS are combined. For example, major roads based on the one source may be classified as minor according to the other, introducing complexity and uncertainty in the modelling process. Consequently, roads should – or could – be classified based on the traffic and not ownership[68] as it has also been pointed out by Xia et al. (1999) who faced similar issues.

6.2.2. Drivers of AADT

In chapter 4, a comprehensive set of variables created at an earlier stage has been examined, to address one of the two aims of the thesis – to understand the impact these factors have on traffic

[67] The RMSEs discussed here mainly refer to the two ML algorithms (i.e., SVR and RF). RMSEs produced by OLS do not apply to this statement, since they have been found to be higher in most cases, as shown in Table 3-3.
[68] Ownership refers to the corresponding authority that is responsible for road improvement and maintenance and assigning road classifications. National agencies such as Highways England and local highway authorities normally comply with government guidance on road classification and therefore, discrepancies in the classification between these two sources should be minimal, if any. However, other sources, such as the OS used in this thesis may exhibit different classifications for the same roads, a matter that is normally identified due to data update issues.

volumes. The analysis has been undertaken for five different vehicle types and four different road classes in England and Wales, where traffic counters have been subdivided into five groups based on specific land use, socioeconomic, public transport and roadway characteristics in the vicinity of each counter as modelled in chapter 3. The results produced by Lasso reveal patterns for specific explanatory variables across vehicle types and road classes. In some cases, heterogeneous results across estimated models have been reconciled by looking at the characteristics of the counters and areas in each model. In this section, the outcomes are summarised as presented and discussed in sections 4.4 and 4.5 and potential conclusions and ways to improve the understanding on traffic volumes are discussed.

Overall, it can be concluded that variables related to roadway characteristics, such as the road nature and road category are almost always present indicating strong predictors of traffic volumes. Similarly, specific socioeconomic variables are associated with traffic volumes in numerous ways. Income is associated with the increased use of private vehicles which indicates decreased use of buses. Population and workplace population densities – associated with the distinction between urban and rural environments – also relates with the traffic volumes of all vehicle types in a different manner. An interesting finding is that population density is positively correlated with high bus volumes, although these patterns are more significant for higher class roads (e.g., ‘A’ or ‘B’). Buses normally drive through major arteries where bus lanes are more likely to occur to facilitate traffic flow along important routes, normally linking important places. The distinction between urban and rural environments and the respective relation with traffic volumes is also clear from other associated characteristics. Variables such as public transport presence and accessibility and the set of distances to urban/major urban areas are associated with all vehicle types although cars, buses and two-wheeled vehicles are more associated with factors related to densely populated urban centres, while LGVs and HGVs are more likely to be present at industrialised areas close to urban centres and their outskirts.

However, specific variables related to specific vehicle types only can also be distinguished. For example, it has been observed that warehouses and factories are always positively correlated with LGVs and HGVs, registered vehicles mainly associate with bus and two-wheeled vehicle volumes, while train accessibility is related with car volumes. Similar conclusions can be drawn when investigating the association of specific variables with traffic volumes on different road types. In particular, it has been noticed that some variables are associated with traffic volumes on ‘A’ roads only, due to the different nature of these roads and use by the vehicles. For example, ring roads correlate with traffic volumes only on ‘A’ roads, although the coefficient signs vary significantly when investigating different vehicle types.

The latter also leads to the consideration of the number of data points within each road class and subgroup. For ‘A’ roads the 14670 points represent 99% of the total road links for this road class in the study area, while the proportions for the other road classes are 13.1% for ‘B’ roads, 0.7% for ‘C’ roads and 0.15% for ‘U’ roads respectively. Moreover, counts for ‘C’ and ‘U’ roads are undertaken manually, and the number of counted vehicles is low and therefore traffic volume modelling and resulting coefficients is challenging and potentially less reliable. This can also be confirmed by the coefficients extracted in the case of ‘U’ roads where no specific patterns have been revealed. Furthermore, ‘U’ roads are often located in entrances to industrial areas or within private properties (e.g., warehouse courtyards) and can introduce bias to the models. Hence, it can be safely assumed that results for ‘A’ roads are more reliable.

Finally, one has to consider the spatial dataset used to distinguish between urban and rural areas and calculate the respective distances. The dataset considers all build up areas as urban (Bibby & Brindley, 2014), indicating that large cities are classified together with villages and small towns, where significant differences in land use, population, public transport and other indicators occur. Even in the case of the six major urban areas there are considerable differences among them. Specifically, larger cities such as London, Manchester or Birmingham, where industrial areas can be surrounded by residential neighbourhoods can add complexity to the analysis.

6.2.3. Emissions

Spatial distribution of AADT, VKT and associated emissions is of high importance for research in urban, transport as well as health and environmental planning. The methodology presented in chapter 5 can provide a significant improvement in estimating emissions since all street segments of the study area are modelled, therefore delivering a better understanding of the spatial distribution of pollution levels in these areas. Considering that the emission modelling results are comparable with the models used by DfT, both at city (i.e., London) and national (i.e., England and Wales) levels as well as on base and projected years, indicates that this approach can provide valid but also more detailed and granular results, since it can be used for both micro (e.g., street level) and macro (e.g., countries or states) analysis as opposed to aggregated models. However, it is important to highlight that this method estimates and presents the spatial distribution of average daily emissions and not concentration of pollutants. Therefore, it can be argued that the modelling presented introduces an alternative method, which due to its ability to model emissions in multiple levels of granularity can be useful for policy makers and planners.

Comparing and validating the presented approach with other – established – models implies that these approaches are trustworthy and can provide accurate estimations. This is contradicted by the fact that, as it has been discussed, estimations among different approaches can vary significantly and as a result the reliability of every model can be questionable. If accurate validation can only be conducted when ground truth values are known, a complete monitoring of emissions across the full study area, which is in practice infeasible, would be required. Hence, uncertainty in the projections of any model should be expected, normally increasing with the length of the projection and the disaggregation of the model (Department for Transport, 2018d).

Nonetheless, the modelling and its outputs presented in this thesis provide considerable insights on the current and potential future contribution of road transport on emissions, although it should be noted that the projected results for England and Wales cannot be utilized to assess the UK's path towards meeting its emission mitigation targets. Firstly, the targets set by the government apply to the whole of the UK, while the thesis has only focused on England and Wales. Secondly, the targets to reduce emissions are overall set for all contributing sectors such as manufacturing and construction, electricity generation and agriculture (Climate Change Committee, 2020a), while the thesis has focused on road transport emissions. In addition to the latter, the CCC's latest report (Climate Change Committee, 2020b) that focuses on transport emissions explore road transport as a section of surface transport (also including rail and active transport). In the report, several options to reduce emissions are generally discussed, although clear targets for this sector are not indicated, and therefore there can be no modelling to capture these effects. For example, a ban on new petrol and diesel cars sales by 2030-2035 is considered as an option, and a shift to alternative lower-carbon transport modes, such as walking, cycling or public transport, is assumed to occur by 2035 (Climate Change Committee, 2020b), although further details and data are not provided.

The presented model is a useful tool to assess emissions and potential gains from road transport given the current knowledge and available data. Moreover, due to its varied nature (i.e., both aggregated and disaggregated), policy implementation processes can be facilitated both at national as well as at more targeted local levels. For example, although the overall UK's mitigation targets cannot be assessed by the model's output, it can be concluded that CO_2 emissions from road transport have the potential to be reduced by 21% considering scenario 1 and by 18% considering scenario 2 in 2035 compared to 1990 levels in the study area, as shown in Table 6-1. Considering that England and Wales comprise approximately 88% of the UK's road transport CO_2 emissions (Department for Transport, 2020), respective inferences can be drawn and related policies can be designed and implemented.

Table 6-1: CO_2 (million tonnes) emission change in England and Wales from 1990 to 2035

Year	Scenario 1		Scenario 2	
	Author's method	**DfT**	**Author's method**	**DfT**
1990	107.13		107.13	
2035	84.68	77.15	87.58	81.08
Change (%)	-20.96%	-27.98%	-18.25%	-24.31%

On the other hand, in the case of localized policies such as ULEZ and its extension, several reports contain contradicting and frequently abstract statements. For example, it has been argued that the extension of the zone would have benefits both in terms of air pollution reductions and the economy (Laybourn-Langton and Quilter-Pinner, 2016), while on the other hand it has been argued that it would only result into minimal reductions in harmful emissions (Jacobs, 2014). Certainly, none of the above can yet be ratified. From a research perspective, one can only expect for the policies to be enforced and data to become available. However, as opposed to the England and Wales study, ULEZ and ULEZ extension are easier and straightforward to assess, considering the assumptions made. Introduction of new technology vehicles with lower emission factors across a specific spatial extent (e.g., ULEZ) and assuming equal use of these vehicles as implemented in the model, is anticipated to reduce emissions and consequently, the spatial extension of the area is also expected to have additional effects.

This thesis has highlighted both the complexity and importance of road transport, as well as its interdependence with numerous sectors, such as urban planning and the economy. Considering the international increase in many measures of people and goods movement (National Academy of Sciences, 2010), it is vital to retain, expand and modernize road transport capacity through technological developments and policy implementation. This will allow not only to meet the constantly increasing demand for transportation and achieve social and economic growth, but also to counterbalance the associated damaging environmental effects at national and international levels. However, technological development and its consideration and integration within relevant policies, indicates that extended research and significant investment on infrastructure – such as electric vehicles' charging points – has to occur, which is not the focus of this thesis. The thesis has presented a method to produce disaggregated estimates of traffic volumes and associated emissions, and provided insights on how these estimates could be used to evaluate the impacts of related policies. Thus, the method presented can be used to inform further decisions.

6.3. Limitations and uncertainties

The modelling processes that have been followed to meet the aims of the thesis have resulted into sensible and comparable results with similar studies. Still, as every other process, the methodology exhibits some limitations and uncertainties that need to be addressed.

Firstly, combining the datasets to fit the purposes of the thesis can introduce uncertainty in the model and analysis. It has been seen that during the transport and emission modelling process, datasets from various sources have been used. Specifically, in the case of road transport modelling, data have been extracted from nine different sources to create over forty variables. The different datasets can be created using different methods, some datasets may have important missing information, while some others are collected during different time periods. For example, there are several ports and airports of different size, function and importance in England and Wales, and one cannot be certain if all the facilities were included in the ports and airports datasets. Another important and common example is related to the collection of socioeconomic data in relation to traffic. While AADT base year data correspond to 2015, the socioeconomic data related to population and corresponding densities extracted from ONS is based on the 2011 census. Hence, one would expect the socioeconomic characteristics to have changed – probably slightly – within the course of 4 years, something that it would be expected to magnify when doing the projections. Similarly, one would expect the land use information from VOA to change every year, specifically considering the type and number of facilities included in this dataset.

Secondly, the transport model presented in chapter 3 focused on the estimation of total traffic volumes without taking into consideration the different vehicle types, although the analysis of traffic determinants (chapter 4) and emissions estimation (chapter 5) has been conducted on a vehicle-type level. This can result into some uncertainty in the computation of emissions in chapter 5 and can partially also explain the deviations reported in Table 5-18, Table 5-19 and Table 5-20. This also implies that a recalibration of the model for each vehicle type would allow to better inform the analysis of traffic determinants conducted in chapter 4 – the latter also explaining the fact that some – although very few – of the coefficients were not precisely clarified.

Thirdly, it should be mentioned that the thesis estimated only tailpipe (i.e., exhaust) emissions from vehicles and did not consider emissions produced by other parts of the vehicles, specific activities or other factors that would affect the produced emissions. For example, tyre wear produced by friction, drivers' behaviour and weather conditions all affect the amount of emissions produced by the vehicles. However, emissions from tyres are minimal and cannot be modelled by the methodology presented in this thesis. In addition, behavioural issues can only be modelled with simulation models – found to be unsuitable for the purposes of this research. Moreover, weather conditions that can also affect behaviour and overall emission production incorporates temporal trends, that on the one hand are again out of the scope of my thesis, but on the other it can be assumed that the AADT values can capture temporal traffic deviations at an aggregated level.

Finally, it is worth commenting on the estimated emission projections. In chapter 5, data gaps extracted from official published policies and governmental gateways were reported, that in turn have been reflected in the model outputs. For example, the latest official reports refer to options to reduce emissions, such as active travel (e.g., walking and cycling), although there are no available data on the amount of walking or cycling that is projected to replace motorised transportation. This led to the linear assumptions related to the use of each vehicle type that had to take place as discussed in section 5.5, which have probably affected the projected emission estimations. For instance, the projected distance travelled (i.e., VKT) has been estimated based only on the expected traffic growth (Table 5-2), without considering any potential shift from motorised transport to active travel. This is also related to uncertainties that can occur when combining different datasets as discussed above, since the projected traffic values are extracted from DfT, while the latest report on mitigating transport emissions comes from CCC, also published in different dates. Moreover, the assumptions made are likely to have affected the use of electric vehicles the most. Taking into account that substantial infrastructure needs to be integrated for the use of these vehicles and that behavioural issues – such as people's interest to purchase electric vehicles – have to be considered, it is safe to assume that the projected number of electric vehicles that have been integrated into the model may be unreliable.

6.4. Future research

Based on the outcomes of the thesis further research may be conducted to address the limitations discussed in section 6.3 as well as to enrich and improve the findings and the methodology presented.

For example, the collection of additional data and incorporation of explanatory variables has the potential to improve performance of ML algorithms (Domingos, 2012; Junqué de Fortuny et al., 2013) and thus can be explored in future studies if computational processing allows it.

Moreover, in the case of transport modelling, several different approaches and methods can be applied to investigate potential improvement of the model outputs. For example, implementation of other clustering and validation techniques – e.g., automated weighting clustering algorithms proposed in other studies (e.g. Chen and Wang, 2013; Huang et al., 2005) and k-fold cross validation (Koul et al., 2018) – could reveal different patterns of traffic flows and corresponding driving factors as well as potentially provide more accurate error measurement.

Furthermore, in the case of identifying the impact of several factors on traffic volumes, one has to consider the spatial dataset used to distinguish between urban and rural areas and calculate the respective distances. The dataset considers all build up areas as urban (Bibby & Brindley, 2014), indicating that large cities are classified together with villages and small towns, something which would be helpful to differentiate in future studies. Even in the case of the six major urban areas there are considerable differences among them. Specifically, cities such as London or Manchester, where industrial areas can be surrounded by residential neighbourhoods, can add complexity to the analysis and should be examined individually. Hence, it can be identified that further sampling and acquisition of more reliable data can potentially lead to an improved model and understanding of traffic volumes. Moreover, similar regularisation methods such as ridge regression (Hoerl and Kennard, 1970) and elastic net (Zou and Hastie, 2005) can be applied to compare – and potentially combine – the results and conclude with more meaningful outcomes. In addition, the application of interpretable machine learning can be explored, where the 'black-box' nature of the algorithms can be unfolded so as to the behaviour and output of the algorithms can be explained (Doshi-Velez and Kim, 2017).

In terms of emission modelling, again methodological aspects can be explored. Firstly, although three classification algorithms have been tested, other probabilistic classification models can be trialled to explore the potential to achieve higher classification accuracy and emission estimation if possible. The model can also be applied to other areas where clean air zones have been introduced (e.g., Birmingham) to assess the impact of these policies, while it would also be interesting to assess the findings of this thesis for ULEZ and ULEZ extension when official data will become available.

Finally, considering data availability and computational capacity a spatio-temporal modelling approach for a detailed and comprehensive assessment of AADT and changes in AADT across space and time should be considered in future studies.

References

Abadi, A., Rajabioun, T., Ioannou, P.A., 2015. Traffic Flow Prediction for Road Transportation Networks With Limited Traffic Data. IEEE Trans. Intell. Transp. Syst. 16, 653–662. https://doi.org/10.1109/TITS.2014.2337238

Abdallah, C., Afif, C., Sauvage, S., Borbon, A., Salameh, T., Kfoury, A., Leonardis, T., Karam, C., Formenti, P., Doussin, J.F., Locoge, N., Sartelet, K., 2020. Determination of gaseous and particulate emission factors from road transport in a Middle Eastern capital. Transp. Res. Part D Transp. Environ. 83, 102361. https://doi.org/10.1016/j.trd.2020.102361

Abdelwahab, H.T., 2017. Traffic micro-simulation model for design and operational analysis of barrier toll stations. Ain Shams Eng. J. 8, 507–513. https://doi.org/10.1016/j.asej.2016.05.010

Aditjandra, P.T., Cao, X., Mulley, C., 2012. Understanding neighbourhood design impact on travel behaviour: An application of structural equations model to a British metropolitan data. Transp. Res. Part A Policy Pract. 46, 22–32. https://doi.org/10.1016/j.tra.2011.09.001

Aditjandra, P.T., Galatioto, F., Bell, M.C., Zunder, T.H., 2016. Evaluating the impacts of urban freight traffic: Application of micro-simulation at a large establishment. Eur. J. Transp. Infrastruct. Res. 16, 4–22. https://doi.org/10.18757/ejtir.2016.16.1.3110

Adler, P., Falk, C., Friedler, S.A., Nix, T., Rybeck, G., Scheidegger, C., Smith, B., Venkatasubramanian, S., 2018. Auditing black-box models for indirect influence. Knowl. Inf. Syst. 54, 95–122. https://doi.org/10.1007/s10115-017-1116-3

Ajtay, D., Weilenmann, M., 2004. Compensation of the exhaust gas transport dynamics for accurate instantaneous emission measurements. Environ. Sci. Technol. 38, 5141–5148. https://doi.org/10.1021/es0353513

Akaike, H., 1974. A New Look at the Statistical Model Identification. IEEE Trans. Automat. Contr. 19, 716–723. https://doi.org/10.1109/TAC.1974.1100705

Akhanli, S.E., Hennig, C., 2017. Some Issues in Distance Construction for Football Players Performance Data. Arch. Data Sci. 2, 1–17. https://doi.org/10.5445/KSP/1000058749/09

Alotaibi, R., Bechle, M., Marshall, J.D., Ramani, T., Zietsman, J., Nieuwenhuijsen, M.J., Khreis, H., 2019. Traffic related air pollution and the burden of childhood asthma in the contiguous United States in 2000 and 2010. Environ. Int. 127, 858–867. https://doi.org/10.1016/j.envint.2019.03.041

Anda, C., Erath, A., Fourie, P.J., 2017. Transport modelling in the age of big data. Int. J. Urban Sci. 21, 19–42. https://doi.org/10.1080/12265934.2017.1281150

Anderson, M.D., Sharfi, K., Gholston, S.E., 2006. Direct Demand Forecasting Model for Small Urban Communities Using Multiple Linear Regression. Transp. Res. Rec. 1981, 114–117.

André, M., Joumard, R., Vidon, R., Tassel, P., Perret, P., 2006. Real-world European driving cycles, for measuring pollutant emissions from high- and low-powered cars. Atmos. Environ. 40, 5944–5953. https://doi.org/10.1016/j.atmosenv.2005.12.057

Andre, M., Keller, M., Sjodin, A., Gadrat, M., Mc Rae, I., 2008. The Artemis European Tools for Estimating the Pollutant Emissions From Road Transport and Their Application in Sweden and France 118–127.

Anselin, L., Syabri, I., Kho, Y., 2010. GeoDa: An Introduction to Spatial Data Analysis, in: Handbook of Applied Spatial Analysis. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 73–89.

https://doi.org/10.1007/978-3-642-03647-7_5

Aoki, Y., 2017. Evaluation of in vivo mutagenesis for assessing the health risk of air pollutants. Genes Environ. 39, 1–12. https://doi.org/10.1186/s41021-016-0064-6

Apronti, D., Ksaibati, K., Gerow, K., Hepner, J.J., 2016. Estimating traffic volume on Wyoming low volume roads using linear and logistic regression methods. J. Traffic Transp. Eng. (English Ed. 3, 493–506. https://doi.org/10.1016/j.jtte.2016.02.004

ApSimon, H.M., Warren, R.F., Wilson, J.J.N., 1994. The abatement strategies assessment model-ASAM: Applications to reductions of sulphur dioxide emissions across Europe. Atmos. Environ. 28, 649–663. https://doi.org/10.1016/1352-2310(94)90042-6

Arentze, T.A., Timmermans, H.J.P., 2004. A learning-based transportation oriented simulation system. Transp. Res. Part B Methodol. 38, 613–633. https://doi.org/10.1016/j.trb.2002.10.001

Arnott, R., Inci, E., 2006. An integrated model of downtown parking and traffic congestion. J. Urban Econ. 60, 418–442. https://doi.org/10.1016/j.jue.2006.04.004

Arnott, R., Williams, P., 2017. Cruising for parking around a circle. Transp. Res. Part B Methodol. 104, 357–375. https://doi.org/10.1016/j.trb.2017.07.009

Artelt, S., Kock, H., König, H.P., Levsen, K., Rosner, G., 1999. Engine dynamometer experiments: platinum emissions from differently aged three-way catalytic converters. Atmos. Environ. 33, 3559–3567. https://doi.org/10.1016/S1352-2310(99)00109-0

Aslam, N., Cheng, T., 2018. Smart Card Data and Human Mobility, in: Longley, P., Cheshire, J., Singleton, A. (Eds.), Consumer Data Research. UCLPress. https://doi.org/10.14324/111.9781787353886.CITATIONS

Ayaz, S., Khattak, K., Khan, Z., Minallah, N., Khan, M., Khan, A., 2022. Sensing technologies for traffic flow characterization: From heterogeneous traffic perspective. J. Appl. Eng. Sci. 20, 29–40. https://doi.org/10.5937/jaes0-32627

Bacher, J., Wenzig, K., Volger, M., 2004. SPSS TwoStep Cluster - a first evaluation, Arbeits- und Diskussionspapiere / Universität Erlangen-Nürnberg, Sozialwissenschaftliches Institut, Lehrstuhl für Soziologie. Amsterdam. Available at: https://nbn-resolving.org/urn:nbn:de:0168-ssoar-327153 (Accessed: 3 May 2019)

Balac, M., Axhausen, K.W., 2016. Activity Rescheduling within Multi Agent Transport Simulation Framework (MATSim), in: TRB 96th Annual Meeting Compendium of Papers. Transportation Research Record. https://doi.org/https://doi.org/10.3929/ethz-b-000118787

Bao, Q., Kochan, B., Shen, Y., Creemers, L., Bellemans, T., Janssens, D., Wets, G., 2018. Applying FEATHERS for travel demand analysis: Model considerations. Appl. Sci. 8. https://doi.org/10.3390/app8020211

Barlow, T., Latham, S., Mccrae, I., Boulter, P., 2009. A reference book of driving cycles for use in the measurement of road vehicle emissions, TRL Published Project Report. TRL.

Basak, D., Pal, S., Patranabis, D.C., 2007. Support Vector Regression. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 10634 LNCS, 699–708. https://doi.org/10.1007/978-3-319-70087-8_72

Baškovic, K., Knez, M., 2013. A Review of Vehicular Emission Models. Science (80-.). 216, 3000.

Batty, M., Besussi, E., Chin, N., 2003. Traffic, Urban Growth and Suburban Sprawl, Centre for Advanced Spatial Analysis. https://doi.org/10.1103/PhysRevE.78.016110. Available at:

https://discovery.ucl.ac.uk/id/eprint/216/ (Accessed: 31 July 2017)

Baumgartner, R., Ryner, L., Richter, W., Summers, R., Jarmasz, M., Somorjai, R., 2000. Comparison of two exploratory data analysis methods for fMRI: Fuzzy clustering vs. principal Component analysis. Magn. Reson. Imaging 18, 89–94. https://doi.org/10.1016/S0730-725X(99)00102-2

Becker, F., Axhausen, K.W., 2017. Literature review on surveys investigating the acceptance of automated vehicles. Transportation (Amst). 44, 1293–1306. https://doi.org/10.1007/s11116-017-9808-9

Beckx, C., Panis, L.I., Vankerkom, J., Janssens, D., Wets, G., Arentze, T., 2009. An integrated activity-based modelling framework to assess vehicle emissions: Approach and application. Environ. Plan. B Plan. Des. 36, 1086–1102. https://doi.org/10.1068/b35044

Bellucci, P., Cipriani, E., 2010. Data accuracy on automatic traffic counting: the SMART project results. Eur. Transp. Res. Rev. 2, 175–187. https://doi.org/10.1007/s12544-010-0039-9

Ben-Akiva, M., 1973. Structure of Travel Demand Models. PhD Thesis, Massachusetts Institute of Technology, Boston, Accessed: 25 May 2018, http://hdl.handle.net/1721.1/14790

Ben-Akiva, M., Bierlaire, M., 1999. Discrete Choice Methods & Their Applications to Short Term Travel Decisions, in: Handbook of Transportation Science. Springer, Boston, pp. 5–33.

Ben-Akiva, M., Bowman, J.L., 1998. Integration of an activity-based model system and a residential location model. Urban Stud. 35, 1131–1153.

Bento, A., Cropper, M.L., Mobarak, A.M., Vinha, K., 2005. The Effects of Urban Spatial Structure on Travel Demand in the United States. Rev. Econ. Stat. 87, 466–478. https://doi.org/10.1162/0034653054638292

Bergstra, A., Brunekreef, B., Burdorf, A., 2018. The mediating role of risk perception in the association between industry-related air pollution and health. PLoS One 76, 601–603. https://doi.org/10.1378/chest.76.5.601

Bernard, Y., Miller, J., Wappelhorst, S., Braun, C., 2020. Impacts of the Paris low-emission zone and implications for other cities, International Council on Clean Transportation. Washington DC: TRUE—The Real Urban Emissions Initiative

Bharadwaj, N., Kumar, P., Arkatkar, S., Maurya, A., Joshi, G., 2016. Traffic data analysis using image processing technique on Delhi-Gurgaon expressway. Curr. Sci. 110, 808–822. https://doi.org/10.18520/cs/v110/i5/808-822

Bhat, C., Guo, J., Srinivasan, S., Sivakumar, A., 2004. Comprehensive Econometric Microsimulator for Daily Activity-Travel Patterns. Transp. Res. Rec. J. Transp. Res. Board 1894, 57–66. https://doi.org/10.3141/1894-07

Bholowalia, P., Kumar, A., 2014. EBK-Means : A Clustering Technique based on Elbow Method and K-Means in WSN. Int. J. Comput. Appl. 105, 17–24.

Bibby, P., Brindley, P., 2014. 2011 Rural-Urban Classification of Local Authority Districts in England: User Guide. London UK: Department for Environment, Food & Rural Affairs

Bickel, P.J., Chen, C., Kwon, J., Rice, J., van Zwet, E., Varaiya, P., 2007. Measuring Traffic. Stat. Sci. 22, 581–597. https://doi.org/10.1214/07-STS238

Bishop, C.M., 2013. Pattern Recognition and Machine Learning, Springer. Cambridge, UK. https://doi.org/10.1117/1.2819119

Bolbol, A., Cheng, T., Tsapakis, I., Haworth, J., 2012. Inferring hybrid transportation modes from sparse GPS data using a moving window SVM classification. Comput. Environ. Urban Syst. 36, 526–537. https://doi.org/10.1016/j.compenvurbsys.2012.06.001

Bolton, R.J., Krzanowski, W.J., 2003. Projection pursuit clustering for exploratory data analysis. J. Comput. Graph. Stat. 12, 121–142. https://doi.org/10.1198/1061860031374

Bontje, M., 2019. Shenzhen: satellite city or city of satellites? Int. Plan. Stud. 24, 255–271. https://doi.org/10.1080/13563475.2019.1657383

Borge, R., de Miguel, I., de la Paz, D., Lumbreras, J., Pérez, J., Rodríguez, E., 2012. Comparison of road traffic emission models in Madrid (Spain). Atmos. Environ. 62, 461–471. https://doi.org/10.1016/j.atmosenv.2012.08.073

Boughedaoui, M., Joumard, R., Kerbachi, R., 2008. On-Board Emission Measurement of High-Loaded Light-Duty Vehicles in Algeria. J. Air Waste Manage. Assoc. 58, 45–54. https://doi.org/10.3155/1047-3289.58.1.45

Boulter, P., McCrae, I., Barlow, T., 2007. A review of instantaneous emission models for road vehicles. Wokingham, UK: Transportation Research Laboratory

Boulter, P.G., McCrae, I.S., 2007. ARTEMIS : Assessment and Reliability of Transport Emission Models and Inventory Systems – final report, 2007. Available at: http://worldcat.org/issn/09684093 (Accessed: 27 June 2018)

Brand, C., 2010. Transport Carbon Model Reference Guide. Oxford. Available at: https://ukerc.ac.uk/publications/uk-transport-carbon-model-reference-guide (Accessed: 25 February 2019)

Brathwaite, T., Vij, A., Walker, J.L., 2017. Machine Learning Meets Microeconomics: The Case of Decision Trees and Discrete Choice. arXiv Prepr. arXiv:1711. Available at: http://arxiv.org/abs/1711.04826 (Accessed: 7 June 2018)

Breiman, L., 2001. Random Forests. Mach. Learn. 45, 5–32. https://doi.org/10.1017/CBO9781107415324.004

Breiman, L., 1996. Bagging predictors. Mach. Learn. 24, 123–140. https://doi.org/10.1007/BF00058655

Brimblecombe, P., Townsend, T., Lau, C.F., Rakowska, A., Chan, T.L., Močnik, G., Ning, Z., 2015. Through-tunnel estimates of vehicle fleet emission factors. Atmos. Environ. 123, 180–189. https://doi.org/10.1016/j.atmosenv.2015.10.086

Brown, I., Mues, C., 2012. An experimental comparison of classification algorithms for imbalanced credit scoring data sets. Expert Syst. Appl. 39, 3446–3453. https://doi.org/10.1016/j.eswa.2011.09.033

Brown, P., Wakeling, D., Pang, Y., Murrells, T., 2018. Methodology for the UK's Road Transport Emissions Inventory Version for the 2016 National Atmospheric Emissions Inventory. London, UK: The Department for Business. Energy & Industrial Strategy.

Browne, M., Allen, J., Nemoto, T., Visser, J., 2010. Light goods vehicles in urban areas. Procedia - Soc. Behav. Sci. 2, 5911–5919. https://doi.org/10.1016/j.sbspro.2010.04.006

Brustlin, V.H., Cambridge Systematics, Gallop Corporation, Bhat, C.R., Shapiro Transportation Consultin, 2012. Travel Demand Forecasting: Parameters and Techniques. The National Academies Press, Washington, DC. https://doi.org/10.17226/14665

Bunker, B.J., Franchek, M.A., Thomason, B.E., 1997. Robust multivariable control of an engine-dynamometer system. IEEE Trans. Control Syst. Technol. 5, 189–199. https://doi.org/10.1109/87.556024

Burgard, D.A., Bishop, G.A., Stadtmuller, R.S., Dalton, T.R., Stedman, D.H., 2006. Spectroscopy Applied to On-Road Mobile Source Emissions. Appl. Spectrosc. 60, 135A-148A. https://doi.org/10.1366/000370206777412185

Burns, C., Thomason, J., Tansey, W., 2020. Interpreting Black Box Models via Hypothesis Testing, in: Proceedings of the 2020 ACM-IMS on Foundations of Data Science Conference. pp. 47–57. https://doi.org/https://doi.org/10.1145/3412815.3416889

Butcher, L., 2017. Speed limits in England: Briefing Paper - Number CBP00468. London, UK: House of Commons Library.

Caceres, N., Romero, L.M., Benitez, F.G., 2012. Estimating Traffic Flow Profiles According to a Relative Attractiveness Factor. Procedia - Soc. Behav. Sci. 54, 1115–1124. https://doi.org/10.1016/j.sbspro.2012.09.826

Caceres, N., Romero, L.M., Morales, F.J., Reyes, A., Benitez, F.G., 2018. Estimating traffic volumes on intercity road locations using roadway attributes , socioeconomic features and other work-related activity characteristics. Transportation (Amst). 45, 1449–1473. https://doi.org/10.1007/s11116-017-9771-5

Calinski, T., Harabasz, J., 1974. A dendrite method for cluster analysis. Commun. Stat. - Theory Methods 3, 1–27. https://doi.org/10.1080/03610927408827101

Cardozo, O.D., García-Palomares, J.C., Gutiérrez, J., 2012. Application of geographically weighted regression to the direct forecasting of transit ridership at station-level. Appl. Geogr. 34, 548–558. https://doi.org/10.1016/j.apgeog.2012.01.005

Carli, R., Dotoli, M., Epicoco, N., Angelico, B., Vinciullo, A., 2015. Automated evaluation of urban traffic congestion using bus as a probe. IEEE Int. Conf. Autom. Sci. Eng. 2015-Octob, 967–972. https://doi.org/10.1109/CoASE.2015.7294224

Carslaw, D.C., Rhys-Tyler, G., 2013. New insights from comprehensive on-road measurements of NOx, NO2 and NH3 from vehicle emission remote sensing in London, UK. Atmos. Environ. 81, 339–347. https://doi.org/10.1016/j.atmosenv.2013.09.026

Caselli, M., de Gennaro, G., Marzocca, A., Trizio, L., Tutino, M., 2010. Assessment of the impact of the vehicular traffic on BTEX concentration in ring roads in urban areas of Bari (Italy). Chemosphere 81, 306–311. https://doi.org/10.1016/j.chemosphere.2010.07.033

Castiglione, J., Bradley, M., Gliebe, J., 2014. Activity-Based Travel Demand Models: A Primer. Transportation Research Board, Washington, DC. https://doi.org/10.17226/22357

Castro-Neto, M., Jeong, Y., Jeong, M.K., Han, L.D., 2009. AADT prediction using support vector regression with data-dependent parameters. Expert Syst. Appl. 36, 2979–2986. https://doi.org/10.1016/j.eswa.2008.01.073

Cervero, R., 2006. Alternative approaches to modeling the travel-demand impacts of smart growth. J. Am. Plan. Assoc. 72, 285–295. https://doi.org/10.1080/01944360608976751

Cervero, R., 1994. Transit-based housing in California: evidence on ridership impacts. Transp. Policy 1, 174–183. https://doi.org/10.1016/0967-070X(94)90013-2

Cervero, R., Kockelman, K., 1997. Travel demand and the 3Ds: Density, diversity, and design. Transp. Res. Part D Transp. Environ. 2, 199–219. https://doi.org/10.1016/S1361-9209(97)00009-6

Chaabouni, S., Saidi, K., 2017. The dynamic links between carbon dioxide (CO_2) emissions, health spending and GDP growth: A case study for 51 countries. Environ. Res. 158, 137–144. https://doi.org/10.1016/j.envres.2017.05.041

Chan, T.L., Ning, Z., Leung, C.W., Cheung, C.S., Hung, W.T., Dong, G., 2004. On-road remote sensing of petrol vehicle emissions measurement and emission factors estimation in Hong Kong. Atmos. Environ. 38, 2055–2066. https://doi.org/10.1016/j.atmosenv.2004.01.031

Chapman, L., 2007. Transport and climate change: a review. J. Transp. Geogr. 15, 354–367. https://doi.org/10.1016/j.jtrangeo.2006.11.008

Chatterton, T., Barnes, J., Wilson, R.E., Anable, J., Cairns, S., 2015. Use of a novel dataset to explore spatial and social variations in car type, size, usage and emissions. Transp. Res. Part D Transp. Environ. 39, 151–164. https://doi.org/10.1016/j.trd.2015.06.003

Chauke, V., 2015. Variations in Traffic Count Data, in: 34th Southern African Transport Conference. Johannesburg, South Africa, pp. 788–800.

Chen, C., Huang, C., Jing, Q., Wang, H., Pan, H., Li, L., Zhao, J., Dai, Y., Huang, H., Schipper, L., 2007. On-road emission characteristics of heavy-duty diesel vehicles in Shanghai. Atmos. Environ. 41, 5334–5344. https://doi.org/10.1016/j.atmosenv.2007.02.037

Chen, L., Wang, S., 2013. Central clustering of categorical data with automated feature weighting. IJCAI Int. Jt. Conf. Artif. Intell. 1260–1266.

Chen, P., Hu, S., Shen, Q., Lin, H., Xie, C., 2019. Estimating Traffic Volume for Local Streets with Imbalanced Data. Transp. Res. Rec. 2673, 598–610. https://doi.org/10.1177/0361198119833347

Chen, X., Chen, C., Cai, Y., Wang, H., Ye, Q., 2018. Kernel Sparse Representation with Hybrid Regularization for On-Road Traffic Sensor Data Imputation. Sensors 18, 2884. https://doi.org/10.3390/s18092884

Cheng, Y., He, L., He, W., Zhao, P., Wang, P., Zhao, J., Zhang, K., Zhang, S., 2019. Evaluating on-board sensing-based nitrogen oxides (NOX) emissions from a heavy-duty diesel truck in China. Atmos. Environ. 216, 116908. https://doi.org/10.1016/j.atmosenv.2019.116908

Chhorn, V., Prabnasak, J., Chompoo-Inwai, C., 2018. An evaluation of public truck terminal implementation and truck ban policy in a medium-size city of Thailand: A case study of Khon Kaen city, in: Hamontree, C. (Ed.), MATEC Web of Conferences. p. 02060. https://doi.org/10.1051/matecconf/201819202060

Chiu, Y., Bottom, J., Mahut, M., Paz, A., Balakrishna, R., Waller, T., Hicks, J., 2011. Dynamic Traffic Assignment. Transp. Netw. Model. Comm. 1–39. https://doi.org/10.1016/j.trd.2016.06.003

Cho, S., Lee, W. Do, Hwang, J.H., Kochan, B., Knapen, L., Bellemans, T., Choi, K., Joh, C.H., 2015. Validation of activity-based travel demand model using smart-card data in Seoul, South Korea. Procedia Comput. Sci. 52, 707–712. https://doi.org/10.1016/j.procs.2015.05.080

Choi, H.-W., Frey, H.C., 2010. Estimating Diesel Vehicle Emission Factors at Constant and High Speeds for Short Road Segments. Transp. Res. Rec. J. Transp. Res. Board 2158, 19–27. https://doi.org/10.3141/2158-03

Choudhary, A., Gokhale, S., 2019. On-road measurements and modelling of vehicular emissions during traffic interruption and congestion events in an urban traffic corridor. Atmos. Pollut. Res. 10, 480–492. https://doi.org/10.1016/j.apr.2018.09.008

Churpek, M., Yuen, T., Winslow, C., Meltzer, D., Kattan, M., Edelson, D., 2017. Multicenter Comparison of Machine Learning and Conventional Regression for Predicting Clinical Deterioration on the

Wards. Crit. Care Med. 44, 368–374. https://doi.org/10.1097/CCM.0000000000001571

Çodur, M.Y., Tortum, A., 2015. An Artificial Neural Network Model for Highway Accident Prediction: A Case Study of Erzurum, Turkey. PROMET - Traffic & Transportation 27, 217–225. https://doi.org/10.7307/ptt.v27i3.1551

Committee on Climate Change, 2015. Sectoral scenarios for the Fifth Carbon Budget. London, UK: Committee on Climate Change. Available at: https://www.theccc.org.uk/publication/sectoral-scenarios-for-the-fifth-carbon-budget-technical-report/(Accessed: 25 February 2019)

Committee on Cliate Change, 2020a. The Sixth Carbon Budget: The UK's path to Net Zero. London, UK: Committee on Climate Change. Available at: https://www.theccc.org.uk/publication/sixth-carbon-budget/ (Accessed: 24 July 2021)

Committee on Climate Change, 2020b. The Sixth Carbon Budget: Surface Transport. London, UK: Committee on Climate Change. Available at: https://www.theccc.org.uk/publication/sixth-carbon-budget/ (Accessed: 24 July 2021)

Corsmeier, U., Imhof, D., Kohler, M., Kühlwein, J., Kurtenbach, R., Petrea, M., Rosenbohm, E., Vogel, B., Vogt, U., 2005. Comparison of measured and model-calculated real-world traffic emissions. Atmos. Environ. 39, 5760–5775. https://doi.org/10.1016/j.atmosenv.2005.06.048

Cortes, C., Vapnik, V., 1995. Support-Vector Networks. Mach. Learn. 20, 273–297. https://doi.org/10.1023/A:1022627411411

Cuddeback, G., Wilson, E., Orme, J.G., Combs-Orme, T., 2004. Detecting and Statistically Correcting Sample Selection Bias. J. Soc. Serv. Res. 30, 19–33. https://doi.org/10.1300/J079v30n03_02

Dallmann, T.R., DeMartini, S.J., Kirchstetter, T.W., Herndon, S.C., Onasch, T.B., Wood, E.C., Harley, R.A., 2012. On-Road Measurement of Gas and Particle Phase Pollutant Emission Factors for Individual Heavy-Duty Diesel Trucks. Environ. Sci. Technol. 46, 8511–8518. https://doi.org/10.1021/es301936c

Daly, H.E., Fais, B., 2014. UK TIMES Model Overview. London, UK: University College London, Energy Institute. Available at: https://iea-etsap.org/Applications/UK%20TIMES%20Model%20Overview.pdf (Accessed: 25 February 2019)

Das, S., Tsapakis, I., 2019. Interpretable machine learning approach in estimating traffic volume on low-volume roadways. Int. J. Transp. Sci. Technol. 9, 76–88. https://doi.org/10.1016/j.ijtst.2019.09.004

Davison, J., Bernard, Y., Borken-Kleefeld, J., Farren, N.J., Hausberger, S., Sjödin, Å., Tate, J.E., Vaughan, A.R., Carslaw, D.C., 2020. Distance-based emission factors from vehicle emission remote sensing measurements. Sci. Total Environ. 739, 139688. https://doi.org/10.1016/j.scitotenv.2020.139688

De Blasiis, M.R., Di Prete, M., Guattari, C., Veraldi, V., Chiatti, G., Palmieri, F., 2013. Investigating the influence of highway traffic flow condition on pollutant emissions using driving simulators. WIT Trans. Ecol. Environ. 174, 171–181. https://doi.org/10.2495/AIR130151

DEFRA, 2020. Clean Air Zones Framework. London, UK: Department for Environment, Food & Rural Affairs. Available at: https://www.sheffield.gov.uk/home/pollution-nuisance/clean-air-zone (Accessed: 16 June 2021)

DEFRA, 2019. Clean Air Strategy 2019. London, UK: Department for Environment, Food & Rural Affairs. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/770715/clean-air-strategy-2019.pdf (Accessed: 16 June 2021)

DEFRA, 2018. Emissions of Air Pollutants in the UK , 1970 To 2016. London, UK: Department for Environment, Food & Rural Affairs. Available at: https://www.burnright.co.uk/wp-content/uploads/2019/02/Statistics-Defra-air-pollution-graph.pdf (Accessed: 27 June 2018)

Degraeuwe, B., Pisoni, E., Christidis, P., Christodoulou, A., Thunis, P., 2021. SHERPA-city: A web application to assess the impact of traffic measures on NO_2 pollution in cities. Environ. Model. Softw. 135, 104904. https://doi.org/10.1016/j.envsoft.2020.104904

Delhaye, A., Marot, L., 2015. The Motorcycling Community in Europe. Brussels, Belgium: Federation of European Motorcyclists' Associations (FEMA). Available at: http://www.fema-online.eu/riderscan/IMG/pdf/deliverable9_motocyclingcommunityineurope.pdf (Accessed: 12 May 2020)

Department for Transport, 2020a. Vehicle Licensing Statistics: Annual 2019. London, UK: Department for Transport. Available at: https://www.gov.uk/government/statistics/vehicle-licensing-statistics-2019 (Accessed 27 September 2020)

Department for Transport, 2020b. Road Lengths in Great Britain 2019. London, UK: Department for Transport. Available at: https://www.gov.uk/government/statistics/road-lengths-in-great-britain-2019 (Accessed 22 September 2020)

Department for Transport, 2020c. Transport Statistics for Great Britain. London, UK: Department for Transport. Available at: https://www.gov.uk/government/collections/transport-statistics-great-britain (Accessed: 28 June 2021)

Department for Transport, 2019. Road Lengths in Great Britain 2016. London, UK: Department for Transport. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/611185/road-lengths-in-great-britain-2016.pdf (Accessed: 21 January 2020)

Department for Transport, 2018a. Transport energy and environment statistics 2011. London, UK: Department for Transport. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/8947/energy-2011.pdf (Accessed: 7 January 2020)

Department for Transport, 2018b. All traffic data metadata. London, UK: Department for Transport. Available at: https://data.dft.gov.uk/gb-traffic-matrix/all-traffic-data-metadata.pdf (Accessed: 4 March 2022)

Department for Transport, 2018c. Traffic Statistics Methodology Review - Alternative Data Sources. London, UK: Department for Transport. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/722515/traffic-statistics-methodology-review-alternative-data-sources.pdf (Accessed: 2 March 2022)

Department for Transport, 2018d. Road Traffic Forecasts 2018. London, UK: Department for Transport. Available at: https://www.gov.uk/government/publications/road-traffic-forecasts-2018 (Accessed: 20 May 2021)

Department for Transport, 2017. Road safety statistics: notes and definitions. London, UK: Department for Transport. Available at: https://www.gov.uk/government/publications/road-accidents-and-safety-statistics-guidance (Accessed: 25 February 2019)

Department for Transport, 2016a. Notes and definitions: Road Traffic Estimates. London, UK: Department for Transport. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_dat a/file/527807/road-traffic-notes-definitions.pdf (Accessed: 8 April 2021)

Department for Transport, 2016b. National Travel Survey: Motorcycle use in England. London, UK: Department for Transport. Available at: https://www.gov.uk/government/collections/national-travel-survey-statistics (Accessed: 12 May 2020)

Department for Transport, 2014. Road Traffic Estimates. London, UK: Department for Transport. Available at: https://www.gov.uk/government/collections/road-traffic-statistics (Accessed: 5 June 2019)

Department for Transport, 2013. Road Traffic Estimates Methodology Note. London, UK: Department for Transport. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/524848/annual-methodology-note.pdf (Accessed: 11 January 2018)

Department for Transport, 2012. Road Transport Forecasts 2011: Results from the Department for Transport's National Transport Model. London, UK: Department for Transport.

Department for Transport, 2010. Discrepancies between Road Freight and Road Traffic HGV traffic estimates. London, UK: Department for Transport. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/192486/hgv-traffic-statistics-note.pdf (Accessed: 29 March 2022)

Department for Transport (DfT), 2020. Transport Statistics Great Britain. London, UK: Department for Transport. Available at: https://www.gov.uk/government/collections/transport-statistics-great-britain (Accessed: 24 July 2021)

Diao, M., Ferreira, J., 2014. Vehicle miles traveled and the built environment: Evidence from vehicle safety inspection data. Environ. Plan. A 46, 2991–3009. https://doi.org/10.1068/a140039p

Diez, D., Barr, C., Cetinkaya-Rundel, M., 2012. OpenIntro Statistics, 3rd ed. OpenIntro, Boston, MA. https://doi.org/10.19232/uv4pb.2016.2.90 (Accessed: 12 May 2020)

Dombalyan, A., Kocherga, V., Semchugova, E., Negrov, N., 2017. Traffic Forecasting Model for a Road Section. Transp. Res. Procedia 20, 159–165. https://doi.org/10.1016/j.trpro.2017.01.040

Domingos, P., 2012. A few useful things to know about machine learning. Commun. ACM 55, 78–87. https://doi.org/10.1145/2347736.2347755

Doshi-Velez, F., Kim, B., 2017. Towards A Rigorous Science of Interpretable Machine Learning. arXiv Prepr. 1702, 1–13.

Doustmohammadi, M., Anderson, M., 2016. Developing Direct Demand AADT Forecasting Models for Small and Medium Sized Urban Communities. Int. J. Traffic Transp. Eng. 5, 27–31. https://doi.org/10.5923/j.ijtte.20160502.01

Drucker, H., Burges, C.J.C., Kaufman, L., Smola, A., Vapnik, V., 1997. Support Vector Regression Machines, in: 9th International Conference on Neural Information Processing Systems. MIT Press, Cambridge, MA, USA, Denver, Colorado, pp. 155–161.

Efron, B., 1979. Bootstrap Methods: Another Look at the Jackknife. Ann. Stat. 7, 1–26.

El Esawey, M., Mosa, A.I., Nasr, K., 2015. Estimation of daily bicycle traffic volumes using sparse data. Comput. Environ. Urban Syst. 54, 195–203. https://doi.org/10.1016/j.compenvurbsys.2015.09.002

Elkafoury, A., Bady, M., Aly, M.H.F., Negm, A.M., 2013. Emissions Modeling for Road Transportation in Urban Areas : State-of-Art Review, in: 23rd International Conference on "Environmental

Protection Is a Must." pp. 1–16. Available at: https://www.researchgate.net/publication/239729900_Emissions_Modeling_for_Road_Transportation_in_Urban_Areas_State-of-Art_Review (Accessed: 27 June 2018)

Elkafoury, A., Negm, A., Aly, M., Bady, M., Ichimura, T., 2015. Develop dynamic model for predicting traffic CO emissions in urban areas. Environ. Sci. Pollut. Res. 1–12. https://doi.org/10.1007/s11356-015-4319-8

Ellison, R.B., Greaves, S.P., Hensher, D.A., 2013. Five years of London's low emission zone: Effects on vehicle fleet composition and air quality. Transp. Res. Part D Transp. Environ. 23, 25–33. https://doi.org/10.1016/j.trd.2013.03.010

Eom, J.I.N.K.I., Man, S.I.K.P., Heo, T.-Y., Huntsinger, L.F., 2006. Improving the prediction of annual average daily traffic for nonfreeway facilities by applying a spatial statistical method. Transp. Res. Rec. 22–29. https://doi.org/10.3141/1968-03

Esteves-Booth, A., Muneer, T., Kubie, J., Kirby, H., 2002. A review of vehicular emission models and driving cycles. Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci. 216, 777–797. https://doi.org/10.1243/09544060260171429

European Environmental Agency, 2014. Costs of air pollution from European industrial facilities 2008–2012 — an updated assessment, EEA Technical report No 20/2014. Luxemburg. Availableat: http://www.worldatlas.com/articles/the-world-leaders-in-coconut-production.html (Accessed: 21 January 2020)

Facanha, C., Horvath, A., 2007. Evaluation of life-cycle air emission factors of freight transportation. Environ. Sci. Technol. 41, 7138–7144. https://doi.org/10.1021/es070989q

Fagnant, D.J., Kockelman, K., 2016. A direct-demand model for bicycle counts: the impacts of level of service and other factors. Environ. Plan. B Plan. Des. 43, 93–107. https://doi.org/10.1177/0265813515602568

Faith, D.P., Minchin, P.R., Belbin, L., 1987. Compositional dissimilarity as a robust measure of ecological distance. Vegetatio 69, 57–68. https://doi.org/10.1007/BF00038687

Fallahshorshani, M., André, M., Bonhomme, C., Seigneur, C., 2012. Coupling Traffic, Pollutant Emission, Air and Water Quality Models: Technical Review and Perspectives. Procedia - Soc. Behav. Sci. 48, 1794–1804. https://doi.org/10.1016/j.sbspro.2012.06.1154

Fecht, D., Hansell, A.L., Morley, D., Dajnak, D., Vienneau, D., Beevers, S., Toledano, M.B., Kelly, F.J., Anderson, H.R., Gulliver, J., 2016. Spatial and temporal associations of road traffic noise and air pollution in London: Implications for epidemiological studies. Environ. Int. 88, 235–242. https://doi.org/10.1016/j.envint.2015.12.001

Fitz, D.R., Bumiller, K., Bufalino, C., James, D.E., 2020. Real-time PM10 emission rates from paved roads by measurement of concentrations in the vehicle's wake using on-board sensors part 1. SCAMPER method characterization. Atmos. Environ. 230, 117483. https://doi.org/10.1016/j.atmosenv.2020.117483

Fitz, D.R., Bumiller, K., Etyemezian, V., Kuhns, H.D., Gillies, J.A., Nikolich, G., James, D.E., Langston, R., Merle, R.S., 2021. Real-time PM10 emission rates from paved roads by measurement of concentrations in the vehicle's wake using on-board sensors Part 2. Comparison of SCAMPER, TRAKER™, flux measurements, and AP-42 silt sampling under controlled conditions. Atmos. Environ. 256, 118453. https://doi.org/10.1016/j.atmosenv.2021.118453

Fontaras, G., Ciuffo, B., Zacharof, N., Tsiakmakis, S., Marotta, A., Pavlovic, J., Anagnostopoulos, K., 2017. The difference between reported and real-world CO 2 emissions: How much improvement can be expected by WLTP introduction? Transp. Res. Procedia 25, 3933–3943. https://doi.org/10.1016/j.trpro.2017.05.333

Fontaras, G., Franco, V., Dilara, P., Martini, G., Manfredi, U., 2014. Development and review of Euro 5 passenger car emission factors based on experimental results over various driving cycles. Sci. Total Environ. 468–469, 1034–1042. https://doi.org/10.1016/j.scitotenv.2013.09.043

Franco, V., Kousoulidou, M., Muntean, M., Ntziachristos, L., Hausberger, S., Dilara, P., 2013. Road vehicle emission factors development: A review. Atmos. Environ. 70, 84–97. https://doi.org/10.1016/j.atmosenv.2013.01.006

Freitas, B.C.B., Morais, M.G., Costa, J.A.V., 2017. Chlorella minutissima cultivation with CO_2 and pentoses: Effects on kinetic and nutritional parameters. Bioresour. Technol. 244, 338–344. https://doi.org/10.1016/j.biortech.2017.07.125

Fricker, J., Saha, S., 1987. Traffic Volume Forecasting Methods for Rural State Highways. West Lafayette, Indiana. Available at: https://onlinepubs.trb.org/Onlinepubs/trr/1988/1203/1203-002.pdf (Accessed: 6 May 2018)

Friedman, B., Gordon, S., Peers, J., 1994. Effect of Neotraditional Neighborhood Design on Travel Characteristics, Transportation Research Record. Transportation Research Board, Washington, DC.

Friedman, J.H., 2001. Greefy Function Approximation: A Gradient Boosting Machine. Ann. Stat. 29, 1189–1232.

Friedman, J.H., Meulman, J.J., 2004. Clustering objects on subsets of attributes. J. R. Stat. Soc. Ser. B Stat. Methodol. 66, 815–839. https://doi.org/10.1111/j.1467-9868.2004.02059.x

Fu, M., Kelly, J.A., Clinch, J.P., 2017. Estimating annual average daily traffic and transport emissions for a national road network: A bottom-up methodology for both nationally-aggregated and spatially-disaggregated results. J. Transp. Geogr. 58, 186–195. https://doi.org/10.1016/j.jtrangeo.2016.12.002

Gadda, S., Magoon, A., Kockelman, K.M., 2007. Estimates of AADT: Quantifying the Uncertainty, in: 86th Annual Meeting of the Transportation Research Board. p. 25.

Gao, S., Wang, Y., Gao, Y., Liu, Y., 2013. Understanding urban traffic-flow characteristics: A rethinking of betweenness centrality. Environ. Plan. B Plan. Des. 40, 135–153. https://doi.org/10.1068/b38141

Gaudry, M., Heinitz, F., Last, J., Mandel, B., 2000. Methodological Developments within the Quasi-Direct Format Demand Structure : the Multicountry Application for Passengers MAP-1. Piublication CRT-98-46, Centre de recherche sur les trasnports, University de Montreal, Montral and Working Paper N° 9815, Bureau d'economie theorique et appliquee (BETA), Universite Louis Pasteur, Strasbourg

Gebotys, C.H., Elmasry, M.I., 1989. Integration of Algorithmic VLSI Synthesis with Testability Incorporation. IEEE J. Solid-State Circuits 24, 409–416. https://doi.org/10.1109/4.18602

Gecchele, G., Rossi, R., Gastaldi, M., Caprini, A., 2011. Data Mining methods for Traffic monitoring data analysis: A case study. Procedia - Soc. Behav. Sci. 20, 455–464. https://doi.org/10.1016/j.sbspro.2011.08.052

Gelhausen, M.C., Berster, P., Wilken, D., 2018. A new direct demand model of long-term forecasting air passengers and air transport movements at German airports. J. Air Transp. Manag. 1–13. https://doi.org/10.1016/j.jairtraman.2018.04.001

Geller, M.D., Sardar, S.B., Phuleria, H., Fine, P.M., Sioutas, C., 2005. Measurements of Particle Number and Mass Concentrations and Size Distributions in a Tunnel Environment. Environ. Sci. Technol. 39, 8653–8663. https://doi.org/10.1021/es050360s

Ghosh, K., Singh, N., Maitra, B., 2019. Traffic Management During a Big Sport Event : A Case Study in Kolkata City. East. Asia Soc. Transp. Stud. 13, 1873–1891.

Gitelman, V., Doveh, E., Bekhor, S., 2017. The Relationship between Free-Flow Travel Speeds, Infrastructure Characteristics and Accidents, on Single-Carriageway Roads. Transp. Res. Procedia 25, 2026–2043. https://doi.org/10.1016/j.trpro.2017.05.398

Gower, J.T., 1971. A General Coefficient of Similarity and Some of Its Properties. Int. Biometric Soc. 27, 857–871.

Greater London Authority, 2019. Central London Ultra Low Emission Zone - Six Month Report. London, UK: Greater London Authority. Available at: https://www.london.gov.uk/sites/default/files/ulez_six_month_evaluation_report_oct19.pdf%0Ahttps://trid.trb.org/view/1673357 (Accessed: 16 June 2021)

Greenacre, M., Primicerio, R., 2013. Multivariate Analysis of Ecological Data. Fundacion BBVA, Bilbao, Spain.

Grieshop, A.P., Lipsky, E.M., Pekney, N.J., Takahama, S., Robinson, A.L., 2006. Fine particle emission factors from vehicles in a highway tunnel: Effects of fleet composition and season. Atmos. Environ. 40, 287–298. https://doi.org/10.1016/j.atmosenv.2006.03.064

Guerreiro, C., Gonzalez Ortiz, A., de Leeuw, F., Viana, M., Horalek, J., 2016. Air quality in Europe — 2016 report, European Environment Agency. Luxemburg. https://doi.org/10.2800/413142 (Accessed: 27 June 2018)

Gutiérrez, J., Cardozo, O.D., García-Palomares, J.C., 2011. Transit ridership forecasting at station level: An approach based on distance-decay weighted regression. J. Transp. Geogr. 19, 1081–1092. https://doi.org/10.1016/j.jtrangeo.2011.05.004

Hamanaka, R.B., Mutlu, G.M., 2018. Particulate Matter Air Pollution: Effects on the Cardiovascular System. Front. Endocrinol. (Lausanne). 9, 1–15. https://doi.org/10.3389/fendo.2018.00680

Hammah, R.E., Curran, J.H., 1999. On distance measures for the fuzzy K-means algorithm for joint data. Rock Mech. Rock Eng. 32, 1–27. https://doi.org/10.1007/s006030050041

Han, B., Zhang, K., Burnham, T., Kwon, E., Yu, X., 2013. Integration and road tests of a self-sensing CNT concrete pavement system for traffic detection. Smart Mater. Struct. 22, 015020. https://doi.org/10.1088/0964-1726/22/1/015020

Han, J., Kamber, M., Pei, J., 2012. Data Mining Concepts and Techniques, 3rd Editio. ed. Morgan Kaufmann. https://doi.org/10.1016/b978-0-12-381479-1.00001-0

Hancock, K.L., 2008. Freight Demand Modeling Tools for Public-Sector Decision Making Summary of A Conference, Freight Demand Modeling Tools for Public-Sector Decision Making. The National Academies Press, Washington, DC. https://doi.org/https://doi.org/10.17226/23090

Handy, S., 1996. Methodologies for exploring the link between urban form and travel behavior. Transp. Res. Part D Transp. Environ. 1, 151–165. https://doi.org/10.1016/S1361-9209(96)00010-7

Hankey, S., Lu, T., Mondschein, A., Buehler, R., 2017. Spatial models of active travel in small communities: Merging the goals of traffic monitoring and direct-demand modeling. J. Transp. Heal. 7, 149–159. https://doi.org/10.1016/j.jth.2017.08.009

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning, 2nd ed. Spr, New York. Springer, https://doi.org/10.1007/978-0-387-84858-7

Hausberger, S., Minarik, S., Fontaras, G., 2022. Proceedings of the 24th International Transport and Air Pollution (TAP) Conference, in: Proceedings of the 24th International Transport and Air Pollution (TAP) Conference. Luxemburg, pp. 1–312. https://doi.org/10.2760/019404

Hausberger, S., Rexeis, M., Zallinger, M., Luz, R., 2009. Emission Factors from the Model PHEM for the HBEFA Version 3, University of Technology, Graz, Report Nr. I-20/2009 Haus-Em.

Hauseberger, S., Rexeis, M., 2017. Update of Emission Factors for EURO4, EURO 5 and EURO 6 Dieasel Passenger Cars for the HBEFA Version 3.3, University of Technology, Graz, Report Nr. I-31/2013 Rex-EM-I. Berne, Switzerland. Available at: https://www.hbefa.net/e/documents/HBEFA3-3_TUG_finalreport_01062016.pdf (Accessed: 27 June 2018)

He, Z., Deng, S., Xu, X., 2006. Approximation Algorithms for K-Modes Clustering, in: Huang, D.-S., Li, K., Irwin, G.W. (Eds.), Computational Intelligence. Kumming, China, pp. 296–302.

Hébrail, G., Hugueney, B., Lechevallier, Y., Rossi, F., 2010. Exploratory analysis of functional data via clustering and optimal segmentation. Neurocomputing 73, 1125–1141. https://doi.org/10.1016/j.neucom.2009.11.022

Hess, D., 2001. Effect of Free Parking on Commuter Mode Choice: Evidence from Travel Diary Data. Transp. Res. Rec. 1753, 35–42. https://doi.org/10.3141/1753-05

Hesse, M., 2013. Cities and flows: Re-asserting a relationship as fundamental as it is delicate. J. Transp. Geogr. 29, 33–42. https://doi.org/10.1016/j.jtrangeo.2012.12.014

HM Government, 2021. Net Zero Strategy: Build Back Greener. London, UK. Available at: https://www.gov.uk/government/publications/net-zero-strategy (Accessed: 31 March 2022)

HM Government, 2020. The Ten Point Plan for a Green Industrial Revolution. London, UK. Available at: https://www.gov.uk/government/publications/the-ten-point-plan-for-a-green-industrial-revolution (Accessed: 19 April 2021)

HM Government, 2011. Implementing the Climate Change Act 2008 : The Government's proposal for setting the fourth carbon budget. London, UK. Available at: https://www.gov.uk/government/publications/implementing-the-climate-change-act-2008-the-government-s-proposal-for-setting-the-fourth-carbon-budget-policy-statement (Accessed: 19 April 2021)

Hoerl, A.E., Kennard, R.W., 1970. Ridge Regression: Applications to Nonorthogonal Problems. Technometrics 12, 69–82. https://doi.org/10.1080/00401706.1970.10488635

Hohn, M.E., 1999. The semivariogram, in: Geostatistics and Petroleum Geology. Springer Netherlands, Dordrecht, pp. 15–80. https://doi.org/10.1007/978-94-011-4425-4_2

Hong, S.P., Min, Y.H., Park, M.J., Kim, K.M., Oh, S.M., 2016. Precise estimation of connections of metro passengers from Smart Card data. Transportation (Amst). 43, 749–769. https://doi.org/10.1007/s11116-015-9617-y

Hoogendoorn, S., Knoop, V., 2012. Traffic flow theory and modelling, in: The Transport System and Transport Policy: An Introduction. Edward Elgar Publishing, pp. 125–159. https://doi.org/10.1680/muen.14.00001

Horowitz, A., 2006. Statewide Travel Forecasting Models. The National Academies Press. https://doi.org/10.17226/13958

Huang, J.Z., Ng, M.K., Rong, H., Li, Z., 2005. Automated variable weighting in k-means type clustering. IEEE Trans. Pattern Anal. Mach. Intell. 27, 657–668. https://doi.org/10.1109/TPAMI.2005.95

Huang, Y., Organ, B., Zhou, J.L., Surawski, N.C., Hong, G., Chan, E.F.C., Yam, Y.S., 2018. Remote sensing of on-road vehicle emissions: Mechanism, applications and a case study from Hong Kong. Atmos. Environ. 182, 58–74. https://doi.org/10.1016/j.atmosenv.2018.03.035

Huang, Z., 1998. Extensions to the k -Means Algorithm for Clustering Large Data Sets with Categorical Values. Data Min. Knowl. Discov. 304, 283–304.

Huang, Z., 1997a. A Fast Clustering Algorithm to Cluster Very Large Categorical Data Sets in Data, in: DMKD. pp. 34–39.

Huang, Z., 1997b. Clustering Large Data Sets with Mixed Numeric and Categorical Values, in: Proceedings of the 1st Asia-Pacific Conference on Knowledge, Discovery and Data Mining (PAKDD). pp. 2303–2307. https://doi.org/10.1109/WCICA.2004.1342001

Hueglin, C., Buchmann, B., Weber, R.O., 2006. Long-term observation of real-world road traffic emission factors on a motorway in Switzerland. Atmos. Environ. 40, 3696–3709. https://doi.org/10.1016/j.atmosenv.2006.03.020

Hummel, M., Edelmann, D., Kopp-Schneider, A., 2017. Clustering of samples and variables with mixed-type data. PLoS One 12, 1–23. https://doi.org/10.1371/journal.pone.0188274

Humphreys, B.R., Pyun, H., 2018. Professional sporting events and traffic: Evidence from U.S. cities. J. Reg. Sci. 58, 869–886. https://doi.org/10.1111/jors.12389

iCET, 2015. Calculating Urban Transportation Emissions: Private Vehicles. Los Angeles: Innovation Center for Energy and transportation. Available at: http://www.icet.org.cn/english/admin/upload/2015101235945161.pdf (Accessed: 27 June 2018)

Inci, E., van Ommeren, J., Kobus, M., 2017. The external cruising costs of parking. J. Econ. Geogr. 17, 1301–1323. https://doi.org/10.1093/jeg/lbx004

Iodice, Paolo, Senatore, A., 2015. Analytical-experimental analysis of last generation medium-size motorcycles emission behaviour under real urban conditions. Int. J. Automot. Mech. Eng. 12, 3018–3032.

Iodice, P., Senatore, A., 2015. Exhaust emissions of new high-performance motorcycles in hot and cold conditions. Int. J. Environ. Sci. Technol. 12, 3133–3144. https://doi.org/10.1007/s13762-014-0741-6

IPCC, 2022. Climate Change 2022: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [H.-O. Portner, D.C. Roberts, M. Tigor, E.S. Poloczanska, K. Mintenbeck, A. Alegria, M. Craig, S. Langsdorf, S. Loschke, V. Moller, A. Okem, B. Rama (eds)]. Cambridge University Press. In Press.

Jacobs, 2014. Ultra Low Emission Zone: Integrated Impact Assessment. London, UK. Available at: https://cleanair.london/app/uploads/CAL-294-ULEZ-HIA_Final_ULEZ-consultation_Reduced-size.pdf (Accessed: 10 June 2021)

Jacyna-Gołda, I., Gołębiowski, P., Izdebski, M., Kłodawski, M., Jachimowski, R., Szczepański, E., 2017. The evaluation of the sustainable transport system development with the scenario analyses procedure. J. Vibroengineering 19, 5627–5638. https://doi.org/10.21595/jve.2017.19275

Jahanshahi, K., Jin, Y., 2016. The built environment typologies in the UK and their influences on travel behaviour: new evidence through latent categorisation in structural equation modelling. Transp.

Plan. Technol. 39, 59–77. https://doi.org/10.1080/03081060.2015.1108083

Jain, M., Korzhenevych, A., Hecht, R., 2018. Determinants of commuting patterns in a rural-urban megaregion of India. Transp. Policy 68, 98–106. https://doi.org/10.1016/j.tranpol.2018.04.018

Jamriska, M., Morawska, L., 2001. A model for determination of motor vehicle emission factors from on-road measurements with a focus on submicrometer particles. Sci. Total Environ. 264, 241–255. https://doi.org/10.1016/S0048-9697(00)00720-8

Jamriska, M., Morawska, L., Thomas, S., He, C., 2004. Diesel Bus Emissions Measured in a Tunnel Study. Environ. Sci. Technol. 38, 6701–6709. https://doi.org/10.1021/es030662z

Järvinen, A., Timonen, H., Karjalainen, P., Bloss, M., Simonen, P., Saarikoski, S., Kuuluvainen, H., Kalliokoski, J., Dal Maso, M., Niemi, J. V., Keskinen, J., Rönkkö, T., 2019. Particle emissions of Euro VI, EEV and retrofitted EEV city buses in real traffic. Environ. Pollut. 250, 708–716. https://doi.org/10.1016/j.envpol.2019.04.033

Jayasinghe, A., Sano, K., Nishiuchi, H., 2015. Explaining Traffic Flow Patterns Using Centrality Measures. Int. J. Traffic Transp. Eng. 5, 134–149. https://doi.org/10.7708/ijtte.2015.5(2).05

Jeong, C.H., Wang, J.M., Hilker, N., Debosz, J., Sofowote, U., Su, Y., Noble, M., Healy, R.M., Munoz, T., Dabek-Zlotorzynska, E., Celo, V., White, L., Audette, C., Herod, D., Evans, G.J., 2019. Temporal and spatial variability of traffic-related $PM_{2.5}$ sources: Comparison of exhaust and non-exhaust emissions. Atmos. Environ. 198, 55–69. https://doi.org/10.1016/j.atmosenv.2018.10.038

Ježek, I., Katrašnik, T., Westerdahl, D., Močnik, G., 2015. Black carbon, particle number concentration and nitrogen oxide emission factors of random in-use vehicles measured with the on-road chasing method. Atmos. Chem. Phys. 15, 11011–11026. https://doi.org/10.5194/acp-15-11011-2015

Jiang, B., Liu, C., 2009. Street-based topological representations and analyses for predicting traffic flow in GIS. Int. J. Geogr. Inf. Sci. 23, 1119–1137. https://doi.org/10.1080/13658810701690448

Jiang, S., Smith, M.H., Kitchen, J., Ogawa, A., 2009. Development of an Engine-in-the-loop Vehicle Simulation System in Engine Dynamometer Test Cell, in: SAE International. pp. 1–9. https://doi.org/10.4271/2009-01-1039

Jianqin, Z., Peiyuan, Q., Yingchao, D., Mingyi, D., Feng, L., 2015. A space-time visualization analysis method for taxi operation in Beijing. J. Vis. Lang. Comput. 31, 1–8. https://doi.org/10.1016/j.jvlc.2015.09.002

Jimenez, J.L., Mcrae, G.J., Nelson, D.D., Zahniser, M.S., Kolb, C.E., 2000. Remote Sensing of NO and NO 2 Emissions from Heavy-Duty Diesel Trucks Using Tunable Diode Lasers. Environ. Sci. Technol. 34, 2380–2387. https://doi.org/10.1021/es9911622

Jin, T., Lu, K., Liu, S., Zhao, S., Qu, L., Xu, X., 2017. Chemical Characteristics of Particulate Matter Emission from a Heavy-Duty Diesel Engine Using ETC Cycle Dynamometer Test. Aerosol Air Qual. Res. 17, 406–411. https://doi.org/10.4209/aaqr.2016.06.0264

Jinturkar, S.P., Pawar, S.J., 2016. Vehicle Detection and Parameter Measurement using Smart Portable Sensor System. Commun. Appl. Electron. 4, 5–9. https://doi.org/10.5120/cae2016652132

Joumard, R., André, J., Rapone, M., Zallinger, M., Kljun, N., André, M., Samaras, Z., Roujol, S., Laurikko, J., Weilenmann, M., 2008. Emission Factor Modelling for Light Vehicles within the European ARTEMIS Model, in: 16th International Symposium of Transport and Air Pollution. HAL, Graz, Austria, pp. 1–10.

Joumard, R., Rapone, M., André, M., 2006. Analysis of the cars pollutant emissions as regards driving cycles and kinematic parameters.

Jung, S., Kim, Jounghwa, Kim, Jeongsoo, Hong, D., Park, D., 2017. An estimation of vehicle kilometer traveled and on-road emissions using the traffic volume and travel speed on road links in Incheon City. J. Environ. Sci. (China) 54, 90–100. https://doi.org/10.1016/j.jes.2015.12.040

Junqué de Fortuny, E., Martens, D., Provost, F., 2013. Predictive Modeling With Big Data: Is Bigger Really Better ? . Big Data 1, 215–226. https://doi.org/10.1089/big.2013.0037

Kang, L., Zhu, X., Sun, H., Wu, J., Gao, Z., Hu, B., 2019. Last train timetabling optimization and bus bridging service management in urban railway transit networks. Omega (United Kingdom) 84, 31–44. https://doi.org/10.1016/j.omega.2018.04.003

Karjalainen, P., Pirjola, L., Heikkilä, J., Lähde, T., Tzamkiozis, T., Ntziachristos, L., Keskinen, J., Rönkkö, T., 2014. Exhaust particles of modern gasoline vehicles: A laboratory and an on-road study. Atmos. Environ. 97, 262–270. https://doi.org/10.1016/j.atmosenv.2014.08.025

Kaufman, L., Rousseeuw, P., 1990. Finding Groups in Data - An Introduction to Cluster Analysis. John Wiley & Sons, Hoboken, New Jersey. https://doi.org/10.1002/9780470316801

Kęder, M., Grzeszczyk, R., Merkisz, J., Fuć, P., Lijewski, P., 2014. DESIGN OF A NEW ENGINE DYNAMOMETER TEST STAND FOR DRIVING CYCLE SIMULATION. J. KONES. Powertrain Transp. 21, 217–224. https://doi.org/10.5604/12314005.1130477

Kelly, J.A., Clinch, J.P., 2009. Temporal variance of revealed preference on-street parking price elasticity. Transp. Policy 16, 193–199. https://doi.org/10.1016/j.tranpol.2009.06.001

Kepaptsoglou, K., Stathopoulos, A., Karlaftis, M.G., 2017. Ridership estimation of a new LRT system: Direct demand model approach. J. Transp. Geogr. 58, 146–156. https://doi.org/10.1016/j.jtrangeo.2016.12.004

Khan, S.M., Islam, S., Khan, M.Z., Dey, K., Chowdhury, M., Huynh, N., Torkjazi, M., 2018. Development of Statewide Annual Average Daily Traffic Estimation Model from Short-Term Counts: A Comparative Study for South Carolina. Transp. Res. Rec. 2672, 55–64. https://doi.org/10.1177/0361198118798979

Kihara, N., Tsukamoto, T., Matsumoto, K., Ishida, K., Kon, M., Murase, T., 2000. Real-Time On-Board Measurement of Mass Emission of NOx, Fuel Consumption, Road Load, and Engine Output for Diesel Vehicles, in: SAE Technical Papers. https://doi.org/10.4271/2000-01-1141

Kim, J.Y., Lee, S., Kim, J.-Y., Park, H., 2017. A Study on Mode Choice of Trips to Sport Facilities Using SP Survey Data. J. Korean Soc. Transp. 35, 197–209. https://doi.org/10.7470/jkst.2017.35.3.197

Kim, S., Park, D., Heo, T.-Y., Kim, H., Hong, D., 2016. Estimating Vehicle Miles Traveled (VMT) in Urban Areas Using Regression Kriging. J. Adv. Transp. https://doi.org/10.1002/atr

Kitamura, R., Fujii, S., 1998. Two Computational Process Models of Activity-Travel Behavior. Theor. Found. Travel Choice Model. 251–279.

Kodinariya, T., Makwana, P., 2013. Review on determining number of Cluster in K-Means Clustering. Int. J. Adv. Res. Comput. Sciene Manag. Stud. 1, 90–95.

Koperski, K., Han, J., Stefanovic, N., 1998. An efficient two-step method for classification of spatial data, in: The Eighth International Symposium on Spatial Data Handling (SDH'98). Maine, USA, pp. 45–54

Koul, A., Becchio, C., Cavallo, A., 2018. Cross-validation approaches for replicability in psychology. Front. Psychol. 9, 1–4. https://doi.org/10.3389/fpsyg.2018.01117

Kousoulidou, M., Ntziachristos, L., Fontaras, G., Martini, G., Dilara, P., Samaras, Z., 2012. Impact of biodiesel application at various blending ratios on passenger cars of different fueling technologies. Fuel 98, 88–94. https://doi.org/10.1016/j.fuel.2012.03.038

Krause, R.W., Huisman, M., Steglich, C., Sniiders, T.A.B., 2018. Missing network data a comparison of different imputation methods, in: Proceedings of the 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2018. pp. 159–163. https://doi.org/10.1109/ASONAM.2018.8508716

Krecl, P., Harrison, R.M., Johansson, C., Targino, A.C., Beddows, D.C., Ellermann, T., Lara, C., Ketzel, M., 2021. Long-term trends in nitrogen oxides concentrations and on-road vehicle emission factors in Copenhagen, London and Stockholm. Environ. Pollut. 290, 118105. https://doi.org/10.1016/j.envpol.2021.118105

Kwon, S.B., Namgung, H.G., Jeong, W., Park, D., Eom, J.K., 2016. Transient variation of aerosol size distribution in an underground subway station. Environ. Monit. Assess. 188. https://doi.org/10.1007/s10661-016-5373-5

Labib, S.M., Neema, M.N., Rahaman, Z., Patwary, S.H., Shakil, S.H., 2018. Carbon dioxide emission and bio-capacity indexing for transportation activities: A methodological development in determining the sustainability of vehicular transportation systems. J. Environ. Manage. 223, 57–73. https://doi.org/10.1016/j.jenvman.2018.06.010

LAEI, 2013. London Atmospheric Emissions Inventory (LAEI) 2013 Methodology. London, UK. Available at: https://data.london.gov.uk/dataset/london-atmospheric-emissions-inventory-2013 (Accessed: 24 November 2020)

Lakkaraju, H., Kamar, E., Caruana, R., Leskovec, J., 2017. Interpretable & Explorable Approximations of Black Box Models, in: ArXiv Preprint. pp. 1–5.

Lam, W.H.K., Lo, H.K., 2004. Traffic assignment methods. Handb. Transp. Geogr. Spat. Syst. 609–625. https://doi.org/doi:10.1108/9781615832538-034

Larose, D.T., 2004. Discovering Knowledge in Data: An Introduction to Data Mining. John Wiley & Sons, Inc., Hoboken, NJ, USA. https://doi.org/10.1002/0471687545

Latake, P.T., Pawar, P., 2015. The Greenhouse Effect and Its Impacts on Environment. Int. J. Innov. Res. Creat. Technol. 1, 333–337.

Laurent, E., 2021. The Well-being Transition. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-030-67860-9

Law, T.H., Hamid, H., Goh, C.N., 2015. The motorcycle to passenger car ownership ratio and economic growth: A cross-country analysis. J. Transp. Geogr. 46, 122–128. https://doi.org/10.1016/j.jtrangeo.2015.06.007

Laybourn-Langton, L., Quilter-Pinner, H., 2016. London: Global Green City, Institute for Public Policy Research Briefing. London, UK. Available at: www.ippr.org/files/publications/pdf/london-global-green-city_Apr2016.pdf (Accessed: 10 June 2021)

Leatherman, J.M., 2018. Comparison of emissions measurement between a sensor-based compact emissions meter and a standard PEMS. West Virginia University Libraries. https://doi.org/10.33915/etd.3683

Leduc, G., 2008. Road Traffic Data : Collection Methods and Applications. EUR Number Tech. Note JRC 47967 JRC 47967, 55. https://doi.org/JRC 47967 - 2008

Lee, W. Do, Cho, S., Bellemans, T., Janssens, D., Wets, G., Choi, K., Joh, C.H., 2012. Seoul activity-based model: An Application of Feathers Solutions to Seoul Metropolitan Area. Procedia Comput. Sci.

10, 840–845. https://doi.org/10.1016/j.procs.2012.06.109

Lieberthal, B., Gardner, A.M., 2021. Connectivity, reproduction number, and mobility interact to determine communities' epidemiological superspreader potential in a metapopulation network. PLoS Comput. Biol. 17, 1–22. https://doi.org/10.1371/JOURNAL.PCBI.1008674

Liu, C., Susilo, Y.O., Karlström, A., 2016. Estimating changes in transport CO_2 emissions due to changes in weather and climate in Sweden. Transp. Res. Part D Transp. Environ. 49, 172–187. https://doi.org/10.1016/j.trd.2016.09.004

Liu, Z., Ge, Y., Johnson, K.C., Shah, A.N., Tan, J., Wang, C., Yu, L., 2011. Real-world operation conditions and on-road emissions of Beijing diesel buses measured by using portable emission measurement system and electric low-pressure impactor. Sci. Total Environ. 409, 1476–1480. https://doi.org/10.1016/j.scitotenv.2010.12.042

Lowry, M., 2014. Spatial interpolation of traffic counts based on origin-destination centrality. J. Transp. Geogr. 36, 98–105. https://doi.org/10.1016/j.jtrangeo.2014.03.007

Loyd, C.D., 2007. Local Models for Spatial Analysis. CRC Press, Boca Raton, Florida.

Lozhkina, O. V., Lozhkin, V.N., 2016. Estimation of nitrogen oxides emissions from petrol and diesel passenger cars by means of on-board monitoring: Effect of vehicle speed, vehicle technology, engine type on emission rates. Transp. Res. Part D Transp. Environ. 47, 251–264. https://doi.org/10.1016/j.trd.2016.06.008

Lu, T., Buehler, R., Mondschein, A., Hankey, S., 2017. Designing a bicycle and pedestrian traffic monitoring program to estimate annual average daily traffic in a small rural college town. Transp. Res. Part D Transp. Environ. 53, 193–204. https://doi.org/10.1016/j.trd.2017.04.017

Ma, X., Lei, W., Andréasson, I., Chen, H., 2012. An Evaluation of Microscopic Emission Models for Traffic Pollution Simulation Using On-board Measurement. Environ. Model. Assess. 17, 375–387. https://doi.org/10.1007/s10666-011-9296-9

Mahesh, S., Ramadurai, G., Nagendra, S.M.S., 2019. On-board measurement of emissions from freight trucks in urban arterials: Effect of operating conditions, emission standards, and truck size. Atmos. Environ. 212, 75–82. https://doi.org/10.1016/j.atmosenv.2019.05.030

Mahesh, S., Ramadurai, G., Shiva Nagendra, S.M., 2018. Real-world emissions of gaseous pollutants from diesel passenger cars using portable emission measurement systems. Sustain. Cities Soc. 41, 104–113. https://doi.org/10.1016/j.scs.2018.05.025

Mamakos, A., Carriero, M, Bonnel, P, Demircioglu, H., Douglas, K., Alessandrini, S., Forni, F., Montigny, F., Lesueur, D., Bonnel, Pierre, Carriero, Massimo, 2011. EU-PEMS PM EVALUATION PROGRAM - Third Report – Further Study on Post DPF PM/PN Emissions, JRC European Commission. Luxemburg. https://doi.org/10.2788/34274

Manisalidis, I., Stavropoulou, E., Stavropoulos, A., Bezirtzoglou, E., 2020. Environmental and Health Impacts of Air Pollution: A Review. Front. Public Heal. 8, 1–13. https://doi.org/10.3389/fpubh.2020.00014

Martin, W., McGuckin, N., 1998. Travel Estimation Techniques for Urban Planning. Transportation Research Record, Washington, DC. Available at: https://onlinepubs.trb.org/onlinepubs/nchrp/nchrp_rpt_365.pdf (Accessed: 25 May 2018)

Martinet, S., Liu, Y., Louis, C., Tassel, P., Perret, P., Chaumond, A., André, M., 2017. Euro 6 Unregulated Pollutant Characterization and Statistical Analysis of After-Treatment Device and Driving-Condition Impact on Recent Passenger-Car Emissions. Environ. Sci. Technol. 51, 5847–5855.

https://doi.org/10.1021/acs.est.7b00481

Martínez, L.M., Viegas, J.M., Silva, E.A., 2009. A traffic analysis zone definition: A new methodology and algorithm. Transportation (Amst). 36, 581–599. https://doi.org/10.1007/s11116-009-9214-z

Mascia, M., Hu, S., Han, K., North, R., Van Poppel, M., Theunis, J., Beckx, C., Litzenberger, M., 2017. Impact of Traffic Management on Black Carbon Emissions: a Microsimulation Study. Networks Spat. Econ. 17, 269–291. https://doi.org/10.1007/s11067-016-9326-x

Mathew, S., Pulugurtha, S.S., 2021. Comparative Assessment of Geospatial and Statistical Methods to Estimate Local Road Annual Average Daily Traffic. J. Transp. Eng. Part A Syst. 147, 04021035. https://doi.org/10.1061/jtepbs.0000542

Matzer, C., Hausberger, S., Lipp, S., Rexeis, M., 2017. A New Approach for Systematic use of PEMS Data in Emission Simulation, in: Journal of Earth Sciences and Geotechnical Engineering. pp. 1792–9660.

McCord, M.R., Yang, Y., Jiang, Z., Coifman, B., Goel, P.K., 2003. Estimating Annual Average Daily Traffic from Satellite Imagery and Air Photos Empirical Results. Transp. Res. Rec. 136–142. https://doi.org/10.3141/1855-17

McCormick, R.L., Ryan, L.B.A., Daniels, T.L., Yanowitz, J., Graboski, M.S., 1998. Comparison of Chasis Dynamometer In-Use Emissions with Engine Dynamometer FTP Emissions for Three Heavy-Duty Diesel Vehicles, in: SAE Technical Papers. pp. 1980–1988. https://doi.org/10.4271/982653

McCune, B., Grace, J., 2002. Distance Measures, in: McCune & Grace, Analysis of Ecological Communities. MJM Software, Gleneden Beach, Oregon pp. 45–57.

McNally, M.G., 2007. The Four-Step Model, in: Handbook of Transport Modelling. Emerald Group Publishing Limited, Oxford, UK, pp. 35–53. https://doi.org/10.1108/9780857245670-003

Medeiros, C.B., Joliveau, M., Jomier, G., De Vuyst, F., 2010. Managing sensor traffic data and forecasting unusual behaviour propagation. Geoinformatica 14, 279–305. https://doi.org/10.1007/s10707-010-0102-7

Melkumova, L.E., Shatskikh, S.Y., 2017. Comparing Ridge and LASSO estimators for data analysis. Procedia Eng. 201, 746–755. https://doi.org/10.1016/j.proeng.2017.09.615

Mellios, G., Hausberger, S., Keller, M., Samaras, C., Ntziachristos, L., Hausberger, S., Samaras, C., Mellios, G., Nziachristos, L., 2011. Parameterisation of fuel consumption and CO_2 emissions of passenger cars and light commercial vehicles for modelling purposes, Jrc. Luxemburg.

Meyer, T., 2012. Root mean square error compared to, and contrasted with, standard deviation. Surv. L. Inf. Sci. 72, 107–108.

Milenković, M., Stepanović, N., Glavić, D., Tubić, V., Ivković, I., Trifunović, A., 2020. Methodology for determining ecological benefits of advanced tolling systems. J. Environ. Manage. 258, 110007. https://doi.org/10.1016/j.jenvman.2019.110007

Mohamad, I. Bin, Usman, D., 2013. Standardization and its effects on K-means clustering algorithm. Res. J. Appl. Sci. Eng. Technol. 6, 3299–3303. https://doi.org/10.19026/rjaset.6.3638

Mohamad, D., Sinha, K., Kuczek, T., Scholer, C., 1998. Annual Average Daily Traffic Prediction Model for County Roads. Transp. Res. Rec. J. Transp. Res. Board 1617, 69–77. https://doi.org/10.3141/1617-10

Mokhtarian, P.L., Cao, X., 2008. Examining the impacts of residential self-selection on travel behavior: A focus on methodologies. Transp. Res. Part B Methodol. 42, 204–228.

https://doi.org/10.1016/j.trb.2007.07.006

Morawska, L., Jamriska, M., Thomas, S., Ferreira, L., Mengersen, K., Wraith, D., McGregor, F., 2005. Quantification of Particle Number Emission Factors for Motor Vehicles from On-Road Measurements. Environ. Sci. Technol. 39, 9130–9139. https://doi.org/10.1021/es050069c

Morawska, L., Ristovski, Johnson, Jayaratne, Mengersen, K., 2007. Novel Method for On-Road Emission Factor Measurements Using a Plume Capture Trailer. Environ. Sci. Technol. 41, 574–579. https://doi.org/10.1021/es060179z

Moreno-Torres, J.G., Saez, J.A., Herrera, F., 2012. Study on the impact of partition-induced dataset shift on k-fold cross-validation. IEEE Trans. Neural Networks Learn. Syst. 23, 1304–1312. https://doi.org/10.1109/TNNLS.2012.2199516

Morley, D.W., Gulliver, J., 2016. Methods to improve traffic flow and noise exposure estimation on minor roads. Environ. Pollut. 216, 746–754. https://doi.org/10.1016/j.envpol.2016.06.042

Munigety, C.R., Mathew, T. V., 2016. Towards Behavioral Modeling of Drivers in Mixed Traffic Conditions. Transp. Dev. Econ. 2, 1–20. https://doi.org/10.1007/s40890-016-0012-y

Munir, S., Mayfield, M., Coca, D., Mihaylova, L.S., 2020. A nonlinear land use regression approach for modelling NO2 concentrations in urban areas-Using data from low-cost sensors and diffusion tubes. Atmosphere (Basel). 11. https://doi.org/10.3390/atmos11070736

Murena, F., Prati, M.V., Costagliola, M.A., 2019. Real driving emissions of a scooter and a passenger car in Naples city. Transp. Res. Part D Transp. Environ. 73, 46–55. https://doi.org/10.1016/j.trd.2019.06.002

Musoro, J.Z., Zwinderman, A.H., Puhan, M.A., Ter Riet, G., Geskus, R.B., 2014. Validation of prediction models based on lasso regression with multiply imputed data. BMC Med. Res. Methodol. 14, 1–13. https://doi.org/10.1186/1471-2288-14-116

Nakaya, T., 2014. GWR4 User manual. Kyoto, Japan. Available at: https://www.semanticscholar.org/paper/GWR-4-.-09-User-Manual-GWR-4-Windows-Application-Nakaya/80959614149d242970797725838a614f0300c76d (Accessed: 9 August 2021)

National Academy of Sciences, 2010. Understanding the changing planet: Strategic directions for the geographical sciences. https://doi.org/10.17226/12860

Neveu, A.J., 1983. Quick-Response Procedures to Forecast Rural Traffic. Transp. Res. Rec. 944, 47–53.

Nghiem, T.-D., Nguyen, Y.-L.T., Le, A.-T., Bui, N.-D., Pham, H.-T., 2019. Development of the specific emission factors for buses in Hanoi, Vietnam. Environ. Sci. Pollut. Res. 26, 24176–24189. https://doi.org/10.1007/s11356-019-05634-9

Nine, R.D., Clark, N.N., Daley, J.J., Atkinson, C.M., 1999. Development of a heavy-duty chassis dynamometer driving route. Proc. Inst. Mech. Eng. Part D J. Automob. Eng. 213, 561–574. https://doi.org/10.1243/0954407991527107

Nosal, T., Miranda-Moreno, L.F., Krstulic, Z., 2014. Incorporating Weather. Transp. Res. Rec. J. Transp. Res. Board 2468, 100–110. https://doi.org/10.3141/2468-12

Nostikasari, D., 2015. Representations of everyday travel experiences: Case study of the Dallas-Fort Worth Metropolitan Area. Transp. Policy 44, 96–107. https://doi.org/10.1016/j.tranpol.2015.06.008

Ntziachristos, L., Gkatzoflias, D., Kouridis, C., 2009. Information Technologies in Environmental Engineering 5. https://doi.org/10.1007/978-3-540-88351-7

Ntziachristos, L., Samaras, Z., 2000. COPERT II Computer Programme to calculate Emissions from Road Transport. Copenhagen, Denmark: European Environment Agency. Available at: https://www.eea.europa.eu/publications/TEC06 (Accessed: 27 June 2018)

Nugmanova, Arndt, Hossain, Kim, 2019. Effectiveness of Ring Roads in Reducing Traffic Congestion in Cities for Long Run: Big Almaty Ring Road Case Study. Sustainability 11, 4973. https://doi.org/10.3390/su11184973

Oakil, A.T., Manting, D., Nijland, H., 2018. The role of individual characteristics in car ownership shortly after relationship dissolution. Transportation (Amst). 45, 1–12. https://doi.org/10.1007/s11116-018-9857-8

Office for National Statistics, 2019. Road transport and air emissions: Contribution of road transport to greenhouse gas and air pollutant emissions. London, UK. Available at: https://www.ons.gov.uk/economy/environmentalaccounts/articles/roadtransportandairemissions/2019-09-16 (Accessed: 7 January 2020)

Ogutu, J.O., Schulz-Streeck, T., Piepho, H.P., 2012. Genomic selection using regularized linear regression models: ridge regression, lasso, elastic net and their extensions, in: In BMC Proceedings. BioMed Central. pp. 1–6.

Olea, R.A., 1999. The Semivariogram, in: Olea, R.A. (Ed.), Geostatistics for Engineers and Earth Scientists. Springer US, Boston, MA, pp. 67–90. https://doi.org/10.1007/978-1-4615-5001-3_5

Olivier, J.G.J., Peters, J.A.H.., 2019. Trends in Global CO_2 and Total Greenhouse Gas. The Hague, Netherlands. Available at: https://www.pbl.nl/sites/default/files/downloads/pbl-2020-trends-in-global-co2-and-total-greenhouse-gas-emissions-2019-report_4068.pdf (Accessed: 16 June 2021)

Oluwatobi, A.N., Tayo, A.O., Oladele, A.T., Adesina, G.R., 2021. The design of a vehicle detector and counter system using inductive loop technology. Procedia Comput. Sci. 183, 493–503. https://doi.org/10.1016/j.procs.2021.02.089

Ong, H.C., Mahlia, T.M.I., Masjuki, H.H., 2011. A review on emissions and mitigation strategies for road transport in Malaysia. Renew. Sustain. Energy Rev. 15, 3516–3522. https://doi.org/10.1016/j.rser.2011.05.006

Oprešnik, S.R., Seljak, T., Bizjan, F., Katrašnik, T., 2012. Exhaust emissions and fuel consumption of a triple-fuel spark-ignition engine powered passenger car. Transp. Res. Part D Transp. Environ. 17, 221–227. https://doi.org/10.1016/j.trd.2011.08.002

Opsahl, T., Panzarasa, P., 2009. Clustering in weighted networks. Soc. Networks 31, 155–163. https://doi.org/10.1016/j.socnet.2009.02.002

Ordnance Survey, 2018. ITN Layer Urban Paths Theme [WWW Document]. URL https://digimap.edina.ac.uk/webhelp/os/data_information/os_products/os_mastermap_itn_urban_paths.htm (accessed 3.22.22).

Orru, H., Ebi, K.L., Forsberg, B., 2017. The Interplay of Climate Change and Air Pollution on Health. Curr. Environ. Heal. reports 4, 504–513. https://doi.org/10.1007/s40572-017-0168-6

Ortuzar, J. de D., Willumsen, L.G., 2011. Modelling Transport, 4th ed. John Wiley & Sons, Chichester, UK.

Osei, L.K., Ghaffarpasand, O., Pope, F.D., 2021. Real-World Contribution of Electrification and Replacement Scenarios to the Fleet Emissions in West Midland Boroughs, UK. Atmosphere (Basel). 12, 332. https://doi.org/10.3390/atmos12030332

Oxley, T., Apsimon, H., Dore, A., Sutton, M., Hall, J., Heywood, E., Gonzales, T., Campo, D., Warren, A.R., 2003. The UK Integrated Assessment Model, UKIAM: A National Scale Approach to the Analysis of Strategies for Abatement of Atmospheric Pollutants Under the Convention on Long-Range Transboundary Air Pollution. Integr. Assess. 4, 236–249. https://doi.org/10.1080/1389517049051538$16.00

Page, M., Kelly, C., Bristow, A., 2004. Exploring Scenarios to 2050 for Hydrogen Use in Transport in the UK, in: European Transport Conference. Strasbourg, France, pp. 49–67. https://doi.org/10.1258/itt.2010.100803

Pang, Y., Tsagatakis, I., Murrells, T., Brace, S., 2016. Methodology and Changes made in the 2014 NAEI Road Transport Inventory: A Briefing Note Produced for DECC on Changes in Fuel Consumption. Available at: https://www.gov.uk/government/publications/naei-road-transport-inventory-changes-made-in-2010-a-briefing-note-produced-by-aea-on-changes-in-fuel-consumption (Accessed: 7 June 2018)

Papagiannakis, A., Baraklianos, I., Spyridonidou, A., 2018. Urban travel behaviour and household income in times of economic crisis: Challenges and perspectives for sustainable mobility. Transp. Policy 65, 51–60. https://doi.org/10.1016/j.tranpol.2016.12.006

Paredes, M., Hemberg, E., O'Reilly, U.-M., Zegras, C., 2017. Machine learning or discrete choice models for car ownership demand estimation and prediction?, in: 2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS). IEEE, pp. 780–785. https://doi.org/10.1109/MTITS.2017.8005618

Patarasuk, R., Gurney, K.R., O'Keeffe, D., Song, Y., Huang, J., Rao, P., Buchert, M., Lin, J.C., Mendoza, D., Ehleringer, J.R., 2016. Urban high-resolution fossil fuel CO_2 emissions quantification and exploration of emission drivers for potential policy applications. Urban Ecosyst. 19, 1013–1039. https://doi.org/10.1007/s11252-016-0553-1

Pendyala, R.M., Kitamura, R., Kikuchi, A., Yamamoto, T., Fujii, S., 2005. Florida Activity Mobility Simulator. Transp. Res. Rec. J. Transp. Res. Board 1921, 123–130. https://doi.org/10.1177/0361198105192100114

Pierson, W.R., Brachaczek, W.W., 1982. Particulate Matter Associated with Vehicles on the Road. II. Aerosol Sci. Technol. 2, 1–40. https://doi.org/10.1080/02786828308958610

Pierson, W.R., Gertler, A.W., Robinson, N.F., Sagebiel, J.C., Zielinska, B., Bishop, G.A., Stedman, D.H., Zweidinger, R.B., Ray, W.D., 1996. Real-world automotive emissions—Summary of studies in the Fort McHenry and Tuscarora mountain tunnels. Atmos. Environ. 30, 2233–2256. https://doi.org/10.1016/1352-2310(95)00276-6

Pinjari, A., Eluru, N., Srinivasan, S., Guo, J., Copperman, R., Sener, I., Bhat, C., 2008. Cemdap: Modeling and microsimulation frameworks, software development, and verification., in: TRB 87th Annual Meeting Compendium of Papers DVD. Washington, DC, pp. 13–17.

Pinjari, A.R., Bhat, C.R., 2011. Activity-based Travel Demand Analysis, in: A Handbook of Transport Economics. Edward Elgar Publishing, pp. 213–248. https://doi.org/10.4337/9780857930873.00017

Pitaksringkarn, J., Tanwanichkul, L., Yamthale, K., 2018. A correlation between pavement skid resistance and wet-pavement related accidents in Thailand. MATEC Web Conf. 192, 02049. https://doi.org/10.1051/matecconf/201819202049

Pointer, G., 2005. The UK's major urban areas, in: Chappell, R. (Ed.), Focus on People and Migration. Palgrave Macmillan UK, London, pp. 45–60. https://doi.org/10.1007/978-1-349-75096-2_3

Pouresmaeili, M.A., Aghayan, I., Taghizadeh, S.A., 2018. Development of Mashhad driving cycle for passenger car to model vehicle exhaust emissions calibrated using on-board measurements. Sustain. Cities Soc. 36, 12–20. https://doi.org/10.1016/j.scs.2017.09.034

Prud'homme, R., Bocarejo, J.P., 2005. The London congestion charge: A tentative economic appraisal. Transp. Policy 12, 279–287. https://doi.org/10.1016/j.tranpol.2005.03.001

Puliafito, S.E., Allende, D., Pinto, S., Castesana, P., 2015. High resolution inventory of GHG emissions of the road transport sector in Argentina. Atmos. Environ. 101, 303–311. https://doi.org/10.1016/j.atmosenv.2014.11.040

Pulugurtha, S.S., Mathew, S., 2021. Modeling AADT on local functionally classified roads using land use, road density, and nearest nonlocal road data. J. Transp. Geogr. 93, 103071. https://doi.org/10.1016/j.jtrangeo.2021.103071

Pun, L., Zhao, P., Liu, X., 2019. A Multiple Regression Approach for Traffic Flow Estimation. IEEE Access 7, 35998–36009. https://doi.org/10.1109/ACCESS.2019.2904645

Rahman, M.M., Najaf, P., Fields, M.G., Thill, J.C., 2021. Traffic congestion and its urban scale factors: Empirical evidence from American urban areas. Int. J. Sustain. Transp. 0, 1–16. https://doi.org/10.1080/15568318.2021.1885085

Raja, P., Author, C., Doustmohammadi, M., Anderson, M.D., 2018. Estimation of Average Daily Traffic on Low Volume. Int. J. Traffic Transp. Eng. 7, 1–6. https://doi.org/10.5923/j.ijtte.20180701.01

Rajab, S., Al Kalaa, M., Refai, H., 2011. Classification and speed elimination of vehicles via tire detection using single-element piezoelectric sensor. J. Adv. Transp. 50, 1366–1385. https://doi.org/10.1002/atr

RAND, 2004. Motorcycles and Congestion : The Effect of Modal Split Full Summary, Department for Transport (with support from Transport for London). London, UK.

Raparthi, N., Debbarma, S., Phuleria, H.C., 2021. Development of real-world emission factors for on-road vehicles from motorway tunnel measurements. Atmos. Environ. X 10, 100113. https://doi.org/10.1016/j.aeaoa.2021.100113

Rasouli, S., Timmermans, H., 2014. Activity-based models of travel demand: Promises, progress and prospects. Int. J. Urban Sci. 18, 31–60. https://doi.org/10.1080/12265934.2013.835118

Rawi, R., Mall, R., Kunji, K., Shen, C.H., Kwong, P.D., Chuang, G.Y., 2017. PaRSnIP: Sequence-based protein solubility prediction using gradient boosting machine. Bioinformatics 34, 1092–1098. https://doi.org/10.1093/bioinformatics/btx662

Ren, W., Xue, B., Geng, Y., Lu, C., Zhang, Y., Zhang, L., Fujita, T., Hao, H., 2016. Inter-city passenger transport in larger urban agglomeration area: Emissions and health impacts. J. Clean. Prod. 114, 412–419. https://doi.org/10.1016/j.jclepro.2015.03.102

Rexeis, M., Hausberger, S., Kühlwein, J., Luz, R., 2013. Update of Emission Factors for EURO 5 and EURO 6 vehicles for the HBEFA Version 3 . 2. Univ. Technol. Graz, Rep. Nr. I-31/2013 Rex-EM-I 43, 1–31. https://doi.org/Report No. I-25/2013/ Rex EM-I 2011/20 679

Ricardo Energy & Environment, 2019. Air Quality damage cost update 2019. Hanwell, UK. Available at: https://uk-air.defra.gov.uk/assets/documents/reports/cat09/1902271109_Damage_cost_update_2018_FINAL_Issue_2_publication.pdf (Accessed: 21 January 2021)

Rich, J., Hansen, C.O., Vuk, G., Nielsen, O.A., Korzenewycg, A., Brocker, J., 2009. Traffic flow: Scenario, Traffic Forecast and Analysis of Traffic on the TEN-T, Taking into Consideration the External Dimension of the Union. Final Report. European Commission, Copenhagen, Denmark.

Roess, R., Prassas, E., McShane, W., 2011. Summary for Policymakers, in: Intergovernmental Panel on Climate Change (Ed.), Climate Change 2013 - The Physical Science Basis. Cambridge University Press, Cambridge, pp. 1–30. https://doi.org/10.1017/CBO9781107415324.004

Rokach, L., Maimon, O., 2014. Data Mining with Decision Trees, Series in Machine Perception and Artificial Intelligence. World Scientific. https://doi.org/10.1142/9097

Rousseeuw, P.J., 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. 20, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Rubino, L., Bonnel, P., Hummel, R., Krasenbrink, A., Manfredi, U., De Santi, G., 2008. On-road Emissions and Fuel Economy of Light Duty Vehicles using PEMS: Chase-Testing Experiment. SAE Int. J. Fuels Lubr. 1, 2008-01–1824. https://doi.org/10.4271/2008-01-1824

Rudin, C., 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat. Mach. Intell. 1, 206–215. https://doi.org/10.1038/s42256-019-0048-x

Sarlas, G., Axhausen, K.W., 2014. Towards a direct demand modeling approach, in: 14th Swiss Transport Research Conference. Ascona, Switzerland, pp. 1–15.

Schmitz, S., Weiand, L., Becker, S., Niehoff, N., Schwartzbach, F., von Schneidemesser, E., 2018. An assessment of perceptions of air quality surrounding the implementation of a traffic-reduction measure in a local urban environment. Sustain. Cities Soc. 41, 525–537. https://doi.org/10.1016/j.scs.2018.06.011

Schwarz, G., 1978. Estimating the Dimension of a Model. Ann. Stat. 6, 461–464. https://doi.org/10.1214/aos/1176344136

Seaver, W.L., Chatterjee, A., Seaver, M.L., 2000. Estimation of Traffic Volume on Rural Local Roads. Transp. Res. Rec. J. Transp. Res. Board 1719, 121–128. https://doi.org/10.3141/1719-15

Sekhar, C.R., Minal, Madhu, E., 2016. Mode Choice Analysis Using Random Forrest Decision Trees. Transp. Res. Procedia 17, 644–652. https://doi.org/10.1016/j.trpro.2016.11.119

Selby, B., Kockelman, K.M., 2013. Spatial prediction of traffic levels in unmeasured locations: Applications of universal kriging and geographically weighted regression. J. Transp. Geogr. 29, 24–32. https://doi.org/10.1016/j.jtrangeo.2012.12.009

Seo, J., Park, Jisu, Park, Junhong, Park, S., 2021. Emission factor development for light-duty vehicles based on real-world emissions using emission map-based simulation. Environ. Pollut. 270, 116081. https://doi.org/10.1016/j.envpol.2020.116081

Seriani, S., Fernández, R., 2015. Planning guidelines for metro–bus interchanges by means of a pedestrian microsimulation model. Transp. Plan. Technol. 38, 569–583. https://doi.org/10.1080/03081060.2015.1039235

Setyawan, A., Kusdiantoro, I., Syafi'i, 2015. The effect of pavement condition on vehicle speeds and motor vehicles emissions. Procedia Eng. 125, 424–430. https://doi.org/10.1016/j.proeng.2015.11.111

Shamo, B., Asa, E., Membah, J., 2015. Linear Spatial Interpolation and Analysis of Annual Average Daily Traffic Data. J. Comput. Civ. Eng. 29, 04014022. https://doi.org/10.1061/(asce)cp.1943-5487.0000281

Sharma, S., Lingras, P., Xu, F., Kilburn, P., 2001. Application of Neural Networks to Estimate AADT on Low-Volume Roads. J. Transp. Eng. 127, 426–432. https://doi.org/10.1061/(ASCE)0733-947X(2001)127:5(426)

Shen, L., Gao, T., Zhao, J., Wang, Limao, Wang, Lan, Liu, L., Chen, F., Xue, J., 2014. Factory-level measurements on CO_2 emission factors of cement production in China. Renew. Sustain. Energy Rev. https://doi.org/10.1016/j.rser.2014.03.025

Shimbel, A., 1953. Structural parameters of communication networks. Bull. Math. Biophys. 15, 501–507. https://doi.org/10.1007/BF02476438

Shojaeshafiei, M., Doustmohammadi, M., Subedi, S., Anderson, M., 2017. Comparison of Estimation Methodologies for Daily Traffic Count Pediction in Small and Medium Sized Communities. Int. J. Traffic Transp. Eng. 6, 71–75. https://doi.org/10.5923/j.ijtte.20170604.01

Shorter, J.H., Herndon, S., Zahniser, M.S., Nelson, D.D., Wormhoudt, J., Demerjian, K.L., Kolb, C.E., 2005. Real-Time Measurements of Nitrogen Oxide Emissions from In-Use New York City Transit Buses Using a Chase Vehicle. Environ. Sci. Technol. 39, 7991–8000. https://doi.org/10.1021/es048295u

Shoup, D.C., 2006. Cruising for parking. Transp. Policy 13, 479–486. https://doi.org/10.1016/j.tranpol.2006.05.005

Silva, J. de A., Morency, C., Goulias, K.G., 2012. Using structural equations modeling to unravel the influence of land use patterns on travel behavior of workers in Montreal. Transp. Res. Part A Policy Pract. 46, 1252–1264. https://doi.org/10.1016/j.tra.2012.05.003

Singer, B.C., Harley, R.A., 1996. A Fuel-Based Motor Vehicle Emission Inventory. J. Air Waste Manage. Assoc. 46, 581–593. https://doi.org/10.1080/10473289.1996.10467492

Smit, R., Bainbridge, S., Kennedy, D., Kingston, P., 2021. A decade of measuring on-road vehicle emissions with remote sensing in Australia. Atmos. Environ. 252, 118317. https://doi.org/10.1016/j.atmosenv.2021.118317

Smit, R., Dia, H., Morawska, L., 2009. Road Traffic Emission and Fuel Consumption Modelling: Trends, New Developments and Future Challenges, in: Demidov, S., Bonnet, J. (Eds.), Air Water and Soil Science and Technology. Nova Science Publishers, New York, pp. 29–68.

Smit, R., Ntziachristos, L., Boulter, P., 2010. Validation of road vehicle and traffic emission models – A review and meta-analysis. Atmos. Environ. 44, 2943–2953. https://doi.org/10.1016/j.atmosenv.2010.05.022

Sookun, A., Boojhawon, R., Rughooputh, S.D.D.V., 2014. Assessing greenhouse gas and related air pollutant emissions from road traffic counts: A case study for Mauritius. Transp. Res. Part D Transp. Environ. 32, 35–47. https://doi.org/10.1016/j.trd.2014.06.005

Sorensen, A., 2001. Subcentres and satellite cities: Tokyo's 20th century experience of planned polycentrism. Int. Plan. Stud. 6, 9–32. https://doi.org/10.1080/13563470120026505

Stead, D., 2001. Relationships between land use, socioeconomic factors, and travel patterns in Britain. Environ. Plan. B Plan. Des. 28, 499–528. https://doi.org/10.1068/b2677

Stedman, D.H., Bishop, G.A., 1990. An Analysis of On-Road Remote Sensing as a Tool for Automobile Emissions Control. Springfield, Illinois.

Stein, A., Corsten, L.C.A., 1991. Universal Kriging and Cokriging as a Regression Procedure. Biometrics 47, 575. https://doi.org/10.2307/2532147

Steindl, G., Pfeiffer, C., 2017. Comparison of Black Box Models for Load Profile Generation of District Heating Networks, in: 12th Sustainable Development of Energy, Water and Environment Systems. pp. 1–10.

Strecht, P., Cruz, L., Soares, C., Mendes-Moreira, J., Abreu, R., 2015. A Comparative Study of Classification and Regression Algorithms for Modelling Students' Academic Performance, in: Proceedings of the 8th International Conference on Educational Data Mining. pp. 392–395.

Sturm, P.J., Kirchweger, G., Hausberger, S., Almbauer, R.A., 1998. Instantaneous emission data and their use in estimating road traffic emissions. Int. J. Veh. Des. 20, 181–191. https://doi.org/10.1504/IJVD.1998.001844

Sun, X., Das, S., 2019. Estimating Annual Average Daily Traffic for Low-Volume Roadways: A Case Study in Louisiana, in: 12th International Conference on Low-Volume Roads. Transportation Research Board, Kalispell, Montana, pp. 190–201.

Taghizadeh-Hesary, Farhad, Taghizadeh-Hesary, Farzad, 2020. The Impacts of Air Pollution on Health and Economy in Southeast Asia. Energies 13, 1812. https://doi.org/10.3390/en13071812

Tawfeek, M.H., El-Basyouny, K., 2019. Estimating Traffic Volume on Minor Roads at Rural Stop-Controlled Intersections using Deep Learning. Transp. Res. Rec. 2673, 108–116. https://doi.org/10.1177/0361198119837236

Tibshirani, R., 1996. Regression Shrinkage and Selection via the Lasso. J. R. Stat. Soc. 58, 267–288.

Tietge, U., Mock, P., Franco, V., German, J., Bandivadekar, A., Ligterink, N., Lambrecht, U., 2015. From laboratory to road fuel consumption and CO_2 for passenger cars in Europe, International Council on Clean Transportation.

Transport for London, 2019. HGV safety permit guidance For operators entering London. London, UK. Available at: https://content.tfl.gov.uk/hgv-safety-permit-guidance-for-operators-entering-london2.pdf (Accessed: 12 May 2020)

Troy, A., Morgan Grove, J., O'Neil-Dunne, J., 2012. The relationship between tree canopy and crime rates across an urban-rural gradient in the greater Baltimore region. Landsc. Urban Plan. 106, 262–270. https://doi.org/10.1016/j.landurbplan.2012.03.010

Tsagatakis, I., Brace, S., Passant, N., Pearson, B., Kiff, B., Richardson, J., Ruddy, M., 2017. UK Emission Mapping Methodology 2007, AEA, Harwell, UK. Didcot, UK.

Tzouras, P.G., Lázaro, C.P., 2018. Illegal parking in urban streets: connection with the geometric characteristics and its mitigation through traffic calming measures, in: 5th Panhellenic Conference of Planning and Regional Development. pp. 45–63.

U.N, 2015. Conference of the Parties, Adoption of Paris Agreement. https://doi.org/10.1201/9781351116589-2

UK Government, 2020. UK sets ambitious new climate target ahead of UN Summit [WWW Document]. Dep. Business, Energy Ind. Strateg. URL https://www.gov.uk/government/news/uk-sets-ambitious-new-climate-target-ahead-of-un-summit (Accessed: 19 April 2021)

Unal, A., Frey, H.C., Rouphail, N.M., 2004. Quantification of Highway Vehicle Emissions Hot Spots Based upon On-Board Measurements. J. Air Waste Manage. Assoc. 54, 130–140. https://doi.org/10.1080/10473289.2004.10470888

US Environmental Protection Agency, 2017. Inventory of U.S. Greenhouse Gas Emissions and Sinks, Nihon Hōshasen Gijutsu Gakkai zasshi. https://doi.org/10.6009/jjrt.2015_JSRT_71.10.1027

Valach, A.C., Langford, B., Nemitz, E., Mackenzie, A.R., Hewitt, C.N., 2015. Seasonal and diurnal trends in concentrations and fluxes of volatile organic compounds in central London. Atmos. Chem. Phys. 15, 7777–7796. https://doi.org/10.5194/acp-15-7777-2015

Van Wee, B., 2009. Self-selection: A key to a better understanding of location choices, travel behaviour and transport externalities? Transp. Rev. 29, 279–292. https://doi.org/10.1080/01441640902752961

Vanhulsel, M., Degraeuwe, B., Beckx, C., Vankerkom, J., De Vlieger, I., 2014. Road transportation emission inventories and projections - Case study of Belgium: Methodology and pitfalls. Transp. Res. Part D Transp. Environ. 27, 41–45. https://doi.org/10.1016/j.trd.2013.12.002

Vasic, A.M., Weilenmann, M., 2006. Comparison of real-world emissions from two-wheelers and passenger cars. Environ. Sci. Technol. 40, 149–154. https://doi.org/10.1021/es0481023

Vaughan, A.R., Lee, J.D., Misztal, P.K., Metzger, S., Shaw, M.D., Lewis, A.C., Purvis, R.M., Carslaw, D.C., Goldstein, A.H., Hewitt, C.N., Davison, B., Beevers, S.D., Karl, T.G., 2016. Spatially resolved flux measurements of NO_x from London suggest significantly higher emissions than predicted by inventories. Faraday Discuss. 189, 455–472. https://doi.org/10.1039/c5fd00170f

Vojtisek-Lom, M., Zardini, A.A., Pechout, M., Dittrich, L., Forni, F., Montigny, F., Carriero, M., Giechaskiel, B., Martini, G., 2020. A miniature Portable Emissions Measurement System (PEMS) for real-driving monitoring of motorcycles. Atmos. Meas. Tech. 13, 5827–5843. https://doi.org/10.5194/amt-13-5827-2020

Wang, H., Fu, L., Bi, J., 2011. CO_2 and pollutant emissions from passenger cars in China. Energy Policy 39, 3005–3011. https://doi.org/10.1016/j.enpol.2011.03.013

Wang, H., McGlinchy, I., 2009. Review of vehicle emission modelling and the issues for New Zealand, in: 32nd Australasian Transport Research Forum. Transportation Research Board, Auckland, New Zealand, pp. 1–13.

Wang, J., Hankey, S., Wu, X., Lindsey, G., 2016. Monitoring and Modeling of Urban Trail Traffic. Transp. Res. Rec. 2593, 47–59. https://doi.org/10.3141/2593-06

Wang, T., Gan, A., Alluri, P., 2013. Estimating annual average daily traffic for local roads for highway safety analysis. Transp. Res. Rec. 5, 60–66. https://doi.org/10.3141/2398-07

Wang, X., Kockelman, K., 2009. Forecasting Network Data. Transp. Res. Rec. 2105, 100–108. https://doi.org/10.3141/2105-13

Wardman, M., Whelan, G.A., Toner, J.P., 1994. Direct Demand Models of Air Travel: A Novel Approach to the Analysis of Stated Preference Data (No. 421), White Rose Research Online. Leeds, UK. Available at: http://eprints.whiterose.ac.uk/2157/ (Accessed: 25 May 2018)

Wardrop, J.G., 1952. Some Theoretical Aspects of Road Traffic Research. Proc. Inst. Civ. Eng. 1, 325–362. https://doi.org/10.1680/ipeds.1952.11259

Watson, J., Tetteh, A., Dutton, G., Bristow, A., Kelly, C., Page, M., 2004. UK Hydrogen Futures to 2050 UK (No. 46), Futures. Available at: https://www.researchgate.net/publication/228701890_UK_Hydrogen_Futures_to_2050 (Accessed: 25 February 2019)

Webb, E., Laverty, A., Mindell, J., Millett, C., 2016. Free bus travel and physical activity, gait speed, and adiposity in the english longitudinal study of ageingxs. Am. J. Public Health 106, 136–142. https://doi.org/10.2105/AJPH.2015.302907

Weiss, M., Bonnel, P., Hummel, R., 2011. Analyzing on-road emissions of light-duty vehicles with portable emission measurement systems (PEMS), JRC Scientific and Technical Reports. Luxemburg.

Weller, K., Lipp, S., Röck, M., Matzer, C., Bittermann, A., Hausberger, S., 2019. Real World Fuel

Consumption and Emissions From LDVs and HDVs. Front. Mech. Eng. 5, 1–22. https://doi.org/10.3389/fmech.2019.00045

Wen, Y., Wang, H., Larson, T., Kelp, M., Zhang, S., Wu, Y., Marshall, J.D., 2019. On-highway vehicle emission factors, and spatial patterns, based on mobile monitoring and absolute principal component score. Sci. Total Environ. 676, 242–251. https://doi.org/10.1016/j.scitotenv.2019.04.185

Westerdahl, D., Wang, X., Pan, X., Zhang, K.M., 2009. Characterization of on-road vehicle emission factors and microenvironmental air quality in Beijing, China. Atmos. Environ. 43, 697–705. https://doi.org/10.1016/j.atmosenv.2008.09.042

WHO, 2016. Ambient air pollution: A global assessment of exposure and burden disease. Geneva, Switzerland. Available at: https://apps.who.int/iris/handle/10665/250141 (Accessed: 16 June 2021)

Wihersaari, H., Pirjola, L., Karjalainen, P., Saukko, E., Kuuluvainen, H., Kulmala, K., Keskinen, J., Rönkkö, T., 2020. Particulate emissions of a modern diesel passenger car under laboratory and real-world transient driving conditions. Environ. Pollut. 265, 114948. https://doi.org/10.1016/j.envpol.2020.114948

Wilde, M.M., 2002. Distance Measures, in: Quantum Information Theory. Cambridge University Press, Cambridge, pp. 203–231. https://doi.org/10.1017/CBO9781139525343.010

Wong, K.I., 2013. An Analysis of Car and Motorcycle Ownership in Macao. Int. J. Sustain. Transp. 7, 204–225. https://doi.org/10.1080/15568318.2013.710057

Wong, T.T., 2015. Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. Pattern Recognit. 48, 2839–2846. https://doi.org/10.1016/j.patcog.2015.03.009

Wu, J., Xu, H., 2019. Annual Average Daily Traffic Prediction Model for Minor Roads at Intersections. J. Transp. Eng. Part A Syst. 145, 04019041. https://doi.org/10.1061/jtepbs.0000262

Wyatt, D.W., 2017. Assessing Micro Scale Carbon Dioxide (CO_2) Emission on UK Road Networks using a Coupled Traffic Simulation and Vehicle Emission Model. PhD Thesis, University of Leeds. Leeds, UK. Accessed: 2 May 2018, https://etheses.whiterose.ac.uk/19238/

Xia, Q., Zhao, F., Chen, Z., Shen, L.D., Ospina, D., 1999. Estimation of Annual Average Daily Traffic for Nonstate Roads in a Florida County. Transp. Res. Rec. 1660, 32–40. https://doi.org/10.3141/1660-05

Xie, Y., Dai, H., Zhang, Y., Wu, Y., Hanaoka, T., Masui, T., 2019. Comparison of health and economic impacts of $PM_{2.5}$ and ozone pollution in China. Environ. Int. 130, 104881. https://doi.org/10.1016/j.envint.2019.05.075

Yadav, S., Shukla, S., 2016. Analysis of k-Fold Cross-Validation over Hold-Out Validation on Colossal Datasets for Quality Classification. Proc. - 6th Int. Adv. Comput. Conf. IACC 2016 78–83. https://doi.org/10.1109/IACC.2016.25

Yang, B., Wang, S.G., Bao, Y., 2011. Efficient local AADT estimation via SCAD variable selection based on regression models. Proc. 2011 Chinese Control Decis. Conf. CCDC 2011 1898–1902. https://doi.org/10.1109/CCDC.2011.5968510

Yang, Y., Zou, H., 2015. A fast unified algorithm for solving group-lasso penalize learning problems. Stat. Comput. 25, 1129–1141. https://doi.org/10.1007/s11222-014-9498-5

Yang, Z., Deng, B., Deng, M., Huang, S., 2018. An Overview of Chassis Dynamometer in the Testing of Vehicle Emission. MATEC Web Conf. 175, 02015.

https://doi.org/10.1051/matecconf/201817502015

Yanowitz, J., McCormick, R.L., Graboski, M.S., 2000. In-Use Emissions from Heavy-Duty Diesel Vehicles. Environ. Sci. Technol. 34, 729–740. https://doi.org/10.1021/es990903w

Yonar, F., 2019. Piezoelectric Devices in Traffic Information Collection, in: International Symposium on Engineering, Natural Sciences and Architecture. Kocaeli, Turkey, pp. 1–5.

Yu, X., Prevedouros, P.D., Sulijoadikusumo, G., 2010. Evaluation of Autoscope, SmartSensor HD, and Infra-Red Traffic Logger for Vehicle Classification. Transp. Res. Rec. J. Transp. Res. Board 2160, 77–86. https://doi.org/10.3141/2160-09

Yu, Y.J., Cho, M.-G., 2008. A Short-Term Prediction Model for Forecasting Traffic Information Using Bayesian Network. 2008 Third Int. Conf. Converg. Hybrid Inf. Technol. 242–247. https://doi.org/10.1109/ICCIT.2008.355

Zahir, M.H., Haron, S., 2019. The study of headway patterns and level of service in myBAS ridership. J. Phys. Conf. Ser. 1349. https://doi.org/10.1088/1742-6596/1349/1/012082

Zhang, J., Li, Z., Pu, Z., Xu, C., 2018. Comparing prediction performance for crash injury severity among various machine learning and statistical methods. IEEE Access 6, 60079–60087. https://doi.org/10.1109/ACCESS.2018.2874979

Zhang, M., 2007. The Role of Land Use in Travel Mode Choice: Evidence from Boston and Hong Kong. J. Am. Plan. Assoc. 70, 344–360. https://doi.org/10.1080/01944360408976383

Zhang, S., Zhao, P., He, L., Yang, Y., Liu, B., He, W., Cheng, Y., Liu, Y., Liu, S., Hu, Q., Huang, C., Wu, Y., 2020. On-board monitoring (OBM) for heavy-duty vehicle emissions in China: Regulations, early-stage evaluation and policy recommendations. Sci. Total Environ. 731, 139045. https://doi.org/10.1016/j.scitotenv.2020.139045

Zhang, X., Chen, M., 2020. Enhancing statewide annual average daily traffic estimation with ubiquitous probe vehicle data. Transp. Res. Rec. 2674, 649–660. https://doi.org/10.1177/0361198120931100

Zhang, Y., Cheng, T., Aslam, N.S., 2019. Exploring the Relationship Between Travel Pattern and Social-Demographics Using Smart Card Data and Household Survey. ISPRS - Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. XLII-2/W13, 1375–1382. https://doi.org/10.5194/isprs-archives-XLII-2-W13-1375-2019

Zhao, F., Chung, S., 2001. Contributing Factors of Annual Average Daily Traffic in a Florida County: Exploration with Geographic Information System and Regression Models. Transp. Res. Rec. 1769, 113–122. https://doi.org/10.3141/1769-14

Zhao, F., Park, N., 2004. Using Geographically Weighted Regression Models to Estimate Annual Average Daily Traffic. Transp. Res. Rec. 1879, 99–107. https://doi.org/10.3141/1879-12

Zhao, J., Deng, W., Song, Y., Zhu, Y., 2014. Analysis of Metro ridership at station level and station-to-station level in Nanjing: An approach based on direct demand models. Transportation (Amst). 41, 133–155. https://doi.org/10.1007/s11116-013-9492-3

Zhao, S., Zhao, P., Cui, Y., 2017. A network centrality measure framework for analyzing urban traffic flow: A case study of Wuhan, China. Physica A 478, 143–157. https://doi.org/10.1016/j.physa.2017.02.069

Zheng, P., Mike, M., 2012. An Investigation on the Manual Traffic Count Accuracy. Procedia - Soc. Behav. Sci. 43, 226–231. https://doi.org/10.1016/j.sbspro.2012.04.095

Zheng, Y., Weng, Q., 2016. Evaluation of the correlation between remotely sensing-based and GIS-based anthropogenic heat discharge in Los Angeles County, USA. 4th Int. Work. Earth Obs. Remote Sens. Appl. EORSA 2016 - Proc. 324–328. https://doi.org/10.1109/EORSA.2016.7552822

Zhou, X., Tanvir, S., Lei, H., Taylor, J., Liu, B., Rouphail, N.M., Frey, H.C., 2015. Integrating a simplified emission estimation model and mesoscopic dynamic traffic simulator to efficiently evaluate emission impacts of traffic management strategies. Transp. Res. D Transp. Environ. 37, 123–136. https://doi.org/10.1016/j.trd.2015.04.013

Zhuge, C., Shao, C., Wang, S., Hu, Y., 2017. Sensitivity analysis of integrated activity-based model: using MATSim as an example. Transp. Lett. 7867, 1–11. https://doi.org/10.1080/19427867.2017.1286772

Zoico, E., Rossi, A., Di Francesco, V., Sepe, A., Olioso, D., Pizzini, F., Fantin, F., Bosello, O., Cominacini, L., Harris, T.B., Zamboni, M., 2010. Adipose tissue infiltration in skeletal muscle of healthy elderly men: Relationships with body composition, insulin resistance, and inflammation at the systemic and tissue level. Journals Gerontol. - Ser. A Biol. Sci. Med. Sci. 65 A, 295–299. https://doi.org/10.1093/gerona/glp155

Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. J. R. Stat. Soc. 134, 1080. https://doi.org/10.1016/S0140-6736(02)29256-9

Appendix

A. Supplementary material for chapter 1

Perhaps reflecting the several disciplines investigating transport, terminology is sometimes used with contradicting meanings. In fact, a great deal of confusion is related to the terms 'traffic volume' and 'traffic flow', sometimes used interchangeably. With regard to the former, I follow Zhao & Park (2004), so that traffic volume is defined as "the number of vehicles that pass a point on a highway for a given lane or a given direction of a highway and during a specified time interval". With regard to the latter, I follow Hoogendoorn & Knoop (2012), so that traffic flow is defined as the average number of vehicles passing a cross section in one unit of time, essentially traffic volume (n) as defined above divided by the length of the observation period (T) so that traffic flow q can be represented as $q = \frac{n}{T}$. The confusion between traffic volume and traffic flow implies that Annual Average Daily Traffic (AADT), one of the main variables used in my study, is sometimes used to measure traffic flows, e.g. Leduc (2008) and Pang et al. (2016), while other authors, e.g. Roess et al. (2011) and Sharma et al. (2001), use AADTs to measure traffic volumes. In fact, Leduc (2008) mention that the term traffic volume may be used to frame analysis of AADT, particularly in the US. For the sake of clarity, I will use AADT as a measure of traffic volume, my choice reflecting the formalisation of AADT below which is taken from Leduc (2008):

$$AADT_i = \sum_{j=1}^{365} \frac{TC_{i,j}^{24}}{365} \tag{24}$$

where $TC_{i,j}^{24}$ is the 24-hour traffic count on road link i at day j. Similarly, speed is defined by Hoogendoorn & Knoop (2012), as:

$$u = \frac{q}{k} \tag{25}$$

where q is flow and k is density with density defined as the number of vehicles per distance unit:

$$k = \frac{m}{X} = \frac{m}{\sum_{i=1}^{m} s_i} = \frac{1}{\bar{s}} \tag{26}$$

where X is the length of the street segment and m is the number of vehicles.

B. Supplementary material for chapter 2

Table B-1: Road categories used in the calculation of expansion factors

Category	Description
01	Motorways in holiday areas
02	Motorways in other rural areas with an estimated AADF of up to 59,999
03	Motorways in other rural areas with an estimated AADF of 60,000 or more
04	Motorways in part rural and part urban areas and conurbations
05	Motorways in mostly urban areas and Greater London
06	Rural ‘A’ roads in holiday and very rural areas with an estimated AADT of up to 4,999
07	Rural ‘A’ roads in holiday and very rural areas with an estimated AADT of between 5,000 and 7,999
08	Rural ‘A’ roads in holiday and very rural areas with an estimated AADT of 8,000 or more
09	Rural ‘A’ roads in all other areas with an estimated AADT of up to 13,999
10	Rural ‘A’ roads in all other areas with an estimated AADT of 14,000 or more
11	Urban ‘A’ roads in holiday areas
12	Urban ‘A’ roads in all other areas except Greater London with an estimated AADT of up to 19,999
13	Urban ‘A’ roads in all other areas except Greater London with an estimated AADT of 20,000 or more
14	Urban ‘A’ roads in Outer London
15	Urban ‘A’ roads in Inner London
16	Urban ‘A’ roads in Central London
50	Minor rural roads in holiday areas with an estimated AADT of up to 399
51	Minor rural roads in holiday areas with an estimated AADT of 400 or more
52	Minor rural roads in all other areas with an estimated AADT of up to 2,499
53	Minor rural roads in all other areas with an estimated AADT of 2,500 or more
54	Minor urban roads in all areas except Greater London
55	Minor urban roads in Greater London

C. Supplementary material for chapter 3

Table C-1: Outline of used datasets

Dataset	Source	Description	Spatial	Date
1. Traffic count points	Department for Transport (DfT)	Geocoded count points in England and Wales	N	2015
2. Integrated Transport Network (ITN)	Ordnance Survey (OS)	road network in Great Britain (GB) – roads and road junctions	Y	2015
3. Urban Paths (ITNUP)	Ordnance Survey (OS)	man-made footpaths, subways, steps, footbridges, and cycle paths in Britain's urban areas	Y	2015
4. Lower Super Output Areas (LSOAs)	Ordnance Survey (OS)	Designated areas for England and Wales with minimum 1000 population	Y	2011
5. Socioeconomic Characteristics	Office for National Statistics (ONS)	population, population density, workplace population, workplace density, number of households and median income at each LSOA	N	2011
6. Number of registered vehicles	Office for Low Emission Vehicles (OLEV)	Number of registered cars and vans for each LSOA	N	2011
7. Urban Area polygons	Ordnance Survey (OS)	Urban Areas boundaries	Y	2016
8. Bus stops and stations	National Public Transport Access Nodes (NaPTAN) database	Geolocated bus stops and stations in Britain	N	2016
9. Train and light train stations	National Public Transport Access Nodes (NaPTAN) database	Geolocated train and light rail stations in Britain	N	2016
10. Ports	British Port Association	Geolocated passenger and commercial ports	N	2015
11. Airports	Civil Aviation Authority	Geolocated passenger airports	N	2015
12. Charging points	Office for Low Emission Vehicles (OLEV)	Geolocated charging points for electric vehicles	N	2016
13. Non-domestic properties	Valuation Office Agency (VOA)	Geotagged and classified properties in England and Wales	N	2017

Table C-2: VOA re-classified land-use categories

CLASS	ELEMENTS
Research, Education and Training	Schools, Colleges, Libraries, Universities, Language and Music Schools, etc.
Factories, Workshops, and Industrial Activity	Energy Production Facilities, Factories, Workshops, Mines, Oil Fields, Recycling Plants, Shipyards, Scrap Yards
Healthcare	Hospitals, GPs, Surgeries, Clinics
Leisure	Public Houses, Bars, Nightclubs, Restaurants, Art Galleries, Cinemas and Theatres, Coffee shops
Office and Business Space	Offices, Banks, Business Units
Public Services, Infrastructure and Buildings	Post Offices, Community Centres, Police and Fire Stations, Prisons, Courts
Shops, Stalls, Kiosks and Markets	Shops, Kiosks, Showrooms, Stores
Super/Hyper Stores	Superstores, Malls
Sport	Stadia, Sport Centres, Golf Courses, Tennis Centres, Football Grounds
Vacation Sites, Accommodation and Facilities	Campsites, Caravan Sites, Hotels, Guest Houses, Holiday Units, Hostels, Motels, Beach Houses
Petrol Stations	Petrol Stations
Vehicle Infrastructure	Vehicle Repair Workshop, Garages, Car Wash
Warehouse and Storage	Warehouses, Depots, Storage Depots, Land Used for Storage
Parking Space	Car/Vehicle Park Sites and Park Spaces, Motorcycle Bays
Animal Husbandry, Farming and Agriculture	Aviaries, Farms, Animal Shelters, Stud Farms
Marine Infrastructure	Mooring Sites, Quays, Wharfs, Lifeboat Stations, Marine Control Centres
Under (re)construction	Properties and Premises Undergoing (re)Construction

Table C-3: Traffic values by road class and cluster

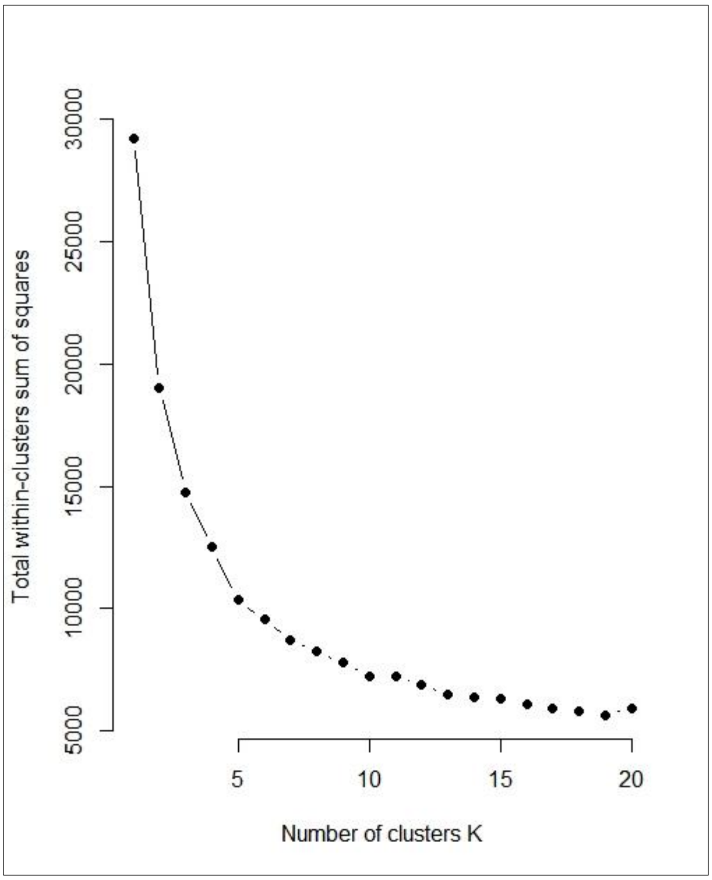
Road Class	Cluster	Number of points	Traffic Sum	Traffic per point	Share
A Roads	1 – red	521	39,224,718	75,287.37	14.2%
	2 – yellow	2,170	76,883,890	35,430.36	27.8%
	3 – blue	1,672	25,533,749	15,271.38	9.2%
	4 – white	5,627	92,841,986	16,499.38	33.6%
	5 – green	4,680	41,880,043	8,948.73	15.2%
	Total	**14,670**	**276,364,386**	**Share of traffic in sample**	**95.5%**
B roads	1 – red	86	1,613,758	18,764.63	22.6%
	2 – yellow	216	2,516,417	11,650.08	35.2%
	3 – blue	194	1,429,658	7,369.37	20.0%
	4 – white	252	1,090,089	4,325.75	15.2%
	5 – green	284	501,520	1,765.92	7.0%
	Total	**1,032**	**7,151,442**	**Share of traffic in sample**	**2.5%**
C roads	1 – red	59	886,950	15,033.05	24.7%
	2 – yellow	207	1,561,951	7,545.66	43.5%
	3 – blue	147	660,926	4,496.10	18.4%
	4 – white	218	192,124	881.30	5.3%
	5 – green	427	290,185	679.59	8.1%
	Total	**1,058**	**3,592,136**	**Share of traffic in sample**	**1.2%**

Road Class	Cluster	Number of points	Traffic Sum	Traffic per point	Share
U Roads	1 – red	31	269,195	8,683.71	12.3%
	2 – yellow	187	594,486	3,179.07	27.2%
	3 – blue	557	739,595	1,327.82	33.9%
	4 – white	1,070	509,433	476.11	23.3%
	5 – green	196	69,499	354.59	3.2%
	Total	**2,041**	**2,182,208**	**Share of traffic in sample**	**0.8%**
ALL ROADS	**Overall Total**	**18,801**	**289,290,172**	**Share of all road traffic**	**100%**

Table C-4: Total Traffic Share by road class

Road Class	Traffic Volume – Vehicle Miles Travelled (VMT) in billions	Share
A	144.9	57%
B	22.9	9%
C	52.5	20%
U	35	14%
Total	**255.3**	**100%**

Figure C-1: "Elbow" method indicating K=5



- Choice of weights

The choice of weights has an impact on clustering formation and therefore on the performance of a given methodology. In the literature, clustering applications are normally used as exploratory analyses (e.g. Baumgartner et al., 2000; Bolton and Krzanowski, 2003; Hébrail et al., 2010), with the purpose to separate the data into "similar" groups of objects (Larose, 2005). Considering this exploratory process and the concept of similarity, my choice of using weights is based on the pioneering work of Hastie et al., (2009), stating that "Variables that are more relevant in separating the groups should be assigned a higher influence in defining object dissimilarity". In addition, Friedman and Meulman (2004) state that optional weights can be assigned to certain variables to raise their importance when those variables are considered a priori more critical in cluster formation than others. During my experiments, I have tested a number of weight combinations so as to identify clusters with similar values of traffic. I eventually noticed that results improved when the clustering process was dominated by the dependent variable which is not entirely surprising as the values of the dependent variable are those over which MAPEs are computed. As an example, setting three different weights for the dependent variable (weights of 2, 5 and 10), returned higher errors for weights 2 and 5 as compared to the application of weight 10. A sample of runs for weights 2 and 5 can be seen in Table C-5 and Table C-6 respectively, where the errors are higher compared with weight 10.

Table C-5: MAPE for AADT weight = 2

Dependent variable weight: 2				
A3200	Cluster	Regression	Random Forest	SVR
	1	44.6%	44.5%	43.93%
	2	37.4%	35.9%	46.8%
	3	70.3%	60.8%	70.1%
	4	54.9%	49.7%	129.7%
	5	43.3%	39.7%	49.8%
A2000	1	68.9%	59.5%	69.2%
	2	44.85%	46.1%	61.3%
	3	62.4%	61.3%	100%
	4	39.4%	37.1%	48.5%
	5	45.2%	41.4%	48.8%

A1600	1	51.7%	42.9%	55.3%
	2	53.7%	55.5%	130.4%
	3	58.6%	55.2%	88.1%
	4	39.6%	37.2%	44.5%
	5	65.1%	55.8%	64.1%
A1000	1	50.2%	48.4%	56%
	2	41.3%	38.9%	48.7%
	3	66.6%	66.5%	130.1%
	4	64.2%	58.7%	90.8%
	5	76.1%	95%	82.8%
A800	1	58.8%	57.1%	120.1%
	2	40.3%	37.6%	45.2%
	3	57.2%	42.2%	65.8%
	4	49.8%	42.8%	48.5%
	5	64.7%	61.1%	91.3%
A500	1	39.7%	37.8%	48.5%
	2	67.5%	63.9%	96.4%
	3	58.6%	52.7%	59.6%
	4	61.5%	61.8%	74.8%
	5	64.8%	61.9%	118.3%

Table C-6: MAPE for AADT weight = 5

Dependent variable weight: 5				
A3200	Cluster	Regression	Random Forest	SVR
	1	41.9%	38.6%	39%
	2	13%	11.6%	11%
	3	45.3%	38.5%	36%
	4	30.1%	27.1%	28.8%
	5	45%	42.6%	43.2%
A2000	1	58.2%	53.2%	52.8%
	2	10.7%	9.3%	9.1%
	3	32%	28.9%	30.7%
	4	40.9%	39.2%	39.4%
	5	69.5%	67.1%	68.5%
A1600	1	40.7%	37.7%	37.5%
	2	29%	27.7%	28.4%
	3	13.1%	12.4%	11.8%
	4	51.7%	46%	47.2%
	5	73.9%	63.8%	63.1%
A1000	1	13%	12.4%	12.1%
	2	46.8%	44.4%	44.5%
	3	61%	55.7%	55.3%
	4	43.2%	41.7%	40.4%
	5	30.5%	30.1%	30.1%

A800	1	55.8%	46.1%	44.8%
	2	51.6%	47.9%	48.1%
	3	30.1%	29.6%	29.7%
	4	11.8%	12.3%	11.3%
	5	37.2%	35.8%	35.5%
A500	1	5.9%	5.8%	5.6%
	2	46.8%	43.6%	44.4%
	3	23.5%	23%	23.2%
	4	7.8%	7.3%	7.5%
	5	81.4%	43.1%	43.4%

It might be that even higher weights for the dependent variable would deliver lower errors although it has not been tested. This relationship between the weights and the returned MAPEs is in my mind an experimental process and of course it is unlikely that the weights are "the best" as they are by no means optimised. I think this needs to be rightly pointed out, as it implies that there might be more gains from my approach in terms of predictive accuracy something which is explored in the following chapters.

D. Supplementary material for chapter 4

Table D-1: Independent Variables

Group	Variable	Abbreviation	Description	Type	Unit
Roadway Characteristics	1. Urban/Rural	Urb_Rur	Whether a count point is located at urban or rural environment	Categorical	N/A – dummy (0-1), where 0 is rural and 1 is urban
	2. Distance to Urban Area	Dist_Urb	Distance of count point to the edge of the nearest edge of spatial polygon indicating urban area	Numerical	Meters
	3. Distance to Major Urban Area[69]	Dist_M_Urb	Distance of count point to the edge of the nearest edge of spatial polygon indicating Major urban area[69]	Numerical	Meters
	4. Distance to Urban Area Centroid	Dist_Urb_C	Distance of count point to the geometrical centroid of the nearest spatial polygon indicating urban area – indicating urban (city/town/village) centre	Numerical	Meters
	5. Distance to Major Urban Area[69] Centroid	Dist_M_U_C	Distance of count point to the geometrical centroid of the nearest spatial polygon indicating major urban area[69] – indicating major urban (city) centre	Numerical	Meters
	6. Toll Road	Toll	Whether a count point is located on a road with tolls	Categorical	N/A – dummy (0-1), where 0 indicates free road and 1 a toll road
	7. Ring Road	Ring_Road	Whether the count point is located on a ring road	Categorical	N/A – dummy (0-1), where 0 is a regular road and 1 is a ring road
	8. Road Nature	Nature	Whether the count point is located on a single or dual carriageway, slip road or roundabout	Categorical	N/A – categorical with multiple categories, depending on road class
	9. Road Category	RCat	Whether the point is located on a primary or trunk road (mainly for higher class roads) in an urban or rural area	Categorical	N/A – categorical with multiple – usually 4 – categories.
	10. Junction Accessibility	Junction	Whether the point is located on a road with access to motorway within the specified service area	Categorical	N/A – dummy (0-1), where 0 indicates no access and 1 indicates access to motorway
Public Transport	11. Bus Stops	Bus_stops	Number of bus stops within the specified service area	Numerical	count
	12. Bus Stations	Bus_stat	Number of bus stations within the specified service area	Numerical	Count
	13. Train Accessibility	Perc_SA	Indicating train station accessibility within the LSOA[70] where the count point is located.	Numerical	Decimal

[69] These are the six largest urban agglomerations in England and Wales defined by Pointer, (2005). The urban agglomerations are: Greater London, West Midlands (Birmingham, Wolverhampton, Coventry), Greater Manchester, West Yorkshire (Leeds and Bradford), Tyneside (Newcastle and Sunderland) and Liverpool Urban Areas.
[70] As a remined, the Lower Super Output Areas (LSOAs) are approximately 35,000 areas designed by the Office for National Statistics (ONS) for England and Wales, with population minimum of 1000.

Group	Variable	Abbreviation	Description	Type	Unit
Socioeconomic	14. Population	Population	Total population of a count point's intersecting LSOAs[70]	Numerical	Count
	15. Population Density	pd	Average population density of a count point's adjacent LSOAs[70]	Numerical	Decimal (people per hectare)
	16. Workplace Population	workpop	Total number of registered employed people of a count point's adjacent LSOAs[70]	Numerical	Count
	17. Workplace Population Density	work_dens	Average of registered employers' density around a count points' adjacent LSOAs[70]	Numerical	Decimal (people per hectare)
	18. Workplace plus Population Density	w_n_p_d	Average workplace plus population density of a count point's adjacent LSOAs[70]	Numerical	Decimal (people per hectare)
	19. Income	Income	Average median income of a count point's adjacent LSOAs70	Numerical	British Pound Sterling (in thousands)
	20. Households	Household	Total number of households of a count point's adjacent LSOAs[70]	Numerical	Count
	21. Registered Vehicles	car_van	Total number of registered cars and vans of a count point's adjacent LSOAs[70]	Numerical	Count
Land Use	22. Charging Points	Charge_p	Number of charging points within each service area	Numerical	Counts
	23. Ports	Port	Whether there is a port within the specified service area around the count point	Categorical	N/A – dummy (0-1), where 0 indicates no port and 1 indicates port occurrence
	24. Airports	Airport	Whether there is an airport within the specified service area around the count point	Categorical	N/A – dummy (0-1), where 0 indicates no airport and 1 indicates airport occurrence
	25. Research, Education and Training	Research	Total number of Schools, Colleges, Libraries, Universities, Language and Music Schools, etc. within the specified service area	Numerical	Count
	26. Factories, Workshops, and Industrial Activity	Factories	Total number of Energy Production Facilities, Factories, Workshops, Mines, Oil Fields, Recycling Plants, Shipyards, Scrap Yards within the specified service area	Numerical	Count
	27. Healthcare	Healthcare	Number of Hospitals, GPs, Surgeries, Clinics within the specified service area	Numerical	Count
	28. Leisure	Leisure	Number of Public Houses, Bars, Nightclubs, Restaurants, Art Galleries, Cinemas and Theatres, Coffee shops within the specified service area	Numerical	Count
	29. Office and Business Space	Office	Number of Offices, Banks, Business Units within the specified service area	Numerical	Count
	30. Public Services, Infrastructure and Buildings	Public	Number of Post Offices, Community Centres, Police and Fire Stations, Prisons, Courts within the specified service area	Numerical	Count
	31. Shops, Stalls, Kiosks and Markets	Shops	Number of Shops, Kiosks, Showrooms, Stores within the specified service area	Numerical	Count
	32. Super/Hyper Stores	Superstore	Number of Superstores, Malls within the specified service area	Numerical	Count

Group	Variable	Abbreviation	Description	Type	Unit
	33. Sport	Sport	Number of Stadia, Sport Centres, Golf Courses, Tennis Centres, Football Grounds within the specified service area	Numerical	Count
	34. Vacation Sites, Accommodation and Facilities	Vacation	Number of Campsites, Caravan Sites, Hotels, Guest Houses, Holiday Units, Hostels, Motels, Beach Houses within the specified service area	Numerical	Count
	35. Petrol Stations	Petrol	Number of Petrol Stations within the specified service area	Numerical	Count
	36. Vehicle Infrastructure	Vehicle	Number of Vehicle Repair Workshop, Garages, Car Wash within the specified service area	Numerical	Count
	37. Warehouse and Storage	Warehouses	Number of Warehouses, Depots, Storage Depots, Land Used for Storage within the specified service area	Numerical	Count
	38. Parking Space	parking	Number of Car/Vehicle Park Sites and Park Spaces, Motorcycle Bays within the specified service area	Numerical	Count
	39. Animal Husbandry, Farming and Agriculture	Animals	Number of Aviaries, Farms, Animal Shelters, Stud Farms within the specified service area	Numerical	Count
	40. Marine Infrastructure	Marine	Number of Mooring Sites, Quays, Wharfs, Lifeboat Stations, Marine Control Centres within the specified service area	Numerical	Count
	41. Under (re)construction	Under_cons	Number of Properties and Premises Undergoing (re)Construction within the specified service area	Numerical	Count

Table D-2: Base Categories for Categorical Variables

Categorical Variable	Abbreviation	Base Category			
		Road Class			
		A	B	C	U
Road Nature	Nature	Single Carriageway			
Road Category	RCat	Primary Rural (PR)	B Rural (BR)	C Rural (CR)	U Rural (UR)
Presence of Motorway Junction	Junction	Junction = 0 – No Junction is present			
Toll Road	Toll	Toll = 0 – No tolls on this road			
Presence of Port	Port	Port = 0 – No Ports are present			
Presence of Airport	Airport	Airport = 0 – No airports are present			
Ring Road	Ring_Road	Ring_Road = 0 – The road is not a ring road			

Table D-3: 'A' Roads coefficients for Cars, Buses and Two-wheeled vehicles

Vehicle Type	**Cars**					**Buses**					**Two-wheeled**				
Group	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
optimal lambda	0.00337	0.00152	0.00103	0.00045	0.00003	0.00195	0.00362	0.00944	0.00045	0.00233	0.00402	0.00091	0.0005	0.00004	0.00093
Variables															
(Intercept)	56153.08	27532.37	10051.43	16634.74	6260.01	206.9	154.2	443.0	140.5	46.5	619.2	254.4	223.7	152.4	62.0
Animals	-1.0%	-0.3%	-0.8%	0.2%	0.8%	0.1%	.	-1.6%	0.7%	-0.2%	0.4%	2.3%	-1.0%	1.6%	0.7%
Bus_stat	-1.3%	0.6%	0.1%	-0.1%	-2.3%	-4.5%	0.7%	2.2%	3.8%	-0.2%	-2.3%	-0.7%	-2.1%	-1.6%	-3.3%
Charge_p	0.3%	.	3.3%	0.9%	1.4%	6.6%	1.9%	-1.9%	3.5%	2.8%	-0.1%	0.0%	-2.2%	-1.0%	1.3%
Factories	-1.1%	.	2.2%	-1.2%	-2.0%	-3.2%	-13.9%	-5.3%	-2.3%	-3.7%	-4.0%	.	-0.8%	-2.7%	0.5%
Healthcare	.	-1.0%	-0.5%	0.4%	0.6%	3.9%	2.9%	-2.1%	4.3%	0.6%	0.9%	0.8%	-2.1%	-1.5%	-2.2%
Leisure	.	2.3%	-1.6%	-4.4%	-4.2%	-9.5%	.	0.3%	-6.2%	.	.	1.5%	-0.8%	-0.6%	-4.5%
Marine	0.0%	-0.1%	-2.5%	-0.8%	-1.8%	.	-0.1%	-4.3%	-2.0%	-3.0%	-0.1%	-1.1%	-4.0%	-0.4%	0.2%
Offices	.	-0.6%	.	-0.9%	-1.0%	.	0.4%	-8.2%	-7.1%	.	12.2%	0.0%	12.3%	-4.2%	-1.6%
Parking	.	.	-4.0%	1.0%	0.0%	-9.3%	-4.4%	-10.3%	0.3%	1.0%	-3.5%	-5.2%	-13.4%	1.5%	.
Petrol	.	0.8%	5.2%	2.0%	1.7%	4.4%	5.7%	4.6%	3.2%	3.4%	.	2.2%	5.4%	5.7%	3.4%
Public	.	-0.7%	1.6%	-0.5%	-3.6%	-3.4%	7.7%	7.4%	0.3%	-0.4%	1.4%	3.5%	6.9%	-1.1%	-2.1%
Research	.	.	0.6%	.	1.9%	-7.0%	.	-0.3%	4.5%	2.3%	.	-3.4%	-0.3%	1.2%	2.3%
Shops	.	.	-12.0%	-0.7%	1.9%	.	.	12.0%	9.7%	-4.6%	1.1%	-2.2%	-19.7%	0.0%	5.1%
Sport	0.5%	-0.6%	2.9%	0.9%	0.4%	4.8%	2.0%	2.3%	1.5%	0.9%	3.4%	3.7%	5.2%	2.2%	0.8%
Superstore	.	-0.2%	1.3%	1.4%	0.3%	-0.6%	5.3%	.	-1.0%	-0.8%	.	.	6.8%	1.2%	-0.6%
Under_cons	.	.	3.6%	0.2%	3.1%	-1.4%	.	-2.1%	0.4%	.	-5.9%	.	2.3%	0.3%	2.5%
Vacation	-0.9%	.	0.3%	-0.7%	1.3%	10.5%	1.3%	-0.2%	2.0%	4.2%	.	0.4%	-4.3%	2.0%	1.7%
Vehicle	.	.	1.8%	0.5%	1.0%	13.9%	-3.9%	-10.2%	1.5%	-2.1%	.	-8.3%	-4.3%	-1.6%	0.3%
Warehouses	.	-2.0%	-0.3%	1.6%	-0.4%	-3.2%	-1.2%	.	-2.1%	.	0.2%	1.7%	7.4%	3.2%	-3.2%
Bus_stops	-2.0%	.	-2.8%	-2.7%	5.6%	8.0%	11.0%	29.5%	22.5%	32.9%	-6.7%	-3.2%	-8.6%	-2.7%	6.8%
Population	.	.	16.9%	.	8.7%	46.1%	60.1%	16.1%	19.3%	22.2%	.	-16.2%	11.6%	-6.3%	3.9%
Income	0.2%	2.2%	4.8%	3.5%	12.7%	11.6%	2.4%	10.8%	7.5%	.	22.1%	18.9%	81.8%	19.7%	11.1%

Vehicle Type	Cars					Buses					Two-wheeled				
Group	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
optimal lambda	0.00337	0.00152	0.00103	0.00045	0.00003	0.00195	0.00362	0.00944	0.00045	0.00233	0.00402	0.00091	0.0005	0.00004	0.00093
Variables															
Households	.	.	-16.3%	-0.3%	7.2%	-11.8%	.	.	7.1%	.	.	23.6%	18.3%	14.8%	3.8%
Perc_Sa	-1.0%	-0.3%	5.9%	0.3%	-2.3%	1.4%	.	13.5%	0.7%	-3.2%	5.6%	8.4%	13.5%	1.0%	-3.6%
car_van	-0.1%	.	12.4%	-0.4%	-10.2%	-15.3%	-17.0%	-2.0%	-20.8%	-16.5%	-6.0%	-0.5%	-20.5%	-9.4%	-6.4%
workpop	.	.	4.9%	1.8%	3.7%	.	.	10.3%	2.4%	4.4%	1.4%	.	10.7%	4.5%	4.3%
work_dens	-0.4%	.	4.2%	.	-3.4%	5.2%	0.3%	.	0.2%	-1.0%	.	.	5.0%	-0.7%	-1.8%
w_n_p_d	.	5.1%	0.8%	2.6%	2.9%	10.1%	.	.	1.5%	.	4.6%	15.7%	7.3%	3.9%	.
pd	3.6%	.	.	1.0%	2.6%	5.2%	-2.1%	17.7%	0.5%	.	37.5%	17.1%	33.7%	4.4%	3.5%
Dist_Urb	-1.5%	-0.7%	0.8%	.	-12.9%	0.1%	.	-1.0%	0.0%	-11.8%	-6.1%	-4.7%	-1.2%	-3.1%	-6.1%
Dist_M_Urb	-2.8%	-1.4%	1.8%	0.0%	-42.5%	44.0%	32.0%	.	30.0%	.	-28.9%	-17.4%	6.5%	-20.5%	-45.4%
Dist_Urb_C	6.8%	.	2.8%	1.7%	3.4%	3.9%	10.8%	9.9%	9.0%	-0.9%	35.2%	14.3%	22.6%	13.0%	1.7%
Dist_M_U_C	.	.	-2.2%	0.4%	66.8%	-24.6%	-18.9%	-5.7%	-19.4%	5.0%	42.5%	38.2%	-2.2%	48.5%	83.8%
RCatPU	.	-1.7%	.	-17.6%	-30.4%	65.0%	11.3%		.	-20.0%	20.7%	-7.2%		-28.5%	-34.9%
RCatTR	.	4.4%	.	-7.6%	10.1%	9.5%	.		-13.2%	5.0%	-10.1%	-12.1%		-22.6%	1.2%
RCatTU	3.5%	9.0%	.	-7.7%	12.0%	19.0%	.	.	2.0%	.	-9.6%	6.6%	.	-19.2%	.
Toll1	.	.	-26.7%	.	33.4%	.	.	.	.	.	.	.	94.8%	119.8%	.
Urb_Rur1	.	.	.	-14.4%	-25.5%	6.5%	.	.	-4.3%	-33.2%	.	-2.3%	.	-20.4%	-22.1%
Port1	.	.	-48.7%	.	-35.7%	.	.	.	0.5%	.	.	.	-72.5%	.	.
Junction1	.	.	.	.	29.6%	-2.9%	.	.	.	11.0%	-9.6%	-2.7%	30.8%	-5.1%	11.3%
Ring_Road1	2.6%	4.1%	19.5%	14.5%	2.5%	-11.8%	-15.7%	-21.7%	-29.7%	.	-11.4%	2.6%	13.0%	0.2%	.
Nature: Dual Carriageway	.	.	39.1%	16.9%	48.9%	32.2%	.	16.1%	6.2%	27.8%	.	.	32.4%	21.6%	24.8%
Nature: Roundabout	.	.	0.0%	.	-0.9%	.	.	.	38.5%	.	.	.	-31.5%	30.8%	.
Nature: Slip Road	.	-7.2%	.	-11.5%	-52.1%	30.9%	21.1%	.	-19.6%	-58.4%	8.0%	.	-16.7%	-4.3%	-57.6%

Table D-4: 'A' Roads coefficients for LGVs and HGVs

Vehicle Type	**LGVs**					**HGVs**				
Group	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
optimal lambda	0.001300	0.000759	0.003736	0.000261	0.000442	0.003262	0.001162	0.004827	0.001312	0.000047
Variables										
(Intercept)	10040.5	4561.5	1658.2	2763.1	1222.3	3237.4	1242.0	292.4	570.8	275.1
Animals	-0.3%	0.1%	-1.6%	0.6%	0.8%	-0.5%	-2.8%	-1.8%	-0.7%	-0.7%
Bus_stat	-1.9%	-1.3%	-0.3%	-1.0%	-2.2%	-0.8%	-2.0%	-1.7%	-1.1%	0.5%
Charge_p	0.8%	0.4%	2.7%	1.0%	0.7%	0.2%	-2.0%	.	-0.4%	0.9%
Factories	-2.1%	1.4%	4.2%	0.9%	0.5%	-5.1%	-0.9%	1.6%	0.0%	2.8%
Healthcare	-0.5%	.	-1.0%	-0.3%	-0.3%	.	0.2%	.	-0.3%	-1.2%
Leisure	1.9%	0.7%	.	-1.8%	-3.2%	.	0.6%	0.9%	0.4%	-3.6%
Marine	-1.3%	-0.6%	-2.9%	0.0%	-2.3%	-3.8%	1.0%	-4.9%	.	-0.6%
Offices	2.4%	-0.3%	.	0.0%	-2.6%	4.5%	5.9%	.	-0.2%	-1.8%
Parking	.	-3.5%	-5.4%	0.4%	0.5%	0.0%	-8.4%	-8.2%	0.4%	1.7%
Petrol	-1.1%	1.3%	4.8%	1.7%	1.6%	.	1.9%	5.7%	2.9%	2.6%
Public	-0.8%	-1.9%	-0.5%	-0.9%	-3.0%	1.0%	-4.1%	0.8%	-2.4%	-4.4%
Research	0.0%	.	0.2%	-0.8%	1.2%	.	2.3%	-1.3%	-2.0%	-0.8%
Shops	.	-3.2%	-12.3%	-2.7%	3.6%	5.7%	-16.3%	-13.7%	-4.4%	2.7%
Sport	.	0.0%	1.4%	0.2%	-0.2%	-5.0%	-2.7%	3.5%	.	-1.2%
Superstore	-1.3%	-2.1%	-0.4%	0.0%	-0.7%	-2.5%	-1.4%	-0.1%	.	-2.2%
Under_cons	0.2%	3.1%	2.8%	.	3.3%	0.2%	7.3%	1.9%	-1.1%	2.2%
Vacation	-3.9%	-0.3%	-1.6%	-0.4%	-0.9%	-3.5%	-2.9%	-3.6%	-0.7%	-6.0%
Vehicle	-0.6%	-0.2%	0.4%	1.0%	1.5%	-2.2%	-0.1%	-0.2%	0.3%	0.7%
Warehouses	0.7%	4.1%	3.1%	2.4%	1.1%	4.6%	14.0%	10.3%	5.7%	4.9%
Bus_stops	-6.1%	-4.0%	-1.6%	-3.0%	2.6%	-17.0%	-9.7%	-1.1%	-9.0%	-10.4%
Population	.	-7.0%	2.4%	0.2%	9.6%	.	.	12.1%	.	5.1%
Income	1.2%	2.0%	4.6%	0.2%	6.8%	-4.9%	5.6%	18.0%	-4.0%	-3.9%

Vehicle Type	LGVs					HGVs				
Group	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
optimal lambda	0.001300	0.000759	0.003736	0.000261	0.000442	0.003262	0.001162	0.004827	0.001312	0.000047
Variables										
Households	3.8%	13.7%	0.0%	2.6%	0.8%	1.4%	19.4%	.	-3.6%	-19.4%
Perc_Sa	0.9%	.	6.1%	0.1%	-2.2%	.	-2.3%	8.9%	-1.3%	-4.9%
car_van	-3.8%	-4.6%	3.5%	-3.0%	-5.5%	-0.8%	-18.4%	-7.0%	-0.7%	17.3%
workpop	-1.4%	.	4.6%	1.2%	1.7%	8.9%	21.1%	9.6%	6.8%	10.4%
work_dens	-2.9%	3.7%	.	-0.1%	-0.5%	-12.5%	-11.0%	.	.	.
w_n_p_d	.	.	.	-1.8%	-0.3%	.	.	0.8%	-7.3%	-1.4%
pd	8.0%	.	4.9%	-0.7%	.	16.0%	.	6.0%	-2.7%	-1.4%
Dist_Urb	-2.2%	-1.1%	0.1%	0.3%	-10.4%	5.7%	2.1%	-0.9%	2.8%	-6.0%
Dist_M_Urb	-20.1%	-0.1%	.	-5.6%	-46.9%	-26.7%	.	3.0%	-3.6%	-53.1%
Dist_Urb_C	12.9%	1.9%	0.2%	2.2%	2.2%	14.1%	3.6%	11.2%	2.8%	-2.2%
Dist_M_U_C	19.3%	.	-1.0%	8.7%	85.9%	23.7%	-6.6%	-5.2%	.	94.7%
RCatPU	-1.0%	-1.1%	.	-18.9%	-27.0%	-14.2%	-20.4%	.	-37.5%	-31.2%
RCatTR	9.3%	22.2%	.	2.9%	14.4%	57.7%	111.8%	.	55.4%	67.7%
RCatTU	12.2%	23.9%	.	-1.8%	11.0%	35.3%	59.4%	.	7.0%	37.8%
Toll1	.	.	.	.	.	.	.	.	.	96.0%
Urb_Rur1	-0.5%	.	.	-15.3%	-21.6%	-6.8%	.	.	-19.0%	-19.9%
Port1	.	.	.	-4.1%	-3.1%	.	10.3%	.	.	55.6%
Junction1	-0.3%	2.3%	.	6.7%	27.0%	.	1.4%	.	10.6%	49.9%
Ring_Road1	7.8%	6.4%	20.6%	16.8%	.	36.9%	10.3%	32.1%	22.2%	6.0%
Nature: Dual Carriageway	.	.	32.5%	18.1%	43.6%	.	.	40.4%	30.8%	76.0%
Nature: Slip Road	8.9%	-8.1%	.	-15.8%	-51.9%	.	-20.2%	.	.	-59.4%
Nature: Roundabout	.	.	.	.	.	.	.	.	.	-14.7%

Table D-5: 'B' Roads Coefficients for Cars, Buses and Two-wheeled vehicles

Vehicle Type			**Cars**					**Buses**					**Two-wheeled**		
Group	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
optimal lambda	0.01133	0.01075	0.0037	0.01728	0.05504	0.0795	0.03645	0.13435	0.02457	0.11736	0.04305	0.01842	0.02467	0.02508	0.0405
<u>Variables</u>															
(Intercept)	15542.6	9677.0	6296.4	3465.8	1224.3	163.0	114.5	35.3	20.2	6.7	143.6	84.9	56.5	33.9	14.0
Animals	.	.	1.2%	.	.	-8.4%	.	.	-2.9%	.	-2.1%	.	.	-4.8%	3.3%
Bus_statio	0.6%	.	0.0%	.	.	.	0.6%	.	.	.	.	1.0%	.	-1.6%	8.4%
Charge_p	.	-0.1%	-1.3%	.	-0.1%	.	12.3%	.	-13.7%	-1.0%	.	.	.	.	.
Factories	.	.	.	.	.	.	.	.	-2.7%	.	-9.6%	.	7.9%	12.2%	.
Healthcare	1.8%	.	.	.	.	.	12.1%	.	.	.	.	-2.6%	.	.	.
Leisure	.	-1.1%	.	.	.	12.7%	.	.	.	.	.	.	.	.	.
Marine		.	.	.	.	.	1.1%	.	3.0%	.	.	3.3%	.	4.0%	.
Office	.	.	.	.	.	.	3.0%	-1.9%	.	.	.	.	.	.	.
Parking	.	.	.	.	-2.2%	-0.2%	.	-13.1%	.	.	0.9%	-6.9%	.	.	.
Petrol	0.1%	0.9%	.	.	.	7.3%	.	.	-7.0%	.	.	-2.8%	.	.	-0.5%
Public	.	-0.2%	-0.3%	.	.	.	.	.	.	.	-1.3%	0.9%	.	-1.3%	.
Research	-1.7%	.	-0.2%	-0.4%	.	.	.	.	5.6%	.	.	.	.	12.1%	.
Shops	.	.	-1.4%	.	.	.	3.3%	.	.	.	-2.4%	.	0.1%	.	.
Sport	0.7%	.	2.0%	.	.	7.6%	3.0%	.	.	.	3.5%	.	.	.	.
Superstore	-0.4%	.	.	.	.	7.8%	-6.8%	.	-4.3%	.	-1.0%	0.6%	.	-1.5%	.
Under_cons	.	.	0.0%	.	.	.	.	.	.	.	.	.	.	.	.
Vacation	0.3%	.	.	.	.	.	.	.	4.1%	.	.	.	-1.6%	.	.
Vehicle	-1.2%	.	.	.	.	.	.	.	-2.6%	.	.	.	.	-10.2%	.
Warehouses	.	.	.	.	.	.	-0.8%	.	-4.3%	.	.	.	.	.	2.4%
Bus_stops	0.2%	-1.2%	.	.	.	22.1%	.	24.0%	23.0%	.	-4.4%	-13.4%	-14.0%	.	.
Population	.	.	.	.	.	.	8.7%	.	16.0%	.	.	.	.	.	.
Income	2.0%	.	-1.1%	.	7.7%	.	-5.7%	-15.9%	-17.8%	-2.8%	20.5%	11.3%	2.9%	6.8%	25.1%

Vehicle Type	Cars					Buses					Two-wheeled				
Group	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
optimal lambda	0.01133	0.01075	0.0037	0.01728	0.05504	0.0795	0.03645	0.13435	0.02457	0.11736	0.04305	0.01842	0.02467	0.02508	0.0405
Variables															
Households	-1.5%	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Perc_Sa	.	.	-3.1%	-1.6%	.	.	8.8%	.	-20.0%	.	.	5.5%	.	3.7%	.
car_van	.	.	.	.	.	1.9%	.	.	.	.	.	-2.8%	-14.0%	-2.5%	.
workpop	4.9%	.	.	.	.	.	.	.	.	-0.3%	.	.	.	3.3%	.
work_dens	.	.	.	-6.6%	.	.	.	.	.	.	.	.	.	5.5%	.
w_n_p_d	.	.	.	.	.	.	.	.	.	.	.	37.9%	16.5%	.	.
pd	.	.	-1.2%	.	.	.	29.3%	2.4%	.	.	11.8%	11.4%	12.2%	.	.
Dist_Urb	-2.0%	.	-2.2%	.	.	-3.2%	.	-2.6%	-13.3%	-6.4%	.	2.6%	.	12.3%	.
Dist_M_Urb	-1.0%	.	.	.	.	.	.	.	.	.	.	1.8%	2.0%	.	.
Dist_Urb_C	-3.4%	.	0.6%	.	.	32.8%	11.7%	.	.	.	.	0.1%	.	.	4.5%
Dist_M_U_C	.	.	.	.	.	-0.9%	.	.	9.7%	3.5%	14.3%	14.9%	.	1.7%	.
RCatBU	.	.	-1.6%	.	.	.	.	.	.	.	.	.	-2.9%	.	.
Toll1		.	.	.	.	.	.	.	.	.	.	.	.	.	.
Urb_Rur1	.	.	-6.8%	.	.	.	.	.	.	.	.	.	.	.	.
Port1						.	.	.	.	.	.	.	.	.	.
Junction1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Nature: Dual Carriageway	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

Table D-6: 'B' Roads coefficients for LGVs and HGVs

Vehicle Type	**LGVs**					**HGVs**				
Group	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
optimal lambda	0.024127	0.021667	0.016022	0.010409	0.027577	0.116876	0.010734	0.023223	0.026079	0.061834
Variables										
(Intercept)	2181.1	1352.4	941.3	579.9	239.1	252.0	191.1	129.0	75.4	39.3
Animals	0.4%	.	.	-0.1%	1.8%	.	8.2%	-2.2%	.	.
Bus_statio	.	.	0.3%	-3.1%	.	.	.	.	-3.7%	.
Charge_p	.	-1.5%	.	.	.	.	.	.	.	.
Factories	.	3.7%	3.3%	2.0%	.	.	4.5%	3.1%	1.4%	.
Healthcare	.	-0.5%	.	.	.	.	.	5.8%	.	.
Leisure	.	.	.	.	.	.	-0.9%	.	.	.
Marine	.	.	.	0.1%	.	.	-1.0%	-2.0%	.	.
Office	.	.	.	.	.	.	.	.	.	.
Parking	1.2%	.	.	.	-5.1%	.	.	.	.	.
Petrol	.	.	.	.	.	.	-2.9%	-7.1%	.	.
Public	.	.	.	-2.7%	.	.	4.3%	.	-1.3%	.
Research	-4.1%	.	.	.	.	.	-1.4%	.	1.9%	.
Shops	.	.	.	.	.	.	-5.3%	.	9.1%	.
Sport	.	.	.	.	.	.	0.1%	-0.8%	.	.
Superstore	.	.	-2.8%	.	-5.6%	.	-6.6%	-0.6%	.	-0.2%
Under_cons	.	.	.	.	.	.	2.3%	.	.	.
Vacation	-1.0%	.	.	.	.	.	-1.0%	.	.	.
Vehicle	.	.	.	.	.	.	.	.	.	.
Warehouses	3.2%	2.0%	1.4%	6.6%	1.7%	.	15.4%	10.5%	22.2%	.
Bus_stops	.	-2.0%	-0.9%	-1.4%	.	.	-9.6%	.	-20.1%	-13.5%
Population	.	.	.	.	.	.	3.1%	.	.	.
Income	.	.	0.7%	1.2%	6.2%	.	.	.	.	5.5%

Vehicle Type	LGVs					HGVs				
Group	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
optimal lambda	0.024127	0.021667	0.016022	0.010409	0.027577	0.116876	0.010734	0.023223	0.026079	0.061834
Variables										
Households	.	.	.	.	.	-0.8%	1.8%	-14.7%	.	.
Perc_Sa	.	.	.	.	.	.	0.7%	.	.	.
car_van	.	-0.5%	-2.6%	.	.	-6.2%	.	-0.5%	-4.1%	.
workpop	.	0.5%	.	.	.	.	10.4%	.	.	.
work_dens	.	.	.	.	.	.	.	.	.	.
w_n_p_d	.	.	.	.	.	.	.	.	.	.
pd	.	.	-2.4%	.	.	.	.	-4.7%	.	.
Dist_Urb	.	.	.	3.3%	.	.	7.7%	5.1%	24.1%	.
Dist_M_Urb	.	.	.	.	.	.	.	.	2.7%	.
Dist_Urb_C	-1.6%	.	-3.1%	-3.6%	.	.	-3.6%	-1.8%	-13.4%	.
Dist_M_U_C	.	3.4%	0.6%	7.7%	2.0%	.	.	-3.3%	.	.
RCatBU	.	.	-6.1%	.	.	.	-25.9%	-20.9%	.	.
Toll1	.	.	.	.	.	.	.	.	.	.
Urb_Rur1	.	.	.	.	.	.	.	.	.	.
Port1	.	.	.	.	.	.	.	.	.	.
Junction1	.	.	.	.	.	.	39.6%	.	.	.
Nature: Dual Carriageway	.	.	.	.	.	.	.	.	.	.

Table D-7: 'C' Roads coefficients for Cars, Buses and Two-wheeled vehicles

Vehicle Type	**Cars**					**Buses**					**Two-wheeled**				
Group	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
optimal lambda	0.01287	0.00499	0.01306	0.04309	0.09038	0.17908	0.03766	0.08565	0.09927	0.04089	0.05294	0.01428	0.03763	0.17618	0.08544
<u>Variables</u>															
(Intercept)	12086.8	6252.4	3767.7	480.0	308.3	109.6	53.5	26.5	2.8	1.8	100.2	54.0	31.6	3.8	3.7
Animals	.	.	.	7.9%	.	.	.	-7.3%	.	.	0.6%	.	.	.	.
Bus_statio	2.5%	.	.	.	.	.	.	12.0%			.	.	.	.	.
Charge_p	-1.2%	2.7%	0.9%	.	.	.	-12.4%	1.9%	5.1%	.	.	.	.	.	.
Factories	-2.6%	.	-0.2%	8.0%	.	.	.	.	.	.	.	-0.5%	.	.	.
Healthcare	.	1.2%	.	.	0.6%	.	9.7%	.	16.1%	.	.	.	.	.	.
Leisure	.	-0.7%	.	.	.	.	.	11.5%	.	.	.	.	.	.	.
Marine	.		.	.	.	.	.	.	.	.	.	.	.	.	.
Office	.	-2.1%	-6.7%	.	.	.	0.1%	2.3%	.	.	.	7.9%	.	.	.
Parking	4.7%	.	0.1%	.	.	5.2%	.	-19.6%	4.7%	.	.	-3.5%	.	.	.
Petrol	-4.7%	0.9%	.	0.6%	.	.	5.6%	.	.	.	.	.	.	.	.
Public	.	-6.6%	.	.	.	2.3%	-9.4%	11.8%	.	-0.5%	-2.5%	4.2%	.	.	-0.9%
Research	.	3.2%	.	4.5%	.	.	7.3%	.	5.3%	.	1.0%	0.0%	.	.	.
Shops	-1.9%	.	.	.	.	.	.	.	.	.	.	-17.4%	.	.	.
Sport	.	1.5%	.	9.3%	.	.	.	.	.	.	.	2.5%	-1.5%	.	.
Superstore	.	0.4%	-1.4%	.	.	.	-1.1%	-1.7%	.	.	.	.	.	.	.
Under_cons	.	.	.	.	.	.	1.7%	.	.	.	.	.	2.4%	.	.
Vacation	.	2.3%	.	.	.	.	.	5.6%	.	0.2%	2.9%	2.7%	.	.	0.8%
Vehicle	.	2.3%	.	4.9%	.	.	-10.8%	-21.9%	-12.8%	3.0%	-3.6%	.	.	.	.
Warehouses	-1.9%	0.3%	.	0.9%	.	.	.	.	.	.	.	5.2%	.	.	.
Bus_stops	-0.2%	.	.	.	.	6.7%	32.0%	36.9%	2.3%	.	.	-3.2%	-15.7%	.	.
Population	.	.	-2.9%	12.3%	21.5%	.	.	.	.	8.5%	.	.	.	.	1.5%
Income	0.5%	1.0%	.	12.8%	.	.	-3.2%	.	.	.	.	9.6%	4.6%	.	.

Vehicle Type	Cars					Buses					Two-wheeled				
Group	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
optimal lambda	0.01287	0.00499	0.01306	0.04309	0.09038	0.17908	0.03766	0.08565	0.09927	0.04089	0.05294	0.01428	0.03763	0.17618	0.08544
Variables															
Households	.	.	.	.	8.7%	.	-28.1%	.	.	-1.9%	.	5.9%	.	.	.
Perc_Sa	0.6%	.	-0.2%	.	.	.	.	.	9.9%	.	.	18.7%	.	.	.
car_van	.	0.9%	.	1.8%	.	.	.	-4.3%	.	.	.	-8.7%	.	.	.
workpop	2.6%	.	.	.	.	.	.	.	-0.2%	-9.8%	.	.	.	.	.
work_dens	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
w_n_p_d	.	.	-0.5%	.	.	.	2.4%	.	.	-0.8%	.	13.7%	29.3%	.	.
pd	-1.9%	.	.	.	.	6.9%	39.2%	.	8.4%	-7.1%	1.7%	7.5%	20.3%	.	.
Dist_Urb	-2.8%	-1.7%	.	-23.6%	.	.	-9.3%	-22.4%	-7.0%	.	10.1%	6.1%	3.7%	.	.
Dist_M_Urb	.	.	.	-9.6%	.	.	.	.	.	-7.3%	6.7%	.	.	.	.
Dist_Urb_C	-0.6%	-1.6%	.	.	-13.2%	13.4%	22.6%	.	-13.1%	.	.	5.7%	.	.	-5.9%
Dist_M_U_C	.	.	.	.	.	.	23.4%	.	.	.	.	18.1%	8.2%	.	.
RCatCU	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Urb_Rur1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Junction1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Nature: Dual Carriageway	19.3%	1.8%	.	.		.	67.0%	.	.	.	.	22.2%	.	.	.

Table D-8: 'C' Roads coefficients for LGVs and HGVs

Vehicle Type	**LGVs**					**HGVs**				
Group	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
optimal lambda	0.037421	0.034117	0.017805	0.034064	0.071351	0.269453	0.054199	0.096501	0.102070	0.091427
<u>Variables</u>										
(Intercept)	1610.4	815.4	489.2	92.4	64.9	167.8	73.8	36.7	9.2	8.7
Animals	.	.	3.2%	4.5%	.	.	.	.	.	.
Bus_statio	.	.	-5.6%	.		.	-1.3%	.	.	
Charge_p	-0.1%	.	-1.4%	.	.	.	.	.	.	.
Factories	1.6%	.	1.7%	12.6%	.	.	.	14.9%	4.0%	.
Healthcare	.	.	.	.	5.3%	.	.	.	.	.
Leisure	.	.	.	.	.	.	.	.	-0.7%	.
Marine	.	.	0.9%	.	.	.	.	1.7%	.	.
Office	.	.	.	.	.	.	.	.	.	.
Parking	.	.	.	.	.	.	.	.	.	.
Petrol	.	.	.	.	.	.	.	.	0.4%	.
Public	.	.	.	.	.	.	.	.	.	.
Research	.	.	-0.5%	.	.	.	.	.	.	.
Shops	-4.2%	.	-3.4%	.	.	.	.	.	.	.
Sport	.	.	.	9.3%	.	.	.	.	.	.
Superstore	.	.	.	-2.0%	.	.	.	.	.	.
Under_cons	.	.	.	.	.	.	.	.	.	.
Vacation	.	.	0.3%	.	.	.	.	.	.	.
Vehicle	.	.	.	2.4%	.	.	.	.	.	.
Warehouses	9.6%	6.3%	4.1%	4.1%	1.4%	.	25.8%	4.8%	9.1%	.
Bus_stops	.	.	.	.	.	.	.	.	.	.
Population	.	.	.	5.5%	12.2%	.	.	.	.	.
Income	-1.5%	.	.	6.4%	.	.	.	.	.	.

Vehicle Type	LGVs					HGVs				
Group	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
optimal lambda	0.037421	0.034117	0.017805	0.034064	0.071351	0.269453	0.054199	0.096501	0.102070	0.091427
Variables										
Households	.	.	.	.	8.0%	.	.	.	.	.
Perc_Sa	.	0.4%	-0.9%	.	.	.	7.9%	.	.	.
car_van	-4.3%	.	-0.8%	3.5%	.	.	-17.5%	-7.2%	.	.
workpop	1.8%	.	.	.	.	2.7%	1.9%	.	13.2%	.
work_dens	.	.	.	-3.2%	.	.	.	.	-6.9%	.
w_n_p_d	.	.	.	.	.	.	.	.	-14.8%	.
pd	.	.	.	.	.	.	.	.	.	.
Dist_Urb	.	.	3.6%	-10.9%	.	.	.	2.3%	.	-6.6%
Dist_M_Urb	3.5%	.	.	-7.9%	.	.	-1.8%	.	.	.
Dist_Urb_C	.	.	-0.8%	-5.0%	-5.9%	.	0.1%	.	.	.
Dist_M_U_C	.	.	1.8%	.	.	.	.	.	.	.
RCatCU	.	.	.	.	.	.	.	.	.	.
Urb_Rur1	.	.	.	.	.	.	.	.	.	.
Junction1	.	.	.	.	.	.	.	.	.	.
Nature: Dual Carriageway	.	.	.			.	.	.		

Table D-9: 'U' Roads coefficients for Cars, Buses and Two-wheeled vehicles

Vehicle Type	Cars					Buses					Two-wheeled				
Group	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
optimal lambda	0.04174	0.03112	0.01683	0.01728	0.01602	0.26058	0.09733	0.04008	0.02086	0.06023	0.31563	0.06228	0.01317	0.02585	0.09611
<u>Variables</u>															
(Intercept)	6939.1	2606.5	1029.9	326.2	129.4	81.1	9.8	3.3	1.5	1.2	59.5	15.6	6.9	2.3	2.0
Animals	2.6%	.	.	1.1%	11.5%	.	-10.6%	.	.	.	.	.	1.9%	.	.
Bus_statio	-8.6%	.	-3.0%	.		.	.	-4.9%	.	.	.	.	.	.	.
Charge_p	.	.	.	-7.0%	7.2%	.	.	.	-0.3%	.	.	.	2.0%	-4.2%	1.3%
Factories	.	.	-5.7%	.	.	0.4%	-22.2%	.	.	.	.	.	-7.4%	.	.
Healthcare	.	.	.	.	-1.6%	.	.	-9.1%	-1.1%	.	.	.	-8.0%	.	.
Leisure	.	.	.	.	0.5%	.	3.4%	.	.	.	.	.	0.7%	-0.2%	.
Marine	-3.0%	-3.3%	0.0%	.	.	.	3.2%	.	.	.	.	1.1%	6.7%	.	.
Office	.	.	-0.2%	-0.3%	21.9%	.	.	.	.	.	1.5%	.	.	.	1.1%
Parking	.	.	.	.	-5.9%	.	0.9%	.	.	.	.	.	-5.9%	-0.7%	.
Petrol	.	.	-1.8%	1.7%	1.0%	.	.	-2.8%	.	.	.	-2.9%	-9.6%	.	.
Public	.	.	.	.	6.3%	.	.	-0.6%	-0.8%	.	.	.	4.2%	.	.
Research	.	.	.	.	8.1%	.	.	.	.	.	.	.	0.4%	.	.
Shops	.	.	.	-0.4%	-6.2%	.	.	-1.7%	-3.4%	.	.	.	-1.9%	.	.
Sport	.	.	-0.8%	0.3%	-3.0%	.	24.7%	.	-1.0%	.	.	5.6%	.	.	.
Superstore	.	.	-1.6%	0.4%		36.1%	.	.	-2.3%	.	.	.	-2.8%	-1.0%	.
Under_cons	.	.	-1.0%	.	3.3%	.	.	.	-0.3%	.	.	.	-8.7%	.	.
Vacation	.	.	1.2%	-1.5%	0.4%	.	-5.6%	.	-0.2%	.	.	.	4.0%	-0.5%	.
Vehicle	.	.	.	-4.5%	6.6%	.	.	-15.7%	-1.5%	.	.	.	-3.1%	-2.3%	.
Warehouses	7.0%	.	-0.7%	.	.	70.2%	-14.2%	-9.3%	.	.	.	.	.	1.4%	.
Bus_stops	.	.	.	.	33.3%	.	86.4%	9.6%	1.6%	.	.	.	-4.5%	-1.0%	12.7%
Population	.	.	.	.	-35.9%	.	-13.9%	.	.	.	.	.	-24.3%	.	.
Income	.	.	.	3.8%	.	.	-9.7%	-4.8%	-6.0%	.	.	.	11.4%	.	.

Vehicle Type	Cars					Buses					Two-wheeled				
Group	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
optimal lambda	0.04174	0.03112	0.01683	0.01728	0.01602	0.26058	0.09733	0.04008	0.02086	0.06023	0.31563	0.06228	0.01317	0.02585	0.09611
Variables															
Households	.	1.7%	-1.4%	.	.	.	.	-0.4%	-7.2%	.	.	10.1%	65.6%	.	.
Perc_Sa	.	.	-12.0%	-0.8%	13.4%	.	.	-2.8%	-2.3%	.	.	19.4%	.	-5.2%	.
car_van	.	0.1%	.	4.1%	65.6%	.	.	-22.3%	-2.6%	.	.	.	-7.3%	.	.
workpop	.	.	.	2.0%	0.3%	.	.	.	.	.	14.1%	.	.	0.8%	.
work_dens	.	.	.	.	.	.	.	.	.	.	.	.	26.1%	.	7.3%
w_n_p_d	.	.	.	.	26.3%	.	.	.	.	.	.	.	.	.	1.5%
pd	.	.	-10.3%	-5.7%	12.5%	.	.	.	.	.	.	.	.	3.6%	.
Dist_Urb	.	.	.	-0.4%	-16.2%	.	-4.4%	-3.1%	.	.	.	2.3%	-3.6%	0.4%	.
Dist_M_Urb	-5.3%	.	.	.	-0.9%	.	.	1.6%	.	.	.	.	.	.	.
Dist_Urb_C	0.5%	.	.	-3.0%	4.0%	-9.4%	-19.7%	.	0.0%	.	.	.	15.6%	1.6%	.
Dist_M_U_C	.	-0.2%	.	.	.	.	.	.	.	.	.	2.9%	15.0%	10.2%	1.6%
RCatUU	.	.	.	.	46.1%										
Urb_Rur1	.	.	.	.	.										
Airport1	.		.		.										
Port1	.	.	.	.	.										
Junction1	.	.	.	.	.										
Nature: Dual Carriageway	.	.	.	.	.	.	.	.	.	.					

Table D-10: 'U' Roads coefficients for LGVs and HGVs

Vehicle Type	**LGVs**					**HGVs**				
Group	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
optimal lambda	0.152836	0.044376	0.006068	0.024257	0.082841	0.058570	0.075876	0.040810	0.019935	0.236722
Variables										
(Intercept)	823.3	323.7	133.9	43.7	27.1	152.2	21.0	7.1	2.4	3.0
Animals	.	.	.	1.4%	.	6.8%	.	-2.4%	0.5%	.
Bus_statio	.	.	-5.1%	.		.	.	.	.	
Charge_p	.	.	-1.9%	-4.2%	.	.	.	.	-2.3%	.
Factories	.	.	-4.9%	1.8%	.	.	.	.	11.6%	.
Healthcare	.	.	-2.6%	.	.	.	.	.	.	.
Leisure	.	.	0.4%	.	.	.	.	.	.	.
Marine	.	.	.	.	.	-14.2%	1.5%	.	.	.
Office	.	.	.	.	10.3%	.	.	.	.	.
Parking	.	.	-5.7%	.	.	.	.	.	3.9%	.
Petrol	.	.	-1.8%	2.0%	.	.	.	.	.	.
Public	.	.	4.2%	.	1.0%	.	.	.	-5.3%	.
Research	.	.	0.5%	.	0.6%	.	.	.	-0.5%	.
Shops	.	0.0%	.	-1.5%	.	.	.	.	-1.0%	.
Sport	.	.	0.1%	.	.	.	.	.	.	.
Superstore	.	.	-0.8%	.		.	.	.	-3.8%	
Under_cons	.	.	.	.	.	.	.	.	.	.
Vacation	.	.	3.1%	-0.8%	.	.	.	.	-0.8%	.
Vehicle	.	.	.	-0.1%	7.9%	43.8%	2.4%	0.0%	-4.3%	.
Warehouses	.	11.1%	10.4%	0.8%	3.8%	6.7%	35.4%	17.7%	5.7%	.
Bus_stops	.	-0.3%	-0.3%	.	14.0%	.	-14.2%	-8.6%	-0.1%	.
Population	.	.	-26.2%	.	.	.	.	-0.5%	.	.
Income	.	.	-2.2%	.	.	44.8%	.	.	0.8%	.

Vehicle Type	LGVs					HGVs				
Group	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
optimal lambda	0.152836	0.044376	0.006068	0.024257	0.082841	0.058570	0.075876	0.040810	0.019935	0.236722
Variables										
Households	.	.	13.7%	.	.	.	.	.	-5.7%	.
Perc_Sa	.	.	-5.2%	-0.5%	2.1%	.	.	.	.	.
car_van	.	.	.	.	.	-30.2%	-16.6%	-22.3%	.	.
workpop	.	.	.	.	3.1%	.	15.2%	.	7.0%	.
work_dens	.	.	6.9%	.	.	.	.	5.3%	.	.
w_n_p_d	.	.	.	.	14.3%	.	.	.	.	.
pd	.	.	.	-5.4%	.	.	-2.2%	.	-5.1%	.
Dist_Urb	.	.	2.0%	.	-8.4%	6.8%	10.7%	4.1%	0.4%	.
Dist_M_Urb	.	.	1.8%	.	.	.	.	.	.	.
Dist_Urb_C	.	.	.	-3.7%	.	31.8%	-0.2%	.	-0.7%	.
Dist_M_U_C	.	.	4.0%	2.5%	.	-9.9%	.	6.8%	0.2%	.
RCatUR	.	.	.	.	.	.	.	.	.	.
RCatUU	.	.	.	.	.	-58.6%	.	.	.	.
Urb_Rur1	.	.	.	.	.	.	.	.	.	.
Airport1	.	.	.	.	.	.	.	.	.	.
Port1	.	.	.	.	.	.	.	.	.	.
Junction1	.	.	.	.	.	.	.	.	.	.
Nature: Dual Carriageway	.	.	.	.	.	.	.	.	.	.

E. Supplementary material for chapter 5

Table E-1: Classification accuracy for England and Wales

Algorithm	Road Class			
	A	B	C	U
RF	85.56%	63.16%	71.96%	70.32%
GBM	89.58%	66.51%	74.30%	71.29%
KNN	84.50%	45.45%	71.50%	60.10%

Table E-2: Classification accuracy for Greater London

Algorithm	Road Class			
	A	B	C	U
RF	92.05%	50.00%	60.00%	67.57%
GBM	93.18%	50.00%	66.67%	70.27%
KNN	89.77%	37.50%	33.33%	64.86%

Table E-3: Traffic change in Greater London (in billion VKT)

Vehicle Type	VKT 2015 (bVKT)	VKT 2019 (bVKT)	Change (%)
Cars/Taxis	25.6	28.1	+9.77%
LGVs	4.7	5.9	+25.53%
HGVs	1.1	1.0	-9.09%
Two-wheeled	0.7	0.9	+28.57%
Buses	0.6	0.5	-16.67%