Factors Affecting Road Traffic: Identifying Drivers of Annual Average Daily Traffic (AADT) Using LASSO Regression

Alexandros Sfyridis[a,*], Paolo Agnolucci[a,b]

[a]Institute for Sustainable Resources, University College London, 14 Upper Woburn Place, WC1H 0NN, London, United Kingdom

[b]World Bank Macroeconomics, Trade and Investment Global Practice, 1818 H Street, N.W., Washington, D.C. 20006, USA

* Corresponding author

Email addresses: alexandros.sfyridis.15@ucl.ac.uk (A. Sfyridis), p.agnolucci@ucl.ac.uk (P. Agnolucci)

Keywords: Annual Average Daily Traffic (AADT), Least Absolute Shrinkage and Selection Operator (LASSO), Road Traffic, Transport

Acknowledgments

This work was supported by the Natural Environment Research Council (NE/M019799/1) and by the UK Energy Research Centre (Grant Number: EP/L024756/1).

Abstract

Road traffic data is important for various applications in transport studies, such as those related to safety, environmental damages and economic evaluations. Although significant improvement in estimation accuracy has been achieved, less is known about the association of specific factors with road traffic volumes. This paper presents an investigation of the relation of various road, area and socio-economic characteristics with Annual Average Daily Traffic (AADT) in England and Wales for four different road classes and five vehicle types. This is achieved by applying the Least Absolute Shrinkage and Selection Operator (LASSO) regression on a comprehensive set of land use, socioeconomic, public transport and roadway variables. The output reveals that specific socioeconomic and roadway characteristics are those that are mainly associated with traffic volumes across all vehicle types and road classes. Moreover, the association of other variables with traffic volume varies, depending on the road class and vehicle type, creating space for future research. The results can support urban planning and inform policies related to transport congestion and environmental impacts mitigation.

1. **Introduction**

Availability of road traffic data is of high importance for several disciplines, as it is used for Green House Gases (GHG) and air pollutant emission estimation (1), exposure to noise generation and related health impacts (2) as well as economic evaluations of safety projects (3) among others. However, traffic data collection across the full extent of the road network is challenging and costly, and therefore transport departments rely on automatic traffic counters installed on limited sites and sometimes on manual measurements undertaken seasonally on few locations. Consequently, several methods to estimate traffic volumes across the network have been investigated, taking into consideration, to a different extent, the factors associated with traffic. In particular, Annual Average Daily Traffic (AADT) – a measure of traffic volume (traffic volume definition in provided in the Supplementary Information) defined as the average daily traffic on a street segment over a year (4) – has been the research focus for studies on motorised (5) and non-motorised transport (6,7), aiming to address issues such as accident prediction (8) and air pollutants estimation (9).

Whilst research on AADT estimation and related driving factors has been improved with the incorporation of novel modelling approaches, significant limitations are still observed. First, estimations are mainly conducted on total AADT while volumes for different vehicle types is largely unexplored. Secondly, research is preoccupied with AADT estimation on major roads (10), while only a few studies take into account minor roads (3,11). Moreover, the majority of studies focuses on estimation accuracy (12,13), without considering the relationships between several characteristics – such as land use – and road transport. Finally, the explanatory variables used in several models are limited and consequently many driving factors are excluded. 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 paper, we investigate the association of driving factors with traffic volumes (AADT) focusing on England and Wales. The Department for Transport (DfT) in the UK, classifies road network to two categories, major and minor roads (14). 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 (15). 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 (15). Our analysis is focused on four different road classes which are examined individually (i.e., 'A', 'B', 'C' and 'U' roads). Motorways are excluded taking into consideration that these roads are not affected by the surrounding characteristics (30). The analysis is conducted for five different vehicle types (i.e., Cars, buses, Light Good Vehicles, Heavy Good Vehicles and two-wheeled vehicles) where a statistical model is applied, and the most statistically significant variables are identified so that their association with AADT can be assessed. To perform this, we build on the work conducted in (16), where high AADT estimation accuracy is achieved across all road types and a comprehensive set of driving factors is used in the study.

The paper is presented in seven sections. Section 2 provides a literature review on AADT modelling approaches and the findings regarding the driving factors. Section 3 presents the dataset used and section 4 describes the methodology used to understand the association of various factors with AADT. Section 5 presents the summarised results, while the full set of results for 'A', 'B', 'C', and 'U' roads is presented in the Supplementary Information. Section 6 analyses and discusses the findings while section 7 concludes the outcomes of this study and explores future considerations to improve this work to further contribute to the field.

2. **Literature review**

Numerous studies undertake AADT estimation, with the use of various explanatory variables incorporated in Machine Learning (ML) and Data Mining approaches (4,17), Linear Regressions (18) and Spatial statistics (19). Recent rise in applications of Machine Learning algorithms has provided higher prediction accuracy (20–22), although these models may lack strong theoretical basis and are sometimes considered 'black-boxes' (23–25) which means there is a lack of knowledge in the internal process (26,27) making them unsuitable for interpreting the outcomes (20). This is reflected both in AADT estimation studies (4) and other disciplines (28,29), where ML models provide more accurate results compared to other models, but the exact relations of predictors with the response variable cannot be assessed and interpreted. In this section we provide a synopsis of AADT estimation studies, firstly by briefly summarising models aimed at increasing prediction accuracy and secondly by focusing on modelling approaches where statistical coefficients are shown and discussed.

2.1. Prediction-driven AADT models

This subsection presents statistical models intent to increase AADT estimation accuracy rather than interpreting the potential effects of explanatory variables on traffic volumes. For example, in (30) and (31) Kriging interpolation is applied using mean square prediction error (MSPE) and absolute percentage errors (APE) to evaluate the spatial models developed. Similarly, in (19) Kriging and Geographic Weighted Regression (GWR) is also applied and the models are assessed with APE, while in (32) regression Kriging is applied and the model outputs are assessed by using the Mean absolute percentage error (MAPE) evaluation metric.

In the Machine Learning literature, in (12) the K-STAR (K*) and Random Forest models are used and validated with the P-value and Nash-Sutcliffe (N-S) statistic, while in (13) Artificial Neural Networks are applied and validated with MAPE. Root Mean Square Error (RMSE) is used in (4) to validate the Random Forest, Support Vector Regression and K-nearest neighbour models used to estimate AADT in low volume roads. A combination of clustering and regression modelling has been evaluated with Mean Absolute Error (MAE) and MAPE in (33).

2.2. Coefficient-based AADT estimation models

This subsection presents studies where predictor coefficients are extracted, mainly based on standard statistical regression models. Specifically, the authors in (34) used linear regression with 11 independent variables in Indiana, US. The results show that road environment (i.e., the road located in an urban or rural area), accessibility to motorways and population are statistically significant and exhibit positive coefficients. The authors in (35) also apply linear regression on 12 variables, to conclude that the number of lanes, road environment and functional classification (i.e., the road being arterial or collector) exhibit high positive coefficients. Number of lanes and functional classification explain 50% and 26% of the variation respectively. A multiple regression model has been used in (36) to estimate AADT in rural roads in Georgia, US. Among the 45 variables, population and number of farms are statistically significant with positive signs, while distance to large urban areas is also significant with negative impact. In (37) four regression models are tested with eight variables in Florida. The best performing model – validated by MSE – indicates that functional class and number of lanes are the most significant predictors with high positive coefficients. Accessibility to employment centres and motorways, population and workplace population are also statistically significant and have positive signs. In (38) Ordinary Least Squares (OLS) regression is applied on a similar set of variables to find that the number of lanes and access to motorways have higher positive signs. In the same study, the variables of population, employment population and distance to employment centres are statistically significant with positive coefficients.

More recently, the authors in (11) used liner regression to estimate AADT values on low volume roads in Wyoming, US. Six variables are statistically significant with paved roads – as opposed to dirt roads – accessibility to primary roads and population in the vicinity of traffic counters result into positive signs as opposed to land use variables which exhibit negative coefficients. In (39) with a multiple linear regression model for two cities in Alabama, US, the authors identified that functional class and number of lanes are statistically significant with high positive coefficients. Retail employment has also a positive sign. On the contrary, population and non-retail employment have negative signs as opposed to findings from other studies. Finally, the authors in (40) used a Generalised Linear Mixture Model (GLMM) and the Synthetic Minority Over-sampling Technique (SMOTE) with twenty variables in Seattle, US. The findings show that 17 and 15 variables respectively are statistically significant with the "Spatial weighted volume" having the highest positive coefficient. The Spatial weighted volume is defined as the "the sum of other roads' weighted AADT divided by the reciprocal of squared Euclidean distance, in counts/sq. feet" (40). 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. Distance to motorways also has a negative sign.

3. **Data**

We use the dataset presented in (16), which fits our purposes due to low AADT estimation errors achieved in the paper and the rich set of variables introduced in the models. The dataset contains approximately 19,000 geotagged traffic count points located in England and Wales on four different road types – 'A', 'B', 'C' and 'U'. For each road type, AADT for different vehicle types – Cars, Buses, LGVs, HGVs and Two-wheeled – are taken. For each count point, service areas of different sizes are drawn and for each service area four different sets of variables associated with AADT are considered: i) roadway characteristics taking into account attributes of the roadway, ii) socioeconomic characteristics considering attributes such as population and income in the counters' vicinity, iii) land use characteristics related to building use and infrastructure around a count point and iv) public transport facilities including bus stops and train station accessibility. The service areas are essentially "buffers" that consider the road network – instead of using straight Euclidean distances. The measure captures the actual distance (in metres) a vehicle covers on the network (16). The complete set of variables is presented in Table A1.

For each road class, the count points are split into five subcategories (henceforth referred to as groups), which are colour coded in the same way across all road types and each group has a different service area. The groups are a result of the clustering process followed to create the dataset 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 groups and some variables have zero values in some groups. However, a generalised pattern related to traffic volumes emerge – as reported in (16) – with AADT decrease starting from group 1 towards group 5 for all road types (Figure 1). Moreover, the groups appear to be significantly affected by the location of traffic counters, being placed in urban, suburban or rural areas (Table 1). The independent variables are presented in Table A1 in the Supplementary Information. For more information on the process followed to build the dataset, the reader can refer to (16).

[FIGURE 1 HERE]

Figure 1: Total AADT Values by Road Class and Group

Table 1: Indication of Count Point Locations and samples for each Group

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%	

4. Methodology

To understand the association of particular factors with AADT we examine the coefficients obtained from statistical modelling by first transforming the data in a form enabling comparison of coefficients across variables, then by applying the regression model, and lastly by extracting statistically significant coefficients. From Table A1, 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). To allow comparison of the coefficients, we standardised all variables (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, provided that they have finite mean μ and standard deviation (SD) σ :

$$x' = \frac{x - \mu}{\sigma} \tag{1}$$

This transformation will allow to compare and assess the association of each variable with AADT (41).

Moreover, to account for the distribution of the dependent variable that is skewed to the right, we apply logarithmic transformation on the variable (42) for all groups and subgroups. Therefore, one can interpret the coefficients of predictors on the dependent variable as percentage impacts (43), rather than absolute units.

To model the relationship between the independent and dependent variables, we use the Least Absolute Shrinkage and Selection Operator (LASSO) method (44) applied on each vehicle type, road class and subgroup individually. Lasso is both a regression and variable selection method, that aims to produce a set of statistically significant predictors to minimise the estimation error (45) and it has been found to perform well when used for both regression as well as variable selection tasks (46,47). The variable selection is done by imposing a penalty term on the model parameters so that some regression coefficients shrink to zero (48) and consequently these variables are excluded. This is an advantage over similar methods such as Ridge regression, that do not reduce the number of variables (49) and therefore not able to reduce the complexity of a dataset incorporating a large number of variables. 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 Equation 2 below (50):

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

where y_i and x_i are the outcome and predictor variables respectively and β represents the coefficients at the ith observation of the jth independent variable.

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, we apply the k-fold cross validation method (51,52) 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 (53,54).

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 (55). Here we use k=10 and test on 100 repetitions. The average value for optimal lambda across the repetitions is calculated and used as the parameter in Lasso.

5. **Results**

In this section, results are presented for each road and vehicle type. Firstly, we present and briefly discuss the coefficient of determination (i.e., R squared – R^2) for all models applied, indicating the goodness of fit and then, we show the coefficients extracted from Lasso.

5.1. R-squared (R^2)

In Table 2 the R^2 values for all the models applied are shown. Overall, it can be observed that certain patterns emerge across vehicle types, road classes and groups. More specifically, the R^2 values tend to be higher in the 'A' road class while they are normally lower in other road classes with a few exceptions, indicating the higher reliability of the results for 'A' roads. Moreover, R^2 values are normally higher in high traffic volume groups (e.g., groups 1 and 2), and in most cases gradually decrease in low traffic volume groups, a pattern that is again more distinct in the case of 'A' roads. This observation also complies with the finding from (16) – where the data have been extracted from – where the models perform better in high volume road classes and groups. In terms of vehicle types, it can also be observed that the R^2 statistic is in most cases higher for buses and two-wheeled vehicles and are normally lower in the case of LGVs and HGVs, with a few exceptions such as group 1 in 'C' roads (LGVs) and group 1 in 'U' roads (HGVs). The goodness of fit is also normally lower for cars compared to other vehicle types. However, it should be noted that the amount of noise in the data does prevents from reaching very high R^2 values. Low R^2 is also common in cross-sectional data (56) compared to time-series data analysis, where the values tend to be higher.

Table 2: Lasso R-squared values for each road class, group and vehicle type

Group	Cars	Buses	LGVs	HGVs	Two-wheeled
	A				
1	0.22	0.57	0.36	0.37	0.63
2	0.10	0.40	0.22	0.54	0.52
3	0.22	0.23	0.16	0.28	0.65
4	0.20	0.18	0.23	0.35	0.20
5	0.22	0.15	0.18	0.20	0.13
	B				
1	0.27	0.49	0.17	0.10	0.39
2	0.10	0.41	0.17	0.32	0.46
3	0.25	0.18	0.28	0.31	0.28
4	0.21	0.23	0.22	0.32	0.23
5	0.10	0.27	0.10	0.10	0.10
	C				
1	0.46	0.10	0.36	0.10	0.42
2	0.18	0.24	0.10	0.25	0.38
3	0.44	0.24	0.23	0.13	0.41
4	0.24	0.17	0.10	0.18	0.10
5	0.10	0.10	0.10	0.10	0.10
	U				
1	0.38	0.32	0.10	0.66	0.20
2	0.10	0.32	0.11	0.26	0.13
3	0.36	0.11	0.25	0.15	0.21
4	0.10	0.10	0.10	0.10	0.10
5	0.40	0.10	0.14	0.10	0.10

5.2. Lasso coefficients

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. It is worth pointing out that the meaning of coefficients is different across variable types. In the case of continuous variables, the estimated coefficients are standardised semi-elasticities and 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 A2 in the Supplementary Information. In this section we present the Lasso regression outcomes for all road classes. However, due to the large number of variables investigated, we identify and present the variables exhibiting high positive or negative coefficients across all groups and vehicle types for each road class separately. Specifically, the three most statistically significant variables for each road class are shown in Table 3. The complete set of results is presented in the Supplementary Information also including the coefficients in Tables A3 – A10. Discussion and interpretation of our findings are presented in section 6.

Table 3: Lasso coefficients for each road class and vehicle type

Variable	Cars	Buses	LHVs & HGVs	Two-wheeled
	A			
Distance from Urban Area	-12.9% – 0.8%	-11.8% – 0.1%	-10.4% – 5.7%	-6.1% – -1.2%
Dual Carriageway	16.9% – 48.9%	6.2% – 32.2%	18.1% – 76%	21.6% – 32.4%
Ring Road	2.5% – 19.5%	-29.7% – -11.8%	6% – 32.1%	-11.4% – 13%
	B			
Distance from Urban Area	-2.2% – -2%	-17.8% – -2.8%	3.3% – 24.1%	2.6% – 12.3%
Bus stops	-1.2% – 0.2%	22.1% – 24%	-20.1% – -0.9%	-14% – -4.4%
Income	-1.1% – 7.7%	-17.8% – -5.7%	0.7% – 5.5%	2.9% – 25.1%
	C			
Distance from Urban Area	-23.6% – -1.7%	-22.4% – -7%	-10.9% – 3.6%	3.7% – 10.1%
Population	-2.9% – 21.5%	0% – 8.5%	5.5% – 12.2%	0% – 1.5%
Income	0.5% – 12.8%	-3.2% – 0%	-1.5% – 6.4%	4.6% – 9.6%
	U			
Distance from Urban Area	-16.2% – -0.4%	-4.4% – -3.1%	-8.4% – 10.7%	-3.6% – 2.3%
Bus stops	0% – 33.3%	1.6% – 86.4%	-14.2% – 14%	-4.5% – 12.7%
Income	0% – 3.8%	-9.7% – -4.8%	-2.2% – 44.8%	0% – 11.4%

5.2.1. ‘A’ roads

For this road type, we have observed that the variables representing roadway characteristics are those highly associated with large deviations in traffic volume across all groups for the vehicle types we have examined. Specifically, from Table 3, it can be observed that dual carriageways are correlated with high vehicle volumes compared to single carriageways ranging from 6% for buses, up to 76% for HGVs, while the values are also high in the case of cars, which are the vehicles that dominate the streets. The distance from urban areas’ centres is also statistically significant for all vehicle types in the majority of groups, with mixed signs being observed in the case of LGVs and HGVs. In particular, the negative sigs – up to -10% for LGVs and -6% for HGVs – are observed in the low traffic volume group (i.e., group 5 – see Supplementary Information), while positive signs are evident in higher volume groups. On the contrary, the coefficients for cars, buses and two-wheeled vehicles indicate that the further a count point is from an urban area, the lower the traffic volumes for these vehicles, considering all other factors being equal. The values are as high as -13% for cars, -12% for buses and -6% for two-wheeled vehicles, all being observed in the low volume group. The occurrence of ring roads completes the set of roadway characteristics associated with deviations in traffic volumes. Ring roads are related to significantly higher volumes for car, LGV and HGV traffic across all groups with the coefficient signs being as high as 20% for cars, 21% for LGVs and 37% for HGVs respectively, mainly observed in groups 3 and 4. Ring roads are also related to lower two-wheeled vehicle traffic volumes in group 1 although the signs are positive in all other groups, reaching 13% in group 3. On the contrary, these roads are associated with significantly lower bus volumes across all groups ranging from -12% in group 1, up to -30% in group 5.

5.2.2. ‘B’ roads

For ‘B’ roads the set of variables associated with deviations in traffic volumes across all groups and vehicle types are not only related to the roadway characteristics, but also consist of public transport and socioeconomic indicators. The distance from urban areas is again statistically significant and correlated with lower car and bus traffic volumes, as high as -2% and -18% for these vehicle types respectively, although the signs for two-wheeled vehicles are positive, ranging from 3% to 12%. Similarly, the further the count point is from an urban area, the higher LGVs and HGVs traffic volumes, particularly in the case of HGVs, where distance is correlated with higher traffic volumes at 24%. The presence of bus stops, however, is related to lower LGV traffic volume as high as -2%, while the sings indicate significantly smaller number of HGV vehicles, ranging from -10% up to -20% and observed in groups 3, 4 and 5. Bus stops are also correlated with a slightly lower traffic volume for cars in this road class, although the coefficients signs are remarkably higher for two-wheeled vehicles, ranging from -4% to -14%. Unsurprisingly, bus stops are related to higher bus numbers in 'B' roads, ranging from 22% to 24% across the groups. Finally, income is the socioeconomic variable that is found to be correlated with traffic volumes across the groups and vehicle types for this road class. In particular, income is highly correlated with increased numbers of two-wheeled vehicles across all groups, reaching as high as 25%, while it is also correlated with higher number of cars (8%), although the sign is negative in group 3 (-1%). Income is also associated with lower bus traffic volumes indicating a smaller number of vehicles of about -3% in group 5 and up to -16% and -18% in groups 3 and 4 respectively. This socioeconomic variable is related with higher LGV and HGV volumes ranging from 1% to 6% for LGVs and being retained by Lasso in group 5 only for HGVs (6%).

5.2.3. 'C' roads

In the case of 'C' roads, the distance from urban areas variable is retained by Lasso and is correlated with lower car and bus traffic volumes, as well as higher two-wheeled vehicle numbers. Specifically, the coefficient sings indicate fewer cars ranging from -2% to -24% and fewer buses from -7% to -22% the further the count point is from an urban area, with all signs being observed in groups 1-4. On the contrary, distance is associated with more two-wheeled vehicles in groups 1-3, with the signs ranging from 4% to 10%. In the case of LGVs and HGVs mixed signs are observed where the traffic volume is higher in group 3 at 4% and 2%, although the coefficients are negative in low volume groups 4 and 5 at -11% and -7% for each vehicle type respectively. The socioeconomic variable of population is associated with higher number of all vehicle types across all groups, with the exception of cars in group 3, where the coefficient is negative (-3%). In other groups, population is related to higher car volumes up to 22%, higher bus and two-wheeled vehicle volumes at 9% and 2% respectively, as well as increased LGV traffic volumes as high as 12%, with all coefficients being observed in low traffic volume groups, such as group 4 and group 5. Finally, income is again selected by Lasso, where the coefficient signs are mostly similar to the ones observed for the 'B' road class. Income is correlated with higher two-wheeled and car vehicle traffic volume up to 10% and 13% respectively and with lower bus traffic volume of approximately -3%. The variable is also associated with higher LGVs in group 4 at 6%, although the sign is negative in group 1 at -2%.

5.2.4. 'U' roads

For 'U' roads the variable indicating distance from urban areas is again statistically significant, with the signs being similar to the ones observed in other road classes. For cars, distance is associated with lower traffic volumes reaching -16% in group 5, while for buses, the values range from -3% to -4% in groups 2 and 3, similar to road class 'C'. The signs for two-wheeled vehicles are mixed, indicating lower volumes in group 3 (-4%) and higher in groups 2 and 4 reaching 2%. Similarly, the coefficient signs are also mixed in the case of LGVs indicating that distance is associated with fewer LGVs in group 5 by 8%, although the sign is positive in group 3 at 2%. However, distance is associated with higher HGV volumes up to 11%, with the variable being retained in most groups. The variable indicating bus stops is retained by Lasso, as it has been the case in the 'B' road class. The presence of bus stops is again unsurprisingly associated with significantly higher bus traffic volumes up to 86%, although it also indicates higher car and two-wheeled vehicles volumes in group 5 at 33% and 13% respectively. However, the signs are negative for two-wheeled vehicles in other groups ranging from -1% to -5% as are the coefficients for LGVs and HGVs. In particular, bus stops are correlated with lower LGV and HGV traffic volumes up to -14% with the exception of LGVs again in group 5, where the signs are positive at 14%. Finally, income is again retained by Lasso as in 'B' and 'C' road classes with similar coefficients. Income is related to higher car and two-wheeled vehicle volumes (i.e., 4% and 11% respectively) and lower bus traffic volumes ranging from -5% to -10%. Similar to 'B' roads income is correlated with lower LGV volumes by -2%, although the sign is different and indicates considerably higher HGV volumes reaching up to 45% in group 1.

6. Discussion

This study focused on the identification of several characteristics that are associated with traffic volumes of the five different vehicle types, normally used by the Department for Transport (DfT). The coefficients were analysed by road class and vehicle type, further classified in five subgroups. In this section we discuss the results presented in section 5 identifying the causes behind the statistically significant variables produced by Lasso and providing evidence of how these can be of importance to practitioners, such as policy makers, urban and transport planners. Due to the large number of variables, we focus on those exhibiting similar patters across road types and groups and those with the highest impact on each vehicle's traffic volume as indicated by the coefficients.

As a general observation we can see that most coefficients take the expected value across road classes, vehicle types and groups. However, one has to bear in mind the imbalanced number of traffic counters for each road type (Table 1) and that for smaller road classes traffic counts are taken seasonally and manually, and eventually adjusted to estimate AADT (57,58), 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.

We also can observe the significance of some variables across all road classes and groups. For example, one can observe that distance to urban areas is statistically significant for all vehicle types, across all road classes and most groups as shown in Table 3 and Tables A3-A10. This pattern – which is more distinct for 'A' roads – indicates that the further from urban areas the less cars, buses and two-wheeled vehicles will be on the streets. The number of private vehicles (i.e., cars and two-wheeled) is expected to be higher in urban areas considering the larger population concentration and higher economic activity taking place in these areas, while more buses are expected in urban rather than rural areas (59). The coefficients for HGVs and LGVs are mixed, implying further investigation. In particular, for 'A' roads, the higher the distance from urban and major urban areas, the lower the level of HGV and LGV 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 groups (i.e., groups 1, 2 and 3) the signs are similar to 'B' roads. However, the negative coefficients observed in low volume rural groups 4 and 5 particularly for HGVs, indicate that operation of these vehicles is related with urban areas of higher economic activity.

Moreover, carriageway type is significant for 'A' roads, with dual carriageways being associated with higher traffic volumes for all vehicle types. The variable is also statistically significant for the majority of vehicle types in 'B' and 'C' roads, although it is not significant for 'U' roads, due to the fact that there are not many dual carriageways in these roads. More vehicles are expected to drive through dual carriageways – as opposed to single – due to safety and increased speed limits (60–62) and lower design standards for single carriageways (63). Studies have shown that accidents are less likely to occur on dual carriageways (64,65) while the occurrence of these roads is key, particularly in urban environments, due to the fact that they normally reduce travel time of individuals and therefore, providing access to significant services such as schools, workplaces and healthcare facilities (66). However, the design and construction of dual carriageways should not be considered a de facto approach by urban and transport planners due to multiple concerns related with these roads. For example, safety in dual carriageways does not only depend on design standards but also on other factors, such as lighting and weather conditions (67), while the higher number of vehicles normally driving through these roads is associated with higher noise and air pollution levels (68) as well as other health-related conditions such as attention-deficit/hyperactivity disorder (ADHD) (69). In fact, our findings highlight the importance of considering additional factors related to the functionality of dual carriageways. For instance, taking into account the higher traffic volumes, introducing bus lanes to facilitate bus flow or imposing restrictions for HGVs similar to those applied in the Greater London area (70) can be considered.

In a similar manner, ring roads are statistically significant in 'A' roads only, due to the fact that all ring roads in the dataset belong to this road class. These roads are associated with higher traffic volumes for cars, LGVs and HGVs and lower traffic volumes for buses, while the coefficient signs are mixed in the case of two-wheeled vehicles as also mentioned in section 5. This is an expected outcome since ring roads are usually large roads designed to carry and distribute large traffic volumes radially across the network (71) and also maximising the average travel speed (72). This is clearly reflected in Table 3 in the case of cars and large number of HGVs which use ring roads to access large facilities normally located at a distance from the urban core, before transportation switches to smaller vehicles in inner city roads (73). Hence, the importance of ring roads cannot be overstated considering the increased capacity and the capability to reduce traffic congestion (74,75), although transport planners should be aware of major concerns related to the functionality and limitations of ring roads. Ring roads encourage the use of private vehicles , which in turn, can result into extensive urban sprawl (76) and increased levels of air pollution (77,78). Moreover, studies have investigated the economic effects of ring roads, concluding that they decentralise the service sector and industrial activity, as well as displacing population to surrounding regions, normally resulting into reduced inner city GDP (79). Similarly, (80) concludes that ring roads affect the spatial configuration of shops and the economic centres of built environments. Ring roads are also associated with lower bus traffic, and although buses are less likely to drive through these roads, a similar pattern in the coefficient signs is observed for the income variable (Table 3). Income is associated with higher car – and two-wheeled – traffic volumes and lower for buses, particularly in the case of 'B', 'C' and 'U' road classes. Many studies have concluded that households with higher incomes are more likely to own private vehicles, such as cars (81–83), resulting in lower bus traffic, since transport needs are met by private vehicles (84,85). On the contrary, lower income and regularly marginalised communities rely on public transport which in many occasions suffer from poor services (86). Dividing communities is also facilitated by the occurrence of ring roads obstructing the continuity of urban flows (87) and individuals who reside in these areas are normally characterized by lower income and also spend more time commuting as opposed to higher income earners (88).

This highlights the necessity to invest in public transport which has beneficial effects on social quality (89) and economic development (90), as well as the potential to mitigate traffic congestion caused by the excess usage of private vehicles, which again, is associated with higher air pollution levels and adverse impacts on health (91,92). The latter is also clearly reflected in the results, where bus stops are associated with higher bus volumes and lower car and two-wheeled vehicle volumes in the majority of cases with a few exceptions. These exceptions are normally identified in group 5, where counters are located mainly in rural and/or suburban areas. In these areas commuting is more likely to occur with private vehicles, due to lack of public transport (85) and long waiting times (93). The importance of public transport to address the aforementioned issues is also reflected by the coefficients for the train/light rail accessibility (Tables A3-A10), which is associated with lower car and two-wheeled vehicle traffic volumes in the majority of cases. Exceptions –

i.e., positive coefficient signs for private vehicles – are again observed in group 5, and in group 3, where train station accessibility is correlated with higher car volumes in ‘A’ roads due to the fact that train stations are usually located near heavy traffic roads (94). Interestingly, train/light rail station accessibility is also associated with higher bus volumes due to buses connecting to train stations (95–97). Finally, it is worth commenting on the association between bus stops and HGV and LGV traffic volumes where the coefficient signs are mainly negative as expected, considering that these vehicles normally operate close to industrial areas where people commute with private vehicles. Again, an exception is observed in group 5, indicating rural and suburban areas. These areas are significantly different in terms of land use and demographic characteristics compared to urban environments where most of the traffic counters in the dataset are located (Table 1) and thus different traffic patterns are expected.

In terms of the socioeconomic characteristics examined, it is shown that the population variable is statistically significant across all vehicle types and most groups. Again, the model has returned the expected outcomes, with a few exceptions, suggesting that the higher the population the higher the traffic volumes of all vehicle types. Specifically, in the case of cars, the coefficients associated to population – and other socioeconomic factors as shown in Tables A3-A10 – are all associated with higher traffic volumes complying with outcomes from other models (34,37,38). Similar to the case of cars the population variable is associated with higher bus volumes across ‘A’, ‘B’ and ‘C’ roads and most groups, indicating that bus routes are placed and used in heavy populated urban areas, where public transport has a significant effect (98,99). The coefficient signs, which are similar for two-wheeled vehicles, indicate that these vehicles are concentrated in dense urban areas with large populations as it is also stated in (100) and (101), while the fact that population also attracts larger number of LGVs and HGVs is also a reflection that higher economic activity – usually observed in large urban areas – drives demand for LGVs and HGVs.

7. Conclusions and future work

In this paper we have examined a comprehensive set of variables to understand their association 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 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. In this section, we summarise the outcomes as presented and discussed in sections 5 and 6 and reflect on potential conclusions and ways to improve our understanding on traffic volumes.

Overall, we can conclude that specific variables are associated with traffic volumes across almost all groups, road classes and vehicle types. For example, distance to urban areas was found to be statistically significant in all cases, although factors related to densely populated urban centres are mainly associated with car, bus and two-wheeled vehicles’ traffic volumes , while LGVs and HGVs are more present at industrialised areas close to urban centres and their outskirts. Socioeconomic variables are related with traffic volumes in numerous ways. Income is associated with the increased use of private vehicles which contributes to the decreased volumes of buses, while population – and workplace population densities – also associated with the distinction between urban and rural environments – relate to volumes of all vehicle types in a different manner. An interesting finding is that population density relates to higher bus volumes (see Supplementary Information), although these patterns are more significant for higher class roads (e.g., ‘A’ or ‘B’), probably due to the fact that buses normally drive through major arteries where bus lanes are more likely to occur.

The distinction between urban and rural environments and the respective association with traffic volumes is also clear from other characteristics. Variables such as public transport presence and accessibility are statistically significant for all vehicle types in most cases.

However, we can also distinguish specific variables correlated only with vehicle types on specific road classes as it can be seen in Table 3 and Tables A3-A10. For instance, we have noticed that some variables are statistically significant on 'A' roads only, due to the different nature of these roads and use by the vehicles. As an example, carriage way type is found to be correlated with higher traffic volumes of all vehicle types in 'A' roads, although the variable is also retained in other road classes for some vehicle types. The variable indicating ring roads is present only on 'A' roads, due to the fact that all ring roads in England and Wales are of 'A' road type, although it affects vehicle types in a different manner. Similar conclusions can be drawn when investigating the relations of specific variables with different vehicle types. For instance, it has been observed that warehouses and factories are always positively correlated with LGVs and HGVs, and registered vehicles are mainly correlated with bus and two-wheeled vehicle volumes.

The above leads us to also consider the number of data points within each road class and subgroup. For 'A' roads the 14,670 points represent over 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. 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, we can safely assume that results for 'A' roads are more reliable.

Moreover, 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 (102), 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, we can identify that further sampling and acquisition of more reliable data can potentially lead to an improved model and understanding of traffic volumes. Moreover, one could apply similar regularisation methods such as ridge regression (103) and elastic net (49) to compare the results and conclude with more meaningful outcomes.

Finally, considering the availability of the utilized dataset across time, a spatiotemporal analysis could be considered in future research, to provide a more comprehensive investigation of the associations between the driving factors and traffic volumes across space and time. The implementation of such a modelling approach, could be conducted by cross-sectional analysis across a predefined timeframe and by identifying patterns in the extracted coefficients signs for each year or by implementing panel data analysis. The latter would also be beneficial to consider unobserved heterogeneity, although we consider that the presented approach of applying independent models for each road class, group and vehicle type addresses – at least to a certain extent – this issue.

Regardless of the identified limitations and potential ways to improve and enrich research in this area, we have been able to set a number of conclusions on how our findings can be of significance to planners and decision makers. For example, the results confirm that the presence of and accessibility to public transport services can reduce congestion, while it has also been discussed that public transport can facilitate in the air pollution mitigation and contribute towards alleviating social inequalities. On the contrary, ring roads encourage the use of private vehicles and are associated with higher traffic volumes of LGVs and HGVs indicating the occurrence of converse effects. These and all other results presented in this paper strengthen the outcomes produced by related studies providing evidence on the relationship between economic activity, infrastructure and social issues with transport, considering the numerous factors investigated. The utilisation of Lasso estimator on this dataset is able to explain how several indicators correlate with traffic volumes and in turn affect economic activity and the environment among other factors. Consequently, road and urban planners can take into account these relationships and act accordingly to identify optimal ways to improve transportation while also considering the corresponding socioeconomic and environmental effects. This knowledge is vital for sustainable urban development.

We acknowledge that our research does not provide answers or solutions to these vital problems. However, we consider that it contributes towards explaining how several roadway, socioeconomic, and transport related indicators can affect our streets, our societies as well as the economic development and urban environments.

Author Contributions

The authors confirm contribution to the paper as follows: Study conception and design A. Sfyridis, P. Agnolucci; data collection: A. Sfyridis; analysis and interpretation of results: A. Sfyridis, P. Agnolucci; draft manuscript preparation: A. Sfyridis, P. Agnolucci. **All authors reviewed the results and approved the final version of the manuscript.**

References

1. Sookun A, Boojhawon R, Rughooputh SDDV. Assessing greenhouse gas and related air pollutant emissions from road traffic counts: A case study for Mauritius. Transp Res Part D Transp Environ [Internet]. 2014;32:35–47. Available from: http://dx.doi.org/10.1016/j.trd.2014.06.005

2. Shu S, Yang P, Zhu Y. Correlation of noise levels and particulate matter concentrations near two major freeways in Los Angeles, California. Environ Pollut [Internet]. 2014;193:130–7. Available from: http://dx.doi.org/10.1016/j.envpol.2014.06.025

3. Wang T, Gan A, Alluri P. Estimating annual average daily traffic for local roads for highway safety analysis. Transp Res Rec [Internet]. 2013;5(2398):60–6. Available from: http://www.scopus.com/inward/record.url?eid=2-s2.0-84896975450&partnerID=40&md5=d003814edf16bda376f39fdb03cdfce8

4. Das S, Tsapakis I. Interpretable machine learning approach in estimating traffic volume on low-volume roadways. Int J Transp Sci Technol [Internet]. 2019;9(1):76–88. Available from: https://doi.org/10.1016/j.ijtst.2019.09.004

5. Lowry M. Spatial interpolation of traffic counts based on origin-destination centrality. J Transp Geogr [Internet]. 2014;36:98–105. Available from: http://dx.doi.org/10.1016/j.jtrangeo.2014.03.007

6. Lu T, Buehler R, Mondschein A, Hankey S. 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 [Internet]. 2017;53:193–204. Available from: http://dx.doi.org/10.1016/j.trd.2017.04.017

7. Hankey S, Lu T, Mondschein A, Buehler R. Spatial models of active travel in small communities: Merging the goals of traffic monitoring and direct-demand modeling. J Transp Heal [Internet]. 2017;7(January):149–59. Available from: https://doi.org/10.1016/j.jth.2017.08.009

8. Çodur MY, Tortum A. An Artificial Neural Network Model for Highway Accident Prediction: A Case Study of Erzurum, Turkey. PROMET - Traffic&Transportation [Internet]. 2015;27(3):217–25. Available from: http://www.fpz.unizg.hr/traffic/index.php/PROMTT/article/view/1551

9. Sfyridis A, Agnolucci P. Road Emissions in London: Insights from Geographically Detailed Classification and Regression Modelling. Atmosphere (Basel) [Internet]. 2021 Jan 30;12(2):188. Available from: https://www.mdpi.com/2073-4433/12/2/188

10. Puliafito SE, Allende D, Pinto S, Castesana P. High resolution inventory of GHG emissions of the road transport sector in Argentina. Atmos Environ. 2015;101:303–11.

11. Apronti D, Ksaibati K, Gerow K, Hepner JJ. Estimating traffic volume on Wyoming low volume roads using linear and logistic regression methods. J Traffic Transp Eng (English Ed [Internet]. 2016;3(6):493–506. Available from: http://dx.doi.org/10.1016/j.jtte.2016.02.004

12. Shojaeshafiei M, Doustmohammadi M, Subedi S, Anderson M. Comparison of Estimation Methodologies for Daily Traffic Count Pediction in Small and Medium Sized Communities. Int J Traffic Transp Eng. 2017;6(4):71–5.

13. Fu M, Kelly JA, Clinch JP. 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 [Internet]. 2017;58:186–95. Available from: http://dx.doi.org/10.1016/j.jtrangeo.2016.12.002

14. Department for Transport. Notes and definitions: Road Traffic Estimates [Internet]. 2016. Available from:

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/527807/road-traffic-notes-definitions.pdf

15. Department for Transport. All traffic data metadata [Internet]. 2018. Available from: https://www.gov.uk/government/publications/road-traffic-estimates-great-britain-jan-to-mar-q1-2014%5Cnhttps://www.gov.uk/government/collections/road-traffic-statistics

16. Sfyridis A, Agnolucci P. Annual average daily traffic estimation in England and Wales: An application of clustering and regression modelling. J Transp Geogr [Internet]. 2020;83(January):102658. Available from: https://doi.org/10.1016/j.jtrangeo.2020.102658

17. Gecchele G, Rossi R, Gastaldi M, Caprini A. Data Mining methods for Traffic monitoring data analysis: A case study. Procedia - Soc Behav Sci [Internet]. 2011;20:455–64. Available from: http://dx.doi.org/10.1016/j.sbspro.2011.08.052

18. Pun L, Zhao P, Liu X. A Multiple Regression Approach for Traffic Flow Estimation. IEEE Access. 2019;7:35998–6009.

19. Selby B, Kockelman KM. Spatial prediction of traffic levels in unmeasured locations: Applications of universal kriging and geographically weighted regression. J Transp Geogr [Internet]. 2013;29:24–32. Available from: http://dx.doi.org/10.1016/j.jtrangeo.2012.12.009

20. Brathwaite T, Vij A, Walker JL. Machine Learning Meets Microeconomics: The Case of Decision Trees and Discrete Choice. arXiv Prepr [Internet]. 2017;arXiv:1711(04826). Available from: http://arxiv.org/abs/1711.04826

21. Paredes M, Hemberg E, O'Reilly U-M, Zegras C. 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) [Internet]. IEEE; 2017. p. 780–5. Available from: http://ieeexplore.ieee.org/document/8005618/

22. Sekhar CR, Minal, Madhu E. Mode Choice Analysis Using Random Forrest Decision Trees. Transp Res Procedia [Internet]. 2016;17(December 2014):644–52. Available from: http://dx.doi.org/10.1016/j.trpro.2016.11.119

23. Adler P, Falk C, Friedler SA, Nix T, Rybeck G, Scheidegger C, et al. Auditing black-box models for indirect influence. Knowl Inf Syst. 2018;54(1):95–122.

24. Krause RW, Huisman M, Steglich C, Sniiders TAB. 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. 2018. p. 159–63.

25. Steindl G, Pfeiffer C. Comparison of Black Box Models for Load Profile Generation of District Heating Networks. In: 12th Sustainable Development of Energy, Water and Environment Systems [Internet]. 2017. p. 1–10. Available from: https://people.fh-burgenland.at/bitstream/20.500.11790/1017/1/Comparison of Black Box Models.pdf

26. Lakkaraju H, Kamar E, Caruana R, Leskovec J. Interpretable & Explorable Approximations of Black Box Models. In: arXiv preprint [Internet]. 2017. p. 1–5. Available from: http://arxiv.org/abs/1707.01154

27. Burns C, Thomason J, Tansey W. Interpreting Black Box Models via Hypothesis Testing. In: Proceedings of the 2020 ACM-IMS on Foundations of Data Science Conference [Internet]. 2020. p. 47–57. Available from: http://arxiv.org/abs/1904.00045

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

2017;44(2):368–74.

29. Zhang J, Li Z, Pu Z, Xu C. Comparing prediction performance for crash injury severity among various machine learning and statistical methods. IEEE Access. 2018;6:60079–87.

30. Eom JINKI, Man SIKP, Heo T-Y, Huntsinger LF. Improving the prediction of annual average daily traffic for nonfreeway facilities by applying a spatial statistical method. Transp Res Rec. 2006;(1968):22–9.

31. Wang X, Kockelman K. Forecasting Network Data. Transp Res Rec [Internet]. 2009;2105(2105):100–8. Available from: http://trrjournalonline.trb.org/doi/10.3141/2105-13

32. Kim S, Park D, Heo T-Y, Kim H, Hong D. Estimating Vehicle Miles Traveled (VMT) in Urban Areas Using Regression Kriging. J Adv Transp [Internet]. 2016; Available from: http://onlinelibrary.wiley.com/doi/10.1002/atr.144/full

33. Caceres N, Romero LM, Morales FJ, Reyes A, Benitez FG. Estimating traffic volumes on intercity road locations using roadway attributes , socioeconomic features and other work-related activity characteristics. Transportation (Amst). 2018;45(5):1449–73.

34. Mohamad D, Sinha K, Kuczek T, Scholer C. Annual Average Daily Traffic Prediction Model for County Roads. Transp Res Rec J Transp Res Board [Internet]. 1998;1617(98):69–77. Available from: http://trrjournalonline.trb.org/doi/10.3141/1617-10

35. Xia Q, Zhao F, Chen Z, Shen LD, Ospina D. Estimation of Annual Average Daily Traffic for Nonstate Roads in a Florida County. Transp Res Rec [Internet]. 1999 Jan 1;1660(1):32–40. Available from: http://journals.sagepub.com/doi/10.3141/1660-05

36. Seaver WL, Chatterjee A, Seaver ML. Estimation of Traffic Volume on Rural Local Roads. Transp Res Rec J Transp Res Board [Internet]. 2000 Jan 1;1719(1):121–8. Available from: http://journals.sagepub.com/doi/10.3141/1719-15

37. Zhao F, Chung S. Contributing Factors of Annual Average Daily Traffic in a Florida County: Exploration with Geographic Information System and Regression Models. Transp Res Rec [Internet]. 2001;1769(01):113–22. Available from: http://dx.doi.org/10.3141/1769-14

38. Zhao F, Park N. Using Geographically Weighted Regression Models to Estimate Annual Average Daily Traffic. Transp Res Rec [Internet]. 2004 Jan 1;1879(1):99–107. Available from: http://journals.sagepub.com/doi/10.3141/1879-12

39. Doustmohammadi M, Anderson M. Developing Direct Demand AADT Forecasting Models for Small and Medium Sized Urban Communities. Int J Traffic Transp Eng. 2016;5(2):27–31.

40. Chen P, Hu S, Shen Q, Lin H, Xie C. Estimating Traffic Volume for Local Streets with Imbalanced Data. Transp Res Rec. 2019;2673(3):598–610.

41. Diez D, Barr C, Cetinkaya-Rundel M. OpenIntro Statistics [Internet]. 3rd ed. Boston, MA: OpenIntro; 2012. 175 p. Available from: https://d1wqtxts1xzle7.cloudfront.net/34286417/statistics_open_book-with-cover-page-v2.pdf?Expires=1649001318&Signature=LqVNdMCdUmW5LQb6cMLHjRD6OeDHBy1kNld2KKQHyS79lVGqmAzV2A5yd4TuXXBtYMETE6E5HQRs9C7~ghcA8wsI5Y3~mgHzCa-aPOLw3uokNIbp7W8gKh2iES5PAmU-MwXF6i8

42. Zoico E, Rossi A, Di Francesco V, Sepe A, Olioso D, Pizzini F, et al. 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. 2010;65 A(3):295–9.

43. Troy A, Morgan Grove J, O'Neil-Dunne J. The relationship between tree canopy and crime rates across an urban-rural gradient in the greater Baltimore region. Landsc Urban Plan [Internet]. 2012;106(3):262–70. Available from: http://dx.doi.org/10.1016/j.landurbplan.2012.03.010

44. Tibshirani R. Regression Shrinkage and Selection via the Lasso. J R Stat Soc. 1996;58(1):267–88.

45. Musoro JZ, Zwinderman AH, Puhan MA, Ter Riet G, Geskus RB. Validation of prediction models based on lasso regression with multiply imputed data. BMC Med Res Methodol. 2014;14(1):1–13.

46. Chen X (Michael), Zhang S, Li L. Multi-model ensemble for short-term traffic flow prediction under normal and abnormal conditions. IET Intell Transp Syst [Internet]. 2019 Feb 19;13(2):260–8. Available from: https://onlinelibrary.wiley.com/doi/10.1049/iet-its.2018.5155

47. Al-Sahili O, Al-Deek H, Sandt A, Mantzaris A V., Rogers JH, Faruk MO. Investigating and Modeling the Illegal U-Turn Violations at Medians of Limited Access Facilities. Transp Res Rec J Transp Res Board [Internet]. 2018 Dec 21;2672(14):73–84. Available from: http://journals.sagepub.com/doi/10.1177/0361198118778941

48. Ogutu JO, Schulz-Streeck T, Piepho HP. Genomic selection using regularized linear regression models: ridge regression, lasso, elastic net and their extensions. In: In BMC proceedings BioMed Central [Internet]. 2012. p. 1–6. Available from: http://www.biomedcentral.com/1753-6561/6/S2/S10

49. Zou H, Hastie T. Regularization and variable selection via the elastic net. J R Stat Soc. 2005;134(3456):1080.

50. Hastie T, Tibshirani R, Friedman J. Elements of Statistical Learning 2nd ed. [Internet]. Vol. 27, Elements. 2009. 745 p. Available from: http://www-stat.stanford.edu/~tibs/book/preface.ps

51. Melkumova LE, Shatskikh SY. Comparing Ridge and LASSO estimators for data analysis. Procedia Eng [Internet]. 2017;201:746–55. Available from: https://doi.org/10.1016/j.proeng.2017.09.615

52. Yang Y, Zou H. A fast unified algorithm for solving group-lasso penalize learning problems. Stat Comput [Internet]. 2015;25(6):1129–41. Available from: http://dx.doi.org/10.1007/s11222-014-9498-5

53. Wong TT. Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. Pattern Recognit [Internet]. 2015;48(9):2839–46. Available from: http://dx.doi.org/10.1016/j.patcog.2015.03.009

54. Moreno-Torres JG, Saez JA, Herrera F. Study on the impact of partition-induced dataset shift on k-fold cross-validation. IEEE Trans Neural Networks Learn Syst. 2012;23(8):1304–12.

55. Yadav S, Shukla S. Analysis of k-Fold Cross-Validation over Hold-Out Validation on Colossal Datasets for Quality Classification. Proc - 6th Int Adv Comput Conf IACC 2016. 2016;(Cv):78–83.

56. Dancer D, Tremayne A. R-squared and prediction in regression with ordered quantitative response. J Appl Stat [Internet]. 2005 Jul;32(5):483–93. Available from: http://www.tandfonline.com/doi/abs/10.1080/02664760500079423

57. Department for Transport. Road Traffic Estimates Methodology Note [Internet]. London, UK; 2013. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/524848/annual-methodology-note.pdf

58. Department for Transport. Road Traffic Estimates [Internet]. London, UK; 2014. Available from: https://www.gov.uk/government/publications/road-traffic-estimates-great-britain-jan-to-mar-q1-2014%5Cnhttps://www.gov.uk/government/collections/road-traffic-statistics

59. Carli R, Dotoli M, Epicoco N, Angelico B, Vinciullo A. Automated evaluation of urban traffic congestion using bus as a probe. IEEE Int Conf Autom Sci Eng. 2015;2015-Octob:967–72.

60. Butcher L. Speed limits in England: Briefing Paper - Number CBP00468 [Internet]. London; 2017. Available from: http://researchbriefings.files.parliament.uk/documents/SN00468/SN00468.pdf

61. Gitelman V, Korchatov A, Carmel R. Safety-related behaviours of e-cyclists on urban streets: an observational study in Israel. Transp Res Procedia [Internet]. 2022;60(2021):609–16. Available from: https://doi.org/10.1016/j.trpro.2021.12.079

62. Pitaksringkarn J, Tanwanichkul L, Yamthale K. A correlation between pavement skid resistance and wet-pavement related accidents in Thailand. Hamontree C, editor. MATEC Web Conf [Internet]. 2018 Aug 14;192:02049. Available from: https://www.matec-conferences.org/10.1051/matecconf/201819202049

63. Gitelman V, Doveh E, Bekhor S. The Relationship between Free-Flow Travel Speeds, Infrastructure Characteristics and Accidents, on Single-Carriageway Roads. Transp Res Procedia [Internet]. 2017;25:2026–43. Available from: http://dx.doi.org/10.1016/j.trpro.2017.05.398

64. Gray RC, Quddus MA, Evans A. Injury severity analysis of accidents involving young male drivers in Great Britain. J Safety Res [Internet]. 2008;39(5):483–95. Available from: http://dx.doi.org/10.1016/j.jsr.2008.07.003

65. Šliupas T. The impact of road parameters and the surrounding area on traffic accidents. Transport. 2009;24(1):42–7.

66. Ale AS. Assessment of Dual Carriageway for Sustainable Socio-Economic Development in Ado Ekiti, Nigeria. Int J Sustain Dev World Policy [Internet]. 2018;7(1):27–36. Available from: https://archive.conscientiabeam.com/index.php/26/article/view/2056

67. Fountas G, Fonzone A, Gharavi N, Rye T. The joint effect of weather and lighting conditions on injury severities of single-vehicle accidents. Anal Methods Accid Res [Internet]. 2020 Sep;27:100124. Available from: https://doi.org/10.1016/j.amar.2020.100124

68. Sołowczuk AB, Kacprzak D. Effect of Lane Narrowing Implemented on an Urban Dual Carriageway on Speed Reduction and Acoustic Environment. Buildings. 2022;12(1).

69. Saez M, Barceló MA, Farrerons M, López-Casasnovas G. The association between exposure to environmental factors and the occurrence of attention-deficit/hyperactivity disorder (ADHD). A population-based retrospective cohort study. Environ Res [Internet]. 2018 Oct;166(May):205–14. Available from: https://doi.org/10.1016/j.envres.2018.05.009

70. London Councils. London Lorry Control Scheme. London, UK; 2017.

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

72. Van Zuylen H, Taale H. Urban Networks with Ring Roads. Transp Res Rec J Transp Res Board. 2004;1984:180–7.

73. Chhorn V, Prabnasak J, Chompoo-inwai C. 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: MATEC Web of Conferences. 2018. p. 192.

74. Nugmanova, Arndt, Hossain, Kim. Effectiveness of Ring Roads in Reducing Traffic Congestion in Cities for

Long Run: Big Almaty Ring Road Case Study. Sustainability [Internet]. 2019 Sep 11;11(18):4973. Available from: https://www.mdpi.com/2071-1050/11/18/4973

75. Variyam SP, Alex AP, Saraswathy MV. Trip Assignment Modelling for an Indian City to Assess the Benefits of Proposing Ring Roads. Int J TRAFFIC Transp Eng [Internet]. 2020 Feb 2;10(1):69–80. Available from: http://ijtte.com/uploads/2020-02-01/b2dd6d45-dea8-4a48ijtte.2020.10(1).07.pdf

76. Batty M, Bessusi E, Chin N. Traffic, Urban Growth and Suburban Sprawl [Internet]. London, UK; 2003. (Cities). Report No.: 70. Available from: https://www.researchgate.net/publication/32884861_Traffic_urban_growth_and_suburban_sprawl

77. Caselli M, de Gennaro G, Marzocca A, Trizio L, Tutino M. Assessment of the impact of the vehicular traffic on BTEX concentration in ring roads in urban areas of Bari (Italy). Chemosphere [Internet]. 2010;81(3):306–11. Available from: http://dx.doi.org/10.1016/j.chemosphere.2010.07.033

78. Lozhkin V, Gavkalyk B, Lozhkina O, Evtukov S, Ginzburg G. Monitoring of extreme air pollution on ring roads with PM2.5 soot particles considering their chemical composition (case study of Saint Petersburg). Transp Res Procedia [Internet]. 2020;50:381–8. Available from: https://doi.org/10.1016/j.trpro.2020.10.045

79. Baum-Snow N, Brandt L, Henderson JV, Turner MA, Zhang Q. Roads, Railroads, and Decentralization of Chinese Cities. Rev Econ Stat [Internet]. 2017 Jul;99(3):435–48. Available from: https://direct.mit.edu/rest/article/99/3/435-448/58455

80. Van Nes A. The Impact of the Ring Roads on the Location Pattern of Shops in Town and City Centres. A Space Syntax Approach. Sustainability [Internet]. 2021 Apr 1;13(7):3927. Available from: https://www.mdpi.com/2071-1050/13/7/3927

81. Cervero R, Kockelman K. Travel demand and the 3Ds: Density, diversity, and design. Transp Res Part D Transp Environ [Internet]. 1997;2(3):199–219. Available from: http://linkinghub.elsevier.com/retrieve/pii/S1361920997000096

82. Papagiannakis A, Baraklianos I, Spyridonidou A. Urban travel behaviour and household income in times of economic crisis: Challenges and perspectives for sustainable mobility. Transp Policy [Internet]. 2018;65(January 2017):51–60. Available from: https://doi.org/10.1016/j.tranpol.2016.12.006

83. Silva J de A, Morency C, Goulias KG. 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. 2012;46(8):1252–64.

84. Aditjandra PT, Cao X, Mulley C. Understanding neighbourhood design impact on travel behaviour: An application of structural equations model to a British metropolitan data. Transp Res Part A Policy Pract [Internet]. 2012;46(1):22–32. Available from: http://dx.doi.org/10.1016/j.tra.2011.09.001

85. Diao M, Ferreira J. Vehicle miles traveled and the built environment: Evidence from vehicle safety inspection data. Environ Plan A. 2014;46(12):2991–3009.

86. Power A. Social inequality, disadvantaged neighbourhoods and transport deprivation: an assessment of the historical influence of housing policies. J Transp Geogr [Internet]. 2012 Mar;21:39–48. Available from: http://dx.doi.org/10.1016/j.jtrangeo.2012.01.016

87. Amr A. Spatial inertia and separation in a hierarchy-driven structure: four case studies of flow continuity along ring roads. Urban, Plan Transp Res [Internet]. 2020 Jan 1;8(1):171–95. Available from: https://doi.org/10.1080/21650020.2020.1843527

88. Zhao P, Li S. Restraining transport inequality in growing cities: Can spatial planning play a role? Int J Sustain Transp [Internet]. 2016 Nov 25;10(10):947–59. Available from:

http://dx.doi.org/10.1080/15568318.2016.1191693

89. Saif MA, Zefreh MM, Torok A. Public Transport Accessibility: A Literature Review. Period Polytech Transp Eng [Internet]. 2018 May 3;47(1):36–43. Available from: https://pp.bme.hu/tr/article/view/12072

90. Owen D, Hogarth T, Green AE. Skills, transport and economic development: Evidence from a rural area in England. J Transp Geogr [Internet]. 2012;21:80–92. Available from: http://dx.doi.org/10.1016/j.jtrangeo.2012.01.015

91. Basagaña X, Triguero-Mas M, Agis D, Pérez N, Reche C, Alastuey A, et al. Effect of public transport strikes on air pollution levels in Barcelona (Spain). Sci Total Environ [Internet]. 2018;610–611:1076–82. Available from: https://doi.org/10.1016/j.scitotenv.2017.07.263

92. Bauernschuster S, Hener T, Rainer H. When labor disputes bring cities to a standstill: The impact of public transit strikes on traffic, accidents, air pollution, and health. Am Econ J Econ Policy. 2017;9(1):1–37.

93. Becker F, Axhausen KW. Literature review on surveys investigating the acceptance of automated vehicles. Transportation (Amst). 2017;44(6):1293–306.

94. Kwon SB, Namgung HG, Jeong W, Park D, Eom JK. Transient variation of aerosol size distribution in an underground subway station. Environ Monit Assess [Internet]. 2016;188(6). Available from: http://dx.doi.org/10.1007/s10661-016-5373-5

95. Seriani S, Fernández R. Planning guidelines for metro–bus interchanges by means of a pedestrian microsimulation model. Transp Plan Technol [Internet]. 2015;38(5):569–83. Available from: http://dx.doi.org/10.1080/03081060.2015.1039235

96. Kang L, Zhu X, Sun H, Wu J, Gao Z, Hu B. Last train timetabling optimization and bus bridging service management in urban railway transit networks. Omega (United Kingdom) [Internet]. 2019;84:31–44. Available from: https://doi.org/10.1016/j.omega.2018.04.003

97. Hong SP, Min YH, Park MJ, Kim KM, Oh SM. Precise estimation of connections of metro passengers from Smart Card data. Transportation (Amst) [Internet]. 2016;43(5):749–69. Available from: http://dx.doi.org/10.1007/s11116-015-9617-y

98. Bento A, Cropper ML, Mobarak AM, Vinha K. The Effects of Urban Spatial Structure on Travel Demand in the United States. Rev Econ Stat. 2005;87(3):466–78.

99. Friedman B, Gordon S, Peers J. Effect of Neotraditional Neighborhood Design on Travel Characteristics [Internet]. Transportation Research Record. Washington, DC: Transportation Research Board; 1994. 1–137 p. Available from: http://onlinepubs.trb.org/Onlinepubs/trr/1994/1466/1466-009.pdf

100. RAND. Motorcycles and Congestion : The Effect of Modal Split Full Summary, Department for Transport (with support from Transport for London) [Internet]. London, UK; 2004. Available from: http://mwigan.com/mrw/ewExternalFiles/04 0800 wsp dft tfl Motorcycles and congestion final summary report.pdf

101. Wong KI. An Analysis of Car and Motorcycle Ownership in Macao. Int J Sustain Transp. 2013;7(3):204–25.

102. Bibby P, Brindley P. 2011 Rural-Urban Classification of Local Authority Districts in England: User Guide [Internet]. 2014. Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/389780/RUCLAD2011_User_Guide.pdf

103. Hoerl AE, Kennard RW. Ridge Regression: Applications to Nonorthogonal Problems. Technometrics.

1970;12(1):69–82.

104. Hoogendoorn S, Knoop V. Traffic flow theory and modelling. In: The Transport System and Transport Policy: An Introduction [Internet]. Edward Elgar Publishing; 2012. p. 125–59. Available from: http://books.google.com/books?hl=en&lr=&id=_UwJWfgoOeEC&oi=fnd&pg=PA125&dq=Traffic+flow+theory+and+modelling&ots=4dq-GBQZ2I&sig=aobOGIRBMgKjhyafk6qlMhIGfCQ

105. Leduc G. Road Traffic Data : Collection Methods and Applications. EUR Number Tech Note JRC 47967 [Internet]. 2008;JRC 47967(January 2008):55. Available from: http://ftp.jrc.es/EURdoc/EURdoc/JRC47967.TN.pdf

106. Pang Y, Tsagatakis I, Murrells T, Brace S. Methodology and Changes made in the 2014 NAEI Road Transport Inventory: A Briefing Note Produced for DECC on Changes in Fuel Consumption [Internet]. 2016. Available from: 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

107. Roess R, Prassas E, McShane W. Summary for Policymakers. In: Intergovernmental Panel on Climate Change, editor. Climate Change 2013 - The Physical Science Basis [Internet]. Cambridge: Cambridge University Press; 2011. p. 1–30. Available from: http://books.google.com/books?id=fGE6PgAACAAJ&pgis=1

108. Sharma S, Lingras P, Xu F, Kilburn P. Application of Neural Networks to Estimate AADT on Low-Volume Roads. J Transp Eng [Internet]. 2001 Oct;127(5):426–32. Available from: http://ascelibrary.org/doi/10.1061/%28ASCE%290733-947X%282001%29127%3A5%28426%29

109. Pointer G. The UK's major urban areas. In: Chappell R, editor. Focus on People and Migration [Internet]. London: Palgrave Macmillan UK; 2005. p. 45–60. Available from: http://link.springer.com/10.1007/978-1-349-75096-2_3