Exploring distributions of car mileages: new insights into travel patterns using data from every private car registered in Great Britain

S. D. Ball
TRL
Crowthorne House, Nine Mile Ride,
Wokingham, Berks, RG40 3GA, UK
+44 1344 770391

S. Cairns
TRL / UCL
Crowthorne House, Nine Mile Ride,
Wokingham, Berks, RG40 3GA, UK

P. Emmerson
TRL
Crowthorne House, Nine Mile Ride,
Wokingham, Berks, RG40 3GA, UK

R. E. Wilson
University of Bristol
Dept. of Engineering Mathematics,
Woodland Rd, Bristol, BS8 1UB, UK

J. Anable
University of Leeds
Lifton Villas, 1-3 Lifton Place, Leeds,
LS2 9JT, UK

T. Chatterton
University of the West of England
Coldharbour Lane,
Bristol, UK, BS16 1QY, UK

ABSTRACT
Assessment of vehicle use is often based on looking at average mileages, and how those change. However, it is argued that transport policies, or other factors, may impact relatively subtly on different groups within populations, in ways which might be difficult to detect by looking at differences in, or changes to, averages. This paper reports on unique opportunities arising from a research project that is fusing and analysing several UK Government data sources together for new purposes – specifically, data collected for vehicle licensing and from regular vehicle inspections. In particular, the resultant dataset provides a unique opportunity to look at the distribution of mileages travelled by vehicles registered in different areas, and to understand how and why those distributions may differ, and which characteristics of those distributions may be most insightful for identifying and understanding the variation. This paper describes the first stages of analysis being undertaken at a range of spatial scales, calculating a range of statistics for the vehicle distributions found, and assessing the extent to which they differ, and what that might imply.

Keywords
Transport; car mileage distributions; vehicle inspection data; local transport evaluation; spatial analysis; car use.

1. INTRODUCTION

1.1 Overview

This paper describes the development of new techniques for analysing several datasets collected by the UK Government for other purposes, in order to provide insights into a range of different policy areas. Specifically, it reports on the fusing and interpretation of data collected during vehicle licensing, and during periodic vehicle inspection tests (the UK 'MOT' tests) in order to provide new insights into personal car ownership and use, and the related issues of energy use, air quality emissions, vehicle-related climate change contributions and the diffusion of new vehicle technologies. The work has been undertaken largely in a UK context. However, the techniques and processes developed would potentially be of relevance to all states or countries which collect similar types of information via a centralised computerised system, (for some exploration of this, see Cairns et al. 2014a, Diao & Ferreira 2014, Reardon et al. 2016, Chatterton and Anable 2016).

This ongoing work has been funded by the Research Councils UK Energy Programme (EP/K000438/1), and is formally supported by the UK Department of Transport and the UK Department of Energy and Climate Change. It is a collaborative project involving the Transport Research Laboratory, University of Leeds, University of Bristol, University of the West of England, UCL and the University of Aberdeen (www.MOTproject.net). As a case study, this paper focuses on one particular benefit of the dataset – namely the potential to look at variability in the mileages done by vehicles in a particular locality, rather than simply dealing with averages, based on a relatively comprehensive national dataset.

1.2 Background

Information about car use is typically gathered in three ways:

- Travel surveys – such as the UK National Travel Survey – where a sample of the population is asked about the amount that they personally travel during a given period (such as a week or a year), or for details of the use of each of their household vehicles.
- Measurement of vehicles ('traffic') on the road, using a variety of measures (from tube counters and inductive loops through to tracking of GPS devices), used to produce information such as the UK National Road Traffic Estimates.
- Modelling of traffic flows, often generated from underlying assumptions about the influence of socio-demographics and land-use patterns on likely trip rates, as, for example, occurs in the UK's National Transport Model.

All of these methods may be used at a range of spatial scales (including local and national policy making), and have particular strengths and weaknesses. For example, traffic counts are more informative about where traffic is experienced, whilst travel surveys are more helpful for understanding where personal car use is generated.

Weaknesses that they all share are that they require dedicated effort for their generation (which is often expensive), and the real data which they collect, or use, is rarely comprehensive, thereby requiring extrapolation to achieve population-level estimates, and meaning that undertaking analysis at a fine spatial scale is problematic.

Another issue is that the sample sizes achieved (and concerns about sampling methodology) mean that data reporting often focuses on 'average' behaviour. However, for some policy questions, there may be interest in the variability in behaviour across the population – such as whether the mileage generated by vehicles registered in an area is relatively evenly distributed across them or whether a small proportion of vehicles are responsible for a large share of the distance travelled. Such information could help inform how policies are formulated and targeted.

Understanding of vehicle distributions, and how they vary and change, could also be helpful for detecting potentially subtle policy impacts. For example, with car-share schemes (known as car clubs in the UK), it is recognised that zero or low-mileage drivers often increase their car use upon becoming a member, whilst households which reduce their car ownership after joining are likely to moderate their use – shifts which cannot always be readily detected through examination of averages (see, for example, Cairns and Harmer 2012).

Therefore, the data described in this paper is of particular interest, both given the potential to use information which is being collected anyway (without additional cost), and because its relative comprehensiveness makes it possible to examine vehicle mileage distributions, not simply 'average' behaviour.

2. DESCRIPTION OF DATA

2.1 Description of data sources and processes

The core dataset that has been analysed for this paper is drawn from two UK Government sources – the MOT vehicle inspection test records stored by the Driver and Vehicle Standards Agency, and details of all licensed vehicles held by the Driver and Vehicle Licensing Agency (two agencies of the UK Department of Transport).

The MOT records consist of the data collected during the annual vehicle inspection tests for all vehicles (up to 3.5 tonnes) of more than three years old in Great Britain[1]. A computerised system for this information was introduced in 2005, with full implementation by April 2006, and public release of an anonymised version of this information began in 2010. About 35 million tests are recorded every year, meaning that over 300 million records are being analysed for this project. Data are organised by test, and include details of the nature of the vehicle (date of first use, make, model, colour and engine size), an odometer mileage reading, together with details of the test (date, the area where the testing station is located and test results). There are related tables providing more details of test failures, although those have not been utilised in this project.

[1] Note that data from Northern Ireland is excluded.

The vehicle licensing data is, instead, organised by vehicle, and includes details of about 55 million vehicles from 2003 onwards. Unlike the MOT records (where test dates are vehicle dependent), data are reported at regular, quarterly intervals[2]. For this project, key details provided (which are not available from the MOT records) include the location of the registered vehicle keeper, an indication of whether the vehicle is in personal or commercial ownership, and the vehicle's CO_2 rating (according to the manufacturer).

Key activities which have been undertaken to generate usable information include:

- Hosting the data appropriately, specifically on a stand-alone machine held in a secure environment, with access only via a virtual private network, and with data downloads only permitted after independent verification that they do not breach the data protection requirements.
- Fusing and re-organising the data into a master table, which is organised by vehicle, but includes quarterly information for fields that change over time.
- Reconciling inconsistencies between data gathered at different MOT tests that should theoretically be unchanging (such as vehicle engine size).
- Generating an estimate of the mileage travelled by each vehicle for each quarter.

The generation of mileage estimates for individual vehicles has been a major analytical task of the project, and is reported in more detail elsewhere (see, for example, early work on this topic by Wilson et al. 2013a, 2013b, Cairns et al. 2014b). Particular challenges include the fact that although roughly annual, test intervals vary; there are erroneous readings recorded (due to both accidental and deliberate error); there will be variation in vehicle use between test intervals (not least due to changes in the weather and associated activities – e.g. the traditional British summer holiday); there are seasonal patterns of car purchasing (which generate an alternative set of seasonal impacts); and, finally, most vehicles do not provide a mileage until they are at least three years old. Processes for dealing with these range from complex heuristics to identify erroneous values, through to simple assumptions (e.g. that mileage increases linearly from zero to the three year value). Given seasonal issues and the processes used, the mileage data is only considered to be robust for annual estimates. (Quarterly data is generated to enable the specification of different annual periods, depending, for example, on whether calendar or financial year figures are of most value). As the project progresses, mileage estimation processes will be refined further, enabling sensitivity testing of initial conclusions.

2.2 Applications of the data

The nature of the final dataset created enables a wide variety of practical applications. In particular, vehicle keeper location information is provided at a relatively fine geographical scale. In England and Wales, vehicles are linked to their 'Lower-layer Super Output Area' (LSOA) – units comprising an average of about 700 households, of which there are 34,753; whilst in Scotland, they are linked to 6,976 'Data Zones', which are slightly smaller in population. This enables direct linking with area-based

[2] For 2003-2008, data are available for Q4 only. From 2009 onwards, data are available for all four quarters. Reporting occurs at the end of the quarter – e.g. Q1 data are for 31st March.

UK Census information about socio-demographic characteristics, as well as a range of other spatially disaggregate datasets (for example area-based information about average household income and public transport accessibility).

Consequently, activities undertaken on the 'LSOA dataset' by the project team have included: exploring the interrelationships between the variables in the core dataset at different spatial levels (Chatterton et al. 2015); understanding the links between energy use by private cars, and household energy and gas use (Chatterton et al. 2016a); exploring spatial variation in where people experience air pollution as opposed to where those creating vehicle pollution live (Barnes et al. 2014); exploring how variation in motoring costs and pollution vary with income (Chatterton et al. 2016b); and undertaking detailed spatial modelling of the determinants of car ownership and use (Yeboah et al. 2016, Emmerson et al. 2016).

Meanwhile, this paper works with the individual vehicle dataset, to understand intra-zonal variability, and whether such understanding would contribute to our ability to detect and understand transport policy impacts.

3. METHODOLOGY

The research questions specifically being examined in the work reported in this paper are as follows:

- For various measures that can be used to describe the distribution of mileages in a spatial area, how much does each one vary? (and is the variability different at different spatial scales?)
- Is it the case that looking at variations in mean mileage is likely to be as informative as using any other measure? If not, what would be the recommended set of distributional measures that could be used in future analysis to help understand similarities and differences between areas in terms of patterns of vehicle use?

The work described in this paper is being undertaken at five spatial scales using data for Great Britain:

- National (1 unit: Great Britain, excluding Northern Ireland)
- Regions (11 units: North East; North West; Yorkshire; East Midlands; West Midlands; East of England; London; South East; South West; Wales; Scotland)
- Local authority districts / unitary authorities (380 units: 326 in England; 22 in Wales; 32 in Scotland)
- Middle-layer Super Output Areas (MSOAs) (8,480 units: 6,791 in England; 410 in Wales; 1,279 Intermediate Geography Zones in Scotland; typically 2000-4500 households per MSOA and 1000-2500 households per IGZ); and
- Lower-layer Super Output Areas (LSOAs) (41,729 units: 32,844 in England; 1,909 LSOAs in Wales; 6,976 Data Zones in Scotland; typically 500-800 households per LSOA and 200-500 households per DZ)

(MSOAs and LSOAs are areal units defined for the purposes of the UK Census data collection[3].).

In all cases, we have:

- The number of vehicles present in that unit as of 30th June 2011
- The number of vehicles present in that unit as of 30th June 2011 for which a valid mileage reading can be inferred for the period 1st January – 31st December 2011
- A mileage reading for each of those vehicles
- The type of registered keeper: 'Private'; 'Commercial; and 'Between Keepers'.

Note that for this study, we have filtered vehicles to be those in MOT vehicle classes 4 and 4A[4] in private ownership (which are mainly cars).

For each unit, we calculate a range of descriptive statistics about the distribution of vehicle mileages, and to explore how those vary. This is done both on the whole data set, and on particular sub categories of the data (for example, areas from different rural-urban categories; areas with high or low car ownership; areas with particular public transport accessibility characteristics etc.).

At the time of writing, fine-tuning of the master dataset is still taking place, meaning that some of the material in the results section will be modified for the final version. Analysis that has been undertaken is also largely illustrative of the larger data analytics task that will take place over the coming months.

Specifically, this paper reports on:

- Evidence about the variability in average mileages available through this data set and how this changes with spatial scale and urbanization (see Section 4.1)
- Illustrative information about how mileage distributions differ between urban and rural areas, and the range of descriptive statistics that could be used to summarise this (see Section 4.2)
- Comparative information for one of these variables (the proportion of vehicles doing more than 12,000 miles p.a.) which can be considered in relation to the averages analysis reported in section 4.1 (see Section 4.3)
- An initial exploration of how the two variables described above (mean mileage, and the proportion of vehicles doing more than 12,000 miles p.a.) for five sample area types, defined according to levels of car ownership (see Section 4.4)

4. RESULTS

4.1 Differences in mean annual mileage per vehicle

In order to provide the context for later analysis, this section explores variation in average mileages per vehicle and how this changes with spatial scale and level of urbanisation[5] (average household car ownership is not considered at this stage but explored further in Section 4.4.) Specifically, this section presents results for the mean annual mileage per vehicle (in thousands of miles) for the period Jan 2011 to Dec 2011. The following findings emerge:

- Figure 1 shows how mean annual mileage varies for the 11 British regions (top left); 380 local authority districts (top right), 8,480 MSOAs (bottom left) and 41,729

[3] http://www.ons.gov.uk/ons/guide-method/geography/beginner-s-guide/census/super-output-areas--soas-/index.html

[4] https://www.gov.uk/getting-an-mot/mot-test-fees

[5] http://www.ons.gov.uk/ons/guide-method/geography/products/area-classifications/2011-rural-urban/index.html

LSOAs (bottom right). This highlights that London appears to be qualitatively different to most of the other British regions. (A non-trivial difference in average mileage remains, even when it is compared with other major urban conurbations, although that plot is not given here).

- Figure 2 provides a spatial plot of results whilst Figure 3 provides the average MSOA values for groups of MSOAs – specifically, once they have been categorised into a national urban-rural classification generated by the UK Office for National Statistics. This highlights both the importance of urbanisation – a well-known effect that is often captured indirectly in transport models via a population density variable – but, also, potentially another set of factors, which may relate to the relative isolation/self-sufficiency of the locations. Specifically, in Figure 2, both the Isle of Wight, a large island off the south coast of England, and some of the more isolated coastal towns appear to have relatively low average mileages (although the age profile of residents may also be significant here). Meanwhile, in Figure 3, it is notable that setting appears to affect average mileages. Specifically, vehicles located in 'urban cities and towns located within relatively sparsely populated areas' (C2) have lower average mileages than other urban cities and towns (C1). Similarly vehicles in rural towns in sparse settings (D2) have lower mileages than other rural towns (D1).
- Table 1 then provides a summary of the variability in results at the different spatial levels. Figure 1 and Table 1 indicate that, as area size reduces, the range of the average mileages done in different areas increases (as would probably be expected), whilst the distribution of observed values becomes increasingly normalised. The discrepancy between the local authority district and the MSOA level plots, compared with the similarity of the MSOA and LSOA plots, is noticeable.

More broadly, it is clear that consideration of average areal mileages does provide a number of important transport insights, with variability between areas at all spatial scales. This analysis is compared with equivalent analysis undertaken using a different distributional measure in Section 4.3.

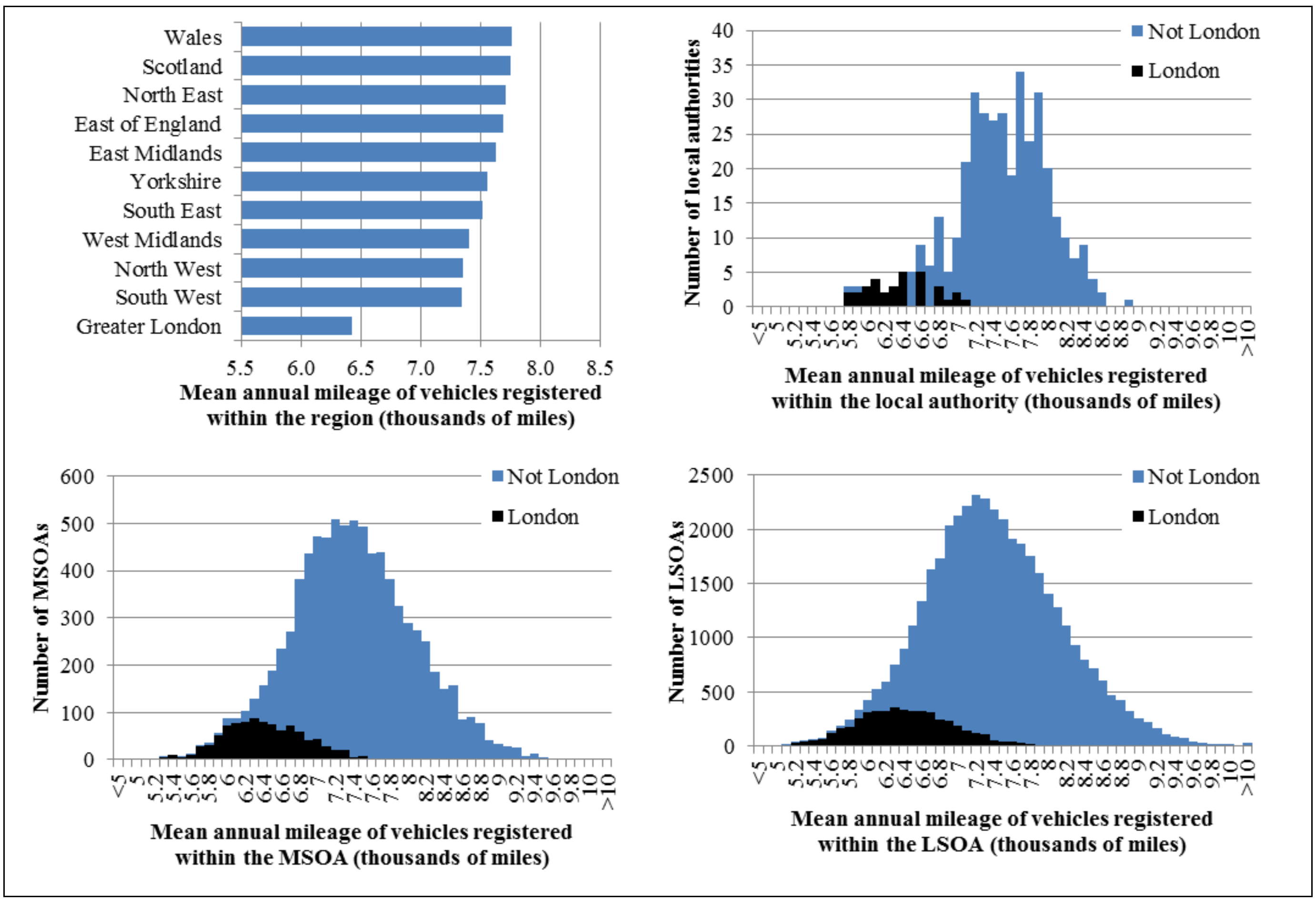

Figure 1: Mean annual mileage per vehicle for different spatial scales (Great Britain, 2011)

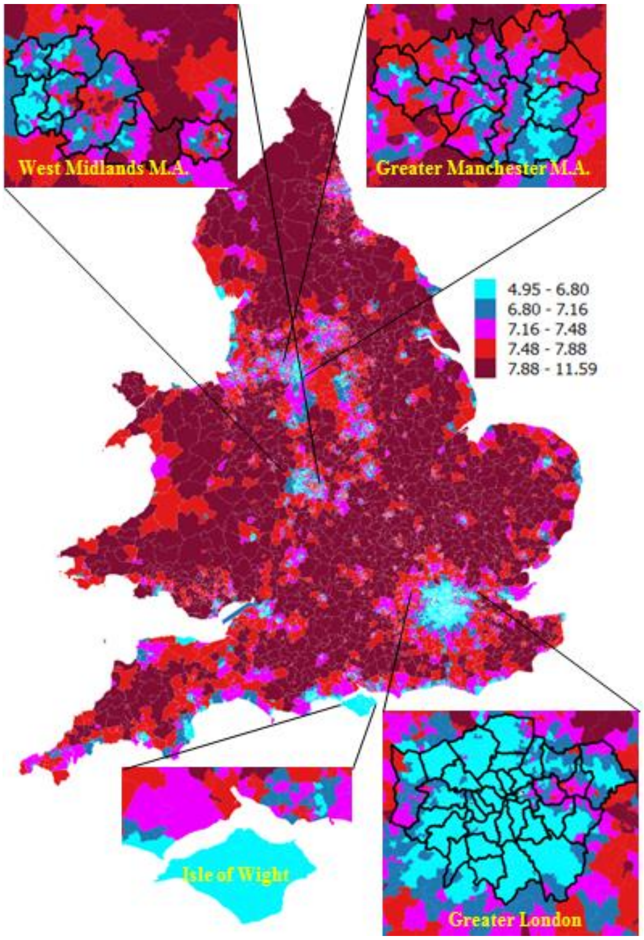

Figure 2: Mean annual mileage per vehicle (England and Wales, MSOA-level, 2011)

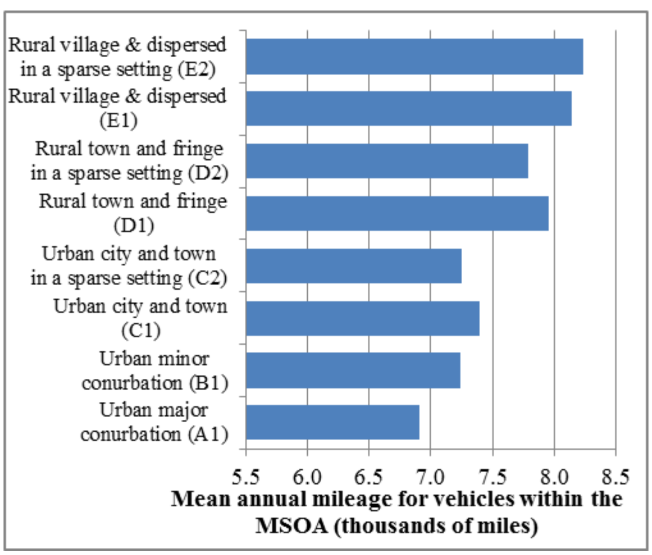

Figure 3: Mean annual mileage per vehicle by ONS rural-urban classification (England and Wales, 2011)

Table 1: Variability in mean annual mileage per vehicle at different spatial scales (Great Britain, 2011)

	Number of units	Mean annual mileage per vehicle		% diff*
		Min	Max	
Region	11	6,428	7,757	21%
Local authority (District and UA)	380	5,793	8,879	53%
MSOA	8,480	4,951	11,586	134%
LSOA	41,729	4,517	14,972	231%

*% difference is calculated as the range divided by the minimum

4.2 Comparison of mileage distributions

The previous section has considered inter-regional variability in average values. In contrast, this section investigates intra-regional variability – specifically, the distribution of individual vehicle mileages within spatial areas. As an exemplar, the individual vehicle data has been filtered for two large areas that are known to have different socio-demographic characteristics: a predominantly urban area with the West Midlands metropolitan area[6] and a predominantly rural area with Wales excluding Cardiff and Swansea. In mid-2011, the total numbers of privately registered cars within these areas were 972,998 and 1,211,505, respectively; valid mileage readings were available for 761,059 (78.2%) and 933,133 (77.0%), respectively. The distributions of the annual mileages per vehicle for these two areas are shown in Figure 4. Meanwhile, Table 2 indicates a number of different statistical measures that can be used to define the two distributions, and Figure 5 provides a breakdown of the different areas within the West Midlands metropolitan area.

The first finding of note is that both the distributions in Figure 4 could be represented by a Gamma distribution or potentially a Generalised Extreme Value (GEV) distribution, and our exploration of the data for other areas has suggested that this is typical (at least as far as local authority district scale).

The second significant finding is that many of the measures that could be used to describe the distributions other than the mean – for example, the distribution's kurtosis ('peakedness'), its skewness or the extent of the tail (i.e. vehicles being driven more than 12,000 miles), all suggest a significantly greater difference between the two areas than a simple comparison of the mean values would suggest.

Meanwhile, Figure 5 illustrates that within one area there will then be further variability, albeit potentially quite subtle in nature.

These results support the rationale for undertaking this work – namely the hypothesis that use of distributional statistics other than the mean may be more sensitive at detecting differences between areas, and therefore may potentially be more useful if trying to detect the impacts of particular local policies

[6] Note that the West Midlands 'metropolitan area' is the main urban conurbation of Birmingham, Wolverhampton and Coventry and is a subset of the wider West Midlands 'Region'

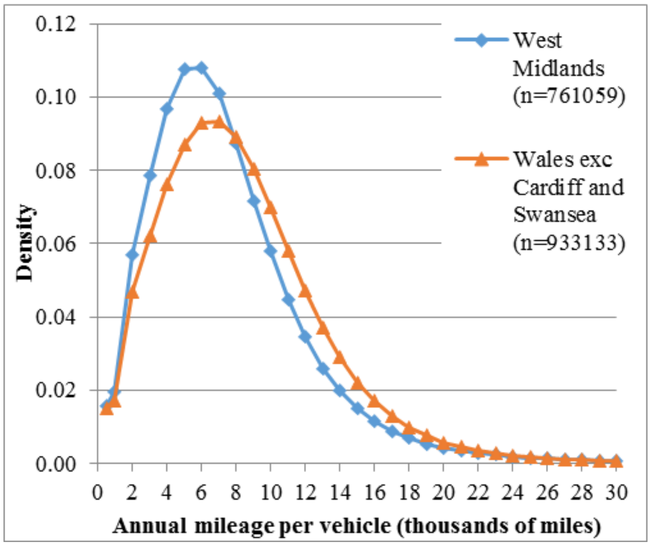

Figure 4: Distribution of average annual mileage per vehicle (West Midlands metropolitan area and Wales, 2011)

Table 2: Statistical measures of distributions given in Figure 4

Measure	West Midlands	Wales exc Cardiff/Swansea	Diff	Diff (%)
Mean	6,886	7,425	-539	8%
Median	5,900	6,581	-681	11%
Standard deviation	4,792	4,978	-186	4%
Kurtosis	2.81	1.52	1.30	60%
Skewness	1.43	1.08	0.35	28%
Coeff of variation	0.70	0.67	0.03	4%
% of vehicles >12k miles	12.0%	16.4%	-4.5%	32%
% of vehicles >15k miles	5.8%	7.6%	-1.8%	26%

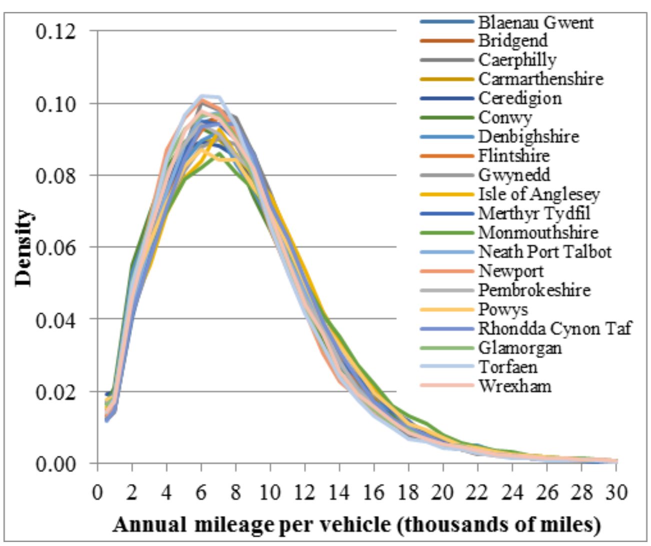

Figure 5: Distribution of average annual mileage per vehicle (Local authority disticts within Wales, 2011)

4.3 Indicators for the tail of the distribution

As a further test of the contention given in Section 4.2, an alternative distributional measure – the proportion of vehicles travelling more than 12,000 miles p.a. – has been chosen, and used for the same analysis as that conducted in Section 4.1 using mean data. (12,000 miles has been chosen based on the mileage divisions used in the UK National Travel Survey – though there are valid arguments as to why a range of different potential thresholds might be the most informative.) A brief summary of the results is given in Figure 6 and Table 3. For brevity, 'the proportion of vehicles travelling more than 12,000 miles' is referred to as the 'high mileage share' in the points below.

Key findings that emerge are:

- As with the mean data, the high mileage share identifies London to be an outlier, in terms of travel behaviour (Figure 6).
- The distributions of the high mileage share values are more asymmetric than those for the mean values, with a tail of areas which have relatively high proportions of high mileage vehicles (Figure 6).
- The degree of variability between areas is considerably greater when compared using high mileage share values (Table 3), as opposed to the mean values. For example, at MSOA level, mean values vary by a factor of three, whilst the high mileage share values vary by a factor of six.

Overall, the implication is that the two measures are likely to provide somewhat different insights into the variability between areas.

Table 3: Variability in proportion of vehicles within an area that travel over 12,000 miles per year at different spatial scales (Great Britain, 2011)

	Number of units	% of vehicles over 12,000 miles per year		Range	% diff
		Min	Max		
Region	11	9.6%	15.9%	6.3%	66%
Local authority (District and UA)	380	6.1%	23.2%	17.1%	180%
MSOA	8,480	4.3%	33.0%	28.7%	567%
LSOA	41,729	0%	46.0%	46.0%	N/A

*% difference is calculated as the range divided by the minimum

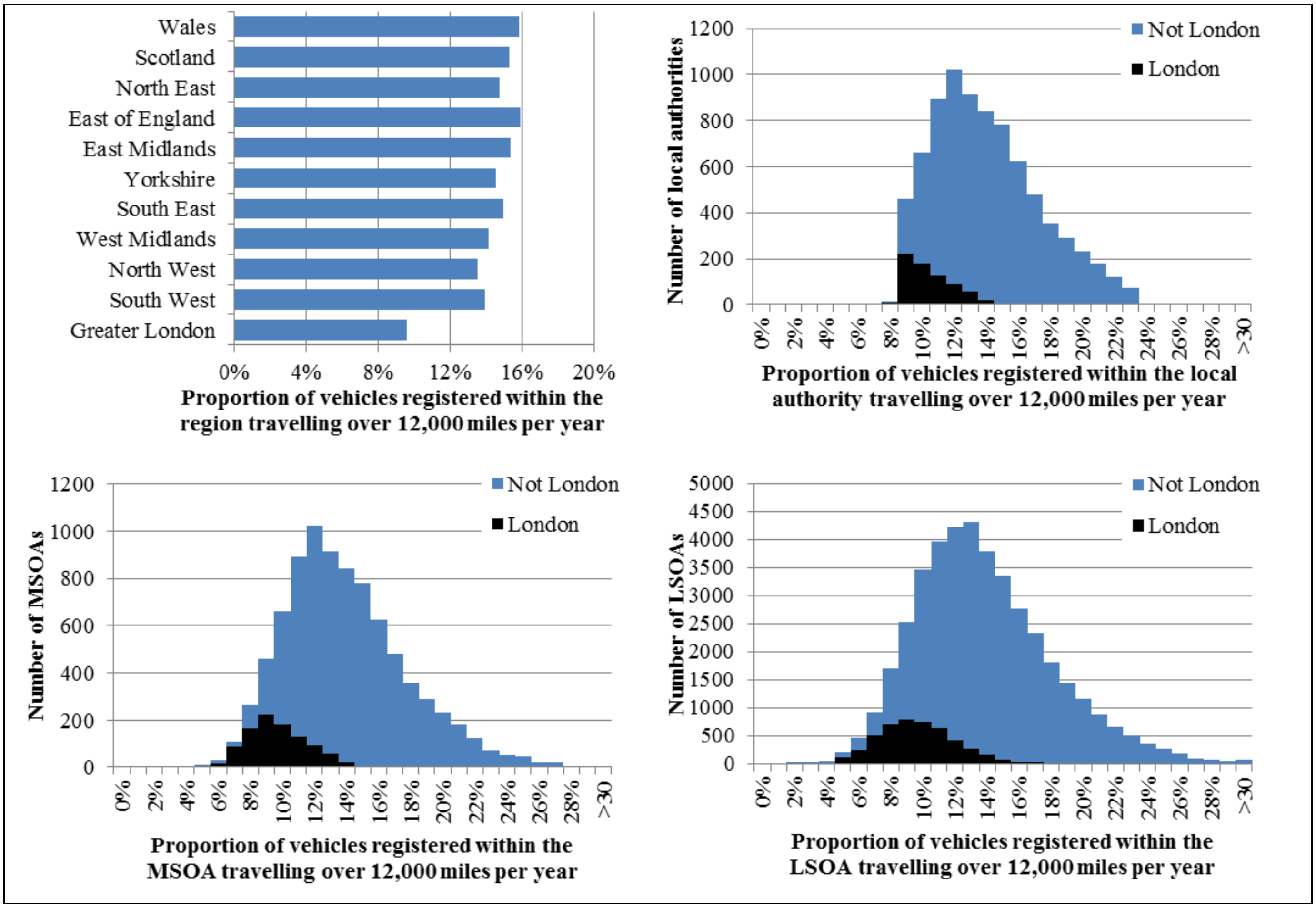

Figure 6: The proportion of vehicles within an area that travel over 12,000 miles per year, for different spatial scales (GB, 2011)

4.4 Combining the car usage data with car ownership data

The annual vehicle mileage data from the MOT vehicle inspection test data has been combined with information from the UK 2011 Census to investigate interactive effects of car usage and car ownership. The UK Census includes the following question, "*How many cars/vans are owned or available for use, by members of this household? (You should include company cars/vans available for private use)*". This question can be used to generate two key measures: the proportion of households with access to at least one car; and the average number of cars per household for those households that have at least one car. The bivariate plot in Figure 7 shows that there is a strong positive relationship between these two measures at the MSOA-level.

In order to undertake some illustrative analysis, five area types have been defined, as marked and labelled on the figure. Thresholds have been chosen based on a brief observation of the data in order to generate some area types where car driving might be expected to be different based on car ownership characteristics. Later work may attempt to define clusters of areas using more sophisticated techniques. The types of areas used here are:

- I – Low proportion of households with at least one car (<40%) and low average number of cars per household (<1.3);
- II – Medium proportion of households with at least one car (60-70%) and low average number of cars per household (<1.3);
- III – Medium proportion of households with at least one car (60-70%) and medium average number of cars per household (1.5-1.7);
- IV – High proportion of households with at least one car (>90%) medium average number of cars per household (1.5-1.7);
- V – High proportion of households with at least one car (>90%) and high average number of cars per household (>2).

The same data is also presented as a stacked bar chart in Figure 8; the benefit of this alternative layout is that it shows the number of MSOAs within each category (it is difficult to achieve this using Figure 7 given the relatively large number of data points that are superimposed on each other.)

Several other data sets about MSOA characteristics have also been linked in this analysis – specifically, the ONS rural-urban classification (already defined in Section 4.1) and ONS data on total mean household incomes[7]. Interestingly, all five area types are relatively distinct, both in terms of how urbanised they are, and in terms of average incomes. Summary information is given in Table 4.

[7] http://www.ons.gov.uk/ons/datasets-and-tables

Table 4: Details of five illustrative area types (differentiated by their car ownership characteristics)

Group	**I**	**II**	**III**	**IV**	**V**
% of HH with at least one car	**Low (<40%)**	**Medium (60-70%)**	**Medium (60-70%)**	**High (>90%)**	**High (>90%)**
Avg num of cars per HH	**Low (<1.3)**	**Low (<1.3)**	**Medium (1.5-1.7)**	**Medium (1.5-1.7)**	**High (>2)**
Num of MSOAs	230	48	56	73	58
Main ONS rural-urban classification	Urban major conurbation (A1)	Urban major conurbation (A1) & Urban city/town (C1)	Urban major conurbation (A1) & Urban city/town (C1)	Urban city/town (C1)	Rural village and dispersed (E1)
Main mean HH income	Very low	Very high	Low	High	Very high
Mean annual vehicle mileage (selected MSOAs)					
Mean of the values	6.6	6.6	7.2	7.8	8.0
Max	11.6	8.2	8.5	9.7	9.2
Min	5.4	5.6	5.4	6.4	6.6
Range	6.2	2.6	3.1	3.3	2.6
Coefficient of variation	0.11	0.10	0.07	0.08	0.07
Proportion of vehicles travelling over 12,000 miles per year (selected MSOAs)					
Mean of the values	11.0%	10.5%	12.7%	15.7%	18.7%
Max	32.7%	18.2%	21.2%	27.0%	26.6%
Min	6.5%	6.1%	6.5%	9.3%	10.0%
Range	26.1%	12.1%	14.7%	17.8%	16.6%
Coefficient of variation	0.27	0.28	0.17	0.23	0.19

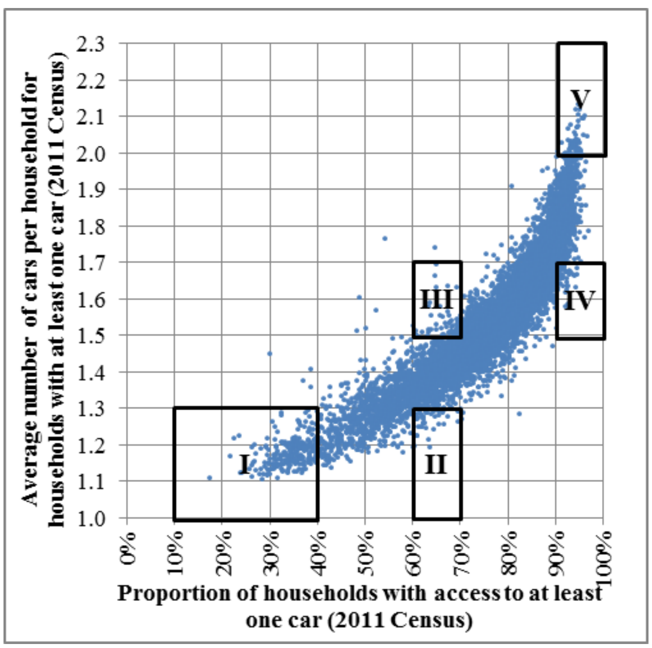

Figure 7: Bivariate plot of average number of cars per household and the proportion of households with access to at least one car (UK 2011 Census, England and Wales, MSOA)

The purpose of identifying the five groups is to investigate the measures of car usage discussed earlier in the paper and how these help to differentiate vehicle use for areas of varying levels of car ownership. Results from this analysis are also given in Table 4.

From the values given, it is clear that information on mean annual mileages (per vehicle) does provide some differentiation of the groups, although, for example, it is relatively poor at differentiating between Group I (urban areas with low car ownership and low incomes); and Group II (urban areas where incomes are high and far more households own cars, although the number of cars per household is relatively low). Equally, it shows relatively little difference between Group IV (urban cities and towns where lots of households own cars but the number of cars per household is relatively average), and Group V (high earning rural communities where lots of households own lots of cars).

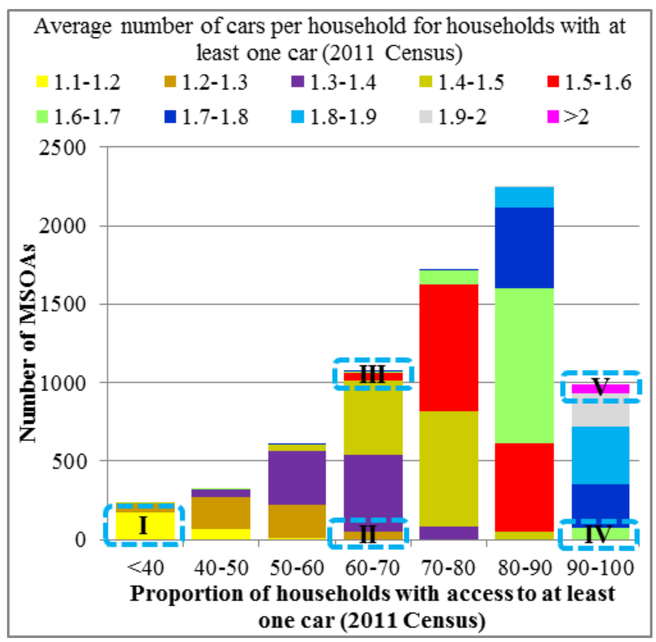

Figure 8: Stacked bar chart of average number of cars per household and the proportion of households with access to at least one car (UK 2011 Census, England and Wales, MSOA)

In contrast, information about the proportion of high mileage vehicles makes greater distinction between the different groups. As an example, comparing Group V (highest car ownership and highest car usage per vehicle) against Group I (lowest car ownership and lowest car usage per vehicle), mean annual mileage as a metric could be used to suggest that "On average Group V has 22% more car usage per vehicle compared to Group I (8.0 to 6.5 thousand miles per year). In comparing the same groups, the proportion of high mileage vehicles metric could be used to suggest that i.e. "Group V has 69% more high mileage vehicles compared to Group I (18.7% doing over 12,000 miles per year, compared to 11.0%). This is not to say that the high mileage metric is necessarily 'better' than average mileage; rather, that it (along with other metrics not yet explored) may be a useful complementary method to better understand travel patterns.

5. CONCLUSIONS

5.1 Summary of work to date

The purpose of this paper has been to describe the generation of a new large-scale dataset that can be used to examine vehicle mileages. Key findings to date include:

- Mean mileages do provide differentiation between areas, and readily demonstrate the importance of factors such as urbanisation.
- Comparison of mileage *distributions* for two disparate areas shows that measures other than the mean mileage typically vary more between areas than the mean does, and may therefore be a more sensitive way of detecting areal variation.
- Exploration of one alternative metric – the proportion of vehicles driven more than 12,000 miles a year – does show considerably greater variation between areas than the mean data does.
- If five distinct area types are considered (as defined according to car ownership characteristics), a metric such as the proportion of vehicles driven more than 12,000 miles a year appears to show greater variation than use of mean values. Given that the areas represented appear to be distinctly different (according to associated income and urbanisation characteristics), this supports the contention that measures other than the mean may be useful if trying to understand how areas vary, or change, in terms of car driving behaviour.

5.2 Further work

This paper reports on work that is currently in progress. Further activities will include investigation of a greater range of descriptive statistics for the mileage distributions (for example, their skewness and kurtosis; goodness of fit to a theoretical shifted gamma distribution; and some measure of the extent to which mileage for vehicles based in any given area is concentrated or spread across the vehicles within it, possibly using some kind of adapted Gini coefficient). It will also explore how these measures could be used in conjunction with population or household size information, given that understanding mileage per person or per household is usually more useful than mileage per vehicle.

The intention is also to look at how distributions vary for areas depending on other key transport criteria such as the provision of bus and rail transit options, and the proximity of key services and facilities. The aim is to identify the best ways in which mileages in any given area can be summarised, and to decide which descriptive statistics are likely to be the most meaningful in terms of attempting to detect differences which may be due to local transport policies.

Another key issue for local transport policy evaluation is to define meaningful benchmark or control areas, which can provide some insight into 'what would have happened without intervention', and the intention is that this study may also help to provide some criteria by which these could be identified.

Finally, it is hoped that, as part of the larger project, the work will be of international relevance, given that computerised recording of vehicle test inspection data is likely to become increasingly ubiquitous, thereby meaning that the analytical methods and techniques developed in this study may be useful elsewhere.

6. ACKNOWLEDGMENTS

The work has been undertaken as part of the MOT project (EP/K000438/1), www.MOTproject.net, funded by the UK Engineering and Physical Sciences Research Council under the Research Councils UK Energy Programme. Grateful thanks to members of DfT, DVSA, DVLA and DECC, who have provided advice and support for this work, and to other members of the project team: Oliver Turnbull, Jo Barnes, Godwin Yeboah, John Lees-Miller, Simon Notley and Fraser McLeod. Census data used is taken from National Statistics data © Crown copyright and database right 2012.

7. REFERENCES

[1] Barnes, J., Chatterton, T., Anable, J., Wilson, R.E., and Cairns, S. 2014. *Air pollution and energy use in England and Wales - A social and environmental justice analysis*. RGS-IGB Annual International Conference - Geographies of Co-Production, Kensington, London, 26-29 August 2014.

[2] Cairns, S., Rahman, S., Anable, J., Chatterton, T., Wilson, R.E. 2014a. *Vehicle Inspections – From Safety Device to Climate Change Tool*. MOT Project Working Paper

[3] Cairns, S., Wilson, R.E., Chatterton, T., Anable, J., Notley, S., and McLeod, F. 2014b. *Using MOT test data to analyse travel behaviour change - scoping report*. ISBN:978-1-910377-01-7.

[4] Cairns, S. and Harmer, C. 2012. *The emission impacts of car clubs in London*. TRL report PPR591, ISBN:978-1-908855-01-5.

[5] Chatterton, T., Barnes, J., Wilson, R.E., Anable, J., and Cairns, S. 2015. *Use of a novel dataset to explore spatial and social variations in car type, size, usage and emissions. Transportation Research D: Transport and* Environment, 39, pp.151-164.

[6] Chatterton, T., Barnes, J., Anable, J., and Yeboah, G. 2016a. *Mapping household direct energy consumption to provide a new perspective on energy justice*. Energy Research and Social Science (Special Edition: Justice, mobility & energy).

[7] Chatterton, T., Anable, J., Cairns, S., Wilson, R.E. and Yeboah, G. 2016b. *Financial Implications of Car Ownership and Use: A Social and Spatial Distributional Analysis*. University Transport Study Group (UTSG) 48th Annual Conference, Bristol, January 2016.

[8] Chatterton, T. and Anable, J. 2016. *Vehicle Odometers and Other Novel Methods of Examining Car Ownership and*

Usage American Association of Geographers Annual Meeting. San Francisco, 1/4/16.

[9] Diao, M. and Ferreira, Jr. J. 2014. Vehicle miles traveled and the built environment: evidence from vehicle safety inspection data *Environment and Planning A*, volume 46, pages 2991 – 3009, DOI=10.1068/a140039p.

[10] Emmerson, P., Anable, J., Chatterton. T., Cairns, S., Wilson., R.E. and Ball, S. 2016. *Analysing MOT/vehicle licensing data and transport model data to generate insights about car use and emissions in Strathclyde*. Scottish Transport Applications And Research (Star) Conference, 18/5/16, Glasgow.

[11] Reardon, T., Irvin, E., Brunton, S., Hari, M., Reim, P., and Gillingham, K. 2016. *Quantifying vehicle miles traveled from motor vehicle 1 inspection data: the Massachusetts Vehicle Census*. TRB 95th Annual Meeting 10-14/1/16.

[12] Wilson, R. E., Cairns, S., Notley, S., Anable, J., Chatterton, T., and McLeod, F. 2013a. *Techniques for the inference of mileage rates from MOT data*. Transportation Planning and Technology. 36 (1), pp.130-143.

[13] Wilson, R.E., Anable, J., Cairns, S., Chatterton, T., Notley, S. and Lees-Miller, J.D. 2013b. *On the estimation of temporal mileage rates*. Transportation Research Part E, 60, 26–139.

[14] Yeboah, G., Anable, A., Emmerson, P., Chatterton, T., Cairns, S., Ball, S. and Wilson, R.E. 2016. *Exploring variance in car ownership levels at the local level: beyond income and density*. Universities Transport Studies Group, Bristol, 6/1/16.