Understanding variation in car use: exploration of statistical metrics at differing spatial scales using data from every private car registered in Great Britain

S. D. Ball
Transport Research Laboratory (TRL)
Wokingham, UK

S. Cairns
Transport Research Laboratory (TRL) / UCL
Wokingham, UK / London, UK

P. Emmerson
Transport Research Laboratory (TRL)
Wokingham, UK

R. E. Wilson
University of Bristol
Bristol, UK

J. Anable
University of Leeds
Leeds, UK

T. Chatterton
University of the West of England
Bristol, UK

***Abstract*— Assessment of vehicle use is often based on looking at average mileages, and how those vary. 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 analysis undertaken at a range of spatial scales and the calculation of a range of statistics to describe the vehicle distributions, 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

I. INTRODUCTION

A. 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 [2], [8], [9] and [11]).

This ongoing work is a collaborative project led by the University of Leeds, and involves the Transport Research Laboratory, University of the West of England, University of Bristol, 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.

Innovative features of the work include the techniques developed to rework a national dataset which is relatively comprehensive, but collected for other reasons; analysis of the core dataset in conjunction with other information sources; undertaking exploratory analysis in a topic area which is usually restricted by data limitations; the ability to look at issues at a range of spatial scales, including very detailed local analysis; and a new visualisation technique for combining car ownership and usage at local authority level, which previously was not possible (as will be discussed in Section 4d and 4e).

B. Background

Information about car use is typically gathered in three ways:

This work has been funded by the Research Councils UK Energy Programme (EP/K000438/1), and is formally supported by the UK Department for Transport and the UK Department of Energy and Climate Change.

- 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.
- Strategic 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 usage 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 that 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 [4]).

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

II. DESCRIPTION OF DATA

A. *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 for 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. 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.

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[1]. 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 reorganising 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.

Data processing has involved importing SQL databases into Matlab and compressing data fields to minimum limits through carefully choosing data types and using lookup dictionaries. Processes have been developed to define vehicle-specific variables using consensus algorithms, in order to reconcile multiple values for the same variable from different years. The data has been exported to R and other software packages to allow analysis at both the individual vehicle level and for different areal units.

[1] 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.

The generation of the 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 [3], [12], [13]). 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). For example, it has also been necessary to define thresholds to screen out unexpectedly high mileages (currently set at 55,000 miles p.a. based on sensitivity testing during the scoping study) since these would skew all future analysis. 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.

B. *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, whilst in Scotland, they are linked to 'Data Zones', which are slightly smaller in population. This enables direct linking with area-based 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 provision).

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 [5]; understanding the links between energy use by private cars, and household energy and gas use [7]; exploring spatial variation in where people experience air pollution as opposed to where those creating vehicle pollution live [1]; exploring how variation in motoring costs and pollution vary with income [6]; and undertaking detailed spatial modelling of the determinants of car ownership and use [10], [14]. 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.

III. METHODOLOGY

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

- How do mileage distributions vary between areas?
- Is it the case that looking at variations in mean mileage is 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?
- How do the distributions vary for different normalisation approaches (per car, per household, per person)?

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, the Channel Islands and the Isle of Man)
- Regions (11 units: East of England; East Midlands; London; North East; North West; Scotland; South East; South West; Wales; West Midlands; Yorkshire)
- 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 (IGZs) in Scotland; typically 2000-4500 households per MSOA and 1000-2500 households per IGZ)
- Lower-layer Super Output Areas (LSOAs) (41,729 units: 32,844 in England; 1,909 LSOAs in Wales; 6,976 Data Zones (DZs) in Scotland; typically 500-800 households per LSOA and 200-500 households per DZ)[2].

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 to 31st December 2011; a mileage reading for each of those vehicles; the type of registered keeper, 'Private', 'Commercial or 'Between Keepers'. For this study, we have filtered vehicles to be those in MOT Classes 4 and 4a[3] (which are mainly cars). The sample size is shown in Table 1, split by privately registered and commercially registered cars. This only includes instances for which the vehicle is registered to an LSOA[4] (excluding cars 'Between Keepers') and only includes cars that have at least one MOT[5]. The MOT dataset also contains information for other vehicles, which are not included here: motorcycles and three-wheeled vehicles (Classes 1, 2 and 3); private passenger vehicles with at least 13 seats (Classes 5, 5a); goods vehicles over 3,000kg up to 3,500kg (Class 7). The remainder of the study focuses on privately

[2] MSOA, IGZ, LSOA and DZ are areal units defined for the UK Census.

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

[4] The LSOA is derived from the postcode of the registered keeper, so it was not possible to assign LSOAs to some vehicles where this was incorrect or incomplete, or where the vehicle was 'Between Keepers'.

[5] The MOT dataset contained information about vehicles until the end of 2014, such that all vehicles that were less than three years old in June 2011 would have 'shown up' in the dataset for their first MOT, and would therefore be included in our sample.

registered cars (that are registered to an LSOA) with valid mileage estimates in the MOT dataset, i.e. n=20,236,605.

TABLE I. SAMPLE SIZE OF MOT DATASET, MEAN ANNUAL MILEAGE PER CAR (GREAT BRITAIN, 2011; MOT DATASET CLASS 4 AND 4A VEHICLES)

Car registration type[a]	Total cars registered to an LSOA	Total cars registered to an LSOA, with valid mileage estimates	Mean annual mileage per car (miles)
Privately registered cars	25,418,355	20,236,605 (79.6%)	7,392
Commercially registered cars	2,362,860	1,966,439 (83.2%)	11,996
Total	27,781,215	22,203,044 (79.9%)	7,799

[a] Registration is based on the 'name' field; e.g. those listed as Mr/Mrs are classed as privately registered

IV. RESULTS

A. Differences in mean annual mileage between areas

1) Mean annual mileage per car

In order to provide the context for later analysis, this section explores variation in average mileages per car and how this changes with spatial scale. Specifically, this section presents results for the mean annual mileage per car (in thousands of miles) for the period Jan 2011 to Dec 2011. The following findings emerge:

- Fig. 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 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).
- Table 2 then provides a summary of the variability in results at the different spatial levels. Fig. 1 and Table 2 indicate that, as area size reduces, the range of the average mileages done in different areas increases (as would 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.

TABLE II. MEAN ANNUAL MILEAGE PER CAR FOR DIFFERENT SPATIAL SCALES (GREAT BRITAIN, 2011; MOT DATASET)

Geographical units	Number of units	Mean annual mileage per car		Max / Min
		Min	*Max*	
Region	11	6,428	7,757	1.21
LA (District and UA)	380	5,793	8,879	1.53
MSOA	8,480	4,951	11,586	2.34
LSOA	41,729	4,517	14,972	3.31

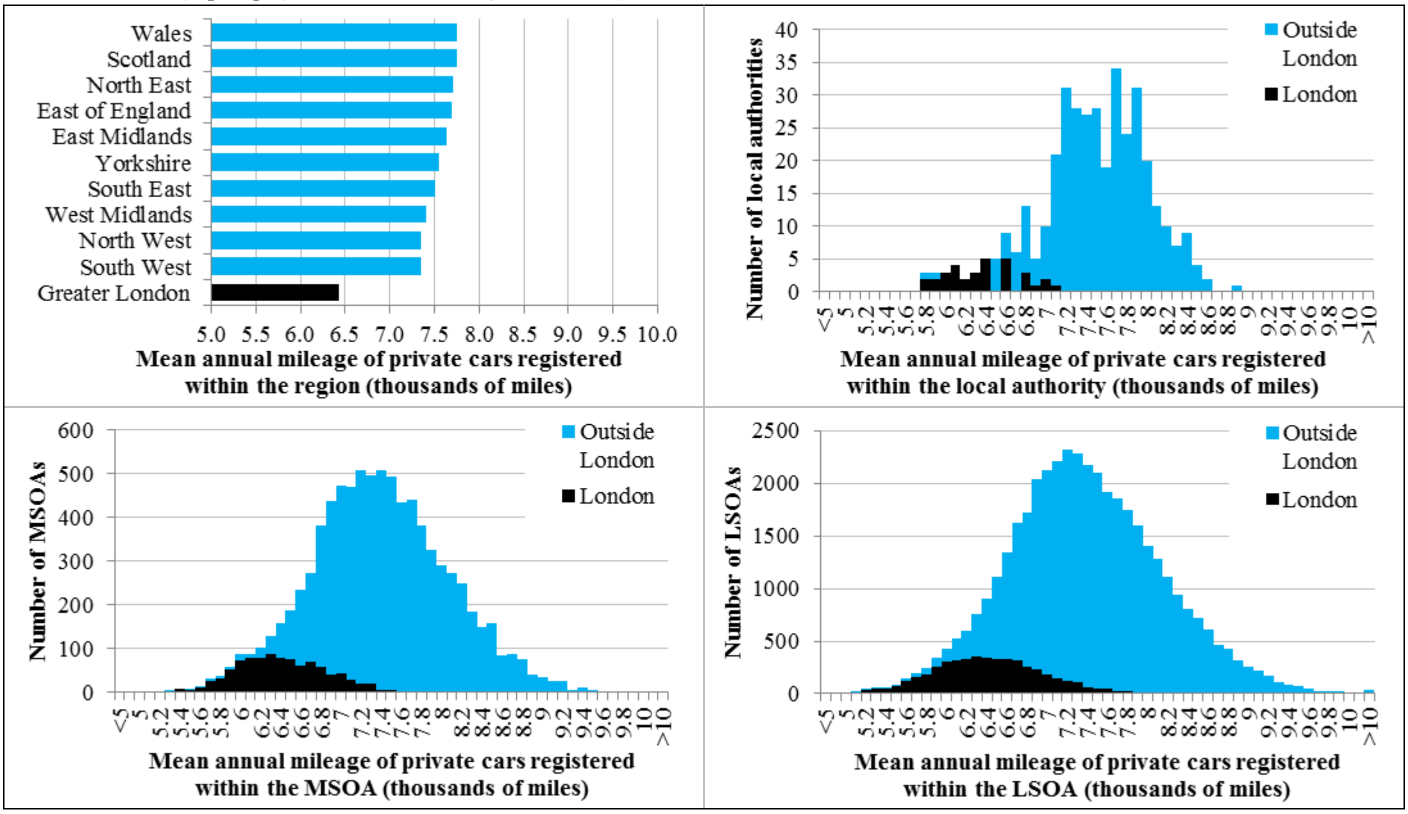


Fig. 1. Mean annual mileage per car for different spatial scales (Great Britain, 2011; MOT dataset)

B. Intra-zonal variation in car use

Having looked at the differences in averages between areas, the next stage of the work involved examining intra-zonal variability. Fig. 2 demonstrates the distribution of mileages by private cars for each of the regions in Great Britain. Similar plots have been created at all spatial scales, though are not shown here due to space constraints. (Specifically, an Excel VBA utility was developed to automatically export similar plots to PowerPoint for all local authorities within each region, all MSOAs within each local authority and all LSOAs within each local authority; this enabled the visualisation of the data for all 50,600 spatial units being considered.)

The primary finding is that most distributions look relatively similar, with the nearest fit likely to be a Gamma distribution. This was not our initial expectation – early discussions focused around whether areas might, for example, have bi-modal distributions, reflecting groups of high mileage and low mileage drivers. However, this was not the case.

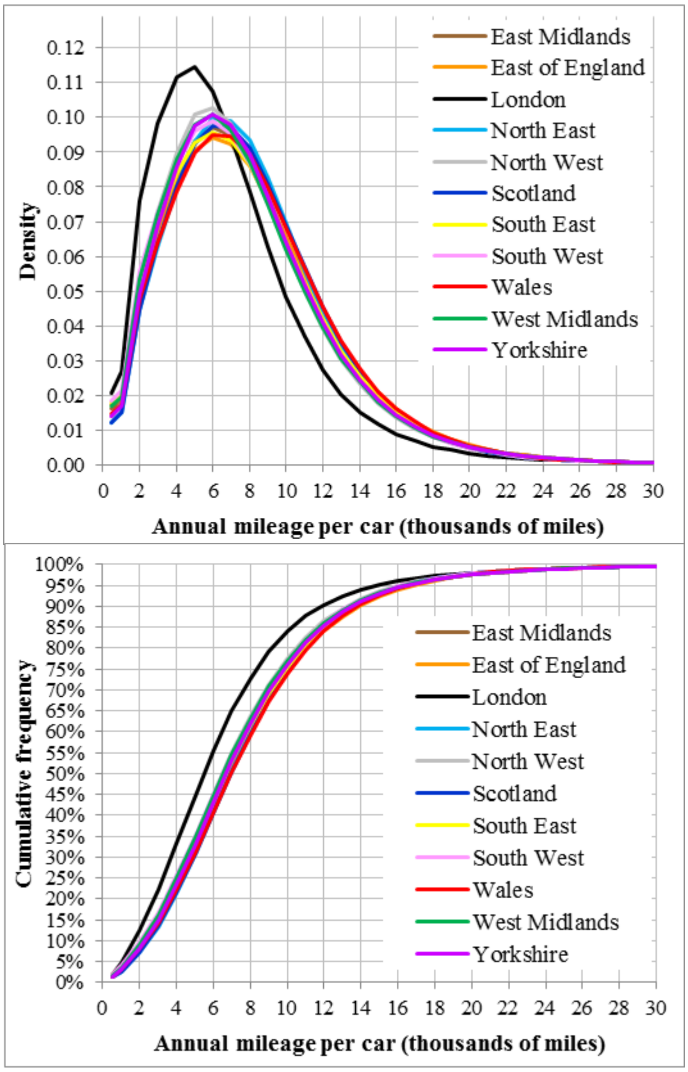


Fig. 2. Distribution of annual mileage per car (Great Britain, Regions, 2011; MOT dataset)

Another key finding was that the distributions often become more ‘spread out’ towards the tail end, implying that measures other than the mean may be more powerful for detecting differences between areas. This is demonstrated empirically in the next section.

C. Assessing areal differences using a measure other than the mean

As an illustration of the value of using a measure other than the mean to distinguish between areas, the analysis described in Section A has been repeated, but this time looking at the proportion of vehicles travelling more than 12,000 miles p.a.. A brief summary of the results is given in Fig. 3 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 (Fig. 3).
- 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 (Fig. 3).
- 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 2.3, whilst the high mileage share values vary by a factor of 7.7.

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

TABLE III. PROPORTION OF CARS WITHIN AN AREA THAT TRAVEL OVER 12,000 MILES PER YEAR FOR DIFFERENT SPATIAL SCALES (GREAT BRITAIN, 2011; MOT DATASET)

Geographical units	Number of units	% of cars over 12,000 miles		Max / Min
		Min	***Max***	
Region	11	9.6%	15.9%	1.66
LA (District and UA)	380	6.1%	23.2%	3.80
MSOA	8,480	4.3%	33.0%	7.67
LSOA	41,729	0%	46.0%	N/A

D. Moving from vehicle data to person data

The analysis described above has thus far focused on vehicle-specific measures – i.e. describing areas according to the distributions of *vehicle* usage. However, in most transport policy making, it is more meaningful to use people-centric metrics, not least to enable an exploration of the interaction between vehicle mileages and levels of car ownership. To enable such analysis, data from the core project dataset has been fused with data from the UK 2011 Census, in order to generate the cumulative plots shown in Fig. 6a and Fig. 6b. The approach is described below.

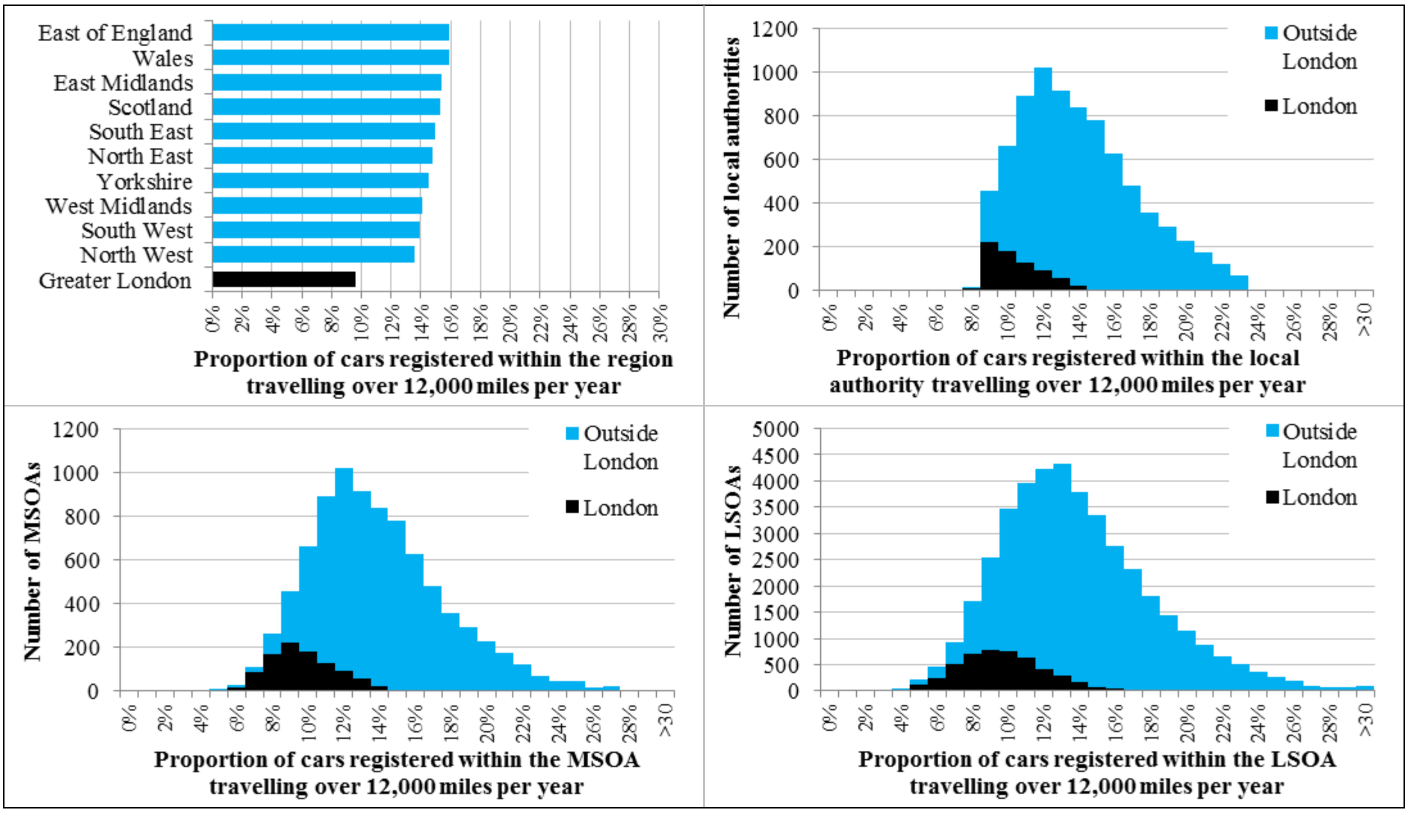


Fig. 3. The proportion of cars within an area that travel over 12,000 miles per year, for different spatial scales (Great Britain, 2011; MOT dataset)

1) Annual mileage 'per household'

As specified in Table 1, valid mileage estimates were available for 79.6% of privately registered cars in Great Britain. Similarly for any given LSOA, valid mileage estimates were typically available for around 70% to 90% of privately registered cars. An infill was applied to those vehicles with no mileage estimate, using the average mileage per car within the respective LSOA; this yielded an estimated total annual mileage for each LSOA.

The 2011 UK Census enabled the number of households within each LSOA to be partitioned into those with no access to a car (HHwNC) and those with access to at least one car (HHwC). The total annual mileage of each LSOA from the core dataset was divided by the number of HHwC in the respective LSOA from the Census dataset. The distribution of this measure is shown in Fig. 4a for both London and outside London. The median LSOA within London had a mean mileage per HHwC of 7,900 miles per year, whilst outside London this was 10,000 miles per year.

The proportion of households with no access to a car was 26.1% (41.6% of households in London and 23.8% of households outside London), see Fig. 5a.

A new visualisation technique is presented in Fig. 6a that combines the data in Fig. 4a and Fig. 5a. This takes the form of a cumulative distribution with the intercept on the y-axis equal to the proportion of households for that area with no access to a car; the LSOAs within that area are then sorted by mean mileage per HHwC and these LSOA means are presented cumulatively. As such the vertical axis can be considered to be the 'cumulative proportion of households'. Given the large number of LSOAs within the two regions shown in Fig. 6a, the curves appear to be smooth. For similar curves at smaller areal levels (with fewer LSOAs), these curves will appear 'stepped', i.e. it is assumed that each person living in a HHwC within a particular LSOA will do the mean mileage per HHwC for that LSOA. Although there is a risk of ecological fallacy, when LSOAs were created they were designed to be relatively homogeneous and thus this is a reasonable assumption. Future work will test the sensitivity of this assumption through applying a distribution of vehicle usage within each LSOA, rather than assuming a flat mean. A further implicit assumption of this technique is that households with no access to a car do no mileage. Whilst this may not always be true given the availability of car rental, it is a necessary simplifying assumption.

The dotted arrows illustrate taking the median value from the vertical axis, which, for London, was 6,300 miles per year; i.e. half of households within London do more than 6,300 miles per year, while the other half of households do less than this amount. Outside London, the median was 9,500 miles per year. For areas where the proportion of households with no access to a car is greater than 50%, the median will be zero.

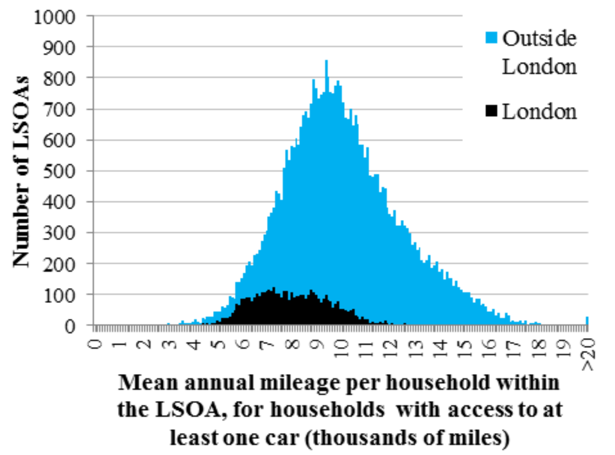
2) Annual mileage 'per person'

As well as information on the household car availability, the UK Census also enables the population to be partitioned by whether they live in a HHwC or a HHwNC. Equivalent graphs for population to those discussed above were generated using this partition. Fig. 4b shows the distribution of mean mileage per capita for people living in HHwC. For the median LSOA within London, this was 2,800 miles per person per year, while for the median LSOA outside London, this was 4,000 miles per person per year. Compared to the household graphs, there is a greater relative difference between London and non-London, which is explained by the fact that LSOAs in London typically have fewer people per household compared to the rest of the country. Fig. 5b shows that the proportion of the *population* living in a HHwNC was 19.7% (33.9% within London and 17.5% outside London). This is substantially less than the proportion of *households* with no access to a car, which is to be expected, because households with no cars typically have fewer people per household compared to households with cars. Fig. 6b combines the data from Fig. 4b and Fig. 5b in a similar approach as before. As expected, given the above discussion, comparing against Fig. 6a shows that the y-axis intercept is lower and the relative gradient is steeper for London. The median values were 2,500 miles per capita in London, i.e. half of the population of London do more than 2,500 miles per year, while the other half of the population do less than this amount. The median value for outside London was 3,800 miles per person.

Having examined the outcomes, the decision has been made to focus on individual metrics rather than household metrics, partly because other measures are often quoted on a per capita basis, and partly because, in Britain, annual data are available for population but not household numbers, which will be relevant when moving to a multi-year dataset.

a) Annual mileage per household | **b) Annual mileage per person**

Fig. 4. Mean annual mileage per household / per person, for households with access to at least one car (Great Britain, LSOA-level, 2011; MOT dataset and Census combined)



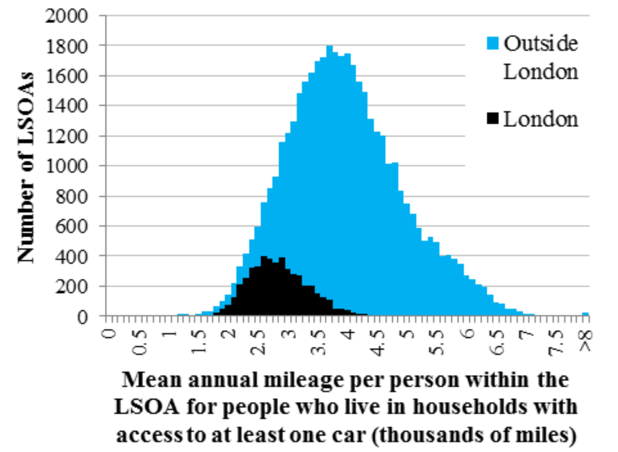
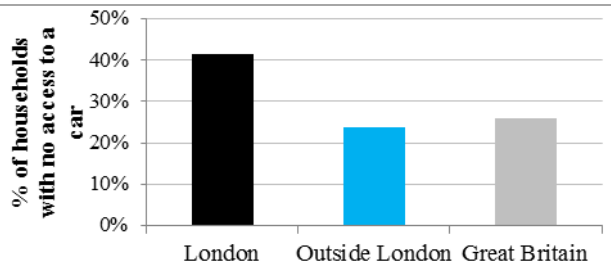


Fig. 5. Proportion of households / proportion of population with no access to a car (Great Britain, 2011; Census)



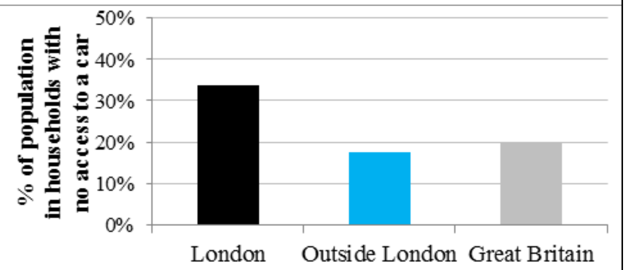
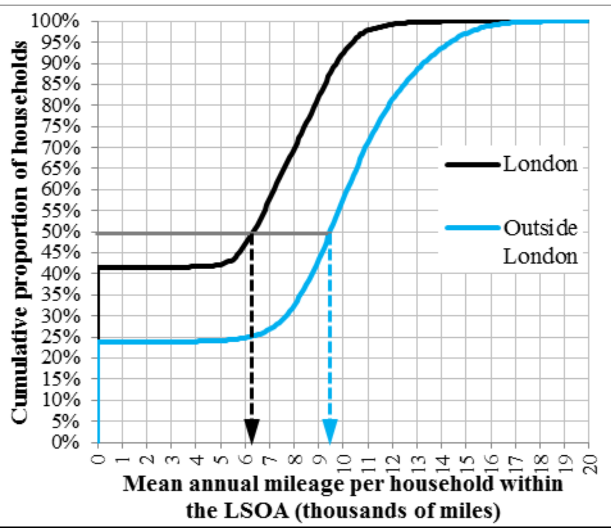
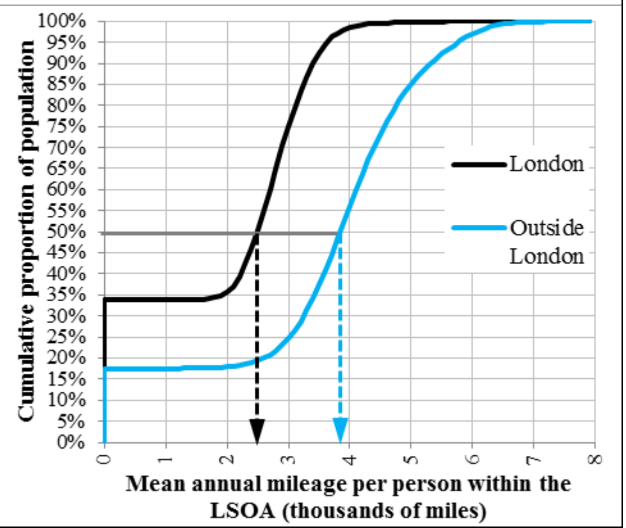


Fig. 6. Distribution of mean annual mileage per household / per person (Great Britain, 2011; MOT dataset and Census combined)





E. Comparing per capita distributions using different metrics

Having developed the processes described above, it then becomes possible to compare different areas on this basis. Fig. 7 shows the distributions created for the 11 regions and reiterates how London is markedly different from the other regions, both in the high intercept (high proportion of population living in HHwNC) and steep gradient (low mean mileage per capita for people living in HHwC). This may be partially explained by the fact that the London region is almost entirely urban, while the other regions are a mixture of urban and rural. Five regions have relatively higher mileage per capita (EM, WAL, SW, SE, EE), while four regions (each with large metropolitan areas) have relatively lower mileage per capita (WM, NW, Y, NE). Scotland is quite different from the other regions, with a high intercept (i.e. a high proportion of population living in HHwNC), although with a gentle gradient, suggesting there is a mixture of areas with low and high mileage per capita.

Fig. 8 illustrates the distributions for the 380 local authorities. These are more clearly 'stepped' as discussed above, although they are sufficient to show the wide range of distributions for different areas. Colour coding has been used depending on the mean value for the distribution. Whilst this shows that, as a general rule, the mean and the shape of the distribution are related, it also demonstrates that there are locations where the same mean is generated through different distributions. Fig. 9 illustrates this point further.

All of the distributions in Fig. 9 represent a mean annual mileage per person of 3,400 miles. However, the profiles are relatively different. For example, Epsom and Ewell is a relatively wealthy area, located on the outskirts of London, with high quality transport links to the capital; here, over 90% of the population own a car, but less than 10% drive more than 4,000 miles p.a.. In contrast, North Ayrshire is a less wealthy area, located in a relatively remote part of Scotland; here, nearly a quarter of people do not own a car, but about half drive at least 4,000 miles p.a.. This highlights the need for any expectations about car ownership and use to take into account the geodemographic characteristics of the area, when considering the scope for policy to make a difference.

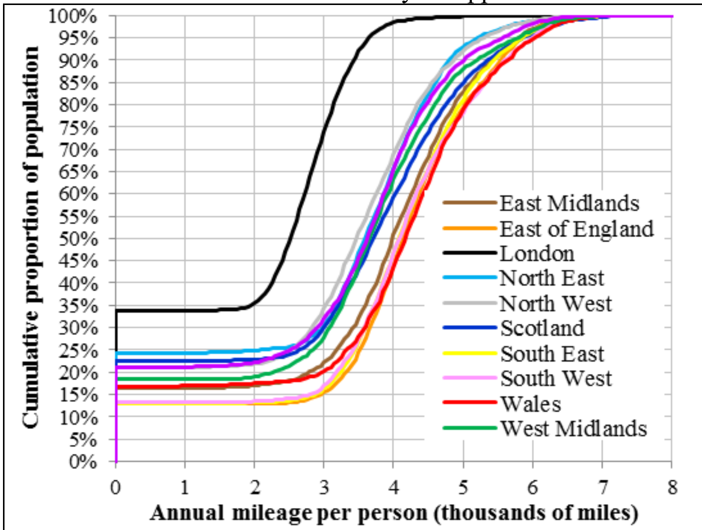


Fig. 7. Distribution of annual mileage per person (Great Britain, 2011, Regions; MOT dataset and Census combined)

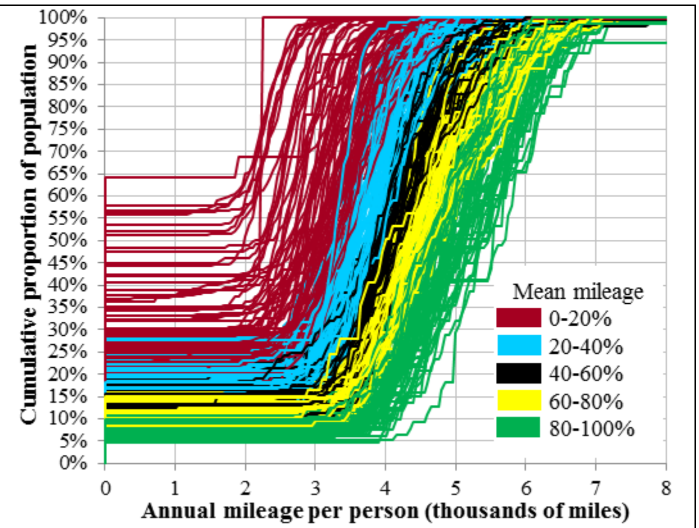


Fig. 8. Distribution of annual mileage per person (Great Britain, 2011, local authorities; MOT dataset and Census combined)

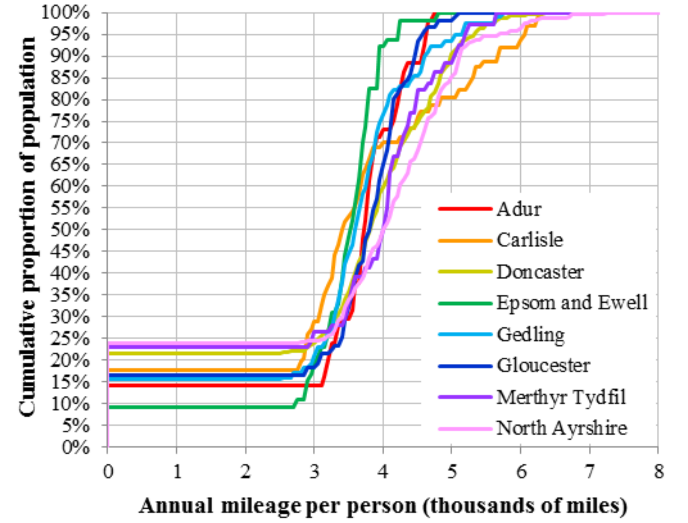


Fig. 9. Distribution of annual mileage per person (2011, selected local authorities; MOT dataset and Census combined)

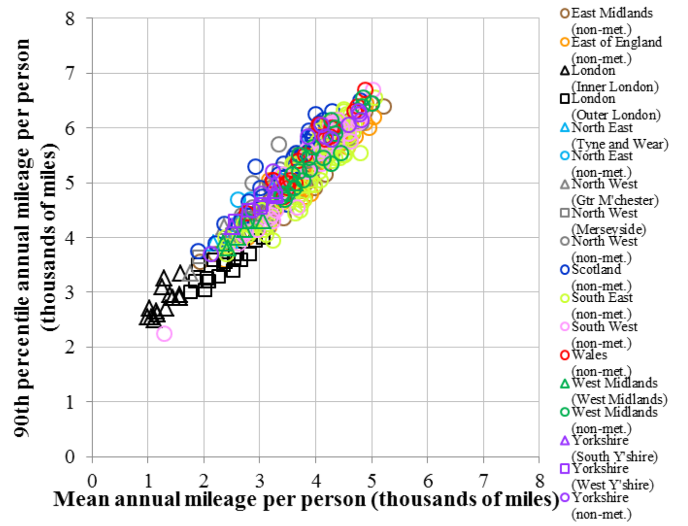


Fig. 10. 80th percentile of annual mileage per person, plotted against mean mileage per person (GB, 2011, local authorities; MOT dataset and Census)

Fig. 10 gives a bivariate plot of the 90th percentile annual mileage per capita (i.e. drawing a horizontal cross section line from 90% on Fig. 8), compared to the mean annual mileage per capita. This describes the lower bound of mileage of the top population decile for each local authority. Although these variables are closely correlated, there are some local authorities that have higher or lower mileage in this decile than the mean alone would otherwise suggest.

Table 4 provides a number of different metrics that can be used to describe those distributions at the regional scale including mean, median, mileage reading of 90th percentile, mileage reading of 95th percentile, kurtosis, skewness, coefficient of variation, proportion of cars travelling over 7,400, 12,000 and 15,000 miles p.a. A notable 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 (e.g. vehicles being driven more than 12,000 miles), all suggest a greater difference between areas than a simple comparison of the mean values would suggest. As an example, for the mean mileage per car at local authority level, the ratio of the 75th pecentile to the 25th percentile was 1.10; the equivalent ratio was 1.34 for the proportion of cars travelling over 12,000 miles. This implies that use of a metric other than the mean would be helpful when the differences to be detected are relatively small.

Table 5 provides the equivalent statistics for annual mileage per person. The ratios of the 75th to 25th percentile were typically higher than in Table 4, which is reflected visually when comparing Fig. 7 with Fig. 2.

TABLE IV. STATISTICAL MEASURES OF DISTRIBUTIONS OF ANNUAL MILEAGE PER CAR (GB, REGIONS AND LAS, 2011; MOT DATASET)

Units and area	Statistical measure	Sample size: Total privately registered cars with mileage	Mean	Median	90th percentile	95th percentile	Kurtosis	Skewness	Coefficient of variation	% of cars >7,400 miles p.a.	% of cars >12,000 miles p.a.	% of cars >15,000 miles p.a.
a) Annual mileage per car (Regions, see Fig. 2)	Minimum	766,491	6.36	5.42	11.72	14.71	9.06	1.64	0.64	31.1%	9.3%	4.7%
	25th %tile	1,605,121	7.33	6.48	13.28	16.14	9.44	1.72	0.65	41.5%	13.8%	6.6%
	Median	1,851,767	7.51	6.64	13.46	16.27	10.00	1.81	0.66	43.0%	14.6%	6.9%
	75th %tile	2,138,443	7.66	6.84	13.63	16.46	10.43	1.84	0.67	44.9%	15.1%	7.1%
	Maximum	3,324,663	7.72	6.94	13.81	16.69	12.94	2.29	0.75	45.8%	15.7%	7.5%
	IQR	533,323	0.33	0.36	0.35	0.33	0.99	0.11	0.02	3.4%	1.3%	0.5%
	Range	2,558,172	1.36	1.52	2.09	1.98	3.88	0.65	0.11	14.7%	6.3%	2.8%
	75th / 25th	1.33	**1.04**	**1.06**	**1.03**	**1.02**	**1.11**	**1.07**	**1.03**	**1.08**	**1.10**	**1.08**
	Max / Min	4.34	**1.21**	**1.28**	**1.19**	**1.13**	**1.43**	**1.40**	**1.17**	**1.47**	**1.68**	**1.59**
b) Annual mileage per car (local authorities)	Minimum	192	5.72	4.63	10.40	12.55	4.41	1.24	0.60	25.4	6.0	2.5
	25th %tile	35,729	7.11	6.26	12.76	15.51	8.51	1.57	0.65	39.0	12.2	5.7
	Median	44,871	7.46	6.65	13.34	16.16	9.79	1.75	0.66	43.1	14.1	6.6
	75th %tile	61,083	7.81	7.01	13.94	16.77	11.44	2.01	0.68	46.6	16.3	7.7
	Maximum	254,230	8.83	8.10	15.62	18.58	18.27	2.69	0.83	56.0	22.9	11.6
	IQR	25,354	0.71	0.75	1.18	1.26	2.93	0.44	0.04	7.6	4.1	2.0
	Range	254,038	3.11	3.46	5.22	6.03	13.85	1.45	0.23	30.5	16.9	9.1
	75th / 25th	1.71	**1.10**	**1.12**	**1.09**	**1.08**	**1.34**	**1.28**	**1.06**	**1.20**	**1.34**	**1.36**
	Max / Min	1324.11	**1.54**	**1.75**	**1.50**	**1.48**	**4.14**	**2.16**	**1.39**	**2.20**	**3.84**	**4.60**

[b] The City of London is a special case local authority with very few residents and a sample size of only 192 cars; 378 of the 380 local authorities had a sample size of at least 7,500 cars.

TABLE V. STATISTICAL MEASURES OF DISTRIBUTIONS OF ANNUAL MILEAGE PER PERSON (GB, REGIONS AND LAS, 2011; MOT DATASET AND CENSUS)

Units and area	Statistical measure	Sample size: Population	Mean	Median	90th percentile	95th percentile	Coeff. of variation	Kurtosis	Skew-ness	% of population >3,268 miles p.a.	% of population >4,500 miles p.a.	% of population >5,000 miles p.a.	% of population >5,500 miles p.a.	% of population >6,000 miles p.a.
a) Annual mileage per person (Regions, see Fig. 7)	Minimum	2,550,818	1.93	2.54	3.45	3.67	0.44	1.70	-1.15	15.7%	0.6%	0.2%	0.1%	0.0%
	25th %tile	4,808,638	3.09	3.63	4.95	5.46	0.48	2.37	-0.84	62.6%	19.0%	9.4%	4.7%	1.6%
	Median	5,196,386	3.30	3.73	5.38	5.83	0.57	3.70	-0.70	65.6%	26.6%	15.5%	8.5%	3.5%
	75th %tile	6,338,213	3.77	4.11	5.52	5.89	0.59	5.22	-0.56	77.7%	35.1%	20.3%	10.4%	4.0%
	Maximum	8,446,500	3.85	4.20	5.66	6.07	0.75	146.17	5.00	81.5%	38.3%	22.3%	12.2%	5.5%
	IQR	1,529,575	0.68	0.48	0.56	0.43	0.11	2.85	0.28	15.1%	16.1%	10.9%	5.7%	2.4%
	Range	5,895,682	1.92	1.65	2.21	2.40	0.31	144.48	6.15	65.7%	37.7%	22.1%	12.1%	5.4%
	75th / 25th	1.32	**1.22**	**1.13**	**1.11**	**1.08**	**1.24**	**2.21**	**0.67**	**1.24**	**1.85**	**2.16**	**2.22**	**2.51**
	Max / Min	3.31	**1.99**	**1.65**	**1.64**	**1.65**	**1.69**	**86.19**	**-4.35**	**5.18**	**65.92**	**107.88**	**125.61**	**137.43**
b) Annual mileage per person (local authorities, see Fig. 8)	Minimum	2,123	1.00	0.00	2.24	2.24	0.25	1.10	-2.82	0.0%	0.0%	0.0%	0.0%	0.0%
	25th %tile	92,512	3.02	3.53	4.43	4.62	0.38	2.70	-1.68	62.6%	8.1%	1.4%	0.0%	0.0%
	Median	124,338	3.64	4.05	5.07	5.43	0.46	3.77	-1.31	79.7%	29.4%	12.1%	4.3%	0.9%
	75th %tile	197,849	4.19	4.49	5.78	6.02	0.55	5.40	-0.97	88.3%	49.3%	30.4%	16.1%	5.5%
	Maximum	1,051,366	5.21	5.69	6.71	16.02	1.49	144.33	10.70	95.2%	90.8%	78.6%	58.8%	34.7%
	IQR	105,337	1.18	0.96	1.35	1.40	0.17	2.70	0.71	25.8%	41.2%	29.0%	16.1%	5.5%
	Range	1,049,243	4.21	5.69	4.47	13.78	1.24	143.23	13.52	95.2%	90.8%	78.6%	58.8%	34.7%
	75th / 25th	2.14	**1.39**	**1.27**	**1.31**	**1.30**	**1.45**	**2.00**	**0.58**	**1.41**	**6.11**	**21.58**	N/A	N/A
	Max / Min	495.23	**5.21**	N/A	**3.00**	**7.16**	**5.92**	**130.81**	**-3.80**	N/A	N/A	N/A	N/A	N/A

V. CONCLUSIONS

A. *Summary of findings*

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. The work demonstrates that mean mileages provide differentiation between areas. Moreover, the internal distribution of mileages within any given area follows a relatively similar distribution, making the mean a more reliable indicator than initially assumed. However, exploration of one alternative metric – the proportion of vehicles driven more than 12,000 miles a year –shows considerably greater variation between areas than the mean data does. Consequently use of metrics other than the mean may be a more sensitive way of detecting areal variation.

By fusing the core dataset with Census data, it is also possible to generate area profiles indicating how vehicle mileages are distributed across the population as a whole, including the significance of non-car owners to vehicle mileage in the area. A range of statistical measures can be generated to describe the subsequent distributions, and, as with the vehicle mileages, measures other than the mean appear to be more sensitive for distinguishing between areas. In addition, it is possible to show that areas with the same mean can have different internal distributions, with potential implications for policy.

B. *Further work*

Further work in this area will primarily involve investigating how the empirically observed differences described here can be explained and/or used to detect underlying areal characteristics and policy measures. For example, it would be interesting to explore whether areas with high levels of public transit provision or use, or cycle mode share, or car sharing initiatives, display distinctive differences in profiles.

From a social justice perspective, it would be interesting to investigate how the Gini coefficient relates to the various statistical measures explored in this paper and furthermore to understand whether areas with a high Gini coefficient for car use also have high levels of inequality in income, or other energy use.

Ultimately, it would be helpful to be able to advise local authorities on which distributional measures they should use to compare car use in their area with car use in other areas, and whether the nature of the distribution in their area has particular implications for the types of policies they should adopt.

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.

REFERENCES

[1] Barnes, J. and Chatterton, T. (2016) An environmental justice analysis of exposure to traffic-related pollutants in England and Wales. In: *Air Pollution XXVI, Proceedings of the 24th International Conference on Modelling, Monitoring and Management of Air Pollution, Greece, 20-22 June 2016*, Crete, Greece, 20-22 June 2016. Southampton, UK: Wessex Institute of Technology Press [In Press]

[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.