National-Scale Spatiotemporal Variation in Driver Navigation Behaviour and Route Choice

Elliot Karikari[1]
Leeds Institute for Data Analytics, University of Leeds, UK

Manon Prédhumeau
School of Geography, University of Leeds, UK

Peter Baudains
ESRC Consumer Data Research Centre, University of Leeds, UK

Ed Manley
School of Geography, University of Leeds, UK

Abstract

Understanding human behaviour is an integral task in GIScience, facilitated by increasingly large and descriptive datasets on human activity. Large-scale trajectory data have been particularly useful in measuring behaviours in different contexts, and understanding the relationship between the built environment and people. Yet, to date, most of these studies have focused on urban or regional scale analyses, with less exploration of behavioural variation at larger spatial scales. Human navigation behaviour is inherently linked to variation in spatial structure, and a study of national variations could help to better understand this variability. In this paper, we analyse GPS data from over 1 million journeys by 50,000 connected cars across the UK. Some key statistics relating to route choice are computed, and their variations are explored over time and space. A k-mean clustering of the trips identifies different types of trips and shows that their distribution varies by time of day and across the country. The insights gained from the data highlight spatio-temporal variations in road navigation, which should be considered in transportation modelling and planning.

2012 ACM Subject Classification Applied computing → Transportation

Keywords and phrases Connected Car, Geospatial big Data, Navigation Behaviour, Cluster Analysis

Digital Object Identifier 10.4230/LIPIcs.GIScience.2023.45

Category Short Paper

Acknowledgements The data for this research have been provided by the Consumer Data Research Centre, an ESRC Data Investment, under project ID CDRC 376, ES/L011840/1; ES/L011891/1.

1 Introduction

The increasing availability of vehicle usage data, made possible by the rise of electric connected vehicles, presents an opportunity for researchers to analyse vast amounts of data related to speed, location, and direction [2]. The ubiquity of the technology means that never before have granular data on navigation behaviour been available at such a large scale. In this study we leverage connected car data to gain novel insights into human mobility patterns and behaviour, scaling analysis up to the national scale. While previous studies have explored various aspects of mobility, such as travel distances, radius of gyration, and visited locations [6], this study specifically examines drivers' routes. Studies have relied on diverse data sources, including GPS tracking devices [6, 5], mobile phone data [4], and transportation surveys [8], revealing insights into the fundamental drivers of navigation behaviour. To date,

[1] corresponding author

12th International Conference on Geographic Information Science (GIScience 2023).
Editors: Roger Beecham, Jed A. Long, Dianna Smith, Qunshan Zhao, and Sarah Wise; Article No. 45; pp. 45:1–45:6
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

there has been no examination of navigation behaviour across an entire country. This study aims to fill this gap by examining the driver navigation behaviour and route choice using national-scale GPS data. The central research question of this study is: to what degree do drivers' navigation behaviours and route choices in the UK vary spatially and temporally ?

In this paper, we explore navigation across the entire UK. We observe how a set of indicators describing navigation behaviour vary over space and time. We outline the methods and data involved in the study, before describing the results and their implications.

2 Method

The methods used to derive insights into navigation behaviour are outlined in our previous work [3]. This paper established a methodology for deriving six key statistical measures - travel distance, travel time, stop time, number of turns, angular deviation and sinuosity - with application to the same data. Here, we extend our previous work by applying these six key statistical measures to 1,224,270 trips, i.e. 66.92% of the entire dataset and analysing their spatio-temporal variations. Section 3 details the processing undertaken to clean up the data retaining only one-way trips over 4 weeks in July. Section 4 then presents the statistical results obtained and a k-mean clustering analysis, in order to identify patterns associated with the different journey types and to examine their spatial and temporal variation.

3 Data

This work uses high-frequency GPS recordings from 50,000 connected cars across the UK (`https://data.cdrc.ac.uk/dataset/wejo-connected-vehicle-trajectories`). The dataset consists of over 400 million GPS data points, collected in July 2022 where an observation was recorded every 3 seconds on average during each journey (over 1.8 million). To ensure anonymity, the first and last 15 seconds of each journey have been removed.

An initial two-stage filtering of the data was applied. We selected a four-week period from 4th to 31st July for the analysis. This timeframe provides a balanced selection of weekdays and weekends. Some trips in the dataset were then identified as round trips, which are defined as trips for which the Haversine distance between origin and destination is less than 800 meters. These round trips were filtered out as they were found to greatly skew sinuosity results. Further analysis of this data could reveal specific behaviour associated with round trips. The present analysis however focused on the behaviours associated with one-way journeys which makes up 66.9% of the entire dataset.

In this study, we used Python and the Scikit-mobility library [7] to process data and generate key statistics such as travel distance and stop time. We also derived additional measures such as travel time, number of turns, cumulative angular deviation, and sinuosity, which were essential in providing insights into human mobility patterns.

4 Results

The results computed on the 1,224,270 trips indicate a diverse set of navigation behaviours within the data. Table 1 shows a summary of descriptive statistics generated per journey.

Results reveal a wide range of variability in distances travelled. On average, drivers tended to travel 13.1 km. However, this average is skewed by a few long trips as half of all trips made were below 5.5 km. It was also shown that stop time accounted for approximately 28% of their overall travel time. Furthermore, we observed that people tended to take routes

Table 1 Descriptive statistics on 1,224,270 trips.

	Min	Max	Mean	Q1	Median	Q3	SD
Travel distance (km)	0.8	715.4	13.1	2.8	5.5	12.6	24.9
Travel time (min)	0.5	747.2	16.7	5.7	10.30	19.5	20.5
Stop time (min)	0.0	646.3	4.7	0.0	2.3	5.7	7.8
Number of turns	0.0	863.0	17.1	8.0	13.0	22.0	15.1
Cumulative angular deviation (°)	0.5	111197.7	2504.3	1189.9	1955.2	3168.1	2074.5
Sinuosity	1.00	273.16	1.59	1.23	1.37	1.60	1.62

with an average of 17 turns per journey. This suggests that the complexity of travel routes should be considered when analysing travel behaviour. Finally, the average sinuosity of 1.59, meaning routes are around 60% longer than the Haversine distance, highlights a considerable amount of inefficiency in navigation behaviour and/or infrastructure. For comparison, [1] reported average sinuosities of 1.377 in Boston and 1.339 in San Francisco for pedestrians.

4.1 Variation over space and time

Next, we accessed spatiotemporal variation, the count of trip starts and sinuosity by time of the day and its geographic variation.

Figure 1 shows high trip starts recorded during all hours of the day on weekdays (in green). Two distinct peaks were identified, at 07:00, and between 14:00 to 16:00. The morning peak is likely indicative of people going to work, while the afternoon peak may be attributed to picking up children from school or people leaving work. The number of trips starts on the weekend (in red), steadily increased during the early hours of the day peaking at 10:00, before steadily declining towards the end of day.

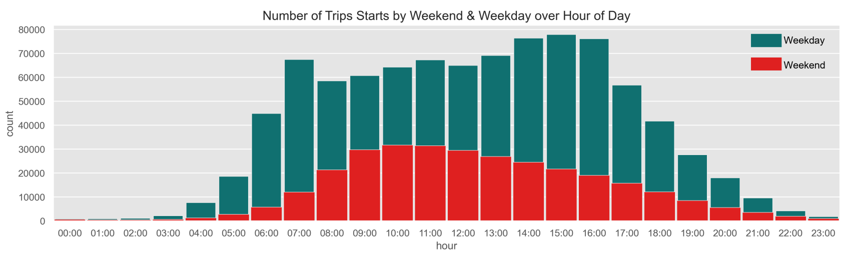

Figure 1 Distribution of connected car trip origins over time.

Geographic visualisations were done using a non-contiguous cartogram at the Local Authority (LA) level from `https://github.com/houseofcommonslibrary`. The LAs have been grouped and scaled in size relative to their populations. Figure 2 shows some regions with a high number of trip origins in Scotland such as Lanarkshire and Falkirk, and Glasgow and Clyde. Cornwall in the Southwest, parts of West Midlands, Merseyside, and West Yorkshire are also highly represented.

The sinuosity variable measures how much a trip deviates from the Haversine distance between the origin and destination. Trips with a sinuosity of 1 are direct and identical to their equivalent Haversine distance. Trips in Bedfordshire, Northamptonshire, Lanarkshire and Falkirk, Wiltshire, and Tyne and Wear have a relatively high average sinuosity (from 2.5 to 2.8). This suggest that drivers in these regions are constraints by infrastructure into driving further to reach their destinations, or that drivers take detours to avoid congestion.

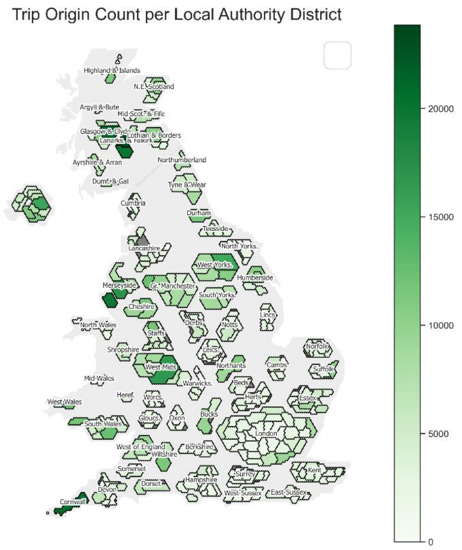

Figure 2 Distribution of connected car trip origins.

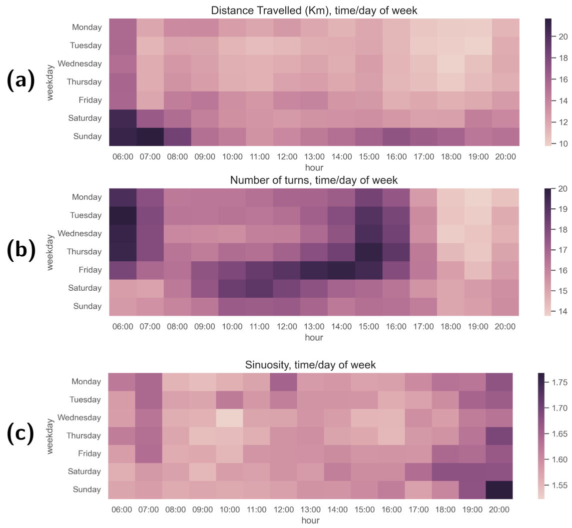

Figure 3 Average (a) distance travelled, (b) number of turns and (c) sinuosity over time and day of week.

Within areas with relatively lower sinuosity (from 1.7 to 1.9), i.e. Gloucestershire, Somerset, Highland and Islands, Oxfordshire, and Mid Wales, more direct routes are possible. Camden in the London area has a very high sinuosity (>8.5), which calls for future investigation.

Distances travelled vary depending on the day of the week and the time of day (Figure 3a). Long trips are more common in the early morning hours (around 06:00) on all days of the week, with more long trips on weekends starting between 06:00 and 07:00. This indicates a self-selection bias, in that people who need to travel further are more likely to be leaving in the early morning, relative to later in the day. Results also indicate a relationship between the number of turns per trip and start time (Figure 3b). It appears people may opt to take more turns during peak weekday and weekend periods. However, there is no clear relationship between time of day and sinuosity (Figure 3c), meaning that the routes do not deviate more significantly than usual during these periods. This is an indication that people seek to avoid traffic congestion during peak periods, but do not deviate widely from the shortest route.

4.2 Clustering analysis

After exploring the variability in the travel behaviour data, we identified different route types using k-means clustering analysis. This machine learning method groups similar data points together based on their features. Highly correlated route attributes (correlation coefficient > 0.7 or variance inflation factor > 2.5) were not used. As a result, only three variables – travel distance, number of turns, and sinuosity – were used to cluster the trips. We evaluated different values of k (i.e., [2-8]), using silhouette scores and silhouette visualisers, and found that k=4 resulted in the most distinct trip types. However, clustering with k=6 produced silhouette scores almost as good as k=4, and may be worth further exploration.

Short one-way trips (Cluster 0) Direct trips with the fewest average number of turns (11). Observed travel distance (7 km) is on average 56% longer than the Haversine distance. Most trips fell within this cluster (79%).

Mid-range one-way trips (Cluster 1) Longer trips with more turns (37). Observed travel distance (22 km) is on average 93% longer than the Haversine distance. This cluster accounted for 18.9 % of all trips.

Long one-way trips (Cluster 2) Very long trips (139 km), on average 40% longer than the Haversine distance. The average distance travelled is 6 times longer than in Cluster 1 but has only 13% more turns. This may suggest that Cluster 2 uses more major roads. 2.4% of all trips were accounted for in this cluster.

Sinuous trips (Cluster 3) 0.1% of clustered trips were identified as round trips, with unusually high sinuosity values (32) compared to the other clusters (1 to 2). This indicates that the simple filtering process used (removing trips with origin-destination distance <800m) could be improved using a filter to remove high sinuosity trips.

As most trips (79%) were short one-way trips (Cluster 0), we ran the clustering on this subset to identify variations within trips (Table 2).

Table 2 Descriptive statistics of re-clustered Cluster 0.

	Average travelled distance	Average number of turns	Average sinuosity	% trips
0A	6.90	17.66	1.45	34.7
0B	4.06	6.81	1.34	51.4
0C	5.22	13.26	2.66	6.6
0D	29.49	12.99	1.33	7.3

Cluster 0A Short sinuous one-way trips, with high number of turns. Observed travel distance is 45% longer than its Haversine distance.

Cluster 0B Short, direct, low sinuosity, one-way trips. Observed travel distance is 34% longer than its Haversine distance.

Cluster 0C Short highly sinuous one-way trips, with moderate number of turns. Observed travel distances is 166% longer than its Haversine distance.

Cluster 0D Mid-range, low sinuosity, one-way trips with moderate number of turns. Observed travel distance is 33% longer than its Haversine distance.

Further analysis found that Fridays had the highest number of short sinuous trips (0A), while Wednesdays had more short direct trips (0B). Drivers may be more willing to take indirect routes on Fridays when they have more time or are less constrained by work schedules. The analysis also showed that short sinuous trips (0A) were more common in major urban conurbations including London, Greater Manchester, and South Yorkshire (Figure 4), while short direct trips (0B) were highly represented in all other areas. Travel behaviour in some areas may be different from others, possibly due to road infrastructure or specific traffic conditions. Further research could explore these variations and identify potential solutions for improving travel efficiency.

5 Conclusion

This study provides insights into road navigation behaviour across the UK, based on connected cars data. The analysis of travel distance, number of turns, and sinuosity revealed patterns that vary by time of day and day of the week. The identification of different trip types further highlights the variability in navigation behaviour across the UK. This new perspective on navigation behaviour can supplement the outputs of classic surveys and will be used to create synthetic trip datasets that are representative of observed behaviours.

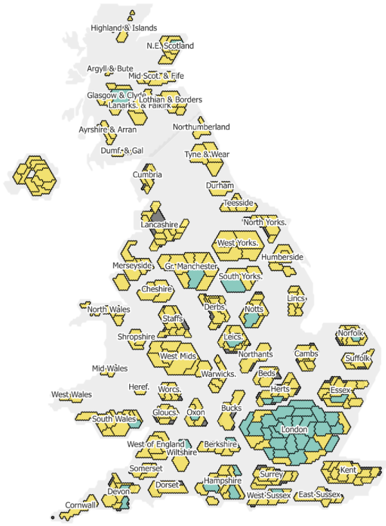

Figure 4 Most frequent cluster for each local authority: short sinuous one-way trips (0A) in green and short direct one-way trips (0B) in yellow.

Results from this study can inform transportation planning and policy. For example, the finding that 25% of the analysed car trips are shorter than 2.8 km can guide the development of zero-emission local policies by identifying where and when drivers make short trips. Moreover, insights from this study may be used to refine transport models with new behavioural patterns, and help to predict drivers' behaviour. However, the lack of socio-demographic data prevents the assessment of the representativeness of the data. This study is purely observational and further exploration of causation is required. Stop time and point of interest data could enable the investigation of the trip purposes. Overall, this study underscores the potential of using vehicle trajectory data to understand travel decisions.

References

1. C Bongiorno, Y Zhou, M Kryven, D Theurel, A Rizzo, P Santi, J Tenenbaum, and C Ratti. Vector-based pedestrian navigation in cities. *Nat. Comput. Sci*, 1(10):678–685, 2021.
2. R Coppola and M Morisio. Connected car: technologies, issues, future trends. *ACM Comput. Surv.*, 49(3):1–36, 2016.
3. E Karikari, M Prédhumeau, P Baudains, and E Manley. Analysing connected car data to understand vehicular route choice. In *31st Annual Geographical Information Science Research UK Conference (GISRUK), Glasgow, Scotland*, 2023.
4. L Li, S Wang, and F-Y Wang. An analysis of taxi driver's route choice behavior using the trace records. *IEEE Trans. Comput. Soc. Syst.*, 5(2):576–582, 2018.
5. EJ Manley, JD Addison, and T Cheng. Shortest path or anchor-based route choice: a large-scale empirical analysis of minicab routing in London. *J. Transp. Geogr.*, 43:123–139, 2015.
6. L Pappalardo, S Rinzivillo, Z Qu, D Pedreschi, and F Giannotti. Understanding the patterns of car travel. *Eur. Phys. J. Spec. Top.*, 215:61–73, 2013.
7. L Pappalardo, F Simini, G Barlacchi, and R Pellungrini. scikit-mobility: A Python library for the analysis, generation and risk assessment of mobility data. *J. Stat. Softw.*, 2022.
8. B Yin and F Leurent. What are the multimodal patterns of individual mobility at the day level in the Paris region? A two-stage data-driven approach based on the 2018 Household Travel Survey. *Transportation*, pages 1–30, 2022.