Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 00 (2018) 000–000

Transportation Research Procedia

www.elsevier.com/locate/procedia

World Conference on Transport Research - WCTR 2019 Mumbai 26-31 May 2019

Investigation of Driver Route Choice Behaviour using Bluetooth Data

Sreedevi Somadas Kottayil*, Panagiotis Tsoleridis, Kacper Rossa, Richard Connors, Charles Fox

The Institute for Transport Studies, University of Leeds, Leeds, LS2 9JT, United Kingdom

Abstract

Many local authorities use small-scale transport models to manage their transportation networks. These may assume drivers' behaviour to be rational in choosing the fastest route, and thus all drivers behave the same given an origin and destination, leading to simplified aggregate flow models, fitted to anonymous traffic flow measurements. Recent price falls in traffic sensors, data storage, and compute power now enable Data Science to empirically test such assumptions, by using per-driver data to infer route selection from sensor observations and compare with optimal route selection. A methodology is presented using per-driver data to analyse driver route choice behaviour in transportation networks. Traffic flows on multiple measurable routes for origin-destination pairs are compared based on the length of each route. A driver rationality index is defined by considering the shortest physical route between an origin-destination pair. The proposed method is intended to aid calibration of parameters used in traffic assignment models e.g. weights in generalized cost formulations or dispersion within stochastic user equilibrium models. The method is demonstrated using raw sensor datasets collected through Bluetooth sensors in the area of Chesterfield, Derbyshire, UK. The results for this region show that routes with a significant difference in lengths of their paths have the majority (71%) of drivers using the optimal path but as the difference in length decreases, the probability of optimal route choice decreases (27%). The methodology can be used for extended research considering the impact on route choice of other factors including travel time and road specific conditions.

Keywords: Data Science; Bluetooth; Driver Rationality; Traffic flow models

* Corresponding author. Tel.: +0-000-000-0000 ; fax: +0-000-000-0000 .
E-mail address: sreesomadas@gmail.com

1. Introduction

Most transport practitioners today, including local authorities and transport consultancies, rely on traditional Transport Modelling assumptions of User Equilibrium to assess the driving behaviour in their case studies. There are many cases where transport models and their parameters are transferred without calibration to a new application context with undesirable outcomes. Therefore, there is a need for more realistic calibration of these models to real-world driver behaviour.

Emerging technologies of data collection through sensors, supported by a continuous decrease in their installation costs and an increase of data storage and computing capacity, can provide the necessary framework for data-driven applications in the field of transport modelling. Currently, due to the sheer amount of incoming real-time data, there is a growing need for developing new methods of data analysis to address the limitations of traditional practices. This study utilizes some of these techniques to investigate the rationality of drivers in a transportation network from traffic sensor data, which in turn provides new parameters to calibrate flow models based on stochastic routing. More generally, this illustrates how the new Transport Data Science approach can be used as a complimentary tool and provide evidence to support, and enhance traditional transport models.

1.1. Transport Sensor Data

The use of passively collected data from ubiquitous sensors is becoming increasingly popular in Transport Planning, due to its significant advantages over traditional transport surveys. It provides inexpensive and continuous data collection on a 24-hour basis, resulting in a greater spatio-temporal resolution. On the other hand, traditional methods typically include data collected from censuses, not frequently updated, or from expensive and infrequent transport-related surveys, with limited population coverage, which in many cases are prone to errors (e.g. respondents' cognitive fatigue-sampling bias) or might cause traffic disruption (Yang et al., 2015). "Big" data is characterized by its volume, velocity, variety, veracity and value (Ishwarappa and Anuradha, 2015). Subsequent analysis of data collected through mobile phones (Iqbal et al., 2014; Bwambale et al., 2017), GPS and Bluetooth sensors (Barceló et al., 2010; Martchouk et al., 2011; Gong et al., 2016), Automated Fare Collection (AFC) systems (Mahrsi et al., 2014), Automated Number-Plate Recognition (ANPR) systems (Fox et al., 2010), and Location-based Social Networking (LSBN) data from social media (Yang et al., 2015), among others, has the potential of providing useful insights, uncovering previously hidden mobility patterns at a high resolution level. This level of analysis has been supported by trends and advances in computer power, storage space and cloud computing, since the beginning of the century (El-Seoud et al., 2017).

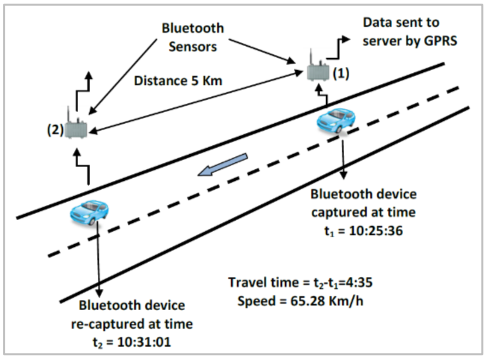

Fig. 1. Bluetooth sensor monitoring system [Source: (Barcelo et al., 2010)]

Bluetooth sensors, specifically, have been primarily used for Origin-Destination and travel time estimation (Barceló et al., 2010) or for driver classification (Crawford et al., 2018). Traditional methods of estimating travel times, involving inductive loop detectors or probe vehicles combined with GPS devices, are limited in providing a complete dataset with multiple time periods during the day and different driver classes (Martchouk et al., 2011). On the contrary, data derived from Bluetooth sensors can provide real time monitoring of travel time and speed for each individual passing vehicle being detected by them (fig. 1). The main drawback of Bluetooth data is the ability to capture only a subset of the total flow, as many drivers might opt to disable Bluetooth devices for battery conservation (Yang et al., 2015). The likelihood also exists that a passing vehicle with a Bluetooth-enabled device will not get detected by the sensor. An experiment conducted by Araghi et al. (2014) indicated that on average only 80% of detectable vehicles are actually being detected while passing from a Bluetooth sensor zone. To overcome the limitations of both approaches, there is a growing research interest in fusing together Bluetooth and loop detector data (Bachmann et al., 2013).

1.2. Traffic Models

Rational decision-making is often a key assumption in social sciences from economics to transport. The central model of economics is the homo economicus which assumes that agents behave rationally and seek to maximise their utility (Persky, 1995). Simon (1991) argued that due to imperfect and incomplete information it is rather reasonable to assume that rationality is “bounded”. Bounded rationality was assumed in a variety of transport models (Di et al., 2013) and in road safety (Sivak, 2002) or transport policy making (Marsden et al., 2012) as well as route choice studies (Nakayama et al., 2001, Watling et al., 2018). Specifically for the latter, cost indifference bands were assumed where a cost excess compared to the minimum cost route, up to a threshold, does not influence individuals’ route choice (Watling et al., 2018). Transport models usually have some uncertainty effects introduced as random variables (Ben-Akiva and Lerman, 1985). The error term as suggested by Daganzo and Sheffi (1979) can be interpreted as the uncertainty of travel time. Throughout the years, different route choice models were proposed based on different assumptions, i.e. that a driver bases the decision to find the least costly route on the weighted average of time from the past (Horowitz, 1984). Hyunmyung (2012) implied that drivers’ behaviour and choices tend to be habitual and repetitive. User Equilibrium (UE) models, however, do not take the past experiences into account, so UE solutions can be more sensitive to changes in the network than drivers actually are. Kobayashi (1994) implied that the reason for thinking that drivers are rational is due to drivers’ learning process. Nakayama et al. (2001), however, argued that even after a long learning process, there is heterogeneity in driver’s perceptions of routes and therefore behaviour in the network is heterogenous. Rationality of drivers can be measured by travelling behaviour between origins and destinations. This is one of the key components of route choice problems and many plausible solutions have been developed.

In Deterministic UE, users are assumed to have perfect knowledge of the network conditions and make rational decisions based solely on the generalized costs of each available route. This assignment model has been particularly successful at describing traffic flows in congested networks, for example in the canonical morning peak period that stereotypically comprises drivers with perfect network knowledge, strongly motivated to minimize their travel time. The travel time along each link depends on the traffic volume, as described by a Volume-Delay function, and link costs add together to give origin-destination route costs. A range of different Volume-Delay functions have been formulated in the literature based on their specific context, with the one defined by the Bureau of Public Roads (BPR) in the USA (1964), being the most widely used (Eq. 1) (Ortuzar and Willumsen, 2011).

$$t_a = t_a^0\left(1+\alpha\left(\frac{x_a}{c_a}\right)^{\beta}\right) \tag{1}$$

where t_a (min/km) is the travel time on link a, having free flow travel time t_a^0 (min/km). The link flow is x_a (veh/hr) and c_a (veh/hr) is the link capacity and a and β are calibration parameters.

UE can be formulated as a linear programming minimization problem (Beckmann et. al, 1956) with the following objective function (for fixed demand),

$$min\ z_{UE}(\boldsymbol{x}) = \sum_a \int_0^{x_a} t_a(w)dw \tag{2}$$

where $\boldsymbol{x} = [x_a]$ is the vector of link flows. Denoting the flow on the k-th path connecting OD pair ij as f_k^{ij} and the total OD demand as Q, applicable flow conservation and non-negativity constraints are as follows:

$$\sum_k f_k^{ij} = Q_{ij} \tag{3}$$

$$f_k \geq 0 \tag{4}$$

Various algorithms have been proposed in the literature for solving the UE minimization problem, such as Origin-Based Assignment algorithms (OBA) can achieve convergence even for large networks (Bar-Gera, 2002).

Stochastic User Equilibrium (SUE), first defined by Daganzo and Sheffi (1977), generalises UE by including an error component in the route choice model which can be interpreted as representing incomplete knowledge of network conditions by users. This route choice model does not assign all users to minimum cost route(s), hence those users not minimizing their generalized cost could be considered not (strictly) rational. It is worth noting that in fact users may be heterogeneous regarding the evaluation of route attributes, such as route distance and travel time, while there may be even additional parameters influencing route choices (Ortuzar and Willumsen, 2011) and this would give rise to a distribution of *perceived* route costs. Whereas UE provides a good model for congested urban networks, SUE is mostly applicable in non-congested and larger scale networks where strict cost minimization is less relevant and individual preferences can influence the perceived generalized costs and choice of preferred routes. In traffic networks, where congestion steadily increases, traffic flow patterns obtained from SUE gradually converge to the UE equivalent patterns (Florian and Hearn, 2008). Compared to UE, SUE provides a framework to add flexibility to the assignment model and incorporate uncertainty, with the trade-off being an increase in computation time. Two different methods are commonly adopted for solving SUE, namely Monte-Carlo simulation and application of the Multinomial Logit model (Ortuzar and Willumsen, 2011). Nonetheless, despite the widespread use of MNL due to its simplicity (Prashker and Bekhor, 2004), route choice often consists of numerous overlapping alternatives, which violates the basic IIA principal and makes the use of MNL unsuitable. One of the major concerns in SUE implementation, besides their computational complexity, is the quality of data used as input. Parameters for SUE models are typically obtained from Revealed (RP) or Stated Preference (SP) surveys with both being prone to errors leading in many cases to model misspecification. The purpose of the present study is to show how we can instead inform these parameters and drivers' rationality specifically, directly from passively collected data.

2. Methods

In the following, "rationality" (strict cost minimization) of drivers in their route selection is tested in the area of Chesterfield, Derbyshire, UK, making use of its extensive Bluetooth sensor network to collect sensor data, and OpenStreetMap data to find optimal routing. Actual driver routes are inferred though a mixture of sensor detections and local-scale shortest path assumptions, and optimal routes are inferred from a global shortest-path assumption. Actual and optimal routes are compared as a function of route length to measure driver rationality. Chesterfield region was chosen as it is a major area of interest for its local authority, Derbyshire County Council (DCC), due to traffic congestion issues making it a bottleneck for drivers. Therefore, DCC wishes to understand whether drivers using the local road network act rationally in order to suggest new interventions.

Traffic data from Bluetooth Sensors (BS) were obtained from the DCC database system. For computational reasons, this study uses traffic flow for a single weekday, 14 February 2017. The database preparation and management was carried out using the relational database PostgreSQL. Data processing was performed in accordance with DCC and University of Leeds privacy policies.

2.1. Bluetooth Data Cleaning Algorithm

The Bluetooth data for 31 sites (Table 1) contained detections for each vehicle as unique Media Access Control (MAC) addresses along with the timestamp for each detection, at every chosen site. There was significant noise in the Bluetooth data caused by repeated detections of a single vehicle at very close instances of less than 60 seconds of time. The dataset was cleaned of these unwanted detections during the import, by removing the erroneous detections happening for each site separately. The repetitive detections within each site, which were removed during the data cleaning process, were about 48% of the actual raw data. The repetitive detections might be due to long queues or slow-moving traffic caused by unforeseen traffic conditions (Fox, 2018), resulting in the same vehicle being detected more than once by the same sensor. It was also observed that certain vehicles were detected at different sites at very short time-windows, even if the sites were fairly distant from each other. However, calibration of Bluetooth sensors was not within this study's scope so data used were purely based on the detection outputs received.

Table 1. Bluetooth sensor sites

site ID	Location description
DCCJT_MAC000010100	A632 Chesterfield Road, Duckmanton
DCCJT_MAC000010101	A632 Chesterfield Road, Calow
DCCJT_MAC000010102	A619 Chatsworth Road
DCCJT_MAC000010103	B6057 Hollywell Street, Chesterfield
DCCJT_MAC000010104	A61 Derby Road, Chesterfield
DCCJT_MAC000010105	A619 Worksop Road, Mastin Moor
DCCJT_MAC000010106	A632 Chesterfield Road, Duckmanton
DCCJT_MAC000010107	A632 Chesterfield Road, Arkwright Town
DCCJT_MAC000010108	A619 Chatsworth Road, Chesterfield
DCCJT_MAC000010109	A619 Markham Road, Chesterfield
DCCJT_MAC000010110	A619 Markham Road, Chesterfield
DCCJT_MAC000010111	B6507 Sheffield Road, Stonegravels
DCCJT_MAC000010112	Lockoford Road, Whittington Moor
DCCJT_MAC000010113	A619 Rother Way, Chesterfield
DCCJT_MAC000010114	A61 Tesco Roundabout
DCCJT_MAC000010115	B6051 Newbold Road, Newbold
DCCJT_MAC000010116	A619 Market Street, Staveley
DCCJT_MAC000010117	A619 Chesterfeld Road, Brimington
DCCJT_MAC000010118	A619 Chatsworth Road, Chesterfield
DCCJT_MAC000010119	C327 Whitecotes Lane, Walton
DCCJT_MAC000010120	Whittington Moor Roundabout Dunston Road
DCCJT_MAC000010121	Whittington Moor Roundabout Station Road
DCCJT_MAC000010122	A6 Dronfield Bypass
DCCJT_MAC000010123	A619 Markham Road / Lordsmill Street, Chesterfield
DCCJT_MAC000010124	A632 Lordsmill St / Chesterfield Road, Chesterfield
DCCJT_MAC000010125	M1 Jnc 29 Heath Roundabout
DCCJT_MAC000010126	M1 Jnc 29 Heath Roundabout A6175 Clay Cross
DCCJT_MAC000010127	A61 High Street Clay Cross
DCCJT_MAC000010128	A61 Alfreton Rd / Derby Rd
DCCJT_MAC000010129	HornsBridge RNDBT South side
DCCJT_MAC000010130	HornsBridge RNDBT North side

2.2. Map Data

The routes in this study are defined as paths connecting sensor location origins and destinations via road links. Each sensor is located by latitude and longitude. These locations were converted to 2D geometric projections of Cartesian coordinates, which was associated with the World Geodetic System, 1984 (WGS84) and the British National Grid (BNG). These projections were used for the study to locate the sensors on a 2D map taken from OpenStreetMap data as illustrated in Fig 2. Optimal routes were developed using Dijkstra's Shortest Path algorithm considering each sensor location as a node to create links between them and the algorithm utilised physical length between sensor-to-sensor distance as its cost parameter.

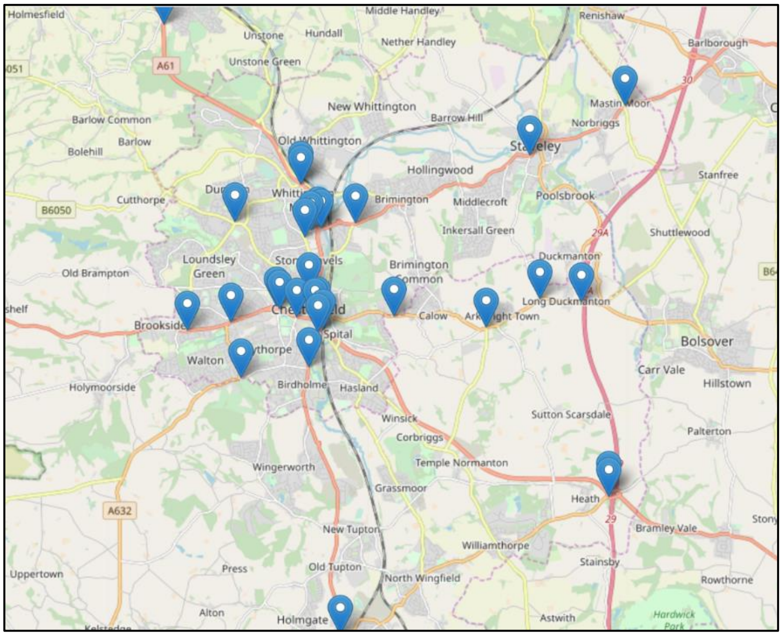

Fig. 2. Detector Locations

2.3. Baseline Traffic Estimate

A simple estimate of traffic flows over the network was obtained by aggregating flows along origin-destination routes. Here, a set of origins and destinations is defined by Bluetooth sensor locations at the peripheries of the network. Journeys on each origin-destination route are then retrieved by selecting matching Bluetooth MACs at the origin and destination, such that the destination detection occurs within a 1-hour time window of the origin detection. Assuming that the route taken is the shortest Dijkstra path between the origin and destination, each journey on each route can then be added to the flow at each segment of its Dijkstra path, to produce a flow estimation map (Fig. 5). This flow estimation is based on the assumption that all drivers choose the shortest route, which is the assumption tested empirically in this study.

2.4. Rationality Analysis

The study's purpose is to examine whether drivers are strict cost minimizers. Our methodology is briefly illustrated in the flowchart of Fig. 3.

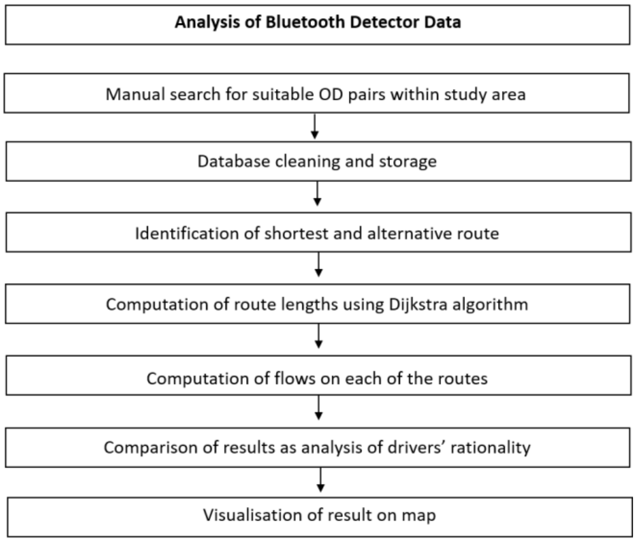

Fig. 3. Methodology Flowchart

Analysis of Bluetooth detector data was performed to count journeys of min-cost and other routes as follows:

- Define an *OD pair*, as illustrated in Fig. 4, as any pair of sensor locations (A, B) to represent an origin A and destination B.
- Define a *Measurable Route (MR)*, as illustrated in Fig.4, from a set of three sensor locations (A, C, B) such that C lies along at least one route for OD pair (A, B), such that the route comprises the Dijkstra path from A to C and the Dijkstra path from C to B.
- Define a *Measurable Route Pair (MRP)* as a pair of routes ((A, C, B) (A, D, B)) from A to B such that (A, C, B) and (A, D, C) are measureable routes, and (A, C, B) is the shortest route from A to B. We define (A, C, B) the optimal measurable route of the pair, and (A, D, B) the suboptimal measurable route of the pair. A measurable route pair enables us to compare sets of per-driver journeys from A to B via two different observable midpoints C and D.
- Define an *Alternative Route (AR)* any route from A to B when this route does not contain any other sensor.

A search over OD pairs was performed to find measurable route pairs, from the 31 available Bluetooth sites as origins and destinations. A total of five measurable route pairs were identified. The small number of detected measurable route pairs was due to the small network size of the town of Chesterfield.

Flows on each of the measurable route pairs were calculated by matching the MAC address detections from Bluetooth sensors, which also contained the timestamps of each detection. To filter spurious matches, the timestamp of a vehicle at its destination was required to be greater than at the mid-point locations which were in turn greater than at the origin location. To filter further spurious matches, the time interval between origin-to-midpoint and midpoint-to-destination was required to be less than 30 minutes for each, making the origin-to-destination journey time interval less than 60 minutes as an average limit for a weekday condition. Finally, multiple detections of the same MAC on a route within these time windows were filtered to only the first sightings of the MAC at each location.

Measurable route flows between an origin sensor, A, and a destination sensor, B, having mid-point sensors, C and D, were obtained by counting the number of vehicles that travelled on those routes or paths using the detections (Fig. 3).

OD flows from origin A to destination B were also computed for each measurable route pair, using the same filtering process as above to count all filtered detection matches at A and B, but without considering midpoints. OD flow thus contained the aggregate vehicle flows from both measurable routes plus all possible alternative routes. The traffic flow component on alternative routes was then computed by subtracting the shortest or optimal route flow from the OD flow.

These flows were used to analyse the percentage of people taking optimal and suboptimal measurable routes and aggregate alternative routes to give an insight of how drivers chose their routes when length is considered as the main constraint.

To measure behaviour rationality, we propose the Driver Rationality Index (DRI) defined as:

$$DRI^{ij} = \frac{Traffic\ Volume^{ij}_{Shortest\ Path}}{Traffic\ Volume^{ij}_{Total}} \tag{5}$$

where, ***i*** and ***j*** are the origin and destination sensors, respectively. The DRI was calculated using MR and OD flows for each OD pair.

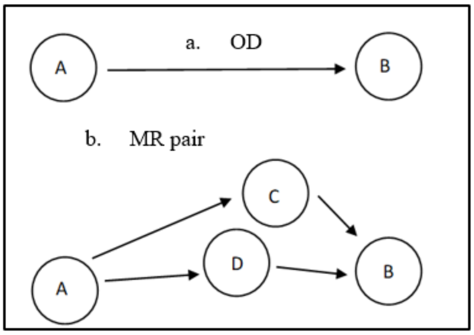

Fig. 4. (a) OD Route, (b) Measurable Route

3. Results

3.1. Baseline Traffic Estimates

`To set the rationality decisions in context, it is useful to examine some baseline traffic estimates. The traffic estimate from the Bluetooth OD route flows and for the specific day examined is shown in Fig. 5. This gives an overview of the traffic in the whole area, including some regions which have no sensors themselves. It should be mentioned, however that these are biased flow estimates because links containing more Bluetooth sensors will show more traffic.

3.2. Bluetooth Sensor Analysis: Measurable Route Pairs

The detected vehicles passing through the 5 selected measurable route pairs were filtered based on the methodology described and their total flows were calculated. In Table 2, the total volumes travelling between the selected OD pairs are presented. The third OD pair, MAC000010119-MAC000010130, contained the highest daily volume (for the selected day) of 532 veh/day, while the least number of vehicles were observed in the fourth and fifth OD pairs, MAC000010121-MAC000010124 and MAC000010123-MAC000010120, with 2 veh/day and 13 veh/day, respectively. The low total volume in those routes has deemed the results obtained from their analysis insignificant. In addition, for the two latter pairs, it should be noted that they represent routes mostly on the same links but on different directions of the main entrance of Chesterfield. The low volumes on those routes for that particular day is an indication of abnormal road conditions, such as a partial road closure for maintenance works.

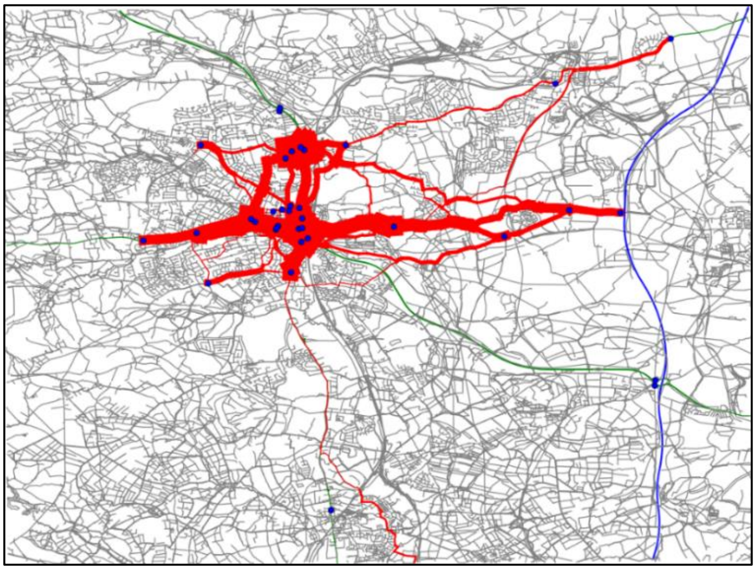

Fig. 5. Flow map between Bluetooth sensors

In Fig. 6 to Fig. 10, the maps for each OD pair from Table 2 are illustrated. Table 3 shows the respective distance and volume of chosen OD routes. Furthermore, in Fig. 10 the relation of traffic flow on each route with its respective distance is depicted.

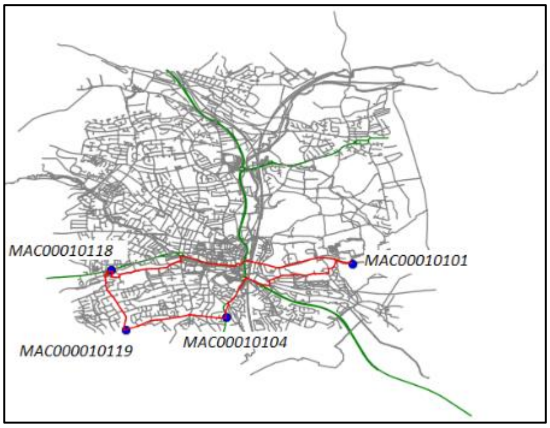

Fig. 6. MRs between MAC00010101-MAC00010119

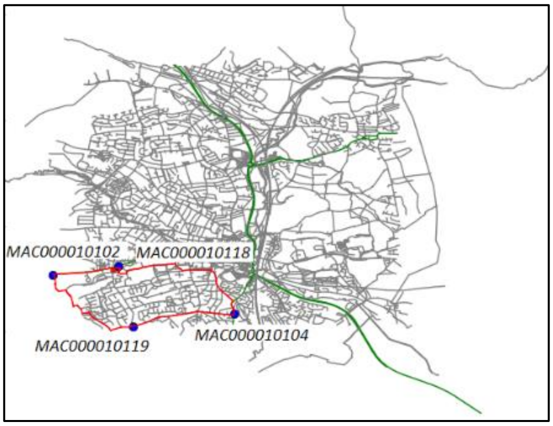

Fig.7. MRs between MAC00010102-MAC00010104

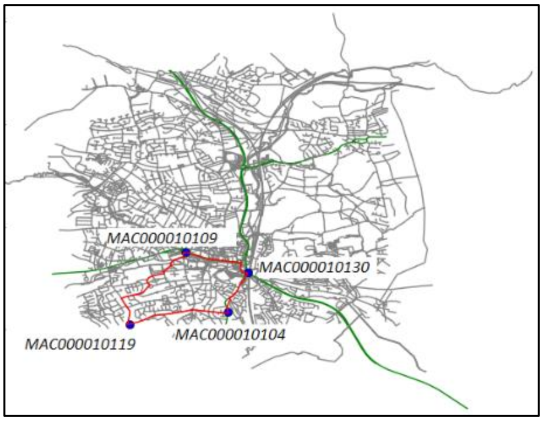

Fig. 8. MRs between MAC00010119-MAC00010130

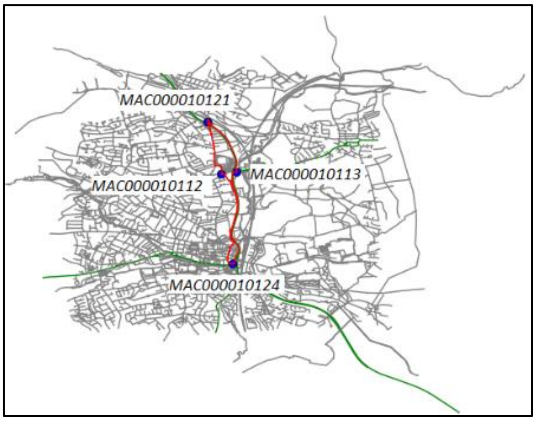

Fig. 9. MRs between MAC00010121-MAC00010124

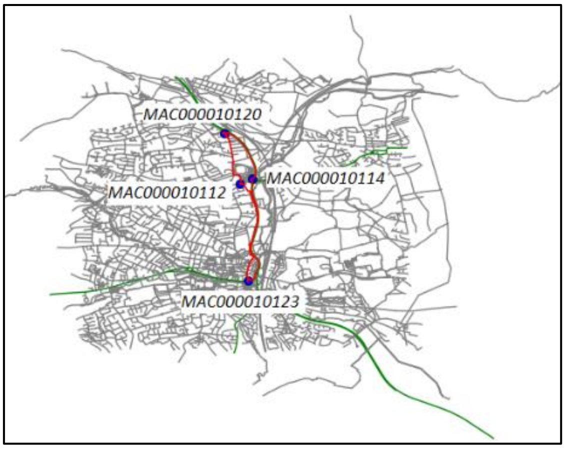

Fig. 10. MRs between MAC00010123-MAC00010120

Table 2. Total volume for OD pairs

OD index	Selected OD pairs	Daily Volume (veh)
1	MAC000010101-MAC000010119	74
2	MAC000010102-MAC000010104	166
3	MAC000010119-MAC000010130	532
4	MAC000010121-MAC000010124	2
5	MAC000010123-MAC000010120	13

Table 3. Distance and volume for MR

OD Index	Selected OD pairs	Bluetooth sensors			Length(m)	Volume (veh/day)
		Origin sensors	Midpoint sensors	Destination sensors		
1	MAC000010101-MAC000010119	MAC000010101	MAC000010104	MAC000010119	4690	20
		MAC000010101	MAC000010118	MAC000010119	5587	3
2	MAC000010102-MAC000010104	MAC000010102	MAC000010119	MAC000010104	3738	66
		MAC000010102	MAC000010118	MAC000010104	3931	59
3	MAC000010119-MAC000010130	MAC000010119	MAC000010104	MAC000010130	2625	379
		MAC000010119	MAC000010109	MAC000010130	3457	10
4	MAC000010121-MAC000010124	MAC000010121	MAC000010112	MAC000010124	3426	0
		MAC000010121	MAC000010113	MAC000010124	3476	2
5	MAC000010123-MAC000010120	MAC000010123	MAC000010112	MAC000010120	3317	0
		MAC000010123	MAC000010114	MAC000010120	3410	12

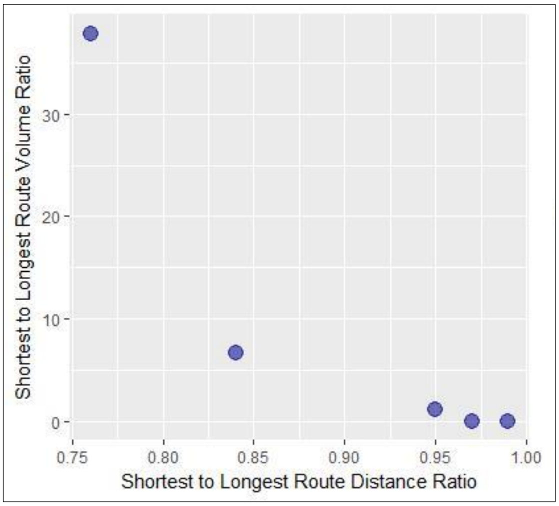

Fig. 11. Relation of traffic flow versus distances on MRs

In Fig. 11, the distance and volume ratios of shortest to longest route pairs are plotted. It can be observed that when the ratio of route distances is much less than unity, the traffic volume is strongly in favour of the shortest route. Specifically, for the two routes with the most significant distance ratio, namely “MAC00010119-MAC000010130” and “MAC000010101-MAC000010119” with distance differences between their respective routes of 832 meters and 901 meters, a significant portion of drivers choose the shortest routes. Drivers on the OD pair “MAC000010102-MAC000010104”, with distance ratio close to unity, were observed to be equally distributed between their chosen MRs. As stated above, due to the low flow numbers on the remaining two OD pairs it is not possible to derive useful insights.

3.3. Bluetooth Sensor Analysis: Comparison with Alternative Routes

The traffic flows on the optimal measurable routes were then compared with the aggregate flows on all possible alternative paths. As presented in Table 4 and illustrated in Fig.12, only in the OD pair MAC000010119-MAC000010130, a large majority of drivers, 71.24%, chooses the measurable shortest path. Nonetheless, the same is not documented in the remaining pairs. Specifically, for MAC000010101-MAC000010119 and MAC000010102-MAC000010104, the percentage of "rational" drivers is 27.03% and 39.76%, respectively. In this study, the other alternative paths cannot be explicitly examined as they do not contain a mid-point Bluetooth sensor.

Table 4. Percentage of drivers on measurable shortest path and on alternative paths

Selected OD pair	Shortest path (%)	Alternative path (%)
MAC000010101-MAC000010119	27.03	72.97
MAC000010102-MAC000010104	39.76	60.24
MAC000010119-MAC000010130	71.24	28.76
MAC000010121-MAC000010124	0.00	100
MAC000010123-MAC000010120	0.00	100

Excluding the last two pairs from the analysis, the DRI for each route was examined and their average values are presented in Table 5. On average, 46% of drivers captured from Bluetooth sensors followed the shortest path, indicating that more than half of drivers might deviate from the shortest route. It should be highlighted, that this analysis is highly dependent on the current sensor locations inside the network. The three selected OD pairs captured only a portion of daily traffic flows in Chesterfield with the vast majority of drivers not being included in the analysis. The results might be significantly different, provided that a larger number of measurable routes with sensors on origin-midpoint-destination could be identified, highlighting the importance of the sensor location problem.

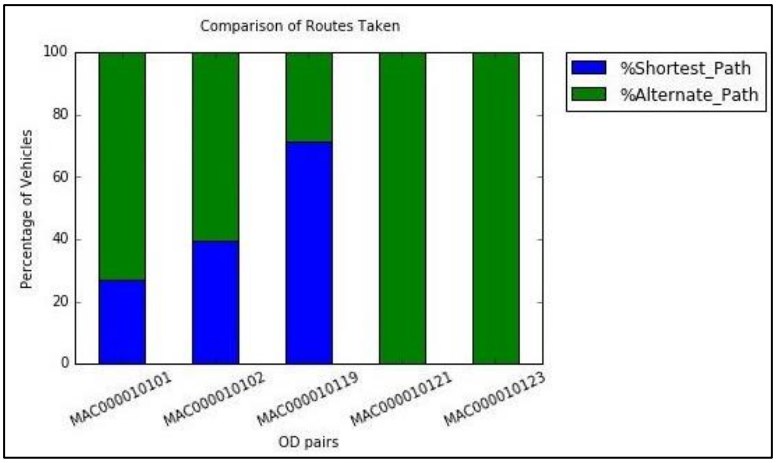

Fig. 12. Percentage of rational drivers on selected OD pairs

Table 5. DRIs for OD pairs

Selected OD pair	Driver rationality index
MAC000010101-MAC000010119	0.27
MAC000010102-MAC000010104	0.40
MAC000010119-MAC000010130	0.71
Average	0.46

4. Discussion

From the individual vehicles detected, on average less than half (46%) showed shortest-route rational driving behaviour (Table 5). Disaggregate analysis on each of the routes examined illustrates the importance of distance difference between an optimal and a suboptimal route. According to the dataset examined, drivers show a high level of rational behaviour when a significant distance difference between the two routes exists (Table 5), with a high percentage (71%) selecting the shortest path, but the percentage of rational drivers decreases (to 27%) as the two route distances become more similar. Therefore, it can be concluded that in the latter case, there are other factors that might influence route choice apart from travel distances. This conclusion can be further validated using a more longitudinal dataset with a higher temporal resolution. Important variables for route choice behaviour are travel cost and travel time. Therefore, models determining the shortest path may also include the aforementioned variables, which can be captured by personalized travel surveys such as mobile application-based or traditional surveys.

These percentages, as functions of route distance ratios or differences, could be used to calibrate parameters for Stochastic Route Choice models. The percentages on optimal routes are surprisingly low and suggest that, if distance was the sole consideration, traffic could be made more efficient in this area by providing more information to drivers about optimal routes, such as via additional signs like Variable Messaging Signs or subsidized satnavs. Nonetheless, the current methodology is limited by the assumption that distance is the only factor considered. The methodology could be extended by using other cost metrics such as travel time, road type and local conditions, such as special events (e.g. accidents). Data fusion of different sources, such as Automatic Traffic Counts (ATC) including Automatic Number Plate Recognition (ANPR) and CCTV footages, data of pavement and weather conditions using advanced Data Fusion techniques like Artificial Intelligence algorithms (e.g. Artificial Neural Networks) and Kalman Filters could be included to reduce the data limitations within the study.

The study was limited by the available data and the size of the network used in the case study to only 5 measurable route pairs, where a general methodology is presented for using passively collected data to directly derive insights on route choice. These were capable of producing meaningful results but much more could be inferred from a similar study on a larger scale. Computational power limitations also restricted the research to use daily flow observations for only one day. Future research should include traffic flow observations for a longer period of time and more measurable route pairs. An important conclusion from this study is the necessity for local authorities to install additional sensors specifically to increase the number of measurable route pairs for their specific networks. In addition, researchers studying general mobility behaviour could combine measurable route pairs from multiple local authorities' data to examine spatial differences of driver rationality. Furthermore, the derived DRI, proposed in this study, is an aggregate measure that can be used to evaluate the cost indifference bands in a bounded rationality route choice model. Future research could focus on an individual level analysis, as well, where a disaggregated DRI could act as a proxy of the individual's rational behaviour, which is the basic assumption made in Random Utility models.

Acknowledgments

The authors would like to thank Derbyshire County Council (DCC), specially Dean Findlay and Neill Bennett for provision of unrestricted access to their traffic database and for their interest in the research. Thanks also to Lawrence Duncan, PhD researcher at the Institute for Transport Studies, University of Leeds, for his support and suggestions.

References

Araghi, B.N., Olesen, J.H., Krishnan, R., Christensen, L.T., Lahrmann, H., 2014. Reliability of Bluetooth technology for travel time estimation. Journal of Intelligent Transportation Systems: Technology, Planning and Operations 19, 240-255.

Bachmann, C., Roorda, M.J., Abdulhai, B., Moshiri, B., 2013. Fusing a Bluetooth Traffic Monitoring System with Loop Detector Data for Improved Freeway Traffic Speed Estimation. Journal of Intelligent Transportation Systems: Technology, Planning and Operations. 17.2, 152-164.

Barceló, J., Montero, L., Marques, L., Carmona, C., 2010. Travel time forecasting and dynamic origin-destination estimation for freeways based on Bluetooth traffic monitoring. Transportation Research Record 2175, 19-27.

Beckmann, M.J., McGuire, C.B., and Winsten, C.B., 1956. Studies in the Economics of Transportation. Yale University Press, New Haven, Conn.

Ben-Akiva, M., Lerman, S. R., 1985. Discrete choice analysis: Theory and application to travel demand. MIT Press.

Bwambale, A., Choudhury, C.F., Hess, S., 2017. Modelling trip generation using mobile phone data: A latent demographics approach. Journal of Transport Geography.

Crawford, F., Watling, D.P., Connors, R.D., 2018. Identifying road user classes based on repeated trip behaviour using Bluetooth data. Transport Research Part A 113, 55-74.

Daganzo, C. F., Sheffi, Y., 1977. On stochastic models of traffic assignment. Transportation Science 11.3, 253-274.

Di, X., Liu, H. X., Pang, J., Ban, X., 2013. Boundedly rational user equilibria (BRUE): Mathematical formulation and solution sets. Procedia – Social and Behavioral Sciences 80, 231-248.

Ezlsäßer, M., Wirtz, B. W., 2017. Rational and emotional factors of customer satisfaction and brand loyalty in a business-to-business setting. Journal of Business & Industrial Marketing 32.1, 138-152.

El-Seoud, S.A., El-Sofany, H.F., Abdelfattah, M., Mohamed, R., 2017. Big Data and Cloud Computing: Trends and Challenges. International Journal of Interactive Mobile Technologies 11.2, 34-52.

Fox, C., Billington, P., Paulo, D., Cooper, C., 2010. Origin-Destination analysis on the London orbital automated number plate recognition network. European Transport Conference, Glasgow, UK.

Fox, C., 2018. Data Science for Transport. A Self-Study Guide with Computer Exercises. Springer, UK.

Gong, L., Liu, X., Wu, L., Liu, Y., 2016. Inferring trip purposes and uncovering travel patterns from taxi trajectory data. Cartography and Geographic Information Science 43.2, 103-114.

Horowitz, J. L., 1984. The stability of stochastic equilibrium in a two-link transportation network. Transportation Research Part B 18.1, 13–28.

Hyunmyung, K., Yongtaek, L., 2012. A Day-to-day route choice model based on drivers' past experience. KSCE Journal of Civil Engineering 16.7, 1267-1279.

Iqbal, S.M., Choudhury, C.F., Wang, P., Gonzalez, M.C., 2014. Development of origin-destination matrices using mobile phone call data. Transportation Research Part C Emerging Technologies 40, 63-74.

Ishwarappa, Anuradha, J., 2015. A Brief Introduction on Big Data 5Vs Characteristics and Hadoop. Procedia Computer Science 48, 319-324.

Kobayashi, K., 1994. Information, rational expectations and network equilibria — an analytical perspective for route guidance systems. The Annals of Regional Science 28.4, 369-393.

Mahrsi, M.K.E., Côme, E., Baro, J., Oukhellou, L., 2014. Understanding passenger patterns in public transit through smart card and socioeconomic data. 3rd International Workshop on Urban Commuting. August 14, 2014, New York City, USA.

Marsden, G., Frick, K. T., May, A. D., Deakin, E., 2012. Bounded rationality in policy learning amongst cities: lessons from the transport sector. Environment and Planning Part A, 44.4, 905-920.

Martchouk, M., Mannering, F.L., Bullock, D.M., 2011. Analysis of Freeway Travel Time Variability Using Bluetooth Detection. Journal of Transportation Engineering 137.10, 697-704.

Nakayama, S., Kitamura, R., Fujii, S., 2001. Drivers' route choice rules and network behavior: Do drivers become rational and homogeneous through learning? Transportation Research Record 1752.1, 62-68.

Ortuzar J.D.D., Willumsen, L.G., 2011. Modelling Transport. John Wiley & Sons.

Persky, J., 1995. Retrospective: The ethology of homo economicus. The Journal of Economic Perspectives 9.2, 221-231.

Prashker, J.N., Bekhor, S., 2004. Route Choice Models used in the Stochastic User Equilibrium Problem: A Review. Transport Reviews: A Transnational Transdisciplinary Journal 24.4, 437-463.

Simon, H., 1991. Bounded rationality and organizational learning. Organization Science. 2.1, 125–134.

Sivak, M., 2002. How common sense fails us on the road: Contribution of bounded rationality to the annual worldwide toll of one million traffic fatalities. Transportation Research Part F 5.4, 259-269.

Yang, F., Jin, P.J., Cheng, Y., Zhang, J., Ran, B., 2015. Origin-Destination Estimation for Non-Commuting Trips Using Location-Based Social Networking Data. International Journal of Sustainable Transportation 9, 551-564.

Yin, Y., Ieda, H., 2001. Assessing performance reliability of road networks under nonrecurrent congestion. Transportation Research Record Journal of Transportation Research Board 1771.1, 148-155.