Measuring spatial intrapersonal variability of road users using Bluetooth data

F. Crawford[a] *, D. P. Watling[b] and R. D. Connors[b]
[a] *Centre for Transport and Society, University of the West of England, Coldharbour Lane, Bristol, U.K.*
[b] *Institute for Transport Studies, University of Leeds, U.K.*
* *Email: fiona.crawford@uwe.ac.uk*

Keywords: Bluetooth data; intrapersonal variability; regularity; spatial variability; spatial diversity; sequence alignment; detection rate; data mining; association rules

In reviewing developments in transportation research, both Schlich and Axhausen (2003) and Heinen and Chatterjee (2015) have highlighted the disproportionate amount of attention paid to the variability between people, in contrast to the variability in an individual's behaviour from day-to-day. The latter, known as intrapersonal variability, is important because the regularity (or irregularity) of an individual's trips can provide an insight into their transport needs (Schlich and Axhausen, 2003). This research focuses on spatial aspects of intrapersonal variability as this provides information about the familiarity of travellers with different parts of the network, for example alternative routes or access to destinations in different parts of the city. This could inform models of traveller response to network disruptions. Information about the different parts of the network used over time could also inform the development and modelling of transport policies involving spatial boundaries such as fare zones or congestion charging zones. Measures of spatial intrapersonal variability also provide insights into the predictability of traveller behaviour and thus could inform the parameter values for day-to-day dynamical models which include learning mechanisms, for example the "switching choice probability" (Cantarella and Cascetta, 1995) which relates to travellers reconsidering, but not necessarily changing, their previous route choice. Understanding intrapersonal variability can also inform the development of user classes, for example based on attitude to risk (Shao, Lam and Tam, 2006) or information availability (Han *et al.*, 2018), and the corresponding parameters for modelling choices.

Spatial intrapersonal variability is increasingly relevant due to the impacts of Information and Communication Technologies (ICT) on travel. Changes in ICT have meant that more people can work (at least occasionally) remotely (Felstead, 2012) and real-time information now plays a greater role in route choice. ICT is also enabling new mobility services, many of which operate on an on-demand basis. For such services to be economically viable, it is crucial that everything from the charging structures to the organisation of vehicles and staff are designed based on traveller needs, which includes their multiday behaviour. Telecommunication technologies are also having an impact upon the data we can collect relating to mobility.

For this research, Bluetooth data was identified as the most suitable source of data as it can be passively collected in large quantities, it is closely linked to the road network and it is a relatively cheap source of data which has been implemented in many cities. Data from fixed Bluetooth detectors is becoming increasingly popular for measuring travel times on the road network (Haseman, Wasson and Bullock, 2010; Hainen *et al.*, 2011; Moghaddam and Hellinga, 2013), particularly in urban areas, and has also been used in OD estimation (Barcelö *et al.*, 2010; Carpenter, Fowler and Adler, 2012).

To the knowledge of the authors, Bluetooth data has only been used to examine intrapersonal variability (or more correctly intra-device variability) by Crawford et al. (2018). In that paper, a method was proposed for estimating road user classes based on repeated trip behaviour including trip frequency, and both spatial and time of day intrapersonal variability. Spatial variability was examined by firstly grouping together similar trip sequences, by comparing trip sequences using Sequence Alignment then using k-means clustering to obtain a more manageable number of spatial categories. Spatial variability for each device was then measured by counting the number of spatial categories their trips were classified into and the percentage of their trips which were in their most commonly used spatial category.

Previous work has not, however, taken into account the fact that different sensor locations will result in different probabilities of detecting a passing Bluetooth device. Also, whilst Crawford et al. (2018) used real-world data to demonstrate the clustering of trip sequences, the findings related to broader road user classifications based on repeated trip behaviour and so methods of gaining useful insights from the spatial variability data alone were not considered. The current research seeks to fill these two gaps in the literature.

Methodological approach

This research builds upon the spatial clustering methodology presented in Crawford et al. (2018). Firstly, a model is developed to estimate the probability of detection at a sensor location based on the road characteristics. The model is developed by measuring detection rates at the central sensor within 'trios' of Bluetooth detectors, where the shortest path between the two outermost sensors (A and C) passes the central sensor (B), and then relating these detection rates to road characteristics at the central sensor. A modified version of the Sequence Alignment procedure used in Crawford et al. (2018) was then developed which incorporates these sensor-specific probabilities of detection when comparing trip sequences. The Sequence Alignment procedure was used to measure the distance between trip sequences so that they could be clustered into groups with similar spatial characteristics.

One way to examine the spread of each person's trips across the set of spatial categories is to use the Herfindahl-Hirschman Index (HHI). The HHI is often used for examining the market share of different businesses and it is also known as Simpson's Diversity Index in ecology, where it is used to measure species diversity. The HHI has also been used to measure intrapersonal variability in travel choices (Heinen and Chatterjee, 2015; Susilo and Axhausen, 2014). In this research, a normalised HHI is calculated for each person separately. The values for the normalised HHI fall within the interval from 0 to 1, where a value of 0 equates to making an equal number of trips in each possible spatial category during the year, and a value of 1 equates to all trips being within a single spatial category.

The normalised Herfindahl-Hirschman Index can be calculated as follows:

$$H^* = \frac{\left(\sum_{i=1}^{M} {s_i}^2\right) - {}^{1}/_{M}}{1 - {}^{1}/_{M}}$$

where M is the total number of spatial categories, and s_i is the proportion of (this traveller's) trips belonging to spatial category i.

By having a single measure of spatial diversity for each traveller, intrapersonal comparisons can be made, for example the spatial diversity of trips on weekdays versus weekend days.

An alternative way to examine the spatial nature of the trips is to assess the relationships between the spatial categories, not in terms of geographic proximity (as was considered in the Sequence Alignment process), but in terms of the proportion of users who frequently make trips from both of two given categories, for example.

This can be examined using Association Analysis (Tan, Steinbach and Kumar, 2014, Chapters 6 and 7). This approach, also known as Market Basket Analysis, has been used previously in transportation research, for example by Pande and Abdel-Aty (2009) to examine relationships between crash characteristics. Traditionally, the method has been used to identify a set of association rules which provide insights into the products which are commonly purchased together in shops. In the current application, the method is used to examine whether regular travellers who commonly make trips within spatial cluster X also make trips within another (given) spatial cluster Y. If an association is identified, this rule would be denoted by: $X \rightarrow Y$. The association rules are then sorted based on their confidence measures (namely the proportion of people who make trips in cluster X who also make trips in cluster Y).

Expected results

A case study application was undertaken using data from eight sites over a one year period (1/1/2015 to 31/12/2015). The sites are in and around the town of Wigan in northwest England (Figure 1). Devices which recorded 52 or more trips within the case study area during the year were retained for the analysis, resulting in 9,564 devices in total. Estimated sensor-specific detection rates are shown in Table 1.

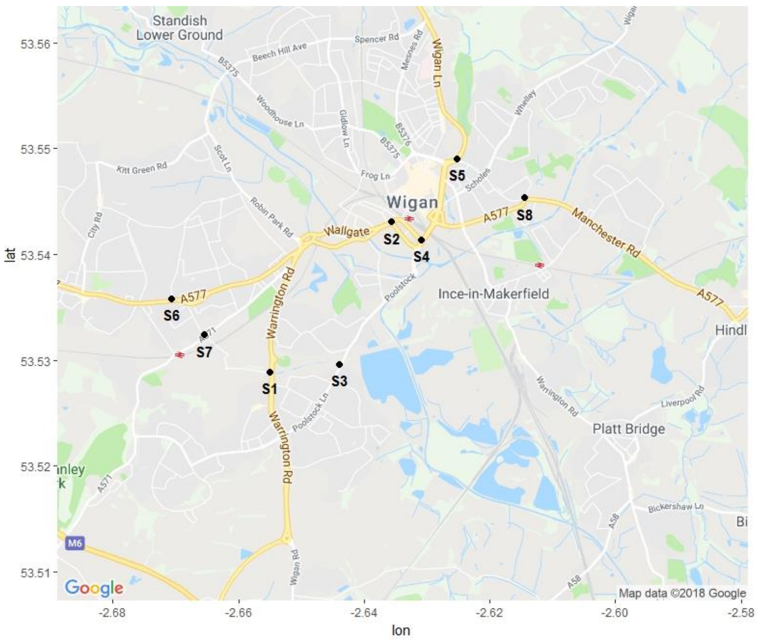

Table 1: Estimated detection rates

Ref	Estimated detection rate (%)
S1	73
S2	71
S3	88
S4	76
S5	66
S6	88
S7	74
S8	57

Figure 1: Case study area including Bluetooth detector locations

After grouping trip sequences as discussed above, the spatial diversity in trips undertaken by each person was calculated. Figure 2 shows the distribution of the HHIs across all devices. The peak has a HHI of approximately 0.08 and one way of obtaining this value is to use 10 of the 55 spatial categories and to use them all equally often.

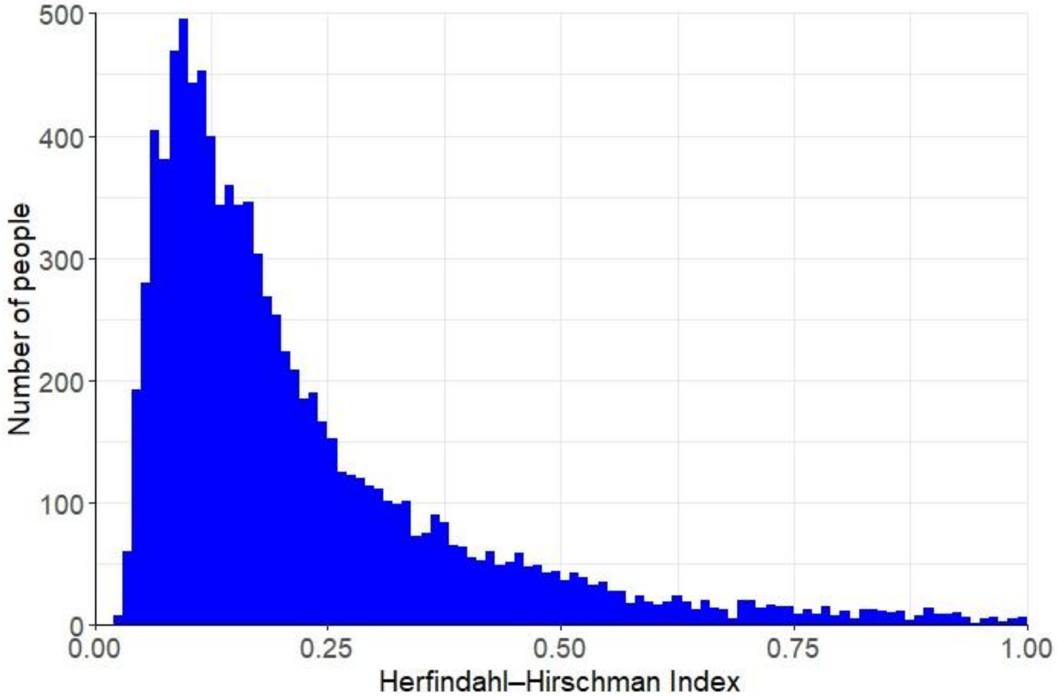

Figure 2: Histogram of the normalised Herfindahl–Hirschman Index

Of the people making at least 10 trips on weekdays and at least 10 trips on weekdays (6,664 people), 57% had a larger HHI for weekends than weekdays. This suggests that people are more likely to have a more even distribution of trips across multiple spatial categories on weekdays.

Association rules were also constructed for the relationships between spatial categories. Additional processing was required as 16 of the categories had a paired category which approximately related to the reverse journey. Unsurprisingly, the strongest associations were identified between these pairs of trips. After accounting for these relationships, the analysis highlighted a key group of trips (one spatial category) in the town centre which provided the only spatial overlap between people travelling to and from the town centre from the south and travellers using a combination of all other routes into and within the centre.

References

Arentze, T. and Timmermans, H. (2005) Modelling learning and adaptation in transportation contexts. *Transportmetrica* [online]. 1 (1), pp. 13–22. Available from: https://doi.org/10.1080/18128600508685638doi:10.1080/18128600508685638.

Barcelö, J., Montero, L., Marqués, L. and Carmona, C. (2010) Travel Time Forecasting and Dynamic Origin-Destination Estimation for Freeways Based on Bluetooth Traffic Monitoring. *Transportation Research Record* [online]. 2175 (1), pp. 19–27. Available from: https://doi.org/10.3141/2175-03doi:10.3141/2175-03.

Cantarella, G.E. and Cascetta, E. (1995) Dynamic Processes and Equilibrium in Transportation Networks: Towards a Unifying Theory. *Transportation Science* [online]. 29 (4), pp. 305–329. Available from: https://doi.org/10.1287/trsc.29.4.305doi:10.1287/trsc.29.4.305.

Carpenter, C., Fowler, M. and Adler, T.J. (2012) Generating Route-Specific Origin–Destination Tables Using Bluetooth Technology. *Transportation Research Record* [online]. 2308 (1), pp. 96–102. Available from: https://doi.org/10.3141/2308-10doi:10.3141/2308-10.

Crawford, F., Watling, D.P. and Connors, R.D. (2018) Identifying road user classes based on repeated trip behaviour using Bluetooth data. *Transportation Research Part A: Policy and Practice*. 113 pp. 55–74. doi:10.1016/j.tra.2018.03.027.

Felstead, A. (2012) Rapid change or slow evolution? Changing places of work and their consequences in the UK. *Journal of Transport Geography* [online]. 21 pp. 31–38. Available from: http://www.sciencedirect.com/science/article/pii/S0966692311001530doi:https://doi.org/10.1016/j.jtrangeo.2011.10.002.

Hainen, A.M., Wasson, J.S., Hubbard, S.M.L., Remias, S.M., Farnsworth, G.D. and Bullock, D.M. (2011) Estimating Route Choice and Travel Time Reliability with Field Observations of Bluetooth Probe Vehicles. *Transportation Research Record* [online]. 2256 (1), pp. 43–50. Available from: https://doi.org/10.3141/2256-06doi:10.3141/2256-06.

Han, L., Sun, H., Wang, D.Z.W. and Zhu, C. (2018) A stochastic process traffic assignment model considering stochastic traffic demand. *Transportmetrica B: Transport Dynamics* [online]. 6 (3), pp. 169–189. Available from: https://doi.org/10.1080/21680566.2016.1240051doi:10.1080/21680566.2016.1240051.

Haseman, R.J., Wasson, J.S. and Bullock, D.M. (2010) Real-Time Measurement of Travel Time Delay in Work Zones and Evaluation Metrics Using Bluetooth Probe Tracking. *Transportation Research Record* [online]. 2169 (1), pp. 40–53. Available from: https://doi.org/10.3141/2169-05doi:10.3141/2169-05.

Heinen, E. and Chatterjee, K. (2015) The same mode again? An exploration of mode choice variability in Great Britain using the National Travel Survey. *Transportation Research Part A: Policy and Practice* [online]. 78 pp. 266–282. Available from: http://www.sciencedirect.com/science/article/pii/S0965856415001433doi:https://doi.org/10.1016/j.tra.2015.05.015.

Moghaddam, S.S. and Hellinga, B. (2013) Quantifying Measurement Error in Arterial Travel Times Measured by Bluetooth Detectors. *Transportation Research Record* [online]. 2395 (1), pp. 111–122. Available from: https://doi.org/10.3141/2395-13doi:10.3141/2395-13.

Pande, A. and Abdel-Aty, M. (2009) Market basket analysis of crash data from large jurisdictions and its potential as a decision support tool. *Safety Science* [online]. 47 (1), pp. 145–154. Available from: http://www.sciencedirect.com/science/article/pii/S0925753507001877doi:https://doi.org/10.1016/j.ssci.2007.12.001.

Schlich, R. and Axhausen, K.W. (2003) Habitual travel behaviour: Evidence from a six-week travel diary. *Transportation* [online]. 30 (1), pp. 13–36. Available from: https://doi.org/10.1023/A:1021230507071doi:10.1023/A:1021230507071.

Shao, H., Lam, W.H.K. and Tam, M.L. (2006) A Reliability-Based Stochastic Traffic Assignment Model for Network with Multiple User Classes under Uncertainty in Demand. *Networks and Spatial Economics* [online]. 6 (3), pp. 173–204. Available from: https://doi.org/10.1007/s11067-006-9279-6doi:10.1007/s11067-006-9279-6.

Susilo, Y.O. and Axhausen, K.W. (2014) Repetitions in individual daily activity--travel--location patterns: a study using the Herfindahl--Hirschman Index. *Transportation* [online]. 41 (5), pp. 995–1011. Available from: https://doi.org/10.1007/s11116-014-9519-4doi:10.1007/s11116-014-9519-4.

Tan, P.-N., Steinbach, M. and Kumar, V. (2014) *Introduction to Data Mining*. Harlow, Essex: Pearson.