Assessing people travel behavior using GPS and open data to validate neighbourhoods characteristics

Oliveti, M; van der Spek, Stefan; Quak, Wilko

Publication date
2016

Document Version
Final published version

Published in
Proceedings of the the 19th AGILE International Conference on Geographic Information Science

Assessing people travel behavior using GPS and open data to validate neighbourhoods characteristics

Matilde Oliveti	Stefan van der Spek	Wilko Quak
Delft University of Technology	Delft University of Technology	Delft University of Technology
Julianalaan 134,	Julianalaan 134,	Julianalaan 134,
2628 BL Delft,	2628 BL Delft,	2628 BL Delft,
the Netherlands	the Netherlands	the Netherlands
matilde.oliveti@libero.it	S.C.vanderSpek@tudelft.nl	c.w.quak@tudelft.nl

Abstract

With the help of technologies such as GPS tracks, GIS and open data is now possible to study people travel behaviour in a new way. Nowadays, large datasets can be easily handled thanks to databases and better visualized using GIS. Moreover, the availability of GPS data, open data and VGI makes accessible a lot of new information, which was not obtainable before. This research is about the analysis of mobility patterns in different neighbourhoods in three cities in the Netherlands. The study is based on the validation of mobility theoretical performances by the actual performances measured through the analysis of GPS logs of households, who live in the neighbourhoods. A series of spatial indicators based on proximity, density and accessibility are computed in order to assess the theoretical performances of the neighbourhoods. To achieve this task, information about built environment characteristics and infrastructure networks is retrieved from OpenStreetMap, and other datasets. In the end the neighbourhoods are classified into five classes, according to the different levels of performances in terms of sustainable mobility. In such a way, it is possible to better understand the key factors that influence actual people travel patterns, providing policy makers with accurate information about the real movement of people.

Keywords: GIS, OpenStreetMap, GPS tracking, Actual people movement, Mobility

1 Introduction

Sustainable urban development tops political and research agendas worldwide. [1]. The existing land use and mobility patterns of our urban environment are not sustainable. Today more and longer trips are mostly by private car, with high consumption of energy and air pollution, further leading to sprawl, and ultimately bringing the need for new policies and strategies for our cities [2].

Policy makers and urban planners demand to understand, to measure and to monitor the dynamics of the contemporary patterns of mobility. Describing urban areas using aggregate statistics at neighbourhood or city levels using general characteristics does not lead to a complete understanding of actual mobility patterns. Furthermore, the efforts and the studies carried out in this field until now are not consistent [2] and often lack attention to the particular problems regarding knowledge application in spatial planning [3].

With the help of technologies such as GPS tracks, GIS and open data is now possible to study people travel behaviour in a new way. Nowadays, large datasets can be easily handled thanks to databases and better visualized using GIS. Moreover, the availability of GPS data, open data and VGI makes accessible a lot of information, which was not obtainable before.

The general scope of this research is comparing the performance of different neighbourhoods in terms of mobility patterns, where mobility patterns stands for “where do people actually go?”, “which mode of transportation do they use?”, “what are their main destinations?”. In particular, one of the aims of this study is to assess mobility patterns measuring the following spatial characteristics:

- proximity: the distance to the nearest element of the mobility network infrastructure of each mode.
- density: the intensity of a given mobility mode or land use activity.
- accessibility: the importance of a location based on the distance to other locations and to opportunities associated with activities.

This study intends to develop a standard procedure that can be applied and re-used in urban studies, integrating the current technology in order to assess people travel behaviour in the real world. The final scope of this research is both technical and societal. Technical, since the research aims to understand to which extent GPS tracks and OpenStreetMap data can be integrated within the framework; and societal, since the result of this study will have societal implications, as neighbourhood performances are directly related to people and their lifestyles. In the end, the outcome of this research will be a measure of what is ‘sustainable’ and what is ‘unsustainable’ in each neighbourhood in terms of mobility. Specifically, we assess and compare mobility patterns in 10 different neighbourhoods in the Netherlands.

2 Related work

Nowadays, tracking technologies, such as GPS, offers huge possibilities for studying human activity patterns in time and space and in new ways [4]. Many studies in this field often lack attention to the particular problems regarding knowledge application in the field of planning [5]. Furthermore, not so many works about people travel behaviour use GPS tracks data, since most of them tend to use traditional methods, like paper or phone recall surveys. However, it has been shown that data collected using these methods deviate systematically from actual behaviour [5]. On the contrary, tracking technology provides real behaviour, not just indicators based on stated preferences. In this context, the book 'Urbanism on Track' [6], derived by an international expert meeting held in 2007 at TUDelft, deeply reflects on the state of the art of the applications of new tracking technologies, such as GPS and cell phone, in urban design and spatial planning processes. Among others, Shoval [4] presents two cases in which he used GPS data for research on outdoor mobility of elderly people with cognitive disorders and research on the user-density of an Israeli heritage site.

In the past, efforts were attempted by governments in order to reduce car mobility. For instance, in the '90s the Dutch government introduced the Vinex policy with the hope to influence peoples' travel behaviour by creating urban landscapes that invite people to use alternative modes of transportation. However, the results of this policy have not been very successful, as today the new districts developed are still too much oriented towards auto mobility [7].

Recently, several researches have been carried out to study travel patterns of the inhabitants of a small number of neighbourhoods in order to investigate to what extent certain spatial features of neighbourhoods provide an explanation for mobility [8, 9]. The present research is built on this literature.

Travel behaviour research shows that the characteristics of the built environment in residential areas influence peoples' daily travel behaviour; however, the extent of this influence is subject to debate [5]. Cervero and Kockelman [10] in their work came to the conclusion that compact neighbourhoods can degenerate vehicle trips and encourage non-motorized travel. On the contrary, Meurs and Haaijers [8] stated that mobility is a secondary matter, influenced by several important elements, such as lifestyles, spatial characteristics, and accessibility.

In the literature, different assessment models and criteria are used to measure mobility patterns and neighbourhood characteristics. In most of the cases, indicators and composite indexes are computed, and statistical analyses are performed. Factor analysis [10,11,12], regression analysis [7,13] and Likert-type scaling [5] are found to be really popular among these studies. The choice of the statistical method really depends on the variables utilized [14].

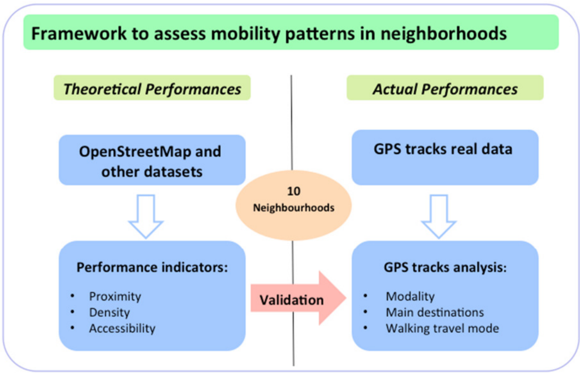
Cervero and Kockelman [10] in their work used factor analysis to linearly combined variables like 'average sidewalk width', and 'intensity of planting strips and street trees' to represent the dimension of 'pedestrian-oriented design'. They classified the indicators according to three principal dimensions (density, diversity, and design) with which the built environment influence travel demand. Hilbers and Snellen [7] used a series of parameters for evaluating the mobility impacts of the Dutch Vinex policy, like distance to several daily facilities, land use mix within the surrounding area, and distance to nearest motorway exit.

3 Methodology

3.1 Framework

The main steps carried out in this work are presented in a framework used to assess mobility patterns in different neighbourhoods (Figure 1). Performance indicators, namely quantitative and qualitative spatial measurements, are used to achieve this task. On the left side, theoretical performances are addressed, using data from OpenStreetMap and other datasets to compute a series of mobility indicators; on the right side, actual mobility patterns, derived from the analysis of GPS real data, are used to validate the theoretical performances.

Figure 1: Framework for assessing mobility patterns: performance indicators validation by GPS real track data.



Theoretical performance indicators were chosen after an extensive literature review. In the end, 17 indicators (Table 1) were selected and organized into three groups (proximity, density, and accessibility), based on the classification Gil and Read [13] made in their work.

Table 1 Theoretical Performance Indicators.

Measure	Indicator description
Proximity	Distance (shortest path) to closest railway station (km)
	Distance (shortest path) to the closest bus stop (m)
	Distance (shortest path) to closest motorway exit (km)
	Distance (shortest path) to several daily facilities (supermarket, school, etc.) (m)
	Distance (shortest path) to city centre (km)
Density	Population density (residents/ km2)
	Road, public transport, cycle and walk network density (km/km2)

	Parks and green areas density (parks/ km2)
	Buildings density (buildings/ km2)
	Land use mix per neighbourhood (% land use type/total area)
	Buildings function density (office, residential, industrial, etc.) (building function type/total n° buildings)
Accessibility	City centre accessibility (travel distance/travel time)
	Ratio n° buildings with railway station within 1 km (Network/Euclidean distance)
	% Buildings with railway station within 10 min travel time by car, bike and walking
	Ratio n° buildings with bus stop within 500 m (Network/Euclidean distance)
	N° shops within 10 min travel time by car, bike and walking
	N° schools within 2 and 5 km (Euclidean distance)

Proximity indicators are mainly related to measures like the distance to the closest railway station, bus stop, supermarket, etc. Density indicators are measures of intensity, such as the land use mix, green area density, buildings density, etc. Accessibility indicators represent the mean distance to activities and facilities, like the percentage of buildings within a railway station, etc.

3.2 Datasets

For this research, the GPS tracks are derived from a previous GPS survey conducted in 2012 [5,15] for an urban analysis project by Paul van de Coevering of the Urban and Regional Development Section of TUDelft. The data consisted of 40 million GPS points with coordinates (x, y, z) and time (t), recorded every 5 seconds by more than 800 people living in Amersfoort, Veenendaal and Zeewolde. The reasons behind the choice of these neighbourhoods were the consistency of the data of the GPS survey, and the diversity of the three cities in size, urban form and mobility facilities.

In addition, OpenStreetMap (OSM), a digital map database of the world built through crowdsourced Volunteered Geographic Information (VGI), was chosen for retrieving information about the infrastructure networks. OSM data is freely available, it has universal coverage and a rich feature set that covers all modes of transportation. In the case of the Netherlands, respect to other datasets, OSM seemed the most appropriate choice since it offers better semantic accuracy and it can have a very good level of completeness. Finally, several Dutch datasets were used in this study: BAG (Basisregister Adressen en Gebouwen) for addresses and buildings data in order to geolocate households, BBG (Bestand Bodemgebruik) for land use information, and CBS (Central Bureau of Statistics) for retrieving data about the population (income, age, house value, etc.).

4 Data processing and analysis

4.1 Theoretical Performance Indicators

Before being able to compute theoretical performance indicators, it was necessary to extract the information related to the infrastructure network. OSM data was accessed through the database PostreSQL and in the end three separated networks were created (car, cycle and walk networks), by selecting the specific osm_highway type. The theoretical performance indicators were computed using different tools and plugins in PostgreSQL and QGIS (e.g. road graph, buffer, distance matrix, etc.). Moreover, for accessibility indicators a PostgreSQL/PostGIS extension called PgRouting, which provides routing functionalities, was used.

All the indicators were quantified considering different mode of transportation (car, cycle and walk) and using the different networks created through OSM.

For proximity indicators two types of distances were implemented: network distance, which measures the length of the shortest street network linking an origin and a destination; and Euclidean distance, which consists in the length of the straight geometric line, linking an origin and a destination. The network distance to the closest destination was computed using the Road Graph plugin in QGIS.

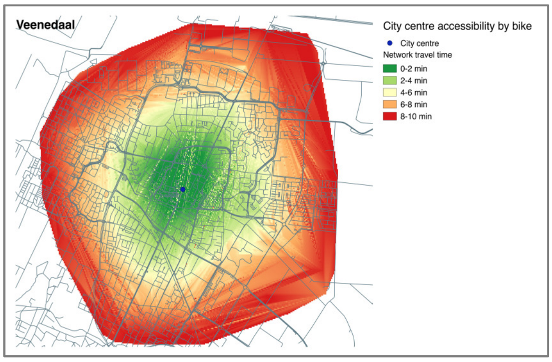
Regarding density indicators, a land use entropy index was used to quantify homogeneous land use in a given area. In total, over 30 land use classes were considered, such as residential, retail, industrial, parks, sport, etc.

The measure of land-use mix follows [11]:

$$(LUM = -\sum_{i=1}^{n} p_i \ln p_i / \ln n)$$

Finally, in order to compute accessibility indicators, a PgRouting function called pgr_drivingdistance was used to query the database and to get as output a cost attribute for all the nodes of the network, based on the distance to a certain location. Not only travel distance was implemented, but also travel time was taken into account. In the end, all the nodes, each one with a certain distance/time cost, were interpolated in QGIS and a heatmap was obtained (Figure 2).

Figure 2: Heatmap showing city centre accessibility by bike in Veenendaal.



5 Validation

5.1 GPS tracks analysis

After the performance indicators implementation, the next step consisted in validating the theoretical performances using GPS real data. The GPS tracks used in this research were already pre-processed and classified in an interpretation-validation process made in previous studies [9,16]. No effort was spent here in GPS data classification and trip segmentation, since it was out of the scope of this study. GPS data was already split in travel modes by an algorithm that uses mean and maximum speed [16]. Before the GPS log was analysed, a series of cleaning operations needed to be performed. The amount of data was reduced, selecting only the GPS track points of the residents within the neighbourhoods.

Travel modes were investigated counting GPS track points for each modality (car, bicycle, walk, etc.). By filtering the track points based on the postal code and on the timestamp, only households' single visits were selected and analysed. In such a way, the most visited locations could be highlighted on a map, showing the main destinations of households per neighbourhood, using a colour scale to display the different intensities (Figure 3).

Figure 3: Main destinations of inhabitants of Amersfoort city centre.

5.2 Statistical analysis

In order to be able to compare the various indicators with the output of the GPS analysis, variable transformation was necessary. The first step comprised the data normalization: the data was normalized using the z-scores method which is one of the most commonly used normalization method in data mining. This method converts all indicators to a common scale with an average of zero and standard deviation of one. The operation was performed in the software SPSS.

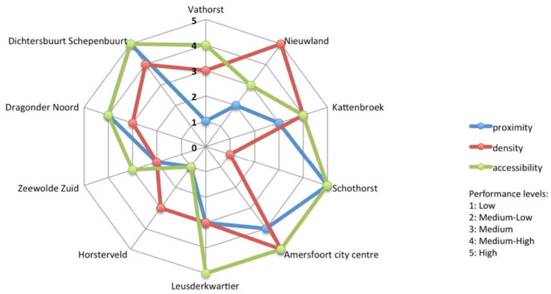
The second step consisted of data clustering, assembling data in classes. Different clustering methods were tested, but in the end no large differences were found and so Natural Breaks clustering method was chosen.

Finally, a correlation test was carried out in order to see if the indicators and the results of the GPS analysis were correlated somehow. Several methods exist, but Spearman correlation was used in this case since there were only 10 cases and most of the variables were not normally distributed.

With the aim to assess the overall performances of each neighbourhood, a score between 1 and 5 was assigned to each group of indicator (proximity, density and accessibility). Each score represented a different level of performances: low, medium-low, medium, medium-high and high.

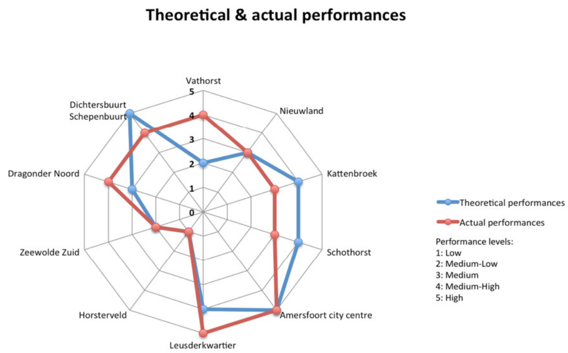
As we can clearly see in the spider diagram (Figure 4), neighbourhoods have different scores in term of proximity, density and accessibility.

Figure 4: Theoretical performances in terms of proximity, density and accessibility.



The same procedure was repeated with the actual performances derived by the GPS analysis. In this way it was possible to directly compared theoretical performances with actual performances (Figure 5).

Figure 5: Comparison between theoretical performances and actual performances.



6 Discussion and results

The analysis presented in the sections above has led to several conclusions and interesting results.

The output of the GPS analysis of actual performances is almost in line with the one coming from the theoretical performances. In 4 out of 10 neighbourhoods actual and theoretical performances do perfectly match; in the other cases they are slightly different, since the values are assigned to neighbouring classes (Figure 5). This small mismatch can be due to an overestimation of the services and facilities presented in the neighbourhood, or maybe it can be related to the classification method itself.

The comparison between network and Euclidean distance has given a good indication of how efficient is the infrastructure network; however, this only works if considering each indicator separately as no general trend was found. For instance, in Nieuwland the network seems really efficient if looking at the distance to the closest school, but it is really inefficient in case of the distance to the closest supermarket. These differences are probably related to the structure of the infrastructure network extracted from OpenStreetMap and to the type of roads selected (see Section 4.1).

BBG land use dataset, utilized for retrieving information about land use is barely valid for the analysis at the neighbourhood level. Since in all the neighbourhoods the prevalent land use is residential, a much more detailed dataset for land use is desired to detect more diverse land use mixes and get more differences.

Considering the main destinations, it seems that people tend to go shopping in their own neighbourhood, despite the city centre is also an important target for shopping for all the inhabitants of all neighbourhoods (see Figure 3). About railway station, people do not always go to take the train to the closest railway station. Central stations in Amersfoort and Veenendaal are much more popular destinations since there are more trains passing and for more destinations.

In order to guarantee a good accessibility to the city centre, the size of the city should not be too large; otherwise the suburban neighbourhoods are found to be very far from the city centre. The proximity to the city centre is an important factor which can encourage the use of non-motorized travel modes, as people who live near the city centre tend to use car less.

7 Conclusions

This paper makes a contribution to the existing body of knowledge in mobility studies by comparing a series of GIS-based neighbourhood indicators with the actual people travel behaviour detected by GPS survey. OpenStreetMap, the most relevant example of the Volunteered Geographic Information (VGI), and other datasets were used to retrieve information about built environment characteristics, car, bicycle and pedestrian networks.

GPS technology has the ability to detect the actual travel patterns and the real use of the infrastructure network. Furthermore, GPS is much more accurate than traditional surveys (e.g. counting, paper travel diaries, etc.) and provides also additional information like route choice, path and speed.

Overall, this work aimed to improve the understanding of mobility performances in different neighbourhoods. The general purpose of the research has been successfully achieved. Of course, limitations are always presented and improvements are always possible. Mobility patterns are not easy to quantify and there is no best approach to measure them.

The findings reported in this study are generally consistent with the existing literature. In order to reduce car travel and to promote the use of public transport, factors like closeness to railway station, high building density, and high pedestrian and cycle network density, have to co-exist to a certain degree. The synergy of efficient performance in terms of proximity, density and accessibility in combination is likely to yield more appreciable impacts in sustainable mobility.

One of the strengths of the research is that the indicators are implemented using open-data and therefore this method can be replicated in other areas, using the same procedure and extracting information related to other neighbourhoods from OpenStreetMap. Moreover, the indicators were chosen in such a way that could be easily understood and interpreted by researchers, planners and policy makers. In doing so, they are more likely to be used in evaluation studies of mobility and to have impact on the policy making process. In fact, the GIS-based indicators can be used as an evaluation method of the sustainable mobility potential of neighbourhoods during planning stages of new neighbourhoods, but also for monitoring performance, propose policy and planning interventions on existing neighbourhoods. After having fully investigated the key factors yielding to a more sustainable people mobility, trajectory data may not be needed anymore in coming researches.

8 Future research

In the future this study can be enhanced and some of its limitations may be overcome. First of all, the computation of the shortest path can be improved using additional datasets to add information to the network leading to a more precise measure of distance. For instance, one-way streets were not taken into account in the computation, as this information was poorly presented in OpenStreetMap. In addition, traffic lights should also be included in the calculation of travel time as they can have a big influence on it.

Second, the list of the theoretical performance indicators can always be changed and improved in future research. Some indicators like distance to the closest bus stop and the number of schools within 2 and 5 km, were found to have limited impact on the analysis, and therefore new indicators may be added to the list. Furthermore, as indicators may have different influence on the final results, a series of weights may be applied in order to level them.

Finally, the GPS survey could be integrated with information derived by other technologies such as Wi-Fi, Bluetooth or RFID, which are often used in indoor environments. In this way, people could be easily tracked not only outdoor but also indoor, providing a better picture of the actual travel behaviour.

References

[1] European Commission. Sustainable Urban Transport Plans - Preparatory Document in relation to the follow-up of the Thematic Strategy on the Urban Environment (Technical report No. 2007/018). European Commission, Luxembourg, 2007.

[2] Gil, J. Urban form and the multi-modal mobility network structure. Evaluating the sustainable accessibility of urban areas in the city-region. PhD review report, TU Delft, 2010.

[3] Van der Spek, S., Van Schaick, J., De Bois, P., and De Haan, R. Sensing Human Activity: GPS Tracking. Sensor Journal, 2009.

[4] Shoval, N. The GPS revolution in spatial research. In J.V.Schaick, & S.C.V.D. Spek (Eds.), Urbanism on track (pp. 17-24), Amsterdam, The Netherlands: IOS Press BV, 2008.

[5] Bohte, W. Residential self-selection and travel. The relationship between travel-related attitudes, built environment characteristics and travel behaviour. In *Sustainable Urban Areas*, Delft University Press, 2010.

[6] Van Schaick, J., Van Der Spek, S.C. Urbanism on track. Application of Tracking Technologies in Urbanism. Delft University Press, 2008.

[7] A. Snellen, D., and Hilbers, H. Evaluation of the mobility impacts of the Dutch Vinex policy. Netherlands Institute for Spatial Research (Ruimtelijk Planbureau), 45th Congress of the European Regional Science Association. Amsterdam, 2005.

[8] Meurs, H., and Haaijer, R. Spatial structure and mobility. Transportation Research Part D 6 429±446, Elsevier, 2001.

[9] Bohte, W., Maat, K., and Quak, W. A method for deriving trip destinations and modes for GPS-based travel surveys. In J. Van Schaick, & S. Van der Spek (Eds.) Urbanism on Track, chap. 10, (pp. 129-145). IOS Press, 2008.

[10] Cervero, R. and Kockelman, K. Travel demand and the 3ds: density, diversity, and design. TranspnRes.-D, Vol. 2, No. 3, pp. 199-219, 1997.

[11] Frank, L.D., Saelens, B.E. Powell, K.E., and Chapman, J.E. Stepping towards causation: Do built environments or neighborhood and travel preferences explain physical activity, driving, and obesity? Social Science & Medicine 65 (2007) 1898–1914, 2007.

[12] Yigitcanlar, T., and Dur, F. Developing a Sustainability Assessment Model: The Sustainable Infrastructure, Land-Use, Environment and Transport Model. Sustainability 2010, 2, 321-340; doi:10.3390/su2010321, 2010.

[13] Gil, J., and Read, S. Patterns of sustainable mobility and the structure of modality in the Randstad city-region. A|Z ITU Journal of the Faculty of Architecture 11, 2014.

[14] McDonald, N.C. Travel and the social environment: Evidence from Alameda County, California. Transportation Research Part D 12 (2007) 53–63, 2007

[15] Onderzoek verplaatsingsgedrag, Research on travel behaviour, May 2105. URL: http://www.verplaatsingsgedrag.nl/index5.html

[16] Biljecki, F. Automatic segmentation and classification of movement trajectories for transportation modes, Master's thesis TU Delft, 2010.