Assessing the Accuracy of Pathfinding Algorithms for Scottish Children's Home-to-School Commutes: A Comparison with GPS Trajectories

Hyesop Shin[1]
MRC/CSO Social and Public Health Sciences Unit, 90 Byres Road, Glasgow, UK, G12 8TB

Rich Mitchell
MRC/CSO Social and Public Health Sciences Unit, 90 Byres Road, Glasgow, UK, G12 8TB

Abstract

Walking to and from school has significant implications for children's physical and mental well-being. This study aims to investigate the accuracy of routing engines (Google Maps, Mapbox, and OSRM) in replicating GPS trajectories and explore potential associations with gender and socioeconomic status. The study analysed GPS data from 227 children aged 10-11 years old in Scotland. The results indicated that OSRM exhibited the highest accuracy with a mean GPS track overlap of 56%. However, no substantial differences were found between the routing engines. Additionally, the accuracy of the engines did not vary based on gender or socioeconomic status. These findings provide reassurance that potential biases do not arise when using these navigation tools, as their accuracy remains consistent across different demographic groups.

2012 ACM Subject Classification General and reference → General conference proceedings

Keywords and phrases Pathfinding algorithms, GPS, Navigation tools, Children, Physical Activity

Digital Object Identifier 10.4230/LIPIcs.CVIT.2016.23

Category Short Paper

Supplementary Material The data that we used will not be shared with the public due to the non-disclosure agreement on personal information. However, the complete set of codes is available on GitHub.

Funding *Hyesop Shin*: This work is funded by the UK Medical Research Council (MRC) Places and Health Programme, Grant/Award Number: MC_UU_00022/4; Chief Scientist Office (CSO), Grant/Award Number: SPHSU19

1 Introduction

Walking to and from school can greatly impact the overall physical and mental health of children [1]. Previous research has emphasised the importance of active commuting to school, as it encourages increased physical activity, benefiting bone and muscular fitness, mental well-being, and even saving time during drop-offs and pick-ups while raising awareness of traffic safety [1, 2, 3]. Furthermore, active commuting helps to reduce (non) tailpipe emissions and air pollution, both of which have far-reaching effects on public health and the environment [4].

Understanding the routes children take to school is vital for their safety [5], fostering a sense of familiarity [6, 7], and effectively managing time during active travel [8]. Additionally, route choices hold particular significance for children from lower socioeconomic backgrounds

[1] Corresponding author

who may have limited access to transportation options and fewer opportunities to engage with enriching resources like museums and libraries.

To measure these school routes, a combination of GPS (Global Positioning System) trajectories and GIS (Geographic Information System) estimates have gained popularity. While GPS trajectories offer high accuracy for individual children, they come with challenges such as the expensive and labour-intensive process of collecting GPS data [9]. As a result, a compelling solution emerges in exploring the potential of routing algorithms (engines) to replicate these GPS trajectories, eliminating the need for extensive data collection efforts. These algorithms have gained attention for their ability to generate optimised routes based on factors such as distance, traffic conditions, the presence of CCTV cameras, and road conditions [10, 11].

However, despite their widespread usage, the accuracy of these pathfinding algorithms in replicating actual GPS trajectories remains a topic of investigation [12, 13, 14, 15, 16]. One reason for this is that many “GPS vs GIS” routing studies often compare only a single shortest-path tool in GIS, without taking into account the various routing methods available between two locations [12, 14, 15, 17, 18]. Additionally, these studies often rely on GIS layers that guide routing algorithms based on road-based polylines, potentially disregarding smaller alleyways, park trails, or roads that have not yet been updated [14]. As a result, such approaches can lead to an oversimplified conclusion that GPS is a superior instrument for correctly measuring commute patterns without considering the broader context. Further, it is important to understand the potential consequences of demographic and socio-economic bias in performance-operating algorithms, particularly when replacing GPS data with routing algorithms. If these algorithms display poor performance for marginalised groups, it could introduce bias into subsequent actions and distort our understanding of the broader context [19].

The objective of this study is to investigate and compare the accuracy of path-finding algorithm route selections to GPS trajectories. Specific questions are described below:

1. Which routing engine is the closest to the GPS data?
2. Does the accuracy of the routing algorithms, when compared to the GPS trajectories, vary 68 by the distance to school, gender, and socio-economic characteristics?

2 Methodology

2.1 GPS data

The study used information from 227 children in Scotland drawn from the “SPACES (Studying Physical Activity In Children’s Environments)” study [2], which collected GPS data for children aged 10 and 11. The participant’s home and school locations, their activity measures, gender, and socio-economic data were provided. We used the Scottish Index of Multiple Deprivation (SIMD)[3] as a proxy for an individual’s socioeconomic status in this study. During the pre-processing stage of the GPS data, the following inclusion criteria were applied in order to isolate GPS tracks which represented children walking to school in the morning: 1) only

[2] Please visit the following link for more information: https://www.gla.ac.uk/schools/healthwellbeing/research/mrccsosocialandpublichealthsciencesunit/programmes/places/movementurbanlandscapes/spaces/

[3] SIMD is a ranked tool for determining a relative measure of deprivation across 6,976 small areas, with 1 being the most deprived and 6,976 being the least deprived. For more information, visit https://www.gov.scot/collections/scottish-index-of-multiple-deprivation-2020/

points recorded on weekdays were considered, 2) points recorded between 07:30 and 09:00 were selected, 3) the study focused on the home-to-school trajectory, not vice versa, as some children go to different places depending on their parents' working conditions, and 4) points recorded with a speed of less than 5 km/h were included [20], indicating children whose mode of travel was most likely walking.

To process the GPS data, each participant's GPS track was randomly sampled from one of seven days in which the child was walking to school and had a valid track. The recorded points were then cleaned and interpolated to form a polyline.

2.2 Modelling routes from home to school

The provided GPS track data was processed by filtering out points with incorrect coordinates or overlapping locations over time. Next, we interpolated the individual data points to create a polyline representing the travel path of each child to school. To measure the overlapping percentage between the modelled polylines, we created a 30-meter buffer around the GPS track polyline, creating a polygon [12, 14]. These are the steps from 1 to 3 illustrated in Figure 1.

For the routing engines, we generated routes for the children's journeys between home and school using the "walking" mode. This study compared three popular routing models: Google API, Mapbox API, and Open Source Routing Machine (OSRM). The main objective of these routing models is to determine the shortest and most efficient path to school [12, 14]. To ensure consistency in the data cleaning and analysis process, we utilized specific R packages: `mapsapi` to access Google API's routing engine, `mapboxapi` for Mapbox API, and `osrm` for OSRM. Once the routes for each child were generated, we applied a 30-meter buffer to the routes obtained from the three routing engines, creating polygons.

Then, we performed a spatial intersection between the GPS polygons and each of the modelled routes, to determine the extent of polygon overlap, measured as percentage. The computation of the comparison between navigation routes and GPS tracks utilised the concept of spatial intersection, as described in previous studies [12, 14, 15, 16]. The resulting percentage of spatial intersection served as an indicator of the similarity between the two routes. A complete mismatch between the two routes would result in an error rate of 100%, indicating no overlap, while a perfect match between the routes would yield a difference of 0%. These are the steps from 5 to 8 illustrated in Figure 1.

It is important to note that different route engines can produce varying results due to factors such as pathfinding algorithms, road structure prioritisation, and an incomprehensible road database. As an example, one of the navigation methods identified that the child's actual path went through a park, whereas the other two methods only provided detour routes along the streets (refer to Figure 2).

3 Results

A total of 227 participants contributed data on their walking routes to school. The characteristics of these participants can be found in the provided Table Go to Link here. Among the participants, there were 99 boys and 128 girls within the age range of 10 to 11. An analysis of the trip length data indicates that the majority of participants walked a distance of less than 2 kilometres to reach school.

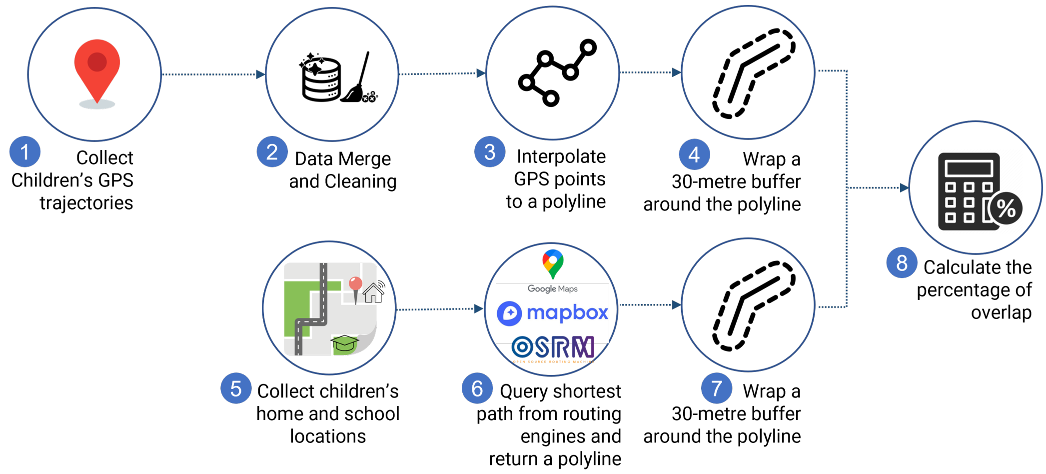

Figure 1 The extraction of GPS and GIS data and the calculation of overlapping percentages

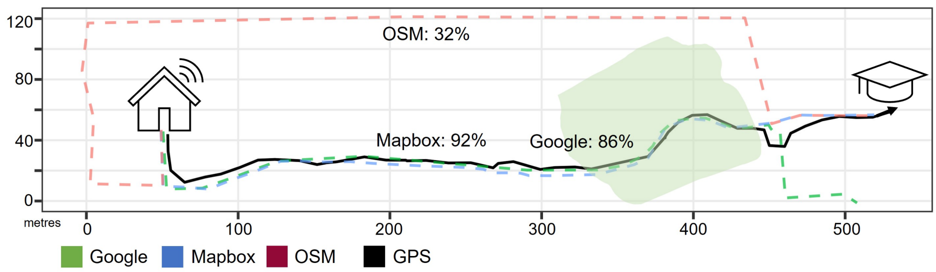

Figure 2 A comparison example of the accuracy of three routing engines against the GPS trajectory in a school journey

3.1 Which routing algorithm is the closest to the GPS data?

Our findings show that the mean GPS track overlap varied between the three routing models. OSRM exhibited the highest value at 56%, followed by Mapbox at 52%, and Google at 47% (see Figure 3A). However, we also observed significant variation in the overlapping percentage between participants, as indicated by the standard deviation, which can be influenced by the distance to school. Furthermore, the differences between the routes provided by the models were found to be small.

When examining the percentage of children who had the highest accuracy with the GPS tracks (<10% error rate from Figure 3B), the results were as follows: OSRM - 12%, Google - 11%, and Mapbox - 8%. Considering a 30% error margin, the differences in accuracy between routing algorithms were approximately 42% for OSRM, 38% for Mapbox, and 30% for Google.

3.2 Does the accuracy of the routing algorithm, when compared to the GPS trajectories, vary by the distance to school, gender, and socio-economic characteristics

Figure 4 illustrates the relationship between the accuracy of modelled routes and GPS tracks and the distribution of children's distance to school. This visual representation offers valuable insights into whether children residing closer to school tend to exhibit higher accuracy

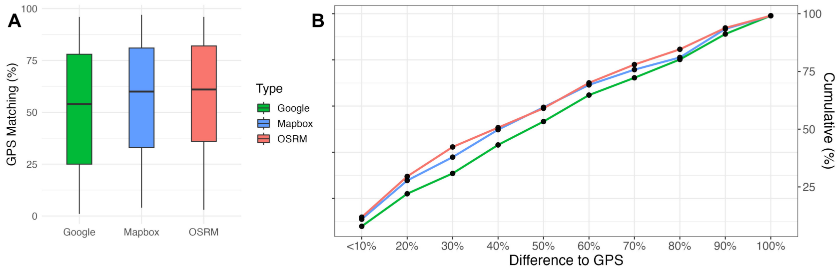

Figure 3 A: Comparing the accuracy of the routing engines to GPS tracks and modelled routes. B: Cumulative line plot for each group among the 227 children. The x-axis percentage shows the error rate between GPS tracks and provides a cumulative error rate as the x-axis increases.

compared to those living farther away. It also enables us to explore potential correlations between distance and factors such as gender (Figure 4A) or socioeconomic status (Figure 4B).

Our analysis, based on a sample of 227 children, revealed that among those with distances to school under 1km, over 60 of them achieved a remarkable accuracy rate of 75% in relation to the GPS tracks. Interestingly, we did not observe any discernible systematic differences between genders or across SIMD groups. Despite the fact that the sample size was skewed towards wealthier children, these findings provide strong evidence to confidently conclude that there are no systematic disparities in GPS accuracy based on gender or socioeconomic groups.

4 Conclusion

In our study, we conducted a comprehensive analysis to compare the accuracy of GPS tracks between home and school, employing three route estimation engines: Google Maps, Mapbox, and OSRM. The results showed that OSRM had the highest accuracy of 56%, which did not show a meaningful difference from the other two engines in the overall context. However, it is important to note that the accuracy of GPS tracks varied on an individual basis, influenced by factors such as the complexity of the built environment and the availability of neighbourhood amenities such as parks. Furthermore, our analysis demonstrated that the errors produced by these engines had no important association with gender or socioeconomic status, and only a weak relationship with the distance to school. These findings are particularly reassuring as they suggest that potential biases do not arise when utilising the aforementioned navigation tools. The accuracy of these platforms remains consistent regardless of socioeconomic status, indicating that the accuracy does not vary based on whether the child is from a disadvantaged background or not.

References

1 Roger L Mackett and James Paskins. Children's physical activity: The contribution of playing and walking. 22(5):345–357. Publisher: Wiley Online Library.

2 Anne Martin, Paul Kelly, Jessica Boyle, Fenella Corlett, and John J Reilly. Contribution of walking to school to individual and population moderate-vigorous intensity physical activity: systematic review and meta-analysis. 28(3):353–363. Publisher: Human Kinetics, Inc.

3 Andreia Nogueira Pizarro, José Carlos Ribeiro, Elisa Amélia Marques, Jorge Mota, and Maria Paula Santos. Is walking to school associated with improved metabolic health? 10(1):1–7. Publisher: BioMed Central.

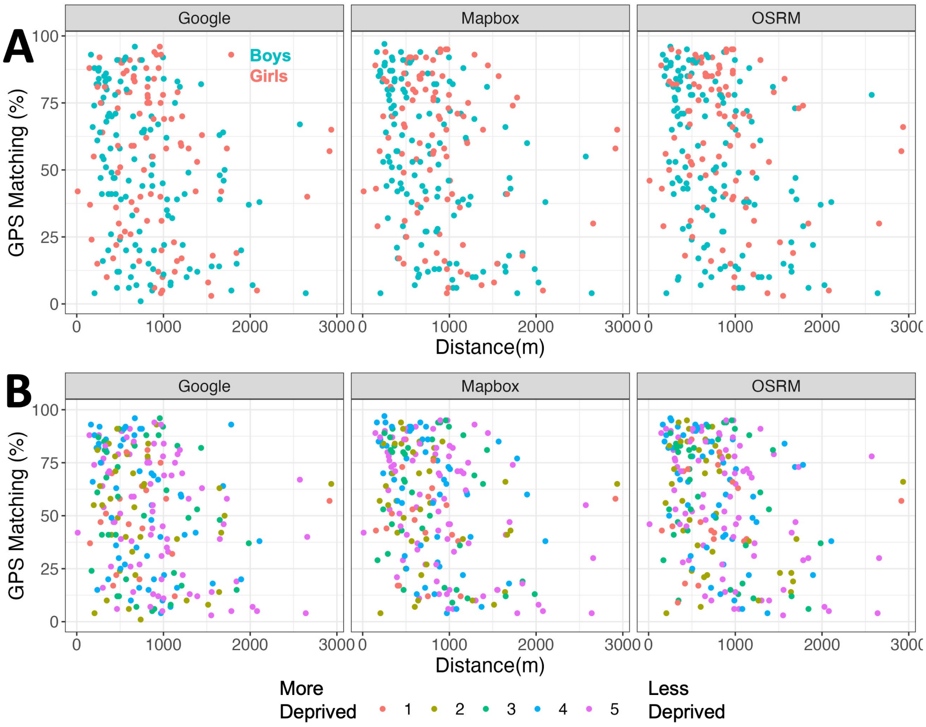

Figure 4 Both scatter plots illustrate the accuracy between modelled routes and GPS tracks based on the distance to school. The upper plot (A) shows the difference between boys and girls and the bottom plot (B) shows the difference based on the child's socioeconomic status. In summary, there is no discernible pattern regarding the distance to school among different genders or socioeconomic groups.

4 Shiraz Ahmed, Muhammad Adnan, Davy Janssens, and Geert Wets. A route to school informational intervention for air pollution exposure reduction. 53:101965.

5 Jenna R. Panter, Andrew P. Jones, Esther M.F. Van Sluijs, and Simon J. Griffin. Neighborhood, Route, and School Environments and Children's Active Commuting. *American Journal of Preventive Medicine*, 38(3):268–278, March 2010.

6 Ilse M. Harms, Bridget R.D. Burdett, and Samuel G. Charlton. The role of route familiarity in traffic participants' behaviour and transport psychology research: A systematic review. *Transportation Research Interdisciplinary Perspectives*, 9:100331, March 2021.

7 UK Government. Education recovery in schools: spring 2022. April 2022.

8 Noreen C. McDonald and Annette E. Aalborg. Why Parents Drive Children to School: Implications for Safe Routes to School Programs. *Journal of the American Planning Association*, 75(3):331–342, June 2009.

9 A Larkin and P Hystad. Towards Personal Exposures: How Technology Is Changing Air Pollution and Health Research. *Current environmental health reports*, 4(4):463–471, 2017.

10 Jeffrey Linwood. *Turn-by-Turn Navigation with Mapbox*, pages 221–234. Apress.

11 Lauri Sintonen, Hannu Turtiainen, Andrei Costin, Timo Hämäläinen, and Tuomo Lahtinen. OSRM-CCTV: CCTV-aware routing and navigation system for privacy and safety. pages 279–288. Springer.

12 Alice M. Dalton, Andrew P. Jones, Jenna Panter, and David Ogilvie. Are GIS-modelled routes a useful proxy for the actual routes followed by commuters? 2(2):219–229.