

TRANSPORT FINDINGS

Prediction of the Deviation between Alternative Routes and Actual Trajectories for Bicyclists

Haotian Wang¹ , Emily Moylan¹ , David M. Levinson¹ ¹ School of Civil Engineering, The University of Sydney

Keywords: choice set generation, travel behaviour, route choice, bicycle

<https://doi.org/10.32866/001c.35701>

Findings

This study estimates a panel regression model to predict bicyclist route choice. Using GPS trajectories of 600 trips from 49 participants in spring 2006 in Minneapolis, we calculate deviation, the average distance between alternative routes and actual trajectories, as the dependent variable. Trip attributes, including trip length, Vehicle Kilometres Travelled (VKT), the number of traffic lights per kilometer, and the percentage of bike trails and separated bike lane, are included as independent variables. F-tests indicate that both fixed entity and time effect panel regression models offer better fits than the intercept-only model. According to our results, routes with shorter length and higher share of bike trails tend to have less deviation in their trajectories. Traffic lights per km, VKT, and share of bike lane are not significant at the 95% confidence level in this data set.

1. Questions

Route choice modelling, especially for bicyclists, quantifies the factors determining route selection and, based on those factors, can be used for facility location and network improvement. For revealed choice datasets, the plausibility of the models is undermined by the challenge of creating realistic choice sets. Moreover, the routes that were actually considered by the traveller may be difficult to identify. Previous studies (Li, Muresan, and Fu 2017; Zhu and Levinson 2015) have used ratio of overlap to evaluate the quality of the alternatives, but in dense urban networks, this ratio is low. In this study, we introduce 'deviation' which captures the closeness or average distance between alternative routes and actual trajectories. High values for deviation indicate that the route alternative was not similar to the chosen route, and when all alternatives for a particular trip have high deviation, the choice set has not captured the attribute(s) that are most important to the route choice decision.

The hypotheses are:

1. Shorter lengths are more likely to result in smaller deviation.
2. Less traffic signals are more likely to result in smaller deviation.
3. Lower traffic flow is more likely to result in smaller deviation.
4. Routes with high percentage of bike trail are more likely to result in less deviation.
5. Route with high percentage of dedicated bike lane (or on-street bike lane) are more likely to result in less deviation.

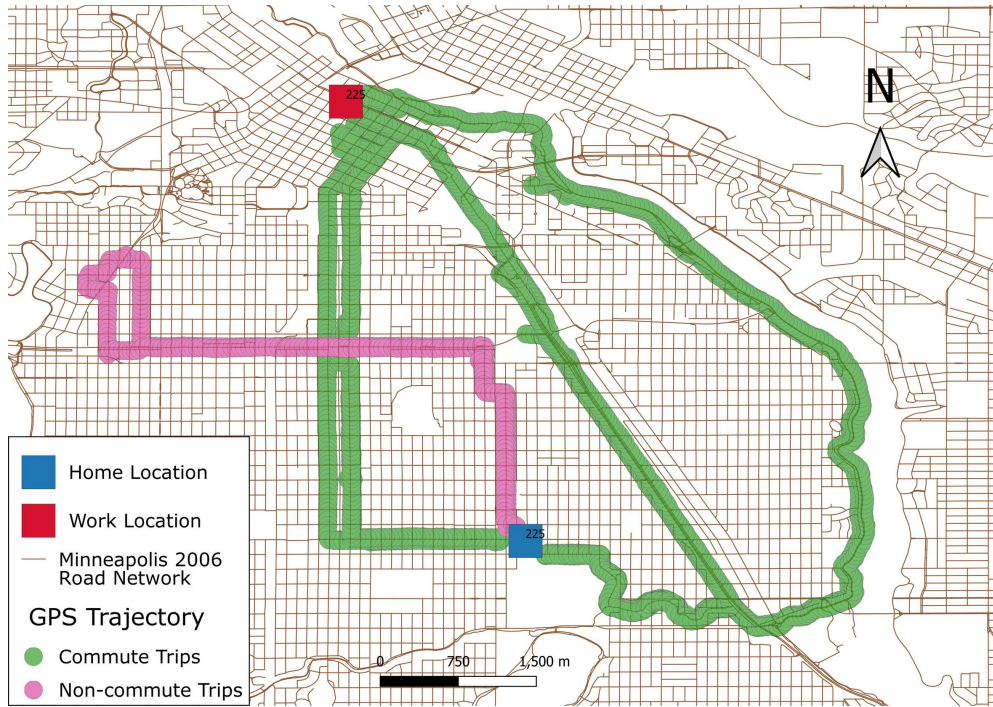


Figure 1. Example of Commute and Non-Commute Trips

2. Methods

2.1. Data collection

In this study, all GPS data are obtained from Harvey et al. (2008) and Menard et al. (2009), who collected repeated data over 2 weeks from 49 regular bicyclists living in South Minneapolis in 2006. Small GPS dataloggers, which record location and elevation every 2 seconds with roughly 3 meters accuracy, are attached to participants' bicycles. All GPS points which are located in a 100m radius of the participant's home or workplace location, are removed and the recorded home and workplace locations have been randomized within the same radius to protect the privacy of participants. In addition, as recorded GPS for home and workplace were jittered, and participants might not start recording their trips exactly at the home or work place, there is a discrepancy between home or workplace and the start point or end point of some trips. A 250-m radius around the home and workplace is used to filter out origins and destinations of non-commute trips such as shown in [Figure 1](#). After filtering, 600 of 831 trips remain.

2.2. Map matching

As the map data are downloaded from OpenStreetMap version 2021 and the trajectories were recorded in 2006, adjustments based on archival Google Maps StreetView are carried out to reproduce the historical road network before matching GPS data to network. In addition, the 2021 cycleways have been categorized as three different types of cycleway in 2006.

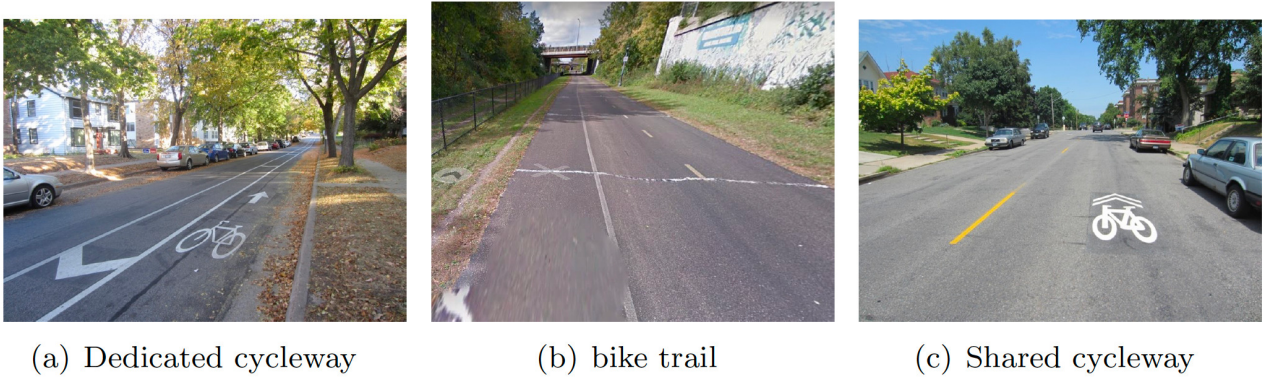


Figure 2. Examples of three types of cycleways.

Source: Google Streetview of Minneapolis

As shown in [Figure 2](#), the first type is *dedicated cycleway*, which are approximately 5 ft-wide (1.5 m) and located on the edge of the roadway. *Bike trails* are another facility designed for bicycles but do not share the right of way with cars to provide safer travelling conditions for cyclists. The last type is *shared cycleways*, (typically sharrows) which are shared with other transport modes like buses or cars.

In this study, we use KD-Tree which is a geometric measure for matching trajectories to the network. To populate network points set P , a point was generated every 55 meters along each segment. The basic idea of this method is setting P as reference layer and trajectory points set R as target layer. For each point in target layer R , the algorithm finds the closest point in reference layer P and records it into the matched trip set M .

2.3. Choice set generation

The size and quality of the choice set influence the route choice modelling (Prato 2009). To gain a choice set which better captures actual trajectories, a link labelling approach (Ben-Akiva et al. 1984), which generates alternative paths by optimizing different criteria, is implemented. According to previous studies (Bernardi, La Paix-Puello, and Geurs 2018; Broach, Dill, and Gliebe 2012; Lin and Fan 2020; Zimmermann, Mai, and Frejinger 2017), distance, traffic flow, and bike facilities are important factors for cyclists planning their trips.

To understand the preference of bike facilities for 49 participants, the constitution of each trip is analysed. For each type of bike facility, the number of trips for which the length percentage of that facility is higher than other road types is listed in [Table 1](#). Based on the lack of trips dominated by shared cycleways, this facility is not as important as the other two when people choose travel routes. Thus paths with maximized bike trail proportion and paths with maximized dedicated cycleway proportion are included in the choice set.

Table 1. Length proportion of Bike facilities

Road type Class	Number of Trips, constituted mostly by the type class	Max length proportion
Bike trail	225	0.877
Dedicated cycleway	73	0.705
Shared cycleway	0	0.200

However, paths with maximized percentage of bike trails sometimes result in extremely long routes (approximate 3 times the shortest path length) which might be unrealistic for most commute trips. To include a more realistic percentage of the length on bike trails, a set of factors from 0.1 to 1 with 0.1 increase each time are used to weight the bike trails' length. In addition, as recommended in previous study (Ghanayim and Bekhor 2018; Sobhani, Aliabadi, and Farooq 2019; Tilahun, Levinson, and Krizek 2007), the shortest path and the fastest path are generated for each OD pair.

All generated path are then included into a set G , and alternative routes for a OD pair with more than 80% similarity, which is measured by the length of common links, are removed from G . The size of G is too large for efficient modeling, and the difference of the size of choice sets between travelers might also create bias. Therefore, for each traveller, 6 paths are defined

- shortest path ($A_{i=0}$),
- path with maximum proportion of dedicated cycleways ($A_{i=1}$),
- path with maximum proportion of on-street cycleways ($A_{i=2}$),
- path with maximum proportion of secondary road ($A_{i=3}$),
- path with maximum proportion of cycleway ($A_{i=4}$), and
- path with maximum proportion of bike trail but total length within 1.3 times shortest path length ($A_{i=5}$),

have been selected to form choice set A . The reason for containing $A_{i=3}$ is that secondary roads are found have a relative high proportion of recorded trips, and fewer turns are needed on secondary roads. For $A_{i=5}$, scenarios with length within 1.1 to 1.6 times shortest path length are tested, and 1.3 gives the highest similarity to observed trajectories.

2.4. Deviation

Deviation measures the similarity between paths. It can be applied to compare the generated alternative routes and actual trajectories. To measure the dependent variable 'deviation', D , we construct the convex hull, which is the polygon formed by the set of all points in the alternative route and trajectory as shown in [Figure 3](#). D equals to the square root of that area.

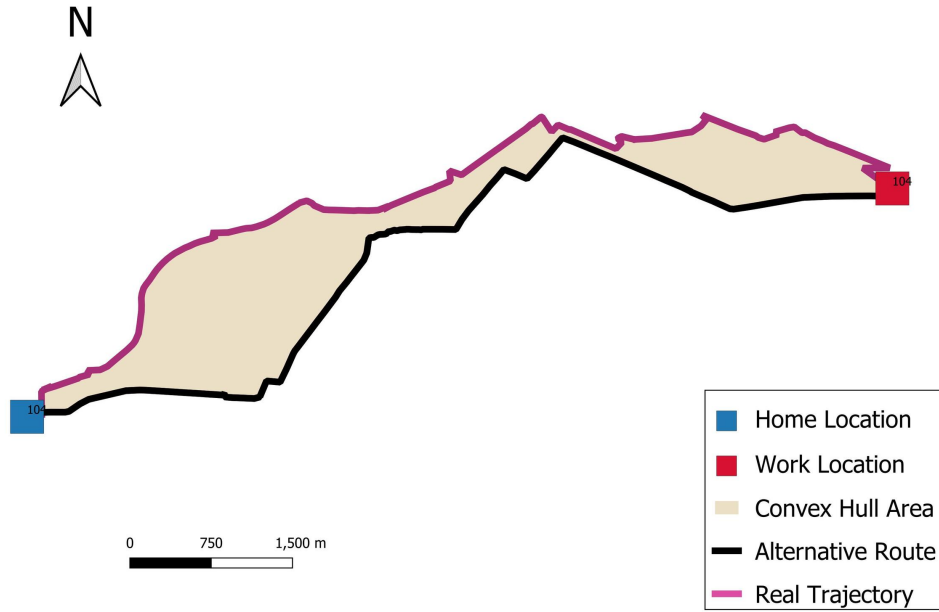


Figure 3. Convex Hull Example

Table 2. Variables in regression model

Variables	Description	Symbol
Length	Total length of the route	L
VKT	Total vehicle kilometres travelled for the route	V
Traffic light per meter	Total number of traffic light on the route divide by total length of the route	T
Length percentage of bike trail	Total length of bike trail in the route divide by total length of the route	C
Length percentage of dedicated cycleway	Total length of dedicated cycleway in the route divide by total length of the route	B

2.5. Panel regression model

The alternative route with the lowest D best captures the features of the selected route. Since 600 trips are collected from 49 travelers during a period of time, to control the correlation of the errors due to unobserved variables associated with panel data, a Fixed Effect (FE) regression model is applied to model the deviation D , and the results are compared to pooled ordinary least squares (OLS) regression model. The variables in regression models are presented in [Table 2](#).

3. Findings

Overall, as shown in [Table 3](#), routes with shorter length and higher percentage of bike trail minimize deviation with actual trips. So the hypotheses 1 and 4 made in section 1 are corroborated. Percentage of on-street bike lane, VKT, and traffic lights per km are not statistically significant at 95% confidence level in this data set.

Table 3. Outputs for Convex Hull: Deviation between alternative routes and actual route

Variables	Entity and Time Fixed Effects	Pooled OLS
constant	515.23** (250.31)	775.90** (77.40)
Length	242.03*** (25.35)	230.23*** (6.78)
VKT	-0.0018 (0.0011)	-0.0029 (-0.0007)
Traffic light per km	72.29 (51.823)	32.25 (26.304)
Percentage of bike trail	-724.18*** (190.25)	-832.10*** (120.95)
Percentage of dedicated cycleway	-285.29 (190.25)	-210.07* (122.94)
F-test for Poolability	0.00	0.00
Adjusted R^2	0.349	0.345
Durbin-Watson	2.12	1.56

(standard error) ; *: $P \leq 0.1$; **: $P \leq 0.05$; ***: $P \leq 0.01$

Submitted: March 24, 2022 AEST, Accepted: May 14, 2022 AEST



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-SA-4.0). View this license's legal deed at <https://creativecommons.org/licenses/by-sa/4.0> and legal code at <https://creativecommons.org/licenses/by-sa/4.0/legalcode> for more information.

REFERENCES

- Ben-Akiva, Moshe, MJ Bergman, Andrew J. Daly, and Rohit Ramaswamy. 1984. "Modelling Inter Urban Route Choice Behaviour." In *Papers Presented during the Ninth International Symposium on Transportation and Traffic Theory Held in Delft the Netherlands, 11-13 July 1984*.
- Bernardi, Silvia, Lissy La Paix-Puello, and Karst Geurs. 2018. "Modelling Route Choice of Dutch Cyclists Using Smartphone Data." *Journal of Transport and Land Use* 11 (1): 883–900. <https://doi.org/10.5198/jtlu.2018.1143>.
- Broach, Joseph, Jennifer Dill, and John Gliebe. 2012. "Where Do Cyclists Ride? A Route Choice Model Developed with Revealed Preference Gps Data." *Transportation Research Part A: Policy and Practice* 46 (10): 1730–40. <https://doi.org/10.1016/j.tra.2012.07.005>.
- Ghanayim, Muhammad, and Shlomo Bekhor. 2018. "Modelling Bicycle Route Choice Using Data from a Gps-Assisted Household Survey." *European Journal of Transport and Infrastructure Research* 18 (2).
- Harvey, Francis, Kevin J. Krizek, and Reuben Collins. 2008. "Using Gps Data to Assess Bicycle Commuter Route Choice." Technical report.
- Li, Siyuan, Matthew Muresan, and Liping Fu. 2017. "Cycling in Toronto, Ontario, Canada: Route Choice Behavior and Implications for Infrastructure Planning." *Transportation Research Record* 2662 (1): 41–49.
- Lin, Zijng, and Wei (David) Fan. 2020. "Bicycle Ridership Using Crowdsourced Data: Ordered Probit Model Approach." *Journal of Transportation Engineering, Part A: Systems* 146 (8): 04020076.
- Menard, Jason, Francis Harvey, and Kevin J. Krizek. 2009. "Improving Gps Data Collection of Human Spatial Behavior." Technical report.
- Prato, Carlo Giacomo. 2009. "Route Choice Modeling: Past, Present and Future Research Directions." *Journal of Choice Modelling* 2 (1): 65–100. [https://doi.org/10.1016/s1755-5345\(13\)70005-8](https://doi.org/10.1016/s1755-5345(13)70005-8).
- Sobhani, Anae, Hamzeh Alizadeh Aliabadi, and Bilal Farooq. 2019. "Metropolis-Hasting Based Expanded Path Size Logit Model for Cyclists' Route Choice Using Gps Data." *International Journal of Transportation Science and Technology* 8 (2): 161–75. <https://doi.org/10.1016/j.ijtst.2018.11.002>.
- Tilahun, Nebiyu Y., David M. Levinson, and Kevin J. Krizek. 2007. "Trails, Lanes, or Traffic: Valuing Bicycle Facilities with an Adaptive Stated Preference Survey." *Transportation Research Part A: Policy and Practice* 41 (4): 287–301.
- Zhu, Shanjiang, and David Levinson. 2015. "Do People Use the Shortest Path? An Empirical Test of Wardrop's First Principle." *PloS One* 10 (8): e0134322. <https://doi.org/10.1371/journal.pone.0134322>.
- Zimmermann, Maëlle, Tien Mai, and Emma Frejinger. 2017. "Bike Route Choice Modeling Using Gps Data without Choice Sets of Paths." *Transportation Research Part C: Emerging Technologies* 75 (February): 183–96. <https://doi.org/10.1016/j.trc.2016.12.009>.