A link-based bicycle perturbed utility route choice model for Copenhagen[†]

Mads Paulsen[1], Mirosława Łukawska[*2], Thomas Kjær Rasmussen[3], and Mogens Fosgerau[4]

[1]Postdoc, Department of Technology, Management and Economics, Technical University of Denmark, Denmark
[2]PhD Student, Department of Technology, Management and Economics, Technical University of Denmark, Denmark
[3]Associate Professor, Department of Technology, Management and Economics, Technical University of Denmark, Denmark
[4]Professor, Department of Economics, University of Copenhagen, Denmark

SHORT SUMMARY

This paper estimates a bicycle perturbed utility route choice model from a large dataset of GPS traces. Results quantify the influence of a range of factors on bicycle route choice. The model allows very fast parameters estimation by ordinary linear regression. In addition, the model uses the complete network – no choice set generation is required.

Keywords: Bicycle traffic; GPS data; Perturbed utility; Route choice

1. INTRODUCTION

There is a growing consensus that bicycle as a transport mode contributes to positive impacts on the environment and individual health, and can help to lower traffic congestion (Banerjee et al., 2021). To encourage road users to shift to bicycle, policy makers seek to understand the behaviour of the current cyclists. One of the aspects to consider is route choice preferences - which factors determine cyclists' decisions, what is the magnitude of the influence, and how interdependent are these factors. There is therefore a considerable interest in this topic and corresponding strong motivation for quantifying the above-mentioned preferences.

Pritchard (2018) reviews the already extensive bicycle route choice literature based on observed behaviour (e.g. GPS trajectories). Estimating such route choice models allows the factors that influence route choice to be identified and quantified, this includes infrastructure type (Menghini et al., 2010; Hood, Sall & Charlton, 2011; Broach, Dill & Gliebe, 2012; Zimmermann, Mai & Frejinger, 2017; Ghanayim & Bekhor, 2018; Prato, Halldórsdóttir & Nielsen, 2018; Skov-Petersen et al., 2018) surface type (Prato, Halldórsdóttir & Nielsen, 2018), gradient (Menghini et al., 2010; Hood, Sall & Charlton, 2011; Broach, Dill & Gliebe, 2012; Zimmermann, Mai & Frejinger, 2017; Prato, Halldórsdóttir & Nielsen, 2018), and land-use (Ghanayim & Bekhor, 2018; Prato, Halldórsdóttir & Nielsen, 2018; Skov-Petersen et al., 2018). However, the vast majority of existing studies are based on traditional maximum likelihood estimation of discrete choice models that require generating choice sets of route alternatives for each trip.

[†]This research is partly funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 740369).

Several methods for generating such choice sets exist (Halldórsdóttir et al., 2014), but creating realistic choice sets that contains the observed routes without containing a lot of irrelevant alternatives remains very difficult, especially for bicycle route choice models (Ton et al., 2018). Bicycle networks are typically more fine-grained than car road networks, as cyclists also use small streets, and even paths where cars are not allowed. This makes the number of reasonable route alternatives very high and the creation of suitable choice sets becomes very difficult (Ton et al., 2018; Rasmussen, Lukawska & Paulsen, 2021). Consequently, researchers have been forced to discard almost half of the observed trips due to inadequate similarities with any of the generated alternatives (Koch, Knapen & Dugundji, 2021).

Link-based route choice models such as the Recursive Logit (RL) model (Fosgerau, Frejinger & Karlstrom, 2013) and Nested Recursive Logit (NRL) (Mai, Fosgerau & Frejinger, 2015) avoid choice set generation, and have been successfully applied to bicycle route choice (Zimmermann, Mai & Frejinger, 2017). However, calculation times were enormous as it took, respectively, 15 days and 43 hours, to estimate the NRL and RL models with link-size attributes (Zimmermann, Mai & Frejinger, 2017). Furthermore, Koch & Dugundji (2021) were not able to obtain identification with these models.

The present paper is the first to apply the recently developed Perturbed Utility Route Choice (PURC) model (Fosgerau, Paulsen & Rasmussen, 2022) to bicycle route choice. The PURC model has several important advantages. It realistically allows zero probabilities on links, reflecting that most links are unused for any given origin-destination (OD) pair; it does not require choice set generation; it implies realistic substitution patterns by incorporating the actual network structure; and, not least, allows estimation of parameters through ordinary linear regression, making it possible to explore many model specifications within a short time even when using large datasets of observed routes and highly disaggregate networks.

2. METHODOLOGY

A network consists of nodes $\mathscr{V}$ and edges $\mathscr{E}$, and a network link incidence matrix A. Furthermore, a set of OD pairs $\mathscr{B}$ is given in terms of OD demand vectors b with elements

$$b_v = \begin{cases} -1, & v \text{ is the origin node} \\ 1, & v \text{ is the destination node} \\ 0, & \text{all other nodes} \end{cases}, \qquad v \in \mathscr{V}, b \in \mathscr{B}. \tag{1}$$

For each OD b, the PURC model formulates the route choice as a utility maximisation problem (2) for a representative traveler, who distributes non-negative flow x_b across the entire network to maximize the perturbed utility $U(x_b)$, while satisfying flow conservation ($Ax_b = b$) at every node,

$$\begin{array}{lll} \max & U(x_b) = l^\intercal\left[(u \circ x_b) - F(x_b)\right], & \\ \text{s.t.} & Ax_b \quad = b, & x_b \geq 0. \end{array} \tag{2}$$

Here, vector l contains the length of the links of the network, u is a corresponding vector of utility rates (utility per distance), and $F(x_b)$ is a strictly convex and increasing perturbation function applied to each component of x_b, specified using the entropy-like form

$$F(x_b) = (1 + x_b)\ln(1 + x_b) - x_b. \tag{3}$$

The perturbation induces the traveler to distribute flow on more links than just those on the shortest path (Fosgerau, Paulsen & Rasmussen, 2022).

Specifying the utility vector $u = z\beta$, where z is a matrix of link characteristics and β is a parameter vector, Fosgerau, Paulsen & Rasmussen (2022) show that the first-order conditions for the utility maximization problem lead to the following linear regression equation, for each $b \in \mathscr{B}$,

$$(I - BA^{\intercal}C)B(l \circ F'(x_b)) = (I - BA^{\intercal}C)B(l \circ z)\beta + \varepsilon. \tag{4}$$

The left-hand side is the dependent variable vector, whereas the right-hand side is linear in the parameters β and a vector of residuals ε. The matrix B is an identity matrix with rows corresponding to links with zero flows removed, and C is the Moore-Penrose inverse of $BA^{\intercal}$. By precomputing and storing $(I - BA^{\intercal}C)B$, estimating the parameters of a new model specification is as simple as adding/removing/replacing columns of z, performing a matrix multiplication, and applying linear regression.

Figure 1: Heatmap of GPS trajectories from Hövding, Greater Copenhagen Region

3. RESULTS AND DISCUSSION

Data description

We used a large-scale crowd-sourced GPS dataset of observed route choices from Hövding[*] airbag helmets. As can be seen from Figure 1 the dataset is very extensive, containing 277,559 preprocessed bicycle GPS trajectories in Metropolitan Copenhagen. Another benefit of the Hövding dataset is that the data collection process is less user-dependent and therefore less error-prone than in case of smartphone-based GPS data. The study furthermore utilizes a highly detailed bicycle network (420,973 links) and takes into account characteristics such as land-use, elevation, surface conditions, cycle highways, and bicycle infrastructure, retrieved from OpenStreetMap (2020), among other sources.

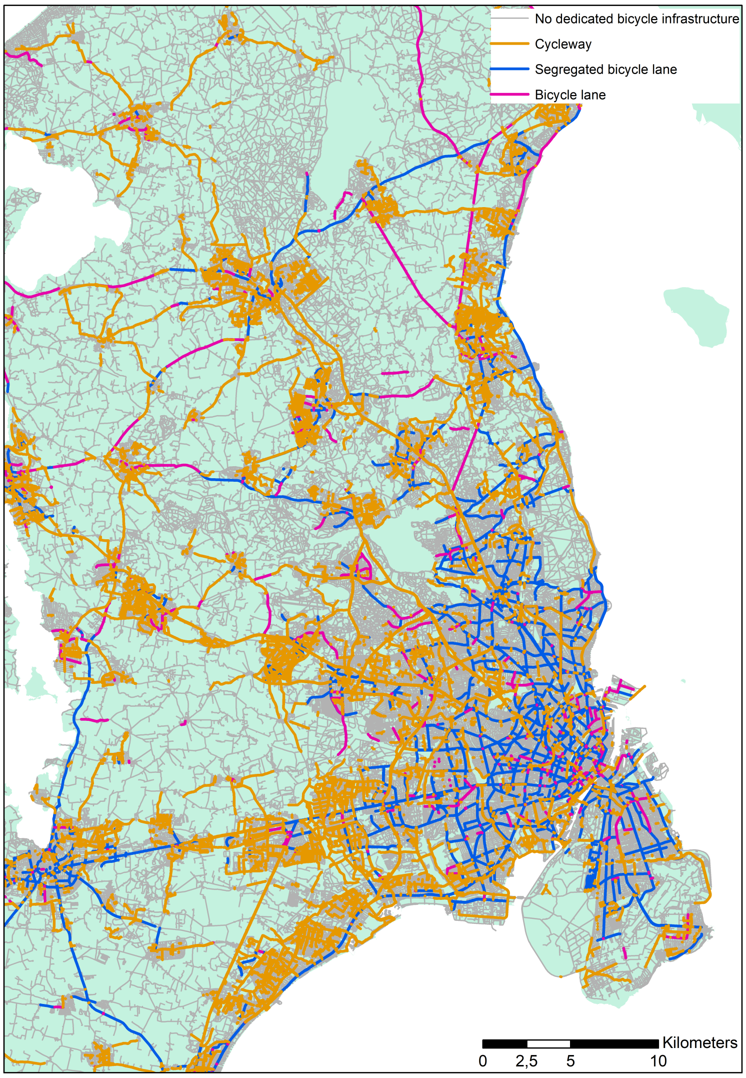

Figure 2: Network of bicycle infrastructure in Greater Copenhagen Region

200 origins and 200 destinations were selected based on the number of trajectories traveling through them as in Fosgerau, Paulsen & Rasmussen (2022), and included trips were cropped to match the first origin and last destination encountered on the trip. The estimation was performed based on a total of 85,529 valid cropped trajectories across 8,283 OD pairs. The average number of used links per OD pair was 198.4, resulting in 1,643,487 observations for the linear regression.

[*]https://hovding.com

	Unit	**Coef.**	**Std. err.**	**z**
General network attributes	[m]			
Length		-0.2236	0.006	-38.022
Wrong way		-0.1882	0.001	-151.981
Surface type	[m]			
Cobblestones		-0.1143	0.007	-17.027
Elevation gain	[vertical m]			
0-3.5 m/km		—	—	—
3.5-50 m/km		-6.3647	0.778	-8.176
$>$ **50 m/km**		-19.2973	0.812	-23.766
Infrastructure	[m]			
Cycleways		0.0946	0.005	19.818
Footways		-0.0740	0.003	-27.614
Shared paths		-0.0042	0.002	-1.747
Pedestrian zones		-0.0820	0.006	-14.176
Living streets		-0.0440	0.009	-4.951
Residential roads w/o bike infrastructure		—	—	—
Residential roads w/ bike lanes		0.0314	0.005	5.823
Residential roads w/ segregated bike tracks		0.0752	0.004	18.095
Medium other roads w/o bike infrastructure		0.0263	0.003	8.955
Medium other roads w/ bike lanes		0.0799	0.006	13.249
Medium other roads w/ segregated bike tracks		0.0848	0.002	40.440
Large other roads w/o bike infrastructure		-0.0122	0.002	-4.969
Large other roads w/ bike lanes		0.0046	0.008	0.553
Large other roads w/ segregated bike tracks		0.0818	0.002	39.022
Steps		-0.5953	0.031	-19.446
Land-use (Cycleways)	[m]			
Green areas		0.1466	0.007	20.497
Areas near water		-0.0439	0.006	-6.772
Low urban areas		—	—	—
High urban areas		0.0488	0.005	9.445
Industrial areas		0.1520	0.007	21.623
Open landscape		0.0398	0.014	2.772
Land-use (All other infrastructure types)	[m]			
Green areas		0.0232	0.004	5.478
Areas near water		0.0607	0.004	13.175
Low urban areas		—	—	—
High urban areas		0.0468	0.003	16.491
Industrial areas		0.0122	0.005	2.530
Open landscape		-0.0226	0.011	-2.062
Bicycle route classification	[m]			
Cycle superhighway		0.0171	0.001	12.260

Table 1: The estimated parameters for the bicycle route choice model.

Model estimation

The computational efficiency of using linear regression allowed a large bank of different model specifications to be tested, and Table 1 shows the parameter estimates for the final model specification for the case study, along with heteroskedasticity-consistent standard errors. Generally, parameters are found to be reasonable, and are predominantly highly significant. The parameters show that cyclists have a negative preference for riding longer distances, even more when going against the direction of traffic or on cobblestones. Regarding elevation, the steeper the gradient, the more nuisance each vertical meter causes.

Cycleways (in own trace) are the most preferred infrastructure type, and are highly preferred over other infrastructure types without motorised traffic, i.e. footways, shared paths, pedestrian zones. When motorised traffic is present, segregated bicycle tracks are preferred over bicycle lanes or no bicycle infrastructure. Large roads are deemed less attractive than smaller roads, however to a lesser extend for roads with segregated bicycle tracks.

Generally, preferences for different types of land-use differ considerably between cycleways and other infrastructure types (except regarding high urban areas). Cyclists generally do not prefer cycleways near water. In contrast, cycleways are highly preferred over other infrastructure types in green, industrial areas and in open landscape – perhaps due to isolation from heavy traffic in the latter two cases. Finally, all else equal, cyclists are found to have a slight preference for cycle super highway routes.

Modelled vs observed link flow shares

Figure 3 illustrates the estimated model's prediction for a single OD-relation as well as the observed flow share for the same OD-relation. As can be seen, the model functions as intended by allocating most links in the network zero flow share (i.e. inactive) and furthermore distributes flow realistically among used links, reproducing (to a large extent) the actual behaviour.

(a) Modelled **(b) Observed**

Figure 3: Modelled and observed link flows of a single OD pair

4. CONCLUSIONS

We have estimated a bicycle perturbed utility route choice model from a large dataset of GPS traces and a highly disaggregate network. Taking advantage of the efficiency that this approach offers, we tested a wide range of model specifications and we presented a model that best explains the route choice behavior of cyclists in the case study area. The results show consistency with the existing literature, indicating e.g. a negative preference for riding against the direction of traffic or the preference for segregated bicycle tracks in the presence of motorized traffic. We also showed how preferences for various bicycle infrastructure types differ, depending on the land-use characteristics of the route. Furthermore, we have demonstrated on an example that applying the estimated model does indeed result in inactive links between OD-relations and (by comparison to observed link flow shares) realistic distribution of flow between used links.

Further work could pursue an even better model specification by testing different attribute combinations and interaction effects in a range of models. This is facilitated by the computational efficiency of the estimation technique. Furthermore, additional network attributes could be added and their relevance for bicycle route choice evaluated. This could e.g. be information on intersection types or the presence of signalized intersections. Also, information about the planned network of cycle highways in the Greater Copenhagen Region could be used as a factor. This would reveal whether super cycle highways are preferred because they are super cycle highways, or whether bicycle superhighways are simply proposed and implemented along already preferred corridors.

Additional research could also look into the analysis of bicycle travel demand, i.e. how the number of bicycle trips in an OD-combination depends on the corresponding generalised cost associated with the optimal flow between this pair.

Finally, it would be relevant to perform counterfactual analyses based on various scenarios, eg. related to bicycle infrastructure investments or reductions, and quantify the corresponding benefits/disbenefits. Coupled with the relationship between bicycle demand and generalized costs, it would even be possible to provide an estimate of the increase/decrease in bicycle use when upgrading/degrading the network. This kind of analysis would address the concern of policy makers for evaluating ways to encourage more people to travel by bicycle.

REFERENCES

Banerjee, A., Łukawska, M., Jensen, A. F. & Haustein, S. (2021). "Facilitating bicycle commuting beyond short distances: insights from existing literature". *Transport Reviews*, pp. 1–25.

Broach, J., Dill, J. & Gliebe, J. (2012). "Where do cyclists ride? A route choice model developed with revealed preference GPS data". *Transportation Research Part A: Policy and Practice* 46.10, pp. 1730–1740.

Fosgerau, M., Frejinger, E. & Karlstrom, A. (2013). "A link based network route choice model with unrestricted choice set". *Transportation Research Part B: Methodological* 56, pp. 70–80.

Fosgerau, M., Paulsen, M. & Rasmussen, T. K. (2022). "A perturbed utility route choice model". *Transportation Research Part C: Emerging Technologies* 136, p. 103514.

Ghanayim, M. & Bekhor, S. (2018). "Modelling bicycle route choice using data from a GPS-assisted household survey". *European Journal of Transport and Infrastructure Research* 18.2.

Halldórsdóttir, K., Rieser-Schüssler, N., Axhausen, K. W., Nielsen, O. A. & Prato, C. G. (2014). "Efficiency of choice set generation methods for bicycle routes". *European Journal of Transport and Infrastructure Research EJTIR Issue* 14.144, pp. 332–348.

Hood, J., Sall, E. & Charlton, B. (2011). "A GPS-based bicycle route choice model for San Francisco, California". *Transportation Letters* 3.1, pp. 63–75.

Koch, T. & Dugundji, E. (2021). "Limitations of recursive logit for inverse reinforcement learning of bicycle route choice behavior in Amsterdam". In: *Annual Meeting of the Transportation Research Board*. Waschington DC, USA.

Koch, T., Knapen, L. & Dugundji, E. (2021). "Path complexity and bicyclist route choice set quality assessment". *Personal and Ubiquitous Computing* 25.1, pp. 63–75.

Mai, T., Fosgerau, M. & Frejinger, E. (2015). “A nested recursive logit model for route choice analysis”. *Transportation Research Part B: Methodological* 75, pp. 100–112.

Menghini, G., Carrasco, N., Schüssler, N. & Axhausen, K. (2010). “Route choice of cyclists in Zurich”. *Transportation Research Part A: Policy and Practice* 44.9, pp. 754–765.

OpenStreetMap (2020). *https://www.openstreetmap.org*.

Prato, C. G., Halldórsdóttir, K. & Nielsen, O. A. (2018). “Evaluation of land-use and transport network effects on cyclists’ route choices in the Copenhagen Region in value-of-distance space”. *International Journal of Sustainable Transportation*, pp. 1–12.

Pritchard, R. (2018). “Revealed Preference Methods for Studying Bicycle Route Choice—A Systematic Review”. *International Journal of Environmental Research and Public Health* 15.3, p. 470.

Rasmussen, T. K., Lukawska, M. & Paulsen, M. (2021). *An easily interpretable and efficient choice set generation method for bicycle route choices*. Presentation at the 5th Cycling Research Board Annual Meeting (CRBAM2021), Copenhagen, Denmark.

Skov-Petersen, H., Barkow, B., Lundhede, T. & Jacobsen, J. B. (2018). “How do cyclists make their way? - A GPS-based revealed preference study in Copenhagen”. *International Journal of Geographical Information Science* 32.7, pp. 1469–1484.

Ton, D., Duives, D., Cats, O. & Hoogendoorn, S. (2018). “Evaluating a data-driven approach for choice set identification using GPS bicycle route choice data from Amsterdam”. *Travel Behaviour and Society* 13, pp. 105–117.

Zimmermann, M., Mai, T. & Frejinger, E. (2017). “Bike route choice modeling using GPS data without choice sets of paths”. *Transportation Research Part C: Emerging Technologies* 75, pp. 183–196.