Congestion effect on bicycle route choices in Copenhagen

Andrés Fernández Palao, Anders Fjendbo Jensen, Thomas Kjær Rasmussen and Otto Anker Nielsen

Transport Modelling Division, Department of Management, Technical University of Denmark

Bygningstorvet 116B, 2800 Kgs. Lyngby, Denmark

anfpa@dtu.dk

1 Introduction

Bicycling is becoming more and more popular in cities around the world. For instance, Copenhagen is going to invest from 1.1 to 1.8 bio. DKK in bicycle infrastructure by 2025 (Municipality of Copenhagen, 2017). Due to this trend in urban cycling, bicycles are starting to be taken into account in transport modelling (e.g., Tønning and Vuk, 2017). However, bicycle transport is still included in a simplistic manner in most transport models, which could lead to a misunderstanding of cyclists' behaviours and preferences, as well as to bias when evaluating potential infrastructure investments.

Several bicycle route choice models have been estimated based on stated preference (SP) data (e.g. Bovy and Bradley, 1985; Hunt and Abraham, 2007; Sener, Eluru, and Bhat, 2009). Recently, the development of the GPS technology have made it easier to observe actual route choices, and consequently more studies based on revealed preference (RP) data have been carried out in the recent years (e.g., Prato, Halldórsdóttir, and Nielsen, 2018; Bernardi, Paix-Puello, and Geurs, 2018). Apart from the length of the trip, these studies found that steepness, turns, or cycling in the wrong direction have a significant negative effect. Furthermore, land-use attributes as well as the infrastructure also showed to have an influence on route choice preferences.

Beside these traditional attributes, the literature suggests that individual stress levels can influence cyclists' behaviour. Moreover, "bicycle stress level" is a theory developed by Geelong Bike Plan Team in Australia in 1978 (Harkey, Reinfurt & Knuiman, 1998; Sorton & Walsh, 1994). This theory assumes that cyclists, besides focusing on reducing travel time and physical effort, also to a great extent seeks to reduce their individual stress level e.g. caused by motor traffic. Since then, different concepts for cyclists stress level have been employed as indicators and used in simplified calculations when prioritising bicycling infrastructure projects – most often with the goal of being able to evaluate bicycling behaviour and prioritise bicycling projects with a minimum of input (Mekuria, Furth & Nixon, 2012). For example, Sorton & Walsh (1994) used the bicycling stress level theory to simplify complex physical and psychological information to basic classifications from 1-5. However, the literature is missing examples of quantitative modelling of stress factors.

The intention of our study is to contribute to the literature of bicycle behaviour by estimating how stress factors affect cyclist behaviour in a detailed route choice model. In particular, we seek to explain how bicycle congestion affects cyclists at an individual level. Especially in Copenhagen, there are many locations where cyclists experience congestion and this might mean that cyclists avoid these locations if they feel less safe or the density of cyclists is too high. The route choice model is based on RP data and carried out in the Greater Copenhagen area, Denmark. While the GPS data is the same as the one used in Prato, Halldórsdóttir, and Nielsen (2018), the bicycle network has been through a major update. The structure of the paper is as follows. Section 2 presents the data, while Section 3 presents the estimation methodology. Section 4 presents and discusses the results and Section 5 presents the conclusion and potential avenues for new research.

2 Data

2.1 Bicycle Network

The bicycle network is comprised by 77,455 unidirectional bicycle links, 58,181 nodes, covers the Greater Copenhagen area, and contains very detailed information regarding different link attributes. For instance, although land-use information along the links was already included in previous studies (e.g., Prato, Halldórsdóttir, and Nielsen, 2018), this information only expressed the presence or not of each land-use category near to each link. Therefore, it was not possible to investigate differences in land-use distances within a link. In this new network, we count with the exact number of meters of each land-use category on both sides of each link. In addition, our bicycle network includes information regarding the hourly capacity and average flow of every link. Assuming that the relation between flow and speed accords with the Bureau of Public Roads (BPR)-formula, the differences in speed and travel time between the congested and the free flow scenarios can be calculated. Therefore, the links were classified in three different congestion levels using the relation between the congested and uncongested speeds.

2.2 GPS data gathering, processing and map-matching

The GPS data was gathered in the Greater Copenhagen area in 2012 and 2013 by giving GPS trackers to a total of 318 cyclists. The raw data was filtered and bicycle trips were subsequently map-matched to the bicycle network by applying the algorithm described in Nielsen and Jørgensen (2004). An additional filtering after the map matching to remove trips erroneously map-matched resulted in a dataset containing a total of 2,496,328 GPS points and 4,630 bicycle trips, made by 301 different individuals.

2.3 Similarity between most-used links and congested links

In Figure 1a below we present the number of cyclists per link in order to show the links that are most used by the observed routes. Clearly, the most-used routes are following the larger main roads in the network. In Figure 1b, the level of congestion for each link is shown and it is clear that there is a strong overlap between busy links and congested links. This indicates that cyclists like to follow the most direct routes, maybe despite discomforts such as congestion.

Figure 1a - Busiest observed links

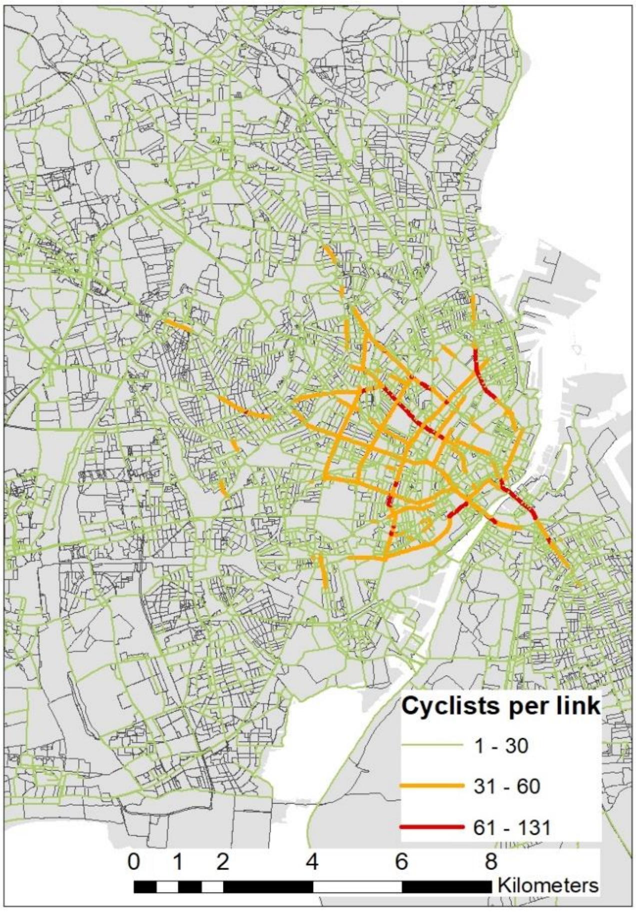
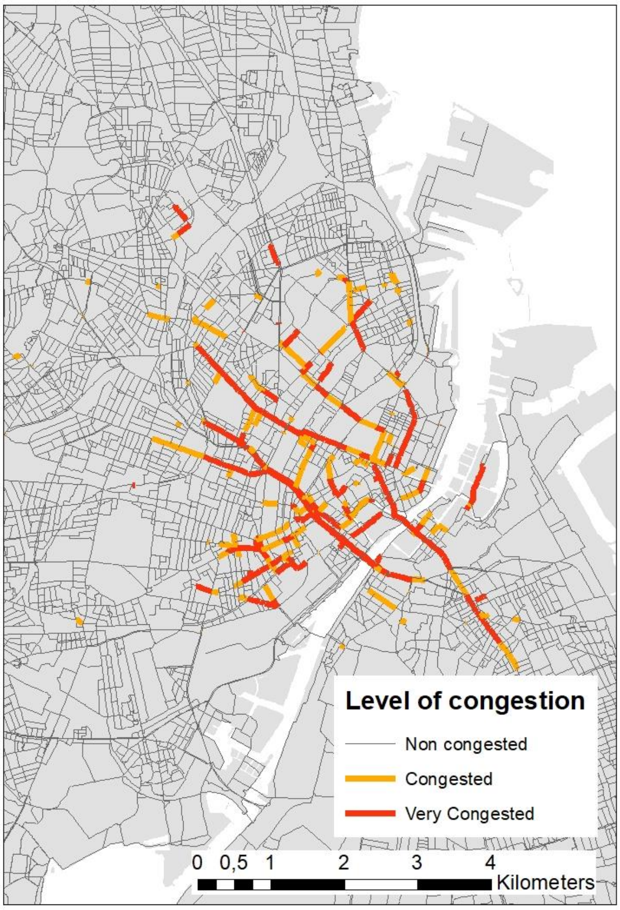


Figure 1b - Level of congestion



Due to this result, we found it necessary to include the motorised road type as part of the modelling to avoid implausible results due to correlation between road type and bicycle congestion variables.

3 Model development and methodology

3.1 Choice set generation

The present study applied a doubly stochastic generation (DSGF) method (Bovy and Fiorenzo-Catalano, 2007) with a multi-attribute cost function for generating the choice set.

Apart from considering route length, the cost function includes link type (i.e., road, road with bicycle lane, road with bicycle path, bicycle path, footpath, steps), motorized traffic lanes (i.e., zero, one, two, three, and four), surface types (i.e., paved or unpaved), land-use information (i.e., nature or not nature), and cycling in the wrong direction or not. As a result, the link cost function was defined as follows:

$$C_a = \beta_T TT_a + \beta_D D_a + \sum_{\{k=1\}}^{K} \beta_{P_k} P_{ak} D_a + \sum_{\{h=1\}}^{H} \beta_{S_h} S_{ah} D_a + \sum_{\{l=1\}}^{L} \beta_{R_l} R_{alD_a} + \beta_U U_a D_a + \beta_W W_a D_a + \varepsilon_a (1)$$

where C_a is the cost of link a , TT_a is the total travel time of link a, D_a is the length of link a, P_{ak} is equal to 1 when link a is link type k $(k = 1, \dots, K)$, and 0 otherwise, S_{ah} is equal to 1 when link a is surface type h $(h = 1, \dots, H)$, and 0 otherwise, R_{al} is equal to 1 when link a counts with l motorized traffic lanes l $(l = 1, \dots, L)$, and 0 otherwise, U_a is equal to 1 when there is some kind of natural element (i.e., water, parks, forests, sand) at the right or left side of link a, and zero otherwise, W_a is equal to 1 when cycling in the wrong direction, and 0 otherwise. Due to the doubly stochastic nature of the DSFG method, the different parameters(i.e., β_T, β_D, β_{P_k}, which is different for each link type k, β_{S_h}, which is different for each surface type h, β_{R_t}, which is different for each number of motorized traffic lanes t, β_U and β_W) were log-normally distributed, while the error term ε_a, was gamma distributed.

3.2 Model formulation

With the intention of accounting for similarity among alternatives, a Path-Size logit (PSL) model (Ben-akiva and Bierlaire, 2009) was estimated. Additionally, as all previous studies concluded that length is the attribute which affects route choices the most, the attributes defined by length were modelled in value of distance (VoD) space, obtaining the influence of such attributes with respect to the length (distance) of the route. The utility function is defined as follows:

$$U_{nj} = \gamma\left(d_{nj} + \boldsymbol{\beta_y y_{nj}}\right) + \boldsymbol{\beta_z z_{nj}} + \varepsilon_{nj} \qquad (2)$$

where n is the cyclist, and j is a route alternative included in the choice set J. As the route length parameter was fixed to 1, γ is a parameter which represents the original preference space parameter for route length d_{nj}. $\boldsymbol{y_{nj}}$ and $\boldsymbol{z_{nj}}$ are vectors of the attributes defined and not defined by length, respectively, while $\boldsymbol{\beta_y}$ and $\boldsymbol{\beta_z}$ are the corresponding parameters to be estimated, and ε_{nj} is the error term.

As a result, the probability P_{nj} that cyclist n is going to choose route j was defined as follows:

$$P_{nj} = \frac{exp(\gamma(d_{nj} + \boldsymbol{\beta_y}\, \boldsymbol{y_{nj}}) + \boldsymbol{\beta_z}\, \boldsymbol{z_{nj}} + \varepsilon_{nj} + \beta_{PS}\, \ln PS_j)}{\sum_{\{l \in C\}} \exp(\gamma\left(d_{nl} + \boldsymbol{\beta_y}\, \boldsymbol{y_{nl}}\right) + \boldsymbol{\beta_z}\, \boldsymbol{z_{nl}} + \varepsilon_{nl} + \beta_{PS} \ln PS_l)} \qquad (3)$$

where the terms $\ln PS_j$ and $\ln PS_l$ are the path size factors of routes j and l, respectively, while the β_{PS} is the parameter related with them that has to be calculated.

4 Results

4.1 Generated route choice sets

For a coverage threshold of 80%, the choice set was able to replicate the 60.91% of the observed routes. The maximum number of generated alternatives per observed route was defined as 100, obtaining an average of 83.59 and a standard deviation of 37.03. Comparing the observed routes to the generated alternatives, the observed routes are, reasonably, on average shorter, have less kilometres cycled in the wrong direction, less number of turns, and lower cumulative elevation gain.

4.2 Model estimation

Initially, models including the path size term were tested. However, a negative estimate value for the path size term was obtained, contradicting the theoretically expected positive value (Ben-Akiva and Bierlaire, 2009). As a result, it was decided to not include the path size term, reducing the model estimated to a Multinomial Logit (MNL) model.

When we included both road type and level of (bicycle) congestion in the model, we found that cyclists prefer larger roads and roads with more congestion. Furthermore, the latter result appeared to have the greatest effect, which we found implausible. Instead, we tested each variable in separate models and found that the model including road type obtained the best fit when looking at Log-likelihood. Also, we found it more plausible that the main roads were chosen due to practical features related to each road type (such as more direct routes, fewer stops, or bicycle path quality) rather than that they are congested. Thus, the model with road types was chosen for further work. Unfortunately, this means that we so far are not able to present a reliable result for bicycle congestion.

Table 1 presents the estimates for the best model specification. The interpretation of the attribute parameters differs depending on if modelled in VoD space or in preference space. On one hand, VoD space estimates are the rates of substitution, so when the sign is positive (negative) the route is perceived as longer (shorter) because of the effect of that attribute. For instance, the estimated value of cycling in the wrong way for females is equal to 1.30, meaning that for females the route is perceived as 130% longer when cycling in the wrong direction. On the other hand, when the estimates are obtained by modelling in preference space, a positive (negative) sign means that this attribute affects route choices in a positive (negative) way.

Table 1: Model estimates of the final Multinomial Logit (MNL) model

Parameter	**Estimate**	**St.err.**	**t-test**	***p***
Not modelled in VoD space				
Turns				
Right turns	-0.059	0.013	-4.65	0.00
Left turns	-0.058	0.013	-4.29	0.00
Cumulative elevation gain (m)				
Above 35 vertical meters/km	-0.032	0.009	-3.62	0.00
Modelled in VoD space				
Length (m)	1.000	-	-	-
Wrong way (m)	1.30	0.096	13.46	0.00
Wrong way-males (m)	-0.267	0.089	-2.98	0.00
Bicycle infrastructure type (m)				
Road without bicycle facilities	-	-	-	-
Road with bicycle lanes	-0.076	0.061	-1.24	0.21**
Road with segregated bicycle path	-0.296	0.025	-11.67	0.00
Bicycle path in own trace	0.278	0.106	2.63	0.01
Bicycle path in own trace – medium cyclists	-0.458	0.107	-4.27	0.00
Bicycle path in own trace – fast cyclists	-0.382	0.110	-3.48	0.00
Footpath	0.233	0.146	1.59	0.11**
Steps	0.500	0.824	0.61	0.54**
Number of motorised traffic lanes (m)				
0 lanes	0.339	0.056	6.05	0.00

1 lane	-1.16	0.294	-3.94	0.00
2 lanes	-	-	-	-
3-4 lanes	0.115	0.038	3.03	0.00
Motorised road type (m)				
Small roads	-	-	-	-
Medium roads	-	-	-	-
Large roads	0.205	0.075	2.72	0.01
Large roads – medium speed cyclists	-0.260	0.082	-3.18	0.00
Large roads – fast cyclists	-0.452	0.0853	-5.30	0.00
Land-use attributes / right side (m)				
Hydro	-	-	-	-
Green restricted areas	-	-	-	-
Green areas/ Parks	-	-	-	-
Green areas/ Forests	0.644	0.157	4.11	0.00
Green areas/ Forests - males	-0.086	0.187	-0.46	0.65**
Industrial and technical areas	0.145	0.066	2.20	0.03
Urban / High residential and centre areas	0.198	0.052	3.82	0.00
Urban / Low residential areas	0.263	0.059	4.43	0.00
Model parameters				
Gamma	-0.002	0.0001	-19.61	0.00
Number of estimated parameters				24
Number of observations				3461
Null log-likelihood				-15,036.860
Final log-likelihood				-11,270.134
Adjusted rho-square				0.250

Note: * not significant at the 95% level; ** not significant at the 90% level.

According to the results, turn frequency has a negative influence and cumulative elevation gain is also seen as a negative factor when the steepness is higher than 35 vertical meters per kilometre. Cyclists clearly dislike cycling in the wrong direction and especially females, who perceive the route as 130% larger, while for males this perception is 103.3%.

With regard to the bicycle infrastructure, cyclists perceive routes as 29,6% shorter when cycling in roads with segregated bicycle paths, always in comparison to the reference (roads without bicycle facilities). Roads with bicycle lanes were not significantly different from the reference, which contradicts the findings of Prato, Halldórsdóttir, and Nielsen (2018). In addition, it was found that for slow cyclists (average cycling speed lower than 15.5km/h) bicycle paths in own trace are very unattractive (27.8% longer), while for medium and fast cyclists they are more attractive than the reference (18% and 10.4% shorter, respectively).

In line with the preference for roads with segregated bicycle paths, it was found that roads with 2 motor traffic lanes are preferred over paths without motor traffic lanes, showing a clear cyclists' preference for bicycle facilities next to roads. Besides, roads with two motor traffic lanes are considered less attractive than roads with one, but more attractive than roads with three or four, meaning that cyclists could be affected by motor traffic levels, as concluded by Prato, Halldórsdóttir, and Nielsen (2018).

Only the land-use attributes at the right side showed a significant effect. Routes next to water, green restricted areas, and parks were used as reference and were shown to be the preferred ones over all the others. In addition, routes next to low residential areas, and high residential and centre areas, are perceived as longer, but the latter in a lesser extent than the former (26.3% and 19.8% larger, respectively). The reason could be that cycling in high residential and centre areas is usually more intuitive and attractive than in more remote areas. Finally, there is a clear gender heterogeneity regarding cycling next to forests, which could be explained by the generally poor lighting and surface condition in such places.

Finally, the effect of motorised road type, which was found to be correlated with congestion, showed a fast cyclists' strong preference (24,7% shorter), and a medium fast cyclists' smaller preference (5,5% shorter), both for large roads compared to small and medium roads. In contrast, for slow cyclists the strong preference was for small and medium roads over large ones (20,5% shorter). The reason behind could be that the faster cyclists usually follow their mental maps composed by large and main roads. These cyclists could be fast due to individual characteristics or due to the purpose of the trip. We need to test further, whether other of the characteristics present in the data could explain this behaviour better than cycling speed. However, the results indicate that for this group, congestion seems not to be an issue. Also, the results show a clear heterogeneity across cyclists when it comes to the preference for road types. There seem to be a slower group of cyclists who prefer smaller roads which also often are less congested.

5 Conclusion

This study presents results from a bicycle route choice model based on RP data, which can be used by decision makers as a tool for evaluating the effectiveness of cycling-related investments and policies. Similar to previous studies, we find that cyclists prefer shorter routes, low turn frequency, less steepness and routes with good cycling infrastructure, especially roads with segregated bicycle paths. Based on improved land-use variables we also find that cyclists prefer green surroundings and that high residential and centre areas are preferred to low residential areas.

The main goal of the study was to analyse whether also stress factors have an effect on cyclist's behaviour and in particular how congestion in Copenhagen might affect route choice behaviour. As the initial analysis of the data indicated that the most used links in the network were also those with the most congestion, we found it necessary to control for road type in the model as the most used routes were following the larger (most direct) roads. Due to large correlation between road type and the level of congestion, we did not manage to include the congestion attribute in the final model. The modelling results confirm that cyclists overall prefer larger roads, but only for cyclists categorized as medium and fast speed. The speed of the cyclists might indirectly be related to individual characteristics but also to trip purpose, which we unfortunately do not have in the data. Fast cyclists might prefer the direct routes and better cycling conditions which are usually related to the main roads, and with the current modelling results they seem unaffected by congestion. This leaves us with the slowest group of cyclists, who associate large roads with a 20.5% addition in route length. Further analysis with more advanced models should indicate whether this behaviour is due to a congestion effect or other factors related with large roads.

6 References

Ben-akiva, M. and Bierlaire, M. “Discrete Choice Methods And Their Applications To Short Term Travel Decisions”. 2009. doi: 10.1.1.37.2711.

Bernardi, S., Paix-Puello, L. L., and Geurs, K. “Modelling route choice of Dutch cyclists using smartphone data”.Journal of Transport and Land Use, 11(1), 2018.ISSN 1938-7849. URL https://www.jtlu.org/index.php/jtlu/article/view/1143.

Bovy, P. H. L. and Bradley, M. A. “ROUTE CHOICE ANALYZED WITH STATED-PREFERENCE APPROACHES”.Transportation Research Record, (1037):1–11,1985. ISSN 21694052, 03611981.

Bovy, P. H. L. and Fiorenzo-Catalano, S. “Stochastic route choice set generation: Behavioral and probabilistic foundations”.Transportmetrica, 3(3):173–189, 2007.ISSN 19440987, 18128602.

Harkey, D., Reinfurt, D., & Knuiman, M. “Development of the Bicycle Compatibility Index”. Transportation Research Record: Journal of the Transportation Research Board, 1636, 13–20, 1998. https://doi.org/10.3141/1636-03

Hunt, J. D. and Abraham, J. E. “Influences on bicycle use”.Transportation, 34(4):453–470, Jul 2007. ISSN 1572-9435. doi: 10.1007/s11116-006-9109-1. URLhttps://doi.org/10.1007/s11116-006-9109-1.

Municipality of Copenhagen. “Cykelstiprioriteringsplan 2017-2025”. Technical report, KøbenhavnsKommune, 2017.

Mekuria, M. C., Furth, P. G., & Nixon, H. “Low-Stress Bicycling and Network Connectivity”. MINETA Transportation Institute publications, Report 11-19, 2012.

Nielsen, O. A. and Jørgensen, R. M. “Map-matching algorithms for GPS-data -methodology and test on data from the AKTA road pricing experiment in Copenhagen”. World Conference on Transport Research Society (wctrs), 2004.

Prato, C. G., Halldórsdóttir, K., and Nielsen, O. A. “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, 0(0):1–12, 2018. doi: 10.1080/15568318.2018.1437236. URL https://doi.org/10.1080/15568318.2018.1437236.

Sener, I. N., Eluru, N., and Bhat, C. R. “An analysis of bicycle route choice preferences in Texas, US”.Transportation, 36(5):511–539, 2009. ISSN 15729435,00494488. doi: 10.1007/s11116-009-9201-4.

Sorton, A., & Walsh, T. (1994). “Bicycle stress level as a tool to evaluate urban and suburban bicycle compatibility”. Transportation Research Record, (1438), 17–24.

Tønning, A. and Vuk, G. “Modelling of bicycle transport in OTM”.Trafikdage på Aalborg Universitet, 2017. ISSN 1603-9696.