Dette resumé er publiceret i det elektroniske tidsskrift
Artikler fra Trafikdage på Aalborg Universitet
(Proceedings from the Annual Transport Conference at Aalborg University)
ISSN 1603-9696
www.trafikdage.dk/artikelarkiv

Using crowd source data in bicycle route choice modeling

Anders Fjendbo Jensen, afjje@dtu.dk; Andrés Fernandez Palao, anfpa@dtu.dk; Thomas Kjær Rasmussen, tkra@dtu.dk; Otto Anker Nielsen, oani@dtu.dk
Transport Modeling Division, Department of Technology, Management and Economics, Technical University of Denmark

Abstract
We present a bicycle route choice model modelled in Value of Distance space based on revealed GPS data and an improved network with a very detailed representation of the bicycle infrastructure and detailed calculations of the related attributes. Beside common attributes, such as trip-length, bicycle infrastructure, elevation gain and land-use, we utilize crowd sourced data to analyse the influence on route choice of intersections where a large number of individuals have indicated large congestion. The results show that the individual's perception of high congestion in an intersection has an effect on the route choice behaviour, and that passing such intersections perceived as congested is connected to a lower utility for a given route.

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' behaviour and preferences, as well as to bias when evaluating potential infrastructure investments.

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.

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 (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.

The intention of our study is to obtain a better understanding of the effect of individual's perception of the bike network on their route choice by utilizing crowd sourced data. Such data is often collected by e.g. municipalities when they would like their citizens to help identify where there are problems in their bicycle network. The potential benefit of such a way of collecting data could be that the users might have a different perception of the network than what can be measured quantitatively. Or, if it is possible to measure quantitatively, it might be expensive to do so.

Data

The bicycle network was originally developed in connection with the previous ACTUM research project and used and improved in Halldórsdóttir (2015) and OTM (Tønning and Vuk, 2017). In connected with a recent project at DTU, the network has been further developed with improved representation of the bicycle infrastructure and more exact calculations of elevation and land-use variables.

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.

A choice set generation process was applied to generate the relevant alternatives to each of the observed trips and calculate the attributes necessary for the model estimation. 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.

The crow sourced data was collected by Copenhagen Municipality in 2016 (Streuli, 2017). In their homepage, they asked interested users to indicate specific locations on an online interactive map where they though there was a missing bikepath, where the current bikepath should be wider or where there is an intersection with large congestion. The data was used by Copenhagen Municipality in their analyses towards developing their bicycle infrastructure plan 2017-2025. Even though most locations indicated by the users fitted the expectations of the administration, the crowd sourced data meant that some locations were given more attention.

Method

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:

$$U_{nj} = \gamma\left(d_{nj} + \boldsymbol{\beta_y y_{nj}}\right) + \boldsymbol{\beta_z z_{nj}} + \varepsilon_{nj} \qquad (1)$$

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:

$$P_{nj} = \frac{exp(\gamma(d_{nj} + \boldsymbol{\beta_y}\, \boldsymbol{y_{nj}}) + \boldsymbol{\beta_z}\, \boldsymbol{z_{nj}} + \varepsilon_{nj} + \beta_{PS} \ln PS_j)}{\sum_{\{l \in C\}} \exp(\gamma(d_{nl} + \boldsymbol{\beta_y}\, \boldsymbol{y_{nl}}) + \boldsymbol{\beta_z}\, \boldsymbol{z_{nl}} + \varepsilon_{nl} + \beta_{PS} \ln PS_l)} \qquad (2)$$

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.

In order to utilize the crowd sourced data, relevant attributes was calculated with geodata processing methods in ArcGIS. At this stage we calculated an attribute defined by the number of users who indicated that an intersection is very congested. The process is illustrated in Figure 1. As each user would indicate an intersection by pointing at a specific location in the interactive map instead of e.g. specifying it by street names, the data consists of a cloud of observations around an intersection. Such a cloud is then identified and the number of observations within each cloud is calculated. This count information is for each cloud joined to each overlapping turn movement in the corresponding intersection. We then labelled all turn movements whether they belonged to an intersection where less than 50 users indicated that this intersection was congested or whether 50 or more users indicated whether this intersection was congested. The final route choice attribute was then describing how many intersections of each type a route was crossing.

Figure 1 – Calculating congested intersection attribute from crowd source data

Results

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.

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 1.28, meaning that for females the route is perceived as 128% 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.

Tabel 1 – eksempel

Parameter	**Estimate**	**St.err.**	**t-test**	***p***
Not modelled in VoD space				
Turns				
Right turns	-0.057	0.013	-4.32	0.00
Left turns	-0.054	0.013	-4.00	0.00
Congested turns	-0.148	0.075	-1.99	0.05
Cumulative elevation gain (m)				
Above 35 vertical meters/km	-0.032	0.009	-3.59	0.00
Modelled in VoD space				
Length (m)	1.000	-	-	-
Wrong way (m)	1.28	0.097	13.30	0.00
Wrong way-males (m)	-0.263	0.090	-2.91	0.00
Bicycle infrastructure type (m)				
Road without bicycle facilities	-	-	-	-
Road with bicycle lanes	-0.044	0.062	-0.72	0.47**
Road with segregated bicycle path	-0.231	0.032	-7.17	0.00
Bicycle path in own trace	0.303	0.107	2.82	0.00
Bicycle path in own trace – medium cyclists	-0.471	0.109	-4.34	0.00
Bicycle path in own trace – fast cyclists	-0.421	0.111	-3.79	0.00
Footpath	0.166	0.146	1.14	0.26**
Steps	0.704	0.824	0.85	0.39**
Number of motorised traffic lanes (m)				
0 lanes	0.323	0.056	5.74	0.00
1 lane	-1.18	0.296	-3.99	0.00
2 lanes	-	-	-	-
3-4 lanes	0.117	0.039	2.98	0.00
Motorised road type (m)				
Small roads	-	-	-	-
Medium roads	-0.143	0.035	-4.06	0.00
Large roads	-0.016	0.056	-0.29	0.77**
Large roads – medium speed cyclists	-0.085	0.054	-1.57	0.12**
Large roads – fast cyclists	-0.203	0.057	-3.60	0.00
Land-use attributes / right side (m)				

Hydro	-	-	-	-
Green restricted areas	-	-	-	-
Green areas/ Parks	-	-	-	-
Green areas/ Forests	0.657	0.157	4.18	0.00
Green areas/ Forests - males	-0.169	0.190	-0.89	0.37**
Industrial and technical areas	0.156	0.067	2.32	0.02
Urban / High residential and centre areas	0.222	0.053	4.16	0.00
Urban / Low residential areas	0.281	0.061	4.64	0.00
Model parameters				
Gamma	-0.002	0.0001	-19.61	0.00
Number of estimated parameters				26
Number of observations				3461
Null log-likelihood				-15,036.860
Final log-likelihood				-11,270.486
Adjusted rho-square				0.250

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

As expected, a higher number of turns reduces the utility of a route, but interestingly we did not find any difference between left turn or right turn. Utilizing the crowd sourced data, we show that a higher number of intersections with more than 50 user points further reduces the utility of a route greatly, indicating that these intersections are especially unattractive for cyclists.

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 128% larger, while for males this perception is 101.7%. With regard to the bicycle infrastructure, cyclists perceive routes as 23.1% 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 (30.3% longer), while for medium and fast cyclists they are more attractive than the reference (16.8% and 11.8% 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 lanes are considered less attractive than roads with one, but more attractive than roads with three or four, possibly meaning that cyclists are 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 (28.1% and 22.2% larger, respectively). The reason could be that bicycle facilities in high residential and centre areas are usually better maintained. 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.

The effect of motorised road type, which was found to be correlated with congestion, showed a fast cyclists' strong preference (21.9% shorter) for large roads compared to small and medium roads, while for medium and slow cyclists the difference was not significant. In addition, cyclists also show a preference for medium roads over small ones. 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.

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.

Halldórsdóttir, K. . Behavioural models for cycling - Case studies of the Copenhagen Region (Ph.D. thesis). Department of Transport, Technical University of Denmark, Kgs. Lyngby, Denmark, 2015.

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.

Streuli, M. H., “Ny cykelstriprioriteringsplan for København”, Trafik og Veje, May, 2017.

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.