

Evaluation of land-use and transport network effects on cyclists' route choices in the Copenhagen Region in value-of-distance space

Prato, Carlo Giacomo; Halldórsdóttir, Katrín; Nielsen, Otto Anker

Published in: International Journal of Sustainable Transportation

Link to article, DOI: 10.1080/15568318.2018.1437236

Publication date: 2018

Document Version Peer reviewed version

Link back to DTU Orbit

Citation (APA):

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*, *12*(10), 770-781. https://doi.org/10.1080/15568318.2018.1437236

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.

- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Evaluation of land-use and transport network effects on cyclists' route choices in the Copenhagen Region in value-of-distance space

Carlo Giacomo PRATO ^{a*}, Katrín HALLDÓRSDÓTTIR ^b, Otto Anker NIELSEN ^b

^a School of Civil Engineering, The University of Queensland, St Lucia, 4072 Brisbane, Australia

^b Department of Management Engineering, Technical University of Denmark Bygningstorvet 116B, 2800 Kgs. Lyngby, Denmark

* Corresponding author School of Civil Engineering, The University of Queensland, St Lucia, 4072 Brisbane, Australia Phone: +61.7.33651569, E-Mail: c.prato@uq.edu.au

Abstract

Growing interest in sustainable transportation systems has driven decision-makers towards policies and investments aimed at promoting cycling, but little to no effort has been made towards incorporating bicycle transport in transport planning models. This study contributes towards this direction by estimating a bicycle route choice model in value-of-distance space from a large sample of 3384 cycling trips that were traced with GPS devices in the Copenhagen Region. The novelty of this study lies in (i) observing cyclists' behaviour in a cycling-oriented country, (ii) exploiting rich data about the cycling environment, (iii) estimating the model in value-of-distance rather than preference space, and (iv) not focusing only on preferences for traditional variables (e.g., distance, turns, hilliness, intersections, motorised road characteristics), but also on perceptions and preferences for bicycle facilities (e.g., bicycle lanes, bicycle paths, bicycle traces) and land-use designations (e.g., residential, industrial, sports, scenic areas). The findings from the model show that: (i) cyclists exhibit heterogeneous preferences for avoiding right and left turns, cycling the wrong way, using roundabouts and bridges, and cycling alongside residential and scenic areas; (ii) cyclists dislike cycling on unpaved and hilly surfaces and alongside larger roads; (iii) cyclists have clear perceptions about different types of bicycle facility, with a preference for bicycle lanes and segregated paths; (iv) cyclists have clear perceptions about land-use designations, with a preference for cycling alongside sports and scenic areas; (v) time-of-day and air temperature contribute to the perceptions of cyclists and their preferences for bicycle facilities and land-use designations.

Keywords: Cycling; Route choice; Land-use; Bicycle infrastructure; Doubly stochastic generation function; Generalised mixed path size logit.

1. Introduction

In recent years, growing interest in sustainable transportation systems has driven decision-makers towards policies and investments aimed at promoting cycling. However, while notable effort has been posed towards representing motorised transport in planning models, little to no effort has been made towards incorporating bicycle transport in those models with the consequent unavailability to decision-makers of quantitative forecasts of the impact of those policies and investments. This study contributes towards the incorporation of cycling into planning models by providing knowledge about cyclists' preferences not only in terms of traditional variables (e.g., distance, left turns, right turns, hilliness, road characteristics), but also in terms of bicycle facilities and land-use designations that decision-makers could invest and decide upon. Moreover, this study proposes rates of substitutions that express the preferences for various factors as a measure per unit of distance via the estimation of a model in value-of-distance space.

Existing literature on cyclists' route choice behaviour is based on either stated preferences (SP) data (e.g., Bovy & Bradley, 1985; Axhausen & Smith, 1986; Hopkinson & Wardman, 1996; Stinson & Bhat, 2003; Krizek, 2006; Hunt & Abraham, 2007; Tilahun et al., 2007; Sener et al., 2009) or revealed preferences (RP) data (e.g., Aultman-Hall et al., 1997; Shafizadeh & Niemeier, 1997; Hyodo et al., 2000; Howard & Burns, 2001; Larsen et al., 2013; Snizek et al., 2013; Yeboah & Alvanides, 2015; Zimmerman et al., 2017). As SP data bear the advantage of controlling for the experimental environment but the disadvantage of not observing actual behaviour, large-scale RP data are to be preferred because the benefit of recording cyclists' actual preferences surpasses possible technical problems related to tracing and map-matching, as well as possible modelling issues associated with generating plausible alternative routes for model estimation. When combined with the formulation and estimation of route choice models, large-scale RP studies create the opportunity for providing quantitative forecasts of cyclists' behaviour.

Large-scale RP studies on cyclists' route choices have been performed in cities with both a cycling-oriented culture like Zurich (Menghini et al., 2010) as well as North American cities with a car-oriented culture like San Francisco (Hood et al., 2011), Portland (Broach et al., 2012), Waterloo (Casello & Usyukov, 2014), and Eugene (Zimmermann et al., 2017). These studies revealed that cyclists have a preference to a

greater extent for shorter distances (Menghini et al., 2010; Hood et al., 2011; Broach et al., 2012; Casello & Usyukov, 2014; Zimmermann et al., 2017), and to a lesser extent for lower gradients (Menghini et al., 2010; Hood et al., 2011; Broach et al., 2012; Casello & Usyukov, 2014; Zimmermann et al., 2017), fewer traffic lights (Menghini et al., 2010; Broach et al., 2012), lower numbers of turns (Broach et al., 2012; Zimmermann et al., 2017) and shorter distances cycling the wrong way in one-way streets (Hood et al., 2011). These studies uncovered also that cyclists have a preference for designated lanes rather than paths in San Francisco (Hood et al., 2011), for segregated paths and dedicated bridges rather than lanes in Portland (Broach et al., 2012), for lanes and boulevards dedicated to cycling in Eugene (Zimmermann et al., 2017), and more in general for marked routes in Zurich (Menghini et al., 2011). While it is understandable that distance was found to be the most relevant factor for cyclists' route choice (especially for commuters), it is plausible that other factors emerged as less relevant or completely irrelevant because of the samples being most likely affected by self-selection bias (e.g., cycling enthusiasts) and the environment description being most likely impaired from sufficient observations of bicycle infrastructure to have significant estimates. It should be noted that a recent study proposed a bi-objective optimisation model where travel time was considered alongside the suitability of route to cycling (Ehrgott et al., 2012), a qualitative measure of how a route would appeal and consequently enter the route choice set of heterogeneous cyclists. Moreover, evidence from previous RP studies (Menghini et al., 2010; Hood et al., 2011; Broach et al., 2012; Casello & Usyukov, 2014; Zimmermann et al., 2017) suggested that cost functions containing variables other than travel time should be considered. However, route choice models were not estimated with these variables.

This study presents a large-scale RP study on cyclists' route choices from the perspective of an established cycling city. With high cycling market shares that reach about 37% of the commuting trips (Pucher & Buehler, 2012) and a highly developed and connected bicycle network from significant investments in dedicated infrastructure, the Copenhagen Region offers the ideal setting for a large-scale study aimed at understanding the effects of infrastructure and land-use on the choices of cyclists across the entire population. Accordingly, this study expands the existing body of literature by formulating and estimating a route choice model of cyclists in the Copenhagen Region. A generalised mixed path size logit model was estimated to uncover the determinants of

3384 cycling route choices while accounting for heterogeneity across the 291 cyclists, preference consistency within the cyclists, and similarities across routes. Specifically, the route choice model was not estimated in the traditional preference space, but rather in value-of-distance (VoD) space in order to evaluate the contribution of each factor per unit of distance. Notably, the estimation in VoD space echoes the estimation in willingness-to-pay (WTP) space (or value-of-time space) in that the model is estimated while accounting for the ratio with respect to the distance parameter (for details about WTP space model estimation, see Train and Weeks, 2005; Scarpa et al., 2008; Thiene and Scarpa, 2009; Hensher and Greene, 2011).

Findings might be useful for emerging cycling cities when looking in particular at the effects of infrastructure and land-use on the cyclists' route choices. This study analysed a rich set of bicycle network attributes such as distance, number of turns, number and type of intersections (e.g., give way, stops, roundabouts, traffic lights), gradient and distance cycling in the wrong way. Alongside the bicycle network, this study investigated a set of motorised network attributes describing road type, number of traffic lanes, speed limits, bridges and tunnels. Most relevantly, this study looked at a detailed set of bicycle infrastructure attributes, such as facility type (i.e., segregated bicycle path, bicycle lane, bicycle path in own traces, footpath, and steps), bridge availability, and surface type, as well as land-use attributes along the route that have been shown relevant to positive cycling experiences (e.g., Snizek et al., 2013), but have not been incorporated in route choice models. Last, this study checked whether preferences in the route choice relate to personal attributes (e.g., gender, age, average speed profile), trip related attributes (e.g., trip purpose, time-of-day, type of day, darkness), and weather attributes (e.g., temperature, sunshine, wind, precipitation).

The remainder of this paper is structured as follows. Section 2 describes the methodology for collecting and processing GPS data as well as formulating and estimating the generalised mixed path size logit in VoD space for the representation of cyclists' route choice behaviour. Section 3 illustrates the case-study with emphasis on the bicycle network and the sample characteristics. Section 4 presents the estimation results and illustrates the findings of the study. Section 5 draws conclusions from this large-scale RP study.

2. Methods

2.1 GPS data collection and processing

The data collection relied on GPS devices that were sent to a sample of participants in the Danish National Travel Survey (in Danish, TU Transportvaneundersøgelsen) that collects travel diaries of a representative sample of the Danish population between 10 and 84 years old via the administration of about 1000 interviews per month. The sampling criteria were that the respondents participated in the TU survey within 6 to 12 months prior to data collection, reported the use of the bicycle in their travel diary, lived in the Copenhagen Region, and were at least 16 years old (according to privacy regulations imposed by the Danish Data Protection Agency). Moreover, the sampling occurred over three rounds with the intention of obtaining a representative sample of the cyclists in the TU survey as well as covering different seasons: (i) from October to December 2012, (ii) from June to July 2013, and (iii) from August to October 2013. In addition to the GPS traces over a period of eight days, the participants were invited to fill a questionnaire about their socio-economic characteristics and a travel diary for one of the days they were wearing the GPS device.

Extensive data processing was required to obtain data that could be suitable for cyclists' route choice models, since the GPS traces were collected for all modes of transport. The data processing was carried out according to a four-step procedure (Schüssler and Axhausen, 2009): (i) GPS data cleaning to remove systematic and random errors from the data; (ii) trip and activity identification where the GPS device had been stationary and/or the spatial density of observations had been high for a period of time; (iii) trip segmentation into single-mode trip legs; (iv) mode identification with fuzzy logic rules that were adapted and extended to the travel behaviour of the Danish population (Rasmussen et al., 2015). The data processing method was validated by comparing the resulting trips and modes with the collected travel diaries.

The data processing resulted in the identification of bicycle trips that were mapped to a high-resolution bicycle network using the map-matching algorithm developed by Nielsen and Jørgensen (2004). Map-matching led to identify trips with the exclusion of cases where no GPS traces could be map-matched because of activities scattered around that were wrongly classified as trips or large deviations (over 20%) between the distances calculated according to the map-matching and the ones calculated from the

processed GPS traces. The characteristics of the identified bicycle trips were defined according to the attributes of the bicycle network and then route choice models were formulated and estimated on the basis of these observed trips.

2.2 Choice set generation

The first stage of modelling route choice behaviour of cyclists was the generation of plausible alternatives to the observed routes. It is well known that model estimates are strongly influenced by choice set size and composition and that biased parameter estimates and choice probabilities are possible consequences of an incorrectly specified choice set (see, e.g., Bekhor et al., 2006; Prato and Bekhor, 2007; Bliemer and Bovy, 2008; Rasmussen et al., 2017). As enumerating all the paths in a highly-detailed network is unrealistic, Halldórsdóttir et al. (2014) tested three choice set generation methods suitable to the task of generating relevant alternatives: (i) breadth first search on link elimination (BSF-LE) (Rieser-Schüssler et al., 2012); (ii) a doubly stochastic generation function (DSGF) (Nielsen, 2000; Bovy and Fiorenzo-Catalano, 2007); (iii) a branch & bound algorithm (B&B) (Hoogendoorn-Lanser et al., 2006; Prato and Bekhor, 2006). As detailed by Halldórsdóttir et al. (2014), the tests focused on different multiattribute cost functions that considered not only route length or time, but also bicyclespecific factors such as road types, bicycle infrastructure types, and land-use designations. The tests showed a significant increase in performances of the choice set generation methods when extending the cost functions to factors other than travel time and specific to the bicycle context.

This study built on those tests by generating the alternatives to the observed routes via the best performing method, namely the DSGF that accounts for variations in travellers' link cost and differences in travellers' attribute preferences by drawing random costs and random parameters from probability distributions. Moreover, this study built on those tests by searching for the best specification of the cost function that would maximise the coverage of the observed routes (Ramming, 2002) in the high-resolution network, namely maximise the number of observed routes that were reproduced at least once at the 80% overlap level in the generation process (see, e.g., Ramming, 2002; Prato and Bekhor, 2007; Rasmussen et al., 2017).

The cost function used for the DSGF in this study extended the one presented by Halldórsdóttir et al. (2014) in that it included information regarding different bicycle path types (i.e., roads without any bicycle infrastructure, roads with bicycle lanes, roads

with segregated bicycle paths, bicycle paths in own trace, footpaths in own trace, steps), surface type (i.e., paved, unpaved), cycling the wrong way with respect to the motorised traffic in a one-way street (i.e., yes, no), and land-use designation (i.e., scenic paths, nonscenic paths). It should be noted that, in the choice set generation, scenic paths were alongside coast, lake, wetland, stream, sand, heath, forests, and parks. Accordingly, the link cost function in this study was defined as follows:

 $C_a = \beta_T T_a + \beta_D D_a + \sum_{k=1}^{K} \left(\beta_{P_k} P_{ak} D_a \right) + \sum_{h=1}^{H} \left(\beta_{S_h} S_{ah} D_a \right) + \beta_U U_a D_a + \beta_W W_a D_a + \varepsilon_a$ (1) where T_a is the travel time on link a, D_a is the length of link a, P_{ak} is equal to 1 when link *a* is of path type k (k = 1,...,K) and 0 otherwise, S_{ah} is equal to 1 when link *a* has surface type *h* (h = 1,...,H) and 0 otherwise, U_a is equal to 1 when link *a* has a scenic land-use on its right-hand side, W_a is equal to 1 when link *a* is in the wrong way with respect to the traffic and 0 otherwise, and ε_a is an error term associated with link *a*. It should be noted that Copenhagen has a right-hand side driving rule, which means that bicycle paths are on the right-hand side of the road with respect to the direction of traffic. In the DSGF, the parameters related to each variable (β_T , β_D , β_{P_k} for each path type k, β_{S_h} for each surface type h, β_{U} , and β_{W}) were coefficients to be drawn from random distributions to express the heterogeneity in cyclists' perceptions of the attributes of the routes, and the error term ε_a was a coefficient to be drawn from a random distribution to express the heterogeneity in cyclists' perceptions of the overall cost. The selection of the distributions and their parameters was imposed by an iterative process searching for the highest coverage of the observed routes (see, e.g., Ramming, 2002; Prato and Bekhor, 2007; Rasmussen et al., 2017).

2.3 Model formulation and estimation

The second stage of modelling route choice behaviour of cyclists was the specification of the utility function and the model itself. Given the focus in this study not only on preferences for traditional variables (e.g., distance, turns, hilliness, intersections, motorised road characteristics), but also on preferences for bicycle facilities and land-use designations, a utility function for the alternative routes was defined to exploit the rich amount of information about the cycling environment. Moreover, given the evident relevance of distance in cycling route choices (Menghini et al., 2010; Hood et al., 2011; Broach et al., 2012; Casello & Usyukov, 2014; Zimmermann

et al., 2017), a model formulation was chosen to evaluate the contribution of other factors with respect to the unit of distance.

Accordingly, the simplest specification of the utility U_{njt} associated by cyclist n with the alternative route j in a choice set of J alternatives was defined as:

$$U_{njt} = \boldsymbol{\beta}'_n \boldsymbol{x}_{njt} + \boldsymbol{\delta}' \boldsymbol{z}_{njt} + \varepsilon_{njt}$$
(2)

where the vector \mathbf{x}_{njt} contains a subset of characteristics of alternative route j as perceived by cyclist n in choice situation t, for which the preferences are expressed by a vector $\boldsymbol{\beta}_n$ of random parameters, and the vector \mathbf{z}_{njt} contains a subset of characteristics of alternative route j as perceived by cyclist n in choice situation t, for which the preferences are expressed by a vector $\boldsymbol{\delta}$ of fixed parameters. The error terms ε_{njt} are extreme value distributed and allow capturing the panel effect of repeated observations t for the same cyclist n.

The generalised mixed logit model (Fiebig et al., 2009) separated out the possibility that taste heterogeneity is independent of scale heterogeneity, and that taste heterogeneity is proportional to scale heterogeneity, by defining the random parameters as follows:

$$\boldsymbol{\beta}_{n}^{\prime} = \sigma_{n}\boldsymbol{\beta} + [\gamma + \sigma_{n}(1 - \gamma)]\boldsymbol{\Gamma}\boldsymbol{\nu}_{n}$$

$$\sigma_{n} = exp(-\tau^{2}/2 + \tau\omega_{n})$$
(3)

where $\boldsymbol{\beta}$ is a vector that provides the average values of the preference parameters in vector $\boldsymbol{\beta}_n$, \boldsymbol{v}_n is a vector of random variables that provide the stochastic element of the preference parameters, $\boldsymbol{\Gamma}$ is a lower triangular matrix that provides the standard deviations and covariances of $\boldsymbol{\beta}_n$, γ is a weighting parameter that expresses how the variance of residual taste heterogeneity varies with scale ($0 \le \gamma \le 1$), σ_n is the individual specific standard deviation of the idiosyncratic error term, w_n is the individual specific taste preference resulting in taste heterogeneity, and τ is the coefficient on the unobserved scale heterogeneity. It was assumed without loss of generality that $var[\boldsymbol{v}_n] = \boldsymbol{I}$, so that $var[\boldsymbol{\beta}_n] = \boldsymbol{\Gamma}\boldsymbol{\Gamma}'$ (Fiebig et al., 2009, Hensher and Greene, 2011), and that $\boldsymbol{\beta}_n$ is unchanged for the same cyclist *n* between different choice situations *t*.

Given the interest in estimating a model in VoD space, it was assumed that the utility is separable in distance d_{njt} (with preference parameter λ_n) and other non-distance characteristics:

$$U_{njt} = \lambda_n d_{njt} + \boldsymbol{\beta}'_n \boldsymbol{x}_{njt} + \boldsymbol{\delta}' \boldsymbol{z}_{njt} + \varepsilon_{njt}$$
(4)

Moreover, given the similarity problem in route choice, it was assumed that similarity across alternative routes is captured by a path size term (Ben-Akiva and Bierlaire, 1999; Prato, 2009):

$$ps_j = -\sum_{a \in \Phi_j} \left(\frac{d_a}{d_j} ln \sum_{j \in C_n} \delta_{aj} \right)$$
(5)

where d_a is the length of link a, d_j is the length of route j, Φ_j is the set of links composing route j, and δ_{aj} is the link-route incidence dummy (equal to one if link a is part of Φ_j and zero otherwise). It should be noted that a path size term p_{Snjt} was then calculated for each choice situation t of each cyclist n.

Accordingly, it was possible to specify the utility in eq. (4) in VoD space so that parameters could be obtained as direct estimates of the marginal rates of substitution between the parameters of the observed attributes and the parameter of the distance:

$$U_{njt} = \lambda_n [d_{njt} + (1/\lambda_n) \boldsymbol{\beta}'_n \boldsymbol{x}_{njt} + (1/\lambda_n) \boldsymbol{\delta}' \boldsymbol{z}_{njt}] + \beta_{ps} p \boldsymbol{s}_{njt} + \varepsilon_{njt} = \lambda_n [d_{njt} + \boldsymbol{\theta}'_n \boldsymbol{x}_{njt} + \boldsymbol{\varphi}' \boldsymbol{z}_{njt}] + \beta_{ps} p \boldsymbol{s}_{njt} + \varepsilon_{njt}$$
(6)

As suggested for the estimation in WTP space (Thiene and Scarpa, 2009; Hensher and Greene, 2011), the model in VoD space was obtained by setting $\gamma = 0$, defining the parameter $\lambda_n = \lambda_d \exp (\lambda_0 + \tau w_n)$ where the coefficient λ_d for the preference for distance is equal to 1, making the parameter λ_n the normalising constant in the VoD space representation, and relaxing the restriction $\lambda_0 = -\tau^2/2$.

Accordingly, the model in VoD space was estimated by using a form of the generalised mixed logit model where the probability *P*_{nit} of cyclist *n* choosing route *i* among the alternative routes *j* in situation *t* is expressed as follows (Hensher and Greene, 2011):

$$P_{nit} = \frac{exp(\lambda_n[d_{nit} + \theta'_n x_{nit} + \varphi' z_{nit}] + \beta_{ps} ps_{nit})}{\sum_{j=1}^{J_{nt}} exp(\lambda_n[d_{njt} + \theta'_n x_{njt} + \varphi' z_{njt}] + \beta_{ps} ps_{njt})}$$
(7)

where the estimated parameter β_{ps} does not belong to the VoD space. The model was estimated in NLogit 5 by simulating the log-likelihood function (Greene and Hensher, 2010):

$$LogL = \sum_{n=1}^{N} log \left\{ \frac{1}{R} \sum_{r=1}^{R} \prod_{t=1}^{T_n} \prod_{i=1}^{J_{nt}} \left[\frac{exp(\lambda_n[d_{nit} + \theta_{nr}x_{nit} + \varphi'z_{nit}] + \beta_{ps}ps_{nit})}{\sum_{j=1}^{J_{nt}} exp(\lambda_n[d_{njt} + \theta_{nr}x_{njt} + \varphi'z_{njt}] + \beta_{ps}ps_{njt})} \right]^{d_{nit}} \right\}$$
(8)

where d_{nit} is equal to 1 if cyclist *n* chooses route *i* in choice situation *t* and 0 otherwise, θ_{nr} are simulated draws of θ_n where the parameters are expressed as a function of draws of σ_{nr} and v_{nr} under the restriction that $\gamma = 0$. It should be noted that there are *N* cyclists and *R* draws are extracted for T_n choice situations for each cyclist *n* and J_{nt} alternatives for each choice situation *t* for each cyclist *n*.

3. Case-study

3.1 Data processing

The acceptance rate from the participants in the TU survey that were invited to wear a GPS device was about 65%, and a total of 318 cyclists from the Copenhagen Region were provided with the devices for an average of eight days. As planned, the 318 cyclists filled a travel diary and answered a socio-economic questionnaire.

The four-step data processing cleaned the GPS traces from systematic errors, identified activities and trips, and segmented the trips by single modes. At this third step, a total of 6,378,651 GPS points and 14,557 single-mode stages were identified. After mode identification, a total of 2,681,108 GPS points and 5,027 bicycle trips were retained for further analysis. Then, the map-matching procedure mapped these trips to the high-resolution bicycle network prepared for this study from the combination of various sources: the two topographic networks TOP10DK (Kort & Matrikelstyrelsen, 2001) and FOT-kort10 (FOT-Kort10, 2010), the network of the Danish National Transport Model (Rich et al., 2010), the OpenStreetMap network (OSM, 2015), and the NAVTEQ network (NAVTEQ, 2010). The compilation of the sources allowed to create a high-resolution geographic network of roads and paths used by cyclists, comprising 361,053 directional links and 268,762 nodes for the study area as illustrated in figure 1. Additional details on the bicycle network are provided by Halldórsdóttir et al. (2014) and Halldórsdóttir (2015).

The map-matching procedure allowed identifying 3,443 trips from 291 cyclists for further analysis. The large number of removed trips was related to the requirement of having high confidence for matching the trips to the network, and Rasmussen et al. (2015) showed that the confidence increased from 69% to 91% by using more restrictive rules that, in turn, identified with certainty a lower number of trips. The richness of the bicycle network allowed calculating the attributes for the routes: network attributes (e.g., distance, number of turns, number and type of intersections, gradient, distance cycling in the wrong way); road attributes (e.g., road type, number of traffic lanes, speed limits, bridges, tunnels); bicycle attributes (e.g., bicycle

infrastructure type, bridge availability, surface type); land-use attributes (e.g., high residential, low residential, industrial, sports, park, forest).



Figure 1: Bicycle network of the Copenhagen Region, with overview of the region (left) and detail of the Copenhagen city centre (right)

Data from the Danish Meteorological Institute (DMI) were joined with the bicycle trips in order to grasp whether differences in route preferences existed according to different weather attributes (i.e., temperature, sunshine, wind, precipitation).

The sample of 291 cyclists was almost equally divided in gender (45.0% males and 55.0% females) and quite widespread distributed in age (see figure 2). It would appear that the sample of cyclists did not suffer from self-selection bias as it compared reasonably well with the sample of cyclists participating in the TU survey (which in turn is representative of the Danish population). Commuting trips (31.8%) were identified in the peak hours between 7am and 9am as well as between 3pm and 6pm, while other trips (68.2%) were observed in the off-peak hours. The majority of the trips (80.6%) were in weekdays rather than weekends, and an even greater majority (85.7%) were in daylight rather than darkness. Additional details on the sample are provided by Halldórsdóttir (2015).



Figure 2: Age distribution of the 291 cyclists

3.2 Generated choice sets

The DSGF method with the cost function in eq. (1) was applied by defining the distributions of the beta parameters and the error term ε_a in order to obtain the maximum coverage of the observed routes (see, e.g., Ramming, 2002; Prato and Bekhor, 2007; Rasmussen et al., 2017) and evaluating graphically the stochasticity of the generated alternatives.

The maximum coverage, calculated as the percentage of observations for which the best generated route overlaps at least a certain threshold with the observed route (Ramming, 2002), was obtained with the beta parameters being log-normally distributed and the error term ε_a being gamma distributed with the distribution parameters presented in table 1. It should be noted that the distribution of the error term ε_a was chosen to be gamma also to allow for the error terms in the link to be additive over the route (for a discussion, see Prato, 2009).

For model estimation purposes, observations where the observed route was reproduced at least at the 80% overlap threshold were considered consistent with the observed behaviour (e.g., Ramming, 2002; Prato and Bekhor, 2007) and the chosen routes were added to the choice set when they were not reproduced at the 100% overlap threshold. The maximum coverage with the presented distribution parameters for the DSGF method replicated over 80% of the observed routes with an 80% coverage threshold and generated between 1 and 100 alternatives to the observed route (mean = 65.6, st. dev. = 40.2). The 59 observations without alternatives to the observed one were discarded and, accordingly, the number of bicycle trips considered for model estimation was 3384 for the 291 cyclists.

Parameter	Variable	Distribution	Mean	Variance
β_{DS}	Length	No distribution	1	0
eta_{TT}	Free time	Log-normal	1	0.25
ε	Error term	Gamma	0*	2
β_{PT_1}	Road with no bicycle infrastructure	Log-normal	1.25	1.5625
β_{PT_2}	Road with bicycle lane	Log-normal	0.75	0.5625
$eta_{{ extsf{PT}_3}}$	Road with segregated bicycle path	Log-normal	0.5	0.25
$eta_{{ extsf{PT}_4}}$	Bicycle path in own trace	Log-normal	0.5	0.25
$eta_{{\scriptscriptstyle PT}_5}$	Footpath in own trace	Log-normal	1.5	2.25
$eta_{{ extsf{PT}_6}}$	Steps	Log-normal	1.5	2.25
β_{ST_1}	Paved surface	Log-normal	0.75	0.5625
β_{ST_2}	Unpaved surface	Log-normal	1.25	1.5625
β_{LU_1}	Scenic land-use (e.g., park, forest, coast)	Log-normal	0.5	0.25
β_{LU2}	Non-scenic land-use	Log-normal	1.5	2.25
β_{WW}	Cycling the wrong way in one-way streets	Log-normal	1.5	2.25

Table 1: Predefined cost parameters, plus the added scale parameters

Note: * The input is set to zero in the program in order to enable the program to centre the gamma distribution to the mean cost obtained with equation 1.

Table 2 presents means and standard deviations of the network and land-use attributes for the chosen and the alternative routes. The chosen routes appear on average shorter and with fewer turns, lower gradient, and fewer intersections. It should be noted that almost all the variables are defined in relation to the distance (e.g., distance on a bicycle lane, distance alongside a road with two lanes, distance alongside a high residential area) because of the intention to estimate the model in VoD space. The network variables were defined according to the bicycle infrastructure on the road, while the land-use designations were divided into three categories according to being on the left-hand side, the right-hand side, or both sides of the road.

		Chosen routes		Alternative routes	
Variable	Unit	mean	st. dev.	mean	st. dev.
Network attributes					
Distance	[km]	3.759	4.659	5.374	5.496
Wrong way	[km]	0.070	0.261	0.092	0.252
Direction					
Left turns	-	2.993	4.053	6.436	6.427
Right turns	-	2.711	2.953	4.901	3.914

Table 2: Variable description of the network and land-use attributes for the chosen and alternative routes

Bicycle specific attributes					
Bicycle infrastructure type					
Motorised road without bicycle facilities	[km]	0.927	1.363	1.326	1.722
Motorised road with bicycle lane	[km]	0.150	0.374	0.209	0.450
Motorised road with segregated bicycle path	[km]	2.326	3.627	2.643	3.620
Bicycle path in own trace	[km]	0.311	0.958	0.955	2.243
Footpath in own trace	[km]	0.045	0.193	0.235	0.583
Steps	[km]	0.001	0.008	0.005	0.026
Bicycle facility type					
Bicycle bridge	[km]	0.002	0.025	0.005	0.038
Motorised traffic bridge crossing water	[km]	0.015	0.068	0.030	0.099
Motorised traffic tunnel	[km]	0.000	0.010	0.000	0.011
Cumulative elevation agin	<u>L</u>				
0-10 meters/km	[km]	0.004	0.006	0.006	0.006
10-35 meters/km	[km]	0.009	0.013	0.013	0.016
35-50 meters/km	[km]	0.002	0.003	0.003	0.005
Above 50 meters/km	[km]	0.003	0.007	0.007	0.013
Surface tune		0.005	0.007		0.015
Paved	[km]	3 658	4.540	1.021	5 007
Cobblestone	[km]	0.006	0.077	0.027	0.217
Uppaved	[lum]	0.000	0.077	0.037	0.217
Number of intersections		0.090	0.439	0.400	1.157
Cive were	[no]	0 6 1 7	1 7/2	0.756	1 000
Give way	[110]	0.017	1.745	0.750	1.908
Roundabout	[110]	0.516	2.095	0.741	2.400
Materiad network attributes	נווס]	7.915	11.794	9.029	11./55
Motorised network attributes					
Motorised roud type	[]rm]	1047	2 200	2027	2142
Large motorised roads	[KIII] []]	1.047	3.390	2.027	3.14Z
Medium motorised roads	[KIII] []]	0.603	1.198	0.000	1.125
	[KM]	0.006	0.091	0.008	0.080
	[KM]	1.266	1./45	2.604	3.294
	[кт]	0.037	0.187	0.050	0.215
Motorisea free speed		0 0 0 0	0.044	0.050	1 000
Below 11 km/h	[km]	0.308	0.866	0.870	1.889
11-30 km/h	[km]	0.293	0.634	0.641	1.040
31-50 km/h	[km]	2.478	3.021	3.030	3.047
51-70 km/h	[km]	0.613	1.729	0.736	1.654
71-90 km/h	[km]	0.067	0.769	0.096	0.695
91-100 km/h	[km]	0.000	0.000	0.001	0.029
Over 100 km/h	[km]	0.000	0.000	0.000	0.016
Number of motorised traffic lanes					
1 lane	[km]	0.010	0.073	0.009	0.063
2 lanes	[km]	2.937	3.631	3.810	3.713
3 lanes	[km]	0.045	0.159	0.043	0.149
4 lanes	[km]	0.262	1.172	0.339	1.114
5 lanes	[km]	0.047	0.209	0.065	0.242
6 lanes	[km]	0.086	0.426	0.138	0.535
7 lanes	[km]	0.000	0.000	0.000	0.000
8 lanes	[km]	0.007	0.054	0.008	0.057

Land-use designations					
High residential area on the right side	[km]	0.497	0.804	0.722	1.023
High residential area on the left side	[km]	0.496	0.809	0.723	1.025
High residential area on both sides	[km]	1.413	1.927	1.726	1.866
Low residential area on the right side	[km]	0.509	0.981	0.730	1.214
Low residential area on the left side	[km]	0.469	0.904	0.681	1.153
Low residential area on both sides	[km]	0.787	1.596	1.067	1.709
Industry on the right side	[km]	0.251	0.588	0.312	0.624
Industry on the left side	[km]	0.248	0.585	0.306	0.612
Industry on both sides	[km]	0.164	0.516	0.222	0.555
Technical on the right side	[km]	0.173	0.446	0.209	0.420
Technical on the left side	[km]	0.166	0.415	0.209	0.418
Technical on both sides	[km]	0.104	0.283	0.150	0.338
Park on the right side	[km]	0.333	0.618	0.555	0.900
Park on the left side	[km]	0.366	0.628	0.607	0.919
Park on both sides	[km]	0.223	0.711	0.636	1.262
Sport on the right side	[km]	0.056	0.205	0.112	0.326
Sport on the left side	[km]	0.054	0.183	0.101	0.297
Sport on both sides	[km]	0.017	0.123	0.021	0.123
Forest on the right side	[km]	0.155	0.529	0.275	0.719
Forest on the left side	[km]	0.150	0.540	0.268	0.696
Forest on both sides	[km]	0.146	0.581	0.355	1.144
Scenic on the right side	[km]	0.185	0.540	0.346	0.794
Scenic on the left side	[km]	0.154	0.473	0.315	0.744
Scenic on both sides	[km]	0.095	0.315	0.168	0.462

4.2 Route choice model estimates

Given the observed routes from the data processing and the alternative routes from the choice set generation, generalised mixed logit models in VoD space were estimated with iterative testing of variable significance for both single attributes and their combinations. The asymptotic t-test was used to verify whether parameter estimates were different from zero at the 95% confidence level, the t-test for estimate differences was employed to decide whether parameter estimates were to be aggregated for various attributes, and correlation values were checked to observe whether parameter estimates were to be excluded because of multi-collinearity. Distributions were tested for the various attributes in the utility function: bounded distributions (e.g., lognormal, constrained triangular) were considered for attributes for which cyclists were expected to have a clear negative preference (e.g., number of turns, elevation gain); unbounded distributions (e.g., normal, triangular) were contemplated for attributes for which different cyclists could have different preferences (e.g., cycling the wrong way, land-use designations).

Table 3 presents the results for the best model specification where the parameter estimates are significantly different from zero at the 95% confidence level while accounting for parameter combinations and attribute interactions. Among the parameter combinations, attributes that did not exhibit significantly different estimates and hence were combined were bicycle lanes and paths, roads with 3 and 4 lanes, roads with 5 lanes or more, industrial and technical areas, as well as park and forest and scenic areas. Taste heterogeneity across cyclists was found significant for number of turns (lognormal distributions), cycling the wrong way (normal distribution), number of roundabouts (normal distribution), cycling on a motorised bridge (normal distribution), cycling alongside sports areas both on one and two sides of the road (normal distribution), cycling alongside scenic areas with medium to high temperature both on one and two sides of the road (normal distribution). Parameters expressing the taste heterogeneity are reported with the mean and the standard deviation of their respective distributions. Notably, a likelihood ratio test between the generalised mixed logit model in VoD scale (LogL = -8,685.44, number of parameters = 46) and the equally specified mixed logit model in preference space (LogL = -8,543.65, number of parameters = 48) shows that the former is to be preferred over the latter (LRT = 283.58, df = 2, p = 0.0000). Moreover, scale heterogeneity is observed as the variance parameter for scale is equal to 0.461 and significant at the 99% confidence level. Lastly, the correction for the similarity across alternative routes has the expected positive sign indicating that utility is reduced for similar routes, and it has a value slightly higher than the 1 that is theoretically expected (Ben-Akiva and Bierlaire, 1999).

It should be noted that table 3 presents the estimates while fixing the estimate of the distance equal to 1 in mean and 0 in standard deviation. This means that the interpretation of the estimates (namely the rates of substitutions) needs to consider that a positive (negative) sign implies that the attribute contributes to a perception of a route being longer (shorter) per unit of distance. For example, cycling in the wrong way has a parameter estimate with a mean of 2.412 which suggests that, on average, the route is perceived 241.2% longer when cyclists choose to ride their bicycle against the motorised traffic. Similarly, cycling on bicycle lanes and paths in the peak hours has a parameter estimate of -0.249 which suggests that the route is perceived 24.9% shorter when cyclists choose to ride their bicycle on lanes and paths in those hours.

Parameter	estimate	st. err.	t-test	р
Network attributes				
Distance (mean)	1.000	-	-	-
Distance (st. dev.)	0.000	-	-	-
Wrong way (mean, normal distribution)	2.412	0.257	9.40	0.000
Wrong way (st. dev., normal distribution)	1.597	0.196	8.14	0.000
Direction				
Straight	-	-	-	-
Left (mean, lognormal distribution)	0.990	0.066	14.89	0.000
Left (st. dev., lognormal distribution)	0.508	0.086	5.93	0.000
Right (mean, lognormal distribution)	1.668	0.070	23.90	0.000
Right (st.dev., lognormal distribution)	0.565	0.101	5.57	0.000
Cumulative elevation gain				
0-10 meters/km	-	-	-	-
10-35 meters/km	1.841	0.764	2.41	0.016
35-50 meters/km	2.091	0.851	2.46	0.014
Above 50 meters/km	4.893	1.777	2.75	0.006
Bicycle infrastructure type				
Motorised road without bicycle facilities	-	-	-	-
Segregated bicycle path / lane - peak hours (mean, normal distribution)	-0.249	0.082	-3.05	0.002
Segregated bicycle path / lane - peak hours (st.dev., normal distribution)	0.096	0.047	2.02	0.042
Segregated bicycle path / lane - off-peak hours	-0.195	0.071	-2.73	0.006
Bicycle path in own trace - peak hours	0.366	0.081	4.53	0.000
Bicycle path in own trace - off-peak hours	0.146	0.044	3.31	0.001
Footpath in own trace	2.167	0.170	12.75	0.000
Steps	10.281	3.137	3.28	0.001
Surface type				
Paved	-	-	-	-
Not paved	0.212	0.071	2.97	0.003
Infrastructure				
Bicycle bridge	-2.879	1.409	-2.04	0.041
Motorised traffic bridge, crossing water/sea (mean, normal distribution)	1.275	0.507	2.51	0.012
Motorised traffic bridge, crossing water/sea (st.dev., normal distribution)	0.430	0.180	2.39	0.017
Number of intersections – yield	0.093	0.041	2.25	0.024
Number of intersections – roundabout (mean, normal distribution)	-0.129	0.040	-3.23	0.001
Number of intersections – roundabout (st.dev., normal distribution)	0.255	0.086	2.96	0.003
Number of intersections – traffic lights	0.208	0.077	2.72	0.007
Number of motorised traffic lanes				
1 lane	-1.390	0.480	-2.90	0.004
2 lanes	-	-	-	-
3 to 4 lanes	0.234	0.064	3.67	0.000
5 or more lanes	0.449	0.101	4.46	0.000
Land-use attributes				
Low residential area (on the right side and both sides)	-	-	-	-
High residential area on the right side - peak hours	0.141	0.070	2.02	0.042
High residential area on the right side - off-peak hours	0.288	0.084	3.45	0.001
Industrial area on the right side	0.235	0.102	2.31	0.026

Table 3: Model estimates of the generalised mixed logit in VoD space

Sports area on the right side (mean, normal distribution)	0.503	0.226	2.22	0.026
Sports area on the right side (st. dev., normal distribution)	0.274	0.134	2.04	0.041
Scenic area on the right side: low temperature	0.183	0.073	2.49	0.013
Scenic area on the right side: medium-high temperature (mean, normal distribution)	-0.330	0.114	-2.89	0.004
Scenic area on the right side: medium-high temperature (st.dev., normal distribution)	0.152	0.068	2.23	0.026
High residential area on both sides: peak hours	0.101	0.043	2.34	0.019
High residential area on both sides: off-peak hours	0.232	0.078	2.98	0.003
Industrial area on both sides	0.353	0.136	2.60	0.009
Sports area on both sides (mean, normal distribution)	-0.369	0.160	-2.31	0.021
Sports area on both sides (st. dev., normal distribution)	0.182	0.076	2.39	0.017
Scenic area on both sides: low temperature	0.478	0.218	2.19	0.029
Scenic area on both sides: medium-high temperature (mean, normal distribution)	-0.383	0.189	-2.03	0.042
Scenic area on both sides: medium-high temperature (st.dev., normal distribution)	0.121	0.058	2.09	0.037
Correction for similarity across routes				
ln(Path-size)	1.238	0.018	67.62	0.000
Parameters of the generalised mixed logit model				
Variance parameter $ au$	0.461	0.062	7.48	0.000
Weighting parameter γ	0.000	-	-	-
Sample mean σ_n	1.000	-	-	-
Sample st.dev. σ_n	0.164	-	-	-
Distance (preference space)	-1.213	0.165	-7.36	0.000
Number of estimated parameters:				48
Number of observations:				3,384
Number of individuals:				291
Null log-likelihood:			-12	2,827.84
Final log-likelihood:			-{	3,543.65
Adjusted rho-square:				0.330

4.2.1 Network attributes

The estimate of the distance in preference space shows that cyclists prefer shorter routes, while the estimates in VoD space indicate that a higher number of left and right turns is related to the perception of longer routes. These results confirm previous findings about both the distance (Menghini et al., 2010; Hood et al., 2011; Broach et al., 2012; Casello & Usyukov, 2014; Zimmermann et al., 2017) and the relation between left and right turns (Hood et al., 2011; Broach et al., 2012). Moreover, these results suggest that a left turn is perceived on average as about a 423m detour and a right turn as about a 221m detour (reasonably, in countries with left-hand side driving the right turn would be expected to be more penalised). Also, taste heterogeneity across cyclists exists and the tail of the distribution suggests that a percentage of cyclists heavily dislike any turn movement.

Interestingly, estimates in VoD space for cycling the wrong way shows that significant heterogeneity emerges: while the mean of the parameter distribution shows

a clear negative preference as the distance in the wrong way is perceived as 2.4 times higher than the one in the right way, the standard deviation of the estimated distribution suggests that about 6.6% of the cyclists have a positive preference for shortcuts involving being on the wrong direction of a one-way street. While the issue of cycling the wrong way was observed by Hood et al. (2011), taste heterogeneity was not looked into. It should be noted that the 6.6% value emerges from the normal distribution having a mean of 2.412 and a standard deviation of 1.597 that gives the probability of a value being positive equal to 93.4%, and the consideration that the ratios are with respect to the negative preference for distance.

The best model specification suggests that going downhill or on a flat terrain do not imply variations in the perceptions of distance, while going uphill implies that cyclists feel as if their distances increase in line with an increase in the gradient. It should be noted that the analysis of slope involved looking into both gains and losses with different categorisations. Interestingly, an average elevation gain over 5.0% implies that the perception of the distance is about 4. 9 times the one if the terrain was flat or downhill. The dislike for steep slopes was qualitatively assessed also in San Francisco (Hood et al., 2011) and Portland (Broach et al., 2012), although it was not quantified. Also, no differences were found between males and females or across age categories, in contrast with findings from San Francisco indicating that females are more sensitive to average uphill slopes than males (Hood et al., 2011).

Model estimates in VoD space suggest that the bicycle infrastructure type has an effect on the perception of distance by cyclists, and specifically bicycle lanes and paths have a positive effect in the feeling of a reduced distance as suggested by previous literature (Menghini et al., 2010; Hood et al., 2011; Broach et al., 2012). Moreover, the best model specification suggests that the benefit is higher in peak hours with respect to off-peak hours, and it is in both cases a moderate one with a perception of decreased distance in the order of about 24.9% and 19.5%. Unlike previous studies, the best model specification indicates that cyclists are not willing to take detours towards bicycle paths in their own trace, and they are even less willing to do so during peak hour. A possible explanation is that the classical Copenhagen configuration (road – curb – bicycle lane/path – curb – sidewalk) does not entail a path in its own trace, but a path integrated within the road network. Another possible explanation is that some paths in their own trace are not paved, another element that contributes to cyclists perceiving

their routes as longer. Lastly, a peculiarity of this study consists in looking at dedicated cycling bridges that provide a significant relief in the perception of distance and indicate that the policy of the municipality of Copenhagen of investing in dedicated bridges is absolutely embraced by cyclists.

Model estimates in VoD space indicate that cyclists in the Copenhagen Region are sensitive to the number of intersections, with yielding signs and traffic lights contributing to the perception of longer routes, in line with previous findings in Zurich (Menghini et al., 2010). Interestingly, roundabouts are the only type of intersection with a significant random parameter that shows an interesting positive effect in average, which can only be explained with the fact that cyclists have the right of way with respect to the motorised traffic in roundabouts combined with the fact that drivers have high awareness of cycling traffic and do not violate this right. However, the distribution for the parameter about roundabouts shows that about 30.7% of cyclists perceived instead that their routes are longer because of the presence of this type of intersection. Notably, tests for traffic variables did not find a significant effect for speed limits and traffic volumes, in line with findings from San Francisco (Hood et al., 2011) but in opposition with findings from Portland (Broach et al., 2012). Possibly, bicycle lanes and paths are mainly where roads have high traffic volumes, which makes the effect of the latter already captured in the effect of the former. Also possibly, the number of lanes was a proxy also for the traffic as narrower roads seem to have a positive effect on the perception of lower distances, while larger roads appear to have the opposite effect.

4.2.2 Land-use attributes

The best model specification suggests that cyclists are sensitive to both what is on the right-hand side as well what is around them on both sides in terms of land-use destinations. A low residential area was selected as the category of reference, given that the Copenhagen Region is characterised by this type of area especially in the outskirts of the municipality.

Estimates in VoD space suggest that high residential areas imply the perception of longer distances, in particular when the areas are to the right of the cyclists and even more so during off-peak hours. This finding could be related to the higher number of possible conflicts, since in these areas there are more public transport stops, shopping opportunities, and office premises. The higher penalty for off-peak hours could be associated with the fact that commuting during peak hours compensates for these

negative effects by getting a better perception of distance given a higher focus on direct routes to work.

Estimates in VoD space indicate that industrial areas are not liked by cyclists that clearly are willing to take detours to avoid them in both cases that they are on the right-hand side or both sides of the cyclists. The same does not apply for sports areas, and interestingly a vast majority (96.7%) of the cyclists seem to perceive longer routes when cycling alongside sports areas that are on the right-hand side of the road, while the vast majority (97.9%) of the cyclists appears to have the opposite feeling when the sports areas are on both sides of the road where they cycle.

Scenic areas are preferred by cyclists, but only depending on the air temperature. It should be noted that models were estimated with several categorisation of the air temperature, and the best model specification was obtained when distinguishing low (below 5°C), medium (between 5°C and 15°C) and high (over 15°C) temperature and then combining medium and high temperatures because their parameter estimates were not significantly different. Cyclists clearly perceive routes being shorter when passing through parks, forests, coastal locations, and other scenic areas unless the temperature becomes low and then they are not willing to take detours. Interestingly, air temperature was the only weather attribute that entered the best model specification, although it is plausible that weather attributes such as wind and precipitation affect more the choice of cycling since for example rain would be equally affecting every alternative route.

5. Discussion and conclusions

Given the lack of consideration for non-motorised traffic in transport planning models, this study contributes to expand the knowledge about cyclists' preferences by formulating and estimating a generalised mixed logit model in VoD space that takes into account not only network attributes, but also bicycle specific attributes as well as landuse designations. The behavioural model in this study is a fundamental step towards the specification of a cost function that can be used in a cycling traffic assignment model to be combined with the vehicle and public transport assignment models in the Copenhagen Region.

The model provides indications about the different factors affecting cyclists' route choices as well as quantitative evaluations of their effect on the perception of distance. The estimation of the model in VoD space indicates that cyclists minimise left turns,

right turns, and uphill slopes. It is also evident that cyclists perceive routes being shorter when they are able to separate themselves from traffic, avoid large roads, and possibly use the bridges that the municipality is heavily investing in. More interestingly, it is apparent that cyclists select their routes in low residential areas and within sports areas or scenic areas (provided that it is not too cold). The value of the estimation of the model in VoD space not only indicate the preference direction, but also quantify the effects of the various factors via the estimation of the rates of substitution that are essential to support planners in making informed decisions. Although the values for the Copenhagen Region might not be transferable "as is" to other regions, the order of magnitude appears sensible and further research in different regions appears valuable for planners to make informed decisions. Recent studies in other areas have shown the value of calculating rates of substitution such as the value of congestion (e.g., Liu et al., 2007; Wardman and Ibañez, 2012; Prato et al., 2014) for car drivers or the value of transfer penalties (e.g., Raveau et al., 2011; Anderson et al., 2014; Raveau et al, 2014) for public transport users.

The findings from this study suggest that thinking about new bicycle infrastructure does not imply only reflecting on the type of infrastructure and its separation from the motorised traffic, but entails also pondering on its location and its configuration when it comes to the possibility of choosing a location alongside a land-use destination and the possibility of including a bicycle bridge. Cyclists seem to consider shorter routes in scenic areas and across sports areas, which indicates the need for planning not only to cater to cyclists, but also to indulge the population at large that would possibly consider cycling if more dedicated infrastructure allowed them to cycle in pleasant conditions. Hence, the major finding from this study consists in the derivation of a utility function that can be easily translated into a cost function for the evaluation of competing bicycle projects: by knowing which type of infrastructure would be constructed where, the calculation of the distance perceived by the Copenhagen Region cyclists could derive from the estimated parameters in the utility function and could help decision-makers in the evaluation of the effectiveness of the projects in terms of the predicted share of cyclists using the piece of infrastructure. Moreover, although the parameters might not apply "as is" to another context, those values might allow decision-makers in other areas of the world to grasp approximately whether alternative pieces of infrastructure could be more or less attractive.

Avenues for future research emerge from this study. Firstly, the utility function should be used within a cycling traffic assignment that should consider the formulation of volume-delay functions at the very least in the centre of Copenhagen where congestion on bicycle lanes is observed in the morning. Secondly, a procedure for having the utility function from model estimation to mirror the cost function in the traffic assignment model could be considered to ensure consistency between the observed behaviour and the predicted behaviour.

Acknowledgments

The authors gratefully acknowledge the financial support of the Danish Road Directorate to the project "The effect of cycling policies", and the technical support of the Data Model Centre at DTU Transport for the assistance of Morten Aabrink and Svend J. Senstius in managing the data collection and processing, and Søren H. Pedersen and Svend J. Senstius in constructing the bicycle network. Lastly, the authors acknowledge the insightful comments of four reviewers on earlier versions of the manuscript.

References

- Anderson, M.K., Nielsen, O.A., Prato, C.G., 2014. Multimodal route choice models of public transport passengers in the Greater Copenhagen Area. *EURO Journal in Transportation and Logistics*, doi:10.1007/s13676-014-0063-3.
- Aultman-Hall, L., Hall, F.L., & Baetz, B.B. (1997). Analysis of bicycle commuter routes using Geographic Information Systems: Implications for bicycle planning. *Transportation Research Record*, 1578, 102-110.
- Axhausen, K.W., & Smith, R.L. (1986). Bicyclist link evaluation: a stated preference approach. *Transportation Research Record*, 1085, 7-25.
- Bekhor, S., Ben-Akiva, M.E., & Ramming, M.S. (2006). Evaluation of choice set generation algorithms for route choice models. *Annals of Operations Research*, 144, 235-247.
- Ben-Akiva, M.E., & Bierlaire, M. (1999). Discrete choice methods and their applications to short-term travel decisions. In Hall, R. (Ed.), *Handbook of Transportation Science*, Kluwer: Dordrecht, The Netherlands, pp. 5-34.
- Bliemer, M.C.J., & Bovy, P.H.L. (2008). Impact of route choice set on route choice probabilities. *Transportation Research Record*, 2076, 10-19.
- Bovy, P.H.L., & Bradley, M.A. (1985). Route choice analyzed with stated preference approaches. *Transportation Research Record*, 1037, 11-20.

- Bovy, P.H.L., & Fiorenzo-Catalano, S. (2007). Stochastic route choice set generation: Behavioral and probabilistic foundations. *Transportmetrica*, 3, 173-189.
- 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, 1730-1740.
- Casello, J., & Usyukov, V. (2014). Modeling cyclists' route choice based on GPS data. *Transportation Research Record*, 2430, 155-161.
- Ehrgott, M., Wang, J. Y., Raith, A., & Van Houtte, C. (2012). A bi-objective cyclist route choice model. *Transportation Research Part A: Policy and Practice*, 46, 652-663.
- Fiebig, D., Keane, M., Louviere, J., & Wasi, N. (2009). The Generalized Multinomial Logit: Accounting for scale and coefficient heterogeneity. *Marketing Science*, 29, 393-421.
- FOT-Kort10 (2010). Information on the new Kort10 and FOT-Kort10 (in Danish). Available at: support.kortforsyningen.dk/attachments/download/292/informationkort10.pdf [Accessed 18 June 2015].
- Greene, W.H., & Hensher, D.A. (2010). Does scale heterogeneity across individuals matter? An empirical assessment of alternative logit models. *Transportation*, 37, 413-428.
- Halldórsdóttir, K. (2015). Behavioural models for cycling Case studies of the Copenhagen Region. Ph.D. thesis, Department of Transport, Technical University of Denmark, Kgs. Lyngby, Denmark.
- 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*, 14, 4, 332-348.
- Hensher, D.A., & Greene, W.H. (2011). Valuation of travel time savings in WTP and preference space in the presence of taste and scale heterogeneity. *Journal of Transport Economics and Policy*, 45, 505-525.
- Hood, J., Sall, E., & Charlton, B. (2011). A GPS-based bicycle route choice model for San Francisco, California. *Transportation Letters: The International Journal of Transport Research*, 3, 63-75.
- Hoogendoorn-Lanser, S., van Nes, R., & Bovy, P.H.L. (2006). A rule-based approach to route choice set generation. Presented at the *11th International Conference on Travel Behaviour Research*, Kyoto, Japan.

- Hopkinson, P., & Wardman, M. (1996). Evaluating the demand for new cycle facilities. *Transport Policy*, 3, 241-249.
- Howard, C., & Burns, E.K. (2001). Cycling to work in Phoenix: Route choice, travel behavior, and commuter characteristics. *Transportation Research* Record, 1773, 39-46.
- Hunt, J.D., & Abraham, J.E. (2007). Influences on bicycle use. *Transportation*, 34, 453-470.
- Hyodo, T., Suzuki, N., & Takahashi, K. (2000). Modeling of bicycle route and destination choice behavior for bicycle road network plan. *Transportation Research Record,* 1705, 70-76.
- Kort & Matrikelstyrelsen (2001). Top10DK Geometric characteristics (in Danish). Kort & Matrikelstyrelsen, Copenhagen, Denmark, Available at:

ftp://ftp.kms.dk/download/pdf/top10dkspec320.pdf [Accessed 08 June 2015].

- Krizek, K.J. (2006). Two approaches to valuing some of bicycle facilities' presumed benefits. *Journal of the American Planning Association*, 72, 309-320.
- Larsen, J., Patterson, Z., & El-Geneidy, A. (2013). Build it. But where? The use of geographic information systems in identifying locations for new cycling infrastructure. *International Journal of Sustainable Transportation*, 7, 299-317.
- Liu, H., He, X., & Recker, W. (2007). Estimation of the time-dependency of values of travel time and its reliability from loop detector data. *Transportation Research Part B: Methodological*, 41, 448-461.

Menghini, G., Carrasco, N., Schüssler, N., & Axhausen, K.W. (2010). Route choice of cyclists in Zurich. *Transportation Research Part A: Policy and Practice*, 44, 754-765.

NAVTEQ (2010). NAVSTREETS Street Data Reference Manual v2.8.

- Nielsen, O.A. (2000). A stochastic transit assignment model considering differences in passengers utility functions. *Transportation Research Part B: Methodological*, 34, 377-402.
- Nielsen, O.A., & Jørgensen, R.M. (2004). Map-matching algorithms for GPS data methodology and test on data from the AKTA road-pricing experiment in Copenhagen. Proceedings of the 10th World Conference on Transport Research, Istanbul, Turkey.
- OSM (2015). © OpenStreetMap contributors. Available at: http://www.openstreetmap.org/ [Accessed 11 August 2015].

- Prato, C.G. (2009). Route choice modeling: past, present and future research directions. *Journal of Choice Modelling*, 2, 65-100
- Prato, C.G., & Bekhor, S. (2006). Applying branch-and-bound technique to route choice set generation. *Transportation Research Record*, 1985, 19-28.
- Prato C.G., & Bekhor, S. (2007). Modeling route choice behavior: How relevant is the composition of choice set? *Transportation Research Record*, 2003, 64-73.
- Prato, C.G., Rasmussen, T.K., & Nielsen, O.A. (2014). Estimating value of congestion and of reliability from the observation of route choice behavior of car drivers. *Transportation Research Record*, 2412, 20-27.
- Pucher, J., & Buehler, R. (2012). City Cycling. The MIT Press, Cambridge, MA.
- Ramming, S. (2002). *Network Knowledge and Route Choice*. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA.
- Rasmussen, T.K, Anderson, M.K., Nielsen, O.A., & Prato, C.G. (2017). Timetable-based simulation method for choice set generation in large-scale public transport networks. *European Journal of Transport and Infrastructure Research*, 17, 1-24.
- Rasmussen T.K., Ingvardson, J.B., Halldórsdóttir, K., & Nielsen, O.A. (2015). Improved methods to deduct trip legs and mode from travel surveys using wearable GPS devices: A case study from the Greater Copenhagen area. *Computers, Environment and Urban Systems*, 54, 301-313.
- Raveau, S., Guo, Z., Muñoz, J.C., & Wilson, N.H.M. (2014). A behavioural comparison of route choice on metro networks: Time, transfers, crowding, topology and sociodemographics. *Transportation Research Part A: Policy and Practice*, 66, 185-195.
- Raveau, S., Muñoz, J.C., & de Grange, L. (2011). A topological route choice model for metro. *Transportation Research Part A: Policy and Practice*, 45, 138-147.
- Rich, J.H., Nielsen, O.A., & Hansen, C.O. (2010). Overall design of the Danish National Transport Model. Proceedings of the *Annual Danish Transport Conference*, Aalborg University, Aalborg, Denmark.
- Rieser-Schüssler, N., Balmer, M., & Axhausen, K.W. (2012). Route choice sets for very high-resolution data. *Transportmetrica*, 9, 825-845.
- Scarpa, R., Thiene, M., & Train, K. (2008). Utility in Willingness to Pay space: A tool to address confounding random scale effects in destination choice to the Alps. *American Journal of Agricultural Economics*, 90, 994-1010.

- Schüssler, N., & Axhausen, K.W. (2009). Processing GPS raw data without additional information. *Transportation Research Record*, 2105, 28-36.
- Sener, I.N., Eluru, N., & Bhat, C.R. (2009). An analysis of bicycle route choice preferences in Texas, US. *Transportation*, 36, 511-539.
- Shafizadeh, K., & Niemeier, D. (1997). Bicycle journey-to-work: Travel behavior characteristics and spatial attributes. *Transportation Research Record*, 1578, 84-90.
- Snizek, B., Sick Nielsen, T. A., & Skov-Petersen, H. (2013). Mapping bicyclists' experiences in Copenhagen. *Journal of Transport Geography*, *30*, 227-233.
- Stinson, M.A., & Bhat, C.R. (2003). Commuter bicyclist route choice: Analysis using a stated preference survey. *Transportation Research Record*, 1828, 107-115.
- Thiene, M., & Scarpa, R. (2009). Deriving and testing efficient estimates of WTP distributions in destination choice models. *Environmental and Resource Economics*, 44, 379-395.
- Tilahun, N.Y., Levinson, D.M., & Krizek, K.J. (2007). Trails, lanes, or traffic: Valuing bicycle facilities with an adaptive stated preference survey. *Transportation Research Part A: Policy and Practice*, 41, 287-301.
- Train, K., & Weeks, M. (2005). Discrete Choice models in preference space and willing to-pay space. In: Scarpa, R., & Alberini, A. (Eds.), *Applications of Simulation Methods in Environmental and Resource Economics*, Springer Publisher, Dordrecht, The Netherlands, pp. 1-16.
- Wardman, M., & Ibañez, J.N. (2012). The congestion multiplier: Variations in motorists' valuations of travel time with traffic conditions. *Transportation Research Part A: Policy and Practice*, 46, 213-225.
- Yeboah, G., & Alvanides, S. (2015). Route choice analysis of urban cycling behaviors using OpenStreetMap: Evidence from a British urban environment. In Arsanjani, J.J., Zipf, A., Mooney, P., & Helbich, M. (eds.), *OpenStreetMap in GIScience*, Springer International Publishing Switzerland, pp. 189-210.
- Zimmermann, M., Mai, T., & Frejinger, E. (2017). Bike route choice modelling using GPS data without choice sets of paths. *Transportation Research Part C: Emerging Technologies*, 75, 183-196.