Transportation Research Part A 44 (2010) 754–765

Contents lists available at ScienceDirect

Transportation Research Part A

journal homepage: www.elsevier.com/locate/tra

Route choice of cyclists in Zurich

G. Menghini [a,1], N. Carrasco [b,2], N. Schüssler [b,3], K.W. Axhausen [b,*]

[a] Andrau 9, CH-7013 Domat/Ems, Switzerland
[b] IVT, ETH, CH-8093 Zürich, Switzerland

ARTICLE INFO

Article history:
Received 20 September 2009
Received in revised form 16 June 2010
Accepted 24 July 2010

Keywords:
Route choice
Discrete choice modelling
Bicycle lanes
Map matching

ABSTRACT

This paper presents the first route choice model for bicyclists estimated from a large sample of GPS observations and overcomes the limitations inherent in the generally employed stated preference approach. It employs an improved mode detection algorithm for GPS post-processing to determine trips made by bicycle, which are map matched to an enriched street network. The alternatives are generated as a random sample from an exhaustive, but constrained search. Accounting for the similarity between the alternatives with the path-size factor the MNL estimates show that the elasticity with regards to trip length is nearly four times larger than that with respect to the share of bike paths. The elasticity with respect to the product of length and maximum gradient of the route is small. No other variable describing the routes had an impact. The heterogeneity of the cyclists is captured through interaction terms formulated on their average behaviour.

1. Introduction

The encouragement of cycling is a central element in just about all current plans for urban and suburban travel behaviour change. The advantages of cycling are obvious, as it is healthy, energy efficient, quiet and compatible with the urban scale. Well designed, continuous and safe cycling networks are the method of choice to encourage cycling, but their design requires an in depth understanding of the trade-offs bicyclists make in their route choice: gradients versus length, traffic lights versus roundabouts, traffic volume versus speeds, bicycle lanes versus direct connections. Next to better design guidance this improved understanding would also improve the generalised cost estimates needed in mode choice modelling, where the current representation of cycling is rudimentary at best.

Our current understanding of these trade-offs relies more or less completely on stated preference (SP) surveys, as previous revealed preference (RP) studies were with one exception (Aultman-Hall et al., 1997) not able to trace or did not try to trace the routes, which the cyclists travelled between origin and destination. In the absence of RP-based route choice models, our knowledge is not as soundly based as one would hope. The recent availability of on-going and long-duration GPS-based observation of travellers has changed the situation fundamentally. It is now possible to trace the route choice of travellers in great detail across all modes with lightweight, unobtrusive and cheap devices over multiple days (e.g. Wolf, 2000; Stopher, 2008). For large samples this comes at the price of the (automatic) processing of the GPS points and especially with the difficulty of identifying the modes used (e.g. Tsui and Shalaby, 2006; Schüssler and Axhausen, 2009a,b). For small samples it is

* Corresponding author. Tel.: +41 44 633 39 43; fax: +41 44 633 10 57.
E-mail addresses: gianluca@menghini.biz (G. Menghini), nelsonc@ethz.ch (N. Carrasco), schuessler@ethz.ch (N. Schüssler), axhausen@ethz.ch (K.W.

[1] Tel.: +41 79 564 80 90.
[2] Tel.: +41 44 633 30 87; fax: +41 44 633 10 57.
[3] Tel.: +41 44 633 30 85; fax: +41 44 633 10 57.

0965-8564/$ - see front matter
doi:10.1016/j.tra.2010.07.008

possible to post-process the observations by hand and with the support of the respondents, but given the costs involved, this is out of reach for large samples of respondents and trips.

The purpose of this paper is to report the first route choice model of bicyclists estimated on the basis of a very large sample of GPS-observed person days. Its methods will help to close an important gap in the literature. It will highlight the difficulties involved in creating the choice sets for estimation, will report the discrete choice estimation results for its example data set and will conclude with a discussion of policy implications of the results, but also of their possible biases.

The literature identifies a large set of attributes of the route and the cyclist as relevant for their choices, but it highlights especially length/travel time, gradient, existence of cycle lanes, type of intersections, presence of parking, traffic volume and age and cycling experience among the characteristics of the cyclists. A recent paper by Sener et al. (2008) provides an exhaustive review of the English language literature since the mid 1970s up to 2007 covering both revealed and stated preference studies (e.g. Axhausen and Smith, 1986; Bovy and Stern, 1990; Shafizadeh and Niemeier, 1997; Stinson and Bhat, 2003; Hunt and Abraham, 2007; Hyodo et al., 2000, but see Tilahun et al., 2007), which they cite only in an earlier version. None of the studies they cite is employing comparable data making comparisons and definite conclusions difficult. There is no German or French literature to speak of since the early work of Leutzbach et al., 1986, which focused on rural bicycle paths.

Aultman-Hall et al. (1997) geo-coded the cycle-stage[4] of 397 commute trips reported by residents in Guelph, Ontario. They compared the chosen paths against the shortest paths, but due to then missing algorithms were then not able to develop a route choice model. The closest study and data set to the one presented here is Harvey and Krizek (2007), which instrumented a small sample of volunteer cyclists recruited from relevant organisations for three weeks in Minneapolis. The about 50 participants undertook nearly 1000 cycle trips during that period. The authors highlight that the participants did not always chose the shortest route, but their analysis remains descriptive. No formal choice models were estimated.

The literature on route choice of car drivers and public transport users cannot be reviewed here, but see Prato (2009) for a recent overview. Basically, the same problem applies: many SP – based studies and few RP results. Given the larger policy interest here, there are earlier RP-based studies, but they have remained rare until the recent availability of GPS data. See Ben-Akiva et al. (1984) for the first RP-based car route choice model.

The remainder of the paper is organised as follows. The next section will describe the work involved in obtaining the routes from the available GPS observations. As the routes are matched to navigation networks a new approach described in Section 3 must be employed to generate the choice sets. A descriptive analysis of the data is followed by the choice modelling results. The final section provides the outlook and policy implications.

2. Data

Choice modelling requires both observed choices and matching sets of non-chosen alternatives. To construct the set of alternatives a suitable network model is required, which had to be constructed, as none of the locally existing network models or maps contained all relevant attributes. The chosen routes were identified in a large scale GPS-data set made available to us, which included 2435 person weeks tracking 73,493 trips in Zürich (see Schüssler and Axhausen, 2008, 2009a,b).

2.1. Street network

The street network was compiled from four different sources: VECTOR25 landscape model of Switzerland (SwissTopo, 2008), a complete (navigation-quality) digital street network of Canton Zurich (Kanton Zürich, Amt für Raumordnung und Vermessung, 2007), the recommended bike routes of Zurich (Stadt Zürich, Tiefbauamt, 2007), and the built bicycle facilities of Zurich's communal master plan (Stadt Zürich, Tiefbauamt, 2007; Menghini, 2008). It merges the relevant characteristics from each network resulting in a more detailed network, covering especially the marked bicycle routes and gradients, which are a crucial consideration in Zürich, which is situated along the valley of the river Limmat and its neighbouring hill sides. It consists of 24,680 links and 8686 nodes. The gradient of each link was calculated as the ratio of the height difference between the start and end point of each divided by the distance between them using the Digital Terrain Model DTM-AV of Cadastral Survey from the Swiss Federal Office of Topography – Swisstopo. This height model is based on high precision laser scanning from aeroplanes flying at 2000 m. The accuracy corresponds to ±0.5 m 1σ and the density is of about 1 point per 2 m^2.

2.2. GPS data

The choices and routes were extracted from a GPS study which was originally conducted by a private sector company with the aim to explore how often participants pass specific advertising billboards. The study observed a representative sample of 2435 Zürich residents for an average of 6.99 days in 2004. No personal characteristics were made available.

[4] Axhausen (2008) defines a stage (unlinked trip) as the uninterrupted movement with one mode or means of transport. A trip is the sequence of stages between two activities. In general a vehicle trip involves two walk stages and the stage with the vehicle.

Mode had to be determined by the mode detection algorithm of Schüssler and Axhausen (2009a,b). For further analysis, only GPS points for which the most probable mode was a bicycle were used here. At this point it is not possible to quantify the share of false positives and wrong negatives in the processed data, i.e. walk stage identified as cycling or cycling classified as car driving, but we employed optimised parameter sets (see Schüssler and Axhausen, 2008 or Schüssler and Axhausen, 2009a,b). Still, the overall quality of the automatic processing is so high, that we do not assume that this is a problem. The automatic processing identifies the stages of a trip (i.e. unlinked trips in American usage). As with all GPS-data the problem sometimes arises that a stage is split into multiple parts due to interruptions of the trace in urban canyons, signal shadows etc. The processing attempts to link such parts into the whole stage, but it will never capture all of these cases.

The original dataset depicted in Fig. 1 includes bike stages made by persons living in Zurich anywhere in Switzerland. In a further step the data was filtered to include only those stages taking place in the city of Zurich (the silhouette of the city is visible in the figure below). From a total of 1101,421 points and 9047 stages, 320,576 points and 3387 stages were retained.

After the filtering process these were mapped to the street network (in Navteq® format) using the map matching algorithm developed by Schüssler and Axhausen (2009a,b) extending Marchal et al. (2006).

2.3. Results of the map matching

Fig. 2 depicts the results of the map matching procedure. The links used by bicyclists in the sample are displayed in red. Grey links correspond to the rest of the street network.

Some errors were observed in the matching procedure, especially along links where a good point flow was not available, or where a dense scatter of points was present. The errors were minimized by means of the data filtering described previously and by calibrating the parameter values for the matching algorithm (for further discussion see Menghini (2008)).

The processing identified 2498 unique origin–destination pairs. Additionally, person statistics for the entire GPS point sample were generated (see below), as well as the link sequences of the chosen routes.

3. Generation of alternative routes

The alternative (non-chosen) routes for the origin–destination pairs were generated using the infrastructure provided the multi-agent transport simulation toolkit (MATSim) (see www.matsim.org). The cost attribute considered was the length of the link, which is consistent with the assumption that the speed of the cyclists depends in the main on their own choice. For alternatives (see also Park and Rilett, 1997; Ramming, 2002; Van der Zijpp and Fiorenzo-Catalano, 2005; Prato and Bekhor, 2006 or Bovy and Fiorenzo-Catalano, 2006).

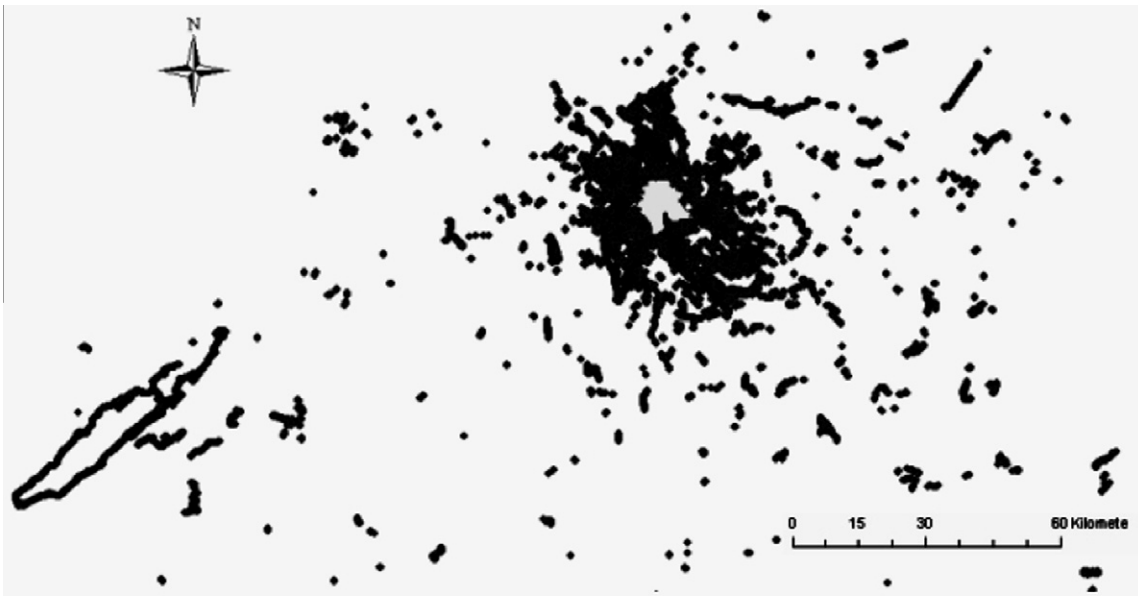

Fig. 1. Original GPS-data set: bicycle stages of persons living in Zurich.

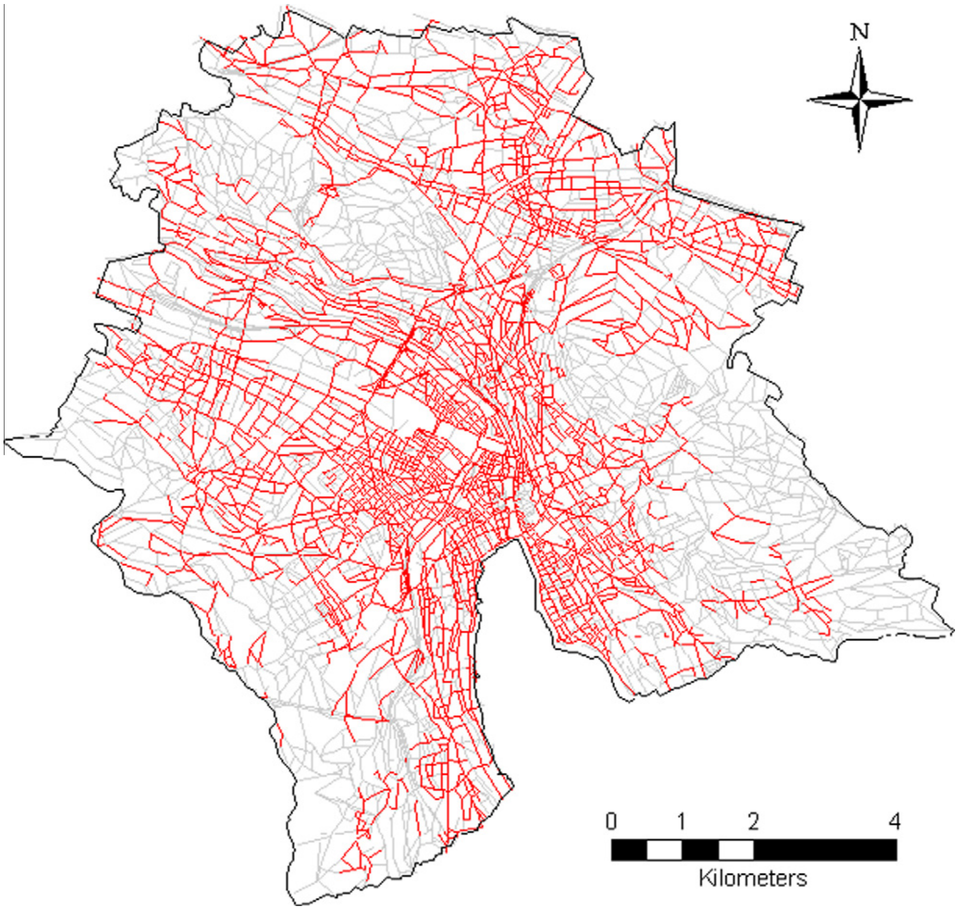

Fig. 2. Results of the map matching procedure (darker links) – City of Zurich.

3.1. Alternative path generation

Schüssler et al. (2009) describe in detail the choice-set generation used. It employs a breadth-first search link elimination approach. It searches for the shortest path between origin and destination and removes the links in turn. These shortest paths become in turn the starting points for the next iteration of link elimination. The algorithm keeps track of the networks generated and retains only unique and connected networks and in turn shortest paths for the choice set. The depth, i.e. number of links removed, is increased until the desired number of distinct routes in the choice set has been generated or the original shortest path is exhausted. Schüssler et al. (2009) discuss various run time experiments and run-time inspired optimisations of the algorithm. Fig. 3 illustrates the algorithm with an example.

3.2. Additional route characteristics

In order to characterise the route and the trade-offs of the cyclists further route characteristics were calculated. Table 1 describes the variables added. It was not possible to add average or even time-specific traffic volumes, as there was no (dynamic) transport model of the required spatial and network resolution available yet.

The *path-size* factor of Ben-Akiva and Bierlaire (1999) was chosen to capture the similarity between the alternatives (for an alternative treatment of the independence of irrelevant alternatives (IIA) property of the multinomial logit model see Cascetta et al., 1996) (see below for further discussion of the IIA property):

$$PS_{in} = \sum_{a \in \Gamma_i} \frac{l_a}{L_i} \frac{1}{\sum_{j \in C_n} \delta_{aj} \frac{L^*_{C_n}}{L_j}}$$

with l_a, length of link a from Route i; L_i, length of route i; Γ_i, set of links of route i; δ_{aj}, indicator with value = 1 if the link is part of route j, otherwise = 0; $L^*_{C_n}$, length of the shortest route for the OD pair n; C_n, total number of routes between OD pair n.

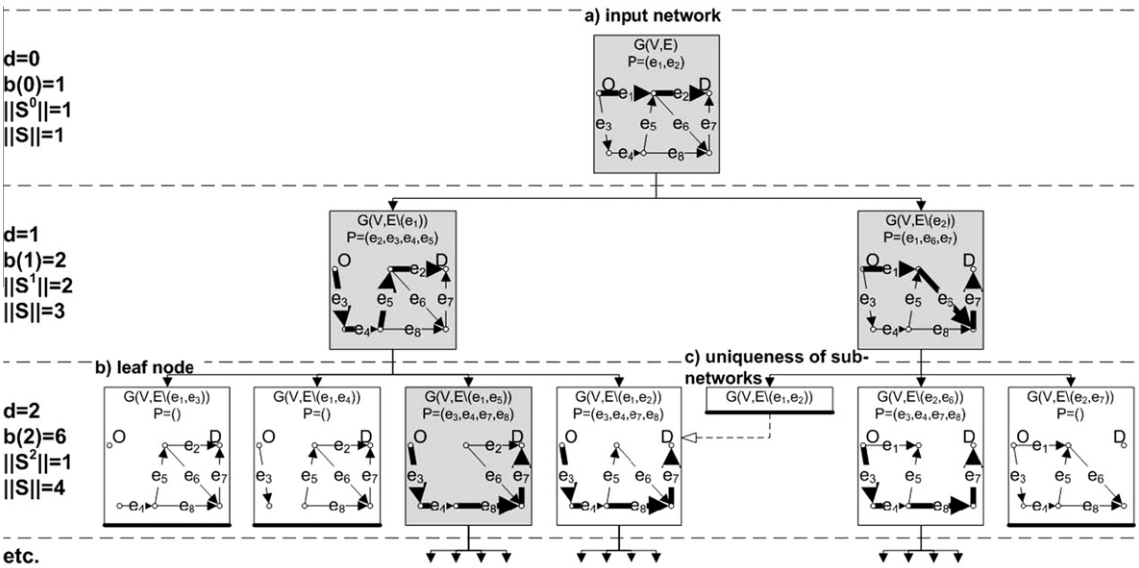

Fig. 3. Illustration of the breadth-first search link elimination algorithm for choice-set generation. *Source*: Schüssler et al. (2009). d: depth; S^n: additional alternatives found at depth n; S: size of the choice set; $b(d)$: Number of candidate networks at depth d;

Table 1
Variables included in the choice set describing the routes.

Variable	Description
Length	Route length [m]
RiseAv	Average absolute gradient [m/100 m]
RiseMax	Maximum gradient [m/100 m]
BikePath	Percentage of marked bike paths along the route [0–100]
TLights	Number of traffic lights
PS	Path size measure [R+]

The reason for excluding average traffic volume along the routes was explained above. The other omission is the degree of parking, but inside the built-up area of Zürich all roads have curb-line parking. The few observed or alternative routes through the wooded hillsides did not justify the data collection effort involved.

4. Descriptive analysis

The chosen routes are in comparison shorter, less steep, involve fewer traffic lights and more marked bike paths, which are not necessarily special facilities, as these could not be identified in the basic maps available (see Table 2). In 35.9% of the cases, the chosen route was also the shortest route.

The distributions are generally similar, but the chosen routes accentuate the skew in comparison with the non-chosen alternative (see the examples in Fig. 4).

The multi-day nature of the data sets allows calculating personal characteristics for each cyclist, such as average speed of the cycling stages. These can be used in-lieu of the unavailable socio-demographics to describe and differentiate the

Table 2
Comparison of the chosen and non-chosen routes.

		Chosen			Alternatives		
Variable	Unit	Mean	Median	St. dev	Mean	Median	St. dev
Route length	[m]	999	742	852	1325	1163	718
Average gradient	[m/100 m]	0.90	0.35	1.38	1.02	0.55	1.18
Maximum gradient	[m/100 m]	3.27	1.45	4.27	5.24	3.14	5.58
Percentage of marked bike paths along the route	[0–100]	78.6	86.1	25.1	68.7	72.0	22.4
Number of traffic lights	[]	2.83	2.00	3.91	2.92	2.00	3.63
Path size measure	[]	0.293	0.160	0.375	0.361	0.271	0.288

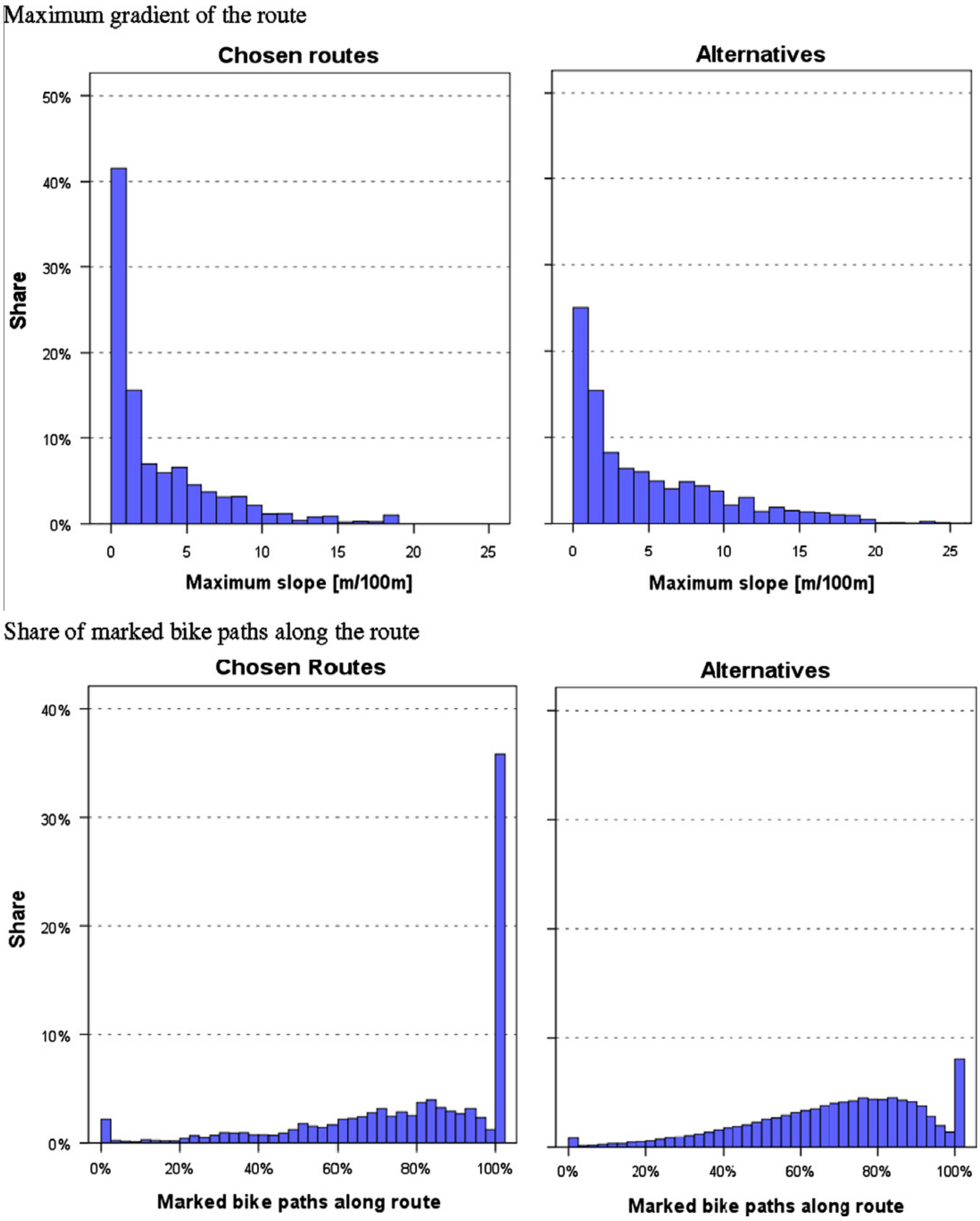

Fig. 4. Distributions of selected characteristics of the chosen and non-chosen routes.

Table 3
Characteristics of the observed cyclists.

Variable	Unit	Mean	Median	St. dev
Number of observed bicycle stages	[]	3.04	2.00	3.95
Average speed by cyclist	[m/s]	2.80	2.76	0.62
Median speed by cyclist	[m/s]	2.80	2.76	0.62
Average stage length	[km]	0.999	0.917	0.515
Median stage length	[km]	0.875	0.788	0.451

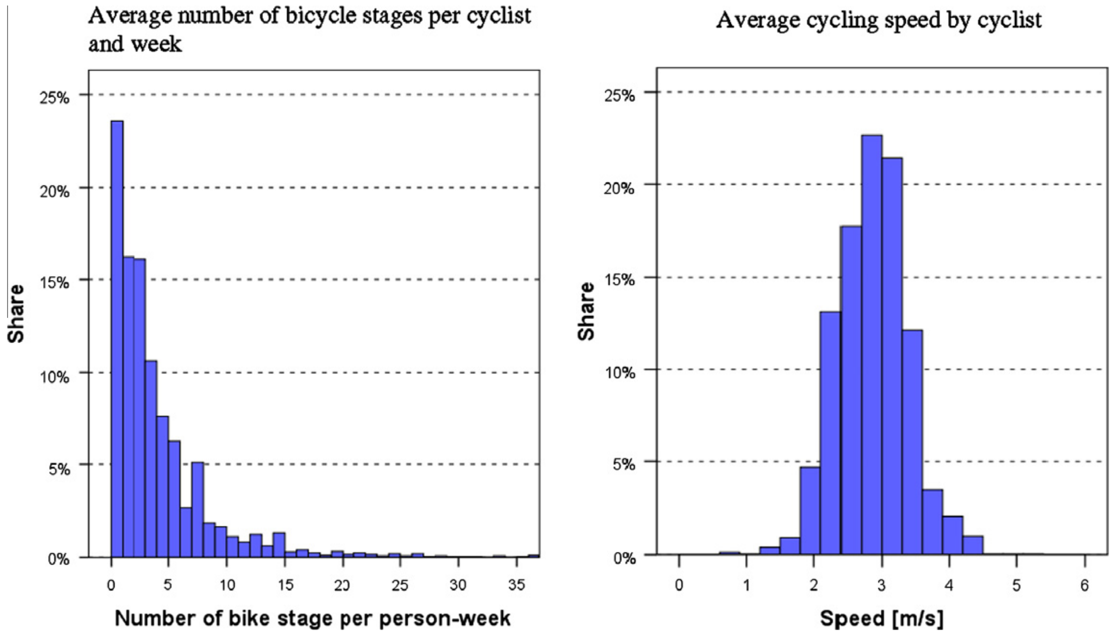

Fig. 5. Distribution of the cyclists' characteristics.

respondents (Table 3). While the average speeds follow a roughly normal distribution (Fig. 5.), the distribution of the number of trips by person are very left skewed, with only a small number undertaking most of their travel in Zürich by bicycle. The range and shape of the Zürich distribution are comparable to the same distributions from two 6-week diary studies shown in Fig. 6 (2003 Thurgau and 1999 Mobidrive) (see Axhausen et al., 2007, 2002).

5. Model estimation

The preferred model for discrete choices is the multinomial logic model (MNL) and its extensions within the GEV model family (Domencich and McFadden, 1975; Ben-Akiva and Lerman, 1985; Train, 2003). The pervasive IIA – problem (independence of irrelevant alternatives) can be accounted for through either explicit models of the error variance–covariance

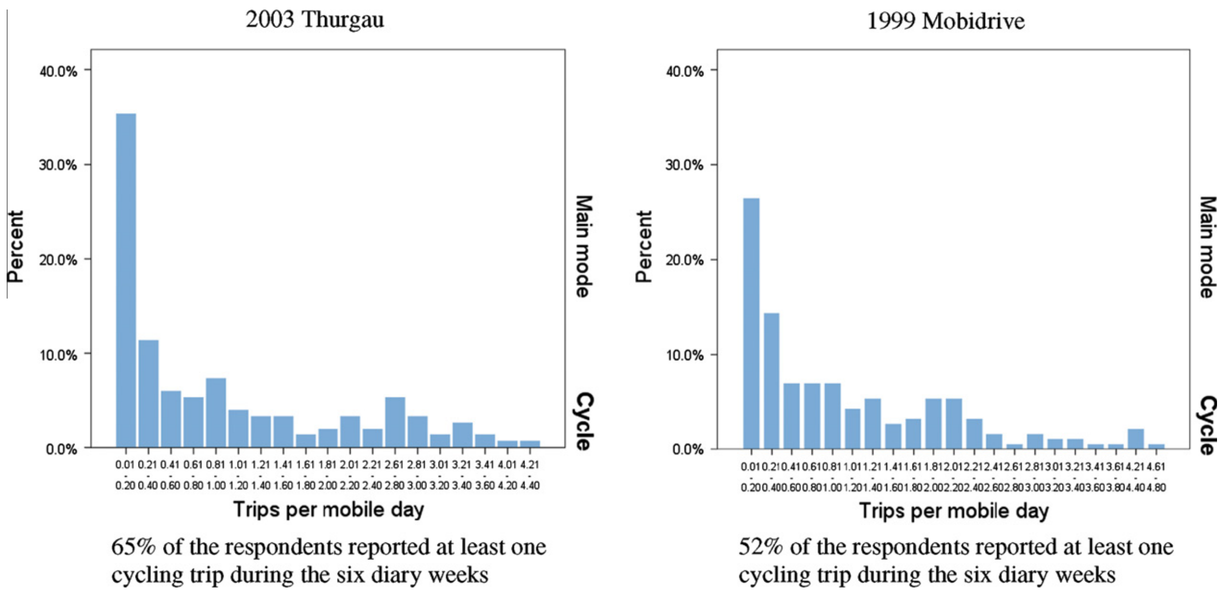

Fig. 6. Distribution of the number of cycle trips per mobile day and cyclist.

structure or through explicit similarity measures, which is the only practicable for route choice with its very large choice sets (for a review see Axhausen and Schüssler (2007)). As mentioned above, the path-size factor of Bierlaire and Ben-Akiva is used here as it is has advantages to its alternatives. Models including the log of the Path Size (*LogPS*) were also estimated, but consistently yielded worse model results. See model estimation results (Tables 5–8).

The models presented here are based on an initial analysis by Menghini (2008), but they are expanded with the derived person characteristics (see Table 3). The models were estimated using the software BIOGEME (Bierlaire, 2003, 2008). The two basic models are shown below in Table 4. Although a logarithmic transformation is recommended by Ben-Akiva and Bierlaire (1999), there was prior evidence that the unconstrained Box-Cox-transformation can improve the model fit. This was adopted here as well. The two models distinguish themselves through the addition of an interaction term between the maximum gradient and the length of the route. The other variables describe the effort and comfort of the route: length, maximum gradient rather then average gradient, number of traffic lights and the share of the length which is sign posted, marked or a built cycle path. These basic models were expanded with additional terms: (b) with a personal speed interaction term, (c) with a personal average length interaction term, (d) with both interacting terms.

The speed and length interactions added to models (b) and (c) respectively are:

$$+\beta_{BikePath} * \left(\frac{average\_speed\_person}{average\_speed\_all\_person}\right)^{\lambda_A vSpeed} * BikePath \quad \text{(b)}$$

$$+\beta_{Length} * \left(\frac{1}{observation\_person}\right) * \left(\frac{Length}{average\_length\_person}\right)^{\lambda\_Path\_Length} \quad \text{(c)}$$

Table 4
Utility functions of the two basic Models (a).

U=

$$\beta_{Length} * Length + \beta_{RiseMax} * RiseMax + \beta_{TLights} * TLights + \beta_{BikePath} * BikePath + \beta_{PS} * \left(\frac{PS^{\lambda} - 1}{\lambda}\right) \quad \text{(1a)}$$

$$\beta_{Length} + \beta_{RiseMax} * RiseMax + \beta_{TLights} * TLights + \beta_{BikePath} * BikePath + \beta_{PS} * \frac{PS^{\lambda} - 1}{\lambda} + \beta_{IRiseLength} * RiseMax * Length \quad \text{(2a)}$$

Table 5
Results for Model 1 (1/2).

	(a) Basic model						(b) With SPEED interaction					
	PS			*LogPS*			*PS*			*LogPS*		
Parameters	6			5			7			6		
Observations	2498			2498			2498			2498		
Final LOG likelihood	−7516.19			−7597.91			−7510.65			−7592.17		
Adjusted ρ^2	0.2647			0.2568			0.2651			0.2572		
	Value	Std. error	*t*-test	Value	Std. error	*t*-test	Value	Std. error	*t*-test	Value	Std. error	*t*-test
Estimated parameters												
β_Length	−0.01	0.0001	−43.10	−0.01	0.0001	−36.73	−0.01	0.0001	−43.17	−0.01	0.0001	−36.85
β_RiseMax	−23.33	1.3079	−17.84	−23.69	1.3261	−17.87	−23.32	1.3077	−17.83	−23.69	1.3261	−17.87
β_Tlights	0.08	0.0151	5.33	0.10	0.0154	6.29	0.08	0.0151	5.32	0.10	0.0154	6.28
β_PS	1.01	0.0712	14.20	–	–	–	1.01	0.0712	14.18	–	–	–
Lambda_PS	1.61	0.1003	16.02	–	–	–	1.61	0.1004	16.02	–	–	–
β_BikePath	2.67	0.1327	20.10	2.70	0.1336	20.20	0.54	0.2742	1.95	0.53	0.2695	1.95
β_LogPS	–	–	–	−0.24	0.1095	−2.23	–	–	–	−0.25	0.1092	−2.25
β_AvSpeed	–	–	–	–	–	–	0.94	0.2928	3.22	0.96	0.2928	3.28
β_PathLength	–		–	–		–	–	–	–	–	–	–

Table 6
Results for Model 1 (2/2).

	(c) With LENGTH interaction						(d) With BOTH interactions					
	PS			*LogPS*			*PS*			*LogPS*		
Parameters	7			6			8			7		
Observations	2498			2498			2498			2498		
Final log likelihood	−7359.94			−7609.48			−7352.59			−7602.62		
Adjusted ρ^2	0.2798			0.2556			0.2805			0.2561		
	Value	Std. error	*t*-test	Value	Std. error	*t*-test	Value	Std. error	*t*-test	Value	Std. error	*t*-test
Estimated parameters												
β_Length	−2.98	0.2523	−11.79	−3.79	0.2855	−13.26	−3.00	0.2523	−11.87	−3.79	0.2843	−13.33
β_RiseMax	−26.32	1.3216	−19.91	−28.01	1.3422	−20.87	−26.29	1.3210	−19.90	−27.99	1.3419	−20.86
β_Tlights	0.09	0.0153	5.86	0.08	0.0152	5.51	0.09	0.0153	5.87	0.08	0.0152	5.51
β_PS	−0.04	0.0076	−5.32	–	–	–	−0.04	0.0077	−5.32	–	–	–
Lambda_PS	−1.89	0.0878	−21.51	–	–	–	−1.89	0.0877	−21.50	–	–	–
β_BikePath	2.69	0.1393	19.32	2.82	0.1382	20.42	0.40	0.2134	1.88	0.52	0.2525	2.05
β_LogPS	–	–	–	−1.15	0.1074	−10.75	–	–	–	−1.16	0.1073	−10.82
β_AvSpeed	–	–	–	–	–	–	1.11	0.3005	3.70	0.99	0.2776	3.58
β_PathLength	2.26	0.1096	20.62	2.20	0.0993	22.18	2.26	0.1090	20.71	2.21	0.0989	22.30

Table 7
Results for Model 2 (1/2).

	(a) Basic model						(b) With SPEED interaction					
	PS			*LogPS*			*PS*			*LogPS*		
Parameters	7			6			8			7		
Observations	2498			2498			2498			2498		
Final log likelihood	−7490.55			−7571.08			−7484.72			−7564.95		
Adjusted ρ^2	0.2671			0.2593			0.2676			0.2598		
	Value	Std. error	*t*-test	Value	Std. error	*t*-test	Value	Std. error	*t*-tesat	Value	Std. error	*t*-test
Estimated parameters												
β_Length	−0.01	0.0001	−44.00	−0.01	0.0001	−37.93	−0.01	0.0001	−44.09	−0.01	0.0001	−38.07
β_RiseMax	−34.75	2.0867	−16.65	−35.30	2.0776	−16.99	−34.82	2.0866	−16.69	−35.38	2.0771	−17.04
β_Tlights	0.08	0.0151	5.54	0.10	0.0154	6.43	0.08	0.0151	5.53	0.10	0.0154	6.43
β_PS	1.00	0.0710	14.10	–	–	–	1.00	0.0710	14.07	–	–	–
Lambda_PS	1.60	0.1012	15.85	–	–	–	1.61	0.1013	15.85	–	–	–
β_BikePath	2.65	0.1328	19.94	2.68	0.1337	20.04	0.50	0.2598	1.93	0.49	0.2526	1.93
β_lRiseLength	0.01	0.0011	7.73	0.01	0.0011	8.02	0.01	0.0011	7.78	0.01	0.0011	8.09
β_LogPS	–	–	–	−0.23	0.1092	−2.12	–	–	–	−0.23	0.1090	−2.15
β_AvSpeed	–	–	–	–	–	–	0.98	0.2956	3.30	1.00	0.2952	3.38
β_PathLength	–	–	–	–	–	–	–	–	–	–	–	–

This formulation was first suggested by Mackie et al. (2003), and had been successfully used since for various Swiss analyses capturing traveller heterogeneity explicitly (see Axhausen et al., 2008 or Hess et al., 2008).

In order to account for the possibility of a person reporting several observations, a weighting term was added in (c). No panel effect was calculated because 68% of the persons contributed only one observation.

As mentioned before, the two models (d) include both interaction terms, and in models (e) the insignificant terms of the previous models were excluded, i.e. the number of traffic signals in Model 1 and again the same and the maximum gradient in Model 2. The results for all the model estimations are shown in Tables 5–8.

Due to the non-linear nature of the estimated models, elasticity values were calculated for the different parameters using the following expression.

$$E_{P_i,X_i} = e_i \cdot \beta_i \cdot x_i^{e_i} \cdot (1 - P_i)$$

Additionally, elasticity was calculated in two different ways for each parameter. The first value resulted from calculating the elasticity for each observation and then averaging across all observations. The second value was calculated by first averaging the variable and probability values, and subsequently calculating the elasticity. Results are shown in Table 9.

Finally, the trade-offs between the variables were calculated and the results shown in Table 10.

The overall model fit is high. The length interaction term improves the model substantially, capturing the heterogeneity of the sample and their differences in destination choice. In both models the formulations with only the significant variables are not or only barely significantly different in terms of the goodness of fit. The introduction of the maximum rise by length

Table 8
Results for Model 2 (2/2).

	(c) With LENGTH interaction						(d) With BOTH interactions					
	PS			*LogPS*			*PS*			*LogPS*		
Parameters	8			7			9			8		
Observations	2498			2498			2498			2498		
Final Log Likelihood	−7359.89			−7608.77			−7352.55			−7601.97		
Adjusted ρ^2	0.2798			0.2555			0.2804			0.2561		
	Value	Std. error	*t*-test	Value	Std. error	*t*-test	Value	Std. error	*t*-test	Value	Std. error	*t*-test
Estimated parameters												
β_Length	−2.97	0.2527	−11.76	−3.77	0.2861	−13.17	−2.99	0.2527	−11.84	−3.77	0.2849	−13.24
β_RiseMax	−25.80	2.0754	−12.43	−26.05	2.1176	−12.30	−25.85	2.0716	−12.48	−26.11	2.1143	−12.35
β_Tlights	0.09	0.0153	5.85	0.08	0.0152	5.47	0.09	0.0153	5.86	0.08	0.0152	5.48
β_PS	−0.04	0.0076	−5.31	–	–	–	−0.04	0.0077	−5.31	–	–	–
Lambda_PS	−1.89	0.0879	−21.49	–	–	–	−1.89	0.0878	−21.49	–	–	–
β_BikePath	2.69	0.1392	19.33	2.82	0.1381	20.45	0.40	0.2139	1.88	0.52	0.2548	2.05
β_IRiseLength	0.00	0.0012	−0.32	0.00	0.0013	−1.19	−0.34	1.2219	−0.28	0.00	0.0013	−1.14
β_LogPS	–	–	–	−1.15	0.1076	−10.64	–	–	–	−1.15	0.1075	−10.72
β_AvSpeed	–	–	–	–	–	–	1.11	0.3005	3.69	0.99	0.2772	3.56
β_PathLength	2.26	0.1098	20.56	2.20	0.1000	21.99	2.26	0.1093	20.65	2.20	0.0995	22.12

Table 9
Elasticities for parameters in combined Models 1 and 2.

	Model (1d) both interactions				Model (2d) both interactions			
Final LL	−7352.59				−7352.55			
Adjusted ρ^2	0.2805				0.2804			
Parameters	Value	*t*-test	Average across observations	Of average values	Value	*t*-test	Average across observations	Of average values
β_length	−3.00	−11.87	−14.025	−5.428	−2.99	−11.84	−13.981	−5.419
β_RiseMax	−26.29	−19.90	−0.741	−0.690	−25.85	−12.48	−0.728	−0.679
β_Tlights	0.09	5.87	0.229	0.205	0.09	5.86	0.228	0.204
β_BikePath	0.40	1.88	0.275	0.275	0.40	1.88	0.275	0.275
β_IRiseLength	–	–	–	–	−0.34	−0.28	−0.013	−0.009

Table 10
Trade offs between different model parameters for final Models 1 and 2.

Model 1(d)		Model 2(d)	
Quotient	Value	Quotient	Value
$\frac{\beta_{length}}{\beta_{BikePath}}$	−7.452 [%/m]	$\frac{\beta_{length}}{\beta_{BikePath}}$	−7.425 [%/m]
$\frac{\beta_{length}}{\beta_{RiseMax}}$	0.114 [1/100 m]	$\frac{\beta_{length}}{\beta_{RiseMax}}$	0.116 [1/100 m]
$\frac{\beta_{Tlight}}{\beta_{RiseMax}}$	−0.003 [m/100 m]	$\frac{\beta_{Tlight}}{\beta_{RiseMax}}$	−0.003 [m/100 m]
		$\frac{\beta_{length}}{\beta_{IRise\_Length}}$	8.835 [m/100 m]
		$\frac{\beta_{BikeAv}}{\beta_{IRise\_Length}}$	−1.190 [m^2/%.100 m]

interaction term (Model 2) improves the fit significantly, and dominates the maximum rise term. Surprisingly, the path size term and its Box-Cox parameter are never significant. We tested it in its logarithmic form and it remained insignificant.

As the mean elasticity shows (Elasticity 1), it is length, which dominates the choices of the Zürich cyclists, but the share of bicycle path has also a substantial, but much smaller impact. The gradient has a strong impact on route choice as well, as shown by the elasticity of the maximum gradient. Still, it is small in comparison with the impact of the length. It would interesting test this again in a city, where the hillside could be detoured around, or even better in the context of destination choice. The interaction term with the average length strengthens the dominance of the path length for those who travel beyond their mean trip length and therefore reduce the deviation from the shortest path. Surprisingly, given the literature, the faster cyclists in Zürich prefer the marked routes, although these tend to follow minor roads.

The two elasticity formulae give strikingly different results in the case of length, which highlights the danger of evaluating elasticities at the mean of the underlying variable.

6. Perspectives and future work

This paper has shown that it is now possible to estimate high quality route choice models, here for cyclists, from GPS data. The effort involved in cleaning the points and in identifying the modes is currently still substantial, but the fast progress in the automation of these processes will soon make such data sets the rule and not the exception (e.g. Czerniak, 2002; Chung and Shalaby, 2005; Zhou and Golledge, 2006; Quddus et al., 2007 or Schüssler and Axhausen, 2009a,b). This will have to be matched by improved network databases, which will have to include a richer set of attributes, especially about the junctions, parking and cycling facilities.

The results underline the importance of direct and marked routes for cyclists, in conjunction with an aversion for steep maximum gradients in conjunction with long routes. The cyclist also avoid signal-controlled junctions. The other factors, as far as we could estimate them, were non-significant, especially the average gradient. The cyclist consider the maximum gradient and not the average gradient, which is reasonable.

The data limitations forced us to forgo testing further possibly relevant characteristics of the street network: average or maximum traffic volume along the route, which we can only add when we have employed an assignment model of the study area with a navigation – level street network; types of intersections, especially the impact of roundabouts, which in the case of the city of Zurich are nearly completely absent, but a very prominent form of intersection control in Swiss suburbia and small towns; the type of bike path, here in particular the effect of marked path versus built and separated paths.

While we are able to capture some of the heterogeneity of the cyclists through interaction terms with their mean trip characteristics, it would be desirable to repeat this work with a data set including the socio-demographics of the respondents, especially age, sex, body-mass index and a measure of risk aversion. The data set should, if the budget can afford this, additionally contain a relatively balanced number of trips per persons, so as to be able to control for the panel nature of the data set.

Finally, the results show again that a policy, which aims to increase the amount and length of cycling, will have to provide direct, preferably marked, paths between origins and destinations of the travellers. Detours are only acceptable, if at all, for short trips.

Acknowledgements

The authors are very grateful for the implementation of the choice set generation algorithm by M. Balmer. Prof. Robert Weibel (Universität Zürich) provided very generous and helpful comments during the first phase of this work. The GPS data was generously provided by mappuls, Luzern. The comments of the anonymous reviewers helped us to improve the paper. All remaining errors are our own.

References

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., 2008. Definition of movement and activity for transport modelling. In: Hensher, D.A., Button, K.J. (Eds.), Handbook of Transport Modelling, second ed. Elsevier, Oxford, pp. 329–344.

Axhausen, K.W., Smith, R.L., 1986. Bicyclist link evaluation: a stated preference approach. Transportation Research Record 1085, 7–25.

Axhausen, K.W., and Schüssler, N., 2007. Similarity, visibility and similar concepts: a note, Arbeitsberichte Verkehrs- und Raumplanung, 458, IVT, ETH Zürich, Zürich.

Axhausen, K.W., Hess, S., König, A., Abay, G., Bates, J.J., Bierlaire, M., 2008. Income and distance elasticities of values of travel time savings: new Swiss results. Transport Policy 15 (3), 173–185.

Axhausen, K.W., Löchl, M., Schlich, R., Buhl, T., Widmer, P., 2007. Fatigue in long duration surveys. Transportation 34 (2), 143–160.

Axhausen, K.W., Zimmermann, A., Schönfelder, S., Rindsfüser, G., Haupt, T., 2002. Observing the rhythms of daily life: a six-week travel diary. Transportation 29 (2), 95–124.

Ben-Akiva, M.E., Bierlaire, M., 1999. Discrete choice methods and their applications to short term travel decisions. In: Hall, W. (Ed.), Handbook of Transportation Science. Kluver, Dordrecht, pp. 5–35.

Ben-Akiva, M.E., Lerman, S.R., 1985. Discrete Choice Analysis: Theory and Application to Travel Demand. MIT Press, Cambridge.

Ben-Akiva, M.E., Bergman, M.J., Daly, A.J., Ramaswamy, R., 1984. Modeling inter-urban route choice behaviour. In: Volmuller, J., Hamerslag, R.E. (Eds.), Proceedings of the 9th International Symposium on Transportation and Traffic Theory. VNU Science Press, Utrecht, pp. 299–330.

Bierlaire, M., 2008. An Introduction to BIOGEME (Version 1.6), EPF Lausanne, Lausanne. <http://transp-or.epfl.ch/page63023.html>.

Bierlaire, M., 2003. BIOGEME: a free package for the estimation of discrete choice models. Presented at the 3rd Swiss Transport Research Conference, Ascona, March 2003.

Bovy, P.H.L., Fiorenzo-Catalano, S., 2006. Stochastic route choice set generation: behavioral and probabilistic foundations. Paper Presented at the 11th International Conference on Travel Behaviour Research, Kyoto, August 2006.

Bovy, P.H.L., Stern, E., 1990. Route Choice and Way Finding in Transport Networks. Kluwer, Dordrecht.

Cascetta, E., Nuzzola, A., Russo, F., Vitetta, A., 1996. A modified logit route choice model overcoming path overlapping problems: specification and some calibration results for interurban networks. In: Lesort, J.B. (Ed.), Proceedings of the 13th International Symposium on Transportation and Traffic Theory. Pergamon, Oxford, pp. 697–711.

Chung, E., Shalaby, A., 2005. A trip reconstruction tool for GPS-based personal travel surveys. Transportation Planning and Technology 28 (5), 381–401.

Czerniak, R.J., 2002. NCHRP Synthesis of Highway Practice 301: Collecting, Processing, and Integrating GPS Data into GIS, NCHRP Report, 301, National Research Council, Washington, DC.

Domencich, T.A., McFadden, D., 1975. Urban Travel Demand: A Behavioural Analysis. North Holland, Amsterdam.

Harvey, F., Krizek, K.J., 2007. Commuter Bicyclist Behavior and Facility Disruption, Research Report, 2007/15, Minnesota Department of Transportation, St. Paul.

Hess, S., Erath, A., Axhausen, K.W., 2008. Estimated value of savings in travel time in Switzerland: analysis of pooled data. Transportation Research Record 2082, 43–55.

Hunt, J.D., Abraham, J.E., 2007. Influences on bicycle use. Transportation 34 (4), 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.

Kanton Zürich, Amt für Raumordnung und Vermessung (ARV), 2007. Geodaten des GIS-ZH, Strassennetz. <http://www.giszh.zh.ch/internet/bd/arv/gis/de/gisdaten/nach_thema/strassennetz.html>.

Leutzbach, W., Buck, A., Axhausen, K.W., 1986. Möglichkeiten und Grenzen der Führung des Radverkehrs auf Radfahrstreifen von anbaufreien Straßen (Bicycle Paths on Rural Roads), Schriftenreihe Straßenbau und Straßenverkehrstechnik, 488, Bundesministerium für Verkehr, Bonn.

Mackie, P., Wardman, M., Fowkes, A.S., Whelan, G., Nellthorp, J., Bates, J.J., 2003. Values of Travel Time Savings in the UK, Institute for Transport Studies. University of Leeds and John Bates Services, Leeds and Abingdon.

Marchal, F., Hackney, J.K., Axhausen, K.W., 2006. Efficient map-matching of large GPS data sets – tests on a speed monitoring experiment in Zurich. Transportation Research Record 1935, 93–100.

Menghini, G., 2008. Routenwahl von Fahrradfahrern: Diskrete Entscheidungsmodellierung mittels GPS-Daten, Masterarbeit. Universität Zürich, Zürich.

Park, D., Rilett, L.R., 1997. Identifying multiple and reasonable paths in transportation networks: a heuristic approach. Transportation Research Record 1607, 31–37.

Prato, C.G., 2009. Route choice modeling: past, present and future research directions. Journal of Choice Modelling 2 (1), 65–100.

Prato, C.G., Bekhor, S., 2006. Applying branch & bound technique to route choice set generation. Transportation Research Record 1985, 19–28.

Quddus, M.A., Ochieng, W.Y., Noland, R.B., 2007. Current map matching algorithms for transport applications: state-of-the art and future research directions. Transportation Research Part C 15 (5), 312–328.

Ramming, M.S., 2002. Network Knowledge and Route Choice, Dissertation, Massachusetts Institute of Technology, Cambridge.

Schüssler, N., Axhausen, K.W., 2009a. Processing GPS raw data without additional information. Transportation Research Record 2105, 28–36.

Schüssler, N., Axhausen, K.W., 2008. Identifying trips and activities and their characteristics from GPS raw data without further information. Paper Presented at the 8th International Conference on Survey Methods in Transport, Annecy, May 2008.

Schüssler, N., Axhausen, K.W., 2009. Map-matching of GPS Points on High-resolution Navigation Networks using Multiple Hypothesis Technique, Arbeitsberichte Verkehrs- und Raumplanung, 568, IVT, ETH Zürich.

Schüssler, N., Balmer, M., Axhausen, K.W., 2009. Route choice sets for very high-resolution data. Paper Submitted to the 89th Annual Meeting of the Transportation Research Board, Washington, DC.

Sener, I.N., Eluru, N., Bhat, C.R., 2008. An analysis of bicycle route choice preferences using a web-based survey to examine bicycle facilities. Working Paper, Department of Civil Engineering, University of Texas, Austin.

Shafizadeh, K., Niemeier, D., 1997. Bicycle journey-to-work: travel behavior characteristics and spatial attributes. Transportation Research Record 1578, 84–90.

Stadt Zürich, Tiefbauamt, 2007. Velorouten – das städtische Veloroutennetz. <http://www.stadt-zuerich.ch/internet/taz/home/mobilitaet/fuss_veloverkehr_/velorouten.html>.

Stinson, M.A., Bhat, C.R., 2003. Commuter bicyclist route choice: analysis using a stated preference survey. Transportation Research Record 1828, 107–115.

Stopher, P.R., 2008. The travel survey toolkit: where to from here? Paper Presented at the 8th International Conference on Survey Methods in Transport, Annecy, May 2008.

Swisstopo, 2008. VECTOR 25 Landschaftsmodell. <http://www.Swisstopo.admin.ch/internet/Swisstopo/de/home/products/landscape/vector25.html>.

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 41 (4), 287–301.

Train, K.E., 2003. Discrete Choice Methods with Simulation. Cambridge University Press, Cambridge.

Tsui, S.Y.A., Shalaby, A.S., 2006. Enhanced system for link and mode identification for personal travel surveys based on global positioning systems. Transportation Research Record 1972, 38–45.

Van der Zijpp, N.J., Fiorenzo-Catalano, S., 2005. Path enumeration by finding the constrained k-shortest paths. Transportation Research Part B 39 (6), 545–563.

Wolf, J., 2000. Using GPS Data Loggers to Replace Travel Diaries in the Collection of Travel Data, dissertation, Georgia Institute of Technology, School of Civil and Environmental Engineering, Atlanta.

Zhou, J., Golledge, R., 2006. A three-step general map matching method in the GIS environment: travel/transportation study perspective. International Journal of Geographical Information System 8 (3), 243–260.