ETH zürich

ETH Library

# Route choice of cyclists in Zurich

## GPS-based discrete motion models

**Working Paper**

**Author(s):**
Menghini, Gianluca; Carrasco, Nelson; Schüssler, Nadine; Axhausen, Kay W.

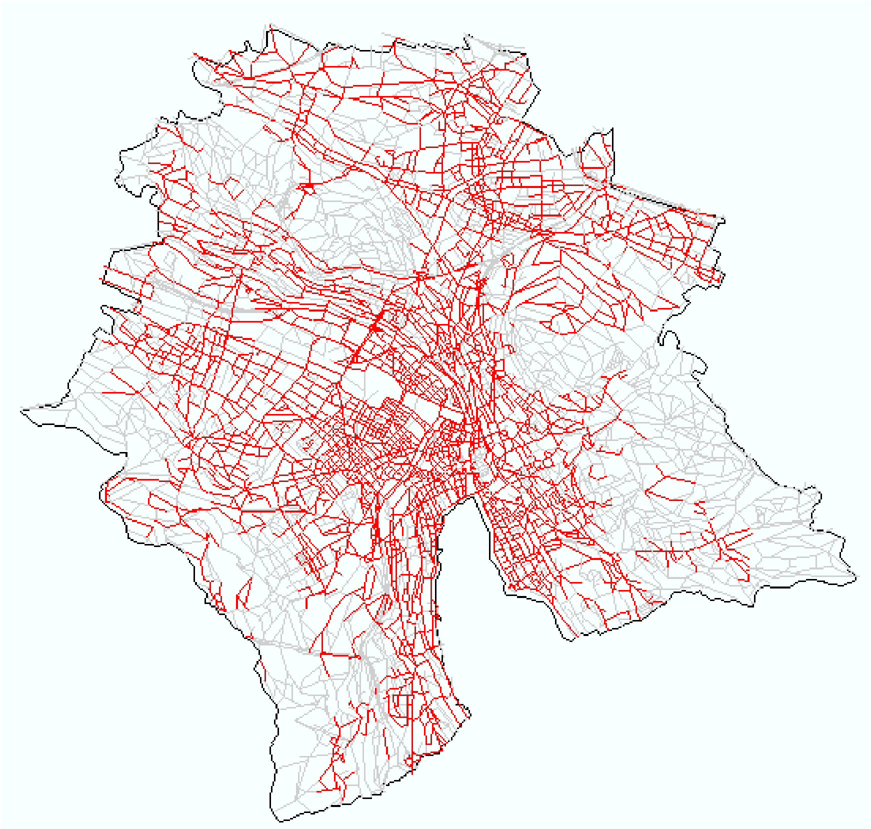
**Publication date:**
2009

**Permanent link:**
https://doi.org/10.3929/ethz-a-005729195

**Rights / license:**
In Copyright - Non-Commercial Use Permitted

**Originally published in:**
Arbeitsberichte Verkehrs- und Raumplanung 544

This page was generated automatically upon download from the ETH Zurich Research Collection.
For more information, please consult the Terms of use.



# Route choice of cyclists in Zurich:
## GPS-based discrete choice models


**G Menghini**
**N Carrasco**
**N Schüssler**
**KW Axhausen**



**Arbeitsberichte Verkehrs- und Raumplanung 544**

**January 2009**



Institut für Verkehrsplanung und Transportsysteme
Institute for Transport Planning and Systems

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Arbeitsberichte Verkehrs- und Raumplanung

# Route choice of cyclists in Zurich

| G Menghini | N Carrasco | N Schüssler | KW Axhausen |
|---|---|---|---|
| Andrau 9 | IVT | IVT | IVT |
| CH - 7013 Domat/Ems | ETH | ETH | ETH |
| | CH – 8093 Zürich | CH – 8093 Zürich | CH – 8093 Zürich |
| +41 79 564 80 90 | +41 44 633 30 87 | +41 44 633 30 85 | +41 44 633 39 43 |
| | Fax:+41 44 633 10 57 | Fax:+41 44 633 10 57 | Fax:+41 44 633 10 57 |
| gianluca@menghini.biz | nelsonc@ethz.ch | schuessler@ethz.ch | axhausen@ethz.ch |

January 2009

## 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 by 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.

### Key Words

Route choice; Discrete choice modelling; Bicycle lanes; Map Matching

### Preferred Citation Style

Menghini, G., N. Carrasco, N. Schüssler and K.W. Axhausen (2009) Route choice of cyclists: discrete choice modelling based on GPS-data, *Arbeitsberichte Verkehrs- und Raumplanung*, **544**, IVT, ETH Zurich, Zurich.

# 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 their current representation for cycling is rudimentary at best.

Our current understanding of these trade-offs relies more or less completely on stated preference surveys, as previous revealed preference studies were not able to trace or did not try to trace the routes, which the cyclists travelled between origin and destination.  In the absence of actual route choice data, 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). For small samples it is 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. It will highlight the difficulties involved in creating the choice sets for estimation, will report the discrete choice estimation results 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, Eluru and Bhat (2008) provides an exhaustive review of the English language literature since the mid 1970's 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, Suzuki and Takahashi, 2000), but see Tilahun, Levinson and Krizek, 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, Buck and Axhausen, 1986, which focused on rural bicycle paths. 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 1'000 cycle trips during that period. The authors highlight that the participant did not always chose the shortest route, but their analysis remains descriptive. No formal choice models were estimated.

# 2 Data

Choice modeling requires both the 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 11'000 trips in Zürich (See Schüssler and Axhausen, 2008 and 2009).

## 2.1 Street network

The street network was compiled from four different sources: VECTOR25 landscape model of Switzerland (SwissTopo, 2008), 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 includes the most relevant characteristics of each network resulting in a more detailed network for the purpose of bicyclists' route choice, 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 the neighbouring hill sides. It consists of 24'680 links and 8'686 nodes.

## 2.2 GPS data

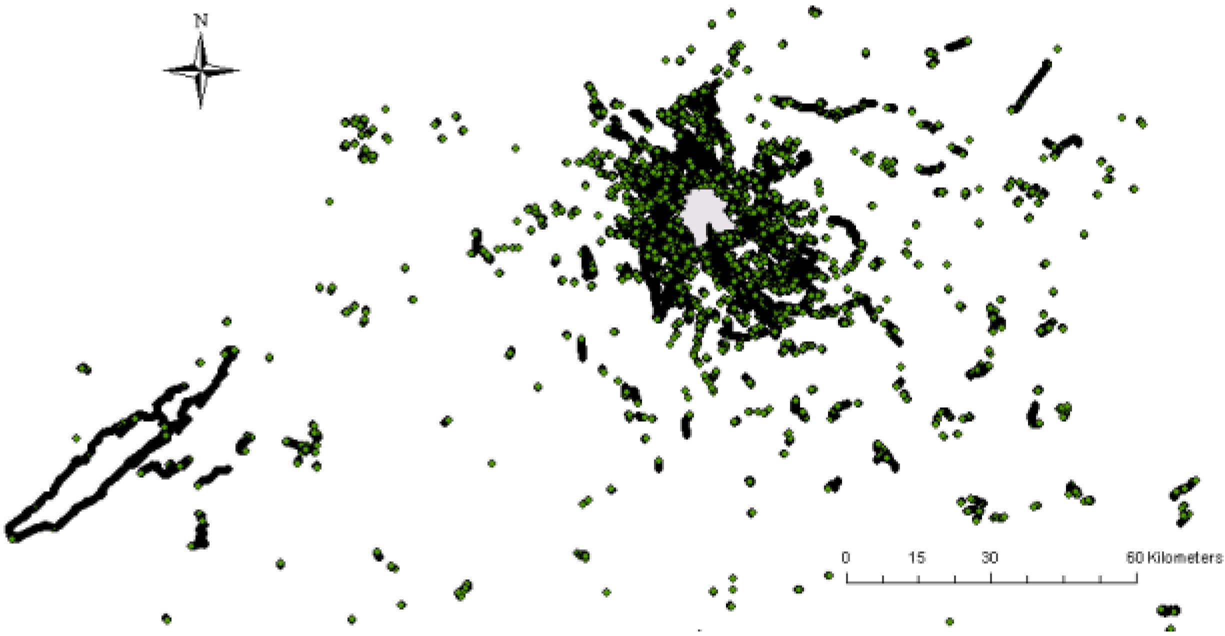
The choices and routes were extracted from a GPS study which was originally conducted by a private sector company with an aim to explore how often participants pass specific publicity billboards. The study observed a representative sample of 2435 Zürich residents for an average of 6.99 days in 2004. No personal characteristics were made available. Mode had to be determined by the mode detection algorithm of Schüssler and Axhausen (2009). 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 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 2009). Still, the overall quality of the automatic processing is so high, that we do not assume that these will be a problem. The automatic processing identifies the stages of a trip (see Axhausen, 2008) (i.e. unlinked trips). As with all GPS-data the problem sometimes arises that the true stage is split in multiple parts due to interruptions of the records 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 Figure 1 includes bike trips made by persons living in Zurich anywhere in Switzerland. In a further step the data was filtered to include only those trips taking place in the city of Zurich (the silhouette of the city is visible in the figure below). From a total of 1’101.421 points and 9’047 stages, 320’576 points and 3’387 stages were retained.

In a second step, a 26 m wide buffer along every street axis was created to include only those points located up to this distance from the street network. Because the map matching procedure assigns every point to a network link, eliminating those points located outside the buffer is an attempt to remove scatter and to reduce the error measures in the process (distance between GPS point and assigned link). Roughly 23% of the original GPS point data set and 36% of the stages (260’845 points and 3’315 stages) were retained.

In a final filtering step, polylines for each individual stage were generated from the GPS points using the Person, Stage and Trip ID of each point. The result can be seen on Figure2 (left). Longer routes are marked in darker colours.

Figure 1 Original GPS data set. Registered bike trips of persons living in Zurich



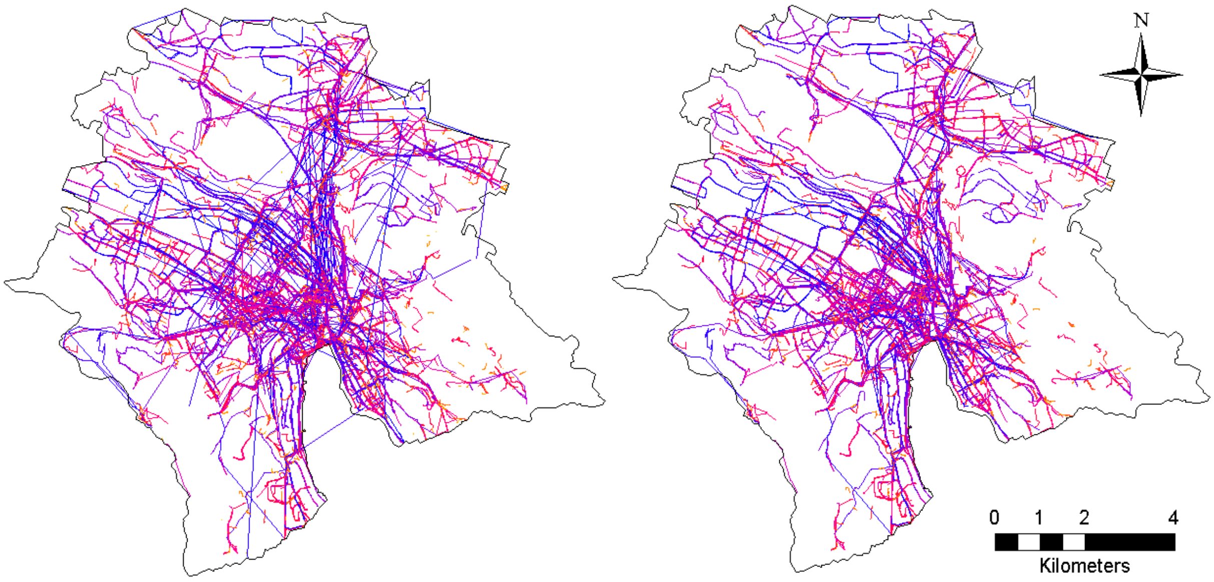

It can be appreciated that in some cases odd jumps occur from one point of the city to another (e.g. a line crossing lake Zurich). These polylines were revised manually, deleting those which displayed random crossings and odd behaviour. Additionally, those polylines with a length of less than 150 m were also deleted. It was assumed the mode detection algorithm erroneously classified these very short trips as bike trips, when most likely they were walking trips.

Figure 4 shows the polylines before (left) and after the cleaning process. This filtering stage left roughly 23% of the original GPS points and 29% of the stages for the map matching process (247’863 points and 2’657 stages). Table 1 summarizes the result of the filtering process. From the original GPS point dataset, 22.5% of the points and 29.4% of the stages were kept for the map matching procedure.

Table 1 Results of the GPS data filtering – Summary

| Filtering stage | Number of points | Number of stages | % left after filtering | |
|---|---|---|---|---|
| | | | Points | Stages |
| Before filtering | 1’101’421 | 9’047 | 100.0 | 100.0 |
| Clip with city limits | 320’576 | 3’387 | 29.1 | 37.4 |
| Clip with the street buffer | 260’845 | 3’315 | 23.7 | 36.6 |
| Polyline removal | 247’863 | 2’657 | 22.5 | 29.4 |

Figure2 Polylines from GPS points: before and after filtering




After the filtering process these were allocated to the street network (in Navteq® format) using the map matching algorithm developed by Marchal, Hackney and Axhausen (2006) employing the postGIS database management software (pgAdmin III) (see www.pgadmin.org). The parameters of the map matching had been calibrated in a series of runs (Table 2).

Table 2 Final parameters of the map matching algorithm

| Parameter | Description | Value |
|---|---|---|

| | | |
|---|---|---|
| orientedLink | Consideration of the link orientation | false |
| keepOnes | Paths with only one link are kept? | false |
| minCruisedRatio | Minimum distance which must be covered to keep a back and forth link | 0.5 |
| minPoints | Minimum number of points | 8 |
| maxDist | Max distance between two points to be considered in the same path | 200 |
| maxTimeGap | Maximum time gap between two points [ms] | 120000 |
| minCandidates | Minimum number of candidates for matching | 8 |
| maxCandidates | Maximum number of candidates for matching | 100 |
| adaptativeFactor | Factor to determine the number of candidates | 0.001 |
| coeffDistance | Coefficient of line distance (difference between two points and the two matched points) in comparison of distance (point to link) | 0.25 |

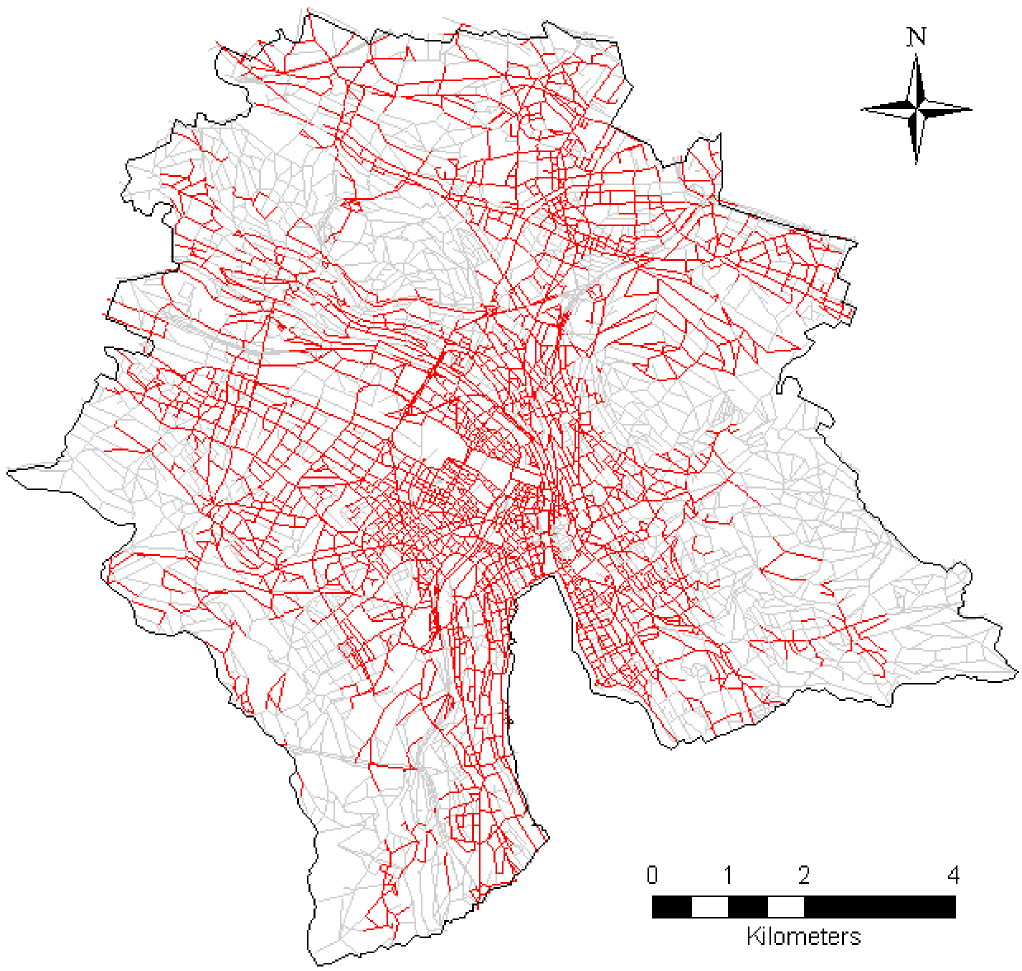
## 2.3 Results of the map matching

Figure 6 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 636 unique origin-destination pairs. Additionally, person statistics for the entire GPS point sample were generated (see below), as well as the link sequence of the chosen routes.

Figure3 Results of the map matching procedure (red links) – City of Zurich



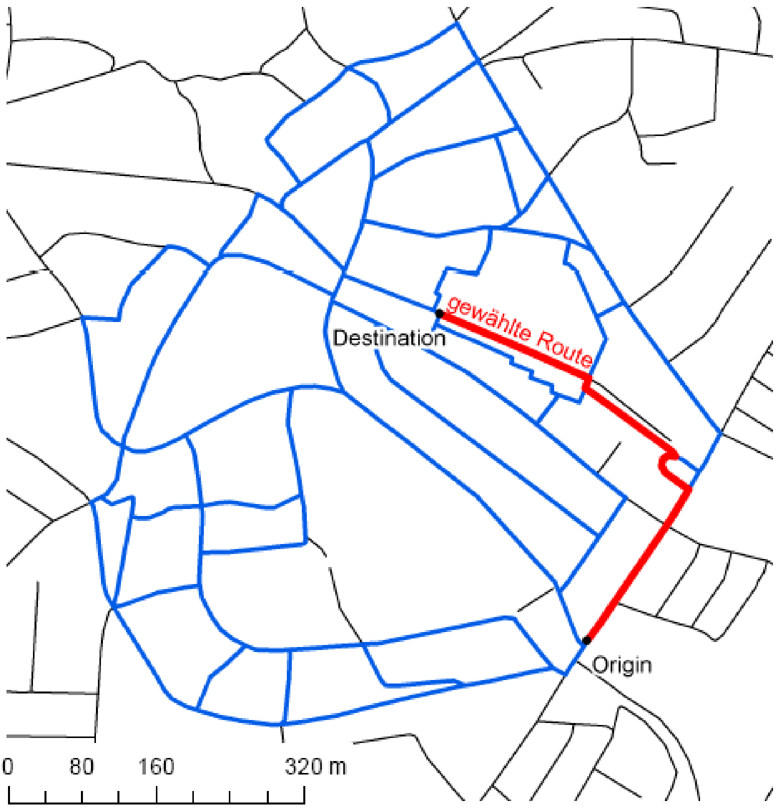

# 3 Generation of alternative routes

The alternative (non-chosen) routes for the origin-destination pairs were generated using 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 a cyclist depends in the main only on his own choices. 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) Applying branch & bound technique to route choice set generation, Transportation Research Record, 1985, 19-28.

## 3.1 Alternative path generation

A broad search algorithm was implemented within MATSim. An exhaustive search was implemented up to a certain detour factor and controlling for overlap by removing up to three links within previously found shortest paths. A total of 60 routes were determined between each of the final OD pairs. A random sample of 20 routes was picked form the 60 and the chosen route added. The broad search exhausted the alternatives, which should ensure that all relevant alternatives are included in the choice set. Figure 4 illustrates the results for a given OD pair.

Figure 4 Example for the alternatives and chosen route for a given OD pair




Source: Menghini (2008); Alternatives in blue; chosen route in red and slightly thicker.

## 3.2 Additional route characteristics

In order to characterise the route and the trade-offs of the cyclists' further route characteristics were calculated using ArcGIS, MATSim and SPSS. Table 3 describes the variables added. It was not possible to add average or even time-specific traffic volumes, as there is no (dynamic) transport model of the required spatial and network resolution.

The *path-size* factor of Bierlaire and Ben-Akiva (1999) was chosen to capture the similarity between the alternatives using BIOROUTE, a utility available with BIOGEME (www.biogeme.epfl.ch) (For an alternative see Cascetta, Nuzzola, Russo and Vitetta, 1996). A modified logit route choice model overcoming path overlapping problems: Specif. The similarity was calculated based on the path length, since travel time (another usual measure of similarity) was assumed to be linearly related to distance:

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

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 effort involved.

Table 3 Variables included in the choice set describing the routes

| Variable | Description |
|---|---|
| Length | Route length [m] |
| RiseAv | Average absolute gradient [m/100m] |
| RiseMax | Maximum gradient [m/100m] |
| BikeAv | Percentage of marked bike paths along the route [0 – 1] |
| TLights | Number of traffic lights [R+] |
| PS | Path Size measure [R+] |

# 4 Descriptive Analysis

The chosen routes are in comparison shorter, less steep, involve fewer traffic lights and more bike marked paths, which are not necessarily special facilities, as these could not be identified in the basic maps available (See Table 4).

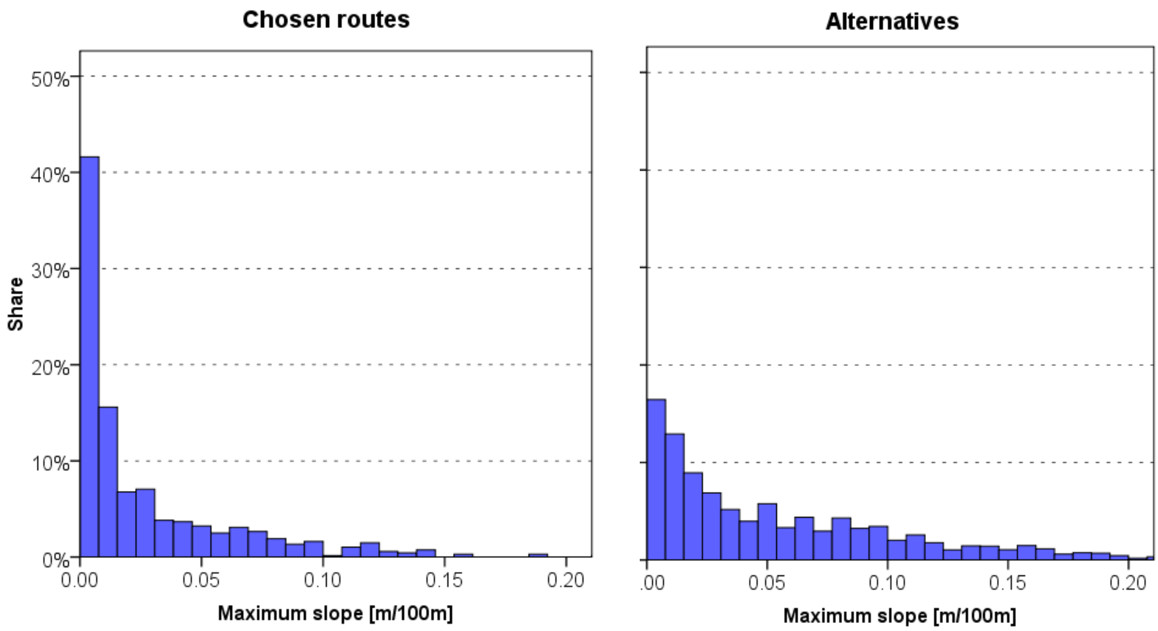
Table 4 Comparison of the chosen and non-chosen routes

| | | Chosen | | | Non-chosen | | |
|---|---|---|---|---|---|---|---|
| Variable | Unit | Mean | Median | St.Dev | Mean | Median | St.Dev |
| Route length | [m] | 586 | 489 | 382 | 1412 | 1299 | 682 |
| Average gradient | [m/100m] | 1.05 | 0.40 | 1.64 | 1.13 | 0.71 | 1.19 |
| Maximum gradient | [m/100m] | 2.70 | 1.07 | 3.61 | 5.62 | 3.83 | 5.62 |
| Percentage of marked bike paths along the route | [0 – 1] | 74.5 | 83.0 | 30.3 | 64.6 | 66.9 | 23.0 |
| Number of traffic lights | [] | 1.05 | 0 | 1.74 | 1.89 | 1 | 2.30 |
| Path Size measure | [] | .0731 | .0381 | .1057 | .0734 | .0713 | .0248 |

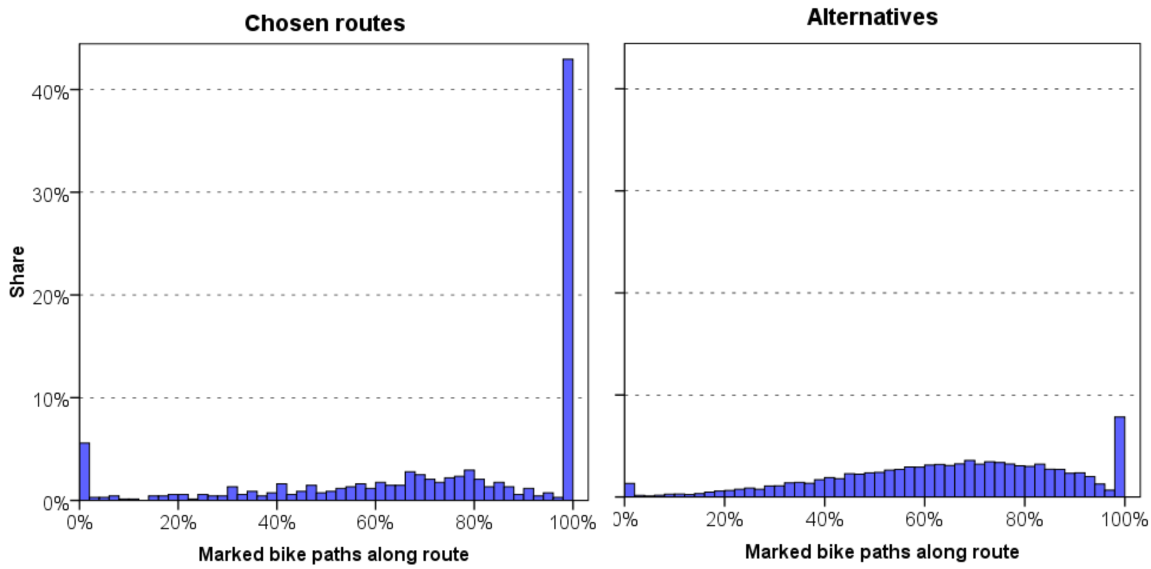
The distributions are generally similar, but the chosen routes accentuate the skew in comparison with the non-chosen alternative (See the examples in Figure5).

Figure5 Distributions of selected characteristics of the chosen and non-chosen routes

Maximum gradient




Percentage of marked bike paths along the route



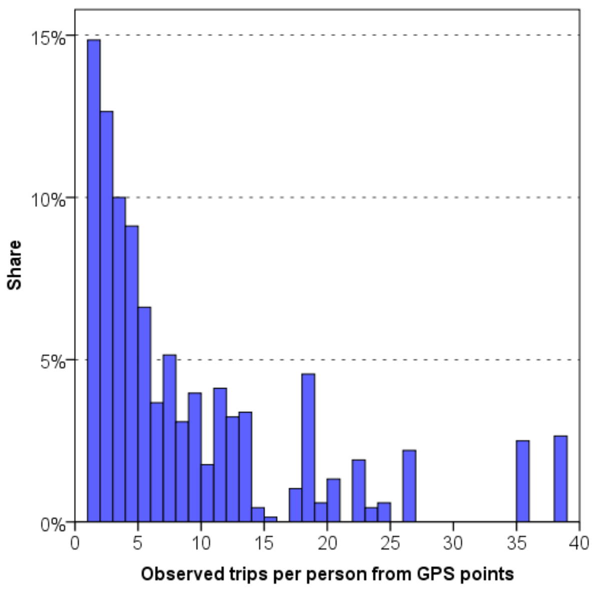

The multi-day nature of the data sets allows calculating personal characteristics for each cyclist, such as average speed of the cycling stages. These can be used in-lieu of the non available socio-demographics to describe and differentiate the respondents (Table 5). While the average speeds follow a roughly normal distribution (Figure 6), 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 (Figure 7) are comparable to the same distributions from two six week diary studies (2003 Thurgau and 1999 Mobidrive) (see Axhausen, Löchl, Schlich, Buhl and Widmer, 2007; Axhausen, Zimmermann, Schönfelder, Rindsfüser and Haupt, 2002).

Table 5 Sample characteristics of the cyclists

| Variable | Unit | Mean | Median | St.Dev |
|---|---|---|---|---|
| Number of observed bicycle trips | [] | 2.31 | 1.00 | 2.50 |
| Average speed by person | [km/h] | 11.78 | 11.76 | 3.73 |
| Median speed by person | [km/h] | 10.89 | 10.90 | 3.33 |
| Average trip length | [m] | 574.3 | 522.6 | 285.6 |
| Median trip length | [m] | 551.8 | 502.6 | 276.2 |

Figure 6 Distribution of the cyclists characteristics

Number of bicycle trips per cyclist and week (9% of the sample cycling)




Average cycling speed by cyclist

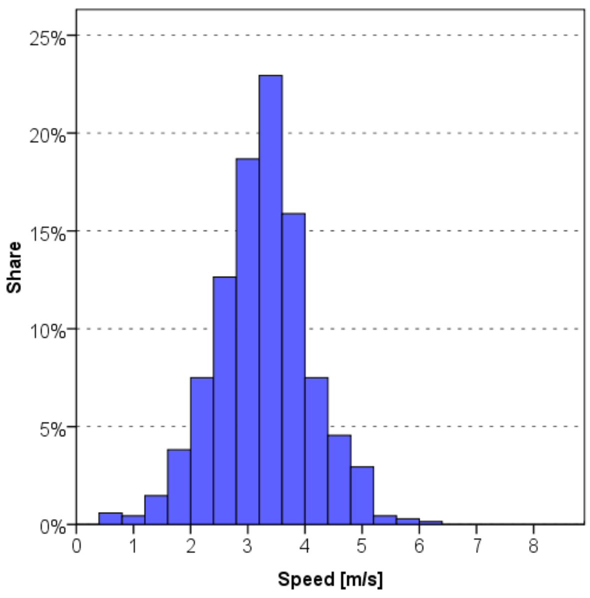
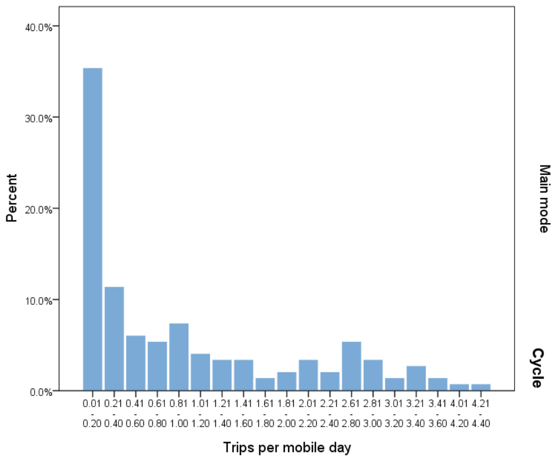
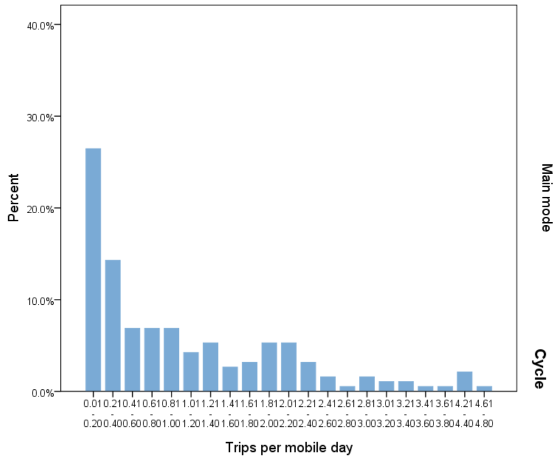


Figure 7 Distribution of the number of cycle trips per mobile day

2003 Thurgau (65% of sample cycling during the six weeks of reporting)




1999 Mobidrive (52% of the sample cycling during the six weeks of reporting)




# 5 Model estimation

The preferred model for discrete choices is the multi-nominal 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 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.

The models presented here are based on an initial analysis by Menghini (2008), but they are expanded both with the derived person characteristics, as well as by an explicit accounting of the repeated observation per person through a suitable random variable. The models were estimated using the software BIOGEME (Bierlaire, 2003 and 2008). The two basic models are shown below in Table 6. Although a logarithmic transformation is recommended by Bierlaire and Ben-Akiva (1999), there was prior evidence that allowing 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 a sign posted, marked or 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. Finally, (e) only significant variables of the previous models were included.

Table 6 Utility functions of the two basic models (a)

| U= | |
|---|---|
| $\beta_{Length} * Length + \beta_{RiseMax} * RiseMax + \beta_{TLights} * TLights + \beta_{BikeAv} * BikeAv + \beta_{PS} * \left(\frac{PS^{\lambda}-1}{\lambda}\right)$ | 1(a) |
| $\beta_{Length} * Length + \beta_{RiseMax} * RiseMax + \beta_{TLights} * TLights +$ $\beta_{BikeAv} * BikeAv + \beta_{PS} * \left(\frac{PS^{\lambda}-1}{\lambda}\right) + \beta_{WRiseLength} * RiseMax * Length$ | 2(a) |

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

$$\beta_{BikeAv} * \left(\frac{average\_speed\_person}{average\_speed\_all\_persons}\right)^{\lambda\_AvSpeed} * BikeAv \quad \text{(b)}$$

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

This formulation was first suggested by Mackie, Wardman, Fowkes, Whelan, Nellthorp and Bates (2003), and had been successfully used since for various Swiss analyses capturing traveller heterogeneity explicitly (see Axhausen, Hess, König, Abay, Bates and Bierlaire (2008) or Hess, Erath, Vrtic and Axhausen (forthcoming)).

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 register 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 Model 1 and the same and the maximum gradient in Model 2. The results are shown in Tables 5 and 6.

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

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

Additionally, elasticity was calculated in two different ways for each parameter. The first value resulted from calculating the value elasticity for each person and the averaging for all persons. The second value was calculated by first averaging the variable and probability values, and subsequently calculating the elasticity. Results are shown in Tables 7 and 8.

Finally, a trade off analysis is carried out and the results shown in Table 9.

Table 7 Results for Model 1

| | a) Basic model | | | b) With speed interaction | | | c) With length interaction | | | d) With both interactions | | | e) Only significant variables | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Number of Parameters | 6 | | | 7 | | | 7 | | | 8 | | | 5 | | |
| Final Log Likelihood | -944.80 | | | -944.02 | | | -899.53 | | | -889.44 | | | -897.84 | | |
| Adjusted Rho Squared | 0.5090 | | | 0.5088 | | | 0.5318 | | | 0.5365 | | | 0.5337 | | |
| ***Estimated Parameters*** | *Value* | *Std. Error* | *t-Test* | *Value* | *Std. Error* | *t-Test* | *Value* | *Std. Error* | *t-Test* | *Value* | *Std. Error* | *t-Test* | *Value* | *Std. Error* | *t-Test* |
| β_length | -0.01 | 0.00 | -20.06 | -0.01 | 0.00 | -20.05 | -0.80 | 0.13 | -6.21 | -0.88 | 0.14 | -6.31 | -0.84 | 0.13 | -6.50 |
| β_RiseMax | -18.60 | 3.46 | -5.37 | -18.66 | 3.48 | -5.37 | -21.26 | 3.64 | -5.85 | -21.97 | 3.72 | -5.90 | -20.31 | 3.55 | -5.72 |
| β_Tlights | -0.06 | 0.05 | -1.21 | -0.06 | 0.05 | -1.21 | -0.05 | 0.06 | -0.84 | -0.04 | 0.06 | -0.80 | - | - | - |
| β_PathSize | -51.29 | 35.95 | -1.43 | -49.65 | 34.79 | -1.43 | 0.00 | 0.00 | -0.35 | -55.72 | 81.74 | -0.68 | - | - | - |
| Lambda_PS | 11.20 | 6.66 | 1.68 | 10.89 | 6.46 | 1.69 | -1.11 | 0.61 | -1.83 | 15.55 | 21.17 | 0.73 | - | - | - |
| β_BikeAv | 1.61 | 0.27 | 5.96 | 1.66 | 0.27 | 6.04 | 1.65 | 0.28 | 5.95 | 1.75 | 0.30 | 5.93 | 1.72 | 0.29 | 5.96 |
| λ_AvSpeed | - | - | - | 0.80 | 0.69 | 1.16 | - | - | - | 1.58 | 0.68 | 2.32 | 1.71 | 0.66 | 2.59 |
| λ_PathLength | - | - | - | - | - | - | 4.45 | 0.24 | 18.45 | 4.36 | 0.24 | 18.16 | 4.45 | 0.24 | 18.92 |

Table 8 Results for Model 2

| | a) Basic model | | | b) With speed interaction | | | c) With length interaction | | | d) With both interactions | | | e) Only significant variables | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Number of Parameters | 7 | | | 8 | | | 8 | | | 9 | | | 5 | | |
| Final Log Likelihood | -944.48 | | | -943.73 | | | -880.04 | | | -878.98 | | | -886.10 | | |
| Adjusted Rho Squared | 0.5086 | | | 0.5085 | | | 0.5414 | | | 0.5414 | | | 0.5398 | | |
| ***Estimated Parameters*** | *Value* | *Std. Error* | *t-Test* | *Value* | *Std. Error* | *t-Test* | *Value* | *Std. Error* | *t-Test* | *Value* | *Std. Error* | *t-Test* | *Value* | *Std. Error* | *t-Test* |
| β_length | -0.01 | 0.00 | -17.55 | -0.01 | 0.00 | -17.55 | -0.75 | 0.14 | -5.58 | -0.75 | 0.13 | -5.62 | -0.72 | 0.12 | -5.79 |
| β_RiseMax | -15.70 | 5.08 | -3.09 | -15.85 | 5.09 | -3.11 | -2.70 | 5.76 | -0.47 | -2.55 | 5.71 | -0.45 | - | - | - |
| β_Tlights | -0.07 | 0.05 | -1.25 | -0.07 | 0.05 | -1.24 | -0.05 | 0.06 | -0.87 | -0.05 | 0.05 | -0.82 | - | - | - |
| β_PathSize | -50.97 | 36.07 | -1.41 | -49.32 | 34.87 | -1.41 | -446.0 | 206.86 | -2.16 | -45.93 | 55.81 | -0.82 | - | - | - |
| Lambda_PS | 11.25 | 6.75 | 1.67 | 10.93 | 6.54 | 1.67 | 100 | 1.8E+308 | 0.00 | 14.23 | 15.49 | 0.92 | - | - | - |
| β_BikeAv | 1.61 | 0.27 | 5.96 | 1.66 | 0.27 | 6.04 | 1.68 | 0.29 | 5.91 | 1.72 | 0.29 | 5.84 | 1.68 | 0.29 | 5.85 |
| β_RiseLength | 0.00 | 0.00 | -0.77 | 0.00 | 0.00 | -0.75 | -0.02 | 0.01 | -4.13 | -0.02 | 0.01 | -4.09 | -0.02 | 0.00 | -6.49 |
| λ_AvSpeed | - | - | - | 0.79 | 0.69 | 1.14 | - | - | - | 1.48 | 0.69 | 2.14 | 1.61 | 0.68 | 2.38 |
| λ_PathLength | - | - | - | - | - | - | 4.35 | 0.27 | 16.41 | 4.36 | 0.26 | 16.50 | 4.43 | 0.26 | 17.09 |

Table 9 Elasticity values for parameters in Model 2

| | d) With both interactions | | | | e) Only significant variables | | | |
|---|---|---|---|---|---|---|---|---|
| Final LL | -878.98 | | | | -886.10 | | | |
| Adjusted $\rho^2$ | 0.5414 | | | | 0.5398 | | | |
| **Parameters** | *Value* | *t-Test* | *Elasticity 1* | *Elasticity 2* | *Value* | *t-Test* | *Elasticity 1* | *Elasticity 2* |
| β_length | -0.75 | -5.62 | -4.012 | -0.197 | -0.72 | -5.79 | -4.184 | -0.185 |
| β_RiseMax | -2.55 | -0.45 | -0.049 | -0.046 | - | - | - | - |
| β_Tlights | -0.05 | -0.82 | -0.035 | -0.033 | - | - | - | - |
| β_BikeAv | 1.72 | 5.84 | 1.162 | 1.124 | 1.72 | 5.83 | 1.219 | 1.160 |
| β_RiseLength | -0.02 | -4.09 | -0.0003 | -0.0002 | -0.02 | -6.49 | -0.0003 | -0.0002 |

Table10 Trade off between different model parameters for final models

| Model 2(d) | | Model 2(e) | |
|---|---|---|---|
| Quotient | Value | Quotient | Value |
| $\frac{\beta_{length}}{\beta_{BikeAv}}$ | -0.437 | $\frac{\beta_{length}}{\beta_{BikeAv}}$ | -0.429 |
| $\frac{\beta_{length}}{\beta_{Rise\_Length}}$ | 34.13 | $\frac{\beta_{length}}{\beta_{Rise\_Length}}$ | 30.86 |
| $\frac{\beta_{BikeAv}}{\beta_{Rise\_Length}}$ | -78.03 | $\frac{\beta_{BikeAv}}{\beta_{Rise\_Length}}$ | -72.01 |

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 are the formulations with only the significant variables not or only barely significantly different in terms of the goodness of fit. The introduction of the maximum rise by length interaction term (model 2) improves the fit significantly, and dominates the maximum rise term. Surprisingly, the path size term and it 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 hardly any impact on route choice, but this would need to be tested again in a city, where the hill side could be detoured around, or 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 they 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 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, Ochieng and Noland, 2007 or Schüssler and Axhausen, 2009). This will have to be matched by improved network data bases, 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 other factors, as far as we could estimate them, were non-significant, especially the number of traffic lights or the average gradient. The results will be integrated as a first, but good estimate of the generalised costs in the agent-based micro-simulation MATSim to improve our assignment and our model of daily scheduling.

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 MATSim 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.

# 7 Acknowledgements

The authors are very grateful for the implementation of the choice generation algorithm in MATSim 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.

# 8 Literature

Axhausen, K.W. (2008) Definition of movement and activity for transport modelling, in D.A. Hensher and K.J. Button (eds.) *Handbook of Transport Modelling*, 2nd edition, 329-344, Elsevier, Oxford.

Axhausen, K.W. und R.L. Smith (1986) Bicyclist link evaluation: A stated preference approach, *Transportation Research Record*, **1085**, 7-25.

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

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

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

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

Ben-Akiva, M.E. und M. Bierlaire (1999) Discrete choice methods and their applications to short term travel decisions, in W. Hall (ed.) *Handbook of Transportation Science*, 5-35, Kluver, Dordrecht.

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

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. und S. Fiorenzo-Catalano (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. und E. Stern (1990) *Route Choice and Way Finding in Transport Networks*, Kluwer, Dordrecht.

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

Chung, E. und A. Shalaby (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 D.C.

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

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

Hess, S., A. Erath, M. Vrtic and K.W. Axhausen (forthcoming) Joint valuation of travel time savings estimation on four separate Swiss data sets, *Transportation Research Record*

Hunt, J.D. und J.E. Abraham (2007) Influences on biycle use, *Transportation*, **34** (4) 453-470.

Hyodo, T., N. Suzuki and K. Takahashi (2000) Modeling of bicyle 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., A. Buck and K.W. Axhausen (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., M. Wardman, A. S. Fowkes, G. Whelan, J. Nellthorp and J. J. Bates (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., J.K. Hackney und K.W. Axhausen (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. and L.R. Rilett (1997) Identifying multiple and reasonable paths in transportation networks: a heuristic approach, *Transportation Research Record*, **1607**, 31-27.

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

Quddus, M.A., W.Y. Ochieng und R.B. Noland (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. and K.W. Axhausen (2009) Processing GPS raw data without additional information, paper presented at the *88th Annual Meeting of the Transportation Research Board*, Washington, D.C., January 2009

Schüssler, N. und K.W. Axhausen (2008) Identifying trips and activities and their characteristics from GPS raw data without further information, Vortrag, *8th International Conference on Survey Methods in Transport*, Annecy, May 2008.

Schüssler, N. and K. W. Axhausen (2007) Recent developments regarding similarities in transport modeling, presented at the *6th Swiss Transport Research Conference*, Ascona, September 2007.

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

Shafizadeh, K. and D. Niemeier (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. and C.R. Bhat (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., D.M. Levinson und K.J. Krizek (2007) Trails, lanes, or traffic: Valuing bicycle facilities with an adaptive stated preference survey, *Transportation Research A*, **41** (4) 287-301.

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

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

Upadhyay, D., N. Schüssler, K.W. Axhausen, M. Flamm and V. Kaufman (2008) Optimal parameter values for GPS post-processing: An experiment, *Working Paper*, **507**, IVT, ETH Zurich, Zurich.

Van der Zijpp, N.J. und S. Fiorenzo-Catalano (2005) Path enumeration by finding the constrained k-Shortest paths, *Transportation Research 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. und R. Golledge (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.