DTU Library

Behavioural models for cycling - Case studies of the Copenhagen Region.

Halldórsdóttir, Katrín

Publication date:
2015

Document Version
Publisher's PDF, also known as Version of record

Citation (APA):
Halldórsdóttir, K. (2015). *Behavioural models for cycling - Case studies of the Copenhagen Region.* DTU Transport.

BEHAVIOURAL MODELS FOR CYCLING

CASE STUDIES OF THE COPENHAGEN REGION

PhD Thesis

Katrín Halldórsdóttir
2015

DTU Transport
Department of Transport

BEHAVIOURAL MODELS FOR CYCLING

CASE STUDIES OF THE COPENHAGEN REGION

PhD thesis

Katrín Halldórsdóttir

Technical University of Denmark
Department of Transport

Supervisor:

Professor Otto Anker Nielsen

Technical University of Denmark
Department of Transport

Co-supervisor:

Professor Carlo Giacomo Prato

Technical University of Denmark
Department of Transport

Kgs. Lyngby 2015

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

PREFACE

This PhD thesis, "*Behavioural models for cycling - Case studies of the Copenhagen Region*", concludes my PhD study. The study is financed by the Danish Road Directorate, under the project "The effect of cycling policies", which has been conducted from December 2010 to September 2015 at the Department of Transport, at the Technical University of Denmark. Professor Otto Anker Nielsen and Professor Carlo Giacomo Prato supervised the PhD study.

The main contributions of this study are the six papers listed below:

Halldórsdóttir et al. (2011): Halldórsdóttir, K., Christensen, L., Jensen, T.C. and Prato, C.G. (2011). Modelling mode choice in short trips - Shifting from car to bicycle. Presented at the *39th European Transport Conference*, 10th – 12th October, Glasgow, Scotland, UK.

Prato et al. (2015): Prato, C.G., Halldórsdóttir, K. and Nielsen, O.A. (2015). Latent lifestyle and mode choice decisions when travelling short distances. Presented at the *14th International Conference on Travel Behaviour Research*, 19th – 23rd July 2015, Windsor, England. Submitted to *Transportation* in September 2015.

Halldórsdóttir et al. (2015a):	Halldórsdóttir, K., Nielsen, O.A. and Prato, C.G. (2015). Home-end and activity-end preferences for access to and egress from train stations in the Copenhagen Region. Submitted to *International Journal of Sustainable Transportation* in September 2015.
Rasmussen et al. (2015):	Rasmussen, T.K., Ingvardson, J.B., Halldórsdóttir, K. and 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*. Available online 4th May 2015 from: doi:10.1016/j.compenvurbsys.2015.04.001
Halldórsdóttir et al. (2014):	Halldórsdóttir, K., Rieser-Schüssler, N., Axhausen, K.W., Nielsen, O.A. and Prato, C.G. (2014). Efficiency of choice set generation methods for bicycle routes. *European Journal of Transport and Infrastructure Research*, vol. 14, no. 4, pp. 332-348.
Halldórsdóttir et al. (2015b):	Halldórsdóttir, K., Nielsen, O.A. and Prato, C.G. (2015). Land-use and network effects on bicycle route choice in the Greater Copenhagen area. Presented at the *2nd Symposium of the European Association for Research in Transportation (hEART), 4th - 6th September 2013, Stockholm, Sweden.* Working paper.

Katrín Halldórsdóttir, 28th of September 2015

ACKNOWLEDGMENTS

My journey through this PhD study has been interesting, challenging, and very educating. It has contributed to great personal growth and invaluable experience. I have met many interesting and highly skilled researchers and PhD students, whose research and enthusiasms has been a great inspiration for my PhD work.

First and foremost, I would like to thank my supervisor Professor Otto Anker Nielsen for good ideas, interesting discussions, and methodological input. I would also like to thank my co-supervisor Professor Carlo G. Prato for numerous inspiring discussions, his patience with my never-ending questions, and support when it was needed. Both Otto's and Carlo's engagement and competences have been crucial, which has facilitated my learning process. Thank you both for showing me great patience, understanding, and encouragement.

The Danish Road Directorate is gratefully acknowledged for the financial support of the research project "Effects of cycling policies" that this PhD study forms a part of. I would also like to thank Linda Christensen for applying for the support and for encouraging me to apply for the PhD scholarship.

During my PhD study, I was fortunate enough to visit the Swiss Federal Institute of Technology in Zürich (ETHZ), at the Institute for Transport Planning and Systems (IVT), from January 2012 to April 2012. I would like to thank Professor Kay W. Axhausen for the supervision during my stay, interesting discussions, and methodological input, which led to the cooperation on one of the articles in my thesis. I would especially like to thank Dr. Nadine Rieser as she showed me exceptional attention and helped me to gain the most from my stay, which led to the cooperation on one of the articles in my thesis. Thank you for all your support, inspiring conversations, both professional and personal, and for all the laughter and good times. My heartfelt gratitude goes to Ilka and Javier for their long-time friendship and to the rest of the crew for the wonderful friendship they offered me and for making my stay in Zürich one of my best experiences. I would also like to thank Julie and Gianni for their incredible hospitality and for making me feel at home. The Idella Foundation, the Nordic Road Association (NVF), and IDA’s & Berg-Nielsens Foundation are acknowledged for financially supporting my visit at the Swiss Federal Institute of Technology in Zürich.

I am very grateful to all the people who participated in the GPS data-collection that formed an important part of my PhD study. This thesis would not have been the same without the contributions from the participants.

I would like to thank my colleagues and fellow PhD students at DTU Transport. Studying at DTU Transport has been of great value to me, as the learning environment has been very inspiring and I have met a number of fantastic people from whom I have learned a lot. Specifically, I would like to thank Linda Christensen and Thomas C. Jensen for their collaboration on a conference paper and I would especially like to thank Thomas K. Rasmussen and Jesper B. Ingvardson for our successful work on a journal paper. I would like also to thank Morten Aabrink and Svend Jacob Senstius for their assistance with the GPS data-collection, from the data collection, GPS post-processing, and their help with map-matching. I would like to thank Søren H. Pedersen as well for his help with everything GIS related. I especially want to thank Søren and Jacob for their assistance on constructing the bicycle network. I would also like to thank Morten, Jacob, and Søren for commenting on a specific chapter in the thesis. I would like to thank Hjalmar Christiansen for providing assistance with everything TU related and for assisting with data collection on parking availability at train stations. Finally, special thanks go to Ulla Salado-Jimena for assisting me with the Danish text in this thesis.

I wish to thank my friends for all their support and patience during the PhD work. A special thank goes to my friend Kristín for her support and interest in my work during the years of my PhD study and for proofreading at the final stages before submission. I would like to thank the three musketeers Bergrún, Sigrún, and Harpa for always being there and I look forward to making up for lost time. I would like to thank Eddi and Anna for always reminding me to have fun, especially in the earlier stages of my PhD work.

Finally, I would like to thank my family for their endless and loving support throughout the years. I would like to especially thank my parents, Sumarlína and Halldór, who have always been there for me. I am grateful for their continuous support throughout my education, for always believing in me, and for teaching me to never give up in the face of adversity. Special thanks go to my three brothers, Haukur, Halldór Örn, and Unnar Helgi, and to my sister-in-law Ester for endless and fruitful talks about nothing and everything.

Last but certainly not least, my deepest appreciation goes to my best friend and loving husband Sverrir. Thank you for your endless patience, understanding, and constant support through the ups and downs of this challenging journey. Thank you for coping with my long study hours, for your constant encouragement, and for your confidence in me. To our son Ragnar Þór, you make it all worthwhile. Together you embody my idea of happiness and joy and I look forward to catching up on quality time with you both.

ABSTRACT

Bicycle transport has traditionally been underrepresented in traffic models, because historically the main focus has been on modelling more resource-intensive investments in motor traffic roads and public transport. In order to decrease road congestion and to reduce the related health and societal problems, there is a growing interest in promoting more sustainable transport systems, with a particular emphasis on the bicycle as a sustainable transport alternative. Accordingly, the objective of this PhD study is to expand the knowledge about travellers' choices of the bicycle as a mean of transport above other alternatives, as well as to create knowledge on the interaction between infrastructure and cyclists' route choices. In this study, the focus is on the traditional approaches to mode choice modelling, where the focus is on all transport modes, as well as the modelling of cyclists' route choices. The study focuses on identifying which conditions can: (i) promote bicycle use, with an emphasis on everyday cycling; (ii) influence the shift from motorised private transport to a more sustainable transport alternative; and (iii) find methods that make cycling more attractive, e.g., improving accessibility.

The private car is the most dominant mode of transportation in cities throughout the world, even for short trips where it could easily be replaced by more sustainable transport options, such as bicycles. Like the car, a bicycle provides flexibility when travelling. It also cost much

less and, in some cases, is even a faster and more efficient choice of transport, especially in highly congested areas. While the car might be a more popular alternative, especially in suburban or rural areas where activities are dispersed over large distances, short trips appear more receptive to a decrease in the use of the car.

In this PhD study, the mode choice behaviour when travelling short distances was analysed, in the Copenhagen Region, in order to identify factors that affect the travellers' choices. Mixed logit models were estimated, in order to capture taste variations and differentiate travel time parameters across modes, on a dataset including trip information and socio-economic variables for 7,958 individuals and 10,982 trip chains with five available alternative modes (i.e., walking, cycling, car driver, car passenger, and public transport). The results showed that travellers' have heterogeneous preferences regarding the travel time of non-motorised modes, and more homogeneous preferences regarding the travel time of motorised modes. The results also showed that mode choice behaviour in short distance travelling is related to travellers' personal characteristics (i.e., gender, occupation, income, and having a public transport monthly pass) and their household characteristics (i.e., number of cars and family composition). Finally, the results showed that the mode choice is also related to the trip characteristics (i.e., hilliness, temperature, trip purpose, urban characteristics, and parking availability). Lastly, the results showed that cyclists have a heterogeneous preference towards temperature and hilliness, meaning that some cyclists do not mind hilly areas or lower temperatures, while others do. In order to encourage the shift from private cars to more sustainable transport alternatives, decision-makers need to address specific population groups for specific trip purposes and focus on factors that are able to make cycling more attractive. The results suggested that further investigation of heterogeneity might uncover whether different population groups exhibit different preference structures.

In previous literature on short trips, the focus has mainly been on mode choice models to uncover the determinants of choice between car and sustainable transport alternatives. Generally, the focus has been more on the characteristics of the alternatives and less on the socio-economic characteristics of the travellers, while considering the population to have homogeneous preferences and the same probability of shifting mode, regardless of their characteristics. However, this assumption appears rather unrealistic. In this thesis, in the first study on travel behaviour when travelling short distances, it was concluded that the choice between transport alternatives is not only related to the level-of-service characteristics of the alternatives, but also to a large extent the socio-economic characteristics of the travellers.

Based on that study, a more suitable methodological approach was adopted, namely a latent class choice model, to identify lifestyle groups and to understand how lifestyle affects mode choice decisions when travelling short distances. The model allows linking observable characteristics of the individual with the probability of them having chosen a certain lifestyle and then the probability of individuals, with that specific lifestyle, choosing a specific transport mode for short distances. Short trip chains in the Copenhagen Region were investigated, on a data sample with 10,982 observations with five available alternative modes (i.e., walking, cycling, car driver, car passenger, and public transport). The results highlight the importance of investigating the heterogeneity of the population, when analysing the potential for switching from the car to sustainable travel modes. The results showed that the population is split into four lifestyle groups: auto-oriented, bicycle-oriented, public transport-oriented, and public transport-averse. This population split is according to several characteristics (i.e., gender, age, family composition, number of cars, income, occupation, and residence location). Each lifestyle group has a heterogeneous perception of travel time, where the rates of substitution between alternative transport modes were extremely different. In addition, each lifestyle group weighs the dispreferences for public transport transfers differently, has a different perception of weather conditions on active travel modes, and selects a transport mode depending on the trip purpose. When thinking about measures to increase the attractiveness of sustainable transport options in short distance travelling, decision-makers should: (i) propose traditional or creative solutions to encourage car-oriented individuals out of their cars; (ii) direct public transport-averse individuals with policies that make the car unattractive; and (iii) hinder the attractiveness of cars in the future for bicycle- and walk-oriented individuals. When thinking about bicycle infrastructure improvements, the reduction of cycling travel time has little effect on car-oriented individuals, unless the time savings are very high, and bicycle-oriented individuals will only modify their routes as they already consider bicycles the fastest choice.

Efforts to increase the use of public transport, with the aim of improving the sustainability of cities, usually focus on the service of the public transport system itself, while the accessibility to and from the public transport network receives less attention. This PhD study contributes to the existing literature by investigating the choice of access and egress modes to and from train stations in the Copenhagen Region. This study adopted a mixed logit model that distinguished between the preference structure at the home-end and activity-end for travellers who have chosen trains as their main transport mode. The model accounted for the heterogeneity in the travellers' preferences and alternative mode perceptions, while investigating the effect of policy variables such as car parking availability, park & ride opportunities, bicycle parking availability and type, and the possibility of carrying bicycles on trains. The choices between five alternative transport modes was analysed (i.e., walking, cycling, being a car driver, being a car passenger, and riding a bus) for 2,921 observations of trips at the home-end of journeys, and 3,658 trips at the activity-end of journeys. The results showed that the choice of access and egress mode is affected by travel time and trip characteristics (i.e., travelling with someone or in the city centre), as well as underlining the relevance of bicycle parking and the possibility of carrying bicycles on trains to the choice of cycling to the train station. Most importantly, the results showed that travellers' have heterogeneous preferences with regard to travel time and perception of the alternatives, as well as their preference structure relates more to their socio-economic characteristics (i.e., gender, season ticket, occupation and trip purpose, along with the number of cars and other motorists in the household) than the trip characteristics. The study successfully identified factors that can contribute to the sustainability of the travel choices after selecting a train as the main transport mode, e.g., by improving bicycle parking availability at train stations, but focusing on specific population groups might also contribute further, especially when considering travellers' occupation and trip purpose.

Bicycle route choice models provide measures to search for factors that make cycling more attractive. In this study, the findings from the model estimates depend on the observation of actual route choices and the generation of realistic alternatives. While collecting data on actual route choices has greatly profited from enhancements in GPS device technology, the post-processing of such large data is still difficult. In this study, a fully automatic post-processing procedure was proposed and applied to extract relevant information for further analysis. It makes it possible to process raw individual-based GPS data, with no additional information required from the respondent, by combining fuzzy logic- and GIS-based methods. By applying this method it is possible to automatically identify trips, trip-stages, and the most probable transport mode used on each trip-stage. The method was validated on a dataset consisting of raw individual-based GPS logs, collected from 183 respondents living in the Greater Copenhagen area, with a total of 427 trip legs, thereof 113 bicycle trips legs. The method was validated through the application of a control-questionnaire. The study showed that using the proposed method: (i) correctly linked 82% of the reported trip legs to corresponding trip legs, (ii) avoided classifying non-trips such as scatter around activities as trip legs, (iii) correctly identified the transport mode for more than 90% of the trip legs, and (iv) was robust through the specification of the model parameters and thresholds. The results document that using the proposed method enabled the possibility of using individual-based GPS units to collect travel surveys in large-scale multi-modal networks.

The literature on the generation of alternative route sets has mainly focused on the implementation of path generation methods for cars or public transport, which are normally generated on a simplified network. Only few studies have focused on bicycle route choice sets, which require a highly detailed network. In this study, the efficiency of choice set generation methods was analysed by their ability to generate relevant and heterogeneous bicycle routes in a high-resolution network by using different evaluation methods, such as replicating the observed routes while also generating realistic alternatives that take into account taste heterogeneity across cyclists. Three choice set generation methods for bicycle route choice were examined: A doubly stochastic generation function, a breadth first search on link elimination, and a branch & bound algorithm. The dataset used to evaluate the methods consisted of 778 bicycle trips traced by GPS and carried out by 139 persons. In addition, the extension of cost functions was proposed with bicycle-oriented factors not limited to distance and time, but also other factors considered relevant to cyclists, i.e., scenic routes, dedicated cycle lanes, and road types. The results showed that both the doubly stochastic generation function and the breadth first search on link elimination generated realistic routes, while the first produced more heterogeneous routes and the latter outperformed in computation cost. The two methods revealed similar performances in terms of coverage, i.e., almost 64% and 68%, respectively. The branch & bound method had lower coverage compared to the other two methods, as it reproduced approximately 40% of the observed routes. As to be expected, shorter routes resulted in a very good coverage for all methods, where there are typically (much) less possible alternative routes, while longer routes exhibited larger differences across algorithms, with the doubly stochastic generation function performing best. The results also indicated the heterogeneous and complex preference structure for cyclists when considering routes, thus emphasising the importance of realistic and heterogeneous alternative route sets for model estimation.

Based on the above data, cyclists' route choices were analysed by estimating a path-size logit model, accounting for similarities between the alternative routes. A large sample of GPS observations was estimated, comprised of 3,363 bicycle trips total. The logarithm of the path-size variable was significant and positive, thus correctly accounts for route overlap. The results showed that cyclists are sensitive to the effects of distance, cycling the wrong way, turn frequency, hilliness, different bicycle facility types, bicycle bridges, surface type, intersection type (i.e., cyclists prefer roundabouts over other crossing types), the number of motorised traffic lanes, and crossing water/sea on motorised traffic bridges. Whereas motorised traffic type, speed limit, annual average daily traffic (AADT), time dependent traffic volumes, and accident patterns had no statistically significant effect on cyclists' route choices. Most importantly, the results showed that cyclists appear to place relatively high value on different land-use conditions along the routes, that is, dispreference for high residential area and/or town centre and industrial areas, a willingness to take detours to cycle in recreational areas or parks when they are on both sides of the path, but avoidance of these detours when such areas are on one side of the path. Previous model estimates showed that the parameters describing paths along a scenic area and in forests did not have a significant effect on cyclists' route choices. The results also showed that personal characteristics influence the route choice (i.e., gender and type of cyclist), that there were differences in route choice preferences depending on the time of day and whether it was weekday or weekend, and also different weather conditions (i.e., temperature, rain, and sunshine). The route choice model can be used to forecast future travel behaviour. However, the interaction between the bicycle route choice model and the mode choice models needs to be investigated. By focusing on the interaction between infrastructure and route choice of cyclists, it is possible to contribute to the understanding of which factors influence cyclists' route choices.

The work conducted in the PhD study contributes to the current literature on bicycle transport by investigating the choice of the bicycle as a transport alternative and cyclists' route choices. Problems related to the modelling of cyclists' route choices were successfully solved, i.e., by collecting actual route choices using individual-based GPS units, post-processing the raw GPS data in order to get usable information on observed bicycle routes, and effectively generating realistic alternatives in a high resolution network. It was possible to analyse travel behaviour on extensive revealed preference data and the study showed that it is possible to estimate quite advanced models on an elaborate set of variables and utility functions. The findings showed that it is important to take into consideration the heterogeneity of individuals and that decision-makers should focus on specific individuals or groups within the population when thinking about measures to increase the appeal of sustainable travel options. The findings also showed the importance of well-built bicycle facilities and the importance of choosing the location of such facilities carefully.

DANSK ABSTRAKT

Cykeltrafik har traditionelt været underrepræsenteret i trafikmodeller, idet fokus historisk set hovedsageligt har været på modellering af mere ressourcekrævende investeringer i vejanlæg og kollektiv transport. For at mindske trængsel på vejene og reducere relaterede sundheds- og miljøproblemer er der en voksende interesse i at fremme mere bæredygtige transportsystemer. Formålet med denne ph.d.-afhandling er at udbygge viden om trafikanternes valg af cykel som transportmiddel frem for andre transportmiddel. Formålet er også at skabe viden om samspillet imellem infrastrukturen og cyklisters rutevalg. Afhandlingen fokuserer på de traditionelle metoder til modellering af transportmiddelvalg, med fokus på alle transportmidler, samt på modellering af cyklisters rutevalg. Afhandlingen fokuserer på at identificere, hvilke forhold der kan (i) fremme cykling med fokus på hverdagscykelture, (ii) påvirke et skifte fra privatbilisme til et mere bæredygtigt transportalternativ og (iii) finde metoder, der gør cykling mere attraktivt, f.eks. ved at forbedre tilgængeligheden.

Privatbilisme er den mest fremherskende transportform i byer i hele verden, selv til korte ture, hvor bilen nemt kunne erstattes af mere bæredygtige transportvalg, f.eks. cykling. Ligesom bilen er cyklen en fleksibel transportform. Det er også meget billigere, og i visse tilfælde er det også et hurtigere og mere effektivt transportvalg, især i stærkt befærdede

områder. Selvom bilen nok er det mest populære alternativ, især i forstæder eller landområder, hvor aktiviteterne er spredt over store afstande, ser det ud til, at de ture, hvor folk vil være mest tilbøjelige til at bruge bilen mindre, er de korte ture.

Afhandlingen undersøger derfor adfærden i forbindelse med transportmiddelvalg for korte ture i Københavnsområdet for at identificere faktorer, der påvirker de rejsendes valg. Der er estimeret "mixed logit"-modeller for at afdække præferenceændringer og differentiere rejsetidsparametre på tværs af transportmidler på et datasæt af observerede ture med tilknyttede socioøkomiske data, der omfatter 7.958 personer og 10.982 turkæder, med fem transportmidler i valgsæt (dvs. gang, cykel, bilfører, bilpassager og kollektiv transport). Resultaterne viser, at de rejsende har heterogene præferencer med hensyn til rejsetiden, når det drejer sig om ikke-motoriserede transportmidler, og mere homogene præferencer med hensyn til rejsetiden, når det drejer sig om motoriserede transportmidler. Resultaterne viste også, at adfærden ved valg af transportmiddel i tilfælde af korte ture afhænger af de rejsendes personlige karakteristika (dvs. køn, beskæftigelse, indkomst, og periodekort) og deres husstands karakteristika (dvs. antal biler og familiestruktur). Endelig viste resultaterne, at transportmiddelvalget også afhænger af karakteristika ved selve turen (dvs. kuperet terræn, temperaturen, rejseformål, bymæssige karakteristika og parkeringsmuligheder). Endelig viste resultaterne, at cyklister har forskellige præferencer overfor temperatur og kuperet terræn, hvilket betyder, at nogle cyklister ikke har noget imod kuperet terræn eller lave temperaturer, mens andre har. Når beslutningstagerne skal tilskynde rejsende til at skifte fra privatbiler til mere bæredygtige transportalternativer, er de nødt til at adressere specifikke befolkningsgrupper for så vidt angår specifikke rejseformål samt fokusere på faktorer, der gør cykling mere attraktivt. Resultaterne indikerer, at yderligere forskning i heterogenitet vil kunne afdække, hvorvidt forskellige befolkningsgrupper har forskellige præferencestrukturer.

Tidligere litteratur om bæredygtig transport har især fokuseret på transportmiddel-valgsmodeller til at afdække afgørende faktorer i forbindelse med valget mellem private motoriserede transportalternativer og bæredygtige transportalternativer. Der har generelt været mere fokus på alternativernes karakteristika og mindre på de rejsendes socioøkonomiske karakteristika, mens befolkningen er blevet anset for homogen med hensyn til præferencer og med hensyn til sandsynligheden for at skifte transportmiddel uanset deres karakteristika. Denne formodning må imidlertid anses for at være temmelig urealistisk. Afhandlingen har i den første undersøgelse om rejseadfærd i forbindelse med korte ture vist, at valget mellem transportalternativer ikke kun afhænger af alternativernes karakteristika, men også i vidt omfang af de rejsendes socioøkonomiske karakteristika.

På basis heraf valgte denne afhandling en mere passende metodisk tilgang, dvs. en “latent class”- valgmodel til at analysere livsstilsgrupper og forstå, hvordan livsstil påvirker beslutningen om rejseform for korte ture. Modellen gør det muligt at sammenkæde individers observerbare karakteristika med sandsynligheden for at have valgt en bestemt livsstil og derefter sandsynligheden for, at en person med denne specifikke livsstil vælger et specifikt transportmiddel for korte turkæder. Afhandlingen undersøgte korte turkæder i Københavnsområdet på grundlag af et datasæt med 10.982 observationer, med fem transportmidler i valgsæt (dvs. gang, cykel, bilfører, bilpassager og kollektiv transport). Resultaterne understreger vigtigheden af at undersøge befolkningens heterogenitet, når man undersøger muligheden for at ændre transportmiddelvalg til bæredygtige transportmidler. Resultaterne viste, at befolkningen kan opdeles i fire livsstilsgrupper, dvs. de bilorienterede, de cykelorienterede, dem, der er orienteret mod kollektiv transport, samt dem, der er modstandere af kollektiv transport. Denne opdeling af befolkningen kan forklares ved forskellige karakteristika (dvs. køn, alder, familiestruktur, antallet af biler, indkomst, beskæftigelse, og bopæls placering). De enkelte livsstilsgrupper har heterogene rejsetidspræferencer, hvor substitutionsgraden mellem transportmiddelalternativerne var signifikant forskellig mellem hver livsstilgruppe. Endvidere har alle livsstilsgrupperne alt-andet-lige-præferencer for at skifte transportmiddel i kollektivt transport, de er følsomme overfor vejrforholdene og deres transportmiddelvalg afhænger af turformålet. Når der overvejes foranstaltninger til at øge interessen for bæredygtige transportmuligheder i tilfælde af korte ture, skal beslutningstagerne: (i) foreslå traditionelle eller kreative løsninger for at få bil-orienterede individer ud af deres biler; (ii) rette politikker, der gør bilen uinteressant, mod personer, der utilbøjelige til at anvende kollektiv transport; og (iii) gøre biler mindre attraktive i fremtiden for cykel- og gangorienterede individer. Når der overvejes cykelinfrastrukturforbedringer, har en reduktion af cykelrejsetiden ringe effekt på bil-orienterede individer, medmindre tidsbesparelser er meget høje, og cykel-orienterede enkeltpersoner vil kun ændre på deres ruter, da de allerede betragter cykler som det hurtigste alternativ.

Indsatsen for at øge brugen af kollektiv transport, med det formål at forbedre byernes bæredygtighed, fokuserer normalt på det kollektive transportsystems serviceniveau, mens tilgængeligheden til og fra det kollektive transportnet får mindre opmærksomhed. Denne ph.d.-afhandling bidrager til den eksisterende litteratur ved at undersøge valg af tilbringer/frabringer-transportmiddel til og fra togstationer i Københavnsområdet. Afhandlingen estimerede “mixed logit”-modeller, der skelner mellem præferencestrukturer for de to turender (hjemme-enden og aktivitets-enden) for rejsende, der har valgt tog som deres hovedtransportmiddel. I modellerne blev der taget hensyn til heterogeniteten i de rejsendes præferencer og opfattelse af alternative transportmidler, mens effekten af politiske variabler såsom parkeringstilgængelighed, ”Park & Ride”-muligheder, tilgængelighed og type af cykelparkering, samt muligheden for at tage cyklen med på toget blev undersøgt. Valget mellem fem transportformer blev analyseret (gang, cykel, bilfører, bilpassager og bustransport), hvor datasættet til brug for modelestimeringen bestod af 2.921 observationer af ture til/fra de rejsendes bopæl, og 3.658 ture ved aktivititets-enden (herunder ture mellem aktiviteter). Resultaterne viste, at valget af tilbringere/frabringere påvirkes af rejsetid og tur karakteristika (dvs. om man rejser sammen med nogen eller er i bymidten), samt understreger relevansen af cykelparkering og muligheden for at tage cykler med på toget for valg cykling til togstationen. Resultaterne viste endvidere, at de rejsende har heterogene præferencer for så vidt angår rejsetid og vurdering af de forskellige transportmidler, og at deres præferencestruktur relaterer mere til deres socioøkonomiske karakteristika (dvs. køn, periodekort, erhverv og rejseformål samt antallet af biler og andre bilister i husstanden) end turens karakteristika. Det lykkedes i forbindelse med undersøgelsen at identificere faktorer, der kan bidrage til bæredygtigheden af de foretagne rejsevalg efter valg af tog som hovedtransportmiddel, f.eks. ved at forbedre cykelparkeringen på stationerne, men fokus på bestemte befolkningsgrupper kan også bidrage yderligere, især når man overvejer rejsendes beskæftigelse og turformål.

Rutevalgsmodeller for cykler giver mulighed for at finde faktorer, der gør cykling mere attraktivt. Estimationen heraf har i afhandlingen benyttet observerede ruter samt beregning af realistiske alternativer. Selvom indsamlingen af data om de aktuelle rutevalg har nydt stor gavn af forbedringerne inden for GPS-teknologien, så er efterbehandlingen af sådanne store datamængder fortsat en udfordring. I afhandlingen præsenteres og benyttes en fuldautomatisk efterbehandlingsprocedure. Dette gør det muligt at behandle rå individbaserede GPS-data, uden behov for yderligere oplysninger fra respondenten, ved at kombinere “fuzzy logic”- og GIS-baserede metoder. Ved hjælp af denne metode var det muligt automatisk at identificere ture, delture og det mest sandsynlige transportmiddel på den enkelte deltur. Metoden blev valideret på et datasæt bestående af rå individbaserede GPS-logs indsamlet fra 183 respondenter boende i Storkøbenhavn, dækkende i alt 427 turben, heraf 113 cykelturen. Til valideringen af metoden anvendtes et kontrolspørgeskema. Undersøgelsen viste, at anvendelse af den foreslåede metode: (i) forbandt 82 % af de rapporterede turben korrekt til de tilhørende turben, (ii) undgik at klassificere ikke-ture såsom “punktsværme” rundt om aktivitetspunkter som turben, (iii) identificerede transportmidlet korrekt for mere end 90 % af turbenene, og (iv) var robust ved specificering af modelparametre og tærskelværdier i modellen. Resultaterne dokumenterede, at anvendelse af den foreslåede metode gjorde det muligt at anvende individbaserede GPS-enheder til at indsamle rejsevaneundersøgelser i storskala multimodale transportnet.

Litteraturen om generering af valgsæt til estimering af rutevalgsmodeller har hovedsageligt fokuseret på implementering af metoder til generering af valgsæt til biler eller offentlig transport, som normalt genereres på et forenklet netværk. Kun få undersøgelser har fokuseret på valgsæt til cykeltrafik, som kræver et meget detaljeret netværk. I denne afhandling blev effektiviteten af metoderne til generering af valgsæt undersøgt med henblik på at fastslå deres evne til at generere relevante og forskellige cykelruter i et detaljeret digitalt kort ved at anvende forskellige evalueringsmetoder, såsom at replikere de observerede ruter samtidig med at der genereres realistiske alternativer, der tager hensyn til cyklisternes heterogenitet i forhold til valg af ruter. Der blev undersøgt tre metoder til generering af valgsæt for cykelrutevalg: En dobbelt-stokastisk metode, en metode baseret på grafsøgning med fravalg af kanter (breadth first search on link elimination) og en forgreningsalgoritme ("branch & bound") algoritme. Datasættet til brug for vurdering af metoderne bestod af 778 GPS-sporede cykelture udført af 139 personer. Afhandlingen foreslår ydermere en udvidelse af nyttefunktionerne med cykelorienterede faktorer, der ikke er begrænset til afstand og tid, men også andre faktorer, der anses for relevante for cyklister, dvs. smukke ruter, dedikerede cykelstier og vejtyper. Resultaterne viste, at såvel den dobbelt-stokastiske metode som grafsøgning med fravalg af kanter genererede realistiske ruter, hvor førstnævnte genererede mere heterogene ruter og sidstnævnte klarede sig bedst med hensyn til regnetid. De to metoder viste sig at have næsten samme performance hvad angår dækning af de observerede ruter, nemlig hhv. 64 % og 68 %. "Branch & bound"-metoden havde en lavere dækning i sammenligning med de to andre metoder, da den kun reproducerede ca. 40 % af de observerede ruter. Som forventet havde alle metoderne en meget god dækning for de korte ruter, hvor der typisk er (meget) færre mulige alternative ruter, mens der for de lange ruters vedkommende var større forskelle på tværs af algoritmerne, med den dobbelt-stokastiske metode som den bedste. Resultaterne indikerede endvidere cyklisternes heterogene og komplekse præferencestruktur, når de overvejer ruter, hvilket understreger vigtigheden af realistiske og heterogene alternative valgsæt til modelestimering.

På basis af ovennævnte data analyserede afhandlingen cyklisternes rutevalg ved at estimere en "path-size logit model" for at tage højde for overlap mellem alternative ruter. Et stort datasæt af GPS- observationer, der bestod af 3.363 cykelture i alt, blev estimeret. Logaritmen for "path-size"-faktoren var signifikant og positiv og tog korrekt højde for ruteoverlap. Resultaterne viste, at cyklisters rutevalg påvirkes af turens længde, af at cykle den forkerte vej, svinghyppighed, kuperet terræn, forskellige former for cykelstier, cykelbroer, overfladetyper, vejkrydstyper (dvs. at cyklister foretrækker rundkørsler fremfor andre former for kryds), antallet af vognbaner til motorkørertøjer og bilbroer, der krydser vand/hav. Derimod påvirkede motoriseret trafik type, hastighedsgrænse, årsdøgnstrafik (ÅDT), tidafhængige trafikmængder og ulykkesmønstre ikke cyklisters rutevalg signifikant. Resultaterne viste især, at arealanvendelsen langs ruten også har stor betydning for cyklisterne, idet de har negative præferencer for boligområder med høj bebyggelse og/eller bymidter og industriområder, mens de er positive overfor at tage omveje for at kunne cykle i rekreative områder og parker, når de ligger på begge sider af ruten, mens cyklisterne undgår disse omveje, når de rekreative områder kun ligger på den ene side af ruten. Tidligere modelestimater viste, at de parametre, der beskriver ruter i et naturskønt område og i skove, ikke havde en signifikant effekt på cyklisternes rutevalg. Resultaterne viste også, at personlige karakteristika påvirker rutevalget (dvs. køn og cyklisttype), at der var forskelle i rutevalgspræferencerne afhængig af tidspunkt på dagen, og om det var hverdag eller weekend, og endelig også forskellige vejrforhold (dvs. temperatur, regn og solskin). Resultaterne understreger vigtigheden af veludformede cykelstier mv. og betydningen af at vælge placeringen af nye cykelstier omhyggeligt. Rutevalgsmodellen kan bruges til at forudsige fremtidig rejseadfærd. Samspillet mellem rutevalgsmodellen for cyklister og transportmiddelvalgsmodellerne bør dog undersøges nærmere. Ved at fokusere på samspillet mellem infrastruktur og cyklisters rutevalg er det muligt at bidrage til forståelsen af, hvilke faktorer der påvirker cyklisternes rutevalg.

Det arbejde, der er blevet udført i forbindelse med denne ph.d.-afhandling, bidrager til den nuværende litteratur om transport på cykel ved at undersøge valg af cyklen som transportalternativ og cyklisternes rutevalg. Afhandlingen har med succes løst problemer relateret til modellering af cyklisters rutevalg, dvs. ved at indsamle faktiske rutevalg med individbaserede GPS-enheder, efterbehandling af de rå GPS-data med henblik på at få brugbare oplysninger om observerede cykelruter og effektivt generere realistiske alternativer i et højopløsningsnetværk. Vi var i stand til at analysere rejseadfærd på baggrund af et omfattende sæt af afslørede præferencedata, og afhandlingen viste, at det er muligt at estimere ganske avancerede modeller på et udførligt sæt variabler og nyttefunktioner. Resultaterne viste, at det er vigtigt at tage hensyn til, at befolkningen er en uhomogen gruppe, og at beslutningstagerne bør fokusere på bestemte personer eller grupper i befolkningen, når de overvejer foranstaltninger for at gøre bæredygtige transportformer mere attraktive. Resultaterne påviste også vigtigheden af veludformede cykelstier og betydningen af at vælge placeringen af sådanne cykelstier omhyggeligt.

TABLE OF CONTENTS

BEHAVIOURAL MODELS FOR CYCLING I

PREFACE I
ACKNOWLEDGMENTS III
ABSTRACT VII
DANSK ABSTRAKT XIII
TABLE OF CONTENTS XIX
LIST OF FIGURES XXV
LIST OF TABLES XXIX

PART I 1

INTRODUCTION 1

CHAPTER 1 3

INTRODUCTION 3
1.1 BACKGROUND 3
1.2 RESEARCH OBJECTIVES 5

1.3 OUTLINE OF THE THESIS 6

PART II **9**

MODE CHOICE MODELLING **9**

CHAPTER 2 **11**

INTRODUCTION **11**
2.1 MODELLING MODE CHOICE IN SHORT TRIPS – SHIFTING FROM CAR TO BICYCLE 12
2.2 LATENT LIFESTYLE AND MODE CHOICE DECISIONS WHEN TRAVELLING SHORT DISTANCES 13
2.3 HOME-END AND ACTIVITY-END PREFERENCES FOR ACCESS TO AND EGRESS FROM TRAIN STATIONS IN THE COPENHAGEN REGION 14

CHAPTER 3 **15**

DATA DESCRIPTION **15**
3.1 THE DANISH NATIONAL TRAVEL SURVEY 16
3.2 DATA DESCRIPTION 16
3.3 DATA PREPARATION 37
3.4 ADDITIONAL INFORMATION 49
3.5 REFERENCES 51

CHAPTER 4 **53**

MODELLING MODE CHOICE IN SHORT TRIPS – SHIFTING FROM CAR TO BICYCLE **53**
ABSTRACT 55
4.1 INTRODUCTION 57
4.2 DATA 59
4.3 METHODOLOGY 60
4.4 RESULTS 62
4.5 CONCLUSIONS 70
4.6 REFERENCES 72

CHAPTER 5 **77**

LATENT LIFESTYLE AND MODE CHOICE DECISIONS WHEN TRAVELLING SHORT DISTANCES **77**
ABSTRACT 79
5.1 INTRODUCTION 81
5.2 METHODS 83
5.3 ESTIMATION RESULTS 89

5.4 Summary and conclusions 97
Acknowledgments 98
5.5 References 98

CHAPTER 6 **105**

Home-end and activity-end preferences for access to and egress from train stations in the Copenhagen Region **105**
Abstract 107
6.1 Introduction 109
6.2 Earlier research 110
6.3 Methods 111
6.4 Results 115
6.5 Conclusions 124
Acknowledgements 125
6.6 References 125

PART III **129**

Route choice modelling **129**

CHAPTER 7 **131**

Introduction **131**
7.1 Improved methods to deduct trip legs and mode from travel surveys using wearable GPS devices: A case study from the Greater Copenhagen area 132
7.2 Efficiency of choice set generation methods for bicycle routes 133
7.3 Land-use and transport network effects on bicycle route choice in the Greater Copenhagen area 134

CHAPTER 8 **137**

Data description **137**
8.1 Data collection 138
8.2 Bicycle network database 154
8.3 References 157

CHAPTER 9 **159**

Improved methods to deduct trip legs and mode from travel surveys using wearable GPS devices: A case study from the Greater Copenhagen area **159**

Abstract 161
9.1 Introduction 163
9.2 Literature review 164
9.3 Methodology 167
9.4 Case study: Greater Copenhagen area 172
9.5 Results 178
9.6 Transferability and sensitivity 184
9.7 Conclusions 189
Acknowledgements 190
9.8 References 191

CHAPTER 10 195

Efficiency of choice set generation methods for bicycle routes 195
Abstract 197
10.1 Introduction 199
10.2 Data 200
10.3 Methods 202
10.4 Results 210
10.5 Discussion and conclusions 215
Acknowledgements 217
10.6 References 217

CHAPTER 11 221

Land-use and network effects on bicycle route choice in the Greater Copenhagen area 221
Abstract 223
11.1 Introduction 225
11.2 Data 227
11.3 Model development 231
11.4 Results 235
11.5 Conclusion 243
Acknowledgment 244
11.6 References 244

PART IV 251

Conclusions 251

CHAPTER 12 253

CONCLUSIONS **253**
12.1 MAIN CONTRIBUTIONS 253
12.2 RECOMMENDATION FOR FUTURE WORK 255
12.3 MAIN CONCLUSIONS 259

PART V **261**

APPENDIX **261**

APPENDIX A **263**

BICYCLE NETWORK DATABASE – DATA DESCRIPTION **263**

APPENDIX B **269**

RESULTS FROM HALLDÓRSDÓTTIR ET AL. (2015B) **269**
B.1 GENDER 272
B.2 TYPE OF CYCLIST 274
B.3 DARKNESS 276
B.3 PEAK HOURS 278
B.4 WEEKEND 280
B.5 AIR TEMPERATURE 282
B.6 RAIN 284
B.7 SUNSHINE 286

List of Figures

Figure 1: Overview of the four-step travel modelling process 4
Figure 2: The structure of the PhD thesis 6
Figure 3: Distribution of trips on bicycle across trip distance in Denmark, 2012-2014. *Source:* Skougaard and Christiansen (2015) 17
Figure 4: Cycled kilometres per capita in the respondents' area in Denmark, 2012-2014. *Source:* Skougaard and Christiansen (2015) 18
Figure 5: Bicycle mode share in commuting trips in large cities, 2009. *Source:* Pucher and Buehler (2012, page 292) 18
Figure 6: Average daily travel on a bicycle (km/person/day) divided into age groups in Denmark, 2012-2014. Source: Skougaard and Christiansen (2015) 19
Figure 7: Share of total bicycle transport by trip purpose (km) in Denmark, 2012-2014. Source: Skougaard and Christiansen (2015) 20
Figure 8: Travel distance (km) per person, divided by main occupation in Denmark, 2012-2014. Source: Skougaard and Christiansen (2015) 20
Figure 9: Definition of a trip chain and trips 21
Figure 10: Mode share when travelling short distances in the Copenhagen Region, 2006-2010 22
Figure 11: Distribution of mode share per distance when travelling short distances in the Copenhagen Region, 2006-2010 23

Figure 12: Number of observations per distance when travelling short distances in the Copenhagen Region, 2006-2010 23
Figure 13: Distribution of journeys per urban area when travelling short distances in the Copenhagen Region, 2006-2010 24
Figure 14: Age group distribution when travelling short distances in the Copenhagen Region, 2006-2010 24
Figure 15: Gender share when travelling short distances in the Copenhagen Region, 2006-2010 25
Figure 16: Share of gender and choice of transport mode when travelling short distances in the Copenhagen Region, 2006-2010 25
Figure 17: Share of main occupation and choice of transport mode when travelling short distances in the Copenhagen Region, 2006-2010 26
Figure 18: Share of trip purpose and choice of transport mode when travelling short distances in the Copenhagen Region, 2006-2010 27
Figure 19: Mode share at the home-end when accessing passenger trains in the Copenhagen Region, 2006-2011 29
Figure 20: Mode share at the activity-end when accessing passenger trains in the Copenhagen Region, 2006-2011 29
Figure 21: Distribution of mode share per distance at home-end when accessing passenger trains in the Copenhagen Region, 2006-2011 30
Figure 22: Distribution of mode share per distance at activity-end when accessing passenger trains in the Copenhagen Region, 2006-2011 30
Figure 23: Number of observations per distance when accessing passenger trains in the Copenhagen Region, 2006-2011 31
Figure 24: Age group distribution when accessing passenger trains in the Copenhagen Region, 2006-2011 32
Figure 25: Gender share when accessing passenger trains in the Copenhagen Region, 2006-2011 32
Figure 26: Share of gender and choice of access and egress mode at home-end when accessing passenger trains in the Copenhagen Region, 2006-2011 33
Figure 27: Share of gender and choice of access and egress mode at activity-end when accessing passenger trains in the Copenhagen Region, 2006-2011 33
Figure 28: Share of main occupation and choice of access and egress mode at home-end when accessing passenger trains in the Copenhagen Region, 2006-2011 34
Figure 29: Share of main occupation and choice of access and egress mode at Activity-end when accessing passenger trains in the Copenhagen Region, 2006-2011 34
Figure 30: Share of trip purpose and choice of access and egress mode at home-end when accessing passenger trains in the Copenhagen Region, 2006-2011 36
Figure 31: Share of trip purpose and choice of access and egress mode at activity-end when accessing passenger trains in the Copenhagen Region, 2006-2011 36
Figure 32: Definition of zones and zone centroids 39
Figure 33: Definition of trip and stages 39
Figure 34: The NAVTEQ network used in Halldórsdóttir et al. (2011) and Prato et al. (2015). Left: Overview of Denmark, Right: Zoom of the Copenhagen Region 41

Figure 35: The NAVTEQ network used in Halldórsdóttir et al. (2015a). Left: Overview of Denmark, Right: Zoom of the Copenhagen Region 42
Figure 36: Definition of a zone connector 43
Figure 37: NTM public transport network. Left: Overview of Denmark, Right: Zoom of the Copenhagen Region 46
Figure 38: TOP10DK network. Left: Overview of Denmark, Right: Zoom of the Copenhagen Region .48
Figure 39: The process flow in part III 132
Figure 40: Number of respondents in each municipality in the Greater Copenhagen area, in each round and over the total sample 141
Figure 41: Gender share in the GPS dataset 142
Figure 42: Age group distribution in the GPS dataset 143
Figure 43: Share of respondents with a driving license in the GPS dataset 143
Figure 44: Share of education level in the GPS dataset (HF is abbreviation for higher preparatory certificate, HHX is higher commercial certificate, and HTX is higher technical certificate) 144
Figure 45: Share of main occupation in GPS dataset 144
Figure 46: Percentage distribution of cycling respondents per municipality in the Greater Copenhagen area. Left: GPS dataset. Right: TU-survey....147
Figure 47: Hourly split of cycling, during weekdays, according to the TU-survey....149
Figure 48: Hourly split of cycling, during weekdays, in the collected GPS data 149
Figure 49: Average speed of cyclists divided into age-groups and gender 150
Figure 50: Average speed, the number of cyclists in each interval 151
Figure 51: Number of bicycle trips in each rainfall (mm/hour) interval 152
Figure 52: Number of bicycle trips in each sunshine (minutes) interval 152
Figure 53: Number of trips in each air temperature (°C) interval 153
Figure 54: Number of bicycle trips in each wind speed (m/s) interval 153
Figure 55: Bicycle network. Left: Zoom of the Copenhagen Region, Right: Illustration of the network details 155
Figure 56: Approach used in this study. Boxes highlighted in light red denote steps that are similar to corresponding steps in Schüssler and Axhausen (2009). Boxes highlighted in dark red are steps where this paper contributes with new methods 168
Figure 57: The stepwise mode classification algorithm. Continuous arrows denote mode classification whereas dotted arrows denote no change from previous step. Step 2 is directly adopted from Schüssler and Axhausen (2009), but with adapted fuzzy logic rules 169
Figure 58: Example of a rail trip leg (shown with red) on the Danish S-train ring line (Ringbanen) identified by the rail proximity algorithm. The railway network is highlighted by bold black lines, and stations with green dots 174
Figure 59: The distributions of the 95th percentiles of speed and acceleration and the median speed for the subset of trip legs for which the mode is known (rail excluded) from the control questionnaire 175
Figure 60: Example of results from the bus stop algorithm. The lines represent relevant bus lines, while black and green dots represent GPS points and stops, respectively 177
Figure 61: Classification of trip legs identified by the algorithms 180

Figure 62: Distribution of the observations over distance travelled 201
Figure 63: Distribution of coverage over 778 observations 213
Figure 64: Average maximum coverage over distance 213
Figure 65: Percentage of chosen route completely replicated as a function of distance 214
Figure 66: Path size distribution for the resulting route sets 215
Figure 67: Bicycle network database in the Greater Copenhagen area. Left: Overview of the Greater Copenhagen area network. Right: Illustration of the network details 228
Figure 68: Age group distribution in the GPS dataset 230
Figure 69: Feedback mechanism in the four-step travel modelling process 257
Figure 70: Feedback mechanism for access/egress modes to passenger trains 258

LIST OF TABLES

Table 1: Number of observations per main occupation and the choice of transport mode when travelling short distances in the Copenhagen Region, 2006-2010 26
Table 2: Number of observations per trip purpose and the choice of transport mode when travelling short distances in the Copenhagen Region, 2006-2010 27
Table 3: Number of observations per main occupation and the choice of access and egress mode at home-end when accessing passenger trains in the Copenhagen Region, 2006-2011 35
Table 4: Number of observations per main occupation and the choice of access and egress mode at activity-end when accessing passenger trains in the Copenhagen Region, 2006-2011 35
Table 5: Number of observations per trip purpose and the choice of access and egress mode at home-end when accessing passenger trains in the Copenhagen Region, 2006-2011 37
Table 6: Number of observations per trip purpose and the choice of access and egress mode at activity-end when accessing passenger trains in the Copenhagen Region, 2006-2011 37
Table 7: Time periods for the road traffic assignment 40
Table 8: Time periods for the public transport assignment 45
Table 9: Category variables for personal characteristics 63
Table 10: Category variables for trip characteristics 64
Table 11: Continuous variables 65
Table 12: The results from the mixed logit model estimates 66
Table 13: Sample characteristics 88

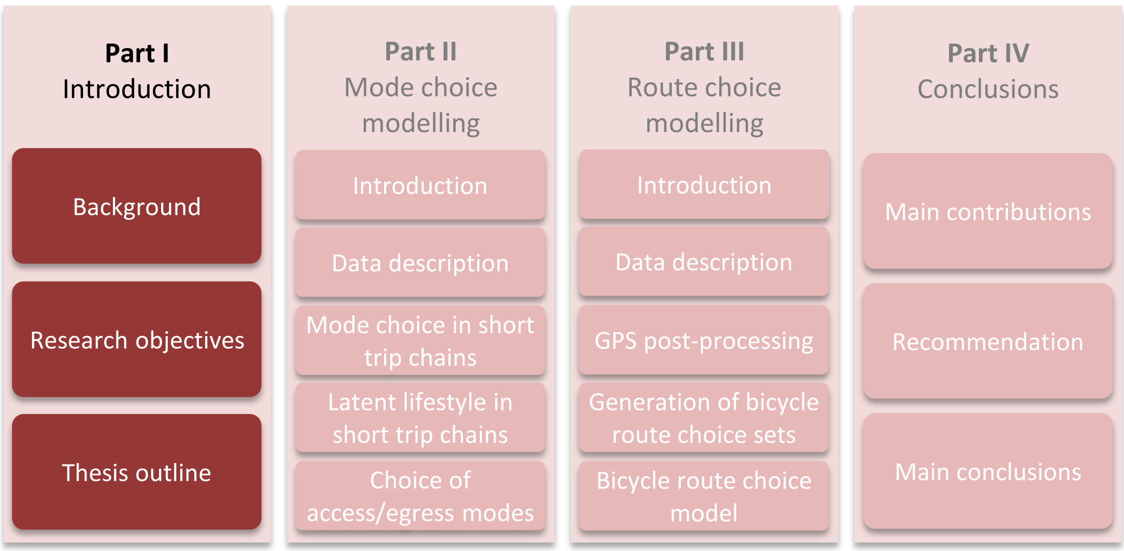
Table 14: Performances of the estimated choice models 89
Table 15: Estimates of the EC logit model (without latent lifestyle segmentation) 90
Table 16: Estimates of the class specific choice model 92
Table 17: Estimates of the class membership model 95
Table 18: Trip characteristics 116
Table 19: Travellers' characteristics 118
Table 20: Mixed logit model estimates 120
Table 21: Comparison the socio-demographics of the GPS data sample together with the TU-survey (HF is abbreviation for higher preparatory certificate, HHX is higher commercial certificate, and HTX is higher technical certificate) 146
Table 22: Fuzzy logic rules applied 176
Table 23: Feedback algorithms joining trip legs. ΔTime and ΔSpace refer to temporal and spatial distances between end of trip A and start of trip B (C for 3 leg cases), respectively. Order is the classification of the 2 (3) trip legs and Classification is the mode assigned to the merged trip leg 178
Table 24: The results of the mode identification when using Baseline algorithm (compared to reported mode use) 182
Table 25: The results of the mode identification when using Algorithm 1 without map matching (compared to reported mode use) 182
Table 26: The results of the mode identification when using Algorithm 2 with map matching (compared to reported mode use) 183
Table 27: Sensitivity tests for step 1: rail proximity. Comparison of classified rail trip legs to reported mode for different thresholds for (i) the share of observations in vicinity, (ii) the distance to the network, and (iii) the length of the trip legs 186
Table 28: Sensitivity tests for step 2: fuzzy logic rules. Number of trip legs correctly/wrongly classified as walk or bike. Each version denotes different specifications for intervals of speed and acceleration profiles and different fuzzy logic rules 187
Table 29: Sensitivity tests for step 3:bus line alignment. Fit with observed mode for different thresholds for the share of bus stops stopped at and threshold for time in the vicinity of the bus stops 188
Table 30: Sensitivity analysis of map matching component to remove non-trips. Number of correctly/wrongly classified trip legs remaining for various thresholds of the minimum share being matched by using GPS observations. Also listed is the number of non-trips remaining in the data set 189
Table 31: Implementation of the branch & bound generation technique 204
Table 32: Overview of the calibrated parameters and variation factors 205
Table 33: Coverage measures of the applied algorithms 210
Table 34: Computational costs and techniques consistency 212
Table 35: Predefined cost parameters, plus the added scale parameters 233
Table 36: Variable description of the network- and land-use attributes for the chosen and non-chosen routes 236
Table 37: The results from the basis path-size logit model estimates 238

Table 38: Bicycle network - attributes overview 265

PART I

INTRODUCTION

INTRODUCTION TO THE PHD STUDY



Chapter 1

INTRODUCTION

INTRODUCTION TO THE PHD STUDY

1.1 BACKGROUND

It is a common practise to use the support of traffic models in the overall planning process of implementing a new infrastructure or a political initiative. There are many compelling arguments for added political efforts aimed at increasing cycling. However, bicycle transport has traditionally been underrepresented in traffic models. Historically, the primary focus has mainly been on modelling more resource-intensive investments in motor traffic roads and public transport. In addition, there are various aspects in modelling that need to be taken into account, such as how extensive the transport system is, that individuals are not always rational in their choices, and also that they have heterogeneous preferences.

Figure 1 gives an overview of the four-step travel modelling process, which is a standard urban transportation planning system. Decision-makers can use this tool to analyse current travel behaviour and to forecast future behaviour, in order to prioritise different projects. The four steps of the model are:

- *Trip production*, which calculates the frequency of trips in each zone, categorised as trip generation and trip attraction, as a function of land-use, household demographics, and other socio-economic characteristics.
- *Trip distribution*, where the number of trips between the origins and destinations are calculated, often using a gravity model function.
- *Mode choice*, which estimates the ratio of transport modes, used between each origin and destination pair.
- *Route choice* (route assignment), which assigns the trips by each transport mode to a route between an origin and destination.

Mode choice models enable decision-makers to forecast how the transport mode share will shift when building a new infrastructure or implementing a political initiative. Route choice models can be used to determine the need for a new infrastructure or to forecast how a political initiative would influence current route choice behaviour. First, the number of travellers on each network link can be estimated by assigning the trips between each zone pair. The generated routes thus represent the present traffic patterns. Then, decision-makers can apply route assignments to forecast how the distributions of routes will possible change when building a new infrastructure or when implementing a political initiative. After applying different scenarios with these common models, the cost and/or benefit of the proposed future projects can be evaluated through different decision criteria.

The PhD study focuses on the two last steps of the four-step travel modelling process, while concentrating on cycling, namely the choice of the bicycle as a transport alternative and modelling the route choices of cyclists.

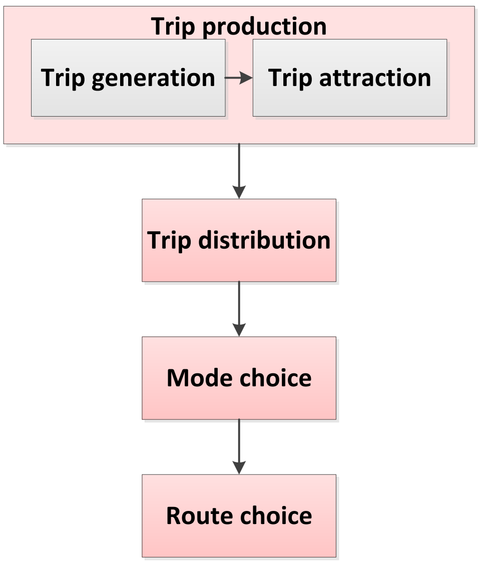


FIGURE 1: OVERVIEW OF THE FOUR-STEP TRAVEL MODELLING PROCESS

1.2 Research objectives

The objective of this PhD study is first to expand the knowledge of travellers' choice of the bicycle as a means of transport above other alternatives. The objective is also to create knowledge on the interaction between infrastructure and cyclists' route choices. Accordingly, the main research objective of this study was divided into two main themes.

In the first theme, focus was on modelling the choice of bicycle as a transport mode. By investigating different transport modes, it is possible to identify which conditions can promote bicycle use, with an emphasis on everyday cycling, and which conditions can influence the shift from motorised private transport to a more sustainable transport alternative, such as bicycles. In this theme, mode choice in short distance travelling was investigated, as bicycles are an ideal transport alternative when travelling short distances. In addition, combining bicycles with public transport is a realistic alternative to cars when travelling longer distances. The following two main objectives were set out to be completed within the mode choice modelling theme:

(i) To estimate advanced mode choice models in order to identify the determinants of choice between the private car and sustainable travel alternatives in short distance travelling;
(ii) To understand the preference related to the choice of access and egress modes to and from train stations.

The second theme dealt with the modelling of cyclists' route choices. Bicycle route choice models provide measures to search for factors that make cycling more attractive, by focusing on the actual route choices of cyclists. In order to model the route choices of cyclists, numerous parts need to be taken into consideration. Consequently, the focus in this theme was on the following four main objectives:

(iii) To collect data using GPS technology to register geographical points, recording the behaviour of a sample of bicyclists from different municipalities in the Greater Copenhagen area;
(iv) To develop a fully automatic post-processing procedure, making it possible to process raw individual-based GPS data with no additional information required from the respondent;
(v) To analyse the efficiency of choice set generation methods to generate realistic bicycle routes;
(vi) To develop a model to analyse cyclists' route choices and evaluate their trade-offs.

1.3 Outline of the thesis

This PhD thesis consists of four main parts: (I) introduction, (II) mode choice modelling, (III) route choice modelling, and (IV) conclusions. An appendix provides additional information. The thesis is structured so to reflect the overall work process during the PhD study. There are twelve chapters, where each chapter consists of different sections and subsections. Figure 2 provides a general overview of the thesis structure. The main contributions of the PhD study are six papers, enclosed in the two main parts, Part II and Part III.

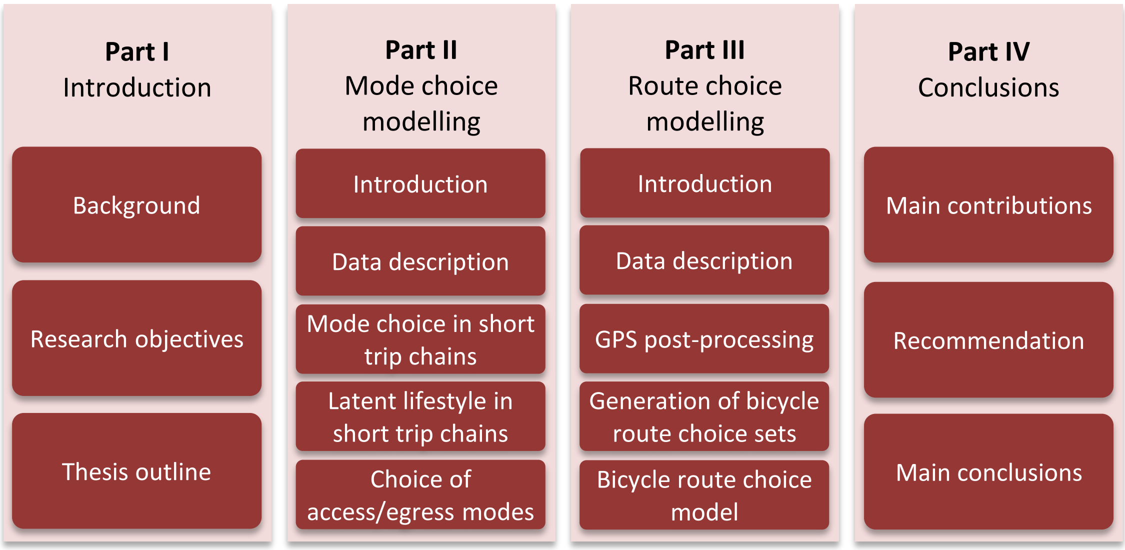


Figure 2: The structure of the PhD thesis

The structure of the thesis is as follows:

Part I – Introduction

Chapter 1 introduces the background to the PhD study, the main research objectives, and finally outlines the PhD thesis' contents.

Part II – Mode choice modelling

Part II deals with the first theme of this study, which focuses on modelling the choice of bicycle as a transport mode. Part II contains the first three papers that form the basis of this study. They are presented in ***chapter 4*** to ***chapter 6***.

Chapter 2 The main contents of Part II are introduced and the objectives of the three papers, enclosed in this part, are outlined.

Chapter 3 The Danish National Travel Survey, which formed the basis for the model estimation in all three papers, is presented. In this chapter, some general data description is presented, followed by a description of the data preparation conducted for all three papers.

Chapter 4 The conference paper "Modelling mode choice in short trips – Shifting from car to bicycle", which was presented at the *39th European Transport Conference* (ETC), in October 2011 in Glasgow, Scotland, is presented. In the paper, an investigation of the mode choice behaviour of individuals in the Copenhagen Region is presented. Focus is on short trip chains, with particular emphasis on shifting from private motorised transport to a more sustainable mode, such as bicycle.

Chapter 5 The paper "Latent lifestyle and mode choice decisions when travelling short distances" is presented. This paper was presented at the *14th International Conference on Travel Behaviour Research* (IATBR), in July in Windsor, England. The paper was then submitted to *Transportation* in September 2015. The paper proposes a latent class analysis, in order to understand how lifestyle affects the decision on how to travel short distances, in the Copenhagen Region.

Chapter 6 The paper "Home-end and activity-end preferences for access to and egress from train stations in the Copenhagen Region" is presented. The paper was submitted to *International Journal of Sustainable Transportation* in September 2015. In this paper, the multimodal travelling of individuals in the Copenhagen Region is investigated, with focus on the access and egress mode choice for trips where the main transport mode is passenger trains.

Part III – Route choice modelling

In Part III focus is on the second theme, the modelling of the route choices of cyclists. Part III contains the last three papers, which are presented in ***chapter 9*** to ***chapter 11***. These papers also form the basis of this study.

Chapter 7 Presents the main subjects of Part III, which deal with bicycle route choice modelling, and outlines the main objectives of the three papers.

Chapter 8 discusses the GPS data collection conducted for the bicycle route choice model and presents some findings from an analysis conducted on the data, followed by a description of the bicycle network constructed for this study.

Chapter 9 presents the journal paper "Improved methods to deduct trip legs and mode from travel surveys using wearable GPS devices: A case study from the Greater Copenhagen area". The paper was published online in *Computers, Environment and Urban Systems* in May 2015. A fully automated method is proposed in order to post-process large raw individual-based GPS datasets, collected in the highly complex large-scale multi-modal network, with no additional information requested from the respondent. The method combined already established methods to identify trips, trip legs, and assign the most probable mode of transport, together with a combined fuzzy logic- and GIS-based algorithm.

Chapter 10 presents the journal paper "Efficiency of choice set generation methods for bicycle routes", which was published in the *European Journal of Transport and Infrastructure Research* in December 2014. The paper analysed the efficiency of choice set generation methods for bicycle routes and proposed the extension of cost functions to bicycle-oriented factors not limited to distance and time.

Chapter 11 presents the working paper "Land-use and transport network effects on bicycle route choice in the Greater Copenhagen area". A preliminary version was presented at the *2nd Symposium of the European Association for Research in Transportation (hEART),* in September 2013, Stockholm, Sweden. The paper analyses cyclists' route choices and evaluates their trade-offs in an established bicycle city, thus providing inspiration for emerging cycling cities by focusing in particular on the interaction between infrastructure and cyclists' route choice.

Part IV – Conclusions

Chapter 12 summarises the main contributions of the PhD study and then discusses how the different models, used in this study, can be put together in an overall model framework, that improves bicycle modelling and supports policies, followed by the main conclusions of the overall study.

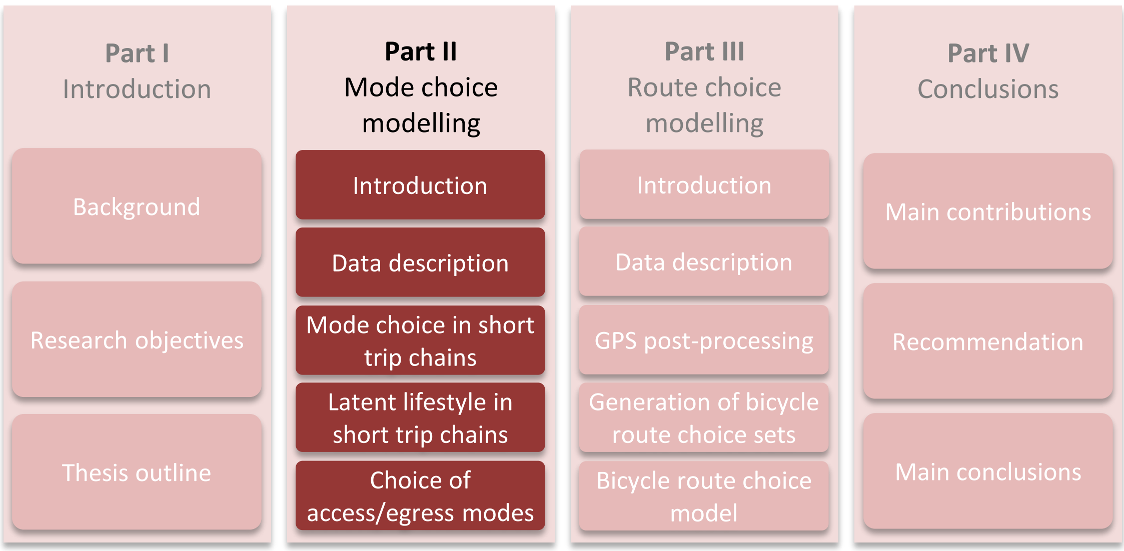
Part V – Appendix

The appendix gives an overview and a description of the attributes in the bicycle network database and presents some additional findings from Halldórsdóttir et al. (2015b).

PART II

MODE CHOICE MODELLING

IDENTIFY CONDITIONS THAT PROMOTE BICYCLE USE



Chapter 2

INTRODUCTION

INTRODUCTION TO PART II

Part II focuses on modelling the choice of bicycles as a transport mode, by focusing on all major transport alternatives. By doing so, it is possible to identify factors that influence travellers' choices and to identify which conditions can promote bicycle use, with an emphasis on everyday cycling, and encourage the shift from motorised private modes to sustainable alternative options.

Short trips are extremely frequent in everyday life for multiple purposes, such as commuting, shopping, and picking up or dropping off children. Still, even in short trips, private cars are the most dominant mode of transportation in cities throughout the world. However, the use of private cars could easily be replaced by bicycles, when travelling short distances, as they provide flexibility when travelling, cost much less and, in some cases, are even faster and a more efficient choice of transport, especially in highly congested areas. In addition, with the aim of improving the sustainability of cities, combining the use of bicycles with public transport could be a realistic alternative to cars, when travelling longer distances. However, efforts to increase the use of public transport usually focus on the service of the public transport system itself, while the accessibility to and from the public transport network receives less attention.

Accordingly, Part II addresses the first two research objectives, listed in the introduction in Part I. The first two papers investigate the mode choice behaviour in short distance travelling, while the third paper investigates the choice of access and egress modes to and from train stations.

In order to investigate the mode choice behaviour in the Copenhagen Region, in short distance travelling and when accessing train stations, level-of-service information is needed on the actual mode choices, as well as the alternative modes. Chapter 3 describes the Danish National Travel Survey, which formed the basis for the model estimation in all three papers, and some general data description, followed by a description of the data preparation conducted for all three papers. The following subsections briefly describe the papers and clarify which research objectives, and additional contributions, they address.

2.1 MODELLING MODE CHOICE IN SHORT TRIPS – SHIFTING FROM CAR TO BICYCLE

Title:	Modelling mode choice in short trips – Shifting from car to bicycle
Author(s):	Katrín Halldórsdóttir, Linda Christensen, Thomas C. Jensen, and Carlo G. Prato
Presented:	The *39th European Transport Conference (ETC)*, 10^{th} – 12^{th} October 2011, Glasgow, Scotland, UK
Abbreviated:	Halldórsdóttir et al. (2011)

The purpose of this paper is to identify factors that affect the choice of transport modes in short distance travelling. Investigating short distance travelling is particularly important, as the number of short trips is especially high for multiple purposes, where the private car may easily be replaced by the bicycle. In this paper, the mode choice behaviour of individuals from the Copenhagen Region when travelling short distances was investigated, with a particular emphasis on uncovering the determinants of choice between cars and sustainable transport alternatives, such as cycling. In this paper, the first research objective is tackled:

(i) To estimate advanced mode choice models in order to identify the determinants of choice between the private car and sustainable travel alternatives in short distance travelling.

In addition, this paper has the following aims:

(i.a) To estimate mixed logit models, that are used to capture taste variations and differentiate travel time parameters across modes;

(i.b) To investigate mode choice by focusing on the characteristics of the alternatives, on the socio-economic characteristics of the travellers, and the trip characteristics;

(i.c) To uncover factors that can make cycling more attractive when travelling short distances;

(i.d) To understand which policies might be effective in promoting the choice of sustainable transport alternatives, in order to reduce car traffic.

2.2 LATENT LIFESTYLE AND MODE CHOICE DECISIONS WHEN TRAVELLING SHORT DISTANCES

Title:	Latent lifestyle and mode choice decisions when travelling short distances
Author(s):	Carlo G. Prato, Katrín Halldórsdóttir, and Otto A. Nielsen
Presented:	Presented at the *14th International Conference on Travel Behaviour Research*, 19th – 23rd July 2015, Windsor, England
Submitted:	Submitted to *Transportation* in September 2015
Abbreviated:	Prato et al. (2015)

The purpose of this paper is to identify lifestyle groups and to understand how lifestyle affects mode choice decisions when travelling short distances, by adopting a latent class choice model. In the existing literature on short distance travelling, the main focus has been on mode choice models when investigating which factors influence the choice between private motorised alternatives and sustainable travel alternatives. The characteristics of the alternatives are usually the main focus, while the socio-economic characteristics of the travellers receive less attention. More importantly, it is assumed that the population as homogeneous in its preferences. However, this assumption seems rather unrealistic and thus this paper offers a different approach. This paper proposes a latent class analysis that helps to unravel how lifestyle affects the decision of how to travel in short trip chains in the Copenhagen Region. This paper also tackles the first research objective in the PhD study:

(i) To estimate advanced mode choice models in order to identify the determinants of choice between the private car and sustainable travel alternatives in short distance travelling.

More specifically, this paper also has the following aims:

(i.e) To estimate a latent class choice model to identify latent lifestyle groups and choice specific travel behaviour;

(i.f) To investigate the probable heterogeneity preferences across individuals;

(i.g) To investigate how lifestyle influences the short-term choices of travel mode rather than long-term choices.

2.3 Home-end and activity-end preferences for access to and egress from train stations in the Copenhagen Region

Title:	Home-end and activity-end preferences for access to and egress from train stations in the Copenhagen Region
Author(s):	Katrín Halldórsdóttir, Otto A. Nielsen, and Carlo G. Prato
Submitted:	Submitted to *International Journal of Sustainable Transportation* in September 2015.
Abbreviated:	Halldórsdóttir et al. (2015a)

The purpose of this paper is to understand the determinants of the choice of access and egress mode for travellers who have chosen trains as their main transport mode. Generally, the efforts to increase the use of public transport usually focus on the service of the public transport system itself, while the accessibility to and from the public transport network receives less attention. Previous literature shows that the access and egress mode to and from the railway network is an important factor when travelling by train. Accordingly, this paper focuses on the second research objective:

(ii) To understand the preference related to the choice of access and egress modes to and from train stations.

In addition, the paper has four broad aims:

(ii.a) To estimate a mixed logit model, to account for heteroscedasticity across alternatives and heterogeneity across travellers;

(ii.b) To differentiate between the preference structure at the home-end and activity-end for travellers who have chosen train as their main transport mode;

(ii.c) To investigate the effect of policy variables such as car parking availability, park & ride opportunity, bicycle parking availability and type, the possibility to carry a bicycle on the train, alongside socio-economic characteristics of the travellers and level of service measures of the travel modes.

Chapter 3

DATA DESCRIPTION

THE DANISH NATIONAL TRAVEL SURVEY

This chapter presents the data used for the model estimations in all three papers in Part II. In order to analyse current travel behaviour or to forecast future travel behaviour, the observed choices and the related choice sets of the non-chosen alternatives for each traveller are necessary. The observed choices were obtained from the Danish National Travel Survey (abbreviated TU-survey, in Danish Transportvaneundersøgelsen) (Christiansen and Skougaard, 2015), while route choice models and simulation methods were used to calculate the attributes of the alternatives within the choice set of each traveller. The history of the TU-survey is briefly described in subsection 3.1, while subsection 3.2 describes travel behaviour in Denmark based on Skougaard and Christiansen (2015). Subsection 3.2 also briefly describes travel behaviour in the Copenhagen Region, when travelling short distances and when accessing passenger trains. Subsection 3.3 describes the data preparation and how the level-of-service variables were calculated for each transport alternative. Finally, subsection 3.4 briefly describes additional information, collected for Halldórsdóttir et al. (2015a), on parking facilities at train stations and the possibility to bring bicycles on trains.

3.1 THE DANISH NATIONAL TRAVEL SURVEY

The TU-survey characterises the travel behaviour of a representative sample of the Danish population, combining detailed information on actual travel behaviour with a large variety of background variables. Information on the actual travel, includes information on how much, how, where, when, and why the trips take place, for how long, etc. In order to describe travel behaviour for selected groups, background variables are also collected, e.g., age, gender, income, education, car availability, etc.

The respondents are asked about their travel behaviour on a single day. The survey runs all year round, collecting data throughout the year, thus characterising the differences in travel behaviour between seasons and different days of the week. The survey also collects highly detailed geographical information, such as precise home address information and destination location. Approximately 95% of all locations are geographically coded directly by the respondents, through a “search and select” procedure. In the remaining cases, the respondents complete the locations description as a free text, which is post-processed afterwards such that 97% can be geo-coded at a coordinate level and 99.5% at a zonal level (Christiansen and Skougaard, 2015).

A new questionnaire was implemented in February 2009, with a thorough mapping of the public transport travel routes. This includes information on precise address of activities, as well as information on bus stops and stations (Anderson, 2013). This highly detailed information makes it possible to analyse the public transport routes more accurately and the characteristics of the mode choice depending on the route.

DTU Transport organises the data collection, on behalf of the Ministry of Transport and several other ministries, and carries the overall responsibilities for the survey. The TU-survey collects information on trips, travelled on the day before the interview of Danes, between the age of 10 and 84, that are randomly selected. The response rate is on average approximately 60%, which is considered quite satisfactory for such an extensive survey of this type. The data collection is mainly conducted as telephone interviews, approximately 80% of the data, while self-reported internet interviews are approximately 20% (Christiansen and Skougaard, 2015).

3.2 DATA DESCRIPTION

The following subsections describe some findings from a data analysis conducted on the TU-survey. First, subsection 3.2.1 describes the travel behaviour in Denmark, based on data analysis conducted by Skougaard and Christiansen (2015) is discussed. Then, subsections 3.2.2 to 3.2.3 present descriptive statistics on travel behaviour in the Copenhagen Region, when travelling short distances and when accessing passenger trains.

3.2.1 Travel behaviour in Denmark

Skougaard and Christiansen (2015) found that in Denmark most bicycle trips are less than 5km (see Figure 3). In Denmark, 86% of all trips are less than 5km, while the total number of cycled kilometres per trip is 54%. Some Danes also cycle long distances, 22% of the total cycle kilometres are more than 11 km, despite the fact that they only constitute 4% of the trips.

By looking at the cycled kilometres per capita in the respondents' area (Figure 4), Skougaard and Christiansen (2015) found that Danes living in the larger cities generally cycle more than Danes living in the small- and medium-sized cities. Cyclists in Copenhagen cycle the most, where small- and medium-sized cities only cycle around half or less than that of the daily bicycle transportation in Copenhagen.

The level of bicycle use in Denmark is relatively high when compared to other countries, especially in Copenhagen (see Figure 5). Accordingly, Copenhagen and the capital region were used as a case study in this PhD thesis, with the objective to develop models where focus is on the mode- and route choice characteristics in an established bicycle city.

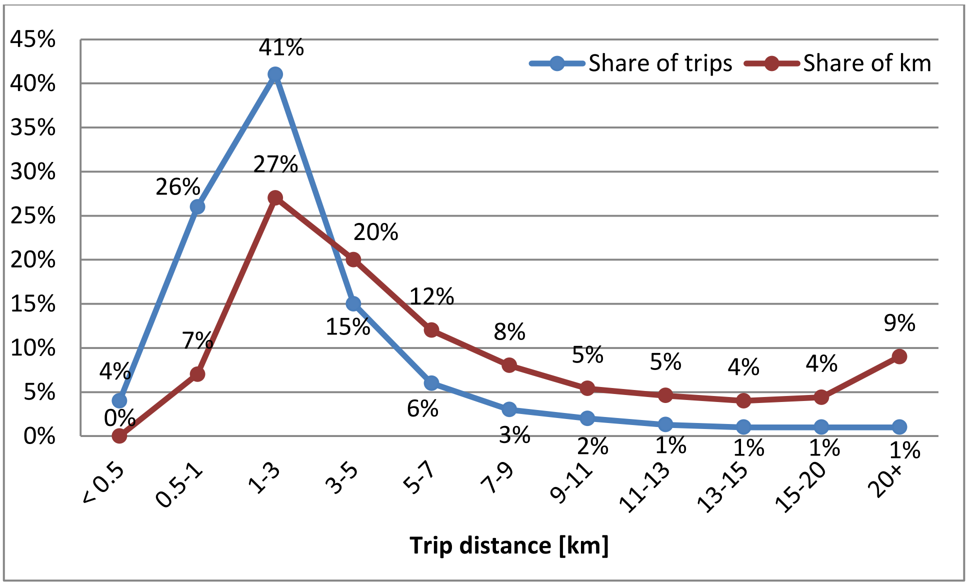


Figure 3: Distribution of trips on bicycle across trip distance in Denmark, 2012-2014. ***Source:*** **Skougaard and Christiansen (2015)**

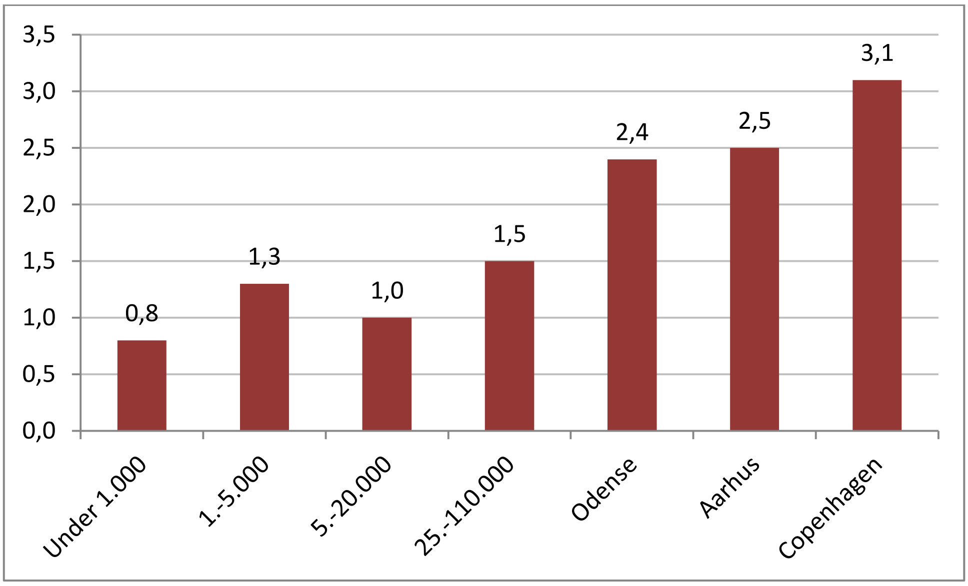


FIGURE 4: CYCLED KILOMETRES PER CAPITA IN THE RESPONDENTS' AREA IN DENMARK, 2012-2014. *SOURCE:* SKOUGAARD AND CHRISTIANSEN (2015)

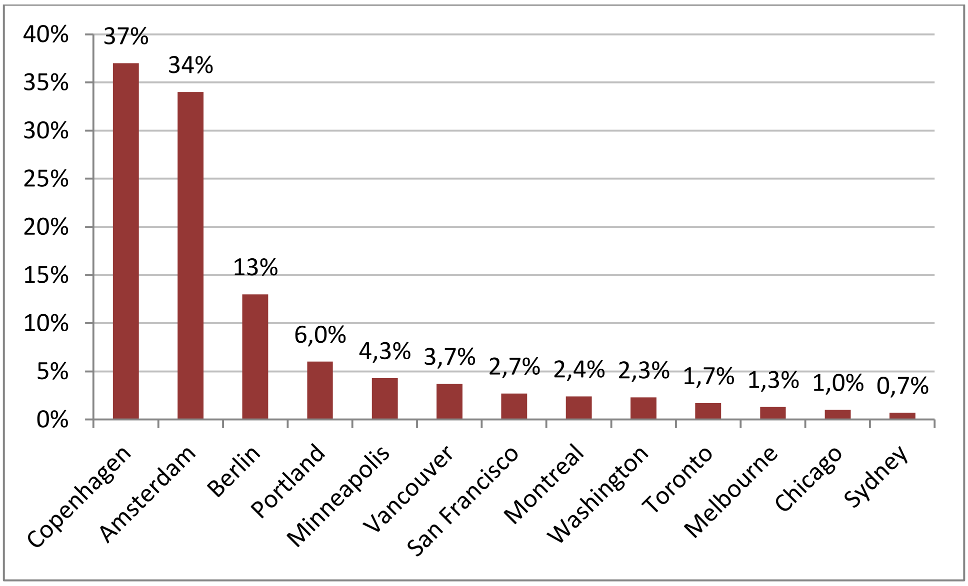


FIGURE 5: BICYCLE MODE SHARE IN COMMUTING TRIPS IN LARGE CITIES, 2009. *SOURCE:* PUCHER AND BUEHLER (2012, PAGE 292)

There is a clear correlation between age and daily cycled kilometres as shown by Skougaard and Christiansen (2015). By comparing these two factors (Figure 6), it can be seen that young respondents cycle the most. The number of cycled kilometres rises slightly until the age of 30 and remains unchanged onwards until the age of 60, after which it starts to decrease.

Skougaard and Christiansen (2015) also found that a large share of bicycle trips, measured in terms of cycled kilometres, consists of commuting trips to work or a place of study (Figure 7), with 34% work trips and 12% to a place of study. Therefore, it is not surprising that employees, students, and school pupils cycle the most (Figure 8). Bicycle transport as a leisure activity also has a large share of cycled kilometres, or a total of 37%.

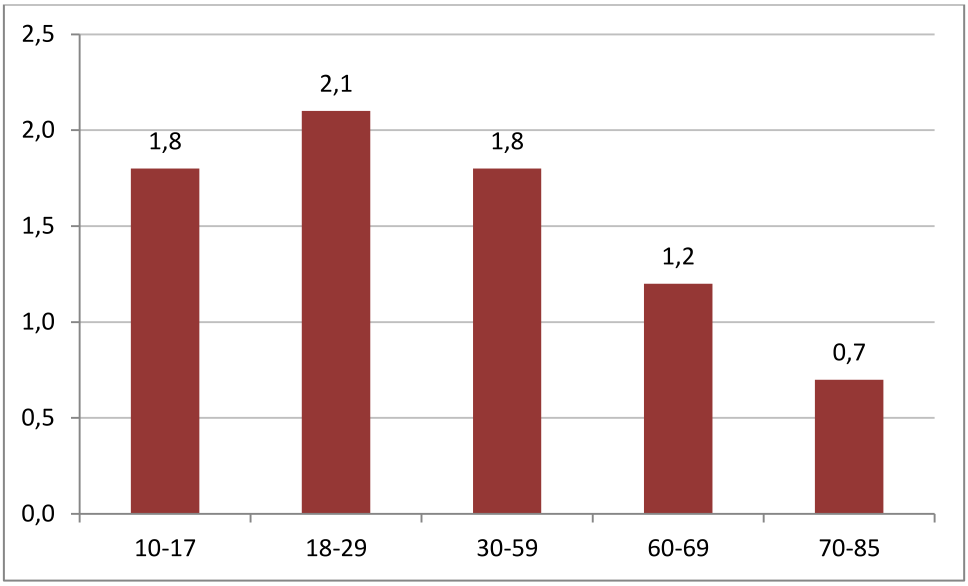


FIGURE 6: AVERAGE DAILY TRAVEL ON A BICYCLE (KM/PERSON/DAY) DIVIDED INTO AGE GROUPS IN DENMARK, 2012-2014. SOURCE: SKOUGAARD AND CHRISTIANSEN (2015)

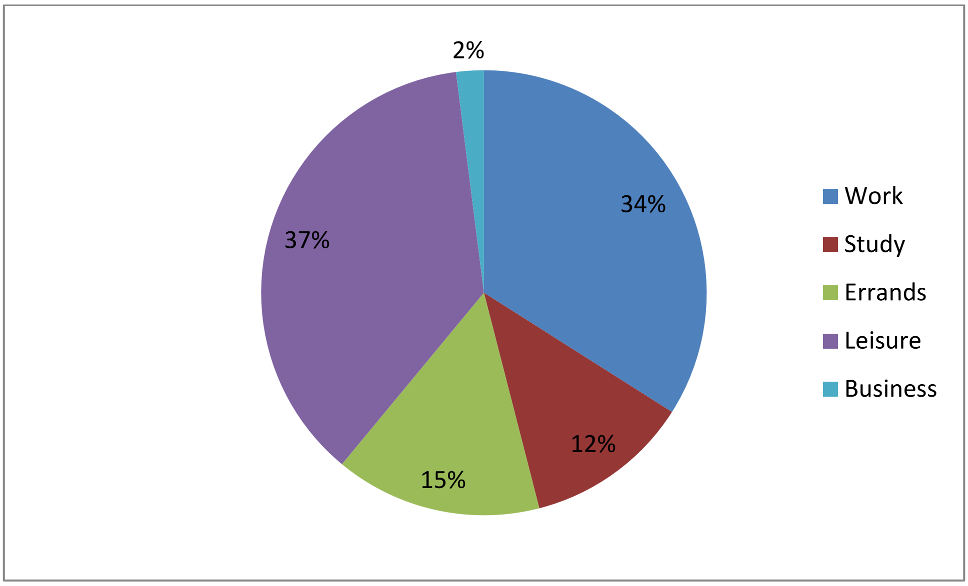


FIGURE 7: SHARE OF TOTAL BICYCLE TRANSPORT BY TRIP PURPOSE (KM) IN DENMARK, 2012-2014. SOURCE: SKOUGAARD AND CHRISTIANSEN (2015)

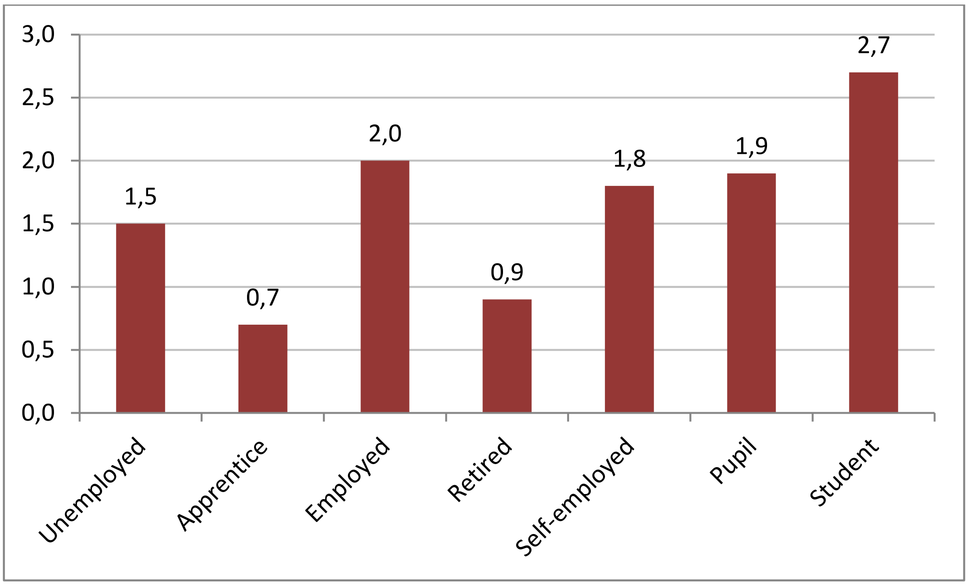


FIGURE 8: TRAVEL DISTANCE (KM) PER PERSON, DIVIDED BY MAIN OCCUPATION IN DENMARK, 2012-2014. SOURCE: SKOUGAARD AND CHRISTIANSEN (2015)

3.2.2 TRAVEL BEHAVIOUR WHEN TRAVELLING SHORT DISTANCES IN THE COPENHAGEN REGION

Halldórsdóttir et al. (2011) and Prato et al. (2015) estimated the choice of transport mode in journeys, i.e., the total travel in a chain of trips, which start and end at home (Figure 9). This definition was chosen as normally it is not possible to choose between a car and a bicycle in the middle of a trip chain, after leaving home with one means of transport. Christensen and Jensen (2008) analysed the choice of transport mode of car owners in Denmark. The paper showed that car owners occasionally choose to cycle on journeys up to 22 km, while they rarely cycle longer distances. Thus, Halldórsdóttir et al. (2011) and Prato et al. (2015) investigated journeys shorter than 22 km, as this travel distance was considered the limit for when it is appropriate to influence the shift from car to a bicycle. The data sample only included individuals above 18 years, as it is the driving age limit in Denmark. Observations were excluded from the dataset if respondents opted not to provide information or if other relevant information was missing for the analysis. Given the restrictions, the data sample was extracted from the TU-survey including 7,958 individuals and 10,982 journeys conducted in the Copenhagen Region, during the period 2006-2010. Only selected results are presented in this section, given the extensive data collected from the TU-survey. More information can be found in Halldórsdóttir et al. (2011) and Prato et al. (2015).

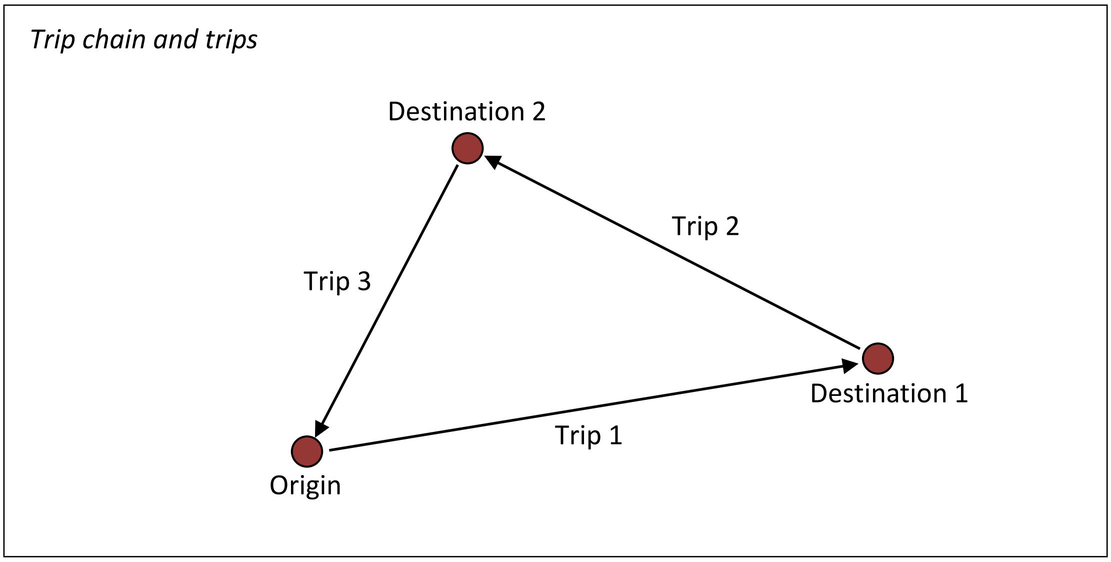


FIGURE 9: DEFINITION OF A TRIP CHAIN AND TRIPS

Figure 10 shows that the share of respondents using each mode is 25% walking, 28% cycling, 35% driving, 5% being driven, and 7% taking public transport. By looking at the distribution of mode share per distance (Figure 11), it can be seen that most travellers walk up to 2-4 km, from where the share starts to decrease with increasing distance. The share of cyclists increases up to 1-2 km, where it remains steady with an average of 30% until 14-16 km, from where it starts to slowly decrease again. The share of car drivers rises slowly with increasing distance and at 6-8 km car drivers start to dominate with approximately 50% of the journeys. The share of car passengers is not particularly high, with less than 10%. The public transport share rises slowly and after 20 km it becomes higher than the bicycle share. Approximately 80% of the journeys are less than 10-12 km (Figure 12) while most observations are from the Greater Copenhagen area (including the city centre), i.e., approximately 80% of the journeys (Figure 13).

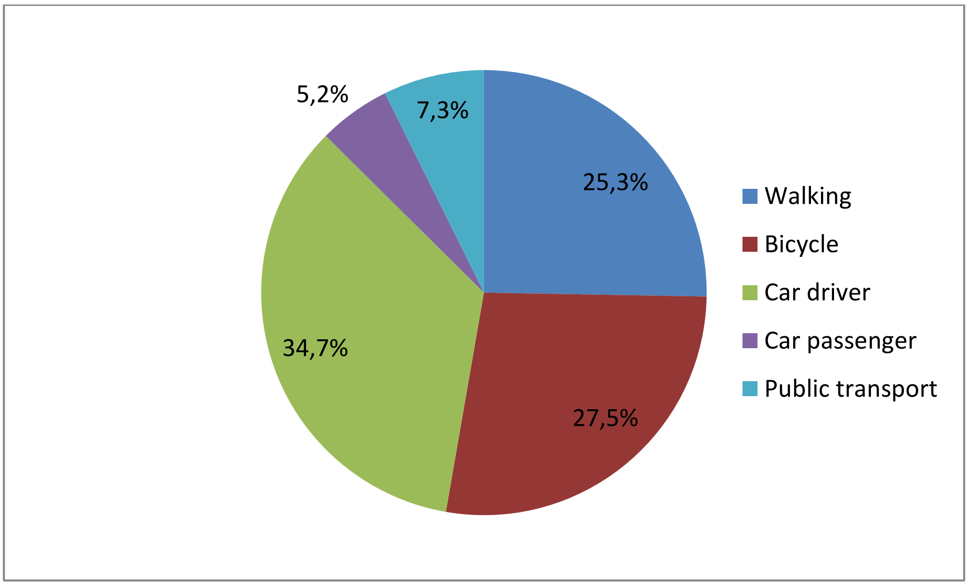


Figure 10: Mode share when travelling short distances in the Copenhagen Region, 2006-2010

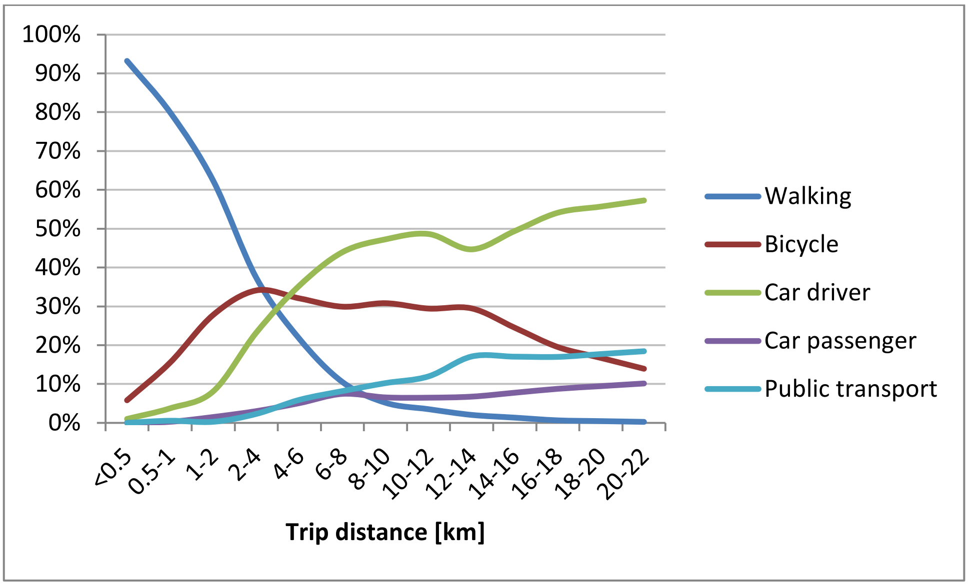


FIGURE 11: DISTRIBUTION OF MODE SHARE PER DISTANCE WHEN TRAVELLING SHORT DISTANCES IN THE COPENHAGEN REGION, 2006-2010

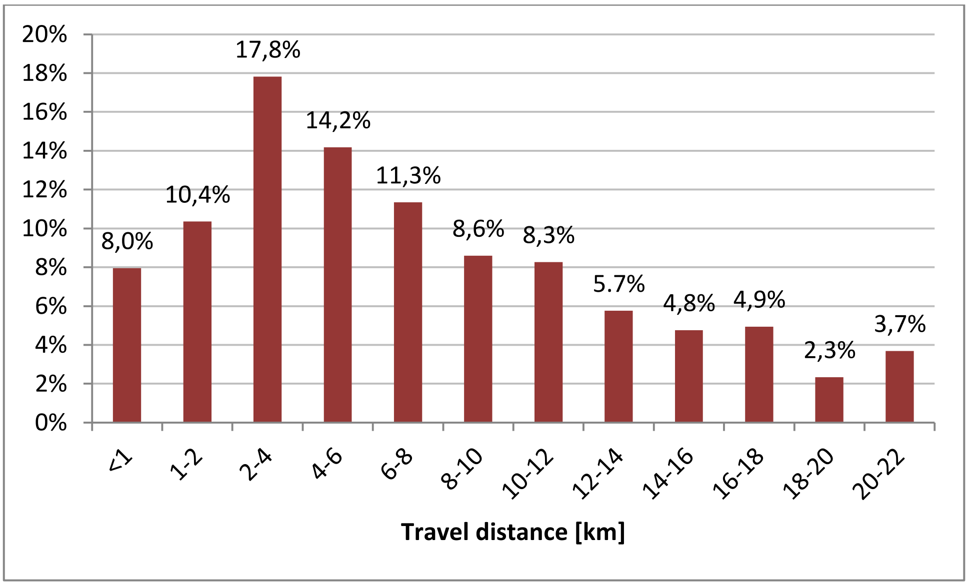


FIGURE 12: NUMBER OF OBSERVATIONS PER DISTANCE WHEN TRAVELLING SHORT DISTANCES IN THE COPENHAGEN REGION, 2006-2010

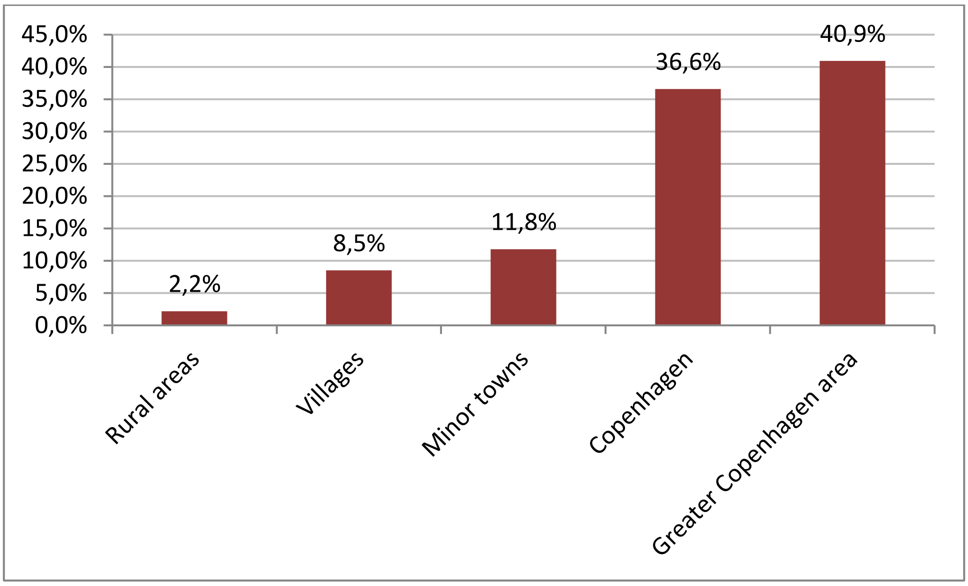


Figure 13: Distribution of journeys per urban area when travelling short distances in the Copenhagen Region, 2006-2010

The average age in the dataset is 48 years (Figure 14), where 47% are male (Figure 15). Figure 16 shows that the share of females is slightly higher in the walking and bicycle journeys. More males are car drivers, while females are frequently car passengers or public transport travellers.

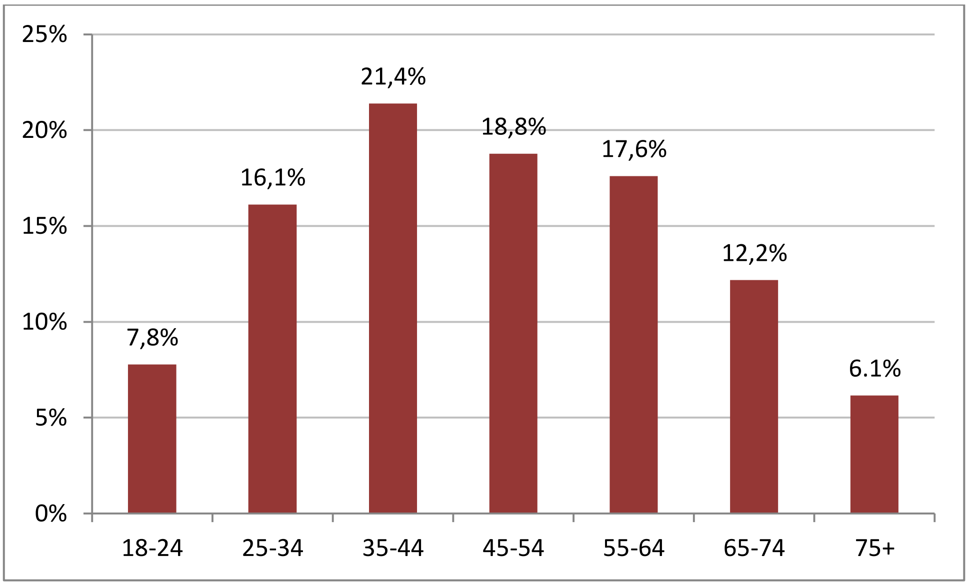


Figure 14: Age group distribution when travelling short distances in the Copenhagen Region, 2006-2010

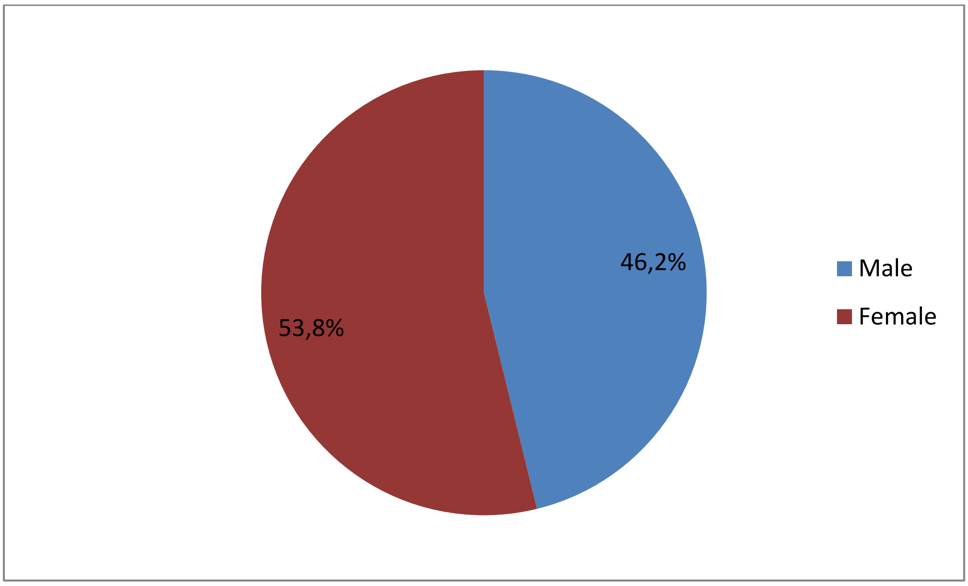


Figure 15: Gender share when travelling short distances in the Copenhagen Region, 2006-2010

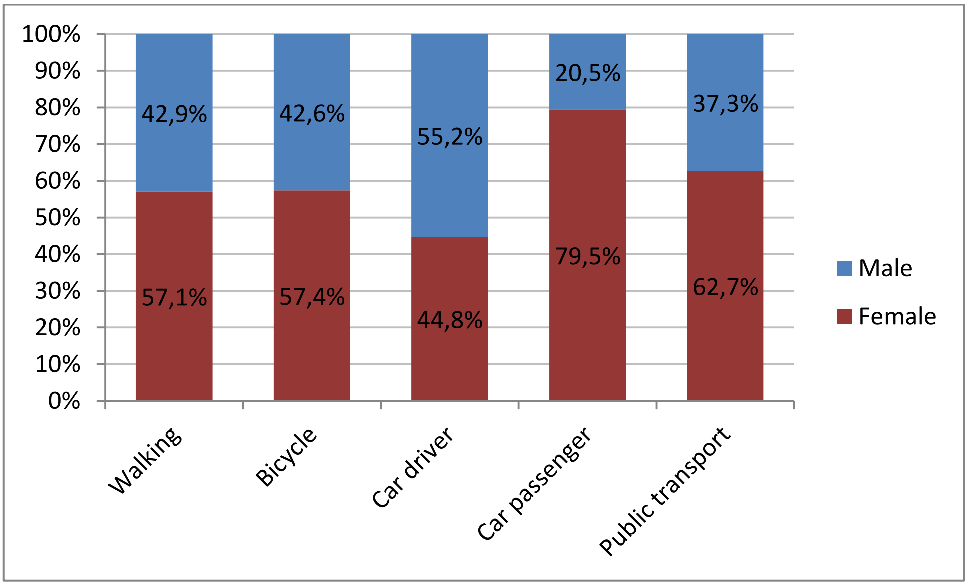


Figure 16: Share of gender and choice of transport mode when travelling short distances in the Copenhagen Region, 2006-2010

Figure 17 illustrates the share of main occupation and the choice of transport mode when travelling short distances in the Copenhagen Region, while Table 1 lists the number of observations in each category. The homemaker occupation category only had two observations, so it was not included in the analysis. The share of non-motorised- and motorised alternatives is more or less equal for employed, self-employed, and retired respondents, or approximately 50%. Un-employed respondents have a share of 64% choosing non-motorised transport alternatives, while students have the highest share at 70%. Self-employed respondents have the highest share of car drivers at 48%, while students have the highest bicycle share at 45%.

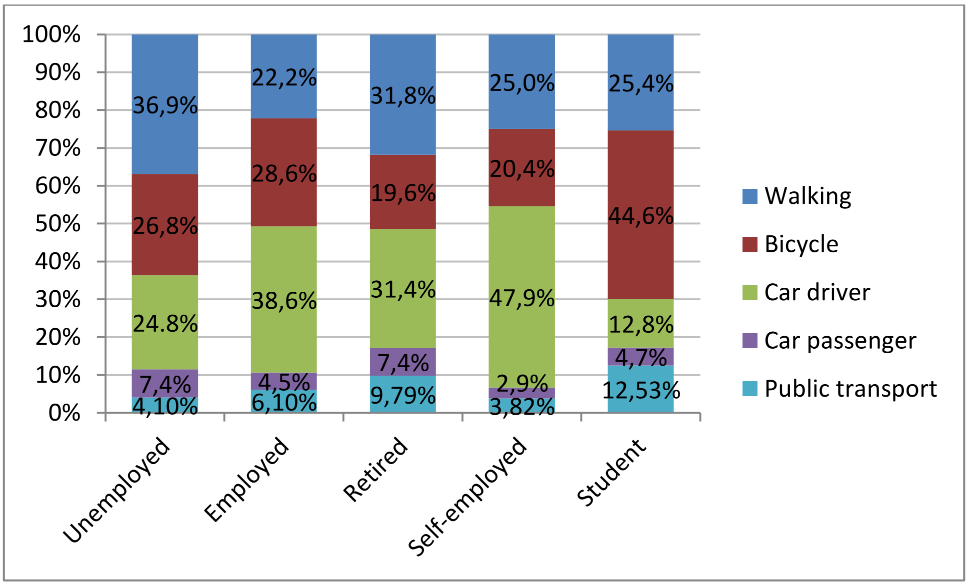


Figure 17: Share of main occupation and choice of transport mode when travelling short distances in the Copenhagen Region, 2006-2010

Table 1: Number of observations per main occupation and the choice of transport mode when travelling short distances in the Copenhagen Region, 2006-2010

	Unemployed	Employed	Retired	Self-employed	Student
Walking	135	1,465	802	131	243
Bicycle	98	1,888	495	107	427
Car driver	91	2,553	793	251	123
Car passenger	27	299	187	15	45
Public transport	15	403	247	20	120
	366	**6,608**	**2,524**	**524**	**958**

Figure 18 shows the share of trip purpose and the choice of transport mode when travelling short distances in the Copenhagen Region, while Table 2 lists the number of observations per trip purpose and the choice of transport mode. It can be seen from Figure 18 that many bicycle journeys are commuting trip chains, or approximately 50%, when the trip purpose is commuting or commuting in combination with other. The bicycle share is also approximately 50% when trip chains are business related. The share of bicycle is a great deal lower, or approximately 20%, when the trip purpose is shopping, escorting, or leisure. The share of car drivers is dominating when escorting, or 60%.

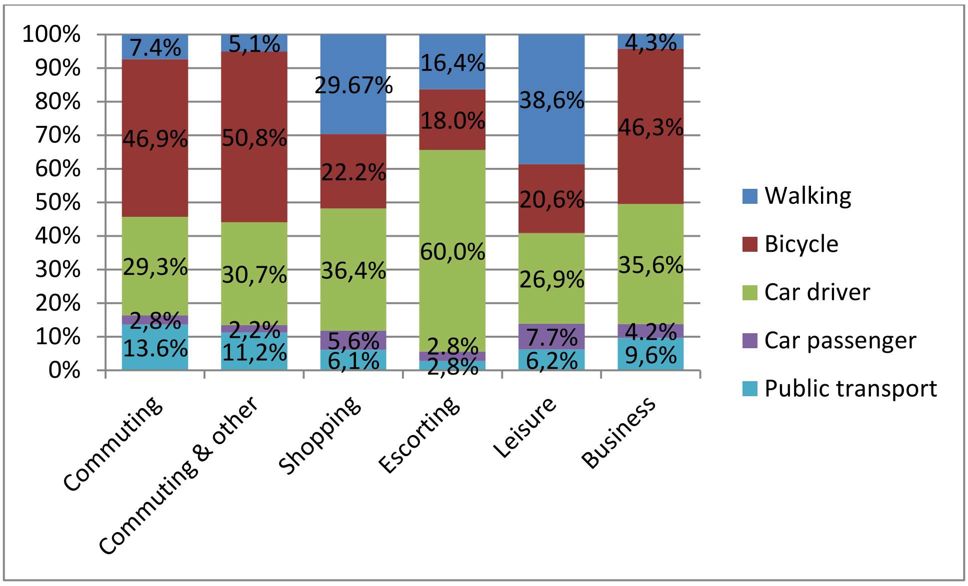


Figure 18: Share of trip purpose and choice of transport mode when travelling short distances in the Copenhagen Region, 2006-2010

Table 2: Number of observations per trip purpose and the choice of transport mode when travelling short distances in the Copenhagen Region, 2006-2010

	Commuting	Commuting & other	Shopping	Escorting	Leisure	Business
Walking	124	40	1,060	212	1,332	8
Bicycle	792	399	791	234	712	87
Car driver	495	241	1,301	777	930	67
Car passenger	47	17	201	37	265	8
Public transport	230	88	219	36	214	18
	1,688	**785**	**3,572**	**1,296**	**3,453**	**188**

3.2.3 Travel behaviour when accessing passenger trains in the Copenhagen Region

Halldórsdóttir et al. (2015a) investigated the choice of access and egress transport mode in the Copenhagen Region during the period 2006-2011, where passenger trains were the main transport mode. The dataset used for the analysis was extracted from the TU-survey, including the new questionnaire with mapping of the public transport trips as described in section 3.1. The home-end and the activity-end were analysed separately, to account for the different availability of transport modes. The data sample included 1,743 individuals and 2,921 trips, at the home-end, along with 1,909 individuals and 3,658 trips, at the activity-end. Trip chains are not always homebound, which partly explains why there is a different number of individuals and trips at each end. Furthermore, in one trip chain there can be more than one activity. Given the extensive data collected, only selected results are presented here. Further information can be found in Halldórsdóttir et al. (2015a).

In Figure 18 and Figure 19, the share of access and egress mode choice is illustrated, at home-end and the activity-end, respectively. The figures clearly show that the availability of transport mode affects mode choice, with walking being the most chosen transport alternative at both ends. At the home-end, travellers have a higher access to private transport alternatives, thus, being a car driver is only available at the home-end and the share of bicycles is larger. In Figure 20 and Figure 21, the mode choice at different distances is shown, with cut-off at 5 km. At the home-end, the share of travellers walking is dominant for trips less than 1-1.5 km long, but this dominance decreases rapidly with increasing distance. The share of bicycles slowly grows with the largest share at 1 to 2.5 km, at approximately 40% from where it starts to diminish again. The share of motorised transport alternatives grows with increasing distance, with busses being the most dominant transport mode, with approximately 50% of the trips after 3 km. The mode share at the activity-end is very similar to the home-end. The main difference is that travellers walk further and travellers have limited access to private transport alternatives. Consequently, the share of busses is larger at the activity-end and the bicycle share is less than 20%. By looking at the number of observations per distance (Figure 22) it can be seen that most access and egress trips are less than 2 km, or approximately 75-85%.

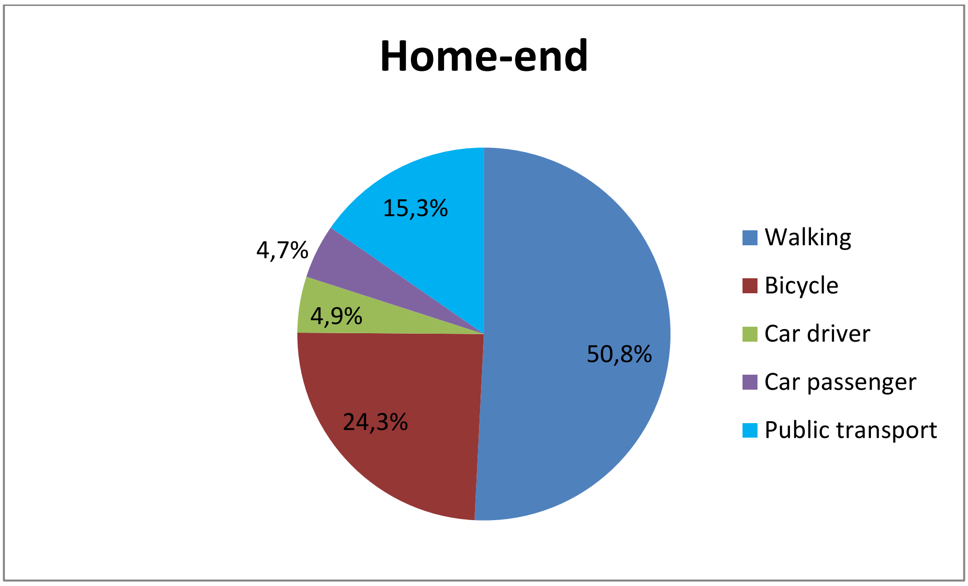


FIGURE 19: MODE SHARE AT THE HOME-END WHEN ACCESSING PASSENGER TRAINS IN THE COPENHAGEN REGION, 2006-2011

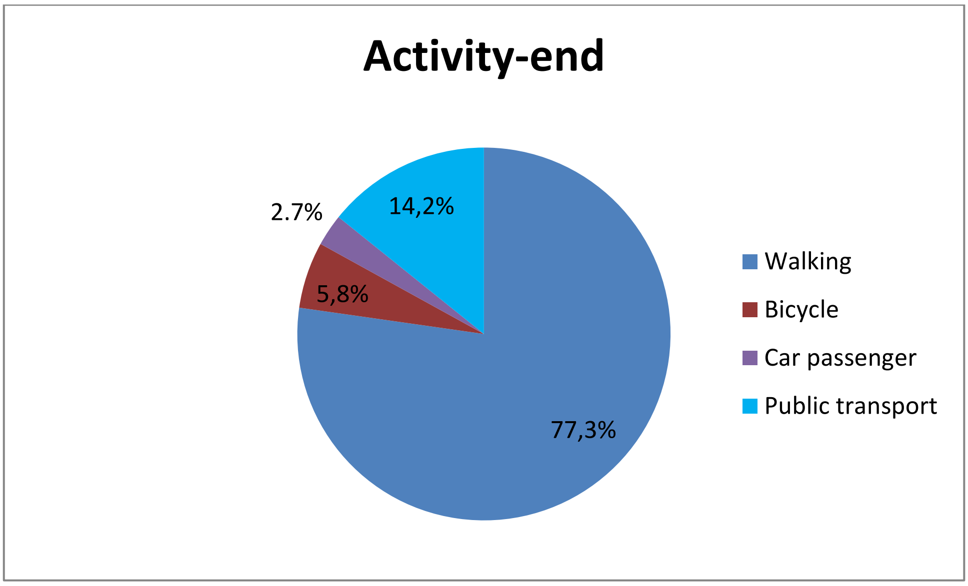


FIGURE 20: MODE SHARE AT THE ACTIVITY-END WHEN ACCESSING PASSENGER TRAINS IN THE COPENHAGEN REGION, 2006-2011

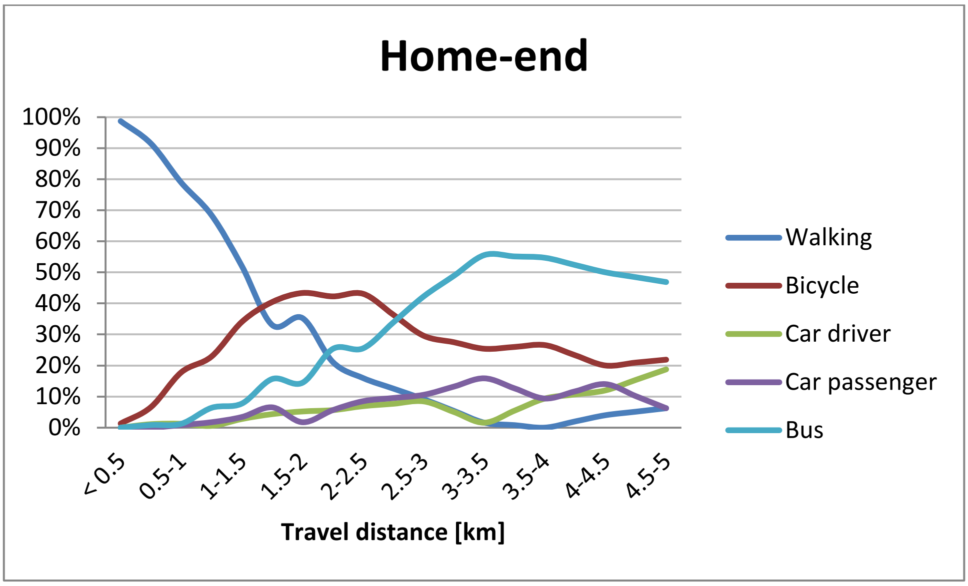


FIGURE 21: DISTRIBUTION OF MODE SHARE PER DISTANCE AT HOME-END WHEN ACCESSING PASSENGER TRAINS IN THE COPENHAGEN REGION, 2006-2011

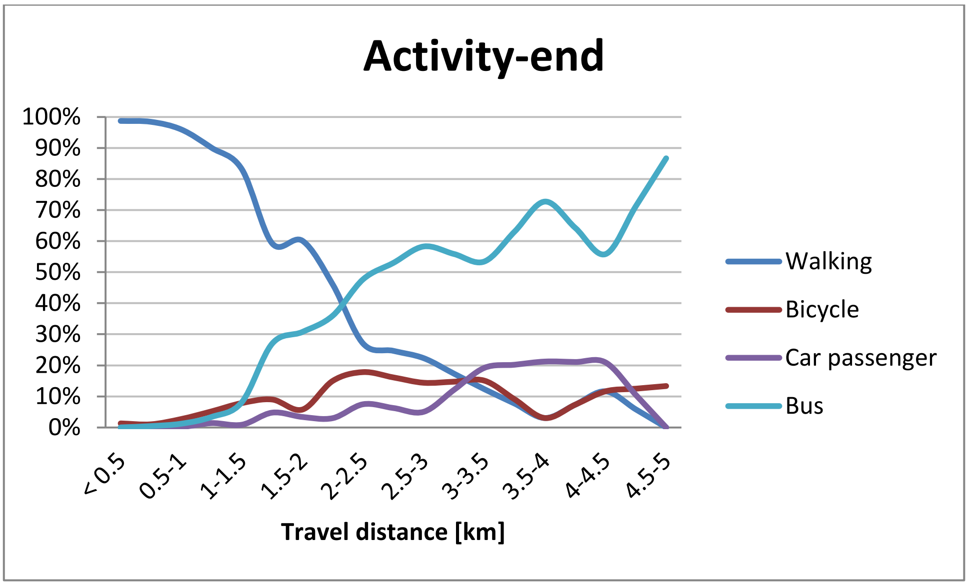


FIGURE 22: DISTRIBUTION OF MODE SHARE PER DISTANCE AT ACTIVITY-END WHEN ACCESSING PASSENGER TRAINS IN THE COPENHAGEN REGION, 2006-2011

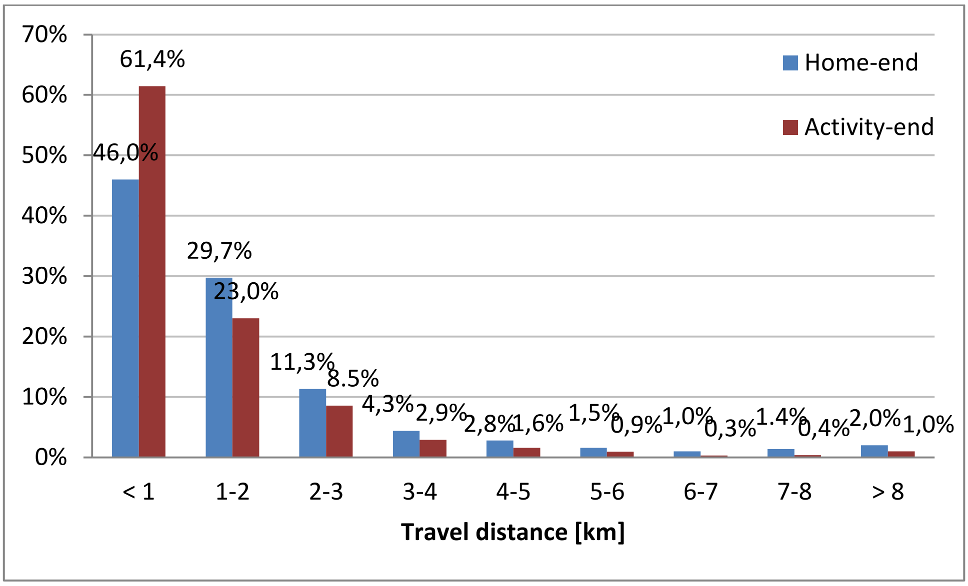


FIGURE 23: NUMBER OF OBSERVATIONS PER DISTANCE WHEN ACCESSING PASSENGER TRAINS IN THE COPENHAGEN REGION, 2006-2011

In the dataset, the average age is 39 years (Figure 24), where males constitute 41% of the sample (Figure 25). By looking at the share of gender and choice of access and egress mode (Figure 26 and Figure 27) there is no distinct difference. The share of females is slightly larger in all transport modes, as the sample contains more females.

In Figure 28 and Figure 29, the share of main occupation and the access and egress mode choice, at the home- and activity-end respectively, is illustrated. Table 3 and Table 4 list the number of observations in each category, at the home- and activity-end respectively. At the home-end, the share of walking is dominant with unemployed having the largest share of 70% and self-employed have the smallest at approximately 45%. Students, employed, and self-employed travellers have the largest bicycle share, or approximately 20-30%. Self-employed also have the largest share of car drivers at 13%, while students and retired travellers have the largest share of busses, approximately 20%. At the activity-end, there is little difference in mode share between the occupation groups. The share of walking is the dominant transport mode with approximately 75-80%.

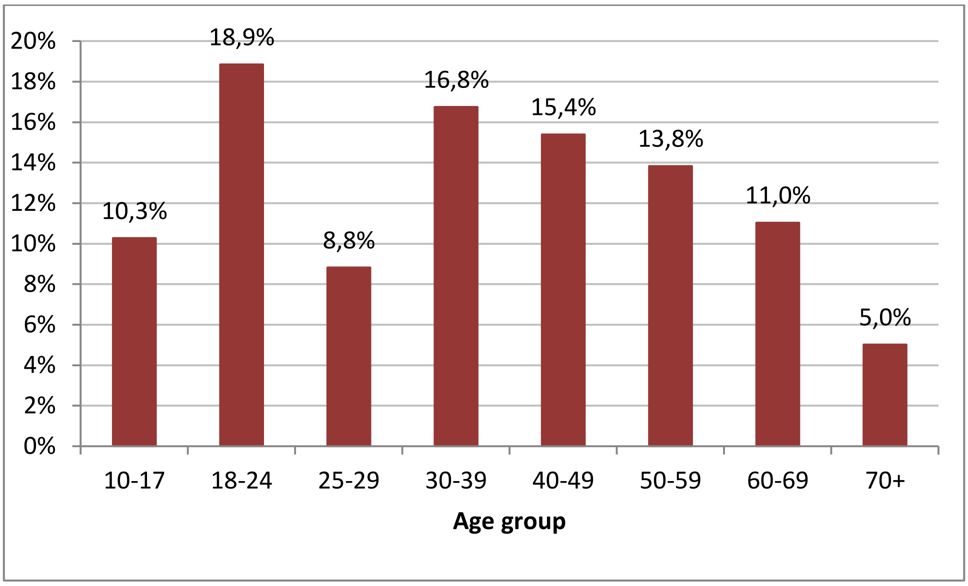


FIGURE 24: AGE GROUP DISTRIBUTION WHEN ACCESSING PASSENGER TRAINS IN THE COPENHAGEN REGION, 2006-2011

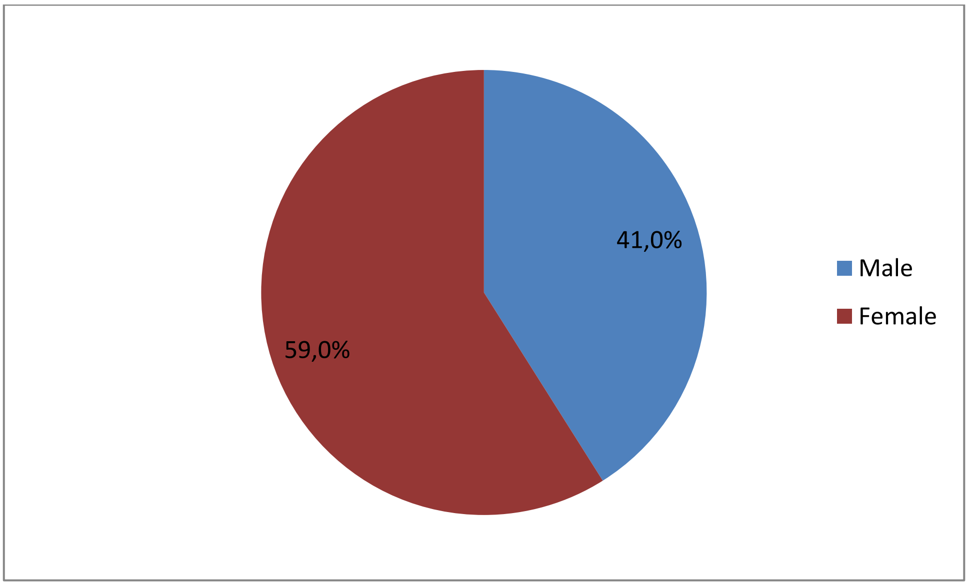


FIGURE 25: GENDER SHARE WHEN ACCESSING PASSENGER TRAINS IN THE COPENHAGEN REGION, 2006-2011

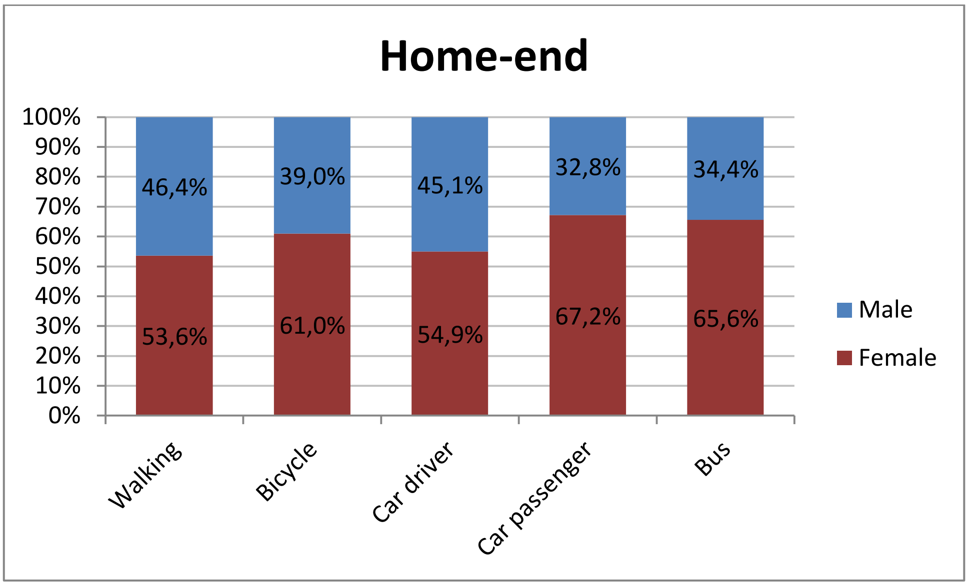


FIGURE 26: SHARE OF GENDER AND CHOICE OF ACCESS AND EGRESS MODE AT HOME-END WHEN ACCESSING PASSENGER TRAINS IN THE COPENHAGEN REGION, 2006-2011

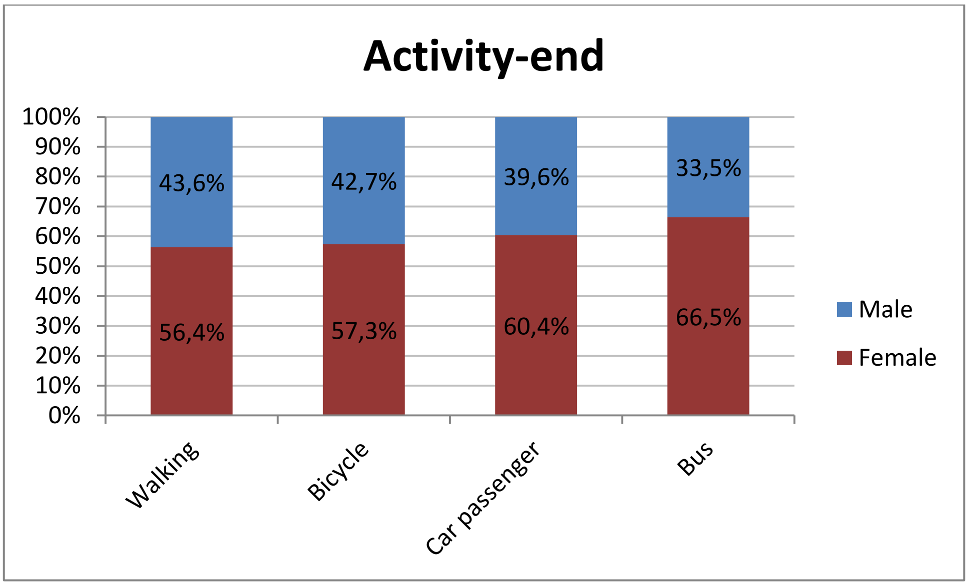


FIGURE 27: SHARE OF GENDER AND CHOICE OF ACCESS AND EGRESS MODE AT ACTIVITY-END WHEN ACCESSING PASSENGER TRAINS IN THE COPENHAGEN REGION, 2006-2011

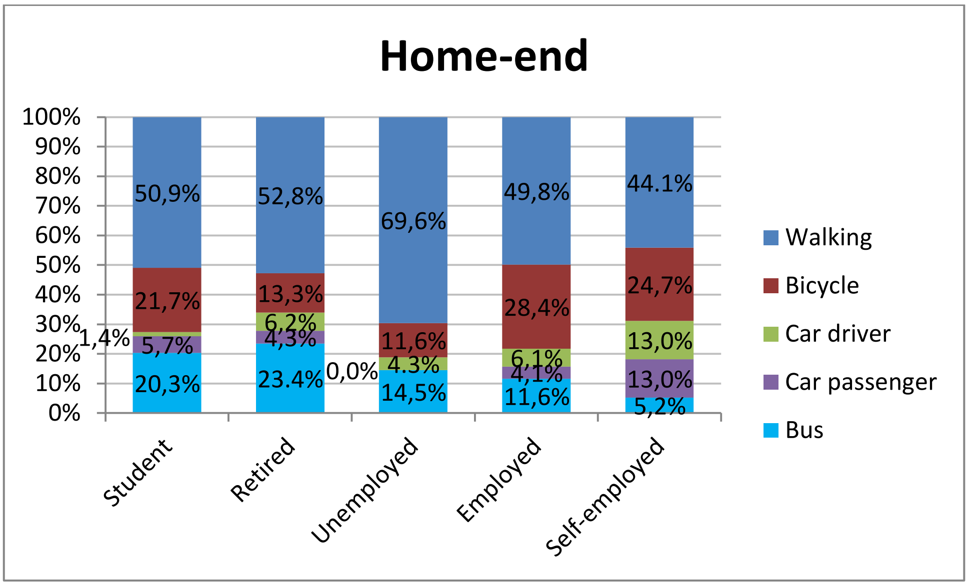


FIGURE 28: SHARE OF MAIN OCCUPATION AND CHOICE OF ACCESS AND EGRESS MODE AT HOME-END WHEN ACCESSING PASSENGER TRAINS IN THE COPENHAGEN REGION, 2006-2011

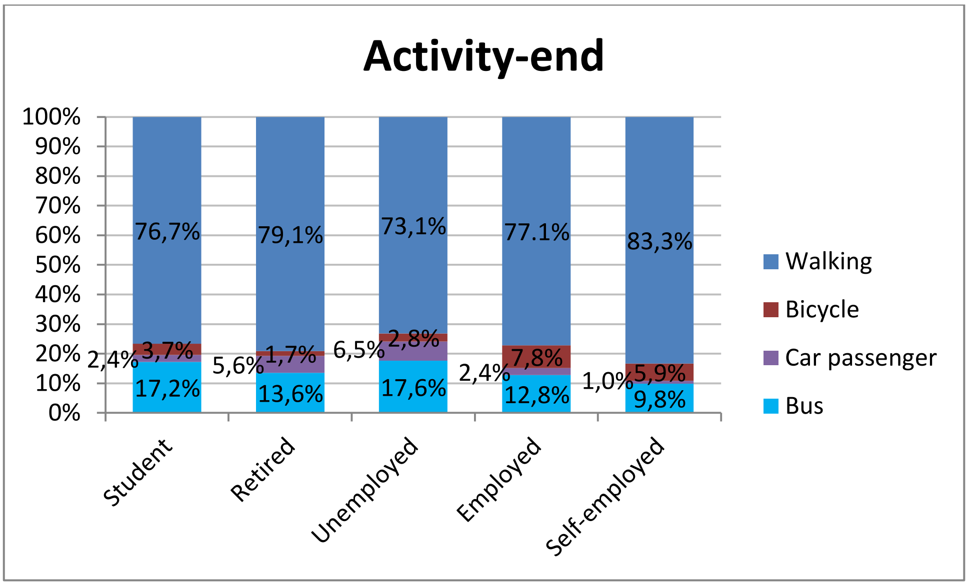


FIGURE 29: SHARE OF MAIN OCCUPATION AND CHOICE OF ACCESS AND EGRESS MODE AT ACTIVITY-END WHEN ACCESSING PASSENGER TRAINS IN THE COPENHAGEN REGION, 2006-2011

TABLE 3: NUMBER OF OBSERVATIONS PER MAIN OCCUPATION AND THE CHOICE OF ACCESS AND EGRESS MODE AT HOME-END WHEN ACCESSING PASSENGER TRAINS IN THE COPENHAGEN REGION, 2006-2011

	Student	Retired	Unemployed	Employed	Self-employed
Walking	432	171	48	799	34
Bicycle	184	43	8	456	19
Car driver	12	20	3	97	10
Car passenger	48	14	0	65	10
Bus	172	76	10	186	4
	848	**324**	**69**	**1,603**	**77**

TABLE 4: NUMBER OF OBSERVATIONS PER MAIN OCCUPATION AND THE CHOICE OF ACCESS AND EGRESS MODE AT ACTIVITY-END WHEN ACCESSING PASSENGER TRAINS IN THE COPENHAGEN REGION, 2006-2011

	Student	Retired	Unemployed	Employed	Self-employed
Walking	839	280	79	1,543	85
Bicycle	41	6	3	155	6
Car passenger	26	20	7	47	1
Bus	188	48	19	255	10
	1,094	**354**	**108**	**2,000**	**102**

In Figure 30 and Figure 31, the share of trip purpose and the access and egress mode choice are illustrated, while Table 5 and Table 6 list the number of observations per trip purpose and the access and egress mode choice. There is some correlation between the choice of access and egress mode and the trip purpose. At the home-end, walking has the largest share with approximately 45-70%. It can also be seen that many bicycle trips are work and study related, or approximately 25-30%. Busses are also mostly used in study trips, with a share of 26%. At the activity-end, the share of walking is dominant with approximately 70-85%. The bus share is still the largest in study trips, at 24%. The share of bicycles is a good deal smaller at the activity-end, with errand trips having the largest share at 12%.

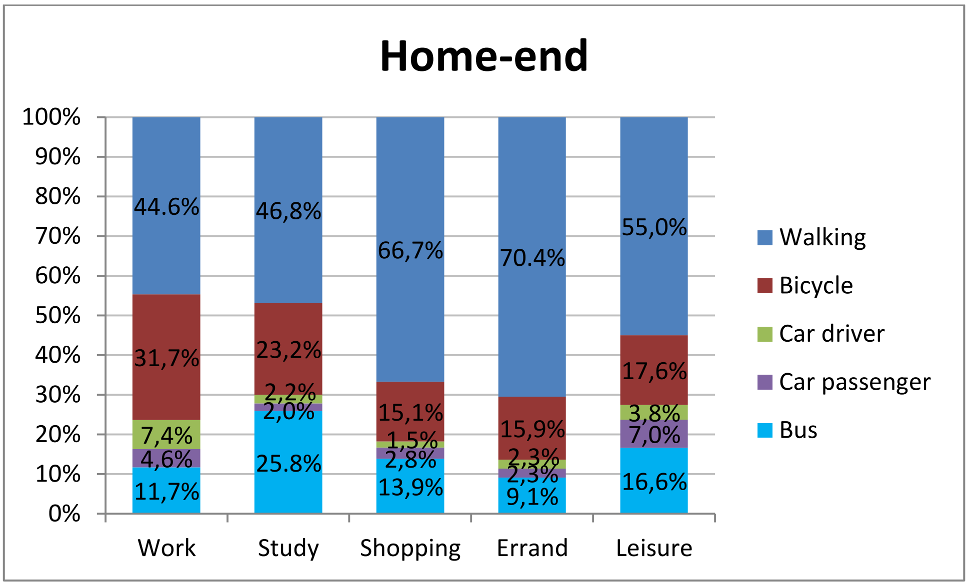


FIGURE 30: SHARE OF TRIP PURPOSE AND CHOICE OF ACCESS AND EGRESS MODE AT HOME-END WHEN ACCESSING PASSENGER TRAINS IN THE COPENHAGEN REGION, 2006-2011

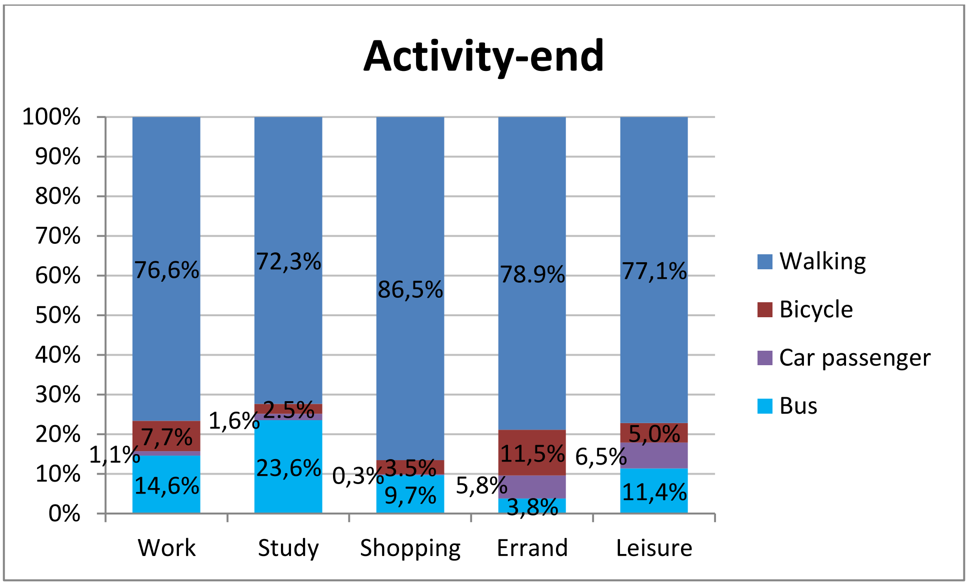


FIGURE 31: SHARE OF TRIP PURPOSE AND CHOICE OF ACCESS AND EGRESS MODE AT ACTIVITY-END WHEN ACCESSING PASSENGER TRAINS IN THE COPENHAGEN REGION, 2006-2011

TABLE 5: NUMBER OF OBSERVATIONS PER TRIP PURPOSE AND THE CHOICE OF ACCESS AND EGRESS MODE AT HOME-END WHEN ACCESSING PASSENGER TRAINS IN THE COPENHAGEN REGION, 2006-2011

	Work	Study	Shopping	Errand	Leisure
Walking	576	192	216	31	469
Bicycle	409	95	49	7	150
Car driver	95	9	5	1	32
Car passenger	59	8	9	1	60
Bus	151	106	45	4	142
	1,290	**410**	**324**	**44**	**853**

TABLE 6: NUMBER OF OBSERVATIONS PER TRIP PURPOSE AND THE CHOICE OF ACCESS AND EGRESS MODE AT ACTIVITY-END WHEN ACCESSING PASSENGER TRAINS IN THE COPENHAGEN REGION, 2006-2011

	Work	Study	Shopping	Errand	Leisure
Walking	1,238	368	339	41	840
Bicycle	124	13	14	6	54
Car passenger	18	8	1	3	71
Bus	236	120	38	2	124
	1,616	**509**	**392**	**52**	**1,089**

3.3 DATA PREPARATION

In this section, the preparation of the data is described and how the level-of-service variables were calculated for each transport alternative. When individuals have chosen to travel to a specific destination, they first have to make a choice on when to travel, which transportation mode to use, and finally they optimise their route between the origin and destination depending on their preferences.

The attributes of the alternatives, within the choice set of each traveller, were calculated using route choice models and simulation methods. The transport alternatives could only be estimated generally, as the travellers do not have as much information on possible mode choices as the researcher. For the mode choice estimation five alternatives, i.e., walking, bicycle, car driver, car passenger, and public transport, were considered in all three papers (i.e., Halldórsdóttir et al., 2011; Prato et al., 2015; and Halldórsdóttir et al., 2015a). In Halldórsdóttir et al. (2011) and Prato et al. (2015) the public transport alternative included busses, trains, and the metro, while in Halldórsdóttir et al. (2015a) the only public transport alternative considered was busses. This is because the paper investigated the choice of access and egress transport mode, when accessing passenger trains, so trains and the metro are already included as the main transport alternative.

In Halldórsdóttir et al. (2011) and Prato et al. (2015) the level-of-service variables were calculated for each trip in the trip chains. Then, the level-of-service variables were summed over the trip chains. Hence, the level-of-service in the overall chain (defined by its main mode) was used for model estimation in both papers. In Halldórsdóttir et al. (2015a) the level-of-service variables were calculated for each access- and egress stage of the passenger train trip.

Subsection 3.3.1 and subsection 3.3.2 describe how the level-of-service variables for the car- and public transport alternatives were calculated using route choice models. Subsection 3.3.3 describes how the level-of-service variables were calculated for the walk and bicycle alternatives, using simulation methods.

3.3.1 Road traffic assignment

In this section, it is described how the level-of-service variables for the car alternatives were calculated. The variables were calculated through assignment procedures available for each period of the day in which the trip was conducted. This allowed for the consideration of congestion conditions similar in average to the ones encountered by the travellers. For the car alternatives, the most important variables that needed to be calculated were the travel times and the travel distance.

Traffic Analyst (Rapidis, 2013) is a transportation planning and a modelling extension for ArcGIS Desktop (ESRI, 2015). A tool within Traffic Analyst, called Road traffic assignment, was used to calculate the level-of-service variables for the car alternatives. The tool can be used to simulate the flow of car traffic through a road network. Using this transportation model it is also possible to calculate the related effects of the simulated traffic flow, such as travel costs, which are described as the resulting average travel distance and average travel time between each pair of zones. Extensive data preparation was needed in order to digitalise the input data and to formulate the model parameters needed to run the assignment. Even though it was not theoretically complicated to prepare all the input data, practically it was very time consuming. The following subsections describe in more detail how the inputs for the assignment procedures were defined and how they were generated.

3.3.1.1 Zone centroids

In transportation models, zones are used to describe the area being modelled, by dividing it into smaller areas. Then, each zone is represented by a zone centroid that is in the centre of the zone (Figure 32). In order to calculate the level-of-service variables required for the mode choice estimation, each location point needed to be digitalised as a zone centroid. Then, each location point corresponds to the different zones and zone centroids. The observed data included coordinates and precise address information, which allowed for the location points to be digitalised in high accuracy. Thus, the routes between the location points could be analysed more accurately and the characteristics of the mode choice depending on the route. Figure 33 shows how a trip and stages were defined, with the corresponding location points. In Halldórsdóttir et al. (2011) and Prato et al. (2015) there were two location points in each trip (i.e., an origin and a destination point). Halldórsdóttir et al. (2015a) only focus on the access- and the egress stage of the passenger train trip. Consequently, there were four location points that needed to be digitalised (i.e., an origin and a destination point, plus the location of the first- and the last station in the train trip).

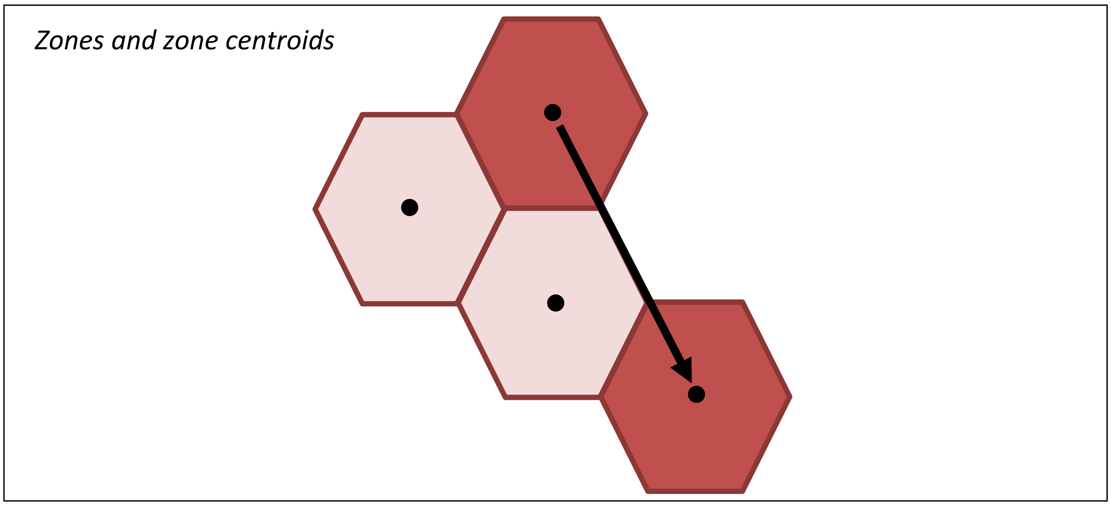


FIGURE 32: DEFINITION OF ZONES AND ZONE CENTROIDS

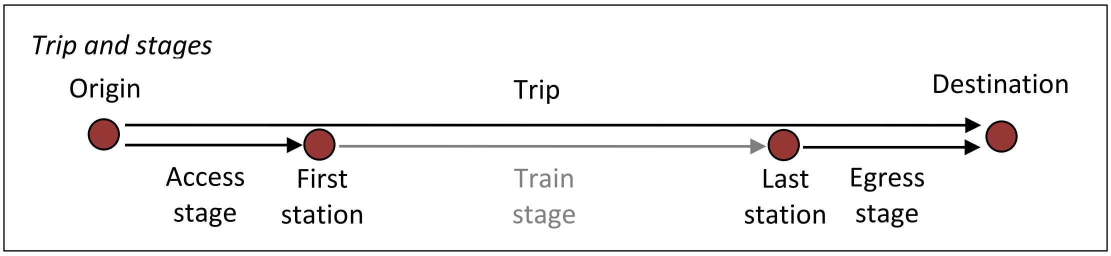


FIGURE 33: DEFINITION OF TRIP AND STAGES

3.3.1.2 TRIP MATRICES

After the location points had been digitalised as zone centroids, trip matrices were constructed. Trip matrices describe the travel patterns, i.e., the number of trips between each zone pair. In Halldórsdóttir et al. (2011) and Prato et al. (2015) there were multiple trips per observation (Figure 33) and only one trip chain (Figure 9), while there were two stages (i.e., the access- and egress stage) in the Halldórsdóttir et al. (2015a). The train stage was excluded from the calculation in Halldórsdóttir et al. (2015a), as the paper only focused on the feeder modes. The assignment procedures are available in each period of the day in which each trip was conducted. By dividing the trips into time periods, congestion conditions, similar in average to the ones encountered by the travellers, could be included. The road traffic assignment was divided into seven time periods and each trip was assigned to the appropriate time period using the trip's departure time. Table 7 lists the time periods and specifies which time intervals were within each period. Additionally, each trip was assigned a category ID, depending on the purpose of the trip. The trips were divided into three categories, i.e., commuting trips, business trips, and all other trips. Each category was then assigned a set of route choice parameters.

TABLE 7: TIME PERIODS FOR THE ROAD TRAFFIC ASSIGNMENT

Time period ID	Time interval	Description
1	21-05	Evening/night non-peak hours
2	05-07	Morning non-peak hours
3	07-08	Morning peak hours
4	08-09	Morning peak hours
5	09-15	Daytime non-peak hours
6	15-18	Afternoon peak hours
7	18-21	Evening non-peak hours

3.3.1.3 ROAD NETWORK

In Halldórsdóttir et al. (2011) and Prato et al. (2015), the NAVTEQ road network (NAVTEQ, 2010) was used to calculate the level-of-service variables for the car alternatives. Geometrically, the network has high accuracy and has a very complete detailed coverage and a high number of attributes, which describe travel speed (both free- and queue speed), link type, number of lanes, one-way roads, preload on links, etc. Due to the high number of observations in the papers, only major roads were included to limit calculation time. Thus, the network consisted of 26,154 possibly bidirectional links, limited to the Copenhagen Region. The level of detail was still quite high as can be seen in Figure 34.

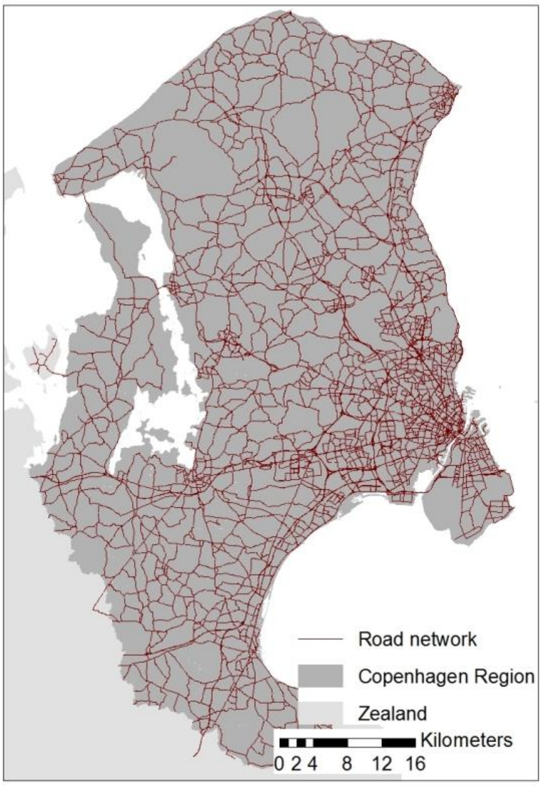


FIGURE 34: THE NAVTEQ NETWORK USED IN HALLDÓRSDÓTTIR ET AL. (2011) AND PRATO ET AL. (2015). LEFT: OVERVIEW OF DENMARK, RIGHT: ZOOM OF THE COPENHAGEN REGION

The level of geometric detail was very important in Halldórsdóttir et al. (2015a), when calculating the level-of-service variables, as the travel distances were very short. Accordingly, a full version of the NAVTEQ road network was used, including 158,443 possibly bidirectional links (see Figure 35). The study area was the Copenhagen Region, but the road network covers the eastern part of Denmark (Zealand). The number of observations was not as high as in the other two studies, thus including the full version of the network was not troublesome in relation to calculation time.

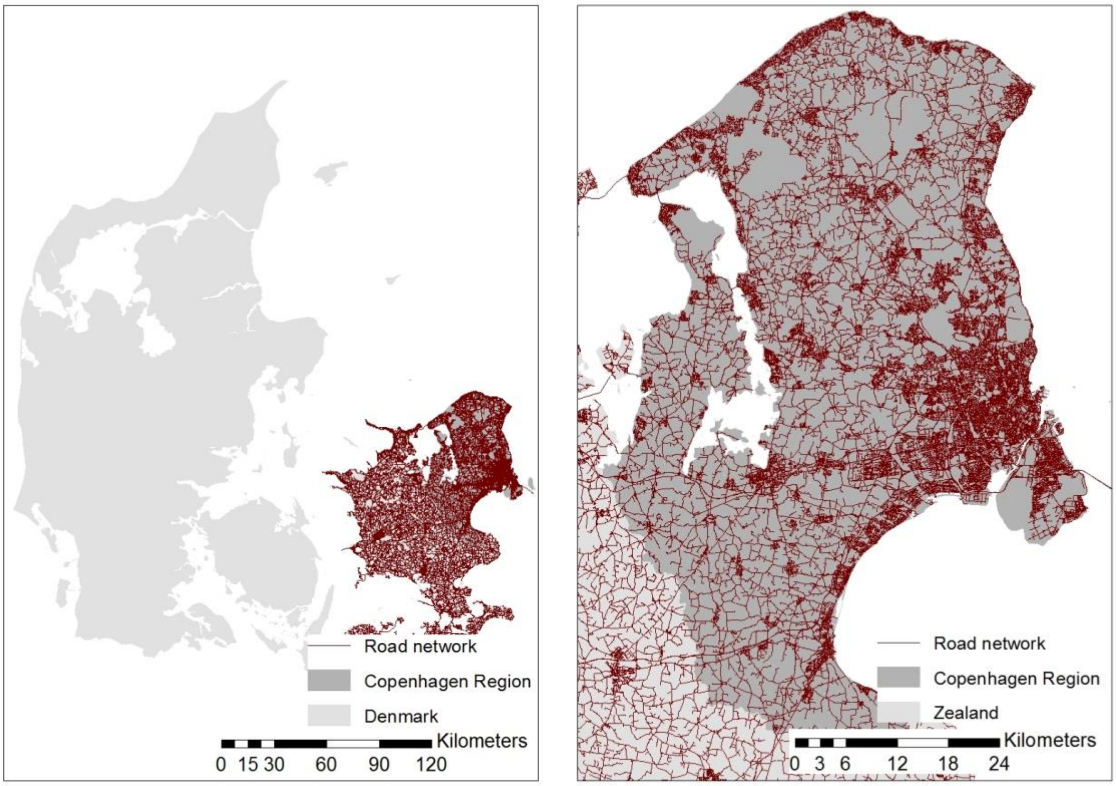


FIGURE 35: THE NAVTEQ NETWORK USED IN HALLDÓRSDÓTTIR ET AL. (2015A). LEFT: OVERVIEW OF DENMARK, RIGHT: ZOOM OF THE COPENHAGEN REGION

3.3.1.4 ZONE CONNECTORS

In order to assign the trips onto the road network, zone connectors needed to be generated, linking the zone centroids to the road network (Figure 36). The zone connectors were generated with a path search through the road network, connecting the zone centroids to the road network through the nearest node. A highly detailed network was used to generate the connectors, in order to ensure accurate travel distances. The motorways were not included in the network, in order to ensure that the road traffic assignment would use the surrounding roads to connect to the motorways.

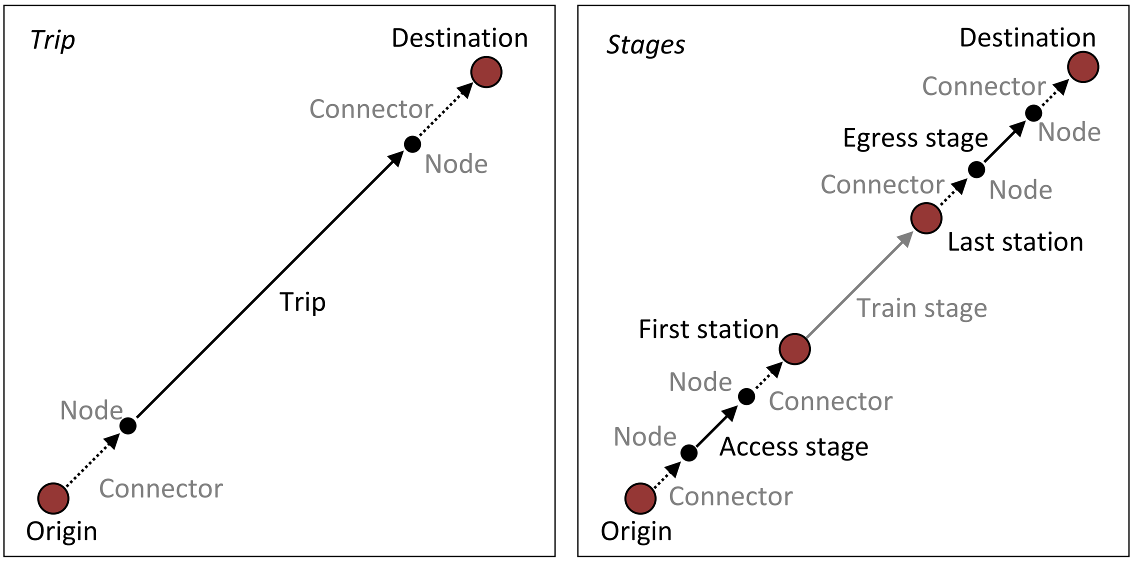


FIGURE 36: DEFINITION OF A ZONE CONNECTOR

After the zone connections were digitalised, the travel time on each connector was calculated. In the early stages of the PhD study it was assumed that the zone connectors were only used to connect the location points to the nearest nodes. Thus, a very low speed was chosen for the car connectors in Halldórsdóttir et al. (2011) and Prato et al. (2015). The travel time was then calculated as a function of length.

In the later stages of the PhD study, the travel speed for the zone connectors was extended with a piecewise linear function in Halldórsdóttir et al. (2015a). The travel speeds were analysed from the TU-survey, resulting in the following speed function for the car alternatives (Anderson, 2013):

$$\text{Speed}\left(\text{km/h}\right) = 15 + 25 \cdot \text{Length} / 8000\ , \qquad (1)$$

where the length is given in meters. The speed function shows that the greater the travel distance is, the faster travellers drive.

3.3.1.5 ROUTE CHOICE ASSIGNMENT

The final step of the data preparation, for the road traffic assignment, was to run the road traffic assignment. By running the assignment, it is possible to combine the modelling of road congestion effects with stochastic simulation of route choice parameters and travel costs, through the use of iterative calculation methods. The level of detail of the input data was quite high. The road network contained information on traffic load on each link in the different time periods, which enabled calculating the optimal route depending on the congestion restraints in the different periods. An all-or-nothing assignment was performed, by using deterministic route choice parameters and only running one iteration. Afterwards, the optimal travel route through the network was generated, for each zone pair, and the resulting outputs calculated. The car travel time included the free flow travel time, plus the added travel time due to congestion.

3.3.2 PUBLIC TRANSPORT ASSIGNMENT

This section describes how the level-of-service variables for the public transport alternatives were calculated. Assignment procedures, available for each period of the day, were used to calculate the variables. For the public transport alternatives, the most important variables that needed to be calculated were the travel distance and the different travel time parameters, e.g., the waiting time, the access- and egress time, and the in-vehicle time.

A tool within Traffic Analyst, called Public transport assignment, was used to calculate the variables for the public transport alternatives. The tool simulates the traffic flow of public transport passengers. The input data for the public transport assignment was constructed in the same way as for the road traffic assignment, with few alterations. In the following sub-sections, the alterations are described in more detail.

3.3.2.1 ZONE CENTROIDS AND TRIP MATRICES

The zones and zone centroids were constructed in the same way as for the road traffic assignment. The trip matrices were also defined in the same way, however with minor alterations. There were more detailed time periods available for the public transport assignment. Therefore, the assignment was divided into ten time periods, instead of seven. Table 8 lists the time periods for the public transport assignment and specifies which time intervals were within each period.

TABLE 8: TIME PERIODS FOR THE PUBLIC TRANSPORT ASSIGNMENT

Time period ID	Time interval	Description
1	05-06	Morning non-peak hours
2	06-07	Morning non-peak hours
3	07-08	Morning peak hours
4	08-09	Morning peak hours
5	09-15	Daytime non-peak hours
6	15-16	Afternoon peak hours
7	16-17	Afternoon peak hours
8	17-18	Afternoon peak hours
9	18-21	Evening non-peak hours
10	21-05	Night non-peak hours

3.3.2.2 PUBLIC TRANSPORT NETWORK

The route choice assignment was implemented using a detailed public transport network. The network used, in Halldórsdóttir et al. (2011), Prato et al. (2015), and Halldórsdóttir et al. (2015a), was the National Transport Model (NTM) network (Rich et al., 2010). The network is based on complete timetable information available, without any aggregation or simplification, containing, for example, a combination of buses, trains, and other public transport modes. The network has high detail and covers all of Denmark (Figure 37). In this PhD thesis, the Copenhagen Region was only investigated. In Halldórsdóttir et al. (2015a), the only relevant public transport alternative was bus, thus the train lines were excluded from the public transport network for the route choice calculations in that paper.

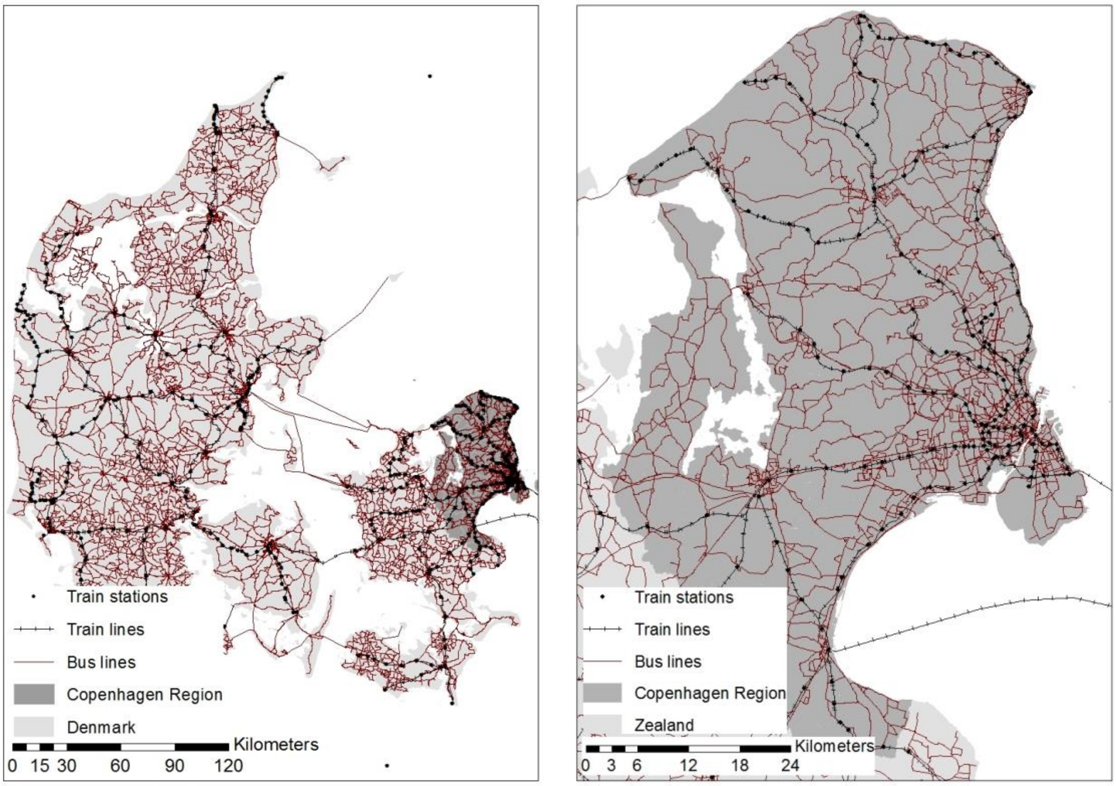


FIGURE 37: NTM PUBLIC TRANSPORT NETWORK. LEFT: OVERVIEW OF DENMARK, RIGHT: ZOOM OF THE COPENHAGEN REGION

3.3.2.3 PUBLIC CONNECTORS

The path generation program, described in subsection 3.3.1.4, was used to create the public connectors. The zone centroids were connected to the nearest stops or stations, as opposed to the nearest nodes like in the road traffic assignment. The public connectors signified the stage from a location point to the nearest bus stops or train stations. Travellers are often prepared to travel further to a stop or a station, in order to get better service, e.g., higher frequency or more connections. To ensure that enough relevant connection points were considered in the assignment procedure, the path generation program searched for the ten nearest stops/stations. To minimise calculation time, any unnecessary connectors, that had a travel distance larger than 5 km, were removed. The assignment procedure selects the best route through the public transport network and as a result automatically excludes any unfeasible connectors. In Halldórsdóttir et al. (2015a) the access- and egress transport modes to passenger trains were investigated, so the only public transport alternative was bus. Accordingly, in that paper the zone centroids were only connected to the nearest bus stops.

When running the public transport assignment it was assumed that walking was the only alternative on the public connectors. It was assumed that if travellers would decide to cycle, they would most likely choose to cycle all the way to the access point as the travel distances were short. An average walking speed was applied to the public connectors in Halldórsdóttir et al. (2011) and Prato et al. (2015) and the travel time was calculated as a function of length. In Halldórsdóttir et al. (2015a), the travel speed on the connectors was extended with a piecewise linear function, analysed in the TU-survey. The speed function for walking was (Anderson, 2013):

$$\text{Speed}\left(\text{km/h}\right) = 4 + 4 \cdot \text{Length} / 8000\,, \tag{2}$$

where the length was specified in meters. The speed function shows that there seems to be a trend that only the more fit travellers walk the long distances, so the greater the travel distance is, the faster the travellers walk.

3.3.2.4 Route choice assignment

After all the inputs had been generated, the route choice assignment was executed. The optimal travel route through the network was calculated, for each zone pair, by minimising travel cost (travel time and distance) for each traveller. Before initialising the calculations, the number of launches and iterations were defined. The number of launches controls how often traffic is placed in the transport system. Since the number of travellers in each zone pairs was one, the number of launches was also set to one. The number of iterations was set to ten, so the output was the average travel time from all ten iterations.

3.3.3 Shortest path generation

In this section, it is described how the level-of-service variables for walking and bicycle were calculated. Simulation methods were used to calculate the attributes of the non-motorised alternatives within the choice set of each traveller. The most important variables that needed to be calculated, for these two modes, were the travel distance and consequently the travel time. The following subsections describe in more detail the transport network and the simulation method used in the calculations.

3.3.3.1 Bicycle and pedestrian network

A complete bicycle network was not available at the time of the analysis. Thus, in Halldórsdóttir et al. (2011), Prato et al. (2015), and Halldórsdóttir et al. (2015a), a preliminary network was used for the route calculations for the walking- and bicycle alternatives. The network used is called TOP10DK (Kort & Matrikelstyrelsen, 2001) and is a very detailed geographical network covering all of Denmark, containing the entire road network and close to all paths.

Two major advantages of the network are that the registration of the entire network follows one base instruction and that the network is topologically coherent between municipalities. The network however does not contain information on bicycle paths along roads. Paths along roads are only digitalised if they run in their own track and are at least ten meters from the road network. Since the majority of bicycle paths along roads are within ten meters from the road network, there are virtually no bicycle paths digitalised along the road network. Thus, for the analysis, the road network was included in the bicycle and pedestrian network. Bicycle or pedestrian paths along major roads are usually in their own tracks or along smaller roads parallel to the major roads. Motorways or large traffic roads were removed from the network, as it is illegal to cycle on those roads. In addition, it was not possible to distinguish between bicycle paths and other small paths that are for example reserved for pedestrians. Generally, this does not cause a problem since the bicycle- and pedestrian paths are positioned along each other. There are nonetheless some exceptions, for example where in some parks cycling is not allowed. The preliminary bicycle and pedestrian network can be seen in Figure 38. At a later stage of the PhD study, a more complete bicycle network was constructed (see further section 8.1.3).

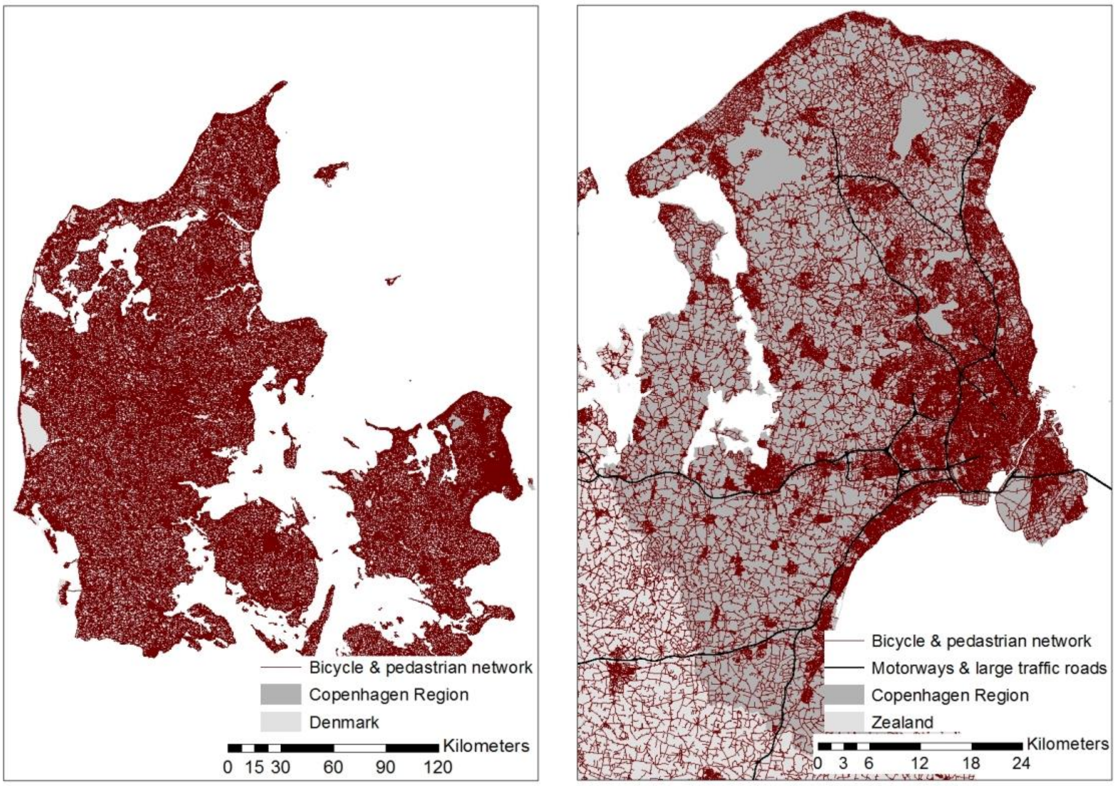


FIGURE 38: TOP10DK NETWORK. LEFT: OVERVIEW OF DENMARK, RIGHT: ZOOM OF THE COPENHAGEN REGION

3.3.3.2 Route choice assignment

The travel distance for the walk- and bicycle alternative was calculated, for each traveller, with a shortest path simulation method, using the route analyst tool within the ArcGIS desktop extension Network Analyst. The route analyst applies a Dijkstra's algorithm to find the optimal route, from one location to another.

In Halldórsdóttir et al. (2011), Prato et al. (2015), and Halldórsdóttir et al. (2015a), length was used as the cost attribute. It is known that cyclists choose their route depending on other attributes than distance, although cyclists are mostly affected by the route length (Menghini et al., 2010). Other negative attributes are slope, turn frequency, intersection control, and bicycle facility types (Hood et al., 2011; Broach et al., 2012). In addition, cyclists' route choice also appears to be influenced by traffic volumes (Broach et al., 2012). Analysing these parameters required more detailed behavioural and geographical data, which was not available at the time of the analysis.

There was no available information on bicycle- or pedestrian travel speed on different parts of the network. Thus, it was assumed that path characteristics do not influence the walking- or the bicycle speed. Accordingly, in Halldórsdóttir et al. (2011) and Prato et al. (2015), the travel speed for the two modes was estimated as an average travel speed. The travel time was thus dependent on the travel distance. In Halldórsdóttir et al. (2015a) the travel speed for the walking and bicycle was also extended with a piecewise linear function, as described in subsection 3.3.1.4. The speed function for walking was described in equation (2) and the speed function for bicycles was (Anderson, 2013):

$$\text{Speed}\left(\text{km/h}\right) = 6 + 14 \cdot \text{Length} / 8000 \,, \tag{3}$$

where the length was given in meters. Again, the speed function expresses that the greater the travel distance is, the faster the travellers bicycle. Thus, indicating that only the more fit cyclists travel the long distances.

3.4 Additional information

The following subsections describe additional information, collected for Halldórsdóttir et al. (2015a), on parking facilities at train stations and the possibility to bring bicycles on trains. The policy variables were retrieved by on site investigation and analysts' knowledge of the study area.

3.4.1 Parking facilities at train stations

Car parking availability was defined for each train station as the offer in terms of car parking spaces and average occupancy on the basis of time-of-day. The basic philosophy of Park & Ride is that the car is used on the part of the journey where there is, for example, greater need for flexibility or poor public transport service. The car is then parked at the Park & Ride facility and the traveller changes to a train (or a bus). Public transport then brings the traveller quickly to the destination, thus avoiding traffic congestion and uncertainties about finding a parking space for the car at the activity-end.

There are about 8,000 Park & Ride spaces at train stations in the Copenhagen Region outside the inner city. Two out of three of these parking spaces are used daily, so approximately 5,000 motorists park their car each day and take the train or bus in the metropolitan area. This gives an average occupancy rate of approximately 65% (Anon., 2009). The positions of the Park & Ride facilities were geocoded in GIS, whereby access to the road network and closeness to stations could be calculated. Then the Park & ride availability was defined for each of the parking spaces available at train stations outside the inner city in the Copenhagen Region while considering their average occupancy rate on the basis of statistics provided by the region.

It is estimated that on average 90,000 people a day use a bicycle in combination with another transport mode, generally train or bus (Lindboe et al., 2003), i.e., 18 times more frequently than the motorised Park & Ride. Many policy strategies have thus been implemented to encourage travellers to use the combination of bicycle and public transport, including adding locked- and covered bicycle parking areas at train stations. The availability of locked- and covered bicycle areas was extracted from various data sources and defined for each train station as the offer for different types of options, namely open bicycle racks, covered bicycle racks, and locked bicycle parking places.

3.4.2 Bring bicycles on the train

Travellers are allowed to bring their bicycle for free on the suburban and the local trains in the Copenhagen Region, whilst they have to pay an extra fee for bringing their bicycle on the metro and the regional trains. Accordingly, it was assumed that if a traveller would want to bring their bicycle on the train, he/she would choose the optimal alternative available for each trip. Thus, the availability of different train types at the train stations was considered (i.e., suburban train, local train, regional train, IC train, and metro), at both ends of the trip, and considered if the travellers had the possibility to bring their bicycle along for free, or if they had to pay a fee. In addition, time restrictions were controlled for, since it is prohibited to bring a bicycle on the metro during peak hours, or to embark or disembark with a bicycle at Nørreport station, also during peak hours.

3.5 References

Anderson, M.K. (2013). *Behavioural Models for Route Choice of Passengers in Multimodal Public Transport Networks*. PhD Thesis, DTU Transport, Kgs. Lyngby, Denmark.

Anonymous. 2009. "*Parker & Rejs*." Work notes. Version 1.1. Region Hovedstaden.

Broach, J., Dill, J. and Gliebe, J. (2012). Where do cyclists ride? A route choice model developed with revealed preference GPS data. *Transportation Research Part A*, 46, pp. 1730–1740.

Christensen, L. and Jensen, T.C. (2008). *Korte ture i bil - Kan bilister ændre adfærd til gang eller cykling?* DTU Transport, 2800 Kgs. Lyngby, Denmark.

Christiansen, H. and Skougaard, B.Z. (2015). *Documentation of the Danish National Travel Survey.* Report 10, DTU Transport, Data- and Modelcenter, Kgs. Lyngby, Denmark.

Skougaard, B. Z. and Christiansen, H., (2015). *Transportvaneundersøgelsen - Faktaark om cykeltrafik i Danmark*. [pdf] DTU Transport, Data- and Modelcenter, Kgs. Lyngby. Available at: http://www.modelcenter.transport.www7.sitecore.dtu.dk/Transportvaneundersoegelsen/TU-udgivelser/Faktaark-om-cykeltrafik-i-Danmark-2014 [Accessed 05 June 2015].

ESRI (2015). *ArcGIS Desktop: Release 10.3.1*. Environmental Systems Research Institute. Redlands, California.

Halldórsdóttir, K., Christensen, L., Jensen, T.C. and Prato, C.G. (2011). Modelling mode choice in short trips - Shifting from car to bicycle. Presented at the *39th European Transport Conference*, 10th – 12th October, Glasgow, Scotland, UK.

Halldórsdóttir, K., Nielsen, O.A. and Prato, C.G. (2015). Home-end and activity-end preferences for access to and egress from train stations in the Copenhagen Region. Submitted to *International Journal of Sustainable Transportation* in September 2015.

Hood, J., Sall, E. and Charlton, B. (2011). A GPS-based bicycle route choice model for San Francisco, California. *Transportation Letters: The International Journal of Transport Research*, 3, pp. 63–75.

Kort & Matrikelstyrelsen (2001). *TOP10DK - Geometrisk registrering*. Specifikation udgave 3.2.0. Kort & Matrikelstyrelsen, Copenhagen, Denmark. [pdf] Available at: ftp://ftp.kms.dk/download/pdf/top10dkspec320.pdf [Accessed 08 June 2015].

Lindboe, E., Kærn, L. and Lützen, N. (2003). "*Cykel Parker og Rejs*." Hovedstadens Udviklingsråd.

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

NAVTEQ (2010). *NAVTEQ NAVSTREETS Street Data Reference Manual*. Version 2.8. NAVTEQ, US.

Prato, C.G., Halldórsdóttir, K. and Nielsen, O.A. (2015). Latent lifestyle and mode choice decisions when travelling short distance. Presented at the *14th International Conference on Travel Behaviour Research*, 19th – 23rd July, in Windsor, England. Submitted to *Transportation* in September 2015.

Pucher, J. and Buehler, R. eds. (2012). *City Cycling*. The MIT Press, Cambridge, Massachusetts.

Rapidis (2013). Traffic Analyst til ArcGIS. [online] Available at: http://www.rapidis.com/traffic_analyst [Accessed 08 June 2015].

Rich, J.H., Nielsen, O.A. and Hansen, C.O. (2010). Overall Design of the Danish National Transport Model. In proceedings for the *Annual Transport Conference in Aalborg*, Aalborg University.

Chapter 4

MODELLING MODE CHOICE IN SHORT TRIPS – SHIFTING FROM CAR TO BICYCLE

Title:	Modelling mode choice in short trips – Shifting from car to bicycle
Author(s):	Katrín Halldórsdóttir, Linda Christensen, Thomas C. Jensen, and Carlo G. Prato
Presented:	The *39th European Transport Conference (ETC)*, 10th – 12th October 2011, Glasgow, Scotland, UK
Abbreviated:	Halldórsdóttir et al. (2011)

A preliminary version of this paper was presented at the 39th European Transport Conference in October 2011, Glasgow, Scotland, UK and published in the conference proceedings. In this chapter the paper has been updated with improved model estimates and revised.

ABSTRACT

This paper investigates the mode choice behaviour of individuals from the Copenhagen Region, when travelling short distances. Data from the Danish National Travel Survey identify the travel behaviour of the Danish population through interviews, collecting travel diaries, and socio-economic variables of a representative sample of the population. Short trip chains were investigated on a data sample with 7,958 individuals and 10,982 observations.

The model considers five alternatives (i.e., car driver, car passenger, public transport, walking, and bicycle), for which level-of-service variables are calculated through assignment procedures, available for each period of the day in which the trip chains were conducted. The present study estimates a mixed logit model, which is able to capture taste variations and differentiate travel time parameters across modes. The mixed logit model allows the investigation of the effect of level-of-service variables, individual characteristics of the travellers, purpose of the trip chains, and environmental conditions.

Results suggest heterogeneity among cyclists in the sensitivity to travel time, temperature, and hilliness. The cost parameter is negative and significant. However, it is estimated very low, possibly because of lower relevance of the cost for short trip chains. Expectedly, the selection of sustainable transport modes for short trip chains is negatively linked to owning one or more cars. Urban density also has a positive correlation with the selection of sustainable transport modes, whereas motorised private cars are negatively correlated. The results show that the choice between transport alternatives is not only related to the level-of-service characteristics of the alternatives, but also to a large extent the socio-economic characteristics of the travellers.

4.1 INTRODUCTION

Over the years, demand for faster and more flexible transport has grown, as production, income, number of trips, and travel distances have increased. Consequently, the use of motorised private modes has steadily increased in Denmark. As a result, road congestion has become a major problem, especially in the Copenhagen Region. It is necessary to decrease the congestion and to solve the related health problems. One method is the promotion of sustainable transport modes, which has led to a particular emphasis on the shifting from motorised private transport to cycling.

Many projects have sought ways to increase walking and cycling instead of driving a car for shorter trips. Two large EU-financed projects addressed this issue and studied the differences in the share of walking and cycling in different European cities (Hydén et al., 1999; Forward, 1998). Various studies have also been conducted in Denmark, where applied methods were implemented to promote cycling (Troelsen et al., 2004; Jensen, 2001, 2004). In Aarhus, the second-largest city in Denmark, located on the east coast of Jutland, a bicycle and a free bus pass were made available to car drivers during a test period to see whether they would switch to more sustainable transport modes when travelling to work (Trafikforsknings-gruppen Aalborg Universitet, 2001). In order to increase people's possibilities to cycle, Egetoft et al. (2002) motivated people to plan their trips better. Finally, a study based on the Danish National Travel Survey (Jensen and Thost, 1999) concluded that the hilliness and the size of the city (with less than 10,000 inhabitants) is important for the bicycle share. In addition, the size of the city is important for the length of the cycling trip, and the share of people working in their own municipality.

A number of studies have investigated the potential for a mode shift from car use to more sustainable transport options, either through frequency analysis or mode choice models. Mackett (2003) analysed why people use their cars in short trips, through a revealed preference survey. The results showed that the main reasons for using the car were: carrying heavy goods, escorting someone, time constraints, and because the car was needed to run errands. Rodríguez and Joo (2004) analysed the importance of the physical surroundings on the choice of transport mode and showed that the availability of bicycle lanes and hilliness are especially important. Vågana (2006) presented a data analysis, based on the Norwegian Transport Survey, describing walking and cycling trips, whereas Vågana (2007) investigated whether it is possible to transfer short car trips to walking or cycling by means of logistic regression analysis. The results showed that gender, age, size of the city, season, length of trip chain, and certain purposes are significant for the choice of transport mode. Wardman et al. (2007) studied the effect of various policies intended to increase the share of cycling on commuting trips, by means of combining revealed- and stated preference data. Results showed that segregated bicycle paths and on-road bicycle lanes were found to be highly effective in increasing the bicycle share, as well as paying employees to cycle to work. Kim and Ulfarsson (2008) estimated a mode choice model, analysing the choice of transportation mode of short home-based trips. The results showed that there is a negative correlation for cycling with increased age, vehicle availability, being single with no children, and when being accompanied by others on trips, whereas having a bus pass, going to school, and social/recreational is positively correlated. de Nazelle et al. (2010) examined how shifting short car trips to other transportation modes reduces emissions, while also identifying what characteristics influence mode choice for short trips through logistic regression analysis. The study showed that the likelihood of driving increased with age and that family composition also has an effect. Socio-economic characteristics, particularly vehicle ownership and extra car in the household, increase the odds of driving on short trips, while people are less likely to drive during warm months and if the trip purpose is social, personal, or family related.

Nankervis (1999) and Bergström and Magnusson (2003) studied the significance of the weather on bicycle commuting, where the main focus was on the maintenance of bicycle paths during winter. Rietfeld and Daniel (2004) analysed to which extent municipality policies matter in relation to variation in bicycle share. The results showed that the most important factors for the choice of bicycling are the physical aspects, such as altitude difference, city size, and the share of young people in the population. In addition, the results concluded that differences in ethnic compositions is also important, as well as policy-related variables, such as the number of stops per km on the route and the risk of accidents. Parkin et al. (2008) presented a logistic regression model, based on aggregate data, that shows that the quality of main roads and the annual rainfall, as well as the temperature, are important in commuting trips. Also, segregate bicycle paths have a significant relation to bicycle share, even though the elasticity is low. This is in contrast to a paper by Wardman et al. (2007) that illustrates that segregate bicycle paths and bicycle lanes have a very high effect in the preference when choosing to cycle in commuting trips.

The Department of Transport at the Technical University of Denmark, in collaboration with the Danish Road Directorate, investigated short trips by car and examined whether it is possible to make car drivers shift to cycling or walking (Christensen and Jensen, 2008). The study presents a multinomial logit (MNL) model, investigating trip chains shorter than 22 km based on data from the Danish National Travel Survey (TU-survey) and demonstrated how three types of conditions influence the choice of transport mode: (i) conditions concerning the purpose of the trip chain and the road user, where car ownership and number of children in the family are the most important factors; (ii) conditions concerning the environment of the trip chain, where differences of hilliness and temperature have proved to be greatly relevant; and (iii) conditions concerning the travelling circumstances, where the project describes the effect of speed for car drivers and cyclists; the parking conditions; as well as a general effort to promote cycling.

The study showed that cycling policies would reduce the short trip chains by car in favour of cycling. 90% of short car trip chains would be transferred to bicycle, in the case where travellers would transfer to sustainable transport modes.

This paper aims to extend the previous choice model with a fresh and up-to-date perspective. Public transport was not considered in the MNL model of the previous study, due to the lack of data on the level-of-service of the public transport. The present study includes public transport in the choice model, since the evaluation of the transfer from car to bicycle could be biased by the exclusion of public transport as a possible option. In addition, the MNL model of the previous study did not allow the consideration of heteroscedasticity across alternatives and heterogeneity across travellers. To be able to capture taste variations through a specification, that expresses randomly distributed parameters and differentiates the travel time parameters across modes to express different values of time for different modes, the present study estimates a mixed logit model.

The remainder of the paper is structured as follows: section 4.2 describes the data used in the study; section 4.3 describes the methods applied to measure and model the behaviour of travellers; section 4.4 presents the results of a mixed logit model; and section 4.5 summarises the major findings of this study.

4.2 DATA

The data used in the study were extracted from the Danish National Travel Survey (abbreviated TU-survey) (Christiansen, 2009). For model estimation, 7,958 individuals and 10,982 observations constitute the sample of short trip chains in the Copenhagen Region.

The TU-survey identifies the travel behaviour of a representative sample of the Danish population through interviews where travel diaries and socio-economic variables (e.g., age, gender, income, education, car availability, etc.) are collected. DTU Transport conducts the survey on behalf of the Ministry of Transport and several other government departments. The TU-survey investigates travel during the day before the interview of Danes, between the age of 10 and 84, who are randomly selected. The respondents are asked about why they travel and by what means of transport they travelled during the day in question. In addition, the respondents are asked about the trips, when and where they take place, for how long, etc.

The survey reflects the diversity of the Danish population, collecting travel behaviour and sensitive personal information, such as name, coordinates, and precise address information. Data are collected each day throughout the year, thus characterising the differences in travel behaviour across seasons, weekdays, etc. The study is the only large Danish survey combining actual travel behaviour with a wide range of background variables. It gives a good description of the average travel behaviour of each person, by asking the respondents about their travel behaviour on a single day.

4.2.1 Hilliness, parking, and weather

Terrain ratio is calculated as the average gradient of all journeys undertaken within a radius of 5 km from the respondent's home. Hence, it indicates how hilly the area is and thus how difficult it is to cycle. The average parking is calculated within a radius of 5 km from the person's destination, therefore describing how difficult it is to park in the area. The temperature is obtained from the Danish Meteorological Institute (DMI). All three variables were implemented as continuous variables in the model specification.

4.3 Methodology

4.3.1 Mode choice attributes

To analyse current travel behaviour, or to forecast future travel behaviour, observed choices and alternatives composing the choice set of each traveller are necessary. The TU-survey collects the current travel behaviour, i.e., the observed choices. Route choice models and simulation methods were used to calculate the attributes of the alternatives within the choice set of each traveller. Five alternatives were considered, i.e., car driver, car passenger, public transport, walk, and bicycle.

The level-of-service variables for car driver, car passenger, and public transport were calculated through assignment procedures available for each period of the day in which the trip chain was conducted. The calculation of the level-of-service variables allows considering congestion conditions similar in average to the ones encountered by the travellers. The car travel time includes free flow travel time plus the added travel time due to congestion. The public transport travel time includes waiting time, access- and egress time, walking time, and in-vehicle time. The cost for car drivers was calculated with values from the Danish Transportation Economic Unit Prices (Modelcenter, 2010). The cost for public transport was estimated from the TU-survey as an average cost per km travelled, limited to the minimum- and maximum cost for the public transport as it is set in the Danish public transport pricing system.

There is no available information on travel speed on different parts of either the bicycle- or the pedestrian network. Therefore, the travel speed for the two modes was estimated as an average travel speed. The travel time was thus dependent on the travel distance. The travel distance was calculated with a shortest path simulation method. The cost for bicycles was also calculated with values from the Danish Transportation Economic Unit prices. It was assumed that the travel cost for walking is zero.

4.3.2 MODEL SPECIFICATION

The present study estimates a mixed logit model (for a detailed discussion see Train, 2003). The mixed logit probability can be derived from utility-maximising behaviour based on random coefficients. The decision maker has a choice set of J alternatives. The utility of decision maker n from alternative j is specified as:

$$U_{nj} = \beta'_n x_{nj} + \varepsilon_{nj}, \tag{4}$$

where x_{nj} are observed variables that associate with the alternative and decision maker, β_n is a vector of coefficients of these variables for decision maker *n,* representing the individuals' preferences, and ε_{nj} is a random term that is *iid* extreme value distributed over alternatives and decision makers. The coefficients vary over decision makers with density $f(\beta)$, which is a function of its parameters θ (e.g., mean and covariance of the β*'s* in the population). In the standard logit the β is fixed, while in the mixed logit the β varies over decision makers.

The researcher can only observe the x_{nj}*'s* but not β_n or the ε_{nj}*'s*. If the β_n would be observed by the researcher then the choice probability would be standard logit, given that the ε_{nj}*'s* are *iid* extreme value. Then the probability restricted on β_n is:

$$P_{ni}(\beta) = \frac{\exp(\beta'_n x_{ni})}{\sum_j \exp(\beta'_n x_{nj})}. \tag{5}$$

However, the researcher cannot condition on β, since the β_n is unknown. The unrestricted choice probability, which is the mixed logit probability, is therefore the integral of $P_{ni}(\beta_n)$ over all possible variables of β_n:

$$P_{ni}(\beta) = \int \frac{\exp(\beta'_n x_{ni})}{\sum_j \exp(\beta'_n x_{nj})} f(\beta)\, d\beta\ . \tag{6}$$

The present study specifies log-normal distribution, where $ln\beta \sim N(b,W)$, with parameters b and W that were estimated, for time variables that are supposed to be negative, and normal distribution for variables that are not expected to have a specific sign.

4.4 RESULTS

4.4.1 DATA ANALYSIS

The data analysis in this paper focuses on trip chains shorter than 22 km. Cases were excluded because respondents opted not to provide information or because other relevant information was missing for the analysis. Given the restrictions the sample includes 7,958 individuals and 10,982 trip chains. Given the extensive data collected from the survey, only selected results are presented here.

Table 9 and Table 10 present the category variables for personal characteristics and the trip characteristics, respectively. The continuous variables are presented in Table 11. The transport mode share is: 25% walking, 28% cycling, 35% driving, 5% being driven, and 7% taking public transport. The dataset has 3,674 males (46%) and the average age is 48.

TABLE 9: CATEGORY VARIABLES FOR PERSONAL CHARACTERISTICS

Variable	**Total**
Personal characteristics	(*N = 7,958*), *N(%)*
Age group	
18 - 24	619 (8)
25 - 34	1,283 (16)
35 - 44	1,702 (21)
45 - 54	1,494 (19)
55 - 64	1,401 (18)
65 - 74	969 (12)
75 and older	490 (6)
Gender	
Male	3,674 (46)
Female	4,284 (54)
Main occupation	
Student	711 (9)
Welfare	1,858 (23)
Unemployed	249 (3)
Employed	4774 (60)
Self-employed	364 (5)
Homemaker	2 (0)
Respondent has a driving licence	
Yes	6,653 (84)
No	1,305 (16)
Respondent has a bicycle	
Yes	6,275 (79)
No	1,683 (21)
Respondent has a monthly pass for public transport	
Yes	1,475 (19)
No	6,483 (81)
Vehicle availability	
Zero car	2,289 (29)
One car	4,309 (54)
Two cars	1,235 (16)
Three cars or more	125 (2)
Household category – Number of children	
Children between the age 0-4	996 (13)
No children between the age 0-4	6,962 (87)
Children between the age 5-9	1,135 (14)
No children between the age 5-9	6,823 (86)
Children between the age 10-15	1,314 (17)
No children between the age 10-15	6,644 (83)

TABLE 10: CATEGORY VARIABLES FOR TRIP CHARACTERISTICS

Variable	**Total**
Trip characteristics	(*N = 10,982*), *N(%)*
Mode choice	
Walking	2,776 (25)
Bicycle	3,015 (28)
Car driver	3,811 (35)
Car passenger	575 (5)
Public transport	805 (7)
Urban characteristics	
Copenhagen centre	4,016 (37)
Greater Copenhagen area	4,492 (41)
Minor town	1,296 (12)
Village	935 (9)
Rural area	243 (2)
Trip purpose	
Business	188 (2)
Commuting	1,688 (15)
Combination of commuting and other	785 (7)
Leisure	3,453 (31)
Shopping	3,572 (33)
Escorting	1,296 (12)

TABLE 11: CONTINUOUS VARIABLES

Variable	N	Mean	Std. dev.	Min	Max
Personal characteristics					
Income DKK/1,000	7,958	270.1	184.3	0.0	1,500.0
Trip characteristics					
Walk					
Travel time [min]	2,776	21.6	21.9	0.1	186.9
Travel distance [km]	2,776	1.7	1.7	0.0	15.0
Travel cost [DKK/km]	2,776	0	0	0	0
Bicycle					
Travel time [min]	3,015	22.7	17.6	0.5	88.0
Travel distance [km]	3,015	5.7	4.4	0.1	22.0
Travel cost [DKK/km]	3,015	1.9	1.5	0.04	7.5
Car driver					
Travel time [min]	3,811	11.9	7.2	0.0	44.7
Travel distance [km]	3,811	8.7	4.4	0.0	22.0
Travel cost [DKK/km]	3,811	20.9	12.9	0.1	53.0
Car passenger					
Travel time [min]	575	11.9	7.1	0.2	36.0
Travel distance [km]	575	8.7	5.4	0.1	21.9
Travel cost [DKK/km]	575	10.4	6.5	0.2	26.5
Public transport					
Waiting time [min]	805	14.7	12.5	1.5	143.4
Access/egress time [min]	805	22.7	10.9	0.8	79.5
In vehicle time [min]	805	23.7	13.0	0.4	68.8
Travel distance [km]	805	11.2	5.2	1.3	22.0
Travel cost [DKK/km]*	805	69.3	20.7	24	108

*Limited to the minimum and maximum cost for the public transport as it is set in the Danish public transport pricing system.

4.4.2 Model estimates and discussion

Table 12 summarises the results from the mixed logit model. The asymptotic t-test is primarily used to test whether a specific parameter in a model differs from a known constant, often zero. Not all coefficient variables, obtained in the survey, proved to be statistically significant at the 90% level. In addition, some variables that are considered interesting preference indicators cannot be included in the estimations, because they are correlated with other more important variables. According to these considerations, some variables were deleted to increase the reliability of the model. The final model is constituted by 10,982 observations, where there are 7,958 individuals and 45 estimated parameters. The alternative specific constant (ASC) for walking is fixed to zero for identification purposes.

Table 12: The results from the mixed logit model estimates

Variable	**Category**	**Transport mode**	**Value**	**t-test**	
Alternative specific constant		Walk	0		
		Bicycle	-2.60	-15.28	***
		Car driver	-9.17	-26.63	***
		Car passenger	-5.41	-32.58	***
		Public transport	-5.59	-22.02	***
Travel cost [DKK/km]		All modes	-0.008	-2.12	**
Time [min]	Travel time - μ	Walk	-1.81	-44.66	***
	Travel time - σ		0.55	26.76	***
	Travel time - μ	Bicycle	-2.28	-39.61	***
	Travel time - σ		0.40	8.71	***
	Travel time	Car driver	-0.10	-9.49	***
	Travel time	Car passenger	-0.12	-10.03	***
	Waiting time	Public transport	-0.03	-7.25	***
	In-vehicle time		-0.02	-2.71	***
	Access/egress time		-0.04	-6.89	***
Number of transfers		Public transport	-0.27	-3.03	***
Personal characteristics					
Monthly pass	Yes	Public transport	2.31	20.60	***
Gender	Male	Car driver	0.57	7.18	***
		Car passenger	-1.22	-9.70	***
		Public transport	-0.30	-2.60	***
Car ownership	One car	Car driver	2.98	22.02	***
	Two cars	Car driver	3.49	22.59	***
	Three cars or more	Car driver	3.76	12.61	***
Children in household	Ages of 5 to 9	Car driver	0.14	1.52	
	Ages of 10 to 15	Car driver	0.18	2.24	**

Income [DKK/1,000]		Car driver	0.001	3.53	***
Main occupation	Student	Bicycle	0.40	2.88	***
		Public transport	0.35	1.95	**
	Welfare	Bicycle	-0.27	-2.82	***
		Public transport	0.24	1.77	*
	Unemployed	Car driver	-0.54	-2.93	***
Trip characteristics					
Temperature	μ	Bicycle	-0.03	-5.51	***
	σ		0.08	8.64	***
Trip purpose	Commuting	Bicycle	0.37	2.58	***
		Car driver	-0.48	-3.63	***
		Car passenger	-0.49	-2.58	***
	Leisure	Bicycle	-2.06	-17.28	***
		Car driver	-1.60	-17.68	***
		Public transport	-1.36	-10.16	***
	Shopping	Bicycle	-0.98	-9.06	***
	Escorting	Car driver	1.15	9.97	***
Urban characteristic	Copenhagen centre	Car driver	-0.56	-5.43	***
		Car passenger	-1.27	-10.25	***
Hilliness	μ	Bicycle	-0.03	-3.12	***
	σ		0.08	6.45	***
Parking availability		Car driver	3.07	10.82	***
Number of estimated parameters:				45	
Number of observations:				10,982	
Number of individuals:				7,958	
Null log-likelihood:				-17,150.04	
Final log-likelihood:				-9,514.80	
Adjusted rho-square:				0.443	

*** Significant at a 99% level, ** significant at a 95% level;* significant at a 90% level.

4.4.2.1 TIME- AND TRAVEL COST VARIABLES

The travel time variables are important trip characteristics, considered a good indicator of individuals preferences. Wardman et al. (2007) have documented that travel time while cycling is considered to be three times more unpleasant than travel time by other transport modes. With an increase in the travel time the perceived convenience of bicycle trips declines, while it does not with other transport modes (Noland and Kunreuther, 1995). Studies have also shown that experienced cyclists are more sensitive to travel times (Stinson and Bhat, 2005; Hunt and Abraham, 2007). However, some cyclists may prefer slightly longer commuting distances, increasing the travel time, because of health and fitness reasons.

The model shows that the coefficients for the travel time variables are consistently negative for all transport modes. The in-vehicle time, the waiting time, and the access and egress times for public transport mode are significant and have the expected negative sign. Walk travel time and bicycle travel time are log-normally distributed, where both mean and standard deviation are significant. If this non-homogeneous effect is overlooked in the estimation, the model could compensate by making the public- and car transport modes more attractive. After all, it can be assumed that individuals use public transport and car more for longer distance travel than shorter distance.

The coefficient variable describing the number of transfer in public transport has a negative sign. This indicates that the higher the numbers of changes individuals have to take, the less they are willing to travel by public transport.

The cost of a transport mode is important when choosing a type of mode (e.g., Noland and Kunreuther, 1995; Rietveld and Daniel, 2004; Rodríguez and Joo, 2004; Pucher and Buehler, 2006). According to Bergström and Magnussen (2003) one reason why commuters choose to cycle is because it is economical. The results show that there is a negative correlation between travel cost and the choice of transport alternative.

4.4.2.2 Personal characteristics

The results show that having a monthly pass is positively correlated to travelling by public transport, thus confirming Kim and Ulfarsson (2008). The results also show that males are more likely to drive a car, while females are more likely to be car passengers and use public transport. A preliminary result, which is not included in the final model, showed that there was no statistical difference between males and females in relation to cycling. This contradicts other research, which concludes that males cycle more than females (e.g., Pucher et al., 1999; Howard and Burns, 2001; Dickinson et al., 2003; Rietveld and Daniel, 2004; Rodríguez and Joo, 2004; Moudon et al., 2005; Plaut, 2005; Stinson and Bhat, 2005; Dill and Voros, 2007). This difference could be because the share of female cyclists is much higher in the Copenhagen Region than other capital regions.

Mode choice behaviour is also strongly linked to household characteristics. Studies have shown that having a car in the household has a strong negative effect on the share of bicycles as a mode choice (e.g., Cervero, 1996; Stinson and Bhat, 2004, 2005; Plaut, 2005; Pucher and Buehler, 2006; Dill and Voros, 2007; Guo et al., 2007; Parkin et al., 2008) and that it increases the frequency of cycling to own fewer cars (Stinson and Bhat, 2004). This is consistent with the findings of the present paper. The coefficient variables describing car availability in the household is positive for all three categories (i.e., one car, two cars, and three cars or more). The higher the number of cars in the household is, the more likely individuals are to drive a car (e.g., Kim and Ulfarsson, 2008; de Nazelle et al., 2010). The results also show that individuals with children are more likely to be car drivers, when compared to other transport modes (e.g., de Nazelle et al., 2010).

The results show that there is a positive correlation between driving a car and income, indicating that the higher their income, the more individuals are willing to drive. The coefficient variable for income is not statistically significant for bicycles and is thus not included in the final model. This is consistent with other studies that found that income has no significant effect on bicycle share (Dill and Carr, 2003; Zacharias, 2005).

Another important indicator of personal preferences is the main occupation. Employment is used as a reference variable. The coefficient variable for students is positive for the bicycle and public transport, indicating that the higher the value is the more individuals are willing to travel by each mode. Bicycles are the likeliest choice of transport mode for students, followed by public transport. The model estimates shows that individuals on welfare are more likely to use public transport than to travel with other transport modes. The results also show that unemployed are less likely to drive a car in short trip chains.

The coefficient variables for the age categories are correlated with the main occupation variables. The main occupation variable is considered a better indicator of personal characteristics with respect to age, and to avoid multicollinearity, the age variables are not considered in the final model.

4.4.2.3 Trip characteristics

The primary trip characteristics are the time variables (discussed in subsection 4.4.2.1). The estimation results show that other attributes describing the trip characteristics are also related to mode choice.

The model results show that temperature is normally distributed for cycling. The results show that individuals are more likely to cycle with increasing temperature (Parkin et al., 2008), but the significant standard deviation suggests that some cyclists do not mind lower temperatures or that they cycle regardless of the temperature level, because it is their only transport alternative. It could be that these cyclists are regular commuters, which are less influenced by temperature than other cyclists (Bergström and Magnussen, 2003; Brandenburg et al., 2004).

Trip purpose is also a good indicator of personal preferences. There is a positive correlation between cycling and commuting, while in commuting trip chains individuals are less likely to use cars. The results show that individuals are less likely to cycle, drive a car, or use public transport in leisure trip chains. Also, individuals are less likely to use a bicycle when shopping and when travellers are escorting someone, they are more likely to drive a car (e.g., Mackett, 2003; Kim and Ulfarsson, 2008).

Studies have shown that the bicycle share grows with higher residential densities (e.g., Pucher and Buehler, 2006; Guo et al., 2007; Parkin et al., 2008; Zahran et al., 2008). The model results show that urban characteristics also affect mode choice. In central Copenhagen individuals are more likely to choose sustainable transport modes, reflected in the coefficient variables for car driver and car passenger being negative. This is also an indication of the difficulty of accessing and finding parking in the city centre. Shown as well by the coefficient variable for the minimum level of car parking at the destination being estimated positive in relation to driving a car.

Hilliness has been found to have negative effect on the bicycle share (e.g., Rietveld and Daniel, 2004; Rodríguez and Joo, 2004; Timperio et al., 2006; Parkin et al., 2008). The results show that hilliness is normally distributed for bicycles. The presence of slopes has a negative impact on cycling, but the standard deviation is also significant and suggests that some cyclists do not mind riding up-hill.

4.5 Conclusions

This paper analysed the mode choice behaviour, in the Copenhagen Region, in trip chains shorter than 22 km on the basis of data from the Danish National Travel Survey. The sample for model estimation included 7,958 individuals and 10,982 trip chains. Route choice models and simulation methods were used to calculate the alternative attributes within the choice set of each traveller. Five alternatives were considered, i.e., car driver, car passenger, public transport, walk, and bicycle. The level-of-service variables for car and public transport were calculated through assignment procedures available for each period of the day in which the trip chain was conducted. The travel attributes for walk and bicycle were calculated with a shortest path simulation method. A mixed logit model was estimated. The results help identifying important factors that affect the mode choice, i.e., level-of-service variables, the socio-economic characteristics of the travellers, trip characteristics, and environmental conditions.

Firstly, the paper shows that travel time is as important for cyclists, as other transport modes. The coefficients for the travel time variables are consistently negative and significant, for all transport modes. Thus, travellers aim to minimise the travel time for all transport modes. The results also show that travellers prefer not to transfer between public transports modes. Walk- and bicycle travel times are log-normally distributed, and significant standard deviation indicates that individuals have a heterogeneous preference for the travel time of non-motorised modes, and a more homogeneous preference for the travel time of motorised modes. The model estimates also show that there is a negative correlation between travel cost and the choice of transport alternative. However, it is estimated very low, possibly because of lower relevance of the cost for short trip chains.

Secondly, certain attributes of the personal characteristics are related to the mode choice. The study concludes that males are more likely to drive a car than females, while females were more likely to be car passengers and use public transport. The results show that income has no significant effect on bicycle share in mode choice. However, the variable for car driver is significant. The results indicate that the higher their income, the more individuals are willing to drive a car. Although income has no significant effect on bicycle share, main occupation has, as well as on other transport modes.

The study shows that mode choice behaviour was also strongly linked to household characteristics. Having a monthly pass has a positive influence on public transport share, and the higher the number of cars in a household, the more likely individuals are to drive a car. This could indicate that there is a negative effect on share of bicycles as a mode if there is a car in the household and that having fewer cars could increase cycling frequency. The study concludes that individuals with children were more likely to be car drivers, when compared to other transport modes.

Finally, the study shows that attributes describing the trip characteristics are also related to the mode choice. Hilliness is normally distributed for cyclists and the presence of slopes has in average a significant negative impact on cycling. The results also show that individuals are on average more likely to cycle and walk with increasing temperatures, but the significant standard deviation suggests that some cyclists do not mind lower temperatures. The study concludes that in commuting trip chains individuals are less likely to use cars, whereas they are more likely to drive a car when escorting others. Individuals are less likely to cycle, drive a car, or use public transport in leisure trip chains. Also, individuals are less likely to use a bicycle when shopping. In central Copenhagen individuals are more likely to choose sustainable transport modes and the minimum level of car parking at the destination is positively related to driving a car.

The objective of this study was to uncover the determinants of choice between car and sustainable transport alternatives in short distance travelling. This study helps uncover factors that are able to make cycling more attractive, e.g., improving accessibility for bicycles and addressing specific population groups for specific trip purposes.

The model presented in this paper is a work in progress. The model suggests that a further heterogeneity investigation, possibly with a latent class approach, might uncover whether different population groups exhibit different preference structures. Lastly, scenario simulations would allow further evaluation of the effects of possible policy instruments intending to convert short car trips to bicycles or walking.

4.6 References

Bergström, A. and Magnusson, R. (2003). Potential of transferring car trips to bicycle during winter. *Transportation Research Part A*, 37, pp. 649-666.

Brandenburg, C., Matzarakis, A. and Arnberger, A. (2004). The effects of weather on frequencies of use by commuting and recreation bicyclists, in: A. Matzarakis, C. R. De Freitas and D. Scott (Eds) *Advances in Tourism Climatology*, Vol. 12, pp. 189–197 (Freiburg: Berichte des Meteorologischen Instituts der Universität Freiburg).

Cervero, R. and Radisch, C. (1996). Travel choices in pedestrian versus automobile oriented neighbourhoods. *Transport Policy*, 3(3), pp. 127-141.

Christiansen, H. (2009). Modernisering af Transportvaneundersøgelsen. Trafikdage på Aalborg Universitet. [online] Available at: <http://www.trafikdage.dk/td/papers/papers09/295_HjalmarChristiansen.pdf> [Accessed 8 Mars 2010].

Christensen, L. and Jensen, T.C. (2008). *Korte ture i bil – Kan bilister ændre adfærd til gang eller cykling?* Rapport 3. DTU Transport, Kgs. Lyngby, Denmark.

de Nazelle, A., Morton, B.J., Jerrett, M. and Crawford-Brown, D. (2010). Short trips: An opportunity for reducing mobile-source emissions? *Transportation Research Part D: Transport and Environment*, 15, 451-457.

Dickinson, J.E., Kingham, S., Copsey, S. and Hougie, D.J.P. (2003). Employer travel plans, cycling and gender: will travel plan measures improve the outlook for cycling to work in the UK? *Transportation Research Part D*, 8(1), pp. 53–67.

Dill, J. and Carr, T. (2003). *Bicycle Commuting and Facilities in Major US Cities: If You Build Them, Commuters will Use Them.* (Washington, DC: Transportation Research Board).

Dill, J. and Voros, K. (2007). *Factors Affecting Bicycling Demand: Initial Survey Findings from the Portland Region* (Washington, DC: Transportation Research Board).

Egetoft, A., Stuven, H.C. and Kragh, T. (2002). *Familiens transportvaner - kan vi lade bilen stå?, I Persontransport og bæredygtighed.* Notat 2002:3, Danmarks Transport Forskning, Denmark.

Forward, S. (1998). *Behavioural factors affecting modal choice.* Swedish National Road and Transport Institute, Sweden.

Guo, J.Y., Bhat, C.R. and Copperman, R.B. (2007). *Effect of the Built Environment on Motorized and Non-Motorized Trip Making: Substitutive, Complementary, or Synergistic?* (Washington, DC: Transportation Research Board).

Howard, C. and Burns, E.K. (2001). *Cycling to Work in Phoenix: Route Choice, Travel Behavior, and Commuter Characteristics* (Washington, DC: Transportation Research Board).

Hunt, J.D. and Abraham, J.E. (2007). Influences on bicycle use. *Transportation*, 34, pp. 453–470.

Hydén, C., Nilsson, A. and Risser, R. (1999). *WALCYNG—how to enhance walking and cycling instead of shorter car trips and to make those modes safer.* Bulletin 165. Institutionen för Trafikteknik, Lunds Tekniska Högskola, Sweden.

Jensen, S.U. (2001). *Odense - Danmarks Nationale Cykelby, Midtvejsevaluering af transportvaner*, Notat 7:2001. Danmarks Transport Forskning, Denmark.

Jensen, S.U. (2004). *Evaluering af Odense Cykelby.* Trafikdage på Aalborg Universitet, Denmark.

Jensen, S.U. and Thost, P. (1999). *Bystrukturens betydning for cykeltrafikken.* Trafikdage i Aalborg. [online] Available at: <http://www.trafikdage.dk/td/indhold/99/bpot99.htm> [Accessed 5 June 2005].

Kim, S. and Ulfarsson, G.F. (2008). Curbing automobile use for sustainable transportation: analysis of mode choice on short home-based trips. *Transportation*, 35, 723-737.

Mackett, R. (2003). Why do people use their cars for short trips? *Transportation*, 30, 329-349.

Modelcenter (2010). Transportøkonomiske Enhedspriser 2010. [online] Available at: <http://www.dtu.dk/centre/Modelcenter/Nyheder.aspx?guid=%7BA89F9087-C647-4E17-A393-33CAD053C12C%7D> [Accessed 22 August 2011].

Moudon, A.V., Lee, C., Cheadle, A.D., Collier, C.W., Johnson, D., Schmid, T.L. and Weather, R.D. (2005). Cycling and the built environment: a US perspective. *Transportation Research Part D,* 10, p. 245-261.

Nankervis, M. (1999). The effect of weather and climate on bicycle commuting. *Transportation Research Part A*, 33, pp. 417-431.

Noland, R.B. and Kunreuther, H. (1995). Short-run and long-run policies for increasing bicycle transportation for daily commuter trips, *Transport Policy*, 2(1), pp. 67-79.

Parkin, J., Wardman, M. and Page, M. (2008). Estimation of the determinants of bicycle mode share for the journey to work using census data. *Transportation*, 35(1), pp. 93-109.

Plaut, P.O. (2005). Non-motorized commuting in the US. *Transportation Research Part D*, 10, pp. 347-356.

Pucher, J. and Buehler, R. (2006). Why Canadians cycle more than Americans: a comparative analysis of bicycling trends and policies. *Transport Policy*, 13(3), pp. 265–279.

Pucher, J., Komanoff, C. and Schimek, P. (1999). Bicycling renaissance in North America? Recent trends and alternative policies to promote bicycling. *Transportation Research Part A,* 33(7/8), pp. 625–654.

Rietveld, P. and Daniel, V. (2004). Determinants of bicycle use: do municipal policies matter? *Transportation Research Part A*, 38, pp. 531–550.

Rodríguez, D.A. and Joo, J. (2004). The relationship between non-motorized mode choice and the local physical environment, *Transportation Research Part D*. 9(2), pp. 151-173.

Stinson, M.A. and Bhat, C.R. (2004). Frequency of bicycle commuting: internet-based survey analysis. *Transportation Research Record*, 1878, pp. 122-130.

Stinson, M.A. and Bhat, C.R. (2005). *A Comparison of the Route Preferences of Experienced and Inexperienced Bicycle Commuters* (Washington, DC: Transportation Research Board).

Timperio, A., Ball, K., Salmon, J., Roberts, R., Giles-Corti, B., Baur, C.A. and Crawford, D. (2006). Personal, family, social, and environmental correlates of active commuting to school. *American Journal of Preventive Medicine*, 30(1), p. 45-51.

Trafikforskningsgruppen Aalborg Universitet (2001). *Cykel`buster projektet i Århus - Fra bil til cykel eller bus med positive virkemidler – projektevaluering.* Transportrådet Notat 01-01, Denmark.

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

Troelsen, J., Jensen, S.U. and Andersen, T. (2004). *Evaluering af Odense - Danmarks Nationale Cykelby*. Odense Kommune, Denmark.

Vågane, L. (2006). *Turer til fots og på sykkel.* Den nasjonale reisevaneundersøkelsen 2005. TØI rapport 858/2006, Norway.

Vågane, L. (2007). *Short car trips in Norway: Is there a potential for modal shift?* Available through: <http://www.etcproceedings.org/conference/listall>

Wardman, M., Tight, M. and Page, M. (2007). Factors influencing the propensity to cycle to work. *Transportation Research Part A*, 41, pp. 339-350.

Zacharias, J. (2005). Non-motorized transportation in four Shanghai districts. *International Planning Studies*, 10(3-4), p. 323-340.

Zahran, S., Brody, S.D., Maghelal, P., Prelog, A. and Lacy, M. (2008). Cycling and walking: explaining the spatial distribution of healthy modes of transportation in the United States. *Transportation Research Part D*, 13(7), pp. 462-470.

Chapter 5

LATENT LIFESTYLE AND MODE CHOICE DECISIONS WHEN TRAVELLING SHORT DISTANCES

Title:	Latent lifestyle and mode choice decisions when travelling short distance
Author(s):	Carlo G. Prato, Katrín Halldórsdóttir, and Otto A. Nielsen
Presented:	To be presented at the *14th International Conference on Travel Behaviour Research*, 19th – 23rd July 2015, Windsor, England
Submitted:	Submitted to *Transportation* in September 2015
Abbreviated:	Prato et al. (2015)

A preliminary version of this paper was presented at the 14th International Conference on Travel Behaviour Research, 19th – 23rd July 2015, Windsor, England*. The paper was then submitted to Transportation in September 2016.*

ABSTRACT

In the quest for sustainable travel, short trips appear the most amenable to curbing the use of the automobile. Existing studies about short trips evaluate the potential of shifting from the automobile to sustainable travel options while considering the population as homogeneous in its preferences and its tendency to accept these alternative travel options as realistic. However, this assumption appears quite unrealistic and the current study offers a different perspective: the mode choices when traveling short distances are likely related to lifestyle decisions. Short trip chains of a representative sample of the Danish population in the Copenhagen region were analysed, and more specifically a latent class choice model was estimated to uncover latent lifestyle groups and choice specific travel behaviour. Results show that four lifestyle groups are identified in the population: car oriented, bicycle oriented, public transport oriented, and public transport averse. Each lifestyle group has specific perceptions of travel time (with extremely different rates of substitution between alternative travel modes), transfer penalties in public transport trip chains, weather influence on active travel modes, and trip purpose effect on mode selection. When thinking about measures to increase the appeal of sustainable travel options, decision-makers should consequently look at specific individuals within the population and more sensitive individuals to comfort and level-of-service improvements across the lifestyle groups.

Keywords: *Lifestyle choices; Short trip mode choice; Latent class models; Auto oriented; Bicycle oriented; Public transport oriented.*

5.1 INTRODUCTION

When looking at mode choice decisions, the core assumption is that individuals want to travel from their origin to their destination in the way that guarantees the shortest time, the cheapest cost, the most comfortable travel, and the most flexible opportunity for escorting children to their activities, carrying heavy goods and changing destination or time of travel seamlessly. The automobile provides the answers to these needs, as its large use in cities throughout the world has been observed even for short trips where it has replaced sustainable travel options such as walking, cycling, and public transport (see, e.g., Pucher et al., 1999; Rietveld, 2000; Mackett, 2003; de Nazelle et al., 2010).

The large use of the automobile causes great distress for the sustainability of the cities of the future, given that environmental (see, e.g., Hertel et al., 2008; de Nazelle et al., 2010; Lindsay et al., 2011), climate (see, e.g., Maibach et al., 2009; Fuglestvedt et al., 2010; Borken-Kleefeld et al., 2013), and health (see, e.g., de Nazelle et al., 2011; Rojas-Rueda et al., 2011; Scheepers et al., 2013) concerns urge looking for sustainable travel solutions. While the convenience and swiftness of the automobile might thwart the attempt to reduce its use in suburban or rural areas where activities are dispersed over larger distances and travel alternatives are more scarce, short trips appear more amenable to curbing automobile use (e.g., Frank et al., 2000; Mackett et al., 2003; Loukopoulos and Gärling, 2005; Kim and Ulfarsson, 2008; de Nazelle et al., 2010; Monzon et al., 2011).

Modal shift from the automobile to sustainable travel modes for short trips has been analysed in the literature with a focus on the potential for individuals to benefit from the climate, environmental, and health perspective. However, this potential very rarely translates into individuals actually leaving their cars, as the automobile is very convenient even on short trips for carrying goods, picking up and dropping off spouses and children, staying within time constraints, and enjoying comfort and convenience. Even though emission reduction is an obvious positive effect of the use of public transport and active travel modes (e.g., Frank et al., 2000; de Nazelle et al., 2010; Lindsay et al., 2011), often individuals value the convenience and swiftness of the automobile with respect to these sustainable travel options far more than their potential contribution to solve environmental issues (e.g., Banister, 2008; Monzon et al., 2011; Borken-Kleefeld et al., 2013). Even though active travel present obvious benefits from the health perspective (e.g., Rojas-Rueda et al., 2011; Grabow et al., 2012; Piatkowski et al., 2015), often individuals consider the potential risks in terms of decreased safety and increased accident probability (e.g., de Hartog et al., 2010; Rojas-Rueda et al., 2011; Schepers and Heinen, 2013).

The assessment of the potential for modal shift from automobile use to active travel and/or public transport in short trips usually relies on mode frequency analysis and mode choice models (see, e.g., Mackett, 2003; Loukopoulos and Gärling, 2005; Kim and Ulfarsson, 2008; Maibach et al., 2009; de Nazelle et al., 2010; Rojas-Rueda et al., 2011; Carse et al., 2013; Scheepers et al., 2013). Interestingly, individuals are assumed in these studies to behave homogeneously and to have the same underlying probability of shifting mode regardless of their characteristics. The assumption of homogenous individuals does not seem plausible: (i) travel behaviour has extensive literature on taste heterogeneity across individuals in mode choice models (for recent applications see, e.g., Hess and Stathopoulos, 2013; Forsey et al., 2014; Noland et al., 2014; Paullsen et al., 2014); (ii) long-term decisions such as residential location, workplace location, car ownership, bicycle possession and public transport pass purchase play a role on short-term decisions such as mode choice (for recent applications see, e.g., Pinjari et al., 2011; Vovsha et al., 2013; Zhou, 2014; Guerra, 2015). The assumption of all individuals having the same probability of shifting travel mode does not seem plausible either, since their lifestyle most likely plays a role in the mode choice for short trips just as it is observed to play a part in other decisions: (i) residential location has been related to the lifestyle of households (e.g., Walker and Li, 2007; Smith and Olaru, 2013) and knowledge-workers (e.g., Frenkel et al., 2013a; Frenkel et al., 2013b); (ii) mode choices and mobility styles have been associated with the lifestyle of individuals and their correlated residential locations (e.g., Krizek, 2006; Scheiner and Holz-Rau, 2007; Kitamura, 2009; Vij et al., 2013); (iii) the decision about owning a car and, in that case, selecting a car type has been connected with lifestyle stages (e.g., Choo and Mokhtarian, 2004; Van Acker et al., 2014); (iv) time use patterns, activity participation and neighborhood characteristics have been linked to lifestyle choices (e.g., Krizek and Waddell, 2002; Schwanen and Mokhtarian, 2005; Fan and Khattak, 2012; Sun et al., 2012); (v) risky driving in adolescents has been coupled with the lifestyle of the family where they were raised (Bina et al., 2006).

The current study proposes the analysis of mode choices for short trip chains from a lifestyle perspective. Unlike existing studies about short trips, this study recognizes the heterogeneity across individuals and relates the short-term choices of travel mode with the long-term decisions of lifestyle. Unlike most existing studies about lifestyle in the transportation literature, this study looks at lifestyle influencing short-term choices rather than long-term ones such as residential location or car availability. The current study proposes a latent class analysis that allows inferring how lifestyle affects the decision of how to travel in a short-time horizon and for short distances where sustainable travel options seem realistically feasible.

This study focuses on short trip chains in the Copenhagen region as an example of a large metropolitan area that offers sustainable travel options and yet experiences significant traffic congestion and stalling cycling modal shares. Although policies exist to curb car purchase with extremely high registration taxes that have negative externalities (see, e.g., Mabit and Fosgerau, 2011; Rich et al., 2013) and extensive infrastructure exists for cycling, not only an impasse in the cycling shares has been observed, but also young Danes have expressed the worrisome intentions of using the car instead of the bicycle in the near future (Sigurdardottir et al., 2013). The current study offers a different perspective when looking at the impasse in sustainable travel progress by estimating a latent class model that links observable characteristics of the individual to the likelihood of having chosen a certain lifestyle that then affects the travel choices for short trip chains. Data about short trip chains were available from a representative sample of the Danish population who participated in the Danish National Travel Survey: the sample included 10,982 trip chains with five available alternative modes (i.e., walking, cycling, car driver, car passenger, and public transport), and contained information about the characteristics of the travellers, the trip chains, and the environment.

The remainder of the paper is structured as follows. Section 5.2 introduces the model formulation and the data collection for looking at the effect of lifestyle on the mode choice of Copenhageners traveling short distances. Section 5.3 presents the empirical results in terms of determinants of travel behaviour specific to individuals having different lifestyles and predictors of lifestyle group belonging. Section 5.4 summarizes the conclusions and highlights further research directions.

5.2 METHODS

5.2.1 MODEL

A latent class choice model is the most suitable methodological approach to analyse the effect of lifestyle on mode decisions for short distances. As previously clarified, the model allows to simultaneously uncover lifestyle preferences that are not directly observable from the data and to elicit mode choice preferences that are heterogeneous across the lifestyle groups. Details about latent class choice models are provided by Gopinath (1995), Walker (2001), and Greene and Hensher (2003).

The latent class choice model is composed of two parts: (i) a class membership model that represents the probability of individual *n* to have lifestyle *s*, and (ii) a class specific choice model that represents the probability of individual *n* with a specific lifestyle *s* to choose travel mode *i* for short trip chain *t*. Given the characteristics X_n of the individual and the attributes X_i of the travel modes, the probability of individual *n* to choose mode *i* for short trip chain *t* is expressed as:

$$P\left(i_t \middle| X_n, X_{nit}\right) = \sum_{s=1}^{S} P\left(i_t \middle| X_{nit}, s\right) P\left(s \middle| X_n\right) \tag{7}$$

where $P(s|X_n,)$ is the probability of individual n with characteristics X_n to have lifestyle s, and $P(i_t|X_{nit},s)$ is the probability of individual n, conditional on having lifestyle s, to choose mode i with attributes X_{nit} as perceived by individual n for short trip chain t. It should be noted that the probability of choosing mode i for short trip chain t is equal to the sum over all the S lifestyles of the products of the probability of the class specific choice model (conditional on lifestyle s) and the probability of having that lifestyle.

In the current study, the class specific choice model is specified as an error component logit that captures the correlation between alternative modes (i.e., active travel vs. motorised travel) and the panel effect for multiple trip chains t being performed by individual n with lifestyle s. Given five alternative modes available to individual n (W = walk, C = bicycle, D = car driver, P = car passenger, B = public transport) performing T short trip chains, the utility functions U_{nits} of the travel modes i for short trip chain t of individual n having lifestyle s are expressed as follows:

$$\begin{aligned}
U_{nWts} &= \beta_s X_{nWt} + \sigma^A \eta_n^A + \varepsilon_{nWts} \\
U_{nCts} &= \beta_s X_{nCt} + \sigma^A \eta_n^A + \varepsilon_{nCts} \\
U_{nDts} &= \beta_s X_{nDt} + \sigma^M \eta_n^M + \varepsilon_{nDts} \\
U_{nPts} &= \beta_s X_{nPt} + \sigma^M \eta_n^M + \varepsilon_{nPts} \\
U_{nBts} &= \beta_s X_{nBt} + \sigma^M \eta_n^M + \varepsilon_{nBts}
\end{aligned} \tag{8}$$

where the error components η_n^A and η_n^M capture the correlation across active travel modes A and motorised travel modes M as well as the panel effect across individuals n. The error components η_n^A and η_n^M are i.i.d. normally distributed with mean equal to zero and variance equal to one, the error terms ε_{nWts}, ε_{nCts}, ε_{nDts}, $\varepsilon_{nPts,}$ and ε_{nBts} are i.i.d. extreme value distributed across individuals, trip chains, and lifestyles, and the vectors η (= η_n^M, η_n^A) and ε (=ε_{nWts}, ε_{nCts}, ε_{nDts}, ε_{nPts}, ε_{nBts}) are independent (see Walker, 2001). The column vectors X_{nWt}, X_{nCt}, X_{nDt}, X_{nPt} and X_{nBt} contain the attributes of the travel modes as perceived by individual n for trip chain t, and they are obviously independent of the lifestyle s. The parameters to be estimated are the row vectors β_s, which are specific to each lifestyle s, and the scalars σ^A and σ^M that are equal across lifestyles s to impose a parsimonious specification of the error structure and facilitate model identification (see Walker and Li, 2007).

In the current study, the class membership model is specified as a logit model where the utility function U_{ns} of individual n having lifestyle s is:

$$U_{ns} = \delta_s + \gamma_s X_n + \varepsilon_{ns} \tag{9}$$

where the vector X_n contains the socio-economic-demographic characteristics of the individuals n, δ_s is a class specific constant to be estimated, γ_s is a vector of class specific parameters to be estimated, and ε_{ns} is an i.i.d. extreme value distributed error term. It should be noted that the probabilistic nature of the class membership model allows for each individual to have a different probability of having a different lifestyle s, and hence to have multiple lifestyles in which one might be dominant because of a very high probability.

Given the specification of the two components of the latent class choice model, the probability of individual *n* choosing mode *i* for short trip chain *t* conditional on having lifestyle *s* is expressed as:

$$P\left(i_t \left| X_{nit}, s; \beta_s, \sigma^A, \sigma^M \right.\right) = \int \prod_{t=1}^{T} P\left(i_t \left| X_{nit}, s, \eta; \beta_s, \sigma^A, \sigma^M \right.\right) f\left(\eta\right) d\eta \tag{10}$$

This is the product of the logit probability of each individual *n* choosing mode *i* for each of *T* trip chains (where the number of trip chains per individual varies, thus the panel is unbalanced), conditional on the unknown η and hence integrated over the distribution of η. The probability of individual *n* having lifestyle *s* is expressed as:

$$P\left(s \left| X_n; \delta_s, \gamma_s \right.\right) = \frac{\exp\left(\delta_s + \gamma_s X_n\right)}{\sum_{r=1}^{S} \exp\left(\delta_r + \gamma_r X_n\right)} \tag{11}$$

Accordingly, the probability of individual *n* choosing mode *i* for short trip chain *t* is:

$$P\left(i_t \left| X_n, X_{nit}; \beta_s, \sigma^A, \sigma^M, \delta_s, \gamma_s \right.\right) = \sum_{s=1}^{S} P\left(i_t \left| X_{nit}, s; \beta_s, \sigma^A, \sigma^M \right.\right) P\left(s \left| X_n; \delta_s, \gamma_s \right.\right) \tag{12}$$

and the log-likelihood is expressed as:

$$LL = \sum_t \sum_i d_{nit} \omega_t \ln\left(P\left(i_t \left| X_n, X_{nit}; \beta_s, \sigma^A, \sigma^M, \delta_s, \gamma_s \right.\right)\right) \tag{13}$$

where d_{nit} is equal to 1 if individual *n* chooses mode *i* for trip chain *t* (and 0 otherwise), and ω_t is the weight of short trip chain *t*. The model is estimated via maximum likelihood estimation, and numerical integration is used to evaluate the two-dimension integral in equation (10). The model estimation produces simultaneously the parameter estimates for the elements of the vectors β_s, δ_s, and γ_s, and the scalars σ^A and σ^M, which allow evaluating the different trade-offs made by individuals having different lifestyles. It should be noted that the model is probabilistic in nature, namely each individual *n* has a non-null probability to have latent lifestyle *s*, and the estimate of the size of each lifestyle group is provided. Moreover, the main issue with the model estimation is that the number of lifestyles cannot be estimated endogenously, but the exogenous definition of the number *S* of classes for the estimation of different models can be performed and then the performances of the different models can be compared. In the current study, latent class choice models were estimated with *S* varying from 2 to 6 and the number of classes was selected via a combination of statistical information and interpretation of the estimation results.

5.2.2 Data

Data about short trip chains of a representative sample of the Danish population were obtained from the Danish National Travel Survey (TU, in Danish Transportvane-Undersøgelsen).

The TU survey collects information about the travel behaviour of a representative sample of the Danish population between 10 and 84 years old via the administration of about 1,000 interviews per month (about 80% by telephone and about 20% on the internet) since 2006. The TU survey is administered by the Department of Transport of the Technical University of Denmark with the support of an external consultant for the calibration of the representative sample. The TU survey participants are extracted via a stratified random procedure from the Danish Civil Registration System (in Danish, Det Centrale Personregister) managed by the Danish National Board of Health (in Danish, Sundhedsstyrelsen) with the objective of reaching representativity of the population as listed in the Danish National Register managed by the Danish Census Bureau (in Danish, Danmark Statistik). The Danish Data Protection Agency (in Danish, Datatilsynet) permits the use of sensitive data for research purposes, namely names, addresses, and coordinates of the movements. Approximately 95% of the locations (e.g., home addresses, workplace addresses, trip points) are coded geographically by the respondent with a “search and select” available in the survey. Addresses are identified at the coordinate level in 98% of the cases, and at the zone level in 99.9% of the cases, which implies that absolute confidentiality is guaranteed prior to processing the data.

The 10,982 short trip chains analysed in the current study were the result of the application of the following criteria: (i) the trip chains were below a distance threshold of 22 km, which constitutes the 95% percentile of the trip chains by active travel mode in the Copenhagen Region and accordingly is a realistic distance threshold to be considered for curbing automobile use in short trips; (ii) the trip chains were performed by the population over 18 years of age that constitutes the driving licensing age in Denmark (trip availability considered the car availability in the household). The short trip chains contained detailed information about: (i) the socio-economic-demographic characteristic of the 7,958 individuals between 18 and 84 years old that travelled for short distances; (ii) the level-of-service variables of the trip chains by walking, cycling, driving, being a passenger in a car, and being a passenger in a public transport vehicle (e.g., bus, metro, train); (iii) the context of the trip chain in terms of trip purpose, detailed location characteristics (e.g., station type, parking availability), weather conditions (e.g., temperature, rain); (iv) the weight of each trip chain that guarantees the representativity of the sample given that the participation to the TU survey is voluntary (62.5% complete responses) and self-selection of population strata is observed. The level-of-service variables were calculated by knowing the network conditions at the time of the trip chain and by assuming shortest path choices for walking and cycling, shortest path choices conditional on the congestion conditions for driving and being a passenger in a car, and detailed indication of the route choices by public transport as collected in the dedicated section of the TU survey (Anderson et al., 2014).

Table 13 summarizes the characteristics of the sample analysed with the latent class choice model, corrected by the weights allowing to achieve population representativity. The short trip chains in the Copenhagen region were 18.0% by walking, 28.4% by cycling, 39.3% by driving, 6.1% by being a passenger in a car, and 8.2% by being passenger on a public transport vehicle. Remarkably, almost half of the short trip chains were still done by car in a city like Copenhagen that offers plenty of sustainable transport alternatives. The sample shows almost equal share of men and women, almost equal proportion of age categories, representative percentage of children in the various ages (0-4 are preschool children, 5-9 are children not travel independent, and 10-15 are children with initial travel independence independent), and representative variation across individuals in terms of occupation and income. The sample also shows the characteristics of the trip chains, with heterogeneous composition according to time and cost, and representative share in terms of purpose and location with the majority in the centre of the city or in the immediate neighbouring municipalities.

TABLE 13: SAMPLE CHARACTERISTICS

Variable	Categories (weighted)			
	Individual characteristics			
Gender	Male	48.3%	Female	51.7%
Age	18-24	8.8%	45-54	17.9%
	25-34	19.8%	55-64	15.8%
	35-44	22.2%	65 or older	15.5%
Single	Yes	28.9%	No	71.1%
Children 0-4	Yes	13.9%	No	86.1%
Children 5-9	Yes	14.6%	No	85.4%
Children 10-15	Yes	15.8%	No	84.2%
Occupation	Student	10.3%	Retired	20.1%
	Employed	61.9%	Unemployed	3.1%
	Self-employed	4.6%		
Income (yearly)	Mean	272,950 kr.	St. dev.	230,610 kr.
Bicycle	Yes	80.0%	No	20.0%
Number of cars	None	31.2%	Two	13.7%
	One	53.7%	Three or more	1.4%
Driving license	Yes	83.4%	No	16.6%
Public transport monthly pass	Yes	18.3%	No	81.7%
Parking availability at destination	Yes	90.5%	No	9.5%
Free parking at destination	Yes	54.3%	No	45.7%
	Trip characteristics			
Trip purpose	Commute	20.7%	Shopping	31.6%
	Business	2.3%	Escorting	9.0%
	Leisure	26.0%	Other	10.5%
Location	Copenhagen centre	36.6%	Minor town	11.8%
	Copenhagen area	40.9%	Rural area	10.7%
Travel time (walking)	Mean	23.40 min	St. dev.	23.40
Travel time (cycling)	Mean	22.72 min	St. dev.	19.26 min
Travel time (driving)	Mean	10.01 min	St. dev.	7.98 min
Travel time (car passenger)	Mean	10.01 min	St. dev.	7.98 min
Access time (public transport)	Mean	0.26 min	St. dev.	0.87 min
Waiting time (public transport)	Mean	14.74 min	St. dev.	14.64 min
In-vehicle time (public transport)	Mean	17.36 min	St. dev.	16.72 min
Number of transfers (public transport)	Mean	0.37	St. dev.	0.75
Travel cost	Mean	18.13 kr	St. dev.	26.90 kr
Temperature	Mean	10.28 C	St. dev.	7.52 C
Rain	Yes	22.2%	No	77.8%

5.3 ESTIMATION RESULTS

5.3.1 SELECTION OF THE NUMBER OF CLASSES

As the number of classes cannot be estimated endogenously, latent class choice models were estimated with the number of lifestyle varying between 2 and 6. It should be noted that also an error component logit specification without segmentation of the individuals according to lifestyle was estimated, and that the class specific and class membership models shared the same specification in order to isolate the effect of the varying number of classes.

The model performances of the different models are presented in Table 14 and the statistics supporting the selection of the number of classes were the rho-bar squared, the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). All the statistics are based on the same principle of evaluating the goodness-of-fit of each model as measured by the log-likelihood at estimates with respect to the parsimony as measured by the number of estimated parameters. However, different statistics suggest that different models are preferable in terms of goodness-of-fit vs. parsimony. On the one hand, increasing the number of parameters implies an increase in the goodness-of-fit when the evaluation is based on the rho-bar squared and the AIC, although the rate of improvement in performances significantly diminishes when estimating 5 and 6 latent classes. On the other hand, the same phenomenon is not observed when the evaluation is based on the BIC, as this statistic imposes a harsher penalty on the lack of parsimony. Given that the BIC suggests that the 4-class choice model gives the better balance between goodness-of-fit and parsimony, and that the behavioural interpretation appears easier and logical for class specific behaviour and class membership of the 4-class choice model, estimates for this model are presented in the remainder of this section.

TABLE 14: PERFORMANCES OF THE ESTIMATED CHOICE MODELS

	EC logit	**Latent class choice models**				
number of classes	1	2	3	4	5	6
number of parameters	30	92	137	182	227	272
log-likelihood at zero	-17,675	-17,675	-17,675	-17,675	-17,675	-17,675
log-likelihood at estimates	-11,530	-10,743	-10,269	-10,030	-9,875	-9,804
rho-bar squared	0.346	0.387	0.411	0.422	0.428	**0.430**
AIC	-23,120	-21,670	-20,811	-20,425	-20,203	**-20,153**
BIC	-23,339	-22,342	-21,812	**-21,754**	-21,861	-22,139

For sake of comparison, Table 15 illustrates the estimates of the error component logit model without latent lifestyle segmentation. Parameter estimates show that the sample has comparable sensitivity to travel time by bicycle and car, significant sensitivity to good weather conditions and hilliness when using active travel modes, and preferences for specific modes according to the trip purpose.

TABLE 15: ESTIMATES OF THE EC LOGIT MODEL (WITHOUT LATENT LIFESTYLE SEGMENTATION)

Variables	**estimate**	**t-statistic**
travel time - walk	*-0.085*	*-9.73*
travel time - bicycle	*-0.075*	*-9.25*
travel time - car driver	*-0.073*	*-7.98*
travel time - car passenger	*-0.089*	*-2.58*
waiting time - public transport	*-0.044*	*-2.26*
access/egress time - public transport	*-0.054*	*-2.54*
in vehicle time - public transport	-0.023	-1.84
number of transfers - public transport	*-0.736*	*-10.65*
travel cost	-0.052	-1.92
temperature - walk	*0.028*	*2.25*
temperature - bicycle	*0.077*	*5.67*
precipitation - walk	*-0.237*	*-3.14*
precipitation - bicycle	*-0.141*	*-2.35*
hilliness - bicycle	*-0.098*	*-2.55*
parking availability - car driver	*0.687*	*3.04*
monthly pass - public transport	*0.897*	*3.89*
commuting purpose - bicycle	0.104	1.41
commuting purpose - car driver	*0.522*	*2.68*
commuting purpose - public transport	0.092	0.27
leisure purpose - bicycle	*-0.815*	*-2.93*
leisure purpose - car driver	-0.405	-1.83
leisure purpose - public transport	-0.452	-1.40
shopping purpose - bicycle	*-0.405*	*-2.58*
shopping purpose - car driver	*0.536*	*2.37*
alternative specific constant - walk	*1.185*	*6.60*
alternative specific constant - car driver	*-1.482*	*-8.22*
alternative specific constant - car passenger	*-2.597*	*-13.97*
alternative specific constant - public transport	*-2.590*	*-12.77*
standard deviation on active travel	*1.090*	*11.25*
standard deviation on motorised travel	*1.264*	*11.73*

5.3.2 THE 4-CLASS MODEL: CLASS SPECIFIC BEHAVIOUR

Table 16 presents the parameter estimates of the class specific choice models with the same specification of the error component logit without latent class segmentation. The two parameters capturing the correlation across modes and the panel effect across individuals are restricted to be equal across the four lifestyle groups for parsimony and identification reasons, and are both significant to indicate that indeed unobservable similarities should have been accounted for in the model specification.

It is evident that the 4-class choice model is better than the 1-class model not only from the perspective of the goodness-of-fit, but also from the perspective of unravelling the heterogeneity in the preferences across individuals. It should be noted that several parameters are significant at the 0.05 and 0.10 confidence level (see the estimates in italic), and also that several parameters are significantly different across classes according to a Wald statistic test (see the note to the table). The examination of the estimated parameters allows the definition of the lifestyle groups, especially when looking at the ratios between the level-of-service estimates for the different travel modes across the different latent lifestyles.

TABLE 16: ESTIMATES OF THE CLASS SPECIFIC CHOICE MODEL

	Lifestyle independent		Lifestyle 1		Lifestyle 2		Lifestyle 3		Lifestyle 4	
Variables	estimate	t-statistic	estimate	t-statistic	estimate	t-statistic	estimate	t-statistic	estimate	t-statistic
travel time - walk [a]			*-0.129*	*-2.99*	-0.074	-1.54	*-0.058*	*-2.03*	*-0.085*	*-1.71*
travel time - bicycle [a]			*-0.129*	*-3.36*	*-0.054*	*-3.81*	*-0.087*	*-2.10*	*-0.072*	*-2.01*
travel time - car driver [a]			*-0.049*	*-2.37*	*-0.126*	*-2.74*	*-0.129*	*-2.67*	-0.082	-1.73
travel time - car passenger [a]			-0.056	-1.47	-0.112	-1.45	*-0.150*	*-2.29*	*-0.116*	*-1.78*
waiting time - public transport [a]			*-0.075*	*-2.25*	*-0.022*	*-2.69*	*-0.014*	*-2.85*	*-0.095*	*-1.85*
access/egress time - public transport			*-0.059*	*-1.83*	*-0.052*	*-4.72*	*-0.033*	*-3.28*	*-0.091*	*-2.66*
in vehicle time - public transport			*-0.038*	*-1.75*	-0.023	-1.65	*-0.013*	*-2.62*	-0.051	-1.85
number of transfers - public transport [a]			*-1.157*	*-4.23*	*-0.416*	*-1.69*	*-0.188*	*-1.89*	*-2.165*	*-4.47*
travel cost			-0.077	-1.51	*-0.072*	*-2.55*	-0.032	-1.44	-0.073	-1.65
temperature - walk			0.029	1.01	*0.017*	*1.77*	0.030	0.90	*0.040*	*1.69*
temperature - bicycle			*0.205*	*2.26*	0.017	0.38	*0.126*	*1.88*	*0.088*	*5.11*
precipitation - walk			*-0.364*	*-2.38*	-0.187	-1.29	*-0.192*	*-1.86*	-0.214	-1.34
precipitation - bicycle			*-0.343*	*-2.02*	-0.151	-0.89	-0.162	-0.83	-0.254	-1.58
hilliness – bicycle [a]			*-0.443*	*-2.22*	*0.172*	*3.12*	0.217	1.40	-0.252	-1.49
parking availability - car driver			*0.460*	*2.04*	0.246	1.25	0.155	1.21	*0.774*	*6.72*
monthly pass - public transport [a]			0.359	1.56	0.252	1.59	*1.295*	*10.87*	*0.414*	*3.48*
commuting purpose - bicycle [a]			*-0.525*	*-2.39*	*0.730*	*3.99*	0.143	0.49	*0.287*	*2.30*
commuting purpose - car driver [a]			*1.092*	*4.20*	*-0.637*	*-2.45*	0.152	0.58	0.256	0.99
commuting purpose - public transport [a]			0.188	0.73	0.236	0.68	*0.690*	*2.50*	*-0.463*	*-2.00*
leisure purpose - bicycle [a]			*-1.671*	*-2.67*	-0.465	-1.34	-0.791	-1.70	*-0.838*	*-2.27*
leisure purpose - car driver [a]			*0.623*	*2.35*	-0.502	-1.03	*-0.698*	*-2.10*	*-1.016*	*-2.75*
leisure purpose - public transport [a]			-0.553	-0.98	-0.211	-0.35	*0.907*	*2.56*	*-0.728*	*-1.81*
shopping purpose - bicycle [a]			*-1.249*	*-3.18*	-0.227	-1.26	*-0.422*	*-2.02*	-0.373	-1.58
shopping purpose - car driver [a]			*0.819*	*2.26*	0.289	0.72	0.471	1.25	*0.894*	*1.98*
alternative specific constant - walk [a]			1.396	1.47	-1.113	-1.19	*3.388*	*2.69*	*1.206*	*2.84*
alternative specific constant - car driver [a]			*2.168*	*2.17*	*-5.479*	*-7.15*	*-4.500*	*-3.65*	0.677	1.35
alternative specific constant - car passenger [a]			1.039	1.18	*-5.434*	*-7.18*	*-4.224*	*-3.43*	*-0.895*	*-1.80*
alternative specific constant - public transport [a]			*-4.218*	*-4.18*	*-3.839*	*-3.71*	*4.701*	*3.81*	-1.239	-1.52
standard deviation on active travel	*1.112*	*10.82*								
standard deviation on motorised travel	*1.315*	*10.32*								

Note: [a] the parameters vary significantly across lifestyle groups (Wald statistic at the 0.10 confidence level)

Lifestyle group 1 is oriented towards the use of the automobile, as emerging from the rate of substitution equal to 2.63 between car driver and bicycle parameters and equal to 2.30 between car passenger and bicycle parameters. Specifically, the individuals with this lifestyle evaluate 1 minute of traveling as car drivers as equal to 2.63 minutes of traveling as cyclists, and hence they will likely never cycle to reach their destination. Also, this lifestyle group exhibits high disutility for adverse weather conditions when walking and cycling and for hilliness when needing a bicycle, expresses a positive preference for driving a car regardless of the purpose of the trip chain, and manifests a negative preference for cycling especially for leisure and shopping purposes. Clearly, individuals with lifestyle 1 consider the car as the fastest, cheapest, most comfortable, and most convenient travel mode.

Lifestyle group 2 is oriented toward the use of the bicycle, as unravelling from the rate of substitution equal to 2.33 between bicycle and car driver parameters and equal to 1.78 between bicycle and public transport parameters. Namely, the individuals with this lifestyle perceive 1 minute on the bicycle far better than the time spent in an automobile or a public transport vehicle and rates of substitution almost inverse with respect to the individuals with lifestyle 1. Weather conditions are not significantly related to the choice of walking and cycling, meaning that bicycle oriented individuals would not care whether it is too hot, too cold or too wet when they need to travel. Also hilliness is not significantly related to cycling, most likely because bicycle oriented individuals might enjoy the possibility of exercise that some hills offer. Commuting is the purpose that this lifestyle group perceives as preferable when cycling, while non-significant relations are observed for leisure and shopping trip chains. Evidently, individuals with lifestyle 2 consider the bicycle as the fastest, most direct and most enjoyable travel mode.

Lifestyle group 3 is oriented toward the use of walk and public transport, as transpiring from the rate of substitution equal to 2.14 between public transport and car driver parameters, and equal to 1.43 between public transport and bicycle parameters. Clearly, the individuals with this lifestyle perceive 1 minute in a public transport vehicle better than the time spent in a car, and slightly better than the time spent on a bicycle. Moreover, their perception of the transfer penalty is lower than the one of the previous two lifestyle groups: the penalty is 3.10 minutes per transfer for lifestyle 3, while it is equal to 4.31 minutes per transfer for lifestyle 2 and 6.10 minutes per transfer for lifestyle 1. Individuals in the lifestyle group 3 have made the conscious decision of purchasing a monthly card to use public transport, they are sensitive to adverse weather conditions when cycling, and even though they have positive preferences for public transport commuting and leisure travel, they recognize the convenience of the automobile for shopping trip chains. Noticeably, individuals with lifestyle 3 consider walking and public transport as the preferable options in terms of time saving and comfort.

Lifestyle group 4 does not appear to have a clear orientation towards a specific travel mode, and the rates of substitution between bicycle and car driver parameters, bicycle and car passenger parameters, and bicycle and walking parameters are in the proximity of the unit. However, individuals in this lifestyle group have clearly a high negative sensitivity to public transport as shown by the highest transfer penalty of 9.14 minutes per transfer and by the highest rates of substitution of 3.27 between bicycle and public transport parameters and 2.90 between car driver and public transport parameters. Individuals with this lifestyle clearly dislike public transport, have comparable preferences for the other travel modes regardless of being active or motorised and regardless of the trip purpose. Manifestly, individuals with lifestyle 4 show an aversion for public transport and comparable values of the other modes.

Summarizing, lifestyle group 1 is automobile oriented, lifestyle group 2 is bicycle oriented, lifestyle group 3 is walk and public transport oriented, and lifestyle group 4 is public transport averse.

5.3.3 THE 4-CLASS MODEL: CLASS MEMBERSHIP MODEL

After presenting the estimates of the class specific choice models that allow illustrating the heterogeneity of travel behaviour across individuals and labelling the latent lifestyle groups, Table 17 presents the estimates of the class membership choice model that allows observing whether socio-economic-demographic characteristics of the individuals in the sample are predictors of the latent lifestyle belonging. It should be noted that several parameters are significant at the 0.05 and 0.10 confidence level (see the estimates in italic), and also that most parameters are significantly different across classes according to a Wald statistic test (see the note to the table).

TABLE 17: ESTIMATES OF THE CLASS MEMBERSHIP MODEL

	Lifestyle 1		Lifestyle 2		Lifestyle 3		Lifestyle 4	
	36%		27%		15%		22%	
Variables	estimate	t-statistic	estimate	t-statistic	estimate	t-statistic	estimate	t-statistic
constant [a]	*-0.617*	*-4.09*	*-0.550*	*-1.82*	*1.426*	*2.97*	-0.259	-0.86
male [a]	*0.529*	*5.45*	0.220	1.32	*-0.994*	*-2.69*	0.246	1.39
age 18-30 (piecewise)	0.004	0.74	*0.016*	*4.37*	*-0.024*	*-2.43*	0.003	0.82
age 31-60 (piecewise)	*0.012*	*2.40*	*-0.012*	*-1.71*	0.004	0.36	-0.003	-0.75
age 60 plus (piecewise)	*0.017*	*2.83*	-0.008	-1.04	0.001	0.09	*-0.010*	*-2.00*
adults over 18 years old [a]	*0.450*	*2.18*	*-0.207*	*-1.82*	-0.101	-0.76	-0.142	-0.69
children under 5 years old [a]	*1.615*	*3.29*	*-0.277*	*-2.52*	*-0.971*	*-2.58*	*-0.367*	*-2.31*
children 5-10 years old [a]	*1.300*	*1.96*	*0.216*	*2.01*	*-0.781*	*-2.26*	*-0.734*	*-2.09*
children 10-15 years old [a]	1.145	1.56	*0.176*	*1.71*	*-0.847*	*-1.85*	*-0.473*	*-1.78*
number of cars [a]	*0.625*	*3.15*	*-0.436*	*-3.70*	-0.378	-1.41	0.190	1.50
income [a]	*0.162*	*4.82*	*-0.073*	*-2.63*	*-0.168*	*-1.68*	0.080	1.55
cph centre [a]	-0.502	-1.53	*0.385*	*2.95*	*0.241*	*2.21*	-0.125	-1.11
cph area [a]	*0.659*	*1.92*	-0.161	-1.07	0.169	1.45	*-0.668*	*-1.91*
student [a]	*-0.691*	*-2.06*	0.508	1.22	0.408	1.09	-0.227	-0.63
self-employed [a]	*0.745*	*2.06*	*-1.126*	*-2.78*	-0.396	-1.13	*0.777*	*2.39*
retired [a]	0.191	0.51	*0.754*	*1.70*	0.133	0.31	*-1.080*	*-3.04*
unemployed [a]	*-0.947*	*-2.38*	0.148	0.35	*1.171*	*2.82*	-0.373	-1.03

Note: [a] the parameters vary significantly across lifestyle groups (Wald statistic at the 0.10 confidence level)

Lifestyle group 1 is more likely to be composed of individuals who are male, are over their thirties, are living with other adults in their households, and also have small children. As they are car oriented, the probability of belonging to this lifestyle group logically increases with an increase of the income and of the number of cars in the household, the residence in the municipalities in proximity of Copenhagen rather than in the centre of Copenhagen, and the occupation being self-employed or salary-employed (note that the parameter associated with students is negative with respect to the salary-employed reference category). In other words, individuals with this lifestyle are more affluent, have growing families, have developing careers, and have residence outside the urban core.

Lifestyle group 2 is more probable to be made of individuals who are male, are in their twenties, and are not living with other adults in their households but might have children. As they are bicycle oriented, the likelihood of having this lifestyle reasonably decreases with higher income and higher number of cars, increases with the residence being in the city centre rather than the outskirts of the metropolitan area or the rural parts of the Copenhagen region, and increases with the occupation being a student or a retired worker. In other words, individuals with this lifestyle are both younger students or salary-employed workers who enjoy the vibrant city centre and older retired workers who have grown up children.

Lifestyle group 3 is more likely to be constituted by individuals who are female, and whose family composition, with or without adults and with or without children in the household, is not significantly correlated with the membership in this class. As they are walk and public transport oriented, the probability of being in this lifestyle group increases for residents of the city centre, students, salary-employed with respect to self-employed, and unemployed. In other words, it seems a bit more complex to profile the group that is however mainly constituted by younger female students or salary-employed workers without small children.

Lifestyle group 4 is more probable to be comprised of individuals of both genders who may be in their twenties and thirties, and do not have small children. As they are public transport averse, the likelihood of having this lifestyle does not relate to having higher income or higher number of cars, decreases in the outskirts of the metropolitan area while not significant difference exists for residence in either the city centre or the rural areas of the Copenhagen region, and decreases for retired workers but increases for self-employed workers. In other words, this lifestyle group is heterogeneous in income and residence location and most likely has the highest flexibility in choosing between bicycle and automobile.

The last piece of information estimated with the class membership model is the probability of belonging to the four lifestyle groups that is fairly split in the individuals in the sample: 36% for lifestyle 1, 27% for lifestyle 2, 15% for lifestyle 3, and 22% for lifestyle 4.

5.4 SUMMARY AND CONCLUSIONS

The current study has presented an analysis of the influence of lifestyle on mode choices in short trip chains in the Copenhagen region. The contribution of this study with respect to previous studies about short trips lies in the consideration of the heterogeneity in individual preferences, especially when comparing parameter estimates for travel time variables, and the relevance of lifestyle decisions on short-term choices. Very relevantly, the findings of this study highlight that analysing the potential for switching from the automobile to sustainable travel modes while considering homogeneous population is a simplistic assumption (see, e.g., Banister, 2008; Monzon, 2011).

The findings of this study highlight that the population in the Copenhagen region is composed of four heterogeneous types of individuals: (i) car oriented individuals are likely more affluent and careerist individuals who have made the conscious decision of buying a car, quite an expensive endeavour in Denmark given the high registration tax, and most likely use the car in every trip chain regardless of the distance; (ii) bicycle oriented individuals are likely younger and at a different stage in life with less children and more interest in the vibrant city centre, and most likely they use the bicycle in most of the trip chains regardless of the distance; (iii) walk and public transport oriented individuals have made the conscious decision of not using the automobile for their trip chains and at the same time do not exhibit a strong negative preference for the bicycle even though they prefer public transport for reaching their farthest destinations; (iv) public transport averse individuals form the most heterogeneous group and are the most flexible to the use of either the bicycle or the automobile depending on the lifestyle stage.

When thinking about promoting sustainable travel modes and even active travel modes, the findings of this study suggest that individuals in lifestyle groups 3 and 4 are the most likely to be swayed to move towards these travel options. While lifestyle group 3 already prefers walking as an alternative, it is evident that the probability of using the bicycle for longer distances than the walkable ones would not be too low when looking at the rates of substitution and the lower emphasis on weather conditions. Even more relevantly, while lifestyle group 4 really dislikes public transport, it is clear that the probability of using the bicycle is at least equal to the one of using the car when considering travel time (ceteris paribus). Accordingly, evaluating the potential of shifting from the automobile to sustainable and active travel modes should consider that any intervention should be directed towards these lifestyle groups in primis. Policies against the car use, such as congestion charging and high registration taxes, might shift individuals from having more disposable income and more cars (as having automobiles becomes more expensive) and possibly changing their lifestyle towards a sustainable one. These policies should also be accompanied by measures that make the travel time of modes alternative to the car much more convenient, as car oriented individuals penalize heavily the time spent on anything other than their automobiles.

When thinking about infrastructure improvements, for example looking at the heavy investments of the Copenhagen municipality in improving bicycle infrastructure, the reduction of the cycling travel time is assumed to make cycling more attractive with respect to alternative travel modes. However, the estimated rates of substitution of the travel time parameters for the car oriented lifestyle group show that the reduction should be by a factor of at least 2 in order for the individuals within this group to even consider the bicycle as a plausible alternative. The rates of substitution of the travel time parameters for the bicycle oriented lifestyle group illustrate that the reduction would not have any impact on the individuals in this group, since they already perceive the bicycle as the fastest and most convenient travel mode. Instead, the rates of substitution of the travel time parameters for the public transport oriented and averse lifestyle groups suggest that the balance could be moved towards cycling even with modest reductions.

Further avenues for research are identifiable. Firstly, the current study estimates a latent class choice model only on the basis of traditional travel survey data and hence does not look at the angle of attitudes and perceptions. An extension could involve the preparation of a survey that would capture the psychological aspects behind the lifestyles observed from the travel survey information. Secondly, the current study assumes utility maximisation for all the lifestyle groups and hence does not open to different behavioural paradigm. An extension could entail the estimation of a latent class model with different formulation of the class specific choice models (e.g., regret minimisation, lexicographic).

ACKNOWLEDGMENTS

The authors gratefully acknowledge the Danish Road Directorate for the financial support of the project “Effects of cycling policies” that the current study forms part of. A preliminary version of the study was presented at the IATBR conference in Windsor in July 2015.

5.5 REFERENCES

Anderson, M.K., Nielsen, O.A. and Prato, C.G. (2014). Multimodal route choice models of public transport passengers in the Greater Copenhagen Area. *EURO Journal on Transportation and Logistics*, doi: 10.1007/s13676-014-0063-3.

Banister, D. (2008). The sustainable mobility paradigm. *Transport Policy*, 15, 73-80.

Bina, M., Graziano, F. and Bonino, S. (2006). Risky driving and lifestyles in adolescence. *Accident Analysis and Prevention*, 38, 472-481.

Borken-Kleefeld, J., Fuglestvedt, J.S. and Berntsen, T. (2013). Mode, load, and specific climate impact from passenger trips. *Environmental Science and Technology*, 47, 7608-7614.

Carse, A., Goodman, A., Mackett, R.L., Panter, J. and Ogilvie, D. (2013). The factors influencing car use in a cycle-friendly city: the case of Cambridge. *Journal of Transport Geography*, 28, 67-74.

Choo, S. and Mokhtarian, P.L. (2004). What type of vehicle do people drive? the role of attitude and lifestyle in influencing vehicle type choice. *Transportation Research Part A: Policy and Practice*, 38, 201-222.

de Hartog, J.J., Boogaard, H., Nijland, H. and Hoek, G. (2010). Do the health benefits of cycling outweigh the risks? *Environmental Health Perspectives*, 118, 1109-1116.

de Nazelle, A., Morton, B.J., Jerrett, M. and Crawford-Brown, D. (2010). Short trips: An opportunity for reducing mobile-source emissions? *Transportation Research Part D: Transport and Environment*, 15, 451-457.

de Nazelle, A., Nieuwenhuijsen, M.J., Antó, J.M., Brauer, M., Briggs, D., Braun-Fahrlander, C., Cavill, N., Cooper, A.R., Desqueyroux, H., Fruin, S., Hoek, G., Panis, L.I., Janssen, N., Jerrett, M., Joffe, M., Andersen, Z.J., van Kempen, E., Kingham, S., Kubesch, N., Leyden, K.M., Marshall, J.D., Matamala, J., Mellios, G., Mendez, M., Nassif, H., Ogilvie, D., Peiró, R., Pérez, K., Rabl, A., Ragettli, M., Rodríguez, D., Rojas-Rueda, D., Ruiz, P., Sallis, J.F., Terwoert, J., Toussaint, J.F., Tuomisto, J., Zuurbier, M. and Lebret, E. (2011). Improving health through policies that promote active travel: a review of evidence to support integrated health impact assessment. *Environment International*, 37, 766-777.

Fan, Y. and Khattak, A. (2012). Time use patterns, lifestyles, and sustainability of nonwork travel behavior. *International Journal of Sustainable Transportation*, 20, 26-47.

Forsey, D., Nurul Habib, K., Miller, E.J. and Shalaby, A. (2014). Temporal transferability of work trip mode choice models in an expanding suburban area: the case of York Region, Ontario. *Transportmetrica A: Transport Science*, 10, 469-482.

Frank, L., Stone, B.J. and Bachman, W. (2000). Linking land use with household vehicle emissions in the central Puget sound: Methodological framework and findings. *Transportation Research Part D: Transport and Environment*, 5, 173-196.

Frenkel, A., Bendit, E. and Kaplan, S. (2013a). The linkage between the lifestyle of knowledge-workers and their intra-metropolitan residential choice: A clustering approach based on self-organizing maps. *Computers Environment and Urban Systems*, 39, 151-161.

Frenkel, A., Bendit, E. and Kaplan, S. (2013b). Residential location choice of knowledge-workers in a "startup metropolis": the role of amenities, workplace and lifestyle. *Cities*, 35, 33-41.

Fuglestvedt, J.S., Shine, K.P., Berntsen, T., Cook, J., Lee, D.S., Stenke, A., Skeie, R.B., Velders, G.J.M. and Waitz, I.A. (2010). Transport impacts on atmosphere and climate: Metrics. *Atmospheric Environment*, 44, 4648-4677.

Gopinath, D.A. (1995). *Modeling Heterogeneity in Discrete Choice Processes: Application to Travel Demand*. Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Grabow, M., Spak, S., Holloway, T., Stone, B., Mednick, A. and Patz, J. (2012). Air quality and exercise-related health benefits from reduced car travel in the Midwestern United States. *Environmental Health Perspectives*, 120, 68-76.

Greene, W.H. and Hensher, D.A. (2003). A latent class model for discrete choice analysis: contrasts with mixed logit. *Transportation Research Part B: Methodological*, 37, 681-698.

Guerra, E. (2015). The geography of car ownership in Mexico City: a joint model of households' residential location and car ownership decisions. *Journal of Transport Geography*, 43, 171-180.

Hertel, O., Hvidberg, M., Ketzel, M., Storm, L. and Stausgaard, L. (2008). A proper choice of route significantly reduces air pollution exposure - A study on bicycle and bus trips in urban streets. *Science of the Total Environment*, 389, 58-70.

Hess, S. and Stathopoulos, A. (2013). A mixed random utility - random regret model linking the choice of decision rule to latent character traits. *Journal of Choice Modelling*, 9, 27-38.

Kim, S. and Ulfarsson, G.F. (2008). Curbing automobile use for sustainable transportation: analysis of mode choice on short home-based trips. *Transportation*, 35, 723-737.

Kitamura, R. (2009). Life-style and travel demand. *Transportation*, 36, 679-710.

Krizek, K.J. (2006). Lifestyles, residential location decisions, and pedestrian and transit activity. *Transportation Research Record*, 1981, 171-178.

Krizek, K.J. and Waddell, P. (2002). Analysis of lifestyle choices: Neighborhood type, travel patterns, and activity participation. *Transportation Research Record*, 1807, 119-128.

Lindsay, G., Macmillan, A. and Woodward, A. (2011). Moving urban trips from cars to bicycles: impact on health and emissions. *Australian and New Zealand Journal of Public Health*, 35, 54-60.

Loukopoulos, P. and Gärling, T. (2005). Are car users too lazy to walk? The relationship of distance thresholds for driving to the perceived effort of walking. *Transportation Research Record*, 1926, 206-211.

Mabit, S.L. and Fosgerau, M. (2011). Demand for alternative-fuel vehicles when registration taxes are high. *Transportation Research Part D: Transport and Environment*, 16, 225-231.

Mackett, R. (2003). Why do people use their cars for short trips? *Transportation*, 30, 329-349.

Maibach, E., Steg, L. and Anable, J. (2009). Promoting physical activity and reducing climate change: opportunities to replace short car trips with active transportation. *Preventive Medicine*, 49, 326-327.

Monzon, A., Vega, L.A. and Lopez-Lambas, M.E. (2011). Potential to attract drivers out of their cars in dense urban areas. *European Transport Research Review*, 3, 129-137.

Noland, R., Park, H., Von Hagen, L.A. and Chatman, D.G. (2014). A mode choice analysis of school trips in New Jersey. *The Journal of Transport and Land Use*, 7, 111-133.

Paulssen, M., Temme, D., Vij, A. and Walker, J.L. (2014). Values, attitudes and travel behavior: a hierarchical latent variable mixed logit model of travel mode choice. *Transportation*, 41, 873-888.

Piatkowski, D.P., Krizek, K.J. and Handy, S.L. (2015). Accounting for the short term substitution effects of walking and cycling in sustainable transportation. *Travel Behaviour and Society*, 2, 32-41.

Pinjari, A.R., Pendyala, R.M., Bhat, C.R. and Waddell, P.A. (2011). Modeling the choice continuum: an integrated model of residential location, auto ownership, bicycle ownership, and commute tour mode choice decisions. *Transportation*, 38, 933-958.

Pucher, J., Komanoff, C. and Schimek, P. (1999). Bicycling renaissance in North America? Recent trends and alternative policies to promote bicycling. *Transportation Research Part A: Policy and Practice*, 33, 625-654.

Rich, J., Prato, C.G., Hels, T., Lyckegaard, A. and Kristensen, N.B. (2013). Analyzing the relationship between car generation and severity of motor-vehicle crashes in Denmark. *Accident Analysis and Prevention*, 54, 81-89.

Rietveld, P. (2000). Non-motorised modes in transport systems: a multimodal chain perspective for the Netherlands. *Transportation Research Part D: Transport and Environment*, 5, 31-36.

Rojas-Rueda, D., de Nazelle, A., Tainio, M. and Nieuwenhuijsen, M.J. (2011). The health risks and benefits of cycling in urban environments compared with car use: health impact assessment study. *British Medical Journal*, 343, d4521.

Scheepers, E., Wendel-Vos, W., van Kempen, E., Panis, L.I., Maas, J., Stipdonk, H., Moerman, M., den Hertog, F., Staatsen, B., van Wesemael, P. and Schuit, J. (2013). Personal and environmental characteristics associated with choice of active transport modes versus car use for different trip purposes of trips up to 7.5 kilometers in the Netherlands. *PLOS One*, 8, e73105.

Scheiner, J. and Holz-Rau, C. (2007). Travel mode choice: affected by objective or subjective determinants? *Transportation*, 34, 487-511.

Schepers, J.P. and Heinen, E. (2013). How does a modal shift from short car trips to cycling affect road safety? *Accident Analysis and Prevention*, 50, 1118-1127.

Schwanen, T. and Mokhtarian, P.L. (2005). What affects commute mode choice: neighborhood physical structure or preferences toward neighborhoods? *Journal of Transport Geography*, 13, 83-99.

Sigurdardottir, S.B., Kaplan, S., Møller, M. and Teasdale, T. (2013). Understanding adolescents' intentions to commute by car or bicycle as adults. *Transportation Research Part D: Transport and Environment*, 24, 1-9.

Smith, B. and Olaru, D. (2013). Lifecycle stages and residential location choice in the presence of latent preference heterogeneity. *Environment and Planning A*, 45, 2495-2514.

Sun, Z., Arentze, T. and Timmermans, H. (2012). A heterogeneous latent class model of activity rescheduling, route choice and information acquisition decisions under multiple uncertain events. *Transportation Research Part C: Emerging Technologies*, 25, 46-60.

van Acker, V., Mokhtarian, P.L. and Witlox, F. (2014). Car availability explained by the structural relationships between lifestyles, residential location, and underlying residential and travel attitudes. *Transport Policy*, 35, 88-99.

Vij, A., Carrel, A. and Walker, J.L. (2013). Incorporating the influence of latent modal preferences on travel mode choice behaviour. *Transportation Resarch Part A: Policy and Practice*, 54, 164-178.

Vovsha, P., Vyas, G., Givon, D. and Birotker, Y. (2013). Individual mobility attributes and their impact on modality style comparison across three population sectors in Jerusalem, Israel. *Transportation Research Record*, 2382, 132-141.

Walker, J.L. (2001). *Extended Discrete Choice Models: Integrated Framework, Flexible Error Structures, and Latent Variables*. Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Walker, J.L. and Li, J. (2007). Latent lifestyle preferences and household location decisions. *Journal of Geographical Systems*, 9, 77-101.

Zhou, J. (2014). From better understandings to proactive actions: Housing location and commuting mode choices among university students. *Transport Policy*, 33, 166-175.

Chapter 6

HOME-END AND ACTIVITY-END PREFERENCES FOR ACCESS TO AND EGRESS FROM TRAIN STATIONS IN THE COPENHAGEN REGION

Title:	Home-end and activity-end preferences for access to and egress from train stations in the Copenhagen Region
Author(s):	Katrín Halldórsdóttir, Otto A. Nielsen, and Carlo G. Prato
Submitted:	Submitted to *International Journal of Sustainable Transportation* in September 2015.
Abbreviated:	Halldórsdóttir et al. (2015a)

ABSTRACT

Increasing public transport use with the aim of improving the sustainability of cities should not focus only on enhancing level and quality of the service offered, but also on understanding determinants of the choice of access and egress mode to and from the railway network. This study proposes a model that recognizes the difference in preference structure at the home-end and activity-end for travellers who have chosen train as their main travel mode, investigates the effect of policy variables such as car parking availability, park & ride opportunity, bicycle parking availability and type, and bicycle on train possibility, and accommodates the heterogeneity in the travellers' preferences for the alternative modes. Accordingly, this study analyses the choices between five transport modes (i.e., walking, cycling, being a car driver, being a car passenger, and riding a bus) for 2,921 home-end trips and 3,658 activity-end trips with a mixed logit model that accounts for heteroscedasticity across alternative modes and repeated observations across individuals. Model estimates uncover the importance of travel time and trip characteristics, underline the relevance of bicycle parking to the choice of cycling to the train station, but most importantly reveal that travellers have heterogeneous perceptions of the alternatives and the travel time, as well as their preference structure relates more to their socio-economic characteristics rather than the trip characteristics. In a nutshell, improving bicycle parking might certainly improve the accessibility to train stations, but addressing specific population groups with specific trip purposes might surely provide an even higher boost in the sustainability of the travel choices after selecting train as main transport mode.

Keywords: *access and egress mode; train stations; home-end; activity-end; public transport; mixed logit model*.

6.1 INTRODUCTION

Cities and metropolitan areas aim at curbing the use of private transport in favour of active travel modes and public transport solutions when targeting their sustainable future. Transit oriented development, namely the development process of housing, employment, activities and public services around existing or new railway stations served by frequent, efficient and high quality intra-urban rail services (see, e.g., Cervero, 1998; Knowles, 2012), appears as an important part of a smart growth approach to urban development in Europe, Asia, and the U.S. (see, e.g., Fullerton and Knowles, 1991; Cervero, 1998; Cervero and Murakami, 2009; Loo et al., 2010; Knowles, 2012; Goetz, 2013). Accordingly, increasing public transport use should not focus only on improving level and quality of the service offered, but also on understanding determinants of accessibility to and from railway stations. While the importance of the choice of access and egress mode to and from the railway network has been discussed in the literature (e.g., Keijer and Rietveld, 2000; Krygsman et al., 2004; Givoni and Rietveld, 2007; Brons et al., 2009; Bergman et al., 2011; Chakour and Eluru, 2014; Róman et al., 2014), this study adds to the growing body of knowledge by proposing a model of access to and egress from train stations in a transit oriented development region.

The Copenhagen Region is an example of transit oriented development around five suburban train lines (see, e.g., Knowles, 2006; Knowles, 2012) that constitute the preferred mode of public transport alongside the metro when considering passengers' perception in the complex public transport network including four types of buses (i.e., regular, high frequency, suburban, and express), four types of train services (i.e., suburban, local, regional, and intercity) and a metro system (Anderson et al., 2014). The current study focuses on the choice of access to and egress from train stations in the Copenhagen Region by building upon travel diaries collected within the Danish National Travel Survey (TU, in Danish Transportvaneundersøgelse). Most relevantly, unlike most of the existing literature focusing on the access to train stations, the current study proposes the analysis of the trips from the perspective that travellers do not have different preferences about accessing to or egressing from stations, but rather have different preference at the home-end and the activity-end of their train trips. Specifically, the current study hypothesizes that differences exist between the home- and the activity-end of their trips because of the different travel mode availability between home and activity location (e.g., access from home to the station with a bicycle would mean egress from the station to home with a bicycle), and the higher knowledge about road network, parking availability and station characteristics at the home-end than at the activity-end. Accordingly, the current study analyses the choices between five transport modes (i.e., walking, cycling, being a car driver, being a car passenger, and riding a bus) for 2,921 home-end trips and 3,658 activity-end trips with a mixed logit model that accounts for heteroscedasticity across alternative modes and repeated observations across individuals.

The current study has a threefold contribution: (i) the model recognizes the difference in preference structure at the home-end and activity-end for travellers who have chosen train as their main travel mode; (ii) the model investigates the effect of policy variables such as car parking availability, park & ride opportunity, bicycle parking availability and type, bicycle on train possibility, alongside socio-economic characteristics of the travellers and level of service measures of the travel modes; (iii) the model accommodates the heterogeneity in the travellers' preferences and alternative mode perceptions.

The remainder of this paper is structured as follows. Section 6.2 summarizes existing literature on the access and egress mode choice to train stations. Section 6.3 describes the data for model estimation and the model specification. Section 6.4 presents and discusses the model estimates, while section 6.5 presents the main conclusions and suggests further research directions.

6.2 EARLIER RESEARCH

Most of the existing literature has focused on the access to train stations with either a descriptive approach (e.g., Keijer and Rietveld, 2000; Rietveld, 2000; Martens, 2004; Martens, 2007) or a modelling approach (e.g., Wardman and Tyler, 2000; Krygsman et al., 2004; Givoni and Rietveld, 2007; Brons et al., 2009; Bergman et al., 2011; Chakour and Eluru, 2014; Martin et al., 2014).

Descriptive studies focused on the observation of specific aspects of travel behaviour. The effect of distance on the ability to walk and cycle to train station and the differences in the travel modes used at the home- and the activity-end were observed for Dutch commuters (Keijer and Rietveld, 2000). The bicycle was looked at as both a feeder mode and a combined mode with public transport (Rietveld, 2000; Martens, 2004; Martens, 2007), and the analysis of Dutch commuters' behaviour revealed that the bicycle was used more at the home-end (Rietveld, 2000), although bicycle lockers were not used much because expensive and the opportunity to bike and ride as well to flexibly rent bikes was offered (Martens, 2004; 2007).

Modelling studies focused on the determinants of mode choice to access train stations. Distance was the most relevant factor when considering between active travel and motorised modes (Krygsman et al., 2004; Givoni and Rietveld, 2007), but when exceeding a certain threshold it could dissuade travellers from travelling by train altogether (Krygsman et al., 2004; Brons et al., 2009), and when being part of a longer trip it could become less relevant (Wadrman and Tyler, 2000). Travel time was also a very relevant factor for the choice of access mode (Givoni and Rietveld, 2007; Chakour and Eluru, 2014) and travel time savings were evaluated as relevant with respect to the total travel time in Australia (Hensher and Rose, 2007) and Spain (Román et al., 2014). Car availability was not found relevant to the choice of access mode (Givoni and Rietveld, 2007), while habit was observed in the access to train stations (Bergman et al., 2011). Findings from a study that modelled jointly the choice of train station and access mode highlighted how the egress mode and the destination of the main train trip play a role into the travellers' preferences for access mode, while the distance and the parking availability play a role into the travellers' decisions for station choice (Chakour and Eluru, 2014).

The current study extends the body of literature by looking not only at the access mode, but also at the egress mode. Most relevantly, acknowledging some research observing differences in the behaviour of Dutch commuters at the home-end and the activity-end of their trip, the current study proposes a modelling approach that looks into the different preference structures at the two ends of the trips by estimating two mixed logit models for home-end and activity-end trips. Lastly, the current study specifies the models with a large variety of variables that extend the ones traditionally used in the literature (e.g., distance, travel time) with others sporadically used (e.g., socio-economic characteristics of travellers) and policy variables never used in the literature on access mode choice such as the availability of car parking, the possibility of park & ride, the opportunity and type of bicycle parking, and the possibility of riding the train with the bicycle. These additional policy variables are of great interest when searching for triggers of increased attractiveness of train stations in particular, and public transport in general.

6.3 Methods

6.3.1 Data

This section describes the data for the study of home-end and activity-end mode choice of access to and egress from the train stations of the Copenhagen Region. The home-end and activity-end trips were extracted from the travel diaries collected via the TU survey.

The TU survey collects travel diaries and socio-economic information of a representative sample of the Danish population between 10 and 84 years old via 1,000 interviews per month that are split into about 80% by telephone and about 20% on the internet since 2006 (Christiansen, 2009). The participants to the TU survey are extracted via a stratified random procedure from the Danish Civil Registration System (in Danish, Det Centrale Personregister) managed by the Danish National Board of Health (in Danish, Sundhedsstyrelsen) with the objective of reaching representativity of the population as listed in the Danish National Register managed by the Danish Census Bureau (in Danish, Danmarks Statistik). An external consultant supports the Department of Transport of the Technical University of Denmark in the administration and calibration of the representative sample. TU data are available for research purposes under an agreement with the Danish Data Protection Agency (in Danish, Datatilsynet).

The TU survey has implemented since 2009 an advanced data collection method for public transport trips that allows mapping every single part of the trip chain including details about stations for access, transfer, and egress, as well as details about transport modes for access, main, and egress part of the public transport trip (Anderson et al., 2014). The current study extracted the trips relevant for the analysis according to the following criteria: (i) the main public transport mode was train; (ii) the home-end part of the trip was identified (i.e., home – train station, and train station - home); (iii) the activity-end part of the trip was identified (i.e., train station – activity, and activity – train station).

The trips contained detailed information about: (i) characteristics of the trips in terms of mode, purpose, time-of-day, season, other travellers during the journey; (ii) level-of-service variables of the trips by walking, cycling, driving, being a passenger in a car, and riding a bus; (iii) socio-economic characteristics of the 1,743 travellers who performed 2,921 trips at the home-end, and the 1,909 travellers who performed 3,658 trips at the activity-end; (iv) policy variables detailing the availability of car parking, park & ride, bicycle parking, and the possibility to carry the bicycle on train.

Characteristics of the trips and the travellers were retrieved from the TU survey, while the level-of-service and the policy variables required calculation and on site investigation.

The level-of-service variables were calculated by knowing the network conditions at the time of the trip and calculating shortest path distances for walking and cycling, shortest path travel times conditional on the congested travel time from the Danish National Transport Model for driving and being a passenger in a car, and timetables supported by the TU survey for riding a bus (Anderson et al., 2014). It should be noted that the shortest path choices considered a speed function as a function of distance for walking and cycling to capture the fact that travellers walk or cycle faster for longer distances according to the observations in the TU survey. The speed function for walking is $4 + 4\cdot\text{distance} / 8000$, while the speed function for cycling is $6 + 14\cdot\text{distance} / 8000$ (where the speed is in km/h and the distance is in m).

The policy variables were retrieved by on site investigation and analysts' knowledge of the study area. Car parking availability was defined for each train station as the offer in terms of car parking spaces and average occupancy on the basis of time-of-day. Park & ride availability was defined for each of the 8,000 spaces available at train stations outside the inner city in the Copenhagen Region while considering an evaluation of their average occupancy on the basis of statistics provided by the Region. Bicycle parking availability was defined for each train station as the offer for different types of options, namely open bicycle racks, covered bicycle racks, locked bicycle parking places. Carrying a bicycle on a train was considered according to the specific public transport mode chosen (i.e., suburban train, local train, regional train, IC train, and metro) and the time-of-day restrictions (e.g., no bicycle carrying on metro or disembarking with a bicycle at highly congested train stations in rush hour). It should be noted that travellers are allowed to carry their bicycle at no additional costs for suburban and local trains in the Copenhagen Region, while an extra bicycle ticket should be bought for metro and regional trains.

6.3.2 MODEL FORMULATION AND SPECIFICATION

The current study formulates a mixed logit model specification (McFadden and Train, 2000) to represent the choice of travel mode at the home-end and activity-end of access and egress to train stations in the Copenhagen Region. The five alternatives of the choice model are walking, cycling, driving, being a car passenger, and riding a bus.

The mixed logit model specification was formulated in order to allow investigating whether heterogeneity exists in terms of preference structure across travellers as well as whether heteroscedasticity exists as travellers might perceive alternative modes differently and hence the error terms related to the various alternatives might be possibly characterised by different variance. Traveller n is assumed to maximise its utility U^E_{ni} at either end E of the trip (i.e., home-end H, and activity-end A) by choosing travel mode i:

$$U^E_{ni} = \alpha^E_i + \sum_{k=1}^{K} \beta^E_{ik} X^E_{nik} + \varepsilon^E_{ni} \tag{14}$$

where α^E_i is a vector of alternative specific constants for alternative modes i at end E, X^E_{nik} is a vector of characteristics k of alternative mode i as perceived by traveller n at end E, β^E_{ki} is a vector of parameters to be estimated, and ε^E_{in} is a vector of independently and identically distributed Gumbel error.

The probability P^E_{ni} of traveller n choosing alternative mode i at end E is formulated according to the well-known multinomial logit model, conditional on values for the alternative specific constants α_i and the taste parameters β^E_{ki}:

$$P^E_{ni} = \frac{\exp\left(\alpha^E_i + \sum_{k=1}^{K} \beta^E_{ik} X^E_{nik}\right)}{\sum_{j=1}^{J} \exp\left(\alpha^E_j + \sum_{k=1}^{K} \beta^E_{jk} X^E_{njk}\right)} \tag{15}$$

In the current study, the alternative specific constants α^E_i are assumed to be distributed according to a normal distribution $f(\alpha^E_i)$ that allows expressing heteroscedasticity across the alternatives, and the parameters β^E_{ki} are distributed according to a distribution $f(\beta^E_{ki})$ that allows representing heterogeneity across travellers. For model identification purposes, four alternative specific constants are formulated as $\alpha^E_i \sim \mathrm{N}(\mu^E_i, \sigma^E_i{}^2)$, where μ^E_i is the mean and $\sigma^E_i{}^2$ is the variance of the normal distribution of each constant. For heterogeneity representation purposes, parameters are tested with different distributions also on the basis of expected sign restrictions (e.g., lognormal distributions for time parameters). Accordingly, the probability of traveller n selecting alternative i may be integrated over the distributions $f(\alpha^E_i)$ and $f(\beta^E_{ki})$ of the random parameters:

$$P^E_{ni} = \int \frac{\exp\left(\alpha^E_i + \sum_{k=1}^{K} \beta^E_{ik} X^E_{nik}\right)}{\sum_{j=1}^{J} \exp\left(\alpha^E_j + \sum_{k=1}^{K} \beta^E_{jk} X^E_{njk}\right)} f\left(\alpha^E_i\right) f\left(\beta^E_{ik}\right) d\alpha^E_i \, d\beta^E_{ik} \tag{16}$$

As the probability does not have a closed-form expression and the integral is multi-dimensional, the maximisation of the log-likelihood function requires simulation that consists in maximising the simulated log-likelihood SLL over the sample of travellers:

$$SLL = \sum_{n=1}^{N} \sum_{i=1}^{J} d_{ni} \ln \left\{ \frac{1}{R} \sum_{r=1}^{R} \left[\frac{\exp\left(\alpha^E_i + \sum_{k=1}^{K} \beta^E_{ik} X^E_{nik}\right)}{\sum_{j=1}^{J} \exp\left(\alpha^E_j + \sum_{k=1}^{K} \beta^E_{jk} X^E_{njk}\right)} \right] \right\} \tag{17}$$

where N is the number of travellers, J is the number of alternative modes, d_{ni} is equal to 1 if traveller n has selected alternative mode i and 0 otherwise, r is one of the R random draws required for integral simulation, and the superscript r represents the instance of a draw of the random parameters α^E_i and β^E_{ki}.

The parameters are estimated in the present study by using 500 random draws from a Modified Latin Hypercube Sampling (MLHS) method (Hess et al., 2006) that allows overcoming the correlation patterns that would emerge with Halton draws for the multi-dimensional integral. The freeware software BIOGEME (Bierlaire, 2008) was used for model estimation thanks to its simplicity and versatility in specifying the model formulated for this analysis.

6.4 RESULTS

6.4.1 SAMPLE CHARACTERISTICS

The sample included 1,743 travellers performing 2,921 trips at the home-end, and 1,909 travellers performing 3,658 trips at the activity-end. The different number of travellers and trips is obviously related to the fact that some travellers might have used the train as main transport mode only in either end. Table 18 presents the characteristics of the trips and Table 19 introduces the characteristics of the travellers at both the home- and the activity-end. Table 18 shows that the access and egress mode counts more cycling at the home-end and more walking at the activity-end, in line with previous findings for Dutch commuters (Keijer and Rietveld, 2000; Rietveld, 2000). The distribution of the trips by purpose is comparable between the two ends, and the bicycle is rarely brought on the train although there are opportunities and policies to promote them similar to the Dutch ones (Martens, 2004; Martens, 2007). Although the majority of the travellers have a driving license, the use of the car is quite scarce for accessing train stations, also in line with previous results (Givoni and Rietveld, 2007).

TABLE 18: TRIP CHARACTERISTICS

		Home-end (N = 2,921) N (%)	Activity-end (N = 3,658) N (%)
Trip characteristics			
Main train mode	S-train	1,406 (48)	1,720 (47)
	Local or regional train	555 (19)	705 (19)
	Metro	370 (13)	527 (14)
	Multiple trains	590 (20)	706 (19)
Access/egress mode choice	Walk	1,484 (51)	2,826 (77)
	Bicycle	710 (24)	211 (6)
	Car driver	142 (5)	0 (0)
	Car passenger	137 (5)	101 (3)
	Bus	448 (15)	520 (14)
Trip purpose	Work	1,290 (44)	1,616 (44)
	Study	410 (14)	509 (14)
	Shopping	324 (11)	392 (11)
	Errand	44 (2)	52 (1)
	Leisure	853 (29)	1,089 (30)
Chosen train/bicycle combination	Took the bicycle on the train	124 (4)	144 (4)
	Lockable cycle parking	56 (2)	12 (0)
	Covered bicycle rack	176 (6)	21 (1)
	Bicycle rack in the open	262 (9)	28 (1)
	On-street bicycle parking	92 (3)	6 (0)
	Did not cycle	2,211 (76)	3,447 (94)
Fellow travellers	Yes	614 (21)	822 (22)
	No	2,307 (79)	2,836 (78)
Urban characteristics	Copenhagen and Frederiksberg	1,172 (40)	2,376 (65)
	Copenhagen Region	1,749 (60)	1,282 (35)
Policy variables			
Park & Ride	Yes	423 (14)	309 (8)
	No	2,498 (86)	3,349 (92)
Bicycle parking in other stations	Locked bicycle parking	1,220 (50)	1,354 (47)
	Covered bicycle rack	827 (34)	708 (25)
	Bicycle racks	384 (16)	790 (28)
Bicycle parking in metro stations	Locked bicycle parking	27 (6)	58 (7)
	Covered bicycle rack	285 (58)	512 (64)
	Bicycle racks	178 (36)	236 (29)
Station type	Metro	490 (17)	806 (22)
	Other	2,431 (83)	2,852 (78)

Level-of-service variables			
Walk	Travel time (min) – mean	10.5	10.0
	Travel time (min) – st. dev.	6.3	6.9
Bicycle	Travel time (min) – mean	10.5	11.6
	Travel time (min) – st. dev.	3.9	4.9
Car driver	Travel time (min) – mean	13.7	-
	Travel time (min) – st. dev.	15.4	-
	Parking walk time (min) – mean	2.2	-
	Parking walk time (min) – st. dev.	0.6	-
Car passenger	Travel time (min) – mean	13.2	14.9
	Travel time (min) – st. dev.	17.1	20.3
Public transport	Waiting time (min) – mean	25.6	19.4
	Waiting time (min) – st. dev.	33.8	25.9
	Access/egress time (min) – mean	8.5	8.0
	Access/egress time (min) – st. dev.	4.1	3.5
	In-vehicle time (min) – mean	9.7	8.7
	In-vehicle time (min) – st. dev.	7.3	7.1

When looking at the level-of-service variables for the access and egress modes at both ends, a few observations apply: the car is never selected and is not considered available for driving to travellers at the activity-end; the car driving mode has a parking time that is an average estimation given the size of the parking lot at each train station; the waiting times for the bus are quite high because often these are feeder services with low frequency and hence appear as high average waiting times in the planning model used for the calculations.

The sample contains a majority of women (using train as main transport mode), not surprising when looking at a higher percentage of men using cars as main transport mode. Students and employed are the vast majority, in line with recent results showing that men who are more affluent and in career are mostly car oriented, while females and students are more public transport oriented (Prato et al., 2015).

TABLE 19: TRAVELLERS' CHARACTERISTICS

Variables	Categories	Home-end *(N = 1,743), N (%)*	Activity-end *(N = 1,909), N (%)*
Individual characteristics			
Age group	10-17	181 (10)	194 (10)
	18-24	323 (19)	366 (19)
	25-29	155 (9)	167 (9)
	30-39	291 (17)	321 (17)
	40-49	272 (16)	290 (15)
	50-59	235 (13)	271 (14)
	60-69	195 (11)	208 (11)
	70 and older	91 (5)	92 (5)
Gender	Male	717 (41)	785 (41)
	Female	1,026 (59)	1,124 (59)
Main occupation	Student	519 (30)	564 (30)
	Retired	194 (11)	196 (10)
	Unemployed	46 (3)	56 (3)
	Employed	937 (54)	1,039 (54)
	Self-employed	47 (3)	54 (3)
Traveller has a driving license	Yes	1,167 (67)	1,286 (67)
	No	576 (33)	623 (33)
Traveller has a bicycle	Yes	1,374 (79)	1,483 (78)
	No	369 (21)	426 (22)
Traveller has a public transport pass	Yes	1,139 (65)	1,241 (65)
	No	604 (35)	668 (35)
Household characteristics			
Vehicle availability	Zero car	724 (42)	739 (39)
	One car	794 (46)	897 (47)
	Two cars	208 (12)	243 (13)
	Many cars	17 (1)	30 (2)
Children 9 years old or younger	Yes	295 (17)	324 (17)
	No	1,448 (83)	1,585 (83)
Children 10-17 old years	Yes	497 (29)	544 (28)
	No	1,246 (71)	1,365 (72)

6.4.2 MODEL ESTIMATES

Initially, the availability of the alternative travel modes was considered for having a reasonable model specification: walking, being a car passenger, and riding a bus, was considered available to all travellers at both ends; cycling was deemed available to travellers owning a bicycle at both ends, as especially Danish commuters keep another bicycle at the activity-end; car driving was judged available to travellers owning a car and holding a driving license at the home-end, but it was not considered available at the activity-end.

The best model specification was then selected by iteratively testing the significance of the parameter estimates and their distribution when random parameters were used. Table 20 presents the estimates of the best models for each of the trip ends, where it should be noted that heteroscedasticity was accounted for as all the alternative specific constants had significant estimates for the mean and the standard deviation of their respective normal distributions, and heterogeneity was observed for the travel time parameter that was log-normally distributed. Moreover, the specification with the random parameters allowed taking into account panel effects by imposing the consistency of the preferences within individuals.

TABLE 20: MIXED LOGIT MODEL ESTIMATES

Variable	Categories	Alternative	Home-end Value	Home-end t-test		Activity-end Value	Activity-end t-test	
Alternative specific constant		Walk	0	—		0	—	
	μ	Bicycle	-14.90	-5.64	***	-18.20	-5.20	***
	σ		30.60	6.11	***	14.70	4.37	***
	μ	Car driver	-11.70	-3.33	***	—	—	
	σ		13.30	4.69	***	—	—	
	μ	Car passenger	-20.60	-6.38	***	-22.10	-3.32	***
	σ		6.41	5.77	***	8.27	2.06	**
	μ	Bus	-5.41	-6.88	***	-7.67	-7.39	***
	σ		1.87	4.80	***	2.03	2.86	***
Travel time	Travel time μ	Generic	-0.97	-7.69	***	-0.68	-6.38	***
	Travel time σ		0.13	3.26	***	0.28	4.38	***
	Access/egress time	Bus	-0.26	-5.26	***	-0.30	-5.91	***
Trip characteristics								
Fellow traveller	—	Bicycle	—	—		-5.23	-3.98	***
		Car passenger	4.77	4.41	***	6.47	1.98	**
		Bus	-0.52	-1.69	*	—	—	
Urban characteristics	Copenhagen / Frederiksberg	Bicycle	—	—		-4.60	-3.04	***
		Car driver	—	—		—	—	
		Car passenger	-3.77	-3.25	***	-6.59	-1.81	*
		Bus	1.37	4.18	***	—	—	
Policy variables								
Parking at stations	Bicycle parking *Metro station*	Bicycle	-17.60	-5.23	***	—	—	
	Covered bicycle rack *Other stations*	Bicycle	—	—		2.90	2.06	**
Bring bicycle on train	Pay	Bicycle	—	—		-4.10	-1.66	*
	Restricted	Bicycle	—	—		-6.46	-2.72	***
Travellers' characteristics								
Gender	Male	Bus	-0.39	-1.53		-1.08	-3.27	***
Season ticket	—	Bus	—	—		1.26	3.52	***
Vehicle availability	One car	Car passenger	6.45	4.79	***	—	—	
	Two cars	Car driver	9.37	3.54	***	—	—	
		Car passenger	8.75	4.95	***	—	—	
	Three cars or more	Car driver	22.6	3.01	***	—	—	
Other motorist	—	Car driver	-10.20	-2.54	***	—	—	

Main occupation and trip purpose								
Student	Study	Bike	—	—		-8.83	-3.21	***
		Car passenger	-2.77	-2.39	**	—	—	
		Bus	1.38	3.39	***	1.25	3.51	***
	Errand	Car passenger	—	—		8.86	1.69	*
	Shopping	Bus	1.11	1.45		—	—	
	Leisure	Bike	-9.62	-4.42	***	—	—	
		Car passenger	2.64	1.78	*	—	—	
Retired	Leisure	Bicycle	—	—		-10.30	-2.50	***
		Car passenger	—	—		5.72	3.39	***
		Bus	1.63	3.33	***	—	—	
Unemployed	Shopping	Bus	—	—		2.84	2.57	***
Self-employed	Leisure	Bus	—	—		2.07	1.90	*
Employed	Errand	Bike	—	—		2.69	1.56	
		Car passenger	—	—		8.14	2.15	**
	Leisure	Car driver	-6.52	-2.57	***	—	—	
		Car passenger	—	—		3.30	2.09	**
	Shopping	Car driver	-12.5	-4.84	***	—	—	
		Car passenger	-6.70	-2.30	**	—	—	
Number of estimated parameters:				31			28	
Number of observations:				2,921			3,658	
Number of individuals:				1,743			1,909	
Null log-likelihood:				-4,021.65			-4,668.66	
Final log-likelihood:				-1,916.68			-1,384.20	
Adjusted rho-square:				0.516			0.698	

Note: * significant at the 90% confidence level; ** significant at the 95% confidence level; *** significant at the 99% confidence level.

6.4.2.1 Travel times

Initial model estimation attempted to estimate a travel time parameter specific to each alternative travel mode, but the best specification was obtained when a generic travel time parameter with log-normal distribution was estimated and both the mean and the standard deviation of the distribution were statistically significant. The travel time significance is not surprising as the distance from the train station has been found as the most relevant determinant of access mode in the literature (e.g., Krygsman et al., 2004; Givoni and Rietveld, 2007; Chakour and Eluru, 2014).

As the walking time between park & ride facilities had a logical negative sign but was not statistically significant, it was joined to the car driving time to obtain a total time by driving. Interestingly, the in-vehicle time and the waiting time for the bus were also with logical negative sign but not statistically significant, while the access/egress time for the bus was significant. This suggests that travellers might be more sensitive to the connection time when reaching train stations, rather than to the actual in-vehicle time, and that they might be timing the low frequency feeder buses not to have to wait.

6.4.2.2 Trip characteristics

Model estimates show that the trip characteristics are also related to the access and egress mode choice at both the home-end and the activity-end. Specifically, for both ends the estimates show that having a fellow traveller decreases the likelihood of using the bus at the home-end, and diminishes the probability of cycling at the activity-end as observed also in the Netherlands (Rietveld, 2000; Givoni and Rietveld, 2007). The former might be related to the cost of multiple tickets that might be on top of the train costs, while the latter might be associated with the availability of bicycles for everyone provided that a person left a bicycle at the activity-end as many commuters are used to.

Model estimates show also that travellers are more likely to use the bus to access train stations in central Copenhagen and Frederiksberg at the home-end, most likely because of the higher level of service and frequency in the city centre. Also, travellers are less likely to use the bicycle for reaching train stations in central Copenhagen and Frederiksberg at the activity-end, probably because of the higher walkability of the city centre. Car driving or car passenger is highly discouraged by the congestion and the difficulty in finding parking in the central areas of the metropolitan area.

It should be noted that were not found significant effects of time-of-day and season, as to suggest that the choices to access and egress train stations are not related to the time and period in which the trips are performed, in line with recent evidence suggesting that lifestyle is what drives mode choices in the Copenhagen Region (Prato et al., 2015).

6.4.2.3 Policy variables

Parking availability for cars was not found related to the access and egress mode choice at both ends, given the small amount of travellers that chose to drive a car to the station. The same applies to park & ride facilities that are seldom used by Copenhageners during their trips, as to suggest that owning a car implies its use for the entire trip and not only to access public transport. Initial estimates suggested that travellers were more willing to drive a car at the home-end when there are dedicated park & ride facilities, although the estimate was not statistically significant at the 90% confidence level.

The bicycle parking availability was investigated by differentiating between metro stations and all other stations where open bicycle racks, covered bicycle racks and lockers were available. Estimates for the home-end model indicate clearly that even the presence of parking facilities does not encourage the use of the bicycle on the metro, most likely because travellers would have to pay to bring their bicycle or would have to leave the bicycle in crowded underground facilities as on-street bicycle parking is nearly impossible next to metro stations. Estimates for the activity-end show that the availability of covered bicycle parking increases the probability of cycling, thus suggesting that there might also be a choice of station to be considered, in line with existing literature about the choice of departure railway station being positively affected by the availability of bicycle parking (Debrezion et al., 2009). Initial estimates indicated a positive but non-significant effect also of lockers at both ends.

When considering at both ends the possibility to bring the bicycle on the train, it is clear that travellers would not opt for the bicycle unless it is free or there is no time restriction. These findings indicate strong preferences for cyclists and probably motivate also the selection of the type of train (as it is free for suburban and local trains).

Lastly, it was also investigated whether the popular campaign “bike to work” that occurs every May of every year in Denmark had any significant effect on the choice of cycling to the station. No parameter estimate was found significant, which is interesting since the intention of the campaign is to stimulate cycling among the non-cyclist, for example by proposing to combine a short distance cycling with public transport.

6.4.2.4 Travellers’ characteristics

Model estimates found significant correlations between travellers’ characteristics and mode choice to access/egress train stations. Gender and age effects were far less significant than the car ownership level and the occupation. The only gender effect is the dislike of males towards bus, which is significant at the activity-end and it is not surprising when looking at short trip mode choice in the same region (Prato et al., 2015).

Having a season ticket for public transportation or having the availability of one or more cars obviously affects the mode choice of travellers, even when analysing access to and egress from train stations. It seems that long-term decisions are influencing short-term ones as the ones considered in the current study. It is however interesting that the only significant parameters were found for the home-end of the trips, most likely because walking is the most preferred mode at the activity-end. It should also be noted that car availability was not found correlated with access mode to train stations in a previous study (Givoni and Rietveld, 2007), although that study did not configure the choice as a matter of being at the home- or the activity-end.

Having another motorist in the household was significantly related to a lower probability of driving a car to the train station at the home-end. Having children was not significantly correlated with the choice of access or egress mode, most likely because travelling alone or with someone else was already controlled for.

6.4.2.5 Main occupation and trip purpose

The specification of the model considered plausible combination of main occupation and trip purpose, as some combinations might not be realistic (e.g., being unemployed and having a work trip purpose).

The best model considered linear combinations of the main occupation and the trip purpose in order to take into account the relation between the two. The reference combination was composed of employed travellers who commute to work, while the remaining combinations of occupation (i.e. employed, student, retired, unemployed, and self-employed) and trip purpose (i.e., study, errand, shopping, and leisure).

When looking at the home-end model, students are less likely to be driven and more likely to use the bus when accessing train stations while going to study, most likely because the season ticket comes with a significant discount with respect to the full tariff. Students also prefer to use the bus for shopping purposes and they have a preference to be driven for leisure trips. Employed travellers do not exhibit a preference for the car, either as a driver or as a passenger, when not going to work and instead going for shopping or leisure.

When looking at the activity-end model, students are less likely to cycle and more likely to use the bus to access train stations. Being a car passenger is more probably for students and employed travellers running errands, as well as employed or retired travellers going out for leisure. Cycling is more likely for employed travellers running errands, but less likely for retired travellers during leisure trips.

6.5 Conclusions

The current study proposed the analysis of the mode choice behaviour for the access to and the egress from train station in the Copenhagen Region. Although similarity exists through the sample, the study presented two different models for the home-end and the activity-end, given different preference structures related to the different knowledge of the network and the area.

The model estimates revealed that there is heteroscedasticity, namely the alternative travel modes are perceived differently across individuals, and there is also taste heterogeneity in the perception of travel time across individuals. It should be noted that initial estimation of a Multinomial Logit model revealed some significant effects that were not significant in the best specification of the mixed logit model, and this finding underlines that overlooking the panel effect can lead to overestimating the effects of some variables.

The model estimates also reveal the relative importance of the elements that could make train stations more attractive, and consequently could support transit oriented development and could make public transport more enticing to travellers. Improving station accessibility, parking availability, bicycle parking possibility, and park & ride opportunity are logical policies that should make the train more attractive. However, it is clear that, although travel time was significant and was distributed over the population, the choices of access and egress mode were more related to who the travellers are as persons and how they perceive the alternatives.

Findings from this study reveal that it is important to have bicycle parking and to provide the opportunity to carry the bicycle on the train, but most of the explanatory ability of the model lies in the travellers' characteristics, especially when considering their occupation and their travel purpose. On the one hand it is certainly a good idea to increase bicycle parking availability at train stations, improve the conditions of bicycle parking especially at metro stations, but on the other hand travellers do not seem to be too concerned with locked parking areas as they are used to on-street parking. Most relevantly, finding from this study suggest that, similarly to what observed for short trips of which the access to and the egress from train stations are certainly a part of, a latent lifestyle segmentation approach could reveal what are the preferences of travellers in the Copenhagen Region and could suggest policy makers which population groups should be addressed when intending to improve the integration of active travel modes and public transport.

ACKNOWLEDGEMENTS

The authors acknowledge the support of Hjalmar Christiansen and Thomas Hummelshøj from the Data and Modelcenter at DTU Transport for the data collection and preparation. The authors gratefully acknowledge the financial support of the Danish Road Directorate to the project "Effects of cycling policies".

6.6 REFERENCES

Anderson, M.K., Nielsen, O.A. and Prato, C.G. (2014). Estimation of multimodal public transport route choice models of transport passengers in the Greater Copenhagen Area. *EURO Journal in Transportation and Logistics*, doi: 10.1007/s13676-014-0063-3.

Bergman, A., Gliebe, J. and Strathman, J. (2011). Modeling access mode choice for inter-suburban commuter rail. *Journal of Public Transportation*, 14, 23-42.

Bierlaire, M. 2008. *An introduction to BIOGEME.* Version 1.6, March 2008, Available at: http://biogeme.epfl.ch/

Brons, M., Givoni, M. and Rietveld, P. (2009). Access to railway stations and its potential in increasing rail use. *Transportation Research Part A: Policy and Practice,* 43, 136-149.

Cervero, R. (1998). *The Transit Metropolis: A Global Inquiry*. Island Press,Washington, DC, USA.

Cervero, R. and Murakami, J. (2009). Rail and property development in Hong Kong: experiences and extensions. *Urban Studies*, 46, 2019-2043.

Chakour, V. and Eluru, N. (2014). Analyzing commuter train user behavior: a decision framework for access mode and station choice. *Transportation*, 41, 211-228.

Christiansen, H. (2009). Modernisation of the Transportvaneundersøgelsen (in Danish). Proceedings of *Trafikdage*, Aalborg, Denmark.

Debrezion, G., Pels, E. and Rietveld, P. (2009). Modelling the joint access mode and railway station choice. *Transportation Research Part E: Logistics and Transportation Review*, 45, 270-283.

Fullerton, B. and Knowles, R. (1991). Scandinavia. Paul Chapman Publishing, London

Givoni, M. and Rietveld, P. (2007). The access journey to the railway station and its role in passengers' satisfaction with rail travel. *Transport Policy*, 14, 357-365.

Goetz, A.R. (2013). Suburban sprawl or urban centers: tensions and contradictions of smart growth approaches in Denver, Colorado. Urban Studies, 50, 2178-2195.

Hensher, D.A. and Rose, J.M. (2007). Development of commuter and non-commuter mode choice models for the assessment of new public transport infrastructure projects: A case study. *Transportation Research Part A*, 41, 428-443.

Hess, S., Train, K. and Polak, J.W. (2006). On the use of a Modified Latin Hypercube Sampling (MLHS) approach in the estimation of a Mixed Logit model for vehicle choice, *Transportation Research Part B: Methodological*, 40, 147-163.

Keijer, M.J.N. and Rietveld, P. (2000). How do people get to the railway station? The Dutch experience. *Transportation Planning and Technology*, 23, 215-235.

Knowles, R.D. (2006). Transport shaping space: differential collapse in time-space. *Journal of Transport Geography*, 14, 407-425.

Knowles, R.D. (2012). Transit Oriented Development in Copenhagen, Denmark: from the Finger Plan to Ørestad. *Journal of Transport Geography*, 22, 251-261.

Krygsman, S., Dijst, M. and Arentze, T. (2004). Multimodal public transport: an analysis of travel time elements and the interconnectivity ratio. *Transport Policy*, 11, 265-275.

Loo, B.P.Y., Chen, C. and Chan, E.T.H. (2010). Rail-based transit-oriented development: lessons from New York City and Hong Kong. *Landscape and Urban Planning*, 97, 202–212.

Martens, K. (2004). The bicycle as a feedering mode: experiences from three European countries. *Transportation Research Part D: Transport and Environment*, 9, 281-294.

Martens, K. (2007). Promoting bike-and-ride: The Dutch experience. *Transportation Research Part A: Policy and Practice,* 41, 326-338.

McFadden, D. and Train, K. (2000). Mixed MNL model for discrete response. *Journal of Applied Econometrics*, 15, 447-470.

Prato, C.G., Halldórsdóttir, K. and Nielsen, O.A. (2015). Latent lifestyle and mode choice decisions when travelling short distances. *Proceedings of the 14th International Conference on Travel Behaviour Research*, Windsor, UK. Submitted to Transportation in September 2015.

Rietveld, P. (2000). The accessibility of railway stations: the role of the bicycle in The Netherlands. *Transportation Research Part D: Transport and Environment*, 5, 71-75.

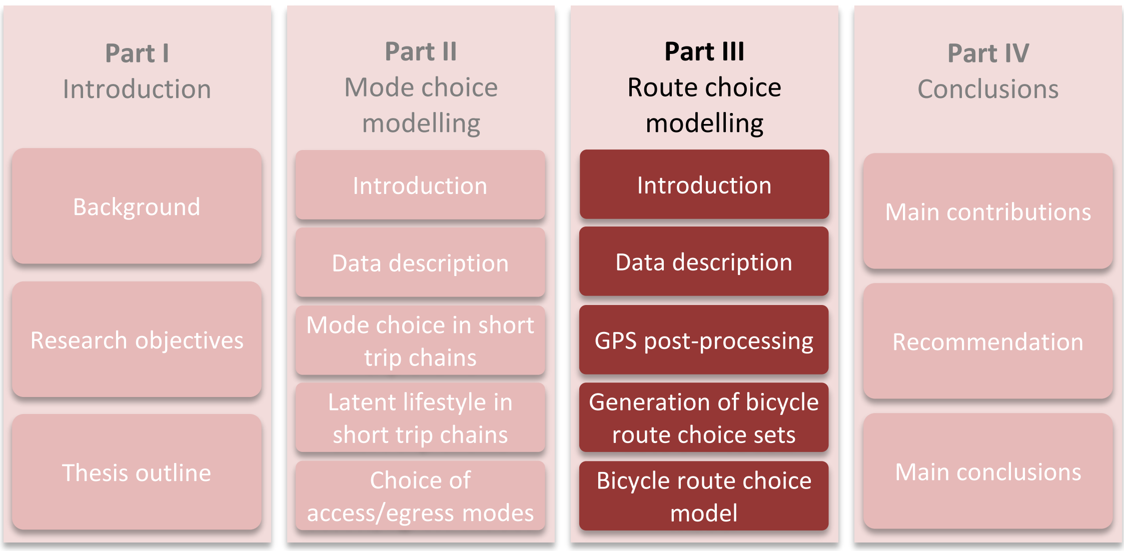
Román, C., Martin, J.C., Espino, R., Cherchi, E., Ortúzar, J.d.D., Rizzi, L.I., Gonzalez, R.M. and Amador, F.J. (2014). Valuation of travel time savings for intercity travel: The Madrid-Barcelona corridor. *Transport Policy*, 36, 105-117.

Wardman, M. and Tyler, J. (2000). Rail network accessibility and the demand for interurban rail travel. *Transport Reviews,* 20, 3-24.

PART III

ROUTE CHOICE MODELLING

THE ROUTE CHOICE OF CYCLISTS



Chapter 7

INTRODUCTION

INTRODUCTION TO PART III

Part III focuses on the modelling of the route choices of cyclists. The motivation behind this part is the growing interest in sustainable transport modes, where bicycle route choice models provide means to search for factors that make cycling more attractive. It is generally assumed that cyclists choose the shortest path by minimizing travel distance between origin and destination in operational travel forecasting models, using a fixed travel speed. Network attributes, such as the presence of dedicated bicycle lanes or separate bicycle boulevards, are usually not considered. Personal attributes are also generally not included, as well as elevation and other environmental attributes that could influence the route choice. Thus, investigating whether these factors significantly influence the route choices of everyday cyclists could be a step in the right direction to promote more cycling.

Results from bicycle route choice models depend on the observation of actual route choices and the generation of realistic alternatives. ***Chapter 8*** gives an overall description of the bicycle network, as well as briefly describing how actual route choices were collected, thus focusing on the third research objective:

(iii) To collect data using global positioning systems (GPS) technology to register geographical points, recording the behaviour of a sample of bicyclists from different municipalities in the Greater Copenhagen area.

The remaining research objectives are tackled within the three papers in Part III. The first paper addresses the issue with processing raw GPS data, the second evaluates methods to generate alternative bicycle route sets for model estimation, and the third paper investigates the route choice preferences in connection with the characteristics of alternative routes. Subsections 7.1 to 7.3 clarify which research objectives each paper concentrates on and briefly describe other contributions.

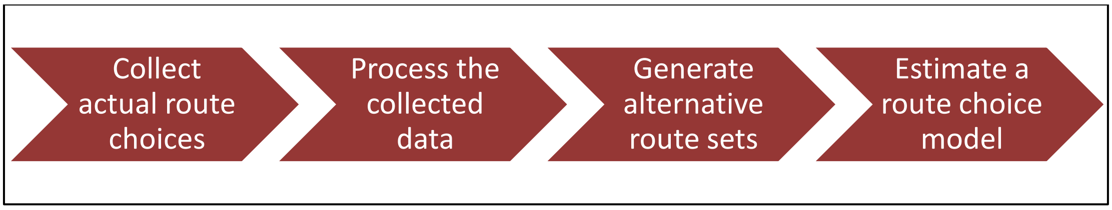


FIGURE 39: THE PROCESS FLOW IN PART III

7.1 IMPROVED METHODS TO DEDUCT TRIP LEGS AND MODE FROM TRAVEL SURVEYS USING WEARABLE GPS DEVICES: A CASE STUDY FROM THE GREATER COPENHAGEN AREA

Title:	Improved methods to deduct trip legs and mode from travel surveys using wearable GPS devices: A case study from the Greater Copenhagen area
Author(s):	Thomas K. Rasmussen, Jesper B. Ingvardson, Katrín Halldórsdóttir, and Otto A. Nielsen
Presented:	The *2nd Symposium of the European Association for Research in Transportation*, 4th – 6th September 2013, Stockholm, Sweden
Published:	*Computers, Environment and Urban Systems*, available online 4 May 2015 from: doi:10.1016/j.compenvurbsys.2015.04.001
Abbreviated:	Rasmussen et al. (2015)

Collecting data on actual route choices has greatly profited from resent enhancements in GPS device technology. However, the post-processing of raw GPS data is still problematic. GPS data collection generates very large data sets, containing a lot of scatter or other irrelevant records. The size of the raw GPS data emphasises the complexity of post-processing such data, as well as other issues related to, e.g., signal loss, lack of qualitative information on routes and activities, etc. Accordingly, processing such large data sets manually is highly unfeasible. The possibility to apply GPS data collection for travel surveys depends heavily on the availability of computer-based analysis tools, which are able to process the raw data set in order to extract relevant information for mode- or route choice analysis. Accordingly, this paper focuses on the fourth research objective, described in part I:

(iv) To develop a fully automatic post-processing procedure, making it possible to process raw individual-based GPS data with no additional information required from the respondent.

This paper also has the following aims:

(v.a) Combining already established methods to identify trips, trip legs, and detect the most probable transport mode, together with a combined fuzzy logic- and GIS-based algorithm;
(v.b) Apply the method in the highly complex large-scale multi-modal network;
(v.c) Validate the method on a raw individual-based GPS logs through the application of a control-questionnaire.

7.2 Efficiency of choice set generation methods for bicycle routes

Title:	Efficiency of choice set generation methods for bicycle routes
Author(s):	Katrín Halldórsdóttir, Nadine Rieser-Schüssler, Kay W. Axhausen, Otto A. Nielsen, and Carlo G. Prato
Presented:	The *1st European Symposium on Quantitative Methods in Transportation Systems (LATSIS),* 4th – 7th September 2012, Lausanne, Switzerland
Published:	*European Journal of Transport and Infrastructure Research*, vol. 14, no. 4, pp. 332-348
Abbreviated:	Halldórsdóttir et al. (2014)

The challenge of modelling the route choices of cyclists, based on the observation of actual route choices, is the generation of realistic alternative routes prior to the model estimation. Recent advances in path generation help confront the challenge of generating plausible alternatives for model estimation, although not without the uncertainty related to the dependency of model estimates on choice set composition. Most studies have focused on applying path generation methods in the car- or public transport context, where simplified networks are normally used. Only few studies have focused on the generation of bicycle route choice sets, which require a highly detailed network. Consequently, the following research objective was tackled in this paper:

(v) To analyse the efficiency of choice set generation methods to generate realistic bicycle routes;

Furthermore, this paper extends the body of knowledge on choice set generation for bicycle route choice by:

(vi.a) Applying three effective path generation methods to the bicycle context, e.g., doubly stochastic generation function, breadth first search on link elimination, and branch & bound algorithm;

(vi.b) Evaluating the efficiency of the three path generation methods to generate realistic heterogeneous alternative routes in a high-resolution network;

(vi.c) Proposing multi-attribute cost functions that account for attributes that are considered relevant in the bicycle route choice context.

7.3 Land-use and transport network effects on bicycle route choice in the Greater Copenhagen area

Title:	Land-use and transport network effects on bicycle route choice in the Greater Copenhagen area
Author(s):	Katrín Halldórsdóttir, Otto A. Nielsen, and Carlo G. Prato
Presented:	*The 2nd Symposium of the European Association for Research in Transportation (hEART), 4th - 6th September 2013, Stockholm, Sweden*
Status:	Working paper
Abbreviated:	Halldórsdóttir et al. (2015b)

This paper focuses on the interaction between infrastructure and cyclists' route choices. To contribute to the current literature on the route choices of cyclists, this paper analysed actual route choices of cyclists, by using a large sample of GPS-observed routes. The paper focuses on the route choice characteristics in the Greater Copenhagen area, an established bicycle city, by focusing in particular on the interaction between infrastructure, land use, and cyclists' route choice. Accordingly, this paper tackles the final research objective:

(vi) To develop a model to analyse cyclists' route choices and evaluate their trade-offs.

Additionally, the paper aims to:

(vi.a) Estimate a path-size logit, to account for similarities between the alternative routes;

(vi.b) Investigate how network attributes and conditions along the routes affect the route choice;

(vi.c) Analyse how personal attributes influence the route choice;

(vi.d) Investigate whether there are differences in route choice preferences between trip related attributes and also between weather attributes.

Chapter 8

DATA DESCRIPTION

DATA COLLECTION FOR THE BICYCLE ROUTE CHOICE MODEL

The TU-survey contains information on the choice of transport mode and socio-demographic variables, but does not collect information on actual route choices. Thus, information on the route choice of cyclists was collected with GPS loggers. The data collection was managed by the Data- and Modelcenter at DTU Transport, while the recruiting of respondents was conducted by Epinion.

Collecting data on travel behaviour, using GPS technology, has become an important research resource. When compared to traditional travel survey methods, collecting data with GPS technology ideally provides far more detailed information on route choice and travel patterns over a longer time period. GPS data provides more accurate information on travel times, such as the start- and end-time of a trip and the duration of a trip. This type of data also provides more accurate and reliable information on actual route choices, travel distance, and accurate geographic locations of activities. Thus, GPS data provides information that is independent from individuals' assessment of travel time, travel distance, and departure time, which can be biased based on their perception. GPS data collection also prevents trips underreporting, which is a common problem in traditional travel surveys. Additionally, carrying a GPS logger is less demanding for the respondents than answering time-consuming questionnaires.

Collecting GPS data over a long time period, often results in large sample sizes, which places high requirements on the post-processing of the data. Extensive post-processing is needed to obtain additional information, such as transport modes, trip purpose, etc. A very detailed digital network is also required to accurately map the routes, which can lead to high computation times during choice set generation, as well as issues with behavioural realism that might produce inconsistent estimates. The data collection is briefly described in subsection 8.1, while subsection 8.2 describes the bicycle network. Rasmussen et al. (2015) describe in detail the post-processing procedure used in order to obtain data that could be used for further analysis. After the GPS post-processing, the identified bicycle trips were mapped to a high-resolution bicycle network using the map-matching algorithm developed by Nielsen et al. (2004). Halldórsdóttir et al. (2014) focus on the generation of route choice sets for cyclists in a highly detailed network and analyse the efficiency of three choice set generation methods in generating realistic bicycle routes.

8.1 DATA COLLECTION

The level of bicycle use is relatively high in Denmark when compared to most other countries. However, there are considerable differences in bicycle use between Danish municipalities, as discussed in section 3.2.1. The PhD study uses GPS technology to register geographical points, recording the behaviour of a sample of cyclists from different municipalities.

Comparing different types of cyclists and evaluating their bicycle routes can give a deeper understanding of what influences individuals' choices for different types of facilities. Accordingly, a dataset on the physical conditions of the bicycle network from different municipalities, which vary in size and geographical location, is used in the study.

The data collection was carried out in the Greater Copenhagen area, where the number of cyclists is most dense. The TU-survey was used to collect possible participants that had permanent residency in the study area. The sampling criteria were that the respondents had previously completed the TU-questionnaire, approximately within the last 6 to 12 months before the data collection, and answered that they had used a bicycle in their reported travel. There are stricter privacy rules within the TU-survey, when interviewing people under 16 years, thus possible recruits were sampled from the TU-survey if they were 16 years or above. These individuals were then contacted and asked whether they would be interested in participating in the project.

Data was then collected by giving the respondents GPS loggers and asking them to carry the logger for seven days. Some participants collected for more than seven days, while other collected for less. The number of days ranged from 1 to 23 days. The final sample contained GPS tracks for an average period of 8.3 days, which resulted in a very large dataset. The median of the sample was 8.0 days, while the standard deviation was 3.13. The respondents were asked to carry the GPS units at all times, to minimise situations where they could forget to bring it. The post-processing procedure, described in Rasmussen et al. (2015), was then used to identify bicycle trips from the collected data.

Three data collection rounds were run in order to capture different seasonal effects in different time periods. Figure 40 shows the study area and the number of respondents in each municipality, per round and in the total sample. In the first round, the respondents picked up the GPS units themselves at the Epinions office in inner Copenhagen. In the last two rounds, the GPS units were sent out by mail, containing a description of the project and how to use the GPS unit to collect data, and also including a return envelope. The change in collection strategy was done in order to optimize time usage, qualification for extracting data, and logistic costs. The need for coordination and communication was also reduced. The first round was carried out from October through December in 2012. Data was mainly collected in the centres of Copenhagen and Frederiksberg, where the number of daily cyclists is the highest. Data was collected from 112 respondents. The second round was carried out from June through July in 2013, where it was prioritised to sample individuals from the areas surrounding the city centre. Total of 71 respondents participated in the second round. The last round was carried out from August through October in 2013, sampling individuals from the Greater Copenhagen area. The final round collected data from 135 respondents. Accordingly, the total number of observations was 318 individuals. The final sample shows that the number of respondents is denser in the city centre, which is in line with the higher number of cyclists in that area. In the first round the respondents were given an incentive to participate (a 200 DKK gift card), while no incentive was given in the last two rounds. Still, the acceptance rate was much higher in the last two rounds, 76% in the second round and 66% in the final round, than in the first round, where the acceptance rate was 53%. On average, the acceptance rate was 65%, which is quite high for a survey of this type.

Travel diaries were also collected from the participants. The participants were contacted on the 3rd or 4th day during the period where they had the GPS loggers. The TU-survey platform was used to collect the travel diaries, thus the participants provided information on their travel on the day before the interview, socio-economic variables, and trip related information, such as trip purpose, location points of activities, etc. Combining in-depth interviews with the GPS data collection enables a more accurate analysis of individual travel activities.

After the post-processing procedure (Rasmussen et al., 2015) had identified all bicycle trips from the collected data and the map-matching algorithm was run, there were 3,443 stages from 291 respondents remaining for further analysis. Halldórsdóttir et al. (2015b) describes the extraction of bicycle trips in more detail, as well as presenting the characteristics of the chosen routes. Subsection 8.1.1 presents some descriptive statistics on the participants of the survey. Subsection 8.1.2 discusses the characteristic of the GPS dataset compared to the TU-survey, while subsection 8.1.3 briefly describes additional information collected for the survey.

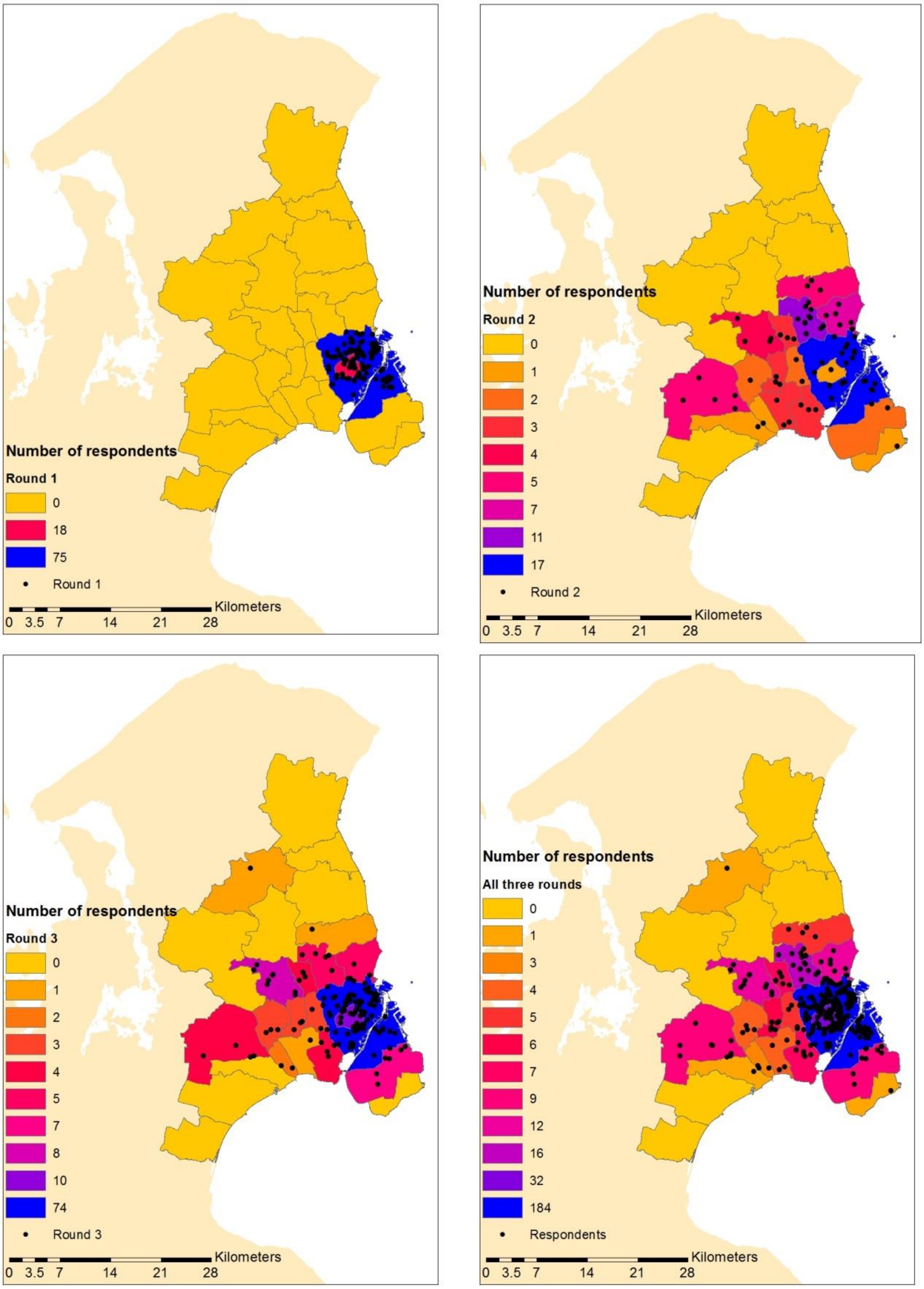


FIGURE 40: NUMBER OF RESPONDENTS IN EACH MUNICIPALITY IN THE GREATER COPENHAGEN AREA, IN EACH ROUND AND OVER THE TOTAL SAMPLE

8.1.1 Data description

The GPS dataset is composed of 45% males (Figure 41). Figure 42 shows the age distribution in the dataset. There are only two individuals in the dataset that are 17 years of age, while no one was at the age of 16. The dataset is composed of individuals between the ages of 17 and 79, where the average age is 40 years. Figure 43 shows the share of respondents that have a driving license. The figure shows that most respondents have a driving license, or approximately 80%.

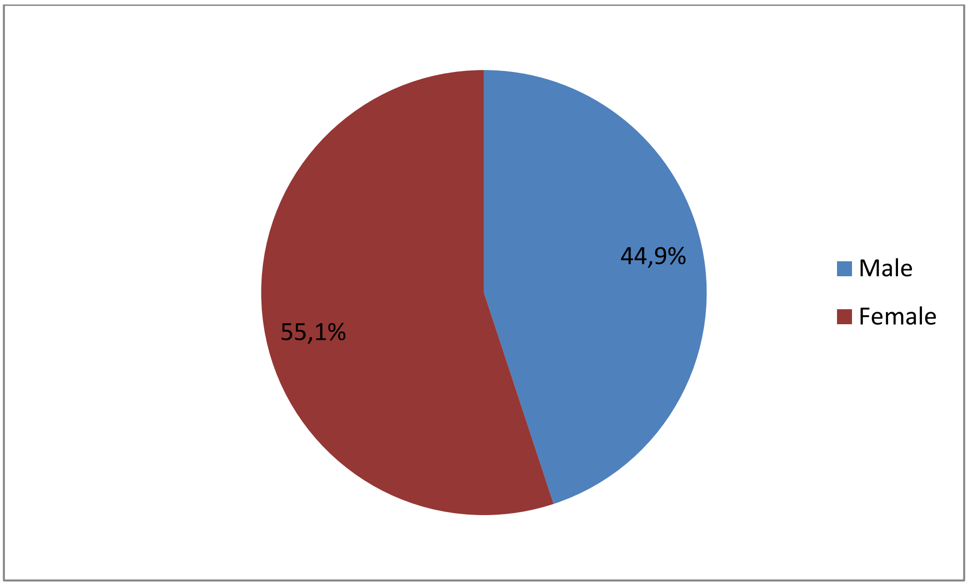


Figure 41: Gender share in the GPS dataset

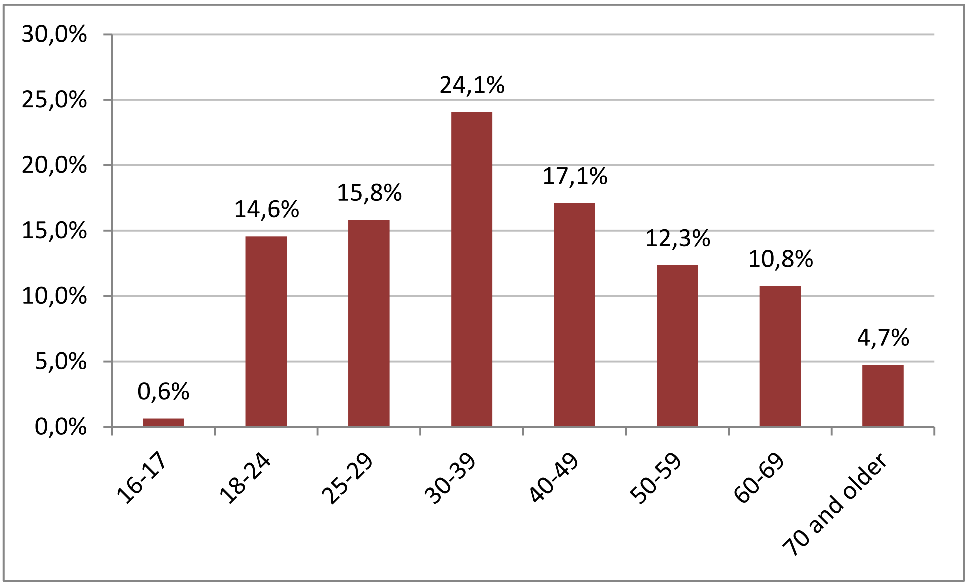


FIGURE 42: AGE GROUP DISTRIBUTION IN THE GPS DATASET

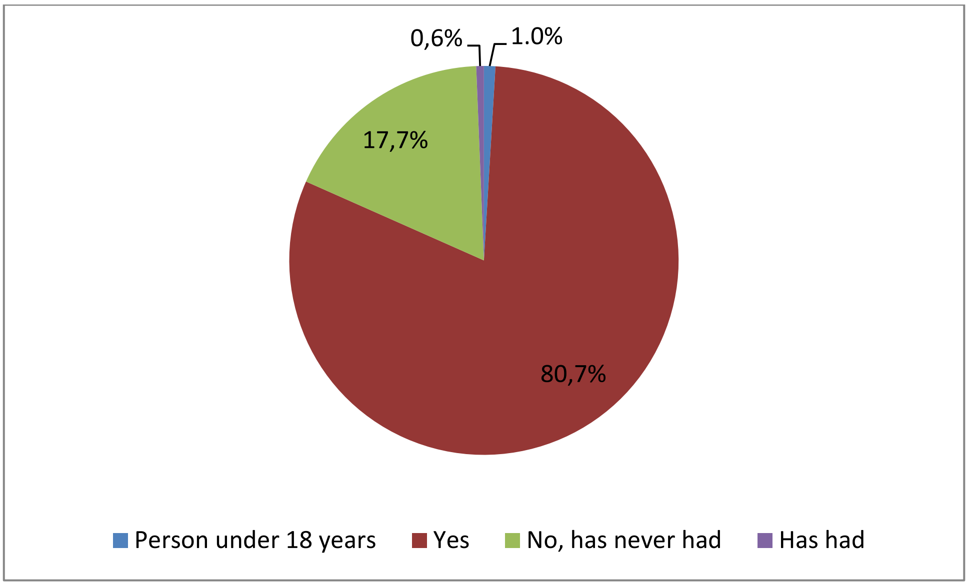


FIGURE 43: SHARE OF RESPONDENTS WITH A DRIVING LICENSE IN THE GPS DATASET

Skougaard and Christiansen (2015) showed that individuals with high level of education cycle the most, along with individuals with upper secondary education. Figure 44 shows the share of the respondents by education level, where most respondents in the GPS dataset have a high level of education. Skougaard and Christiansen (2015) also showed that pupils, students, and employees cycle the most, as discussed in subsection 3.2.1. Figure 45 shows the share of the respondents by main occupation. The GPS dataset has the highest share of employees, or 54%, while students take up approximately 20% of the sample.

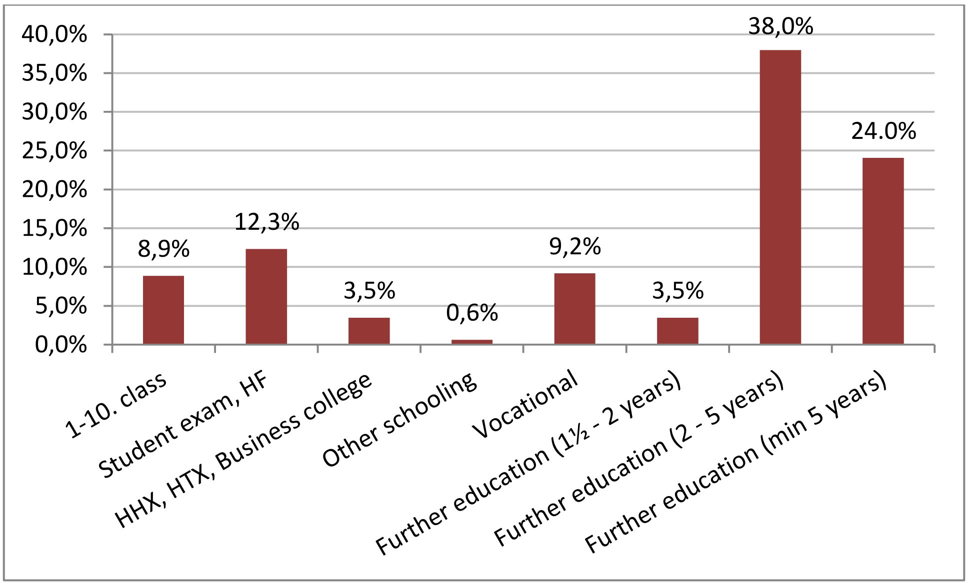


FIGURE 44: SHARE OF EDUCATION LEVEL IN THE GPS DATASET (HF IS ABBREVIATION FOR HIGHER PREPARATORY CERTIFICATE, HHX IS HIGHER COMMERCIAL CERTIFICATE, AND HTX IS HIGHER TECHNICAL CERTIFICATE)

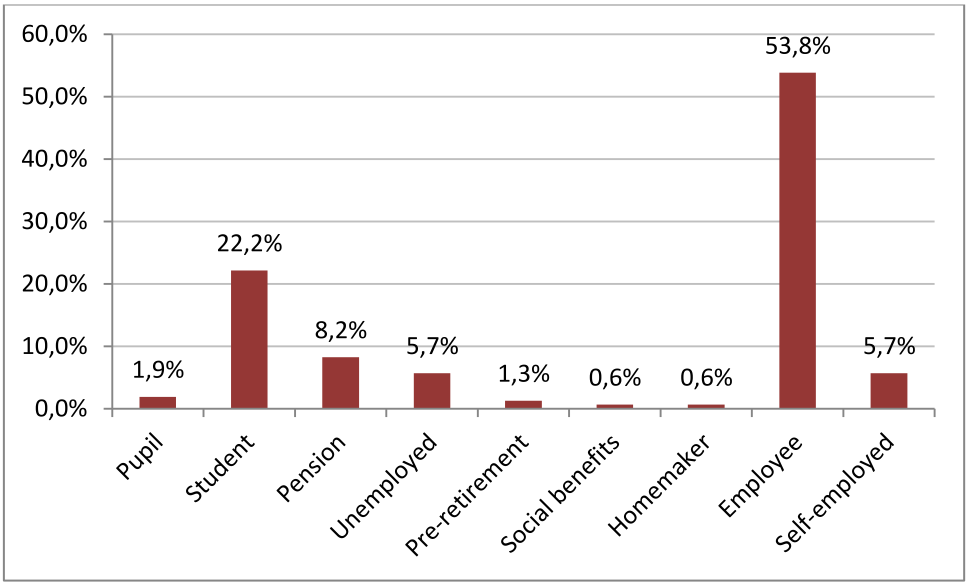


FIGURE 45: SHARE OF MAIN OCCUPATION IN GPS DATASET

8.1.2 CHARACTERISTICS OF THE GPS DATASET COMPARED TO THE TU-SURVEY

The objective of the data collection was to get as many recruits as possible. In the first round, individuals were sampled randomly from the TU-survey, based on their residence location. In order to get a more representative sample of cyclists in the Greater Copenhagen area, individuals were contacted according to a priority list in the following two rounds, based on their residence location. Thus, individuals from specific areas were contacted first, in order to get the best possible sample for the bicycle route choice model, without diminishing the number of observations. If individuals did not answer their phone, Epinion moved on to the next individual on the list, as there was a limited timeframe in which the data could be collected.

The TU-survey is a representative sample of the Danish population. By comparing the collected data to a weighted sample from the TU-survey, of cyclists in the Greater Copenhagen area, it is possible to evaluate if the GPS dataset is a representative sample. Table 21 shows a comparison of the socio-demographics of the GPS data sample together with the TU-survey. The table shows that there is some difference in gender, where the GPS dataset has slightly fewer females than the TU-survey. The GPS dataset does not have many individuals that are 16-17 years of age. The youngest participants in the GPS data collection are 17 years old. Instead there are slightly more participants in the other age groups, especially between the ages 18 to 29. The age groups from 30 to 59 have slightly lower shares. The remaining age groups have higher shares, where the oldest participant in the GPS data collection is 79 year old, while the oldest cyclists in the TU-survey are 85 years old. Compared to the TU-survey, there is a great deal larger share of students and slightly more pensioners, un-employed, and self-employed participants in the GPS data collection. The share of employee is considerable lower in the GPS dataset. There is some deviation between the share of respondents in each education level, where the main difference is in the number of participants in with student exams and with higher education.

TABLE 21: COMPARISON THE SOCIO-DEMOGRAPHICS OF THE GPS DATA SAMPLE TOGETHER WITH THE TU-SURVEY (HF IS ABBREVIATION FOR HIGHER PREPARATORY CERTIFICATE, HHX IS HIGHER COMMERCIAL CERTIFICATE, AND HTX IS HIGHER TECHNICAL CERTIFICATE)

Variables	**GPS dataset**	**TU-survey**
Gender		
Male	44.94%	38.36%
Female	55.06%	61.64%
Age group		
16-17	0.63%	1.38%
18-24	14.56%	12.82%
25-29	15.82%	9.98%
30-39	24.05%	31.86%
40-49	17.09%	22.19%
50-59	12.34%	13.64%
60-69	10.76%	6.24%
70 and older	4.75%	1.90%
Respondents main occupation		
Pupil	1.90%	0.70%
Student	22.15%	5.85%
Apprentice	0.00%	0.01%
Pension	8.23%	2.98%
Unemployed	5.70%	0.24%
Pre-retirement	1.27%	0.07%
Social benefits	0.63%	0.01%
Homemaker	0.63%	0.01%
Employee	53.8%	89.71%
Self-employed	5.7%	0.41%
Respondents education level		
1-10. class	8.86%	3.31%
Student exam, HF	12.34%	7.47%
HHX, HTX, Business college	3.48%	0.66%
Other schooling	0.63%	0.42%
Vocational	9.18%	10.75%
Further education (1½ - 2 years)	3.48%	1.78%
Further education (2 - 5 years)	37.97%	47.39%
Further education (min 5 years)	24.05%	28.22%

Figure 46 shows the percentage distribution of cycling respondents per municipality in the Greater Copenhagen area. The figure, on the left, shows that the GPS dataset has the largest percentage of respondents in the city centre and in the municipalities around the centre. This is in line with the percentage of cyclists, in the study area, from the TU-survey, the figure on the right. When recruiting participants, there were very few candidates in the outer municipalities. This is because there are not as many cyclists in those municipalities. When contacting possible recruits, these municipalities were prioritised. However, either the possible participants were unobtainable or they were not interested in participating in the project, which resulted in no observations in those municipalities.

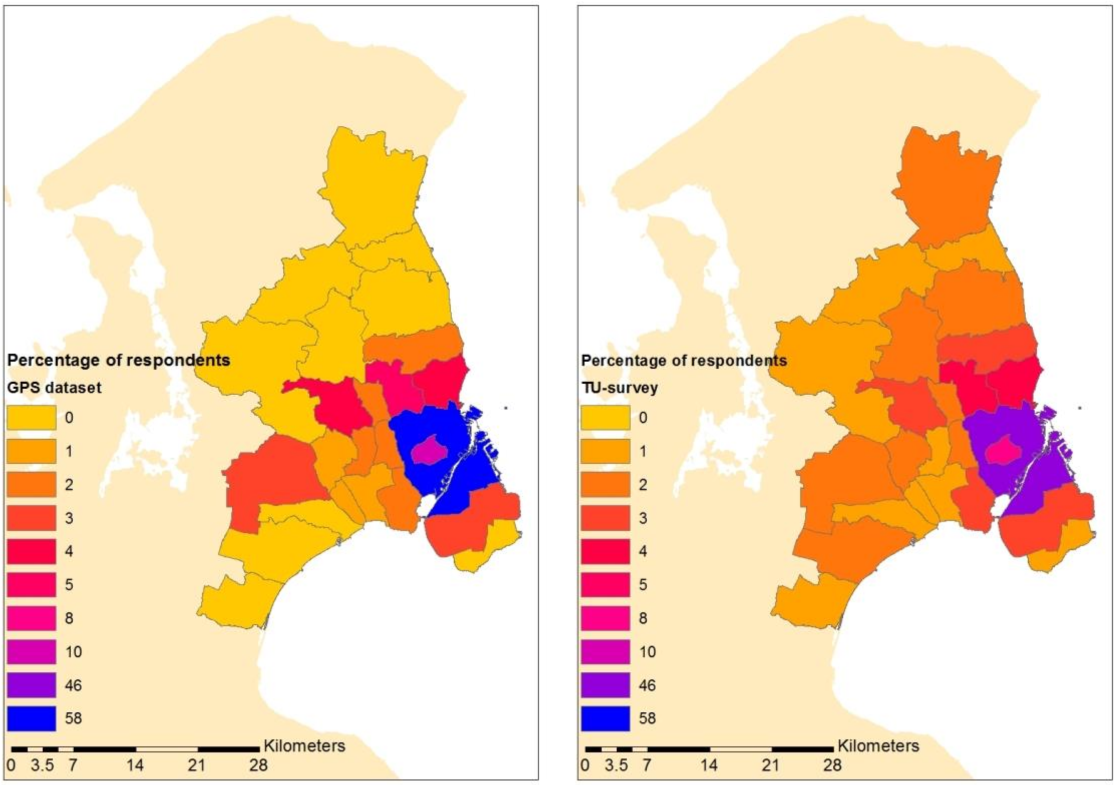


FIGURE 46: PERCENTAGE DISTRIBUTION OF CYCLING RESPONDENTS PER MUNICIPALITY IN THE GREATER COPENHAGEN AREA. LEFT: GPS DATASET. RIGHT: TU-SURVEY

8.1.3 Additional information

Along with collecting information on bicycle routes and travel diaries, additional information was also gathered. Geographical information on the location of respondent's home and work or place of study, collected in the TU-survey, was used to classify commuting trips by using GIS analysis, as described in subsection 8.1.3.1. This proved problematic, so the hourly split of cycling, during weekdays, was also extracted from the TU-survey to identify commuting trips. Subsection 8.1.3.2 describes how the bicycle trips were divided depending on the time of day. Information on sunrise and sundown in the Greater Copenhagen area was also collected in order to analyse whether there were differences in route preference between day and night times. Subsection 8.1.3.3 describes how the cyclists were categorised based on their average cycling speed. Finally, route preference was also analysed in relation to different environmental attributes, i.e., rainfall, temperature, and wind, as described in subsection 8.1.3.4.

8.1.3.1 Commuting trips

The geographic locations were used to classify commuting trips by connecting the location points, identified in the GPS post-processing, with the geographical location of the home address or the place of work/study within a buffer of 200 meters. The trips were categorised into commute trips if the trip started at home and ended at work or vice versa. After extracting all commuting trips with this method, there were only 281 commuting trips out of 3,443 trips, or only 8.2%. In the travel survey, when bicycle was the primary mode selected, the share of trips to work was 33% and 6% to a place of study. Thus, this method proved ineffective in identifying commuting trips. This could be because travellers often combine other trip purposes with their commuting trips, such as dropping off/picking up children or buying groceries.

8.1.3.2 Time of day

The hourly split of cycling, during weekdays, was analysed by looking at the hourly split of bicycle trips for all of Denmark, including all trip purposes as reported in the TU-survey (see Figure 47). The results show that the morning peak hours are very clearly between 7 and 9. The afternoon peak hours are more widespread, most likely because travellers do other errands on the way home as discussed in 8.1.3.1. The maximum afternoon peak was between 15 and 17. The hourly split of bicycle trips, collected by the GPS trackers, follows the similar split (Figure 48).

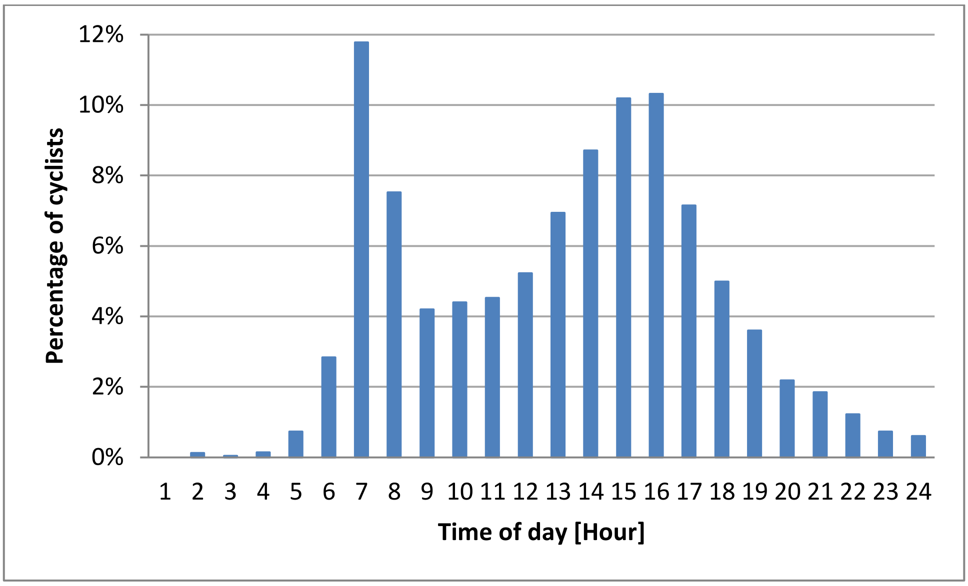


FIGURE 47: HOURLY SPLIT OF CYCLING, DURING WEEKDAYS, ACCORDING TO THE TU-SURVEY

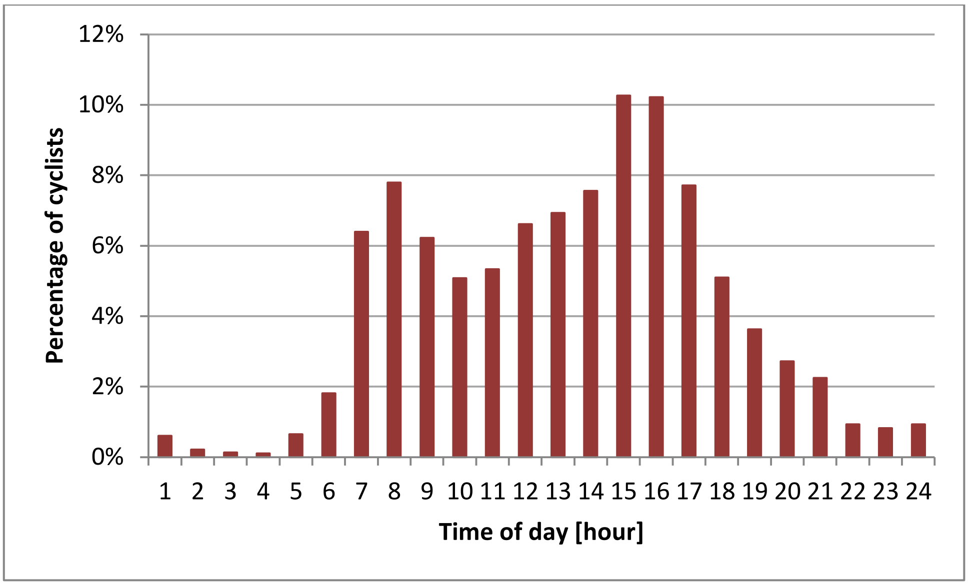


FIGURE 48: HOURLY SPLIT OF CYCLING, DURING WEEKDAYS, IN THE COLLECTED GPS DATA

For estimation purposes, the two peak hours were most interesting for identifying possible *commuting trips*. The stages conducted within these two time periods, during weekdays, were classified as peak hours trips, a total of 1,094 tips, while all others were non-peak hours, a total of 2,349 trips. Finally, stages conducted during weekends were identified in order to investigate differences in route preference between weekends and weekdays. The total number of weekend trips was 669, while there were 2,774 weekday trips.

In order to analyse whether darkness influences route choice, information on sunrise and sundown in the Greater Copenhagen area was collected and compared to the time of day each stage was conducted. There were 491 trips after sundown, while there were 2,952 trips during daylight.

8.1.3.3 Type of cyclist

In the Greater Copenhagen area there are various types of cyclists, e.g., there is the fast sporty cyclist and the slower more leisurely cyclist. Whether there was any correlation between different age- and gender groups and the average speed (see Figure 49) was analysed. However, when averaging over age and gender there was little difference in average speed (12 to 15 km/h in the most extreme cases).

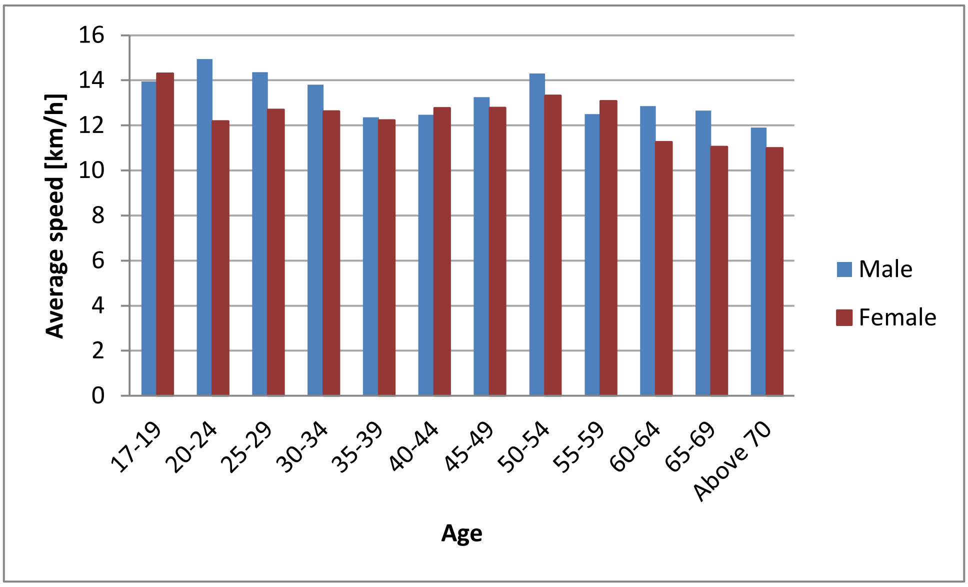


Figure 49: Average speed of cyclists divided into age-groups and gender

By counting the number of cyclists, in each speed interval based on the average speed of each participant, there was quite some variance (see Figure 50). Accordingly, participants were divided into three groups, slow (<10 km/h), medium (10-14 km/h), and fast (>14 km/h). There were 91 trips taken by slow cyclists, 2,633 by medium cyclists, and 719 by fast cyclists.

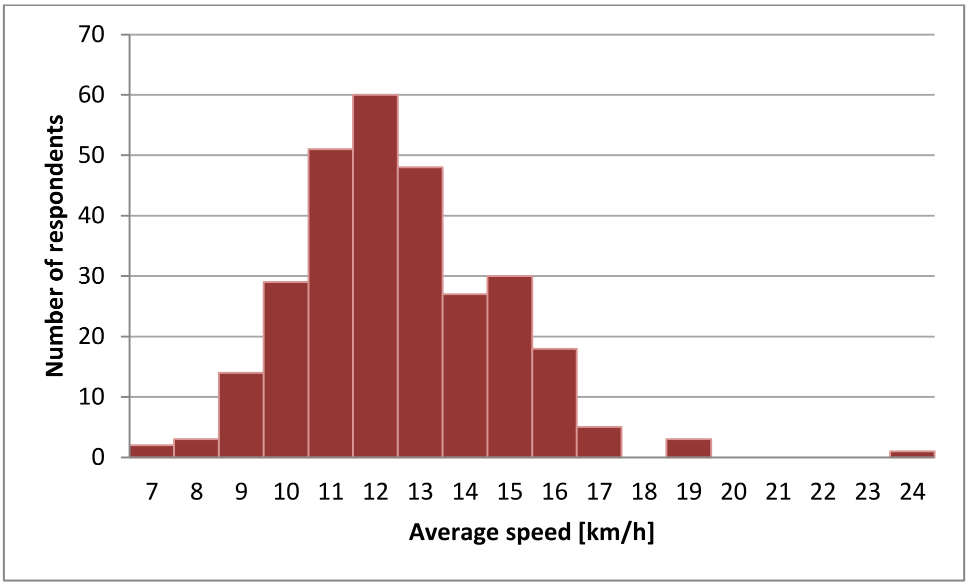


FIGURE 50: AVERAGE SPEED, THE NUMBER OF CYCLISTS IN EACH INTERVAL

8.1.3.4 WEATHER ATTRIBUTES

Weather information, from the Danish Meteorological Institute (DMI), was joined with the bicycle trips in order to analyse whether there were differences in route preference between different weather attributes (i.e., rainfall, temperature, sunshine, and wind). The data was added by joining the closest weather station to the starting point of each trip with data in the time period the trip took place. Figure 51 to Figure 54 show the number of trips with the corresponding weather information.

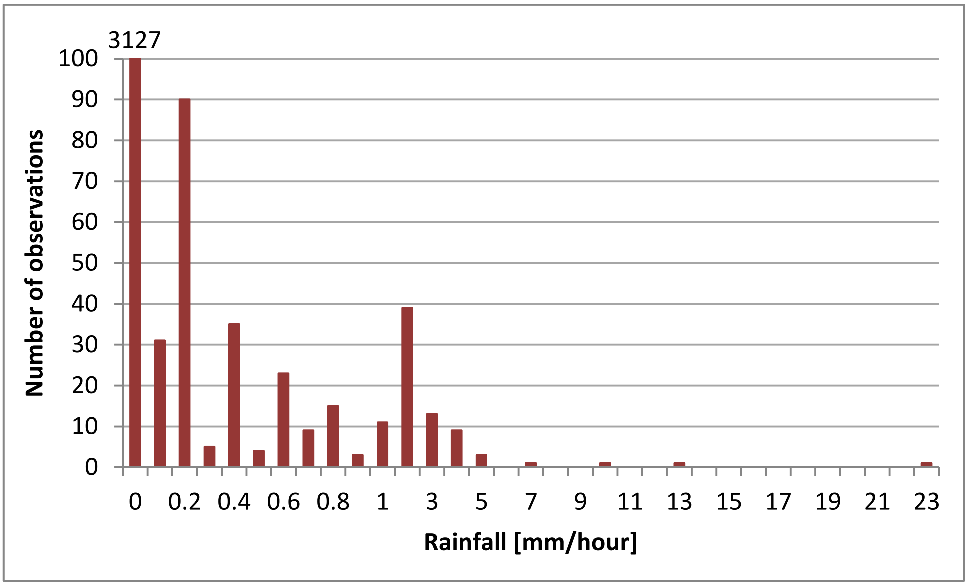


FIGURE 51: NUMBER OF BICYCLE TRIPS IN EACH RAINFALL (MM/HOUR) INTERVAL

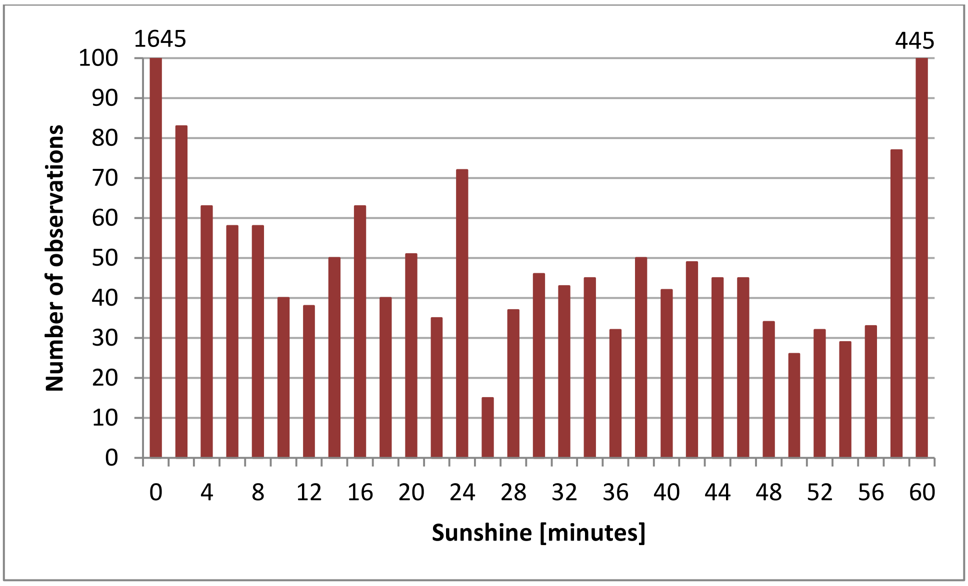


FIGURE 52: NUMBER OF BICYCLE TRIPS IN EACH SUNSHINE (MINUTES) INTERVAL

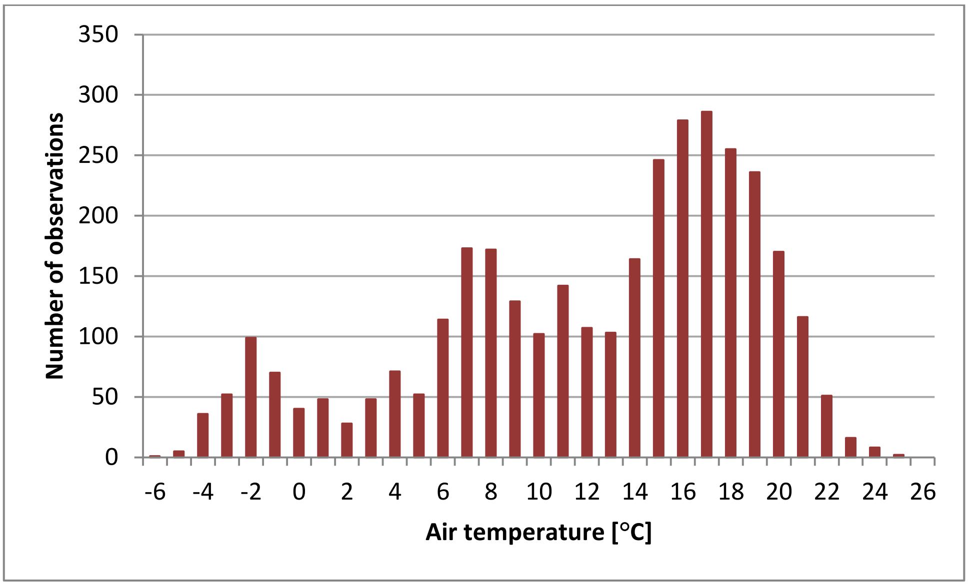


Figure 53: Number of trips in each air temperature (°C) interval

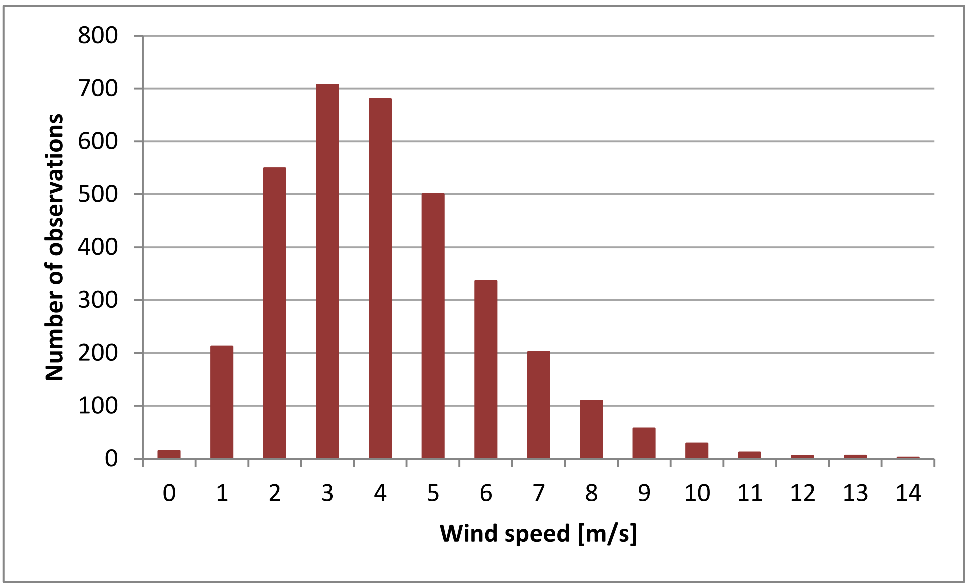


Figure 54: Number of bicycle trips in each wind speed (m/s) interval

8.2 BICYCLE NETWORK DATABASE

The bicycle network database is built on the topographic network FOT-kort10 (FOT-Kort10, 2010) and TOP10DK (Kort & Matrikelstyrelsen, 2001). The two sources were compiled together in order to obtain a detailed bicycle network database, as illustrated in Figure 55. FOT-kort10 is a complete geographic network of roads and paths used by cyclists, containing attributes at a national standard. Most importantly the road network is seamless across municipality borders and was thus chosen as a base network. FOT-kort10 is expected to cover all of Denmark in 2012, but until then the Copenhagen, Tønder, and Åbenrå municipalities are missing. These areas are instead covered by TOP10DK. It is very important to have a highly detailed network, since cyclists are using paths that are not present in standard commercial digital maps made for GPS-based car navigation systems.

The network was constructed together with the Data- and Modelcenter at DTU Transport, which corrected the topological errors in a 2011 version of FOT-kort10. A minor number of links was also added to the network manually, where map-matching results revealed gaps. This resulted in a very detailed network, comprising of 363,252 directional links and 270,018 nodes for the Greater Copenhagen area. A full network is important for the overall completion of the final bicycle network, in order to include attributes describing physical surroundings, e.g., concerning motorised traffic or intersections. Then for the choice set generation, links can be excluded based on an attribute describing bicycle accessibility and roads, where cycling is illegal in Denmark, can be excluded from the calculation, such as motorways and expressways. Dead-ends, which do not contain any of the observed routes, are excluded from the network to minimise calculation time, as well as loops since those links cannot be used in the choice set generation.

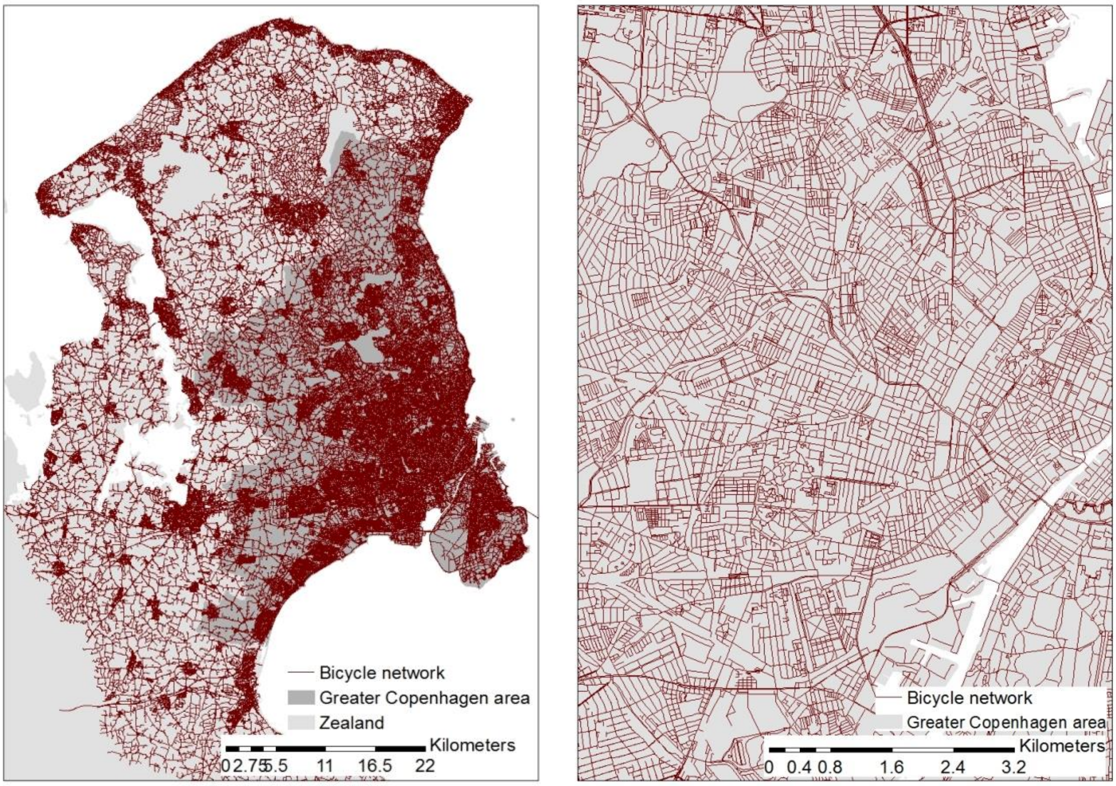


FIGURE 55: BICYCLE NETWORK. LEFT: ZOOM OF THE COPENHAGEN REGION, RIGHT: ILLUSTRATION OF THE NETWORK DETAILS

Relevant information from available data sources was then added to the network database, in order to obtain a more detailed network with characteristics considered important for cyclists (i.e., Open Street Map (OSM), the LTM road network (Rich et al., 2010), NAVTEQ (NAVTEQ, 2010), accidents crash database maintained by the Danish Road Directorate (Vejman), and intersection data (from the Danish Road Directorate)). For further information, see: Halldórsdóttir et al. (2013) and Pedersen and Senstius (2014), while appendix A gives an overview and a description of the network attributes.

A rich set of network attributes was added to the base network. One attribute that is considered important, in analysing the route choice of cyclists, is the effect of different bicycle facility types. The collected information was joined to the base network, on the condition that the information gathered was already digitised. This required extensive work, both in verifying the accuracy of the data and in making it compatible with the base network. Road types are categorised as follows: (i) road, (ii) road with bicycle lane, (iii) road with bicycle path, (iv) bicycle path, (v) footpath, (vi) no bicycle access, and (vii) no access. Also, information on bicycle bridges, such as the one on “Den Grønne Sti”, a green path in

Copenhagen, was digitised in the bicycle network. Accordingly, it can be investigated whether the construction of more tall bridges will be beneficial to everyday cyclists and thereby promote more cycling.

Information on the conditions of the bicycle network was collected, which can affect the choice of bicycle as a transport mode. An important factor on bicycle path condition is pavement type, i.e., whether the path is paved, if it is a gravel path, or whether it is a dirt path.

Driving directions were extracted from various sources, and then the network was manually checked for errors in Copenhagen and the surrounding municipalities, but in all remaining areas of the network all links are consistently open in both directions. This quality control required a lot of work and is not without some errors. For example, not all driving restrictions for cars apply to bicycles. In addition, driving directions do not necessarily reflect actual bicycle behaviour as cyclists often choose to take a short cut by cycling on a pedestrian path, and thus cycle against driving direction.

Even though Denmark has a relatively flat landscape, there are some areas that are considerably hilly. The routes can be analysed to discover how often cyclists choose to detour to avoid particularly hilly terrain and steep hills. The FOT-kort10 network has altitudes (z-coordinates) on each point, including several intermediate points. Appendix A specifies in more detail how the gradients were expressed on each link.

The network conditions along the routes were also included, i.e., motorised traffic type, time-dependent traffic volumes, speed limit, number of motorised traffic lanes, motorised traffic bridges and tunnels, signalised intersections, and roundabouts. Accordingly, it can be investigated whether bicycle route choice is influenced by the conditions along the routes. The information on the main roads was provided through the NTM road network. Information on the remaining roads was collected from the NAVTEQ road network.

Investigating whether land-use has an effect on cyclists' daily travel patterns is also interesting. Thus, land-use information, from FOT-Kort10, was added to the bicycle network, e.g., residential area, city centre, industry, park, forest, and lake. Including information on the right and left side of the path, depending on the direction of the link.

The Danish Road Directorate has a database that collects reported accidents for all of Denmark. Traffic accidents are only officially reported in two ways; by the police, if they witness them or are called to the scene; or if individuals involved in the accident turn to the emergency room. In addition, traffic accidents, where a bicycle was involved, are also only reported if the accident involved a motorised vehicle. This leads to a great deal of underreporting of bicycle accidents. Additionally, not all accidents reported by the hospitals are reported by the police, or only between 6-7% of all slightly injured cyclists are reported by the police and between 14-15% of all severely injured cyclists (Janstrup et al., 2014). Information on accident patterns was added to the bicycle network as an indicator of approximate safety effects, where accidents from the last five years were added, a total of 87,455. There are some challenges in relation to identifying routes that are more unsafe than others. A high level of reported accidents can be an effect of more traffic rather than the route being unsafe. Motorised traffic volumes can also be taken into account as an indicator for safety effects.

After the bicycle network was constructed, the quality of the information was tested and validated by comparing different sources, where these were available. This was done through graphical validation and analysts' knowledge of the study area.

8.3 References

Skougaard, B.Z. and Christiansen, H. (2015). *Transportvaneundersøgelsen - Faktaark om cykeltrafik i Danmark*. [pdf] DTU Transport, Data- and Modelcenter, Kgs. Lyngby. Available at: http://www.modelcenter.transport.www7.sitecore.dtu.dk/Transportvaneundersoegelsen/TU-udgivelser/Faktaark-om-cykeltrafik-i-Danmark-2014 [Accessed 05 June 2015].

FOT-Kort10 (2010). *Information om det nye Kort10 og FOT-Kort10.* [pdf] Available at: support.kortforsyningen.dk/attachments/download/292/information-kort10.pdf [Accessed 18 June 2015].

Halldórsdóttir, K., Rieser-Schüssler, N., Axhausen, K.W., Nielsen, O.A. and Prato, C.G. (2014). Efficiency of choice set generation methods for bicycle routes. *European Journal of Transport and Infrastructure Research*, vol. 14, no. 4, pp. 332-348.

Halldórsdóttir, K., Nielsen, O.A. and Prato, C.G. (2015). Land-use and network effects on bicycle route choice in the Greater Copenhagen area. Presented at the *2nd Symposium of the European Association for Research in Transportation (hEART), 4th - 6th September 2013, Stockholm, Sweden.* Working paper.

Janstrup, K.H., Hels, T., Kaplan, S., Sommer, H.M. and Lauritsen, J. (2014). Understanding traffic crash under-reporting: linking medical records to individual and crash characteristics. Proceedings of *the Transport Research Arena* 2014, April 14-17, Paris, France.

Kort & Matrikelstyrelsen (2001). *TOP10DK - Geometrisk registrering*. Specifikation udgave 3.2.0. Kort & Matrikelstyrelsen, Copenhagen, Denmark. [pdf] Available at: ftp://ftp.kms.dk/download/pdf/top10dkspec320.pdf [Accessed 08 June 2015].

NAVTEQ, 2010. NAVSTREETS Street Data Reference Manual v2.8.

Nielsen, O.A., Würtz, C. and Jørgensen, R.M. (2004). Improved map-matching algorithms for GPS-data - Methodology and test on data from the AKTA roadpricing experiment in Copenhagen. Presented at the *19th European Conference for ESRI Users*, Copenhagen, November.

OSM. © OpenStreetMap contributors. [website]
Available at: http://www.openstreetmap.org/ [Accessed 11 August 2015].

Pedersen, S.H. and Senstius, S.J. (2014). *Landsdækkende cykelnetværk*. Unpublished data dictionary in Danish. Technical University of Denmark, Department of Transport, Kgs. Lyngby, Denmark.

Rasmussen, T.K., Ingvardson, J.B., Halldórsdóttir, K. and 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*. Available online 4 May 2015 from: doi:10.1016/j.compenvurbsys.2015.04.001

Rich, J.H., Nielsen, O.A. and Hansen, C.O. (2010). Overall Design of the Danish National Transport Model. In proceedings for the *Annual Transport Conference in Aalborg*, Aalborg University.

Chapter 9

IMPROVED METHODS TO DEDUCT TRIP LEGS AND MODE FROM TRAVEL SURVEYS USING WEARABLE GPS DEVICES: A CASE STUDY FROM THE GREATER COPENHAGEN AREA

Title:	Improved methods to deduct trip legs and mode from travel surveys using wearable GPS devices: A case study from the Greater Copenhagen area
Author(s):	Thomas K. Rasmussen, Jesper B. Ingvardson, Katrín Halldórsdóttir, and Otto A. Nielsen
Presented:	The *2nd Symposium of the European Association for Research in Transportation*, 4th – 6th September 2013, Stockholm, Sweden
Published:	*Computers, Environment and Urban Systems*, May 2015
Abbreviated:	Rasmussen et al. (2015)

A preliminary version of this paper was presented at the 2nd Symposium of the European Association for Research in Transportation, 4th – 6th September 2013, Stockholm, Sweden. The paper was then updated, rewritten, and published in the Computers, Environment and Urban Systems in May 2015.

ABSTRACT

GPS data collection has become an important means of investigating travel behaviour. This is because such data ideally provide far more detailed information on route choice and travel patterns over a longer time period than possible from traditional travel survey methods. Wearing a GPS unit is furthermore less requiring for the respondents than filling out (large) questionnaires. It places however high requirements to the post-processing of the data. This study developed and tested a combined fuzzy logic and GIS-based algorithm to process raw GPS data. The algorithm is applied to GPS data collected in the highly complex large-scale multi-modal transport network of the Greater Copenhagen area. It detects trips, trip legs, and distinguishes between five modes of transport. The algorithm was validated by comparing with a control questionnaire collected among the same persons and a sensitivity analysis was performed. This showed that the algorithm (i) identified corresponding trip legs for 82% of the reported trip legs, (ii) avoided classifying non-trips such as scatter around activities as trip legs, (iii) identified the correct mode of transport for more than 90% of trip legs, and (iv) were robust towards the specification of the model parameters and thresholds. The method thus makes it possible to use GPS for travel surveys in large-scale multi-modal networks.

Highlights:

- We develop and test a method to process raw individual-based GPS data.
- The method relies on combined fuzzy logic, GIS analyses and feedback algorithms.
- High fit rates in the detection of trips, trip legs and mode for Copenhagen case.
- 92% success rate across five modes with particular good rates for bus and rail.
- The results were not highly sensitive to changes in the specification of the input.

Keywords: *GPS data processing; Revealed preference data; Multi-modal travel survey; Handheld GPS; GIS.*

9.1 INTRODUCTION

Global Positioning Systems (GPS) have been applied in various investigations of transport-related issues over the last 20 years. These applications include, among others, (i) evaluation of system performance, such as measuring historical and real-time congestion and flow levels (Herrera et al., 2010; Li, Guensler, Ogle, & Wang, 2004; Quiroga, 2000; Quiroga & Bullock, 1998), (ii) analysis of travel behaviour, such as response to road pricing schemes (Liu, Andris, & Ratti, 2010; Nielsen, 2004) or deviation from planned route (e.g., Papinski, Scott, & Doherty, 2009), (iii) estimation of route choice parameters in route choice models (Bierlaire, Chen, & Newman, 2013; Prato, Rasmussen, & Nielsen, 2014; Rich & Nielsen, 2007), and (iv) analysis of patterns of physical activity (Mavoa, Oliver, Witten, & Badland, 2011; Rainham et al., 2012).

Much effort has been made in recent years to investigate the use of GPS devices as the data source for travel surveys (Bolbol, Cheng, Tsapakis, & Haworth, 2012; Gong, Chen, Bialostozky, & Lawson, 2011; Stopher, Jiang, & Fitzgerald, 2005; Wolf, 2000, etc.). When compared to traditional travel diaries, collecting data via GPS devices ideally provides the investigator with far more detailed information on travel times, used routes, and locations of activities. Another advantage of using GPS data is that it is not dependent on individuals' (possibly mis-) perception of travel time, travel distance, and departure time (revealed preferences rather than stated preferences). In traditional travel diaries there is often a common problem of underreporting of trips (e.g., Forrest & Pearson, 2005; Stopher, Fitzgerald, & Xu, 2007). This problem is likely to be reduced when using GPS as all movements of participants are logged (Stopher, Clifford, Zhang, & Fitzgerald, 2008). Additionally, far less effort is required by the respondents as answering time-consuming questionnaires can be avoided. This enables larger sample sizes and data collection over a longer time period per respondent.

Today GPS units are sufficiently accurate, lightweight and have long enough battery-life to make multi-day individual-based data collection possible for all conducted trips (e.g., Bolbol et al., 2012; Gong et al., 2011; Stopher & Shen, 2011). Such a data collection facilitates complete analyses and better understanding of individuals' travel patterns. This includes choice of mode of transport, combination of modes, route choices in multi-modal transport networks, and day-to-day variations.

Collection of GPS data generates very large data sets, containing millions of GPS logs and a lot of non-relevant data in the form of e.g. scatter or coordinates collected when the person is not travelling. The size of the data sets combined with issues related to e.g. signal loss and lack of qualitative information on routes and activities underline the truly complex nature of the post-processing of such data. Manual processing of the large data sets is highly unfeasible. The possibility to utilise GPS data for travel surveys thus relies heavily on the availability of computer-based analysis tools which process the raw data set and convert the processed data into a format which is usable in the subsequent analyses of e.g. mode and route choice. This paper introduces such a method to process raw GPS data and to identify trips, trip legs, and mode used. The first part of the method is based on the POSDAP (2012) algorithm developed by Schüssler and Axhausen (2009). This is extended by a component that utilises disaggregate digital information on the infrastructure in combination with a Geographical Information System (GIS) to identify the mode used and to ensure that only actually performed trips are kept in the data set. The extension additionally includes algorithms to detect and correct illogical mode chains and transfers. The new method as well as the original POSDAP (2012) algorithm was applied to a multi-day individual-based GPS data set collected among families living in the Greater Copenhagen area. Corresponding traditional interview-based travel survey data were collected for each of the respondents for one of the days in the survey period. This made it possible to validate the results of the trip and mode identification algorithms.

Section 9.2 of the paper presents a review of the existing literature focused on using GPS as a travel survey method. The new method is introduced in Section 9.3, while Section 9.4 presents the case study and parameter specification used. Section 9.5 reports the results of the application. In Section 9.6 the applicability across case studies is discussed, and a sensitivity analysis towards changes in primary input parameters and thresholds is performed. Section 9.7 relates the results to findings in similar studies and concludes the work. A preliminary version of the work was presented in Rasmussen, Ingvardson, Halldórsdóttir, and Nielsen (2013).

9.2 Literature review

9.2.1 GPS in travel surveys

Technology limited the first travel surveys using GPS to be only vehicle-based, as the devices were large and the power consumption was high (Nielsen & Sørensen, 2008; Wagner, Neumeister, & Murakami, 1996; Yalamanchili, Pendyala, Prabaharan, & Chakravarthy, 1999). These early studies sought mainly to supplement telephone-based travel surveys by collecting additional data to e.g. identify detailed route choices, verify exact time of day as well as detect unreported trips (Wolf, 2000). Additional trip information such as trip purpose was specified by the respondents when starting a trip (Du & Aultman-Hall, 2007; Yalamanchili et al., 1999). This was often done on a connected personal digital assistant (PDA).

As GPS devices have become smaller and lighter, multi-modal GPS based travel surveys have become extensively applied as travel survey method. Draijer, Kalfs, and Perdok (2000) was the first study to expand a GPS-based data collection to support several modes of transport. Respondents were asked to wear a GPS and a PDA device on all trips. There was however a consistent underreporting of trips due to the size and weight of the devices (approx. 2 kg). These trips were walking, cycling, and public transport trips as well as trips with the purpose of shopping and visiting friends. The survey design demanded a constant effort from the participants, as the respondents were asked to turn the device on/off when starting/ending a trip and answer questions on the PDA. Several studies have combined GPS traces with additional information gathered by a travel survey questionnaire. Among these are the studies by de Jong and Mensonides (2003), Bohte and Maat (2009) and Tsui and Shalaby (2006). These used internet-based questionnaires where respondents needed to confirm the trips identified by the trip identification algorithm. Papinski et al. (2009) compared planned and observed route choices for 21 vehicle-based commuting trips in Kitchener–Waterloo in Ontario, Canada. The survey required a high effort by the respondents by performing a pre-interview, GPS data collection, and a post-interview.

Much has been done to reduce the effort needed by the respondents, and many studies today therefore do not ask participants to provide trip information en-route (Schüssler & Axhausen, 2009; Stopher & Shen, 2011). This however sets higher requirements to the post-processing algorithms as these need to identify trip legs and mode of travel from raw GPS data consisting solely of time and space information. Later studies have proposed and analysed fully automatic GPS data processing methods (e.g., Schüssler and Axhausen (2009) and Bolbol et al. (2012)). These do not require any questionnaire data in the post-processing. Schüssler and Axhausen (2009) processed GPS data collected in Switzerland with no additional information provided by the respondents. The data set included 4,882 participants wearing the GPS devices for 6.65 days on average. The results were compared to the existing (national) travel survey. This showed that in aggregate figures the trip and mode identification only deviate slightly from that of the census data. However, the study did not perform any disaggregate evaluation of individual data. This was done in Bolbol et al. (2012), where 81 respondents wore a GPS device for 2 weeks but also answered a travel diary questionnaire. Based on speed and acceleration only, the study designated each trip to one of six different modes. When comparing to the travel diary it was found that most trips by car, train, bicycle, or walk could be inferred correctly. Some modes however had very similar speed and acceleration profiles, making it harder to distinguish between bus and metro, and between bus and bicycle.

Alternative approaches utilising information on local spatial information in the identification have also been proposed (e.g., Bohte & Maat, 2009; Chen, Gong, Lawson, & Bialostozky, 2010; Chung & Shalaby, 2005; Gong et al., 2011; Schüssler, 2010; Stopher et al., 2005; Tsui & Shalaby, 2006). These methods and applications have shortcomings in either (i) not including modes which are important in an application to the Greater Copenhagen area (rail is not included in Chung and Shalaby (2005), and bicycles are not included in Chen et al. (2010) and Gong et al. (2011)), (ii) relying on prompted recall surveys where participants need to verify their trips (Bohte & Maat, 2009; Stopher et al., 2005), and/or (iii) including a very small sample of participants (only 9 participants in Tsui and Shalaby (2006)). These shortcomings were addressed in the present study. Moreover, the study included the five most dominant modes in the Greater Copenhagen area (walk, bicycle, bus, rail, and car) and 183 participants totalling 644 person days of travel. The five modes cover in total 97.5% of all trips undertaken in the Greater Copenhagen area (according to the Danish National Travel Survey (Christiansen, 2012)) and the sample size is sufficiently large to validate the algorithms.

9.2.2 POST-PROCESSING OF GPS DATA

Post-processing of raw GPS data typically involves four steps, namely (i) GPS data cleaning, (ii) trip and activity identification, (iii) trip segmentation into single-mode trip legs, and (iv) mode identification. The approach varies slightly between studies, e.g. steps (ii) and (iii) are performed jointly in Stopher et al. (2005). Some analyses subsequently apply additional steps. Chen et al. (2010), Stopher et al. (2005) and others infer the purpose of the trips identified, while Schüssler and Axhausen (2009) map match the identified trip legs onto the corresponding modal networks.

Most analyses set off with a cleaning and filtering step, where systematic and random errors are removed from the data. This is often conducted by use of the number of satellites visible and the Horizontal Dilution Of Position (HDOP) (Nielsen & Jørgensen, 2004; Stopher et al., 2005). Random errors can be dealt with by including a data smoothing algorithm (Schüssler & Axhausen, 2009).

Trip end points (activity points) are often identified at a location where the device has been stationary for a period of time and/or if the spatial density of observations has been high for a period of time (Schüssler & Axhausen, 2009; Stopher et al., 2005). The result is a number of trips which are defined as being from one activity point to the next. This approach was evaluated in Schüssler (2010). The study finds that the algorithm correctly detected 97% of stated activities without detecting any false activities. Most studies further split trips into trip segments (or trip legs), defined by a change of mode. Correct trip segmentation is crucial for the subsequent identification of the mode of travel of the trip legs. de Jong and Mensonides (2003) divide trips into trip legs whenever the speed drops to 0 km/t, with the option to combine segments again if no mode change occurred. Schüssler (2010) and Tsui and Shalaby (2006) initially identify walking segments if speed and acceleration are low. This is done under the assumption that trip legs of all other modes are preceded or followed by such short walking segments (or by time gaps).

Several studies find that most modes can be identified by only using the speed and acceleration profiles gathered by the GPS device. Moreover, Bolbol et al. (2012) found that using the acceleration profile rather than the speed profile induces better results when distinguishing between modes. The best results were however found when combining the two profiles. This is an easy and efficient approach for the correct identification of some modes. Certain modes can however not be clearly distinguished by such an approach. For example, Bolbol et al. (2012) found that bus and bicycle trips in the Greater London area have similar speed and acceleration profiles. Tsui and Shalaby (2006) found that bus characteristics overlap with characteristics of several other modes.

Other techniques have been proposed to improve the mode detection. Among these are the application of map matching to mode-specific networks by use of GIS-software (e.g., Bohte & Maat, 2009; Chen et al., 2010; Chung & Shalaby, 2005; Gong et al., 2011; Schüssler, 2010; Stopher et al., 2005; Tsui & Shalaby, 2006). In Gong et al. (2011) rail and bus trip legs are identified based on the proximity of start and end locations to rail stations and bus stops. A similar approach for bus trip legs is proposed in Schüssler (2010). Using the proximity to bus stops of start and end locations to identify bus trips seems insufficient in urban areas where the bus network is extensive; trip legs starting and ending near bus stops might have been done by e.g. bicycle rather than by bus. Another approach is to utilise available information about the respondents implicitly in the identification of modes. Stopher et al. (2008) allow car or bicycle as mode for a trip leg only if the household has a car or bicycle at its disposal, respectively. However, these approaches have limitations if applied to a typical Scandinavian city where the bus network is extensive and the ownership and use of bicycles is relatively high.[1]

9.3 METHODOLOGY

This study develops and tests a fully automatic method to post-process GPS data without requiring any information about or from the person carrying the GPS unit. The method performs, and iterates between, a series of steps. These steps identify activities (trip ends), trip legs and the most probable mode chosen. The method is based on the automatic trip and mode detection algorithm developed in Schüssler and Axhausen (2009). This is modified in order to improve the results in three ways; (i) GIS analyses are used to better distinguish

[1] 80% of the inhabitants in Copenhagen have access to a bicycle (2012), and bicycling constitutes 36% (2012) of the trips (Municipality of Copenhagen, 2014). In the city of Odense bicycling constitutes 24% (2012) of the trips (Municipality of Odense, 2014).

between modes with similar speed and acceleration characteristics, (ii) advanced feedback loops between steps are used, allowing inconsistent mode-sequences to alter the trip leg detection algorithm, and (iii) map matching is used to exclude non-trips and hinder wrongly splitting trips on motorways. The method consists of a 6-step process as shown in Figure 56. The following subsections present a detailed description of the steps of the method.

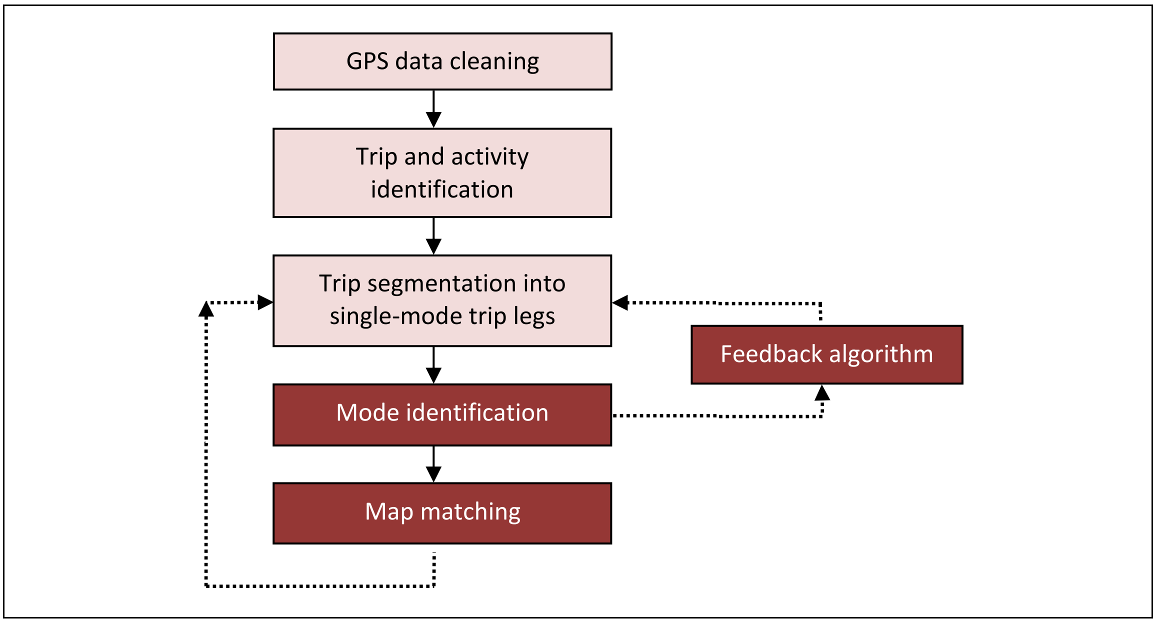


FIGURE 56: APPROACH USED IN THIS STUDY. BOXES HIGHLIGHTED IN LIGHT RED DENOTE STEPS THAT ARE SIMILAR TO CORRESPONDING STEPS IN SCHÜSSLER AND AXHAUSEN (2009). BOXES HIGHLIGHTED IN DARK RED ARE STEPS WHERE THIS PAPER CONTRIBUTES WITH NEW METHODS

9.3.1 GPS DATA CLEANING, TRIP IDENTIFICATION AND TRIP SEGMENTATION

The method initially cleans the data to remove erroneous observations, identifies activities thereby allowing the derivation of the trips, and subsequently segments these trips into trip legs. These three steps are directly adopted from Schüssler and Axhausen (2009). The data are cleaned based on the altitude level as well as the number of satellites visible and their dispersion (HDOP value). A Gauss kernel smoothing approach is applied to remove systematic errors and perform data smoothing. Activities (trip ends) are identified as locations where the bearer of the GPS device has been stationary for a period of time. Stationarity is not defined as 'complete' stationarity with no movement at all, since observations may be caught e.g. when walking within office-buildings which are not actual trips. Consequently, activities may be identified when the GPS unit (i) has not logged positions for a period of time (GPS units turn off if completely stationary), (ii) has been moving at a very low speed during a time period, or (iii) has been located within a limited area for a period of time. A trip between two activities might involve several trip legs with different modes of transport or with different vehicles of the same mode (e.g. changing between train lines). Trip legs are identified by assuming that a short walking segment is needed between modes or when changing vehicles, and the unique characteristics of walking (low acceleration and low speed) are used to identify these walking segments.

9.3.2 Mode identification

Each trip leg is associated with a most probable mode of transport. In Schüssler and Axhausen (2009) and in Bolbol et al. (2012) this is done based on speed and acceleration profiles. Driving conditions in metropolitan areas however range from being slow moving traffic through congested urban areas to fast moving traffic on motorways. Additionally, in urban areas it is hard to distinguish whether the respondent is driving in a bus or following close behind it in a car, or even biking next to it. These factors make it hard to distinguish modes solely based on the speed and acceleration profiles. The method proposed in the present study seeks to improve the mode identification by proposing the three-step mode identification process illustrated in Figure 57. The process is step-wise and utilises the speed and acceleration profiles as well as conduct advanced analyses in GIS software using digital representations of the infrastructure. The steps are explained further in the following subsections.

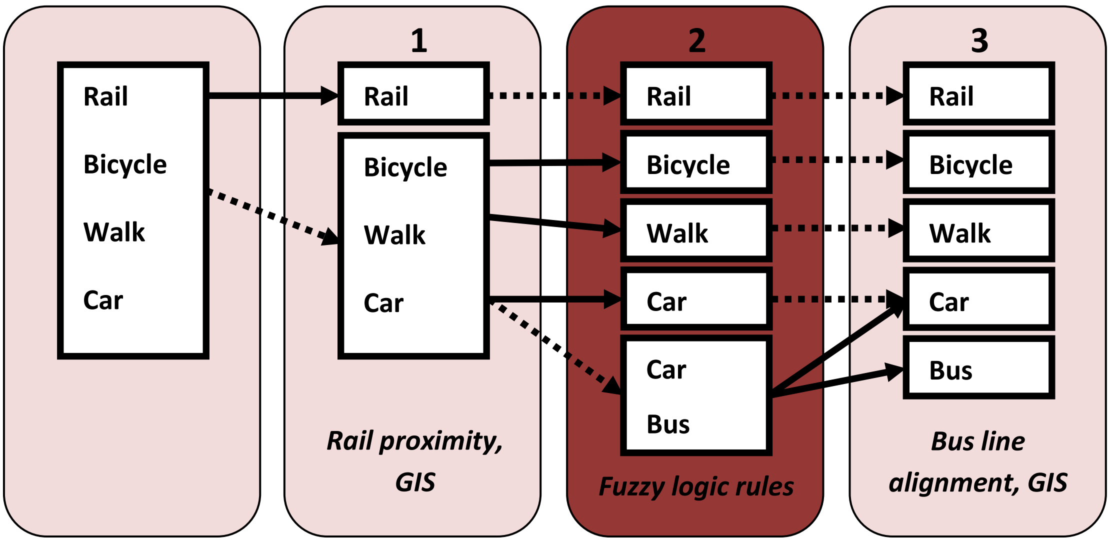


Figure 57: The stepwise mode classification algorithm. Continuous arrows denote mode classification whereas dotted arrows denote no change from previous step. Step 2 is directly adopted from Schüssler and Axhausen (2009), but with adapted fuzzy logic rules

9.3.2.1 Step 1: Rail proximity

Rail networks are typically characterised by not having the same spatial location as the street and path network (with the exception of on-street light rail and tram lines). This provides a way to, with a high chance of success, determine the correct mode for rail trip legs based on their proximity to the alignment of the rail network. The method performs a spatial analysis in GIS to identify the proximity of GPS observations to the rail network. If a large share of the observations of a trip leg is located within the proximity of the rail network, the trip leg is classified as a rail trip leg. Note that not 100% of the observations have to be identified as being in the proximity of the network. This accounts for potential small errors in the digital representation of the rail network and possible measurement errors of the GPS units. Also, a trip leg can only be classified as rail if it is longer than the length of the longest station platform in the network. This is required to avoid classifying within-platform walking trips as rail trip legs.

9.3.2.2 Step 2: Fuzzy logic rules

The next step assigns a mode of transport (walk, bicycle, car, or bus) to the remaining trip legs based on the speed and acceleration characteristics of the trip legs, as also done in Schüssler and Axhausen (2009). Hence, a mode of transport is assigned to each trip leg by dividing the profiles representing the median speed and peak values, e.g., 75–95 percentiles, of speed and acceleration into certain intervals and applying fuzzy logic rules across these for each trip leg (Gong et al., 2011; Schüssler & Axhausen, 2009; Stopher et al., 2005; Tsui & Shalaby, 2006). This means that a trip leg with a very high median as well as peak speed and acceleration is classified as car, since these characteristics are unique for car trips. Similarly, walk and bicycle trip legs have much lower acceleration and maximum/average speed making it possible to classify these.

Applying fuzzy logic rules to the speed and acceleration profiles however do not uniquely separate all modes (e.g., Bolbol et al., 2012; Tsui & Shalaby, 2006); car trips and bus trips may have highly overlapping speed and acceleration profiles in congested urban areas. More advanced analyses are thus needed to separate bus and car trips, and the method uses a twofold approach for this. Initially, the acceleration and speed profiles are used to distinguish between trip legs which can be assumed to definitely be car trips and trip legs which are either car or bus trips. Trip legs which are definitely car trips have very high speed and/or acceleration which cannot be obtained by buses. Trip legs which are either car or bus trip legs (defined as potential bus trips) have characteristics which do not allow a separation based on the profiles. Through this, the set of potential bus trips includes all actual bus trips and a large subset of car trips. Subsequently, the identification of actual bus trips among this set of initially classified potential bus trips are done in step 3 (Section 9.3.2.3). The trip legs not classified as bus trips in step 3 can be assumed to be car trips and added to the set of already identified car trip legs.

9.3.2.3 STEP 3: BUS LINE ALIGNMENT

The subset of potential bus trips are analysed in terms of coherence between GPS-recorded stop locations and bus line bus stops in order to identify whether the trip follows the stopping pattern of any bus line.

An initial step of the identification specifies that if at least a certain number of GPS observations are located within the proximity of a bus stop, the algorithm flags the trip leg as stopping at the bus stop. The distance and number of observations should reflect (i) the time a bus typically stops at a bus stop (longer than when just passing), and (ii) the length of two buses to account for the instance of two buses stopping at the same bus stop at the same time. Next, if the GPS trace stops at a certain share of potential bus stops between boarding and alighting stops on any bus line, the trip leg is flagged as a probable bus trip (on bus line(s) fulfilling this criterion). The threshold for the share of bus stops should reflect the characteristics of the bus line, e.g., a low threshold should be applied for bus lines with few passengers where the bus often does not stop at all bus stops and a higher threshold should be used for high demand bus lines (high level-of-service).

Subsequently, the trip legs in the sample of probable bus trip legs are analysed with regards to the location of their start and end. The start point and end point of each trip leg are analysed to see if they are both located within a certain distance from a bus stop on any of the bus lines identified previously. If this is the case, the trip leg is classified as a bus trip. Otherwise the trip leg is classified as a car trip. An additional benefit of applying this method is the identification of the most probable actual bus line used. Also note that while the rail proximity analysis (step 1) cannot capture trips using on-street rail lines such as e.g. trams or light rail, these can be identified by including trams and light rail in the bus line alignment algorithm.

9.3.3 ALGORITHMIC FEEDBACK

The trip leg and mode identification is improved by identifying and correcting illogical mode shift patterns in a subsequent feedback step. This is done to avoid wrong trip leg splitting and modal classification due to irregular changes in speed and/or acceleration for a trip leg. Such irregularities could e.g. arise in congested stop and-go traffic or if a bus has a long dwell time at a stop. The feedback algorithm uses simple rules to identify irregular mode changes and is based on a set of probable mode transfers. For example, it is likely that a bicycle trip leg follows or precedes a bus trip leg as some passengers might bicycle to and from the bus stop.

It is however not very likely that a bicycle trip leg is followed by a car trip leg with only a short time gap between the trip legs.

9.3.4 Map matching

The initial data cleaning seeks to remove scatter observations. The cleaning may however have failed to remove all such observations, and the subsequent trip leg and mode identification may have identified some non-trips. Non-trips are e.g. short trip legs generated as a consequence of the GPS device being turned on when no trip was actually undertaken (e.g. when persons are walking inside their home). The method removes these non-trips by map matching all trip legs (except rail trip legs) to a disaggregate network representation of the infrastructure. This map matching is done using an algorithm developed by Nielsen and Jørgensen (2004). Trip legs which cannot be map matched properly are classified as non-trips and removed from the data set (if none or only a little share of the mapped route is found by mapping of actual GPS observations[2]). The map matching also allows detecting and correcting trip legs which are wrongly split due to congestion on motorways; if the matched last link and first link of two consecutive trip legs are a motorway or a ramp then the two trip legs are merged into one. The mode of the merged trip leg is determined by re-running the mode classification on the merged trip leg.

9.4 Case study: Greater Copenhagen area

9.4.1 GPS data

The study area covers the Greater Copenhagen area in which approximately 2 million people live. The study utilised data which were collected as part of the ongoing research project ‘‘Analyses of activity-based travel chains and sustainable mobility’’ (Hansen, 2014). The data set included 53 households, corresponding to 183 persons from 6 to 58 years of age. The households were sampled from the Danish National Travel Survey (Christiansen, 2012). All participants were asked to bring a GPS device on all trips undertaken within a period of 3–5 days. Additionally, each respondent was asked to fill in an internet-based travel diary corresponding to one of the days for which GPS data were also collected. This enabled a validation of the proposed method.

The travel diaries were linked to the recorded GPS observations resulting in travel diaries with corresponding GPS data for 101 persons. Consequently, there were 82 persons for which data could not be linked. An analysis identified that this was due to one of the following three

[2] In cases where only a part of the observed route can be map matched, the map matching algorithm generates the shortest path between links to which observations can be mapped.

reasons; (i) the respondent failed to answer the survey for a day where he/she also carried a GPS device, (ii) no or only few GPS data were collected for the day where the survey was filled in, or (iii) there was a large difference between the number of trips reported and what could be seen in the GPS traces.

The GPS device was the wearable KVM BTT08 M (KVM, 2013). It logged data every second, thereby facilitating a high level of accuracy for the identification of en route travel choices and the location of trip ends. The data set contained in total 6,419,441 collected GPS points (observations) corresponding to 1,783 h of travel (including stationary and error data), and was collected on 644 person days of travelling.

The GIS-based analyses utilised a detailed digital representation of the road and public transport networks of the Greater Copenhagen area. The road network was based on the road network of NAVTEQ (2010) and was in a format that allowed for a complex map matching algorithm to be run (Nielsen & Jørgensen, 2004). The public transport network used for mode identification of rail trip legs was a digital representation of the rail line alignment. The analysis distinguishing between bus and car utilised a disaggregate public transport network representation containing information on bus route alignment and stop locations for all bus lines and bus line variants.

9.4.2 Configuration

9.4.2.1 GPS data cleaning, trip identification and trip segmentation

The method was based on the method of Schüssler and Axhausen (2009). However, some values were adjusted to fit the local setting. In the data cleaning step only observations with altitude levels between 37 m and +201 m were regarded as acceptable as this is the altitude range in Denmark +/30 m.

A trip was identified when one of three criteria were met, as described in Section 9.3.1. In the implementation these three criteria translate into the following; (i) if there was a time gap between consecutive observations of 120 s or more, (ii) if the speed had been lower than 0.01 m/s for at least 60 s, or (iii) if the location of the GPS device was within a limited area for at least 60 s. Trip legs were defined based on the identification of short walking segments between trip legs similar to in Schüssler and Axhausen (2009). A test of different values of the thresholds however led to a slight modification of some of the thresholds, e.g. a lower maximum speed when identifying walking trip legs due to the large variation of speeds of the extraordinarily many bicycle trips performed in Copenhagen.

9.4.2.2 MODE IDENTIFICATION

Rail trip legs are easy to identify as the rail network in the Greater Copenhagen area does not follow the street and path network. Consequently, a trip leg was classified as being a rail trip if more than 75% of the observations in one trip leg were located less than 25 m from the rail network. Additionally, the length of a rail trip leg was required to be at least 250 m. This is less than the shortest distance between railway stops in the Greater Copenhagen area, but longer than the longest platform. An example of a successfully identified rail trip leg is shown in Figure 58. In this example almost 98% of the observations for the trip leg were located within 25 m of the rail network.

Trip leg information	
Date:	09-11-2011
Starting time:	16:07:01
Ending time:	16:22:53
Trip length:	15 min 52 sec
Number of obs.:	947
Number of obs. within 25m:	924
Percentage rail:	97.6%

FIGURE 58: EXAMPLE OF A RAIL TRIP LEG (SHOWN WITH RED) ON THE DANISH S-TRAIN RING LINE (RINGBANEN) IDENTIFIED BY THE RAIL PROXIMITY ALGORITHM. THE RAILWAY NETWORK IS HIGHLIGHTED BY BOLD BLACK LINES, AND STATIONS WITH GREEN DOTS

The study used the 95 percentiles to represent peak speed and acceleration in the fuzzy logic rules. Each profile was divided into three or four (possibly overlapping) intervals. The division was based on an empirical analysis of the sample of trip legs in the data for which the mode was known, see Figure 59. Subsequently, the fuzzy logic rules reported in Table 22 were applied to classify each of the remaining trip leg as either walk, bicycle, car, or potential bus trips.

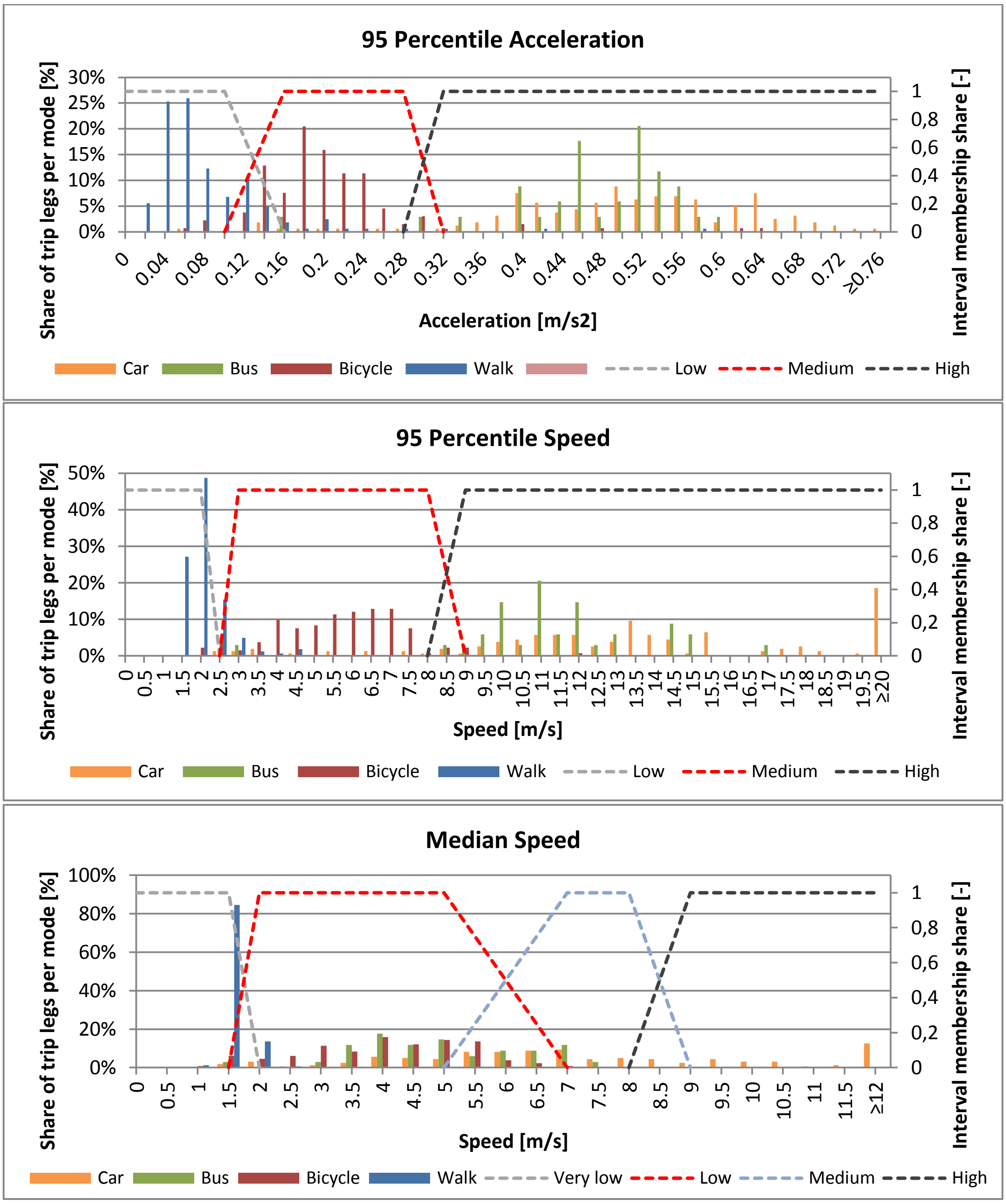


Figure 59: The distributions of the 95th percentiles of speed and acceleration and the median speed for the subset of trip legs for which the mode is known (rail excluded) from the control questionnaire

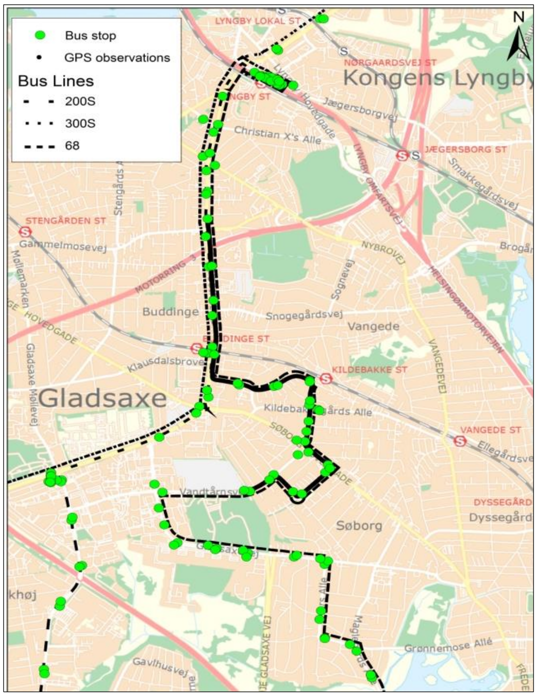
TABLE 22: FUZZY LOGIC RULES APPLIED

95th percentile acceleration	95th percentile speed	Median speed	Mode classification
Low	Low	Very low	Walk
Low	Medium	Very low	Walk
Low	High	Very low	Bike
Medium	Low	Very low	Walk
Medium	Medium	Very low	Walk
Medium	High	Very low	Bike
High	-	Very low	Car
Low	Low	Low	Walk
Low	Medium	Low	Bike
Low	High	Low	Bike
Medium	Low	Low	Bike
Medium	Medium	Low	Bike
Medium	High	Low	Bus
High	Low	Low	Bus
High	Medium	Low	Car
High	High	Low	Bus
Low	Low	Medium	Bike
Low	Medium	Medium	Bike
Low	High	Medium	Car
Medium	Low	Medium	Bike
Medium	Medium	Medium	Bike
Medium	High	Medium	Car
High	Low	Medium	Car
High	Medium	Medium	Car
High	High	Medium	Bus
Low	-	High	Car
Medium	-	High	Car
High	-	High	Car

The bus line alignment algorithm identified bus trip legs among the set of potential bus trips. A trip leg was defined as stopping at a given bus stop if at least 10 observations were logged within 25 m from the bus stop. The threshold for the minimum share of bus stops stopped at was set at 60% for low demand bus lines and 75% for high demand bus lines. 100 m was used as threshold for the distance of start and end point from stops on relevant bus lines.

Figure 60 shows two examples of the application of the method. The example to the left is an actual bus trip, whereas the example to the right is an actual car trip. In the example to the left the GPS carrier stopped at 14 out of 19 bus stops of bus line 68. The trip leg also began and ended close to bus stops on this line. This caused the trip leg to be correctly classified as a bus trip leg. In the example to the right, the GPS carrier stopped at several stops along bus line 161. The trip leg was however correctly classified as a car trip leg. This is because the trip leg started and ended more than 100 m away from a bus stop served by bus line 161.

Actual bus trip



Trip information

Date:	09-11-2011
Starting time:	16:41:24
Ending time:	16:53:18
Trip length:	11 min 54 sec

Bus line stop overlap	# stops	Hit%
200S	2/2	100%
300S	2/2	100%
68	14/19	74%

Origin stop:	68
Destination stop:	200S/300S/68
Classified as:	Bus trip
Actual bus line:	68

Actual car trip

Trip information

Date:	09-11-2011
Starting time:	06:42:26
Ending time:	06:49:07
Trip length:	6 min 41 sec

Bus line stop overlap	# stops	Hit%
161	3/5	60%

Origin stop:	None
Destination stop:	None
Classified as:	Car trip
Actual bus line:	-

FIGURE 60: EXAMPLE OF RESULTS FROM THE BUS STOP ALGORITHM. THE LINES REPRESENT RELEVANT BUS LINES, WHILE BLACK AND GREEN DOTS REPRESENT GPS POINTS AND STOPS, RESPECTIVELY

9.4.2.3 Algorithmic feedback

The algorithmic feedback mechanism searched for sets of two or three consecutive trips legs fulfilling certain criteria indicating non-likely mode-sequences and wrong split of trips, e.g. due to congestion. The search criteria and the resulting mode classification of the joined trip legs are listed in Table 23.

Table 23: Feedback algorithms joining trip legs. ΔTime and ΔSpace refer to temporal and spatial distances between end of trip A and start of trip B (C for 3 leg cases), respectively. Order is the classification of the 2 (3) trip legs and Classification is the mode assigned to the merged trip leg

	ΔTime	ΔSpace	Order	Classification
2 consecutive trip legs A→B	≤ 120sec.	≤ 25m	Car-Bicycle	Rerun mode identification
			Bicycle-Car	
			Bus-Car	*Probable* bus trip
			Car-Bus	
3 consecutive trip legs A→B→C	≤ 300sec.	≥ 25m	Car – Bus – Car	Car trip
			Car – Bicycle – Car	Car trip
			Car – Train – Car	Car trip
			Car – Walk – Car	Car trip
			Bicycle – Bus – Car	Car trip
			Bicycle – Bicycle – Car	Car trip
			Bicycle – Bus – Car	Car trip
			Bicycle – Train – Car	Car trip
			Bicycle – Walk – Car	Car trip
			Car – Bus – Bicycle	Car trip
			Car – Bicycle – Bicycle	Car trip
			Car – Train – Bicycle	Car trip
			Car – Walk – Bicycle	Car trip

9.4.2.4 Map matching

The map matching used the NAVTEQ road network (NAVTEQ, 2010). Non-trips were identified and removed when either no GPS observations could be map matched or if less than 50% of the mapped route was found by mapping of actual GPS observations.

9.5 Results

Two configurations of the method were tested on the available data set;

(i) Algorithm 1 without map matching including trip leg and mode identification as well as feedback algorithm (Sections 9.3.1–9.3.3), but excluding the map matching algorithm,

(ii) (ii) Algorithm 2 with map matching including Algorithm 1 without map matching and the map matching algorithm (Sections 9.3.1–9.3.4).

The effect of including the map matching could be evaluated by comparing the results generated by the two algorithms. A ''traditional'' Baseline algorithm was also tested for comparison. This included trip leg and mode identification as proposed by Schüssler and Axhausen (2009), i.e. with the mode identification step based solely on fuzzy logic rules. Different configurations of intervals as well as rules were tested for each of the three algorithms, and the best configuration was chosen for each algorithm.

The algorithms were run on the full data set consisting of approximately 644 person days of travel. The calculation time on a quad-core computer was less than two hours for the entire algorithm (Algorithm 2 with map matching), and most time was used for step 3 of the mode identification algorithm (the bus line algorithm). The algorithm was run using a batch-script calling Python for spatial analyses (Python scripts call analysis tools of the ArcGIS software package) and SAS for the numerical analyses and aggregation. The remainder of the section only report the results obtained when using a data subset where the travel mode was known from the additional questionnaire. This included trips that were directly connected to the travel diary data supplied by the respondents as well as trips where in-depth investigation made it possible to deduct the travel information manually. The results of the mode identification were evaluated using two assessment measures. The first measure is the success rate which denotes the number of correctly classified trip legs by the algorithm as percentage of the number of observed trip legs of that mode. The second measure is the confidence rate which denotes the number of correctly classified trip legs by the algorithm as percentage of the number of trip legs of that mode identified by the algorithm. Thus, the latter refers to the percentage of trip legs in the output of the algorithm where the mode was correctly identified. Hence, the first measure relates to the observed travel survey trip legs whereas the second measure relates to the trip legs in the output of the algorithm which may also include non-trips (see Section 9.3.4).

9.5.1 Trip leg identification

The total number of trip legs identified was 752, 741 and 427 if using the Baseline algorithm, Algorithm 1 without map matching and Algorithm 2 with map matching, respectively. This compares to the total number of reported trip legs in the subset of the travel survey of 521. Three sources of error influenced these numbers.

(i) There were trip legs in the travel survey where no corresponding GPS trip legs could be identified. This could be due to either the respondent not wearing the

GPS device, the GPS device not being able to get an acceptable signal or the device not functioning properly.

(ii) Some trip legs were identified by the algorithm even though no corresponding trip information was reported by the respondents in the diary. This error was partly due to underreporting by the respondents. Underreporting has also been observed in other studies including Stopher et al. (2007) and Wolf, Oliveira, and Thompson (2003). Another reason was the identification of non-trips (see Section 9.3.4).

(iii) The algorithm sometimes wrongly separated a trip leg into several trip legs due to long dwell times while travelling. This could for example occur in stop-and-go congested traffic. The opposite was also observed, namely that several actual trip legs were identified as one trip leg if the dwell time(s) between trip legs was very low.

Figure 61 illustrates the results of a comparison between the trip legs identified by the algorithms and the trip legs reported by the respondents.

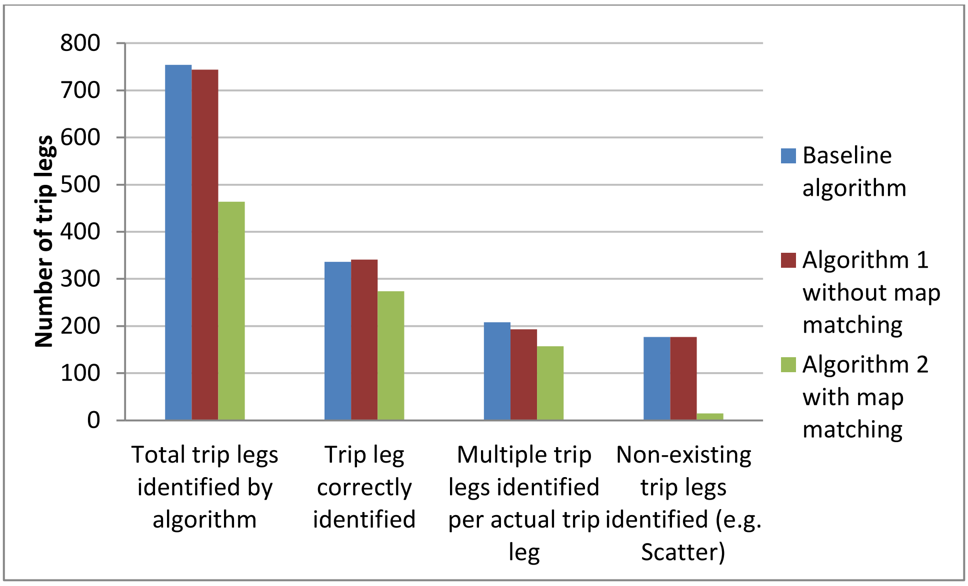


FIGURE 61: CLASSIFICATION OF TRIP LEGS IDENTIFIED BY THE ALGORITHMS

The Baseline algorithm generated trip legs with the correct origin and destination for 45% of the identified trip legs. 28% of the identified trip legs were partial trip legs, i.e. one reported trip leg was identified as two or more trip legs by the algorithm. A further 24% of the identified trip legs were non-trips which should not have been detected as a trip leg. The remaining 4% represents trip legs that either included several actual trip legs (not split correctly) or trips where observations had a too low quality for general usage.

The stepwise mode classification algorithm and the feedback algorithm improved the results (Algorithm 1 without map matching). Fewer trip legs were identified and more trip legs were correctly identified. Additionally, the feedback algorithm caused fewer actual trip legs to be wrongly split into several trip legs, as eleven partial trip legs were successfully connected into actual complete trip legs. The map matching algorithm of Algorithm 2 with map matching connected further nine trip legs successfully into four actual trip legs and 143 trip legs were correctly removed from the sample, cf. Figure 61. However, further analysis showed that some trip legs which should be merged remained unconnected. Overall, the best results were achieved when using Algorithm 2 with map matching as this identified the entire actual trip leg as one trip leg in 59% of the cases, and the entire actual trip leg as one or several trip legs in 93% of the cases. Additionally, the percentage of wrongly identified trip legs (non-trips) dropped to a very low level (3%). An analysis of the distribution of the length of the resulting trip legs found that many of the trip legs removed by Algorithm 2 with map matching were short non-trips. This induces that Algorithm 2 with map matching generated the best coherence with the length of the reported trip legs.

9.5.2 Mode identification

This section presents the methods' capability to identify the correct mode of transport of the trip legs. Table 24 reports the results of a disaggregate comparison between the mode identified by the algorithm and the actual chosen mode (success rate) for the Baseline algorithm. Approximately 82% of the trip legs were assigned the correct mode of transport when only considering trip legs which were actual trip legs. Table 25 reports the corresponding results for Algorithm 1 without map matching. The success rate obtained is 90%. Consequently, results were improved considerably by including the stepwise mode classification algorithm and feedback algorithm. Especially the method to identify rail trips was very efficient – using the fuzzy logic rules caused 24% of the rail trip legs to be correctly identified, whereas the corresponding number for Algorithm 1 without map matching was 97%. Applying the method proposed to identify car and bus trips also improved the results considerably, as the success rate for bus rose from 38% to 76% while the success rate for car rose from 82% to 93%. The success rates for walking and bicycling reduced slightly for Algorithm 1 without map matching when compared to the Baseline algorithm.

Table 24: The results of the mode identification when using Baseline algorithm (compared to reported mode use)

	Observed							
Algorithm	Walk	Bicycle	Bus	Car	Rail	Non-trips		Confidence rate
Walk	184	12	2	6	-	118		57.1%
Bicycle	9	121	-	13	-	53		61.7%
Bus	-	1	14	9	-	2		53.8%
Car	-	4	21	143	25	2		73.3%
Rail	-	-	-	3	8	1		66.7%
Other	-	-	-	-	-	1		-
Total	**193**	**138**	**37**	**174**	**33**	**177**		**62.2%**
Success rate	**95.3%**	**87.7%**	**37.8%**	**82.2%**	**24.2%**	**-**	**81.7%**	

Table 24 and Table 25 however also highlight a weakness of the two approaches. Both approaches identified many trip legs which were not reported in the diary (non-trips generated due to e.g. scatter). This induced the overall confidence rates to be 62% and 69% for the Baseline algorithm and Algorithm 1 without map matching, respectively. The stepwise mode classification algorithm classified trip legs considerably better, especially for bicycle, bus (no generated bus trip legs were wrongly classified) and rail. Summarising, the comparison between the Baseline algorithm and Algorithm 1 without map matching showed that applying the stepwise mode classification algorithm and the feedback algorithm improved the mode classification (especially the success rate) considerably. The two algorithms however – as also found in Section 9.5.1– identified many non-trips.

Table 25: The results of the mode identification when using Algorithm 1 without map matching (compared to reported mode use)

	Observed							
Algorithm	Walk	Bicycle	Bus	Car	Rail	Non-trips		Confidence rate
Walk	180	11	2	2	-	117		57.7%
Bicycle	2	114	-	6	-	16		82.6%
Bus	-	-	28	-	-	-		100.0%
Car	4	8	7	156	1	41		71.9%
Rail	3	-	-	-	33	2		86.8%
Other	3	1	-	3	-	1		-
Total	**192**	**134**	**37**	**167**	**34**	**177**		**69.2%**
Success rate	**93.8%**	**85.1%**	**75.7%**	**93.4%**	**97.1%**	**-**	**90.6%**	

Many of such non-trips were removed when adding the map matching algorithm of Algorithm 2 with map matching (see Section 9.5.1). This improved the overall confidence rate from 69% to 91% (Table 26). The improvement in confidence rate was however at the cost of also removing a large number of generated trip legs for which a corresponding observed trip leg exists. Specifically, many actual trip legs undertaken by foot or bicycle were discarded by the map matching algorithm (e.g., short walking trips through parks, etc.). This was probably a consequence of the map matching being conducted on a road network not including bicycle paths and footpaths. The row denoted by ''Success rate (all)'' highlights this issue. The measure represents the share of the total number of observed trip legs for which a generated trip leg with the correct mode was identified.

All trip legs identified as bus by the proposed algorithms were correctly classified (confidence rate), but some actual bus trips were wrongly classified (success rate). A disaggregate analysis identified two primary reasons for this. The first reason was problems associated with the trip leg identification algorithm and the feedback algorithm. The trip leg identification algorithm caused some actual bus trips to be wrongly split into several trip legs due to congestion, longer dwell times, etc. The feedback algorithm subsequently failed to identify and reconnect these. The other reason was the actual stopping pattern of the buses. At times some buses may have skipped a large percentage of stops, e.g. during evening hours where fewer passengers board the bus.

TABLE 26: THE RESULTS OF THE MODE IDENTIFICATION WHEN USING ALGORITHM 2 WITH MAP MATCHING (COMPARED TO REPORTED MODE USE)

	Observed							
Algorithm	Walk	Bicycle	Bus	Car	Rail	Non-trips		Confidence rate
Walk	75	5	1	1	-	3		88.2%
Bicycle	1	100	-	5	-	-		94.3%
Bus	-	-	29	-	-	-		100.0%
Car	1	7	7	151	1	4		88.3%
Rail	-	-	-	-	33	-		100.0%
Other	1	1	-	1	-	-		-
Total	**78**	**113**	**37**	**158**	**34**	**7**		**90.9%**
Total (all)	**192**	**134**	**37**	**167**	**34**	**-**		
Success rate	**96.2%**	**88.5%**	**78.4%**	**95.6%**	**97.1%**	**-**	**92.4%**	
Success rate (all)	**39.1%**	**74.6%**	**78.4%**	**90.4%**	**97.1%**	**-**	**(68.8%)**	

9.5.3 FURTHER WORK

The method to identify trips and trip legs were adopted directly from POSDAP (2012). We found that at times the approach wrongly split trips into several trip legs, and that this also influenced the results of the mode classification. Though the algorithmic feedback captured some of these wrongly split trip legs, more research is needed in the correct detection of trip legs. New methods could be developed which use the available disaggregate digital representation of the infrastructure, possibly in combination with congestion measures. Such an approach could e.g. hinder that trips are wrongly split into several trip legs when queuing at intersections.

The present study analysed all generated trip legs and found that many of these did not have a corresponding observed trip leg reported in the travel diary. This can partly be because of underreporting, but the analysis found that many non-trips were identified around activity locations. While the method developed and tested in this present study aimed at removing such non-trips (through the map matching step), the other studies reviewed did not seem to explicitly deal with these (important) non-trips. The map matching algorithm succeeded in removing most of the non-trips, however at the cost of also removing trip legs which were actually performed. These wrongly removed trip legs were primarily walking and bicycle trips. The incorrect removal of these could be partly explained by the use of the street network for the map matching. Additional research could test whether expanding the network to also include bicycle paths and footpaths would further improve the results.

The GPS units did not log positions when travelling underground, e.g. in some parts of the metro network and the suburban rail network. Signal was lost when the train runs underground (or when the traveller entered an underground station) and then reappeared when the train returned to the surface (or when traveller returned to street level at a station). There are however only three stretches where trains run underground in the rail network of the Greater Copenhagen area (with only two obvious transfer locations), allowing the underground trips to be correctly reproduced. More research is needed for the reproduction of trips in more complex underground networks where multiple change possibilities exist (e.g. the subway network of New York City).

9.6 TRANSFERABILITY AND SENSITIVITY

The proposed method was successfully applied to a case study of the highly complex and multi-modal transport network of the Greater Copenhagen area. The method is however generally applicable across case studies for which there are disaggregate digital representations of the infrastructure available. When transferring the model across case studies it is important to consider the impact of the differences in the characteristics of the built environment and the transportation network. This impact may be considerable, and the parameters and thresholds of the method may therefore need to be adapted correspondingly to ensure good model performance. In this fitting process it is important to have available a GPS data set with corresponding travel diaries. However, transferring and fitting the method across case studies are easier if the results are not highly sensitive to changes in the specification of parameters and thresholds. To investigate this further, the remainder of this section reports the results of sensitivity analyses towards the specification of various key parameters and thresholds. These relate to the main contributions of this study, namely the mode identification and the removal of non-trips, and we test different setups of these components on the Greater Copenhagen area case study.

9.6.1 MODE IDENTIFICATION

The parameters and thresholds used in the rail trip leg identification process is (i) the threshold for the share of observations in the vicinity of the rail network (distance $\leqslant$ 25 m, length $\geqslant$ 250 m), (ii) the maximum distance to the rail network (share of observations $\geqslant$ 75%, length $\geqslant$ 250 m), and (iii) the minimum trip leg length (share of observations $\geqslant$ 75%, distance $\leqslant$ 25 m). Table 27 reports the results of the sensitivity tests of these three thresholds.

Table 27: Sensitivity tests for step 1: rail proximity. Comparison of classified rail trip legs to reported mode for different thresholds for (i) the share of observations in vicinity, (ii) the distance to the network, and (iii) the length of the trip legs

	Observed				
	Walk	**Bicycle**	**Bus**	**Car**	**Rail**
Share of observations in vicinity of rail network (≥) [%]					
0	33	54	24	94	33
25	10	5	2	5	33
35	9	1	2	1	33
45	5	1	0	0	33
50	3	1	0	0	33
55	3	1	0	0	33
60	2	1	0	0	33
65	1	1	0	0	33
70	0	1	0	0	33
75	0	0	0	0	33
80	0	0	0	0	33
85	0	0	0	0	31
90	0	0	0	0	31
95	0	0	0	0	30
100	0	0	0	0	25
Maximum distance to rail network (≤) [m]					
5	0	0	0	0	8
10	0	0	0	0	25
15	0	0	0	0	31
20	0	0	0	0	33
25	0	0	0	0	33
30	0	1	0	0	33
40	1	1	0	0	33
50	2	1	1	0	33
75	8	2	1	0	33
Minimum trip leg length (≥) [m]					
0	4	0	0	0	33
50	4	0	0	0	33
100	3	0	0	0	33
150	1	0	0	0	33
200	1	0	0	0	33
250	0	0	0	0	33
800	0	0	0	0	33
900	0	0	0	0	30
1000	0	0	0	0	30

The results are not highly sensitive towards the choice of any of the three thresholds. Extremes should however be avoided in all cases: reasonable results are generated when specifying a threshold between 60% and 80% for the minimum share of observations in the vicinity of the alignment, a maximum distance of 15–40 m to the network, and a minimum trip leg length longer than the platforms (200 m in the Greater Copenhagen area) and shorter than the smallest distance between stations.

Fuzzy logic rules are used to distinguish walk, bicycle, potential bus trips and trips which are definitely undertaken by car. Sensitivity tests were performed towards the definition of the fuzzy logic rules and the intervals of the profiles.

Table 28 reports the results, namely the number of correctly/ wrongly classified trip legs obtained for different combinations of varying the definitions of the fuzzy logic rules and the intervals of the profiles. The detailed configuration of the combinations are too comprehensive to report, but include the removal or addition of rules and different degrees of variation of the intervals around the definitions reported in Figure 59. The table includes only results for walk and bicycle trips since these are classified in this step while car and bus are distinguished in the subsequent bus line alignment step.

TABLE 28: SENSITIVITY TESTS FOR STEP 2: FUZZY LOGIC RULES. NUMBER OF TRIP LEGS CORRECTLY/WRONGLY CLASSIFIED AS WALK OR BIKE. EACH VERSION DENOTES DIFFERENT SPECIFICATIONS FOR INTERVALS OF SPEED AND ACCELERATION PROFILES AND DIFFERENT FUZZY LOGIC RULES

	Walk	**Bike**
Specification 1	174/18	120/14
Specification 2	174/18	120/16
Specification 3	165/26	106/34
Specification 4	163/25	124/16
Specification 5	164/26	122/16
Specification 6	168/23	121/17
Specification 7	183/9	116/18

No specification results in confidence rates of 100%. The results vary across specifications, but in general the correct identification only differs slightly between specifications. The method does thus not seem to be highly sensitive towards the specification of the fuzzy logic rules and the intervals. Consequently, while it is important to validate the results via comparisons to travel diaries it is not problematic to use a specification which deviates slightly from the optimal specification.

Two different thresholds were tested for the bus line alignment step, and the results are reported in Table 29. It can be seen that the threshold for the share of bus stops stopped at should not be too low to hinder classifying car and bicycle trips as bus trips. It should not be too high either as this translates into wrong classification of actual bus trips. A comparison of the 0.6/0.6, 0.75/0.75 and the 0.75/0.6 (Table 29) highlights that distinguishing between low and high level-of-service (LoS) bus lines improves the result. A sensitivity analysis of the threshold for the time in vicinity of the stops shows that the threshold should be approximately 10 s to allow the correct classification of most bus trip legs and hinder the wrong classification of trip legs of other modes as bus trips. The thresholds used in the bus line alignment should thus be chosen with parsimony, preferably supported by a sensitivity analysis comparing classified and corresponding reported mode choices.

TABLE 29: SENSITIVITY TESTS FOR STEP 3:BUS LINE ALIGNMENT. FIT WITH OBSERVED MODE FOR DIFFERENT THRESHOLDS FOR THE SHARE OF BUS STOPS STOPPED AT AND THRESHOLD FOR TIME IN THE VICINITY OF THE BUS STOPS

	Observed **Walk**	**Bicycle**	**Bus**	**Car**	**Rail**
Share of bus stops high/low LoS (≥)					
0.5/0.4		2	30	12	
0.6/0.6		1	29	1	
0.75/0.6			29		
0.75/0.75			19		
0.9/0.75			16		
Time vicinity stops (≥) [s]					
5		2	35	38	
10			32		
15			29		
20			16		
30			10		
45			0		

9.6.2 MAP MATCHING

The map matching was found successful in removing non-trips, however at the cost of also removing some actual trips from the data set. Table 30 highlights the results for different values of the threshold for the minimum share of a trip which has to be mapped based on GPS observations. An influence is seen, but in general the results are not highly sensitive to the threshold, as long as it is specified as below 60%. Values above this induce the exclusion of many actual trips.

TABLE 30: SENSITIVITY ANALYSIS OF MAP MATCHING COMPONENT TO REMOVE NON-TRIPS. NUMBER OF CORRECTLY/WRONGLY CLASSIFIED TRIP LEGS REMAINING FOR VARIOUS THRESHOLDS OF THE MINIMUM SHARE BEING MATCHED BY USING GPS OBSERVATIONS. ALSO LISTED IS THE NUMBER OF NON-TRIPS REMAINING IN THE DATA SET

	Observed				
	Walk	**Bicycle**	**Bus**	**Car**	**Non Trip**
	[correct/wrong classification by algorithm]				
Share fully matched (≥) [%]					
0	76/3	100/13	29/8	151/7	8
30	76/3	100/13	29/8	151/7	8
40	76/3	100/13	29/8	151/7	7
50	75/3	100/13	29/8	151/7	7
60	75/3	97/12	29/8	150/6	6
70	74/3	92/11	29/8	150/6	6
100	73/3	72/9	28/6	121/6	5

9.7 CONCLUSIONS

Automated post-processing procedures are essential to facilitate the use of GPS data for transport surveys. This paper has presented a fully automated and disaggregate method to process raw GPS data and classify trips, trip legs, and the most probable mode of transport used. The method is applicable to all cases where data is collected as individual-based GPS traces and where detailed digital information on the local infrastructure is available. This study applied the method to GPS data collected in the Greater Copenhagen area.

The method performs, and iterates between, a series of steps. While being based on the automatic trip and mode detection algorithm developed in Schüssler and Axhausen (2009), the method contributes by utilising (i) available disaggregate information on the local infrastructure to conduct GIS analyses to better distinguish between modes with similar speed and acceleration characteristics, (ii) advanced feedback loops between steps, allowing inconsistent mode-sequences to alter the trip leg detection, and (iii) map matching to exclude non-trips and hinder wrongly splitting of trips on motorways.

Two variants of the method proposed were tested, one algorithm with the map matching step and one without it. This showed that including map matching improves the confidence and success rates by removing many non-trips, however at the cost of also removing some actually performed (primarily walking) trips. Both variants produced success rates above 90% when comparing to the control sample. These results are promising in comparison to the overall success rates obtained in other studies. Gong et al. (2011), Chen et al. (2010) and

Bolbol et al. (2012) obtained success rates of 82.6%, 79.1% and 87.4%[3] respectively. Chung and Shalaby (2005) obtained a success rate of 91.6% in their study including 60 trips. Especially the success rates of 78% for bus and 97% for rail are high when compared to other studies; Gong et al. (2011) and Bolbol et al. (2012) obtained success rates of 35.7% and 84.1% for rail, and 62.5% and 58.3% for bus. The current study also applied the method proposed by Schüssler and Axhausen (2009) on the same data set. This allowed evaluating whether the high success rates were generated due to special circumstances related to the case study rather than improvements in the methodology. Success rates of 24% and 38% were obtained for rail and bus, respectively, when using this existing algorithm. This verified that the high success rates for the two proposed algorithms were generated as a result of applying the suggested advanced feedback algorithm and utilising the available disaggregate network data.

The deployment of the method does not require the respondents to provide any information beyond the GPS traces. It is however important to note that the parameters and thresholds used may need to be adapted/calibrated to fit the characteristics of the specific case. To do this, it is necessary to have available a control sample of corresponding revealed or stated information of trips undertaken by the respondents. A sensitivity analysis revealed that the correct removal of non-trips and correct identification of rail, walk, and bicycle trip legs are not highly sensitive to the specification of the thresholds and parameters. The specification of the parameters used in the distinction between car and bus trip legs should however be done with parsimony, preferably via sensitivity analyses comparing with corresponding observed mode choices.

The study has contributed to literature by demonstrating much improved fit rates in the detection of trips, trip legs, and mode of transport used. Through this we believe that the abilities of automatic post-processing methods are causing travel surveys based on GPS data collection to be highly attractive, even for complex multi-modal study areas.

ACKNOWLEDGEMENTS

The authors would like to thank the Danish Research Council for financing the project 'Analysis of activity-based travel chains and sustainable mobility' (ACTUM), during which the work presented in this paper has been conducted. In addition, the authors would like to thank Dr. Nadine Rieser, from ETH Zürich, Switzerland for providing support for the Java-based software used in parts of the analysis.

[3] Success rate has been calculated as the rate is not presented in Bolbol et al. (2012).

9.8 REFERENCES

Bierlaire, M., Chen, J. and Newman, J. (2013). A probabilistic map matching method for smartphone GPS data. *Transportation Research Part C: Emerging Technologies*, 26, 78–98.

Bohte, W. and Maat, K. (2009). Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: A large-scale application in the Netherlands. *Transportation Research Part C: Emerging Technologies*, 17(3), 285–297.

Bolbol, A., Cheng, T., Tsapakis, I. and Haworth, J. (2012). Inferring hybrid transportation modes from sparse GPS data using moving window SVM classification. *Computers, Environment and Urban Systems*, 36(6), 526–537.

Chen, C., Gong, H., Lawson, C.T. and Bialostozky, E. (2010). Evaluating the feasibility of a passive travel survey collection in a complex urban environment: Lessons learned from the New York City case study. *Transportation Research Part A: Policy and Practice*, 44(10), 830–840.

Christiansen, H. (2012). *Documentation of the Danish National Travel Survey*. Note 2012:5, DTU Transport, Department of Transport, Kgs. Lyngby, Denmark.

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

de Jong, R. and Mensonides, W. (2003). *Wearable GPS device as a data collection method for travel research*. Working Paper, ITS-WP-03-02, The University of Sydney, Sydney, Australia.

Draijer, G., Kalfs, N. and Perdok, J. (2000). Global positioning system as data collection method for travel research. *Transportation Research Record*, 1719, 147–153.

Du, J. and Aultman-Hall, L. (2007). Increasing the accuracy of trip rate information from passive multi-day GPS travel datasets: Automatic trip end identification issues. *Transportation Research Part A: Policy and Practice*, 41(3), 220–232.

Forrest, T. and Pearson, D. (2005). Comparison of trip determination methods in household travel surveys enhanced by GPS. *Transportation Research Record*, 1917, 63–71.

Gong, H., Chen, C., Bialostozky, E. and Lawson, C.T. (2011). A GPS/GIS method for travel mode detection in New York City. *Computers, Environment and Urban Systems*, 36(2), 131–139.

Hansen, C.O. (2014). *Analysis of activity-based travel chains and sustainable mobility*. [accessed 16.11.14].

Herrera, J.C., Work, D.B., Herring, R., Ban, X.J., Jacobson, Q. and Bayen, A.M. (2010). Evaluation of traffic data obtained via GPS-enabled mobile phones: The Mobile Century field experiment. *Transportation Research Part C: Emerging Technologies*, 18(4), 568–583.

KVM (2013). *GPS-BTT08M. Product description*. [accessed 25.02.14].

Li, H., Guensler, R., Ogle, J. and Wang, J. (2004). Using global positioning systems data to understand day-to-day dynamics of morning commute behavior. *Transportation Research Record*, 1895, 78–84.

Liu, L., Andris, C. and Ratti, C. (2010). Uncovering cabdrivers' behavior patterns from their digital traces. *Computers, Environment and Urban Systems*, 34(6), 541–548.

Mavoa, S., Oliver, M., Witten, K. and Badland, H. (2011). Linking GPS and travel diary data using sequence alignment in a study of children's independent mobility. *International Journal of Health Geographics*, 10(64), 1–10.

Municipality of Copenhagen (2014). *København Cyklernes By – Cykelregnskabet 2012*. Municipality of Copenhagen, Center for Traffic [in Danish].

Municipality of Odense (2014). *Mobilitetsplan Odense 2014–15. Municipality of Odense* [in Danish].

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

Nielsen, O.A. (2004). Behavioral responses to road pricing schemes: Description of the Danish AKTA experiment. *Journal of Intelligent Transportation Systems: Technology*, *Planning, and Operations*, 8(4), 233–251.

Nielsen, O.A. and Jørgensen, R.M. (2004). Map-matching algorithms for GPS data – methodology and test on data from the AKTA roadpricing experiment in Copenhagen. In *5th TRI annual symposium on transportation analysis (TRISTAN V)*, Le Gosier, Guadeloupe.

Nielsen, O.A. and Sørensen, M.V. (2008). *The AKTA road pricing experiment in Copenhagen*. In: Road pricing, the economy and the environment, advances in spatial science (pp. 93–109).

Papinski, D., Scott, D.M. and Doherty, S.T. (2009). Exploring the route choice decision-making process: A comparison of planned and observed routes obtained using person-based GPS. *Transportation Research Part F*, 12, 347–358.

POSDAP (2012). *Position data processing*. [accessed 16.11.14].

Prato, C.G., Rasmussen, T.K., & Nielsen, O.A. (2014). Estimating value of congestion and value of reliability from observation of car drivers' route choice behavior. *Transportation Research Record*, 2412, 20–27.

Quiroga, C.A. (2000). Performance measures and data requirements for congestion management systems. *Transportation Research Part C: Emerging Technologies*, 8(1–6), 287–306.

Quiroga, C.A. and Bullock, D. (1998). Travel time studies with global positioning and geographic information systems: an integrated methodology. *Transportation Research Part C: Emerging Technologies*, 6(1–2), 101–127.

Rainham, D.G., Bates, C.J., Blanchard, C.M., Dummer, T.J., Kirk, S.F. and Shearer, C.L. (2012). Spatial classification of youth physical activity patterns. *American Journal of Preventive Medicine*, 42(5), 87–96.

Rasmussen, T.K., Ingvardson, J.B., Halldórsdóttir, K. and Nielsen, O.A. (2013). Enhancing the use of Global Positioning Systems data from wearable GPS devices as travel survey method: A case study in the Copenhagen Region. In *2nd Symposium of the European Association for Research in Transportation*, September 4–6, Stockholm, Sweden.

Rich, J. and Nielsen, O.A. (2007). A socio-economic assessment of proposed road user charging schemes in Copenhagen. *Transport Policy*, 14(4), 330–345.

Schüssler, N. (2010). *Accounting for similarities between alternatives in discrete choice models based on high-resolution observations of transport behaviour*. Ph.D. thesis. ETH Zurich, Switzerland.

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

Stopher, P.R., Jiang, Q. and Fitzgerald, C. (2005). Processing GPS data from travel surveys. In *2nd international colloquium on behavioral foundations of integrated land-use and transportation models: frameworks, models and applications*, Toronto, Ontario, Canada.

Stopher, P.R., Clifford, E., Zhang, J. and Fitzgerald, C. (2008). *Deducing mode and purpose from GPS data*. Working Paper, ITLS-WP-08-06, The University of Sydney, Sydney, Australia.

Stopher, P.R., Fitzgerald, C. and Xu, M. (2007). Assessing the accuracy of the Sydney Household Travel Survey with GPS. *Transportation*, 34(6), 723–741.

Stopher, P.R. and Shen, L. (2011). In-depth comparison of global positioning system and diary records. *Transportation Research Record*, 2246, 32–37.

Tsui, S.Y.A. and 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.

Wagner, D.P., Neumeister, D.M. and Murakami, E. (1996). *Global positioning system for personal travel surveys*. FHWA, U. S. Department of Transportation.

Wolf, J. (2000). *Using GPS data loggers to replace travel diaries in the collection of travel data*. Ph.D. thesis, Georgia Institute of Technology.

Wolf, J., Oliveira, M. and Thompson, M. (2003). Impact of underreporting on mileage and travel time estimates: Results from global positioning system-enhanced household travel survey. *Transportation Research Record*, 1854, 189–198.

Yalamanchili, L., Pendyala, R.M., Prabaharan, N. and Chakravarthy, P. (1999). Analysis of Global Positioning System-based data collection methods for capturing multistep trip-chaining behavior. *Transportation Research Record*, 1660, 58–65.

Chapter 10

EFFICIENCY OF CHOICE SET GENERATION METHODS FOR BICYCLE ROUTES

Title:	Efficiency of choice set generation methods for bicycle routes
Author(s):	Katrín Halldórsdóttir, Nadine Rieser-Schüssler, Kay W. Axhausen, Otto A. Nielsen, and Carlo G. Prato
Presented:	The *1st European Symposium on Quantitative Methods in Transportation Systems (LATSIS Symposium),* 4th – 7th September 2012, Lausanne, Switzerland
Published:	*European Journal of Transport and Infrastructure Research (EJTIR)*, vol. 14, no. 4, pp. 332-348.
Abbreviated:	Halldórsdóttir et al. (2014)

A preliminary version of this paper was presented at the 1st European Symposium on Quantitative Methods in Transportation Systems in September 2012, Lausanne, Switzerland. The paper was then updated, rewritten, and published in the European Journal of Transport and Infrastructure Research in December 2014.

ABSTRACT

The current study analyses the efficiency of choice set generation methods for bicycle routes and proposes the extension of cost functions to bicycle-oriented factors not limited to distance and time. Three choice set generation methods for route choice were examined in their ability to generate relevant and heterogeneous routes: doubly stochastic generation function, breadth first search on link elimination, and branch & bound algorithm. Efficiency of the methods was evaluated for a high-resolution network by comparing the performances with four multi-attribute cost functions accounting for scenic routes, dedicated cycle lanes, and road type. Data consisted of 778 bicycle trips traced by GPS and carried out by 139 persons living in the Greater Copenhagen Area, in Denmark. Results suggest that both the breadth first search on link elimination and the doubly stochastic generation function generated realistic routes, while the former outperformed in computation cost and the latter produced more heterogeneous routes.

Keywords: *bicycle route choice, bicycle route generation, branch and bound, breadth first search, choice set generation, stochastic generation function.*

10.1 Introduction

The growing interest in the transition towards the use of sustainable transport modes motivates the search for factors determining the selection of the bicycle as a viable alternative to the car. Bicycle route choice models provide directions to this search, but findings from these models depend on the observation of actual route choices and the generation of plausible alternatives. While the former is a challenge that recent enhancements in technology and software help tackling, the latter is a challenge that recent advances in path generation help confronting, not without the uncertainty related to the dependency of model estimates on choice set composition (see, e.g., Bekhor et al., 2006; Bliemer and Bovy, 2008; Prato and Bekhor, 2007).

The literature in bicycle route choice shows that most studies have used stated preference (SP) data (e.g., Axhausen and Smith, 1986; Bovy and Bradley, 1985; Hopkinson and Wardman, 1996; Hunt and Abraham, 2007; Krizek, 2006; Sener et al., 2009; Stinson and Bhat, 2003; Tilahun et al., 2007), while only a few studies have used revealed preference (RP) data (e.g., Aultman-Hall et al., 1997; Broach et al., 2011; Broach et al., 2012; Hood et al., 2011; Howard and Burns, 2001; Hyodo et al., 2000; Menghini et al., 2010; Shafizadeh and Niemeier, 1997). On the one hand, SP data trade the easiness in individuating alternative routes with the possible bias in predefining the factors being relevant to route choices of cyclists. On the other hand, RP data trade the realism of observing actual cyclists' behaviour with the challenge of generating plausible alternative routes prior to model estimation.

While the challenge of collecting RP data (i.e., actual route choices) has greatly benefitted from enhancements in GPS device technology, GPS data post-processing (e.g., Schüssler and Axhausen, 2008, 2009b; Stopher, 2009; Tsui and Shalaby, 2006), and highly detailed network digitalization, the challenge of generating plausible alternative routes is still testing. Most studies have focused on implementing path generation methods for cars or public transport, which are normally generated on a simplified network, and only few studies have focused on bicycle route choice sets, which require a highly detailed network. Menghini et al. (2010) applied a Breadth First Search on Link Elimination (BFS-LE) method (Rieser-Schüssler et al., 2012) while limiting the cost function to the route length. Broach et al. (2010) compared a modified route labelling method to a K-shortest paths link penalty (Cascetta et al., 1996; de la Barra et al., 1993; Ramming, 2001), a simulated shortest paths (Bekhor et al., 2006; Ramming, 2001), and labelled routes method (Bekhor et al., 2006; Ben-Akiva et al., 1984; Ramming, 2001). The modified route labelling method performed best out of all four methods, however obtained lower coverage in comparison with studies focusing on car route choice. This is likely because the network used in the study was a realistic "all streets" network, while most car route choice studies cited used much coarser auto networks. Hood et al. (2011)

implemented a doubly stochastic generation function (DSGF) (Bovy and Fiorenzo-Catalano, 2007; Nielsen, 2000) with a multi-attribute cost function while observing better performance than a BFS-LE method with single-attribute cost function. They obtained a slightly higher coverage than Broach et al. (2010). Notably, the first two studies defined cost functions either containing only one attribute or containing only travel time and distance. Although Hood et al. (2011) included a multi-attribute cost function, they only manage to reproduce one-third of the observed routes. This emphasises the importance of both identifying factors that are important in the choice set generation process for bicycle route choice as well as exploring algorithmic issues in generating a plausible set of path choice alternatives in a highly detailed network.

The current study extends the body of knowledge on choice set generation for bicycle route choice. Firstly, the current study applies to the bicycle context the three most effective path generation methods in the car context: BFS-LE, DSGF, and branch & bound (B&B) (Hoogendoorn-Lanser et al., 2006; Prato and Bekhor, 2006). These methods are chosen in order to investigate their transferability in the context of choice set generation for bicycle routes choice, given their proven ability in reproducing observed car route choices. Secondly, the current study evaluates the efficiency of the three path generation methods in a high-resolution network by using different evaluation methods, such as replicating the observed routes while also generating realistic alternatives that take into account taste heterogeneity across cyclists. Lastly, the current study extends the path generation literature by proposing multi-attribute cost functions that account for scenic routes, dedicated cycle lanes, and road type. These attributes are relevant to the route choice of cyclists and are important when intending to estimate models providing insights into the factors relating to cyclists' choices of routes.

The remainder of this paper is organised as follows. Section 10.2 describes the data collected for this study. Section 10.3 describes the path generation methods applied in this study, the bicycle-tailored multi-attribute cost functions, and the methods used to evaluate the efficiency of the methods. Section 10.4 presents the results from the implementation and the comparison of the path generation methods. Section 10.5 discusses the results and draws conclusions.

10.2 DATA

The current study uses a dataset consisting of 778 bicycle trips, traced by GPS and carried out by 139 persons living in the Greater Copenhagen Area in Denmark. In addition to collecting GPS tracks, travel diaries were collected from a sample of the participants by means of a web-based survey.

Extensive data processing was required to obtain data that could be used for choice set generation of bicycle routes, as the GPS data collection and the travel diaries focused on all modes of transport in the Greater Copenhagen Area. The post-processing procedure used is described in detail in Schüssler and Axhausen (2008, 2009b). Initially, various criteria were first used to filter the data, e.g., the number of satellites in view, altitude value, the Horizontal Dilution of Precision (HDOP) value, sudden jumps in position, etc., followed by a Gauss kernel smoothing approach to remove random errors. Then, different criteria were applied to identify trips and activities and trips were divided into single-mode stages. Last, modified fuzzy logic rules were applied to identify the transport mode for each stage using the median speed, 95th percentile acceleration, and 95th percentile speed. The original fuzzy logic rules had to be altered in order to better fit the travel behaviour of Danes and extended with more disaggregate components using GIS software (Rasmussen et al., 2013). The travel diaries were used for validation of the post-processing procedure.

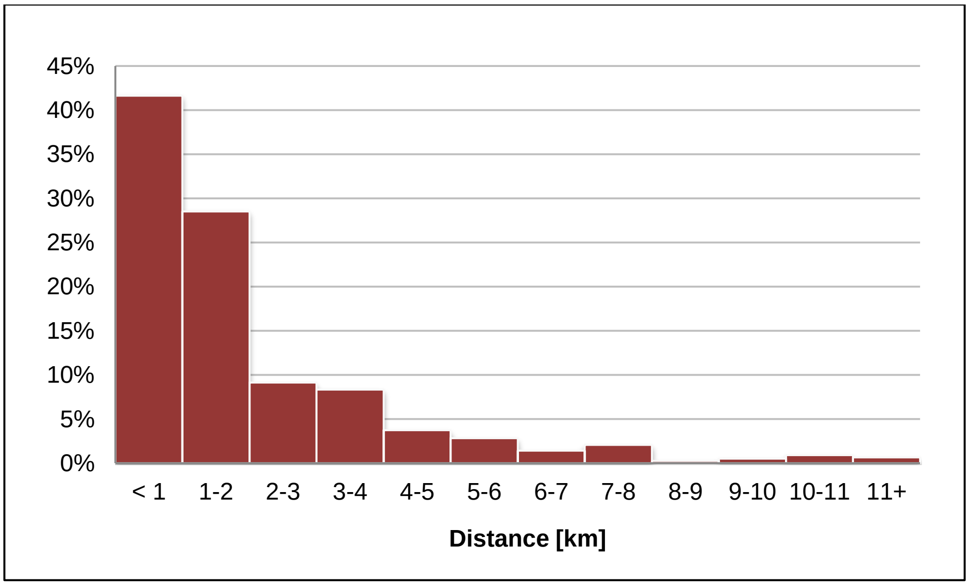


FIGURE 62: DISTRIBUTION OF THE OBSERVATIONS OVER DISTANCE TRAVELLED

After the data post-processing, a total of 1,824,034 GPS points and 4,552 stages were identified, and 490,062 points and 1,026 stages were retained for further analysis after the travel mode was identified as bicycle.

After the filtering, the processed bicycle trips were mapped to a high-resolution network using the map-matching algorithm developed by Schüssler and Axhausen (2009a) that extended a previous algorithm developed by Marchal et al. (2005). The network consists of

110,893 nodes and 272,586 directional links for the study area, and the high-resolution is the result of the compilation of various sources in order to consider attributes that were considered relevant to bicycle route choice such as road type, segregated cycle paths, and land use information. The network includes small paths only accessible by bicycles and pedestrians, and does not include motorways and expressways where cycling is illegal in Denmark.

After the map-matching algorithm was run, there were 778 stages remaining for the choice set generation. In some cases, stages were filtered because of missing network links or because the traveller was cycling off-road, which resulted in the lower number of stages. The average distance was approximately 2 km and approximately 90% of the trips were less than 5 km. Figure 62 shows the distribution of the observations over distance travelled.

10.3 METHODS

This section introduces the applied path generation methods, the defined cost functions and the implemented comparison methods. It should be noted that a maximum choice set size of 20 alternative routes was defined prior to choice set generation, and a time abort threshold was predefined in order to move on to the next observation after processing origin-destination pairs for which the choice set generation could not be completed within the time interval.[4]

10.3.1 CHOICE SET GENERATION METHODS

Doubly stochastic generation function

The DSGF (Bovy and Fiorenzo-Catalano, 2007; Nielsen, 2000) accounts for variation in travellers' link cost and differences in travellers' attribute preferences by drawing random costs and random parameters from probability distributions. Advantages of this method are the inherent heterogeneity of the generated alternatives and its computational efficiency in large networks, even though the randomisation of link costs and the parameters can be time consuming in a high-resolution network.

In the DSGF method, a shortest path search is carried out iteratively using an implementation of the Dijkstra´s algorithm (Dijkstra, 1959) on a realisation of the network. At each iteration, the realisation of the network is obtained by randomly drawing the cost of each link from a probability distribution and extracting attribute preferences for each traveller from another probability distribution. After each iteration, only unique routes not generated in previous

[4] The alternative routes were generated using a tool developed in Java, originally developed for the automatic processing of GPS tracks to reconstruct travel diaries (POSDAP) (see www.sourceforge.net/projects/posdap).

ones are added to the route choice set as the same route may be found several times during the process, even though the realisations of the network are obtained from different costs and preference parameters. The shortest path search is repeated iteratively until the preselected maximum choice set size is achieved or the predefined time abort threshold is reached.

Although the literature reports the implementation of a multi-attribute cost function for the DSGF method in the bicycle route choice context (Hood et al., 2011), the current study extends this implementation by testing and reporting the results of four different cost functions that consider not only route length and time, but also bicycle-oriented factors such as preferences for road types, dedicated cycle paths, and land use. As the correct definition of choice sets is a necessary condition for obtaining meaningful parameter estimates, including these bicycle-oriented factors is essential to the in-depth understanding of cyclists' preferences.

Breadth first search on link elimination

The BFS-LE method (Rieser-Schüssler et al., 2012) combines a breadth first search with topologically equivalent network reduction. The procedure concentrates on generating a route set, which afterwards can be reduced to an individual choice set. Its advantage is a high computational efficiency in a high-resolution network while ensuring a significant level of route variety.

The algorithm searches for the shortest path between origin and destination. Consecutively, the links of the shortest path are removed one by one and the shortest path for the resulting network is determined. Once all links of the original shortest path have been processed, the algorithm proceeds to the next level, where two links at a time are eliminated. The algorithm monitors the generated networks and only keeps unique and connected routes. The algorithm continues until the maximum number of unique routes has been generated, the time abort threshold is met or there are not more feasible routes between origin and destination.

As the literature reports only single-attribute cost functions for the implementation of the BFS-LE method in the bicycle route choice context (Hood et al., 2011; Menghini et al., 2010), the current study proposes to examine the efficiency of the BFS-LE method with the same multi-attribute cost functions implemented for the DSGF method. The same input parameters are applied for the two methods, although obviously for the BFS-LE the error components and the error terms are not extracted from a probability distribution.

Branch & bound

The B&B method (Hoogendoorn-Lanser et al., 2006; Prato and Bekhor, 2006) constructs a connection tree between the origin and destination of a trip by processing link sequences according to a branching rule accounting for logical constraints devised to increase route likelihood and heterogeneity. The strength of this algorithm is the realism and the heterogeneity of the generated routes, but a disadvantage is the high computation time in a high-resolution network. Rieser-Schüssler et al. (2012) tested the method with GPS observed car trips in a high-resolution network where the average number of links per chosen route was 65.69. This proved to be too high for the B&B algorithm, as within a reasonable computation time it only managed to produce alternatives for origin-destination pairs connected by very short paths. However, empirical results have shown that the algorithm can be applied to networks of different sizes by applying different design parameters (Bekhor and Prato, 2009).

In the current study, different behavioural constraints were tested to exclude links that: (i) take the cyclists farther from the destination and closer to the origin (directional constraint); (ii) cause the travel time to be excessively high when compared to the shortest path (temporal constraint); (iii) cause the cyclist to have a detour larger than an acceptable value (loop constraint); (iv) are shared by other routes that would not be considered as separate alternatives (similarity constraint). Table 31 presents the input parameters. Firstly, basis values were implemented according to the indications by Prato and Bekhor (2006). Then, relaxed values were applied to allow for higher travel times and greater overlap, and restrictive values were used to allow for lower travel times and less overlap. It should be noted that the time abort threshold was applied differently than the BFS-LE and the DSGF, namely the B&B searched through the network tree before checking whether the time restriction has been met. As the search tree was rather large because of the highly detailed network, at times the algorithm took longer to compute and the choice set size exceeds the maximum. Consequently, this method could not be compared to the other two algorithms in relation to computation time or number of unique routes created for each chosen route. However, the structure of the derived choice set could be compared.

TABLE 31: IMPLEMENTATION OF THE BRANCH & BOUND GENERATION TECHNIQUE

	Basis	Restricted	Relaxed
Directional factor	1.10	1.10	1.10
Temporal factor	1.50	1.33	1.67
Loop factor	1.20	1.20	1.20
Similarity factor	0.80	0.75	0.85

10.3.2 COST FUNCTIONS

Four cost functions were tested for the BFS-LE and the DSGF methods with various input parameters being considered with the aim to get realistic route alternatives, to obtain the best coverage over the average of the sample, and to capture heterogeneous preferences across cyclists. The parameters accounted for preferences of road types, cycle lanes, and land use. Table 32 presents an overview of the calibrated parameters and the variation factors applied in the cost functions. The following sup-sections describe in more detail the tested cost functions and the calibration of the parameters.

TABLE 32: OVERVIEW OF THE CALIBRATED PARAMETERS AND VARIATION FACTORS

	Cost function			
	Road type	**Cycle lanes**	**Land use**	**Road type, cycle lanes, and land use**
Parameters				
Large roads	0.167	–	–	0.167
Small roads	0.333	–	–	0.333
Other roads	0.500	–	–	0.500
Segregated cycle lanes	–	0.400	–	0.400
No segregated cycle lanes	–	0.600	–	0.600
Scenic roads	–	–	0.125	0.125
Non-scenic roads	–	–	0.250	0.250
Forest roads	–	–	0.500	0.500
Non-forest roads	–	–	0.125	0.125
Variation factors				
$\xi_{Large\ roads}$	2	–	–	2
$\xi_{Small\ roads}$	10	–	–	10
$\xi_{Other\ roads}$	10	–	–	10
$\xi_{Segregated\ cycle\ lanes}$	–	2	–	2
$\xi_{No\ segregated\ cycle\ lanes}$	–	10	–	10
$\xi_{Scenic\ roads}$	–	–	2	2
$\xi_{Non\text{-}scenic\ roads}$	–	–	3	3
$\xi_{Forest\ roads}$	–	–	7	7
$\xi_{Non\text{-}forest\ roads}$	–	–	2	2
ε_a	2	2	2	2

Road type

Cost functions should consider information regarding different road types for cyclists, and in the case of absence of specific information, regarding different road types in general. For example, three different road types may be considered (e.g., large traffic roads, small traffic roads, other roads). Larger roads are usually properly equipped with segregated cycle paths and cycle lanes (e.g., Hunt and Abraham, 2007; Stinson and Bhat, 2003) in the Greater Copenhagen area, while smaller roads have usually fewer or non-segregated cycle lanes, thus resulting in a mixed traffic scenario between cyclists and motorists, but in return have lower speed limits and traffic volumes (e.g., Axhausen and Smith, 1986; Antonakos, 1994). Other roads are a mixture of pedestrian paths or shared bicycle and pedestrian paths.

The first cost function represented individuals with different preference for the three road types:

$$C_a = \sum_k \left(\left(\beta_{RoadType_k} + \xi_{RoadType_{ak}} \right) \cdot RoadType_{ak} \cdot Length_a \right) + \varepsilon_a \tag{18}$$

where C_a is the random cost of link a, $Length_a$ is the length of link a, $RoadType_{ak}$ is the road type k that link a belongs to, $\xi_{RoadType_{ak}}$ are error components related to road type k of link a, $\beta_{RoadType_k}$ are coefficients related to road type k, and ε_a is the random error term for link a.

The implementation of the cost function differs between DSGF and BFS-LE. In the DSGF method, the error terms ε_a express the random link costs and the error components $\xi_{RoadType_{ak}}$ express the heterogeneous preferences. Each error term ε_a was computed as the link length multiplied by a standard normal distribution and a variation factor that determined the “width” of the distribution from which the link lengths were drawn. Each error component $\xi_{RoadType_{ak}}$ was calculated as the corresponding parameter β multiplied by a standard normal distribution and a distribution factor that determined the “width” of the distribution from which the utility parameter itself is drawn in the doubly stochastic case. Calibration of the parameters resulted in the large traffic roads having the lowest cost and capturing preferences for high quality bicycle facilities, and the small and the other traffic roads having double and triple cost and capturing preferences for the most direct route and for paths not shared with pedestrians. In the DSGF method, each error term ε_a was computed with a variation factor of 2, and each error component $\xi_{RoadType_{ak}}$ were computed with a variation factor of 2 for large roads and 10 for small and other roads. In the BFS-LE method, variation in travellers’ link cost and differences in travellers’ attribute preferences are not considered, and hence each error term ε_a is equal to zero for every link a and each error component $\xi_{RoadType_{ak}}$ is equal to one for every road type k and every link a.

Cycle lanes

Cost functions could consider information on cycle lanes, and in particular on Copenhagen-style lanes that are segregated lanes with raised curbs separating the cycle lane from the motorised traffic on one side and from the pedestrians on the other side. Studies have shown that some cyclists prefer routes that are separated from motorised traffic (e.g., Hunt and Abraham, 2007; Stinson and Bhat, 2003) and are willing to take detours to travel on bicycle paths, while others prefer the most direct route (e.g., Stinson and Bhat, 2005).

The second cost function characterised individuals with different perspectives to use segregated bicycle paths:

$$C_a = \sum_k \left(\left(\beta_{BikeLanes_k} + \xi_{BikeLanes_{ak}} \right) \cdot BikeLanes_{ak} \cdot Length_a \right) + \varepsilon_a \tag{19}$$

where C_a is the random cost of link a, $Length_a$ is the length of link a, $BikeLanes_{ak}$ indicates the presence of cycle lane configuration k on link a, $\xi_{BikeLanes_{ak}}$ are error components related to cycle lane configuration k in link a, $\beta_{BikeLanes_k}$ are coefficients related to cycle lane configuration k, and ε_a is the random error term for link a.

Calibration of the parameters resulted in roads with no segregated cycle lanes having almost two times larger cost than roads with segregated bicycle lanes. In the DSGF method, each error term ε_a was computed with a variation factor of 2, and each error component $\xi_{BikeLanes_{ak}}$ was calculated with a variation factor of 2 for roads with segregated cycle paths and 10 for roads without segregated cycle paths. In the BFS-LE, each error term ε_a was equal to zero for every link a and each error component $\xi_{BikeLanes_{ak}}$ was equal to one for every cycle lane configuration k and every link a.

Land use

Cost functions could also consider land use attributes, in particular when referring to scenic areas such as the lakes in the Greater Copenhagen Area, as scenic routes are considered attractive for cyclists (e.g., Antonakos, 1994; Axhausen and Smith, 1986). Major forest areas should also be considered, as the bicycle paths in forests are usually dirt paths and in some cases only suitable for mountain bicycles.

The third cost function assumed that cyclists have different preferences when travelling in different areas:

$$C_a = \sum_k \left(\left(\beta_{LandUse_k} + \xi_{LandUse_{ak}} \right) \cdot LandUse_{ak} \cdot Length_a \right) + \varepsilon_a \tag{20}$$

where C_a is the random cost of link a, $Length_a$ is the length of link a, $LandUse_{ak}$ indicates the land use type k associated to link a, $\xi_{LandUse_{ak}}$ are error components related to land use type k for link a, $\beta_{LandUse_k}$ are coefficients related to land use type k, and ε_a is the random error term for link a.

Calibration of the parameters resulted in non-scenic roads (i.e., not alongside lakes) having double the cost compared to scenic roads, thus capturing the preference for scenic routes. Also, calibration resulted in links in forest areas having four times the cost of roads in non-forest areas, thus capturing the disutility of cycling on gravel or dirt paths. In the DSGF method, each error term ε_a was computed with a variation factor of 2, and each error component $\xi_{LandUse_{ak}}$ was calculated with a variation factor of 2 for roads alongside lakes and not in forests, 3 for roads not alongside lakes and 7 for roads in forests. In the BFS-LE, each error term ε_a was equal to zero for every link a and each error component $\xi_{LandUse_{ak}}$ was equal to one for every land use type k and every link a.

Road type, cycle lanes, and land use

The fourth cost function included all three cost attributes, thus capturing a multi-attribute heterogeneous preference structure across individuals:

$$\begin{aligned} C_a = &\sum_k \left(\left(\beta_{RoadType_k} + \xi_{RoadType_{ak}} \right) \cdot RoadType_{ak} \cdot Length_a \right) \\ &+ \sum_k \left(\left(\beta_{BikeLanes_k} + \xi_{BikeLanes_{ak}} \right) \cdot BikeLanes_{ak} \cdot Length_a \right) \\ &+ \sum_k \left(\left(\beta_{LandUse_k} + \xi_{LandUse_{ak}} \right) \cdot LandUse_{ak} \cdot Length_a \right) + \varepsilon_a \end{aligned} \tag{21}$$

Calibration of the parameters resulted in small traffic roads having double and other roads having triple the cost of large traffic roads, roads with segregated cycle lanes having slightly over two times higher cost than large traffic roads, and roads with no segregated cycle lanes having almost four times larger cost. Moreover, roads alongside lakes resulted having slightly lower cost than large traffic roads, while not having lakes along the route implied almost double cost. Also, roads in forests resulted having a cost three times higher than large traffic roads. In the DSGF method, each error term ε_a was computed with a variation factor of 2, and each error component $\xi_{RoadType_{ak}}$, $\xi_{BikeLanes_{ak}}$ and $\xi_{LandUse_{ak}}$ was calculated with the same variation factors applied to the cases with a single attribute in the cost function. In the BFS-LE, each error term ε_a is equal to zero for every link a and each error component $\xi_{RoadType_{ak}}$, $\xi_{BikeLanes_{ak}}$ and $\xi_{LandUse_{ak}}$ was equal to one.

10.3.3 EVALUATION METHODS

The effectiveness of the three choice set generation methods was evaluated by comparing the generated choice sets to the observed routes. If the choice sets contained the actual chosen route among paths produced with the generation methods, they were considered consistent with the observed behaviour. The consistency of a route choice set generation method was evaluated with respect to the observed behaviour by considering the length of the links shared between the generated route and the observed route for each choice set:

$$O_{nr} = \frac{L_{nr}}{L_n} \tag{22}$$

where O_{nr} is the overlap measure, L_{nr} is the overlapping length between the path generated by choice set generation method r and the observed path for cyclist n, and L_n is the length of the observed path for cyclist n. The coverage is defined as the percentage of observations for which an algorithm generates a route that satisfies a particular threshold for the overlap measure:

$$\max_r \sum_{n=1}^{N} I\left(O_{nr} \geq \delta\right) \tag{23}$$

where $I(\cdot)$ is the coverage function, where when its argument is true it is equal to 1 and when false it equals to 0, and δ is a threshold for the overlap measure.

In order to investigate the heterogeneity of the choice set composition, the path size was calculated for each route in each choice set (Ben-Akiva and Bierlaire, 1999):

$$PS_{in} = \sum_{a \in \Gamma_i} \left(\frac{l_a}{L_i}\right) \frac{1}{\sum_{j \in C_n} \delta_{aj}} \tag{24}$$

where PS_{in} is the path size factor, Γ_i is the set of all links of path i, l_a is the length of link a, L_i is the length of path i, C_n is the generated choice set for cyclist n, and δ_{aj} equals 1 if link a is on path i and 0 otherwise. The path size ranges between 0 and 1, indicating the portion of the route that constitutes a completely independent alternative. Thus, a unique route will have a path size equal to one, while two duplicate routes will each have a path size factor of ½, three routes that completely overlap will each has a size of ⅓, and so on. The path size distribution over the choice sets C_n indicates whether the route choice sets contains heterogeneous routes by representing their average degree of independence.

The behavioural consistency of the route choice set generation methods with respect to an ideal algorithm was measured with a consistency index (see Bekhor and Prato, 2006). Ideally, a choice set generation method would reproduce the observed behaviour perfectly by replicating link by link all the routes collected and would result in 100% coverage for a 100% overlap threshold. However, the actual choice set generation methods only partially reproduce the observed behaviour and generate different numbers of routes. The index measures the behavioural consistency of the methods by accounting for the total overlap over all the observations:

$$CI_r = \frac{\sum_{n=1}^{N} O_{nr,\max}}{N \cdot O_{\max}} \qquad (25)$$

where CI_r is the consistency index of choice set generation method r, $O_{nr,max}$ is the maximum overlap measure obtained with the paths generated by method *r* for the observed choice of each cyclist n, and O_{max} is the overlap 100% overlap over all the N observations for the ideal method.

10.4 Results

Table 33 presents the coverage results for the three choice set generation methods according to different overlap thresholds varying from complete replication to the reproduction of 70% of the collected routes. The first four rows show the results for the BFS-LE method, followed by the results for the DSGF method, and the results from the B&B method.

Table 33: Coverage measures of the applied algorithms

	Overlap threshold			
Algorithm	**100%**	**90%**	**80%**	**70%**
BFS-LE method				
Road type	62.2	67.9	72.8	78.8
Segregated bicycle path	66.1	72.0	78.0	82.6
Land use	62.0	67.0	74.6	81.9
All three cost attributes	67.9	74.8	80.1	84.8
Doubly stochastic generation function				
Road type	62.2	69.0	75.3	82.4
Segregated bicycle path	58.6	64.0	70.7	78.3
Land use	58.9	64.5	70.8	76.2
All three cost attributes	63.5	71.1	75.2	79.2
Branch and bound algorithm				
Basis	38.0	40.1	45.2	50.4
Relaxed	38.3	40.2	43.6	49.7
Restricted	40.4	42.7	46.5	51.3

The BFS-LE method duplicated 62% and up to almost 68% of the chosen routes, whereas at least 79% and up to almost 85% were reproduced with a 70% overlap threshold. The DSGF method replicated approximately 59% and up to almost 64% of the chosen routes, whereas more than 76% and up to more than 82% were reproduced with a 70% overlap threshold. The cost function with all three cost attributes had the highest coverage percentage at 100% overlap threshold for both methods and this finding showed the correctness of the hypothesis of selecting attributes other than distance and time. All three tests with the B&B method had very low coverage, and Table 34 suggests the reason.

Table 34 shows the consistency of the applied methods and their computational costs. The B&B method did not manage to generate any alternatives for a large percentage of the observations within the time abortion threshold, which resulted in a very low coverage as seen in Table 33. The majority of these observations were longer than 4 km. In few cases the DSGF method also did not generate any alternative routes. This is not necessarily considered a problem as these are relatively short trips and thus not applicable in route choice modelling.

Both the BFS-LE method and the DSGF method performed quite well in relation to the consistency index, where the BFS-LE with the cost function with all the attributes performed the best, followed by the DSGF with road type as an attribute. Not surprisingly, the B&B did not perform well.

The BFS-LE had a very low computational time, while the DSGF had a lot higher run time. Since the DSGF had larger computational costs, the method did not produce the maximum choice set size in some of the cases because of the time abort threshold. Consequently, this resulted in a lower number of unique routes. Since the B&B method did not follow the same restrictions as the other two methods in terms of time abort threshold and maximum choice set size, the method produced very high number of unique routes for some of the observations, and also resulted in high computation time. Consequently, the method was not comparable to the other two and no further results are presented relatively to the B&B.

TABLE 34: COMPUTATIONAL COSTS AND TECHNIQUES CONSISTENCY

	Consistency index	Number of unique routes	Number of observations with no alternative	Computational time*
BFS-LE method				
Road type	86.3	15,560	0	0h 5m
Segregated	88.2	15,560	0	0h 4m
Land use	86.9	15,560	0	0h 4m
All three cost	89.5	15,560	0	0h 4m
Doubly stochastic generation function				
Road type	88.5	13,603	0	24h 55m
Segregated	85.8	13,570	2	23h 30m
Land use	84.8	12,333	7	27h 51m
All three cost	87.3	11,613	7	38h 58m
Branch and bound algorithm				
Basis	54.8	49,911	345	33h 26m
Relaxed	54.5	47,411	347	33h 08m
Restricted	55.6	53,184	343	33h 28m

*Computations performed on an Intel(R) Xeon(R) COU E5-1650 0 @ 3.20GHz with 32 GB RAM running Windows 7 Professional.

To visualize the consistency of the BFS-LE and DSGF methods with respect to the observed behaviour, the distribution of coverage over the cumulative percentage of observations is presented in Figure 63. It can be seen that the methods follow a very similar pattern and do not exhibit a significant difference.

Figure 64 shows the average of the maximum coverage over the choice sets as a function of the length of the chosen route. The figure shows that there is a general trend, namely that the results are good (and obvious) for shorter routes, but there is a large difference for longer routes. The DSGF performed better up to 10 km, where the average coverage started slightly decreasing.

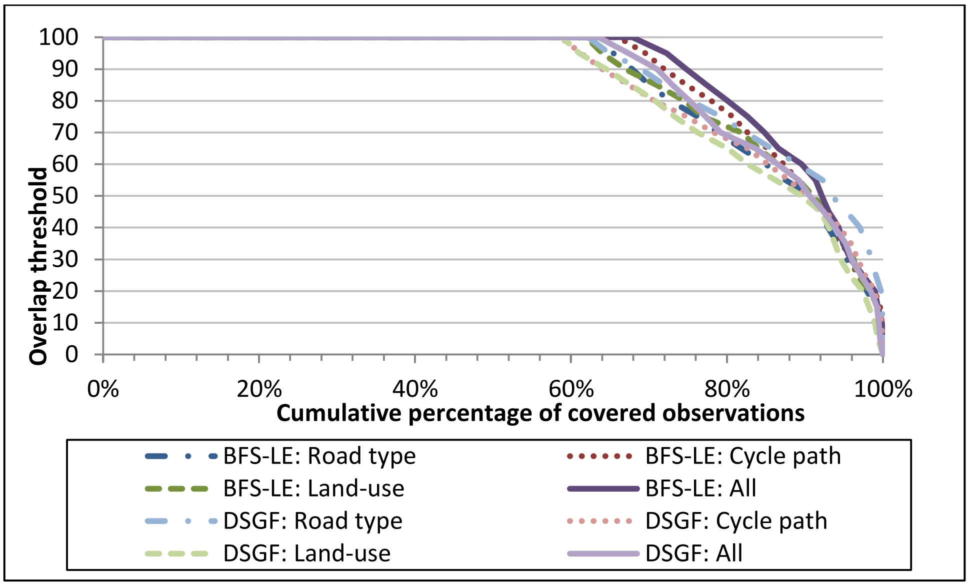


FIGURE 63: DISTRIBUTION OF COVERAGE OVER 778 OBSERVATIONS

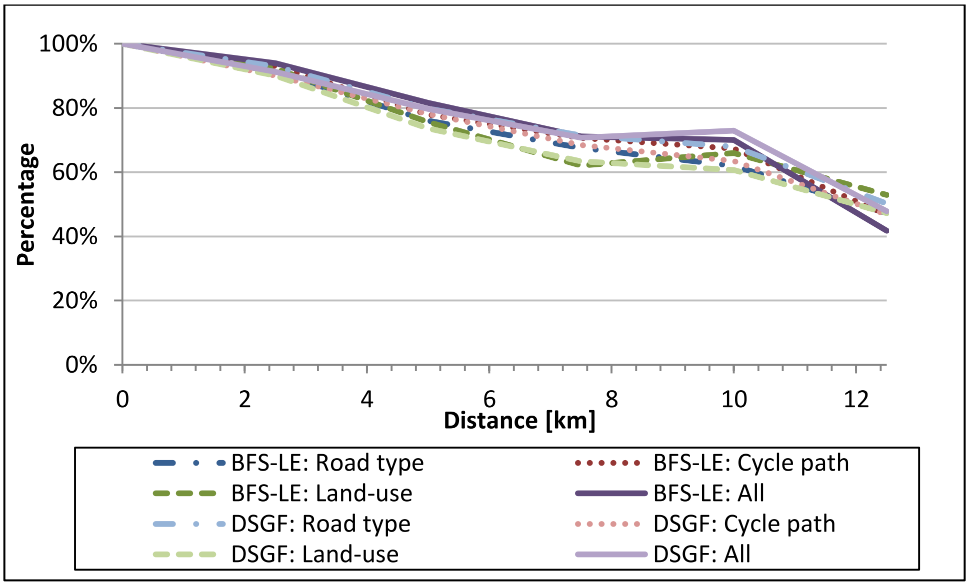


FIGURE 64: AVERAGE MAXIMUM COVERAGE OVER DISTANCE

Figure 65 shows the percentage of completely replicated chosen routes as a function of distance. It can be observed a general trend, namely the number of completely reproduced chosen routes decreases with increasing distance. As intuitively logical, longer routes appear problematic, as routes longer than 8 km were not completely reproduced by any of the choice set generation methods with the exception of one observation getting reproduced by the BFS-LE with the cost function accounting for preferences for specific road types.

Figure 66 shows the distribution of the average path size of all routes in the choice sets generated by the different methods. The path size distribution indicates considerably more diversity between routes generated with the DSGF method than the ones generated with the BFS-LE method.

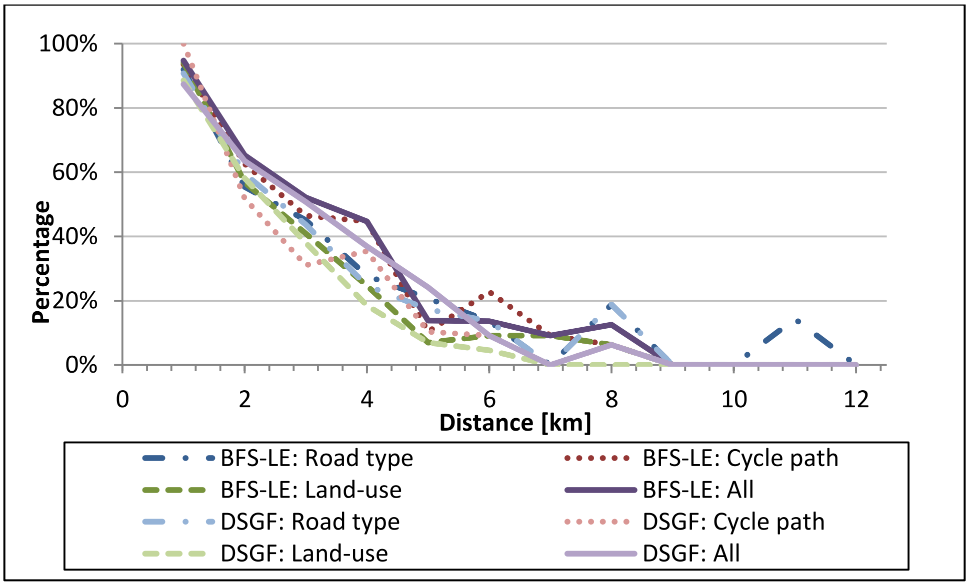


FIGURE 65: PERCENTAGE OF CHOSEN ROUTE COMPLETELY REPLICATED AS A FUNCTION OF DISTANCE

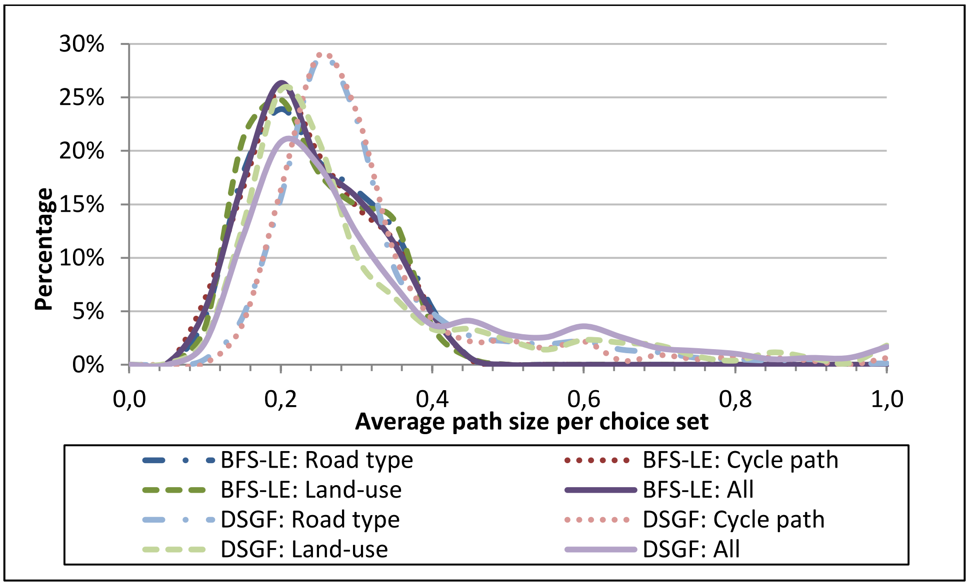


FIGURE 66: PATH SIZE DISTRIBUTION FOR THE RESULTING ROUTE SETS

10.5 DISCUSSION AND CONCLUSIONS

The current study focused on the efficiency of choice set generation of bicycle routes in a high-resolution network. Bicycle routes were collected with GPS devices and three choice set generation methods were implemented while designing cost functions that included not only distance and time related terms, but also other factors that cyclists would consider relevant, such as scenic routes, dedicated cycle lanes, and road types.

The efficiency of the methods was measured with respected to the observed behaviour of cyclists and the relative ability of generating relevant and heterogeneous alternative routes.

The efficiency of choice set generation methods for bicycle routes revealed similar performances in terms of coverage for the BFS-LE and the DSGF methods, with the BFS-LE method replicating almost 68% of the chosen routes and the DSGF method replicating almost 64%. In perspective, it should be noted that Broach et al. (2010) reproduced only less than one fourth of observed routes, and Hood et al. (2011) only replicated about one-third of observed routes, while using networks with significantly lower number of links (i.e., 3 and 8 times less detailed networks). Clearly, extending the cost function to factors that are relevant to cyclists and that do not pertain only distance and time, but also their preference for scenic routes and safe paths, provided a significant increase in performances when compared to previous findings. Among the cost functions, the fourth function with all three additional cost attributes had the highest coverage percentage at 100% overlap threshold for both methods, which indicated the heterogeneous and complex preference structure for cyclists when considering routes. The B&B method had lower coverage compared to the BFS-LE and the DSGF, as it reproduced approximately 40% of the observed routes and was more in line with previous findings. The problem with this method is the high computational time that did not allow reaching the destination within the time abort threshold for a large percentage of the observations.

When looking at the average maximum coverage over distance, a general trend emerged from the results. Specifically, shorter routes illustrated expected results in having very good coverage for all methods, while longer routes exhibited larger differences across algorithms, with the DSGF method performing better up to 10 km routes and the average coverage decreasing. Hood et al. (2011) found similar results of better coverage with the DSGF for longer observations. Moreover, the number of completely reproduced chosen routes decreased with increasing distance, and routes longer than 8 km were not reproduced.

When looking at the distribution of the average path size of the routes in the choice sets generated with the different methods, the DSGF method produced more heterogeneous alternative routes.

When looking at computational costs, the BFS-LE clearly outperformed the DSGF and the B&B. Hood et al. (2011) found instead that the DSGF (using a uniform probability distribution to draw the random error term) had lower computation time than the BFS-LE in a far less detailed network. A possible reason for this difference is that using a uniform probability distribution might cause the randomly drawn travel cost to deviate substantially from the real travel cost and hence might increase substantially the likelihood of generating unrealistic routes. In the current study, error terms and error components were randomly drawn from a standard normal distribution function while discarding the instances where a negative number was drawn that would explain the longer computation time. A more obvious reason for this difference is that algorithms have been programmed with different resources for diverse stopping criteria, observed routes, and considered networks. Two limitations of this study are indeed that a fairer comparison would entail the same programs being developed for the same dataset, and that the coding of the algorithms does not necessarily characterise the definite computation cost, but works only as a proxy. However, these are still relevant results and the better ability in reproducing routes recorded in this study clearly shows the importance of the random draws being from a normal distribution and even more of the cost function being multi-attribute and not only distance and time related.

Results suggest possible directions for further investigation, especially for longer routes. Possible improvements would be to include turn restrictions in the choice set generation, on the line of the turn restrictions constraint included in the original formulation of the B&B method (Prato and Bekhor, 2006). The cost function could also be extended with other attributes considered important for cyclists. Also, the availability of intersection data could contribute to improving further the cost function. Drawing from non-negative distributions (e.g., lognormal, gamma) could help reducing the computation time relating to discarding negative numbers. Last, comparison of model estimation performance and prediction accuracy could be carried out with datasets from the different choice set generation methods.

ACKNOWLEDGEMENTS

The authors acknowledge the support of Lara Montini at ETH-IVT in Zurich, Michael Balmer at Senozon, and Søren Hasling Pedersen and Jesper Bláfoss Ingvardson at DTU Transport. The authors gratefully acknowledge the financial support of the Danish Road Directorate to the project "The effect of cycling policies". The authors are thankful to the anonymous reviewers who provided insightful comments to the initial version of this paper.

10.6 REFERENCES

Antonakos, C.L. (1994). Environmental and travel preferences of cyclists. *Transportation Research Record*, 1438, 25-33.

Aultman-Hall, L., Hall, F.L. and 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. and Smith, R.L. (1986). Bicyclist link evaluation: a stated preference approach. *Transportation Research Record*, 1085, 7-25.

Bekhor, S., Ben-Akiva, M.E. and Ramming, M.S. (2006). Evaluation of choice set generation algorithms for route choice models. *Annals of Operations Research*, 144, 235-247.

Bekhor, S. and Prato, C.G. (2009). Methodological transferability in route choice modelling. *Transportation Research Part B*, 43, 422-437.

Ben-Akiva, M.E., Bergman, M.J., Daly, A.J. and Ramaswamy, R. (1984). Modelling inter urban route choice behavior." Proceeding of the *Ninth International Symposium on Transportation and Traffic Theory*, VNU Science Press, Delft, Netherlands, pp. 299-330.

Ben-Akiva, M.E. and 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. and Bovy, P.H.L. (2008). Impact of route choice set on route choice probabilities. *Transportation Research Record*, 2076, 10-19.

Bovy, P.H.L. and Bradley, M.A. (1985). Route choice analyzed with stated preference approaches. *Transportation Research Record*, 1037, 11-20.

Bovy, P.H.L. and Fiorenzo-Catalano, S. (2007). Stochastic route choice set generation: behavioral and probabilistic foundations. *Transportmetric*a, 3, 173-189.

Broach, J., Gliebe, J.G. and Dill, J.L. (2010). Calibrated labeling method for generating bicyclist route choice sets incorporating unbiased attribute variation. *Transportation Research Record*, 2197, 89-97.

Broach, J., Gliebe, J.G. and Dill, J.L. (2011). Bicycle route choice model developed using revealed preference GPS data. Proceedings of the *90th Annual Meeting of the Transportation Research Board*, Washington, D.C.

Broach, J., Gliebe, J.G. and Dill, J.L. (2012). Where do cyclists ride? A route choice model developed with revealed preference GPS data. *Transportation Research Part A*, 46, 1730-1740.

Cascetta, E., Nuzzolo, A., Russo, F. and 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*, Lyon, France, pp. 697–711.

de la Barra, T., Perez, B. and Anez, J. (1993). Multidimensional path search and assignment. Proceedings of the *21st PTRC Summer Meeting*, Manchester, UK, pp. 307-319.

Dijkstra, E.W. (1959). A note on two problems in connexion with graphs. *Numerische Mathematik*, 1, 269-271.

Hood, J., Sall, E. and 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. and 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. and Wardman, M. (1996). Evaluating the demand for new cycle facilities. *Transport Policy*, 3, 241-249.

Howard, C. and 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. and Abraham, J.E. (2007). Influences on bicycle use. *Transportation*, 34, 453-470.

Hyodo, T., Suzuki, N. and Takahashi, K. (2000). Modeling of Bicycle route and destination choice behavior for bicycle road network plan. *Transportation Research Record*, 1705, 70-76.

Krizek, K.J. (2006). Two approaches to valuing some of bicycle facilities' presumed benefits. *Journal of the American Planning Association*, 72, 309-320.

Marchal, F., Hackney, J. and Axhausen, K.W. (2005). Efficient map matching of large Global Positioning System data sets: tests on speed-monitoring experiment in Zürich. *Transportation Research Record*, 1935, 93-100.

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

Nielsen, O.A. (2000). A stochastic transit assignment model considering differences in passengers utility functions. *Transportation Research Part B*, 34, 377-402.

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

Prato, C.G. and Bekhor, S. (2007). Modeling route choice behavior: how relevant is the composition of choice set? *Transportation Research Record*, 2003, 64-73.

Ramming, S. (2001). *Network Knowledge and Route Choice*. PhD thesis. Massachusetts Institute of Technology, Cambridge, MA.

Rasmussen T.K., Ingvardson, J.B., Halldórsdóttir, K. and Nielsen, O.A. (2013). Enhancing the use of Global Positioning Systems data from wearable GPS devices as travel survey method: a case study in the Copenhagen Region. Presented at the *2nd Symposium of the European Association for Research in Transportation*, Stockholm, Sweden.

Rieser-Schüssler, N., Balmer, M. and Axhausen, K.W. (2012). Route choice sets for very high-resolution data. *Transportmetrica*, 9, 825-845.

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

Schüssler, N. and Axhausen, K.W. (2009a). *Map-matching of GPS traces on high-resolution navigation networks using the Multiple Hypothesis Technique (MHT)*. Working Paper, 568, IVT, ETH Zurich, Zurich.

Schüssler, N. and Axhausen, K.W. (2009b). Processing GPS raw data without additional information. *Transportation Research Record*, 2105, pp. 28-36.

Sener, I.N., Eluru, N. and Bhat, C.R. (2009). An analysis of bicycle route choice preferences in Texas, US. *Transportation*, 36, 511–539.

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

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

Stinson, M.A. and Bhat, C.R. (2005). A comparison of the route preferences of experienced and inexperienced bicycle commuters. Proceedings of the *84th Annual Meeting of the Transportation Research Board*, Washington, D.C.

Stopher, P.R. (2009). Collecting and processing data from mobile technologies. In Bonnel, P., Zmud, J., Lee-Gosselin, M.E.H. and Madre, J.L. (Eds.), *Transport Survey Methods: Keeping Up with a Changing World*, Emerald, Bingley, pp. 361–391.

Tilahun, N.Y., Levinson, D.M. and Krizek, K.J. (2007). Trails, lanes, or traffic: valuing bicycle facilities with an adaptive stated preference survey. *Transportation Research Part A*, 41, 287-301.

Tsui, S.Y.A. and 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.

Chapter 11

LAND-USE AND NETWORK EFFECTS ON BICYCLE ROUTE CHOICE IN THE GREATER COPENHAGEN AREA

Title:	Land-use and network effects on bicycle route choice in the Greater Copenhagen area
Author(s):	Katrín Halldórsdóttir, Otto A. Nielsen, and Carlo G. Prato
Presented:	*The 2nd Symposium of the European Association for Research in Transportation (hEART), 4th - 6th September 2013, Stockholm, Sweden*
Status:	Working paper
Abbreviated:	Halldórsdóttir et al. (2015b)

A preliminary version of this paper was presented at the 2nd Symposium of the European Association for Research in Transportation (hEART), in September 2013, Stockholm, Sweden. This working paper has been updated with improved model estimates and revised.

ABSTRACT

This study presents a bicycle route choice model that was estimated from a large sample of GPS observations, which revealed cyclist route choice preferences. Choice alternatives were generated for the model estimation using a doubly stochastic generation function, accounting for variation in travellers' link cost perception as well as differences among their preferences. A path-size logit model was estimated, capturing taste variations across cyclists. The results showed that cyclists are sensitive to the effects of distance, cycling the wrong way, turn frequency, hilliness, different bicycle facility types, bicycle bridges, pavement conditions, intersection type, the number of motorised traffic lanes, and crossing water/sea on motorised traffic bridges. The cyclists also valued different land-use conditions relatively highly. The results also showed that personal characteristics influence the route choice, that there were differences in route choice preferences depending on the time of day and whether it was a weekday or weekend, and also in different weather conditions.

Keywords: *Bicycle; Route choice; Land-use; Transport network effect; Global positioning system (GPS); Doubly stochastic generation function; Discrete choice model.*

11.1 Introduction

Over the years, non-motorised travel alternatives have been largely underrepresented in most transport models. Growing interest in sustainable transportation systems has increased the focus on policies and investments that encourage more cycling. However, such investments are often made without being able to carry out a quantitative forecast of the impacts of the investment – as opposed to, e.g., road investments. The study contributes to the field of cyclists' route choice modelling by using a large sample of GPS-observed routes and relating this to potential transport network and land-use factors that might influence cyclists' route choices.

Several recent studies focus on bicycle route choice as outlined in the review in Sener et al. (2009). Most studies have been based on stated preference (SP) data (e.g., Bovy and Bradley, 1985; Axhausen and Smith, 1986; Hopkinson and Wardman, 1996; Stinson and Bhat, 2003; Krizek, 2006; Hunt and Abraham, 2007; Tilahun et al., 2007; Sener et al., 2009), and a few on revealed preference (RP) data (e.g., Aultman-Hall et al., 1997; Shafizadeh and Niemeier, 1997; Hyodo et al., 2000; Howard and Burns, 2001; Ehrgott et al., 2012; Larsen et al., 2013; Snizek et al., 2013; Yeboah and Alvanides, 2015). Although SP data have the benefits of making a controlled experimental environment, they are not based on actual observed behaviour.

Cyclists' route choice models, based on revealed preference (RP) data, have had technical problems caused by the difficulty of observing routes. This has been overcome by the use of small transportable GPS-devices. Another challenge has been to generate plausible alternative routes for the model estimation.

There have been few studies on bicycle route choice models estimated from GPS observations. Menghini et al. (2010) estimated the route choice of cyclists in Zürich, Switzerland, and concluded that cyclists are mostly affected by the route length, with little effect of other factors, i.e., bicycle paths, maximum gradient, and traffic lights. The heterogeneity of the cyclists was captured through interaction terms, formulated from their average length and speed, confirming the strong preference for direct routes and showing that faster cyclists prefer marked routes.

Hood et al. (2011) showed that cyclists in San Francisco, USA, prefer bicycle lanes to other bicycle facility types, i.e., paths and routes. Especially infrequent cyclists appear to have a strong preference for this type of bicycle facility. The study also showed that cyclists have a negative preference for length, as well as turns, and they avoid cycling the wrong way down a one-way street. Average up-slopes were disfavoured, especially by women and when commuting, whereas traffic volume, traffic speed, number of lanes, crime rates, rain, and nightfall had no significant effect.

Broach et al. (2012) estimated the route choice of cyclists in the Portland metropolitan area, USA. The study confirms the main findings of the previous two studies, i.e., cyclists are sensitive to the effects of distance and increasing upwards slopes. In addition, in the Portland metropolitan area, cyclists prefer off-street bicycle paths, bicycle boulevards, and bridge facilities. The interaction between bicycle lanes and traffic volumes had significant effect (contradicting Hood et al. (2011) who concluded that traffic volume had no effect), showing increased dispreference for this facility type with increasing traffic. Cyclists are also sensitive to turns and intersection control (i.e., traffic signals and stop signs). Finally, the study concluded that commuting cyclists were more sensitive to distance and less sensitive to most other variables.

The route choice of cyclists in the Waterloo Region, Canada, was estimated in Casello and Usyukov (2014). The study estimated a limited number of variables, only showing the significance of length, bicycle lanes, and gradient, whereas motor traffic speed and traffic volume could not be incorporated in the model.

The level of bicycle use in Denmark is relatively high when compared to most other countries, particularly in the Copenhagen capital, where about 37% of the commuting trips are carried out on bicycles (Pucher and Buehler, 2012). None of the previous studies, on revealed preference models of cyclists' route choices estimated from GPS data, are constructed in established bicycle cities. The main objective of this study is thus to develop a route choice model that focuses on the route choice characteristics in the Greater Copenhagen area, thus providing inspiration for emerging cycling cities by focusing in particular on the interaction between infrastructure, land use, and cyclists' route choice. The current study expands the body of literature by analysing a rich set of *network attributes* describing distance, wrong way, turn frequency, elevation, bicycle facility type (i.e., segregated bicycle path, bicycle lane, bicycle path in own traces, footpath, and steps), bicycle bridges, surface type, intersection type (i.e., traffic lights, roundabouts, and give way), and accident patterns. The current study also includes network attributes describing *motorised* road type, speed, number of traffic lanes, bridge (crossing water/sea), and tunnel. Most relevantly, the current study also expands the body of literature by analysing different *land-use attributes* along the route, as previous studies have shown the importance in positive cycling experience (e.g., Snizek et al., 2013). It is investigated how *personal attributes* influence the route choice (i.e., gender, age, and cyclists' average speed profile), whether there are differences in route choice preference between *trip related attributes* (i.e., trip purpose, peak hours, weekends, and darkness), and also between *weather attributes* (i.e., temperature, rainfall, sunshine, and wind).

The remainder of this paper is structured as follows. Section 11.2 describes the data used in the current study. Section 11.3 describes the choice set generation methods used in this study and the applied modelling methodology. Section 11.4 presents the descriptive statistics of the data used and the results from the model estimation and discussion. Finally, conclusions are provided in section 11.5.

11.2 Data

11.2.1 Bicycle network database

The bicycle network database is built on the topographic network FOT-kort10 (FOT-Kort10, 2010) and TOP10DK (Kort & Matrikelstyrelsen, 2001). The two sources were compiled together in order to obtain a detailed bicycle network database, as illustrated in Figure 67. This is necessary, since cyclists are using paths that are not present in standard commercial digital maps made for GPS-based car navigation systems. The combination of sources created a complete geographic network of roads and paths used by cyclists, comprising of 361,053 directional links and 268,762 nodes for the study area. The network includes small paths only accessible by bicycles and pedestrians and does not include roads where cycling is illegal in Denmark, such as motorways and expressways.

Various sources were added to the bicycle network database with characteristics considered important for cyclists (i.e., Open Street Map (OSM), the LTM road network (Rich et al., 2010), NAVTEQ (NAVTEQ, 2010), where accidents from the last five years were added, a total of 87,455, from the crash database maintained by the Danish Road Directorate (Vejman), and intersection data (from the Danish Road Directorate)). For further information, see: Halldórsdóttir et al. (2013) and Pedersen and Senstius (2014).

A rich set of network attributes was analysed, i.e., distance, turn frequency, gradient, bicycle facility type, bicycle bridges, surface type, intersection type, and accident patterns. The conditions along the routes were also analysed, i.e., motorised traffic type, time-dependent traffic volumes, number of motorised traffic lanes, motorised traffic bridges and tunnels, motorised free speed, and land-use information.

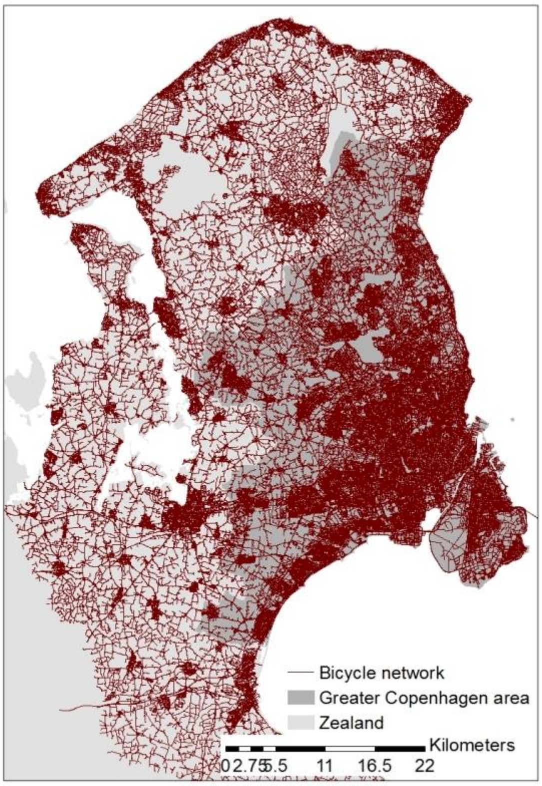


FIGURE 67: BICYCLE NETWORK DATABASE IN THE GREATER COPENHAGEN AREA. LEFT: OVERVIEW OF THE GREATER COPENHAGEN AREA NETWORK. RIGHT: ILLUSTRATION OF THE NETWORK DETAILS

11.2.2 GPS DATA

The data collection was carried out for the Greater Copenhagen area, in Denmark. A total of 318 cyclists were provided with GPS trackers for an average period of eight days. The individuals were sampled from the Danish National Travel Survey (Christiansen, 2009), abbreviated the TU-survey (in Danish Transportvaneundersøgelsen). The sampling criteria were that the respondents had previously been interviewed, approximately within the last 6 to 12 months before the data collection, that they had used a bicycle in their reported travel, and that they lived in the study area. In addition, because of strict privacy rules within the TU-survey, possible recruits were sampled if they were 16 years or older. In order to capture different seasons, three data collection rounds were run; the first round was carried out from October through December in 2012, the second round was carried out from June through July in 2013, and the last round was carried out from August through October the same year. The average acceptance rate was 65%, which is quite high for a survey of this type. In addition to collecting GPS data, travel diaries were collected from the participants, providing their travel information on a selected day, trip purpose in their reported trips, and socio-economic variables.

11.2.2.1 GPS POST-PROCESSING AND MAP-MATCHING

Extensive data processing was required to obtain data that could be used for further analysis on cycling specifically, since the GPS data was delivered as raw GPS-traces and collected all modes of transport. The post-processing of the data was carried out as described in Schüssler and Axhausen (2008, 2009). The method is divided into four steps: (i) *GPS data cleaning*, where systematic and random errors are removed from the data; (ii) *trip and activity identification*, where the GPS tracker had been stationary for a period of time and/or if the spatial density of observations has been high for a period of time; (iii) *trip segmentation into single-mode trip legs* (sub-components of the trip); and (iv) *mode identification*, where modified fuzzy logic rules were applied as described in Rasmussen et al. (2015). The post-processing method was validated by comparing the resulting trips and modes with travel diaries. In order to better fit the travel behaviour of the Danish population, the original fuzzy logic rules had to be altered and extended with more disaggregated components using GIS software (Rasmussen et al., 2015).

After the data post-processing, a total of 6,378,651 GPS points and 14,557 single-mode stages were identified. After the travel mode was identified as a bicycle, 2,681,108 GPS-points and 5,027 stages were retained for further analysis. The identified bicycle trips were mapped to a high-resolution network using the map-matching algorithm developed by Nielsen and Jørgensen (2004). Non-trips were identified and removed when either no GPS observations could be map-matched, such as scatter around activities wrongly classified as trip legs, or if there were large deviations between the calculated map-matched distance and the distances calculated from the processed GPS points. These large deviations were, for example, caused by random errors in the network, such as missing links. Rasmussen et al. (2015) showed that by removing such non-trips, the overall confidence rate improved from 69% to 91%. However, the improvement in confidence rate was at the cost of also removing a large number of actual trip legs. After the map-matching algorithm was run and non-trips were removed, there were 3,443 stages from 291 respondents remaining for further analysis.

11.2.2.2 PERSONAL CHARACTERISTICS

The GPS dataset was composed of 45% males and individuals between the ages of 17 and 79. Figure 68 shows the age distribution of the dataset. For more information on the characteristics of the data sample, see section 8.1.2.

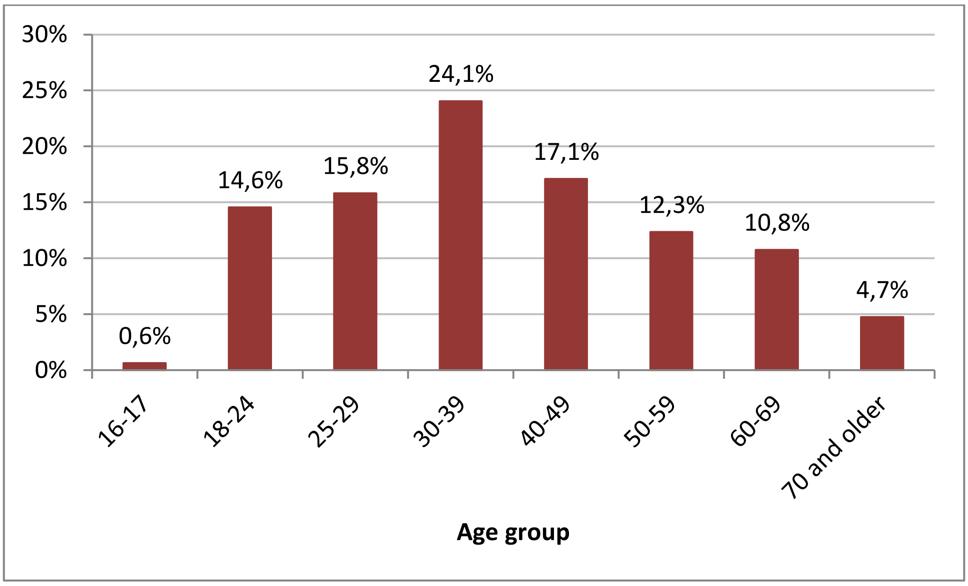


FIGURE 68: AGE GROUP DISTRIBUTION IN THE GPS DATASET

In addition to considering age and gender, participants were also divided into three groups of cyclists, depending on their average speed profiles; slow (<10 km/h), medium (10-14 km/h), and fast (>14 km/h). There were 91 trips taken by slow cyclists, 2,633 by medium cyclists, and 719 by fast cyclists.

11.2.2.3 TRIP INFORMATION

Commuting trips were extracted through GIS analysis, described in section 8.1.3.1. The method proved ineffective in identifying commuting trips as there were only 281 commuting trips out of 3,443 trips. This could be because travellers often combine other trip purposes with their commuting trips, such as dropping off/picking up children or buying groceries. In order to identify possible commuting trips, weekday trips conducted within the morning peak hours, between 7 and 9, and the afternoon peak hours, between 15 and 17, were classified as peak hours trips. A total of 1,094 trips were peak hours trips, while all others were non-peak hours, 2,349 trips in total.

Trips conducted during weekends were identified in order to investigate differences in route preference between weekends and weekdays. The total number of weekend trips was 669, while there were 2,774 weekday trips. In addition, information on sunrise and sundown in the Greater Copenhagen area was also collected and compared to the time of day each stage was conducted. There were 491 trips after sundown, while there were 2,952 trips during daylight.

Finally, weather information, from the Danish Meteorological Institute (DMI), was joined with the bicycle trips in order to analyse whether there were differences in route preference between different weather attributes (i.e., rainfall, temperature, sunshine, and wind). For more information on how the additional information was collected for the survey, see section 8.1.3.

11.3 MODEL DEVELOPMENT

Route choice models are essential to identify which network attributes influence cyclists' route choice and evaluate the trade-offs among these attributes. Section 11.3.1 presents the method used to generate the alternative routes, while section 11.3.2 presents the model used for estimation.

11.3.1 CHOICE SET GENERATION

Bicycle route choice models depend on the observation of actual route choice and the generation of plausible alternatives. Model estimation results are strongly influenced by the size and composition of the choice set (Bekhor et al., 2006; Prato and Bekhor, 2007; Bliemer and Bovy, 2008; Anderson et al., 2014; Rasmussen et al., 2014). An incorrectly specified choice set can lead to biased parameter estimates and choice probabilities, especially when accounting for correlation between alternatives (Bliemer and Bovy, 2008). Preferably, the choice set should only include relevant alternatives. Including all paths in the network is unrealistic.

For this study, three path generation methods were tested, i.e., breadth first search on link elimination (BSF-LE) (Rieser-Schüssler et al., 2012), doubly stochastic generation function (DSGF) (Nielsen, 2000; Bovy and Fiorenzo-Catalano, 2007), and the branch & bound (B&B) (Hoogendoorn-Lanser et al., 2006; Prato and Bekhor, 2006) algorithm, in their ability to generate relevant and heterogeneous bicycle routes (for further information see Halldórsdóttir et al., 2013). Different multi-attribute cost functions were tested, where not only route length or time is considered, but also bicycle-oriented factors, such as preferences for road types, dedicated cycle paths, and land use. The study showed that extending the cost function to factors that are relevant to cyclists provided a significant increase in performances when compared to previous choice set generation methods in the bicycle context.

In the current study, the alternative routes were generated using DSGF. The method accounts for variations in travellers' link cost and differences in travellers' attribute preferences by drawing random costs and random parameters from probability distributions. The advantages of this method are the heterogeneity of the generated alternatives and its computational efficiency in large networks, even though the randomisation of link costs and the parameters can be time consuming in a high-resolution network. The route set alternatives were generated using a modified version of the *Road Traffic Assignment* (RTA) tool in the *Traffic Analyst* (TA) software developed by Rapidis (www.rapidis.com), which is a transportation planning and modelling extension for ArcGIS Desktop (ESRI, 2015).

11.3.1.1 Generation (cost) function

In this study, besides considering information regarding route length and travel time, the cost function considers information regarding different bicycle path types (i.e., motorised roads without any bicycle infrastructure, motorised roads with bicycle paths, motorised roads with bicycle lanes, dedicated bicycle path, dedicated footpath, and steps), surface types (i.e., paved or not paved) and land-use information (i.e., scenic paths, including forests and parks, or non-scenic paths). Travelling the wrong way was also given a penalty factor, as cycling the wrong way on an one-way street is illegal in Denmark. Nonetheless, many cyclists do so in order to avoid long detours. The generation (cost) function was defined as:

$$
\begin{aligned}
\overline{C_a} = & \sum_k \left(\left(\beta_{PathType_k} + \xi_{PathType_{ak}} \right) \cdot PathType_{ak} \cdot Length_a \right) \\
& + \sum_k \left(\left(\beta_{SurfaceType_k} + \xi_{SurfaceType_{ak}} \right) \cdot SurfaceType_{ak} \cdot Length_a \right) \\
& + \sum_k \left(\left(\beta_{LandUse_k} + \xi_{LandUse_{ak}} \right) \cdot LandUse_{ak} \cdot Length_a \right) \\
& + \sum_k \left(\left(\beta_{WrongWay_k} + \xi_{WrongWay_{ak}} \right) \cdot WrongWay_{ak} \cdot Length_a \right) \\
& + \left(\beta_{TravelTime} + \xi_{TravelTime_a} \right) \cdot TravelTime_a + \left(\beta_{\mathrm{Length}} + \xi_{\mathrm{Length}_a} \right) \cdot Length_a
\end{aligned}
\tag{26}
$$

where: $\overline{C_a}$ is the random cost of link *a*; $\beta_{RoadType_k}$ are coefficients related to road type k with error components $\xi_{RoadType_{ak}}$ related to road type k of link *a*; $\beta_{SurfaceType_k}$ are coefficients related to surface type k with error components $\xi_{SurfaceType_{ak}}$ related to surface type k of link *a;* $\beta_{LandUse_k}$ are coefficients related to land-use k with error components $\xi_{RoadType_{ak}}$ related to land-use type k of link *a*; $\beta_{WrongWay_k}$ are coefficients related to wrong way type k with error components $\xi_{WrongWay_{ak}}$ related to wrong way type k of link *a*; $\beta_{TravelTime}$ is a coefficient related to travel time with error component $\xi_{TravelTime_a}$ related to the travel time on link *a*; β_{Length} is a coefficient related to length with error component ξ_{Length_a} related to the length of link *a*.

In order to guarantee non-negative draws, the betas β_i are distributed according to a log-normal distribution:

$$
\beta_i \sim \ln\left(Variable, \beta_i \cdot Variable \right). \tag{27}
$$

The error terms are distributed according to a gamma distribution:

$$C_a \sim \Gamma\left(\frac{\overline{C_a}}{\gamma}, \gamma\right), \tag{28}$$

where $\overline{C_a}$ is the generalized cost calculated in equation (26) and γ is the scale parameter (Nielsen and Jovicic, 1999).

There are two cost types that are predefined in the RTA tool, i.e., length and travel time (the average speed was set at 15 km/hour), and then the error term. Various test runs were conducted in order to select suitable parameters for the generation (cost) function. After each run, the stochasticity of the generated alternatives was evaluated graphically, as well as the maximum coverage over the data sample. The parameters were chosen so as to provide a good amount of variation between the different routes, without them being unrealistic. Table 35 lists the predefined cost parameters, plus the final scale parameters that were used to generate the alternative choice set for the bicycle routes.

TABLE 35: PREDEFINED COST PARAMETERS, PLUS THE ADDED SCALE PARAMETERS

ID	Parameter	Distribution	Mean	Variance
Predefined cost parameters				
β_{Length}	Length	No distribution	1	0
$\beta_{FreeTime}$	Free time	Log-normal	1	0.25
ε	Error term	Gamma	0	2
Scale parameters				
$\beta_{RoadType1}$	Road with no bicycle infrastructure	Log-normal	1.25	1.5625
$\beta_{RoadType2}$	Road with bicycle lane	Log-normal	0.75	0.5625
$\beta_{RoadType3}$	Road with segregated bicycle path	Log-normal	0.5	0.25
$\beta_{RoadType4}$	Bicycle path	Log-normal	0.5	0.25
$\beta_{RoadType5}$	Footpath	Log-normal	1.5	2.25
$\beta_{RoadType6}$	Steps	Log-normal	1.5	2.25
$\beta_{SurfaceType1}$	Surface type – Paved	Log-normal	0.75	0.5625
$\beta_{SurfaceType2}$	Surface type – Not paved	Log-normal	1.25	1.5625
$\beta_{LandUse1}$	Land use – Scenic routes	Log-normal	0.5	0.25
$\beta_{LandUse2}$	Land use – All other routes	Log-normal	1.5	2.25
$\beta_{WrongWay}$	Wrong way	Log-normal	1.5	2.25

Observations over the 80% overlap threshold were considered consistent with the observed behaviour, a value often used in the literature (e.g., Ramming, 2002; Prato and Bekhor, 2007). The DSGF method duplicated 80% of the observed routes with a 80% coverage threshold. The method produced on average 65.60 alternative routes for each bicycle trip, with a standard deviation of 40.19. The number of alternatives varied across choice situations, ranging from 1 to 100 alternatives, and directly attributed to increasing trip distance. Zero alternatives were generated for 59 out of 3,443 trips. These trips were, in most cases, extremely short, or the network density was limited. These trips were not included in the estimated dataset, along with one respondent that was missing travel survey information.

11.3.2 MODEL SPECIFICATION

The Multinomial Logit (MNL) and the Nested Logit (NL) are the most common models used when modelling discrete choices in travel behaviour. However, the MNL model does not account for similarities among alternatives, which is very important when modelling route choice since alternative routes often overlap. Also, routes share links with hundreds of other routes in real-size networks, which makes the NL model not applicable as it assumes that each alternative route belongs only to one nest. Path-size logit (PSL) models, presented by Ben-Akiva and Bierlaire (1999), account for similarities among alternative routes. In this study, PSL models were estimated and performances for different route choice specifications were examined:

$$P_k = \frac{\exp\left(V_k + \beta_{PS} \cdot \ln PS_k\right)}{\sum_{l \in C} \exp\left(V_l + \beta_{PS} \cdot \ln PS_l\right)}, \tag{29}$$

where C is the choice set. The probability of choosing the specific route k is P_k, and V_k and V_l are the utility functions of routes k and l, respectively. Furthermore, β_{PS} is the parameter to be estimated while the PS_k and PS_l are the path-size factors of route k and l, respectively. The path-size specifies the path fraction that constitutes a "full" alternative. Accordingly, a unique path has a size equal to 1 and identical paths N share the size $1/N$.

In order to capture the similarities between the alternative routes, the path-size factor was calculated as:

$$PS_k = \sum_{a \in \Gamma_k} \frac{L_a}{L_k} \frac{1}{\sum_{l \in C} \delta_{al}}, \tag{30}$$

where PS_k is the path-size factor, Γ_k is the set of all links of path k, L_a is the length of link a, L_k is the length of path k, and δ_{al} equals 1 if link a is on path i and equals 0 otherwise. The path-size ranges between 0 and 1, with 0 indicating complete overlap and 1 indicating complete independence.

11.4 RESULTS

11.4.1 DESCRIPTIVE STATISTICS

Table 36 presents a comparison of the means and standard deviations of the network attributes and the land-use variables for the chosen and the non-chosen routes, respectively. In comparison to the alternatives, the chosen routes are on average shorter. As a result, all length related attributes are shorter, to a varying degree. The chosen routes have less turns, are not as steep, and have fewer intersections. Bicycle paths and footpaths in own traces are noticeably shorter for the chosen routes. This is because very few choose these alternatives and therefore it is to be expected that there will be very strong dispreference in the model estimation results. The land-use variables are the total distance travelled along a specific land-use type. The variables are divided into three categories, i.e., if the land-use category is on both sides of the link or if the information is either on the right side of the link or on the left side. In central Copenhagen, and other highly populated areas, it is difficult to differentiate between high residential areas and town centres, as high buildings are often contained in both. Thus, these two variables were joined together.

TABLE 36: VARIABLE DESCRIPTION OF THE NETWORK- AND LAND-USE ATTRIBUTES FOR THE CHOSEN AND NON-CHOSEN ROUTES

		Chosen route		Alternatives	
Variable	**Unit**	**Mean**	**St. dev**	**Mean**	**St. dev**
Network attributes					
ln(Path-size)	-	-2.043	1.348	-2.900	0.882
Length	[km]	3.759	4.659	5.374	5.496
Wrong way	[km]	0.070	0.261	0.092	0.252
Number of turns					
Left turns	-	2.993	4.053	6.436	6.427
Right turns	-	2.711	2.953	4.901	3.914
Cumulative elevation gain					
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
Bicycle facility type					
Motorized road without any bicycle facilities	[km]	0.927	1.363	1.326	1.722
Motorized road with bicycle lane	[km]	0.150	0.374	0.209	0.450
Motorized 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 bridge	[km]	0.002	0.025	0.005	0.038
Surface type					
Paved	[km]	3.658	4.540	4.921	5.007
Cobblestone	[km]	0.006	0.077	0.037	0.217
Unpaved	[km]	0.090	0.439	0.400	1.157
Only MTB	[km]	0.004	0.067	0.016	0.127
Number of intersections					
Give way	-	0.617	1.743	0.756	1.908
Roundabout	-	0.518	2.093	0.741	2.480
Traffic signal	-	7.913	11.794	9.629	11.753
Motorized road type					
Large motorized roads	[km]	1.847	3.390	2.027	3.142
Medium motorized roads	[km]	0.603	1.198	0.686	1.125
Large local roads	[km]	0.006	0.091	0.008	0.086
Small local roads	[km]	1.266	1.745	2.604	3.294
Traffic calmed roads	[km]	0.037	0.187	0.050	0.215
Motorized free speed					
Below 11 km/hour	[km]	0.308	0.866	0.870	1.889
11-30 km/hour	[km]	0.293	0.634	0.641	1.040
31-50 km/hour	[km]	2.478	3.021	3.030	3.047
51-70 km/hour	[km]	0.613	1.729	0.736	1.654
71-90 km/hour	[km]	0.067	0.769	0.096	0.695
91-100 km/hour	[km]	0.000	0.000	0.001	0.029
Above 101 km/hour	[km]	0.000	0.000	0.000	0.016

Number of motorized 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
Motorized traffic bridge, crossing water/sea	[km]	0.015	0.068	0.030	0.099
Motorized traffic tunnel	[km]	0.000	0.010	0.000	0.011
Land-use influence					
Cemetery on the right side	[km]	0.050	0.194	0.052	0.206
Cemetery on the left side	[km]	0.047	0.197	0.047	0.190
Cemetery on both sides	[km]	0.011	0.105	0.024	0.169
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
High residential area/town centre on the right side	[km]	0.497	0.804	0.722	1.023
High residential area/town centre on the left side	[km]	0.496	0.809	0.723	1.025
High residential area/town centre on both sides	[km]	1.413	1.927	1.726	1.866
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
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
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
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
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
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

11.4.2 MODEL ESTIMATION AND DISCUSSION

To investigate the attributes that are relevant in bicycle route choice, various PSL models were estimated. The asymptotic t-test was primarily used to test whether a specific model parameter differed from a known constant. Not all estimated parameters proved to be statistically significant at the 90% level. In addition, some variables, which were considered interesting preference indicators, could not be included in the estimation, either because they were correlated with other more important variables or because of identification problems. Consequently, some variables were removed to increase the reliability of the model. The model was estimated using the Biogeme package (Bierlaire, 2008).

The results from the model estimation are reported in Table 37. The estimated values are presented along with the rates of substitution, which are scaled to the length. The rates of substitution specify the degree to which the average cyclist prefers one type over the other compared to the length. The final model is constituted by 3,363 observations and 26 estimated parameters.

TABLE 37: THE RESULTS FROM THE BASIS PATH-SIZE LOGIT MODEL ESTIMATES

		Model estimates			Rates of
Parameter	**Unit**	**Value**	**t-test**		**substitution**
Network attributes					
ln(Path-size)	-	1.36	20.25	***	-
Length	[km]	-0.91	-2.79	***	-1
Wrong way	[km]	-1.19	-7.36	***	-1.31
Number of turns					
Left	-	-0.32	-19.71	***	-0.36
Right	-	-0.20	-12.11	***	-0.22
Cumulative elevation gain					
0-10 meters/km	*ref.*				
10-35 meters/km	[km]	-33.30	-3.72	***	-36.63
35-50 meters/km	[km]	-37.60	-2.49	***	-41.36
Above 50 meters/km	[km]	-49.40	-5.60	***	-54.35
Bicycle facility type					
Motorised road without any bicycle facilities	*ref.*				
Motorised road with segregated bicycle path/ lane	[km]	0.11	2.98	***	0.12
Bicycle path in own trace	[km]	-0.19	-2.51	***	-0.21
Footpath in own trace	[km]	-1.92	-9.22	***	-2.11
Steps	[km]	-12.70	-4.33	***	-13.97
Bicycle bridge	[km]	2.09	1.60		2.30
Surface type					
Paved	*ref.*				
Not paved	[km]	-0.22	-2.13	**	-0.24
Number of intersections - Roundabout	-	0.05	3.14	***	0.05

Number of motorised traffic lanes					
1 lane	[km]	1.27	3.00	***	1.40
2 lanes	*ref.*				
3 to 4 lanes	[km]	-0.22	-3.77	***	-0.24
5 lanes and above	[km]	-0.42	-5.77	***	-0.46
Motorised traffic bridge, crossing water/sea	[km]	-1.34	-2.53	***	-1.47
Land-use influence					
Low residential area on the right side	*ref.*				
Low residential area on both sides	*ref.*				
High residential area and/or town centre on one side	[km]	-0.22	-3.49	***	-0.24
High residential area and/or town centre on both sides	[km]	-0.60	-9.56	***	-0.66
Industry on both sides	[km]	-0.30	-2.65	***	-0.33
Sport on one side	[km]	-0.56	-4.37	***	-0.61
Sport on both sides	[km]	0.86	3.94	***	0.94
Park on one side	[km]	-0.11	-1.98	**	-0.12
Park on both sides	[km]	0.16	1.52		0.18
Number of estimated parameters:					26
Number of observations:					3,363
Null log-likelihood:					-12,761.08
Final log-likelihood:					-9,687.62
Adjusted rho-square:					0.239

Note: * significant at the 90% level; ** significant at the 95% level; *** significant at the 99% level.

11.4.2.1 NETWORK ATTRIBUTES

The logarithm of the path-size variable is statistically significant and positive, as expected, and thus correctly accounts for route overlap. Consistent with the findings in the literature (e.g., Menghini et al., 2010; Hood et al., 2011; Broach et al., 2012) the results show that cyclists prefer shorter routes and avoid cycling the wrong way down a street (e.g., Hood et al., 2011), unless they, e.g., avoid a large detour, as observed from some of the chosen routes. Turn frequency has a significant negative effect on cyclists' route choice and the model estimates show that cyclists prefer straight routes, and dislike left turns over right turns, also shown in Hood et al. (2011) and Broach et al. (2012).

Many variations of elevation change were tested, i.e., maximum and average slope, and cumulative gain and loss in elevation. The best performing specification was the sum of elevation gain on all subparts of a link categorised with ‰. More specifically, it consists of the length with gain per thousand in the range of 0 to 10 meters (i.e., with slope up to 1%), 10 to 35 meters, 35 to 50 meters, or above 50 meters. The estimation results show that cyclists' route choice is strongly affected by increasing elevation and large in comparison with the impact of the length, indicating that average cyclists are willing to take a detour around the hillside if possible. Hood et al. (2011) and Broach et al. (2012) also found a strong dispreference for steep slopes, while Menghini et al. (2010) only found a small effect.

The parameters describing segregated bicycle paths and bicycle lanes were joined together, as the estimated coefficient parameters were not statistically different from each other. On average, cyclists prefer segregated bicycle paths and bicycle lanes and they are willing to take a detour in order to cycle on these facility types. The results show that, on average, cyclists are not willing to take a detour to travel on bicycle paths in their own trace. In Denmark, it is illegal to cycle on footpaths, which is confirmed by the model estimate as cyclists perceived footpaths more than double more burdensome in ratio to the length. Not surprisingly, there is also a high resistance to routes that have steps, as it can be very inconvenient to step off the bicycle and carry it up or down the staircases, although most staircases in Copenhagen have bicycle rails to assist. This pronounced dispreference is also because the dataset contains very few observed choices for these types, while there are numerous alternative routes. The literature shows that cyclists prefer increased separation from motorised traffic (e.g., Menghini et al., 2010; Hood et al., 2011; Broach et al., 2012) as confirmed in this study. It should be noted that none of the comparable literature investigates bicycle paths in their own trace, footpaths, or steps. The coefficient estimates show that cyclists have a high preference to dedicated bicycle bridges, indicating that cyclists might change their route choices if such facilities are built.

The estimated coefficients for the unpaved surface types (i.e., cobblestone, unpaved, and mountain bicycle paths) were not statistically significant from each other and were thus joined together, showing that cyclists prefer to cycle on paved surfaces.

The model estimates for the number of traffic signals was not statistically significant, thus contradicting Menghini et al. (2010) and Broach et al. (2012) that found that cyclists are sensitive to this type of intersection control. The number of stop signs did not have a significant effect on cyclists' route choices. The number of roundabouts crossed was the only significant intersection parameter. It entered the model with a positive sign, which indicates that cyclists prefer to travel in a roundabout. This is probably because in Denmark bicycles have the right of way to motorised traffic in roundabouts and they do not need to stop when crossing it, opposite to signalised intersections or left turning in any intersection. Various different motorised conditions were tested (i.e., traffic type, speed limit, annual average daily traffic (AADT), and time dependent traffic volumes), however, these parameters did not have statistically significant model estimates. In Denmark, it is recommended to increase separation from mixed traffic, first to bicycle lanes and then to segregated bicycle paths, with increasing traffic volume and speed. Accordingly, segregated bicycle paths and bicycle lanes are highly correlated with motorised speed and traffic volumes, which could explain why these parameters were not statistically significant. These findings correspond with the findings of Hood et al. (2011), while they contradict Broach et al. (2012). The parameters describing the number of traffic lanes were the only parameters estimated statistically significant, showing that cyclists prefer roads with one traffic lane, all else being equal, and that cyclists' disutility towards traffic lanes increases, as the number of traffic lanes increases. The coefficient for motorised traffic bridges, crossing water/sea, was estimated negative and the rates of substitution show that cyclists consider traffic bridges, over water/sea, more than 40% more burdensome to travel on in relation to length.

11.4.2.2 Land-use attributes

The estimated coefficients for the land-use information on the right side of the link and on the left side were not statistically different from each other and were thus joined together. The estimated parameters for high residential area and/or town centre on one side and both sides were negative and statistically significant. This could be an indication that cyclists are sensitive to high congestion levels on bicycle paths, as the bicycle paths are often highly congested in these areas. The negative effect could also be explained by the availability of car parking along these roads (e.g., Sener et al., 2009) and pedestrians crossing the bicycle paths to access shops, therefore these areas become more cumbersome or risky for bicycles. As expected, the parameter for industrial area on both sides was estimated negative and statistically significant.

The model estimates indicate that cyclists are willing to take detours to cycle in recreational areas or parks, when they are on both sides of the path, while they avoid such detours when these areas are only on one side of the path. Previous model estimates showed that the parameter describing paths along a scenic area was estimated positive. However, it was not statistically significant and was thus not included in the final model. The parameter for paths in forest areas were estimated negative but not statistically significant and was thus not included. This is because forests generally have gravel paths, and accordingly were highly correlated with the parameter describing the surface type.

11.4.2.3 Additional model estimates

It was investigated how personal- and trip attributes influence the route choice of cyclists, where each category was estimated separately through linear combinations. The final results are presented in Appendix B with the rates of substitution, scaled to the length.

11.4.2.3.1 Personal attributes

The model estimates in Appendix B show that on average, females have a higher negative preference for increased elevation than males. Hood et al. (2011) also found that females are more sensitive to average up-slopes. Males prefer not to cycle on paths in their own trace. The results also show that females prefer to cycle in park areas, while males prefer to avoid such detours. Interactions with other socio-economic variables were also tested, e.g., age and bicycle path type, but none of them were significant.

The cyclists' average speed profile appears to influence their route choice (also concluded by Menghini et al., 2010), as faster cyclists have a negative preference for cycling on bicycle paths in their own trace, all else being equal, and a positive preference to cycle in parks. The results show that slow and medium fast cyclists have a negative preference to cycle in park areas.

11.4.2.3.2 TRIP ATTRIBUTES

Trip attributes' influence on the route choice of cyclists was investigated as well, through linear combinations. The results are presented in Appendix B with the rates of substitution, scaled to the length.

Cyclists have a higher dispreference to cycling on a bicycle path in own their trace and in parks after sundown. This is not surprising, as these paths are often isolated and poorly lit. The results also indicate that cyclists prefer not to cycle alongside parks during daytime, but instead prefer to cycle through parks.

During peak hours, cyclists have a higher dispreference for bicycle paths in their own trace than during off-peak hours. They also avoid cycling alongside parks during peak hours and prefer to cycle in parks during off-peak hours. During weekends, there is a higher preference for segregated bicycle paths and bicycle lanes. Cyclists also appear to avoid bicycle paths in own their trace on weekdays. The results show that cyclists prefer routes through parks on weekdays, when they are not as crowded with pedestrians. During weekends, cyclists avoid parks as there are more pedestrians, which can be troublesome for cyclists as they need to cycle quite slowly and be more alert and swerve around people. In addition, it is not very popular amongst pedestrians to share paths with cyclists as they find it uncomfortable when a bicycle passes. Most importantly, it is not allowed to cycle through some parks, e.g., in the Copenhagen Municipality. Nonetheless, some cyclists ignore these restrictions.

Several interactions with the weather variables were tested. As there are not many observations during rain, only 289 out of 3,363 in total, there was no measurable difference in route choice preferences in this weather condition, confirming Hood et al. (2011). There are considerably more observations during sunshine, a total of 1740 trips. However, the main difference is that cyclists have a higher dispreference for bicycle paths in their own trace when there is no sunshine.

Trips were divided into three groups, depending on the air temperature: low when the air temperature was below 5°C, medium when between 5°C and 15°C, and high when above 15°C. In previous model estimates, the medium- and high air temperatures were not statistically significant from each other and were thus joined together. The results show that cyclists' route choices are more affected by the air temperature than other weather conditions. There seems to be a correlation between low temperatures and avoiding bicycle paths in their own trace and parks on both sides. These paths are often cleaned later when the weather is snowy or freezing and thus these paths are more unattractive than roads in the wintertime. Several interactions with wind speed were also tested. However, none of them could be incorporated into the final model with statistically significant coefficients.

11.5 CONCLUSION

In this study it was examined which factors relate to the route choice of cyclists in the Greater Copenhagen area. A large sample of GPS observations was estimated, comprised of 3,363 bicycle trips total, with related socio-economic attributes of the cyclists. Comparing different types of cyclists and evaluating their routes provided a deeper understanding of what affects cyclists' route choices. In this study, a PSL model was estimated, accounting for similarities between the alternative routes. The model had a good explanatory power, with the logarithm of the path-size variable significant and positive, thus correctly accounts for route overlap.

The findings of this study report on the sensitivity of cyclists to the effects of distance, cycling the wrong way, turn frequency, and hilliness. When thinking about new bicycle infrastructure, the findings suggest that cyclists are sensitive to the effects of different bicycle facility types and emphasise the importance of well-built bicycle facilities, i.e., segregated bicycle path and bicycle lanes, as well as bicycle bridges and that cyclists prefer paved surface types.

In the Greater Copenhagen area, cyclists' favour having the right of way to motorised traffic at intersections, as the findings show that cyclists' prefer roundabouts over other intersection types. The findings also show that cyclists' are sensitive to the number of motorised traffic lanes and crossing water/sea on motorised traffic bridges, whereas motorised traffic type, speed limit, annual average daily traffic (AADT), and time dependent traffic volumes had no statistically significant effect. Accident patterns also had no significant effect on cyclists' route choices.

Most importantly, the findings show that cyclists appear to place relatively high value on different land-use conditions along the routes, i.e., prefer to avoid high residential areas and/or town centres and industrial areas, whereas they are willing to take detours to cycle in recreational areas or parks when they are on both sides of the path. Cyclists avoid these detours when recreational- or park areas are on one side of the paths. Paths along scenic areas and in forests did not have a significant effect on cyclists' route choices.

The findings of this study reveal, through linear combinations, that there is some heterogeneity among the cyclists in relation to different route preferences (i.e., gender and type of cyclist) and that there are differences in route choice preferences depending on the time of day (i.e., peak hours and darkness) and whether it is a weekday or weekend. The findings also show that cyclists route choices are more affected by the air temperature than rain or sunshine.

This study shows that by focusing on the interaction between infrastructure and route choice of cyclists, it is possible to contribute to the understanding of which factors influence cyclists' route choices. The findings help decision-makers to improve strategies in policy making and prioritise new infrastructure plans, aimed to improve cycling conditions. In addition, the findings can be used to improve travel demand models, which are mostly based on shortest-paths calculations, and thus enable traffic planners to predict bicycle travel more accurately and to forecast future travel behaviour.

Avenues for future research are needed. Firstly, the method used to identify commuting trips from raw GPS data proved not to be very successful. Further research is needed in this area, as investigating route choice preferences depended on different trip purposes in an obvious extension. Secondly, the current study estimates a PSL model only on bicycle trips collected with GPS trackers. An extension could be to estimate the model with the weight of each trip, that guarantees the representability of the sample, and therefore avoid self-selection of population levels, given that the participation to the data collection was voluntary. Thirdly, in order to only include realistic choice set alternatives for the model estimation, the estimated utility function could be used to generate a new set of alternatives for the route choice model and estimate the model again. This procedure should then be repeated iteratively to ensure consistency across model components.

Acknowledgment

The authors gratefully acknowledge the financial support of the Danish Road Directorate to the project "The effect of cycling policies". The authors acknowledge the support of Morten Aabrink and Svend Jacob Senstius at DTU Transport for their assistance with managing the data collection, post-processing the raw GPS data, and map-matching. We would also like to thank Søren H. Pedersen and Svend Jacob Senstius at DTU Transport for their help on constructing the bicycle network.

11.6 References

Anderson, M.K., Nielsen, O.A. and Prato, C.G. (2014). Multimodal route choice models of public transport passengers in the Greater Copenhagen Area. *EURO Journal of Transportation and Logistics*. Available: 10.1007/s13676-014-0063-3

Aultman-Hall, L., Hall, F.L. and Baetz, B.B. (1997). Analysis of Bicycle Commuter Routes Using Geographic Information Systems: Implications for Bicycle Planning. *Transportation Research Record: Journal of the Transportation Research Board, No. 1578*, TRB, National Research Council, Washington, D.C., pp. 102–110.

Axhausen, K.W. and Smith, R.L. (1986). Bicyclist link evaluation: a stated preference approach. *Transportation Research Record, No. 1085*, TRB, National Research Council, Washington, D.C., pp. 7–25.

Bekhor, S., Ben-Akiva, M.E. and Ramming, M.S. (2006). Evaluation of choice set generation algorithms for route choice models. *Annals of Operations Research*, 144, pp. 235–247.

Ben-Akiva, M.E. and Bierlaire, M. (1999). Discrete Choice Methods and Their Applications to Short-Term Travel Decisions. R. Hall (Ed.), *Handbook of Transportation Science*, Kluwer, pp. 5–34.

Bierlaire, M. (2008). *An introduction to BIOGEME.* Version 1.6, March 2008, Available at: http://biogeme.epfl.ch

Bliemer, M.C.J. and Bovy, P.H.L. (2008). Impact of Route Choice Set on Route Choice Probabilities. *Transportation Research Record: Journal of the Transportation Research Board, No. 2076*, Transportation Research Board of the National Academies, Washington, D.C., pp. 10–19.

Bovy, P.H.L. and Bradley, M.A. (1985). Route Choice Analyzed with Stated Preference Approaches. *Transportation Research Record, No. 1037*, TRB, National Research Council, Washington, D.C., pp. 11–20.

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

Broach, J., Dill, J. and Gliebe, J. (2012). Where do cyclists ride? A route choice model developed with revealed preference GPS data. *Transportation Research Part A,* 46, pp. 1730–1740.

Casello, J. and Usyukov, V. (2014). Modeling Cyclists' Route Choice Based on GPS Data. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2430, Transportation Research Board of the National Academia, Washington, D.C., pp.155-161. DOI: http://dx.doi.org/10.3141/2430-16

Christiansen, H. (2009). Modernisering af Transportvaneundersøgelsen. Presented at *Trafikdage* at Aalborg University.

Ehrgott, M., Wang, J. Y., Raith, A. and Van Houtte, C. (2012). A bi-objective cyclist route choice model. *Transportation research part A: policy and practice*, 46(4), 652-663.

ESRI (2015). *ArcGIS Desktop*: Release 10.3.1. Environmental Systems Research Institute. Redlands, California

FOT-Kort10 (2010). *Information om det nye Kort10 og FOT-Kort10.* [pdf] Available at: support.kortforsyningen.dk/attachments/download/292/information-kort10.pdf [Accessed 18 June 2015].

Halldórsdóttir, K., Rieser-Schüssler, N., Axhausen, K.W., Nielsen, O.A. and Prato C.G. (2014). Efficiency of choice set generation methods for bicycle routes. *European Journal of Transport and Infrastructure Research*, Vol. 14, No. 4, pp. 332–348.

Halldórsdóttir, K., Pedersen, S.H. and Senstius, S.J. (2013). *Bicycle Network for Greater Copenhagen area.* Unpublished data dictionary. Technical University of Denmark, Department of Transport, Kgs. Lyngby, Denmark.

Hess, S., Bierlaire, M. and Polak, J.W. (2005). Estimation of value of travel-time savings using mixed logit models*. Transportation Research Part A*, Volume 39, Issues 2–3, pp. 221–236.

Hess S., Train, K.E. and Polak, J.W. (2006). On the use of a Modified Latin Hypercube Sampling (MLHS) method in the estimation of a Mixed Logit Model for vehicle choice. *Transportation Research Part B,* 40, pp. 147–163.

Hood, J., Sall, E. and Charlton, B. (2011). A GPS-based bicycle route choice model for San Francisco, California. *Transportation Letters: The International Journal of Transport Research*, 3, pp. 63–75.

Hoogendoorn-Lanser, S., van Nes, R. and 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. and Wardman, M. (1996). Evaluating the demand for new cycle facilities. *Transport Policy*, Vol. 3, No. 4, pp. 241–249.

Howard, C. and Burns, E.K. (2001). Cycling to Work in Phoenix: Route Choice, Travel Behavior, and Commuter Characteristics. *Transportation Research Record: Journal of the Transportation Research Board, No. 1773, Part 2: Bicycle and Pedestrian Research*, TRB, National Research Council, Washington, D.C., pp. 39–46.

Hunt, J.D. and Abraham, J.E. (2007). Influences on bicycle use. *Transportation*, Vol. 34, pp. 453–470.

Hyodo, T., Suzuki, N. and Takahashi, K. (2000). Modeling of bicycle route and destination choice behavior for bicycle road network plan. *Transportation Research Record, No. 1705*, pp. 70–76.

Kort & Matrikelstyrelsen (2001). *TOP10DK - Geometrisk registrering*. Specifikation udgave 3.2.0. Kort & Matrikelstyrelsen, Copenhagen, Denmark. [pdf] 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*, Vol. 72, No. 3, pp. 309–320.

Larsen, J., Patterson, Z. and 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(4), 299-317.

Menghini, G., Carrasco, N., Schüssler, N. and Axhausen, K.W. (2010). Route choice of cyclists in Zurich. *Transportation Research Part A*, 44, pp. 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*, 34(5), pp. 377–402.

Nielsen, O.A. and Jovicic, G. (1999). A large-scale stochastic Timetable-based transit assignment model for route and sub-mode choices. In *Proceedings of Seminar F, Transportation Planning Methods*. (Vol. P 434). Cambridge.

Nielsen, O.A. and Jørgensen, R.M. (2004). Map-matching algorithms for GPS data – methodology and test on data from the AKTA roadpricing experiment in Copenhagen. In *5th TRI annual symposium on transportation analysis (TRISTAN V)*, Le Gosier, Guadeloupe.

OSM. © OpenStreetMap contributors. [website] Available at: http://www.openstreetmap.org/ [Accessed 11 August 2015].

Pedersen, S.H. and Senstius, S.J. (2014). *Landsdækkende cykelnetværk.* Unpublished data dictionary in Danish. Technical University of Denmark, Department of Transport, Kgs. Lyngby, Denmark.

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

Prato C.G. and Bekhor, S. (2007). Modeling Route Choice Behavior: How Relevant Is the Composition of Choice Set? *Transportation Research Record: Journal of the Transportation Research Board, No. 2003*, Transportation Research Board of the National Academies, Washington, D.C., pp. 64–73.

Pucher, J. and Buehler, R. eds. (2012). *City Cycling*. The MIT Press, Cambridge, Massachusetts. pp. 292.

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. and Prato, C.G. (2014). Timetable-based simulation method for choice set generation in large-scale public transport networks . *Paper submitted to Transportmetrica A*

Rasmussen T.K., Ingvardson, J.B., Halldórsdóttir, K. and 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*, Available at: http://dx.doi.org/10.1016/j.compenvurbsys.2015.04.001

Rich, J.H., Nielsen, O.A. and Hansen, C.O. (2010). Overall Design of the Danish National Transport Model. In proceedings for the *Annual Transport Conference in Aalborg*, Aalborg University.

Rieser-Schüssler, N., Balmer, M. and Axhausen, K.W. (2012). Route choice sets for very high-resolution data. *Transportmetrica*, 9, 825-845.

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

Schüssler, N. and Axhausen, K.W. (2009). Processing GPS raw data without additional information. *Transportation Research Record, No. 2105,* pp. 28–36.

Sener, I.N., Eluru, N. and Bhat, C.R. (2009). An analysis of bicycle route choice preferences in Texas, US. *Transportation*, 36, pp. 511–539.

Shafizadeh, K. and Niemeier, D. (1997). Bicycle Journey-to-Work: Travel Behavior Characteristics and Spatial Attributes. *Transportation Research Record, No. 1578*, pp. 84–90.

Snizek, B., Sick Nielsen, T. A. and Skov-Petersen, H. (2013). Mapping bicyclists' experiences in Copenhagen. *Journal of Transport Geography*, 30, 227-233. Available at: 10.1016/j.jtrangeo.2013.02.001

Stinson, M.A. and Bhat, C.R. (2003). Commuter Bicyclist Route Choice: Analysis Using a Stated Preference Survey. *Transportation Research Record: Journal of the Transportation Research Board, No. 1828*, TRB, National Research Council, Washington, D.C., pp. 107–115.

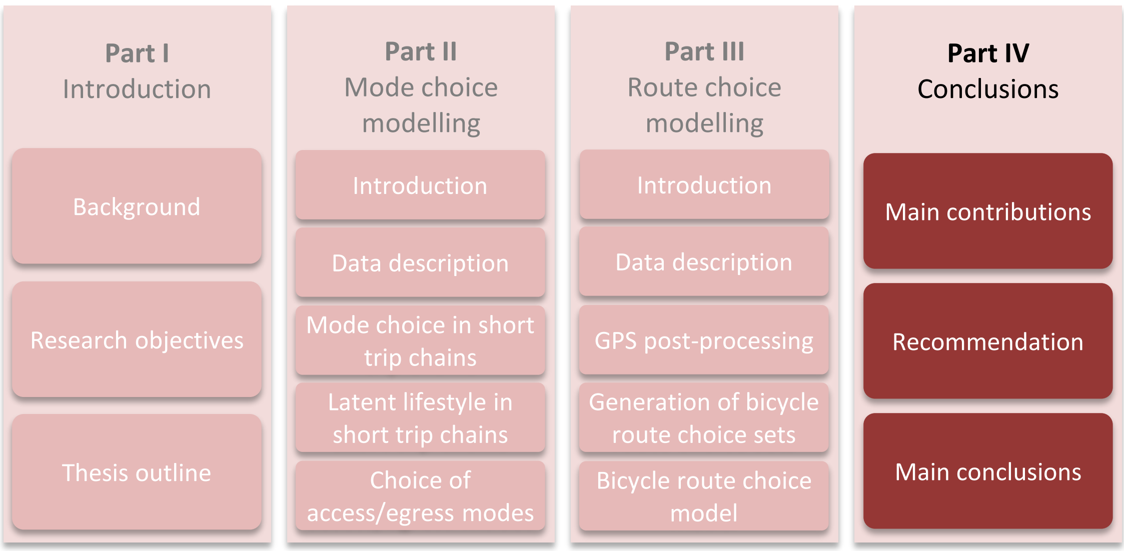
Tilahun, N.Y., Levinson, D.M. and Krizek, K.J. (2007). Trails, lanes, or traffic: Valuing bicycle facilities with an adaptive stated preference survey. *Transportation Research Part A*, Vol. 41, pp. 287–301.

Yeboah, G. and Alvanides, S. (2015). Route Choice Analysis of Urban Cycling Behaviors Using OpenStreetMap: Evidence from a British Urban Environment. *OpenStreetMap in GIScience* (pp. 189-210). Springer International Publishing.

Part IV

Conclusions

Conclusions to the PhD study



Chapter 12

CONCLUSIONS

CONCLUSIONS OF THE PHD STUDY

The research methods and the results have been discussed in the papers throughout this PhD thesis. This chapter briefly summarises the main contributions of the PhD study, in section 12.1. Section 12.2 then discusses how the different models, used in this study, can be put together in an overall model framework that improves bicycle modelling and supports policies. Finally, section 12.3 gives the overall conclusions of the study.

12.1 MAIN CONTRIBUTIONS

Through the application of advanced transport models, the following six general contributions have been completed on the subject of the modelling of bicycle behaviour in the Copenhagen Region:

- The *first* contribution regards the use of a mixed logit model to estimate the *mode choice preferences* in *short distance travelling*. The model proved effective in identifying the heterogeneity among cyclists in the sensitivity to travel time, temperature, and hilliness. The model estimates suggested that further heterogeneity investigation might uncover whether different population groups exhibit different preference structures.

- The *second* contribution regards the application of a *latent class choice model* to uncover different *lifestyle groups and choice specific travel behaviour* in *short distance travelling*. The findings highlight the importance of analysing the possible shift from the car to sustainable travel modes by considering the heterogeneous preferences of travellers, especially when comparing parameter estimates for travel time variables, and the significance of lifestyle decisions on short-term choices. Model estimates showed that four lifestyle groups were identified in the population: car oriented, bicycle oriented, public transport oriented, and public transport averse, where each group had heterogeneity in relation to travel time preferences with extremely different rates of substitution between alternative travel modes.
- The *third* contribution regards the analysis of the *mode choice behaviour* for the *access to and the egress from train stations*. The mixed logit model, used for estimation, successfully accounted for the heterogeneity in the travellers' preferences and heteroscedasticity across alternative modes. The findings emphasise the importance of estimating two different models for the home-end and the activity-end, given different preference structures related to the different knowledge of the network and the area. It is furthermore found that travellers' perceptions of the alternatives and the travel time are heterogeneous. The findings showed that the choice of bicycles as a transport alternative is to a large extent related to the travel time, but also to the policy variables, i.e., parking at stations and the possibility of carrying bicycles on trains. In addition, their preference structure relates more to their socio-economic characteristics as opposed to the trip characteristics.
- The *fourth* contribution regards the *collection of individual-based GPS data* and the importance of *automated post-processing procedures* to classify trips, trip legs, and the most probable mode of transport. A method, combining fuzzy logic and GIS-based algorithm to process raw GPS data, was developed and tested. The proposed method proved successful in processing the raw GPS data, thus making it possible to use GPS loggers to collect the actual route choices of cyclists, in large-scale multi-modal networks.
- The *fifth* contribution regards the *analysis of the efficiency of three choice set generation methods for bicycle routes*, i.e., a doubly stochastic generation function, a breadth first search on link elimination, and a branch & bound algorithm. The *extension of cost functions to bicycle-oriented factors,* not limited to distance and time, was proposed. The findings showed that both the doubly stochastic generation function and the breadth first search on link elimination generated realistic routes, while the first produced more heterogeneous routes and the latter outperformed in

computation cost. The doubly stochastic generation function was chosen for further analysis in this study, due to the complex nature of route choice behaviour in general and the importance of generating relevant and heterogeneous routes for model estimation.

- The *sixth* contribution regards the *route choice behaviour of cyclists*, where a path-size logit model was estimated, accounting for similarities between overlapping alternative routes. The logarithm of the path-size variable was positive and significant, hence it correctly accounted for route overlap. The model successfully identified the importance of numerous attributes and how they influence cyclists' route choices. The model estimates highlight the importance of well-built bicycle facilities and that cyclists also place a relatively high value on different land-use conditions along the route, emphasising the importance of choosing the location of new bicycle paths carefully.

12.2 Recommendation for future work

As discussed in Chapter 1, the problem with existing traffic models is that bicycle transport has been largely simplified in the models. This PhD study focused on the two last steps of the four-step travel modelling process, i.e., the mode choice and the route choice.

All three mode choice models, estimated in the study, emphasise the importance of not only considering the importance of travel time and trip characteristics, but also that travellers have heterogeneous preferences. The mixed logit models are a good means to investigate the effect of policy variables, while taking into consideration the heterogeneous preferences of travellers. The results from the mixed logit models showed that travellers' socio-economic characteristics are highly connected to their preference structure, both when travelling short distances as well as when accessing train stations. The findings from the mixed logit model, analysing short distance mode choices, suggested that further investigation of heterogeneity might uncover whether different population groups exhibit different preference structures. The latent class choice model proved to be a more suitable methodological approach to uncover the determinants of the choice between cars and sustainable transport alternatives when travelling short distances. The model allows identifying lifestyle groups and to understand how lifestyle affects mode choice decisions when travelling short distances. The finding from the mixed logit model, analysing the mode choice behaviour for the access to and the egress from train stations, also indicated that a latent lifestyle segmentation approach could reveal which population groups should be addressed when intending to improve the integration of active travel modes and public transport.

By estimating a path-size logit model, to analyse the route choice of existing cyclists, the rates of substitution between different route variables can be calculated. Consequently, decision-makers are able to weight cyclists' preferences for different route characteristics. Most importantly, the findings from the route choice model can be used to forecast future travel behaviour by investigating further the interaction between the bicycle route choice model and the mode choice models. Figure 69 gives an overview of this feedback mechanism in the four-stage modelling process, where the models studied in this thesis are shown in dark red. The route choice model can be implemented into the mode choice model, e.g., by using the logsum as a measure of consumer surplus or by using the estimated utility function, from the route choice model, to calculate the level-of-service variables for the bicycle alternatives in the mode choice models. Accordingly, the rather simple shortest path simulation method, from Section 3.3.3 – Shortest path generation, can be replaced and more complex route assignment used to calculate the attributes of the bicycle alternative within the choice set of each traveller, similar to the description in Section 3.3.1 – Road traffic assignment. In addition, by adding relevant variables from the bicycle route choice model, to the utility function in the mode choice models, scenario analysis can be conducted to further evaluation of the effects of possible policy instruments intending to increase cycling in the Copenhagen Region.

Furthermore, the interaction between the route choice in public transport, when the main transport mode is passenger trains, and the results from the access and egress mode choice model, needs to be investigated further. Figure 70 shows the feedback mechanism for access and egress modes to passenger trains. As shown in the figure, it is quite complex to model the access and egress mode choice and it entails a lot of bookkeeping to the model framework. The level-of-service variables for the feeder modes would need to be re-calculated and then a new demand assigned. Even though it looks quite complicated, it is fairly straightforward since all the components are available from this study.

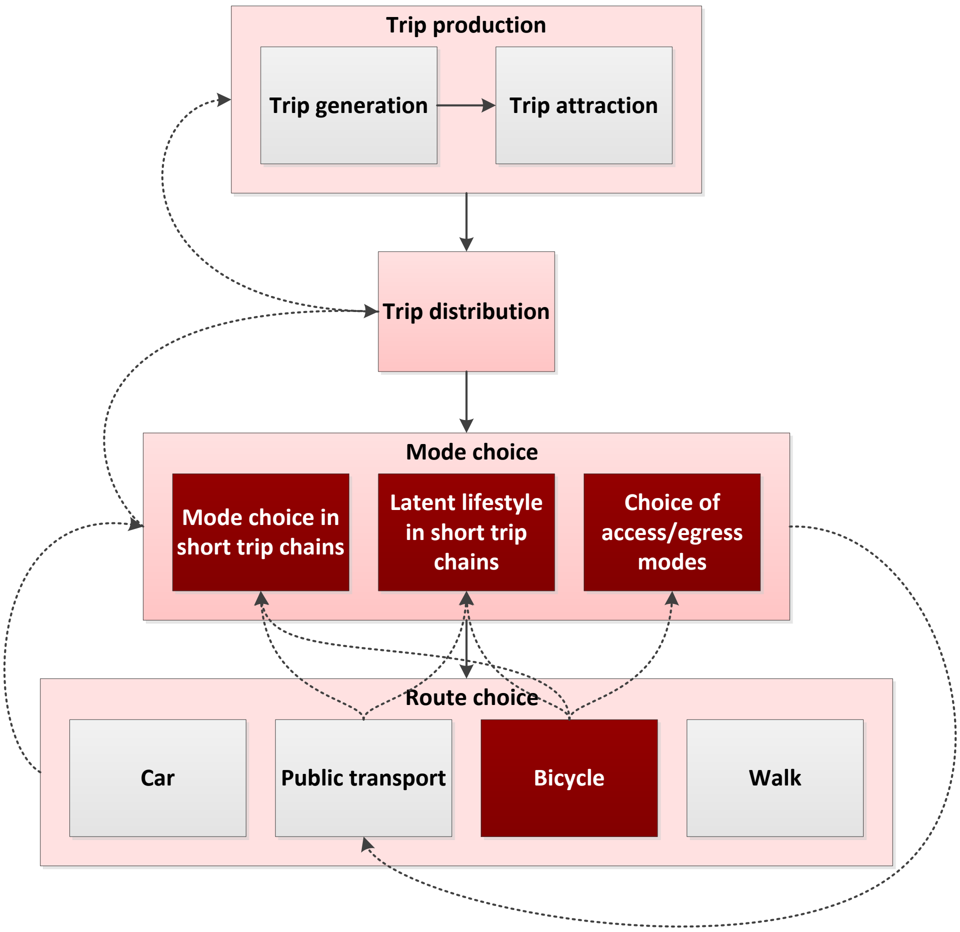


FIGURE 69: FEEDBACK MECHANISM IN THE FOUR-STEP TRAVEL MODELLING PROCESS

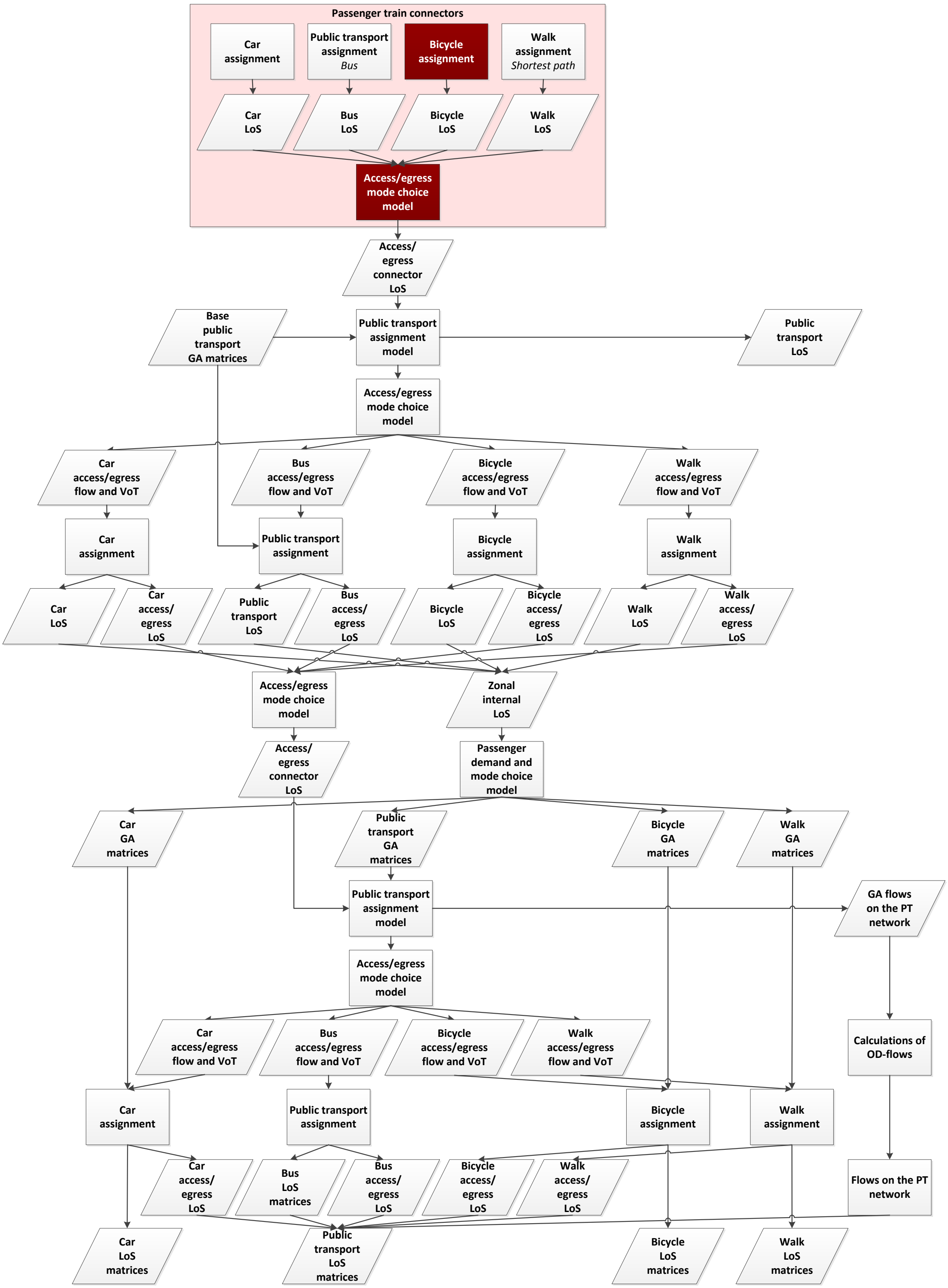


FIGURE 70: FEEDBACK MECHANISM FOR ACCESS/EGRESS MODES TO PASSENGER TRAINS

12.3 MAIN CONCLUSIONS

The main conclusion of this PhD study is that it is possible to estimate quite advanced models: (i) to evaluate the potential of shifting from the private car to sustainable travel options, when travelling short distances; (ii) to investigate the choice of access and egress modes to and from the railway network, by taking into consideration the difference in preference structure at the home-end and activity-end; and (iii) to analyse cyclists' route choice behaviour. These models included a much more elaborate set of variables and utility functions than the very simple models that are often used in practice. It was possible to analyse travel behaviour on extensive revealed preference data, both the preferences in the choice of transport mode and cyclists' route choice behaviour.

Firstly, when thinking about measures to increase the attractiveness of bicycles as a sustainable transport option in short distance travelling, decision-makers should address specific population groups for specific trip purposes and focus on factors that are able to make cycling more attractive in order to encourage the shift from private cars to more sustainable transport alternatives. They should propose traditional or creative solutions to encourage car-oriented individuals out of their cars and minimise the attractiveness of cars in the future for bicycle- and walk-oriented individuals. They should also direct public transport-averse individuals towards the bicycle, with policies that make the car unattractive.

Secondly, when thinking about bicycle infrastructure improvements to increase the use of bicycle as a transport alternative in short distance travelling, the findings showed that decreasing the travel time on bicycles has little effect on car-oriented individuals, unless the time savings are very high. Bicycle-oriented individuals will only modify their routes as they already consider cycling as the fastest means of transport.

Thirdly, when thinking about factors that can contribute to the sustainability of the travel choices after selecting a train as the main transport mode, it is important to improve train station accessibility, e.g., by improving the bicycle network infrastructure. Increasing bicycle parking availability at train stations is certainly helpful, especially improving the conditions of bicycle parking at metro stations. The availability of covered bicycle parking also increases the probability of cycling. Alternatively, travellers do not seem to be too concerned with locked parking areas as there is already a tradition to use on-street parking. It is also important to provide the opportunity to carry the bicycle on the train.

Finally, when thinking about bicycle infrastructure improvements to make cycling more attractive for existing cyclists, decision-makers should focus on minimising excessive detours, by improving the connectivity of the bicycle network and making the infrastructure seamless, while avoiding hilly areas. The findings also highlight the importance of well-built bicycle facilities, i.e., segregated bicycle paths and bicycle lanes, as well as bicycle bridges. The cyclists also place relatively high value on different land-use conditions, emphasising the importance of choosing the location of new bicycle paths carefully.

The bicycle route choice model, developed in this study, is an important component in improving trip assignment models for cyclists. The estimated utility function, from the bicycle route choice model, can be used to generate realistic bicycle routes between each origin-destination pair. In addition, it can be used to calculate more accurately the level-of-service variables for the bicycle alternatives in mode choice models, by including, e.g., the preference for different bicycle facility types and the negative effect of hilliness. Avenues for future research are needed. In order to forecast future travel behaviour, the interaction between the bicycle route choice model and the mode choice models needs to be investigated further. Scenario simulations would also allow further evaluation of the effects of implementing a new infrastructure or a political initiative, intending to increase cycling in the Copenhagen Region, and to forecast future behaviour.

PART V

APPENDIX

APPENDIX A

BICYCLE NETWORK DATABASE – DATA DESCRIPTION

ATTRIBUTE TABLE

A Bicycle network database for the Greater Copenhagen area

This appendix gives an overview and a description of the attributes in the bicycle network database, described in section 8.2. For further information on how the network was constructed, see: Halldórsdóttir et al. (2013) and Pedersen and Senstius (2014).

Table 38: Bicycle network - attributes overview

<table>
<tr><th>Name</th><th>Description</th></tr>
<tr><td>ID</td><td>Unique id</td></tr>
<tr><td>From_node</td><td rowspan="2">Unique node identification</td></tr>
<tr><td>To_node</td></tr>
<tr><td>OpenFor</td><td rowspan="2">Driving directions</td></tr>
<tr><td>OpenBack</td></tr>
<tr><td>Type</td><td>11 Road
12 Road with bicycle lane
13 Road with bicycle path
21 Bicycle path
22 Footpath
23 Steps
31 No bicycle access
32 No access</td></tr>
<tr><td>Surface</td><td>1 Paved
2 Paved, cobblestone
3 Unpaved
4 Unpaved, only mountain bikes</td></tr>
<tr><td>LanduseRight</td><td rowspan="2">Low residential
High residential
Industry
Town centre
Park
Forrest
Heath
Cemetery
Sport facilities
Sand
Technical facilities
Gravel pit
Coast
Lake
Wetland
Stream</td></tr>
<tr><td>LanduseLeft</td></tr>
</table>

Cum_elev_gain		Sum of elevation gain/loss on all subparts of a link, where gain and loss is relative to drawing direction of the link.	
Cum_elev_loss			
Cum_elev_gain_0_10		Sum of elevation gain/loss on all subparts of a link categorized in ‰, where gain and loss is relative to drawing direction of the link.	
Cum_elev_gain_10_35			
Cum_elev_gain_35_50			
Cum_elev_gain_above_50			
Cum_elev_loss_0_10			
Cum_elev_loss_10_35			
Cum_elev_loss_35_50			
Cum_elev_loss_above_50			
FromIntersectionLegsAll		Count of legs in an intersection.	Dead end links are excluded in the count.
ToIntersectionLegsAll			
FromIntersectionLegsRoad			Dead end links and paths (type 21 and 22) are excluded in the count.
ToIntersectionLegsRoad			
FromIntersectionType	-1	Unknown	
	0	No intersection (pseudo-nodes and path intersecting road)	
	1	Giveaway junction (da: vigepligt)	
ToIntersectionType	2	Roundabout	
	3	Traffic Signal	
MotorTrafficSpeedLimit		Based on NavTeq Streets	For paths with type 13 values refers to corresponding road.
MotorTrafficLanes			
MotorTrafficFunctionalClass	1	Used for roads with high volume, maximum speed traffic movement between and through major metropolitan areas.	
	2	Used to channel traffic to MotorTrafficFunctionalClass = 1 roads.	
	3	Roads which interconnect MotorTrafficFunctionalClass = 2 roads and provide a high volume of traffic movement at a lower level of mobility than these.	
	4	Roads which provide for a high volume of traffic movement at moderate speeds between neighbourhoods.	
	5	Roads with low volume and traffic movement. In addition, walkways, Parking lanes etc.	
	-1	Non-navigable links	
LTM_ID		Reference to LTM version 1.05 road network	
LTM_LinkType			
LTM_WDT		LTM modelled average week day total traffic	
LTM_TruckShare		Truck share of LTM_WDT	

<table>
<tr><td>LTM_FreeSpeed</td><td colspan="2">Uncongested speed used in LTM.</td></tr>
<tr><td>AccidentsVejman08_12</td><td colspan="2">Count of reported accidents 2008 to 2012.</td></tr>
<tr><td>AccVejmann08_12_NoIntersect</td><td colspan="2">Count of reported accidents not related to intersections.</td></tr>
<tr><td>Dead_end</td><td>Value is 1 for dead end link, while links with values greater than 1 is dead ends, only if links with lower values are removed.</td><td rowspan="2">Possibly to be used for limiting the number of links.</td></tr>
<tr><td>Serviceroad</td><td>Road of lesser importance.</td></tr>
<tr><td>Urban</td><td colspan="2">Urban area is defined as place with more than 200 inhabitants and less than 200 meters between houses.</td></tr>
<tr><td>Shape_Length</td><td colspan="2">Length in meters.</td></tr>
</table>

Appendix B

Results from Halldórsdóttir et al. (2015b)

Path-size logit model estimates, with linear combinations

B Results from Halldórsdóttir et al. (2015b): the model estimates, with linear combinations

This appendix supplements Halldórsdóttir et al. (2015b). The following sections present the results from the path-size logit model estimates, with linear combinations. For further information, see Halldórsdóttir et al. (2015b).

B.1 Gender	Page 272
B.2 Type of cyclist	Page 274
B.3 Darkness	Page 276
B.4 Peak hours	Page 278
B.5 Weekend	Page 280
B.6 Air temperature	Page 282
B.7 Rain	Page 284
B.8 Sunshine	Page 286

B.1 Gender

Parameter	Model estimates Value	t-test		Rates of substitution
Network attributes				
ln(Path-size)	1.36	21.17	***	-
Length	-0.95	-3.01	***	-1.00
Wrong way	-1.21	-7.47	***	-1.28
Number of turns				
Left	-0.32	-19.58	***	-0.34
Right	-0.20	-12.09	***	-0.22
Cumulative elevation gain				
0-10 meters/km	*ref.*			
10-35 meters/km - Female	-40.20	-3.00	***	-42.41
10-35 meters/km - Male	-24.00	-1.99	**	-25.32
35-50 meters/km - Female	-51.70	-2.47	***	-54.54
35-50 meters/km - Male	-22.90	-1.17		-24.16
Above 50 meters/km - Female	-60.20	-4.58	***	-63.50
Above 50 meters/km - Male	-40.60	-3.29	***	-42.83
Bicycle facility type				
Motorised road without any bicycle facilities	*ref.*			
Motorised road with segregated bicycle path/bicycle lane	0.13	3.63	***	0.13
Bicycle path in own trace - Female	-			-
Bicycle path in own trace - Male	-0.30	-3.18	***	-0.31
Footpath in own trace	-1.87	-9.12	***	-1.97
Steps	-12.90	-4.46	***	-13.61
Bicycle bridge	2.13	1.59		2.25
Surface type				
Paved	*ref.*			
Not paved	-0.20	-2.13	**	-0.21
Number of intersections - Roundabout	0.05	3.22	***	0.05
Number of motorised traffic lanes				
1 lane	1.29	2.99	***	1.36
2 lanes	*ref.*			
3 to 4 lanes	-0.23	-4.14	***	-0.24
5 lanes and above	-0.42	-5.73	***	-0.44
Motorised traffic bridge, crossing water/sea	-1.27	-2.37	**	-1.34
Land-use influence				
Low residential area on the right side	*ref.*			
Low residential area on both sides	*ref.*			
High residential area and/or town centre on one side	-0.24	-3.71	***	-0.25
High residential area and/or town centre on both sides	-0.60	-9.66	***	-0.63
Industry on both sides	-0.30	-2.68	***	-0.32
Sport on one side	-0.53	-4.20	***	-0.56
Sport on both sides	0.78	3.55	***	0.82
Park on one side	-0.13	-2.11	**	-0.14
Park on both sides - Female	0.24	1.71	*	0.26
Park on both sides - Male	-0.10	-0.90		-0.11

Number of estimated parameters:	30
Number of observations:	3,363
Null log-likelihood:	-12,761.075
Final log-likelihood:	-9,675.673
Adjusted rho-square:	0.239

Note: * significant at the 90% level; ** significant at the 95% level; *** significant at the 99% level.

B.2 Type of cyclist

Parameter	Model estimates Value	t-test		Rates of substitution
Network attributes				
ln(Path-size)	1.36	21.23	***	-
Length	-0.901	-2.89	***	-1.00
Wrong way	-1.22	-7.59	***	-1.35
Number of turns				
Left	-0.328	-19.94	***	-0.36
Right	-0.205	-12.31	***	-0.23
Cumulative elevation gain				
0-10 meters/km	*ref.*			
10-35 meters/km	-33.5	-3.66	***	-37.18
35-50 meters/km	-38.5	-2.56	***	-42.73
Above 50 meters/km	-49.9	-5.62	***	-55.38
Bicycle facility type				
Motorised road without any bicycle facilities	*ref.*			
Motorised road with segregated bicycle path/bicycle lane	0.104	3.16	***	0.12
Bicycle path in own trace - Slow	-			-
Bicycle path in own trace - Medium	-			-
Bicycle path in own trace - Fast	-0.454	-4.49	***	-0.50
Footpath in own trace	-1.83	-9.14	***	-2.03
Steps	-12.8	-4.32	***	-14.21
Bicycle bridge	1.91	1.48		2.12
Surface type				
Paved	*ref.*			
Not paved	-0.264	-2.87	***	-0.29
Number of intersections - Roundabout	0.049	3.13	***	0.05
Number of motorised traffic lanes				
1 lane	1.42	3.36	***	1.58
2 lanes	*ref.*			
3 to 4 lanes	-0.179	-3.03	***	-0.20
5 lanes and above	-0.421	-5.9	***	-0.47
Motorised traffic bridge, crossing water/sea	-1.37	-2.57	***	-1.52
Land-use influence				
Low residential area on the right side	*ref.*			
Low residential area on both sides	*ref.*			
High residential area and/or town centre on one side	-0.241	-3.82	***	-0.27
High residential area and/or town centre on both sides	-0.612	-9.95	***	-0.68
Industry on both sides	-0.308	-2.74	***	-0.34
Sport on one side	-0.539	-4.15	***	-0.60
Sport on both sides	0.888	3.6	***	0.99
Park on one side - Slow	-1.31	-2.06	**	-1.45
Park on one side - Medium	-			-
Park on one side - Fast	-0.19	-2.54	***	-0.21

Park on both sides - Slow	-			-
Park on both sides - Medium	-0.219	-1.78	*	-0.24
Park on both sides - Fast	0.659	4.59	***	0.73
Number of estimated parameters:				28
Number of observations:				3,363
Null log-likelihood:				-12,761.075
Final log-likelihood:				-9,658.176
Adjusted rho-square:				0.241

Note: * significant at the 90% level; ** significant at the 95% level; *** significant at the 99% level.

B.3 Darkness

Parameter	Model estimates Value	t-test		Rates of substitution
Network attributes				
ln(Path-size)	1.35	20.36	***	-
Length	-0.916	-2.84	***	-1.00
Wrong way	-1.2	-7.43	***	-1.31
Number of turns				
Left	-0.324	-20	***	-0.35
Right	-0.201	-12.1	***	-0.22
Cumulative elevation gain				
0-10 meters/km	*ref.*			
10-35 meters/km	-32.1	-3.56	***	-35.04
35-50 meters/km	-32.6	-2.17	**	-35.59
Above 50 meters/km	-47.9	-5.53	***	-52.29
Bicycle facility type				
Motorised road without any bicycle facilities	*ref.*			
Motorised road with segregated bicycle path/bicycle lane	0.107	2.94	***	0.12
Bicycle path in own trace - Darkness	-0.727	-4.11	***	-0.79
Bicycle path in own trace - Light	-0.17	-2.22	**	-0.19
Footpath in own trace	-1.92	-9.26	***	-2.10
Steps	-12.6	-4.32	***	-13.76
Bicycle bridge	2.23	1.7	*	2.43
Surface type				
Paved	*ref.*			
Not paved	-0.176	-1.82	*	-0.19
Number of intersections - Roundabout	0.048	3.09	***	0.05
Number of motorised traffic lanes				
1 lane	1.27	2.99	***	1.39
2 lanes	*ref.*			
3 to 4 lanes	-0.219	-3.85	***	-0.24
5 lanes and above	-0.415	-5.69	***	-0.45
Motorised traffic bridge, crossing water/sea	-1.31	-2.49	***	-1.43
Land-use influence				
Low residential area on the right side	*ref.*			
Low residential area on both sides	*ref.*			
High residential area and/or town centre on one side	-0.209	-3.34	***	-0.23
High residential area and/or town centre on both sides	-0.591	-9.55	***	-0.65
Industry on both sides	-0.288	-2.56	***	-0.31
Sport on one side	-0.546	-4.23	***	-0.60
Sport on both sides	0.848	3.99	***	0.93
Park on one side - Darkness	-			-
Park on one side - Light	-0.119	-2.11	**	-0.13
Park on both sides - Darkness	-0.261	-1.24		-0.28
Park on both sides - Light	0.196	1.78	**	0.21

Number of estimated parameters:	28
Number of observations:	3,363
Null log-likelihood:	-12,761.075
Final log-likelihood:	-9,675.422
Adjusted rho-square:	0.240

Note: * significant at the 90% level; ** significant at the 95% level; *** significant at the 99% level.

B.3 Peak hours

Parameter	Model estimates Value	t-test		Rates of substitution
Network attributes				
ln(Path-size)	1.36	21.08	***	-
Length	-0.93	-2.95	***	-1.00
Wrong way	-1.2	-7.45	***	-1.29
Number of turns				
Left	-0.32	-19.7	***	-0.35
Right	-0.2	-12.1	***	-0.22
Cumulative elevation gain				
0-10 meters/km	*ref.*			
10-35 meters/km	-34	-3.73	***	-36.68
35-50 meters/km	-37.4	-2.5	***	-40.35
Above 50 meters/km	-49.4	-5.58	***	-53.29
Bicycle facility type				
Motorised road without any bicycle facilities	*ref.*			
Motorised road with segregated bicycle path/bicycle lane	0.115	3.14	***	0.12
Bicycle path in own trace - Peak hours	-0.26	-2.45	***	-0.28
Bicycle path in own trace - Off-peak hours	-0.17	-1.94	**	-0.18
Footpath in own trace	-1.92	-9.17	***	-2.07
Steps	-12.6	-4.29	***	-13.59
Bicycle bridge	2.1	1.59		2.27
Surface type				
Paved	*ref.*			
Not paved	-0.21	-2.03	**	-0.23
Number of intersections - Roundabout	0.047	3.06	***	0.05
Number of motorised traffic lanes				
1 lane	1.24	2.97	***	1.34
2 lanes	*ref.*			
3 to 4 lanes	-0.23	-4.01	***	-0.25
5 lanes and above	-0.45	-6.56	***	-0.48
Motorised traffic bridge, crossing water/sea	-1.43	-2.67	***	-1.54
Land-use influence				
Low residential area on the right side	*ref.*			
Low residential area on both sides	*ref.*			
High residential area and/or town centre on one side	-0.23	-3.58	***	-0.24
High residential area and/or town centre on both sides	-0.6	-10.1	***	-0.65
Industry on both sides	-0.32	-2.88	***	-0.35
Sport on one side	-0.56	-4.41	***	-0.60
Sport on both sides	0.869	3.99	***	0.94
Park on one side - Peak hours	-0.2	-2.83	***	-0.21
Park on one side - Off-peak hours	-			-
Park on both sides - Peak hours	-			-
Park on both sides - Off-peak hours	0.25	1.86	*	0.27

Number of estimated parameters:	27
Number of observations:	3,363
Null log-likelihood:	-12,761.075
Final log-likelihood:	-9,682.419
Adjusted rho-square:	0.239

Note: * significant at the 90% level; ** significant at the 95% level; *** significant at the 99% level.

B.4 WEEKEND

Parameter	Model estimates Value	t-test		Rates of substitution
Network attributes				
ln(Path-size)	1.36	20.71	***	-
Length	-0.919	-2.89	***	-1.00
Wrong way	-1.19	-7.39	***	-1.29
Number of turns				
Left	-0.324	-19.72	***	-0.35
Right	-0.201	-12.11	***	-0.22
Cumulative elevation gain				
0-10 meters/km	*ref.*			
10-35 meters/km	-34.2	-3.77	***	-37.21
35-50 meters/km	-38.1	-2.52	***	-41.46
Above 50 meters/km	-49.7	-5.63	***	-54.08
Bicycle facility type				
Motorised road without any bicycle facilities	*ref.*			
Motorised road with segregated bicycle path/bicycle lane - Weekday	0.09	2.58	***	0.10
Motorised road with segregated bicycle path/bicycle lane - Weekend	0.209	3.6	***	0.23
Bicycle path in own trace - Weekday	-0.229	-2.83	***	-0.25
Bicycle path in own trace - Weekend	-			-
Footpath in own trace	-1.91	-9.25	***	-2.08
Steps	-12.7	-4.33	***	-13.82
Bicycle bridge	2.07	1.58		2.25
Surface type				
Paved	*ref.*			
Not paved	-0.226	-2.15	**	-0.25
Number of intersections - Roundabout	0.05	3.19	***	0.05
Number of motorised traffic lanes				
1 lane	1.26	2.93	***	1.37
2 lanes	*ref.*			
3 to 4 lanes	-0.209	-3.57	***	-0.23
5 lanes and above	-0.421	-5.88	***	-0.46
Motorised traffic bridge, crossing water/sea	-1.37	-2.58	***	-1.49
Land-use influence				
Low residential area on the right side	*ref.*			
Low residential area on both sides	*ref.*			
High residential area and/or town centre on one side	-0.218	-3.48	***	-0.24
High residential area and/or town centre on both sides	-0.601	-9.6	***	-0.65
Industry on both sides	-0.311	-2.78	***	-0.34
Sport on one side	-0.561	-4.36	***	-0.61
Sport on both sides	0.854	3.87	***	0.93
Park on one side - Weekday	-0.128	-2.25	**	-0.14
Park on one side - Weekend	-			-

Park on both sides - Weekday	0.233	2.07	**	0.25
Park on both sides - Weekend	-0.234	-1.25		-0.25
Number of estimated parameters:				28
Number of observations:				3,363
Null log-likelihood:				-12,761.075
Final log-likelihood:				-9,681.773
Adjusted rho-square:				0.239

Note: * significant at the 90% level; ** significant at the 95% level; *** significant at the 99% level.

B.5 Air temperature

Parameter	Model estimates Value	t-test		Rates of substitution
Network attributes				
ln(Path-size)	1.35	20.08	***	-
Length	-0.892	-2.73	***	-1.00
Wrong way	-1.19	-7.32	***	-1.33
Number of turns				
Left	-0.325	-19.8	***	-0.36
Right	-0.202	-12	***	-0.23
Cumulative elevation gain				
0-10 meters/km	*ref.*			
10-35 meters/km	-33.1	-3.72	***	-37.11
35-50 meters/km	-36.5	-2.43	***	-40.92
Above 50 meters/km	-49.3	-5.64	***	-55.27
Bicycle facility type				
Motorised road without any bicycle facilities	*ref.*			
Motorised road with segregated bicycle path/bicycle lane - Low	-0.135	-1.63	***	-0.15
Motorised road with segregated bicycle path/bicycle lane - Medium/high	0.119	3.23	***	0.13
Bicycle path in own trace - Low	-1.44	-3.56	***	-1.61
Bicycle path in own trace - Medium/high	-0.165	-2.14	**	-0.18
Footpath in own trace	-1.89	-9.2	***	-2.12
Steps	-12.7	-4.34	***	-14.24
Bicycle bridge	2.38	1.81	*	2.67
Surface type				
Paved	*ref.*			
Not paved	-0.209	-2.07	**	-0.23
Number of intersections - Roundabout	0.048	3.08	***	0.05
Number of motorised traffic lanes				
1 lane	1.24	2.94	***	1.39
2 lanes	*ref.*			
3 to 4 lanes	-0.213	-3.73	***	-0.24
5 lanes and above	-0.412	-5.69	***	-0.46
Motorised traffic bridge, crossing water/sea	-1.31	-2.46	***	-1.47

Land-use influence				
Low residential area on the right side	*ref.*			
Low residential area on both sides	*ref.*			
High residential area and/or town centre on one side	-0.209	-3.36	***	-0.23
High residential area and/or town centre on both sides	-0.597	-9.48	***	-0.67
Industry on both sides	-0.298	-2.66	***	-0.33
Sport on one side	-0.55	-4.33	***	-0.62
Sport on both sides	0.84	3.93	***	0.94
Park on one side - Low	-0.226	-1.31		-0.25
Park on one side - Medium/high	-0.112	-1.99	**	-0.13
Park on both sides - Low	-0.776	-1.48		-0.87
Park on both sides - Medium/high	0.196	1.84	*	0.22
Number of estimated parameters:				30
Number of observations:				3,363
Null log-likelihood:				-12,761.08
Final log-likelihood:				-9,666.949
Adjusted rho-square:				0.24

Note: * significant at the 90% level; ** significant at the 95% level; *** significant at the 99% level.

B.6 RAIN

Parameter	Model estimates Value	t-test		Rates of substitution
Network attributes				
ln(Path-size)	1.36	62.01	***	-
Length	-0.912	-39.6	***	-1
Wrong way	-1.19	-15.1	***	-1.30
Number of turns				
Left	-0.324	-38.1	***	-0.36
Right	-0.202	-16.6	***	-0.22
Cumulative elevation gain				
0-10 meters/km	*ref.*			
10-35 meters/km	-33.7	-4.87	***	-36.95
35-50 meters/km	-38.5	-3.03	***	-42.21
Above 50 meters/km	-49.8	-7.93	***	-54.61
Bicycle facility type				
Motorised road without any bicycle facilities	*ref.*			
Motorised road with segregated bicycle path/bicycle lane - Rain	-			-
Motorised road with segregated bicycle path/bicycle lane - No rain	0.12	4.16	***	0.13
Bicycle path in own trace - Rain	-			-
Bicycle path in own trace - No rain	-0.19	-4.23	***	-0.21
Footpath in own trace	-1.9	-17.5	***	-2.08
Steps	-12.9	-4.66	***	-14.14
Bicycle bridge	2.01	1.79	*	2.20
Surface type				
Paved	*ref.*			
Not paved	-0.227	-3.18	***	-0.25
Number of intersections - Roundabout	0.049	3.03	***	0.05
Number of motorised traffic lanes				
1 lane	1.28	2.79	***	1.40
2 lanes	*ref.*			
3 to 4 lanes	-0.22	-5.57	***	-0.24
5 lanes and above	-0.425	-5.84	***	-0.47
Motorised traffic bridge, crossing water/sea	-1.35	-2.48	***	-1.48
Land-use influence				
Low residential area on the right side	*ref.*			
Low residential area on both sides	*ref.*			
High residential area and/or town centre on one side	-0.223	-4.82	***	-0.24
High residential area and/or town centre on both sides	-0.597	-13.9	***	-0.65
Industry on both sides	-0.299	-3.69	***	-0.33
Sport on one side	-0.564	-4.93	***	-0.62
Sport on both sides	0.848	3.75	***	0.93
Park on one side - Rain	-			-
Park on one side - No rain	-0.103	-2.3	**	-0.11

Park on both sides - Rain	-			-
Park on both sides - No rain	0.161	2.75	***	0.18
Number of estimated parameters:				26
Number of observations:				3,363
Null log-likelihood:				-12,761.075
Final log-likelihood:				-9,686.528
Adjusted rho-square:				0.239

Note: * significant at the 90% level; ** significant at the 95% level; *** significant at the 99% level.

B.7 SUNSHINE

Parameter	Model estimates Value	t-test		Rates of substitution
Network attributes				
ln(Path-size)	1.35	61.89	***	-
Length	-0.88	-38.2	***	-1
Wrong way	-1.2	-15.2	***	-1.36
Number of turns				
Left	-0.327	-38.7	***	-0.37
Right	-0.208	-17.1	***	-0.24
Cumulative elevation gain				
0-10 meters/km	*ref.*			
10-35 meters/km	-34.8	-5.1	***	-39.55
35-50 meters/km	-38.8	-3.07	***	-44.09
Above 50 meters/km	-49.1	-7.89	***	-55.80
Bicycle facility type				
Motorised road without any bicycle facilities	*ref.*			
Motorised road with segregated bicycle path/bicycle lane - Sunshine	0.137	4.16	***	0.16
Motorised road with segregated bicycle path/bicycle lane - No sunshine	-0.296	-5.79	***	-0.34
Bicycle path in own trace - Sunshine	-0.156	-2.94	***	-0.18
Bicycle path in own trace - No sunshine	-			-
Footpath in own trace	-1.92	-17.5	***	-2.18
Steps	-12.8	-4.62	***	-14.55
Bicycle bridge	2.13	1.9	*	2.42
Surface type				
Paved	*ref.*			
Not paved	-0.198	-2.73	***	-0.23
Number of intersections - Roundabout	0.047	2.89	***	0.05
Number of motorised traffic lanes				
1 lane	1.33	2.87	***	1.51
2 lanes	*ref.*			
3 to 4 lanes	-0.207	-5.64	***	-0.24
5 lanes and above	-0.423	-5.78	***	-0.48
Motorised traffic bridge, crossing water/sea	-1.37	-2.52	***	-1.56
Land-use influence				
Low residential area on the right side	*ref.*			
Low residential area on both sides	*ref.*			
High residential area and/or town centre on one side	-0.203	-4.74	***	-0.23
High residential area and/or town centre on both sides	-0.591	-13.9	***	-0.67
Industry on both sides	-0.279	-3.43	***	-0.32
Sport on one side	-0.572	-4.99	***	-0.65
Sport on both sides	0.871	3.85	***	0.99
Park on one side - Sunshine	-0.127	-2.34	**	-0.14
Park on one side - No sunshine	-			-

Park on both sides - Sunshine	0.233	3.39	***	0.26
Park on both sides - No sunshine	-			-
Number of estimated parameters:				27
Number of observations:				3,363
Null log-likelihood:				-12,761.075
Final log-likelihood:				-9,687.079
Adjusted rho-square:				0.239

Note: * significant at the 90% level; ** significant at the 95% level; *** significant at the 99% level.5

DTU Transport performs research and provides education on traffic and transport planning. It advises the Danish Ministry of Transport on infrastructure, economic appraisals, transport policy and road safety and collects data on the transport habits of the population. DTU Transport collaborates with companies on such topics as logistics, public transport and intelligent transport systems.

DTU Transport
Department of Transport
Technical University of Denmark

Bygningstorvet 116B
DK-2800 Kgs. Lyngby
Tel. +45 45 25 65 00
Fax +45 45 93 65 33

www.transport.dtu.dk