Unravelling Mode and Route Choice Behaviour of Active Mode Users

Ton, Danique

DOI
10.4233/uuid:bb07b47f-9e2c-448a-a235-9f29baed2d5d

Publication date
2019

Document Version
Final published version

Citation (APA)
Ton, D. (2019). *Unravelling Mode and Route Choice Behaviour of Active Mode Users*. [Dissertation (TU Delft), Delft University of Technology]. TRAIL Research School. https://doi.org/10.4233/uuid:bb07b47f-9e2c-448a-a235-9f29baed2d5d

Unravelling Mode and Route Choice Behaviour of Active Mode Users

Danique Ton

Delft University of Technology, 2019

This research is part of the Allegro project (*Unravelling active mode traveling and traffic: with innovative data to a new transportation and traffic theory for pedestrians and bicycles*), which is financed by the European Research Council and Amsterdam Institute for Advanced Metropolitan Solutions (no. 669792).

Cover photo: Danique Ton, design: Lara-Britt Zomer

Unravelling Mode and Route Choice Behaviour of Active Mode Users

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus Prof. dr. ir. T.H.J.J. van der Hagen,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op dinsdag 17 September 2019 om 15:00

door

Danique TON
Master of Science in Civil Engineering, Technische Universiteit Delft, Nederland
geboren te Vlissingen, Nederland

This dissertation has been approved by the
Promotors: Prof. dr. ir. S.P. Hoogendoorn and Dr. O. Cats
Copromotor: Dr. ir. D.C. Duives

Composition of the doctoral committee:

Rector Magnificus	chairperson
Prof. dr. ir. S.P. Hoogendoorn	promotor
Dr. O. Cats	promotor
Dr. ir. D.C. Duives	copromotor

Independent members:

Prof. dr. G.P. van Wee	Delft University of Technology, Netherlands
Prof. dr. E. Cherchi	University of Newcastle, United Kingdom
Prof. dr. S. Bekhor	Technion – Israel Institute for Technology, Israel
Prof. dr. P.L. Mokhtarian	Georgia Institute of Technology, United States of America

Reserve member:

Prof. dr. M.P. Hagenzieker	Delft University of Technology, Netherlands

TRAIL Thesis Series no. T2019/12, the Netherlands Research School TRAIL

TRAIL
P.O. Box 5017
2600 GA Delft
The Netherlands
E-mail: info@rsTRAIL.nl

ISBN: 978-90-5584-254-4

Printed in the Netherlands

"Luctor et Emergo"

Acknowledgements

My PhD journey could have started right after finishing my MSc degree in April 2014 on invitation by Serge. Luckily it didn’t. In hindsight, it was not the right timing and topic for me. Sorry again, Serge, for waiting so long before declining the offer! During my time at NS Stations I thoroughly enjoyed working in practice on interesting projects and with lovely colleagues, however deep down I still wanted to do more in-depth research. In October 2015, I ran into Serge again after he had been awarded a grant for the Allegro project. This big project on active modes (walking and cycling) contained work for eight PhD projects, of which one really caught my attention. The idea of working with such a big group of people on a similar topic, matching with my interests, was eventually enough to make me come back to Delft (thank you, Serge, for granting me this second (last) chance)! During the last three and a half years, I have worked, laughed, and talked with a lot of people that have inspired, helped, and supported me. Therefore, I would like to take this opportunity to thank everyone that has in any way been involved during my PhD.

Serge, thank you for the good discussions, your creative mind, and your trust. I hope we will continue to have these discussions in the future. Oded, thank you for being structured like me, you have helped me solve many problems. I appreciate that your door was always open for a quick question or a more thorough discussion on research or other things. Dorine, thank you for completing this supervising team with an out-of-the-box mentality. You have opened my mind multiple times for a broad variety of things. Thank you for being my guide during conferences and allowing me to treat you like my rubber duck (for debugging), in times I was confused. I’d like to thank my committee members for thoroughly reading, commenting on, and assessing my thesis. Thank you Bert, Elisabetta, Shlomo, Patricia, and Marjan.

Being part of the Allegro team has made my PhD a very good experience. We had a lot of the great coffee moments, outings, activities, conference visits, vrijdagmidagborrels, discussions, and fun. Thank you, Lara, Florian, Alexandra, Vincent, Marie-Jette, Alphonse, Giulia, Tim, Ilse, Xiaochen, Martijn, Yan, Yufei, Victor, Bernat, Winnie, Flurin, Oded, Dorine, and Serge. Allegro also resulted in a collaboration with Lara and Florian on an interesting project. A special thanks goes to some Allegro members. Lara, I want to thank you for being there when I needed, which already started when writing our Master thesis. I hope I have returned the favour. Your enthusiasm and creativity are much appreciated. Florian, thank you for the many discussions that have helped both our research progress. The supermarket breaks were always a welcome distraction. Tim, the car rides to Rotterdam are never boring. Alphonse, we had many great discussions on the train, car, bicycle, Felyx, and during drinks or dinner. Alexandra, thank you for sharing your happiness (and your cookies). Giulia, you are a great opponent during tennis, but a less good tennis partner. I highly appreciate your energy and mindset. Marie-Jette, I will never forget the conference dinner at hEART in Athens. Vincent, What? Your ability to listen in to conversations at the right time, keeps on surprising me.

Many other colleagues at the Department of Transport and Planning have made my PhD a more enjoyable experience during shared lunches, coffee breaks, discussions, and laughs. Thank you, Paul, Maria, Konstanze, Yihong, Pablo, Bahman, Nikola, Jishnu, Malvika, Xavi, Peter, Mignon, and Sanmay. Special thanks goes to Niharika, for helping me out when coding was difficult. Ding and Panchamy, thank you for allowing me to spend my breaks in your office (on the floor!), discussing whatever comes to mind. Edwin, I am glad I got to know you better. Thank you for the coffee with advice! Niels, the discussions we had about the future were a huge help, I hope we can continue our collaboration. Thank you Priscilla, for your help in organising many things related to the PhD, I appreciate spending my mini-breaks talking to you. Thank you, Moreen and Dehlaila.

Outside of the department there were more people that I would like to acknowledge. Sascha, thank you for allowing me to work with the MPN data. Our discussions were always very useful (with wine and without), both work related and otherwise. Shlomo, thank you for the many constructive Skype calls discussing work and future perspectives. I am glad you hosted me for a brief visit to Technion. I hope we can continue to collaborate in the future. Michel, thank you for hosting me at EPFL after the submission of my thesis. Thanks Tim, Mila, Janody, Rico, Meri, Stefano, and Gael for improving the quality of my stay in Lausanne. This visit helped me define my future plans. Ray, our Skype discussions helped me relativizing writing the thesis. I am glad we finally got to meet each other in person. Conchita, thank you for helping me in the process of printing my thesis. Jeroen, our meetings every once in a while were always good reflection on work, life, and PhD. Let’s continue this in the future.

My friends offered plenty of welcome distractions of PhD life. Thank you, Maxim, Marlon, Jorick, Niharika, Elisabeth, and Marisa for sharing weekends away, dinners, games, lunches, sports, drinks, laughs, tears, and talks about everything and nothing. I would like to thank my father, mother, and Milou, for supporting me in my decisions and cheering for me when needed. I might have not been around that much, but the phone calls, dinners, and visits were always much appreciated. The last person I want to thank is Rudolf. Nol, thank you for listening to me when I needed to talk, being there for me when needed, and sharing in this adventure. We have discovered a bit more of the world during and because of the PhD.

Danique Ton
August 2019

Summary

Due to increasing urbanisation rates worldwide combined with growing transportation demand, liveability of the urban environment is under pressure (UN, 2018). In response, many governments worldwide have set goals for increasing the share of trips made using sustainable modes of transport, such as walking and cycling. The use of active modes (i.e. walking and cycling) provides health benefits for individuals due to increased activity levels, and on a network level these modes (standalone or in combination with public transport) can potentially reduce traffic jams and the associated externalities (including air and noise pollution) when substituting the car. To achieve the desired increase in active mode shares, targeted policies need to be implemented. This requires a better understanding of who currently uses these modes, who could be persuaded to switch to active modes, and which determinants are driving active mode choice.

This intended change towards active modes requires an adequate representation of walking and cycling in the transportation planning models in order to assess the effect of active mode policies on modal shares and distribution over the network. However, this is often not the case. Moreover, integration of active modes in these models occurs very slowly. Walking and cycling are often missing in transportation planning models, treated as a 'rest' category, or combined into slow/active modes, all of which result in incorrect estimates of the active mode shares, making it impossible to correctly identify the impact of potential policy measures on active mode shares. Examples of these policy measures are introduction of new infrastructure or changes to existing infrastructure, which impact route choice and distribution over the network, and reimbursement of using the bicycle to go to work, which impacts the mode choice of individuals.

Investigating mode and route choice of active mode users increases the knowledge on active mode choice behaviour. By bridging this gap, the transportation planning models can potentially be improved. The objective of this thesis is '*to understand and model mode and route choice behaviour of active mode users'*. We identify six topics that are imperative to

achieve this objective that are related to mode choice, route choice, and the integration of both travel choices. First, we investigate the daily mobility patterns of individuals in relation to attitudes towards modes, because attitudes are considered to influence travel behaviour (Chapter 2). Afterwards, we zoom in on individual trips. We aim to understand which determinants drive the choice to walk or cycle (Chapter 3). In this topic we define the mode choice set as all feasible modes per individual and trip. However, not all feasible modes are used by individuals. Therefore, the third topic focuses on modes used over a long period of time, which we coin the experienced choice set. We investigate which determinants are relevant for including or excluding modes in this choice set (Chapter 4). Regarding cyclists' route choice, we investigate the determinants influencing this choice (Chapter 5). This research is based on the experienced choice set. Accordingly, we compare this method to frequently used choice set generation methods to identify the added value of the experienced choice set (Chapter 6). Finally, we perform a literature review on how mode and route choice can be modelled simultaneously (Chapter 7). The following paragraphs detail the findings of this research with respect to these six topics.

The relationship between daily mobility patterns and attitudes towards modes

The daily mobility pattern of individuals is investigated in relation to their attitude towards modes, which can represent their satisfaction with using a mode in terms of for example comfort, safety, and fun (Chapter 2). Data of the Netherlands Mobility Panel (MPN) is used to perform a latent class cluster analysis on the daily mobility patterns. We identify five classes: 1) car and bicycle users, 2) exclusive car users, 3) car, walk, and bicycle users, 4) public transport+ users, and 5) exclusive bicycle users.

We found that individuals are more positive towards the modes they use on a daily basis compared to unused modes, which results in significant differences in attitudes between classes. Individuals that perceive their used mode(s) most positive (consonant users) are potentially less inclined to switch modes. However, some individuals have a better perception of modes they currently do not use (dissonant users). The classes of exclusive car users and car and bicycle users have relatively high shares of dissonant users. These individuals can potentially be persuaded to change to other sustainable modes, for example via reimbursement by employers for cycling or public transport to work. The multimodal classes (1, 3, and 4) already incorporate active modes of transport, which could potentially be further increased. Finally, a large share of the exclusive bicycle users does not use their best perceived mode and 7% uses their least perceived mode. This can trigger an undesirable change in their mobility pattern, because the car is often perceived best by these dissonant users.

Determinants of the active mode choice behaviour

We zoom in on individual trips and investigate the determinants that are relevant for active mode choice. These relevant determinants have been identified by means of discrete choice modelling using data from the MPN (Chapter 3). Contrary to findings from literature, individual characteristics, specifically socio-demographics, and season and weather are of limited influence for active mode choice in the Netherlands. This might be due to respectively the very diverse cycling population and the relatively mild climate in the Netherlands. The most important categories of determinants for cycling are trip characteristics, built environment, and employment conditions. Being reimbursed by the employer for using the bicycle to go to work has a strong positive association with cycling in general. For walking the most important determinants are trip characteristics, built environment, and household characteristics. Both active modes are influenced by different determinants and if they are influenced by the same determinants, the impact of these determinants differs. Consequently, these modes should be considered independently. Policy measures should thus target either walking or cycling when the aim is to increase the modal share of either, but not both modes simultaneously.

The experienced mode choice set and its determinants

A feasible mode (like in Chapter 3) is not necessarily a used mode, because an individual might own a mode but not use it for a trip. Consequently, when the aim is a modal switch over an enduring period of time, it is instead more relevant to know the likelihood of including or excluding a mode in the mode choice set. We propose to evaluate this by investigating the experienced choice set, which is the set of modes used over a long period of time (Chapter 4). This choice set might differ for different trip purposes, therefore we focus on commuting trips. Many individuals only use one mode for their work trip (83.5%), suggesting habitual and/or captive behaviour, which will not be captured when specifying the feasible choice set. We estimate a discrete choice model to identify which determinants are relevant for the formation of the experienced choice set.

We find that the probability for including the bicycle in the experienced mode choice set increases for higher education, owning a bicycle, and being reimbursed by the employer for using the bicycle. It decreases for low urban density, working fulltime, or when the car or public transport is reimbursed by the employer. The probability for including walking (for the full commute) increases with the presence of children under the age of 12 in the household or when an individual lives in a one- or two person household. It decreases when the individual owns a bicycle, and when the individual is reimbursed for using the car. The inclusion of cycling in the mode choice set is thus affected by different determinants compared to walking. Furthermore, the inclusion of these modes in the choice set depends on more determinants than ownership and availability, which are generally used to identify the feasible choice set.

Determinants of cyclists' route choice

To improve the representation of cyclists' route choice in existing transportation planning models, it is important to know which determinants influence this choice. Using GPS data from the inner-city of Amsterdam, the relevant network-related and context determinants are identified (Chapter 5). Distance is valued negatively, which is in line with findings in literature. However, often the impact is higher elsewhere, potentially due to the mixed land use of Amsterdam. Furthermore, the number of intersections per kilometre is valued negatively and overlap between routes is valued positively. The share of cycle path has a different (positive) impact depending on the choice set identification method used. When using the experienced route choice set in the estimation of the route choice model, it is found to be insignificant. Because this method is based on observed routes of individuals, these routes are already optimised to a certain extent. It is likely that all routes include a relatively high share of cycle path, making them irrelevant for route choice. Furthermore, if this is not the case, the street design is such that it does not induce negative impacts for cyclists. Regarding the context determinants, we found that in the morning peak hour distance is valued more negatively compared to other times of the day. This might be due to scheduling constraints in the morning.

The added value of the experienced route choice set

The added value of the experienced route choice set, coined data-driven path identification approach (DDPI) is investigated in comparison to two frequently used choice set generation algorithms: breadth-first search on link-elimination and labelling (Chapter 6). The success of these two algorithms depends largely on the criteria used to generate routes (e.g. distance and/or share of cycle paths), the complexity of the network, and the quality of the network information that is available. If any of these criteria is insufficient, the resulting generated choice set is not fully able to reproduce the observed behaviour, as is the case for Amsterdam.

A route choice model was estimated using the choice sets resulting from these three methods, using the same set of network-related attributes. On the whole, the signs of the parameters of the route choice models are similar between route choice generation methods. A downside of the model using the experienced choice set is that it has lower parameter values than the models using the other methods, which is mostly due to the limited variability in the choice set, resulting in lower elasticities and model fit. Furthermore, it has a very low performance when predicting using out-of-sample data, suggesting that it is not suitable for prediction. A positive aspect of the model using the experienced choice set is that it offers an advantage in case the dependent variables of the choice set generation algorithms are of insufficient quality (criteria, network complexity, or network information). This is because the experienced choice set is able to provide behavioural insights, while it does not depend on any of these issues.

Integrating mode and route choice

Many theoretical frameworks, such as the four-step model, assume relationships between travel choices. This thesis provides evidence that this is also the case for mode and route choice, for example because they are influenced by several similar determinants. Thus, ideally these two choices should be modelled simultaneously. A literature review is performed to study how previous research has handled this integration, because it is yet unknown how this can be adequately addressed (Chapter 7). We focus on discrete choice models. Because only four studies investigate mode and route choice we broadened the scope to include other travel choice dimensions (trip chaining, destination choice, and departure time choice).

The literature study illustrates that very basic modelling structures are used in mode-route studies, namely Multinomial Logit and Nested Logit (NL). The first modelling structure assumes a fully simultaneous choice between mode and route, where each of the joint alternatives are independent. The second modelling structure is used to introduce correlation between modes, meaning that routes are substituted before modes when changes are introduced, which already increases the realism. However, these models do not account for overlap between routes. Consequently, several advancements are imperative to allow for the simultaneous modelling of mode and route choice.

Two requirements for integrating mode and route choice follow from the literature review. First, it is essential that overlap between routes can be accounted for (e.g. via Path-Size Logit). Substitution patterns can vary per person and per trip, as increasing evidence is found that decision-making is heterogeneous. Therefore, second, ideally the model structure incorporates a flexible correlation structure and is able to account for heterogeneity in the decision-making process. Several more advanced modelling structures are mentioned in literature, which could be applied to mode and route choice, such as Cross-Nested Logit, Probit, Mixed Logit (ML), and segmentation approaches. Currently only the segmentation approaches meet the latter of the two requirements. A combination of the segmentation approaches with ML or Probit could largely increase the behavioural realism of the modelled choice dimensions. A downside of these more behaviourally realistic models is that they are less applicable in practice, as increased complexity means reduced interpretability. This reduces the likelihood that these models are adopted in practice. Therefore, more research into behaviourally realistic and interpretable model structures is needed, to allow for adoption in practice.

Implications of this thesis

This thesis extents the body of knowledge on mode and route choice behaviour of active mode users. The conclusions of this thesis imply the following:

- Walking and cycling should be targeted separately via policy measures when the aim is to increase the modal share of either
- The specification of the mode and route choice sets is non-trivial and requires more emphasis, both in research and practice
    - The composition of the mode choice set depends on more determinants than previously assumed, as it does not only depend on ownership and availability, but also on socio-demographics, employment conditions and urban density
    - Walking and cycling should be included as distinct alternatives in the mode choice set, as their inclusion depends on different determinants
    - The choice set generation method influences the relevance of determinants in route choice modelling phase
- The choice set can be defined based on revealed behaviour, where it has most added value if insufficient information is available and when it is used in estimation of the choice set
- Policy measures that target a mode switch towards active modes might not reach certain users, because they are already satisfied with their current mobility pattern
- Both desired (from car to active modes) and undesired (from bicycle to other modes) mode switches are expected based on dissatisfaction with the current daily mobility pattern
- Mode and route choice are related, therefore these travel choices should be investigated and modelled simultaneously

Samenvatting

De leefbaarheid van de stedelijke omgeving staat onder druk, dit wordt veroorzaakt door de groeiende vraag naar vervoer in combinatie met de toenemende verstedelijking (UN, 2018). Wereldwijd hebben veel overheden daarom doelen gesteld om het aandeel verplaatsingen met duurzame modaliteiten, zoals lopen en fietsen, te vergroten. Het fysieke karakter van deze actieve modaliteiten biedt gezondheidsvoordelen voor het individu. Tevens kan het aantal files en de hoeveelheid lucht- en geluidsvervuiling worden verminderd wanneer men niet de auto nemen, maar er voor kiezen om te lopen of te fietsen, eventueel in combinatie met openbaar vervoer. Teneinde diverse gerichte beleidsmaatregelen te kunnen implementeren, is een beter begrip vereist van fietsers, voetgangers en de factoren die de keuze voor deze modaliteiten beïnvloeden.

Om het effect van gerichte beleidsmaatregelen te bepalen, dienen lopen en fietsen adequaat te worden opgenomen in transportmodellen. Vaak zijn deze modaliteiten hierin niet opgenomen, of worden ze gecombineerd (actief/traag/rest) opgenomen. Dit maakt het onmogelijk het effect van een beleidsmaatregel precies door te rekenen. De bouw van nieuwe infrastructuur of het introduceren van een vergoeding door de werkgever voor fietsen naar het werk zouden mogelijk goede beleidsmaatregelen kunnen zijn die invloed hebben op routekeuze en modaliteitskeuze.

Onderzoek naar modaliteits- en routekeuze voor lopen en fietsen vergroot de kennis omtrent het keuzegedrag van deze actieve modaliteiten. Met de overbrugging van dit kennishiaat kunnen transportmodellen mogelijk worden verbeterd. Het doel van deze thesis is '*het begrijpen en modelleren van modaliteits- en routekeuzegedrag van gebruikers van actieve modaliteiten*'. We identificeren zes onderwerpen die hiertoe zijn onderzocht, gerelateerd aan modaliteitskeuze, routekeuze en de integratie van beide keuzes. Eerst, hebben we onderzoek gedaan naar dagelijkse mobiliteitspatronen in relatie tot attitudes naar modaliteiten, omdat attitudes worden gezien als belangrijke voorspellers van reisgedrag (Hoofdstuk 2). Daarna zoomen we in op individuele verplaatsingen en onderzoeken we welke factoren bepalend zijn voor actieve modaliteitskeuze (Hoofdstuk 3). In dit onderzoek is de keuzeset gedefinieerd als de set van beschikbare modaliteiten per individu. Maar niet iedereen overweegt om alle beschikbare modaliteiten ook echt te gebruiken. Middels een longitudinaal onderzoek is vervolgens een set van gebruikte modaliteiten geidentificeerd die we de gebruikte keuzeset noemen. In het bijzonder is onderzocht welke factoren relevant zijn bij de formatie van deze keuzeset (Hoofdstuk 4). Tevens hebben we op basis van gebruikte routes onderzocht welke factoren de routekeuze van fietsers beïnvloedt (Hoofdstuk 5). Om de toegevoegde waarde te bepalen van een gebruikte keuzeset te kunnen bepalen, vergelijken we deze methode vervolgens met twee andere vaak gebruikte methodes om keuzesets te genereren (Hoofdstuk 6). Tot slot onderzoeken we middels een literatuuronderzoek hoe modaliteitskeuze en routekeuze tegelijkertijd kunnen worden gemodelleerd (Hoofdstuk 7). De volgende paragrafen beschrijven de belangrijkste bevindingen van deze thesis omtrent deze zes onderwerpen.

De relatie tussen dagelijkse mobiliteitspatronen en attitudes naar modaliteiten

Het dagelijkse mobiliteitspatroon van mensen is onderzocht in relatie tot hun attitude naar verschillende modaliteiten (Hoofdstuk 2). De attitude naar modaliteiten kan worden geïnterpreteerd als de mate van tevredenheid, gemeten in bijvoorbeeld: comfort, veiligheid en plezier van het gebruik van een modaliteit. Data van het Mobiliteitspanel Nederland (MPN) is gebruikt om een latente klasse clusteranalyse uit te voeren op de dagelijkse mobiliteitspatronen. We identificeren vijf klassen: 1) auto en fiets gebruikers, 2) alleen-auto gebruikers, 3) auto, loop en fiets gebruikers, 4) openbaar vervoer+ gebruikers en 5) alleen-fiets gebruikers.

Het onderzoek toont significante verschillen in attitudes van de vijf mobiliteitsklassen. Zo wordt bijvoorbeeld aangetoond dat mensen over het algemeen positiever zijn over modaliteiten die zij zelf gebruiken dan over modaliteiten die zij niet gebruikten. Mensen met een betere perceptie van de gebruikte modaliteit(en) (consonant gebruik) staan mogelijk minder open voor het wisselen van modaliteit. Sommige mensen hebben echter een betere perceptie van niet-gebruikte modaliteit(en) (dissonant gebruik). De alleen-auto gebruiker en auto en fiets gebruiker kennen een relatief hoog aandeel dissonante gebruikers. Deze mensen kunnen potentieel worden overtuigd te gaan lopen of fietsen, bijvoorbeeld middels een werkgeversvergoeding voor het lopen of fietsen naar werk. De mensen in multimodale klassen (1, 3 en 4) fietsen en/of lopen al, maar zouden dit aandeel kunnen vergroten. Tot slot is een groot deel van de alleen-fiets gebruikers *niet* het meest positief over de fiets, 7% is zelfs het meest ontevreden over de fiets. Omdat de auto binnen deze groep vaak de voorkeur geniet, kan dit leiden tot een onwenselijke verandering van het mobiliteitspatroon.

Drijfveren van lopen en fietsen

We zoomen in op individuele verplaatsingen en onderzoeken welke factoren bepalend zijn voor actieve modaliteitskeuze. Door middel van discrete keuzemodellen zijn de relevante factoren voor lopen en fietsen geïdentificeerd met data van het MPN (Hoofdstuk 3). In tegenstelling tot bevindingen in andere onderzoeken, zijn het weer, het seizoen en sociaal-demografische kenmerken in dit onderzoek weinig bepalend gebleken voor de actieve mobiliteitskeuze. Mogelijk is het relatief milde klimaat in Nederland en het algemene fietsgedrag van Nederlanders hier de reden van. De keuze voor fietsen wordt het meest beïnvloed door kenmerken van de verplaatsing, de omgeving en de werkcondities. Zo heeft het krijgen van een vergoeding van de werkgever om te fietsen naar werk, een sterke positieve invloed heeft op fietsen in het algemeen. Voor lopen geldt dat de kenmerken van de verplaatsing, omgeving en het huishouden het belangrijkst zijn. De verschillende modaliteiten worden beïnvloed door verschillende factoren, daarnaast verschilt het belang dat aan factoren wordt gehecht per modaliteit. Dit bevestigd dat lopen en fietsen onafhankelijke modaliteiten zijn, die onafhankelijk van elkaar onderzocht moeten worden. Effectieve beleidsmaatregelen teneinde een toename van het aandeel verplaatsingen van voetgangers of fietsers, dienen gericht te zijn op lopen of op fietsen maar niet op allebei tegelijk.

Invloedfactoren van de gebruikte modaliteitskeuzeset

Een beschikbare modaliteit (Hoofdstuk 3) is niet per se een gekozen modaliteit, aangezien iemand een modaliteit wel kan bezitten maar niet hoeft te gebruiken voor een verplaatsing. Als het doel is om een modaliteitsverandering te realiseren is het relevanter om te weten wat de kans is dat een modaliteit wordt opgenomen in de modaliteitskeuzeset, dan de keuze per verplaatsing. We onderzoeken dit middels de gebruikte keuzeset, die we definieren als de set van gebruikte modaliteiten gedurende een lange periode (Hoofdstuk 4). Omdat deze keuzeset zou kunnen verschillen per reismotief, focussen wij op de woon-werkverplaatsing. Veel mensen gebruiken maar één modaliteit voor hun woon-werkverplaatsing (83,5%). Dit lijkt op een gewoonte of het niet hebben van alternatieven (gevangen gebruiker), wat niet kan worden gevangen met de beschikbare keuzeset. We schatten een discreet keuzemodel om te identificeren welke factoren relevant zijn voor de formatie van de gebruikte keuzeset.

De kans dat de fiets wordt opgenomen in de keuzeset neemt toe wanneer iemand een hoog opleidingsniveau heeft, zelf een fiets heeft en wanneer iemand een vergoeding krijgt van de werkgever voor het gebruiken van de fiets voor de woon-werk verplaatsing. Deze kans wordt echter kleiner wanneer iemmand in laag-stedelijk gebied woont, fulltime werkt, of de werkgever het gebruik van de auto of het openbaar vervoer vergoedt. De kans dat lopen onderdeel uitmaakt van de keuzeset, neemt toe wanneer kinderen jonger dan 12 jaar aanwezig zijn in het huishouden en wanneer iemand in een een of twee persoons-huishouden woont. Deze kans neemt af wanneer iemand zelf een fiets heeft of wanneer iemand een vergoeding ontvangt voor het gebruiken van de auto voor de woon-werk verplaatsing. Kortom, het opnemen van de fiets in de modaliteitskeuzeset wordt beïnvloed door andere factoren dan het opnemen van lopen. Daarnaast zien we dat meer factoren dan alleen eigenaarschap en beschikbaarheid van modaliteiten (beschikbare keuzeset) relevant zijn voor de gebruikte keuzeset.

Drijfveren van routekeuze van fietsers

Om de weergave van de routekeuze van fietsers in de bestaande transportmodellen te verbeteren is het belangrijk om te weten welke factoren relevant zijn voor deze keuze. Met gps-data van fietsroutes in het centrum van Amsterdam hebben we onderzocht welke netwerk- en contextfactoren belangrijk zijn (Hoofdstuk 5). Afstand heeft een negatieve relatie met routekeuze. De impact van afstand is echter kleiner dan in andere onderzoeken, mogelijk in verband met het gemengde ruimtegebruik in Amsterdam. Ook het aantal kruispunten per kilometer beïnvloedt routekeuze negatief, terwijl overlap van routes een positieve relatie heeft. De impact van het percentage fietspad is altijd positief maar verschilt per keuzesetgeneratie methode. Als de gebruikte routekeuzeset wordt gebruikt is deze factor niet significant. Waarschijnlijk omdat deze methode is gebaseerd op geobserveerd gedrag, wat betekent dat routes al tot op zekere hoogte geoptimaliseerd zijn. Het lijkt erop dat alle routes grotendeels via fietspaden gaan, waardoor deze factor onbelangrijk is. Ook zorgt het straatontwerp in Nederland ervoor dat er geen effect wordt gevonden van de aan- of afwezigheid van een fietspad. In de ochtendspits (contextfactor) wordt afstand negatiever gezien dan tijdens de rest van de dag, mogelijk door afspraken met vaste tijden in de ochtend.

De toegevoegde waarde van de gebruikte routekeuzeset

De toegevoegde waarde van de gebruikte routekeuzeset, ook wel de *data-driven path identification* methode (DDPI) genoemd, is onderzocht in vergelijking met twee vaak gebruikte keuzesetgeneratie algoritmes: de *breadth-first search on link-elimination* en *labelling*

(Hoofdstuk 6). Het succes van deze twee methodes hangt af van de criteria die gebruikt zijn bij het genereren van routes (bijvoorbeeld afstand of percentage fietspad). Daarnaast hangt het af van de complexiteit van het netwerk en de kwaliteit van de netwerkinformatie die beschikbaar is. Als een van deze criteria niet voldoet zullen de gegenereerde routes niet het geobserveerde gedrag reproduceren. Voor het network in Amsterdam waren deze methodes niet succesvol, doordat de kwaliteit van het network te laag was.

De drie gegenereerde keuzesets zijn gebruikt om routekeuzemodellen te schatten met dezelfde netwerkfactoren. De richting van de parameters in de routekeuzemodellen zijn gelijk bij gebruik van de verschillende keuzesets. Een nadeel van het model op basis van de gebruikte keuzeset is dat de parameterwaarden lager zijn, vooral doordat deze methode weinig variabiliteit kent in de netwerkfactoren. Dit resulteert in lagere elasticiteiten en lagere model fit. Daarnaast scoort de gebruikte keuzeset slecht in het validatieproces, wat indiceert dat het geen bruikbare methode is om mee te voorspellen. Een voordeel van de gebruikte keuzeset is dat het inzicht biedt in de voorkeuren van fietsers, wanneer netwerkcomplexiteit en netwerkinformatie een te laag kwaliteitsniveau hebben. In tegenstelling tot veelgebruikte keuzesetgeneratie algoritmes, is deze methode hier niet van afhankelijk.

Integratie van modaliteits- en routekeuze

Veel theoretische raamwerken nemen aan dat modaliteits- en routekeuze gerelateerd zijn. Deze thesis ondersteund deze aanname. Zo hebben wij een aantal factoren geïdentificeerd die zowel de modaliteits-, als de routekeuze bëinvloeden. Idealiter worden deze keuzes daarom tegelijkertijd worden gemodelleerd. Omdat nog niet duidelijk is hoe dit op een adequate manier kan worden gedaan, hebben wij een literatuuronderzoek uitgevoerd naar wijzen waarop deze integratie kan worden behandeld (Hoofdstuk 7). We focussen op de discrete keuzemodellen. Omdat er slechts vier studies zijn gevonden omtrent modaliteits- en routekeuze, hebben we ook andere reiskeuzes opgenomen (tourvorming, bestemmingskeuze en vertrektijdkeuze).

Het literatuuronderzoek toont aan dat simpele modelstructuren zijn gebruikt om modaliteits- en routekeuze te modelleren, namelijk Multinomial Logit en Nested Logit (NL). De eerste modelstructuur neemt aan dat keuzes volledige tegelijkertijd worden gemaakt, waarbij elk gecombineerd alternatief van een modaliteit en een route onafhankelijk is van de rest. De tweede structuur is gebruikt om correlatie tussen modaliteiten toe te laten, wat betekent dat eerder tussen routes wordt gewisseld dan tussen modaliteiten. Deze modellen houden echter geen rekening met overlappende routes. Daarom zijn een aantal verbeteringen in de huidige modellen noodzakelijk zijn om modaliteits- en routekeuze tegelijkertijd te kunnen modelleren.

Twee aspecten zijn belangrijk gebleken voor de mogelijke integratie tussen modaliteits- en routekeuze. Het is belangrijk dat overlap tussen routes wordt opgenomen (bijvoorbeeld via Path-Size Logit). Daarnaast is het belangrijk dat de modelstructuur in staat is om flexibele correlatie structuren op te nemen en om heterogeniteit in de beslissingsstructuur op te nemen. In recent onderzoek wordt steeds vaker aangetoond dat de beslissingsstructuur heterogeen is (verschillende substitutiepatronen per individu). Een aantal geavanceerdere modellen die worden genoemd in de literatuur kunnen worden toegepast op modaliteits- en routekeuze, zoals Cross-Nested Logit, Probit, Mixed Logit (ML) en segmentatie methoden. Op het moment voldoen alleen de segmentatie methoden aan de laatste eis. Modaliteits- en routekeuze in modellen kan realistischer worden via een combinatie van de segmentatie methoden met ML of Probit. Een nadeel van deze modelstructuren is dat deze minder makkelijk toepasbaar zijn in de praktijk, omdat toename in complexiteit ten koste gaat van interpretatie. Dit verlaagt de kans dat deze modellen worden geïmplementeerd in de praktijk. Er is meer onderzoek nodig naar modellen die zowel realistisch als interpreteerbaar zijn, om implementatie in de praktijk mogelijk te maken.

Implicaties van deze thesis

Deze thesis verrijkt de kennis over modaliteits- en routekeuzegedrag van gebruikers van actieve modaliteiten. De conclusies in deze thesis leiden tot de volgende implicaties:

- Lopen en fietsen moeten apart worden opgenomen in beleidsmaatregelen wanneer het doel is om een toename van het aandeel verplaatsingen van één van beide te realiseren
- De specificatie van de modaliteits- en routekeuzesets is niet-triviaal en vereist meer aandacht, zowel in wetenschap als praktijk
    - De samenstelling van de modaliteitskeuzeset is afhankelijk van meer factoren dan voorheen gedacht. Het is niet alleen afhankelijk van eigenaarschap en beschikbaarheid van modaliteiten, maar ook van sociaal-demografische en werk gerelateerde factoren en de stedelijkheidsgraad van de woonplaats
    - Lopen en fietsen moeten als losse alternatieven worden opgenomen in de modaliteitskeuzeset, de opname van deze modaliteiten hangt af van andere factoren
    - De keuzeset generatie methode beïnvloedt de impact van factoren in routekeuzemodellen
- De keuzeset kan worden gespecificeerd op basis van geobserveerd gedrag. Vooral wanneer te weinig informatie beschikbaar is en wanneer het keuzesets worden geschat heeft het speciferen op basis van geobserveerd gedrag toegevoegde waarde.
- Beleidsmaatregelen die zich richten op een modaliteitsverandering zullen mogelijk niet de consonante reiziger bereiken, omdat deze tevreden is met het huidige mobiliteitspatroon
- Zowel gewenste (van auto naar actieve modaliteiten) als ongewenste (van fiets naar andere modaliteiten) modaliteitsveranderingen worden verwacht op basis van de ontevredenheid met het huidige mobiliteitspatroon
- Modaliteits- en routekeuze zijn gerelateerd, daarom worden deze keuzes idealiter gezamenlijk onderzocht en gemodelleerd

Content

Chapter 1 – Introduction

Worldwide, urbanisation rates increased from approximately 34% in 1960 to 55% in 2017 (TheWorldbank, 2018). The urban population is not equally distributed over the world, as for example North America and Europe have a relatively high share of the population living in urban environments (respectively 82% and 74%), while Africa has a much lower share (43%). The UN predicts that 68% of the world population will be living in urban environments by 2050 (UN, 2018). This trend affects the liveability of the urban environment, as growing population also generally means growing transportation demand. Therefore, it is causing issues related to increasing congestion and resulting emissions.

As a result of increasing urbanisation, governments worldwide aim to sustain or increase the liveability of urban environments by focusing on sustainable modes of transport, like active modes (i.e. walking and cycling). Due to the physical activity required for using these modes, they are known to benefit the health of individuals. Furthermore, if active mode travel, for example in combination with public transport, replaces car travel, congestion and emissions (including noise) can be reduced. As an example, the Pan-European region aims to double cycling in the region by 2030 and increase it in every country (UNECE, 2018). The individual countries have varying goals that help in achieving the aim to double cycling levels of the entire region, where for example France aims to increase from 3% in 2012 to 10% in 2020 and the UK aims to double their cycling share to 4% by 2025 (ECF, 2019).

Currently, active mode shares vary largely across countries and also within countries, where generally active mode use in urban environments is higher than in rural areas (Heinen et al., 2010). Several countries have achieved relatively high cycling trip shares, such as the Netherlands (27%), Denmark (18%), and Germany (10%). However, large variations are observed between cities in these countries. For example, in the Netherlands Groningen has a larger bicycle trip share than Rotterdam (39% vs 16%), while in Germany Muenster has a larger share than Wiesbaden (27% vs 3%). Other countries, such as the USA, Australia, UK, and Canada have a very low cycling share (1-2%). Cities in these countries also show variation in the cycling share, however to a lesser extent compared to the cycling rich countries (Pucher and

Buehler, 2008). Consequently, the penetration rate of active modes is at different levels of advancement across the world.

To achieve the desired increase in active mode shares, policies need to be implemented that aim at increasing active mode use. This requires a better understanding of who could be persuaded to cycle or walk and which determinants are relevant for choosing an active mode. Furthermore, additional requirements on the existing (and future) built environment and infrastructure are imposed, as the utilisation will change due to an increase in active mode use.

Traditionally, transportation planners focus on motorised traffic. Due to for example, the size of infrastructure investments, the size of the time and thus economic losses caused by traffic jams, and the impact on traffic safety, the motorised modes were the logical point of attention. Their models are often based on either the four-step model, for example the Swedish national model (Beser and Algers, 2002) and the Dutch national model (Hofman, 2002; van Cranenburgh and Chorus, 2017) or the activity-based approach, for example the Tel Aviv model (Shiftan and Ben-Akiva, 2011) and the Portland model (Bowman et al., 1998), visualised in Figure 1.1. Due to the fact that shares of cycling and walking (standalone or in combination with public transport) are also related to the shares of the car and public transport, one would expect incorporation of active mode behaviour in the transportation planning models. Unfortunately, this is currently not the case and integration occurs very slowly. Active mode choice behaviour is often missing, treated as a rest category, or combined into slow/active modes, which results in incorrect estimates for the active modes and it makes it impossible to derive the impact of potential policy measures (De Jong et al., 2007).

One could agree that one of the reasons why active modes have not been incorporated correctly into these models, is that data and thus information and knowledge on these modes, has been scarce. In recent years, developments in large-scale data collection tools, such as for example Wi-Fi and GPS, together with technological advancements such as smartphones, that enable GPS or data collection applications, have started to make it possible to collect (on a larger scale) revealed preference data concerning pedestrians and cyclists. In research, these tools are now increasingly explored for data collection of active modes, therefore increasing the knowledge on active mode behaviour.

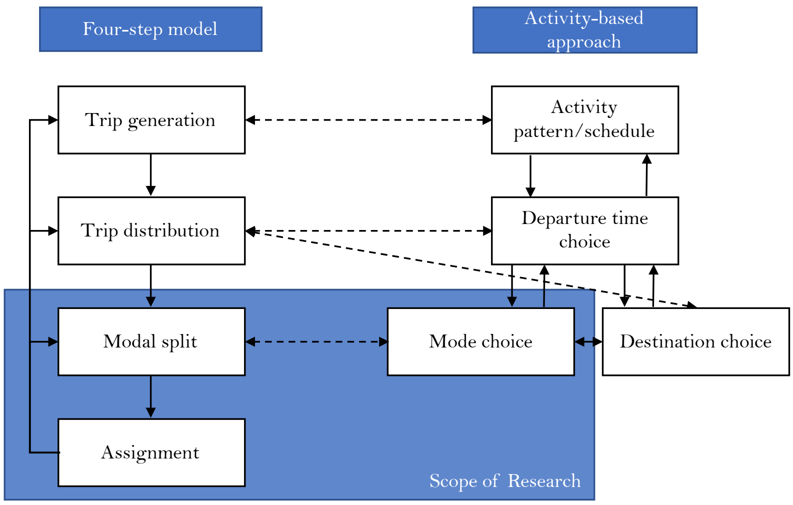


Figure 1.1: Conceptual frameworks of transport modelling and the scope of this thesis

The modules of the four-step model and activity-based approach that are considered most relevant for active modes are mode choice (modal split) and route choice (assignment), because these approaches assume that departure time choice, destination choice, and activity scheduling take place before choosing a mode and route. Some studies have investigated the preceding choices, for example destination choice for active modes (e.g. Borgers and Timmermans, 1986; Hoogendoorn and Bovy, 2004) however the vast majority of research into active modes focusses on mode and route choice (Duives et al., 2018). Because of the aforementioned future goals for active mode shares and current issues in transportation modelling, this thesis aims to increase the knowledge of mode and route choice behaviour of active mode users (see Figure 1.1).

From a scientific point of view, walking and cycling are very different modes from the motorised modes that traditionally have received attention from research and practitioners alike. Both active modes have more degrees of freedom regarding their movement, e.g. a pedestrian can make a 180-degrees turn and make use of infrastructure and non-infrastructure (e.g. grass fields) alike, whereas a cyclist is less flexible compared to the pedestrian and more flexible than motorised traffic. Next to that, both walking and cycling require effort from the individual to move, making it again different from motorised modes. Consequently, it is expected that the behaviour of active mode users is different (more complex) from the behaviour of motorised modes and also that it is driven by different factors.

The bicycle is comparable, in terms of speed and distance travelled, to motorised modes (public transport and car) in the urban environment. Walking, however, fulfils a different function compared to cycling, as it is mostly used locally. This has resulted in a different research focus for each active mode. For cycling, the mode and route choices described above are very relevant. The potential of the bicycle as competitor for motorised modes in urban environments, has resulted in increased investment budgets from governments (for example in the UK (Department for Transport, 2017)). Walking is mostly investigated on a local scale, e.g. an event terrain, city centre, or train station. This different scope for walking also results in investments taking place on a more local level. However, when policies are designed and investments plans are made, often walking and cycling are combined, for example in the UK (Department for Transport, 2017) and Australia and New Zealand (CWANZ, 2018). To understand whether it is valid to combine walking and cycling in the context of mode and route choices, more research is needed.

In this thesis, all modes (car, public transport, bicycle, and on foot) are investigated regarding mode choice, but the focus lies with better understanding the choice for walking and cycling. Due to the aforementioned scope differences between walking and cycling, this research addresses the route choice of cyclists only. The larger scale and competition with motorised modes, make this a more stringent topic to solve for research and practice alike. Various different types of bicycles are present, of which the normal bicycle and electric bicycle form the largest shares. In this research, the focus lies with normal bicycles, as these form the majority of the fleet.

1.1. State-of-the-Art in Mode and Route Choice of Active Mode Users

Before presenting in detail the research objective and questions, first the current state-of-the-art in active mode choice research and cyclists' route choice is briefly elaborated upon. This overview of the literature helps in identifying the research gaps that are relevant for this thesis.

1.1.1. Active Mode Choice

The growing interest of governments worldwide towards active modes, has led to a significant increase in research on active mode choice. Studies aim to identify what makes people walk and cycle, so that policy measures might be derived. To investigate this, both stated preference and revealed preference studies are conducted. The stated preference studies are either interested in the effect of currently non-existing features on mode choice or want to capture attitudes and preferences of individuals or both. The evaluation of non-existing features generally takes place in environments where no or limited investments have been made to the active mode infrastructure, which is correlated with the presence of active mode use. Features of interest are bicycle path and side walk designs (dell'Olio et al., 2013; Kamargianni et al., 2015; Wardman et al., 2007). Attitudes and preferences are supposed to be strong predictors of behaviour (Ajzen, 1991). Consequently, many studies aim to understand the choice for an active mode by relating this to attitudes or perceptions, so as to investigate potential willingness to cycle or walk (Fernández-Heredia et al., 2014; Lindelöw et al., 2014; Motoaki and Daziano, 2015). The revealed preference studies examine the current behaviour and determinants of this behaviour, which can be categorised into socio-demographics, social surroundings, trip characteristics, built environment, and employment conditions (Heinen et al., 2013; Maley and Weinberger, 2011). Some of the revealed preference studies also address the perceptions of individuals but relate this to actual trips rather than hypothetical situations (Muñoz et al., 2016b; Sigurdardottir et al., 2013).

Many of these studies take place in environments where active mode use is rare, such as the USA (Motoaki and Daziano, 2015), Spain (dell'Olio et al., 2013; Muñoz et al., 2016b), or Cyprus (Kamargianni et al., 2015). A very limited number of studies on active mode choice originate from countries with high active mode use, such as the Netherlands (Heinen et al., 2013) and Denmark (Sigurdardottir et al., 2013). In the editorial related to the special issue 'Cycling as Transport' in Transport Reviews, Fishman (2016) states that the Dutch are blind to cycling, as it is such an ordinary activity that it is not warranted much attention. The situation in the Netherlands, and other cycling rich environments, is rather extraordinary, as the cycling culture has long been established, the environment is safe for cycling, the cycling population is very diverse, and infrastructure is well-connected. However, Fishman (2016) states that not much is known yet about active mode choice in such environments. Consequently, there is a need to understand the determinants that influence active mode choice in environments where active modes are dominantly present.

Studies investigating active mode choice often apply discrete choice models. In this framework the choice between several alternatives is modelled, where the alternatives need to be identified by the researcher (i.e. the choice set). In case of stated preference data, the alternatives are decided on beforehand. However, in case of revealed preference data, assumptions need to be made regarding the choice set of each individual. In the literature a variety of methods is employed to deal with the choice set specification. Many studies investigate this on a binary level, which translates to walking or not (Maley and Weinberger, 2011; Rodriguez and Vogt, 2009), or cycling or not (Emond and Handy, 2012; Heinen et al., 2013; Motoaki and Daziano, 2015). This approach avoids the specification of the mode choice set, as all other modes are combined and used as a reference. Other studies have incorporated multiple modes when investigating active mode choice, requiring them to identify the choice set. In this situation a variety of methods has been applied, for example including all modes for everyone (Wardman et al., 2007), including only individuals that live within a certain distance from the destination of interest, such as a school, to make sure all modes are available (Kamargianni et al., 2015; Kamargianni and Polydoropoulou, 2013), using logical constraints related to availability of private modes and maximum distance/travel time covered by certain modes (Gehrke and Clifton, 2014), or using a probabilistic method to introduce latent availability and consideration of modes (Calastri et al., 2017). Different compositions of the choice set is known to impact model estimation and is thus also consequential for the results of potential policy measures

(Cantillo and de Dios Ortúzar, 2005; Swait and Ben-Akiva, 1987a). Hence, more knowledge is needed concerning the formation of the mode choice set.

1.1.2. Cyclists' Route Choice

A combination of growing interest towards cycling by governments worldwide and developments in recent years on large-scale data collection methods, such as GPS, have resulted in a significant increase in cyclists' route choice research in the last couple of years. At the start of the 21st century, research into route choice of cyclists was still mainly done using stated preference surveys, due to absence of these large-scale data collection methods (Hunt and Abraham, 2007; Sener et al., 2009; Stinson and Bhat, 2003). Since 2010, most studies have collected and used revealed preference data, with GPS-data being most commonly used (Bernardi et al., 2018; Ghanayim and Bekhor, 2018; Li et al., 2017; Zimmermann et al., 2017), while some still use stated preference studies, for example to identify attitudes or to measure other aspects that are not directly observable (Motoaki and Daziano, 2015; van Overdijk et al., 2017).

Many of these studies take place in environments where cycling is relatively uncommon, such as the USA (Chen et al., 2018; Hood et al., 2011; Khatri et al., 2016), Canada (Casello and Usyukov, 2014; Li et al., 2017), Brazil (González et al., 2016), Switzerland (Menghini et al., 2010; Montini et al., 2017), and Israel (Ghanayim and Bekhor, 2018), where cycling trip shares range from 1% to 6% (Pucher and Buehler, 2008). When using revealed preference data in these situations, extra care needs to be taken regarding representativeness of the data, which in turn influences the potential effect of measures that aim for increasing the cycling share. Several very recent studies investigate route choice in a cycling-rich context, such as in Denmark (Prato et al., 2018; Skov-Petersen et al., 2018) and the Netherlands (Bernardi et al., 2018). The aim of cyclists' route choice research differs between the cycling-rich and low-cycling contexts. In the latter, the aim is to identify determinants of cyclists' route choice, such that substantiated investments can be made regarding cycling infrastructure, whereas in the high-cycling these determinants are identified to investigate how to influence individuals' route choice, such that for example bicycle traffic jams do not occur.

Interestingly, research from countries with a large cycling share, is sparse (and all dating from 2018). Furthermore, cycling is not realistically incorporated in many transport planning models used in practice (e.g. De Jong et al., 2007). It can be altogether absent, be used as the 'rest' category, treated similar to driving, or very simple assignment procedures are applied, such as the all-or-nothing assignment. Consequently, also in countries that have a high share of cycling trips, more knowledge is required on the determinants of cyclists' route choice.

Cyclists' route choice modelling is mostly done using the discrete choice modelling framework. Various methods have been proposed for identifying the route choice set, which can be distributed into roughly four categories: deterministic methods, constrained enumeration methods, stochastic methods and probabilistic methods (Bovy, 2009; Prato, 2009). These are all path-based methods, which are most often applied in the cycling route choice context. An overview of the methods, based on when they were first introduced in route choice modelling, is presented in Figure 1.2.

Most choice set generation methods belong to the deterministic methods, which are based on repeated shortest path searches in the network. These methods differ in the way they compute the choice set, by means of alteration of different input variables such as search criteria, route constraints and link impedance (Prato, 2009). This category of methods is computationally attractive due to the efficiency of shortest path algorithms. Within this category four main groups of methods can be identified: shortest paths, labelling, link penalty and link elimination. The stochastic methods are also based on repeated shortest path searches in the network, but additionally the computation of optimal paths is randomised based on link impedances or individual preferences from probability distributions. Stochastic methods are mostly simulation based. The probabilistic methods generate a probability for each alternative in the choice set. This means that the computational complexity for these methods is much higher than for the other methods. The application of a full probabilistic method on a complex network is therefore prohibitive (Prato, 2009). The constrained enumeration methods are not only based on shortest routes, but rely on the assumption that individuals choose alternatives according to behavioural rules (Prato, 2009).

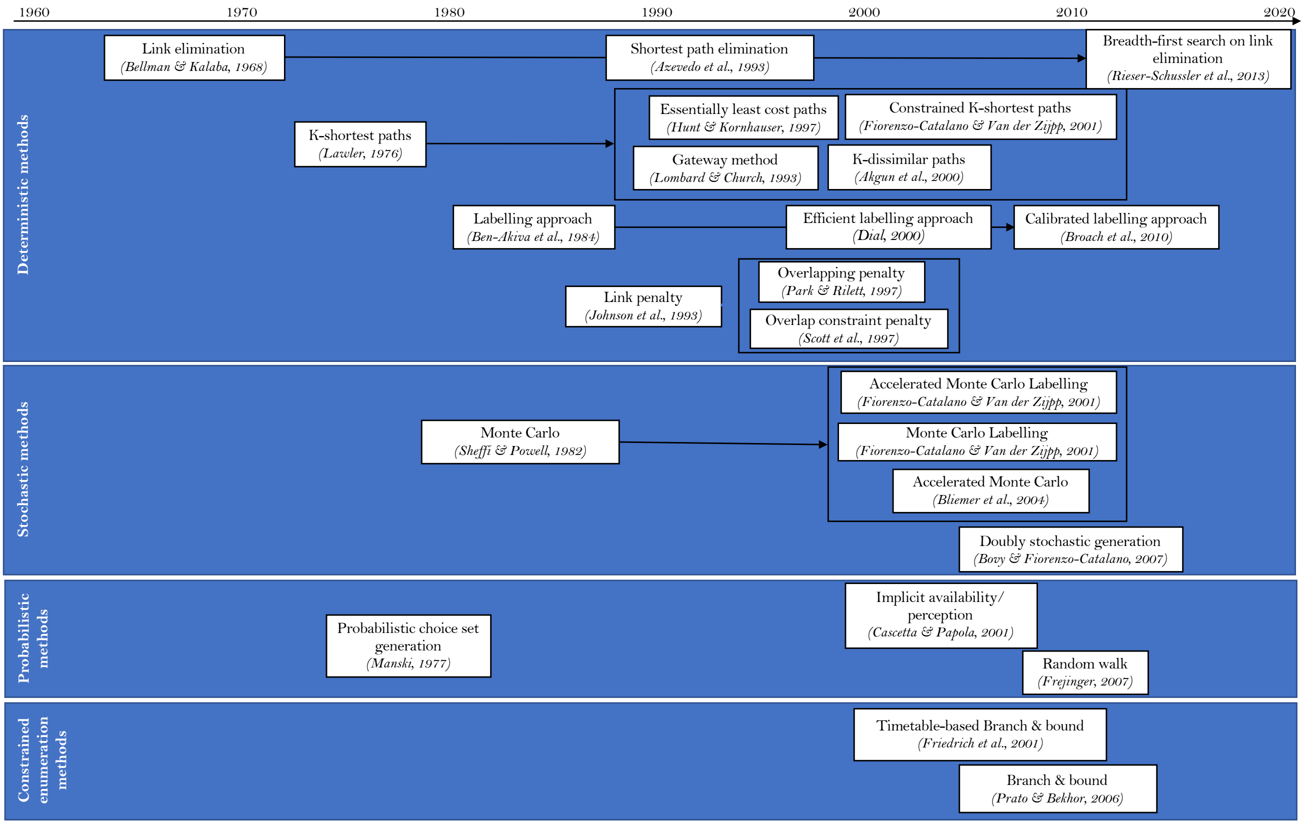


Figure 1.2: Genealogy of route choice set generation algorithms

In the cycling route choice context, a variety of these methods have been applied, where the deterministic methods are prevalent. Labelling is most common, followed by link elimination (mostly the breadth-first search on link elimination), shortest path, and link penalty. Furthermore, the doubly constrained stochastic generation and the branch and bound method are sometimes employed. Recently, a link-based approach has been introduced, which does not depend on identification of the choice set (Fosgerau et al., 2013). This method was applied by Zimmerman et al. (2017) for cyclists' route choice. This method does not use any of the abovementioned discrete choice modelling structures, instead a specialised model, Recursive Logit, is introduced.

Each of the choice set generation methods is prone to include irrelevant routes in the choice set (false positive). Furthermore, not all methods are equally capable of generating the observed route in the choice set, which results with falsely excluding routes from the choice set. This can also be due to not including the right optimisation criteria or a combination of these, which leads to the realisation that bicycle route choice is much more complex than car route choice (for which most methods were developed). In most cases, the observed routes are added to the choice set, however this introduces issues with endogeneity (as observed behaviour is added to the choice set). Consequently, current methods for choice set generation might not suffice for bicycle route choice modelling.

1.2. Research Objective and Questions

Due to increasing interest towards active modes from governments worldwide and advancements in large-scale revealed preference data that benefit active modes, an increase in active mode research, especially related to the choice to use an active mode and route choice related to cycling, can be observed in the last decade. The overview of the current state-of-the art shows that several aspects related to these choices are currently still unknown. These gaps need to be investigated, before further steps in active mode research can be taken. All these issues combined result with the following objective for this thesis:

'To understand and model mode and route choice behaviour of active mode users'

Six research questions are proposed to reach the objective of this thesis. First of all, given the aim to understand behaviour of active mode users, it is necessary to first know who these active mode users and potential active mode users are. This can be evaluated by investigating current behaviour and attitudes towards modes, where the latter reflects satisfaction with mode use. This results in the first research question:

1. What are the mobility patterns and attitudes towards modes of active mode users?

There is a need to model and understand active mode choice. According to the research gaps identified, it is not yet known which factors drive active mode choice in contexts where active modes are dominantly present. Next to that, the knowledge on how the choice set is build up is lacking. This can be explored by looking at the mode choice set which is experienced by individuals. This leads to the following research questions:

2. Which determinants influence active mode choice of individuals in an environment where active modes are dominantly present?
3. What are the determinants of the size and composition of the experienced mode choice set?

Once someone decides to use an active mode, in this case the bicycle, a route needs to be chosen to get from origin to destination. Here, limited knowledge is available regarding which determinants drive this choice in an environment where the bicycle is dominantly present. Furthermore, the choice set generation methodologies that are often used, exhibit several shortcomings. Thanks to the large amount of new large-scale data available, it is possible to infer the choice set from observed data: the experienced choice set. This method needs to be evaluated against the currently used methods, to identify the added value. This leads to the following research questions:

4. Which determinants influence cyclists' route choice behaviour in in an environment where active modes are dominantly present?
5. What is the added value of the experienced route choice set in comparison to frequently used choice set generation algorithms?

As shown in Figure 1, a relationship is assumed in the four-step and activity-based approaches between mode and route choice. Bhat (1998a) mentions that there are several arguments for simultaneously modelling multiple travel choices. One of these arguments is that the choices are influenced by the same determinants. Based on the literature and findings in this thesis, this is the case for (active) mode and route choice. This means that these choices influence each other to a certain extent and impacts of determinants might be different when incorporating both travel choices in a choice model. Consequently, ideally, they should be modelled simultaneously. However, currently it is still uncertain how this can be done best, which leads to the last research question:

6. What are the approaches towards integrating mode and route choice decisions into a single behavioural model?

Due to existing knowledge gaps regarding the individual choices (research question 1-5), these have to be first solved before the direction of simultaneous modelling can be explored. Consequently, this thesis focuses mainly on the individual choices, but investigates directions that can lead to combining these travel choice dimensions. The approach towards answering each of these research questions and reach the overall objective of this thesis is described in Section 1.3.

1.3. Research Approach

This section describes the approach towards reaching the objective of this thesis and answering the research questions. Figure 1.3 shows an overview of this approach, split up into the relevant body of the literature that is reviewed, data that is collected and processed, methodology that is used, and the analyses that are conducted.

To answer *research question 1*, literature is explored regarding daily mobility patterns of individuals and the methods used to investigate these in combination with how attitudes towards modes relate to (daily) mobility patterns. Ajzen (1991) states that attitudes are strong predictors of behaviour, therefore potential active mode users can be identified by relating these two aspects. The data used for this research is the census data from the Netherlands Mobility Panel (MPN), consisting of a three-day travel diary and personal and household surveys. In addition, within the Allegro-project (see details in section 1.5) an extra survey was developed and distributed among the MPN panel members, which investigated perception, attitudes, and wayfinding strategies of active mode users (coined PAW-AM). The daily mobility pattern data, in terms of average trips made per mode per day, is deduced from the travel diary. This is clustered into different classes using a latent class cluster analysis. To compare the attitudes towards modes with these classes, factor analysis is applied to reduce dimensionality. The resulting attitudes and mobility pattern classes are analysed using statistics.

The determinants of active mode choice *(research question 2)* are identified using discrete choice models. All major modes are considered in this research, to allow the comparison between the determinants that are relevant for walking and/or cycling and other modes (public transport and car). The determinants that are found to be relevant in literature are identified and methods for composing the choice set are evaluated. From the MPN data (both travel diary and surveys) the trips and potential determinants are extracted. In this investigation the e-bike and normal bicycle are grouped together, due to similar behaviour found for both modes (e.g. in distance travelled). For the non-chosen modes in the choice set, information on the determinants is extracted from both the MPN data and Google Directions API. Several existing model structures are tested to identify a model with behavioural realism and good explanatory power.

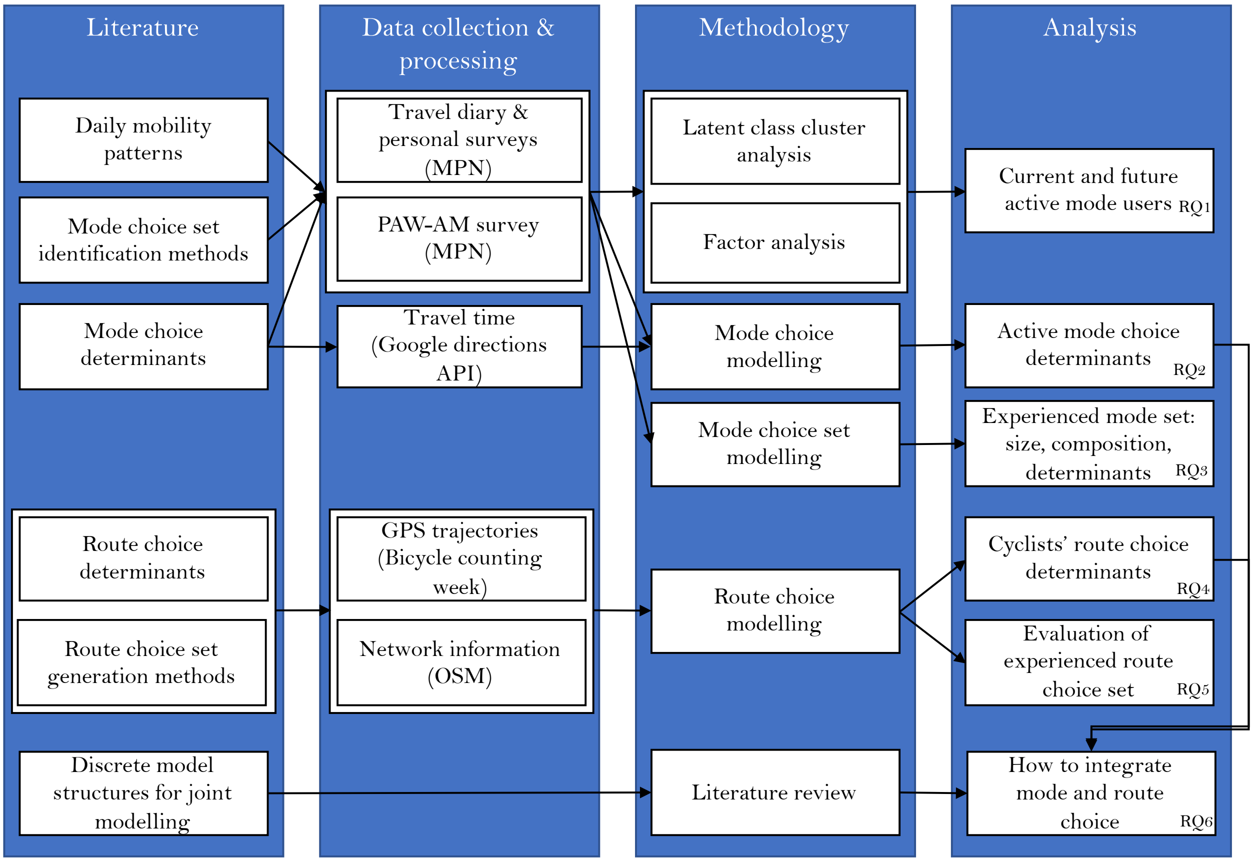


Figure 1.3: Research approach

Based on a question in the PAW-AM survey related to the mode use of individuals in the last half year for different trip purposes, the experienced mode choice set can be identified (which is data-driven). To identify the size, composition and determinants of influence *(research question 3)*, first potential determinants found in literature are evaluated. Data from the PAW-AM, personal, and household surveys is used for this research. The size and composition of the experienced choice set are analysed and compared to reported behaviour in the travel diary. Furthermore, discrete choice models are estimated to identify relevant determinants of the choice set, where different model structures are tested.

To answer *research question 4*, literature on the determinants relevant for cyclists' route choice is reviewed. Furthermore, the currently used route choice set generation methods are evaluated. These methods have several shortcomings regarding false negatives and false positives, consequently the experienced route choice set (data-driven) is explored for this study. The bicycle counting week ('Fietstelweek' in Dutch) collected GPS data from cyclists' in the Netherlands over the course of a week. The data from the city of Amsterdam is available for this research. The experienced choice set is based on observed routes during this seven-day period. Discrete choice models are estimated to identify which determinants are relevant for cyclists' route choice.

Research question 5 relates to the evaluation of the experienced route choice set in comparison to current methods used for choice set generation. Consequently, route choice set generation methods are evaluated and two methods are chosen based on their computational efficiency and presence in the cycling route choice literature: the breadth-first search on link elimination (BFS-LE) and labelling methods. The GPS trajectories of the bicycle counting week are used as input for the experienced route choice set, whereas the network information from Open Street Map (OSM) is used to generate the choice sets based on the choice set generation algorithms. The different choice sets are compared with respect to their composition and ability to reproduce observed choices (false positive and negative), and model estimation and model validation abilities.

Finally, the findings from research questions 2 and 4 show that similar determinants influence active mode and route choice behaviour. Consequently, the simultaneous estimation of mode and route choice models would both benefit and clarify the relationship between these choices. Therefore, a literature review on the current approaches to jointly model these choices is conducted *(research question 6)*. Due to the very limited number of studies on the joint modelling of mode and route choice, the literature review includes multiple travel choices: trip chaining, departure time, destination choice, mode choice, and route choice.

1.4. Research Contributions and Implications

The research in this thesis is expected to contribute to both science and practice in several ways. The most important contributions to science (4.1) and implications for practice (4.2) are hereby discussed.

1.4.1. Contributions to Science

This thesis provides several contributions to science that are related to daily mobility patterns, identification of determinants of mode and route choice, specification of route and mode choice sets, and modelling approaches for simultaneous modelling of multiple travel choices. In this section the main contributions are discussed.

Explicitly Researching the Relationship between the Total Daily Mobility Pattern and Attitudes towards Modes

In this thesis, active mode users and potential users are evaluated by investigating the relationship between behaviour and attitudes. Instead of analysing this relationship at the trip-level, as has been done in previous research, we investigate the daily mobility pattern, in terms of number of trips per mode per day. By explicitly investigating the relationship between behaviour and attitudes, we avoid endogeneity related to attitudinal variables, which would occur if attitudes are used to identify daily mobility patterns, as is done in previous research. This approach allows to provide insights into travel mode consonance and dissonance from a perspective of daily mobility patterns. Preferred mode(s) are thus reflected on in the light of the total daily pattern, instead of on a trip level, providing a complete overview of active mode use. We find a high correlation between mode use on a daily basis and the attitude towards modes. However, this relation was not found for all individuals, which shows potential for switching modes in the future, given the right incentives.

Identification of Determinants of Active Mode Choice in a Cycling-rich Context

A mode choice model is estimated for the Netherlands, which is characterised by its large share of active mode use. This allows for a comparison of relevant determinants for walking and cycling, with existing mode choice literature from environments where active modes are scarce. It provides insights into differences and similarities regarding mode choice determinants between these environments. In the mode choice model various different categories of determinants are included simultaneously, allowing for evaluation of importance of determinants in comparison to other determinants and for calculation of implications of potential policy measures. Travel time is one of the variables that is reported in the travel diary (MPN). We investigate the quality of the self-reported travel time in comparison to calculated travel time, to identify the reliability of self-report data for mode choice research. We find that some explanatory variables are less important in a cycling-rich context compared to the low-cycling context, such as socio-demographics and ownership. Furthermore, we find that walking and cycling are influenced by different explanatory variables and if they are influenced by the same variable, their impact differs.

Identification of Determinants of the Experienced Mode Choice Set

The experienced mode choice set, which consists of used modes that are observed over a longer period of time, provides a rich source of information regarding the size and composition of used modes of individuals. Instead of modelling the choice set in combination with the choice, which is often done in literature, a model is estimated that investigates the formation of the experienced choice set. The determinants that are relevant for the formation of this choice set related to commuting trips are identified in this research. This thesis identifies the relevance of significantly more determinants on the mode choice set, compared to previous research, which can benefit future mode choice research in the choice set specification. Relevant determinants include socio-demographics, ownership characteristics, urban density, and reimbursement by employer.

Identification of Determinants of Cyclists' Route Choice in a Cycling-rich Context

A route choice model is estimated for cyclists in Amsterdam, the Netherlands, which is known for its high share of cycling trips. A comparison between determinants relevant for route choice in environments where cycling is dominant and in those where it is rare can be made. This comparison results in insights on differences and similarities between route choice determinants between both types of environments. We find that, in general, similar determinants are important. However, the impact of these determinants differs, especially related to the distance and percentage of cycle path on a route.

Comparison between Experienced Route Choice Set and Choice Set Generation Algorithms

The route choice model uses a data-driven approach for choice set identification. This experienced route choice set is based on revealed preference data of a large sample of individuals collected over a sufficiently long period of time. All chosen routes are included, resulting with no false negatives and limited false positives. This set is smaller than considered set, but potentially approximates it when observed over a long enough period of time. The data-driven approach is evaluated against commonly used choice set generation algorithms on its added value. The comparison is based on choice set composition, model estimation and model validation. We find that the experienced choice set can be used in analysing composition of choice sets and model estimation, as it provides behavioural knowledge on route choice of cyclists. However, it cannot be used for model validation or prediction purposes.

Systematic Methodology for Identifying Determinants of Active Mode and Route Choice

In this thesis, active mode and route choice are studied by systematically investigating choice set formation and identifying determinants that are relevant for the choice behaviour. The results of these studies are region-specific and can be used there in applications and policy. However, the methodology is general and can be applied in other regions/cities.

Identification of Suitable and Applicable Approaches for Simultaneously Modelling Multiple Travel Choices

A literature review study is performed on the discrete choice modelling structures that have been used to simultaneously model multiple travel choices. The identified structures are assessed with respect to their suitability, reflecting the behavioural realism, and applicability, reflecting the potential for adoption in practice. The review provides an overview of what has been done in the state-of-the-art, assesses the models on their usability given considerations on applicability and suitability, and provides future research directions regarding development of these model structures for the purpose of joint travel choice modelling. We find that the 'ideal' structure in terms of suitability and applicability does not yet exist, however by combining existing approaches, behavioural realism can be largely improved.

1.4.2. Implications for Practice

This thesis presents implications for practice by means of providing (i) insights to how practitioners working on transport planning models can improve the representation of mode and route choice behaviour of active mode users, and; (ii) input on active mode behaviour that is relevant for policy-makers aiming to increase active mode shares. The main implications are discussed in this section.

Practitioners in Transportation Planning

Practitioners working on transport planning models often evaluate policy measures to assess their impact, such as the construction of additional infrastructure or changes to lay-out of streets. Often, they work with the four-step model, consisting of trip generation, distribution, mode choice, and route choice. These models are (still) largely aimed at motorised traffic, consequently active mode choices are insufficiently represented. This thesis shows which determinants are relevant for active mode and cyclists' route choice, which can help enhance the representation of active modes in the mode choice and route choice aspects of transport planning models. Furthermore, this thesis shows that the experienced mode choice set differs largely from choice sets resulting from commonly applied methods. Consequently, the specification of the mode choice set requires more care than previously assumed. Finally, this thesis provides an overview of all the model structures used for the joint modelling of multiple travel choices, that go beyond the currently employed structures in the four-step model. Practitioners can enhance their knowledge on the model structures used for this purpose, so that adoption in practice of more complex models might become a possibility in the future.

Policy-makers

Policy-makers aiming for an increase in active mode shares can benefit from this study in various ways. This thesis presents different classes of daily mobility pattern users that are analysed in combination with their attitudes towards modes, showing consonant or dissonant behaviour. The findings of this research can be used to identify which type of policy measures can be taken to increase the active mode shares. Furthermore, the findings of the mode choice (set) and cyclists' route choice model can be used by policy-makers to create policies that aim for increasing active mode share, related to for example commuting allowances provided by employers or (street) design of new neighbourhoods.

1.5. Research Context

This thesis is part of the ERC Advanced Grant Allegro (Unravelling active mode travelling and traffic: with innovative data to new transportation and traffic theory for pedestrians and bicycles), which has as an overarching research objective:

"To develop and empirically underpin comprehensive behavioural theories, conceptual and mathematical models to explain and predict the dynamics of pedestrians, cyclists, as well as mixed flows at all relevant behavioural levels, including acquiring spatial knowledge, activity scheduling, route choice and operations, within an urban context, with a special focus on the role of ICT on learning, and choice behaviour." – Hoogendoorn (2014)

Allegro consists of eight PhD projects and three Postdoc projects that are part of three themes: an active mode mobility laboratory, transportation and traffic flow theory for active modes in an urban context, and theory and laboratory applications. This thesis is part of the second theme: transportation and traffic flow theory for active modes in an urban context. As discussed before, this thesis aims to understand and model mode and route choice behaviour of active mode users. This thesis has relations with two other projects within this theme, namely the project on analysing activity travel patterns of active mode users (performed by Florian Schneider) and the project on spatial cognition and exploration behaviour in urban environments (performed by Lara-Britt Zomer). This has resulted in collaborations on various topics, of which one is reflected in this thesis (see Chapter 2).

1.6. Thesis Outline

In order to reach the aim of understanding and modelling mode and route choice behaviour of active mode users, this thesis is split up into four parts and eight chapters, see Figure 1.4.

The first part relates to the active mode users. Different modes are used by individuals over the day. To investigate who are current active mode users and who are potential active mode users, a segmentation can be made between user groups. Chapter 2 presents these classes of daily mobility patterns. To investigate the possibilities for changing these patterns towards active modes, the classes are investigated in relation to the attitude towards modes. This part addresses *research question 1*. This chapter is based on the following article:

Ton D., Zomer L.B., Schneider F., Hoogendoorn-Lanser S., Duives D.C., Cats O., and Hoogendoorn S.P. (in press). Latent classes of daily mobility patterns: the relationship with attitudes towards modes. *Transportation.* https://doi.org/10.1007/s11116-019-09975-9

The second part relates to mode choice and is split up in two chapters. In chapter 3, a mode choice model is estimated for the Netherlands, that includes all commonly used modes: car, public transport, bicycle, and walking. In this chapter, determinants are identified that influence walking and cycling. These findings are compared to environments where active mode use is scarce. Furthermore, the relationship between walking and cycling is tested, resulting in information on the (in)dependency between these modes. To get a better understanding of how the mode choice set of individuals is built up, the experienced mode set is evaluated (data-driven). Chapter 4 investigates the size and composition of this choice set and investigates what

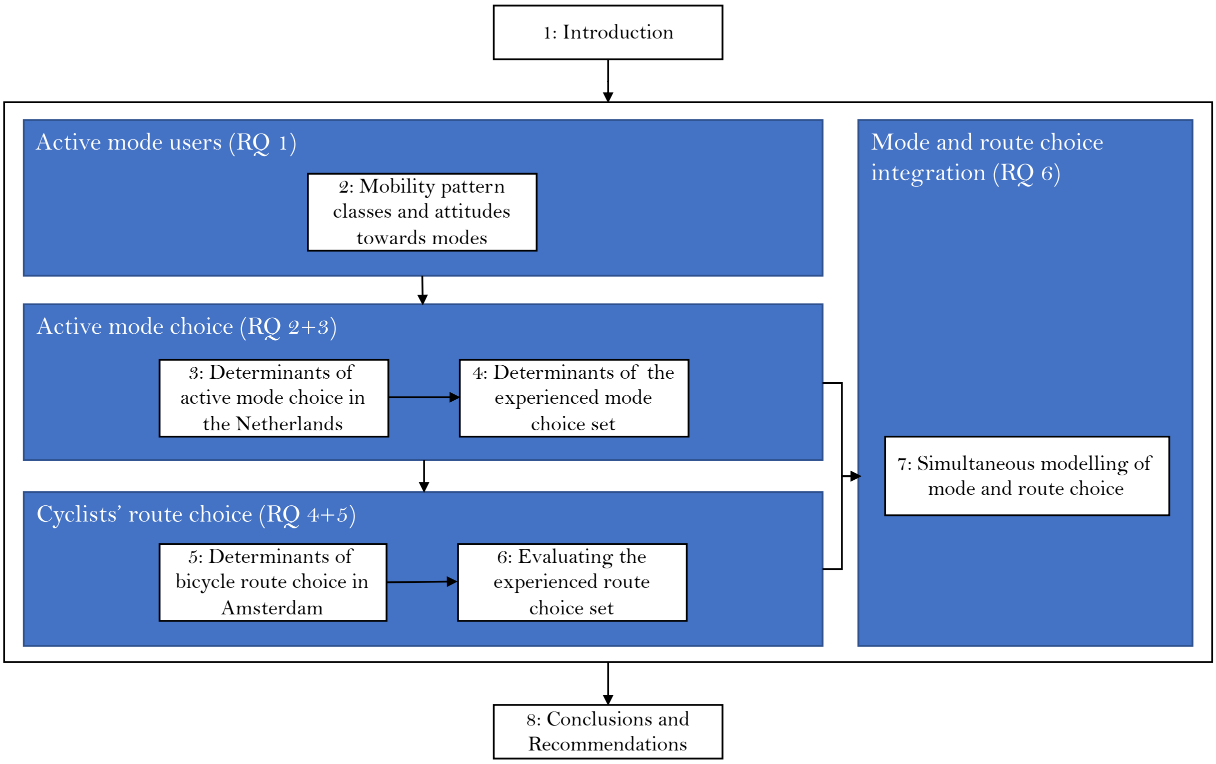


Figure 1.4: Thesis outline

determinants are of this set. This part addresses *research questions 2 and 3*. Chapters 3 and 4 are respectively based on the following articles:

Ton D., Duives D.C., Cats O., Hoogendoorn-Lanser S., and Hoogendoorn S.P. (2019). Cycling or Walking? Determinants of mode choice in the Netherlands. *Transportation Research Part A*. 123:7-23. https://doi.org/ 10.1016/j.tra.2018.08.023

Ton D., Bekhor S., Cats O., Duives D.C., Hoogendoorn-Lanser S., and Hoogendoorn S.P. (submitted). The experienced mode choice set and its determinants: commuting trips in the Netherlands.

The third part pertains to route choice, which is split up in two chapters. In Chapter 5 determinants relevant for cyclists' route choice in Amsterdam, The Netherlands are identified using discrete choice models. The findings are compared against environments where cycling is rare. It uses the experienced choice set (data-driven), which was not previously applied. To evaluate the added value of this data-driven approach, it is compared to two commonly used methods in cyclists' route choice in Chapter 6. The comparison is based on choice set composition, model estimation, and model validation. This part addresses *research questions 4 and 5*. Chapters 5 and 6 are respectively based on the following articles:

Ton D., Cats O., Duives D.C., and Hoogendoorn S.P. (2017). How do people cycle in Amsterdam, the Netherlands? Estimating cyclists' route choice determinants using GPS data from an urban area. *Transportation Research Record: Journal of the Transportation Research Board*. 2662:75-82. https://doi.org/ 10.3141/2662-09

Ton D., Duives D.C., Cats O., and Hoogendoorn S.P. (2018). Evaluating a data-driven approach for choice set identification using GPS bicycle route choice data from Amsterdam. *Travel Behaviour and Society.* 13:105-117. https://doi.org/10.1016/j.tbs.2018.07.001

The fourth part relates to the integration of choice dimensions. According to Bhat (1998) several reasons exist for jointly modelling multiple travel choices. One of these reasons is that both choices are influenced by the same determinants. Parts two and three showed that mode and route choice are both influenced by trip characteristics, such as travel time and distance. Consequently, ideally these two travel choices are also modelled jointly, as to reveal cross-relations between these choice dimensions. To start this research, Chapter 7 provides a literature review of the currently used discrete choice modelling methods for jointly modelling multiple travel choices. Due to the absence of many studies on mode and route choice (only four), other travel choice dimensions are included in this literature review. This part addresses *research question 6*. Chapter 7 is based on the following submitted article:

Ton D., Duives D.C., Cats O., and Hoogendoorn S.P. (submitted). Simultaneous modelling of multiple travel choice dimensions: Assessment of the suitability and applicability of different discrete choice modelling structures.

Finally, Chapter 8 ends with conclusions and recommendations of this thesis. The main findings of the research are provided and answers to the research questions are given. Furthermore, we discuss several of the methodological decisions made in this thesis. Finally, implications for practice and recommendations for future research are provided.

Chapter 2 – Mobility Pattern Classes and Attitudes towards Modes

This chapter is based on the following article:

Ton D., Zomer L.B., Schneider F., Hoogendoorn-Lanser S., Duives D.C., Cats O., and Hoogendoorn S.P. (in press). Latent classes of daily mobility patterns: the relationship with attitudes towards modes. *Transportation.* https://doi.org/10.1007/s11116-019-09975-9

Abstract

Active modes (i.e. walking and cycling) have received significant attention by governments worldwide, due to the benefits related to the use of these modes. Consequently, governments are aiming for a modal shift from motorised to active modes. Attitudes are generally considered to play an important role in travel behaviour. Understanding the relationship between the attitude towards modes and the daily mobility pattern, can support policies that aim at increasing the active mode share. This paper investigates the daily mobility patterns of individuals using a latent class cluster analysis. The relationship between these classes and attitudes towards modes is investigated. Data of the Netherlands Mobility Panel (MPN) of the year 2016 is used, in combination with a companion survey focussing on active modes. This study identifies five classes of mobility patterns: 1) car and bicycle users, 2) exclusive car users, 3) car, walk, and bicycle users, 4) public transport+ users, and 5) exclusive bicycle users. Eight factors of attitudes towards modes are identified: five mode related attitudes, two public transport related attitudes, and one related to the prestige of using modes. The results show that

the majority of the users exhibit a multimodal daily mobility pattern. Generally, individuals are more positive toward used modes, compared to unused modes. Furthermore, a high level of travel mode consonance is found. When this is not the case (dissonance), often active modes or sustainable modes are preferred. Consequently, when the goal is achieving a higher active mode share, some individuals need to be targeted to change their mobility portfolio (exclusive car users and car and bicycle users), whereas others should be encouraged to increase the use of active modes at the cost of car use (public transport+ users and car, walk, and bicycle users).

2.1. Introduction

Walking and cycling (i.e. active modes) have gained significant attention by governments worldwide. They foresee many benefits from high shares of active mode usage. Examples are increased health for individuals, but also reduced emissions and traffic jams if active mode usage replaces car usage. Consequently, many governments have set goals for increasing the active mode share over the next decades (Pan-EuropeanProgramme, 2014). Ideally, this increase in active mode use would be paired with a decrease in car use, so that more sustainable mobility patterns emerge. Kroesen (2014) found that single-mode (habitual) users were less likely to change their mobility pattern over time compared to multi-mode users. Therefore, it is important to evaluate the overall daily mobility pattern of individuals and identify traveller types, as this can provide input for whom to target with policies designed to attain the desired shift towards active modes.

Attitudes are generally considered to play an important role in determining the mode choice and, more general, travel behaviour of individuals (Gärling et al., 1998). An attitude is broadly defined as an affective evaluation, regarding an object or behaviour, which can be positive or negative (Ajzen and Fishbein, 1977). The relationship between attitude and behaviour has been translated to theoretical frameworks, the most prominent being the Theory of Planned Behaviour (Ajzen, 1991), which has often been applied in the travel behaviour domain (e.g. Bamberg et al., 2003; Heinen et al., 2011; Muñoz et al., 2013). Furthermore, attitudes have been introduced into the discrete choice modelling theory, by developing models that can accommodate subjective latent constructs, like the hybrid choice model (Ben-Akiva et al., 1999). This model has also been applied in the travel behaviour domain (e.g. Habib et al., 2014; Kamargianni and Polydoropoulou, 2013; Krueger et al., 2018; Vij et al., 2013). Two main approaches of investigating the relationship between attitudes and travel behaviour have been identified.

The first approach focusses on quantifying the relationship between attitude towards a mode and the mode choice at a trip-level, while using the abovementioned theoretical frameworks and applying discrete choice modelling theory. Several studies investigate this for a single mode, as a binary choice (e.g. to cycle or not to cycle), while other studies research a broader spectrum of modes. Due to the increased interest in active modes, research has often explored walking and cycling, albeit in a binary fashion. Several studies have investigated the impact of the attitude towards walking on the choice to walk (e.g. Lindelöw et al., 2014; Rodriguez and Vogt, 2009). Similarly, past research quantified the impact of the attitude towards cycling on the choice to cycle (e.g. Fernández-Heredia et al., 2014; Heinen et al., 2011; Ma and Dill, 2015). Another body of literature has investigated this relationship for multiple modes, i.e. they investigate the one-on-one relationship but accommodate multiple modes (e.g. Akar and Clifton, 2009; Kamargianni et al., 2015; Maldonado-Hinarejos et al., 2014). Generally, findings suggest that if a person has a positive attitude towards a mode, the probability of using that mode increases. This approach investigates the mode choice at a trip-level, thus ignoring that individuals have a mobility portfolio and use multiple modes on a daily basis. Therefore, this approach does not show the relationship between mobility patterns and attitudes towards different modes.

The second approach focusses on mobility patterns of individuals. This approach also often applies the aforementioned theoretical frameworks. A common way to deal with the many different mobility patterns that can be present in the population, is to reduce the complexity by identifying segments or classes of individuals. In this, often attitudes are used to help classify individuals into classes or segments. Many different studies set out to identify classes of individuals based on mobility patterns, attitudes, and other aspects, which is often referred to as modality styles or mobility styles (e.g. Diana and Mokhtarian, 2009; Krizek and Waddell, 2002; Krueger et al., 2018; Lanzendorf, 2002; Molin et al., 2016). These studies integrate attitudes and observed behaviour into one model, consequently assuming a relationship between attitude and behaviour. An issue arises when deriving input for policies from these segments, as it is argued that attitudes are endogenous to travel choices and should not be used as targets for policy design (Chorus and Kroesen, 2014). Consequently, it might be better to use only observable variables of the mobility pattern for the segmentation analysis, so that policy design can be tailored to population segments.

To overcome the issues related to trip-level research and endogeneity of attitudes, this study investigates the relationship between daily mobility patterns of individuals and the attitudes towards modes by including only observable variables in the segmentation analysis and by explicitly investigating the relationship between attitudes and mobility patterns. This paper presents the findings of a latent class cluster analysis on individuals’ daily mobility patterns, after which the individuals in each class are compared on differences and similarities in their attitudes towards modes. Furthermore, within each class the attitudes towards modes are compared to the observed behaviour to identify whether individuals travel using their most preferred mode. We use census data from the Netherlands. The Netherlands is characterised by a high share of active mode use, consequently we expect a diverse range of mobility pattern clusters. The findings provide insights into how the overall mobility pattern of individuals relates to their attitude towards modes. The results of this study can be used to identify which groups of individuals to target in order to achieve a higher share of active mode usage by means of policy interventions.

The remainder of this paper is organised as follows. Section 2.2 describes the data collected for this study and explains the data filtering process. Section 2.3 details the research methodology. In section 2.4 the results of the analysis are presented and discussed. Finally, section 2.5 concludes this study.

2.2. Data Collection and Filtering

The data that is collected for this research is introduced in section 2.2.1. Then, the data filtering procedure is described in section 2.2.2.

2.2.1. Data Collection

This study uses census data from the Netherlands Mobility Panel (MPN). This is a longitudinal household panel, which was commenced in 2013 with the goal of investigating the changes in travel patterns of individuals and households over a longer period of time. The panel is to a large extent representative of the Dutch population, although teenagers and low-income individuals are slightly underrepresented. Every autumn, panel members fill in a personal survey, household survey and three-day travel diary. The personal survey focuses on personal characteristics and asks questions regarding mode preferences for different activity purposes and the attitude of individuals regarding motorised modes and the bicycle. The household survey contains questions regarding household characteristics and ownership or availability of modes. Finally, in the travel diary every individual is asked to write down all the trips made over the course of three days, including the mode of transport, the trip purpose and the distance covered. We refer the reader to Hoogendoorn-Lanser et al. (2015) for a detailed description of the MPN surveys.

In this study, we investigate the relationship between the daily mobility patterns and attitudes towards modes. The MPN survey contains questions on the attitudes towards motorised modes (i.e. car, train and local public transport) and the bicycle. It does not address walking, even though walking is often used as the main mode of transport in the Netherlands (19% of all trips (CBS, 2018)). In summer 2017, an additional survey on the perceptions, attitudes, and wayfinding styles towards active modes (coined PAW-AM) was distributed among the respondents of the MPN panel, with the goal of enriching the MPN dataset in relation to active modes. We distributed the survey among respondents of the MPN survey, who indicated that they walked or cycled at least once in the last year (consequently excluding 1.3% of the respondents of the MPN panel that did not walk or cycle and are assumed to be largely inactive).

2.2.2. Data Filtering

To perform this research, we need to have data on both the attitude towards modes and mobility patterns. As the attitude towards walking is only measured in 2017, we cannot make use of the longitudinal nature of the MPN dataset. The MPN data of the year 2016 is used, because this dataset contains the most recent travel diaries. Consequently, because we make use of cross-sectional data, we cannot infer causality of the relationship between attitudes and mobility patterns, but only investigate its existence. Next, we merge the MPN surveys (personal, household, and three-day travel diary) and the PAW-AM survey, enabling a complete overview of attitudes towards modes, personal and household characteristics, and daily mobility patterns. Consequently, only individuals that have filled in both the MPN and PAW-AM surveys are included in this study, resulting in a total of 2,871 individuals.

The MPN data collection took place in autumn 2016 (September through November), whereas the PAW-AM survey was distributed in summer 2017 (June). During the elapsed time, several major changes or life events could have occurred. The life events that drastically change mobility patterns, such as changing jobs or moving houses, need to be taken into account. We have therefore excluded individuals that have experienced such a life event. Consequently, the final dataset used in this study consists of 2,425 individuals.

2.3. Methodology

In this section the methodology for analysing the relationship between daily mobility patterns and attitudes towards modes is presented. Section 2.3.1 discusses the definition and classification of the daily mobility patterns. The approach for analysing the attitudes towards modes is described in Section 2.3.2. Finally, in Section 2.3.3 the methodology for analysing the relationship between the classified daily mobility patterns and attitudes towards modes is presented. The research methodology is depicted in Figure 2.1.

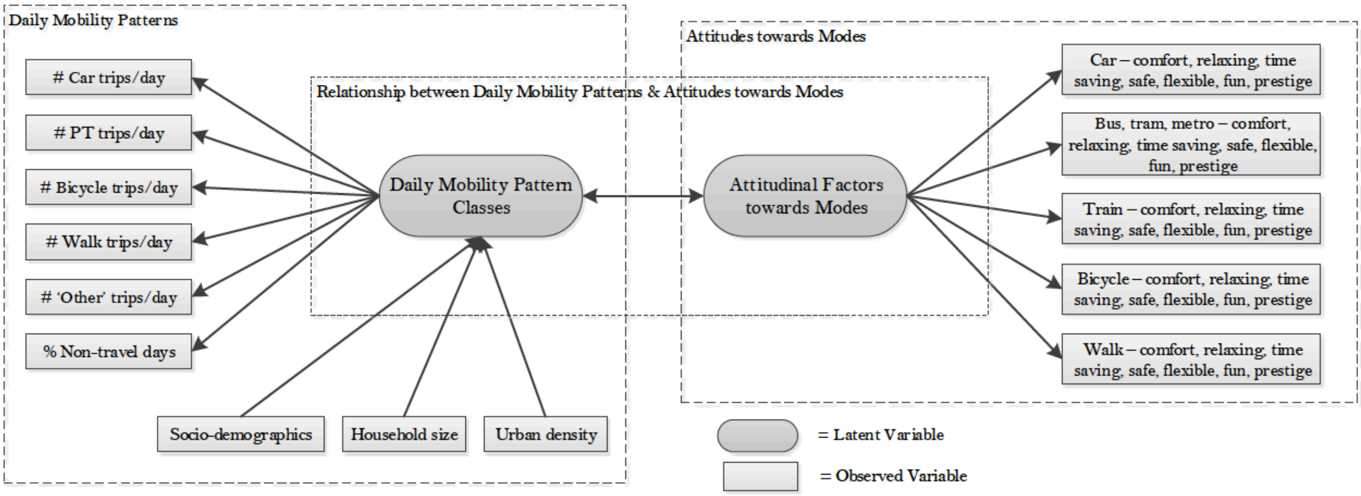


Figure 2.1: Research methodology

2.3.1. Daily Mobility Patterns

The daily mobility pattern can be defined in different ways; therefore, we start by providing the definition used in this study. Afterwards, we describe the approach for classifying the daily mobility patterns.

Defining the Daily Mobility Pattern

The definition of the daily mobility pattern, based on the three-day travel diary, should satisfy two conditions. First, it should reflect the mode choice of individuals. Second, it should take into account a mode use hierarchy of the individual towards different modes. Three possible indicators for defining the daily mobility pattern have been proposed in the literature (e.g. de Haas et al., 2018): distance per mode, travel time per mode, and number of trips per mode. Table 2.1 provides the mean and standard deviations of each of the definitions for the modes in our dataset: car, public transport (PT), bicycle, and walk.

Table 2.1: Characteristics of the average daily mobility pattern of individuals in the data

Mode	Number of Trips Mean (std. dev)	Distance Mean (std. dev)	Travel Time Mean (std. dev)
Car	1.5 (1.6)	27.9 (66.2)	34.8 (44.7)
Public Transport	0.2 (0.5)	6.8 (25.9)	11.1 (38.2)
Bicycle	0.9 (1.3)	2.8 (5.4)	13.4 (22.8)
Walk	0.4 (0.9)	0.4 (1.2)	5.9 (16.4)

The number of trips per mode is found to be the most reliable in self-report studies (de Haas et al., 2018). It also satisfies both conditions. The use of distance would result in the overrepresentation of the car and public transport in the mobility pattern, which might not correctly capture the mode use hierarchy of individuals for this study. It shows that for longer distances these modes are more attractive, but for shorter trips we cannot conclude anything on the mode use hierarchy of individuals. The use of travel time would improve the representation of the mode choices and mode use hierarchy, but here we see very large standard deviations. Especially for public transport this deviation is large, mostly because this mode encompasses both inter-city (train) and intra-city (bus, tram, and metro) trips. Therefore, we define the daily mobility pattern based on the number of trips reported per mode. We average the number of trips per travel day as reported in the three-day travel diary, to identify the daily mobility pattern.

In addition, we identify two other aspects that need to be addressed as part of the classification of the daily mobility pattern. First, trips that are marked as unreliable (e.g. detours, wrong mode assigned) (9.2%), exceptional modes (0.9%), and trips abroad (0.7%) are present in the dataset. Excluding these trips creates incomplete mobility patterns. Therefore, the category "other" is added to classify these trips. Second, on some days an individual does not travel. Non-travel is operationalised by the share of non-travel days out of the reported days.

In sum, a total of six indicators define the daily mobility pattern of individuals: the number of trips by car, public transport (PT), bicycle, or walking, the number of "other" trips, and the share of non-travel days. In these indicators access and egress mode use is excluded, for example a PT trip with walking as access and egress is only counted as a PT trip.

Daily mobility patterns during the week and weekend are very different. Generally, travel during the week is more structured due to work and school. In the weekend individuals travel less, and non-travel occurs more often (de Haas et al., 2017; Hoogendoorn-Lanser et al., 2015). Therefore, we focus on the weekdays for the analysis of daily mobility patterns. This means that some individuals are left with a two- or one-day travel diary, as they have reported trips over (part of) the weekend.

Classifying the Daily Mobility Pattern

The daily mobility patterns are analysed by applying a latent class cluster analysis (LCCA) using *Latent Gold* (Vermunt and Magidson, 2005). This method assigns individuals to classes on a probabilistic basis. It is generally preferred over deterministic clustering techniques, because of the reduction in the misclassification bias (Vermunt and Magidson, 2002). Furthermore, LCCA allows for the use of statistical criteria to determine the optimal number of classes and the significance of model parameters can be assessed.

The LCCA assumes that one latent variable can explain the associations between the indicator variables, which is a categorical variable. Each individual has a probability to belong to each class, based on its characteristics. These characteristics are called covariates and are represented by for example socio-demographics. The covariates are used to predict the probability of class membership. The LCCA model therefore consists of two parts: a structural part where the covariates are used to predict the class membership of individuals, and a measurement part where the latent classes explain the associations between the indicators (Vermunt and Magidson, 2005).

The active covariates cannot be endogenous to the indicators. An example of a non-suitable active covariate in our case is the possession of a driver's license, which is largely endogenous to the number of car trips. Seven suitable active covariates were identified from literature (e.g. de Haas et al., 2018; Molin et al., 2016), namely urban density, occupation, education level, working hours, number of household members, gender, and age.

Furthermore, inactive covariates are included in the model. These do not help in predicting class membership, but can afterwards help understand the composition of each class. In this study we include ownership (endogenous), distance, and the relevant excluded active covariates as inactive covariates.

In the LCCA, the appropriate number of classes is determined by first estimating only the measurement part of the model, thus by only including indicators. This means that no covariates are used yet for determining class membership. The appropriate number of classes to model daily mobility patterns can be decided by using statistical criteria like the Bayesian Information Criterion (BIC) and the relative increase of log-likelihood per added class, which should exceed a threshold of 4% (Nylund et al., 2007). Furthermore, as we want to statistically test the relationship between the daily mobility pattern classes and attitudes towards modes, we need sample sizes per cluster that allow us to do this. To ensure that differences between cluster sizes remain limited, we set the smallest cluster size to 8% of the data. We test models with 1 to 10 classes, where the number of classes is determined based only on the six daily mobility pattern indicators. When the number of classes is decided upon, the model is estimated as a combined measurement and structural model, i.e. with both indicators and active covariates. The initial values of the model were examined; as local optima can be reached in the optimisation process. The best performing setting in the measurement model is used in the combined model, which is tested in terms of stability and performance. The best combination of active covariates, based on improvement in log-likelihood and stability, results in the final model. The combined model is then also estimated for n+1 and n-1-classes, to check if the n-class model is still the best model.

2.3.2. Attitudes towards Modes

De Vos (2018) mentions that to measure the attitude towards modes, statements need to be presented to individuals, which are framed in a way that enables comparison between attitudes towards different modes. He argues that this is achieved by asking about aspects of different modes (e.g. fun) and asking individuals about their opinion on a Likert-scale (with five or seven answers). This method for asking about individuals' attitudes towards modes has been applied by for example Anable and Gatersleben (2005), Molin et al. (2016), Kroesen et al. (2017) and Kroesen and Chorus (2018).

The questions related to the attitudes towards modes in the MPN and PAW-AM surveys are framed according to the method mentioned by De Vos (2018). The respondents were asked seven attitudinal questions per mode. These questions pertain to *comfort, relaxation, time saving, safety, flexibility, fun, and prestige related to using those modes*. Each of these questions was asked in relation to the following five modes: car, bicycle, walking, train (inter-city PT), and bus/tram/metro (BTM – intra-city PT). Because all respondents answer the questions of attitudes towards modes, the public transport modes can be included separately. The questionnaire employed a five point Likert-scale ranging from 'completely agree' to 'completely disagree' (Olde Kalter et al., 2015).

To reduce the size of the analysis and examine whether latent variables underlie the responses to the attitudinal questions, the 35 attitudinal questions are categorised using a factor analysis. We include all questions in the factor analysis, to test if individuals have consistent attitudes towards a mode or consistent attitudinal aspects (e.g. fun) regardless of the mode. This provides insights into how attitudes are formed and also if and for which aspects there is potential for change. A person that has mixed attitudes towards a mode, for example riding a bicycle is fun but unsafe, might change his or her attitude based on changes to the bicycle infrastructure and its related safety (Ma and Dill, 2015). However, if a person is completely positive or negative across the board towards a certain mode, this seems to suggest low potential for change.

A principal axis factoring analysis is applied, which ensures capturing the shared variance of attitudinal questions with latent variables (Field, 2009). Furthermore, varimax rotation is used, which maximises the possibility of capturing each attitudinal question using one factor (Field, 2009). The variables are saved using the regression method. The resulting variables have a mean of zero, however when comparing them to the mobility pattern classes we expect that differences between classes will become visible.

2.3.3. Daily Mobility Pattern Classes versus Attitudinal Factors

The latent classes of daily mobility patterns are compared to the latent attitudinal factors to investigate the presence of a relationship between mobility patterns and attitudes. As mentioned before, many studies include the attitudes in the clustering process (e.g. Diana and Mokhtarian, 2009; Molin et al., 2016), which means that a relationship is assumed between attitudes and daily mobility patterns. Previous research has shown that this relationship is indeed present (e.g. Kroesen et al., 2017). Therefore, we investigate statistical differences and similarities in the attitudes of individuals belonging to different mobility pattern classes. Furthermore, within the latent classes of mobility patterns, a comparison with the attitudes towards modes is made. The goal is to identify to what extent individuals in each class travel with their best perceived travel mode. De Vos (2018) has previously researched this at the trip level and found a high degree of consonance, i.e. travel using the best perceived mode. However, he mentions that it remains unknown whether this also holds for the daily mobility pattern.

As the data does not meet the requirements for performing parametric tests (Field, 2009), the Kruskal Wallis test is used to test whether individuals in different classes have significantly different attitudes (per factor). If this is the case, the Mann Whitney-U test (with a Bonferroni correction, to control for Type 1 errors) shows which classes are significantly different from one another. Consequently, we can conclude on the presence or absence of a relationship between daily mobility patterns classes and attitudinal factors. Furthermore, we know which classes are significantly similar and different in their attitudes.

The comparison within clusters is based on the latent factors towards different modes that arise from the factor analysis on the attitudinal questions. The best perceived mode is identified using the questions that load on the mode specific attitudes, by evaluating the average perception of each mode. It is possible that different modes are perceived equally positive by certain individuals. This is taken into account in the analysis. The goal of this analysis is to identify the extent to which individuals in each class use their best perceived mode, but also to identify the extent to which individuals use their least perceived mode.

2.4. Results and Discussion

This section describes and discusses the results of the LCCA for mobility patterns (2.4.1) and the factor analysis results with respect to attitudes towards modes (2.4.2). Finally, the results of the relationship between mobility patterns and attitudes towards modes are presented (2.4.3).

2.4.1. Latent Classes of Daily Mobility Patterns

A total of 10 models (1-10 classes) were tested for daily mobility patterns, based on the three-day travel diary. The most suitable number of classes is the result of a minimisation of the BIC value, relative increase of log-likelihood of more than 4%, and minimum class size of 8%. Table 2.2 shows the model fit of each of the estimated models. The BIC value decreases with every added class until the 9-class model, the log-likelihood reduction stagnates when exceeding six classes, and the minimum class size is smaller than 8% when more than five classes are introduced. Based on all considerations, we select the 5-class model as the most suitable.

Table 2.2: Evaluation criteria for determining the number of classes of the LCCA

# Classes	# Parameters	Log-likelihood	BIC(LL)	Smallest class (%)
1	14	-18,196	36,502	100.0%
2	26	-6,201	12,605	35.9%
3	38	-1,297	2,891	12.3%
4	50	664	-939	12.2%
5	62	2,553	-4,622	9.9%
6	74	3,654	-6,732	5.4%
7	86	4,693	-8,715	5.2%
8	98	5,742	-10,721	4.4%
9	110	6,704	-12,551	4.3%
10	122	6,388	-11,826	0.4%

The 5-class model was expanded by identifying different combinations of active covariates. Some of the identified covariates are correlated (e.g. age and occupation), consequently we only included one of the correlated covariates in each model. The best combination of active covariates is occupation, urban density, number of household members, gender and education level (log-likelihood = 2,804, improvement log-likelihood = 9.9%). Table 2.3 shows the parameters of the estimated 5-class model, split up in the measurement model and the structural model.

The measurement model consists of an intercept, which can be interpret as a constant that reflects the baseline preference regarding that indicator, while the effect of the classes is taken into account. Furthermore, class-specific parameters reflect the (un)attractiveness of the indicator variables. The intercepts show that bicycle and car trips are most attractive to all individuals. The non-travel ratio is a discrete indicator, that can take five values (as a result of only including weekdays in the analysis), ranging from zero (travel during all days) to one (no travel on any day). For each of the five levels an intercept is calculated which is used as the baseline to calculate the value for each specific class. Following expectations, travelling during all (week)days is preferred. When looking into the classes, several interesting observations are made. For some classes, the intercept is counteracted with the class-specific parameter, for example class 5 and car trips or class 1 and PT trips. This means that this mode is very unattractive for individuals in these classes. Some very positive parameters are also observed, such as bicycle trips for class 5 and walking trips for class 3. Consequently, individuals in those classes find trips using these modes very attractive. Finally, the non-travel ratio has a very positive parameter for class 2, which suggests that the share of non-travel is high for that class.

The structural model also shows an intercept, which reflects the general fit of the population for a class. The indicators show that class 1 has a better fit compared to class 5. The parameters of the covariates show how well each class fits for individuals with those characteristics. Regarding urban density, individuals living in high urban density are more likely to be associated with classes 4 and 5, whereas individuals in low density areas are more prevalent in classes 2 and 3. In many countries, living in a low density area means that one is forced to use the car. In the Netherlands, however, cycling is a very popular mode of transportation, with a modal share of 27% (CBS, 2018). Furthermore, many individuals own bicycles. This means that even in low density areas, bicycles are also available and used, next to the car. The number of household members of an individual influences the daily mobility pattern of individuals. An individual living in a household of 3+ members is more associated with class 1 than class 4, whereas an individual living alone is more associated with class 4. The gender covariate shows that in general more women are present in the population compared to men, however, class 2 and 4 are more associated with men compared to women. Regarding occupation, study/school shows the highest (positive and negative) parameters, meaning that students have the strongest associations. Furthermore, the employed individuals only have a positive association with class 2. Finally, individuals with a high (completed) education level show strong associations with class 4, whereas individuals with a low level show a better fit with class 5.

Table 2.3: Parameters of the LCCA model with 5 classes for weekday daily mobility patterns

Prediction of indicators (Measurement model)									
	Values	**Intercept**	*Wald*	**C.1**	**C.2**	**C.3**	**C.4**	**C.5**	*Wald*
# Car trips		1.19	2,280.1*	0.41	1.03	0.35	-0.60	-1.19	2,470.4*
# PT trips		0.28	1,177.2*	-0.28	-0.28	-0.28	1.11	-0.28	1,178.1*
# Bicycle trips		1.12	1,922.3*	0.09	-1.12	-0.07	-0.50	1.60	2,081.5*
# Walking trips		0.38	909.4*	-0.38	-0.38	1.15	0.00	-0.38	1,221.8*
# Other trip		0.32	824.6*	0.54	-0.32	0.14	-0.05	-0.32	902.1*
% of non- travel days	0	2.61	2,026.1*	-0.49	2.50	-1.08	-0.56	-0.38	298.0*
	1/3	0.15							
	1/2	-0.63							
	2/3	-1.30							
	1	-0.84							

Prediction of latent class membership (Structural model)							
	Values	**C.1**	**C.2**	**C.3**	**C.4**	**C.5**	*Wald*
Intercept		0.48	0.14	0.34	-0.47	-0.48	127.85*
Urban Density	High	-0.09	-0.08	-0.08	0.16	0.10	32.66*
	Medium	0.08	-0.22	-0.01	-0.01	0.15	
	Low	0.00	0.30	0.09	-0.15	-0.25	
# Household members	1	-0.11	-0.14	-0.01	0.25	0.01	23.72*
	2	-0.09	0.02	0.06	0.03	-0.02	
	3+	0.20	0.12	-0.05	-0.28	0.00	
Gender	Female	0.03	-0.13	0.09	-0.03	0.05	15.43*
	Male	-0.03	0.13	-0.09	0.03	-0.05	
Occupation	Study	-0.47	-1.11	-0.72	1.60	0.70	268.15*
	Retired	0.49	0.13	0.33	-0.58	-0.37	
	Unemployed	0.03	0.51	0.45	-0.93	-0.06	
	Employed	-0.05	0.46	-0.06	-0.09	-0.27	
Education Level	Low	0.09	0.08	0.02	-0.36	0.17	21.04*
	Medium	0.01	0.06	0.03	0.04	-0.13	
	High	-0.10	-0.13	-0.04	0.31	-0.04	

*Significant at the 5% level

When applying the models on all individuals in the dataset, profiles can be created for each of the classes. Table 2.4 shows a description of each class and provides the distribution for the population as a whole. The classes are named after the mode use characteristics in the daily mobility pattern. The classes are car and bicycle users (CB), exclusive car users (C), car, walk, and bicycle users (CWB), public transport+ users (PT+), and exclusive bicycle users (B). The CB and C segments together consist of more than half of the sample population. Class CWB is the third class and consists of almost a quarter of the sample population. Consequently, the last two classes are much smaller (PT+ and B). Three classes have a diverse mode use pattern (multimodal users), whereas two classes use on average one mode exclusively (unimodal, habitual users).

Car and Bicycle Users

The CB class is characterised by more than average trips by car, bicycle, and 'other' trips. Individuals in this class have a low share of non-travel days. Furthermore, the CB class has a relatively large share of retired individuals, which results in a higher share of individuals that are 65 and older. These individuals travel further than average by car and bicycle and car ownership is slightly higher than average. Furthermore, this class is comparable to the population as a whole in terms of urban density, number of household members, education level, and gender.

Table 2.4: Within-class distributions of the indicators and covariates

		Class CB	Class C	Class CWB	Class PT+	Class B	Total
Class size	Percentage	27.5%	27.0%	23.7%	12.3%	9.5%	100%
Indicators							
Car trips	Mean	1.6	2.2	1.5	0.6	0	1.5
PT trips	Mean	0	0	0	1.4	0	0.2
Bicycle trips	Mean	1.2	0	1.1	0.6	2.7	0.9
Walking trips	Mean	0	0	1.5	0.4	0	0.4
Other trips	Mean	0.9	0	0.5	0.3	0	0.4
Share of non-travel days	Mean	6%	32%	4%	6%	6%	13%
Active Covariates							
Occupation	Study/School	8%	3%	5%	**34%**	**26%**	11%
	Retired	**22%**	12%	**21%**	9%	12%	16%
	Unemployed	13%	15%	**20%**	4%	13%	14%
	Employed	57%	**70%**	54%	53%	50%	59%
Urban Density	High	49%	47%	50%	**62%**	**57%**	51%
	Medium	23%	16%	21%	17%	22%	20%
	Low	28%	**37%**	29%	22%	22%	29%
# Household members	1	18%	15%	21%	**26%**	18%	19%
	2	32%	32%	**38%**	26%	27%	32%
	3 or more	50%	**53%**	42%	48%	**54%**	49%
Education level	Low	26%	22%	24%	23%	**35%**	25%
	Medium	39%	**42%**	40%	36%	33%	39%
	High	34%	36%	36%	**41%**	32%	36%
Gender	Female	55%	49%	**60%**	54%	58%	55%
	Male	45%	**51%**	40%	46%	42%	45%
Inactive Covariates							
Age	12 - 19	6%	2%	3%	**15%**	**22%**	7%
	20 - 39	25%	34%	28%	**48%**	24%	31%
	40 - 64	46%	**52%**	46%	27%	40%	45%
	65+	**23%**	13%	**23%**	10%	13%	17%
Distance (km/day)	Car	30.6	46.3	24.4	10.2	0	28.0
	PT	0	0	0	55.2	0	6.8
	Bicycle	4.1	0	2.7	1.7	9.0	2.8
	Walk	0	0	1.5	0.4	0	0.4
Ownership	Car	78%	**87%**	75%	41%	48%	72%
	Bicycle	75%	74%	78%	**88%**	**88%**	78%
	PT (subscript.)	25%	19%	29%	**83%**	35%	32%

CB: Car & bicycle users, C: Exclusive car users, CWB: Car, walk, & bicycle users, PT+: Public transport+ users, B: Exclusive bicycle users. Percentages per variable add up to 100%, **Bold** = highest shares for category of variable compared to other classes

Exclusive Car Users

Members of class C only use the car. Furthermore individuals in this class have a high share of non-travel days. Given other characteristics of this class, such as the relatively high share of employed individuals, males, ages between 40 and 64, and low urban densities, this most likely represents working at home days. The individuals in this class mostly live in households with three or more individuals. The individuals often own a car. Finally, they travel farthest (on average 46 km), which might be due to the fact that most live in low urban areas.

Car, Walk, and Bicycle Users
The third class CWB travels by car, on foot, and by bicycle, but does not use public transport. Individuals in this class are relatively often retired or unemployed. Consequently, the population elderly (65+) is well represented in this cluster. Furthermore, a high share of females is present in this class, which are mostly unemployed (including those who are by choice out of the workforce). In this class there are more than average two-person households. Finally, individuals in this class walk further than average. Individuals in this class are comparable to the entire population in terms of the distribution over urban densities, education levels, and ownership levels.

Public Transport+ Users
The PT class users travel most often by public transport, however they also travel by car, bicycle, and on foot. The characteristics of the users in this class are very different from the first three classes. The individuals in this class are mostly studying or working, young (<40 years), highly educated, live quite often alone, and live in high urban areas. The last is not surprising as it is known that the public transport services are more efficient and frequent in densely populated areas in the Netherlands. They travel relatively far by public transport, often involving train travel (inter-city travel). Furthermore, the PT+ users often do not own a car, but do own a bicycle and a public transport subscription.

Exclusive Bicycle Users
The last class are the exclusive bicycle users (B). This group travels frequently and only by bicycle. The users in this class are relatively often school going teenagers, which live with their parent(s). Consequently, they have a low education level (as they have not yet finished their schooling). They live in highly urban areas, where more facilities are reachable within short distances. Furthermore, they travel relatively long total distances by bicycle (9 km). Car ownership is low (also caused by age restrictions), but bicycle ownership is high. During the data collection period, the weather was relatively stable. Potentially, if data was collected for a longer period of time, with more variability in the weather, other modes would have been observed too (e.g. public transport).

To compare the identified classes of daily mobility patterns to other studies, the differences in the research approaches need to be stressed. In other studies, attitudes have been used in the identification of the mobility patterns (Diana and Mokhtarian, 2009; Krueger et al., 2018; Molin et al., 2016) and the objective mobility pattern has been defined differently (de Haas et al., 2018; Diana and Mokhtarian, 2009; Krueger et al., 2018; Molin et al., 2016). Furthermore, several studies have investigated different countries, enabling comparison between countries (Diana and Mokhtarian, 2009; Krueger et al., 2018). Consequently, a one-on-one comparison between the classes identified in different studies is not possible, notwithstanding we hereby identify the noteworthy differences and similarities.

Diana and Mokhtarian (2009) investigated datasets from two countries: USA and France. The modes included in their research differ for each country (the French dataset contained more modes). They identified four groups of users for the French dataset: unimodal car users, car-dominated but multimodal users, highly multimodal users with moderate travel intensity, and highly multimodal users with heavy travel intensity. Bicycle use is very low in this dataset and they did not include walking as a mode, therefore the multimodality is related to public transport and car use. For the USA dataset, they also identified four groups: unimodal car users, moderate travellers which are multimodal but car-dominated, light travellers which are multimodal but car-dominated, heavy travellers which are multimodal but car-dominated.

The first group corresponds to our class C. The other classes mostly show multimodality between car and walking, but the last group also contains a fair share of public transport use. Bicycle use is very low in the USA and was not included in the classification. As a result, both the classifications for France and the USA are very different from our study, because in most of our classes active modes play an important role.

Krueger et al. (2018) investigated mobility patterns in Sydney, Australia. They identified three classes: car-oriented users, public transport-oriented users, and car- and bicycle-oriented users. Class C in our study has large overlap with the car-oriented users, only this class is much larger (50.5%). Furthermore, the car- and bicycle oriented users overlap with our CB class, with the largest difference being that the bicycle is less frequently used. Finally, the public transport-oriented class overlaps with our PT+ class, but is again larger (20.9%). Consequently, all the classes reported by Krueger et al. (2018) have also been identified in our study, while we identify two additional classes and find smaller motorised traffic-dominated classes.

Molin et al. (2016) used data from the Netherlands to classify mobility patterns. They identified five clusters: car multimodal, bicycle multimodal, bicycle and car, car mostly, and public transport multimodal. They did not incorporate walking as a separate mode, consequently their classes are mostly build upon bicycle and car use. Their classes to a large extent correspond to our classes with the exception that they find no bicycle only class.

De Haas et al. (2018) used data from the same panel as our study: the MPN dataset. They identified the daily mobility patterns differently by only including trips per mode, and excluding non-travel and other trips, where they summed the trips over the course of three days. They identified six classes: strict car, car and bicycle, bicycle, car and walk, low mobility, and public transport users. Most classes show good correspondence with the classes identified in this study. Their car and walk class is extended to also include bicycle in our study. Because we only include individuals that have used the bicycle or walked in the last half year, we exclude to a large extent the immobile population, consequently we do not identify a low mobility class.

In general, the results from this study are in line with the findings from other studies in the Dutch context (de Haas et al., 2018; Molin et al., 2016). Differences in the classes with other countries are mostly related to the fact that the Netherlands has a high share of active mode use. Most countries are more car-oriented and lack a high share of active modes to this date. Arguably, the Dutch situation may illustrate what the class distribution of daily mobility patterns could be after achieving a shift towards active modes. Next, we examine the relevance and importance of attitudes in this context using the original PAW-AM survey designed and collected for this study.

2.4.2. Factors of Attitudes towards Modes

For each of the five modes (car, bicycle, walk, BTM, and train) the respondents answered seven attitudinal questions. Figure 2.2 shows in a radar chart format the average scores of the population as a whole on each question for each mode, which provides a first insight into which factors might arise from the factor analysis. The two public transport modes are valued the least, where the train is valued over the urban modes. The Dutch population disagrees on average to the statement that the use of any particular mode relates to one’s prestige. Generally, the car is valued highest, followed by the bicycle. However, regarding relaxation during travel, both walking and cycling are valued more positively than the car.

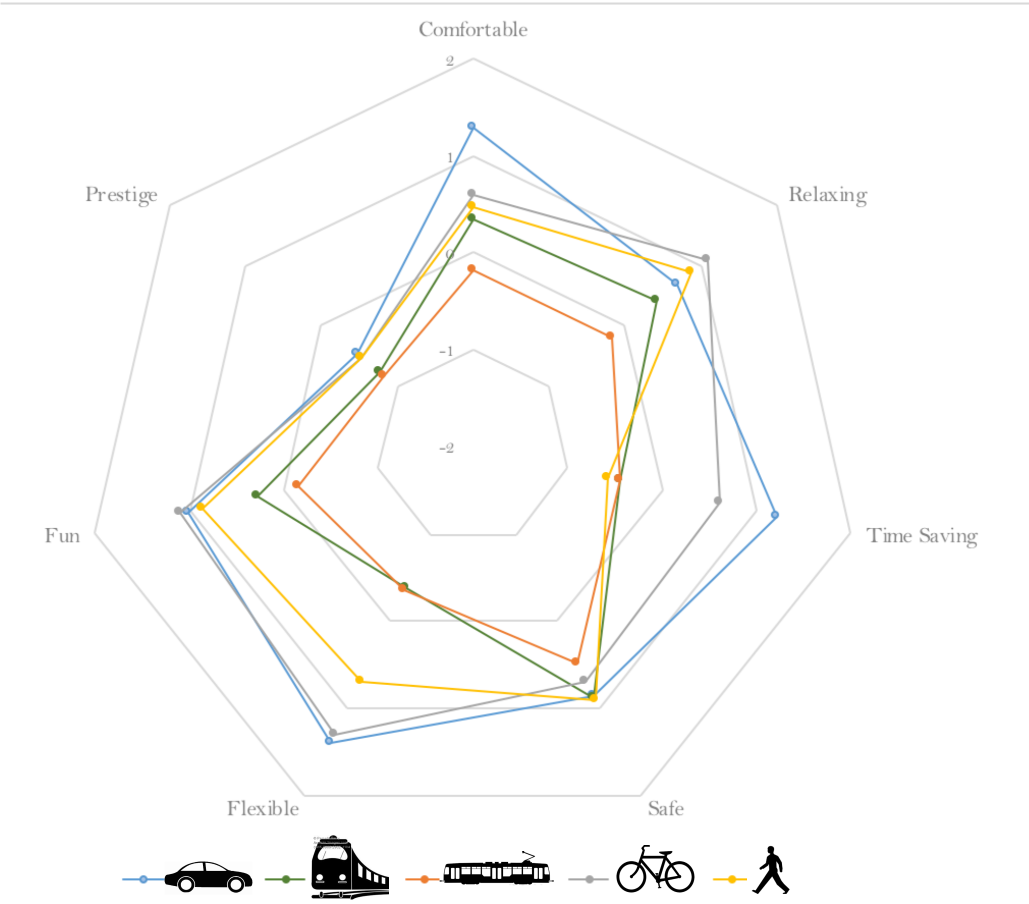


Figure 2.2: Average score per attitude question per mode for the entire population

The data is suitable for factor analysis with a score of 0.865 on the KMO test for sampling adequacy (>0.8) and total variance explained of 59.8%. Two questions related to walking (*prestige* and *time saving*) were excluded from the factor analysis, because they could not be captured by any of the factors (factor loading <0.4). The 33 attitudinal questions were reduced to eight factors; one factor for each mode, one related to the prestige of using modes and two related to PT attitude (combined train and BTM). The eight factors can be characterised as described in Table 2.5. The results of the factor analysis are in line with the expectations based on Figure 2.2. The loading represents how each of the variables load on the factor, where a higher value represents a better fit to the latent factor. Furthermore, the Cronbach's alpha provides a measure of reliability of the resulting latent factors. A value higher than 0.8 is considered good and reflects high internal consistency. A value under 0.7 is questionable, which is observed for the 'public transport safety' factor. This might be attributed to the fact that only two questions are loaded on this factor, where generally at least three are expected. Therefore, the results of this factor need to be interpreted with care.

All questions that are answered in a similar consistent fashion are combined into one latent factor. Interestingly, consistency in answers is exhibited for various attitudes towards a given mode rather than various modes for a given attitude (e.g. comfort). Hence, individuals have relatively strong overall opinions towards different modes and therefore it will be harder to change attitudes via, for example, promotional or information campaigns. This especially holds for the car and bicycle, because six out of the seven attitudinal questions are combined into the attitudinal factor. The train attitude includes only four attitudinal questions, meaning that individuals are more varying in their attitude towards the train. Consequently, the attitude towards the train could potentially be changed using promotional campaigns that focus on the flexibility and time saving aspect (PT efficiency). The only latent factor that strongly represents an attitude covering different modes is the prestige of using modes. This latent factor includes statements on the perceived increase in status associated with using car, train, BTM, or bicycle. Individuals have answered these questions in a similar fashion for each of the modes, disagreeing with the statement that it induces prestige (see Figure 2.2). Consequently, the general tendency is that modes do not increase status for individuals and status is not part of the attitude towards each of the modes.

Table 2.5: Results of the factor analysis on attitudinal questions related to the different modes

Factor	Variables		Loading	Cronbach's Alpha
Car attitude	Travelling by car is …	Comfortable	0.803	0.865
		Relaxing	0.747	
		Time saving	0.622	
		Safe	0.659	
		Flexible	0.678	
		Fun	0.834	
BTM attitude	Travelling by BTM is …	Comfortable	0.794	0.897
		Relaxing	0.800	
		Time saving	0.504	
		Flexible	0.513	
		Fun	0.828	
Bicycle attitude	Cycling is…	Comfortable	0.740	0.827
		Relaxing	0.808	
		Time saving	0.489	
		Safe	0.502	
		Flexible	0.621	
		Fun	0.833	
Walking attitude	Walking is…	Comfortable	0.757	0.816
		Relaxing	0.794	
		Safe	0.437	
		Flexible	0.562	
		Fun	0.822	
Train attitude	Travelling by train is…	Comfortable	0.736	0.849
		Relaxing	0.759	
		Safe	0.454	
		Fun	0.761	
Prestige of using modes	Travelling by … increases status	Car	0.612	0.812
		Train	0.831	
		BTM	0.734	
		Bicycle	0.745	
Public transport efficiency	Travelling by train is…	Time saving	0.654	0.858
		Flexible	0.638	
	Travelling by BTM is…	Time saving	0.609	
		Flexible	0.589	
Public transport safety	Travelling by … is safe	Train	0.583	0.697
		BTM	0.591	

2.4.3. Attitudinal Factors versus Latent Mobility Pattern Classes

The comparison between attitudinal factors and the latent mobility pattern classes is done in two parts. First, a comparison between classes is done, which identifies whether individuals in different classes indeed have different attitudes. Second, a within class comparison is done, which identifies the extent to which individuals use their (least) perceived modes in their daily mobility pattern.

Comparison between Latent Mobility Pattern Classes

A total of five different latent mobility pattern classes has been identified using a LCCA analysis. The attitudinal questions have been reduced in dimension to eight latent factors. In this section we test whether these five groups of individuals have different attitudes (towards various modes). Figure 2.3 shows the attitude scores on each factor for each of the classes. The dashed black line represents the average opinion of all respondents. This is used to reflect the differences in magnitude between factors, as the factors themselves have a mean of zero.

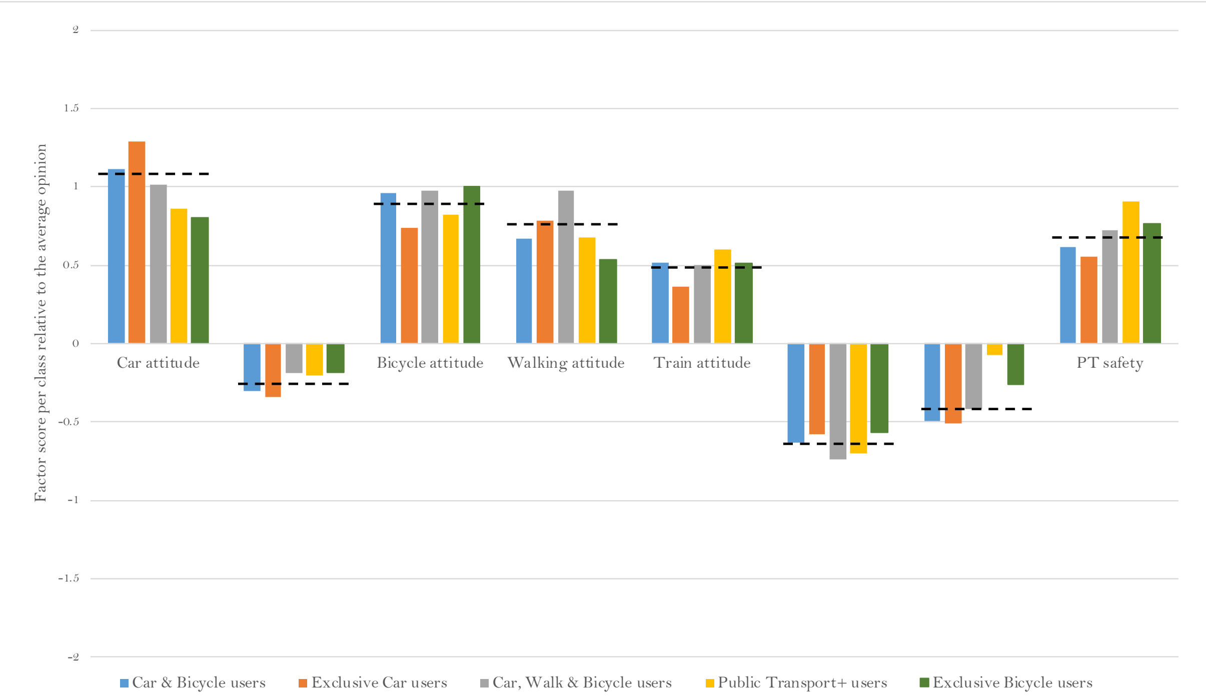


Figure 2.3: Factors representing attitudes of the five classes

Several observations can be made in relation to Figure 2.3. First, the car attitude is highest for the exclusive car users, which only use the car and lowest for the exclusive bicycle users, which do not use the car. Second, the exclusive bicycle users are most positive towards the bicycle, however the other classes that use the bicycle on a daily basis (CB and CWB) are also more positive than average towards the bicycle. Third, the only class that walks on a daily basis (CWB) is most positive towards walking. Fourth, the PT+ users are most positive towards the train. Class C is much less positive towards the train compared to the others. And finally, the PT+ users are respectively most positive and least negative towards PT efficiency and PT safety. In summary, these observations indicate that modes that are actively used by individuals are valued more positively compared to modes that are not or less frequently used.

We test whether the differences observed in Figure 3 are statistically significant. Table 2.6 shows which classes are significantly different from other classes on each of the eight identified attitudinal factors. The Kruskal Wallis test indicated that no significant differences are found in relation to the BTM attitude, which is negative among all user classes. In contrast, statistically significant differences are found for all other attitudes. We then turn to test which classes differ in their attitudes.

Table 2.6: Differences between classes on attitudinal factors

Factor	Class CB *differs from*	Class C *differs from*	Class CWB *differs from*	Class PT+ *differs from*	Class B *differs from*
Car attitude	C,PT+,B	all	C,B	CB,C	CB,C,CWB
BTM attitude	-	-	-	-	-
Bicycle attitude	C	CB,CWB,B	C,PT+	CWB	C
Walking attitude	CWB	CWB,B	all	CWB	C,CWB
Train attitude	C	CB,CWB,PT+	C	C	-
Prestige of using modes	-	CWB	C	-	-
Public transport efficiency	PT+,B	PT+,B	PT+	CB,C,CWB	CB,C
Public transport safety	PT+,B	CWB,PT+,B	C,PT+	CB,C,CWB	C

Several observations can be made in relation to Table 2.6. First, the classes PT+ and B are not statistically different in their attitudes. This might be due to the fact that their socio-demographic profiles are rather similar (young people in high urban areas). Second, the classes CB and CWB are only significantly different in their attitude towards walking, where CWB is more positive than CB. The first class also makes much more use of walking as mode of transport, signifying the largest difference between these two classes. Third, classes CWB and B differ in their attitude towards car and walking. Just like the previous case, this echoes the major difference in the mobility patterns of the two groups. Fourth, classes C and CWB are very different in their attitudes. They are only similar in their BTM attitude and PT efficiency. The multimodal mobility pattern of CWB is therefore generally related to a more positive attitude towards the non-used modes, compared to the unimodal C class. Fifth, the unimodal classes C and B are not significantly different in their BTM attitude, train attitude, and the prestige towards modes. Consequently, they have a similar opinion regarding PT modes, but a different opinion towards the other modes where they are more positive towards the mode they use. And finally, the classes C and PT+ have similar attitudes towards the modes they do not actively use in their daily mobility pattern, namely the active modes: cycling and walking.

Comparison within Latent Mobility Pattern Classes

The most positive attitudinal score for a mode is the best perceived mode for an individual. Ideally, this mode would be used by the individual in their daily mobility pattern. This would reflect travel mode consonance (De Vos, 2018) and would suggest an ideal match between attitudes and behaviour. If this is not the case, i.e. travel mode dissonance, other factors also influence both the daily mobility patterns and attitudes towards modes. This might stem from the fact that perception is not the same as preference. An individual can have comparable measured perceptions for two different modes, but prefer (and thus use) one over the other based on unmeasured characteristics (such as cost or health). Table 2.7 shows the use of the best and least perceived modes for each latent mobility pattern class.

The use of the best perceived mode varies largely between different classes. The PT+ class uses all modes in their daily mobility pattern, consequently they also use their best perceived mode. The CWB class reaches a 91% travel mode consonance. Only 9% of the individuals in this class do not use their best perceived mode, in these cases the train is better perceived. BTM is the only other mode that is not used by the CWB users, however no one has BTM as their best perceived mode. Consequently, other influences drive these individuals to not use the train in the daily mobility pattern. The single-mode classes (B and C) have the lowest levels of travel mode consonance. Potentially, the individuals that do not use their best perceived mode are captive users. Captive users are bound to one mode, meaning that they cannot or do not have the means to make use of another mode. The best perceived modes for class C are walk, bicycle, and train, whereas in class B these are car, walk, and train. Very few individuals have BTM as their best perceived mode.

Table 2.7: Use of best and least perceived modes in the daily mobility pattern: travel mode consonance and dissonance

	Best Perceived Mode					Least Perceived Mode	
	Use	*Not use*	*Top perceptions if not used*			*Use*	*Not use*
Class CB	74%	26%	train	walk	train-walk	9%	91%
Class C	62%	38%	walk	bicycle	train	5%	95%
Class CWB	91%	9%	train			21%	79%
Class PT+	100%	0%	-			100%	0%
Class B	44%	56%	car	walk	train	7%	93%

If an individual uses the least perceived mode in the daily mobility pattern, this shows a larger discrepancy between attitudes and daily mobility patterns compared to not using the best perceived mode. In this case the single-mode classes show the smallest percentages, 5% for class C and 7% for class B. This means that only few people use a single mode, which they perceive least. As these individuals would most likely have deviated from these single modes, they are indeed likely to be captive users. Again, the PT+ class uses all modes in their daily mobility pattern, which means that the least perceived mode is also included in that pattern. The CWB class has a relatively large share of individuals using their least perceived mode, indicating that either car (5%), walk (6%) or bicycle (3%), or a combination of these (7%) is least perceived. However, these individuals are not captive users, as they also deviate from the least perceived mode.

Governments worldwide share the goal of increasing active and PT mode use (sustainable modes). To identify the potential of these modes, we investigate the active mode and PT perceptions of the individuals in each latent mobility pattern class. Table 2.8 shows the best perceived modes categorised in active and PT for the general population and the dissonant users in each class. The latter reflects users that potentially perceive active or PT modes best, but currently do not use these in their mobility pattern. These individuals are therefore potential future users of active or PT modes.

Table 2.8: Active and PT mode perception for each mobility pattern class

	Best perceived mode		Best perceived by dissonant users		
	Active	*PT*	*Dissonant users*	*Active*	*PT*
Class CB	36%	13%	26%	15%	11%
Class C	30%	8%	38%	30%	8%
Class CWB	43%	13%	9%	0%	9%
Class PT+	39%	19%	0%	0%	0%
Class B	44%	15%	56%	16%	11%

Class CB and C have a relatively low share of best perception for both active and PT modes (less than half), suggesting that it might be difficult to persuade the general CB and C users into using (more) active or PT modes. However, the shares of dissonant users are relatively large. These dissonant users perceive active modes (walking in case of CB) or public transport best. The majority of them have a higher perception of active modes than PT. For class CB this is 15% (walking) versus 11%, while for class C it is 30% (cycling and walking) versus 8%. These users can potentially be persuaded to use (more) active modes or public transport, given the right incentives. Class CWB and B have the best perception of active modes. However, most of these users already use active modes, resulting in the potential for more use of active modes, not necessarily switching modes. Class B has a large share of dissonant users, of which 27% has the best perception of walking or PT. Therefore, this class also shows potential for the use of other sustainable modes, besides the bicycle. Class PT+ perceives PT best. Besides all the potential shown for active or PT mode use, from the attitude perspective, it is striking to see how many individuals, across all classes, have the highest perception of the car (almost half).

2.5. Conclusions

This paper presents the findings of a latent class cluster analysis, applied on census data from the Netherlands, with the goal of revealing different daily mobility travel patterns. Furthermore, we explicitly investigate the relationship between the resulting daily mobility travel patterns and the attitudes towards (alternative) modes, to identify potential for increasing the active mode share across the population.

A total of five different daily mobility pattern classes was identified: 1) car and bicycle users, 2) exclusive car users, 3) car, walk, and bicycle users, 4) public transport+ users, and 5) exclusive bicycle users. These user types differ in their socio-demographics, ownership of modes, distance travelled per mode, household sizes, and urban densities. Active mode use is present in most classes, except for the exclusive car users. Furthermore, three classes exercise multimodality (over the days). Classes of individuals that already use active modes or that are multimodal, might be more inclined to use active modes of transport or to increase their active mode use in the future, as they are already familiar with these. It might be hard to convince the exclusive car users to switch to other, more sustainable, modes.

The attitude towards modes was identified by asking individuals seven questions about the comfort, relaxation, time saving, safety, flexibility, fun and prestige associated with using each of the travel modes. A factor analysis was used to reduce the number of dimensions and to identify likeminded attitudinal questions. A factor made out of statements related to one attitudinal question (e.g. fun) would mean that travel in general is seen as fun or not fun. Whereas, a factor made out of statements related to one mode (e.g. the car) would imply that an individual is generally positive or negative towards that mode regardless of the attitudinal aspect. We identified five mode related factors and three attitude related factors. Consequently, the population is generally positive or negative on all aspects for a given mode. This especially holds for the car and bicycle, where six attitudinal questions are included. Consequently, it will be difficult to influence the attitude of individuals, as all aspects are seen as positive or negative.

In this study we investigated the relationship between the attitudinal factors and the daily mobility pattern classes. The findings suggest that an individual is more positive towards the modes that are included the daily mobility pattern, compared to the modes that are not part of his or her mobility pattern. This is consistent with previous findings reported in the literature, which state that unimodal car drivers have a biased or more negative attitude towards public transport modes, compared to multimodal car drivers (Diana and Mokhtarian, 2009; Molin et al., 2016). In our research this statement is confirmed, but we see a much more negative attitude towards all other modes (it scores lowest on bicycle, local public transport and inter-city public transport). In contrast, the multimodal users are very positive towards the used modes and generally also positive towards the unused modes.

We also investigated the degree of travel mode consonance (use of the best perceived mode) within each mobility pattern class. The single-mode classes (exclusive car and exclusive bicycle users) show the lowest shares of travel mode consonance. The individuals that do not use the best perceived mode are dissonant users. We expect that 5% of the exclusive car users and 7% of the exclusive bicycle users are captive users, as they use their least perceived travel mode. A relatively large share of the exclusive car users has a higher perception of active modes and PT (sustainable), showing that there is potential for changing behaviour in all classes.

When the goal is to achieve a higher active (or sustainable) mode share, these findings indicate that there is potential in each of the classes, however the approach towards reaching the goal differs. The car dominated classes (car and bicycle users and exclusive car users) show potential for switching modes towards more sustainable or active modes, as they have relatively large shares of dissonant users (26% and 38%). These mobility pattern classes include many employed individuals, therefore the employer could take a role in changing the behaviour by stimulating or enabling the use of sustainable modes to work. However, more than half of the individuals in these classes have a better perception of the car, which is consonant with the mobility pattern. Therefore, it is expected to be very challenging to change the behaviour of these individuals. However, they might be stimulated to become more sustainable through the use of car-based shared mobility services (for example by the employer). Ride-sourcing, ride-sharing, and car-sharing are examples of car-based services that offer attributes associated to the car, but steer towards more efficient utilisation of vehicle fleets and thus reducing potentially related externalities. Furthermore, the exclusive car users might be unaware or not fully informed of the attributes, such as level-of-service, of active and sustainable modes. Short-term targeted campaigns can be an effective policy measure to expose these users and potentially enlarge their mobility portfolio. The multimodal classes (car, walk, and bicycle users and public transport+ users) already show sustainable behaviour, which is mostly in line with their perceptions. These individuals can increase their active mode use and reduce car use, especially for shorter distance travel. Integration of the sustainable modes, via mobility-as-a-service (MaaS), could help in providing more attractive services that increase the use of sustainable or active modes, at the cost of car use. Finally, the majority of the exclusive bicycle users shows dissonant behaviour. Potentially, this results in a change of behaviour in the future. To ensure active or sustainable mode use in the future, these individuals could also benefit from MaaS schemes, especially because the majority of these individuals lives in dense urban areas.

The data used for this research is cross-sectional. Consequently, we cannot identify how potential policies or campaigns have influenced the behaviour of individuals. Future research entails collecting another wave of data, and identify shifts in behaviour by executing a latent transition analysis (e.g. Kroesen et al., 2017). Also, if another wave of data is available the causality between attitude and behaviour can be investigated. Another interesting aspect that can be investigated when multiple waves of data are available, is the influence of life changing events on individuals' mobility patterns and attitudes towards modes. Next to that, the results found in this study regarding classes of mobility patterns and attitudes towards modes could be used as input for choice models that aim to investigate potential impacts of policies on mode choices. Finally, in this study we regarded the attitudes towards modes in comparison to the latent mobility classes. It would also be interesting to investigate how the covariates of our model, e.g. socio-demographics, explain differences in attitudes towards modes.

Acknowledgements

This research was supported by the Allegro project (no. 669792), which is financed by the European Research Council and Amsterdam Institute for Advanced Metropolitan Solutions. The data was made available by the Netherlands Mobility Panel administered by KiM Netherlands Institute for Transport Policy Analysis. An earlier version of this work was presented during the 15th International Conference on Travel Behaviour Research in Santa Barbara, California.

Chapter 3 – Determinants of Active Mode Choice in The Netherlands

This chapter is based on the following article:

Ton D., Duives D.C., Cats O., Hoogendoorn-Lanser S., and Hoogendoorn S.P. (2019). Cycling or Walking? Determinants of mode choice in the Netherlands. *Transportation Research Part A*. 123:7-23. https://doi.org/10.1016/j.tra.2018.08.023

Abstract

Interest into active modes (i.e. walking and cycling) has increased significantly over the past decades, with governments worldwide ultimately aiming for a modal shift towards active modes. To devise policies that promote this goal, understanding the determinants that influence the choice for an active mode is essential. The Netherlands is country with a large and demographically diverse active mode user population, mature and complete active mode infrastructure, and safe environment. Mode choice research from the Netherlands enables a comparison on relevant determinants with countries that have a low active mode share. Furthermore, it can provide quantitative input for policies aiming at an active mode shift. This paper estimates a mode choice model focusing on active modes, while including a more comprehensive set of modes (i.e. walking, cycling, public transport and car). Based on data from the Netherlands Mobility Panel (MPN) in combination with an additional survey focused on active modes (coined PAW-AM), this study estimates which determinants influence mode

choice. The categories of determinants identified in the literature are individual characteristics, household characteristics, season and weather characteristics, trip characteristics, built environment, and work conditions. The results show that all categories of determinants influence both walking and cycling. However, the choice for cycling or walking is affected by different determinants and to a different extent. In addition, no active mode nest was found in the model estimation. Cycling and walking should thus be regarded as two distinguished alternatives. Furthermore, the results show that active mode use is most sensitive to changes in the trip characteristics and the built environment.

3.1. Introduction

In the past decades, interest into active modes (i.e. walking and cycling) has significantly increased. A high share of active modes in terms of the number of trips has many potential benefits. At the individual level it can provide health benefits due to increased activity levels, and at the network level it might reduce traffic jams and the associated emissions when substituting the car. Governments worldwide have set goals for increasing the active mode share (Pan-EuropeanProgramme, 2014). Ultimately, they are aiming for a modal shift from motorised to active modes. This transition could be achieved by designing effective policies. Understanding which determinants influence the choice for an active mode can serve as valuable input for these policies.

Several countries already have a high share of active mode use, i.e. the Netherlands, Denmark, and Germany (Pucher and Buehler, 2008). In the Netherlands, the active mode share in the number of trips in the year 2017 was 44%, more than half of these are cycling trips (CBS, 2018). Pucher and Buehler (2008) make a distinction between the cycling rich countries and other countries where cycling is uncommon, such as the USA, Canada, and the UK. They identify that even though more kilometres are cycled in the Netherlands, the fatality and accident rates are much lower compared to the cycling poor countries, indicating a very safe cycling environment. Fishman (2016) identifies the Dutch mature and complete cycling infrastructure as the main contributor to the safe environment. Furthermore, Fishman (2016) stresses that in the Netherlands the cycling population is much more diverse in terms of socio-demographic compared to other countries. Women are known to cycle more than men (Heinen et al., 2010) and also elderly people are active bicycle users (Fishman, 2016). Fishman (2016) identifies that there is a knowledge gap concerning active mode choice from countries like the Netherlands, that are mature in terms of infrastructure, safe, and where cyclists' demographics are diverse. This enables the possibility to make a comparison on relevant determinants for active mode choice between cycling rich and cycling poor countries. Furthermore, when investigating active mode choice in Netherlands there is no need to oversample the cycling population, because a representative sample of the population suffices to ensure a large enough sample of cyclists.

Fishman (2016) argues that the Dutch are 'blind to cycling', meaning that cycling is such an ordinary activity that it has not been warranted much attention, both by practitioners and researchers. Only recently this has started to change. Dutch transport planning models, such as LMS (Rijkswaterstaat, 2018), are used by governmental authorities to assess the impact of policies. These models are tailored to the car and public transport. In line with Fishman's (2016) argument, the active modes have not received much attention. Walking and cycling are combined into 'slow modes' and often evaluated as 'rest-category' (De Jong et al., 2007). In order to correctly estimate the impact of policies, it is essential to include behaviourally accurate mode choice models in these transport planning models. Consequently, research on active mode choice in the Netherlands would benefit both practice and research.

The objective of this study is to identify the determinants influencing the choice for an active mode of transport when considering a more comprehensive set of modes (i.e. car, public transport, bicycle and walking) in the Netherlands, which is characterised by high share of active mode use. This paper presents findings from a discrete mode choice model estimated using census data. We investigate the influence of different categories of determinants on the mode choice, e.g. trip characteristics, socio-demographics and the built environment. The results of this study can be used in two ways: first, the findings can facilitate planning and policy measures in other countries that aim for high active mode penetration and second, the model can improve the representation of active modes in the models that are part of the Dutch transport planning models.

In this paper, Section 3.2 identifies and categorizes determinants influencing active mode choice, which serve as input for this study. Section 3.3 gives a description of the data collected for this research, and it explains the data merging and filtering process and provides an overview of the data in terms of descriptive statistics. In Section 3.4, the specification of the mode choice models is detailed, with focus on the identification of the individual mode choice set and the specification of the discrete choice model. In Section 3.5, the results of the model estimation are reported and discussed. Section 3.6 addresses a discussion of the results. Finally, Section 3.7 provides the conclusions of this study.

3.2. Determinants of Active Mode Choice

In this study we focus on the determinants of active mode choice when considered as part of a more comprehensive set of modes. Therefore, this literature review focusses on the determinants that influence walking and cycling. We refer the reader to the literature review sections in for example Buehler (2011) and Paulley et al. (2006) for studies on public transport and car mode choice determinants.

Many studies have investigated which determinants are of importance in active mode choice. It is possible to divide these determinants into six categories (Heinen et al., 2010; Hunt and Abraham, 2007). These are individual characteristics, household characteristics, trip characteristics, built environment, season and weather characteristics, and work conditions. This section briefly discusses the main findings from literature reviews that focus on cycling and walking, with respect to determinants from each category.

Individual Characteristics

The individual characteristics pertain to all determinants related to the person, e.g. socio-demographics, ability to use a mode, and ownership or availability of modes. The socio-demographics have often been investigated, however for both walking and cycling mixed results are found. Often, literature claims that men cycle more often than women (Fraser and Lock, 2011; Muñoz et al., 2016a). Heinen et al. (2010) confirm this for countries with low cycling penetration, however in countries with high cycling penetration, such as the Netherlands and Denmark, women are found to cycle more often than men. Regarding age, mixed results have been found for both walking and cycling (Handy et al., 2014; Heinen et al., 2010; Mitra, 2013). Young people are often found to cycle more (Muñoz et al., 2016a) and old people to cycle less (Fraser and Lock, 2011), albeit the results are inconclusive. Often a higher education level is linked to lower cycling levels (Heinen et al., 2010), while again mixed results have been reported in the literature (Muñoz et al., 2016a).

The availability of a car has a negative association with the probability to walk or cycle (Heinen et al., 2010; Mitra, 2013), whereas the availability of a bicycle has a positive association with cycling (Fraser and Lock, 2011; Handy et al., 2014; Heinen et al., 2010). The relationship between bicycle availability and walking has not been investigated insofar.

Household Characteristics

The household characteristics relate to the other people in the household and their influence on the active mode choice. The size and composition of the household are known to relate to mode choice. For example, the number of children is negatively associated with the choice for walking to the supermarket (Maley and Weinberger, 2011). Hamre and Buehler (2014) confirm this negative association for both walking and cycling (the latter was not significant). Heinen et al. (2010) state that having no children increases the probability of cycling. Income is often identified as a determinant of active mode choice, however mixed results are reported regarding the directionality of the relationship (Handy et al., 2014; Heinen et al., 2010; Mitra, 2013; Muñoz et al., 2016a).

Season and Weather Characteristics

Cyclists and pedestrians are more exposed to the seasonal and weather conditions than a person travelling by car or using public transport. Generally, summer and autumn are mentioned as the most favourable seasons for cycling and walking (Böcker et al., 2013; Heinen et al., 2010). Winter is negatively associated with active mode travel. Wang et al. (2016) report that environments with cold winters and warm summers are less attractive for active mode users. Regarding the daily weather conditions, the impact of mostly rain and temperature have been studied (Böcker et al., 2013; Heinen et al., 2010). Temperature is found to have a non-linear effect, where cold and very hot weather are negatively associated with active mode use. Regarding rain, mixed results have been found (Böcker et al., 2013; Heinen et al., 2010). Other studies do not explicitly mention temperature or rain, but investigate the influence of extreme or adverse weather, which is negatively associated with active mode use (Fraser and Lock, 2011; Wang et al., 2016).

Trip Characteristics

The most investigated trip characteristics are distance and travel time. They are highly correlated and sometimes considered equivalent, however in cycling research, distance is often investigated (Fraser and Lock, 2011; Handy et al., 2014; Heinen et al., 2010; Mitra, 2013; Muñoz et al., 2016a; Winters et al., 2017). Longer distances are found to be negatively associated with active mode use. Heinen et al. (2010) suggest a non-linear relationship between distance and bicycle use, penalising longer distances more adversely. Distance is related to the built environment, because land use and density of the built environment largely determine how far destinations are located in relation to residential areas (Handy et al., 2014). Other trip characteristics are less often investigated. The day of the week was found to influence cycling choice. During weekdays the bicycle has a larger probability to be chosen (Hansen and Nielsen, 2014). Furthermore, a recreational trip purpose is found to have a positive association with cycling (Fraser and Lock, 2011).

Built Environment

The built environment pertains to road infrastructure (e.g. percentage of cycle path or sidewalks along the route), aesthetics (e.g. proximity to parks), and area characteristics (e.g. presence of shops and population density). The built environment is especially relevant for active modes, as they are more (directly) exposed to the surroundings compared to car and public transport users. The presence, density, and continuity of active mode infrastructure (e.g. bicycle lanes or paths and sidewalks) is positively associated with active mode usage (Fraser and Lock, 2011; Handy et al., 2014; Heinen et al., 2010; Mitra, 2013). Facilities related to cycling, such as bicycle parking, are also positively associated with cycling (Heinen et al., 2010). Regarding the aesthetics, the literature states that the presence of among others parks, street plantation, playgrounds, benches and garbage bins are positively associated with both walking and cycling (Fraser and Lock, 2011; Heinen et al., 2010; Wang et al., 2016). Traffic lights were found to have a mixed relationship with cycling and have not been studied in a broader mode choice context (Heinen et al., 2010). Land use is found to be strongly related to active mode use. A mixed land use environment encourages active mode use, whereas low residential density discourages active mode use (Fraser and Lock, 2011; Heinen et al., 2010; Mitra, 2013; Muñoz et al., 2016a; Wang et al., 2016; Winters et al., 2017). At a more aggregate level, small and medium size cities are positively correlated with bicycle use and the city centre is more attractive for cycling compared to the suburbs (Heinen et al., 2010). Furthermore, areas with high population density are attractive for active mode use (Fraser and Lock, 2011; Muñoz et al., 2016a; Wang et al., 2016).

Work Conditions

Finally, the work conditions relate to the facilities that are offered by the employer. This comprises for example facilities at the workplace, reimbursement for travelling to work using a certain mode, and working hours and flexibility thereof. Heinen et al. (2010) and Handy et al. (2014) state that the availability of facilities related to the car, for example (free) parking options, negatively relate to the choice for cycling. Furthermore, Heinen et al. (2010) identify a positive relationship between facilities that are beneficial for cyclists, such as lockers or showers and bicycle choice. Providing incentives or reimbursement for both the bicycle and public transport have a positive association with cycling (Handy et al., 2014; Muñoz et al., 2016a; Winters et al., 2017). Public transport requires access and egress for which both walking and cycling are often used. The use of the bicycle as access and egress mode also boosts the use of the bicycle on other occasions. On the other hand, if the car is incentivised or reimbursed a negative association is found with bicycle use (Handy et al., 2014). Furthermore, if car usage is disincentivised, evidence suggests that this does not benefit bicycle use, but instead increases public transport use (Braun et al., 2016). Finally, regarding working hours, the literature suggests that having a part-time job is more positively associated with cycling compared to a full-time job (Heinen et al., 2010).

Evidently, the significance of determinants belonging to each of the six categories has been previously investigated. Notwithstanding, the directionality and magnitude has not always been conclusive. Furthermore, there is a need to map and perform a more complete analysis of the determinants influencing mode choice (Handy et al., 2014; Heinen et al., 2010), so that trade-offs among determinants and their relative importance can be established by performing a joint model estimation. This ensures that not only the influence of the individual determinants on the mode choice is quantified, but also their relative influence. The latter is essential to support policy makers in determining what to focus on when the goal is increasing the modal share of active modes. This study addresses determinants from all categories to investigate both the individual and relative importance of modal choice determinants.

3.3. Data Collection and Preparation

This section covers the data collection (3.3.1) and preparation of the data for this study (3.3.2). Furthermore, the selection and preparation of the determinants that potentially influence mode choice (3.3.3) is addressed. Finally, the final dataset is described in terms of individual characteristics (3.3.4) and reported trip characteristics (3.3.5).

3.3.1. Data Collection

In this study census data from the Netherlands Mobility Panel (MPN) is used, which is a longitudinal household panel that has started in 2013 and is designed to investigate changes in travel patterns of a fixed panel of individuals and households over a longer period of time. This panel is to a large extent representative for the Dutch population, except for a slightly lower share of low-income individuals and teenagers. Every year, the members of the panel fill in a three-day travel diary, a household survey and a personal survey. In the travel diary they report among other things, the trips made, the modes used and the distances covered. The household survey relates to household characteristics and the ownership and availability of modes, whereas the personal survey focuses on mode preference for certain activities and their attitudes towards motorised modes. The panel comprises about 2,000 households, totalling 4,000 individuals. For more information on the MPN surveys the reader is referred to Hoogendoorn-Lanser et al. (2015).

Even though the MPN census data is a very rich data source, capturing most of the influence categories of determinants identified in Section 3.2, it lacks data on the built environment. Previous research has established the importance of this category, for example positive association with cycling infrastructure (Fraser and Lock, 2011; Handy et al., 2014; Heinen et al., 2010; Mitra, 2013), positive association with presence of parks, street green, playgrounds, benches and garbage bins (Fraser and Lock, 2011; Heinen et al., 2010; Wang et al., 2016), and positive association with population density levels (Fraser and Lock, 2011; Muñoz et al., 2016a; Wang et al., 2016). Consequently, it is essential to also collect data on the built environment. In 2017 an additional survey (coined PAW-AM), which addresses among other things elements of the built environment that are present in the respondents' neighbourhood, was designed to enrich the MPN dataset. Besides the elements of the built environment, the survey focuses on complementary information with respect to active mode use. This survey was distributed among respondents of the MPN survey, who indicated that they walked or cycled at least once in the last year. The goal was to target active mode users, consequently we excluded 1.3% of the respondents of the MPN panel that did not walk or cycle and are assumed to be largely inactive.

3.3.2. Data Preparation

To be able to investigate the influence of all categories of determinants on mode choice, the MPN surveys (household, personal and travel diary) and the PAW-AM survey need to be merged. Only respondents that have filled in both the MPN and PAW-AM surveys are included in this study, resulting in a total of 2,871 respondents.

In the travel diary several filters are applied to identify which of the 26,192 trips (made by all respondents) can be used for this study. Trips are excluded in the following cases: 1) tours in which the origin is also the destination and no intermediate stop is made, 2) trips of which the reporting is unreliable or inconsistent (e.g. due to large detours, incorrect address information or uncertainty about the reported mode), 3) trips that are made as part of professional driving (e.g. truck drivers), 4) trips outside the Netherlands, 5) non-home based trips and 6) trips that are made by rarely chosen or available modes (e.g. skateboard or boat). The reason for excluding tours and professional driving trips is because the motivation for choosing a mode might be different from normal trips. Non-home based trips (i.e. not starting from home) are excluded because of the dependency on the mode of transport that was used before, for example if a person makes the first trip by car, he or she generally needs to return the car back home, which introduces a dependency that results in limited and/or fixed mode choice set. Conversely, trips starting at home provide no limitations in choosing a mode other than the availability of the mode to the person.

Furthermore, the data collection for the MPN surveys took place in autumn 2016 (September – November), whereas the PAW-AM survey was distributed in June 2017. This means that life events (e.g. a new job, different working hours, a new house or the birth of a child) need to be taken into account. That is, if a respondent has experienced a life event, the data from the MPN survey should not be matched to the PAW-AM survey and these respondents are excluded. The reason is that their travel behaviour could have significantly changed due to these events, creating a mismatch in the data. The final dataset that is available for this study consists therefore of 6,368 trips and 1,864 individuals.

3.3.3. Selecting and Processing Potential Determinants

Based on the determinants identified in the literature reviews (Section 3.2) and the availability of data in the MPN and PAW-AM surveys, potential determinants that influence mode choice are selected for this study. Table 3.1 shows an overview of all the determinants selected for this study. Note that all categories of variables are represented in this list, enabling the comparison of the relative importance of various determinants.

Table 3.1: Determinants that are known to influence mode choice in literature and are available in the dataset for inclusion in the model estimation procedure

Individual characteristics	**Trip characteristics**
Gender	No. trips on day of travel
Age	Departure time
Education	Trip purpose
Ethnicity	No. individuals in travel group
Ethnicity parents	Travel time
Occupation	
Driver's license	**Built environment**
Body Mass Index (BMI)	Urbanisation level
Transit subscription	Metropolitan area (Amsterdam, Rotterdam,
Company car	Eindhoven, Den Haag, Utrecht)
Bicycle/car in household	Nature in neighbourhood (green, water, park)
Mode used for going to high school	Street furniture in neighbourhood (garbage bins, playgrounds)
Mode used in the last half year	Traffic related aspects in neighbourhood (speed bumps,
	cycle paths, cycle parking spots, traffic lights)
Household characteristics	Buildings in neighbourhood (shops, restaurants, schools,
No. household members	public buildings, hospitals, sports centres, flats, offices,
No. children in household	industry)
Household income	
	Work conditions
Season and weather	Working hours per week
Extreme weather	Travel compensation (bicycle, public transport, car)
Month of travel	

In the dataset both travel time and distance are known. These two determinants are highly correlated, therefore only one can be included in the model estimation. The distance and travel time are self-reported by the respondents. Regarding public transport, journey planners usually express the trip in travel time, therefore it is expected that the respondents are able to recall the duration of the trip, but they might not know the distance of the trip. Therefore, travel time is preferred over distance.

The travel time is only available for the mode used to make the trip. This means that the travel times need to be calculated for the non-used modes. Furthermore, by analysing the differences between the reported time by the chosen mode and the calculated time for that mode, the quality of reporting can be assessed. Based on this assessment, a decision can be made on whether the use of the reported travel time in the model estimation is valid.

Calculation of the Travel Times for Non-used Modes

Because only the reported mode is provided per trip, travel times of the alternative, non-chosen modes need to be calculated. These are calculated using the Google Directions API. This API does not allow for performing calculations for past events, therefore in order to create similar conditions for most trips, the calculations were made on a weekday during the day. This affects both public transport (PT) and car, as timetables usually change over the day, especially in the evening/night and traffic jams arise in morning and evening peaks. Therefore, regarding PT, all trips made between 22h and 5h were checked in a journey planner to see if there was a PT option available. If no PT option was available, the alternative was marked unavailable. For the car this validation is not possible, as the amount of traffic on the road differs per day and peak-hour period, this means that some discrepancies can be expected in the calculated travel times.

For 1,366 trips, the PT travel time was equal to the walking time, indicating that instead of providing an option to use train, bus, tram or metro, the journey planner advised to walk. Furthermore, in 57 occasions no PT route could be found (and the distance was not walkable). In these situations, PT is not an option and the alternative was marked as unavailable. Next to that, for one trip no car alternative was found (destination was on an island), for five trips no walking alternative was found and for five trips no cycling alternative was found. These alternatives are all marked as unavailable in the choice set. The reason for not finding routes is the availability of roads in the network of the Google Directions API. For active modes it searches for roads where active modes are allowed.

Analysis of Travel Times for Used Modes

To check whether the reported travel times are reasonable and can be used in the model estimation, the travel time of the chosen mode was calculated. Figure 3.1 shows the mean of the calculated minus the reported travel times, plotted against classes of the reported travel times.

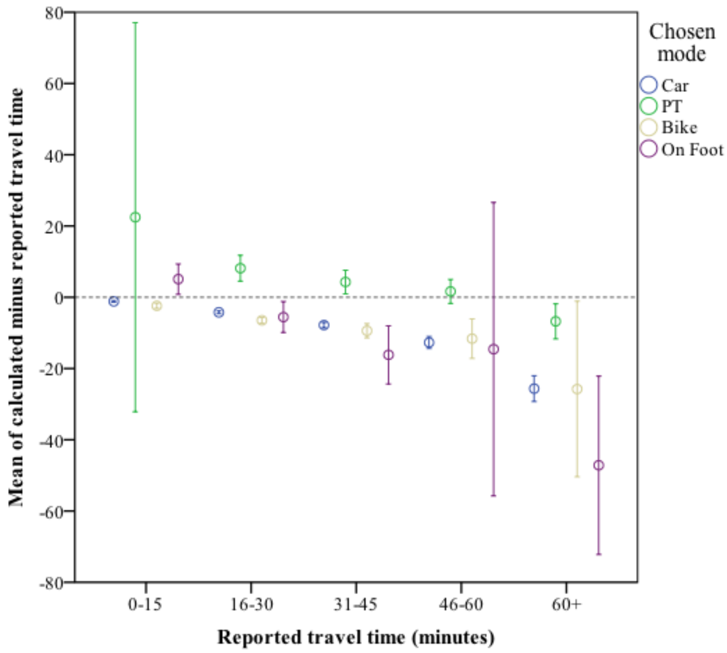


Figure 3.1: Mean and 95% confidence interval of difference between calculated and reported times

Figure 3.1 illustrates that people report shorter PT travel times compared to the calculated travel time, especially for short trips. This might be due to ignoring access and/or egress travel time, but only reporting the in-vehicle time. For the car the reported travel times are over reported, with increasing error for increasing travel time. This could be due to congestion, where the maximum speed cannot be reached. However, it is impossible to check this, because the traffic situation during the trips cannot be recalled. Travel time for cycling trips is also over reported (the extent of overestimation is comparable to the car, but with larger error bars). The calculated travel time is based on the shortest route, but literature suggests that the distance is not the only factor determining cycling route choice (e.g. Menghini et al., 2010; Ton et al., 2017). This indicates that the extent of over reporting could be less severe in reality. The same occurs for walking, although 97% of the trips are below 30 minutes in duration. Summarised, differences between reported and calculated travel times are generally low and can be explained. As the reported travel times also include potential traffic jams or delays, the reported travel time is used in the model estimations.

3.3.4. Characteristics of the Respondents

The final merged and filtered dataset contains 1,864 individuals from all over the Netherlands. The breakdown of their socio-demographics, household size, place of residence and ownership characteristics is described in Table 3.2.

Table 3.2: Characteristics of the individuals in the dataset

		Freq.	Share			Freq.	Share
Gender	Male	852	45.7%	Education level	Low	477	25.6%
	Female	1012	54.3%		Medium	732	39.3%
					High	655	35.1%
Age	12-24	266	14.2%				
	25-34	257	13.8%	No. of household members	1	333	17.9%
	35-44	293	15.7%		2	640	34.3%
	45-54	377	20.2%		3	256	13.7%
	55-64	309	16.5%		4	426	22.9%
	65-74	234	12.5%		5+	209	11.2%
	75+	128	6.8%				
				Urbanisation level	Urban	957	51.3%
Professional situation	Employed	1054	56.5%		Sub-urban	357	19.2%
	Unemployed	249	13.4%		Rural	550	29.5%
	Retired	337	18.1%				
	Student	224	12.0%	Ownership	Car	1359	72.9%
					Bicycle	1696	91.0%
Working Hours	0-12 hours	772	41.4%		PT – subscr.	604	32.4%
	12-35 hours	524	28.1%				
	35+ hours	568	30.5%				

As mentioned before, the sample contains individuals who have cycled and walked at least once during the last year. The distribution of the individuals over age shows that many elderly people are present in the sample (almost 20% is 65 years or older), which indicates that elderly Dutch citizens still use active modes of transport. The surveys are distributed among individuals of 12 years and older, therefore no individuals with a lower age are present in the dataset. The education level shows the highest completed level of education. Consequently, teenagers who are currently studying, have a lower level of education, compared to when they will finish their studies. The low level of education contains finished studies up to the level of pre-vocational secondary education ('VMBO' in Dutch), whereas the medium level of education contains finished studies up to the level of either secondary vocational education ('MBO' in Dutch) or pre-university education ('VWO' in Dutch). The highest level of education includes university education. The result is that a relatively large part of the sample has completed a low level education. More than 82% of the individuals in the dataset live in a household consisting of multiple individuals, which is higher than average for the Dutch population. Furthermore, most individuals live in an urban environment. Finally, most people in the sample own a car and bicycle. The bicycle ownership is high (compared to other countries) with 91%, but in line with the national active mode share. Whereas, only 32% of the respondents in the sample have a type of PT subscription (e.g. travel with discount or travel for free on a fixed line).

3.3.5. Characteristics of the Reported Trips

The final dataset contains 6,368 trips made by car (passenger and driver), public transport (train, tram, bus and metro), bicycle (electric and normal bicycles) and on foot. Table 3.3 provides an overview of the characteristics of the trips made by all individuals in the dataset.

Table 3.3: Characteristics of the trips in the dataset

	Share of trips	**Travel distance [km]**	**Travel time [min.]**	**Largest trip purpose category**	**Trips in the weekend**
	Percentage	*mean (s.d.)*	*mean (s.d.)*	*Purpose (%)*	*Percentage of total trips*
Car	51.6%	18.4 (27.4)	23 (22)	Work (27.4%)	27.5%
PT	5.3%	39.7 (41.1)	66 (39)	Work (42.6%)	14.6%
Bicycle	32.4%	2.9 (3.2)	13 (12)	Leisure (22.6%)	19.0%
On foot	10.7%	0.7 (0.7)	9 (9)	Shopping (37.6%)	26.0%

About 43% of all the trips in the dataset are made using active modes. In line with expectations, travel time and distance are on average higher for car and PT compared to active modes, as the latter are mostly used for short-range trips. On the other hand, the median distance for the car is much shorter with 8km, and for PT this is 27.8km. This shows that overall the car is chosen more often for short distances than PT. Consequently, the car can compete with cycling and walking for short distance trips. The standard deviations for travelled distance are higher than the mean values, which is due to the long tail for travelled kilometres (e.g. maximum travelled distance by car is 260km). PT consists of train, bus, metro and tram. The latter three are mostly found in cities and are generally used for shorter distances. The relatively high mean travel time for PT is therefore mainly due to the large share of train trips in the data, as people cover larger distances and consequently spend a longer time traveling by train. About half of the trips made by PT are work related, which is also reflected in the low percentage of trips made in the weekend. Cycling is mostly used for leisure activities. Notwithstanding, 21.5% of the cycling trips are work related. The relatively low percentages for the main trip purpose of car (27.4%) and bicycle (22.6%) indicate that compared to walking and PT, these modes are used for more diverse trip purposes.

3.4. Specification of the Mode Choice Model

This section describes the mode choice model specification for this study. The approach for identifying the mode choice set for each individual is described in Section 3.4.1. Furthermore, the specification of the discrete mode choice model is presented in Section 3.4.2. Finally, the model estimation process is elaborated upon in Section 3.4.3.

3.4.1. Identification of the Individual Mode Choice Set

When using revealed preference data only the reported mode is known. In choice modelling, the non-chosen but alternative modes need to be identified too. This is a non-trivial task as not every individual has the same set of modes available. Several mode choice studies were identified from literature that focus on active modes, consider the full spectrum of modes, and use revealed preference data.

These studies all use different heuristics for the identification of the individual's mode choice set. Munshi (2016) did not apply any restrictions to the choice set and included all modes for every trip and person, regardless of availability of the modes. Wardman et al. (2007) also did not specify any restrictions in the model estimation, but for forecasting they distinguish between shorter trips (<12km) and longer trips (>12km) and car availability. Kamargianni and Polydoropoulou (2013) applied very strict reasoning as they excluded individuals living more than 2.1 km away from their destination (i.e. school in their case), due to the unavailability of walking for longer distances. So they excluded the individuals to make all alternatives available to the entire sample, with the goal of matching the revealed preference data to their stated preference survey (which included all modes). The most detailed heuristics were introduced by Gehrke and Clifton (2014), who stated that if a person travels alone, a driver's license and car need to be available in the household for the car to be included in the choice set. If a person travels with others, this criterion is not effectuated as this person could travel as a passenger in a car that is not owned by anyone in the household. Regarding PT, they introduced a maximum allowable distance to the nearest stop criterion, which they set to 0.5 mile for bus and 1.0 mile for train. The bicycle also needs to be available in the household, but additionally they allowed for a maximum travel time of 2 hours, assuming that the cycling speed is 10 mph. For walking, they allowed for the same maximum travel time, with a speed of 3 mph.

In this study, the heuristics introduced by Gehrke and Clifton (2014) will be applied and adapted to the Dutch situation. Consequently, the PT, bicycle and walking heuristics are adapted. Gehrke and Clifton (2014) exclude PT trips for individuals who live further than a certain distance. In the Netherlands, people use a variety of access modes and as a result thereof the one-mile boundary is considered too small. In order to avoid a false exclusion of PT from the choice set, we choose to set no distance boundary to PT travel, but as mentioned in Section 3.3.1, the PT route should have been identified in the Google Directions API. For the active modes, a maximum travel time of 2 hours is too generous given the reported travel times in the data. The maximum reported travel time for cycling is 130 minutes, where 99% of the individuals have travel time lower than 60 minutes. For walking these values are respectively 75 minutes and 50 minutes. Therefore, it seems most plausible to adjust to the 99% travel time, as this captures the potential choice for active modes for the vast majority of individuals. Similar to Gehrke and Clifton (2014) this study will use the an equal limit for both modes, which is set to 60 minutes. Summarising, the following heuristics are introduced for identifying the mode choice set:

- Car: Driver (drivers' license and car available), Passenger (travelling with other individual(s));
- PT: Route identified in Google Directions API;
- Bicycle: Bicycle available and calculated travel time <= 60 minutes (mean speed = 16.7 km/h);
- On foot: Calculated travel time <= 60 minutes (mean speed = 4.8 km/h).

The results of implementing these heuristics, in terms of the choice set sizes per trip are reported in Table 3.4. 1.9% of the individuals are captive users for their trip, which are mainly PT trips, but also car and bicycle. These captive users do not influence the choice model, as they will have a 100% probability of choosing the only mode which is available to them. When not all modes are available to a person, mostly walking is excluded. This is due to the fact that the travel time on foot was longer than 1hr. For 31.9% of the trips all modes are available, which means that in these cases all of the above mentioned criteria are met.

Table 3.4: Mode choice set size per individual per trip

Mode choice set size	Frequency	Frequency of including mode in the choice set				
		Car	*PT*	*Bicycle*	*Walk*	***Total***
1	1.9%	14.3%	84.9%	0.8%	0.0%	100%
2	21.5%	74.3%	81.3%	28.3%	16.1%	200%
3	44.8%	84.9%	59.7%	99.1%	56.3%	300%
4	31.9%	100.0%	100.0%	100.0%	100.0%	400%

3.4.2. Discrete Choice Model Specification

In this study, three different mode choice models are estimated with an increasing level of complexity. First, we start with a Multinomial Logit (MNL) model with the utility function for alternative i and observation n at time t specified in the following way (Ben-Akiva and Bierlaire, 1999):

$$U_{int} = V_{int} + \varepsilon_{int}, i \in C_n \tag{3.1}$$

where V_{in} is the deterministic utility for alternative i (which is part of the choice set C_n) and observation n at time t and ε_{int} represents the random error term, which captures uncertainty and is independent and identically (i.i.d.) Gumbel distributed.

Second, since multiple trips per individual are observed, serial correlation can be expected in the error terms of one individual. We therefore test for a panel effect using a mixture of MNL models with a normal distribution of the panel effect error term $\alpha_{in} \sim N(0, \Sigma)$. The utility function for alternative i and observation n at time t is then adapted from Eq. 3.1 in the following way:

$$U_{int} = V_{int} + \alpha_{in} + \varepsilon'_{int}, i \in C_{nt} \tag{3.2}$$

where α_{in}represents the panel effect and ε'_{int} represents the random error term that is independent over observations and time and is i.i.d. Gumbel distributed.

Finally, the population is likely to exhibit taste heterogeneity. Therefore, the third and final model that is estimated is the Mixed MNL (MMNL) model. The utility function for the MMNL is specified according to Eq. 3.2, due to the expected presence of a panel effect. In this model the β's that are part of V_{int} are not fixed (like in the MNL), but varied over all individuals according to a predefined distribution (mostly a Normal distribution). All models are estimated using the Python Biogeme package (Bierlaire, 2016).

The MNL i.i.d. assumption may be violated in case unobserved variables in various alternatives are correlated. In mode choice literature this type of correlation is often found – for example in motorised versus active modes. This calls for the introduction of Nested Logit models (e.g. Barros et al., 2015). To identify whether this type of correlation is also present in the Dutch situation, we tested for the presence of an active mode nest using the Nested Logit model. However, we found no significant results for the nest, indicating that these alternatives are not highly correlated based on unobserved variables in this model.

In addition to MMNL, Latent Class Models (LCM) offer an alternative model structure for capturing taste heterogeneity. The reader is referred to Greene and Hensher (2003) and Hess (Hess, 2014) for a discussion and comparison of model properties and performance. MMNL, unlike LCM, requires Monte-Carlo simulations as part of the model estimation process. In contrast, in the LCM, each parameter needs to be estimated for each class, consequently significantly increasing the number of parameters and the computation time compared to the MMNL. Furthermore, LCM is more sensitive to data quality data, as potential limitations show faster (e.g. confounding, wrong signs or correlations). For this study, we experimented with both methods but experienced that the LCM model did not converge properly, therefore we adopt the MMNL model.

3.4.3. Model Estimation Process

In the estimation process, the significant variables identified in the best MNL model are used as input for the more complex models. The result of this process is that several parameters are found to be insignificant. Therefore, we have chosen to optimize the MMNL model with respect to model fit, by only including parameters that significantly increase the model fit (tested by means of a likelihood ratio test). This means that some insignificant parameters can be present in the model, but model fit decreases if these are fixed to zero.

In the MMNL model the car is the reference alternative, i.e. whenever dummy variables are used in the model, the parameter for the car is fixed to zero.

The comparison on model performance is tested by considering four criteria: the final log likelihood, the adjusted rho-square (compared to the equally likely model), the Bayesian Information Criterion (BIC), and the Akaike Information Criterion (AIC). The goal is to maximize the first two and minimize the latter two criteria.

3.5. Model Estimation Results

This section describes and discusses the results of the model with the highest performance, which is the MMNL model. Section 3.5.1 presents the results of the MMNL model, and compares the findings to general body of literature. The focus of the discussion is on the active modes. Section 3.5.2 addresses the performance of this model in and compares its performance to the other models that are estimated. Section 3.5.3 concludes with the most important findings in comparison to the already existing body of literature.

3.5.1. Identifying the Determinants of Mode Choice Behaviour

In this section, the determinants of mode choice behaviour, as identified in the model estimation, are presented and discussed. The results of the model estimation are presented in Table 3.5. Furthermore, model performance comparison is presented.

Alternative Specific Constants

The alternative specific constants capture the average influence on utility (compared to the reference) of the unobserved variables. The car is taken as the reference case. The bicycle constant is insignificant, which means that the unobserved variables do not favour the bicycle over the car. Public transport has a very negative constant, implying that unobserved characteristics favour the car over public transport. Furthermore, the constant for walking is positive and significant. Consequently, the unobserved variables favour walking over the car. This means that the observed variables impact walking more negatively compared to the car, as the mode shares suggest that the car is favoured over walking.

Table 3.5: Associations of individual, household, season and weather, trip, built environment, and work characteristics with the likelihood of choosing car, PT, bicycle, and walking (a = binary explanatory variable, b = reference alternative, - = not estimated, ** = significant at the 5% level, * = significant at the 10% level)

	Car[b]		PT		Bicycle		Walk	
	coef.	*t-test*	*coef.*	*t-test*	*coef.*	*t-test*	*coef.*	*t-test*
β_{Constant}	-	-	-46.30	-6.29**	0.08	0.16	5.35	6.57**
σ_{Panel}	-	-	-15.20	-5.8**	2.52	11.82**	2.27	8.09**
Individual characteristics								
$\beta_{\text{Student}}^{a}$	-	-	6.91	3.00**	-	-	-	-
$\beta_{\text{High education}}^{a}$	-	-	-	-	0.43	1.89*	-	-
$\beta_{\text{Transit subscription}}^{a}$	-	-	10.80	6.29**	0.54	2.24**	-	-
$\beta_{\text{Company car}}^{a}$	-	-	-	-	-3.22	-1.75*	-	-
$\beta_{\text{Mode used in high school}}^{a}$	2.04	3.72**	-	-	-	-	-	-
$\beta_{\text{Mode use last half year}}^{a}$	2.10	8.09**	7.92	6.81**	2.18	9.79**	2.95	7.91**
Household characteristics								
$\beta_{\text{Household members}}$	-	-	1.98	3.28**	0.21	2.42**	-0.46	-3.66**
$\beta_{\text{Children in household}}$	-	-	-4.01	-2.54**	-	-	-	-
$\beta_{\text{Medium household income}}^{a}$	-	-	-	-	-	-	0.44	1.09
$\beta_{\text{High household income}}^{a}$	-	-	-	-	-	-	0.60	1.75*
Season and weather characteristics								
$\beta_{\text{September}}^{a}$	-	-	-	-	-	-	1.02	2.73**
Trip characteristics								
$\mu_{\text{Travel time}}$	0.70	11.36**	0.24	8.92**	0.12	7.29**	-0.35	-6.97**
$\sigma_{\text{Travel time}}$	-0.27	-10.51**	0.15	7.18**	-	-	0.15	5.76**
$\beta_{\text{Weekday}}^{a}$	-	-	-	-	0.81	3.92**	-	-
$\beta_{\text{Peak hour departure}}^{a}$	-	-	-	-	-	-	-0.56	-2.39**
$\beta_{\text{Travel group size}}$	-	-	-1.80	-2.16**	-0.82	-6.62**	0.47	2.93**
$\beta_{\text{Leisure trip purpose}}^{a}$	-	-	-	-	2.49	6.68**	2.79	6.33**
$\beta_{\text{Work trip purpose}}^{a}$	-	-	6.51	4.95**	2.29	5.65**	2.09	3.57**
$\beta_{\text{School trip purpose}}^{a}$	-	-	10.80	4.95**	5.11	7.34**	-	-
$\beta_{\text{Shopping trip purpose}}^{a}$	-	-	-	-	1.10	3.31**	1.65	3.96**
Built environment								
$\beta_{\text{Amsterdam}}^{a}$	-	-	5.95	2.84**	-	-	2.17	3.55**
$\beta_{\text{Rotterdam}}^{a}$	-	-	-	-	-0.99	-2.18**	-	-
β_{Urban}^{a}	-	-	4.61	3.51**	-	-	-	-
$\beta_{\text{Suburban}}^{a}$	-	-	-	-	0.66	2.56**	-	-
$\beta_{\text{Garbage bins}}^{a}$	-	-	-	-	0.69	2.76**	0.85	2.37**
$\beta_{\text{Playgrounds}}^{a}$	-	-	-	-	-	-	-1.43	-3.16**
$\beta_{\text{Bicycle parking}}^{a}$	-	-	-	-	-	-	0.80	2.53**
β_{Shops}^{a}	-	-	-	-	0.63	2.22**	0.99	2.65**
$\beta_{\text{Public building}}^{a}$	-	-	5.59	2.58**	-	-	-	-
$\beta_{\text{Hospital/GP}}^{a}$	-	-	-	-	-0.23	-0.87	-	-
Work conditions								
$\beta_{\text{Travel compensation}}^{a}$	1.27	4.97**	17.6	4.57**	0.97	2.56**	-	-

Individual Characteristics

Previous research showed that gender and age are very relevant, especially in explaining bicycle mode choice (Fraser and Lock, 2011; Heinen et al., 2010; Muñoz et al., 2016a). Furthermore, Heinen et al. (2010) suggest that in cycling rich countries, such as the Netherlands, women cycle more often than men. In this study, we find that both gender and age are not explanatory variables. Cycling in the Netherlands is truly universal (Pucher and Buehler, 2008). Heinen et al. (2010) also mention that being native Dutch is positively associated with cycling. In this study we do not find a significant relationship between ethnicity and mode choice. The vast majority of our respondents is native Dutch (>95%), presumably this shows not enough diversity to distinguish between native and non-native Dutch citizens. Furthermore, we find that having completed a high level of education (college level) increases the utility for the bicycle, compared to having a lower education level. Regarding education, the general body of literature shows mixed results (Heinen et al., 2010; Muñoz et al., 2016a). Presumably, highly educated people in our sample are more aware of the health benefits related to cycling. A positive significant effect was expected for walking (e.g. Gehrke and Clifton, 2014), but was not affirmed. In line with the literature, being a student is positively related to cycling (Heinen et al., 2010).

Contrary to the majority of active mode choice studies, we find no significant relationship between the availability of car and bicycle in the household and mode choice (Fraser and Lock, 2011; Handy et al., 2014; Mitra, 2013). Pucher and Buehler (2008) show that car ownership has increased sharply in the Netherlands over the past decades, but this has not affected the use of bicycles, potentially explaining the absence of a significant relationship in this study. However, if an individual has a company car available (very small share of the sample), a significant reduction in bicycle utility is found. Trips with the company car are likely to replace trips by bicycle. Next to that, having a PT subscription positively relates to both the PT and cycling probability. The latter could be the result of access and egress transport (Handy et al., 2014; Winters et al., 2017), for which the bicycle is often used (respectively 50% and 10%) in the Netherlands (KiM, 2015). In this study we investigate the main mode choice for a trip. Consequently, the results suggest that the use of the bicycle as access and egress mode is positively associated with bicycle use in general.

Two variables related to past use of modes are tested, referring to the last year of high school and the last half year. For the former only car use has a significant and positive association with car choice. Using a car at an early age (either as passenger or driver) is thus associated with car use at a later age, whereas the other modes do not show this effect. Consequently, whether someone has cycled or walked to high school does not affect the current active mode use. For the latter, all modes test significant and increase the probability of choosing the respective mode, suggesting the formation of habits (Heinen et al., 2010).

Household Characteristics

Previous research has established the relationship between the size and composition of the household and mode choice (Hamre and Buehler, 2014; Heinen et al., 2010; Maley and Weinberger, 2011). The results of this study are mostly in line with previous research. However, regarding the number of children (<12 years) in the household, previous research mentions a negative association with active mode use (Heinen et al., 2010). While we find a significant negative association with PT use, no significant association for the active modes is manifested. This might be due to the Dutch context, as children often cycle from an early age (Pucher and Buehler, 2008). An increasing number of individuals in the household decreases the utility for walking, but increases the utility for cycling and public transport. Pucher and Buehler (2008) state that in the Netherlands cycling is most popular among children and adolescents. More individuals in a household generally means more children and adolescents, which can explain the positive association with bicycle use.

The relationship between household income and walking is not significant at the 95% level, but only at the 90% level. For cycling we find no significant relationship. Previous research has found mixed results for the relationship between income and active mode choice, where positive, negative, as well as insignificant results are reported (Handy et al., 2014; Heinen et al., 2010; Mitra, 2013; Muñoz et al., 2016a). This study therefore adds to these inconclusive results.

Season and Weather Characteristics

The month of travel or seasonality is known to influence active mode choice, with summer and autumn being the most favourable seasons (Böcker et al., 2013; Heinen et al., 2010). The data for this study was collected between September and November, hence late summer and autumn in the Netherlands. We find that September is positively associated with walking, which is in line with previous research. For cycling no significant relationship is identified. The Netherlands has a relatively mild climate (cool summers and warm winters), consequently it is possible that cycling is attractive all-year-round (Wang et al., 2016).

Furthermore, we find no relationship between extreme weather conditions and active mode use. Previous research has asserted that extreme or adverse weather is negatively associated with walking and cycling (Fraser and Lock, 2011; Wang et al., 2016). In the survey the respondents were asked if the weather conditions were extreme, therefore it reflects their subjective interpretation. The reason for not finding a relationship might again be due to the mild climate with frequent rain, which can be considered normal by the Dutch. Consequently, this study suggests that, in contrast with previous research, weather has limited impact on active mode choice.

Trip Characteristics

As mentioned above, most active mode choice literature investigates the impact of distance on active mode choice (Fraser and Lock, 2011; Handy et al., 2014; Heinen et al., 2010; Mitra, 2013; Muñoz et al., 2016a; Winters et al., 2017). Few studies investigated travel time (Heinen et al., 2010; Muñoz et al., 2016a). These variables are highly correlated, so we include only one of them: travel time, which is significantly associated with mode choice. For walking, PT and the car, heterogeneity ($tt_{\text{mode}} \sim N(\mu, \sigma)$) towards travel time is identified. Furthermore, the travel time parameters for car, PT and the bicycle are positive. Hence, the longer a trip (time-wise), the more likely that these modes are chosen. Generally, the literature finds negative associations between travel time and mode choice (Heinen et al., 2010; Muñoz et al., 2016a). However, Heinen et al. (2010) mention that travel times should always be considered in relation to other transport modes. Consequently, these positive values should be interpreted in the context of the modal share per travel time category (Figure 3.2). The shares of bicycle and walking decrease with higher travel times, whereas the shares of the car and public transport increase. Furthermore, as mentioned in Section 3.3.3 there are some differences between calculated and reported times. For example, for the car this means that delays are only present in reported travel time (traffic jams), which is the chosen alternative. The non-chosen alternatives do not register delays. Furthermore, an alternative model was estimated using only travel time and alternative specific constants. The estimation results are two positive travel time parameters (car and PT) and two negative (walking and cycling), which correspond to the modal shares as function of travel time. Finally, the travel time parameters can also stand in (proxy) for other unobserved variables (such as comfort), as suggested by Hess et al. (2005).

Summarising, these arguments explain the initially counterintuitive positive parameter values, but more research is needed to further underpin this finding.

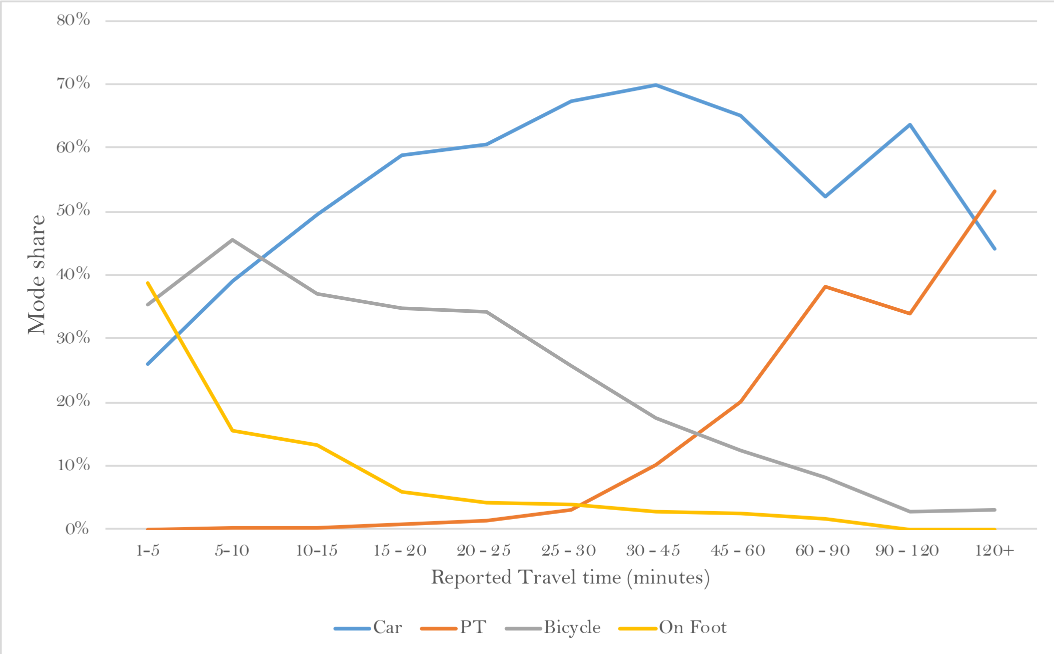


Figure 3.2: Modal share per travel time category

The number of trips made on a day is not significantly related to the mode chosen. Previous findings suggest that if more trips are undertaken on a given day, the car will be preferred over, for example, PT (e.g. Bhat, 1997). Potentially, due to the high active mode share in the Netherlands our dataset does not confirm this relationship.

The moment of travel is not mentioned in the walking and cycling mode choice literature reviews. We find that it has an explanatory power for both walking and cycling. Weekday travel has a positive association with cycling, but is not significant for walking. This is in line with findings from Denmark (Hansen and Nielsen, 2014). Peak hour travel relates negatively to walking, but does not relate to cycling. The latter might be due to time constraints (e.g. appointments or fixed work hours) for which walking is the least efficient mode in terms of speed.

The walking and cycling literature reviews also have not identified a relationship between the number of fellow travellers and mode choice. Our findings assert that for PT and cycling this relation is negative, whereas it is positive for walking. Hence, walking seems to be a good option when traveling in a larger group, but cycling and PT are less appreciated when traveling with other people.

Heinen et al. (2010) mention that the bicycle is more frequently used for recreational trips than for other trip purposes. In this study we investigate four trip purposes: work, school, leisure, and shopping. All trip purposes are associated with mode choice. For cycling all trip purposes are significant and positive, with the highest utility obtained for school trips, followed by leisure trips. For walking no effect is identified for the school purpose and the largest positive impact for leisure trips. Consequently, we can conclude that for the Netherlands more trip purposes are positively associated with active mode use other than only recreational trips.

Built Environment

The respondents were asked about the presence of nature, street furniture, buildings and traffic related aspects in their neighbourhood. The general body of literature has confirmed the relevance of these aspects regarding active mode choice (Fraser and Lock, 2011; Handy et al., 2014; Heinen et al., 2010; Mitra, 2013; Wang et al., 2016). Unlike previous studies, we do not find a relationship between nature and active modes for any of the tested elements (water, green, park). Note that we asked about the characteristics of the neighbourhood not the trip itself, which is mostly mentioned in the literature. Regarding street furniture both garbage bins (positive) and playgrounds (negative) are correlated with active mode choice. The first is in line with literature, however the second contradicts it. Playgrounds are generally found to increase walking and cycling. These variables presumably capture variables that are not directly included in the data (proxy). Playgrounds are more frequently located in sub-urban residential areas, where distances to for example the city centre are larger, consequently negatively correlated to walking. Playgrounds can also be used for walking around, which is mostly referred to in the literature, explaining the contradictory results found (Wang et al., 2016). In general, garbage bins are placed in areas where the streets are used as activity space, i.e. areas associated with a larger density of people passing by. Therefore, this variable could be positively associated with active mode choice. With respect to traffic related aspects, bicycle parking is significantly and positively correlated with walking. This relationship was previously identified for cycling, but not for walking (Heinen et al., 2010; Winters et al., 2017). Other traffic and infrastructure related aspects, such as cycle paths, traffic lights, and speed bumps for cars, do not exhibit explanatory power in this study. Previous research often emphasizes the need for infrastructure to boost active mode use (Fraser and Lock, 2011; Handy et al., 2014; Heinen et al., 2010; Mitra, 2013). In the Netherlands the density and continuity of active mode infrastructure is very high (Fishman, 2016; Pucher and Buehler, 2008). In line with these findings, previous bicycle route choice research in Amsterdam also identified no significant relationship with cycling infrastructure (Ton et al., 2017). Consequently, it is possible that in the Dutch context traffic related aspects are less important for mode choice compared to other countries. Finally, the presence of certain types of buildings (public buildings (e.g. library) and shops) is positively and significantly linked to the utility of the sustainable modes (PT, cycling and walking). This is in line with the literature, where mixed land use is described as one of the attractors of active mode use (Fraser and Lock, 2011; Heinen et al., 2010; Mitra, 2013; Muñoz et al., 2016a; Wang et al., 2016; Winters et al., 2017).

Furthermore, the five largest urban agglomeration areas in the Netherlands are included in the model estimation. Residing in Rotterdam or Amsterdam contributes to explaining mode choice. Amsterdam exercises a positive parameter for walking. Especially the city centre is inviting for walking as distances are short. Rotterdam has a negative parameter for cycling. It is a relatively car-oriented city, as it was reconstructed after the second world war (in the 70s) when the car was booming. Consequently, it is deemed logical that people cycle less, everything else being equal, in Rotterdam as it is less attractive compared to other locations.

Previous research also mentions the importance of population density or urbanisation level (Fraser and Lock, 2011; Heinen et al., 2010; Muñoz et al., 2016a; Wang et al., 2016). We investigate a combination of these, as we measure the density of inhabitants per square kilometre at the municipality level. Consequently, the urban level reflects high population density, which is mostly found in larger cities. Small and medium sized cities are positively correlated with cycling (Heinen et al., 2010). In line with previous research, we find a positive suburban parameter for cycling.

Work Conditions

Reimbursement for using a mode to work is associated with the mode choice in our sample. Receiving a compensation for traveling to work for a certain mode increases the probability of choosing that mode. The effect of the compensation is most pronounced for PT and smallest for the bicycle. This is in line with previous research (Handy et al., 2014; Muñoz et al., 2016a; Winters et al., 2017). Previous research also identifies cross-relationships, e.g. effect of car reimbursement on the use of bicycle (Handy et al., 2014; Winters et al., 2017), however we only investigate the direct relationship in this study.

Previous research suggests that part-time workers are more inclined to cycle compared to full-time workers (Heinen et al., 2010). In this study, we find no relationship between working hours and mode choice. This might be related to the fact that, unlike Heinen et al. (2010), we investigate all trips, not only commuting trips.

3.5.2. Model Performance Comparison

The MMNL model, which includes heterogeneity and panel effects performs best out of the three models on all performance indicators identified in Section 3.4.3 (see Table 3.6). The MNNL model cannot be compared using the likelihood ratio test as some of the parameters in the MNL and Panel effect model were removed due to insignificant results. The removed parameters are BMI (bicycle), September and October (bicycle), number of trips (PT), sport centre (bike), student (bicycle) and industrial area (walk). The model has a relatively high fit compared to the equally likely model. A total of 54% of the proportion of information in the choice data is explained by the model.

Table 3.6: Performance indicators of estimated models (= significant at the 5% level)**

	MNL	Panel effect	MMNL
Initial log likelihood	-6,893.85	-6,893.85	-6,893.85
Final log likelihood	-3,535.89	-3,323.88	-3,110.15
Parameters	61	64	60
Sample size	6,368	6,368	6,368
Adjusted rho-square	0.478	0.509	0.540
BIC	7,606.08	7,208.35	6,745.85
AIC	7,193.78	6,775.77	6,340.31
Likelihood ratio test	-	424.02**	-

3.5.3. Conclusions on Model Estimation Results

The most important findings of this study that are contradicting the general body of literature are summarised here. These contradictions are often attributed to the context of the study, as most literature stems from countries with low cycling penetration (e.g. the USA). Consequently, it could be expected that different determinants influence the mode choice and that their impact differs.

Socio-demographics variables are found less important in explaining active mode choice compared to literature. We find no significant relationship between gender, age, and ethnicity and active mode use. Car and bicycle availability on household level do not influence mode choice, whereas the existing body of literature identifies this as an important variable. Having children is not significantly related to active mode use, whereas other studies suggest a negative relationship. Weather characteristics are not relevant for active mode choice, which contradicts the general body of literature. It could be due to the way we formulated the characteristics (experience of extreme weather), but it could also be that in the Netherlands, due to the mild climate with frequent rain, weather does not impact mode choice. Travel time has a positive association with cycling, which in comparison to the other modes could be explained by the relationship between modal shares in different travel time categories. We find that the travel group size and moment of travel are relevant for mode choice, which has not been identified insofar in the active mode literature. Active mode infrastructure was found to be of limited relevance in explaining active mode choice, whereas literature states this as important determinants for cycling and walking. Mixed land use is found important for active mode use.

3.6. Discussion

This section addresses the relative importance of determinants, discusses their impact on walking and cycling, and reflects on their market shares.

The relative importance of mode choice determinants is displayed in the form of the mean and range of the product of a coefficient and the corresponding variable value. This enables the identification of where the potential lies for increasing the mode shares of walking and cycling. Table 3.7 presents the mean influence of each determinant on the mode choice and the range in terms of utility points. The mean influence for a dummy variable depends on the number of cases in which the value is one, if this is more than half of the cases, the mean is set to one, otherwise it is set to zero. For example, the interpretation of the results displayed in Table 3.7 for the impact of transit subscription on bicycle choice is the following: the range of [0.0,0.5] means that either there is zero impact (absence of transit subscription) or the impact is 0.5 utility points (presence of transit subscription). Furthermore, the mean is zero implying that more than half of the population does not have a transit subscription and overall the impact is zero. In case of other (continuous) independent variables, such as the impact of household size on walking choice, the interpretation is the following: the range of [-4.1,-0.5] means that additional persons in the household influence the choice for walking negatively, where the potential impact ranges between -0.5 (1 person) and -4.1 (9 persons). On average the impact is -0.9, meaning that the average household size is just under two. The values reported in Table 3.7 can thus be used for comparing the importance of each variable with respect to other variables for each mode alternative.

The parameters associated with PT are all relatively high in comparison to the other modes. These high parameters are needed to compensate for the very low alternative specific constant (-46.30) so that it becomes attractive for certain trips and individuals. Most likely, this is the result of the very low share of PT in the sample (5.3%), which makes higher coefficients necessary. Consequently, the determinants that are related to PT have a high mean and range of impact on the PT choice. Potentially, studies investigating mode choice in a car and PT rich environment will find different results for the PT parameters, compared to our study in a context dominated by car and bicycle travel.

Travel time is the most dominant determinant, given all determinants of all modes. The range of travel times for car and PT are larger compared to the active modes, due to their dominant role in long distance travel. This means that the range of impact for these modes is also higher. The impact on cycling and walking is comparable in size, albeit with different parameter signs. Almost all studies on active mode choice take distance or travel time into account (Fraser and Lock, 2011; Handy et al., 2014; Heinen et al., 2010; Mitra, 2013; Muñoz et al., 2016a; Winters et al., 2017). Most studies focus on distance, but as we have mentioned before, distance and travel time are highly correlated and should not be included simultaneously. Often, distance and travel time are related to built environment characteristics, such as mixing of land use and population density. Consequently, our finding of travel time being the most dominant determinant is in line with the general body of literature.

Table 3.7: Mean and range of influence of determinants on mode choice in terms of utility points (a = binary explanatory variable, - = not estimated)

	Car		PT		Bicycle		Walk	
	Mean	*Range*	*Mean*	*Range*	*Mean*	*Range*	*Mean*	*Range*
Individual characteristics								
Student [a]	-	-	0.0	[0.0,6.9]	-	-	-	-
High education [a]	-	-	-	-	0.0	[0.0,0.4]	-	-
Mode used – high school [a]	0.0	[0.0,2.0]	-	-	-	-	-	-
Mode used – last half year [a]	0.0	[0.0,2.1]	0.0	[0.0,7.9]	0.0	[0.0,2.2]	0.0	[0.0,3.0]
Transit subscription [a]	-	-	0.0	[0.0,10.8]	0.0	[0.0,0.5]	-	-
Company car [a]	-	-	-	-	0.0	[-3.2,0.0]	-	-
Household characteristics								
Household size	-	-	4.0	[2.0,17.8]	0.4	[0.2,1.9]	-0.9	[-4.1,-0.5]
No. children	-	-	0.0	[-20.1,0.0]	-	-	-	-
Medium household income [a]	-	-	-	-	-	-	0.0	[0.0,0.4]
High household income [a]	-	-	-	-	-	-	0.6	[0.0,0.6]
Season and weather characteristics								
September [a]	-	-	-	-	-	-	0.0	[0.0,1.0]
Trip characteristics								
Travel time	13.6	[0.0,213.4]	7.9	[0.0,109.9]	2.3	[0.0,15.3]	-5.5	[-56.4,0.0]
Weekday [a]	-	-	-	-	0.8	[0.0,0.8]	-	-
Peak hour departure [a]	-	-	-	-	-	-	0.0	[-0.6,0.0]
Travel group size	-	-	0.0	[-9.0,0.0]	0.0	[-4.1,0.0]	0.0	[0.0,2.4]
Leisure trip purpose [a]	-	-	-	-	0.0	[0.0,2.5]	0.0	[0.0,2.8]
Work trip purpose [a]	-	-	0.0	[0.0,6.5]	0.0	[0.0,2.3]	0.0	[0.0,2.1]
School trip purpose [a]	-	-	0.0	[0.0,10.8]	0.0	[0.0,5.1]	-	-
Shopping trip purpose [a]	-	-	-	-	0.0	[0.0,1.1]	0.0	[0.0,1.7]
Built environment								
Live in Amsterdam [a]	-	-	0.0	[0.0,6.0]	-	-	0.0	[0.0,2.2]
Live in Rotterdam [a]	-	-	-	-	0.0	[-1.0,0.0]	-	-
Live in Urban area [a]	-	-	4.6	[0.0,4.6]	-	-	-	-
Live in Suburban area [a]	-	-	-	-	0.0	[0.0,0.7]	-	-
Garbage bins [a]	-	-	-	-	0.7	[0.0,0.7]	0.9	[0.0,0.9]
Playgrounds [a]	-	-	-	-	-	-	-1.4	[-1.4,0.0]
Bicycle parking [a]	-	-	-	-	-	-	0.0	[0.0,0.8]
Shops [a]	-	-	-	-	0.6	[0.0,0.6]	1.0	[0.0,1.0]
Public buildings [a]	-	-	5.6	[0.0,5.6]	-	-	-	-
Hospitals/GP's [a]	-	-	-	-	-0.2	[-0.2,0.0]	-	-
Work conditions								
Travel compensation [a]	0.0	[0.0,1.3]	0.0	[0.0,17.6]	0.0	[0.0,1.0]	-	-

The impact of determinants on active modes is generally comparable in size, albeit each mode is influenced by a different set of determinants. Very often, in literature, either cycling or walking is investigated (Buehler and Dill, 2016; Fraser and Lock, 2011; Handy et al., 2014; Heinen et al., 2010; Muñoz et al., 2016a) or cycling and walking are treated as being very similar (e.g. active mode travel or physical activity) regarding for example policy development purposes (Mitra, 2013; Wang et al., 2016; Winters et al., 2017). Importantly, this study shows that walking and cycling are influenced by different determinants. Consequently, in policy development it is wise to separate both modes, as otherwise the desired effect might not be reached. In model estimation, this means that both the active modes should be distinguished. When going into more detail in the categories of determinants that influence each mode, we see that the impact of individual characteristics is much stronger for cycling than for walking. Although literature assigns more importance to the individual characteristics, compared to our findings (Handy et al., 2014; Heinen et al., 2010). Moreover, household characteristics are more important in explaining the choice to walk than to cycle. Consequently, it is deemed plausible that the cycling literature reviews did not find household characteristics to be mode choice determinants (Fraser and Lock, 2011; Handy et al., 2014; Heinen et al., 2010; Muñoz et al., 2016a). Trip characteristics influence both walking and cycling, and we find similar impact sizes. Only trip purpose affects cycling more than walking, as more trip purposes seem to be relevant for cycling. Finally, even though different characteristics related to the built environment influence walking and cycling, the overall impact seems comparable in size.

The categories of determinants that have the largest influence on cycling are trip characteristics, individual characteristics and built environment. For walking, the respective categories are trip characteristics, built environment and household characteristics. Consequently, policy measures aimed at increasing the level of walking and cycling are most likely to influence modal usage by targeting trip characteristics. Directly targeting trip characteristics is unfortunately not possible. However, based on our findings, investing in a more liveable built environment may benefit active modes too. For example, by creating a mixed land-use environment with residential and recreational areas, which are equipped with street furniture. Finally, the individual and household characteristics cannot be influenced by means of active mode policy. However, they can provide insight into which segments of the population to target, for example large families or people with high education, which are most prone to response to changes in cycling and walking attributes. The two categories of determinants that have the most limited effect on active mode choice are season and weather characteristics and the work conditions. However, this could be due to the way we define these characteristics and may differ in contexts which exhibit conditions that are not prevalent in the Dutch context (Böcker et al., 2013; Heinen et al., 2010; Muñoz et al., 2016a).

When calculating the impact of altering some of the variables (e.g. as the result of policy measures or campaigns) on the market shares, we see that the trade-off is mostly between the car and active modes. PT shares mostly remain consistently low. The base scenario for this population predicts the following market shares: 44.8% car choice, 1.6% PT choice, 43.5% bicycle choice, and 10.0% walking choice. When altering variables, these shares change. As a first example, if everyone would get a company car, all else being equal, the market shares of the car and walking would increase by 13.5 and 7.8 percentage point respectively, while it would reduce the share of cycling by 21.3 percentage point and not affect PT use. If, as a second example, everyone would be reimbursed for cycling to work, all else being the same, the market shares would change as follows: the share of the car and walking would decrease by 4.1 and 2.2 percentage point respectively, while increasing the bicycle share by 6.3 percentage point and again no effect on the PT share. A third example is providing a transit subscription for everyone, all else being the same. The market share for PT remains again unaffected. The bicycle share increases by 2.8 percentage point, while the share of the car and walking decrease by respectively 1.9 and 0.9 percentage point. Winters et al. (2017) state that literature suggests that promoting PT results in higher bicycle share, which is confirmed in this exercise. These market share calculations suggest that the car and walking act as complementary modes, whereas the car and bicycle are competing modes. Literature often finds that the car and PT are competitors (e.g. Ye et al., 2007). Braun et al. (2016) find that providing incentives not to use the car to work increases PT use instead of active mode use. These studies often use data from countries with low bicycle shares, consequently the two modes with the highest market shares are the car and PT, making them competitors. This also shows that policy measures and incentives, taken from studies in low bicycle countries or cities, cannot be directly transferred to other contexts, such as the Netherlands.

3.7. Conclusions and Future Research Directions

This paper presents the findings of a mode choice model for the Netherlands, focusing on active modes while including a more comprehensive set of modes (i.e. car, public transport, bicycle, walk), aimed at understanding the determinants of choosing a mode in relation to the other modes. The Netherlands has a very high active mode penetration, a safe environment for walking and cycling, mature and complete active mode infrastructure, and a demographically diverse population of active mode users (Fishman, 2016). Consequently, investigating mode choice in the Netherlands can enrich active mode choice literature, which mostly refers to contexts where cycling is an uncommon mode of transport. We compared our findings in terms of the determinants of influence on active mode choice and their relative importance. Choice models were estimated based on travel diary and survey data of the Netherlands Mobility Panel enhanced with a survey that addressed among other things the built environment (coined PAW-AM) comprising of 6,368 trips performed by 1,874 individuals in the year 2016.

Based on a review of the literature on the determinants that influence active mode choice, a total of six categories of determinants were identified: individual characteristics, household characteristics, season and weather characteristics, trip characteristics, built environment and work conditions. In line with previous studies, our findings suggest that determinants belonging to all categories are relevant for explaining modal choices while the extent of their influence varies for the different modes. Contrary to the existing body of literature, we find that the socio-demographic determinants are less important in explaining active mode choice. We do not find significant relationships for gender, age, and ethnicity. This most likely stems from the diverse cycling population in the Netherlands, whereas in other countries often young males are most likely to cycle (Fishman, 2016; Heinen et al., 2010). Furthermore, we find that weather is a less important determinant than suggested in the literature. While previous research reported that rain, high and low temperatures, and hot and cold climates negatively affect active mode use, we do not find a significant relationship between weather and active mode use. This might be due to the definition of our weather variables (i.e. perceived extreme weather), however even the extreme weather does not seem to affect Dutch active mode users. Finally, we conclude that active mode infrastructure, such as bicycle paths, does not influence active mode choice. This contradicts the main body of literature, however given that the active mode infrastructure in the Netherlands is already mature and complete (Fishman, 2016), it is possible that this does not anymore affect the choice among travel alternatives.

Even though all categories of determinants are included in this research, not all potential explanatory variables are included in this study. Especially, determinants related to the work conditions could be enriched in future studies. Furthermore, determinants that are related to the social surroundings of individuals, thus opinions/attitudes and behaviour of the people around the individual, could be investigated, as previous research has showed the potential of these factors (Heinen et al., 2010; Muñoz et al., 2016a). However, if attitudes and opinions are included, different models need to be estimated that can accommodate subjective variables (such as hybrid models (Vij and Walker, 2016)). Furthermore, the built environment is now related to the neighbourhood of the respondent. This could be enriched with information about the neighbourhood of the destination or along the trip.

In most mode choice studies walking and cycling are conjoined. In this study, the presence of an active mode nest was investigated, which represents correlations between modes, suggesting a hierarchical choice structure. No such structure could be identified. Furthermore, the determinants that influence walking and cycling are distinctive. The individual characteristics mostly affect cycling, whereas the household characteristics mostly influence walking. Roughly the same trip characteristics influence both active modes, however their overall impact is different. Besides, regarding the built environment both active modes are influenced, but by different determinants. Both the absence of an active mode nest and the identification of different parameters for walking and cycling suggest that cycling and walking should be considered and treated as two independent alternatives.

From a policy perspective, several of the categories of determinants can be directly or indirectly influenced. If the goal is to stimulate the use of active modes, the trip characteristics and built environment are the most relevant categories. The trip characteristics can only be influenced indirectly. For example, by adapting the infrastructure in such a way that travel time savings are generated (e.g. more crossing locations for pedestrians, so that waiting time is reduced; more passages for pedestrians between street blocks, reducing walking distances, especially in suburban areas). The built environment can be influenced directly, for example by creating mixed land use environment where residential and recreational areas are combined, which are equipped with street furniture. Because walking and cycling are influenced by different determinants, different policy measures might be needed to influence each of the modes. Consequently, targeting active modes might not provide the desired result. Furthermore, this study suggests that the car and bicycle are competitors, whereas in cycling poor countries it is likely that the car and public transport are competitors. Transferring the results to other countries should thus be done with care, also because both this study and the already existing body of active mode choice literature suggest that different determinants are important and to a different extent, for cycling rich and cycling poor countries.

From a modelling perspective, the use of revealed preference data seems promising, given the results that show that there is sufficient variability in the data (otherwise more parameters would have been insignificant). Furthermore, individuals show clear preferences towards modes that are influencing their decision (homogeneity of choice within the individual) based on the panel effect estimates. Finally, the mixed multinomial logit model shows a significant improvement compared to the most commonly used multinomial model, allowing to capture taste heterogeneity towards determinants.

Additional future research directions entail collecting longitudinal data for all identified determinants, so that the causal relationship can be identified. This helps policy-makers by ensuring that investments are most impactful in achieving policy goals. Moreover, longitudinal data can be instrumental in modelling more explicitly how previous experiences with each mode influence current behaviour, e.g. by using Markov chains, which capture the dependency towards previous choices. Next to that, the inclusion of a cost variable in the model could help understand trade-offs between modes in monetary terms. Furthermore, this study only covers trips consisting of a single mode. For a better and more complete picture of the entire mode set also multimodal trips should be included.

Acknowledgements

This research was supported by the Allegro project (no. 669792), which is financed by the European Research Council and Amsterdam Institute for Advanced Metropolitan Solutions. The data was made available by the Netherlands Mobility Panel administered by KiM Netherlands Institute for Transport Policy Analysis.

Chapter 4 – Determinants of the Experienced Mode Choice Set

This chapter is based on the following submitted manuscript:

Ton D., Bekhor S., Cats O., Duives D.C., Hoogendoorn-Lanser S., and Hoogendoorn S.P. (submitted). The experienced mode choice set and its determinants: commuting trips in the Netherlands.

Abstract

Active modes take up an increasingly important place on the global policy-making agenda. In the Netherlands, a country that is well-known for its high shares of walking and cycling, the government aims at achieving a modal shift among 200,000 commuting car drivers towards using the bicycle. To this end, policy measures need to be introduced. When the aim is to achieve a modal switch over an enduring period of time, it is more relevant to know the likelihood of including or excluding a mode in the mode choice set, compared to choosing a mode for a single trip. Therefore, we investigate the formation of the experienced choice set (set of modes used over a long period of time), where the aim is to identify determinants that influence the inclusion or exclusion of a mode in this set. We estimate discrete choice models, based on survey data from the Netherlands Mobility Panel (MPN) and a complementary survey, where individuals were asked to report the frequency of using certain modes of transport for commuting trips over the course of half a year. This study shows that the experienced choice set for commuting is unimodal for the majority of the individuals, and remains constant over

time for most individuals. Reimbursement by the employer for using a certain mode is the most important determinant influencing the experienced choice set, followed by ownership characteristics and urban density. We show that the mode choice set formation depends on more determinants than previously assumed.

4.1. Introduction

Due to increasing urbanisation rates accompanied by growing transportation demands, governments worldwide have been increasingly interested in active modes of transport, i.e. walking and cycling. These modes can reduce congestion and emissions when replacing the car (standalone or in combination with public transport). Furthermore, these modes provide health benefits for the individual. The Netherlands has a very high share of active mode use, with 27% of trips made by bicycle and 17% on foot (CBS, 2018). Notwithstanding, the Dutch government aims at achieving a mode switch from the car to the bicycle for 200,000 commuters (Rijksoverheid, 2018). This aim is supported by the fact that about half of the commuting trips currently travelled by car are shorter than 7.5 km. Consequently, these could be travelled by bicycle. To ensure this aim can be reached, policy measures need to be developed that lead to mode switches among commuters.

Investigating mode use requires observing individuals over a period of time. Kuhnimhof et al. (2006) investigated mode use in Germany using a 7-day travel diary. They found that during the first three to four days the number of modes used increased significantly and thereafter stabilised. Heinen and Chatterjee (2015) investigated the mode use variability of individuals in the UK using a 7-day travel diary. They found that 44% of the respondents were unimodal, the other 56% used a variety of modes over the course of seven days. Ralph (2017) used three years of information from a 1-day travel diary in the USA and clustered young adults based on their mode use. She found that 86% were unimodal car user, of which 4% used the car for long distances. Only 3% of the individuals were multimodal, meaning that they used both car and public transport. A further 11% were car-less, meaning they relied on public transport when they travelled. Ton et al. (2019b) investigated the mode use of individuals in the Netherlands using a 3-day travel diary, where they clustered the individuals based on their frequency of using each mode per day into five clusters. Two clusters were unimodal, namely car only (27%) and bicycle only (9.5%). One cluster (12.3%) used all modes (car, public transport, bicycle, and walking) on a daily basis. The other two multimodal clusters were car and bicycle (27.5%) and car, bicycle, and walking (23.7%). In sum, the number of modes used increases with the number of days for which a travel diary is available, but generally only a selection of modes is used.

When zooming in to the commuting trip purpose, even less variability in the mode use over time has been observed. Kuhnimhof (2009) found that for commuting most individuals repeatedly uses the same mode (72%) over the course of 7-days This is supported by Hensher and Ho (2016), who found that use of a mode increases the likelihood of using that mode again. However, the variation in modes between individuals is higher for commuting compared to other trip purposes (Kuhnimhof et al., 2006). Lavery et al. (2013) investigated commuting trips to McMaster University in Canada. They asked individuals for their primary commuting mode and alternatively asked which modes the respondents considered available/feasible for their commute. When considering both used and available modes, the unimodality varies between 9% for active modes to 55% for public transport. A total of 51% of the individuals states that two modes are used and/or available, 37% mentions three modes, and 4% mentions four modes. Many of these results show routines or habitual behaviour regarding commuting trips.

When the aim is to identify individuals that might switch modes for their commute trips it is essential to understand the composition of their experienced choice set (modes used over a long term) and which determinants drive the inclusion or exclusion of modes from this set. In common mode choice research, discrete choice models are estimated to identify which determinants are associated with mode choice and predict the likelihood of using a mode for a certain trip. Here, the choice set (i.e. between which modes a commuter chooses) is often defined based on deterministic rules, related to travel time, distance, ownership or availability of modes (De Jong et al., 2007; Gehrke and Clifton, 2014; Hamre and Buehler, 2014; Kamargianni and Polydoropoulou, 2013). However, for example Kuhnimhof (2009) and Lavery et al. (2013) show that these assumptions in both science and practice, might not cover all observed behaviour. Consequently, we propose to use discrete choice models to estimate the determinants of the experienced choice set that predict the likelihood of including or excluding a mode for commuting purposes.

In this study we investigate the experienced choice set of individuals in the Netherlands, by observing their mode use for commuting trips over the course of half a year. The data used for identification of the experienced choice set is the Netherlands Mobility Panel (MPN) survey data featuring Dutch-speaking individuals from the Netherlands. This dataset contains personal and household data, and is enriched with a survey that investigates, among others, the modes used by the respondents for commuting trips over a time period of half a year. We identify determinants that are relevant for inclusion of a mode in the experienced choice set by applying discrete choice models. These determinants can be used to identify policy measures that aim to realise a mode switch from the car to the bicycle. Furthermore, we show that more determinants are relevant in the estimation and prediction of the mode choice set than often assumed in mode choice models.

The remainder of this paper is organised as follows. Section 4.2 details the literature on determinants used for identifying the mode choice set. In Section 4.3 the methodology is discussed, together with the modelling approaches applied. Section 4.4 describes the data and shows an overview of the dataset. Then, in Section 4.5 the experienced mode choice set is elaborated upon. Section 4.6 covers the model estimation and validation results. Section 4.7 discusses the experienced choice set in relation to past research and potential applications. Finally, the paper is concluded in Section 4.8 and recommendations are provided.

4.2. Determinants of Mode Choice Sets

In literature, various determinants are used for the specification of the mode choice set. Sometimes studies rely on self-reported availability of modes as perceived by respondents (Lavery et al., 2013; Whalen et al., 2013), thus not relying on determinants for the formation of the choice set. Table 4.1 presents an overview of the determinants that are used to specify the mode set and identifies the operationalisation of these determinants as mentioned in the literature.

Table 4.1 shows that determinants can roughly be divided into four categories: availability of modes, trip characteristics, network characteristics, and individual characteristics. The first two categories are most common in the literature. Contrary to the trip characteristics, the determinants related to availability can be regarded both at the trip- and individual level. Travel time and distance, both trip characteristics, are operationalised in various ways, for example per mode, as aspect of the complete trip or in general. Some studies have incorporated individual characteristics to determine mode availability (Calastri et al., 2017; Vij et al., 2017, 2013). These studies have applied latent class models, where the individual characteristics are used to determine mode availability based on class membership. Calastri et al. (2017) showed that including individual characteristics significantly improves the model fit. Consequently, we expect that also individual and household characteristics are relevant determinants of the experienced mode choice set.

Table 4.1: Operationalisation of determinants used to specify the mode choice set for individuals

Determinant	Operationalisation	Studies
Availability of modes	Car available	(Ben-Akiva and Boccara, 1995; Gehrke and Clifton, 2014; Hensher and Ho, 2016; Kamargianni, 2015; Kamargianni et al., 2015; Rodríguez and Joo, 2004; Ton et al., 2019a; Vij et al., 2017, 2013)
	Driver's license	(Gehrke and Clifton, 2014; Habib et al., 2011; Kamargianni et al., 2015; Ton et al., 2019a)
	Bicycle available	(Gehrke and Clifton, 2014; Kamargianni et al., 2015; Ton et al., 2019a; Vij et al., 2017)
	PT available	(Hensher and Ho, 2016)
	PT subscription (e.g. season pass)	(Vij et al., 2013)
Maximum distance	General maximum distance	(Bhat, 1995; Kamargianni and Polydoropoulou, 2013; Wardman et al., 2007)
	PT stop within walking distance	(Ben-Akiva and Boccara, 1995; Gehrke and Clifton, 2014)
	PT distance	(Rodríguez and Joo, 2004)
	Walking distance	(Calastri et al., 2017; Habib et al., 2011; Kamargianni, 2015; Kamargianni et al., 2015; Rodríguez and Joo, 2004)
	Bicycle distance	(Calastri et al., 2017; Habib et al., 2011; Kamargianni, 2015)
Maximum travel time	Walking time	(Gehrke and Clifton, 2014; Swait and Ben-Akiva, 1987a; Ton et al., 2019a)
	Bicycle time	(Gehrke and Clifton, 2014; Ton et al., 2019a)
	Car time	(Swait and Ben-Akiva, 1987a)
	PT time	(Swait and Ben-Akiva, 1987a)
	Threshold for inclusion of mode	(Cantillo and de Dios Ortúzar, 2005)
Cost	Threshold for inclusion of mode	(Cantillo and de Dios Ortúzar, 2005)
Network connection	Route available using a mode	(Swait and Ben-Akiva, 1987a; Ton et al., 2019a)
	PT stops in neighbourhood	(Habib et al., 2011; Kamargianni et al., 2015)
Individual characteristics as input for mode availability	Gender	(Calastri et al., 2017; Vij et al., 2017, 2013)
	Age	(Calastri et al., 2017; Vij et al., 2017)
	Education level	(Calastri et al., 2017)
	Income	(Vij et al., 2017, 2013)
	Employment	(Vij et al., 2017, 2013)
	Household size	(Vij et al., 2017, 2013)
	Children in household	(Vij et al., 2017)
	Marital status	(Vij et al., 2017, 2013)
	Parenthood status	(Vij et al., 2017, 2013)
	Home ownership	(Vij et al., 2017, 2013)

4.3. Methodology

The experienced choice set is defined as the set of modes used over an enduring period of time. In Section 4.3.1 we present the methodology for retrieving the experienced choice set and identifying the relevant determinants. The model structures used for estimation and validation of the experienced choice set are discussed in Section 4.3.2.

4.3.1. The Experienced Mode Choice Set and its Determinants

The experienced choice set can be retrieved in different ways. One can, for example, observe an individual over a long period of time using a GPS device or a travel diary, which is time consuming and largely impacts the privacy of the individual. This has been done before to study mobility patterns of individuals, where the duration of these data collection efforts range between one day (Ralph, 2017) and six weeks (Vij et al., 2017, 2013). Another method, which is less demanding on the individual, is to use a survey to ask questions related to the mode use of an individual over a long period of time. This method has previously been applied to study mobility patterns of individuals (Lavery et al., 2013; Molin et al., 2016). We apply the latter method and use a survey to collect data (see Section 4.4). The question posed to the respondents is: *which modes have you used at least once in the last half year for commute trips*? Where they could choose multiple modes from a list of the most prominent modes in the Netherlands, namely car, train, bus/tram/metro (BTM), bicycle, and walking. Access and egress modes are not included here. This question provides insights in the modes used by the respondents over a long period of time. By focusing on commuting, i.e. one trip purpose, it is easier for individuals to retrieve their mode use. This question collects aggregated data, consequently the experienced choice set reflects the general experienced choice set and does not directly represent trip-level variations.

The experienced choice set reflects actual observed behaviour. Therefore, it provides a rich source of information, both in terms of choices made by individuals and the composition and size of their experienced mode choice set. We propose to apply discrete choice models to identify determinants that influence the experienced choice set of an individual. The alternatives of the experienced choice set are constructed by combining all historically observed modes into a single alternative. The respondent then chooses between sets of modes, e.g. car-train-walk and bicycle-car. This results with 31 potential experienced choice sets. Given that, for example, individual and household determinants are previously applied and found relevant for choice set specification, we will test the relevance of a number of determinants belonging to ownership, socio-demographics, household characteristics, urban density, and work conditions.

4.3.2. Model Structures Used for Estimation and Validation

The determinants influencing the experienced choice set are revealed by estimating a number of discrete choice models. We start simple by estimating Multinomial Logit (MNL) models. Due to the way in which alternatives are constructed, we expect shared unobserved variables (captured in the error term). This cannot be captured in the MNL model, thus requiring the use of various more complex model structures such as Nested Logit (NL), Cross-Nested Logit (CNL), and Mixed Logit (ML). The utility function for experienced choice set C and individual n is specified according to Eq. 4.1 (Ben-Akiva and Bierlaire, 1999):

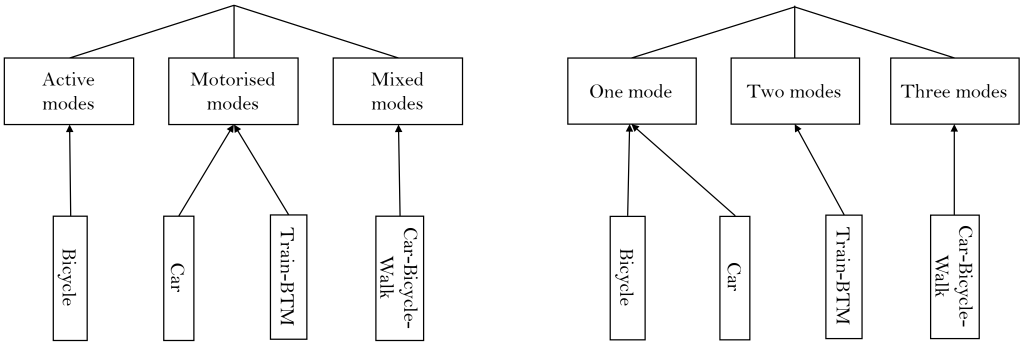
$$U_{Cn} = V_{Cn} + \varepsilon_{Cn}, \qquad C \in G_n \tag{4.1}$$

where V_{Cn} is the deterministic utility for individual n and experienced choice set C, which is part of the feasible choice set of that individual G_n, and where ε_{Cn} represents the random error term capturing the uncertainty. The deterministic part of the utility is composed in the following way for the experienced choice set (the index of individual n is omitted for simplicity):

$$V_C = ASC_C + \sum_{m \in C} \beta_m x_m \tag{4.2}$$

where the alternative specific constant (ASC) is defined per experienced set, and where the parameters are estimated for each mode m that is a member of the experienced set C. As an example, if the alternative is bicycle-walk, this means that for each variable x two parameters are estimated, one for the bicycle and one for walking. A model with alternative specific parameters was estimated, but yielded many insignificant parameters, consequently losing explanatory power for various alternatives.

As mentioned before, the MNL model is only able to capture similarities between alternatives via the observed variables. The NL allows correlations between the unobserved variables of some alternatives, by grouping (nesting) them. In case of the experienced mode choice set, several of these nesting structures can be identified, related to for example size and composition of choice sets. An example of a nesting structure based on composition is shown in Figure 4.1a, where alternatives that contain active modes, motorised modes, or a mixture of these are distinguished (note that this representation does not imply hierarchy). All alternatives can be assigned to one nest. Alternatively, a nesting structure based on size represents the number of modes that are combined into the experienced choice set (Figure 4.1b). A variety of different nesting structures are tested and judged on model fit and behavioural interpretation.



(a) (b)

Figure 4.1: Example of NL structures based on composition (a) and size (b) of the experienced choice set

Due to the way the alternatives are bundled, i.e. combining all used modes into one experienced choice set, it seems plausible that alternatives that have modes in common share correlation in the unobserved variables. For example, the car-train and car-bicycle alternatives might exercise correlations due to the common car-mode. It is not possible to capture this structure in a NL model since nests are mutually exclusive. The CNL model relaxes this assumption by including alternatives as members of multiple nests (Vovsha, 1997). An example of a structure based on mode-nests is shown in Figure 4.2. Often, the membership of a nest is predefined (Bierlaire, 2006), but can also be estimated together with the nesting parameters. This optimises the CNL model further as the degree of membership can vary between alternatives and nests. Again, a variety of different nesting structures was tested and judged on model fit and behavioural interpretation.

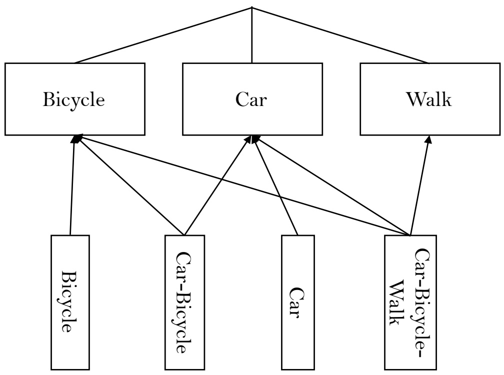


Figure 4.2: Example of a CNL structure based on mode-nests

ML with error component structure has a flexible error structure, and is theoretically able to reproduce the same structure as both the CNL and NL models (McFadden and Train, 2000). It has as extra advantage that it is able to incorporate heterogeneity and heteroscedasticity that can be present in the population.

In the model estimation process, first all variables and parameters are introduced for all alternatives. Afterwards, non-significant parameters were excluded iteratively, so that the model fit (adjusted rho-square compared to the equally likely model, log-likelihood, AIC, and BIC) is optimised. This optimisation is done for the MNL model, the other models are all based on this specification. All models are estimated using the Python Biogeme package (Bierlaire, 2016).

In order to test the predictive power of the best model for the experienced mode choice set, a k-fold cross-validation is performed with five groups. This means that the dataset is randomly distributed into five groups of 20% each. Accordingly, the model is estimated using 80% of the sample and the remaining 20% is used to predict the experienced choice set, given the estimated model. The stability of the parameters is tested, based on model estimations from the different samples. Furthermore, the predictive power is tested by calculating how often the model assigns the highest probability to the actual experienced choice set (hit rate) and the extent to which errors are made. Regarding the latter, if for example the experienced set is bicycle-walk and the predicted set is bicycle only, this is considered a less significant error than predicting train only as the experienced set.

4.4. Data Description

For this study, the data obtained via the Netherlands Mobility Panel (MPN) is used. This is a longitudinal household panel, which commenced in 2013, with the goal of investigating how travel patterns of individuals change over a long period of time. Two surveys focusing on personal and household characteristics and a three-day travel diary are distributed among panel members every autumn. This panel is to a large extent representative of the Dutch population. We refer the reader to Hoogendoorn-Lanser et al. (2015) for a detailed description of the MPN surveys and travel diary.

A companion survey on the perceptions, attitudes, and wayfinding styles towards active modes (coined PAW-AM) was distributed among the MPN panel members in June 2017. This survey addressed among other things the experienced mode choice set of individuals (see Section 3). To identify different determinants that influence the composition of the experienced choice set, we use data from the personal and household surveys. Consequently, the data from the MPN surveys (2016) and the PAW-AM survey (2017) are merged. This study focuses on commuter trips, as such respondents were required to have a job and commute towards their work location. 2,775 respondents fulfilled these requirements. A total of 31 alternatives can be experienced, of which one was never chosen and 18 were rarely chosen (less than 20 times). Therefore, a final filtering was performed to include only experienced choice set alternatives that contain sufficient respondents, which leads to a dataset of 2,652 respondents. A total of 12 experienced mode choice set alternatives are included for model estimation and cross-validation:

1. Car
2. Bicycle
3. Train
4. BTM
5. Walk
6. Bicycle-Train
7. Bicycle-BTM
8. Bicycle-Walk
9. Car-Train
10. Car-Bicycle
11. Car-Bicycle-Walk
12. Car-Bicycle-BTM

Based on the determinants used to specify mode choice sets in literature and the availability of data in the MPN, potential determinants of the experienced mode choice set are selected for this study. Table 4.2 shows an overview of the selected determinants and their operationalisation in the models. Due to the aggregated nature of the experienced mode choice set in this data collection effort, trip-level characteristics are not included. Five categories of determinants are identified, namely socio-demographics, ownership characteristics, work conditions, urban density, and household characteristics. This information is collected in the MPN surveys of the year 2016.

Table 4.2: Potential determinants of the experienced mode choice set

Socio-demographics	**Urban density**
Gender *(m/f)*	Level of urbanisation *(low/medium/high)*
Age *(-49/50+)*	
Education level *(low/medium/high)*	**Household characteristics**
	No. of individuals *(1-2/3+)*
Ownership characteristics	Children *(y/n)*
Driver's license *(y/n)*	Household income *(low/medium/high)*
Car ownership *(y/n)*	
Bicycle ownership *(y/n)*	**Work conditions**
Public transport subscription *(y/n)*	Working hours *(part-time/full-time)*
	Reimbursement for car, public transport, or bicycle *(y/n)*

The characteristics of the respondents in the dataset are presented in Table 4.3. The surveys are only distributed to individuals of 12 years and older. The education level represents the highest completed level of education. Consequently, the younger population that have not finished studies yet, ends up in a lower level of education. The education levels represent the following: low (completion of secondary education), medium (completion of higher secondary education, pre-university education, or secondary vocational education), and high (completion of higher professional education or university education). Many respondents have a medium or high education level, potentially due to the focus on commuting trips. Furthermore, ownership and availability percentages are high. A large share of the respondents lives in a highly urban environment, which represents municipalities of 1,500 inhabitants/km^2 or more. A moderate urban environment is defined as a municipality of 1,000-1,500 inhabitants/km^2 and a low urban environment is defined as less than 1,000 inhabitants/km^2.

Table 4.3: Characteristics of the respondents (N=2,652)

		Percentage (%)			Percentage (%)
Gender	Male	46.8	Children in household	No	76.4
	Female	53.2		Yes	23.6
Age	<=34 years	41.0	Persons in household	1	14.7
	35-49 years	29.3		2	23.2
	50<= years	29.8		3	18.5
				4+	43.7
Education level	Low	19.9			
	Medium	40.8	Reimbursement	Car	35.0
	High	39.3		PT	7.5
				Bicycle	10.5
Ownership	Drivers' license	89.7			
	Car	75.5	Urban density	Urban	52.2
	Bicycle	87.9		Sub-urban	18.6
	PT - subscription	30.5		Rural	29.3

Most respondents have no children (under the age of 12) and live in a four or more-person household. This means that most households have children (over the age of 12) or other inhabitants. Finally, more than half of the respondents are reimbursed by their employer for travelling by a certain mode, where the largest share is reimbursed for the car (e.g. in the form of a company car or kilometre compensation).

4.5. The Experienced Mode Choice Set

This section investigates the experienced mode choice set for commuting trips, using the dataset described in the previous section as derived from the MPN and PAW-AM surveys. In Section 4.5.1 the size and composition of this set are discussed. Section 4.5.2 compares the experienced choice set with a choice set defined based on ownership and availability. Finally, Section 4.5.3 investigates consistency in this experienced set over time.

4.5.1. Size and Composition of the Experienced Mode Choice Set

In the Netherlands, the car and bicycle are the most commonly used modes (CBS, 2018). When analysing the occurrence of different experienced mode choice sets for commuting purposes (Table 4.4), we see that car and bicycle are dominating. Note that access and egress modes are excluded here. The most common sets consist of single-mode alternatives, with the exception of the car-bicycle choice set. Thus, individuals have a relatively small choice set for commuting trips, where most individuals only use one mode for their commute over a period of half a year. This was also found by Kuhnimhof (2009) and Kuhnimhof et al. (2006), however they explored the mode use behaviour over only seven days. Our findings suggest that this unimodality is still largely present over a period of half a year, providing a first indication that individuals are habitual in their mode use for commuting trips.

Table 4.4: Ranking of the experienced mode choice sets

No.	Rank	Alternative	Share (%)	Frequency
1	1	Car	47.1	1248
2	2	Bicycle	26.6	706
3	3	Car-Bicycle	9.8	260
4	4	Train	5.4	142
5	5	Walk	2.2	59
6	5	BTM	2.2	59
7	6	Car-Train	1.8	49
8	7	Bicycle-BTM	1.3	34
9	8	Bicycle-Walk	1.1	29
10	9	Car-Bicycle-Walk	1.0	26
11	10	Bicycle-Train	0.8	20
12	10	Car-Bicycle-BTM	0.8	20

Table 4.5 visualises the modal shares per mode and shows how each mode is part of choice sets of different sizes. When investigating the experienced mode choice sets from this perspective, several observations can be made. First of all, it is confirmed that car and bicycle are the most common modes for commuting trips among individuals in the sample. The other modes; train, BTM, and walking, are used much less for commuting trips. Furthermore, BTM and walking are, relatively, more often part of multimodal choice sets (about 50% of the occurrences) compared to the other modes. Conversely, the car is most often used unimodally. Finally, the majority of the respondents have reported using a single mode for commuting trips in the last half year (83.5%), compared to 14.8% of the respondents that used two modes and 1.7% of the respondents that used three modes. The share of unimodal commuters is higher than the 72% found by Kuhnimhof (2009). One might expect that this percentage decreases when the observation period increases, however our findings show the opposite. Potentially, this is related to the context, i.e. Germany versus the Netherlands. The unimodality that we observe

for commuting trips, does not necessarily mean unimodality in general, as for other trip purposes more or different modes can be used.

Table 4.5: The experienced modes in the past half year depicted against the choice set size

	No. of modes in choice set			
Mode	1	2	3	*Total*
Car	47.1%	11.7%	1.7%	60.4%
Train	5.4%	2.6%	0.0%	8.0%
BTM	2.2%	1.3%	0.8%	4.3%
Bicycle	26.6%	12.9%	1.7%	41.3%
Walk	2.2%	1.1%	1.0%	4.3%

4.5.2. Comparison between Choice Set Definitions

Both in research and practice, the mode choice set is often defined based on deterministic rules (de Jong et al., 2007; Hamre and Buehler, 2014; Gehrke and Clifton, 2014; Kamargianni and Polydoropoulou, 2013). These deterministic show large variations in their rigorousness between studies, which results with different choice sets when applying them on the same data. As an example, for the inclusion of public transport in the choice set, Gehrke and Clifton (2014) state that a bus or train stop should be present within respectively 0.5 and 1.0 mile from the home location and Habib et al. (2011) state that a stop should be available within the neighbourhood. On the other hand, Ton et al. (2019a) use the Google Directions API to identify whether a public transport route is available from home to destination. The nearest stop is not necessarily the best suitable stop for the entire trip, therefore one can argue whether including only the nearest stop will be accurate for a trip. Consequently, these deterministic rules will result with different choice sets, as they are based on different logic and (network) information. In this section we compare the experienced choice set with a choice set defined based on deterministic rules, in this case based on availability and ownership, to identify differences and similarities between rule-based and behaviour-based choice sets.

To define the commuting availability/ownership choice set, we assume that the mode needs to be available to an individual on a daily basis. Therefore, the car and bicycle are included only if the respondent owns the mode. This means that no distinction is made between driver and passenger. Furthermore, public transport is only included if the individual has a subscription (e.g. discount, ticket for a specific line). This seems a plausible assumption for daily commuting trips, because not having this subscription and using it on a daily basis is expensive in the Netherlands (either for the individual or employer). Regarding walking we define no availability assumptions. This will not always hold for commuting trips (see Table 5), because people might not be able to walk or consider is too far to walk. The same applies for the inclusion of bicycle. However, because of the aggregated nature of the data no information on distance or travel time is available for many individuals (47%). Based on the above definition eight different choice set combinations can be identified. Four modes are distinguished in the deterministic choice set: car, public transport, bicycle, and walk. In the experienced choice set public transport is divided into BTM and train, therefore the comparison is not completely one-to-one.

Table 4.6 shows the comparison between the deterministic choice set and experienced choice set, with three exact matches between both sets: walk only, bicycle-walk, and car-bicycle-walk. The total of exact matches (dotted) is 22 out of 2,652 (0.8%). Consequently, when defining a choice set based on availability and ownership many differences are found in excluding and including relevant modes compared to the observed behaviour. The horizontal stripes show the mismatches between the two sets. We found a total of 171 mismatches (6.4%). The largest mismatch in number of respondents is between the car only experienced set and bicycle-walk ownership/availability set (38 out of 2,652). These individuals do not own a car (only a bicycle), but solely use the car as commuting mode. This means that these individuals borrow the car from someone else, are a passenger, or use it via a sharing system. For the majority of the respondents the experienced set is a subset of the ownership/availability set or vice versa (diagonal stripes). For example, in case of the car only experienced choice set and the car-bicycle-walk ownership/availability set (832 out of 2,652), individuals also own a bicycle, but do not use it. Some respondents show a mixture of the ownership/availability and experienced set (white), for example car-bicycle is experienced, whereas car-walk is owned/available. In that case the bicycle was borrowed from someone else or used via a sharing system. Consequently, ownership and availability are not the only explanatory variables for the experienced choice set. As mentioned before, different deterministic rules will result with different choice sets. As these rules are all based on logic and network information, they are likely to mismatch to a certain extent with observed behaviour (experienced choice set).

Table 4.6: Availability/Ownership choice set compared to the experienced choice set. dotted = exact match, diagonal stripes = subset, horizontal stripes = mismatch, white = mixture. w=walk, l= local transit (btm), t=train, b=bicycle, c=car, pt=public transport.

Experienced choice set	Availability/Ownership Choice set								
	w	pt-w	bw	b-pt-w	cw	c-pt-w	cbw	cb-pt-w	Tot.
w	0	0	10	11	7	2	25	4	59
l	0	4	6	31	0	1	7	10	59
t	1	1	2	74	0	11	14	39	142
b	4	5	134	184	27	9	273	70	706
bw	1	0	6	10	2	1	7	2	29
bl	0	1	4	22	0	1	1	5	34
bt	0	0	1	5	1	2	1	10	20
c	17	1	38	37	162	20	832	141	1,248
ct	0	0	0	4	3	5	14	23	49
cb	2	0	17	9	21	3	172	36	260
cbw	0	0	1	2	5	0	16	2	26
cbl	0	0	2	4	0	0	8	6	20
Tot.	25	12	221	393	228	55	1,370	348	2,652

4.5.3. Consistency in the Experienced Mode Choice Set over Time

In the PAW-AM survey, the respondents were asked to recall which modes they have used over the past half year. To investigate consistency over time we compare the experienced choice set from the PAW-AM survey with the choice set containing all reported commuting modes in the three-day travel diary (MPN). In the travel diary, individuals were asked to report all the trips (and modes) made in the course of three days. By filtering the commuting trips from this diary, the experienced choice set based on the three-day travel diary is composed. It is uncertain whether this three-day period captures the whole spectrum of modes used, regardless it will help in identifying (in)consistency over time. The travel diary data was collected in Autumn 2016, whereas the PAW-AM survey covers the first half year of 2017.

Of the 2,652 individuals that filled in the PAW-AM survey, only 1,280 filled in the three-day travel diary and made at least one commuting trip. Approximately two-thirds (67.3%) of these respondents report the same choice set during both periods and are thus considered consistent in their experienced choice set over time. The choice sets that show consistency are the unimodal choice sets and two-mode sets, such as bicycle-walking and car-bicycle. The lack of consistency in the three-mode choice sets, might stem from the fact that only three days were observed. A total of 22.7% of the population reports a subset, either of the experienced set (13.6%) or of the three-day diary set (9.1%). Furthermore, a total of 9.3% of the respondents report a different choice set during the three days compared to the half year. Noticeably, 22.2%

of this group has experienced a life event related to moving jobs or moving homes. One other reason of this shift in behaviour could be the seasonality: autumn versus winter and spring.

Table 4.7: Experienced choice sets of the survey (2017) and the three-day travel diary (2016) compared. dotted = exact match, diagonal stripes = subset, horizontal stripes = mismatch, white = mixture. w=walk, l= local transit (btm), t=train, b=bicycle, c=car.

	2016																					
2017	w	l	wl	t	tw	tl	b	bw	blw	bt	btlw	c	cw	cl	ct	ctw	ctl	cb	cbw	cbl	cbt	Tot.
w	11	0	0	1	0	1	1	0	0	0	0	2	2	0	0	0	0	0	1	0	0	*19*
l	0	14	4	0	0	1	4	0	0	0	0	0	0	4	0	0	0	1	0	0	1	*29*
t	0	1	0	39	4	1	8	0	0	3	0	8	2	0	2	0	1	1	0	0	0	*70*
b	3	1	1	2	0	0	213	10	2	2	0	43	1	1	0	0	0	34	3	3	1	*320*
bw	1	1	0	0	0	0	7	6	0	0	0	0	0	0	0	0	0	0	1	0	0	*16*
bl	0	2	1	0	0	0	11	0	1	1	0	0	0	0	0	0	0	0	0	0	0	*16*
bt	0	0	0	10	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	*13*
c	3	4	1	5	0	2	16	0	0	2	0	553	12	5	4	1	0	13	0	1	0	*622*
ct	0	0	0	8	1	0	0	0	0	0	1	5	1	1	2	0	0	0	0	0	0	*19*
cb	0	0	0	2	0	0	44	0	0	0	0	62	1	0	0	0	0	23	0	1	0	*133*
cbw	0	0	0	0	0	0	5	0	0	0	0	6	1	0	0	0	0	2	0	0	0	*14*
cbl	0	3	0	0	0	0	4	0	0	0	0	2	0	0	0	0	0	0	0	0	0	*9*
Tot.	*18*	*26*	*7*	*67*	*6*	*5*	*314*	*17*	*3*	*8*	*1*	*681*	*20*	*11*	*8*	*1*	*1*	*74*	*5*	*5*	*2*	1,280

When investigating the (in)consistency over time more thoroughly for the 1,280 individuals that are part of both datasets, several observations can be made (see Table 4.7). First, consistency over time occurs most for the unimodal choice sets (96% of the matches). Consequently, the patterns that were uncovered in the half year survey were already present in the year before, confirming the habitual behaviour. The individuals that have a wider spectrum of modes in their experienced set are often partially consistent over time (e.g. expanding the set or shrinking the set). Second, several respondents have (partially) shifted from motorised modes to active modes, which might (again) be due to change in season. For example, 43 individuals shifted from car to bicycle, 34 shifted from car-bicycle to bicycle only, 62 shifted from car to car-bicycle, and 2 shifted from car-walk to walk only. Finally, when a subset of the travel diary set is reported in the survey, the data often suggests habit formation. For example, in case of bicycle-walk, 10 out of 17 individuals shift to bicycle only, for the car-bicycle-walk alternative all individuals become unimodal users, and the car-bicycle alternative shows that the majority of the individuals become either unimodal car or bicycle users.

4.6. Modelling Results

The determinants that are relevant for the experienced choice set are uncovered using discrete choice models. This section details the results of this exercise. Section 4.6.1 describes the overall results of the estimated models. Section 4.6.2 discusses the determinants that are relevant for the experienced mode choice set. Finally, Section 4.6.3 reflects on the suitability of choice sets based on historical data for prediction purposes.

4.6.1. Overall Model Estimation Results

In this study four different model structures are tested: MNL, NL, CNL, and ML. The NL, CNL, and ML models did not produce significantly superior results compared to the MNL model. All the nesting parameters that are estimated for each of the NL models (based on both size and composition) are not significantly different from one, consequently suggesting that the alternatives do not share unobserved variables. The CNL model (based on mode-nests) with variable membership did not converge properly, consequently an iterative process was used to find the best membership for each alternative to the nests. This iterative process consisted of alternately estimating the nesting parameters and attributes, while fixing the alphas, and estimating the alphas, while using the results of the previous iteration as fixed input. This model significantly improved the log-likelihood compared to the MNL, however no significant nesting parameters were found. The ML model reproduces the CNL model with fixed membership, thus performing worse than the CNL model with flexible membership. Consequently, we have to conclude that the MNL model produces the best results. This seems plausible, because we include a relationship in the utility function between alternatives that contain the same modes via the estimation of mode-specific parameters (Eq. 4.2). Consequently, the car-bicycle and unimodal bicycle alternatives contain partially the same parameters. The other model structures did not find a significant effect of shared unobserved variables between alternatives. Table 4.8 shows the overall model fit results. The MNL model is estimated on a random draw of 80% of the data. It has a model fit of 0.542.

Table 4.8: Model fit results

	MNL
Initial log-likelihood (equally likely)	-5307.761
Final log-likelihood	-2383.285
# parameters	45
Likelihood ratio test	-
AIC	4856.6
BIC	5111.5
Adjusted rho-square (compared to the initial model)	0.543
# observations	2,136

4.6.2. Determinants of the Experienced Mode Choice Set

In this section we discuss the different determinants that are relevant for explaining the experienced mode choice set according to the MNL model (see Table 4.9). The utility function consists of alternative specific constants and mode specific parameters. Regarding the first, we have fixed the parameter for car to zero, so that a comparison based on the relative utility can be made. Regarding the second, the model specification implies that if the alternative is car-bicycle, the parameter values for car and bicycle need to be summed up (linear in parameters) to find the combined parameter. We discuss the parameter coefficients per category of variables.

Constants

The constants are alternative-specific and provide information on the average influence of the unobserved variables on the utility (relative to the reference alternative: car-only). The car-only alternative is most frequently chosen, which explains why the parameters values related to most other mode choice alternatives are negative. The constants for walk and cycle are positive, suggesting that these have unobserved variables that favour walking and cycling over the car. This might be due to the shorter distances for which these modes are used, which is not captured in the model (unobserved). Most parameters are significant, which indicates that the mode-specific parameters do not capture all the information in the data. Thus, other influences are present in the experienced mode choice set choice that are currently not observed.

Table 4.9: Determinants of the experienced mode choice set for the MNL model

Determinant	Level	Mode	Coefficient	Robust t-test
Alternative-specific parameters				
Constant	Car		0	-
	Bicycle		2.60	7.65*
	Bicycle-BTM		-0.81	-1.71
	Bicycle-Train		-2.48	-5.33*
	Bicycle-Walk		-0.16	-0.33
	BTM		-0.23	-0.52
	Car-Bicycle		-1.21	-6.10*
	Car-Bicycle-BTM		-3.62	-8.05*
	Car-Bicycle-Walk		-2.96	-7.15*
	Car-Train		-4.69	-13.64*
	Train		-0.62	-1.52
	Walk		0.79	1.84
Mode-specific parameters				
Education level	Medium	Car	0.42	2.50*
(ref=Low)	High	Bicycle	0.28	2.12*
		Car	0.49	2.63*
		Train	0.59	2.70*
Age (ref=<50 years)	50=< years	BTM	-0.64	-2.09*
		Car	-0.24	-1.71
		Train	-0.87	-3.27*
Gender (ref=Female)	Male	Car	0.42	3.31*
Household size (ref=3+ pers.)	1-2 pers.	Walk	0.51	2.08*
Children (ref=No)	Yes	Car	0.37	2.20*
		Walk	0.76	2.39*
Urban density	Low	Bicycle	-0.44	-3.57*
(ref=High)		BTM	-0.75	-2.32*
		Train	-0.88	-3.18*
	Moderate	BTM	-0.70	-1.84
Driver's License		Car	1.71	5.59*
Ownership	Bicycle	Bicycle	0.31	1.83
		Walk	-0.77	-2.48*
	Car	BTM	-0.61	-2.09*
		Car	1.10	6.72*
		Train	0.46	1.95
	PT sub.	BTM	0.97	3.35*
		Car	-0.54	-3.74*
		Train	1.63	6.49*
Working hours	Fulltime	Bicycle	-0.58	-5.12*
Reimbursement	Bicycle	Bicycle	2.04	7.39*
		BTM	0.67	1.79
		Car	-0.78	-3.77*
	Car	Bicycle	-0.89	-6.07*
		Car	1.67	7.96*
		Walk	-1.55	-2.99*
	PT	Bicycle	-0.48	-1.75
		BTM	2.93	8.44*
		Train	3.08	9.95*

*=significant at the 5% level.

Socio-demographics

Three socio-demographic variables are tested: age, gender, and education level. All are found to be significant. This is in line with the research of Vij et al. (2017, 2013) and Calastri et al. (2017), who used among others, these three variables to identify the availability of several modes. The older commuting population (50=< years) is associated with a lower utility for the BTM, car, and train modes compared to the young population (<50 years). The active modes are thus more attractive to the 50+ population than the motorised modes, all else being equal.

Men are more likely to have car included in their experienced mode choice set than women, which is also found by Vij et al. (2013). Furthermore, the education level parameters show that if an individual has a medium and high education level, they are more likely to include the car mode in the experienced choice set compared to an individual with a low education level. The bicycle and train also become more attractive for an individual with a high education level.

Household Characteristics

The household characteristics exhibit a limited association with the experienced mode choice set. The household income does not yield a significant relationship. The salary of the combined household members does not make it more or less likely that a mode is included in the experienced choice set, contrary to Witlox & Tindemans (2004), who found a positive relationship between income and car use. The size and composition of the household do explain the inclusion of walking and using the car in the experienced mode choice set. The presence of children under the age of 12 in the household increases the probability of including walking and the car in the experienced choice set. This is in line with the results found by Vij et al. (2017), as they found that having children increases likelihood that one is dependent on a car. Furthermore, an individual living in a household consisting of one or two persons is more likely to include walking in the experienced choice set compared to an individual living in a larger household.

Urban Density

Low urban density represents municipalities that contain mostly villages and rural areas, moderate density represents municipalities with medium-sized cities, such as Groningen, and high density represents municipalities with large cities, such as Amsterdam and Rotterdam. We took the high urban density as a reference point. With moderate density it is less likely that BTM is included in the choice set. Generally, the density and frequency of BTM is very high in high urban areas, but less so in moderately urban areas. In low urban areas, the bicycle, train, and BTM are less likely to be included in the experienced mode choice set. This finding shows that the experienced choice set is not only related to individual specific determinants, but also to where they live.

Ownership

The ownership variables are all significantly associated with the experienced mode choice set. This concurs with the use of ownership variables in many mode choice studies to identify the mode choice set (Ben-Akiva and Boccara, 1995; Gehrke and Clifton, 2014; Habib et al., 2011; Kamargianni, 2015; Kamargianni et al., 2015; Rodríguez and Joo, 2004; Vij et al., 2017, 2013). Having a driver's license is positively associated with the inclusion of the car in the experienced choice set. It does not have a significant impact on the inclusion of other modes. In contrast, ownership of a certain mode positively relates to inclusion of that mode and at the same time negatively relates to inclusion of other modes in the experienced mode choice set. Owning a bicycle positively relates to including the bicycle in the choice set, which is in line with literature (Heinen et al., 2010; Muñoz et al., 2016a). Moreover, it also reduces the utility of walking for inclusion in the choice set. Owning a car increases the utility of inclusion of car in the choice set, furthermore it increases the utility for including train in the choice set. This suggests that train users often own a car. On the other hand, it reduces the utility of the BTM. Ownership of a public transport subscription results with the expected results: having a subscription increases the probability of including BTM and train in the choice set and reduces the probability of using the car. This suggests that train users are affected differently by car ownership, compared to car users and the ownership of a public transport subscription. The first yields a positive relation, whereas the second yields a negative relation.

Work Conditions

For commuting trips, the work conditions are important determinants. Both the working hours and reimbursement are relevant for the experienced mode choice set, where the latter proves to be more important. Working full-time (more than 35 hours per week) decreases the probability of including the bicycle in the choice set, which is in line with literature (Heinen et al., 2010). This does not hold for the other modes of transport. Regarding the reimbursement, similarly to the trend observed for the ownership of modes: being reimbursed for using a mode to commute to work increases the probability of including that mode in the experienced choice set, whereas it decreases the probability of including another mode in the experienced choice set. One exception is the reimbursement for bicycle, which positively relates to inclusion of BTM in the choice set. Both the bicycle and BTM can be used for similar distance ranges and can thus be used as substitutes. Reimbursement of the car decreases the inclusion of walk and bicycle in the choice set. Finally, a public transport reimbursement increases the utility for BTM and train and reduces the utility of the bicycle.

4.6.3. Using the Experienced Mode Choice Set for Out-of-sample Predictions

To test the performance of the model for predictions, we have partitioned the data into five segments of approximately 20% each. One of the segments is used as the default case to calibrate the parameters involved. The other segments are used for validation. In this validation we investigate several aspects: model performance, stability of the parameters, and prediction accuracy for out-of-sample data.

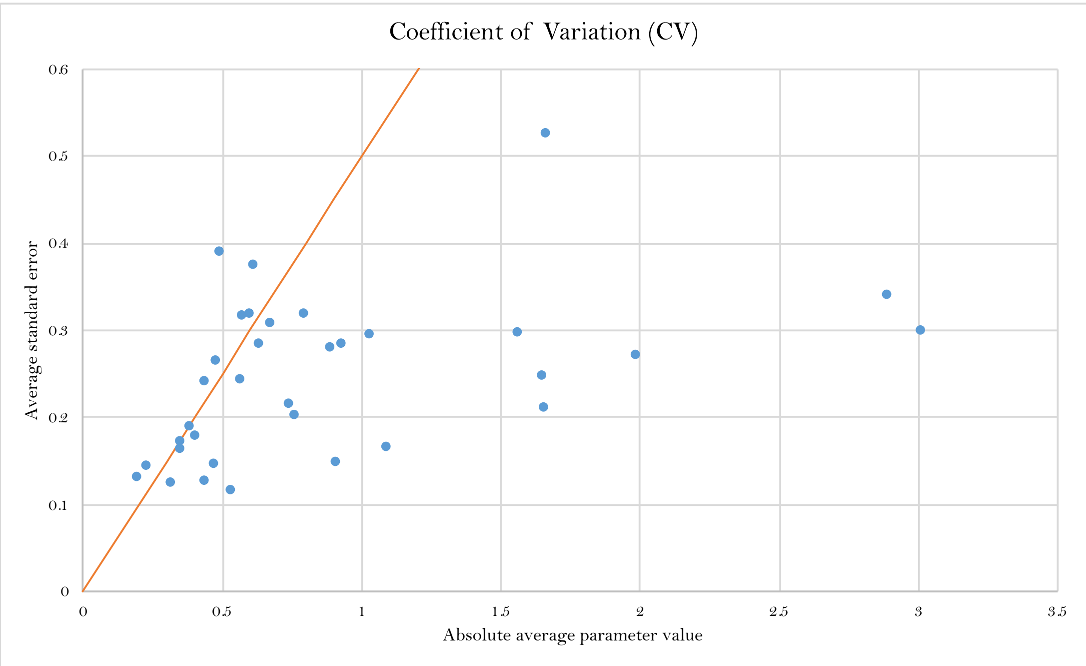


Figure 4.3: Coefficient of variation (mean standard error divided by mean absolute parameter value)

The models perform well for the different segments. The default model produced an adjusted rho-square of 0.542, whereas the validation models show a model fit ranging from 0.526 to 0.541. This means that for all segments between 52% and 54% of the variation of the data can be explained using this model. The AIC and BIC values also show promising results, with some segments showing better results than the default model.

The stability of the parameters is investigated by calculating the coefficient of variation (CV) for each parameter. This is done by dividing the average standard error of the parameter by the average parameter value. If the CV is over 0.5, the parameter is considered less stable. Figure 3 visualises the CV for all parameters. Left of the orange line the CV is over 0.5 and thus less stable. We exclude the constants in this comparison, because these capture the unobserved variables for each model and are therefore different per model. Eight parameters are considered less stable: high education – bicycle, medium education – car, 50+ years – BTM and car, own car – train, low urban – BTM, moderate urban – BTM, and reimbursement bicycle – BTM. These parameters show large variation in their parameter values and standard errors, where the last parameter is the least stable. The majority of the parameters, however, are stable over the different runs.

For investigating the prediction accuracy, we apply the model on out-of-sample data. This results with a probability for every alternative to be chosen. When summing these probabilities over individuals for each alternative, we can calculate what the probability is that the prediction is correct or not. In some cases, the model predicts a choice set that is too small, too large, or mixed. In this case, the prediction is considered better than when the wrong mode(s) are predicted, as it is able to identify part of the choice set. An example is if the prediction is bicycle and the actual choice set was bicycle-car. If the predicted choice set is a total mismatch, we consider it wrong. For example, the predicted set is BTM, whereas the actual choice set is bicycle-walk.

Table 4.10: Predictive power of the experienced mode choice set model on out-of-sample data

	Correct	Too small	Too large	Mixed	Wrong
Run 1	48%	12%	12%	1%	27%
Run 2	50%	11%	12%	1%	27%
Run 3	50%	12%	12%	1%	26%
Run 4	49%	13%	12%	1%	25%
Run 5	49%	12%	11%	1%	27%
Average	*49%*	*12%*	*12%*	*1%*	*26%*

Table 4.10 shows the prediction accuracy based on the above-mentioned situations for each of the model runs. The models predict the correct experienced choice set on average 49% of the time. Three alternatives are especially well predicted: car, bicycle, and train. These alternatives are observed most frequently in the dataset, consequently the model aims to predict these alternatives correctly. In 25% of the cases a choice set is predicted that is smaller, larger, or mixed compared to the observed choice set and in 26% the model produces a wrong choice set. All in all, around 74% of the observations the experienced mode choice set is predicted with sufficient accuracy. This suggests that the relevant determinants are, to a large extent, able to capture the experienced choice set of individuals for commuting purposes. Consequently, these determinants can also be used for the specification of mode choice sets in future mode choice studies.

4.7. Discussion on the Experienced Mode Choice Set

This section discusses our findings of the experienced mode choice set in relation to past research. In particular, we discuss the determinants of the choice set, the unimodal choice sets, and potential applications of the experienced choice set.

In the literature, we identified that most studies specify the mode choice set based on ownership/availability and trip characteristics (e.g. Ben-Akiva and Boccara, 1995; Gehrke and

Clifton, 2014; Hensher and Ho, 2016; Vij et al., 2013). Some studies have investigated the influence of individual characteristics on the availability/inclusion of modes in the choice set (Calastri et al., 2017; Vij et al., 2017, 2013). In this study, we tested both ownership and individual characteristics and found that both influenced the experienced choice set. Our comparison between the ownership/availability choice set and the experienced choice set shows that these sets are very different. A specification based on ownership would falsely exclude alternatives such as car and bicycle, which can also be used as shared modes. This discussion also showed that different deterministic rules result with different choice sets, and that these somewhat 'arbitrary' rules are different from observed behaviour. Furthermore, we found that the urban environment is also relevant for the experienced mode choice set. This means that regardless of the characteristics of an individual, certain modes have a higher or lower probability to be used depending on the type of environment in which one lives. This finding is confirmed by research on mode use and mobility patterns (Kuhnimhof et al., 2006; Ton et al., 2019b). Furthermore, the work conditions prove to be important explanatory variables of the experienced mode choice set for commuting trips. This means that the employer plays an important role in the mode use of its employees. Consequently, more individuals may start cycling when the system of reimbursement for commuting given by the employer is changed to benefit cyclists (Heinen et al., 2013). In sum, a wider variety of characteristics is relevant for the identification of the mode choice set compared to what has been previously assumed.

The majority of individuals has a unimodal experienced choice set for commuting trips. Many of these individuals show habitual behaviour that is consistent over time. This is in line with findings from Lavery et al. (2013) and Kuhnimhof (2009). However, this study offers evidence to suggest that there might be ways to influence travellers in ways that will lead to an increase of the mode set or even a modal shift. The employer can, for example with support from the government, choose to provide reimbursement for sustainable or active modes instead of car use. This increases the probability of including the bicycle or public transport in the experienced choice set. Unimodality is higher in low density urban areas (86.8%) than in moderate (83.8%) or high density urban areas (81.5%). Stimulation of the bicycle use may increase the use of the bicycle in high density urban areas, potentially increasing the experienced choice set size. Consequently, this study shows that several determinants can be used to develop policy measures that might increase the number of modes used by commuters.

A potential application of the experienced choice set lies in the mode choice domain. When embedding this method in the probabilistic approach proposed by Manski (1977), a probability is assigned to each experienced choice set. In a simultaneous model the mode choice is then estimated such that it incorporates the probabilities for the experienced choice sets as shown in equation 4.3.

$$P(i|B,D,X_n) = \sum_{C \in G_n} P(i|C,B,X_n) \cdot P(C|G_n,D,X_n) \tag{4.3}$$

where $P(i|B,D,X_n)$ is the probability that an individual n chooses alternative i, given the parameters B and D and explanatory variables X_n. This depends on the probability of a choice set being chosen by the individual and the choice from this choice set. The set G_n includes all non-empty deterministically feasible modes for individual n. G_n is a subset of the master choice set M that comprises all possible alternatives available for the choice context and population ($G_n \subseteq M$). Eq. 3 thus consists of three parts (Swait and Ben-Akiva, 1987b); 1) a mode choice aspect given a choice set, 2) a deterministic choice set generation aspect to define M_n, and 3) a probabilistic choice set generation aspect that expresses the probability that choice set C is the actual consideration choice set. Among others Ben-Akiva and Boccara (1995) and Cantillo and Ortúzar (2005) have applied variants of this method. This application has to be thoroughly tested with respect to potential endogeneity and bias, as the experienced choice set is based on mode use. Furthermore, this method needs to be benchmarked against commonly used mode choice set specification methods (such as those based on deterministic rules), to adequately identify the (potential) added value of this method.

4.8. Conclusions and Recommendations

This paper presents the findings of an analysis of the experienced mode choice set and identifies the determinants that impact this set. The experienced choice set is the set of modes that is used over a long period of time. We focus on commuting trips in the Netherlands, using data from the Mobility Panel Netherlands (MPN) and a companion survey, in which we ask respondents for their used modes for commuting trips in the past half year. We evaluate the experienced choice set by analysing the size and composition, comparing this set with a choice set based on ownership and availability, and by investigating its consistency over time. The determinants are identified by means of estimation of discrete choice models.

The analysis of the experienced mode choice set for commuting purposes shows that the size of this set is limited. The majority of the respondents only uses one mode in the course of half a year, which indicates habitual behaviour. We investigated which determinants are relevant for the formation of the experienced choice set. Determinants belonging to socio-demographics, household characteristics, urban density, work conditions, and ownership of modes are relevant for choice set formation. The work conditions, especially the reimbursement by the employer for using a specific mode, is particularly important for the experienced mode choice set of commuters. The second group in terms of importance is ownership or availability of modes. We show that more determinants are relevant in the choice set formation than previously assumed. While many studies specify the mode choice set based on ownership and trip characteristics, only some have extended this to including individual characteristics to identify availability of modes. The results are, to a large extent, transferable to out-of-sample data. According to our findings, future research into mode choice could benefit from including a wider variety of variables in the choice set specification.

New modes can potentially be added to the individual's experienced choice set given the right incentives. Policy measures could focus on the reimbursement provided by the employer as this can be used to increase the probability of including the bicycle in the choice set, if this mode is reimbursed and others are not.

Future research may aim to estimate a probabilistic integrated choice set and mode choice model using the experienced choice set. The method needs to be benchmarked against often-used methods to identify its added value in performance and computational effort. This study showed that habitual behaviour is present for the majority of the individuals regarding their commuting trips. When data is available on the experienced choice set over multiple years, the impact of habit formation and life cycle can be investigated in relation to the experienced choice set. Furthermore, it is interesting to investigate if habit formation also arises for other trip purposes or in general in the mobility pattern of the individuals. Next to that, given the increasing evidence that the social environment (i.e. household members and friends) influences mode choice behaviour of individuals (e.g. Pike and Lubell, 2018), additional research could investigate the impact of the social environment on the experienced mode choice set. Furthermore, a potentially interesting addition to the experienced choice set is the way in which the modes are used, i.e. as private mode or via shared systems (and types thereof).

Acknowledgements

This research was supported by the Allegro project (no. 669792), which is financed by the European Research Council and Amsterdam Institute for Advanced Metropolitan Solutions. The data was made available by the Netherlands Mobility Panel administered by KiM Netherlands Institute for Transport Policy Analysis.

Chapter 5 – Determinants of Bicycle Route Choice in Amsterdam

This chapter is based on the following article:

Ton D., Cats O., Duives D.C., and Hoogendoorn S.P. (2017). How do people cycle in Amsterdam, the Netherlands? Estimating cyclists' route choice determinants using GPS data from an urban area. *Transportation Research Record: Journal of the Transportation Research Board*. 2662:75-82. https://doi.org/ 10.3141/2662-09

Abstract

Nowadays, the bicycle is seen as a sustainable and healthy substitute for the car in urban environments. The Netherlands is the leading country in terms of bicycle use, especially in urban environments. Yet route choice models featuring inner-city travel that include cyclists are lacking. This paper estimates a cyclists' route choice model for the inner-city of Amsterdam, based on 3,045 trips collected with GPS data. The main contribution of this paper is the construction of the choice set using an empirical approach which uses only the observed trips in the dataset to compose the choice alternatives. The findings suggest that cyclists are insensitive to separate cycle paths in Amsterdam, which is a city characterised by a dense cycle path network in which cycling is the most prominent mode of travel. In addition, cyclists are found to minimize travel distance and the number of intersections per kilometre. The impact of distance on route choice increases in the morning peak where schedule constraints are more

prevalent. Furthermore, overlapping routes are more likely to be chosen by cyclists given everything else being the same.

5.1. Introduction

Governments worldwide nowadays acknowledge the advantages of cycling as mode of transport. First, there are health benefits for individual cyclists. Second, the bicycle can help reduce emissions when substituting the car (Pan-EuropeanProgramme, 2014). Cycling is most attractive in urban areas without large changes in altitude (e.g. the Netherlands or Denmark), where distances covered are relatively small and car usage is often discouraged and associated with greater travel impedance. Furthermore, most European governments have set goals of increasing the modal share of cycling over the next years (Kuester, 2015).

The Netherlands is the leading country in terms of bicycle use, with 27% of all trips performed by bicycle (Pucher and Buehler, 2008). When focusing on the urban environment, the modal share for bicycles increases further, for example in Amsterdam this was 37% in 2011 (OViN, 2011). Other cities such as Groningen, Delft and Leiden have a comparable share of bicycle trips (Harms et al., 2016; Pucher and Buehler, 2008). Despite the fact that so many people cycle in the Netherlands, models aiming at understanding and predicting cyclists' choice behaviour are lacking (Verkeersnet, 2015).

This shift towards cycling, combined with a lack of models incorporating cycling, calls for the development of models to assess related policy implications. Many cities use forecasting models to estimate if, when and where changes to infrastructure or policy are needed. However, these models are still mainly focused on motorised traffic (Hood et al., 2011). The cycling component is either missing, walking and cycling are combined or the model assumes that cars and cyclists behave similarly. Ideally, in forecasting models mode specific activity and route choices are incorporated. Since both choice processes are currently underdetermined, this study starts out by estimating the route choice determinants for cyclists. Before choosing a route, the traveller has already decided to cycle and which activity to perform, therefore the implications of researching route choice first are expected to be minimal.

Recently, a number of studies have estimated bicycle route choice models for locations where bicycle modal shares range between 1% and 6% (Pucher and Buehler, 2008). Arguably, the determinants of route choice behaviour and their impact might be different from a city such as Amsterdam, where cycling is prominent. These studies used revealed preference (RP) data, more specifically GPS data for estimating the route choice model (Broach et al., 2012; Casello and Usyukov, 2014; Hood et al., 2011; Menghini et al., 2010). Before, most of the data used for model estimation came from stated preference (SP) surveys where the respondents were asked what they would do in a hypothetical situation or route recall surveys where researchers relied on the respondents' ability to recollect chosen routes (e.g. Howard and Burns, 2001; Hunt and Abraham, 2007).

This study aims at estimating cyclists' route choice determinants in a context where cycling is the primary mode of transport. Furthermore, the inner-city of Amsterdam is characterised by a densely built area with well-developed cycling infrastructure. This paper presents the findings from a cyclists' route choice model estimated for the inner-city of Amsterdam, using GPS data to identify the determinants influencing route choice in a network dominantly used by cyclists.

This study contributes to the previous cyclists' RP route choice models by introducing a new approach for choice set identification. Previous RP studies have used choice set generation algorithms to identify the feasible choice set from which the cyclist chooses a route. This approach does not guarantee that the chosen route is generated and may include a large number of alternatives that are not selected by any cyclist. Conversely, an empirical approach is proposed which uses only the observed routes to identify the considered choice set. This implies that the chosen route is by definition included in the choice set. Because all routes in the choice set are chosen at least once, it is likely that the alternative routes are considered by the cyclists in the sample. Furthermore, a behavioural comparison can be made with environments where cyclists make a small minority, because the data is collected in an environment dominantly used by cyclists.

In this paper, Section 5.2 details the data processing phase, going from GPS data to route alternatives and characteristics. In Section 5.3, the processed data is analysed and the results of the estimated route choice models are reported and discussed. Finally, Section 5.4 provides the conclusions of the paper.

5.2. Determining Route Alternatives and Characteristics

This section describes the collection (5.2.1) and map matching (5.2.2) of GPS trajectory data. Furthermore, an empirical approach for identifying the route choice set is proposed (5.2.3), which requires clustering (5.2.4) and filtering (5.2.5) of the data. Finally, the potential determinants for cyclists' route choice are discussed (5.2.6).

5.2.1. Collection of GPS Data

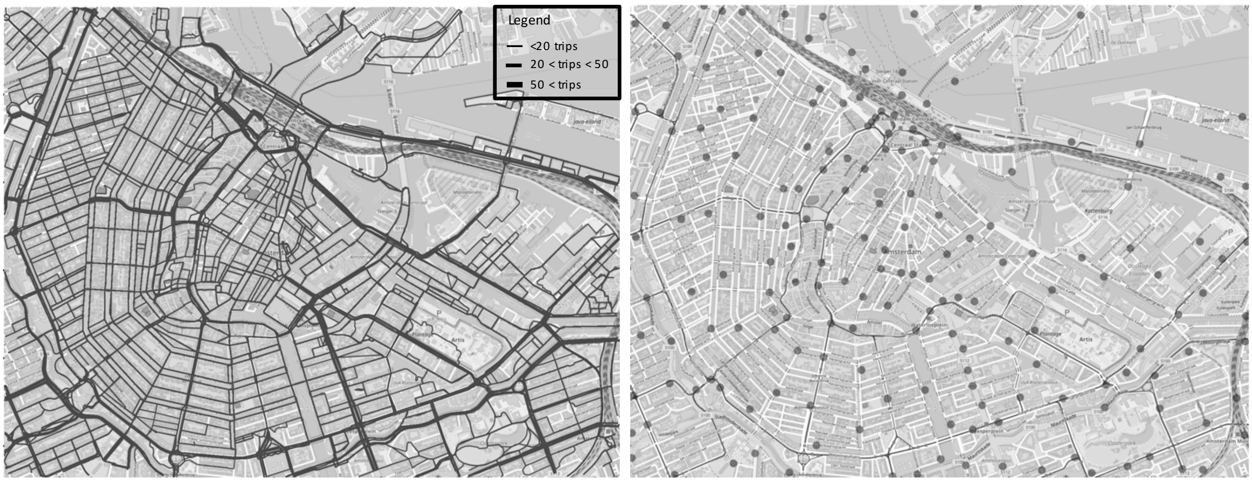
GPS data was collected during a nationwide initiative called the 'Bicycle Counting Week' (BCW), which took place on 14-20 September 2015. The event was organised as a joint initiative of national agencies and companies with the goal of gaining a better insight into the cycling behaviour of Dutch cyclists. Nationwide, a total of 38,000 cyclists participated in this initiative (opt-in sample). Participants' cycling patterns were tracked using an App. In addition, they once filled in a socio-demographic and travel habit survey to complement the GPS data. Several bicycles were put up for raffle under the participants (FietsTelweek, 2015).

During the initiative, data of 377,321 cycling trips was collected nationwide. The respondents' sample includes equal shares of male and female participants. The majority of the participants are in the age group 31-65 (80%), while young people (18-) and old people (65+) are underrepresented. This probably stems from the need for using a smartphone to work with the App and the requirement to have consent from ones' parents if younger than 18 years. Most trips registered are work related (69%), explaining why the group of participants aged 31-65 is overrepresented. Participants could mention multiple reasons for cycling. The most dominant reasons mentioned are health (80%), speed (47%) and comfort (46%) (FietsTelweek, 2015).

As mentioned before, this research focusses on the cycling trips within the city of Amsterdam, where a total of 12,413 trips performed by approximately 5,000 participants were recorded. The Amsterdam sample is similar in terms of gender and age composition to the national sample. However, the share of commuting is higher in Amsterdam (77%). The majority of the cyclists' cycles between 25 and 100km a week (72%), while only 3% cycles less than 10km a week, suggesting that most participants cycle at least to and from their work on a daily basis (FietsTelweek, 2015). All the cycling trips included in this research are superimposed on the map depicted in Figure 5.1a.

5.2.2. Map Matching the GPS Trajectory Data

The map matching is executed by the organizers of the BCW, for a more detailed description on this procedure the reader is referred to Van de Coevering et al. (2014). In the GPS trajectory data, most consecutive GPS data points are measured with an accuracy of 3-4 meters with respect to the infrastructure network. However, outliers up to 50 meters are observed, mainly in dense urban areas. To reduce the impact of these outliers on the analysis, the speed between each two consecutive GPS points is calculated and compared to the actual GPS speed determined by means of Doppler techniques. If the discrepancy between the actual speed and the computed speed is too large, the GPS records are removed from the dataset (van de Coevering et al., 2014).



(a) (b)

Figure 5.1: (a) The network of Amsterdam used for cycling trips. In the centre lies the historical city, surrounded by the ring canal streets and the radial roads heading to and from the city. To the North lies the river IJ, with two ferries connecting its shores. (b) All origin and destination cluster centres resulting from the K-means clustering algorithm.

The remaining GPS trajectories are matched to the OpenStreetMap network. The map matching algorithm deployed by the BCW organizers generates all possible routes from origin to destination and selects the best match for the GPS records. If no match is found, it could be that links are missing (for example in case of desire lines). In that case the route is partitioned and the same procedure is repeated for the sub-routes (van de Coevering et al., 2014).

5.2.3. Identifying the Considered Route Choice Set

In literature, several approaches for choice set identification have been used, with most studies focused on cycling applying a choice set generation algorithm (e.g. Hood et al., 2011; Menghini et al., 2010). The aim of these algorithms is to obtain feasible choice sets (Hoogendoorn-Lanser, 2005), consisting of attractive alternatives. These algorithms however do not guarantee that the chosen route is generated and may include a large number of alternatives that are not chosen by any individual.

An alternative approach for constructing the route choice set is to compile it based on the trips and routes observed in the data. This empirical approach assures that the chosen route is per definition part of the choice set. While the choice set from which each individual cyclist eventually chooses his route (considered choice set (Hoogendoorn-Lanser, 2005)) cannot be observed directly, it is assumed that the observed alternative routes in the data for a given OD pair are included in this set. Unlike the algorithm approach, the empirical approach implies that not all feasible routes are included in the choice set, but rather only routes that are all actually used by the cyclists in the collected dataset. Consequently, the choice set depends on the observed choices and might thus vary for different samples.

A discrete choice model estimated using the realised routes (empirical approach) is expected to have lower explanatory power than a model estimated based on possible routes (algorithm approach). The first approach identifies alternatives that are chosen by at least one cyclist in the data, whereas the second approach also identifies alternatives that are not chosen. As a result, the offset between the chosen route and the alternatives is smaller when estimating a model using only the realised routes.

Two prerequisites exist for applying the empirical approach: each OD pair considered in the analysis should contain multiple trips and at least two distinct routes. For this study a maximum of 19 realised routes for one OD pair is identified.

5.2.4. Clustering of the Origin and Destination GPS Data

Since trip origin and destinations are not likely to be recorded at the exact same geographical location when using high-resolution GPS data (approximately 50% in the BCW database), the GPS origin and destination data points are clustered into larger OD pairs, resulting in more trips and possibly more routes per OD pair.

The k-means clustering method is applied, based on the distance between GPS locations of the origins and destinations (Harrington, 2012). The algorithm minimizes the intra-cluster distances and maximizes the inter-cluster distances. Two downsides of this method are that the solution can get stuck in a local minimum (Harrington, 2012), which results in a suboptimal distribution of GPS locations over the clusters. Furthermore, in case the number of clusters is set too low, the routes in one OD pair cannot be compared, because the origin or destination points are too far apart. The first downside can be (partially) mitigated by setting multiple starting points for the algorithm. This way it is less likely to converge into a local minimum.

This method was applied for different k-values; 150, 200, 250 and 300 clusters. If the number of clusters is set too high, the number of trips per cluster becomes too low and the advantages of clustering the trips diminish. As mentioned before, if the number of clusters is set too low, routes in one OD pair cannot be compared. We find that defining 200 clusters provides the best balance between intra-cluster distance and number of trips per OD pair for the BCW dataset. The number of random starting points is set to 20. Figure 5.1b shows the geographical distribution of the cluster-centres over the inner-city of Amsterdam.

The 200 clusters result in a maximum intra-cluster distance (i.e. diameter) of 444 meters, while the average is 168 meters. The cluster with the largest diameter is located around a park, however the routes chosen are still comparable. Therefore, this is an acceptable diameter for a rather dense network. After clustering, only 30% of the OD pairs consist of one trip, instead of 50% before clustering.

5.2.5. Data Filtering Process

Not all trips in the dataset can be used, mostly because of how the choice set is composed. Therefore, several filtering steps are necessary (see Figure 5.2). In the BCW dataset many cycling trips are made in the inner-city, whereas the density of cycling trips in the suburbs is very low. Therefore, only the trips (partially) traversing the inner-city are used, which limits the available trips to 7,984. Not all trips are included completely, because the boundaries of the inner-city are specified on GPS coordinate level and not on trip level. It is, for example, possible that one trip crosses the inner-city more than once. In this case the trip is split into multiple trips. This demarcation means that some cyclists are observed during the entire trip, whereas others are only observed during part of the trip. We assume that the route choice for a section of the route is not fundamentally different from choosing the complete route.

Due to splitting trips some very short routes are created, for which it is unlikely that route choice is possible. Therefore, a filter is applied on the possibility for route choice, which is defined here as crossing at least two intersections during the trip, resulting in 8,847 trips. When applying the empirical approach to identify the choice set, it is necessary to filter out all OD pairs with only one trip, resulting in 6,208 trips. Also, more than one route needs to be chosen per OD pair. The result is a final dataset of 3,045 trips (see Figure 5.1a). Since other GPS based route choice models have been estimated using less trips (Broach et al., 2012; Casello and Usyukov, 2014; Hood et al., 2011; Menghini et al., 2010), the filtered data set seems large enough to estimate a route choice model for cyclists in inner-city areas. Furthermore, the initial dataset and the final dataset show similar patterns with respect to time of departure and day of travel. The distances covered are slightly larger in the initial sample, due to the geographical demarcation of the inner-city. However, no structural behavioural issues are expected due to the filtering process.

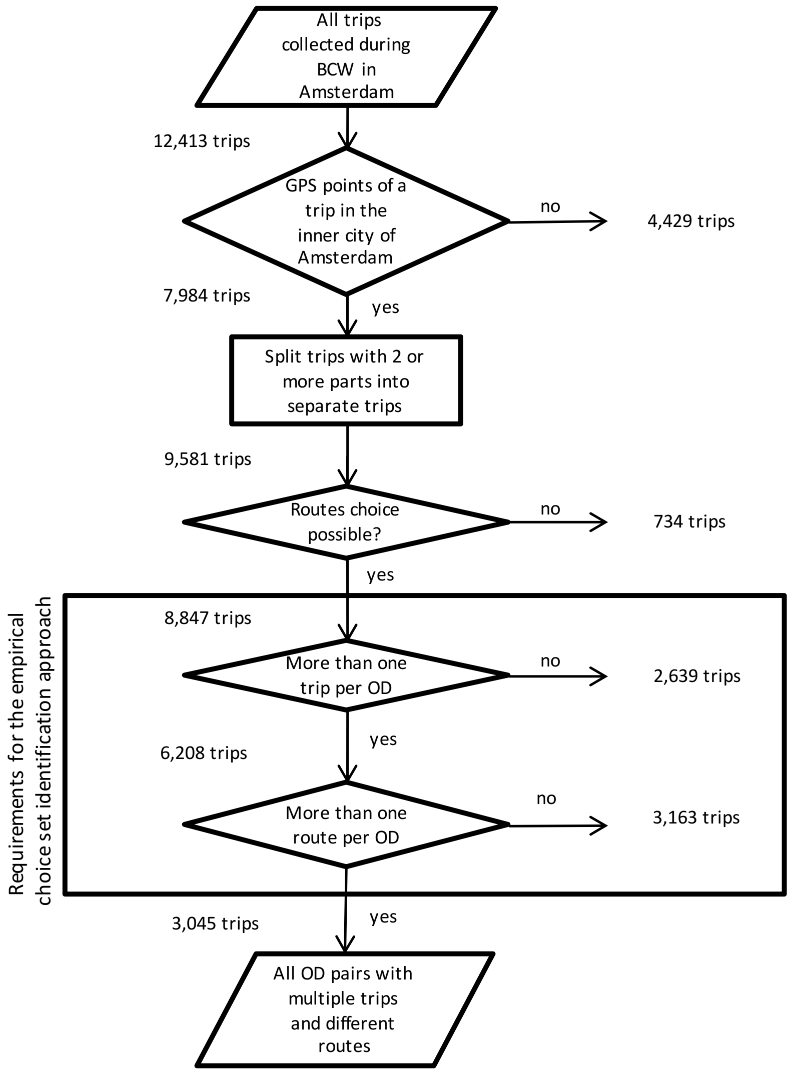


Figure 5.2: Data filtering process

5.2.6. Potential Determinants of Cyclists' Route Choice

Previous research has identified a wide range of attributes that might influence the route choice behaviour of cyclists, where the attributes selected for research mainly depend on the type of data used (RP or SP) and the availability of the data (in case of RP). Both Hunt and Abraham (2007) and Sener et al. (2009) have reviewed many (mostly SP) studies to find attributes that potentially influence bicycle route choice. Based on these reviews (Hunt and Abraham, 2007; Sener et al., 2009) and previous RP studies (Broach et al., 2012; Casello and Usyukov, 2014; Hood et al., 2011; Menghini et al., 2010) three categories of explanatory variables are identified: individual, network and contextual attributes. Table 5.1 shows an overview of all attributes, including how they influence route choice for cyclists.

Table 5.1: Attributes and their influence on cyclists' route choice, based on findings in (Broach et al., 2012; Casello and Usyukov, 2014; Hood et al., 2011; Hunt and Abraham, 2007; Menghini et al., 2010; Sener et al., 2009)

Individual attributes		Network attributes		Contextual attributes	
Attribute	*Influence*	*Attribute*	*Influence*	*Attribute*	*Influence*
Gender	+/-	Distance	Negative	Sunset and Sunrise times	+/-
Age	+/-	% on cycle path	Positive	Weather (rain)	+/-
Cycling experience	+/-	Gradient	Negative	Crime rate / Safety	+/-
Income	+/-	Travel time	Negative	Aesthetics (Canal / Park)	+/-
Household size	+/-	Travel speed	Positive	Sweeping / Snow ploughing	+/-
		Maximum speed (cars)	Negative	Cycling season	+/-
		# Stop signs	Negative	Trip purpose	+/-
		# Intersections	Negative		
		# Bridges	Positive		
		# (Left) turns	Negative		
		% Wrong way	Negative		
		Pavement surface quality	Positive		
		Continuity of cycle paths	Positive		
		Traffic volume (cars)	Negative		
		On-street parking	Negative		
		# Traffic lights	Positive (*8*), Negative (*9*)		

+/- *Not estimated as a separate attribute*

Individual attributes are commonly incorporated in SP studies, mainly as interaction terms, to identify differences in attitude between individuals with respect to network attributes. Looking at RP studies, this means that next to observing actual behaviour, a questionnaire for socio-demographics is necessary. Although, the privacy of the respondent needs to be preserved. In the RP studies, Hood et al. (2011) have included gender and cycling experience in their model, but for example Menghini et al. (2010) did not have these personal attributes at the individual level.

The network attributes that were found to be most influential on route choice behaviour are distance, gradient and cycle path percentage (e.g. Broach et al., 2012; Menghini et al., 2010). Regarding gradient different approaches are applied in literature. For example, Broach et al. (2012) divided sections of the route into different categories of up-slope, whereas Menghini et al. (2010) adopted the maximum gradient of the route. With respect to cycle paths, Furth (2012) identifies four categories: shared streets and lanes, cycling lanes, separate cycle paths and standalone paths. Menghini et al. (2010) only take into account the third category, whereas Hood et al. (2011) take the first, second and fourth category into account.

Contextual attributes are mostly found in SP studies, however also in RP studies trip purpose is found to be influential (e.g. Broach et al., 2012; Hood et al., 2011). Commuting cyclists tend to value distance more negative compared to other purposes.

Based on the literature and the constraints on the availability of information, the following attributes are selected for this study: distance, percentage on separate cycle path (third category in Furth, 2012), number of intersections, rain, sunset and sunrise times and trip purpose. Due to privacy issues, the BCW dataset does not contain any personal information at the individual level.

5.3. Estimating a Cyclists' Route Choice Model

This section provides the analysis of the data collected for the inner-city of Amsterdam. First the descriptive statistics for the trips collected in the inner-city are presented (5.3.1). Then, the specification of the estimated models is described (5.3.2) and the results of the model estimations are discussed (5.3.3).

5.3.1. Analysis of the Trips Cycled in Amsterdam

For this study the selected network attributes are distance, percentage of separate cycle path and number of intersections per km. Table 5.2 shows the range, mean and standard deviation of these attributes for all alternatives. As can be expected from the restriction of the case study area to the inner-city, the average route distance is relatively small. However, the longest route is relatively long as it exceeds 6km. Separate cycle paths are only encountered on roads with a speed limit of 50 km/h or higher. In the inner-city cyclists share roads with motorised traffic and large volumes of pedestrians (Furth, 2012). Therefore, on average, a low percentage of a route's length falls onto a separate cycle path (36%). The number of intersections crossed per km also varies largely and is, as can be expected in a dense urban area, fairly high.

Table 5.2: Descriptive statistics of cycling trips in Amsterdam

Attribute	Description	Range	Mean	Standard Deviation
Distance (km)	Route length	0.13 – 6.69	1.96	1.02
Percentage on separate cycle path	Percentage of the route with a cycle path which is separated from motorised traffic	0% – 100%	36.2%	25.5%
Number of intersections per km	Average number of intersections crossed per km (straight and turn)	1.75 – 50.8	16.8	5.8

The selected contextual attributes are translated into dummy variables. Even though the trip purpose is unknown, two proxy variables can be derived. Firstly, the time of day at which the trip has started is an indicator for commuting to or from work or school (peak hours) versus recreational or social trips (off-peak hours). Secondly, the trip type can be an indicator for cycling only trips or access and egress as part of a multimodal trip. Two train stations are situated within the inner-city boundaries; Amsterdam Centraal and Amsterdam Muiderpoort. Trips starting at one of these stations are considered egress and the trips ending at these stations are considered access, relative to the multimodal trip.

Only 14% of the trips are undertaken in darkness. Most trips are cycled (28%) during the morning peak hours from 7AM to 10AM, followed by trips during daytime from 10AM to 5PM (27%). Almost half of the trips experienced rain showers (46%). Access and egress are equally represented in the dataset (each 9%), implying that most trips in the dataset are cycling only (not directed to or from a train station).

5.3.2. Specification of the Route Choice Models

The most commonly used model to estimate cyclists' route choice, like estimated by Casello and Usyukov (2014), is the MNL model. This model assumes that unobserved variables that influence the utility of a given route are uncorrelated across routes, which is an assumption that is unrealistic when routes overlap. The routes included in this study exercise some degree of overlap, therefore violating this assumption.

To account for overlapping routes, multiple solutions have been proposed in literature. The model structure applied in other cyclists' route choice studies is the PSL model (e.g. Broach et al., 2012; Hood et al., 2011; Menghini et al., 2010), which introduces a similarity measure in the utility function to account for the overlap. This approach maintains the MNL structure, making it easy to compute. For the calculation of the path size (PS) factor, different approaches have been put forward, however no straightforward answer can be provided to the question which performs best. For example, the PS factors developed in a later stage can have illogical route probabilities (Frejinger and Bierlaire, 2007), whereas the earlier versions of the PS factor do not take large differences in route length into account (Ben-Akiva and Bierlaire, 1999). In this study the path size factor put forward by Ben-Akiva and Bierlaire (1999) is adopted, because no large deviations in route lengths are present in our study:

$$PS_{in} = \sum_{a \in \Gamma_i} \left(\frac{l_a}{L_i}\right) \frac{1}{\sum_{j \in C_n} \delta_{aj}} \tag{5.1}$$

where Γ_i is the set of links in route i, l_a is the length of link a, L_i is the length of route i and δ_{aj} the link-route incidence variable which equals one if link a is on route j and zero otherwise. The probability of choosing route i given choice set C_n is specified the following way (Ben-Akiva and Bierlaire, 1999):

$$P(i|C_n) = \frac{e^{\left(\beta_d * \text{Distance}_{in} + \beta_{int} * \frac{\text{Intersections}}{\text{km}}_{in} + \beta_{PS} * \ln PS_{in}\right)}}{\sum_{j \in C_n} e^{\left(\beta_d * \text{Distance}_{jn} + \beta_{int} * \frac{\text{Intersections}}{\text{km}}_{jn} + \beta_{PS} * \ln PS_{jn}\right)}} \tag{5.2}$$

where β_{PS} equals 0 when estimating a MNL model and PS is the path size factor calculated in Equation 1. PS lies between 0 and 1, where 1 means no overlap and 0 full overlap. The latter is not possible, as fully overlapping routes are excluded on beforehand. The natural logarithm of PS is then negative. In this study both the MNL and PSL modelling structure are adopted in order to determine the effect of overlap on the route choice of cyclists. The models are estimated using the Biogeme package (Bierlaire, 2003).

5.3.3. Estimated Cyclists' Route Choice Models

Both a MNL and PSL model are estimated, in order to test for the effect of overlap in the model. To come to these models, all network attributes have been included in the model estimation, and insignificant attributes have been removed to find the most efficient model. For the third model, a stepwise approach is used to add context attributes to the model as interaction terms when a significant and interpretable result is found, resulting in the extended PSL model.

Discussion of the Modelling Results

The results of three model estimations are summarised in Table 5.3. In the MNL model both distance and the number of intersections per km have a significant influence on route choice. Increasing the average distance with one percent results in a 0.50% decrease of being chosen (ceteris paribus) and increasing the average number of intersections per km with one percent results in a decrease of 0.53%. Cyclists prefer fewer intersections per km, our hypothesis is that either they want to reduce interaction with other road users and avoid delays or they want to reduce the cognitive effort during the trip. Translating this to the network of Amsterdam, cyclists avoid the historical city centre and dense residential areas due to the presence of many intersections per km and they prefer the ring streets because of fewer intersections per km.

Cyclists in Amsterdam are willing to cross 7.73 more intersections for a one kilometre shorter route, which is fairly high but reasonable for an urban environment.

Table 5.3: Estimated cyclists' route choice models. *Significant at the 10% level, **Significant at the 5% level, EL = equally likely model

	MNL model		PSL model		Extended PSL model	
	Coef.	*t-stat*	*Coef.*	*t-stat*	*Coef.*	*t-stat*
Distance (km)	-0.255	-2.36**	-0.182	-1.65*	-0.057	-0.48
Morning peak	-	-	-	-	-0.525	-1.77*
# Intersections/km	-0.033	-4.84**	-0.029	-4.00**	-0.029	-4.09**
Ln (Path Size)	-	-	-0.252	-2.13**	-0.248	-2.11**
Adjusted rho-square (compared to EL)	0.003		0.005		0.005	
Likelihood ratio	31.165		43.809		48.694	
Initial Log-likelihood	-4,167.376		-4,167.376		-4,167.376	
Final Log-likelihood	-4,151.794		-4,145.472		-4,143.030	
# Observations	3045		3045		3045	
Number of parameters	2		3		4	

In the PSL model, the path size term is added. As this term is calculated based on overlap in terms of distance, this factor decreases the impact of distance on the total utility, which is now only significant on a 90% confidence level. Increasing the number of intersections by one percent reduces the probability of being chosen by 0.43%, while one percent increase in distance reduces this probability by 0.36%. The impact of the path size factor depends on the degree of overlap. One percent increase for nearly unique routes increases the probability of being chosen by approximately 0.3%, whereas for routes that are almost identical to other routes this is 2.7%. For a route that is one kilometre shorter, cyclists are willing to cross 6.28 intersections, which is slightly lower than in the MNL model.

The PSL model has been extended to include context attributes as interaction terms. The time of day, in particular the morning peak, was found the only significant explanatory variable. The other contextual attributes (rain, sunset and sunrise times and access/egress) did not yield any interpretable significant influence on the network attributes. Morning peak hours (7AM-10AM) are characterised by commuters heading to work, where schedule constraints are more likely. Model estimates show that morning trips are characterised by a significantly greater repelling effect for distance compared to other times of the day. One percent increase in the average distance results in a decrease in the choice probability of 3.4% for cyclists travelling in morning peak and only 0.1% for other times. In this model, cyclists travelling during morning peak are willing to cross 20.07 more intersections for a one kilometre shorter route, whereas during other times this is only 1.97 intersections. The differences in the trade-off clearly show the aversion towards distance of cyclists during morning peak hours.

The path size parameter for the PSL models is significant and negative, indicating that paths that have a high degree of overlap are more likely to be chosen than others (ceteris paribus). Previous studies estimating cyclists' route choice models found a significant positive path size parameter (e.g. Broach et al., 2012; Hood et al., 2011). Therefore, this finding may seem counterintuitive at first, as it does not penalize the routes that overlap but rather increases their choice probability. However, there is evidence that overlapping routes are sometimes valued higher than non-overlapping routes. This because overlap can reduce the uncertainty of the route followed, as was for example found by Lam and Xie (2002) in the context of public transport. Cyclists might prefer routes that offer more downstream decision points to improve route choice robustness. In addition, this might be a result of the characteristics of this case study. In Amsterdam the radial routes provide the backbone of many attractive routes, causing overlapping routes to be valued positively. More behavioural research is needed in order to draw more general conclusions.

Comparison of Model Structures

The PSL model performs significantly better than the MNL model at the 5% level based on the log-likelihood ratio test (12.64 > χ^2). Furthermore, the model fit for the PSL model is higher (compared to the equally likely model). This indicates that including the path size factor to incorporate overlap in the model is beneficial for the interpretation of the results and the prediction of route choice for cyclists. The PSL modelling structure is therefore considered more suitable for estimating route choice models than the MNL structure. Furthermore, the extended PSL model performs significantly better than the PSL model at the 5% level (4.88 > χ^2), meaning that interaction term increases the model fit. The extended PSL model is therefore the best of the three models.

Model Fit

The model fit for all models is very low. In previous studies, where choice set generation algorithms were applied, model fit varied between 23 to 28 percent, significantly better than in this study (Broach et al., 2012; Hood et al., 2011; Menghini et al., 2010). As mentioned before, our hypothesis is that estimating discrete choice models using the empirical approach for composing the choice set, results in a low model fit. Experiments with adding fictional route alternatives that are inferior to the one most commonly selected confirm that model goodness-of-fit improves substantially by artificially enlarging the choice set. This indicates that the application of a generation algorithm leads to over fitting of the data. Furthermore, variance over the alternatives is low in the dataset, most likely due to the fact that cycling costs effort. For example, the shortest route is chosen in 32.6% of the cases, and in 41.4% of the cases the distance of the chosen route is only 10% more than the shortest route, which means on average only 0.2km difference. This implies that it is more difficult to estimate a distance coefficient in the model estimation when constructing the choice set using the empirical approach compared to the algorithm approach.

5.4. Conclusions

This paper presented the findings of a cyclists' route choice model estimated for the inner-city of Amsterdam, aimed at identifying the determinants influencing route choice in a network where cycling is the primary travel mode. Choice models were estimated based on detailed GPS data comprising more than 3,000 trips performed over the course of one week in September 2015.

It is possible to estimate a route choice model for cyclists based on only GPS trajectory data. The results of the estimated route choice models are mostly in line with literature (Broach et al., 2012; Casello and Usyukov, 2014; Hood et al., 2011; Menghini et al., 2010). However, previous cyclists' route choice studies that have used GPS data found that the percentage of the route lying on a separate cycle path is a very important factor for route choice (Broach et al., 2012; Casello and Usyukov, 2014; Hood et al., 2011; Menghini et al., 2010), whereas this study finds no such significant relation. This is presumably due to guidelines for Dutch infrastructure, where cyclists are specifically taken care of. For example, cyclists and motorised traffic are only mixed on streets where the speed limit is 30 km/h and the traffic volume is under 4000 vehicles/day, mitigating the safety risks (Furth, 2012). This finding suggests that when cycling is indeed well-established, separate cycle paths do not necessarily attract cyclists. This might however be due to the location of the cycle paths in the network, which is on the ring streets and not in the centre of the city. More research is necessary for studying how an increasingly dense network of bike paths might lead to a reduction in their importance as route choice determinant. Our results for distance are overall in line with previous studies, although the impact of distance is less pronounced than in other studies (Broach et al., 2012; Casello and Usyukov, 2014; Hood et al., 2011; Menghini et al., 2010). Previous studies calculated more specific attributes related to the number of intersections per km, like number of turns per km, number of signalised intersections per km and number of stop signs per km (e.g. Broach et al., 2012; Hood et al., 2011), they were all found to influence the route choice behaviour of cyclists negatively, however a proper comparison cannot be made. Distance and the number of intersections per km are evidently important regardless of the level of penetration of the cycling. During the morning peak, when people cycle to work or school, distance looms more negatively than during other times of the day, which is consistent with the findings reported by Broach et al. (2012).

In this study both the MNL and PSL modelling structure are adopted in order to determine the effect of overlap on the route choice of cyclists. The effect of taking overlap into account in the model estimation is large as it increases the explanatory value of the model. Routes that are overlapping are valued higher than non-overlapping routes, several explanations can be found for this phenomenon. First of all, an empirical approach for choice set identification is adopted in this study instead of the often used path generation algorithms. This approach allows us to overcome the common shortcomings of not generating the chosen route and having a large number of non-chosen alternatives, by using only the observed routes per OD pair in the dataset in constructing the choice set. Some links are attractive to all cyclists, probably because they form the most direct path to the destination (Bovy and Stern, 2012) or they could have some non-observed advantage. Consequently, it is likely that cyclists choose routes that include these links and because the observed routes form the basis of our choice set, routes with a higher degree of overlap are common and preferred. Another explanation is that the uncertainty of the chosen route is lower when routes overlap and alternatives are present, this can be especially helpful when for example road works are encountered. This explanation also relates to the physical effort needed for cycling. The alternatives available near overlapping routes are usually similar in terms of physical effort, whereas a non-overlapping route might require more physical effort (e.g. longer distance).

The use of the empirical approach for identifying the choice set has its limitations. In particular, the choice set depends on the observed choices and might thus vary for different samples. Filtering is required in data processing, which can be natural in case the circumstances are useful (like here), but this can also be restrictive on generalisability of the results. In addition, the low model fit of the estimated models is attributed to the use of the empirical approach as confirmed by experimenting with the addition of fictive routes. Finally, the positive value found for overlapping routes might be the result of adopting this approach, however this approach should be tested on other datasets in order to draw a more definitive conclusion.

This study was the first to include data from a city where cycling is a well-established and prominent travel mode. Our findings suggest that there are noticeable differences between this case study area which has few comparable cases, and cities where cycling is almost absent.

We recommend also including socio-demographic variables, such as gender, age and cultural background into future data collection and analysis in order to allow identifying their importance. Furthermore, we expect that including more network attributes will help improve interpretation, practical applicability and model fit. Also, for future research we would like to explore more modelling structures, as they might be better suitable for modelling cyclists' route choice. Possible interesting structures are latent class, nested logit and mixed logit. Next to that, we want to explore the sensitivity of the estimation results to the generated choice sets using the empirical approach. Furthermore, we are interested in testing how individual knowledge and familiarity with the network influences route choice when cycling, we expect that this will help understanding the relationship with overlapping routes. Moreover, nowadays more and more people use mobile devices to plan activities and routes, potentially influencing how they travel. Finally, cycling route choice models can be integrated into an activity scheduling and mode choice model, in order to assess their inter-relation with other modes in transport demand forecasting.

Acknowledgements

This research was supported by the Allegro (Unravelling slow mode travelling and traffic: with innovative data to create a new transportation and traffic theory for pedestrians and bicycles) project which is financed by the European Research Council and Amsterdam Institute for Advanced Metropolitan Solutions. The data for this research was provided by the initiators of the 'Fiets Telweek'.

Chapter 6 – Evaluating the Experienced Route Choice Set

This chapter is based on the following article:

Ton D., Duives D.C., Cats O., and Hoogendoorn S.P. (2018). Evaluating a data-driven approach for choice set identification using GPS bicycle route choice data from Amsterdam. *Travel Behaviour and Society.* 13:105-117. https://doi.org/10.1016/j.tbs.2018.07.001

Abstract

Specifying the choice set for travel behaviour analysis is a non-trivial task. Its size and composition are known to influence the results of model estimation and prediction. Most studies specify the choice set using choice set generation algorithms. These methods can introduce two types of errors to the specified choice set: false negative (not generating observed routes) and false positive (including irrelevant routes). Due to increased availability of revealed preference data, like GPS, it is now possible to identify the choice set using a data-driven approach. The data-driven path identification approach (DDPI) combines all unique routes that are observed for one origin-destination pair into a choice set. This paper evaluates this DDPI approach by comparing it to two commonly used choice set generation methods (breadth-first search on link elimination and labelling). The evaluation considers the three main purposes of choice sets: analysis of alternatives in the choice set, model estimation and prediction. The conclusion is that the DDPI approach is a useful addition to the current choice set identification methods. The findings indicate that in analysing alternatives in the choice set, the DDPI approach is most

suitable, as it reflects the observed behaviour. For model estimation the DDPI approach provides a useful addition to the current choice set generation methods, as it provides insights into the preferences of individuals without requiring network-data for additional information or generating routes. In terms of prediction, the DDPI approach is not suitable, as it is not able to perform well with out-of-sample data.

6.1. Introduction

In the context of travel behaviour, many choices must be made by an individual before a trip is made, e.g. destination, mode and route choice. These choices are all discrete in nature, meaning that only one option can be chosen at a time. The choice set from which an individual chooses one, forms an important aspect in the analysis of travel behaviour. Three different purposes of choice sets can be identified. First, it is essential in analysing different travel options in the network (e.g. number of alternatives, characteristics or composition of the alternatives), second it is used for demand model estimation (estimating behavioural parameters), and third it is instrumental in predicting choice probabilities and thereof flow distribution over alternatives/the network (Bovy, 2009). The size and composition of the choice set influence the results of the model estimation and prediction, and consequently the interpretation of the estimated behavioural parameters (Bovy, 2009). This issue is for example relevant in route choice analysis, as many possible alternatives can be identified by the researcher, but only few will be known to the individual, leading to possible mismatches in the choice set identification. Route choice sets are often specified using *choice set generation algorithms* (e.g. k-shortest paths or labelling), which compute a set of routes based on characteristics of the network(-links) (e.g. distance or travel time). The use of these algorithms can introduce two types of errors in the choice set: false negative and false positive errors. False negative errors arise when the algorithm is not able to reproduce the chosen alternatives. The generated alternatives might not match the behaviour and preferences of the individual, and as a result the chosen route is not reproduced. The impact of this error decreases when the ability of the choice set generation algorithm to capture the individuals' behaviour and preferences increases. False positive errors occur when a choice set generation algorithm also generates routes that are not considered by the individual, resulting in a too large choice set. In conclusion, the use of choice set generation algorithms potentially comes with several flaws.

In recent years, large improvements have been made in revealed preference data collection methods. New data sources, such as GPS data that contain detailed spatial and temporal information on the movement pattern of individuals, help creating insights into the individuals' choice behaviour. By combining the GPS records belonging to one individual into separate trips, the observed trips can be used for route choice research (e.g. Hood et al., 2011; Menghini et al., 2010). Next to generating the choice set based on a set of assumptions on network properties, it is then also possible to use the observed trips from GPS data to identify the choice set directly. Every trip between an origin and destination follows a certain route, the unique routes that are observed can then be combined into one choice set. Consequently, the potential false negative error associated with choice set generation algorithms cannot occur and the potential false positive error is negligible because all the routes included in the choice set have been chosen by the individual.

Governments worldwide have shown increasing interest in promoting and understanding cycling usage, due to the potential health, congestion and emissions benefits. Consequently, goals have been set to increase the cycling modal share (Pan-EuropeanProgramme, 2014). Several studies investigated bicycle route choice using GPS data, primarily in areas where cycling is relatively scarce, with the goal of identifying determinants that influence route choice, so that substantiated infrastructure investments can be made

(Broach et al., 2012; Casello and Usyukov, 2014; Chen et al., 2018; Ghanayim and Bekhor, 2018; Hood et al., 2011; Li et al., 2017; Menghini et al., 2010; Montini et al., 2017; Zimmermann et al., 2017). Other studies have taken in place in urban environments with a larger share of cyclists, like Copenhagen (Halldórsdóttir et al., 2014; Prato et al., 2018; Skov-Petersen et al., 2018). These studies have applied different types of choice set generation algorithms, such as labelling, stochastic methods, link elimination, and link penalty. However, none of the studies has applied a data-driven method for choice-set identification as proposed and examined in this study. This approach is applied to a bicycle route choice study for the city of Amsterdam, the Netherlands (Ton et al., 2017). Amsterdam is known for its well-developed bicycle infrastructure and high share of bicycling trips (37%) (OViN, 2011). To evaluate the potential of this data-driven method for choice set identification, we compare the method using the dataset from the city of Amsterdam, to other choice set generation algorithms previously applied in the cycling route choice literature.

This paper evaluates the use of a data-driven approach for choice set identification in travel behaviour analysis. The goal is to investigate whether a data-driven approach can be a valuable addition to the current choice set identification methods. Bicycle GPS data from Amsterdam, the Netherlands, is used to identify the choice set and this choice set is used in the estimation and validation of a route choice model. The evaluation of the data-driven approach is done by means of a comparison study, where it is compared to two commonly used choice set generation methods, to assess and compare their performance and results. Based on computation time, sensitivity to false negative errors and, number of applications, two approaches have been selected: the breadth-first search on link elimination (BFS-LE) introduced by Rieser-Schussler et al. (2013) and the labelling approach introduced by Ben-Akiva et al. (1984). The evaluation is performed on the three abovementioned purposes of choice sets; 1) analysing the composition of the choice set, 2) understanding behaviour (model estimation) and 3) application of the model on out-of-sample data (model validation).

The rest of the paper is outlined as follows. Section 6.2 reviews contemporary choice set generation procedures. In section 6.3, the data-driven approach is elaborated upon in terms of requirements of data, opportunities, limitations of the method, and sensitivity with respect to data collection duration. Section 6.4 describes the methodology for evaluating the specified choice sets as well as the route choice model estimation and validation. Section 6.5 provides background on the data that was collected and prepared for this study. Section 6.6, then details the evaluation of the generated choice sets in comparison to the observed routes and section 6.7 covers the evaluation of the route choice model estimation and validation. Finally, section 6.8 concludes the paper and provides directions for future research.

6.2. Choice Set Generation Methods

This section discusses different choice set generation methods that have been proposed in the past and selects two methods as reference for the evaluation of the data-driven approach.

Many different methods have been proposed for identifying route choice sets (for detailed reviews see Fiorenzo-Catalano (2007) and Ramming (2002)). Bovy (2009) and Prato (2009) identify four categories of choice set generation methods: deterministic methods, stochastic methods, probabilistic methods and constrained enumeration methods. Most choice set generation methods belong to the *deterministic category* and consist of repeated shortest path searches in the network. These shortest path methods have different input variables such as search criteria, route constraints and link impedance (Prato, 2009). They are computationally attractive due to the efficiency of shortest path algorithms. *Stochastic methods* are also based on repeated shortest path searches, but additionally the computation of optimal paths is randomised based on link impedances or individual preferences drawn from probability distributions, mostly done using simulation. These methods have been applied in the bicycle route choice context by Hood et al. (2011), Halldórsdóttir et al. (2014), Ghanayim and Bekhor (2018), and Prato et al. (2018). *Constrained enumeration methods* are not only based on shortest routes, but also make additional behavioural assumptions (Prato, 2009). These assumptions reflect different behavioural thresholds that can be specified, e.g. excluding loops and only including links that bring the individual closer to the destination. These methods have been applied in the bicycle route choice context by Halldórsdóttir et al. (2014), but did not prove to outperform the deterministic or stochastic methods. *Probabilistic methods* assign a probability for each alternative to be included in the choice set. A fully probabilistic approach, as proposed by Manski (1977), which includes the choice set generation and selection in the utility function, is often deemed infeasible due to its computational complexity. As a consequence, these methods have not yet been applied in the bicycle route choice context.

Recently, two alternative approaches have been proposed that address the choice set identification implicitly (i.e. no need for explicit enumeration of alternatives). The first is the *sampling approach* (Flötteröd and Bierlaire, 2013; Frejinger et al., 2009), that assumes a universal choice set and by means of importance sampling selects a subset of these routes. The second approach is the *link-based approach* (Fosgerau et al., 2013), which assumes that individuals make successive choices at each node. The link-based approach was applied in the bicycle route choice context by Zimmerman et al. (2017).

Due to its prevalence in the general and bicycle route choice literature, computational efficiency and deterministic nature (which relates more to the cognitive aspects of the decision-maker rather than being conceived as a computational instrument), *deterministic methods* are selected as reference methods for comparison in this study. Four categories of deterministic methods are identified: shortest paths, link elimination, labelling and link penalty. Previous findings suggest that the *shortest path methods* have the lowest performance in terms of reproducing the observed routes (Bovy, 2009). Furthermore, the *link penalty methods* are known for their large computation times (Bekhor et al., 2006). Therefore, the focus lies with the link elimination and labelling methods.

The *link elimination method* iteratively removes links that are on the shortest path and finds new shortest paths (Bellman and Kalaba, 1960). Prato and Bekhor (2007), Bekhor et al. (2006), and Ghanayim and Bekhor (2018) evaluated this approach and found that in about 40% of the cases false negatives are produced. Azevedo et al. (1993) proposed an alternative approach, where the entire shortest path is eliminated, after which a new shortest path is calculated. This approach is more drastic, as it eliminates overlap but can result in an unrealistic choice set (e.g. large detours). Rieser-Schüssler et al. (2013) adapted the link elimination method by applying a breadth-first search technique on link elimination (BFS-LE), meaning that one starts eliminating links closest to the origin, repeats the shortest path search and moves stepwise towards the destination, before going one level deeper and eliminating two links at once (the one removed in the first level and again the first link of the new shortest route). They found lower error percentages compared to previous implementations of the link elimination method. Furthermore, this method appears to be computationally efficient and is suitable for high density networks (Rieser-Schüssler et al., 2013). It has been applied in different contexts, e.g. cars (Dhakar and Srinivasan, 2014; Montini et al., 2017; Prato et al., 2012; Rieser-Schüssler et al., 2013), bicycles (Halldórsdóttir et al., 2014; Menghini et al., 2010; Montini et al., 2017), heavy goods vehicles (Hess et al., 2015), and public transport (Montini et al., 2017).

Ben-Akiva et al. (1984) introduced the *labelling approach* which searches for the most optimal alternative given a certain label (e.g. distance, time, number of turns etc.). Prato and Bekhor (2007) applied this method to an urban network for cars in which they minimise for distance, free-flow time, travel time and travel delay. They report a false negative rate of 60%. Bekhor et al. (2006) specified and examined 16 different labels in their study. They found that each individual label generates only between 8% and 34% of the observed alternatives, while combined they can reproduce 72% of the observed routes. This method has been applied in the bicycle route choice context by Chen et al. (2018), Li et al. (2017), and Skov-Petersen et al. (2018). Unfortunately, none of them evaluate the performance of this method. Dial (2000) proposed a generalised approach of the labelling method for generating efficient paths. This method minimises a linear combination of labels. Broach et al. (2010) extended the labelling approach by generating multiple optima for one label by varying the label cost function parameter. They applied the method to bicycle traffic and identified eleven different labels, among others the distance of upslope travel and the number of turns. Their method generated more observed alternatives than the labelling method, however, the computation time also increased manifold. They also applied this method in a later study (Broach et al., 2012).

Table 6.1 provides an overview of the performance of the discussed methods in terms of producing false negatives in comparison to the number of alternatives generated. Note that the studies mentioned before are only included in the table if these numbers were provided. In general, when generating more alternatives, the false negative error percentage should decrease (where the false positive error potentially increases). Next to that, computation time of the methods is compared.

Table 6.1: Performance of applied deterministic choice set generation algorithms

Deterministic category	Method	Study	Data	Mode	False negative error	Max no. alternatives	Comp. time
Link elimination method	Link elimination	(Bekhor et al., 2006)	Boston, USA	Car	40%	?	Medium
		(Prato and Bekhor, 2007)	Turin, Italy	Car	42%	10	-
		(Ghanayim and Bekhor, 2018)	Tel-Aviv, Israel	Bicycle	40%	10	-
	Breadth-first search on link elimination	(Rieser-Schussler et al., 2013)	Zurich, Switzerland	Car	37% 27%	20 100	-
		(Hess et al., 2015)	United Kingdom	Trucks	26%	15	-
		(Halldorsdottir et al., 2014)	Copenhagen, Denmark	Bicycle	34%	20	Medium
Labelling approach	Labelling	(Bekhor et al., 2006)	Boston, USA	Car	28% 61%	16 3	Low
		(Prato and Bekhor, 2007)	Turin, Italy	Car	60%	4	-
		(Broach et al., 2010)	Portland, USA	Bicycle	80%	9	Low
	Calibrated labelling	(Broach et al., 2010)	Portland, USA	Bicycle	78%	20	Medium

Because the studies use different datasets, it is hard to objectively compare the results. Most studies have resulted with a relatively high number of alternatives in the choice set, indicating that both relevant and irrelevant alternatives are included in the choice set. The different studies have also addressed different modes; the false negative error percentage is higher for the non-motorised modes compared to the motorised modes for each algorithm. This is most likely due to the higher complexity of the network for bicycles compared to cars and trucks.

From the link elimination methods, the BFS-LE approach introduced by Rieser-Schüssler et al. (2013) is most promising and therefore selected as a reference method in this paper. Several other studies have applied this method and found decent computation times and a lower share of false negatives compared to the original link elimination approach. Furthermore, the original labelling approach introduced by Ben-Akiva et al. (1984) is included as a reference method, because it outperforms the later proposed method of Broach et al. (2010)

in terms of computation time and performs only slightly worse in terms of producing false negative errors.

6.3. Introducing the Data-Driven Path Identification Approach (DDPI)

Due to the increased availability of (passively) collected revealed preference data and the issues associated with current choice set generation algorithms, the opportunity arises to identify choice sets using a data-driven approach. In this section, the data-driven approach coined *Data-Driven Path Identification* (DDPI) which is introduced in a previous study by the authors (Ton et al., 2017), is elaborated upon.

The DDPI approach is based on revealed preference data, like Wi-Fi, Bluetooth or GPS data of a large sample of individuals collected over a longer period. The idea behind this approach is to combine all observed routes from one origin to one destination into a single choice set at the origin-destination level (OD Pair). Using this method, the false negative error (not reproducing the observed route) is resolved. Furthermore, all routes that are included have been chosen by an individual, this means that these routes are optimised to a certain extent. Consequently, it is likely that these routes have been considered by an individual and from this set one route has been chosen. Therefore, the proposed method is expected to be less prone to false positive errors (including routes that are not considered) than choice set generation algorithms. However, because the choice set contains only chosen routes, it is possible that other routes that were considered but not chosen, are excluded, consequently potentially resulting in a choice set that is too small. A counterargument is that if data is collected over a long enough period of time, all relevant and considered routes are part of the data-driven choice set, therefore reducing this issue.

Several requirements need to be met for the DDPI approach to be applicable. First, the data should be collected over a sufficiently long period of time to allow multiple observations per OD pair. Second, it is necessary to have at least two routes per OD pair to facilitate the estimation of a route choice model. However, because of issues with endogeneity, it is preferable to have more than two routes per OD pair. Because the observed routes are optimised to a certain extent by the individual, the variability of the routes is low. By including more routes, the variability of the routes increases and the issue with endogeneity will be less severe. If this is not accounted for, the estimated models will be biased. If there is an OD pair which does not meet these requirements, it needs to either be deleted or aggregated by applying a spatial clustering technique. Clustering of OD pairs can be useful in case of, for example, two neighbours heading for the same destination. It can prevent loss of data, but should be carefully addressed, because the OD pairs still need to be comparable. The impact of these requirements can be small, if they are taken into account in the design phase of the data collection.

The requirements of the method also point to the limitations of the DDPI approach. It imposes additional requirements to the data collection, because if the data is already collected and requirements are not adequately met, a (severe) loss of data and an endogeneity issue can be the result. The endogeneity is the result of including all chosen alternatives in the choice set. The issue is larger if the alternatives are more similar and there are only few. In that case, the method should not be used, as it imposes a bias in the choice model. Similar to other methods, another limitation is found in the generalisability of the results: data is collected for a certain group of people and for a certain region. Consequently, it is per definition uncertain whether the results (modelling or choice set) can be transferred to other groups of people or other regions, similarly to the generalisability issues associated with other methods.

The data collection duration (for example a week versus several months) suitable for the application of the DDPI method depends on the local network and demand properties. It is important to ensure a long enough period so that the routes observed exhibit a sufficient degree of variation.

6.4. Methodology for Evaluating Choice Set Specification Methods

The methodology for assessing the usefulness of the DDPI approach and comparing the different choice set generation methods is presented in this section. Section 6.4.1 details the methodology for comparing the generated choice sets to the observed data. Furthermore, section 6.4.2 discusses the evaluation methodology for estimation and validation of the route choice model. Section 6.4.3 then provides a synthesis of the evaluation methodology.

6.4.1. Evaluating the Specified Choice Sets

The specifications of the algorithms to which the DDPI approach is compared are discussed, and the methodology for comparing the generated choice sets to the observed routes is provided.

Selected Choice Set Generation Algorithms

The BFS-LE and labelling approach have been selected for comparison. Both algorithms use calculations of the shortest path. The algorithm used to calculate the shortest path is Dijkstra (Dijkstra, 1959). The input for Dijkstra's algorithm is a (distance)matrix, which can grow very large, especially when considering bicycles. To decrease the computation time and increase the spatial diversity among routes, a topologically equivalent network reduction is adopted in this study. This means that nodes that connect only two other nodes (i.e. a node degree of two) are removed from the network and the two links are merged into one. Consequently, the network (or matrix) consists of fewer nodes and the resulting shortest path consist of fewer links, thus significantly reducing the computation time.

These choice set generation algorithms can utilise several input variables. Mostly, the algorithms are applied based on travel distance. In the bicycle route choice context, several studies have considered alternative variables. Broach et al. (2012) used an approach that optimised criteria like percentage on designated cycle paths, subject to distance constraints. Haldórsdóttir et al. (2014) search for the shortest route in terms of road type, bicycle paths, and land use. Finally, Chen et al. (2018) used a combination of speed limits, distance, and bicycle facilities to generate routes. Due to limited data availability for the inner-city of Amsterdam (see section 6.5.4), we rely largely on travel distance in the choice set generation algorithms. The two algorithms are specified below.

Breadth-first Search on Link Elimination (BFS-LE)

The BFS-LE algorithm, introduced by Rieser-Schüssler et al. (2013), was developed specifically for high-density networks, e.g. urban networks. The idea behind the approach is to calculate the shortest path (in this paper we adopt calculation based on distance, like in the original study) between an origin and destination, add this path to the choice set and then remove the links of this shortest path step-by-step, starting from the origin node. In each step a new shortest path is calculated and added to the choice set, given that it is unique. A tree structure is adopted to keep track of the removed links and the resulting adapted networks, this means that in the second tree level two links are eliminated (the link that was deleted from the shortest path and the link from the new shortest path).

Maximum computation time, tree-depth, and choice set size can be used as termination measures for the BFS-LE algorithm. In this study, we applied a mix of these measures. Because an individual is not able to remember or consider many routes, we have set the maximum to 20 routes. This seems adequate given the findings from Hoogendoorn-Lanser (2005) indicating that different individuals only know seven alternatives. Since we only search for 20 unique routes, we have applied a tree-depth of one, with a random draw of 20 routes in case more routes are generated. The second level sometimes generated over 1,000 routes, and induced an exponential growth in computation time. The unique routes found in tree-depth one, are added to the choice set resulting from tree-depth zero.

Labelling Approach
The labelling approach proposed by Ben-Akiva et al. (1984) searches for the most optimal route based on different network-related search criteria, e.g. distance, travel time or number of left turns. This method facilitates the composition of a very diverse choice set, given the available data. The number of labels encoded, sets the maximum value of the number of alternatives included in the choice set. The input-matrix required for the Dijkstra's algorithm is adapted for each of the labels considered. In this study, we have identified three labels, resulting in a maximum choice set size of three.

The three labels are the shortest path based on distance, the highest percentage on separate cycle paths and the least amount of intersections on the route. The matrix that serves as input for the Dijkstra algorithm is node-based. Consequently, each link is presented as a connection between two nodes. The algorithm then searches in this matrix to identify the shortest path. Regarding separate cycle paths, each link that has a separate cycle path or a protected lane, has a weight of zero, all other links have a weight of one. The ideal route found by the algorithm consists of 100% separate cycle path, thus maximising the amount of cycle path. Furthermore, regarding intersections, each link is assigned with the same weight, therefore the algorithm searches for the shortest path in terms of the number of links traversed. In the absence of more detailed information, all intersections (with a node degree of at least three) are treated equally.

Evaluation Methodology for Specified Choice Sets
The DDPI approach directly uses the observed routes to identify the choice set, consequently there is no difference between the DDPI approach (after data preparation) and the observed routes, and it is not evaluated separately. The performance of the algorithms is evaluated by comparing the generated choice sets to the observed routes. First, a qualitative analysis is performed, in which two OD pairs are selected and visually compared. This gives an indication on the spatial distribution of the generated routes and potential differences and similarities between the choice sets. Second, a quantitative analysis provides descriptive statistics of three network related variables, based on previous work on bicycle route choice (Ton et al., 2017): percentage on separate cycle paths, distance and number of intersections per kilometre. This analysis shows the general characteristics of the different choice sets compared to the observed routes.

Furthermore, the heterogeneity of the generated choice sets is investigated, quantitatively showing how spatially different the generated routes are. This is done by calculating the path size (PS) factor for each route in the choice set, which is an indicator for overlap between routes (Ben-Akiva and Bierlaire, 1999).

$$PS_{in} = \sum_{a \in \Gamma_i} \left(\frac{l_a}{L_i}\right) \frac{1}{\sum_{j \in C_n} \delta_{aj}} \tag{6.1}$$

where $PS_{in}$is the path size factor, Γ_i is the set of links in route i, l_a is the length (distance) of link a, L_i is the length of route i and δ_{aj} the link-route incidence variable which equals one if link a is on route j and zero otherwise. This means that the PS factor depends largely on the size and composition of the choice set (i.e. including many irrelevant routes affects this factor). The path size factor ranges between zero and one, where one indicates an independent route and zero indicates complete overlap with other routes in the choice set.

The main objective of choice set generation algorithms is to reproduce all observed routes, i.e. resulting with zero false negative errors. False positive errors are mostly considered less important. To test to what extent the algorithm can reproduce the observed routes, the following formula for the reproduction rate is adapted from Prato and Bekhor (2007):

$$RR_r = \sum_{n=1}^{N} I(O_{nr} > \delta) \tag{6.2}$$

where RR_r is the reproduction rate for algorithm r. $I\ (\cdot)$ is the reproduction function, which is equal to one if the argument is true and zero otherwise; O_{nr} is the overlap rate for algorithm r for observation n, and δ is the overlap threshold, which can be set from no overlap (0%) to full overlap (100%). O_{nr} is calculated in the following way:

$$O_{nr} = \frac{L_{nr}}{L_n} \tag{6.3}$$

where L_{nr} is the common distance between the generated route and the observed route for algorithm r and observation n. L_n is the total distance of the observed route for observation n. The reproduction rate (Eq. 6.2) yields how many observed routes are generated when allowing for a certain overlap threshold.

In addition to the reproduction rate, the behavioural consistency of both methods is assessed. The consistency index compares the algorithms to the ideal algorithm that would reproduce all the observed routes, and calculates how well the algorithms perform. The formula used to calculate this index is the following (Prato and Bekhor, 2007):

$$CI_r = \frac{\sum_{n=1}^{N} O_{nr,max}}{N} \tag{6.4}$$

where CI_r is the consistency index for algorithm r; $O_{nr,max}$ is the maximum overlap percentage obtained for observation n using algorithm r, i.e. the best matching generated route to the observed route n; N is the total number of observations in the sample.

6.4.2. Evaluating the Model Estimation and Validation

The specifications of the route choice model that is estimated; the Path-Size Logit (PSL) model is discussed and the methodology to evaluate the model estimation and validation is provided.

Specification of the Route Choice Model

A wide variety of discrete choice models, varying in computational complexity, have been developed that are suitable for route choice. Examples are Cross-Nested Logit (CNL), Paired Combinatorial Logit, C-Logit and PSL. Bliemer and Bovy (2008), Prato and Bekhor (2007) and Bekhor et al. (2006) have compared these models for route choice. They concluded that the CNL and PSL model perform best. Since the CNL model is more complex, requires specialised code and has a higher computation time, we apply the PSL model in this evaluation (Bekhor et al., 2006).

To account for potential correlation among path alternatives (e.g. route overlapping), the PSL model introduces a similarity measure in the utility function. In this study, the path size (PS) factor proposed by Ben-Akiva and Bierlaire (1999) is adopted (Eq. 1). The probability of choosing alternative i given choice set C_n is specified as follows (Ben-Akiva and Bierlaire, 1999):

$$P(i|C_n) = \frac{e^{\left(\beta_d * \text{dist}_{in} + \beta_i * \frac{\text{n\_int}}{\text{km}}_{in} + \beta_{cp} * \%\ \text{sep. cycle path}_{in} + \beta_{PS} * \ln PS_{in}\right)}}{\sum_{j \in C_n} e^{\left(\beta_d * \text{dist}_{jn} + \beta_i * \frac{\text{n\_int}}{\text{km}}_{jn} + \beta_{cp} * \%\ \text{sep. cycle path}_{jn} + \beta_{PS} * \ln PS_{jn}\right)}} \tag{6.5}$$

where based on previous work, three explanatory variables are included per alternative i and observation n: percentage on separate cycle paths ($\%$ sep. cycle path$_{in}$), distance (dist$_{in}$) and number of intersections per kilometre ($\frac{\text{n_int}}{\text{km}}_{in}$). PS is again the path size factor calculated in Eq. 6.1, it ranges between zero and one, where one means no overlap and zero implies complete overlap between routes. The models are estimated using the Python Biogeme package (Bierlaire, 2016).

Evaluation Methodology for Model Estimation and Validation

Three route choice models are estimated and validated, using the two generated choice sets and the choice set that is identified using the DDPI approach. Because for each OD pair routes are generated using the two generation algorithms and multiple routes are observed per OD pair, a union of the observed and generated routes is created for the Labelling and BFS-LE choice sets. Figure 6.1 shows this merging of observed (6.1a) and generated (6.1b and 6.1c) routes for the BFS-LE and labelling method. All observed and generated routes for one method per OD pair are merged into one choice set (6.1d and 6.1e), corrected for the reproduced observed routes.

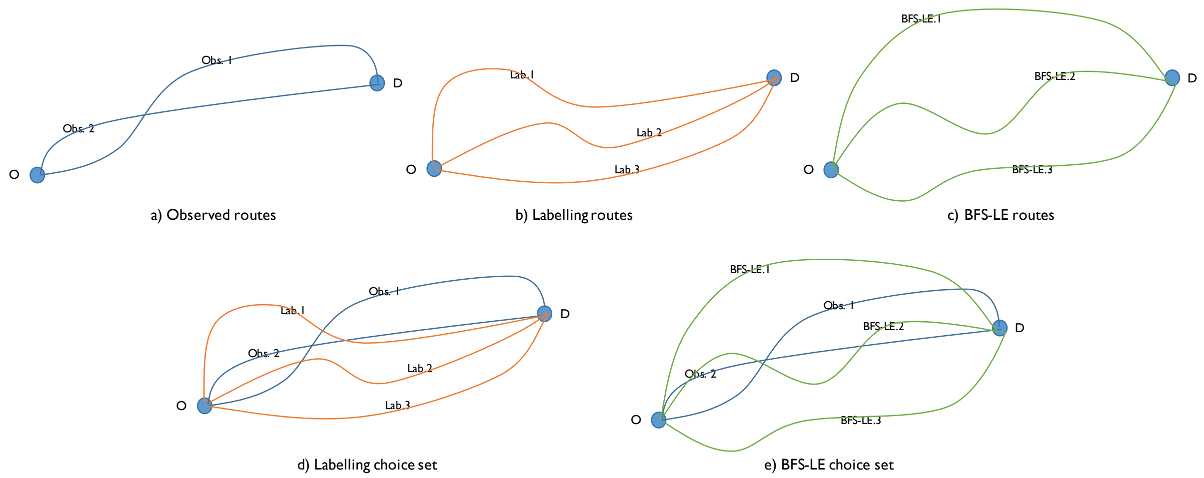


Figure 6.1: Formation of choice sets for labelling and BFS-LE algorithms

The model estimation and validation are done by splitting the data sample into two parts (80/20). The models are estimated using 80% of the observed OD pairs and validated using the remaining 20%. This way, the predictive power of the models can be tested and potential errors can be detected. The model estimation and validation is done for five random draws to test stability of the models. Note that the sampling is done on the OD pairs that result from the DDPI approach, so that the variability in the OD pair remains for the model estimation and the issue with endogeneity is less severe.

Since the models are estimated using different choice sets, a standard comparison based on log-likelihood ratio or model fit (adj. rho-square) cannot be done. The initial log-likelihood is different due the different sizes of the choice sets. Therefore, the comparison is based on the point elasticities of the model's explanatory variables, calculated using the following formula:

$$E_{x_i}^{P_n(i)} = \frac{\partial P_n(i)}{\partial x_i} \frac{x_i}{P_n(i)} \tag{6.6}$$

where $P_n(i)$ is the probability that observation n chooses alternative i and x_i is an attribute (defined in Eq. 6.5) for alternative i. The mean elasticity is then obtained by probability weighting the elasticities for every individual n, where the probability weights relate to the probability of choosing an alternative in the choice set. In the validation phase, the probability for each alternative to be chosen is calculated for the remaining 20% OD pairs. To make a fair comparison between all models, a union of all generated and observed alternatives is generated for each OD pair (in essence a union between Figure 6.1d and 6.1e, corrected for unique routes). The union choice sets for each OD pair are used to assess the predictive power of all models, using three measures. First, the number of times the model assigns the highest utility to the chosen alternative for all observations. This gives an indication about the extent to which the model is able to predict the correct choice. Second, the RMSE value is calculated, which gives an indication of the error that arises between observed probabilities (based on observed routes) and modelled probabilities per OD pair. This value is calculated using the following formula:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N_{OD}} (\hat{P}_i - P_i)^2}{N_{OD}}} \tag{6.7}$$

where $\hat{P}_i$ is the vector of probabilities that is predicted by the model for OD pair i and P_i is the vector of observed probabilities of OD pair i . Finally, the log-likelihood is calculated on the out-of-sample data. As a union of all generated and observed routes is used to define the choice sets, the input is the same for all models. Therefore, a comparison based on log-likelihood is possible. It is calculated using the following formula:

$$\text{Log} - \text{Likelihood} = \sum_{n=1}^{N} \left(\sum_{i \in C_n} y_{in} \ln P(i|C_n) \right) \tag{6.8}$$

where y_{in} is one if n chooses alternative i in choice set C_n, and zero otherwise, and $P(i|C_n)$ is the probability of choosing alternative i in choice set C_n.

6.4.3. Synthesis of the Evaluation Methodology

A concise overview of all the methods introduced for analysis and evaluation of the choice sets, model estimation and model validation is presented in Figure 6.2.

6.5. Data Description and Preparation

The dataset that is used to assess the usefulness of the DDPI approach and benchmark the approach against the BFS-LE and labelling algorithms is a bicycle GPS dataset. This dataset was collected during a nationwide initiative in the Netherlands called the 'Bicycle Counting Week', which took place on 14-20 September 2015. A total of 38,000 cyclists participated using a smartphone application that tracked their cycling movements, recording more than 370,000 trips nationwide. Additionally, a survey was distributed among the participants that used the smartphone application.

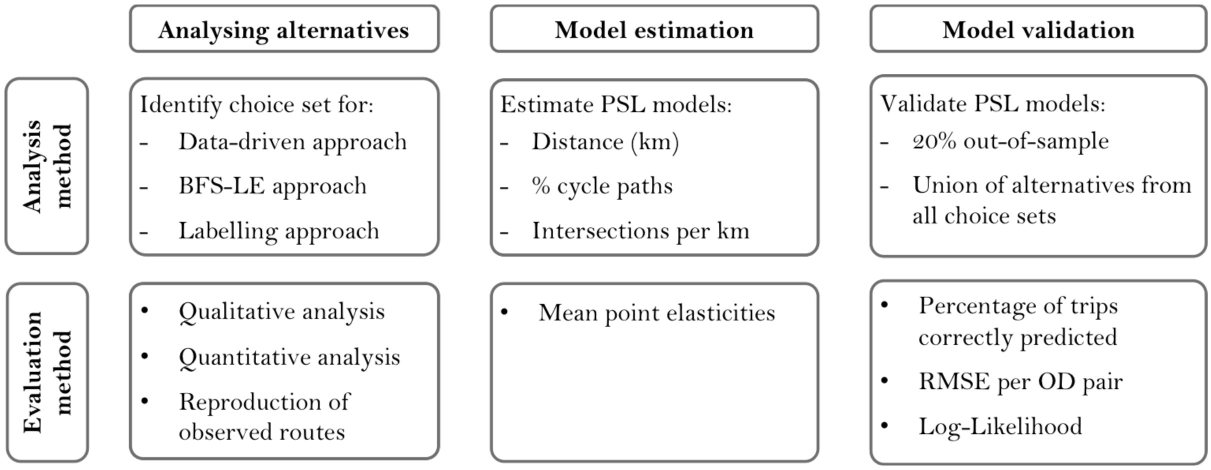


Figure 6.2: Analysis and evaluation methods for analysing the alternatives in the choice set, model estimation and model validation

Section 6.5.1 describes the dataset that is used in this study. Furthermore, section 6.5.2 describes the map matching procedure for matching the GPS trajectory data to the network. Section 6.5.3 provides insights on the clustering procedure applied to the origins and destinations of all the trips made in the dataset. Finally, section 6.5.4 addresses the preparations needed related to the data and network for the choice set generation methods.

6.5.1. GPS Dataset from the Inner-city of Amsterdam

In this evaluation, the focus lies on the inner-city of Amsterdam, which is a densely-built area with well-developed cycling infrastructure. The dataset was used in previous work, where the DDPI approach was applied to estimate a bicycle route choice model for this specific area (Ton et al., 2017). Figure 6.3 shows the network of the inner-city of Amsterdam. In total, 3,045 trips were recorded in the inner-city of Amsterdam. Not all trips could be used in this case study, as some trips were too short to be included and some could not be matched to the topologically equivalent reduced network, resulting in a total of 2,819 trips. The respondents sample consists of equal shares of male and female participants. Most respondents are 31-65 years of age (80%). Most trips are made for commuting purposes (77%). Furthermore, most respondents cycle between 25 and 100 kilometres a week (72%) (FietsTelweek, 2015). The individual characteristics are only available on an aggregate level, due to privacy regulations, therefore it is impossible to link the GPS trajectories to individual travellers. This has two major consequences: (1) individual characteristics cannot be used in the model estimation, whereas several cycling route choice studies have identified the relevance of such variables (Broach et al., 2012; Hood et al., 2011) and (2) it is impossible to identify which trips have been made by which individuals, thus we need to treat each trip as if it was made by a unique individual and cannot therefore test for panel effects in the model estimation.

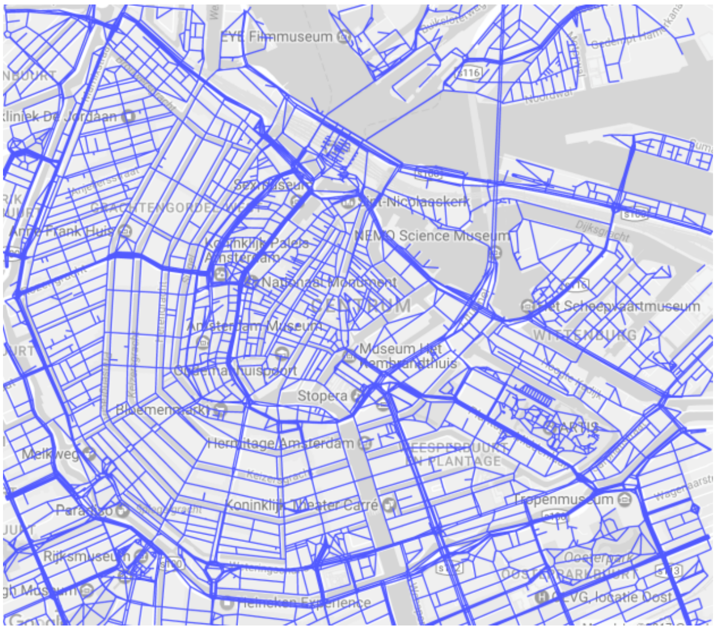


Figure 6.3: Road network of the inner-city of Amsterdam

6.5.2. Map Matching the GPS Trajectory Data

The map matching procedure was conducted by the organizers of the Bicycle Counting Week (van de Coevering et al., 2014). The following is an account of the procedure that has been performed. GPS data points in a trajectory have a maximum accuracy of around 5 meters with respect to the infrastructure. However, outliers are observed in dense urban areas or high building areas, reducing the accuracy by up to 50 meters. In urban areas, this means that the next street can be mistakenly identified. To reduce the impact of these outliers on the observed trajectories, van de Coevering et al. (2014) have calculated the speed between each two consecutive GPS data points and compared it to the actual GPS speed, which was determined by means of Doppler techniques. If a large discrepancy between the actual speed and the calculated speed has been identified, the outlier and two preceding and following GPS data points from the dataset were removed.

The corrected GPS trajectories can afterwards be matched to the network. The entire network is split up in nodes, after which links were divided into smaller segments to determine local differences in network speeds, which helps in determining whether a cyclist was able to cycle on a link. The map matching algorithm they applied generates all possible combinations of origin and destination points in the network, which is necessary because of the inaccuracy of the GPS data points. Routes were then plotted between all the identified combinations of origins and destinations. The goal is to minimise the distance between the GPS trajectory and the network route, which results in routes that best resemble the GPS trajectories. If a match could not be found, this may stem from missing links. In those cases, the route is partitioned and the same procedure is repeated for the sub-routes. For a more detailed description of the map matching procedure, the reader is referred to van de Coevering et al. (2014).

6.5.3. Clustering of the Origins and Destinations of the GPS Trajectories

We applied a clustering method on the observed origins and destinations, to ensure that multiple trips and routes are observed for each OD pair. A k-means clustering approach was applied which minimises the intra-cluster distance and maximises the inter-cluster distance. Different numbers of clusters were tested (150, 200, 250, and 300) to find a good balance between having enough trips per OD pair (high number of clusters) and ability to compare routes in an OD pair (low number of clusters). Finally, a total of 200 clusters provided the best results. For a more detailed description of the clustering, the reader is referred to a previous study by the authors (Ton et al., 2017).

6.5.4. Data and Network Preparations for the Choice Set Generation Methods

As mentioned in section 6.5.1, we cannot identify which individual made which trip, consequently we have to treat every trip-maker as a unique individual. Ideally, the DDPI method would have been applied per individual and OD pair. Given the mentioned restriction in the data, it is not possible to identify individual choice sets. Therefore, this study uses all trips that are observed per OD pair and combines them to form choice sets. Furthermore, data is collected over the course of one week. Consequently, we are not able to test how sensitive this dataset is with respect to the duration of data collection versus the diversity of observed routes. Data would need to be collected over a longer period of time (multiple weeks) in order to test the sensitivity of model performance to the data collection duration.

The choice set generation algorithms use the network of Amsterdam (Figure 6.3) to generate the routes, therefore the network is extracted from OpenStreetMap (OSM). In the road network of OSM the two bicycle/pedestrian ferries crossing the river IJ are not included, therefore two bidirectional links are added to the network with origins and destinations at the ferry landings. Furthermore, the inner-city of Amsterdam contains many one-way streets. Tests with the choice set generation algorithms show that the generated routes contain many detours and illogical routes if these links are not considered to be bi-directional. Therefore, we have converted the entire network into a bi-directional graph. Furthermore, in the OSM network many links that are mainly used by non-motorised modes are not incorporated in the network. Tests with the choice set algorithms show that this affects many OD pairs, therefore these have been added to the network when possible. Still, many links that are used by cyclists, are not included in the network. These links could for example be shortcuts or pedestrian areas, where other modes are not allowed, both of which are not included in the network. Consequently, network-related issues could arise when generating routes. A total of 19,375 nodes is identified in the network. Due to applying topologically equivalent network reduction (as mentioned in Section 6.4.1), the number of nodes decreased to 7,628 nodes (-61%) with a total of 25,135 links.

The insertion of local knowledge regarding the network, to make sure that the majority of the illogical routes will not be generated using the choice set generation methods, underscores a major advantage of the DDPI method. This method relies only on the data that is collected from observed trips and thus does not require any network-information. Consequently, local knowledge is not required for using this method for analysing alternatives, model estimation, and model prediction. Furthermore, the DDPI method can be used as a reference set in adjusting the specification of currently adopted labelling approaches. Next to that, the algorithms use the information from the network or any other data source that is available, which is especially relevant for the labelling algorithm. As mentioned before, only three labels can be identified for this study, due to the limited data availability on the network.

6.6. Generated Choice Set Evaluation

The choice sets that are generated using the BFS-LE and labelling approach are compared to the observed routes according to the methodology described in section 6.4.1. The qualitative analysis for two selected OD pairs is covered in section 6.6.1. Section 6.6.2 details the quantitative analysis on the complete choice sets. Section 6.6.3 provides the results of the analysis on reproduction rate and behavioural consistency of the choice set algorithms. Finally, section 6.6.4 concludes the choice set evaluation.

6.6.1. Qualitative Analysis of the Choice Sets

The observed routes of the two selected OD pairs are plotted on the map in Figure 6.4. Cyclists in the first OD pair (upper OD) travel from the west of the inner-city of Amsterdam to the north side of the central train station and cyclists in the second OD pair (lower OD) travel from the centre (Waterlooplein) to the Vondelpark in the south-west of the inner-city.

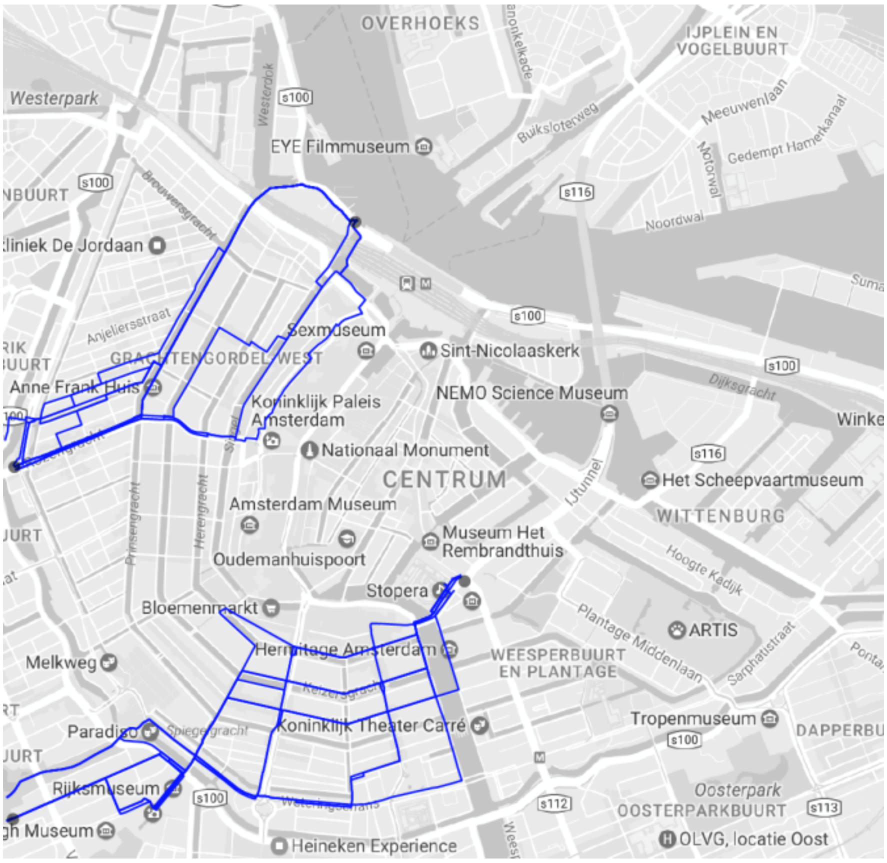


Figure 6.4: Observed routes from two selected OD pairs, plotted on the map of Amsterdam

The routes generated for the first OD pair using the BFS-LE and labelling approach are visualised in Figure 6.5, together with the observed routes. The observed routes (Figure 6.5.1) show a diverse set of routes. The north of the station can only be reached by one of the tunnels underneath the tracks, furthermore the cyclists face the canals that form a ring around the city centre, resulting here in roughly four main routes. The BFS-LE approach (Figure 6.5.2) provides a set of shortest routes, showing less diversity in this case. This approach only shows spatial diversity in the city centre. It avoids following the canals, which is different from the observed behaviour. This indicates that the cyclists are not necessarily aiming for the shortest route. The labelling approach (Figure 6.5.3) shows a more diverse choice set, that mimics the observed behaviour better. It does not provide exact matches, but provides routes that are more spatially different and makes use of the direction of the canals. This first comparison indicates that the labelling approach mimics the observed behaviour better in terms of spatiality and behaviour.

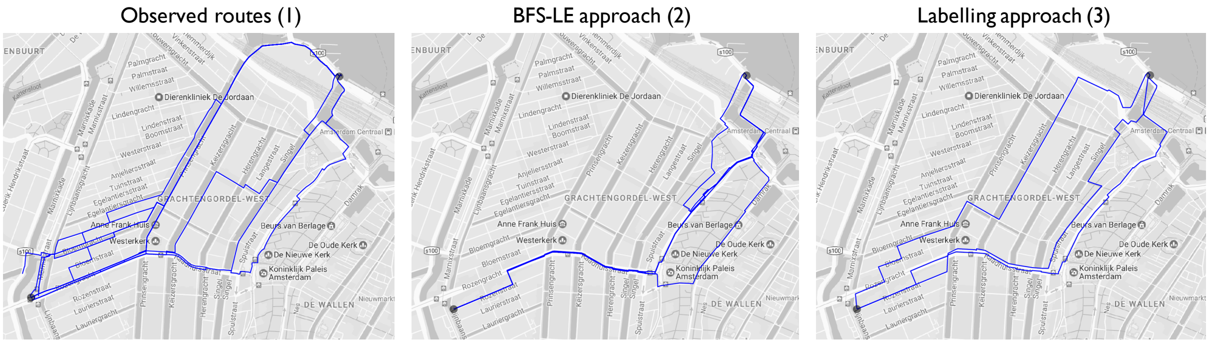


Figure 6.5: Routes generated for a given OD pair from the West of Amsterdam to the central train station, for (1) observed routes, (2) BFS-LE approach and (3) Labelling approach.

The generated choice sets for the second OD pair are visualised in Figure 6.6. The observed routes (Figure 6.6.1) again show a spatially diverse image. For most routes, the number of turns is minimised. The cyclists start northwards, then follow one of the ring roads and continue north, with different turning points. The BFS-LE approach (Figure 6.6.2) shows similar behaviour for the shortest route, however this route turns later than any of the observed routes. The northbound route that is generated is very different from the observed routes. Again, this approach generates a less spatially diverse choice set, that is unable to find all the observed routes. The labelling approach (Figure 6.6.3) is again more spatially diverse than the BFS-LE approach, but shows different routes than to the observed routes. Two of the three generated routes are comparable to the observed routes, in terms of turning. The third route turns often, which is very unlike the observed behaviour. The comparison of the second OD pair shows again that the labelling approach mimics the observed routes better than the BFS-LE approach, however the differences between the choice sets are still large. This qualitative analysis indicates that behaviour of cyclists is not captured based on one objective/label.

Figure 6.6: Routes generated for a given OD pair from Waterlooplein to Vondelpark, for (1) observed routes, (2) BFS-LE approach and (3) Labelling approach.

6.6.2. Quantitative analysis of the choice sets

In this section, the choice sets that are generated by the BFS-LE and labelling approach are compared to the observed routes based on a quantitative analysis. The descriptive statistics are calculated for distance, percentage on separate cycle path and the number intersections per kilometre. Furthermore, the path size factor (Eq. 6.1) is calculated, which is an indicator for heterogeneity of the choice set. Table 6.2 shows the results of the quantitative analysis.

The observed routes show that the mean distance travelled is 1.9 kilometres, whereas the entire area included in the research covers about 6 kilometres. This indicates that the average cyclist does not cross the entire inner-city. Furthermore, the percentage of separate cycle paths encountered on the routes and the amount of intersections per kilometre (all types of intersections) are rather low, the latter was expected from the qualitative analysis. Finally, the path size factor is on average 0.67, which indicates a relatively heterogeneous set of routes, matching the results from the qualitative analysis. The routes chosen by all cyclists are spatially diverse and have a low degree of overlap.

Table 6.2: Descriptive statistics of the explanatory variables and heterogeneity indicator for each choice set identification approach

	Observed routes (N=2,819)			BFS-LE approach (N=12,361)			Labelling approach (N=2,034)		
Variable	*Mean*	*Median*	*St.Dev*	*Mean*	*Median*	*St.Dev*	*Mean*	*Median*	*St.Dev*
Distance (km)	1.93	1.85	1.01	1.92	1.85	0.78	2.82	2.47	1.66
Separate cycle path %	37.9%	34.7%	26.4%	8.3%	6.6%	8.2%	19.4%	9.7%	22.8%
Intersections per km	14.8	14.5	5.0	32.2	32.1	6.9	19.9	15.9	10.5
Path Size factor	0.671	0.704	0.232	0.135	0.090	0.126	0.833	0.864	0.136

The BFS-LE approach optimises for distance, which is reflected in the lower mean distance and standard deviation. However, the difference with respect to observed routes is negligible, which seems to imply that the cyclists prefer shorter routes. As mentioned before, several of the links, found in observed routes, are not included in the network. Inspections of the OD pairs crossing the city centre, showed that 25% of the trips cross these areas even though the network does not include these, indicating that the true shortest path cannot be found by the algorithms. It shows that the true mean distance might be lower than shown in Table 6.2, indicating that the preference for the shorter routes might be less straightforward than appears now. The BFS-LE approach also shows a low percentage of separate cycle paths and a high amount of intersections per kilometre compared to the observed routes. Most likely because the algorithm does not optimise for these variables. Due to the nature of the algorithm, it finds a low variety of routes, leading to a relatively homogeneous set of routes, reflected in the qualitative analysis.

The labelling approach generates a route that optimises for each variable in the descriptive statistics, therefore the standard deviations are large. The mean distance is larger than both other choice sets, whereas the percentage of separate cycle path and number of intersections per kilometre are in between the observed routes and BFS-LE algorithm. Furthermore, due to the optimisation on different variables, the choice set is very heterogeneous and spatially diverse (as was also found in the qualitative analysis).

6.6.3. Reproduction of Observed Routes

This section covers the reproduction rate and behavioural consistency of both the BFS-LE and labelling approach. The reproduction rate is calculated for different levels of overlap between generated and observed routes, varying from 70% to 100%. Table 6.3 shows the results of these analyses.

Table 6.3: Number and percentage of observed routes generated by each choice set generation approach for different threshold levels

	100% Overlap		90% Overlap		80% Overlap		70% Overlap		
Algorithm	*# trips*	*% trips*	*# trips*	*% trips*	*# trips*	*% trips*	*# trips*	*% trips*	**CI**
BFS-LE approach	26	0.9%	53	1.9%	92	3.3%	175	6.2%	0.2701
Labelling approach	38	1.4%	65	2.3%	110	3.9%	183	6.5%	0.3024

Note: the total number of trips is 2,819.

The false negative error for both methods is about 99%, implying that the overwhelming majority of observed routes are not included in the generated choice-sets. The labelling approach is slightly better at reproducing the observed trips and has a higher behavioural consistency compared to the BFS-LE approach. The qualitative analysis showed that the labelling approach could partially reproduce the observed routes, however the overlap between the observed and generated routes is lower than 70%. The BFS-LE approach performs even worse, as was also visible in the qualitative analysis. As mentioned before, network-related issues could impact the choice set generation. This dependency of choice set algorithms on the network shows one advantage of the DDPI method, as this method does not rely on network information.

6.6.4. Conclusions regarding the Evaluated Choice Sets

The choice sets resulting from the BFS-LE and labelling approach differ largely from one another, and they differ largely from the observed routes. The labelling approach is better than the BFS-LE approach in terms of mimicking the observed routes, but shows very large false negative errors (not generating the observed alternative). The quality of the network representation (topology and available label information) that serves as input for the choice set generation methods, which is poor in the bicycle-context, influences the routes that are generated, especially when generating routes based on individual network characteristics. In this case, the observed behaviour is not captured by these characteristics. The differences indicate that cyclists optimise based on more than one network-related objective. Ehrgott et al. (2012) proposed a method for bi-objective optimisation, as they found that cyclists do not optimise based on one objective, like car drivers might do with distance or travel time. Two other methods that might be able to overcome this issue are the link-based approach introduced by Fosgerau et al. (2013) and importance sampling approaches like the Metropolis-Hastings approach (Flötteröd and Bierlaire, 2013), as they approach the choice set generation from the universal choice set.

6.7. Evaluation of Model Estimation and Validation

This section covers the evaluation of the model estimation (6.7.1) and validation (6.7.2). Three route choice models are estimated using the choice sets resulting from the labelling approach, BFS-LE approach and DDPI approach (as shown in Figure 6.1). The evaluation takes place according to the methodology proposed in section 6.4.2. Section 6.7.3 concludes this evaluation section.

6.7.1. Route Choice Model Estimation

When observed routes are not generated using the BFS-LE and labelling approaches, there are no good remedies. Eliminating the entire OD pair would leave very few pairs remaining (approximately 1% of the trips). In practice, the observed routes that have not been generated are added to the choice set (e.g. Broach et al., 2010). Consequently, a union of routes is created based on network characteristics and observed behaviour (like depicted in Figure 6.1). However, this method entails that information/observed behaviour is added to the choice set, which will increase the performance of these choice sets in model estimation and consequently introduces an issue with endogeneity (by including chosen alternatives). The comparison in the model estimation is therefore skewed, due to this poor performance in terms of reproducing observed alternatives.

Five models are estimated for each choice set, every time using a different random sample of 80% of the OD pairs, to investigate the stability of the models. Table 6.4 shows the estimation results for one of the model runs.

Table 6.4: Estimated PSL models using the identified choice sets from data, BFS-LE and labelling

	DDPI model		BFS-LE model		Labelling model	
Variables	**Coef.**	**t-test**	**Coef.**	**t-test**	**Coef.**	**t-test**
Distance (km)	-0.225	-1.72*	-0.341	-2.84**	-1.840	-21.88**
% separate cycle path	0.153	1.00	1.34	9.47**	1.53	11.45**
Intersections/km	-0.018	-2.11**	-0.159	-23.90**	-0.118	-21.61**
(Ln) Path Size	-0.380	-3.94**	1.03	17.11**	0.291	3.77**
N	2,249		2,249		2,249	
Null log likelihood	-3,059.718		-6,921.409		-4,419.422	
Final log likelihood	-3,044.254		-3,539.881		-3,627.528	
Likelihood ratio test	30.928		6,763.057		1,538.788	
Adj. rho square	0.004		0.488		0.178	

** significant at the 5% level, * significant at the 10% level

The signs of distance, separate cycle path percentage and intersections per kilometre are as expected and are the same for each model. However, the parameter and t-test values are different. The DDPI model has lower t-test values compared to the other models, which is due to the endogeneity issue that plays a role in the DDPI choice set. It has the tendency to make attributes less significant. Furthermore, the sign of the path size factor is different for the DDPI model. In this case a route that has more overlap with other routes receives a higher utility. In the context of public transport, Lam and Xie (2002) also found a negative parameter. They argue that overlapping routes can reduce uncertainty by allowing more en-route rerouting possibilities and hence contribute to the robustness of the route taken, which could also hold for the bicycle route choice situation. In case of the BFS-LE and labelling model, adding the observed routes results with a positive PS factor. The generated alternatives overlap with each other, but often the observed alternatives are very different, resulting in a higher utility for the non-overlapping routes. Consequently, the interpretation of the negative PS sign is different from the positive PS sign, showing a difference between observed and generated choice sets.

To compare these models, the average point elasticities for all explanatory variables are calculated (Table 6.5). The elasticity provides information on the impact of marginal changes in each of these variables on the probability of being chosen.

Table 6.5: Mean point elasticities for each explanatory variable for all models

Variable	**DDPI model Elasticity**	**BFS-LE model Elasticity**	**Labelling model Elasticity**
Distance	-0.289	-0.440	-2.577
% separate cycle path	0.042	0.350	0.426
Intersections/km	-0.188	-1.702	-1.316

The interpretation of the elasticities is such that 1% increase in distance results in a decrease in the probability of being chosen of 0.29% for the DDPI model, whereas the BFS-LE model shows a 0.44% decrease and the labelling model shows a decrease of 2.58%. The relative difference between the impact of the BFS-LE model and DDPI model is 52%, but is around 790% with the labelling model. In the labelling model, the impact of marginal changes to all variables, is much higher compared to the other models. The routes generated by the labelling algorithm are very diverse and optimised for different criteria, which indicates that increasing the variability in attributes of the alternatives (labelling routes plus observed route), induces a higher elasticity. This could be confounding the effect of the different parameters on the elasticities with the effect of different attributes.

6.7.2. Route Choice Model Validation

The model validation provides insight into the predictive power of the models. The 20% remaining OD pairs are used to validate the models. For the validation, the alternatives of all three choice sets are combined for each OD pair to make the comparison fair (resulting in a maximum of 41 alternatives for 695 OD pairs, which is the same input for all models). For five random draws the models are estimated and validated. Table 6.6 shows the results of the validation.

Table 6.6: Average validation measures for all 5 estimated models per choice set

	Correct choice predicted	RMSE OD pair	Log-likelihood
DDPI model	1.3%	0.6264	-2,057.083
BFS-LE model	21.1%	0.5677	-1,206.231
Labelling model	27.8%	0.4728	-1,188.331

The DDPI model has lower parameter values compared to the other models. This means for the validation that it does not punish the less attractive alternatives as much as the other models. Consequently, the maximum utility for one alternative is low and similar for all alternatives. This results in a very low percentage of correctly predicted choices. The BFS-LE and labelling models score higher on this validation measure, and are on average able to predict at least one choice correct per OD pair. In terms of prediction per alternative, the two models that were estimated on a generated choice set that has a higher variability and includes both good (observed) routes and bad (generated) routes, perform better.

In terms of the RMSE that is weighted over the OD pairs, the models perform similar (although the BFS-LE and labelling model outperform the DDPI model). This measure gives an indication on the average error that would occur when for example predicting the flows on the network. The DDPI model assigns a rather equal probability to all alternatives, resulting in an average error that is similar to the RMSE of the two other models. These models on the other hand, provide a low probability to the worse (generated) alternatives and a very high probability to the good (observed) alternatives.

The null log-likelihood for this set of alternatives (calculated using $LL(0) = -\sum_n \ln(J_n)$, with J_n being the number of alternatives in choice set C_n) is -1,740.149. The closer the final log-likelihood is to zero, the better the out-of-sample performance is. Both BFS-LE and labelling models improve significantly compared to the null log-likelihood. The DDPI models, which are estimated using only observed information, perform worse on the out-of-sample data in terms of its added value compared to providing equal probabilities to all alternatives (null log-likelihood). Consequently, we can conclude that the DDPI method should not be used for prediction purposes.

6.7.3. Conclusions regarding Model Estimation and Validation

Due to the small number of matches of generated routes with observed routes, the choice sets are enriched with observed routes. Consequently, the choice sets have more information compared to purely generated choice sets, introducing endogeneity. The models that are estimated using the different choice sets differ in their parameter values, t-test values and elasticities. This is in line with expectations as the size and composition of choice set are known to influence the model estimation (Bovy, 2009).

The DDPI model has lower parameter values and t-test values due to small variability in the choice set and issues with endogeneity. Due to the inclusion of the observed alternatives in the BFS-LE and labelling choice set, where they were not generated, these models perform very well as an artefact. The large variability between alternatives (especially in the labelling choice set) and inclusion of both relevant and irrelevant alternatives (especially in the BFS-LE choice set), increases the model fit compared to only using observed routes (DDPI method). The effect of explanatory variables on route choice is higher for the labelling model compared to the other models. The BFS-LE model is a less extreme version of the labelling model, with relatively high parameter and t-test values but elasticities that are more similar to those obtained using the DDPI approach. The reason for this might be the number of alternatives that is included in the BFS-LE approach, which is generally 17 more than the labelling approach.

In terms of predictive powers, the DDPI model was expected to perform less as it is data-driven and might therefore react different to out-of-sample prediction than the labelling and BFS-LE models, which was confirmed by all validation measures. The DDPI method is not suitable for out-of-sample prediction.

6.8. Conclusions and Future Research Directions

This paper presents the findings of an evaluation of a data-driven approach (DDPI) for choice set identification in travel behaviour analysis, performed by comparing the DDPI method to two choice set generation methods: BFS-LE method introduced by Rieser-Schussler et al. (2013) and the labelling approach introduced by Ben-Akiva et al. (1984). Bicycle GPS data from the city of Amsterdam was used a case study. The comparison was based on three aspects. First, an analysis of the choice sets that are identified, which was evaluated by means of a qualitative (visual) analysis, a quantitative analysis, and the reproduction of observed routes. Second, estimation of a route choice models using the three identified choice sets, which were evaluated by means of calculating elasticities. And third, validation these models on out-of-sample data, which were evaluated by means of correctly predicted choices, RMSE per OD pair and the log-likelihood.

In conclusion, the data-driven DDPI method is useful when evaluating or analysing the alternatives in the choice set and can help in understanding the preferences of individuals (using model estimation). The DDPI is not suitable for prediction on out-of-sample data.

The ability of choice set generation algorithms to reproduce observed paths largely depends on the correctness of the underlying network. In this study, the network was intended for motorised traffic (i.e. not validated for bicycle traffic), resulting in choice sets that are not suitable for analysis of the alternatives for cyclists (e.g. in terms of composition and characteristics). Generally, cyclists are allowed to cycle against one-way streets, however this is not included in the network. Furthermore, cyclists do not necessarily comply to the traffic rules in the Netherlands, as exhibited in using links in the network that are not identified for cyclists (e.g. short cuts or pedestrian areas). The first can be incorporated in the network by making all links bi-directional; however, the latter is harder to incorporate. Consequently, a discrepancy arose between the observed routes and the generated routes. The number of generated routes that could be matched to observed routes was very low, partially due to network incompleteness. However, we tested the significance of this shortcoming by removing the affected OD pairs, and found that the number of matched routes was still very low, indicating that generating routes based on single network characteristics (as is done in these algorithms) does not match with the observed behaviour. In conclusion, the choice set based on observed behaviour provides a better source for analysing the alternatives than a generated choice set based on network characteristics.

Given the differences and similarities between the estimated choice models, we conclude that the DDPI method provides useful insights into behaviour. In terms of model fit, it performed worse than the generated choice sets, mostly due to lower variability between routes and their respective attributes. However, no additional network information is required for the DDPI method. Hence, it does not rely on the quality of the underlying network for information or routes that need to be generated. Mostly because of that reason this method is a valuable addition to the existing choice set generation methods, as it does provide insights into preferences of individuals regarding attributes.

The case study analysed in this paper gives first insights into the usefulness of the data-driven DDPI approach for travel behaviour analysis. In this study the data-driven choice set has been applied to bicycle route choice. Future research can test the usefulness of the proposed DDPI method for other types of choice set generation, for example activity scheduling and destination choice, and for route choice models of other modes, for example the car, which potentially exhibits a larger degree of diversity of routes within a shorter time period, due to congestion and traffic lights. Next to that, the model is now estimated on data from one week. It would be very useful to test on a dataset that covers a longer period of time (e.g. a month), because this potentially increases observed variability and thereof reduces the risk of endogeneity. Furthermore, the performance of choice set generation methods depends on the quality of the underlying network. Future studies may match the observed routes and links to the existing network prior to the choice set generation so that missing links can be added to the network. This will potentially result in a higher reproduction of observed routes. Also, this will provide more routes per unique OD pair, therefore reducing the need for clustering. Finally, the methods in this study were tested using random utility theory (specifically the PSL models). A direction for future research could be to apply the method within the random regret framework and test its performance.

Acknowledgements

This research was supported by the Allegro project (no. 669792) which is financed by the European Research Council and Amsterdam Institute for Advanced Metropolitan Solutions. The data from the field experiment Bicycle Counting Week, used for this research was provided by the municipality of Amsterdam.

Chapter 7 – A Review of Modelling Approaches for Simultaneously Modelling Multiple Travel Choices

This chapter is based on the following submitted manuscript:

Ton D., Duives D.C., Cats O., and Hoogendoorn S.P. (submitted). Simultaneous modelling of multiple travel choice dimensions: Assessment of the suitability and applicability of different discrete choice modelling structures.

Abstract

Accurate modelling of simultaneous travel choice dimensions is becoming more important due to the increasing behavioural complexity that is identified in literature regarding the relationships between choice dimensions. Currently, mostly Multinomial Logit and Nested Logit are applied in activity-based and four-step models. These model structures entail some strong assumptions and are found to be largely unrealistic in capturing the relationships between choice dimensions. Consequently, more advanced models are required to capture behavioural realism. In this literature review, we assess the discrete choice modelling structures that have been proposed in literature for simultaneously modelling a combination of trip chain, destination, departure time, mode, and route choices. The assessment relates to the suitability, in terms of flexibility in correlation of model structures, heterogeneity in the decision-making process, and model specification of each choice dimension, and applicability, in terms of computational effort and ease of interpretation, of the model structures. The findings of the

review show that the 'ideal' model structure, which is deemed both behaviourally suitable and practically applicably, does not exist yet. The model structures that are most suitable are less applicable and vice versa. Consequently, directions of future research are increasing suitability of model structures, without ignoring the applicability. A compromise of these two aspects might lead to the adoption of more behaviourally realistic model structures in the activity-based and four-step models.

7.1. Introduction

Travel behaviour encompasses various travel choices made by individuals, such as departure time choice and mode choice. When individual choice outcomes are aggregated, they result in travellers' distribution over modes, time, and the network. Consequently, they are important inputs for demand forecasting and the development of efficient policies. It is widely acknowledged that the travel choice dimensions are inter-related (e.g. Bhat, 1998b; Richards and Ben-Akiva, 1974). According to Bhat (1998a) there are three main reasons for simultaneously modelling these choices. First, the considered choice alternatives are a combination of multiple choice dimensions. Second, observed determinants that influence choice behaviour are related to multiple choice dimensions. And third, the joint choice alternatives share unobserved determinants that influence travellers' sensitivity to changes related to policy measures.

The four-step model and the activity-based model are the best known models that address the modelling of multiple travel choice dimensions (McNally and Rindt, 2007). These models generally cover (a subset of) the following choices: trip chaining, destination, departure time, mode, and route choice. Various modelling approaches can be applied in the four-step and activity-based models, but often discrete choice models are used (Ben-Akiva and Bowman, 1998; de Dios Ortúzar and Willumsen, 2011). Examples of discrete choice modelling in activity-based models are the Portland model (Bowman et al., 1998), the Jakarta model (Yagi and Mohammadian, 2010), and the Tel-Aviv model (Shiftan and Ben-Akiva, 2011). Four-step models using discrete choice modelling are, for example, the Swedish national model (Beser and Algers, 2002) and the Dutch national model (van Cranenburgh and Chorus, 2017;Hofman, 2002).

Discrete choice models can provide behavioural insights into the choice dimension(s) and can be used for forecasting purposes. The most common theoretical framework underlying discrete choice models is random utility theory, which assumes that individuals maximise their utility in the decision-making process (Ben-Akiva and Bierlaire, 1999). Within this framework, a broad range of discrete choice models has been developed. Two of these models, i.e. the Multinomial Logit (MNL) and the Nested Logit (NL) models, are most frequently applied in the activity-based and four-step models. These are relatively simple models which involve some strong assumptions. For example, the MNL model assumes that alternatives are independent of one another, which might not be realistic in the context of multiple choice dimensions (Ben-Akiva and Lerman, 1985). Generally, MNL is used to simultaneously model choice dimensions, for example destination and mode, which then entails that one mode and destination alternative, for example car - supermarket "A", is independent from another, for example car - supermarket "B". The NL model relaxes the independence assumption, by accommodating correlation within part of the alternatives (Williams, 1977), which can be used to calculate substitution patterns, that are often, but need not be, interpreted as implying hierarchy (de Dios Ortúzar and Willumsen, 2011). For example, when the situation changes and an individual changes the route before changing the mode for a trip. However, it is plausible that the substitution pattern holds the other way around (change mode before route), or that it is better to incorporate cross-correlations to reflect a bi-directional relation (Hess et al., 2007b). Especially the latter is not employed in the current four-step and activity-based models. Therefore, these findings suggest that the assumptions in the MNL and NL models prevent to include relationships between choice dimensions realistically.

Furthermore, increasing evidence is found in literature that choice behaviour is heterogenous (e.g. Hensher and Reyes, 2000; Krygsman et al., 2007). Heterogeneity can be found within choice dimensions, where individuals show different preferences. Next to that, it can be found between choice dimensions, where individuals show different decision-making processes. This increasing identified complexity of choice behaviour cannot be accurately captured using the discrete choice models currently adopted in the four-step and activity-based models. Together with the conclusion that currently, relationships between choice dimensions fail to reflect realistic behaviour, this finding leads to the realisation that more advanced discrete choice models are needed in the simultaneous modelling of multiple travel choice dimensions. Many advancements in discrete choice models have been made since the introduction of MNL and NL (e.g. Mixed Logit (McFadden and Train, 2000)). These advancements are aimed at increasing the level of realism of the choice behaviour modelled. In the context of simultaneously modelling multiple travel choice dimensions, this translates to for example, increasing the flexibility of the correlation structure or by allowing a different model specification per choice dimension. Consequently, these more advanced model structures might be more suitable for modelling multiple travel choice dimensions.

On the other hand, MNL and NL are relatively easy to use and require limited computational effort in estimation and forecasting. Especially in practice, these are important determinants of model applicability. Hess et al. (2007a) state that the estimation process of large-scale Mixed Logit models can last 100 days. When the goal is to estimate the effect of different policy measures, it is generally preferred to get an answer within a short period of time. Consequently, advancements in discrete choice models need to be measured against their applicability, if they are to be used and applied in practice.

The objective of this literature review is to assess the suitability and applicability of discrete choice modelling structures for the simultaneous modelling of multiple travel choice dimensions. By assessing the suitability of model structures that are currently in the context of multiple travel choice dimensions, the model structures are judged on their behavioural realism in capturing multiple travel choice dimensions. This helps in understanding what are possibilities and challenges. By also assessing the applicability, it is investigated what is required before adoption of model structures in practice. Suitability is identified via the following key aspects: flexibility of the correlation structure, correlation between choice dimensions, inclusion of heterogeneity in decision-making processes, and model specification of different choice dimensions. Applicability refers to the ease of interpretation and computational efficiency.

The contribution of this review is therefore three-fold: 1) establish which discrete choice modelling structures have been used for modelling multiple travel choices (besides the activity-based and four-step models), 2) provide support regarding which model structure(s) can be used in estimation and forecasting, given suitability and applicability considerations, 3) identify directions for future development in model structures, given the findings of this review.

The remainder of the paper is organised as follows. Section 7.2 details the research scope and describes the literature search methodology. Section 7.3 discusses all modelling approaches and explains how they have been applied. Section 7.4 discusses which choice dimensions have been modelled simultaneously and which model structures have been applied. Then, section 7.5 describes the model assessment indicators, after which section 7.6 provides a discussion on the assessment of the modelling structures. Finally, section 7.7 concludes the review and identifies future research directions.

7.2. Scope and Methodology

In this literature review study, discrete choice modelling structures that are based on the random utility framework and have been used for simultaneously modelling multiple travel choice dimensions are reviewed. We do not intend to provide an extensive review on random utility theory, but rather provide a brief context to support the discussions of different model structures. The interested reader is referred to Ben-Akiva and Lerman (1985) or Ben-Akiva and Bierlaire (1999) for a thorough discussion of random utility theory. Random utility theory acknowledges that the researcher is not aware of the entire decision-making process of individuals. To account for this uncertainty, utility is modelled as a random variable U_{in} (Manski, 1977). The utility that an individual n associates with alternative i is expressed as follows:

$$U_{in} = V_{in} + \varepsilon_{in} \quad \forall\, i \in C_n \tag{7.1}$$

where V_{in} is the deterministic part of the utility representing observed attributes and ε_{in} represents the random error term. The probability that alternative i is chosen by individual n from choice set C_n is (Ben-Akiva and Bierlaire, 1999):

$$\mathrm{P}(i|C_n) = \mathrm{P}\left[U_{in} \geq U_{jn} \;\; \forall j \in C_n\right] = P\left[U_{in} = \max_{j \in C_n} U_{jn}\right] \tag{7.2}$$

In this review, the focus is on studies addressing the simultaneous modelling of multiple travel choice dimensions. We specifically target studies that model two or more travel choice dimensions, using the above mentioned framework. By targeting two or more choice dimensions, also partial activity-based or four-step models are included. Furthermore, the focus lies on random utility theory as this is the most widely known and applied method in practice and research. We consider disaggregate models that represent individuals as decision-makers in the models. The focus lies with the following travel choice dimensions: trip chain, destination, departure time, mode, and route choices. These choice dimensions are commonly modelled in the activity-based and four-step models.

This review examines studies that are published in peer-reviewed articles. The search engines Google Scholar and Scopus are used to identify the studies that fall within the research scope. All possible combinations of the five identified travel choices were used in combination with the term ‘choice model’, for example ‘route mode choice model’. These search combinations resulted with 24 studies. The snowballing method was applied to identify other relevant studies. In total, 31 different studies (32 applications) have been identified which fulfil our selection criteria.

The most common combinations of choice dimensions, which are simultaneously modelled using discrete choice models, are departure time and mode choice (15), trip chain and mode choice (7), destination and mode choice (6), and mode and route choice (4). In contrast, no studies investigating a combination of the following choice dimensions were found: trip chain and destination choice, trip chain and departure time choice, and trip chain and route choice. These combinations might, however, have been investigated using out-of-scope modelling approaches.

7.3. Model Structures

This section introduces the discrete choice modelling structures and discusses how they are adapted to model multiple travel choices simultaneously. In this study the model structures are reviewed in this context, therefore we provide only a small introduction to the concept of each model structure. For reviews on model structures the reader is referred to e.g. Ben-Akiva and Lerman (1985), Ben-Akiva and Bierlaire (1999), and de Dios Ortúzar and Willumsen (2011).

The literature search revealed 31 studies (and 32 applications of model structures) within the defined scope, which can be divided into *eight model structure categories*, based on similarities and differences between the applied approaches. The categories are Multinomial Logit (7.3.1), Nested Logit (7.3.2), Cross-Nested Logit (7.3.3), Probit (7.3.4), Mixed Logit (7.3.5), discrete-continuous models (7.3.6), segmentation approaches (7.3.7), and miscellaneous approaches (7.3.8). The last category captures the studies that could not be assigned to any other category. Appendix 7.A provides an overview of the studies. To better understand the relationships between all the model structures and where they originate from, Figure 7.1 describes the genealogy, which is further elaborated upon in each sub-section.

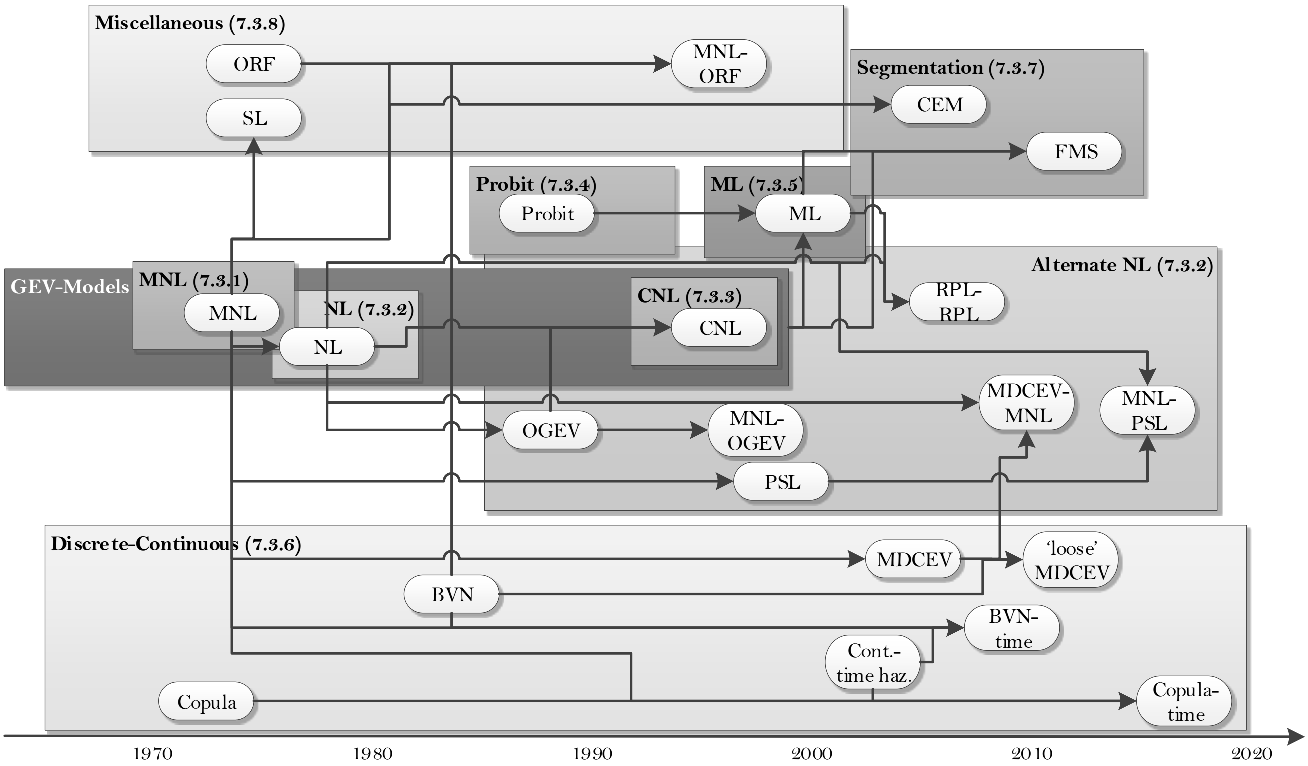


Figure 7.1: The genealogy of model structures, matched to the year of the first mathematical introduction

7.3.1. Multinomial Logit (MNL)

Four studies are identified that applied the MNL structure to model multiple travel choices (Broach and Dill, 2016; Montini et al., 2017; Richards and Ben-Akiva, 1974; Vrtic et al., 2007). The MNL model is the first model in the Generalised Extreme Value (GEV)-family of models. MNL assumes that the error terms ε_{in} introduced in Eq. 7.1 are identically and independently (*iid*) Gumbel distributed. The model introduced by McFadden (1974) is expressed as follows:

$$P(i|C) = \frac{e^{V_i}}{\sum_{j \in C} e^{V_j}} \tag{7.3}$$

where the utility of alternative i is set against the utility of all other alternatives j belonging to choice set C, resulting in the probability P of choosing i. An important property of MNL is independence from irrelevant alternatives (IIA), which means that error terms are not correlated (Ben-Akiva and Lerman, 1985). Due to the simple closed-form mathematical expression, computation effort is low.

Two different MNL approaches have been identified for simultaneously modelling multiple travel choices. The first approach combines all identified alternatives for each of the choice dimensions into joint alternatives ($C_{12} = C_1 * C_2$). Hence, introducing a simultaneous choice for the two decisions, also called Joint Logit (Ben-Akiva and Lerman, 1985). This results in a similar modelling approach for different choices. Richards and Ben-Akiva (1974) have applied this method to destination and mode choice. Furthermore, Montini et al. (2017) modelled route and mode choice.

The second approach only identifies alternatives for one choice and represents the other choice as attributes in V_i, which implies a hierarchy in the choices (e.g. $V_{C_1} = \sum \beta_k x_k + \beta_{C_2} C_2$). This results in individual treatment of different choices, and requires one extra parameter in the utility function. One example in the case of mode and route choice is the inclusion of the attributes of the predicted (or shortest) routes in the mode choice model (Broach and Dill, 2016). Consequently, the route choice is known beforehand. Another example, related to mode, route, and departure time choice is presented by Vrtic et al. (2007). They have identified two mode alternatives, and included for example late and early departure as parameters for the departure time choice.

7.3.2. Nested Logit (NL)

NL is one of the most widely known extensions of the MNL model (Williams, 1977). This model structure is applied most for modelling multiple travel choices, i.e. in eleven studies (Bajwa et al., 2008; Bhat, 1998b; Debrezion et al., 2009; Eluru et al., 2010; González et al., 2016; Hess et al., 2007a; Lizana et al., 2013; Newman and Bernardin, 2010; Paleti et al., 2014; Shakeel et al., 2016; Yang et al., 2016). NL allows for a bundle of alternatives (a nest) to have correlated error terms (e.g. red bus/blue bus paradox), therefore relaxing the IIA property of the MNL model by allowing partial correlation. The model is derived as follows:

$$P(i|C) = P(i|m)P(m|C) = \frac{e^{\mu_m V_i}}{\sum_{j \in C_m} e^{\mu_m V_j}} \frac{e^{\frac{\mu}{\mu_m} \ln \sum_{l \in C_m} e^{\mu_m V_l}}}{\sum_{p=1}^{M} e^{\frac{\mu}{\mu_p} \ln \sum_{l \in C_p} e^{\mu_p V_{lp}}}} \tag{7.4}$$

where $P(i|m)$ is specified according to Eq. 7.3 (there $\mu_m = 1$), and $P(m|C)$ represents the partial correlation within nests m in the choice set. The scaling parameters μ and μ_m need to be estimated, where generally μ is normalised to 1 and μ_m is estimated for each nest. Because correlation is non-negative, in this case $\mu_m \geq 1$. Due to the closed-form mathematical formulation, computational effort is low.

With respect to modelling multiple travel choices, the nests in NL often represent one choice, with the alternatives being a combination of the two choices. For example, in case of mode and route choice, the modes are represented in nests and the mode-route alternatives are assigned to these nests (Shakeel et al., 2016). Therefore, allowing for correlation between the routes of the same mode. Consequently, the choice for a route is dependent on the choice for a mode and one can argue that switching routes is easier than switching modes. Debrezion et al. (2009) used this approach for mode and destination choice, where they represented mode choice in nests. Three studies applied NL on the mode and departure time choice (Hess et al., 2007a;

Lizana et al., 2013; Paleti et al., 2014). They all represented mode choice in nests. Finally, a single application to mode choice and trip chaining was found, where different nesting structures were identified for holidays (mode nests) and normal weekdays (trip chain nests) (Yang et al., 2016).

Five studies have advanced the NL approach, by applying different underlying model structures per choice (alternate NL). First, Bhat (1998b) applied a MNL-OGEV (Ordered GEV) model to mode and departure time choice. In this study, mode choice is represented in nests and the time periods are ordered realistically (Small, 1987). Second, Newman and Bernardin (2010) applied a NL-MNL nesting structure for mode and destination choice, with mode choice nests. Within the mode choice, another nest was introduced. Third, Bajwa et al. (2008) applied a RPL-RPL (Random Parameter Logit, Section 7.3.5) model for mode and departure time choice, where modes are modelled as nests. RPL requires simulation, therefore computation effort is high. In this study, the modelling of multiple travel choices is determined via the NL model structure, therefore it is allocated here. Fourth, Eluru et al. (2010) applied a MDCEV (Multiple Discrete-Continuous Extreme Value) – MNL model to mode, departure time and destination choice. Mode and departure time (nests) are modelled in the MDCEV framework (Bhat, 2005) where they are distributed into episodes that can be assigned to multiple activity times during the day. Finally, a MNL-PSL (Path-Size Logit) model was specified for destination and route choice (González et al., 2016). The destinations are modelled in nests and a path size factor is identified in the utility for the route choice, that corrects for overlap between routes (Ben-Akiva and Bierlaire, 1999).

7.3.3. Cross-Nested Logit (CNL)

Two applications of CNL are identified for simultaneously modelling multiple travel choices (Ding et al., 2014; Yang et al., 2013). The CNL is a direct extension of the NL model, which allows each alternative to be a member of different nests. Therefore, it does not only allow for correlation within a nest but also between nests, resulting full correlation between the alternatives of travel choices (Vovsha, 1997). The model is expressed as follows:

$$P(i|C) = \sum_m P(i|m)P(m|C)$$

$$= \sum_{m=1}^{M} \frac{\alpha_{im}^{\frac{\mu_m}{\mu}} e^{\mu_m V_i}}{\sum_{j\in C_m} \alpha_{jm}^{\frac{\mu_m}{\mu}} e^{\mu_m V_j}} \frac{\left(\sum_{j\in C_m} \alpha_{jm}^{\frac{\mu_m}{\mu}} e^{\mu_m V_j}\right)^{\frac{\mu}{\mu_m}}}{\sum_{n=1}^{M}\left(\sum_{l\in C_n} \alpha_{jn}^{\frac{\mu_n}{\mu}} e^{\mu_n V_j}\right)^{\frac{\mu}{\mu_n}}} \tag{7.5}$$

where it builds upon the NL derivation of Eq. 7.4 by introducing a membership parameter α into the model, which needs to sum up to 1 per alternative. Furthermore, μ is normally constrained to one, and $\mu_m \geq 1$. The non-concave objective function results in medium computation effort. Often, α is predefined in CNL (Bierlaire, 2006).

The first application to multiple travel choices relates to destination and mode choice, where the nests are defined for both destination and mode choices (Ding et al., 2014). The degree of membership is fixed: ½ to a destination nest and ½ to a mode nest. The second application focuses on residential location, mode and departure time choice (Yang et al., 2013). They also modelled each alternative as a nest and also predefined the membership.

7.3.4. Probit

Probit models are different from Logit models in the assumed distribution of the error terms: normal distributions instead of *iid* Gumbel distributions. Two applications of the Probit structure are found for modelling multiple travel choices (Jou, 2001; Ye et al., 2007). Probit is more flexible than Logit (McFadden, 1989). The utility function in Eq. 7.1 is adapted, to include vector notation for all components of the utility function:

$$\boldsymbol{U}_{in} = \boldsymbol{V}_{in} + \boldsymbol{\varepsilon}_{in} \quad \forall\, i \in C_n \tag{7.6}$$

where $\boldsymbol{U}_{in}, \boldsymbol{V}_{in}$ and $\boldsymbol{\varepsilon}_{in}$ are $(J_n \times 1)$ vectors, with J_n being the number of alternatives in C_n. The error term for Probit models is normally distributed: $\boldsymbol{\varepsilon}_{in} \sim N(0, \Sigma)$, and for $J_n = 2, \Sigma$ is given in Eq. 7.7:

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix} \tag{7.7}$$

For Probit models the variance-covariance matrix needs to be estimated. Because of identifiability restrictions, the binary Probit example in Eq. 7.7 results with no freely-estimable parameters. The difference in the error terms between the two alternatives is what matters (difference must also be fixed) and individual variance and covariance parameters cannot be uniquely specified. The covariance represents the correlation between alternatives. Probit does not have a closed-form mathematical formulation, requiring simulation or numerical integration, which results with high computational effort.

In the simultaneously modelling of multiple travel choices two different applications of Probit were found. First, a study by Jou (2001) applies this model for departure time and route choice. He introduces four alternatives, resulting in the estimation of 16 variance-covariance parameters. By estimating the complete variance-covariance matrix (given identifiability restrictions), he allows for full correlation between alternatives of travel choices. Additionally, he includes a parameter that accounts for unobserved heterogeneity: $\upsilon \sim N(0, \sigma)$, based on procedures described in Abdel-Aty et al. (1997).

Ye et al. (2007) have applied a recursive bivariate Probit model to mode choice and trip chain complexity. The recursive model adds an additional element per choice dimension to the utility function in Eq. 6, that represents the second choice C_2 (see Eq. 7.8).

$$\boldsymbol{U}_{\boldsymbol{c1,in}} = \boldsymbol{V}_{c1,in} + \eta C_2 + \boldsymbol{\varepsilon}_{in} \quad \forall\, i \in C_n \tag{7.8}$$

The parameter that is estimated for this dummy represents the impact of the second choice on the first choice. By allowing this for one choice only, basically stating that C_2 is predetermined, a dependency of choices is implied (partial correlation), which could be interpreted as hierarchical decision-making (de Dios Ortúzar and Willumsen, 2011).

7.3.5. Mixed Logit (ML)

The ML is applied to five studies that simultaneously model multiple travel choices (Bhat, 1998a; Börjesson, 2008; De Jong et al., 2003; Hensher and Reyes, 2000; Hess et al., 2007b). This model structure is introduced to bridge the gap between Probit and Logit models (McFadden and Train, 2000). ML introduces a new error term to the utility function described in Eq. 7.1:

$$U_{in} = V_{in} + \eta_{in} + \varepsilon_{in} \quad \forall\, i \in C_n \tag{7.9}$$

where the random error ε_{in} is assumed to follow a *iid* Gumbel distribution and η_{in} is a random term representing one or more additional components of the unobserved part of the utility independent from ε_{in}. The model is expressed as follows:

$$P(i|C) = \int \left(\frac{e^{V_i(\beta)}}{\sum_{j \in C} e^{V_j(\beta)}} \right) f(\beta) d\beta \tag{7.10}$$

where $f(\beta)$ is the probability density function, evaluated for all parameters β. The model does no longer have a closed-form, requiring simulation or numerical integration as part of the estimation process which significantly increases the computational effort. ML can be applied in two different, but mathematically equivalent ways, namely as RPL or as Error Component Logit (ECL).

RPL introduces random taste heterogeneity by estimating a distribution of parameter values, generally assumed to follow a normal distribution $\beta \sim N(\mu, \sigma)$. In the context of modelling multiple travel choices, this approach does not allow for correlation among alternatives, therefore adopting the IIA assumption of MNL models. RPL was applied in a study that addressed mode choice and trip chaining (Hensher and Reyes, 2000). By applying RPL, they assume independence of travel alternatives and have the disadvantage of high computational effort, consequently missing out on the potential of ML for modelling multiple travel choices.

ECL allows alternatives to share the random error term, thus allowing for correlation between alternatives. ECL can closely replicate NL (partial correlation) or CNL (full correlation), therefore it is flexible in its correlation structure. Furthermore, it has the advantage that it can also accommodate for random taste heterogeneity and heteroscedasticity (ECL+RPL). ECL estimates the error component $\sim N(0, \sigma)$ for a set of alternatives. It is applied in three studies on mode and departure time choice (Bhat, 1998a; De Jong et al., 2003; Hess et al., 2007b). All three studies allow for full correlation. Börjesson (2008) applies the ECL+RPL approach to mode and departure time choice, benefiting from the full potential of ML regarding heterogeneity and hierarchy.

7.3.6. Discrete-Continuous Models

Discrete-Continuous models are partially discrete choice and partially continuous. Three studies have applied this model structure for the simultaneous modelling of multiple travel choices (Habib, 2013; Habib et al., 2009; Shabanpour et al., 2017). Regarding travel choices, departure time is sometimes (but not always) modelled as a continuous variable. According to Bhat and Steed (2002) there are several issues when treating departure time as a discrete variable. First, it is difficult to identify the best time interval for discretisation. Second, the times at the boundaries of an alternative are also considered to be distinct alternatives. Third, evaluation of policies needs to be done based on the same discrete time periods. To overcome these issues, one could incorporate departure time as a continuous variable (Vickrey, 1969).

Habib et al. (2009) and Shabanpour et al. (2017) apply the continuous-time hazard model for departure time choice (Bhat and Steed, 2002). It is a log-linear model, which recognizes the dynamics of activity duration via consideration of the conditional probability of termination of the activity. The model is described by the following expression:

$$\ln(t_{in}) = V_{in} + \alpha_{in} \tag{7.11}$$

where $\alpha_{in} \sim N(0, \sigma_{int})$. Both studies apply a MNL model for mode choice, however their method for simultaneously modelling the two choices differs. Habib et al. (2009) estimate the joint model by transforming the random error terms of both individual models into equivalent standard normal variables and they describe the corresponding joint distribution as an equivalent bivariate normal distribution (BVN) (Lee, 1983). Shabanpour et al. (2017), on the other hand, use the copula approach to identify the joint model (Sklar, 1973). The so-called copula C_θ represents the joint probability distribution of random variables with predefined marginal distributions, which can be represented by several functions (Bhat and Eluru, 2009).

The third study, Habib (2013), bases his approach on the MDCEV model (Bhat, 2005). In the MDCEV approach discrete and continuous variables are simultaneously modelled using one utility function with a clear interpretation per choice. He 'loosens' this description by introducing two separate utility functions and by explicitly modelling the correlation between the continuous and discrete choice. To do so, he translates the departure time choice, to the time spent at home before leaving for work. Mode choice follows the utility function introduced in Eq. 7.1 and departure time is defined as follows:

$$U(t_k) = \sum_{k=1}^{2} \frac{1}{\alpha_k} e^{V_k + \varepsilon'_k} (t_k^{\alpha_k} - 1) \tag{7.12}$$

where k=1 indicates the before-departure at home activity, and k=2 is the rest of the day. Furthermore, α_k is the satiation parameter and $t_k^{\alpha_k}$ represents the time expenditure on before-work and rest-of-the-day activities. The joint model describes the correlation using a BVN to allow for full correlation between alternatives, like Habib et al. (2009).

7.3.7. Segmentation Approaches

Segmentation approaches have built upon the Logit models, but add an additional structure to these models that allow for different decision-making processes within the population. Two model structures have been identified that address multiple travel choices: the Co-Evolutionary Model (CEM) (Krygsman et al., 2007) and the Flexible Model Structure (FMS) (Ishaq et al., 2013).

Krygsman et al. (2007) introduce CEM, which combines MNL models for each travel choice with an iterative method that determines the order of decisions for each individual. The expected utility for each travel choice depends on perceived value of the attributes related to that choice. Furthermore, they assume that the availability of an alternative is dependent on other decisions (uncertainty in decisions). They use MNL for the individual decisions, and start their iterative procedure by fixing all probabilities to be equal for all alternatives. At the end of each iteration, the amount of uncertainty in a travel choice is calculated as the entropy of the travel choice and the level of convergence. The latter signifies the difference between calculated probabilities in the current and previous iteration. If the uncertainty of one of the travel choices is lower than an assumed threshold, the decision is fixed by setting the probability of the alternative with the highest utility to 100%. This iterative process continues until all choices have been made. Consequently, this approach introduces a hierarchy in the decision process, as all choices are assumed to be made sequentially. They apply CEM to trip chain and mode choice, where most individuals decide on the trip chain before the mode. Due to iteratively estimating MNL models, the computational effort of this approach is low. Their approach was also applied by Li et al. (2013) to trip chain and mode choice.

FMS allows for more flexibility regarding heterogeneity in individuals' decision-making processes (Ishaq et al., 2013). The population is segmented based on the notion that segments differ in their decision-making processes. The general formulation of the model is as follows:

$$P_{in} = \sum_{s=1}^{S} \sum_{g=1}^{G} P_n\,(i|s)\,P(s|g)P_n(g|x, \alpha_x) \tag{7.13}$$

where $P_n(i|s)$ is the probability that individual n chooses alternative i, given model structure s, and where $P(s|g)$ is the degree of membership of segment g in structure s, and where $P_n(g|x, \alpha_x)$ represents the probability that individual n chooses segment g, given their characteristics x and their weights. As this method is very flexible, each of these parts can be modelled using a different method. This method resembles the latent class model, where segments of individuals can have different model structures assigned (Greene and Hensher, 2003). The authors suggest that the segmentation is modelled using the C-means algorithm, the assignment of segments using ML models, and the choice of alternatives using a NL model. The NL structure suggests that travel choices are partially correlated, with different nesting structures for different segments. They apply FMS to destination and mode choice. Because of the different modelling aspects, one of which is modelled using the ML structure, the computational effort is high.

7.3.8. Miscellaneous Approaches

Two studies in this literature review could not be assigned to the model approaches described hitherto: Simultaneous Logit (SL) (Ye et al., 2007) and MNL-ordered response formulation (Bhat, 1997).

SL is introduced by Schmidt and Strauss (1975), where they assume that travel choices are made simultaneously. This model can be considered an extension of the MNL model. The SL model, for two binary choices, is formulated as follows:

$$\begin{aligned} \ln\left[\frac{P_{in}(X_n = 1|Y_n)}{P_{in}(X_n = 0|Y_n)}\right] &= \gamma' z_n + \alpha Y_n \\ \ln\left[\frac{P_{in}(Y_n = 1|X_n)}{P_{in}(Y_n = 0|X_n)}\right] &= \beta' x_n + \eta X_n \end{aligned} \tag{7.16}$$

where X and Y are travel choices, $\gamma' z_n$ and $\beta' x_n$ the deterministic parts of the utility, and αY_n and ηX_n represent the influence of this choice on the other choice. It is necessary to set $\alpha = \eta$, leaving only α to be estimated (the joint dependence). Ye et al. (2007) applied this method to trip chain and mode choice. Computational effort is low.

Bhat (1997) introduces a joint model for trip chain and mode choice. The mode choice is represented by a MNL model and the trip chain choice is modelled using an ordered response formulation (ORF) (McKelvey and Zavoina, 1975). Trip chain choice is translated to the number of non-work commute stops (ordinal choice) presented in Eq. 7.17:

$$\begin{gathered} s_{in}^* = \gamma'_i x_{in} + \eta_{in}\,, s_{in} = k \\ \text{if } \delta_{i,k-1} < s_{in}^* \leq \delta_{i,k},\, s_{in} \text{ observed for chosen mode } i \end{gathered} \tag{7.17}$$

where s_{in}^* is the stop-making propensity of individual n when using mode i. k is the number of stops and s_{in} is characterised by the stop-making propensity s_{in}^* and thresholds δ, in the standard ordered response fashion. The error terms η_{in} are identically normal-distributed across modes i and individuals n. The joint model can be estimated when the error terms assume the same distribution, requiring transformation of the *iid* Gumbel term into a standard normal variable. A BVN then allows full correlation between alternatives of choices.

7.4. Simultaneous Modelling of Multiple Travel Choice Dimensions

This review focusses on five travel choice dimensions: trip chain, destination, departure time, mode, and route. The studies identified, consider a combination of these choice dimensions. Table 7.1 provides an overview of the models applied for each combination of travel choice dimensions. Please note that in case three choices were modelled simultaneously (Eluru et al., 2010; Vrtic et al., 2007), the model structure was assigned to each combination of these choice dimensions.

Table 7.1: Model structures (and sub-category) applied to model different combinations of travel choice dimensions

	Trip Chain	Destination	Departure Time	Mode	Route
Trip Chain				NL (NL) Probit (RBP) ML (RPL) Misc. (SL) Misc. (MNL-ORF) SA (CEM)	
Destination			NL (MDCEV-MNL)	MNL (JL) NL (NL) NL (NL-MNL) NL (MDCEV-MNL) CNL (CNL) SA (FMS)	NL (MNL-PSL)
Departure Time				MNL (Vj) NL (NL) NL (MNL-OGEV) NL (MDCEV-MNL) NL (RPL-RPL) CNL (CNL) ML (ECL) ML (ECL+RPL) DC (BVN-time) DC (Copula-time) DC ('loose' MDCEV)	MNL (Vj) Probit (Probit)
Mode					MNL (JL) MNL (Vj) NL (NL)
Route					

From Table 7.1 several interesting observations can be made. First, some combinations of choice dimensions have been modelled using many different model structures. In activity-based models, any combination of choices can occur, but often a certain sequence is assumed. For example in the Portland model (Bowman et al., 1998) the hierarchical order (imposed by NL models (de Dios Ortúzar and Willumsen, 2011)) is the following: trip chain, departure time (time-of-day), and combined mode and destination choice. Given this hierarchically imposed order of choice dimensions, which is similar for other activity-based models, it would make sense to find abundance of model structures that have been applied on travel choice combinations that are adjacent in this so-called hierarchy. Table 1 shows that this is the case for some of the adjacent travel choice combinations, such as destination-mode, departure time-destination and departure time-mode. However, one abundantly modelled combination is not part of this so-called hierarchy, namely the trip chain-mode combination. There are, however, sound arguments for modelling this specific combination. Researchers are generally interested in the complexity of trip chaining patterns and how this relates to the modes chosen. For example, Ye et al. (2007) have investigated whether the choice for the car increases the complexity of trip chains in comparison to public transport. Their research focus excludes the need to include departure time in the model. In the four-step model, mode choice generally precedes route choice in the modelling workflow (de Dios Ortúzar and Willumsen, 2011), which is also a recurring combination in this review. Consequently, most combinations of travel choice dimensions that are simultaneously modelled are in line with the order implied by activity-based and four-step models.

Second, a large variation of model structures is found in the departure time-mode choice combination, which is mostly related to the departure time choice dimension. There have been discussions on the representation of departure time as a discrete or continuous variable (Bhat and Steed, 2002). Some of the model structures include discrete departure time (e.g. MNL (Vj) and CNL (CNL)), whereas others model departure time as a continuous variable (discrete-continuous structures). Several arguments were provided by Bhat and Steed (2002) that reason against a discrete representation. There are noticeable efforts to address these limitations, for example by introducing the OGEV model (Bhat, 1998b). In general, the discrete-continuous structures have been applied later than the discrete structures, i.e. before versus after 2010.

Third, the combination mode-route is often modelled at an aggregated level in assignment models (like the four-step model). Prato (2009) provides a thorough review on the different discrete choice models suitable for modelling route choice. Generally, routes partially overlap with one another, which cannot be captured using the MNL model, and is difficult to capture using NL or CNL structures. In the review, only MNL and NL applications are observed, meaning that the overlap of route choices is not (accurately) taken into account (Broach and Dill, 2016; Montini et al., 2017; Shakeel et al., 2016; Vrtic et al., 2007). Consequently, no satisfactory solution is provided for route choice overlap in these model structures. Debrezion et al. (2009) introduce PSL for the route choice (in the destination-route choice combination), which does provide a solution for overlap. Furthermore, the mode-route choice studies all date starting from the late 2000's, suggesting that modelling the mode-route combination by means of discrete choice models, is gaining popularity.

Last of all, in activity-based models the mode-destination combination is often modelled simultaneously (using MNL). In the model structures assessed in this review not only MNL, but also NL, CNL, and a segmentation approach (FMS) are represented. The NL introduces substitution patterns, which imply a sequence in modelling the dimensions, where both hierarchical structures are found (Debrezion et al., 2009; Eluru et al., 2010; Newman and Bernardin, 2010). While CNL allows for correlation over both choice dimensions, results showed a higher substitution for destinations compared to modes (Ding et al., 2014). Ishaq et al. (2013) used the FMS framework and found both hierarchical structures among their population. These findings suggest that the decision might not be 'fully' simultaneous after all, as the substitution patterns suggest one dimension is changed before the other one does.

Summarising, for many combinations of travel choice dimension research has drifted away from the MNL and NL structures that are often applied in activity-based and four-step models. A wide variety of adaptations and improvements have been suggested for different combinations of travel choice dimensions.

7.5. Assessment Indicators

This section discusses the indicators used for assessing the different discrete choice modelling structures. As mentioned before, both the suitability of the model structure for simultaneously modelling multiple travel choice dimensions as well as its applicability by researchers or practitioners are examined. The suitability indicators are discussed in section 7.5.1 and the applicability indicators are discussed in section 7.5.2. Figure 7.2 provides an overview of the indicators and the range of possible outcomes per indicator.

Suitability	Applicability
1. Flexible correlation structure *Yes/No* 2. Correlation between choice dimensions *No/Partial/Full* 3. Heterogeneity in decision-making *Yes/No* 4. Model specification of choice dimensions *Universal/Specific*	1. Computational effort *Low/Medium/High* 2. Parameters related to model structure *e.g.* $\{\mu_{c1}\}$ *for Nested Logit*

Figure 7.2: Indicators for assessment of the model structures

7.5.1. Suitability Indicators

Four different indicators are identified for assessing the suitability of model structures. These indicators relate to the flexibility in the correlation structure of the model, correlation between choice dimensions, heterogeneity in the decision-making process of individuals, and the model specification of different choice dimensions. These indicators are chosen because they are able to capture the degree of realism in decision-making processes of individuals when regarding multiple choice dimensions. Furthermore, these indicators reflect the strong, and potentially unrealistic, assumptions of the MNL and NL models, which are often used in activity-based and four-step models.

First of all, the flexibility of the correlation structure in the model structure is assessed. A flexible correlation structure is defined by the possibility to identify different correlation structures. As an example, a flexible model structure is able to accommodate both full correlation between travel choice dimensions as well as partial correlation, whereas an inflexible model structure only accommodates full correlation. This flexibility allows the researcher to find the best correlation structure between the different choice dimensions for the data available. Consequently, a flexible correlation structure is preferred.

The second indicator, which is related to the first, represents the correlation structure applied by the researcher (ε_{in} in Eq. 7.1). The difference between those two indicators is that the first shows the theoretical possibilities of the correlation structure, whereas the second reflects the application of the correlation structure. In case of modelling multiple travel choice dimensions, this reflects the correlation between alternatives of different travel choices. Three different correlation structures are possible: no correlation (independence), partially correlated, and fully correlated. Partial correlation means that, when the situation changes, alternatives of one choice dimension are substituted before alternatives of the other choice dimension. As an example, the substitution pattern might show that the mode is changed before the departure time. In the literature this is also referred to as hierarchy among choices (de Dios Ortúzar and

Willumsen, 2011). Fully correlated structures allow for calculation of substitution patterns but also emphasize the joint choice, therefore these structures are preferred.

Different correlations structures between travel choices have been found in different studies. Similarly, one can argue that one individual can have a different decision-making process compared to another, resulting in different substitution patterns per individual. The third indicator addresses the segmentation of the population regarding the decision-making process. It reflects whether the model structure allows for segmentation or differentiation within the population in the decision-making process. Due to evidence supporting existence of heterogeneity in decision-making processes (Ishaq et al., 2013; Krygsman et al., 2007), model structures should preferably be able to account for it.

The last indicator addresses the fact that different choice dimensions might need different model structures. For example, route choice ideally is treated differently than departure time choice, as different issues are encountered in determining the alternatives and their correlations. More specifically, in route choice, different paths are available which overlap spatially, whereas in departure time choice this type of overlap is not present. The fourth indicator reflects whether a universal model structure is imposed for all considered choices or a specific structure that is optimised for each choice dimension. For example, route choice and departure time choice are simultaneously modelled using the same model structure or they are modelled using different, more specific, model structures. Consequently, a specific model specification is preferred.

7.5.2. Applicability Indicators

Researchers and, especially, practitioners generally prefer a model structure that is easy to interpret and that is relatively fast to estimate and forecast (Hess et al., 2007a). To assess the practical applicability of the different model structures for modelling multiple travel choices simultaneously, we propose two indicators: computational effort and the additional parameters that need to be estimated due to the selected model structure.

The first indicator is the computational effort. A differentiation between low, medium, and high effort levels is made. A model structure that has low computational effort, has a closed-form mathematical formulation and is concave in its objective function. Lack of closed-form formulation in the model structure requires Monte Carlo simulations or numerical integration in the estimation process, which results in high computational effort for a model structure. At the intermediate level, some model structures have a closed-form formulation, but for example a non-concave objective function (medium). For practical application, low computational effort is preferred. For research purposes, computational effort is generally less of an issue, unless the model is developed for practical purposes.

The second indicator represents the ease of interpretation and is operationalised by means of the number of additional parameters that are attributed to the choice for a model structure. The MNL can be considered a base case since it does not induce any additional parameters. Preferably limited extra parameters are imposed by the model structure, as these have to be estimated and can increase the computational effort as well as reduce the interpretability of estimation results. Furthermore, if many additional parameters need to be estimated, more data is required for model estimation which reflects lack of parsimony.

7.6. Assessment of Model Structures

In this section, the model structures introduced in section 7.3 are assessed with respect to the identified indicators related to suitability and applicability. Section 7.6.1 discusses this assessment related to the suitability indicators and Section 7.6.2 discusses the assessment in relation to the applicability indicators. In Section 7.6.3 we provide a synthesis of the assessment. Appendix 7.A provides an overview of all the reviewed studies and how the applied model structures score on each indicator.

7.6.1. The Suitability of Model Structures

The suitability of the model structures (as implemented in the literature) is assessed using four indicators; flexibility of the correlation structure, correlation between choice dimensions, heterogeneity in decision-making process, and model specification of choice dimensions. Table 7.2 provides an overview of each model structure's score. Please note that some of the categories of model structures have sub-categories which may score differently. Several implementations of model structures are possible, which each have different abilities. For example, RPL (ML) has the power to reflect heterogeneity within a choice dimension towards several variables, but does not accommodate correlations between choice dimensions. ECL (ML), on the other hand, accommodates the latter.

Table 7.2: Assessment of identified model structures based on suitability indicators: model structure (sub-category)

Heterogeneity in decision-making process	Model specification of choice dimensions	Flexible correlation structure					
		No			*Yes*		
		Correlation between choice dimensions			Correlation between choice dimensions		
		No	*Partial*	*Full*	*No*	*Partial*	*Full*
No	*Universal*	MNL (JL)	NL (NL) NL (RPL-RPL) Probit (RBP)	CNL (CNL) Misc. (SL)	ML (RPL)	-	Probit (Probit) ML (ECL) ML (ECL+RPL)
No	*Specific*	MNL (Vj)	NL (MNL-OGEV) NL (NL-MNL) NL (MNL-PSL) NL (MDCEV-MNL)	DC (all) Misc. (MNL-ORF)	-	-	-
Yes	*Universal*	-	SA (CEM)	-	-	SA (FMS)	-
Yes	*Specific*	-	-	-	-	-	-

Given the definition of the indicators by the authors, the top-left quadrant represents model structures that have no flexible correlation structure and do not incorporate heterogeneity in the decision-making process of individuals. In general, model structures are developed for single choice dimensions, in that case the flexibility in correlation and heterogeneity of decision-making are less (or not) relevant. Consequently, many model structures reside in this quadrant. Within this quadrant all MNL and NL structures are situated, which are often used in the activity-based and four-step models. Consequently, this means that flexibility in the correlation structure and heterogeneity in decision-making processes are generally not part of the activity-based and four-step models and are, thus, often compromised.

The lower half of the table refers to heterogeneity in the decision-making process of individuals (or segments of the population). Basically, one person or group of persons could have different substitution patterns of one travel choice compared to another. For example, one group of individuals has a higher sensitivity towards mode changes compared to the destination for shopping. Consequently, this group will change their mode before changing their destination. This might be the reversed for another user group. These two groups can co-exist within the population. In the current activity-based and four-step models, this distinction is not made. Only two model structures are designed to incorporate this heterogeneity: the segmentation approaches CEM (Krygsman et al., 2007) and FMS (Ishaq et al., 2013). These two studies acknowledge that individuals or groups of individuals have different sensitivity levels towards different choice dimensions. Consequently, this indicator offers ample room for improvements in future research.

The right side of the table pertains to model structures with flexible correlation structures. These consists of ML and Probit model structures (FMS incorporates ML in its model structure). FMS is flexible and is able to accommodate different model structures. Therefore, also ML and potentially Probit structures, can be included in FMS while allowing for heterogeneity in the decision-making process. In FMS this happens via a latent class component (Greene and Hensher, 2003). This shows that combining currently existing model structures, might improve the simultaneous modelling of multiple travel choice dimensions.

The grey area in the right-bottom corner corresponds to the 'ideal' model structure in relation to the set of indicators employed in this study. This model structure would be flexible in the correlation structure, include heterogeneity in the decision-making process, and has a choice dimension specific model specification. This 'ideal' structure does not appear in any of the reviewed studies. The model structure that comes closest to it is the segmentation approach FMS (Ishaq et al. 2013). This model structure is flexible and, in theory, allows for different model structures to be used for different choice dimensions. However, it has not been applied in this way, therefore there is no evidence yet to support this claim. Consequently, when the aim is to simultaneously model multiple travel choice dimensions, there is room for improvement regarding the suitability of model structures.

7.6.2. The Applicability of Model Structures

To test the applicability of the model structures identified in this review, the models are assessed using two indicators: one related to the computational effort and one related to the ease of interpretation. The ease of interpretation is translated into an objective indicator: the number of parameters that are induced by the selected model structure.

The number of parameters imposed by the model structure is zero in the case of a MNL model. This model therefore serves as a baseline. Several types of parameters can be imposed by model structures. β's reflecting the impact of other choice dimensions on a particular choice are required by the MNL (Vj) and CEM models. Nesting parameters μ are required in NL, CNL, and segmentation approach FMS (given the specification in the study by Ishaq et al. (2013)). Parameters representing the mean and standard deviations are used in ML and FMS structures. Finally, parameters related to the correlation between two choice dimensions are used in discrete-continuous approaches. Generally, the model structures related to MNL and NL require a limited number of additional parameters. Other structures require more parameters, which in turn makes it more difficult to interpret. The number of additional parameters to be estimated per model is indicated in the overview in Appendix 7.A.

The (qualitative) assessment of the computational effort is visualised in Table 7.3. As mentioned before, most MNL and NL model structures have a low computational effort. This is due to their closed-form formulation. Furthermore, SL and CEM also have a low computational effort, because they are based on the MNL structure. In the medium category, the model structures generally have non-concave formulation resulting with multiple local optima, instead of a global optimum. Consequently, they benefit from closed form formulation, however there is no guarantee that the global optimum is found. All models based on ML or Probit require high computational effort, due to lack of a closed-form formulation, which means that numerical optimisation or Monte Carlo simulation is required to estimate these models. By simulating using a large enough set of draws (e.g. >1000), the global optimum can be found.

Table 7.3: Assessment of identified model structures based on computational efficiency: model structure (sub-category)

Computational effort		
Low	*Medium*	*High*
MNL (all) NL (NL) NL (NL-MNL) NL (MNL-PSL) NL (MNL-OGEV) Misc. (SL) SA (CEM)	NL (MDCEV-MNL) CNL (CNL) DC (all) Misc. (MNL-ORF)	NL (RPL-RPL) Probit (all) ML (all) SA (FMS)

The applicability indicators largely favour the currently applied model structures (MNL and NL). Furthermore, our definition of interpretability of the model estimation is closely related to the computational effort. The model structures that have most parameters imposed, are the ones that are associated with the highest computational effort, but potentially also result with the global optimum (i.e. optimal solution). Consequently, the model structures introduced after MNL and NL seem to be less suitable for practical purposes.

7.6.3. Synthesis of the Assessment of Model Structures

Finally, we bring together the suitability and applicability indicators. This synthesis helps in understanding when a model structure can be used, given considerations on suitability and applicability. We do this by identifying the model structures that are most preferable according to each suitability indicator and assess them based on their applicability. Table 7.4 intersects the assessment of the most suitable model structures with respect to their applicability.

Table 7.4: Assessment of most suitable model structures according to each indicator on computational effort: model structure (sub-category)

	Applicability		
	High	*Medium*	*Low*
Flexible correlation structure			Probit (Probit) ML (ECL) ML (ECL+RPL) SA (FMS)
Full correlation between choice dimensions	Misc. (SL)	CNL (CNL) DC (all) Misc. (MNL-ORF)	Probit (Probit) ML (ECL) ML (ECL+RPL)
Heterogeneity in decision-making process	SA (CEM)		SA (FMS)
Specific model specification per choice dimension	MNL (Vj) NL (MNL-OGEV) NL (NL-MNL) NL (MNL-PSL)	NL (MDCEV-MNL) DC (all) Misc. (MNL-ORF)	

Table 7.4 shows that multiple model structures are considered most suitable according to the individual indicators. Several interesting observations can be made. First, regarding the model structures with flexible correlation structures, we observe that all of them score low on applicability. This leads to the conclusion that currently flexibility of a model structure comes at a high computational cost. Second, the model structures that include full correlation between choice dimensions show a large variety in terms of their applicability (SL (based on MNL) versus the flexible ML and Probit structures). Third, only two model structures are heterogeneous in their decision-making process. FMS includes ML in its model structure, making it less applicable, but it is flexible and has high potential. CEM is based on MNL and thus efficient. Finally, the model structures that incorporate choice dimension specific specification are nearly all NL or discrete-continuous. These are low to medium in their computational effort, making them applicable.

Some model structures occur more often in Table 4. The ML and Probit model structures score high on flexible correlation structures and full correlation between choice dimensions. The FMS structure is flexible and allows for heterogeneity. In theory that model structure could also include full correlation between choice dimensions and potentially also differentiate between model structures per choice dimension. The discrete-continuous and MNL-ORF model structures allow for full correlation and dimension specific specification. All these more suitable model structures are less applicable as they have medium to high computational effort and require the estimation of several parameters that are imposed by the model structure. Consequently, there seems to be a large discrepancy between those model structures which are most suitable according to our indicators, i.e. are considered to be most behavioural realistic, and those most applicable. Activity-based and four-step models encompass the most applicable structures, but these are not necessarily the most suitable.

7.7. Conclusions and Future Research Directions

This literature review study assesses the suitability and applicability of different discrete choice modelling structures when simultaneously modelling of multiple travel choice dimensions. The best known examples of this type of models are the activity-based and four-step models. These models mostly use MNL and NL models, which rely on some strong assumptions and are found to lack the power to accommodate complex relationships realistically. Many advancements have been made in discrete choice models since the introduction of MNL and NL. We reviewed studies that apply discrete choice models to simultaneously model a combination of trip chain, destination, departure time, mode, and route choices. We have assessed the model structures used with respect to their suitability, reflecting the behavioural realism, and the applicability, reflecting the ease of implementation (in practice).

The studies identified for this review sometimes adhere to the hierarchy often introduced in activity-based and four-step models (de Dios Ortúzar and Willumsen, 2011). If they adhere to the order, they often find a variety of substitution patterns that is not always in agreement with the general application of the activity-based and four-step models. If they do not adhere, they do find significant relationships, again with varying substitution patterns. This suggests that a wider variety of relationships occurs between the travel choice dimensions than has been conventionally assumed in the activity-based and four-step models.

This literature review indicates that none of the model structures corresponds to the 'ideal' model based on the identified suitability indicators, i.e. no structure is flexible in the correlation structure, introduces full correlation between choice dimensions, allows for heterogeneity in the decision-making process of individuals, and has specific model specification per choice dimension. Moreover, the models that are considered most suitable (on one or more indicator) are generally less applicable. Consequently, in terms of behavioural realism, the model structures can be improved. Furthermore, this means that currently increased behavioural realism comes at the cost of hampering applicability.

The most promising path for improving the model structures is the latent class model, where several complex model structures can be included, for example ML. The latent class model is able to accommodate heterogeneity in the decision-making process of individuals by assigning individuals to different classes. The ML model is flexible and allows for full correlation between choice dimensions. Consequently, a combination of these models could potentially tackle the suitability component, as it will increase the behavioural realism of the model structure, albeit this does not contribute to its applicability.

Based on this literature review several directions for future research on the simultaneous modelling of multiple travel choices using discrete choice modelling structures are identified. First, heterogeneity in the decision-making process is insufficiently accommodated in current model structures. Research has shown the relevance of this aspect. Accommodating this heterogeneity will contribute to increased behavioural realism. Therefore, more research is needed on how to incorporate this heterogeneity into the model structures. Second, the models that are able to include a specific approach towards each choice dimension do generally not adhere to the other suitability indicators. Consequently, model structures need to be developed that allow to cater for individual choice dimensions while better adhering to the other indicators. Third, latent class models (Greene and Hensher, 2003) seem promising for improving the behavioural realism with respect to user heterogeneity. Finally, often gains made in terms of realism in behaviour come at a high computational cost. A trade-off is made in practice between computational efficiency and behavioural accuracy, resulting in the use of efficient but less nuanced models (like MNL and NL). Consequently, it is important to enable practitioners to use the potential of the current theoretical advancements, by developing models that satisfy or offer a good compromise between suitability on one hand, and applicability on the other hand.

Acknowledgements

This research was supported by the Allegro project (no. 669792), which is financed by the European Research Council and Amsterdam Institute for Advanced Metropolitan Solutions.

Appendix 7.A: Overview of the analysis. TC=trip chain, D=destination, DT=departure time, M=mode, R=route, Cat.=category, spec.=specification, JL=Joined Logit, RBP=recursive bivariate probit, DC=discrete-continuous, BVN=bivariate normal distribution, SA=segmented approach, Misc.=miscellaneous

		Choices modelled					Modelling approach		Suitability				Applicability	
#	Study	TC	D	DT	M	R	Cat.	Sub Category	Flexible correlation structure	Correlation between choices	Heterogeneity in decision-making	Model spec. of choice dimensions	Comp. effort	Required Parameters
1	Richards & Ben-Akiva (1974)	✗	✓	✗	✓	✗	MNL	JL	no	no	no	universal	low	-
2	Montini et al. (2017)	✗	✗	✗	✓	✓	MNL	JL	no	no	no	universal	low	-
3	Broach & Dill (2016)	✗	✗	✗	✓	✓	MNL	in Vj	no	no	no	specific	low	$\{\beta_{c2}\}$
4	Vrtic et al. (2007)	✗	✗	✓	✓	✓	MNL	in Vj	no	no	no	specific	low	$\{\beta_{c2}\}$
5	Shakeel et al. (2016)	✗	✗	✗	✓	✓	NL	NL	no	partial	no	universal	low	$\{\mu_{c1}\}$
6	Debrezion et al. (2009)	✗	✓	✗	✓	✗	NL	NL	no	partial	no	universal	low	$\{\mu_{c1}\}$
7	Hess et al. (2007b)	✗	✗	✓	✓	✗	NL	NL	no	partial	no	universal	low	$\{\mu_{c1}\}$
8	Lizana et al. (2013)	✗	✗	✓	✓	✗	NL	NL	no	partial	no	universal	low	$\{\mu_{c1}\}$
9	Paleti et al. (2014)	✗	✗	✓	✓	✗	NL	NL	no	partial	no	universal	low	$\{\mu_{c1}\}$
10	Yang et al. (2016)	✓	✗	✗	✓	✗	NL	NL	no	partial	no	universal	low	$\{\mu_{c1}\},\{\mu_{c2}\}$
11	Bhat (1998a)	✗	✗	✓	✓	✗	NL	MNL-OGEV	no	partial	no	specific	low	$\{\mu_{c1},\rho_{c2}\}$
12	Newman & Bernardin Jr (2010)	✗	✓	✗	✓	✗	NL	NL-MNL	no	partial	no	specific	low	$\{\mu_{c1},\mu_{c1.1}\}$
13	Bawja et al. (2008)	✗	✗	✓	✓	✗	NL	RPL-RPL	no	partial	no	universal	high	$\{\mu_{c1},\mu_{rp},\sigma_{rp}\}$
14	Eluru et al. (2010)	✗	✓	✓	✓	✗	NL	MDCEV-MNL	no	partial	no	specific	medium	$\{\mu_{c1}\}$
15	González et al. (2016)	✗	✓	✗	✗	✓	NL	MNL-PSL	no	partial	no	specific	low	$\{\mu_{c1},\beta_{PS,c2}\}$
16	Ding et al. (2014)	✗	✓	✗	✓	✗	CNL	CNL	no	full	no	universal	medium	$\{\mu_{c1},\mu_{c2}\}$
17	Yang et al. (2013)	✗	✗	✓	✓	✗	CNL	CNL	no	full	no	universal	medium	$\{\mu_{c1},\mu_{c2}\}$
18	Jou (2001)	✗	✗	✓	✗	✓	Probit	Probit	yes	full	no	universal	high	$\{\Sigma_{c12},\sigma_{\upsilon}\}$
19	Ye et al. (2007)	✓	✗	✗	✓	✗	Probit	RBP	no	part	no	universal	high	$\{\eta_{c1},\rho_{c12}\}$
20	Hensher & Reyes (2000)	✓	✗	✗	✓	✗	ML	RPL	no	no	no	universal	high	$\{\mu_{rp},\sigma_{rp}\}$
21	Bhat (1998b)	✗	✗	✓	✓	✗	ML	ECL	yes	full	no	universal	high	$\{\sigma_{c1},\sigma_{c2}\}$
22	De Jong et al. (2003)	✗	✗	✓	✓	✗	ML	ECL	yes	full	no	universal	high	$\{\sigma_{c1},\sigma_{c2}\}$
23	Hess et al. (2007a)	✗	✗	✓	✓	✗	ML	ECL	yes	full	no	universal	high	$\{\sigma_{c1},\sigma_{c2}\}$
24	Börjesson (2008)	✗	✗	✓	✓	✗	ML	ECL+RPL	yes	full	no	universal	high	$\{\mu_{rp},\sigma_{rp},\sigma_{c1},\sigma_{c2}\}$
25	Habib et al. (2009)	✗	✗	✓	✓	✗	DC	BVN-time	no	full	no	specific	medium	$\{\sigma_{c2},\rho_{c12}\}$
26	Shabanpour et al. (2017)	✗	✗	✓	✓	✗	DC	Copula-time	no	full	no	specific	medium	$\{\sigma_{c2},\theta_{c12}\}$
27	Habib (2013)	✗	✗	✓	✓	✗	DC	'loose' MDCEV	no	full	no	specific	medium	$\{\alpha_{c2},\rho_{c12}\}$
28	Krygsman et al. (2007)	✓	✗	✗	✓	✗	SA	CEM	no	partial	yes	universal	low	$\{\beta_{c1},\beta_{c2}\}$
29	Li et al., (2013)	✓	✗	✗	✓	✗	SA	CEM	no	partial	yes	universal	low	$\{\beta_{c1},\beta_{c2}\}$
30	Ishaq et al. (2013)	✗	✓	✗	✓	✗	SA	FMS	yes	partial	yes	universal	high	$\{\mu_{c1},\mu_{c2},\mu_{rp},\sigma_{rp}\}$
19	Ye et al. (2007)	✓	✗	✗	✓	✗	Misc.	SL	no	full	no	universal	low	$\{\alpha_{c12}\}$
31	Bhat (1997)	✓	✗	✗	✓	✗	Misc.	MNL-ORF	no	full	no	specific	medium	$\{\delta_{c2},\rho_{c12}\}$

Chapter 8 – Conclusions and Recommendations

This thesis investigated mode and route choice behaviour of active mode users and aimed at understanding and modelling this choice behaviour. This final chapter summarises and discusses the main results of the thesis. First, the main findings of the thesis are presented. This is followed by a discussion on several methodological decisions made in this thesis. Next, implications for practice that are resulting from the research in this these are given. Finally, recommendations for future research are provided.

8.1. Main Findings

This section discusses the main findings of this thesis according to scientific contributions discussed in Chapter 1. In this discussion, the research questions are answered.

The Relationship between the Daily Mobility Pattern and Attitudes towards Modes

Walking and cycling are very prominent in the Netherlands, with approximately half of all trips using active modes of transport. However, of course, not everyone uses active modes. In this thesis, the total daily mobility patterns of individuals in the Netherlands is investigated and compared to their attitudes towards modes (Chapter 2). Five different classes of daily mobility pattern users are identified, two of which consist of single mode use (exclusive car and exclusive bicycle users) and three show multi modal use (car and bicycle, public transport+, and car, walk, and bicycle users). Only one of the classes does not include any active mode use (exclusive car users). The bicycle is included in four mobility pattern classes, whereas walking is only present in the two highly multimodal classes.

Attitudes reflect how individuals' think about a mode, based on comfort, safety, fun etc. For most modes, these attitudinal aspects are similar within that mode, i.e. if one aspect of a mode is found to be positive, the others are positive as well. This makes it harder to change individuals' attitudes towards a mode. The attitudes of individuals concerning walking and cycling are generally positive.

We found that the modes used on a daily basis are regarded more positively compared to unused modes and between classes significant differences in attitudes arise (Chapter 2). Individuals that have a preference for their used modes (consonant users) are potentially less inclined to switch modes However, some individuals prefer modes they do not use (i.e. dissonant users). This research determines that the classes of exclusive car users and car and bicycle users have relatively high shares of dissonant users, where the preference lies with active modes of transport. This implies that these individuals might be persuaded to change to active modes via, for example, reimbursement by employers for using the bicycle to work, as this was found to be an important determinant in the choice to cycle (Chapter 3 and 4). The multimodal classes (1, 3, and 4) already contain a share of active mode use, which can be further increased. Finally, a large share of the exclusive bicycle users does not use their preferred mode and 7% uses their least preferred mode. This can trigger an undesirable change in their mobility pattern, because the car is often preferred by these dissonant users. *(answer Research Question 1).*

Determinants of Active Mode Choice Behaviour in a Cycling-rich Context

We zoom in on individual trips. Several categories of mode choice determinants are relevant for active mode choice in the Netherlands: individual characteristics, household characteristics, season and weather characteristics, trip characteristics, built environment, and employment conditions (Chapter 3). The individual characteristics entail for example socio-demographics. These variables have limited association with the choice for active modes in the Netherlands, contrary to findings elsewhere. This might be due to the fact that the Netherlands has a very diverse cycling population, which means that people cycle regardless of their gender and age. Household characteristics are mostly influencing the choice to walk, where the number of children under the age of 12, the number of household members, and household income are relevant determinants. Season and weather are of limited influence on walking and cycling in the Netherlands, which might be due to the relatively mild climate with frequent rain, which is considered normal to the Dutch. Trip characteristics are relevant for both walking and cycling, however the exact determinants and their impact differ. As an example, peak hour travel relates negatively to walking but is unrelated to cycling, whereas weekday travel relates positively to cycling but is unrelated to walking. Features of the built environment are explanatory variables of both walking and cycling, and again the exact determinants and their impact differs. For example, cycling is positively associated with suburban environments, whereas the level of urbanisation has no significant relationship with walking. The employment conditions, especially reimbursement for using a mode, has a positive association with cycling. The most dominant categories of determinants for cycling are trip characteristics, built environment, and employment conditions, whereas trip characteristics, built environment, and household characteristics are most dominant for walking. Walking and cycling are influenced by different determinants and if they are influenced by the same determinants, the impact of these determinants differs *(answer Research Question 2).*

Determinants of the Experienced Mode Choice Set

A feasible mode (like in Chapter 3) is not necessarily a used mode, because an individual might own a mode but not use it for a trip. Consequently, when the aim is a modal switch over an enduring period of time, it is instead more relevant to know the likelihood of including or excluding a mode in the mode choice set. We propose to investigate this using the experienced choice set, which is the set of modes used over a long period of time (Chapter 4). This choice set might differ for different trip purposes, therefore we focus on commuting trips. The experienced choice set shows that many individuals only use one mode for their work trip (83.5%), suggesting habitual and/or captive behaviour, which will not be captured when specifying the feasible choice set. This could, of course, be different when considering multiple trip purposes. When using only one mode in the daily mobility pattern, which can be the case for the car and cycling, we identified that respectively 5% and 7% travel using their least preferred mode, which suggests that these are indeed captive users (Chapter 2). We estimate a discrete choice model to identify which determinants are relevant for the formation of the experienced choice set. (Chapter 4).

The relevant determinants for the experienced mode choice set for commuting can be categorised into individual characteristics, household characteristics, ownership, urban density, and employment conditions. In the dataset, 41% of the individuals has used the bicycle to travel to work, whereas only 4% has walked. The probability for including the bicycle in the experienced mode choice set increases for higher education, higher urban density, working part-time, owning a bicycle, and being reimbursed by the employer for using the bicycle. The probability for including the bicycle decreases when the car or public transport is reimbursed by the employer. The probability to include walking in the experienced choice set (for the full commute) increases with the presence of children under the age of 12 in the household and decreases when more household members (over the age of 12) are present, when the individual owns a bicycle, and when the individual is reimbursed for using the car. The inclusion of cycling in the commuting mode choice set is affected by different determinants compared to walking. Walking is affected by household characteristics (like in Chapter 3), but also by ownership and employment conditions. Cycling is affected by individual characteristics, urban density, ownership, and employment conditions (largely overlapping with the findings in Chapter 3) *(answer Research Question 3)*. Furthermore, the inclusion of these modes in the choice set depends on more determinants than ownership and availability, which are generally used to identify the feasible choice set.

Determinants of Cyclists' Route Choice Behaviour in a Cycling-rich Context

In this thesis, route choice behaviour of cyclists' is investigated in Amsterdam (the Netherlands), where cycling is a dominant mode of transport. The relevant determinants are categorised into network-related and context determinants. Concerning the network-related attributes, distance is valued negatively (Chapter 5). This is in line with findings elsewhere, however often the impact is higher elsewhere. This might be caused by the mixed land use of Amsterdam, which reduces the need to travel longer distances between destinations. Furthermore, the number of intersections is valued negatively and overlap between routes is valued positively (Chapter 5). The share of cycle path has a different impact depending on the choice set identification method used (Chapter 6). When using the experienced route choice set (coined data-driven path identification method (DDPI)) in the estimation of the route choice model, it is not found to be relevant. This method builds upon observed routes of individuals, consequently these routes are already optimised to a certain extent. It is likely that all routes include a relatively high level of cycle path, making them irrelevant for route choice. Furthermore, if cycle paths are largely absent from on a route, the street design is such that it does not induce negative impacts for cyclists. Regarding the context determinants, we found that in the morning peak hour distance is valued more negatively compared to other times of the day (Chapter 5). This might be due to scheduling constraints in the morning. Rain and darkness are not found to affect route choice. Furthermore, no significant relationship is found regarding whether the trip serves as access/egress to the train station versus a standalone trip *(answer Research Question 4)*.

The Added Value of the Experienced Route Choice Set

The added value of the experienced route choice set or DDPI method for route choice modelling is investigated in comparison to two frequently used choice set generation (CSG) algorithms: the breadth-first search on link-elimination (BFS-LE) and labelling algorithms. The DDPI method automatically captures all the chosen routes, because only observed routes are included in the choice set identification. The CSG algorithms generate routes that aim to capture the chosen routes. The success of these algorithms depends largely on the criteria used to generate routes (e.g. distance and/or share of cycle paths), the complexity of the network, and the quality of the network information that is available. If any of these criteria is insufficient, the generated resulting choice set that is generated does not reproduce the observed behaviour. In case of Amsterdam, these criteria could not be met, thus introducing issues concerning the CSG algorithms' ability to capture the chosen routes.

The choice sets from the DDPI method and CSG algorithms are used in a route choice model using the same network-related attributes. On the whole, the signs of the parameters were similar between methods. The only difference is the path-size factor, where the DDPI method produced a positive parameter and the other methods generate a negative parameter. A negative sign is expected based on literature. However, because all chosen routes are optimised to a certain extent it can occur that the most popular street segments are included in many routes (e.g. arterial streets), which results in a positive overlap factor. Consequently, this could be a characteristic of the DDPI method. Next to that, the parameter values for the DDPI method are lower than for the other methods, which is mostly due to the limited variability in the choice set. This makes it more difficult for the model to provide a clear idea of the importance of a determinant. This in turn results in lower elasticities and a lower model fit. The choice sets resulting from the CSG methods show a larger degree of variation per determinant, ensuring that the model is able to provide a clear weight for each determinant, which in turn results in higher explanatory power. To validate the models, out-of-sample data is used on the estimated models. The DDPI method has a very low performance (largely incorrect prediction of routes and low log-likelihood), which means that the DDPI method is not suitable for prediction.

In sum, when the dependent variables of the CSG algorithms are of insufficient quality (criteria, network complexity or network information availability), the DDPI method offers an advantage, as it does not depend on any of these issues. Furthermore, in model estimation it generally shows similar parameter signs and a similar importance hierarchy of attributes as CSG algorithms. This implies that the DDPI method can provide insights into the behavioural preferences of individuals, which is useful for quick insights or when no sufficient network information is available. However, the DDPI method cannot be used for prediction purposes, as the method cannot handle out-of-sample data *(answer Research Question 5).*

Suitable and Applicable Approaches for Simultaneously Modelling Mode and Route Choice

In this thesis mode and route choice are investigated individually to get a better understanding of the determinants that influence these choices, as often done in active mode research. However, several arguments can be made as to why multiple travel choice dimensions should be investigated simultaneously (Bhat, 1998). One of these arguments, which relates to mode and route choice, is when determinants are related to multiple choice dimensions. This is the case in, for example travel time/distance. Therefore, active mode and route choice should also be investigated simultaneously.

In this thesis, we started this investigation by performing a literature review on how previous research has handled the simultaneous modelling of multiple travel choices (Chapter 7). The focus of the literature review is on discrete choice models. Due to the limited number of studies that address simultaneous modelling of mode and route choice using discrete choice modelling (four studies), we broadened the scope to also include the other commonly investigated travel choice dimensions in the four-step and activity-based models: trip chaining, destination choice, and departure time choice.

The mode-route choice studies all date from after 2010, showing that we start moving towards integrated choice modelling in the context of active mode travel research. However, these studies all used very basic modelling structures, such as Multinomial Logit and Nested Logit (NL). The first assumes a fully simultaneous choice between mode and route, where each of the joint alternatives are independent. The second introduces correlation between modes, meaning that routes are substituted before modes when changes are introduced, which already increases the realism. However, when comparing this NL method with state-of-the-art on route choice modelling, one can clearly see issues as for example, overlap between routes is not accounted for in any way. Consequently, many steps need to be made in the current state-of-the-art on simultaneous modelling of mode and route choice.

In the literature review, also several more advanced modelling structures have been introduced that are used to simultaneous model multiple travel choice dimensions, such as Cross-Nested Logit, Probit, Mixed Logit (ML), discrete-continuous (not applicable to mode-route choice), and segmentation approaches. To simultaneously model mode and route choice, it is imperative that overlap between routes can be accounted for (e.g. via Path-Size Logit (Chapter 5 and 6)). Furthermore, substitution patterns can vary per person and per trip, as increasing evidence is presented that decision-making is heterogeneous (Chapter 7). Therefore, ideally the model structure incorporates a flexible correlation structure and is able to account for heterogeneity in the decision-making process. Only the segmentation approaches are currently able to incorporate the latter. A combination of these segmentation approaches with ML or Probit could increase the behavioural realism of the modelled choice dimensions *(answer Research Question 6).* A downside of more behaviourally realistic models is that they are less applicable in practice, as increased complexity means reduced interpretability. This reduces the likelihood that these models are adopted in practice.

8.2. Discussion

In this section we discuss advantages and limitations of the methodological decisions undertaken in this thesis and what impact they have on our findings. The decisions that are discussed are the experienced choice set, trip-level choice behaviour, discrete choice modelling, and single travel choice research.

The Experienced Choice Set

In this thesis, we have studied two applications of the experienced choice set. Identifying determinants that explain the experienced mode choice set (Chapter 4) and using the experience route choice set, coined the data-driven path identification method (DDPI), to estimate a route choice model (Chapter 5 and 6). These two applications show several advantages and limitations for choice modelling in comparison to other CSG methods. Advantages are that the experienced choice set is less prone to falsely include or exclude alternatives and that it is less depended on other sources compared to other CSG methods. Limitations are the issue with endogeneity and bias and additional requirements in data collection compared to other methods.

Frequently employed methods to identify the choice set are prone to misspecifications, e.g. falsely excluding or including alternatives. In case of mode choice, we found for example that not owning a car does not mean not using it, however heuristic based methods might exclude this alternative from the choice set. Furthermore, in case of route choice, if irrelevant criteria are used in the CSG alternatives might be falsely included in the choice set. These misspecifications have implications for model estimation and prediction. The experienced choice set, on the other hand, is based on (revealed) behaviour, reducing the probability of misspecification, as all chosen alternatives are per definition included and non-chosen alternatives (for that trip) are likely to be considered for that trip.

CSG methods rely, next to the collected data, on other sources for generating the choice set. In route choice these sources are related to the network (complexity and quality) and optimisation criteria, for mode choice these sources are often related to ownership and trip length/duration. The experienced choice set does not depend so heavily on other data sources, consequently it is able to provide insights even when other sources are unavailable or of insufficient quality.

One of the limitations of the experienced choice set is that because it is based on chosen alternatives, endogeneity is introduced. This issue is not present in CSG methods, given that non-generated but chosen alternatives are not included in the choice set. We apply the experienced choice set in two ways in discrete choice models, first to estimate the choice set and second to estimate the choice. The issue with endogeneity mostly affects the second application. We argue that the severity of endogeneity would reduce if more observations per person/trip/OD pair are made. However, this issue needs to be considered, as we show that it does indeed impact the model estimation and prediction. Therefore, it affects the usefulness of the second application of the experienced choice set.

When estimating a discrete choice model with a subset of the universal set, per definition a bias is introduced (both in case of CSG and experienced choice set). In case of the two applications of the experienced choice set, bias is introduced differently. In case of estimating the choice set, the potentially introduced bias is directly modelled. Therefore, the bias is included more consciously. In case of estimating the choice using the experienced choice set, bias is a similar issue as in CSG. However, the effect of bias of different choice sets on the modelling outcomes has not been investigated in this thesis.

Finally, applying the experienced choice set imposes additional requirements on the data collection. There are multiple methods available that can result in this dataset. In this thesis two different methods were applied. First, in the route choice context we observed individuals over a longer period of time (one week) using the GPS-traces from the mobile phone. Second, in the mode choice context we asked individuals about the transport modes individuals used for different trip purposes over a longer period of time (half a year). The latter data gathering methodology is less imposing on the individual's privacy and requires less effort, but potentially more prone to contain errors, e.g. in forgetting alternatives or reporting socially desirable behaviour. The first requires that the individual is followed over a sufficiently long period of time, as multiple alternatives need to be observed. In case of route choice, this means that multiple routes have to be observed per origin-destination pair and ideally multiple observations per route are found. If this is not the case, severe loss of data could be the result. When using CSG methods, these additional requirements in data collection are not imposed.

These findings lead to two conclusions. First, using the experienced choice set to estimate the choice set instead of the choice, avoids many of the issues that are related to the latter (endogeneity, bias). Therefore, the choice set application shows higher potential compare to the choice application. Second, in case of applying the experienced choice set in the choice context, it shows added value compared to CSG when limited of insufficient network information is available, as it will help explain behaviour. However, it cannot be used for prediction. Consequently, if sufficient network information is available, it might be beneficial to use another method. Examples are CSG that are optimised for different criteria, sampling methods or a link-based method.

Trip-level Choice Behaviour

The two best known frameworks in transport demand modelling are the four-step model and the activity-based approaches. The four-step model is a trip-based approach which investigates the production and attraction of different zones over time, for different modes and their impact on the network (route choice). In this approach, each of the trips is regarded as independent from the next or previous trip. The activity-based approach suggests that travel only occurs when an activity is performed at a different location, and thus has a strong emphasis on the relationships between activities. This results in viewing the entire agenda of individuals (in tours/days) to understand behaviour, in comparison to viewing one trip at a time to understand travel behaviour.

This thesis investigates mode and route choice behaviour of active mode users using data from the Netherlands. In the Netherlands, the four-step model is applied on a national level (van Cranenburgh and Chorus, 2017; Hofman, 2002). It is imperative to have a match with the modelling framework, to be able to use the findings of this thesis in practice. Therefore, we investigate the mode and route choice behaviour of active mode users on a trip-level.

Several findings in this thesis suggest that adopting the activity-based approach could have produced more insights into the relation between trips, which cannot be captured given the current approach. Chapter 2 investigated the daily mobility patterns of individuals, where we show that for the majority of individuals a variety of modes is used over the day. This multimodal behaviour over the day is not captured when investigating individual trips. Furthermore, in Chapter 3 the previous use of modes largely influenced the current mode choice decision, which illustrates that one trip is related to the previous and the next trip. Thus, active mode research should adopt the activity-based approach, given that no practical implementation issues arise. A related PhD research of Florian Schneider (see Allegro-program, section 1.5) investigates how trips are organised in activity travel patterns. In that way his research provides insights into how active mode travel patterns are different from motorised modes.

Discrete Choice Modelling

In this thesis the discrete choice modelling framework is employed for investigating cyclists' route choice and active mode choice. In recent years, machine learning methods (supervised, unsupervised, and reinforcement learning) have been increasingly applied to research travel behaviour. Both in route choice (albeit not related to active modes) and mode choice these methods have been applied (Hagenauer and Helbich, 2017; Park et al., 2007; Wang and Ross, 2018). Both discrete choice modelling and machine learning have advantages and limitations for the investigation active mode choice behaviour.

Discrete choice modelling is a theory-driven approach, where the researcher imposes the model structure on the data (rich in assumptions), resulting with a model that can be properly interpret. The model produces parameter values, confidence intervals, and statistical measures that can test the model fit on the data. Furthermore, discrete choice models are transparent with respect to how the data is transformed from in- to output (i.e. no black box). On the other hand, machine learning is a data-driven approach, which does not require a priori model structures as it constructs the model structure from the data (very flexible). This means that there is no clear mathematical interpretation of the model in terms of how the data goes from in- to output. Machine learning models often function like a black box. Consequently, when the goal of the research is to understand and interpret the behaviour of decision-makers, discrete choice modelling is very suitable. When the goal of the research is to predict, machine learning is more suitable, because it does not rely on a priori assumptions on the model structure and the data. Furthermore, machine learning is (theoretically) able to achieve a higher performance in prediction. Besides these differences in suitability for reaching different goals, they also differ in the amount of data that is required. Discrete choice models do not require very large amounts of data, whereas in machine learning more data is better.

The data sources employed in this thesis are not sufficient for machine learning applications. Furthermore, the goal of this thesis is to better understand mode and route choice behaviour of active mode users, consequently discrete choice modelling is employed in this thesis. If more data is available, active mode research could benefit from machine learning, especially when advantages of both methods are combined.

Single Travel Choice Research

In this thesis, mode and route choice behaviour of active mode users have been investigated separately. The main reason for doing so is related the lack of knowledge in research regarding these two choices. Active mode research is still in the starting phase, which means that the emphasis lies with understanding the individual travel choices of active mode users. Consequently, the identified research gaps, related to determinants of influence and choice set formation, need to be investigated first. We argue that one first needs to understand individual travel choices before jointly modelling multiple travel choices.

Several arguments can be provided as to why travel choices should be modelled jointly (Bhat, 1998a). First, when the relevant choice alternatives are a combination of multiple travel choices. Second, when the observed determinants that influence the behaviour are related to multiple travel choices. Third, when the joint alternatives share unobserved determinants that influence the sensitivity of individuals to changes related to policy measures. This thesis shows that for example, travel time/distance influences both mode and route choice. Consequently, it seems compelling to jointly investigate these choices.

In this thesis, a first step towards joint modelling of mode and route choice is provided, by investigating existing literature considering discrete choice modelling structures that can model multiple travel choice dimensions simultaneously (Chapter 7). However, several theoretical developments are required regarding discrete choice modelling structures, before mode and route choice behaviour of active mode users can be jointly modelled.

8.3. Implications for Practice

The research presented in this thesis has two main implications for practice. First, evidence is found that walking and cycling are affected by different determinants and to a different extent, resulting in the independent consideration of these modes (Chapter 2, 3, and 4). This has implications for transportation planners and policy makers that are concerned with active mode transport. Second, we show that individuals only use a limited set of modes throughout the day or for specific purposes, often even using a single mode (Chapter 2 and 4). This is not (fully) in line with current practice, resulting with implications for practice.

Walking ≠ Cycling

Governments worldwide have set goals to increase the active mode share. In their plans to promote active modes, walking and cycling are often jointly addressed. In this thesis, we show that walking and cycling are to be considered independently (Chapter 2, 3, and 4). Cycling and walking are mostly affected by different determinants and in the few cases where they are affected by the same determinants, their impact differs (Chapter 3 and 4). Furthermore, no correlation was found between the mode choice alternatives walking and cycling, suggesting independence (Chapter 3 and 4). Moreover, this thesis shows that these two active modes of transport are also used differently. For instance, while the bicycle competes with the car, walking is complementary with the car (Chapter 3). Finally, often only one active mode is included in the daily mobility pattern of individuals (Chapter 2). These findings suggest that cycling and walking are to be considered independently.

These findings suggest that policies should be targeting either walking or cycling instead of both active modes of transport, in order to achieve the desired impact. If these modes are targeted in combination, undesired effects could arise. Different determinants need to be targeted to change the share of walking versus the share of cycling. Also, the potential for increasing active mode use or switching to active mode use varies largely across individuals. This thesis shows that individuals that are more car-oriented require different influencing strategies compared to individuals that already use active modes.

To derive more accurate insights regarding the walking and cycling mode share and to allow for the calculation of the effect of policy measures for walking and cycling separately, both (and independently) active modes should be incorporated in the transport planning models. It is essential that the mode choice models incorporate different factors for walking, mostly trip related, household related, and built environment, compared to cycling, mostly trip related, employment related, and built environment related.

Limited Number of Modes Used per Individual

This thesis shows that individuals generally use a limited set of modes. The majority of the individuals use only one or two modes in their daily mobility pattern (car and bicycle). Furthermore, a limited number of modes is observed in the experienced mode set of individuals, which targets specific trip purposes. Most people use the car to go to work, followed by the bicycle, and a very limited number of individuals uses multiple modes for the same trip purpose. Finally, many individuals have access to a small number of modes (limitations mostly show in car and bicycle). These findings provide implications regarding 1) the potential effect of policies that target an increase in walking and cycling mode shares and 2) the identification of the choice set in transportation planning models that are used to estimate the effect of policy measures.

First, a limited amount of modes is experienced by the individual, which means that it will be harder to change the behaviour of individuals. These individuals might not be aware of the level-of-service and other characteristics other modes are offering. In particular, this holds when the individuals are content with their mode choices (consonant travellers). When these individuals are not content (dissonant travellers), it will be easier to change the behaviour (Chapter 2). This means that they will be more open to explore different modes. As an example, within the group of exclusive car users, a relatively large share is unsatisfied with the car and very positive towards the bicycle, showing potential for change. However, some of these individuals are captive users, and can therefore not change to other modes due to financial reasons or other requirements (e.g. ownership). Consequently, the effectiveness of policy measures that target a switch towards walking or cycling depends on the group of individuals that is targeted.

Second, a limited amount of modes in the experienced choice set means that the specification of the mode choice set requires special attention. In practice, currently the specification of this set often does not receive much attention. Often, no restrictions or only availability restrictions are imposed. This thesis shows that mode availability is not the same as mode adoption (Chapter 4). Consequently, contemporary mode choice models, and the resulting policy measures, could potentially be overestimating the impact on changes in mode choice. Consequently, more refined methods should be employed in practice to generate the choice set. These methods should go beyond availability and ownership, but also consider the impact of for example, socio-demographics, urban density, and employment conditions.

8.4. Recommendations for Future Research

This thesis investigated mode and route choice behaviour of active mode users. While doing this research several new research directions arose that are not addressed in this thesis. These research directions are divided into three topics, namely active mode and route choice (8.4.1), other active mode related choices (8.4.2), and methodological advancements (8.4.3).

8.4.1. Advancements in Active Mode and Route Choice Research

We are currently still in the starting phase of research featuring mode and route choice behaviour of active mode users. The importance of investigating these choices has been acknowledged worldwide and the body of research that focusses on active mode and route choice is growing rapidly. Three research directions are identified and elaborated upon below, that can advance the work presented in this thesis. First, route choice research should feature walking, as it is also considered a main (full-fledged) mode in the urban environment. Second, active mode and route choice behaviour should be studied simultaneously. Third, the impact of smartphones (and their information) should be investigated for active mode and route choice.

Pedestrian Route Choice

This thesis only covers the route choice behaviour of cyclists due to two reasons: 1) increased interest of governments worldwide concerning cycling, and 2) availability and potential of collecting detailed data on cyclists' route choice. We show that cycling and walking should be considered independently (Chapter 2, 3, and 4), consequently we expect that the route choice of pedestrians is also affected by different determinants and to a different extent than the route choice of cyclists. Therefore, it would also be interesting to investigate the route choice behaviour of pedestrians. To perform this type of research, it is necessary that the large-scale data collection efforts that are currently devoted to cyclists, also take place for pedestrians. For example, a pedestrian counting week could be organised that tracks the routes of pedestrians. The information collected can then be used to identify the determinants of pedestrian route choice, where a comparison with other modes can be made. Furthermore, pedestrian route choice models can be integrated with the route choice models of other modes in transport planning models, so as to understand impact on the network of all modes combined.

Integration of Route and Mode Choice Models

In this thesis, we identified the need of integrating mode and route choice into a single model. In Chapter 7, potential suitable modelling approaches for the simultaneous modelling of multiple travel choices are identified. The findings of that review can serve as input for the development of a simultaneous integrated mode and route choice model. To achieve this integrated model, also the non-trivial task of CSG needs to be performed (see for example Chapter 4 and 6). Route choice sets need to also be generated for all modes. Car, bicycle, and walking are private modes that have no external dependencies (such as timetables), thus it might be possible to employ similar generation techniques for each mode, given that suitable criteria are employed for each mode. Public transport, on the other hand, is dependent on timetables with (generally still) fixed lines, which might require a different approach towards the generation of the route choice set. Furthermore, the mode choice set needs to be specified. The experienced choice set could be employed here (like in Chapter 4). In sum, several methodological issues (i.e. modelling approach and CSG) need to be investigated before a useful simultaneous mode and route choice model can be developed.

Influence of ICT on Active Mode and Route Choice Behaviour

In this thesis we investigated how individuals are influenced in their decision-making by several determinants, related to the characteristics of individuals, their surroundings, and their trips. However, in the current era, smartphones are taking up increasingly important roles. Also, regarding mode and route choice behaviour, increasingly information is gathered via smartphones. It is unknown to what extent and how smartphone use influences the choice behaviour of active mode users. Furthermore, it is unknown to what extent familiarity with the environment influences this information acquisition process. More research is required to fully understand this relationship.

8.4.2. Advancements in Other Active Mode Related Travel Choices

Next to mode and route choice, other travel choices, such as departure time and activity related choices, are relevant for walking and cycling. No knowledge is currently present as to what extent this behaviour is different for active modes compared to motorised modes.

Other travel decisions, such as departure time and activity related choices, are currently often excluded from the scope in active mode research. One can, however, imagine situations where these decisions are highly relevant. For example, departure time choice is increasingly relevant in environments where active modes are dominantly present. Bicycle traffic jams are already occurring in the Netherlands. Consequently, a better understanding of the departure time of cyclists and potentially changing it, could help resolve these issues. Furthermore, the activity scheduling of active mode users might differ significantly from the scheduling of motorised modes. In the literature, differences between scheduling of trips using public transport and the car have been studied (e.g. Hensher and Reyes, 2000; Ye et al., 2007), however this is not yet extended to involve active modes. Even though this is also a relevant topic for active modes.

8.4.3. Methodological Advancements

Several methodological aspects related to discrete choice modelling were addressed in this thesis, for example, the development of data-driven choice set formation (Chapter 4 and 6). First steps were taken towards understanding the added value of these methodological aspects. Notwithstanding, further research is needed in relation to the following three topics, which are elaborated upon below. First, it is yet unknown what a long enough observation period is to capture the considered choice set and what it depends on. Second, the potential of integration of the choice set and choice model (probabilistically) has not been identified. Third, heterogeneity in discrete choice models should include the decision-making process to capture choice behaviour more behaviourally realistic.

Observation Period for Experienced Choice Set

The experienced choice set uses observed choices to identify the choice set. In this thesis, the experienced route choice set is based on observations from one-week (Chapter 5 and 6), whereas the experienced mode choice set is based on recalled modes used in the course of half a year (Chapter 4). We hypothesise that when the choices have been observed over a sufficiently long period of time, this choice set approaches the considered choice set. What is currently unknown is what is a sufficiently long period of observation. This can be investigated through a data collection effort that tracks individuals for a long time with GPS, for example half a year. By splitting the observations in time periods (one week, two weeks, three weeks, etc), one can investigate when results become stable and could thus have approached the considered choice set. Ideally, this data collection would take place in different environments as the complexity of the network and variety of mobility patterns will likely influence the satiation time.

Integration of Choice Set and Choice Model

This thesis investigated which determinants influence the experienced mode choice set, which reflects one aspect of the probabilistic method proposed by Manski (1977). This probabilistic method consists of a deterministic identification of feasible choice sets, a probabilistic choice set identification that estimates the probability that the choice set is the considered set (experienced mode choice set) from these feasible sets, and mode choice depended on the considered set. It would be very interesting to take the research of the experienced mode choice set to a next level by integrating it with a mode choice model, similar to the approach proposed by Manski (1977). The simultaneous modelling of the mode choice set and mode choice provides an interesting comparison to current mode choice research. The idea would be that this simultaneous modelling approach provides better model estimates and consequently, provides better predictions compared to frequently employed methods for mode choice modelling. The main concern regarding practical implementation would be the computational effort of such a model, as this will increase with more alternatives and thus choice sets.

Heterogeneity in Decision-making

This thesis provides evidence that heterogeneity should be accounted for in active mode modelling on different levels. Heterogeneity arises in the preferences of individuals towards walking and cycling (taste variation), which results in different attitudes towards modes. Furthermore, the importance of various determinants, such as travel time, varies over individuals regarding active mode choice. Based on Chapter 7, we expect heterogeneity in the decision-making processes of individuals regarding route and mode choice. However, the current discrete choice modelling approaches are largely unable to accommodate heterogeneity in the decision-making process.

References

Abdel-Aty, M.A., Kitamura, R., Jovanis, P.P., 1997. Using stated preference data for studying the effect of advanced traffic information on drivers' route choice. Transp. Res. Part C Emerg. Technol. 5, 39–50.

Ajzen, I., 1991. The theory of planned behavior. Organ. Behav. Hum. Decis. Process. 50, 179–211.

Ajzen, I., Fishbein, M., 1977. Attitude-behavior relations: A theoretical analysis and review of empirical research. Psychol. Bull. 84, 888.

Akar, G., Clifton, K., 2009. Influence of individual perceptions and bicycle infrastructure on decision to bike. Transp. Res. Rec. J. Transp. Res. Board 165–172.

Akgün, V., Erkut, E., Batta, R., 2000. On finding dissimilar paths. Eur. J. Oper. Res. 121, 232–246.

Anable, J., Gatersleben, B., 2005. All work and no play? The role of instrumental and affective factors in work and leisure journeys by different travel modes. Transp. Res. Part A Policy Pract. 39, 163–181.

Azevedo, J., Costa, M.E.O.S., Madeira, J.J.E.R.S., Martins, E.Q.V., 1993. An algorithm for the ranking of shortest paths. Eur. J. Oper. Res. 69, 97–106.

Bajwa, S., Bekhor, S., Kuwahara, M., Chung, E., 2008. Discrete choice modeling of combined mode and departure time. Transportmetrica 4, 155–177.

Bamberg, S., Ajzen, I., Schmidt, P., 2003. Choice of travel mode in the theory of planned behavior: The roles of past behavior, habit, and reasoned action. Basic Appl. Soc. Psych. 25, 175–187.

Barros, A.P., Martinez, L.M., Viegas, J.M., 2015. A new approach to understand modal and pedestrians route in Portugal. Transp. Res. Procedia 10, 860–869.

Bekhor, S., Ben-Akiva, M.E., Ramming, M.S., 2006. Evaluation of choice set generation algorithms for route choice models. Ann. Oper. Res. 144, 235–247.

Bellman, R., Kalaba, R., 1960. On k th best policies. J. Soc. Ind. Appl. Math. 8, 582–588.

Ben-Akiva, M., Bergman, M.J., Daly, A.J., Ramaswamy, R., 1984. Modeling inter-urban route choice behaviour, in: Proceedings of the 9th International Symposium on Transportation and Traffic Theory. pp. 299–330.

Ben-Akiva, M., Bierlaire, M., 1999. Discrete choice methods and their applications to short term travel decisions, in: Handbook of Transportation Science. Springer, pp. 5–33.

Ben-Akiva, M., Boccara, B., 1995. Discrete choice models with latent choice sets. Intern. J. Res. Mark. 12, 9–24.

Ben-Akiva, M., McFadden, D., Gärling, T., Gopinath, D., Walker, J., Bolduc, D., Börsch-Supan, A., Delquié, P., Larichev, O., Morikawa, T., others, 1999. Extended framework for modeling choice behavior. Mark. Lett. 10, 187–203.

Ben-Akiva, M.E., Bowman, J.L., 1998. Activity based travel demand model systems, in: Equilibrium and Advanced Transportation Modelling. Springer, pp. 27–46.

Bernardi, S., La Paix-Puello, L., Geurs, K., 2018. Modelling route choice of Dutch cyclists using smartphone data. J. Transp. Land Use 11.

Beser, M., Algers, S., 2002. SAMPERS—The new Swedish national travel demand forecasting tool, in: National Transport Models. Springer, pp. 101–118.

Bhat, C.R., 2005. A multiple discrete--continuous extreme value model: formulation and application to discretionary time-use decisions. Transp. Res. Part B Methodol. 39, 679–707.

Bhat, C.R., 1998a. Accommodating flexible substitution patterns in multi-dimensional choice modeling: formulation and application to travel mode and departure time choice. Transp. Res. Part B Methodol. 32, 455–466.

Bhat, C.R., 1998b. Analysis of travel mode and departure time choice for urban shopping trips. Transp. Res. Part B Methodol. 32, 361–371.

Bhat, C.R., 1997. Work travel mode choice and number of non-work commute stops. Transp. Res. Part B Methodol. 31, 41–54.

Bhat, C.R., 1995. A heteroscedastic extreme value model of intercity travel mode choice. Transp. Res. Part B Methodol. 29, 471–483.

Bhat, C.R., Eluru, N., 2009. A copula-based approach to accommodate residential self-selection effects in travel behavior modeling. Transp. Res. Part B Methodol. 43, 749–765.

Bhat, C.R., Steed, J.L., 2002. A continuous-time model of departure time choice for urban shopping trips. Transp. Res. Part B Methodol. 36, 207–224.

Bierlaire, M., 2016. PythonBiogeme: a short introduction.

Bierlaire, M., 2006. A theoretical analysis of the cross-nested logit model. Ann. Oper. Res. 144, 287–300.

Bierlaire, M., 2003. BIOGEME: a free package for the estimation of discrete choice models.

Bliemer, M., Bovy, P., 2008. Impact of route choice set on route choice probabilities. Transp. Res. Rec. J. Transp. Res. Board 10–19.

Bliemer, M.C.J., Versteegt, H.H., Castenmiller, R.J., 2004. INDY: a new analytical multiclass dynamic traffic assignment model, in: Proceedings of the TRISTAN V Conference, Guadeloupe.

Böcker, L., Dijst, M., Prillwitz, J., 2013. Impact of everyday weather on individual daily travel behaviours in perspective: a literature review. Transp. Rev. 33, 71–91.

Borgers, A., Timmermans, H., 1986. A model of pedestrian route choice and demand for retail facilities within inner-city shopping areas. Geogr. Anal. 18, 115–128.

Börjesson, M., 2008. Joint RP--SP data in a mixed logit analysis of trip timing decisions. Transp. Res. Part E Logist. Transp. Rev. 44, 1025–1038.

Bovy, P.H., Stern, E., 2012. Route Choice: Wayfinding in Transport Networks: Wayfinding in Transport Networks. Springer Science & Business Media.

Bovy, P.H.L., 2009. On modelling route choice sets in transportation networks: a synthesis. Transp. Rev. 29, 43–68.

Bovy, P.H.L., Fiorenzo-Catalano, S., 2007. Stochastic route choice set generation: behavioral and probabilistic foundations. Transportmetrica 3, 173–189.

Bowman, J.L., Bradley, M., Shiftan, Y., Lawton, T.K., Ben-Akiva, M.E., 1998. Demonstration of an activity based model system for Portland, in: 8th World Conference on Transport Research, Antwerp, Belgium.

Braun, L.M., Rodriguez, D.A., Cole-Hunter, T., Ambros, A., Donaire-Gonzalez, D., Jerrett, M., Mendez, M.A., Nieuwenhuijsen, M.J., de Nazelle, A., 2016. Short-term planning and policy interventions to promote cycling in urban centers: Findings from a commute mode choice analysis in Barcelona, Spain. Transp. Res. Part A Policy Pract. 89, 164–183.

Broach, J., Dill, J., 2016. Using predicted bicyclist and pedestrian route choice to enhance mode choice models. Transp. Res. Rec. 2564, 52–59.

Broach, J., Dill, J., Gliebe, J., 2012. Where do cyclists ride? A route choice model developed with revealed preference GPS data. Transp. Res. Part A Policy Pract. 46, 1730–1740.

Broach, J., Gliebe, J., Dill, J., 2010. Calibrated labeling method for generating bicyclist route choice sets incorporating unbiased attribute variation. Transp. Res. Rec. J. Transp. Res. Board 89–97.

Buehler, R., 2011. Determinants of transport mode choice: a comparison of Germany and the USA. J. Transp. Geogr. 19, 644–657.

Buehler, R., Dill, J., 2016. Bikeway networks: A review of effects on cycling. Transp. Rev. 36, 9–27.

Calastri, C., Hess, S., Choudhury, C., Daly, A., Gabrielli, L., 2017. Mode choice with latent availability and consideration: theory and a case study. Transp. Res. Part B Methodol.

Cantillo, V., de Dios Ortúzar, J., 2005. A semi-compensatory discrete choice model with explicit attribute thresholds of perception. Transp. Res. Part B Methodol. 39, 641–657.

Cascetta, E., Papola, A., 2001. Random utility models with implicit availability/perception of choice alternatives for the simulation of travel demand. Transp. Res. Part C Emerg. Technol. 9, 249–263.

Casello, J., Usyukov, V., 2014. Modeling cyclists' route choice based on GPS data. Transp. Res. Rec. J. Transp. Res. Board 155–161.

Catalano, S.F., der Zijpp, N., 2001. A forecasting model for inland navigation based on route enumeration, in: Proceedings of the AET European Transport Conference, Held 10-12 Septermber, 2001, Homerton College, Cambridge, UK.

CBS, 2018. Personenmobiliteit in Nederland [WWW Document]. Statline, http//statline.cbs.nl [in Dutch].

Chen, P., Shen, Q., Childress, S., 2018. A GPS data-based analysis of built environment influences on bicyclist route preferences. Int. J. Sustain. Transp. 12, 218–231.

Chorus, C.G., Kroesen, M., 2014. On the (im-) possibility of deriving transport policy implications from hybrid choice models. Transp. Policy 36, 217–222.

CWANZ, 2018. Cycling and Walking Australia and New Zealand [WWW Document]. Overv. https//www.transport.wa.gov.au/mediaFiles/active-transport/AT_CYC_P_CWANZ_Overview.pdf.

de Dios Ortúzar, J., Willumsen, L.G., 2011. Modelling transport. John Wiley & Sons.

de Haas, M., Scheepers, E., Hoogendoorn-Lanser, S., 2017. Identifying different types of observed immobility within longitudinal travel surveys, in: 11th International Conference on Transport Survey Methods, Esterel, Canada.

de Haas, M.C., Scheepers, C.E., Harms, L.W.J., Kroesen, M., 2018. Travel pattern transitions: Applying latent transition analysis within the mobility biographies framework. Transp. Res. Part A Policy Pract. 107, 140–151.

De Jong, G., Daly, A., Pieters, M., Miller, S., Plasmeijer, R., Hofman, F., 2007. Uncertainty in traffic forecasts: literature review and new results for The Netherlands. Transportation (Amst). 34, 375–395.

De Jong, G., Daly, A., Pieters, M., Vellay, C., Bradley, M., Hofman, F., 2003. A model for time of day and mode choice using error components logit. Transp. Res. Part E Logist. Transp. Rev. 39, 245–268.

De Vos, J., 2018. Do people travel with their preferred travel mode? Analysing the extent of travel mode dissonance and its effect on travel satisfaction. Transp. Res. part A policy Pract. 117, 261–274.

Debrezion, G., Pels, E., Rietveld, P., 2009. Modelling the joint access mode and railway station choice. Transp. Res. Part E Logist. Transp. Rev. 45, 270–283.

dell'Olio, L., Ibeas, A., Bordagaray, M., Ortúzar, J. de D., 2013. Modeling the effects of pro bicycle infrastructure and policies toward sustainable urban mobility. J. Urban Plan. Dev. 140, 4014001.

Department for Transport, 2017. Cycling and Walking Investment Strategy [WWW Document]. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/603527/cycling-walking-investment-strategy.pdf.

Dhakar, N., Srinivasan, S., 2014. Route choice modeling using GPS-based travel surveys. Transp. Res. Rec. J. Transp. Res. Board 65–73.

Dial, R.B., 2000. Bicriterion traffic equilibrium: T2 model, algorithm, and software overview. Transp. Res. Rec. 1725, 54–62.

Diana, M., Mokhtarian, P.L., 2009. Desire to change one's multimodality and its relationship to the use of different transport means. Transp. Res. Part F Traffic Psychol. Behav. 12, 107–119.

Dijkstra, E.W., 1959. A note on two problems in connexion with graphs. Numer. Math. 1, 269–271.

Ding, C., Xie, B., Wang, Y., Lin, Y., 2014. Modeling the joint choice decisions on urban shopping destination and travel-to-shop mode: a comparative study of different structures. Discret. Dyn. Nat. Soc. 2014.

Duives, D.C., Galama, I., Ton, D., Schneider, F., Daamen, W., Cats, O., Hoogendoorn, S.P., 2018. Independence and sequentiality of cyclists' travel choices: review of determinants of cyclists' mode, route and operational choices (Working paper).

ECF, 2019. European Cyclists' Federation [WWW Document]. Natl. Cycl. policies, https//ecf.com/what-we-do/cycling-all-policies/national-cycling-policies.

Ehrgott, M., Wang, J.Y.T., Raith, A., Van Houtte, C., 2012. A bi-objective cyclist route choice model. Transp. Res. part A policy Pract. 46, 652–663.

Eluru, N., Pinjari, A., Pendyala, R., Bhat, C., 2010. An econometric multi-dimensional choice model of activity-travel behavior. Transp. Lett. 2, 217–230.

Emond, C.R., Handy, S.L., 2012. Factors associated with bicycling to high school: insights from Davis, CA. J. Transp. Geogr. 20, 71–79.

Fernández-Heredia, Á., Monzón, A., Jara-Diaz, S., 2014. Understanding cyclists' perceptions, keys for a successful bicycle promotion. Transp. Res. part A policy Pract. 63, 1–11.

Field, A., 2009. Discopering Statistics Using SPSS, Thrid Edition.

FietsTelweek, 2015. Regiobrochure Nationale Fiets Telweek 2015 [in Dutch].

Fiorenzo-Catalano, S., 2007. Choice set generation in multi-modal transport networks. Delft Univ. Technol. PhD Thesis, TRAIL, Netherlands.

Fishman, E., 2016. Cycling as transport. Transp. Rev. 36, 1–8.

Flötteröd, G., Bierlaire, M., 2013. Metropolis--Hastings sampling of paths. Transp. Res. Part B Methodol. 48, 53–66.

Fosgerau, M., Frejinger, E., Karlstrom, A., 2013. A link based network route choice model with unrestricted choice set. Transp. Res. Part B Methodol. 56, 70–80.

Fraser, S.D.S., Lock, K., 2011. Cycling for transport and public health: a systematic review of the effect of the environment on cycling. Eur. J. Public Health 21, 738–743.

Frejinger, E., 2007. Random sampling of alternatives in a route choice context, in: European Transport Conference.

Frejinger, E., Bierlaire, M., 2007. Capturing correlation with subnetworks in route choice models. Transp. Res. Part B Methodol. 41, 363–378.

Frejinger, E., Bierlaire, M., Ben-Akiva, M., 2009. Sampling of alternatives for route choice modeling. Transp. Res. Part B Methodol. 43, 984–994.

Friedrich, M., Hofsaess, I., Wekeck, S., 2001. Timetable-based transit assignment using branch and bound techniques. Transp. Res. Rec. 1752, 100–107.

Furth, P.G., 2012. Bicycling infrastructure for mass cycling: a trans-Atlantic comparison. City Cycl. 38, 105–140.

Gärling, T., Gillholm, R., Gärling, A., 1998. Reintroducing attitude theory in travel behavior research: The validity of an interactive interview procedure to predict car use. Transportation (Amst). 25, 129–146.

Gehrke, S.R., Clifton, K.J., 2014. Operationalizing land use diversity at varying geographic scales and its connection to mode choice: Evidence from Portland, Oregon. Transp. Res. Rec. 2453, 128–136.

Ghanayim, M., Bekhor, S., 2018. Modelling bicycle route choice using data from a GPS-assisted household survey. Eur. J. Transp. Infrastruct. Res. 18.

González, F., Melo-Riquelme, C., de Grange, L., 2016. A combined destination and route choice model for a bicycle sharing system. Transportation (Amst). 43, 407–423.

Greene, W.H., Hensher, D.A., 2003. A latent class model for discrete choice analysis: contrasts with mixed logit. Transp. Res. Part B Methodol. 37, 681–698.

Habib, K.M.N., 2013. A joint discrete-continuous model considering budget constraint for the continuous part: application in joint mode and departure time choice modelling. Transp. A Transp. Sci. 9, 149–177.

Habib, K.M.N., Day, N., Miller, E.J., 2009. An investigation of commuting trip timing and mode choice in the Greater Toronto Area: Application of a joint discrete-continuous model. Transp. Res. Part A Policy Pract. 43, 639–653.

Habib, K.M.N., Tian, Y., Zaman, H., 2011. Modelling commuting mode choice with explicit consideration of carpool in the choice set formation. Transportation (Amst). 38, 587–604.

Habib, K.N., Mann, J., Mahmoud, M., Weiss, A., 2014. Synopsis of bicycle demand in the City of Toronto: Investigating the effects of perception, consciousness and comfortability on the purpose of biking and bike ownership. Transp. Res. part A policy Pract. 70, 67–80.

Hagenauer, J., Helbich, M., 2017. A comparative study of machine learning classifiers for modeling travel mode choice. Expert Syst. Appl. 78, 273–282.

Halldórsdóttir, K., Rieser-Schüssler, N., Axhausen, K.W., Nielsen, O.A., Prato, C.G., 2014. Efficiency of choice set generation methods for bicycle routes. Eur. J. Transp. Infrastruct. Res. 14, 332–348.

Hamre, A., Buehler, R., 2014. Commuter mode choice and free car parking, public transportation benefits, showers/lockers, and bike parking at work: evidence from the Washington, DC region. J. Public Transp. 17, 4.

Handy, S., Van Wee, B., Kroesen, M., 2014. Promoting cycling for transport: research needs and challenges. Transp. Rev. 34, 4–24.

Hansen, K.B., Nielsen, T.A.S., 2014. Exploring characteristics and motives of long distance commuter cyclists. Transp. Policy 35, 57–63.

Harms, L., Bertolini, L., Brömmelstroet, M. Te, 2016. Performance of municipal cycling policies in medium-sized cities in the Netherlands since 2000. Transp. Rev. 36, 134–162.

Harrington, P., 2012. Machine learning in action. Manning Greenwich.

Heinen, E., Chatterjee, K., 2015. The same mode again? An exploration of mode choice variability in Great Britain using the National Travel Survey. Transp. Res. Part A Policy Pract. 78, 266–282.

Heinen, E., Maat, K., van Wee, B., 2013. The effect of work-related factors on the bicycle commute mode choice in the Netherlands. Transportation (Amst). 40, 23–43.

Heinen, E., Maat, K., Van Wee, B., 2011. The role of attitudes toward characteristics of bicycle commuting on the choice to cycle to work over various distances. Transp. Res. part D Transp. Environ. 16, 102–109.

Heinen, E., Van Wee, B., Maat, K., 2010. Commuting by bicycle: an overview of the literature. Transp. Rev. 30, 59–96.

Hensher, D.A., Ho, C.Q., 2016. Experience conditioning in commuter modal choice modelling--Does it make a difference? Transp. Res. Part E Logist. Transp. Rev. 95, 164–176.

Hensher, D.A., Reyes, A.J., 2000. Trip chaining as a barrier to the propensity to use public transport. Transportation (Amst). 27, 341–361.

Hess, S., 2014. 14 Latent class structures: taste heterogeneity and beyond, in: Handbook of Choice Modelling. Edward Elgar Publishing Cheltenham, pp. 311–329.

Hess, S., Bierlaire, M., Polak, J.W., 2005. Estimation of value of travel-time savings using mixed logit models. Transp. Res. Part A Policy Pract. 39, 221–236.

Hess, S., Daly, A., Rohr, C., Hyman, G., 2007a. On the development of time period and mode choice models for use in large scale modelling forecasting systems. Transp. Res. Part A Policy Pract. 41, 802–826.

Hess, S., Polak, J.W., Daly, A., Hyman, G., 2007b. Flexible substitution patterns in models of mode and time of day choice: new evidence from the UK and the Netherlands. Transportation (Amst). 34, 213–238.

Hess, S., Quddus, M., Rieser-Schüssler, N., Daly, A., 2015. Developing advanced route choice models for heavy goods vehicles using GPS data. Transp. Res. Part E Logist. Transp. Rev. 77, 29–44.

Hofman, F., 2002. Application areas for the Dutch national model, in: National Transport Models. Springer, pp. 81–92.

Hood, J., Sall, E., Charlton, B., 2011. A GPS-based bicycle route choice model for San Francisco, California. Transp. Lett. 3, 63–75.

Hoogendoorn-Lanser, S., 2005. Modelling travel behaviour in multi-modal networks. PhD dissertation. Delft University of Technology.

Hoogendoorn-Lanser, S., Schaap, N.T.W., OldeKalter, M.-J., 2015. The Netherlands Mobility Panel: An innovative design approach for web-based longitudinal travel data collection. Transp. Res. Procedia 11, 311–329.

Hoogendoorn, S.P., 2014. ALLEGRO - ERC Advanced Grant - Horizon 2020.

Hoogendoorn, S.P., Bovy, P.H.L., 2004. Pedestrian route-choice and activity scheduling theory and models. Transp. Res. Part B Methodol. 38, 169–190.

Howard, C., Burns, E., 2001. Cycling to work in Phoenix: Route choice, travel behavior, and commuter characteristics. Transp. Res. Rec. J. Transp. Res. Board 39–46.

Hunt, D.T., Kornhauser, A.L., 1996. Assigning traffic over essentially-least-cost paths. Transp. Res. Rec. 1556, 1–7.

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

Ishaq, R., Bekhor, S., Shiftan, Y., 2013. A flexible model structure approach for discrete choice models. Transportation (Amst). 40, 609–624. https://doi.org/10.1007/s11116-012-9431-8

Johnson, P.E., Joy, D.S., Clarke, D.B., Jacobi, J.M., 1993. HIGHWAY 3.1: An enhanced HIGHWAY routing model: Program description, methodology, and revised users manual.

Jou, R.-C., 2001. Modeling the impact of pre-trip information on commuter departure time and route choice. Transp. Res. Part B Methodol. 35, 887–902.

Kamargianni, M., 2015. Investigating next generation's cycling ridership to promote sustainable mobility in different types of cities. Res. Transp. Econ. 53, 45–55.

Kamargianni, M., Dubey, S., Polydoropoulou, A., Bhat, C., 2015. Investigating the subjective and objective factors influencing teenagers' school travel mode choice--An integrated choice and latent variable model. Transp. Res. Part A Policy Pract. 78, 473–488.

Kamargianni, M., Polydoropoulou, A., 2013. Hybrid choice model to investigate effects of teenagers' attitudes toward walking and cycling on mode choice behavior. Transp. Res. Rec. 2382, 151–161.

Khatri, R., Cherry, C.R., Nambisan, S.S., Han, L.D., 2016. Modeling route choice of utilitarian bikeshare users with GPS data. Transp. Res. Rec. 2587, 141–149.

KiM, 2015. Fietsen en lopen: de smeerolie van onze mobiliteit, in: Ministerie van Infrastructuur En Milieu [in Dutch].

Krizek, K., Waddell, P., 2002. Analysis of lifestyle choices: Neighborhood type, travel patterns, and activity participation. Transp. Res. Rec. J. Transp. Res. Board 119–128.

Kroesen, M., 2014. Modeling the behavioral determinants of travel behavior: An application of latent transition analysis. Transp. Res. Part A Policy Pract. 65, 56–67.

Kroesen, M., Chorus, C., 2018. The role of general and specific attitudes in predicting travel behavior--A fatal dilemma? Travel Behav. Soc. 10, 33–41.

Kroesen, M., Handy, S., Chorus, C., 2017. Do attitudes cause behavior or vice versa? An alternative conceptualization of the attitude-behavior relationship in travel behavior modeling. Transp. Res. Part A Policy Pract. 101, 190–202.

Krueger, R., Vij, A., Rashidi, T.H., 2018. Normative beliefs and modality styles: a latent class and latent variable model of travel behaviour. Transportation (Amst). 45, 789–825.

Krygsman, S., Arentze, T., Timmermans, H., 2007. Capturing tour mode and activity choice interdependencies: A co-evolutionary logit modelling approach. Transp. Res. Part A Policy Pract. 41, 913–933.

Kuester, F., 2015. National Cycling Policies. [WWW Document]. Eur. Cyclists' Fed. https//ecf.com/what-we-do/cycling-all-policies/national-cycling -policies.

Kuhnimhof, T., Chlond, B., Von Der Ruhren, S., 2006. Users of transport modes and multimodal travel behavior: Steps toward understanding travelers' options and choices. Transp. Res. Rec. 1985, 40–48.

Kuhnimhof, T.G., 2009. Measuring and modeling multimodal mode use in the longitudinal section, in: Proceedings of the 88th Annual TRB Meeting.

Lam, S.-H., Xie, F., 2002. Transit path-choice models that use revealed preference and stated preference data. Transp. Res. Rec. 1799, 58–65.

Lanzendorf, M., 2002. Mobility styles and travel behavior: Application of a lifestyle approach to leisure travel. Transp. Res. Rec. 1807, 163–173.

Lavery, T.A., Páez, A., Kanaroglou, P.S., 2013. Driving out of choices: An investigation of transport modality in a university sample. Transp. Res. part A policy Pract. 57, 37–46.

Lawler, E.L., 1976. Combinatorial optimization: Networks and matroids (Holt, Ed.). New York Rinehart Winst.

Lee, L.-F., 1983. Generalized econometric models with selectivity. Econom. J. Econom. Soc. 507–512.

Li, S., Muresan, M.I., Fu, L., 2017. Cycling in Toronto, Ontario, Canada: Route Choice Behavior and Implications for Infrastructure Planning. Transp. Res. Rec. J. Transp. Res. Board pp 41–49.

Li, Z., Wang, W., Yang, C., Jiang, G., 2013. Exploring the causal relationship between bicycle choice and trip chain pattern. Transp. Policy. https://doi.org/10.1016/j.tranpol.2013.06.001

Lindelöw, D., Svensson, Å., Sternudd, C., Johansson, M., 2014. What limits the pedestrian? Exploring perceptions of walking in the built environment and in the context of every-day life. J. Transp. Heal. 1, 223–231.

Lizana, P., Arellana, J., Ortúzar, J. de D., Rizzi, L., 2013. A joint mode/time-of-day choice model using combined revealed preference and stated choice data, in: Proceedings of the 13th World Conference of Transport Research.

Lombard, K., Church, R.L., 1993. The gateway shortest path problem: generating alternative routes for a corridor location problem. Geogr. Syst. 1, 25–45.

Ma, L., Dill, J., 2015. Associations between the objective and perceived built environment and bicycling for transportation. J. Transp. Heal. 2, 248–255.

Maldonado-Hinarejos, R., Sivakumar, A., Polak, J.W., 2014. Exploring the role of individual attitudes and perceptions in predicting the demand for cycling: a hybrid choice modelling approach. Transportation (Amst). 41, 1287–1304.

Maley, D.W., Weinberger, R.R., 2011. Food shopping in the urban environment: Parking supply, destination choice, and mode choice, in: 90th Annual Meeting of the Transportation Research Board, Washington, DC.

Manski, C.F., 1977. The structure of random utility models. Theory Decis. 8, 229–254.

McFadden, D., 1989. A method of simulated moments for estimation of discrete response models without numerical integration. Econom. J. Econom. Soc. 995–1026.

McFadden, D., 1974. Conditional logit analysis of qualitative choice behavior. Front. Econom. 105-142, Acad. Press New York.

McFadden, D., Train, K., 2000. Mixed MNL models for discrete response. J. Appl. Econom. 15, 447–470.

McKelvey, R.D., Zavoina, W., 1975. A statistical model for the analysis of ordinal level dependent variables. J. Math. Sociol. 4, 103–120.

McNally, M.G., Rindt, C., 2007. The activity-based approach, in: Handbook of Transport Modelling. pp. 55–73.

Menghini, G., Carrasco, N., Schüssler, N., Axhausen, K.W., 2010. Route choice of cyclists in Zurich. Transp. Res. part A policy Pract. 44, 754–765.

Mitra, R., 2013. Independent mobility and mode choice for school transportation: a review and framework for future research. Transp. Rev. 33, 21–43.

Molin, E., Mokhtarian, P., Kroesen, M., 2016. Multimodal travel groups and attitudes: A latent class cluster analysis of Dutch travelers. Transp. Res. Part A Policy Pract. 83, 14–29.

Montini, L., Antoniou, C., Axhausen, K.W., 2017. Route and mode choice models using GPS data, in: TRB Annual Meeting Online. pp. 17–3082.

Motoaki, Y., Daziano, R.A., 2015. A hybrid-choice latent-class model for the analysis of the effects of weather on cycling demand. Transp. Res. Part A Policy Pract. 75, 217–230.

Muñoz, B., Monzon, A., Daziano, R.A., 2016a. The increasing role of latent variables in modelling bicycle mode choice. Transp. Rev. 36, 737–771.

Muñoz, B., Monzon, A., Lois, D., 2013. Cycling habits and other psychological variables affecting commuting by bicycle in Madrid, Spain. Transp. Res. Rec. J. Transp. Res. Board 1–9.

Muñoz, B., Monzon, A., López, E., 2016b. Transition to a cyclable city: Latent variables affecting bicycle commuting. Transp. Res. Part a Policy Pract. 84, 4–17.

Munshi, T., 2016. Built environment and mode choice relationship for commute travel in the city of Rajkot, India. Transp. Res. part D Transp. Environ. 44, 239–253.

Newman, J.P., Bernardin, V.L., 2010. Hierarchical ordering of nests in a joint mode and destination choice model. Transportation (Amst). 37, 677–688.

Nylund, K.L., Asparouhov, T., Muthén, B.O., 2007. Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. Struct. Equ. Model. A Multidiscip. J. 14, 535–569.

Olde Kalter, M.-J.J.T., Harms, L., Geurs, K.T., 2015. Elk vervoermiddel heeft zijn voordeel. En zijn nadeel. Over attitudes en voorkeuren ten aanzien van de auto, OV en fiets, in: Colloquium Vervoersplanologisch Speurwerk 2015.

OViN, 2011. Onderzoek Verplaatsingsgedrag Nederland 1986-2012, in: CBS [in Dutch].

Paleti, R., Vovsha, P.S., Givon, D., Birotker, Y., 2014. Joint modeling of trip mode and departure time choices using revealed and stated preference data. Transp. Res. Rec. 2429, 67–78.

Pan-EuropeanProgramme, 2014. Fourth high-level meeting on transport, health and environment, World Health Organisation & United Nations, Paris Declaration.

Park, D., Rilett, L., 1997. Identifying multiple and reasonable paths in transportation networks: A heuristic approach. Transp. Res. Rec. J. Transp. Res. Board 31–37.

Park, K., Bell, M., Kaparias, I., Bogenberger, K., 2007. Learning user preferences of route choice behaviour for adaptive route guidance. IET Intell. Transp. Syst. 1, 159–166.

Paulley, N., Balcombe, R., Mackett, R., Titheridge, H., Preston, J., Wardman, M., Shires, J., White, P., 2006. The demand for public transport: The effects of fares, quality of service, income and car ownership. Transp. policy 13, 295–306.

Pike, S., Lubell, M., 2018. The conditional effects of social influence in transportation mode choice. Res. Transp. Econ. 68, 2–10.

Prato, C., Bekhor, S., 2007. Modeling route choice behavior: How relevant is the composition of choice set? Transp. Res. Rec. J. Transp. Res. Board 64–73.

Prato, C., Bekhor, S., 2006. Applying branch-and-bound technique to route choice set generation. Transp. Res. Rec. J. Transp. Res. Board 19–28.

Prato, C.G., 2009. Route choice modeling: past, present and future research directions. J. choice Model. 2, 65–100.

Prato, C.G., Bekhor, S., Pronello, C., 2012. Latent variables and route choice behavior. Transportation (Amst). 39, 299–319.

Prato, C.G., Halldórsdóttir, K., Nielsen, O.A., 2018. Evaluation of land-use and transport network effects on cyclists' route choices in the Copenhagen Region in value-of-distance space. Int. J. Sustain. Transp. 12, 770–781.

Pucher, J., Buehler, R., 2008. Making cycling irresistible: lessons from the Netherlands, Denmark and Germany. Transp. Rev. 28, 495–528.

Ralph, K.M., 2017. Multimodal millennials? The four traveler types of young people in the United States in 2009. J. Plan. Educ. Res. 37, 150–163.

Ramming, M.S., 2002. Network knowledge and route choice. Ph. D. Thesis, Massachusetts Inst. Technol.

Richards, M.G., Ben-Akiva, M., 1974. A simultaneous destination and mode choice model for shopping trips. Transportation (Amst). 3, 343–356.

Rieser-Schüssler, N., Balmer, M., Axhausen, K.W., 2013. Route choice sets for very high-resolution data. Transp. A Transp. Sci. 9, 825–845.

Rijksoverheid, 2018. Staatssecretaris Van Veldhoven: 200.000 mensen uit de file op de fiets [WWW Document]. https//www.rijksoverheid.nl/actueel/nieuws/2018/06/12/staatssecretaris-van-veldhoven-200.000-mensen-uit-de-file-op-de-fiets [in Dutch].

Rijkswaterstaat, 2018. Nederlands Regionaal Model (NRM) en Landelijk Model Systeem (LMS) [WWW Document]. Rijkswaterstaat, https://www.rijkswaterstaat.nl/wegen/wegbeheer/aanleg-wegen/nederlands-regionaal-model-nrm-en-landelijk-model-systeem-lms.aspx.

Rodriguez, A., Vogt, C.A., 2009. Demographic, environmental, access, and attitude factors that influence walking to school by elementary school-aged children. J. Sch. Health 79, 255–261.

Rodríguez, D.A., Joo, J., 2004. The relationship between non-motorized mode choice and the local physical environment. Transp. Res. Part D Transp. Environ. 9, 151–173.

Schmidt, P., Strauss, R.P., 1975. Estimation of models with jointly dependent qualitative variables: A simultaneous logit approach. University of North Carolina.

Scott, K., Pabón-Jiménez, G., Bernstein, D., 1997. Finding alternatives to the best path.

Sener, I.N., Eluru, N., Bhat, C.R., 2009. An analysis of bicycle route choice preferences in Texas, US. Transportation (Amst). 36, 511–539.

Shabanpour, R., Golshani, N., Derrible, S., Mohammadian, A., Miralinaghi, M., 2017. Joint discrete-continuous model of travel mode and departure time choices. Transp. Res. Rec. J. Transp. Res. Board 41–51.

Shakeel, K., Rashidi, T.H., Waller, T.S., 2016. Choice set formation behavior: joint mode and route choice selection model. Transp. Res. Rec. J. Transp. Res. Board 96–104.

Sheffi, Y., Powell, W.B., 1982. An algorithm for the equilibrium assignment problem with random link times. Networks 12, 191–207.

Shiftan, Y., Ben-Akiva, M., 2011. A practical policy-sensitive, activity-based, travel-demand model. Ann. Reg. Sci. 47, 517–541.

Sigurdardottir, S.B., Kaplan, S., Møller, M., Teasdale, T.W., 2013. Understanding adolescents' intentions to commute by car or bicycle as adults. Transp. Res. part D Transp. Environ. 24, 1–9.

Sklar, A., 1973. Random variables, joint distribution functions, and copulas. Kybernetika 9, 449–460.

Skov-Petersen, H., Barkow, B., Lundhede, T., Jacobsen, J.B., 2018. How do cyclists make their way?-A GPS-based revealed preference study in Copenhagen. Int. J. Geogr. Inf. Sci. 32, 1469–1484.

Small, K.A., 1987. A discrete choice model for ordered alternatives. Econom. J. Econom. Soc. 409–424.

Stinson, M., Bhat, C., 2003. Commuter bicyclist route choice: Analysis using a stated preference survey. Transp. Res. Rec. J. Transp. Res. board 107–115.

Swait, J., Ben-Akiva, M., 1987a. Empirical test of a constrained choice discrete model: mode choice in Sao Paulo, Brazil. Transp. Res. Part B Methodol. 21, 103–115.

Swait, J., Ben-Akiva, M., 1987b. Incorporating random constraints in discrete models of choice set generation. Transp. Res. Part B Methodol. 21, 91–102.

TheWorldbank, 2018. Urban population (% of total) [WWW Document]. The Worldbank: https://data.worldbank.org/indicator/SP.URB.TOTL.IN.ZS.

Ton, D., Cats, O., Duives, D., Hoogendoorn, S., 2017. How Do People Cycle in Amsterdam, Netherlands?: Estimating Cyclists' Route Choice Determinants with GPS Data from an Urban Area. Transp. Res. Rec. 2662, 75–82.

Ton, D., Duives, D.C., Cats, O., Hoogendoorn-Lanser, S., Hoogendoorn, S.P., 2019a. Cycling or walking? Determinants of mode choice in the Netherlands. Transp. Res. Part A Policy Pract. 123, 7–23.

Ton, D., Zomer, L.-B., Schneider, F., Hoogendoorn-Lanser, S., Duives, D., Cats, O., Hoogendoorn, S., 2019b. Latent classes of daily mobility patterns: the relationship with attitudes towards modes. Transportation (Amst). https://doi.org/10.1007/s11116-019-09975-9

UN, 2018. News [WWW Document]. United Nations Dep. Econ. Soc. Aff. https//www.un.org/development/desa/en/news/population/2018-revision-of-world-urbanization-prospects.html.

UNECE, 2018. Economic and Social Council [WWW Document]. Draft pan-European master plan Cycl. Promot. https//thepep.unece.org/sites/default/files/2018-12/1816159E.pdf.

van Cranenburgh, S., Chorus, C.G., 2017. Does The Decision Rule Matter For Large-Scale Transport Models? Transp. Res. procedia 23, 848–867.

van de Coevering, P., de Kruijf, J., Bussche, D., 2014. Bike print. Policy renewal and innovation by means of tracking technology, in: Colloquium Vervoersplanologisch Speurwerk, Eindhoven, Netherlands.

van Overdijk, R., der Waerden, P., Borgers, A., 2017. The Influence of Comfort and Travel Time on Cyclists' Route Choice Decisions.

Verkeersnet, 2015. Onderzoek naar betere fietsmodellen [WWW Document].

http//www.verkeersnet.nl/17820/onderzoek-naar-betere-fietsmodellen/ [in Dutch].

Vermunt, J.K., Magidson, J., 2005. Latent GOLD 4.0 user's guide.

Vermunt, J.K., Magidson, J., 2002. Latent class cluster analysis. Appl. latent Cl. Anal. 11, 89–106.

Vickrey, W.S., 1969. Congestion theory and transport investment. Am. Econ. Rev. 59, 251–260.

Vij, A., Carrel, A., Walker, J.L., 2013. Incorporating the influence of latent modal preferences on travel mode choice behavior. Transp. Res. Part A Policy Pract. 54, 164–178.

Vij, A., Gorripaty, S., Walker, J.L., 2017. From trend spotting to trend'splaining: Understanding modal preference shifts in the San Francisco Bay Area. Transp. Res. Part A Policy Pract. 95, 238–258.

Vij, A., Walker, J.L., 2016. How, when and why integrated choice and latent variable models are latently useful. Transp. Res. Part B Methodol. 90, 192–217.

Vovsha, P., 1997. Application of cross-nested logit model to mode choice in Tel Aviv, Israel, metropolitan area. Transp. Res. Rec. J. Transp. Res. Board 6–15.

Vrtic, M., Schüssler, N., Erath, A., Axhausen, K.W., 2007. Route, mode and departure time choice behaviour in the presence of mobility pricing, paper to be presented at the 86th Annual Meeting of the Transportation Research Board. Washingt. DC.

Wang, F., Ross, C.L., 2018. Machine learning travel mode choices: Comparing the performance of an extreme gradient boosting model with a multinomial logit model. Transp. Res. Rec. 0361198118773556.

Wang, Y., Chau, C.K., Ng, W.Y., Leung, T.M., 2016. A review on the effects of physical built environment attributes on enhancing walking and cycling activity levels within residential neighborhoods. Cities 50, 1–15.

Wardman, M., Tight, M., Page, M., 2007. Factors influencing the propensity to cycle to work. Transp. Res. Part A Policy Pract. 41, 339–350.

Whalen, K.E., Páez, A., Carrasco, J.A., 2013. Mode choice of university students commuting to school and the role of active travel. J. Transp. Geogr. 31, 132–142.

Williams, H.C.W.L., 1977. On the formation of travel demand models and economic evaluation measures of user benefit. Environ. Plan. A 9, 285–344.

Winters, M., Buehler, R., Götschi, T., 2017. Policies to promote active travel: evidence from reviews of the literature. Curr. Environ. Heal. reports 4, 278–285.

Witlox, F., Tindemans, H., 2004. Evaluating bicycle-car transport mode competitiveness in an urban environment. An activity-based approach. World Transp. Policy Pract. 10.

Yagi, S., Mohammadian, A.K., 2010. An activity-based microsimulation model of travel demand in the Jakarta metropolitan area. J. Choice Model. 3, 32–57.

Yang, L., Shen, Q., Li, Z., 2016. Comparing travel mode and trip chain choices between holidays and weekdays. Transp. Res. Part A Policy Pract. 91, 273–285.

Yang, L., Zheng, G., Zhu, X., 2013. Cross-nested logit model for the joint choice of residential location, travel mode, and departure time. Habitat Int. 38, 157–166.

Ye, X., Pendyala, R.M., Gottardi, G., 2007. An exploration of the relationship between mode choice and complexity of trip chaining patterns. Transp. Res. Part B Methodol. 41, 96–

113. https://doi.org/10.1016/j.trb.2006.03.004

Zimmermann, M., Mai, T., Frejinger, E., 2017. Bike route choice modeling using GPS data without choice sets of paths. Transp. Res. part C Emerg. Technol. 75, 183–196.

About the Author

Danique Ton was born in Vlissingen, the Netherlands, on the 29th of September 1990. After finishing her high school in 2008 she came to Delft to study Systems Engineering, Policy Analysis, and Management at Delft University of Technology, where she specialised in Transport, Infrastructure, and Logitics. She completed her BSc. degree in 2011. In 2010, she spent half a year at Queensland University of Technology in Brisbane, Australia, for her minor studies.

In 2011, she started her MSc. in Civil Engineering at Delft University of Technology, where she specialised in Transport and Planning. She obtained here MSc. degree, with distinction, in 2014. Her thesis focussed on route and location choice behaviour of departing passengers in train stations.

After completing her studies, she worked as researcher at NS Stations, where she focussed mainly on the field of pedestrian behaviour in train stations. She investigated the behaviour of pedestrians on both macroscopic and microscopic levels.

In February 2016, she joined the Transport and Planning group at Delft University of Technology for PhD studies. After submitting her thesis in April 2019, she was a visiting scholar at EPFL in Lausanne, Switzerland, to initiate her Postdoc at Delft University of Technology.

Her research interests include travel behaviour research, active modes, sustainable mobility, discrete choice analysis, and policy evaluation.

List of Publications

Journal Articles

- Ton, D., Zomer, L.B., Schneider, F., Hoogendoorn-Lanser, S., Duives, D., Cats, O., & Hoogendoorn, S. (in press). Latent classes of daily mobility patterns: The relationship with attitudes towards modes. *Transportation.* https://doi.org/10.1007/s11116-019-09975-9
- Ton, D., Duives, D.C., Cats, O., Hoogendoorn-Lanser S., & Hoogendoorn, S.P. (2019). Cycling or Walking? Determinants of mode choice in the Netherlands. *Transportation Research Part A*, 7-23. https://doi.org/ 10.1016/j.tra.2018.08.023
- Zomer, L.B., Schneider, F., Ton, D., Hoogendoorn-Lanser, S., Duives, D., Cats, O., Hoogendoorn, S. (2019). Determinants of urban wayfinding styles. *Travel Behaviour and Society*, 72-85. https://doi.org/10.1016/j.tbs.2019.07.002
- Ton, D., Duives, D.C., Cats, O., & Hoogendoorn, S.P. (2018). Evaluating a data-driven approach for choice set identification using GPS bicycle route choice data from Amsterdam. *Travel Behaviour and Society,* 105-117. https://doi.org/10.1016/j.tbs. 2018.07.001
- Ton, D., Cats, O., Duives, D., & Hoogendoorn, S. (2017). How do people cycle in Amsterdam, The Netherlands? Estimating cyclists' route choice determinants using GPS data from an urban area. *Transportation Research Record: Journal of the Transportation Research Board*, 75-82. https://doi.org/ 10.3141/2662-09

Working Papers

- Ton, D., Bekhor, S., Cats, O., Duives, D., Hoogendoorn-Lanser, S., & Hoogendoorn, S. (under review). Determinants of the experienced mode choice set: Commuting trips in the Netherlands.
- Ton, D., Duives, D.C., Cats, O., & Hoogendoorn, S.P. (under review). Simultaneous modelling of multiple travel choice dimensions: Assessment of the suitability and applicability of different discrete choice modelling structures.
- Ton, D., Shelat, S., Nijënstein, S., Rijsman, L., van Oort, N., Hoogendoorn, S. (under review). The role of cycling towards urban transit stations: simultaneously modelling the access mode and station choice.
- Duives D.C., Ton D., Hoogendoorn-Lanser S., Hoogendoorn S., Daamen W. (under review). Who chooses to walk? A study of the impact of the attitude towards walking.
- Schneider, F., Ton, D., Zomer, L.B., Daamen, W., Duives, D., Hoogendoorn-Lanser, S., Hoogendoorn, S. (under review). Trip chain complexity: a comparison among latent classes of daily mobility patterns.
- Duives D.C., Galama, I., Ton, D., Schneider, F., Daamen, W., Cats, O., Hoogendoorn, S.P. (under review). Independence and sequentiality of cyclists' travel choices: review of determinants of cyclists' mode, route, and operational choices.

Peer-reviewed Conference Contributions

- Rijsman, L., van Oort, N., Ton, D., Hoogendoorn, S., Molin, E., & Teijl, T. Walking and bicycle catchment areas of tram stops: factors and insights – *Proceedings of the MT-ITS conference*, June 2019, Kraków, Poland.

- Ton, D., Bekhor, S., Duives, D., Cats, O., Hoogendoorn-Lanser, S., & Hoogendoorn, S. Determinants of the consideration mode choice set – *Presented at the hEART conference,* September 2018, Athens, Greece.
- Ton, D., Zomer, L.B., Schneider, F., Hoogendoorn-Lanser, S., Duives, D., Cats, O., & Hoogendoorn, S. Latent classes of daily mobility patterns: The relationship with attitudes towards modes – *Presented at the International Association of Travel Behaviour Research Conference*, July 2018, Santa Barbara, USA.
- Zomer, L.B., Schneider, F., Ton, D., Hoogendoorn-Lanser, S., Duives, D., Cats, O., & Hoogendoorn, S. Wayfinding styles: The relationship with mobility patterns & navigational preferences – *Presented at the International Association of Travel Behaviour Research Conference*, July 2018, Santa Barbara, USA.
- Schneider, F., Ton, D., Zomer, L.B., Daamen, W., Duives, D., Hoogendoorn-Lanser, S., & Hoogendoorn, S. Trip chains: a comparison among latent mobility pattern classes – *Presented at the International Association of Travel Behaviour Research Conference*, July 2018, Santa Barbara, USA.
- Ton, D., Duives, D.C, Cats, O., & Hoogendoorn, S.P. Who, Why, When & Where do people walk or cycle in the Netherlands? – *Proceedings of the VeloCity Scientific Colloquium*, June 2017, Nijmegen, The Netherlands.
- Ton, D., Duives, D., Cats, O., & Hoogendoorn, S. Evaluating a data-driven approach for choice set identification: Using GPS bicycle route choice data from Amsterdam – *Presented at the International Choice Modelling Conference,* April 2017, Cape Town, South Africa.
- Ton, D., Cats, O., Duives, D., & Hoogendoorn, S. How do people cycle in Amsterdam? Estimating cyclists' route choice determinants using GPS data from an urban area – *Proceedings of the 96th Transportation Research Board,* January 2017, Washington DC, USA.

Other Publications

- Zomer, L.B., Ton, D., van Oijen, T. (2018). Active modes first! *NM Magazine*.
- Duives, D., Ton, D., Hoogendoorn, S. (2017). De toekomst van floating bike data. *NM Magazine*.

TRAIL Thesis Series

The following list contains the most recent dissertations in the TRAIL Thesis Series. For a complete overview of more than 250 titles see the TRAIL website: www.rsTRAIL.nl.

The TRAIL Thesis Series is a series of the Netherlands TRAIL Research School on transport, infrastructure and logistics.

Ton, D., *Unravelling Mode and Route Choice Behaviour of Active Mode Users*, T2019/12, September 2019, TRAIL Thesis Series, the Netherlands

Shu, Y., *Vessel Route Choice Model and Operational Model Based on Optimal Control,* T2019/11, September 2019, TRAIL Thesis Series, the Netherlands

Luan, X., *Traffic Management Optimization of Railway Networks,* T2019/10, July 2019, TRAIL Thesis Series, the Netherlands

Hu, Q., *Container Transport inside the Port Area and to the Hinterland,* T2019/9, July 2019, TRAIL Thesis Series, the Netherlands

Andani, I.G.A., *Toll Roads in Indonesia: transport system, accessibility, spatial and equity impacts,* T2019/8, June 2019, TRAIL Thesis Series, the Netherlands

Ma, W., *Sustainability of Deep Sea Mining Transport Plans,* T2019/7, June 2019, TRAIL Thesis Series, the Netherlands

Alemi, A., *Railway Wheel Defect Identification,* T2019/6, January 2019, TRAIL Thesis Series, the Netherlands

Liao, F., *Consumers, Business Models and Electric Vehicles,* T2019/5, May 2019, TRAIL Thesis Series, the Netherlands

Tamminga, G., *A Novel Design of the Transport Infrastructure for Traffic Simulation Models,* T2019/4, March 2019, TRAIL Thesis Series, the Netherlands

Lin, X., *Controlled Perishable Goods Logistics: Real-time coordination for fresher products*, T2019/3, January 2019, TRAIL Thesis Series, the Netherlands

Dafnomilis, I., *Green Bulk Terminals: A strategic level approach to solid biomass terminal design*, T2019/2, January 2019, TRAIL Thesis Series, the Netherlands

Feng, Fan, *Information Integration and Intelligent Control of Port Logistics System,* T2019/1, January 2019, TRAIL Thesis Series, the Netherlands

Beinum, A.S. van, *Turbulence in Traffic at Motorway Ramps and its Impact on Traffic Operations and Safety,* T2018/12, December 2018, TRAIL Thesis Series, the Netherlands

Bellsolà Olba, X., *Assessment of Capacity and Risk: A Framework for Vessel Traffic in Ports,* T2018/11, December 2018, TRAIL Thesis Series, the Netherlands

Knapper, A.S., *The Effects of using Mobile Phones and Navigation Systems during Driving,* T2018/10*,* December 2018, TRAIL Thesis Series, the Netherlands

Varotto, S.F., *Driver Behaviour during Control Transitions between Adaptive Cruise Control and Manual Driving: empirics and models,* T2018/9, December 2018, TRAIL Thesis Series, the Netherlands

Stelling-Kończak, A., *Cycling Safe and Sound,* T2018/8, November 2018, TRAIL Thesis Series, the Netherlands

Essen, van M.A., *The Potential of Social Routing Advice,* T2018/7, October 2018, TRAIL Thesis Series, the Netherlands

Su, Zhou, *Maintenance Optimization for Railway Infrastructure Networks*, T2018/6, September 2018, TRAIL Thesis Series, the Netherlands

Cai, J., *Residual Ultimate Strength of Seamless Metallic Pipelines with Structural Damage*, T2018/5, September 2018, TRAIL Thesis Series, the Netherlands

Ghaemi, N., *Short-turning Trains during Full Blockages in Railway Disruption Management,* T2018/4, July 2018, TRAIL Thesis Series, the Netherlands

Gun, van der J.P.T., *Multimodal Transportation Simulation for Emergencies using the Link Transmission Model,* T2018/3, May 2018, TRAIL Thesis Series, the Netherlands

Van Riessen, B., *Optimal Transportation Plans and Portfolios for Synchromodal Container Networks,* T2018/2, March 2018, TRAIL Thesis Series, the Netherlands

Saeedi, H., *Network-Level Analysis of the Market and Performance of Intermodal Freight Transport,* T2018/1, March 2018, TRAIL Thesis Series, the Netherlands

Ypsilantis, P., *The Design, Planning and Execution of Sustainable Intermodal Port-hinterland Transport Networks,* T2017/14, December 2017, TRAIL Thesis Series, the Netherlands

Han, Y, *Fast Model Predictive Control Approaches for Road Traffic Control,* T2017/13, December 2017, TRAIL Thesis Series, the Netherlands

Wang, P., *Train Trajectory Optimization Methods for Energy-Efficient Railway Operations,* T2017/12, December 2017, TRAIL Thesis Series, the Netherlands

Weg, G.S. van de, *Efficient Algorithms for Network-wide Road Traffic Control*, T2017/11, October 2017, TRAIL Thesis Series, the Netherlands

He, D., *Energy Saving for Belt Conveyors by Speed Control,* T2017/10, July 2017, TRAIL Thesis Series, the Netherlands

Bešinović, N., *Integrated Capacity Assessment and Timetabling Models for Dense Railway Networks,* T2017/9, July 2017, TRAIL Thesis Series, the Netherlands