Design and Optimization of Road Networks for Automated Vehicles

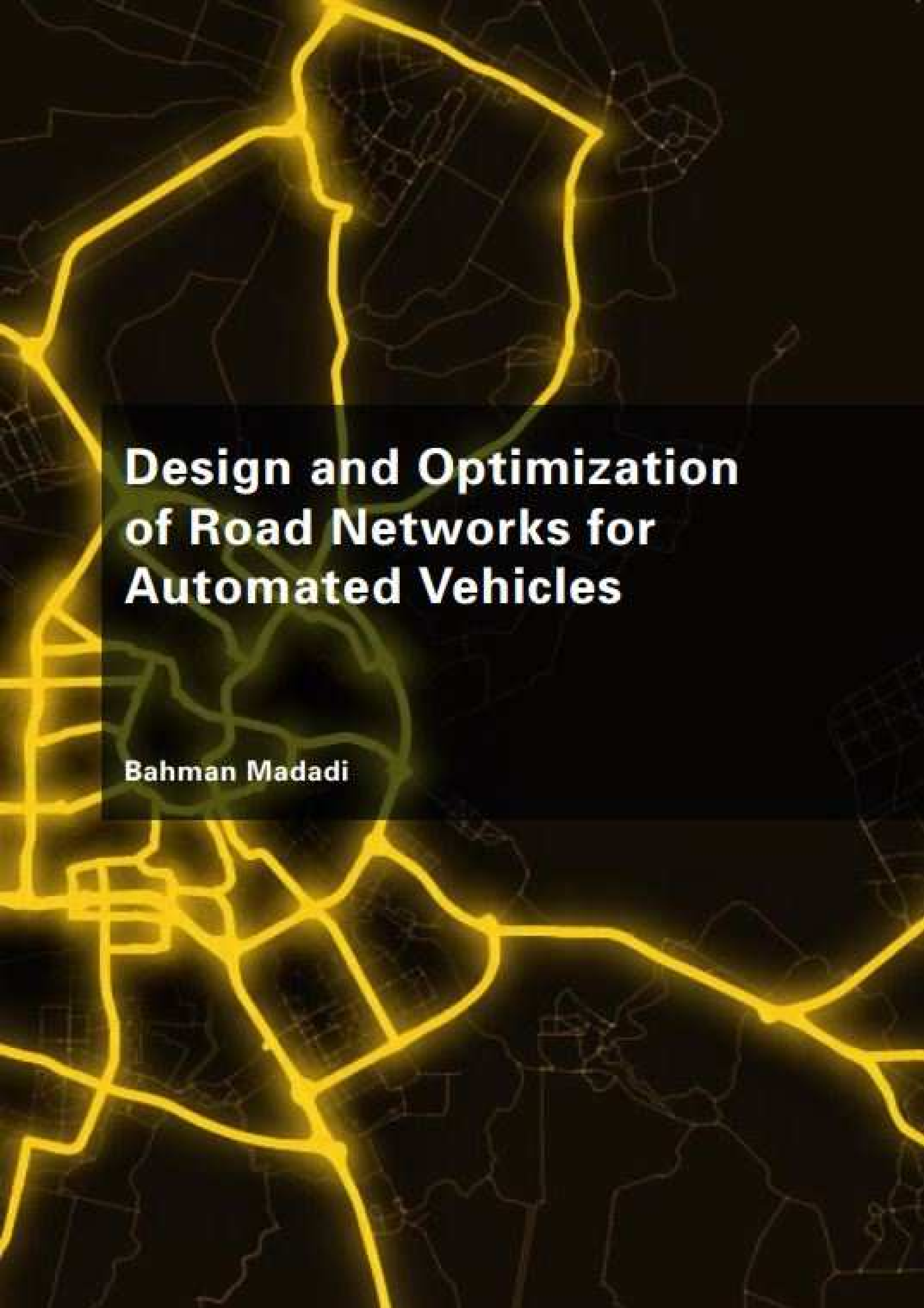
Madadi, B.

DOI
10.4233/uuid:c21d4943-b848-4e77-b5b8-a5423f751dbd

Publication date
2021

Document Version
Final published version

Citation (APA)
Madadi, B. (2021). *Design and Optimization of Road Networks for Automated Vehicles*. [Dissertation (TU Delft), Delft University of Technology]. TRAIL Research School. https://doi.org/10.4233/uuid:c21d4943-b848-4e77-b5b8-a5423f751dbd



Design and Optimization of Road Networks for Automated Vehicles

Bahman Madadi

Design and Optimization of Road Networks for Automated Vehicles

Bahman Madadi

Delft University of Technology, 2021

This research was funded as part of the project Spatial and Transport impacts of Automated Driving (STAD) by the Netherlands Organization for Scientific Research (NWO) under contract 438-15-161.

Cover illustration and design: Bahman Madadi & Solène Jimenez

Design and Optimization of Road Networks for Automated Vehicles

Dissertation

For the purpose of obtaining the degree of doctor

at Delft University of Technology,

by the authority of the Rector Magnificus Prof. dr. ir. T.H.J.J. van der Hagen,

chair of the Board for Doctorates,

to be defended publicly on

Wednesday, 20 January, 2021 at 15:00 o'clock

by

Bahman MADADI

Master of Science in Industrial and Systems Engineering,

Istanbul Sehir University, Turkey

born in Tehran, Iran.

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus	chairperson
Prof. dr. ir. Bart van Arem	Delft University of Technology, promotor
Dr. ir. Rob van Nes	Delft University of Technology, promotor
Dr. Maaike Snelder	Delft University of Technology, copromotor

Independent members:

Prof. dr. ir. Karen Aardal	Delft University of Technology
Prof. dr. ir. Rudy R. Negenborn	Delft University of Technology
Prof. dr. Constantinos Antoniou	Technical University of Munich, Germany
Dr. Risto Kulmala	Traficon, Finland

TRAIL Thesis Series no. T2021/3, the Netherlands Research School TRAIL

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

ISBN: 978-90-5584-272-8

Printed in the Netherlands

“Everything you can imagine is real.”

Pablo Picasso

Preface

This is it! This is the end of my PhD journey. From the excitement of moving to a new country and starting a PhD on a new topic, to home office during a global pandemic and a virtual defense. From struggles to learn a new topic, to presenting my research in international conferences in front of big audiences. From being surprised by the Dutch working culture, to explaining it to the new international colleagues. From the pain of having *een broodje gezond* for lunch, to the pleasure of having *bitterballen* in *borrels*. From long working days to meet deadlines during gray and cold Dutch winters, to joyful bicycle tours around the beautiful canals during the long summer days. From not knowing anyone in this new country, to getting to know many interesting people, and back to staying at home and not meeting any of them. From having my research accepted for presentation in multiple conferences, to not being able to attend them. This journey has been a true roller coaster ride.

This is not how I imagined the last days before my defense. I imagined having my family, my partner and all my friends around, enjoying the ceremonial defense day that marks the end of my PhD journey, having an after-party with all the people I like, and starting a new adventure afterwards to explore new corners of the world in a camping car. But I guess two years ago, nobody could imagine their lives the way they are now.

The Covid-19 pandemic and the subsequent lock-downs have changed the way we live, the way we work, and the way we socialize. I remember reading the acknowledgments of my fellow PhDs who mentioned all the people they were seeing regularly in the department and thanking them all, and I was thinking to myself: what is the big deal about spending time with colleagues? Why do you have to mention every social interaction you have had during your PhD and thank everybody? Now we all appreciate those interactions more than ever and wish to have them back. I have to spend some time and think to remember how I was spending my working days during my PhD. But I slowly remember now; long fresh air and sweet breaks with Florian, talks about life with Jishnu, impulsive discussions and room coffee breaks with Pablo, Reanne, Jeroen, Yihong, Maryna, Goof, Aries, Koen, Rafael, and Meiqi, long evenings working only with me and Jeroen in the room, watching random videos at work with Pablo, Jeroen and Yihong, walks with Reanne, hanging out at the first floor terrace with Javi, dropping by the Allegro room and disturbing Alexandra, Danique, Lara, Vincent, Mari-Jette and Lucia, playing (mini) basketball with Tim, Alphonse and Giulia in their room, long serious discussions with Paul, trying Chinese food with Vincent and Ding, learning Chinese expressions from Yihong, talking about India with Jishnu, Niharika, Freddy, Konstanze, Pablo and Malvika (plus the secret gift mission), speaking Farsi with Azita, Solmaz, Peyman and Ali, listening to Nikola's strange jokes, the regular after lunch coffee breaks, all the nice talks with Maria (remember when you forced me to run to catch the earlier train and save 15 minutes but we took the wrong train and ended up in the wrong city?), Alessandro, Bernat, Martijn, Yan, Nejc, Boudewijn, Panchamy, Tin, Joelle, Hans, Niels, Haneen, Goncalo, Oded and Meng, short unplanned chats

with Rob that could turn into long discussions about pretty much anything, and all the fun department outings, sparks meetings and borrels, interesting TRAIL PhD congresses (special thanks to Conchita for directing me towards the sweets first when I would arrive late), spending time with all these people in conferences, going to Jishnu's wedding in India and Pablo's wedding in Leiden, playing beach volleyball in the summer, and playing football again with DCF Inter thanks to Florian, Vincent and Alphonse who convinced me to join the team. Next time I get a chance for these activities again, I will appreciate them much more.

Of course this is not how the research was done. When I started this PhD, I did not know anything about my topic. I had a background in Industrial and Systems Engineering, so I was familiar with modeling and optimization. Yet I had no academic knowledge of automated vehicles and road networks. This was just an interesting topic for me, and thanks to Rob, Maaike and Bart, and their trust in me, I started this PhD hoping that with my different background, I will bring something new to the table. I tried to combine knowledge from different fields, and to strike a balance between theoretical work that can be published in top journals but might gather dust in a drawer without anyone actually using it in practice, and practical research that can have great societal impacts but might have less scientific value for academics. As a result, my work was sometimes too practical for some people and too theoretical for some others. I hope that both scientists and practitioners can benefit from this research.

Dealing with two supervisors and a promotor was also new for me. In time, I learned to go to Rob for long in-depth discussions, to Maaike for critical decisions that need to be made fast, and to Bart for strategic decisions and wider visions. This is not to say everything was always easy and smooth. Sometimes we were four people with four different opinions in a meeting, even until the last meetings. What I appreciated about my supervisory team was that everybody was (almost) always open to being convinced given good arguments. Rob, Maaike and Bart, thank you for all your support, wisdom, mentorship and knowledge. Without you, this journey would not be possible. I also, thank all my committee members for their constructive feedback, and for accepting to do the hard part of this job without the fun of traveling and enjoying the social events. In retrospect, I enjoyed doing this PhD and this experience confirmed for myself that I like being a researcher.

I owe a special debt of gratitude to Solène, my loving travel companion, not only in this journey, but in life. Thank you for your care when I needed it, for your patience when I was not a good companion (as it happens to all PhD candidates), and for your support throughout this journey.

Finally, I would like to thank my mother and my sister, Behnaz, for being there for me my whole life. Thank you for all your support during all these years. Behnaz, you are a source of strength for me, and I am glad every day to have you in my life. It is a pity that you cannot be here for my defense to see the fruit of all your care and support, but we will celebrate it together soon.

Bahman

Delft, December 2020

Summary

Automated vehicles (AVs) are on the horizon, and they are expected to deliver traffic safety and efficiency benefits to transportation systems. There are different automation levels for AVs based on the functionalities of the automation systems and their operating design domain (i.e., on which conditions these functionalities can be realized). AVs with limited automation functions are already available on the market; however, fully automated vehicles with unlimited operational design domain (ODD) are not expected in the near future. Reaching a high market penetration rate of fully automated vehicles is a gradual process that can take several decades. Thus for a long time, a heterogeneous mix of traffic with AVs of different automation levels and regular vehicles on the roads will be inevitable. During this transition period with mixed traffic, relying on driving automation technology alone without infrastructure support might compromise the potential safety and efficiency gains of AVs. A proper infrastructure can support AVs' functionalities, extend their ODD, and improve safety for all road users, while lack of proper infrastructure can negatively influence these factors. Besides, road infrastructure elements usually have a long lifetime and adjusting them can be costly. Hence, there is a strong need for research and planning to ensure that large infrastructure investments provide the highest societal benefits.

The main objective of this dissertation is to develop a road network design method for strategic planning of infrastructure investment decisions that can be applied in practice on real-sized networks to cope with the uncertain development path of AVs during the transition period to full automation. In order to reach this objective, the following key research questions are formulated.

1. What is a suitable road network configuration for accommodating AVs and what are the impacts of this configuration on network performance?

2. How can network configurations for AVs be optimized and what are the impacts of the optimal configuration on network performance?

3. How should the optimal road network configurations evolve over time and what are the impacts of the network and AV demand evolution on network performance?

4. How can different network configurations for accommodating AVs on road networks be combined and what are the impacts of combining different network configurations on network performance?

In order to address these research questions, a modular approach is used in this thesis to develop a comprehensive mathematical model through four main steps, each one addressing a research question. The general approach is to start with an essential model to understand the fundamental

properties of the problem, gradually add more complexity to the model by considering new dimensions of the problem in each step, and finally develop a model that can realistically capture the main aspects of the problem at a desirable level. Therefore, a new model is developed to address each research question. The applicability of the model developed in response to each research question is demonstrated via a case study. Figure I.1 illustrates the research approach. In the following, we elaborate on how each research question is addressed.

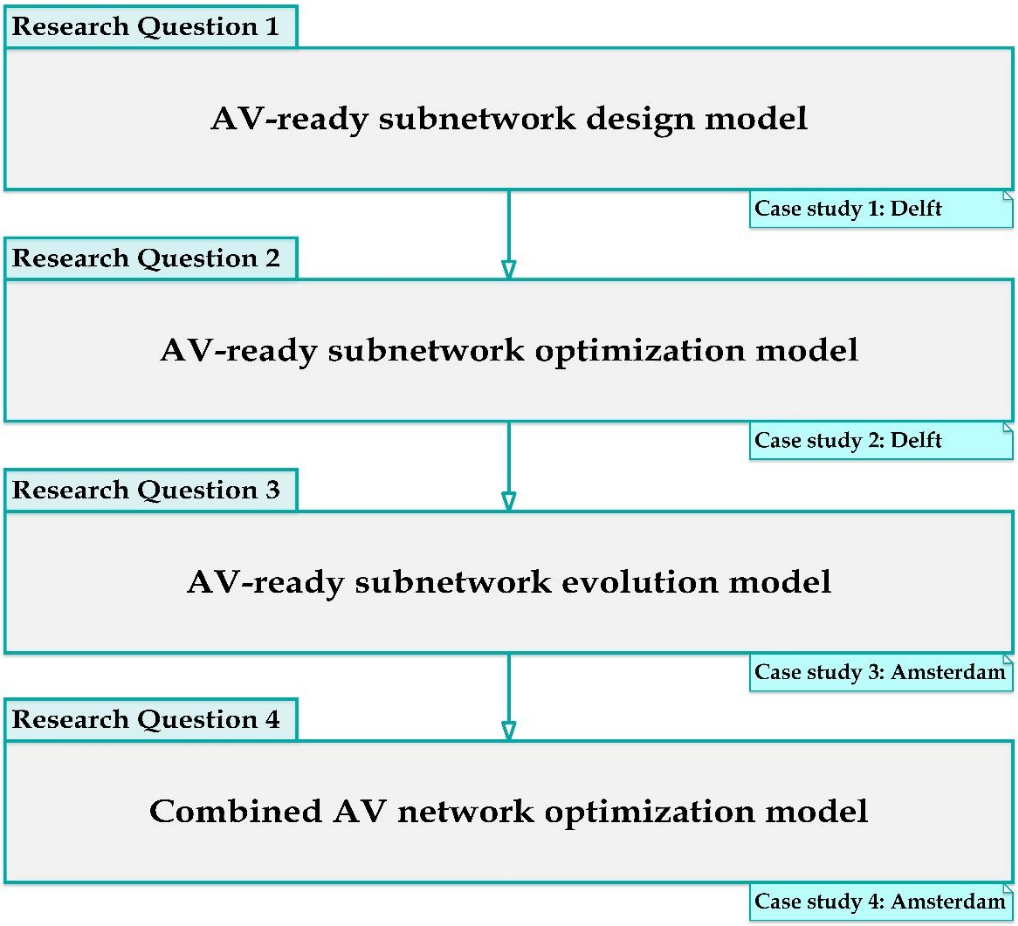


Figure I.1 Research approach

1. What is a suitable road network configuration for accommodating AVs and what are the impacts of this configuration on network performance?

We explored the concept of AV-ready subnetworks for vehicle automation levels 3-4 (according to the SAE classification) in an urban road network having mixed traffic and demonstrated its potential impacts. To evaluate the impacts of this configuration and model different vehicles' route choice behavior in mixed traffic, a static multi-class stochastic user equilibrium traffic assignment with a path-size logit route choice model and a Monte Carlo labeling route-set generation was adapted. The results showed a decrease in total travel cost with the increase in market penetration rate of higher automation levels, a decrease in total travel time, and a minor increase in total travel distance. Although in most cases vehicles with higher automation levels benefited more from the improvements, no deterioration in travel conditions was observed for the rest of the vehicles in any scenario. Furthermore, a noticeable shift of traffic from roads with access function towards roads with flow function and distributors was observed.

2. How can network configurations for AVs be optimized and what are the impacts of the optimal configuration on network performance?

We formulated the problem as a network design problem and presented a bi-level model to optimize the trade-off between infrastructure adjustment costs and network performance benefits of AV-ready subnetwork deployment in road networks. The upper level included the choice of links to be upgraded as part of the AV-ready subnetwork and the lower level entailed the travelers' response to these decisions. We defined solution requirements for the problem, suggested a solution algorithm that meets those requirements, and benchmarked its performance against two solution algorithms for network design problems considering three different performance criteria. Numerical examples for the network of Delft were presented to demonstrate the concept and solution algorithm performances. The results revealed that the algorithm presented in this study has a satisfactory performance and outperforms competing algorithms in all three criteria considered, namely, effectiveness, efficiency and design quality. The design quality criteria is specifically relevant to the formulation of the network design problem introduced in this study, which enforces a connectivity constraint on AV-ready subnetworks. It is specifically this constraint that makes the commonly used solution methods for network design problems less suitable for this problem. Furthermore, our findings indicated that the optimal layout of AV-ready subnetworks depends on the level of AV demand and infrastructure adjustment costs. An example of an AV-ready subnetwork for the network of Delft (case study 2) is depicted in Figure I.2.

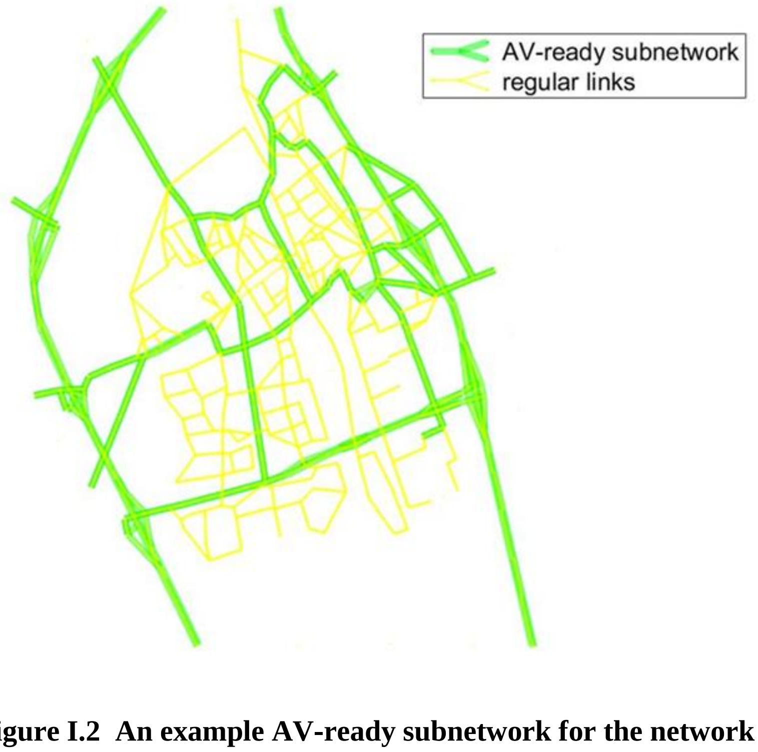


Figure I.2 An example AV-ready subnetwork for the network of Delft (case study 2)

3. How should the optimal road network configurations evolve over time and what are the impacts of network and AV demand evolution on network performance?

We proposed multi-stage deployment of AV-ready subnetworks in road networks and formulated the problem as a time-dependent network design problem, which is a bi-level mixed-integer programming problem. The upper level denoted infrastructure decisions made by authorities in several stages over a finite planning horizon to deploy and update the AV-ready subnetwork, that is, making a selection of links to upgrade as part of the AV-ready subnetwork in each decision stage. The objective of the formulation was to optimize the total societal benefits, which was represented by the summation of total discounted adjustment cost and total discounted travel cost. The lower level involved a travel mode and route choice equilibrium model, which represented travel choices of different user classes in each stage in response to new road network topologies. We presented variational inequality and fixed-point formulations of the lower level problem as a multi-class simultaneous mode and route choice user equilibrium by means of a hierarchical logit model, and solved it using a sequential linear approximation type algorithm. The upper level problem was modeled as a mathematical program with equilibrium constraints and solved for near-optimal solutions. The model was demonstrated on a case study using a realistic representation of the road network of the Amsterdam metropolitan region. In addition, we developed two efficient evolutionary heuristic algorithms that are tailored to the problem structure to solve the problem, and compared their performance to a genetic-algorithm-based heuristic solution method. The results indicated that both proposed algorithms can efficiently solve a large-scale instance of the problem while satisfying constraints in all scenarios, whereas the genetic-algorithm-based solution procedure failed to meet some requirements in scenarios with larger number of periods. In all scenarios, multi-stage deployment of AV-ready subnetworks led to improvements in overall network performance in terms of total travel time and cost. Yet the improvements were always accompanied by increased total travel distance. An example of an AV-ready subnetwork for the network of Amsterdam (case study 3) is depicted in Figure I.3.

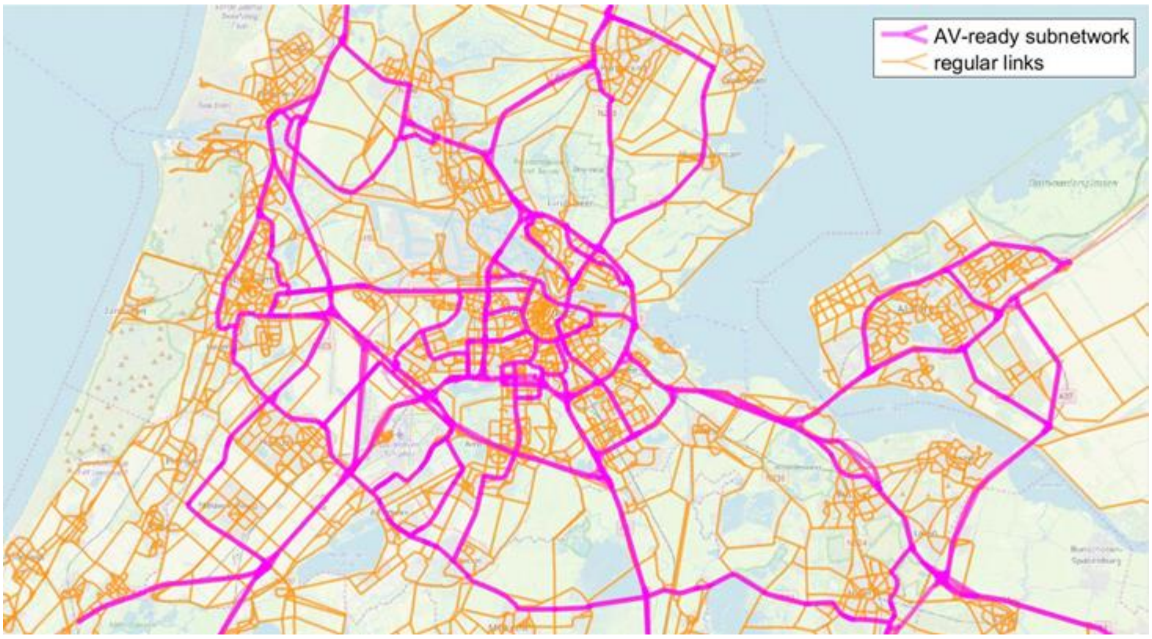


Figure I.3 An example AV-ready subnetwork for the network of Amsterdam (case study 3)

4. How can different network configurations for accommodating AVs on road networks be combined and what are the impacts of combining different network configurations on network performance?

The literature suggests that dedicated infrastructure for AVs and enhanced infrastructure for mixed traffic are the main alternatives for accommodating AVs on road networks during the transition period to full automation. We utilized both alternatives, and proposed a unified mathematical framework for optimizing road networks for AVs by simultaneous deployment of AV-ready subnetworks for mixed traffic, dedicated AV links and dedicated AV lanes. We modeled the problem as a bi-level network design problem where the upper level represents the decisions regarding links to be selected as parts of mentioned subnetworks in order to optimize the trade-off between infrastructure adjustment cost and total system travel cost, and the lower level contains a network equilibrium model that captures the travelers' responses to new network topologies with their route choices. An efficient heuristic solution method was introduced to solve the problem and find coherent network topologies in realistic networks. Applicability of the model on real road networks was demonstrated using a large-scale case study of the Amsterdam metropolitan region. The results indicated that different network configurations are relevant for different market penetration rates of AVs. For low market penetration rates, AV-ready subnetworks, which accommodate AVs in mixed traffic, were the most efficient configuration. After 30% market penetration rate, dedicated AV lanes appeared more often in optimal network topologies, but for very high market penetration rates, AV-ready subnetworks were selected more frequently again.

Based on the developed models and mentioned case studies, six overall conclusions were drawn in this thesis. They are highlighted below.

1. AV market penetration rate is the dominating factor to affect the road network performance. This phenomenon has been observed in all chapters of this thesis.

2. The bi-level modeling framework used in this thesis is essential for design and evaluation of infrastructure plans. This is due to the fact that the performance of infrastructure adjustment decisions depends on the travelers' response to these decisions via their travel choices.

3. The connectivity of subnetworks representing designated infrastructure for AVs is an indispensable requirement for their effectiveness. We established that this connectivity is necessary for finding coherent subnetworks, and we introduced four evolutionary solution methods that are tailored to the structure of the problems studied in this thesis and can efficiently find connected subnetworks in case studies of large-scale road networks. We showed that general metaheuristics with penalty functions are not efficient for dealing with the connectivity constraint and that tailored algorithms have a superior performance.

4. An effective AV-ready subnetwork including an appropriate selection of links to be upgraded with infrastructure adjustments to accommodate AVs can deliver a large proportion of the benefits obtainable from upgrading infrastructure on all links with significantly lower investment cost. Furthermore, deploying AV-ready subnetworks has apparent safety and network performance benefits during the transition period.

5. Different network layouts for accommodating AVs in road networks are relevant for different market penetration rates of AVs. For lower market penetration rates, AV-ready subnetworks, which accommodate AVs in mixed traffic, are found to be the most efficient configuration. However, starting from around 30% market penetration rate, dedicated AV lanes become relevant, and can efficiently host the AV traffic.

6. Road types play a crucial role in the choice of network configuration as well. Motorway on-ramps and off-ramps, single-lane roads and major regional roads that include sections with a single lane have been shown to be appropriate for mixed traffic, while dedicated AV lanes are most suited for motorways.

Future research directions with respect to the methodological improvements include considering new objectives for optimization of the upper level problem, developing new AV diffusion models that are less reliant on scale parameters, using dynamic traffic assignment models to capture the behavioral differences of AVs and regular vehicles, using multimodal traffic assignment models including more mode choices available to travelers, and modeling the travelers' origin and destination choices instead of assuming a fixed demand for each origin-destination pair.

Samenvatting

Automatische voertuigen (AV) doen hun intrede. De verwachting is dat automatische voertuigen de verkeersveiligheid en de netwerkprestatie ten goede komen. Er zijn meerdere automatiseringsniveaus voor automatische voertuigen die verschillen in functionaliteit van de automatiseringssystemen en hun operationeel ontwerpdomein (d.w.z. onder welke voorwaarden de functionaliteiten kunnen worden gerealiseerd). Automatische voertuigen met beperkte automatiseringsfuncties zijn al beschikbaar op de markt; volledig geautomatiseerde voertuigen met een onbeperkt operationeel ontwerpdomein worden echter niet verwacht in de nabije toekomst. Het bereiken van een hoge marktpenetratiegraad van volledig geautomatiseerde voertuigen is een geleidelijk proces dat enkele decennia kan duren. Een heterogene mix van verkeer met automatische voertuigen van verschillende automatiseringsniveaus en conventionele voertuigen zal dus voor lange tijd onvermijdelijk zijn. Tijdens deze overgangsperiode met gemengd verkeer kan het vertrouwen op de automatiseringstechniek zonder infrastructuurondersteuning, de potentiële veiligheids- en efficiëntiewinst van automatische voertuigen in het gedrang brengen. Een goede infrastructuur kan de functionaliteiten van de automatische voertuigen ondersteunen, hun operationeel ontwerpdomein uitbreiden en de veiligheid voor alle weggebruikers verbeteren. Een gebrek aan goede infrastructuur kan deze factoren negatief beïnvloeden. Bovendien heeft wegeninfrastructuur gewoonlijk een lange levensduur en kan het aanpassen ervan bijzonder kostbaar zijn. Daarom is er een grote behoefte aan onderzoek en planning om ervoor te zorgen dat grote investeringen in de infrastructuur de grootst mogelijke maatschappelijke voordelen opleveren.

Het belangrijkste doel van dit proefschrift is het ontwikkelen van een ontwerpmethode waarmee investeringen in fysieke en digitale weginfrastructuur die nodig zijn om automatische voertuigen te kunnen accommoderen, kunnen worden geoptimaliseerd. De methode moet rekening kunnen houden met het onzekere ontwikkelingstraject van automatische voertuigen tijdens de overgangsperiode naar volledige automatisering en moet kunnen worden toegepast op netwerken op ware grootte. Om dit doel te bereiken zijn de volgende onderzoeksvragen geformuleerd.

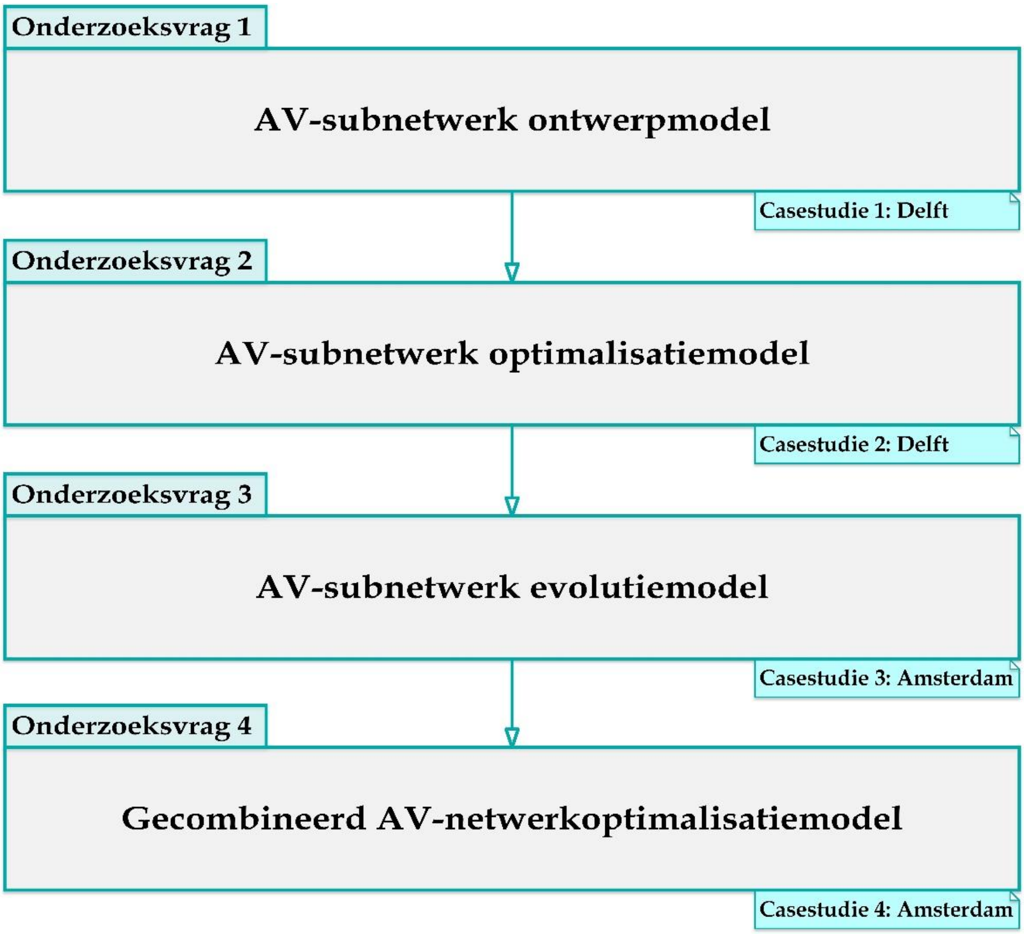
1. Wat is een geschikte wegennetconfiguratie voor het accommoderen van automatische voertuigen en wat zijn de effecten van deze configuratie op de netwerkprestatie?

2. Hoe kunnen netwerkconfiguraties voor automatische voertuigen worden geoptimaliseerd en wat zijn de effecten van de optimale configuratie op de netwerkprestatie?

3. Hoe moeten de optimale wegennetconfiguraties in de loop van de tijd evolueren en wat zijn de effecten van het netwerk en de AV-vraag op de netwerkprestatie?

4. Hoe kunnen verschillende netwerkconfiguraties voor het accommoderen van automatische voertuigen worden gecombineerd en wat is de impact van het combineren van verschillende netwerkconfiguraties op de netwerkprestaties?

Om deze onderzoeksvragen te kunnen beantwoorden, is in deze dissertatie een modulaire aanpak gevolgd om een uitgebreid wiskundig model te ontwikkelen door middel van vier hoofdstappen. Iedere stap behandelt een onderzoeksvraag. In de eerste stap is een basismodel ontwikkeld waarin de fundamentele eigenschappen van het probleem zijn opgenomen. Geleidelijk is in elke stap meer complexiteit aan het model toegevoegd om uiteindelijk tot een model te komen dat op realistische wijze de belangrijkste aspecten van het probleem op een gewenst niveau weergeeft. De toepasbaarheid van het ontwikkelde model wordt voor iedere onderzoeksvraag aangetoond aan de hand van een casestudie. Figuur II.1 illustreert de onderzoeksaanpak. Hieronder gaan we dieper in op de manier waarop elke onderzoeksvraag wordt benaderd.



Figuur II.1 Onderzoeksaanpak

1. Wat is een geschikte wegennetconfiguratie voor het accommoderen van automatische voertuigen en wat zijn de effecten van deze configuratie op de netwerkprestatie?

We hebben het concept van AV-subnetwerken voor voertuigautomatiseringsniveaus 3-4 (volgens de SAE-classificatie) in een stedelijk wegennet met gemengd verkeer onderzocht en de potentiële effecten ervan aangetoond. Een AV-subnetwerk bestaat uit een selectie van wegen waarop automatisch kan worden gereden in gemengd verkeer. Om de effecten van deze configuratie te evalueren en het routekeuzegedrag van verschillende voertuigen in gemengd verkeer te modelleren, is een statisch multi-class stochastisch toedelingsmodel met een path-size logit routekeuzemodel en een Monte Carlo labeling routesetgeneratie ontwikkeld voor automatische voertuigen. De resultaten tonen aan dat als de marktpenetratiegraad van hogere automatiseringsniveaus toeneemt, de totale reiskosten en de totale reistijd dalen terwijl totale reisafstand een kleine toename laat zien. Hoewel voertuigen met hogere automatiseringsniveaus in de meeste gevallen meer profijt hebben van de verbeteringen, wordt voor de rest van de voertuigen in geen enkel scenario een verslechtering van de rijomstandigheden waargenomen. Bovendien treedt een verschuiving op van het verkeer van toegangswegen naar wegen met een ontsluitingsfunctie en een doorstroomfunctie.

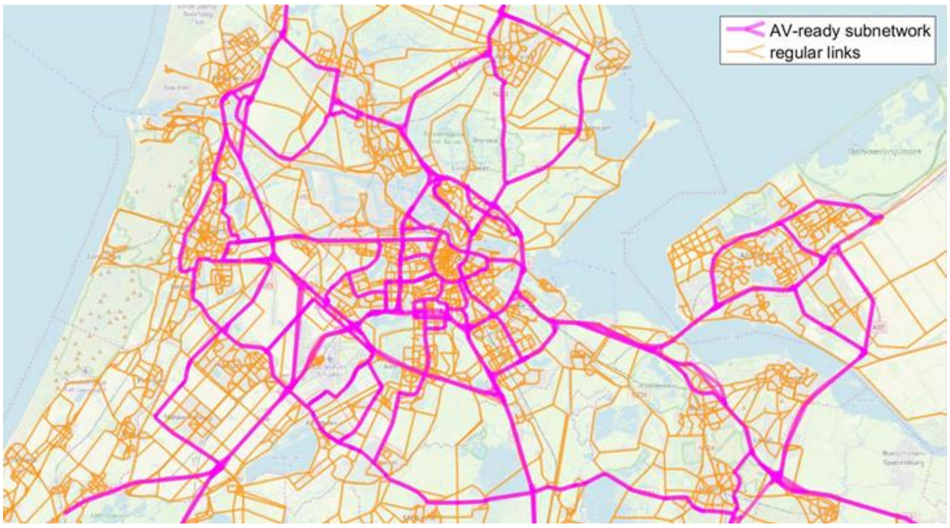
2. Hoe kunnen netwerkconfiguraties voor automatische voertuigen worden geoptimaliseerd en wat zijn de effecten van de optimale configuratie op de netwerkprestatie?

We hebben het probleem als een netwerkontwerpprobleem geformuleerd en een bi-level optimalisatiemodel ontwikkeld dat de kosten voor aanpassing van de infrastructuur minimaliseert en de baten van een betere netwerkprestatie van een AV-subnetwerk maximaliseert. Op het bovenste niveau wordt bepaald welke wegen worden geüpgraded en daarmee onderdeel worden van het AV-subnetwerk en op het onderste niveau wordt de reactie van de reizigers op het aangepaste netwerk gemodelleerd. We hebben de eisen voor het probleem gedefinieerd, een algoritme ontwikkeld dat aan die eisen voldoet en de prestatie ervan vergeleken met twee bestaande algoritmen voor netwerkontwerpproblemen, waarbij drie verschillende prestatiecriteria in aanmerking zijn genomen. De werking en kwaliteit van het ontwikkelde algoritme is gedemonstreerd voor een casestudie van Delft. Uit de resultaten is gebleken dat het ontwikkelde algoritme beter presteert dan concurrerende algoritmen op alle drie de onderzochte criteria, namelijk effectiviteit, efficiëntie en ontwerpkwaliteit. Een belangrijke ontwerpkwaliteitseis die in dit onderzoek is gesteld is dat wegen in het AV-subnetwerk met elkaar verbonden moeten zijn. Voor de meeste bestaande algoritmen voor het netwerkontwerpproblemen bleek het lastig om aan deze eis te voldoen. Uit de resultaten is daarnaast gebleken dat dat de optimale lay-out van AV-subnetwerken afhankelijk is van de vraag naar automatische voertuigen en de kosten voor aanpassing van de infrastructuur. Een voorbeeld van een AV-subnetwerk voor het netwerk van Delft (casestudie 2) is weergegeven in figuur II.2.

Figuur II.2 Een voorbeeld van een AV-subnetwerk voor het netwerk van Delft (casestudie 2)

3. Hoe moeten de optimale wegennetconfiguraties in de loop van de tijd evolueren en wat zijn de effecten van het netwerk en de AV-vraag op de netwerkprestatie?

We hebben een methode ontwikkeld waarmee het AV-subnetwerk in meerdere fases wordt opgebouwd. Het probleem is als een bi-level mixed -integer probleem geformuleerd met een tijdsafhankelijkheid. Op het bovenste niveau wordt bepaald welke infrastructuurbeslissingen autoriteiten in verschillende fases van een planningshorizon nemen om het AV- subnetwerk te ontwikkelen. Dat wil zeggen dat voor elke fase wordt bepaald welke links moeten worden geüpgraded en daarmee onderdeel worden van het AV-subnetwerk. Het doel is om de totale maatschappelijke baten te optimaliseren welke bestaan uit de som van de totale verdisconteerde aanpassingskosten en de totale verdisconteerde reiskosten. Op het onderste niveau wordt voor iedere fase met behulp van een evenwichtstoedelingsmodel de vervoerwijzekeuze en routekeuze van verschillende gebruikersklassen bepaald voor de nieuwe netwerkconfiguraties. Voor het onderste niveau is een ‘variational inequality’ en ‘fixed-point’ formulering opgesteld waarbij een simultaan vervoerwijze- en routkeuzegebruikersevenwicht wordt bepaald met behulp van een hiërarchisch logit-model. Het bovenste niveau is als optimalisatieprobleem met evenwichtsrestricties geformuleerd. We hebben twee efficiënte evolutionaire heuristieken ontwikkeld om het probleem op te lossen. De prestatie van deze heuristieken is vergeleken met de prestatie van een genetisch algoritme voor een casestudie met een realistische weergave van het wegennet van de Metropoolregio Amsterdam. De resultaten hebben aangetoond dat beide voorgestelde algoritmen een grootschalig probleem op efficiënte wijze kunnen oplossen en tegelijkertijd in alle scenario's voldoen aan de eisen, terwijl de op een genetisch algoritme gebaseerde oplossingsmethode in scenario's met een groter aantal fases niet voldoet aan enkele eisen. In alle scenario's leidt de ontwikkeling van een AV-subnetwerk in meerdere fases tot een verbetering van de totale netwerkprestatie in termen van totale reistijd en kosten. Echter, de verbeteringen gaan altijd gepaard met een grotere totale reisafstand. Een voorbeeld van een AV-subnetwerk voor Amsterdam (casestudie 3) is weergegeven in figuur II.3.



Figuur II.3 Een voorbeeld van een AV-subnetwerk voor het netwerk van Amsterdam (casestudie 3)

4. Hoe kunnen verschillende netwerkconfiguraties voor het accommoderen van automatische voertuigen op wegennetwerken worden gecombineerd en wat is de impact van het combineren van verschillende netwerkconfiguraties op de netwerkprestatie?

De literatuur suggereert dat afgeschermde infrastructuur voor automatische voertuigen en verbeterde infrastructuur voor automatische voertuigen in gemengd verkeer de belangrijkste opties zijn om automatische voertuigen op het wegennet te accommoderen tijdens de overgangsperiode naar volledige automatisering. We hebben een raamwerk ontwikkeld waarmee het wegennetwerk voor automatische voertuigen kan worden geoptimaliseerd door gelijktijdig te bepalen welke wegen het beste kunnen worden geüpgraded voor gemengd verkeer (AV-subnetwerk), waar doelgroepstroken zouden moeten worden aangelegd en welke wegen alleen voor automatische voertuigen toegankelijk zouden moeten zijn. We hebben het probleem gemodelleerd als een bi-level netwerkontwerpprobleem waarbij op het bovenste niveau wordt bepaald welke wegen geüpgraded worden voor automatische rijden in gemengd verkeer, op doelgroepstroken of op afgeschermde wegen. Hierbij wordt een afweging gemaakt tussen de kosten voor aanpassing van de infrastructuur en de totale reiskosten van alle gebruikers. Het onderste niveau bevat een netwerkevenwichtsmodel dat de routekeuze van reizigers bepaalt in het aangepaste wegennetwerk. Om dit netwerkontwerpprobleem op te lossen is een efficiënte heuristische oplossingsmethode ontwikkeld waarmee coherente netwerktopologieën ontworpen kunnen worden voor realistische netwerken. De toepasbaarheid van het model is aangetoond met behulp van een grootschalige casestudie voor de Metropoolregio Amsterdam. De resultaten laten zien dat verschillende netwerkconfiguraties relevant zijn voor verschillende marktpenetratiegraden van automatische voertuigen. Voor lage marktpenetratiepercentages zijn AV-subnetwerken die geschikt zijn voor automatische voertuigen in gemengd verkeer de meest efficiënte configuratie. Boven een marktpenetratie van 30% komen doelgroepstroken voor automatische voertuigen vaker voor in de optimale netwerktopologieën, maar bij zeer hoge marktpenetratiegraden worden AV-subnetwerken juist weer vaker geselecteerd.

Op basis van de ontwikkelde modellen en genoemde casestudies zijn in dit proefschrift zes algemene conclusies getrokken. Deze worden hieronder uitgelicht.

1. De penetratiegraad van automatische voertuigen is de meest dominante factor die de prestatie van het wegennet beïnvloedt. Dit fenomeen is in alle hoofdstukken van dit proefschrift waargenomen.

2. Het bi-level modelleringsraamwerk dat in dit proefschrift is gebruikt, is essentieel voor ontwerp en evaluatie van infrastructuurplannen. Dit komt doordat het effect van infrastructuuraanpassingen afhankelijk is van de mate waarin reizigers hun reiskeuze aanpassen als gevolg van de infrastructuuraanpassingen.

3. Connectiviteit van subnetwerken voor automatische voertuigen is een onmisbare voorwaarde voor hun effectiviteit. We hebben vier evolutionaire oplossingsmethoden ontwikkeld die zijn toegesneden op de structuur van de in dit proefschrift bestudeerde problemen en waarmee op efficiënte wijze coherente subnetwerken kunnen worden ontworpen voor casestudies met grootschalige wegennetten. We hebben aangetoond dat algemene metaheuristieken met penaltyfuncties niet efficiënt zijn om met de connectiviteitseis om te gaan en dat de nieuwe specifiek voor dit probleem ontwikkelde algoritmen een superieure prestatie leveren.

4. Een groot deel van alle baten die gerealiseerd zouden kunnen worden door wegen geschikt te maken voor automatisch rijden kunnen behaald worden door een optimale selectie van links in een AV-subnetwerk op te nemen waardoor de investeringskosten veel lager zijn dan als alle links geschikt gemaakt zouden moeten worden voor automatisch rijden. Bovendien heeft de ontwikkeling van AV-subnetwerken duidelijke voordelen op het gebied van veiligheid en netwerkprestatie tijdens de overgangsperiode.

5. Verschillende netwerkconfiguraties om automatische voertuigen in wegennetwerken op te nemen zijn relevant voor verschillende marktpenetratiegraden van automatische voertuigen. Voor een lagere marktpenetratie zijn AV-subnetwerken die geschikt zijn voor automatische voertuigen in gemengd verkeer de meest efficiënte configuratie. Vanaf een marktpenetratiegraad van ongeveer 30% worden doelgroepstroken voor automatische voertuigen ook een efficiënte oplossing.

6. Bij de keuze van de netwerkconfiguratie spelen wegtypen een cruciale rol. Op- en afritten van autosnelwegen, wegen met maar één rijstrook en grote regionale wegen met delen waar maar één rijstrook beschikbaar is, zijn geschikt gebleken voor automatisch rijden in gemengd verkeer. Doelgroepstroken zijn het meest geschikt voor autosnelwegen.

Toekomstige onderzoeksrichtingen met betrekking tot de methodologische verbeteringen omvatten het overwegen van nieuwe doelstellingen voor het bovenste niveau van het netwerkontwerpprobleem, het ontwikkelen van nieuwe diffusiemodellen voor automatische voertuigen die minder afhankelijk zijn van schaalparameters, het gebruik van dynamische verkeerstoedelingsmodellen en het gebruik van multimodale verkeerstoedelingmodellen, inclusief meer vervoerwijzekeuzemogelijkheden voor reizigers.

Contents

List of Acronyms and Abbreviations

ACC	Adaptive cruise control
AD	Automated driving
ADS	Automated driving system
AV	Automated vehicle
BPR	Bureau of public roads
CACC	Cooperative adaptive cruise control
CNDP	Continuous network design problem
CT	Computation time
CTM	Cell transmission model
DDT	Dynamic driving task
DNDP	Discrete network design problem
DTA	Dynamic traffic assignment
DUE	Deterministic user equilibrium
EGS	Evolutionary greedy search
ELS	Evolutionary local search
EPS	Evolutionary policy search
FP	Fixed-point
GA	Genetic Algorithms
IIA	Independence of irrelevant alternatives
KPI	Key performance indicator
LLMP	Lower level mathematical problem
MD	Manual driving
MGA	Modified Genetic Algorithms
MNDP	Mixed network design problem
MNL	Multinomial logit
MPR	Market penetration rate
MSA	Method of successive averages

MUC	Multi-user class
NDP	Network design problem
NDP-T	Time-dependent network design problem
NEM	Network equilibrium model
OD	Origin-destination
ODD	Operational design domain
OEDR	Object and event detection and response
OEM	Original equipment manufacturer
OF	Objective function
PCE	Passenger car equivalent
PCU	Passenger car unit
PSL	Path-size logit
PT	Public transport
RV	Regular vehicle
SUE	Stochastic user equilibrium
TAC	Total adjustment cost
TDTCS	Total discounted travel cost saving
TTC	Total travel cost
TTD	Total travel distance
TTT	Total travel time
ULMP	Upper level mathematical problem
V2I	Vehicle to infrastructure
V2V	Vehicle to vehicle
VI	Variational inequality
VoD	Value of distance
VoT	Value of time
VoTD	Value of travel distance
VoTT	Value of travel time

1 Introduction

1.1 Background

1.1.1 From Futurama to the Future

For almost a century, the introduction of self-driving cars has seemed twenty years away. The first radio-controlled driverless car was introduced to the public on the McCook Air Force test base in Dayton, Ohio, USA on 5 August, 1921 (Kröger, 2016). This experiment indicates that technical efforts for developing self-driving vehicles started at least a century ago.

The first time a vision of a future with autonomous vehicles was introduced to a large audience in spectacular fashion was Norman Bel Geddes' Futurama exhibit at the 1939 World's Fair, sponsored by General Motors. Bel Geddes later described his visionary ideas for automated highways in his book Magic Motorways (Bel Geddes, 1940). From the very onset, whether in a pictorial vision in a popular science magazine (Figure 1.1), or in a descriptive vision published in a technical book (Bel Geddes, 1940), the vision of the automated car was accompanied by automated highways.

The efforts to develop self-driving vehicles continued throughout the following decades and various driving automation features were gradually introduced to new vehicles. However, the hype for self-driving cars gained considerable momentum during the 2010s with Tesla vehicles equipped with Autopilot and Waymo, formerly the Google self-driving car project, which could drive autonomously on certain conditions, yet these vehicles were by no means fully autonomous. In recent years, many car manufacturers have added automated driving features to their new vehicles or announced plans for doing so; nevertheless, a fully autonomous vehicle is not on the horizon yet.

But why should self-driving cars exist? David H. Keller's fiction story of a driverless car published in 1935 (Keller, 1935) is a case in point of the humankind's early interest in self-driving cars:

> "Old people began to cross the continent in their own cars. Young people found the driverless car admirable for petting. The blind for the first time were safe. Parents found they could more safely send their children to school in the new car than in the old cars with a chauffeur." (p. 1470).

On the other hand, the image of self-driving cars is almost perfect for traffic safety campaigns. With many studies showing that human driver error is a contributing factor to car accidents, replacing the error-prone humans with machines practically suggests itself. This may seem plausible at first glance, but the narrative's blind spot is the machine, which is usually portrayed as an infallible wonder-agent. Decades of attempts to perfect the machine have brought us ever closer to an autonomous agent that can drive us everywhere, but we are not quite there yet.

Although fully autonomous cars are not expected in the near future, we are now closer than ever to highly automated vehicles whose implications for mobility and urban planning can be profound. Recently, many studies have predicted substantial gains in safety, traffic efficiency and environmental impacts by replacing regular vehicles (RVs) with automated cars. A review of these studies is provided in (Milakis et al., 2017b).

Figure 1.1. An automated car in an automated highway (Mechanix Illustrated, June 1953, p. 58.)

1.1.2 Driving Automation

Several terms such as autonomous, driverless, self-driving and automated have been used frequently to describe vehicles with automation functions. However, these terms do not always refer to the same concept and they require more clarification. Therefore, this section provides taxonomy and definitions first.

Driving automation systems refer to the hardware and software that are collectively capable of performing part or all of the dynamic driving task (DDT) on a sustained basis (SAE International, 2018). This might include sustained longitudinal and lateral vehicle motion control, object and event detection and response (OEDR), and the fall back of DDT (i.e., handling the situation when there is a system failure) within a certain driving terrain or operational design domain (ODD) within which the vehicle is designed to operate or use the automation system.

Based on whether the driver is responsible for the aforementioned tasks or the system, and the extent of the vehicle's ODD, the Society of Automotive Engineers has defined five levels of automation (SAE International, 2018), which are demonstrated in Figure 1.2. At levels 1-2, the driving automation system provides the driver with longitudinal and lateral control, i.e., adaptive cruise control and lane keeping. Such technologies are already available in the automobile market and they can operate on existing infrastructure. However, at these levels, the driver is still responsible for monitoring the environment. At level 3, the automated driving system (ADS) monitors the environment and executes driving tasks on certain ODDs (e.g., driving in motorways), allowing the drivers to avert their attention from driving tasks while being ready to take back control in case of a failure in ADS or approaching situations beyond ODD of level-3 ADS, i.e., difficult driving conditions that level-3 ADS is unable to handle. Level-4 ADS is expected to handle the fail-safe situation autonomously; however, the ODD would still be limited. This implies that levels 3-4 might need dedicated infrastructure or roads with specific requirements. Finally, at level 5, the ADS is expected to be feasible for all driving modes and completely self-sufficient. This final level of automation signals a major step in the prospect of mobility, but it is not expected in the near future (Shladover, 2016).

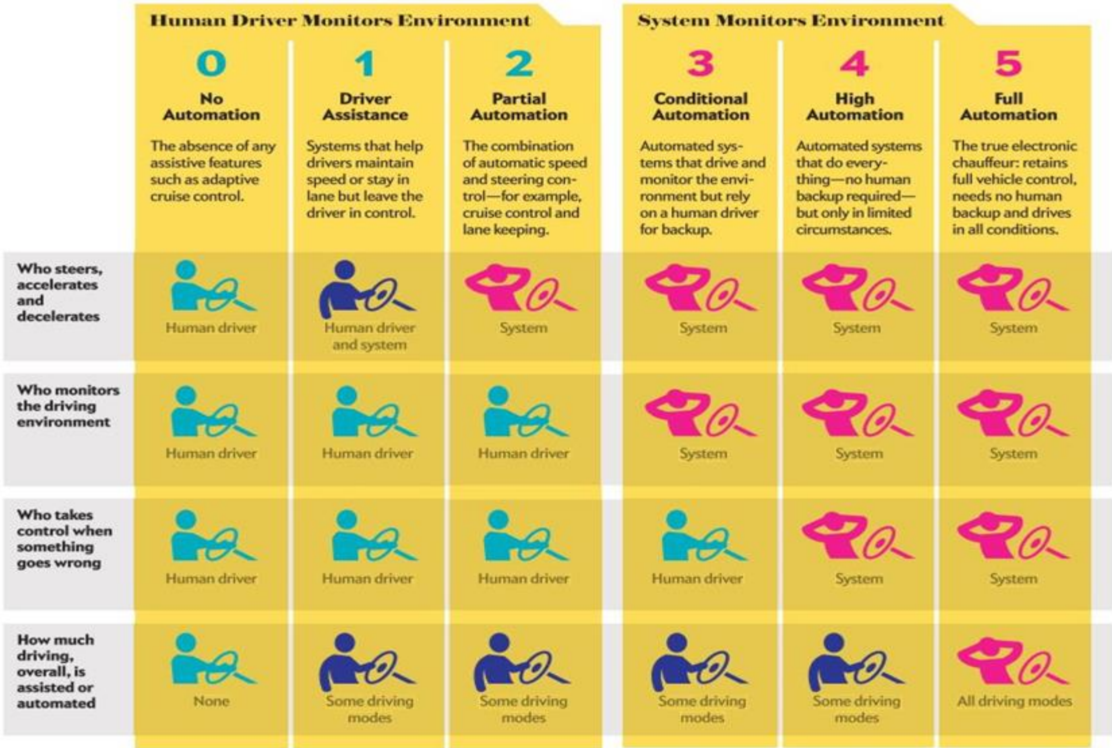


Figure 1.2. Levels of vehicle automation according to SAE International

Note: The presented classification of automation levels was introduced by SAE International; however, this illustration is due to (Shladover, 2016).

Simply stated, an ADS builds a perception of its environment via input from digital maps, cameras and sensors. This perception might be enhanced by communication with other vehicles as well as the infrastructure. The traffic rules are programmed beforehand and its control algorithm determines the best course of action in each situation. Moreover, the ADS may learn from its previous experience, share its data with the software manufacturer, and get periodic software updates from the manufacturer to improve its performance. However, certain environments might be too complex for existing ADSs to fully perceive, which means they might show errors in accurately recognizing objects and patterns or appropriately reacting to events.

Since an ADS must always be capable of coping with its environment, its ODD should be clearly defined for the driver (or the vehicle operator), for the vehicle itself, and for the road operator. However, using field tests on existing level-2 vehicles in the Netherlands, Farah et al. (2020) observed a persistent mismatch between the ODD specified by the original equipment manufacturer (OEM) and by the drivers. This indicates that precisely defining ODDs is crucial for safety of vehicles with a limited ODD. However, the exact extent of level-3 and level-4 ODDs are not clear yet.

The main factors defining the ODDs are the quality of infrastructure, the complexity of the environments the vehicle encounters, and the vehicles' capabilities in terms of sensors and actuation to cope with these environments. The complexity of an environment in this context is defined by traffic state, weather conditions, visibility and types of interactions with regular vehicles as well as other road users, which can also depend on the infrastructure.

Since level-3 and level-4 vehicles' capabilities are unknown at the moment, defining their exact ODD is difficult. Madadi et al. (2018) have found correlations between road attributes and their suitability for level 3-4 ADSs based on expert opinion. Soteropoulos et al. (2020) have proposed an automated drivability index for different roads in the network of Vienna to determine which roads are suitable for level-4 vehicles from a technological standpoint. They conclude that deployment of level-4 vehicles in streets with low drivability index would only be possible with major infrastructure adjustments. These insights can aid in predicting level 3-4 ODDs according to infrastructure characteristics. Nonetheless, clearly defining ODDs remains a challenging task due to the inconsistencies in quality standards of various road types and infrastructure segments within and among different countries.

1.1.3 Infrastructure for Automated Driving

Relying on driving automation technology alone without infrastructure support might compromise the potential safety and efficiency gains of automated vehicles (AVs). Infrastructure support can include physical infrastructure elements such as proper road pavement, clear and harmonized lane markings and road signs, as well as changing topology to minimize complex interactions between AVs and other road users, and digital infrastructure elements such as high-definition digital maps and vehicle to infrastructure (V2I) communication. These road infrastructure elements (can) have a long lifetime and adjusting them can be costly. Hence, there is a strong need for research and planning to ensure that large infrastructure investments provide the highest societal benefits. Numerous recent research projects and action plans regarding road network infrastructure for AVs have recognized the need for infrastructure planning for AVs (Carreras et al., 2018; EU-EIP, 2016; Gyergyay et al., 2018; International Transport Forum, 2015; Johnson, 2017; Lu et al., 2019; Ulrich et al., 2020).

Besides, the analyses of ADS disengagements that have occurred during AV tests in the U.S. provide more evidence in support of road network infrastructure's major influence on safe and efficient operation of AVs. In the states in the U.S. where AV tests are allowed, vehicle manufacturers are obliged to provide publicly available reports of their prototype vehicle's ADS failures and disengagements. Several studies have analyzed these reports to provide insight into safety and efficiency of existing AVs.

According to (Dixit et al., 2016), nearly 10% of ADS disengagements have happened due to road infrastructure (which is the second cause after system failures), and 56% due to system failures, which can also be associated with vehicle's disability to deal with the environment. Additionally, the number of disengagements differs significantly based on road type: it is lower in motorways and higher in urban streets. According to (Favaro et al., 2017), about 90% of the reported accidents involving an AV happened at an intersection, with 32% occurring in city roads. It is stated in (Favarò et al., 2018) that less than 13% of all disengagements happened in motorways, freeways and arterial roads combined, and the rest of the disengagements (around 87%) happened in interstate roads and urban streets. Based on the evidence of existing level-2 AVs, drivers are very likely to use driving automation systems in freeways and unlikely to use them on rural and urban roads (Hardman et al., 2019). These studies signal the need for attention to road network infrastructure and its crucial role in safe operation of AVs.

There is a school of thought that believes ADSs can improve themselves in time via machine learning, thereby denying the need for infrastructure support for ADSs. However, many researchers believe this is an issue that needs to be addressed. Two main approaches can be observed among studies that attempt to address this problem. One approach is adjusting the infrastructure for safe and efficient operation of AVs in mixed traffic. Another is network-wide deployment of dedicated infrastructure for AVs via network design concepts that mainly aim at traffic efficiency. These approaches are discussed in the following paragraphs.

With recent advances in V2I technologies and cloud services, the concept of infrastructure-enabled automated driving as a promising alternative to autonomous driving has been proposed in the literature (Gopalswamy and Rathinam, 2018; Li et al., 2020). The concept includes V2I communication through road side units as well as enhanced GPS, cloud services and broadcast radio. Other studies have specified general infrastructure requirements for safe operation of AVs. A review of these studies is provided in (Farah et al., 2018). However, these requirements can be idealistic, expensive and difficult to meet. In particular, for road types with low volumes, such as small urban streets that primarily serve to provide accessibility to origins and destinations, large investments might be unnecessary.

Another approach suggested in recent studies to address the issue is via network design concepts such as optimal networks of dedicated AV lanes (Chen et al., 2016; Conceição et al., 2020), dedicated AV links (Conceição et al., 2017; Ye and Wang, 2018), and dedicated AV zones (Chen et al., 2017). In these studies, the objective is to maximize the societal benefits, expressed in terms of total system travel time or cost, provided by the deployment of dedicated AV infrastructure (i.e., dedicated AV lanes, links and zones). The decision variables are where to deploy these dedicated lanes, links and zones in road networks in order to observe maximum societal benefits.

Network design methods mentioned in previous paragraph are promising and potentially cost-effective since they utilize strategic approaches that consider the whole network and deal with the problem via optimized design concepts; nonetheless, these design concepts require more details and extensions to become operational. Application of these concepts on large-scale real networks is a crucial yet missing next step, mainly since with real networks, many practical issues and considerations are involved that are not observed with theoretical networks used in these studies. Besides, dedicating parts of a network to one class of vehicles can compromise accessibility of other classes and modes. Furthermore, dedicated AV lanes, links and zones may be underutilized when market penetration rate of AVs is relatively low.

The issue of low AV market penetration rate is particularly relevant for the transition period to full automation. Reaching a high market penetration rate of AVs is a continuing process that can take decades. Thus for a long time, low volumes of AVs and a heterogeneous mix of traffic with AVs and RVs on the roads are inevitable. Network design concepts for this transition period must consider mixed traffic, at least on certain parts of road networks.

Further, designing network concepts for AVs, which optimize the tradeoff between infrastructure investments and their societal benefits, is a gradual process that depends on the level of AV demand. Since the demand for AVs is likely to increase over time, optimal network configurations for AVs should evolve over time as well. An efficient design for a network with a low level of AV demand may not be an efficient design for the same network with a very high level of AV demand. Moreover, future developments regarding AVs are highly uncertain. Relevant design methods need to be able to deal with this uncertainty and to suggest no-regret measures that are applicable for several likely scenarios. As a result, time-dependent network design methods are necessary to cope with the problem.

In addition, the demand for cars and the performance of road networks cannot be assessed without considering alternative modes. Automated driving and new network topologies might affect travel choices, such as mode choice. This might lead to an increase of AV usage. Accordingly, the travelers' response to network configurations and their travel demand elasticity should be captured via multimodal models. Time-dependent network design methods that suggest no-regret measures for AVs and use multimodal traffic assignment models appear to be lacking in the academic literature.

Therefore, the main scientific gap to be dealt with in this thesis is a methodology for determining where and when to invest in road network infrastructure for AVs in order to optimize the network-level traffic performance during the transition period to full automation.

1.2 Research Objective, Research Questions and Approach

The main objective of this thesis is to develop a road network design method for strategic planning of infrastructure investment decisions that can be applied in practice on real-sized networks to cope with the uncertain development path of AVs during the transition period to full automation. This method includes network configurations for mixed traffic and dedicated infrastructure. In addition, its spatial scope and temporal evolution is established. In order to reach this objective, the following key research questions must be addressed.

1. What is a suitable road network configuration for accommodating AVs and what are the impacts of this configuration on network performance?

2. How can network configurations for AVs be optimized and what are the impacts of the optimal configuration on network performance?

3. How should the optimal road network configurations evolve over time and what are the impacts of the network and AV demand evolution on network performance?

4. How can different network configurations for accommodating AVs on road networks be combined and what are the impacts of combining different network configurations on network performance?

In order to address these research questions, a modular and incremental approach is used to develop a comprehensive mathematical model through four main steps. Each step addresses a research question and corresponds to a main chapter of this thesis. The general approach is to start with a primary model to understand the fundamental properties of the problem, and gradually add more complexity to the model by considering new dimensions of the problem in each step to finally develop a model that can realistically capture the main aspects of the problem at a desirable level. The applicability of the method developed in each chapter is demonstrated via a case study. Figure 1.3 illustrates the research approach.

Step 1: First, important road characteristics for the safe operation of AVs are identified. Then, a traffic assignment model is developed to capture the essential behavioral differences of AVs and RVs, and their impacts on the travelers' route choice and the propagation of traffic through the road networks. A network configuration referred to as AV-ready subnetwork is developed in this step. Via several scenarios, its impacts on performance of road networks are assessed using a case study of the road network of the city of Delft, the Netherlands. This exploratory study paves the way for developments in the following steps by providing a research agenda for the topic. This step is represented by **chapter 2**.

Step 2 extends the traffic assignment method developed in step 1 to a bi-level network design method where the upper level represents the infrastructure design decisions made by planners to deploy the AV-ready subnetwork, and the lower level represents the travelers' response to these decisions via their mobility choices that lead to traffic patterns in the network. The objective is to optimize the trade-off between infrastructure investment costs and the societal benefits provided by them. Complexity of the bi-level network design problem along with the extra requirements imposed on solution methods in order to find coherent AV-ready subnetworks necessitate a new solution method. Therefore, a new algorithm is developed in this step and its performance is compared with two existing algorithms in the literature. This model and the solution methods are tested on the network of Delft, and the impacts of the optimal AV-ready subnetworks on road network performance are discussed. **Chapter 3** represents this step.

Step 3 uses the model developed in step 2 as a starting point and incorporates the time dimension into the model as well as the evolution of AV demand over time by considering a long planning horizon and a multi-stage approach. This approach defines the optimal timeline for the evolution of AV-ready networks over time in response to the growing AV demand. In addition, a multi-mode multi-class model for traffic assignment is developed in this step to consider the demand for public transport as well. Solving this problem also calls for a new algorithm. Therefore, two new algorithms are developed for solving this problem and their performances are compared using a case study of a realistic road network of the Amsterdam metropolitan region, which is much larger than the network of Delft. Furthermore, the effects of designs suggested by the method on transport network performance are presented and discussed in this step. This step is represented by **chapter 4**.

Step 4 uses the findings of previous steps as well as new developments in the academic literature to combine the design concept developed in steps 1 and 2 with two other network configurations recently proposed in the literature, namely, dedicated AV lanes and dedicated AV links, to develop a unified modeling framework for designing road networks for AVs including various infrastructure design concepts. Considering combinations of the aforementioned design concepts introduces new types of decision variables to the problem as well as extra constraints to guarantee the logical consistency of the combined network topologies. Therefore, a new solution method that utilizes the specific properties of this problem to find coherent network topologies in an efficient manner is developed in this step to solve the problem for large-scale networks. The applicability of this model on real-sized networks is demonstrated using another case study of the Amsterdam metropolitan region. This step is represented by **chapter 5**.

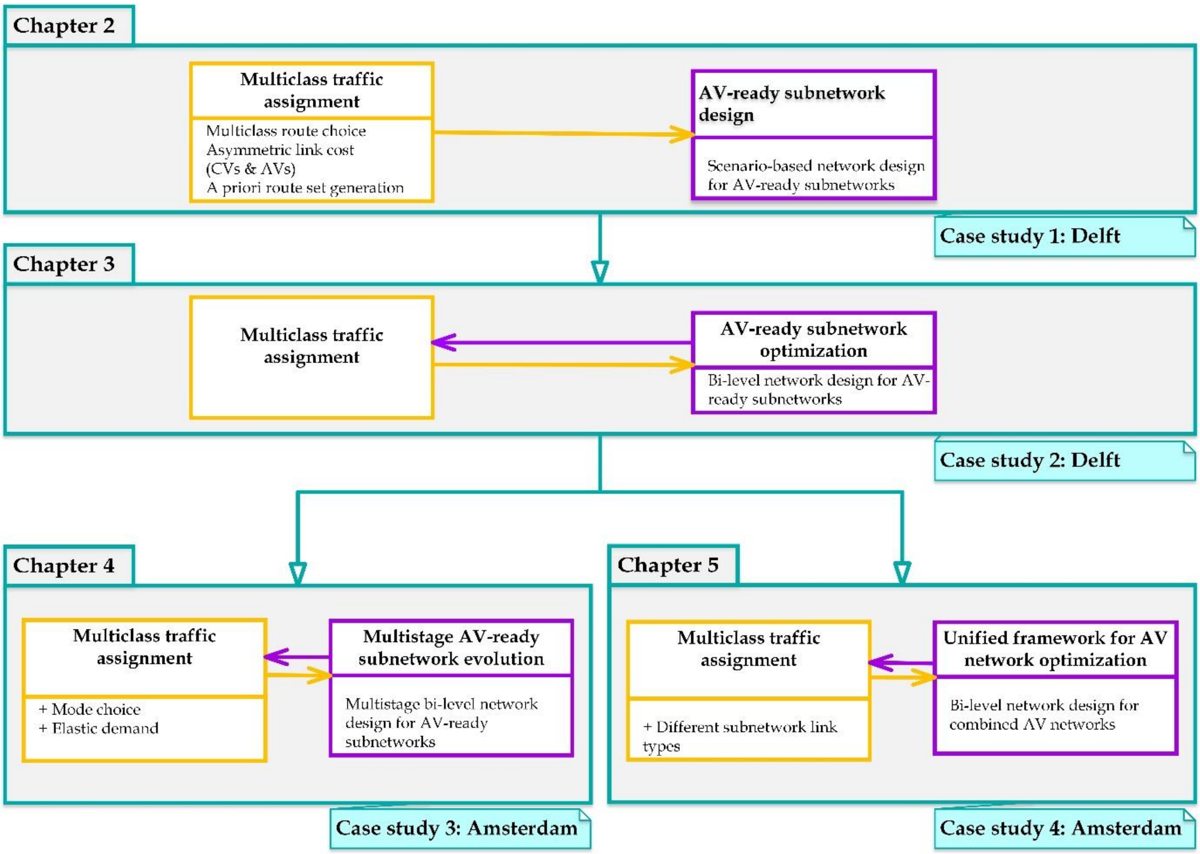


Figure 1.3. Research approach

Note: model components that represent transport demand (i.e., travel behavior) are shown in orange boxes and the ones that represent infrastructure supply decisions are represented in magenta boxes. Colors and directions of arrows signify interactions between model components as well as dependencies between chapters.

1.3 Contributions

The main contributions of this thesis are categorized into two types, namely, scientific contributions and societal relevance. The following sections provide more details for each category of contributions.

1.3.1 Scientific Contributions

In this thesis, a multidisciplinary perspective is applied using several scientific fields such as discrete choice modeling, traffic flow theory, game theory, operations research and automotive engineering. This multidisciplinary perspective culminates in a broad methodology for evaluating network infrastructure decisions regarding AVs. The methodology considers the behavioral aspect of travelers' mobility choices, the policy aspect of urban and regional planners' decisions regarding the road network infrastructure, the technical aspect of AV technologies, and the interactions between aforementioned aspects of the problem. The models developed here include several components that can be utilized in various types of transport models for other applications as well. Each component is represented with a box in Figure 1.3 and briefly discussed below.

Multi-class traffic assignment

A novel traffic assignment model is developed in chapter 2 to capture the behavioral differences of RVs and AVs, and model the traffic propagation through road networks in presence of AV-ready roads. This model is comprised of an a priori route set generation, a multi-class logit-based route choice, and a macroscopic static network loading model with asymmetric link costs for RVs and AVs. The model can be used independently, as it has been used in chapter 2, or to represent the lower level in bi-level road network design problems, as it has been used in chapter 3 of this thesis. In chapter 4, the lower level model is extended to consider demand elasticity via a mode choice component. This component can also be used as an independent traffic assignment model or for multi-stage bi-level road network design methods. Finally, in chapter 5, the traffic assignment model is further extended and used as the lower level in a bi-level design model. This model combines various network configurations for accommodating AVs in road networks. Each one of these versions can be used in various other traffic and transport planning applications as well.

AV-ready subnetwork design

The concept of AV-ready subnetworks is introduced in chapter 2. An AV-ready subnetwork refers to a road network configuration, which is applicable for the transition period to full automation. It includes selecting a subset of roads deemed (potentially) suitable for AVs and upgrading them with infrastructure adjustments to facilitate safe and efficient operation of ADSs in mixed traffic. In this chapter, the impacts of this configuration on network performance are explored via several scenarios using the traffic assignment model described above. As a result of this exploratory study, a research agenda for the topic is presented at the end of this chapter. Many items in that agenda are delivered in subsequent chapters of this thesis.

AV-ready subnetwork optimization

Deploying AV-ready subnetworks entails selecting roads for the subnetwork and upgrading these roads with infrastructure adjustments that can be costly. Therefore, a novel bi-level network design model is developed in chapter 3 to optimize the tradeoff between investment costs and the network performance benefits provided by AV-ready subnetworks. The upper level of the bi-level problem represents infrastructure decisions to deploy AV-ready subnetworks (the supply side), and the lower level represents the travelers’ response to the supply of infrastructure (the demand side).

Optimization of multi-stage AV-ready subnetwork evolution

Since optimal AV-ready subnetworks should change over time in response to the demand for AVs, a novel multi-stage optimization model is developed in chapter 4 to determine the optimal timing for AV-ready subnetwork evolution.

Unified framework for AV network optimization

Recently, other network configurations for accommodating AVs in road networks, such as dedicated AV lanes and links, have been suggested in the literature. Different and incompatible mathematical models are proposed in the literature so far for modeling and optimization of various network design concepts. Consequently, it is not possible for transport planners to compare the performance of different design concepts on their transport networks or to assess the impacts of a combination of these concepts. Therefore, in chapter 5, a general bi-level modeling framework is developed to combine three network design concepts for AVs, namely, AV-ready subnetworks, dedicated AV links and dedicated AV lanes in one unified model.

Algorithmic and computational developments

Applicability of the models developed in four main chapters of this thesis is demonstrated via a case study for each chapter. A semi-real network of Delft is used for the first two main chapters and a real-sized network of the Amsterdam metropolitan region is used for the last two main chapters. Due to the complexity of network design problems, applications of such models on large-scale networks such as the one used for the last two case studies in this thesis are rare in the literature. Nevertheless, these case studies are scientifically relevant and even essential for understanding different dimensions of these problems. Furthermore, this complexity and problem size render most solution methods in the literature ineffective for the case studies considered here. Also, creating coherent AV-ready subnetworks imposes a graph connectivity requirement on solution methods, which has not been considered previously in the literature. Therefore, four novel solution methods (one in chapter 3, two in chapter 4 and one in chapter 5) are developed throughout this thesis to solve case studies 2-4 efficiently, and to suggest coherent networks as solutions. These solution methods can be applied to solve other network design problems in the literature as well.

The algorithm developed in chapter 3 is suitable for network design problems with binary decision variables. The algorithms introduced in chapter 4 are appropriate for multi-stage network design problems with binary decision variables. Finally, the algorithm presented in chapter 5 is proper for network design problems with integer decision variables. The common characteristic of all these solutions is that they are tailored to the structure of the problems considered in this dissertation. They all start with a population of small graphs as initial feasible solutions, and gradually make improvements on these graphs until no further improvement is possible. The improvements are made using operations that are inspired by evolutionary processes yet are carefully crafted to preserve coherence of the graphs for applications in real case studies. The population-based approach of these algorithms is aimed at preventing them from stopping at local optima. Besides, this population-based approach facilitates the use of high-performance parallel computing architectures and contributes to computational efficiency of these algorithms. This efficiency is a necessity when dealing with large-scale realistic road networks as the ones used in case studies of this thesis.

1.3.2 Societal relevance

During this project, regular periodic meetings and workshops were held with practical partners of the project. The partners included planners and authorities in municipalities, metropolitan regions, province offices and the Dutch Ministry of Infrastructure and Environment, along with a number of transport and mobility consultants in the Netherlands. The purpose of these meetings was assuring the practical value of the research, collecting input from the partners and presenting as well as discussing the scientific results with the partners. Stakeholders of this project (excluding academics) are categorized into three main groups and the societal relevance of the thesis is discussed for each group separately.

Policymakers and transport planners

The first contribution of this project for policymakers as well as urban and regional planners is raising awareness of the problem considered in this thesis. Through regular workshops and meetings, discussions were held with transport planners and policymakers in the Netherlands to explain the problem and to make them aware of its significance. Next, this thesis presents possible strategies to cope with the problem. These strategies entail where and when to invest on road networks, possible infrastructure adjustments, and potential regulations to actuate each strategy in order to maximize societal benefits and minimize the total investment cost. Models developed in this thesis provide policymakers and planners with tools to evaluate these strategies, and the case studies provide them with demonstrations to recognize potential impacts of these strategies in practice.

In addition, since the case studies in this thesis represent Dutch cities, they are of special interest for planners in the Netherlands. Nonetheless, their results demonstrate how network design concepts such as AV-ready subnetworks, dedicated AV lanes and dedicated AV links can be deployed in real road networks, and what their potential impacts are in terms of travel choices, road network performance, the usage of different road types by different vehicles, and the distribution of impacts among different classes of users. Therefore, the case studies provide transport planners everywhere with a wide array of options to prepare their network for the transition period to full automation.

Transport modelers and consultants

Transport consultants develop and use transport models to aid policymakers in evaluating decisions and strategies regarding transportation. Therefore, they can use models and solution methods developed in this thesis or incorporate them into their existing models to evaluate transport policies and to guide policymakers in making informed decisions. This means all scientific contributions mentioned in previous section are relevant for transport consultants as well, especially since these models and solution methods are shown to be applicable for complex real-sized networks by means of four case studies.

Road users

The main contribution of this thesis to road users is proposing strategies that consider safety of all road users as the first priority and suggesting no-regret measures that can be beneficial for all road users regardless of the future development path of automated vehicles. Network design concepts and the infrastructure adjustments proposed in this thesis are aimed at guaranteeing safety for all road users. Furthermore, the objective of all design problems formulated in this thesis include minimizing total system travel cost via improving traffic efficiency. This means less generalized travel cost for all road users, thereby less travel resistance and higher accessibility.

1.4 Thesis Organization

The structure of this thesis is as follows. Chapter 2 introduces the concept of AV-ready networks; chapter 3 presents a methodology for designing and optimizing AV-ready subnetworks; chapter 4 introduces a method for the optimal evolution of AV-ready subnetworks over time; and chapter 5 introduces a unified framework for designing road networks for AVs. Chapter 6 summarizes the main findings of this thesis, discusses how these findings aid in achieving the main research objective, and mentions the shortcomings of the methodology developed in this thesis as well as guidelines for its practical implementation.

2 Multi-Class Impact Assessment of Automated-Vehicle-Ready Subnetworks

In order to develop a network design method for AVs, first, a suitable network configuration must be specified and assessed. The main objective of this chapter is to suggest and explore a suitable network configuration, i.e., AV-Ready Subnetworks, to accommodate AVs in road networks. This addresses the first research question raised in this dissertation. This objective has been accomplished through the following steps. First, suitable roads for this network configuration are determined based on their attributes. Next, a traffic assignment model is developed to assess the impacts of the configuration on network performance. Then, variations in roads to include in the subnetwork as well as variations in demand for AVs are explored via several scenarios in a case study, and a sensitivity analysis is performed to evaluate the sensitivity of the model as well as the assessment results and implications to variations in input parameters. Finally, based on the findings of this chapter, a research agenda for the topic is proposed in the conclusions, which steers the direction of research in subsequent chapters of this thesis.

The term "automated driving (AD) subnetwork" used in this chapter refers to the concept of automated-vehicle-ready (AV-ready) subnetwork. Moreover, the term value of time (VoT) used in this chapter refers to the concept of value of travel time (VoTT).

This chapter is based on the following journal article:

Madadi, B., van Nes, R., Snelder, M., & van Arem, B. (2019). Assessing the travel impacts of subnetworks for automated driving : An exploratory study. *Case Studies on Transport Policy*, *7*(1), 48–56. https://doi.org/10.1016/j.cstp.2018.11.006

2.1 Introduction

With recent technological and strategic advancements in automobile industries and transportation sectors, there are new possibilities for the future of mobility. Automated driving (AD) is one of the promises of the future. According to (SAE International, 2016), there are five levels of vehicle automation; at levels 1-2, the driving automation system provides the driver with longitudinal and lateral control (i.e., adaptive cruise control and lane keeping). Such technologies are already available in the automobile market and they can operate on existing infrastructure. At level 3, automated driving system (ADS) monitors the environment and executes driving tasks on certain operating design domains (ODD) (e.g. driving in motorways), allowing the drivers to avert their attention from driving tasks while being ready to take back control in case of a failure in ADS or approaching situations beyond ODD of level-3 ADS (i.e., difficult driving conditions that level-3 ADS is unable to handle). Level-4 ADS is expected to handle the fail-safe situation autonomously; however, the ODD would still be limited. This means that levels 3-4 might require dedicated infrastructure or roads with specific requirements. Finally, at level 5, ADS is expected to be feasible for all driving modes and completely self-sufficient. This last level of automation signals a major evolution in the prospect of mobility, but it is not expected in the near future (Shladover, 2016).

AD is a trend that will evolve over time, both in the level of automation and the market penetration rate of automated vehicles (AVs). Many studies focus on the impacts of AD for the case that the total fleet is fully automated (SAE level 5 with unlimited ODD); however, it might take a long time before this situation is realized. In the transition period, there will be a mix of different levels of automation, including level 0 (i.e., non-automated vehicles). According to (SAE International, 2016), ODD of level 3-4 ADS is limited. That means vehicles equipped with ADS at these levels cannot drive in automated mode everywhere. The question of where can level 3-4 automated vehicles drive safely in automated mode has not received enough attention in the academic literature yet. Furthermore, this can have implications for the usage of transport networks and travel choices.

We envision that AD using ADS levels 3-4 will be allowed on a selection of roads and for the remaining roads, manual driving will be compulsory (albeit supported by various assisting driving automation systems such as collision avoidance systems). In these selected roads, AD will be allowed in mixed traffic conditions (i.e., in the same lanes with non-automated vehicles) and these roads need investments to fulfil certain design requirements to facilitate safe and efficient AD. There is therefore a need for a network design approach to decide which roads should be selected to facilitate AD using ADS levels 3-4 in the transition stage.

The aim of this chapter is to explore the concept of a specific network configuration for ADS levels 3-4 and to estimate its impacts on travel time, distance and cost in urban regions having mixed traffic. Network configuration refers to the selection of links on which level 3-4 ADS is facilitated.

Our focus is on network configuration concepts and their impacts on route choice behavior and consequently, on generalized travel cost. Yet, it should be noted that AD can have several impacts on travel choices including location, destination, mode, departure time, and route choice as well as driving behavior. For an extensive review of existing literature on impacts of AD, the reader is referred to (Milakis et al., 2017b).

We introduce the concept of AD subnetwork as a possible network configuration for the transition period, elaborate on our approach for link selection, and explain relevant operational concepts. In order to evaluate network usage and route choice impacts of this configuration, we introduce a modified static multi-user class (MUC) stochastic user equilibrium (SUE) traffic assignment method with a path-size logit as well as a Monte Carlo labeling combination approach for a priori route set generation. Then, we present a case study to demonstrate the concept of AD subnetwork, and we analyze its impacts on a specific network.

Based on this analysis and an exploratory literature study, we present a research agenda for the development of a network design method to accurately model and assess the impacts of various network configurations for AVs in urban regions.

The rest of this chapter is organized as follows: section 2.2 explains the problem background and an explorative literature study; section 2.3 introduces the concept of AD subnetwork and our evaluation method; section 2.4 presents the case study and numerical results; and the last two sections present discussion, conclusion, and recommendations.

2.2 Background

In order to provide a systematic discussion on the literature, we categorize relevant studies into three major clusters, namely, microsimulations, macro simulations, and network design problem (NDP). Since this is an exploratory study, and there is no other study in the literature that addresses our problem with the same method, we do not place our work within any of these classes. Instead, we refer to the most relevant studies in each category and mention their common aspects with our study with the aim of positioning our work within the literature.

2.2.1 Microsimulations

One of the major envisioned advantages of AD is the possibility of cooperative adaptive cruise control (CACC). Shladover et al. (2015) provide clear definitions and operating concepts of CACC. Main benefits of adaptive cruise control (ACC), i.e., improving traffic flow and fuel efficiency, are expected to be realized with cooperative ACC (CACC) rather than autonomous ACC. CACC with vehicle to vehicle (V2V) communication could reduce the average driving time headway from 1.4 seconds (current average for manual driving) to approximately 0.6 seconds (Nowakowski et al., 2010) which would increase lane capacity. With reduced time headways, at 100% penetration rate of CACC-equipped vehicles, it is possible to increase lane capacity in motorways without bottlenecks from 2200 v/h to about 4000 v/h (Shladover et al., 2012). Using the microscopic MIXIC traffic simulation model of a highway bottleneck, van Arem et al. (2006) conclude that CACC has the potential to improve traffic stability and throughput depending on market penetration rate and traffic volume. The extent of positive impacts becomes greater with higher penetration rates (>60%) and higher traffic volumes.

Several microsimulation studies consider dedicated lanes for CACC-equipped vehicles. Using the MIXIC simulation model on a four-lane highway, van Arem et al. (2006) conclude that only with high CACC penetration rates for the highway stretch before the bottleneck with high traffic volume, the case with dedicated CACC lane has a better performance compared to the case without the special lane. However, in the scenario with 20% CACC penetration, severe congestion is observed before the lane drop. By means of a microsimulation framework including varying behavioral mechanisms for connected and automated vehicles, Mahmassani

(2016) concludes that dedicated lanes for AVs can only be effective if their use is optional and when the market share of AVs is larger than the percentage of nominal capacity represented by that lane.

Microsimulation experiments are flexible tools to study local and specific impacts of AVs under different scenarios using various micro models. However, existing studies focus on specific stretches of motorways. To the best of our knowledge, there is no published microsimulation study on urban streets. Moreover, they are infeasible for network-wide studies.

2.2.2 Macro Simulations

An alternative approach for assessing the impacts of AD at network level is to use macroscopic traffic assignment models. Some researchers have used the expected impacts of AD from the literature along with certain behavioral assumptions for AVs to develop macroscopic models to study their impacts on transport networks. Many studies conclude that deployment of CACC leads to faster reaction times, thereby reducing driving time headways which can increase lane capacity (see, for instance (Mahmassani, 2016; Nowakowski et al., 2010; Shladover et al., 2012, 2015; van Arem et al., 2006)). In a macroscopic static traffic assignment, this can be modeled via lower passenger car unit (PCU) values or higher link capacities. The magnitude of this change depends on the proportion of AVs on the link and is reported for several future scenarios in (Puylaert et al., 2018) where the authors use a system dynamic approach to quantify the impacts of early forms of automation. Another expected impact from CACC is fuel efficiency (Rios-Torres and Malikopoulos, 2017; Shida and Nemoto, 2009; Shladover et al., 2015). This can be included in generalized travel cost used in macroscopic traffic assignment models via value of distance (VoD). Additionally, value of time (VoT) might change in parts of the trips where AD is possible. Although there is no consensus in the literature about the effect of AD on VoT, some studies conclude that the possibility of performing other activities in AVs might lead to lower VoT (Milakis et al., 2017a; Puylaert et al., 2018). These expected effects suggest that AVs could be modelled as a separate user class.

Levin and Boyles (2015) present a multi-class four-step model using a static traffic assignment that includes AV repositioning to avoid parking fees, increases in link capacity as a function of proportion of AVs, and several classes in demand based on VoT and AV ownership. The study focuses only on the differences between no automation and full automation (i.e., level 0 and level 5 SAE). A multi-class cell transmission model (CTM) is developed in (Levin and Boyles, 2016) to model the differences in capacity and backward wave speed in shared regular vehicle (RV) and AV roads. This model is used in (Patel et al., 2016) to study the effects of reservation controls and increased capacity from AVs on highway and arterial networks.

2.2.3 Network Design Problem

In transportation literature, optimal decisions regarding adjustments to and expansions of road network infrastructure are considered within the concept of NDP (Yang and Bell, 1998). So far, very few studies have proposed new network configurations for AVs and considered them within the concept of NDP. Chen et al. (2016) consider the problem of optimal deployment of AV lanes as a bi-level NDP where the upper level includes decisions such as where, when, and how many lanes should be considered as dedicated lanes for AVs and the lower level includes network equilibrium with two classes representing RVs and AVs. The study presents a possible network configuration for AD, and a network-wide assessment of its impacts using a macroscopic static traffic assignment with an MUC deterministic user equilibrium (DUE) route choice model.

Another network configuration is presented in (Chen et al., 2017) where Chen et al. consider the problem of optimal AV zones in transport networks. An AV zone includes links that are adjusted for AVs. RVs are not allowed in AV zones. So, different classes of vehicles encounter different network topologies. As for routing, they consider a deterministic mixed routing model where system optimal routing is applied for AV zone and user equilibrium for the rest of the network.

2.2.4 An Alternative Network Configuration

So far, the only studies considering AD concepts in NDP and offering specific network configurations are (Chen et al., 2016) and (Chen et al., 2017). A possible improvement on both studies is considering a clear definition of AVs (i.e., their automation level) as well as their ODD. Furthermore, in order to apply these methods in practice, some extensions to their network representation are necessary to consider various road types in urban settings.

This chapter offers a more realistic network configuration compared to dedicated lanes and AV zones for the transition period. We select certain parts of the network mainly consisting of roads with flow and distribution function to allow for AD (road functionality is discussed in the next section). Adjustments to these roads include (but are not limited to) improvements in quality of on/off ramps, lane markings, road and traffic signs as well as rearranging intersections with uncontrolled complex conflicts and separating inhomogeneous traffic. For an overview of possible adjustments the reader is referred to (Courbon et al., 2016; Farah et al., 2018; Nitsche et al., 2014; Zhang, 2013). These adjustments can improve safety for all road users, regardless of market penetration rate of AVs and development path of AD in the future.

Therefore, the problem becomes choosing links to adjust in order to construct a subnetwork to allow and facilitate AD in mixed traffic. This chapter presents a qualitative scheme for this selection. For assessing the first-order impacts of rerouting in this configuration, a modified MUC static stochastic traffic assignment model is utilized. The modifications are based on the expected impacts of AD from the literature which was discussed earlier in this section and will be elaborated on in the next section.

2.3 AD Subnetworks

In this section, we introduce and elaborate on the concept of AD subnetwork. Furthermore, we propose a model to evaluate the performance of this configuration. Mathematical formulation and the solution method for this model are discussed in this section as well.

2.3.1 Constructing the AD Subnetwork

In order to envisage a network configuration for AD, it is essential to specify a feasible realm of operation for ADS levels 3-4. Four major criteria are considered in defining the feasibility of roads for AD; roads with limited access, high quality (e.g. pavement, lane marking, traffic signs, and lights), segregated traffic (homogeneity of mass and speed for vehicles in each lane), and grade separated or clear at-grade intersections are regarded as feasible. Additionally, roads with potential for having such standards with reasonable adjustments are added to the set of feasible links. Adjustment costs and optimizing the link choice set are not included in this chapter but debated in the discussion.

Network hierarchy and road function are defining factors for road standards and their potential for accommodating AVs mixed with RVs carrying the least possible risk of conflicts. The categorization of roads based on their function was first introduced in (Buchanan, 1963). Later this became the basis for mono-functionality principle of sustainable safety vision in the Netherlands (Wegman et al., 2008; Wegman and Aarts, 2006). The principle entails that roads must have a single function and their design and use should comply with that function. Facilitating traffic throughput (mobility or flow function) and providing access to destinations (accessibility or access function) are two possible road functions. Distribution function is a third category introduced to offer the appropriate transition between providing access and facilitating throughput. Although finding a clear correspondence between road function and network hierarchy is not always a straightforward task, freeways mainly serve mobility by facilitating throughput (flow function) and most local roads serve accessibility by providing access to adjacent parcels (access function). It should be noted that roads are not always designed to serve a single function.

Road network observations in Delft, the Netherlands reveal that all roads with flow function and the majority of roads with distribution function (potentially) meet mentioned standards. In contrast, none of the roads with access function meets the standards. Then the process is reduced to approving roads with flow function, rejecting roads with access function and examining the distributors to specify AD subnetwork. Figure 2.1 (the case study section) depicts the constructed AD subnetworks for the case of Delft which is discussed in details later in this chapter.

2.3.2 Operational Concepts and Assumptions

Our assumptions regarding the AD subnetwork concept and how it is used by RVs and AVs are summarized below:

- Level-3 and level-4 ADS-equipped vehicles (which will be referred to as AVs for the remainder of this chapter) form CACC platoons (whenever possible) in automated mode only within the AD subnetwork and this concept is referred to as automated driving (AD);

- Level 0-2 vehicles or regular vehicles (RVs) do not form CACC platoons and AD is not possible for them, although they can use assisting driving automation systems which should not be confused with ADS;

- AVs always start manually and proceed in manual driving mode till reaching AD subnetwork (green parts in Figure 2.1);

- Upon arrival to AD subnetwork, the ADS notifies the driver of the possibility of AD and the driver opts for AD;

- When reaching one of the boundaries of the AD subnetwork, ADS notifies the driver again to take back control and resume manually;

- The driver must be ready at all times to take back control, particularly when exiting AD subnetwork and in case of a failure in level-3 ADS;

- Outside the AD subnetwork (blue parts in Figure 2.1) all vehicles drive manually;

- All vehicles are allowed everywhere in the network but AD is only possible inside the AD subnetwork for AVs.

2.3.3 Multi-Class Route Choice and Network Equilibrium

Modeling AVs' routing behavior within the AD subnetwork requires considering the effects of AD on generalized travel cost, which is the determining factor for route choice. The main expected effects of AD levels 3-4 on generalized travel cost are capacity improvements via shorter headways, lower VoT through the use of time for other activities in AVs, and lower VoD due to fuel efficiency of CACC. We adopt a generalized travel cost function consisting of the summation of distance weighted by VoD, and travel time weighted by VoT *(Eq.2.13)*. On the links within the AD subnetwork, lower VoD and VoT values are applied for AV class *(Eq.2.14)*. Link travel time is based on a modified BPR function where total flow is a weighted sum of class-specific flows to capture the correlation between link capacity and the proportion of AVs on the link *(Eq.2.7)*. Weighting is based on PCU values. Lower PCU values are applied for AVs on links within the AD subnetwork to account for the shorter gaps between AVs and their leading vehicles *(Eqs.2.2-3)*. The magnitude of changes in PCU, VoD and VoT are reported in Table 2.1 and discussed in the next section.

We argue that for the transition period, system optimal route choice is highly unlikely, therefore the most realistic approach to frame the network equilibrium problem is an MUC SUE formulation. Fisk (1980) presents the formulation of the single class logit-based SUE assignment as a mathematical programming problem. An early extension of the problem to an MUC SUE is introduced in (Daganzo, 1983). Most common formulations of the SUE problem are based on the multinomial logit (MNL) model due to its closed form and efficient computation times. However, the known issue of independence of irrelevant alternatives (IIA) in MNL models can lead to overestimation of flow for overlapping routes. Several extensions to the MNL model have been discussed in (Chen et al., 2012) where the performance of existing extensions to the MNL model are compared. The path-size logit (PSL) model presented in (Ben-Akiva and Bierlaire, 1999) is one of the extensions that can lead to more realistic flow predictions. In this chapter, an MUC extension of PSL SUE formulation is presented. Different formulations for PSL are reported in the literature. The one adapted here is based on the formulation introduced in (Ben-Akiva and Ramming, 1998). The mathematical formulation of this method is presented in the next subsection. Our main modeling assumptions for route choice are summarized below:

- Route choice is based on an MUC SUE where within the AD subnetwork, AV class has a lower generalized travel cost due to lower VoT and VoD resulting from CACC, but in the remaining parts of the network all classes have identical generalized travel cost;

- Link travel time within the AD subnetwork depends on the proportion of AVs. A weighted sum of class-specific flows based on their PCU value is used to calculate link travel time *(Eqs.2.2,3,7)*.

2.3.4 Mathematical Formulation

The following notation is used throughout this chapter.

W	Set of origin-destination pairs w
R^w	Set of routes r between origin-destination pair w
M	Set of user classes m
A_0	Set of links a not belonging to AD-enabled subnetwork
A_1	Set of links a belonging to AD-enabled subnetwork
A	Set of all links a in the network; $A_0 \cup A_1$
μ_m	Logit choice model parameter for class m
D_m^w	Demand of origin-destination pair w for class m
$PS_m^{w,r}$	Path-size penalty of route r between origin-destination pair w for class m
β_m	Path-size correction parameter for class m
η_m	Value of time for class m
t_a^0	Free flow travel time of link a
θ_m	Driving cost per kilometer for class m (VoD)
l_a	Length of link a
γ_m	PCU value of class m
$\delta_{m,a}^{w,r}$	1 if flow of w from route r for class m uses link a, 0 otherwise (assignment map)
α_a	BPR function parameter for link a
b_a	BPR function parameter for link a
Λ_a	Capacity of link a
$t_a(q_a)$	Travel time of link a
q_a	Total flow of link a based on weighted sum of class-specific flows (PCU equivalent)
$F_m^{w,r}$	Flow in route r between O-D pair w for class m (class-specific route flow)
$f_{m,a}$	Flow in link a for class m (class-specific link flow)
$C_m^{w,r}$	Travel cost of route r between O-D pair w for class m (route cost)
$c_{m,a}$	Travel cost of link a for class m (link cost)
$\overline{f}_{m,a}$	Equilibrium flow of class m in link a
$\overline{t}_a$	Equilibrium travel time of link a
$P_m^{w,r}$	Proportion of travelers of class m taking route r between O-D pair w
TTC	Total travel cost
TTT	Total travel time
TTD	Total travel distance

Traffic assignment: network equilibrium

Our PSL-based MUC SUE formulation for route choice is presented here as a mathematical programming problem.

MP:

Min

$$Z = \sum_{m} \frac{1}{\mu_m} \sum_{w \in W} \sum_{r \in R^w} F_m^{w,r} \ln F_m^{w,r} - \sum_{m} \frac{1}{\beta_m} \sum_{w \in W} \sum_{r \in R^w} F_m^{w,r} \ln PS_m^{w,r}$$

$$+ \sum_{m \in M} \sum_{a \in A} \int_0^{q_a} c_{m,a}(x) dx, \quad 2.1$$

s.t.

$$q_a = \gamma_0 (f_{0,a} + f_{1,a}), \qquad \forall a \in A_0, \quad 2.2$$

$$q_a = \gamma_0 f_{0,a} + \gamma_1 f_{1,a}, \qquad \forall a \in A_1, \quad 2.3$$

$$\sum_{r \in R^w} F_m^{w,r} = D_m^w, \qquad \forall w \in W, \forall m \in M, \quad 2.4$$

$$\sum_{w \in W} \sum_{r \in R^w} F_m^{w,r} \delta_{m,a}^{w,r} = f_{m,a}, \qquad \forall a \in A, \forall m \in M, \quad 2.5$$

$$F_m^{w,r} \geq 0, \qquad \forall w \in W, \forall m \in M, \forall r \in R^w. \quad 2.6$$

First equation introduces the objective function with three terms: the first one maximizing the entropy of flows leading to stochastic route flows, the second one including the path-size penalty for overlapping routes, and the third term minimizing individual generalized travel costs. *Eqs.2.2-3* introduce the PCU equivalent of total flow as a PCU-weighted sum of class-specific flows which is used in calculating link travel time. *Eq.2.4* guarantees total route flows satisfy total demand, and *Eq.2.5* converts route flows to corresponding link flows for each class. Link travel time function is given as:

$$t_a(q_a) = t_a^0 [1 + \alpha_a (\frac{q_a}{\Lambda_a})^{b_a}]. \quad 2.7$$

And link cost per class is:

$$c_{0,a}(q_a) = \theta_0 l_a + \eta_0 t_a(q_a), \quad \forall a \in A, \quad 2.8$$

$$c_{1,a}(q_a) = \theta_0 l_a + \eta_0 t_a(q_a), \quad \forall a \in A_0, \quad 2.9$$

$$c_{1,a}(q_a) = \theta_1 l_a + \eta_1 t_a(q_a), \quad \forall a \in A_1. \quad 2.10$$

Eqs. 2.9-10 guarantee different link travel costs for AV class within and outside the AD Subnetwork using different VoT and VoD values. The solution to the above MP formulation gives the following probability for route proportions:

$$P_m^{w,r} = \frac{\exp(-\mu_m C_m^{w,r} + \beta_m \ln PS_m^{w,r})}{\sum_{r \in R^w} \exp(-\mu_m C_m^{w,r} + \beta_m \ln PS_m^{w,r})} \qquad \forall w \in W, \forall m \in M, \forall r \in R^w. \quad 2.11$$

Path-size penalty is defined as:

$$PS_m^{w,r} = \sum_{a \in r} (\frac{l_a}{l_r})(\frac{1}{\sum_{r \in R^w} \delta_{m,a}^{w,r}}). \quad 2.12$$

And route-based travel cost (generalized travel cost) for two classes are given as:

$$C_0^{w,r} = \sum_{a \in A} \delta_{0,a}^{w,r} F_0^{w,r} (\theta_0 l_a + \eta_0 t_a(q_a)), \quad 2.13$$

$$C_1^{w,r} = \sum_{a \in A_0} \delta_{1,a}^{w,r} F_1^{w,r} (\theta_0 l_a + \eta_0 t_a(q_a)) + \sum_{a \in A_1} \delta_{1,a}^{w,r} F_1^{w,r} (\theta_1 l_a + \eta_1 t_a(q_a)). \quad 2.14$$

Eq.2.14 guarantees that for AV class, parts of the routes within the AD subnetwork have different generalized travel cost via different VoT (η_m) and VoD (θ_m) values, and different link travel times depending on PCU equivalent of class-specific flows (q_a) which is calculated based on *Eqs.2.2-3*.

Impacts: network performance criteria

The most common performance criteria used in the literature for assessing the impacts of road network configurations are total travel cost (TTC), total travel time (TTT), and total travel distance (TTD) (Yang and Bell, 1998). *Eqs. 2.15-17* represent these metrics, respectively. These values are based on equilibrium flows and travel times. Impacts of RVs and AVs in AD subnetwork in equilibrium conditions are based on the following formulae.

$$TTC = \sum_{a \in A_0} (\eta_0 \bar{t}_a + \theta_0 l_a)(\overline{f}_{0,a} + \overline{f}_{1,a}) + \sum_{a \in A_1} [(\eta_0 \bar{t}_a + \theta_0 l_a)\overline{f}_{0,a} + (\eta_1 \bar{t}_a + \theta_1 l_a)\overline{f}_{1,a}], \quad 2.15$$

$$TTT = \sum_{a \in A} \bar{t}_a (\overline{f}_{0,a} + \overline{f}_{1,a}), \quad 2.16$$

$$TTD = \sum_{a \in A} l_a (\overline{f}_{0,a} + \overline{f}_{1,a}). \quad 2.17$$

2.3.5 Route-Set Generation

One particular importance of route-set generation for modeling AV behavior is to capture specific route sets that might become attractive for AVs due to the changes in their VoT, VoD, and PCU value. Considered route sets in traffic models must include these routes as well. For instance, in the case of Delft, any route that is (partially) within the AD subnetwork (potentially) has a lower travel cost for AVs. These changes might make some routes that at first sight seem long and unusual become relevant alternatives for AVs due to their lower costs. This indicates the need for new route-set generation approaches to generate realistic route sets for AVs.

Common route-set generation methods do not generate such routes but some methods have the potential to serve this purpose. In this chapter, the Monte Carlo labeling combination method introduced in (Catalano and van der Zijpp, 2001) is used with some adjustments to generate appropriate route sets for AVs. In addition to common labels (travel cost, travel time, travel distance), a label with a multiplier (with a value between 0 and 1) is used for the cost of links within the AD subnetwork to generate more routes that cross the AD subnetwork but are too expensive for RVs. This is to ensure that the longer routes within AD subnetwork which can become feasible due to higher efficiency of AD are included in the considered route sets for AVs. For generating routes for RVs, no adjustment has been made on the original method.

2.3.6 Solution Algorithm

There are several solution algorithms in the literature for the MUC SUE problem. A review of these algorithms is provided in (Noriega and Florian, 2007). The problem with presented formulation in this chapter can readily be solved using the solution method developed in (Wu et al., 2006) where the authors introduce an MUC extension of MSA algorithm.

2.4 Case Study

A case study is used to demonstrate the impacts of AVs in AD subnetwork modeled with the proposed method. In this case, a network similar to the road network in Delft, The Netherlands is used in order to observe some practical issues related to road types in real networks. The network and demand patterns are based on a tutorial project for the transport modelling software package *OmniTRANS*. It includes 1,151 links, 494 nodes and 22 zones.

Demand for AVs is considered via seven scenarios based on different market penetration rates of AVs. It is assumed that the ratio of AVs to RVs available to travelers is the same as the AV market penetration rate and that the AVs available to travelers are homogeneously distributed in all zones. Three different network configurations and three network variants for the second configuration are used for experiments:

- Base case network: this is the reference point for comparison with all other cases and is the regular Delft network including all the links in Figure 2.1 as non-AD links $(A_0 = A, A_1 = \varnothing)$.

- AD subnetwork: this network is shown in Figure 2.1 $(A_0 \cup A_1 = A, A_0 \cap A_1 = \varnothing)$. The subnetwork for AD (main variant) covers 38% of the overall distance in the network. Moreover, two additional variants of this subnetwork (shown in Figure 2.1) are used in scenario analysis to showcase the impacts of link selection for AD subnetwork.

- AD everywhere network: this is used to demonstrate the extreme impacts for comparisons and it includes all links in Figure 2.1 as AD links $(A_1 = A, A_0 = \varnothing)$.

There are several road types in this network representation. Apart from the connectors which are artificial links connecting zone centroids to the network, four major categories are recognized that signify network hierarchy, namely, freeways, regional roads, main urban roads, and local roads. Mentioned list is in the descending order in terms of network hierarchy. For the (main variant of) AD subnetwork, all local roads (lowest level according to network hierarchy) are considered infeasible for AD subnetwork and all freeways (highest level) are considered feasible. For the remaining road types, a selection is made based on road function, potential quality, traffic segregation, and complexity of relevant intersections.

Studied impacts are total travel cost, total travel time, and total travel distance which were introduced earlier. Furthermore, the distribution of impacts for each network type, network variant, demand scenario, road type, and user class is investigated.

Figure 2.1 AD subnetwork variants (from left, respectively): variant 3, variant 2, and variant 1

Note: links that belong to the AD subnetwork are shown with (bright) green and the rest with (dark) blue.

Input parameters related to modeling AVs are provided in Table 2.1. There is no consensus in the literature on exact magnitude of changes in PCU, VoT, and VoD values as a result of AD and there is no possibility of validating different results at the moment. Therefore, we have chosen similar values to those used in (Puylaert et al., 2018) and performed a sensitivity analysis to demonstrate the sensitivity of outcomes to the input parameters. The parameters for the PSL model are also reported in Table 2.1. The base case demand and network data are from the base case in the Delft project in *OmniTRANS* software introduced earlier.

This case study using the AD subnetwork design method is implemented in MATLAB and the code is available from the authors upon request.

Table 2.1 Important input parameters related to AD for Delft case study

Parameter	Class	Penetration rate		Bandwidth (sensitivity analysis)
		[0%-40%]	[40%-100%]	
PCU	RV	1	1	-
	AV	0.95	0.9	[0.7-1.1]
VoT (€/h)	RV	9	9	-
	AV	8.55	8.55	[6.3-9.9]
VoD (€/km)	RV	0.19	0.19	-
	AV	0.18	0.16	[0.13-0.21]
μ	RV	0.2	0.2	-
	AV	0.29	0.29	-
β	RV	3	3	-
	AV	3	3	-

Ten scenarios using 7 demand patterns, 3 network configurations, and 3 network variants for the second configuration are considered in this chapter, and for each case, key performance indicators are calculated separately for each road type and each user class using the model. Furthermore, sensitivity analyses are performed for AD parameters used in the model. We focus the discussion on key performance indicators (TTC, TTT, TTD) with a distinction per road type, plus a sensitivity analysis. All numbers reported are indexed with respect to the base case scenario, unless stated otherwise in the caption.

Table 2.2 provides a comparison of key network performance indicators for three variants of AD subnetwork using 50% AV penetration rate scenario including separate results for different classes. No AD with 0% AV penetration rate and AD everywhere with 100% AV penetration scenarios represent the two ends of the spectrum with no impacts and highest impacts, respectively, as potential lower and upper bounds. Road types on which AD is facilitated in each scenario are indicated with checkmarks (✓) in relevant columns. Different variants of AD subnetwork as well as AD everywhere scenario are considered in order to provide insight into the impacts of higher distance coverage of AD subnetwork. As demonstrated in Table 2.2, TTC and TTT further improve in AD subnetwork variants that include more road types (more links). Moreover, TTD is slightly higher in variants 1 and 2 compared to the main variant. This can be explained by lower accessibility of AD subnetwork in these variants due to their lower network density that leads to rerouting to longer routes.

Table 2.2 Indexed travel impacts for different variants of AD subnetwork

Network Type		**No AD (base case)**	**AD subnetwork variants**			**AD everywhere**
Freeways		-	✓	✓	✓	✓
Regional roads		-	-	✓	✓	✓
Main urban roads		-	-	-	✓	✓
Local roads		-	-	-	-	✓
Network Variant		**Variant 1 (Main)**	**Variant 1**	**Variant 2**	**Variant 3 (Main)**	**Variant 1 (Main)**
AV Penetration Rate		**0%**	**50%**	**50%**	**50%**	**100%**
Parameter Ratio (X_{AV}/X_{RV})		-	**PCU_{AV}/PCU_{RV} = 0.90 VOT_{AV}/VOT_{RV} = 0.95 VOD_{AV}/VOT_{RV} = 0.85**			
Total Travel Cost	RV	100.00	49.92	49.89	49.88	0.00
	AV	0.00	47.79	46.89	46.50	88.98
	Overall	100.00	97.71	96.78	96.38	88.98
Total Travel Time	RV	100.00	49.83	49.74	49.72	0.00
	AV	0.00	50.03	49.97	49.95	98.50
	Overall	100.00	99.86	99.71	99.67	98.50
Total Travel Distance	RV	100.00	50.00	50.00	50.00	0.00
	AV	0.00	50.30	50.38	50.13	100.23
	Overall	100.00	100.30	100.38	100.13	100.23

Note: Checkmarks (✓) indicate road types on which AD is facilitated in each scenario and indexing is based on the base case scenario.

Table 2.3 summarizes the changes in total travel time, cost and distance for all demand scenarios, network configurations, and main network variants compared to the base case. A significant and steady decrease in total travel cost, a minor decrease in total travel time, and a minor increase in total travel distance are observed with increase in AV market penetration rate. The only exception is the decrease in total travel distance in AD everywhere scenario compared to AD subnetwork with 100% AV penetration rate. This is explained by the fact that most of the induced travel distance in AD subnetwork cases is the result of rerouting towards the subnetwork, whereas in the AD everywhere scenario there is no need for rerouting since AD is possible everywhere. Yet, there is an increase in travel distance in this case compared to the base case due to lower cost of distance and time for AVs.

Table 2.3 Indexed travel impacts for all network types, demand scenarios, and the main AD subnetwork variant

Network Type		No AD (Base Case)	AD Subnetwork (main variant)						AD Everywhere
AV Penetration Rate		0%	10%	30%	50%	70%	90%	100%	100%
Parameter Ratio (X_{AV}/X_{RV})		$PCU_{AV}/PCU_{RV} = 0.95$ $VOT_{AV}/VOT_{RV} = 0.95$ $VOD_{AV}/VOT_{RV} = 0.95$			$PCU_{AV}/PCU_{RV} = 0.90$ $VOT_{AV}/VOT_{RV} = 0.95$ $VOD_{AV}/VOT_{RV} = 0.85$				
Total Travel Cost	RV	100.00	89.97	69.94	49.88	29.90	9.96	0.00	0.00
	AV	0.00	9.71	29.11	46.50	65.05	83.58	92.84	88.98
	Overall	100.00	99.68	99.05	96.38	94.95	93.54	92.84	88.98
Total Travel Time	RV	100.00	89.94	69.86	49.72	29.79	9.92	0.00	0.00
	AV	0.00	10.04	30.08	49.95	69.82	89.65	99.55	98.50
	Overall	100.00	99.98	99.94	99.67	99.61	99.55	99.55	98.50
Total Travel Distance	RV	100.00	90.00	70.00	50.00	30.00	10.00	0.00	0.00
	AV	0.00	10.02	30.05	50.13	70.18	90.24	100.26	100.23
	Overall	100.00	100.02	100.05	100.13	100.19	100.24	100.26	100.23

Note: Indexing is based on the base case scenario.

There is a shift of traffic, as evidenced by total travel distance in Table 2.4, from local roads and freeways to regional roads and main urban roads. The pattern is evident in all scenarios with AVs and is intensified with higher AV penetration rates. On the other hand, travel time and cost in various road types follow a different trajectory. In local roads, travel time and cost are slightly lower compared to the base case but this is only due to less traveled distance. In freeways, the improvements in travel time and cost are more significant as a result of the higher efficiency gained through AD. Finally in regional roads and main urban roads, an improvement in travel cost is observed as a result of AD efficiency despite the increasing travel distance and time.

Since different values for the changes in AD parameters (i.e., PCU, VoT, and VoD) as a result of AD efficiency are reported in the literature and there is no real data for validation, it is appropriate to perform a sensitivity analysis in order to assess possible changes in results with deviations in these parameters. A summary of the sensitivity analyses for PCU, VoT, and VoD is demonstrated in Table 2.5. Rows with even numbers are eliminated; nonetheless, the presented results are sufficient to observe that changes in parameters within a realistic range of values have limited influence on the results. Obviously, the travel costs are the most sensitive measures as they are directly affected by the values for VoD and VoT. However, when looking at TTT and TTD, the impacts are minimal. It is essential to notice that these results are valid for a certain range of parameters that we considered sensible; however, larger deviations might cause deeper and more significant effects, not only on route choice but also on other travel choices which might require more elaborate modeling as well.

Table 2.4 Indexed distribution of impacts for all user classes in different road types

Road Type		Freeways	Regional Roads	Main Urban Roads	Local Roads	All Roads
0% Penetration Rate (Base Case)						
Total Travel Cost		41.12	12.01	9.25	12.45	100.00
Total Travel Time		30.53	10.70	10.08	16.72	100.00
Total Travel Distance		49.86	13.09	8.56	8.92	100.00
50% Penetration Rate in AD Subnetwork (main variant)						
Total Travel Cost	RV	20.47	5.98	4.61	6.23	49.88
	AV	18.00	5.52	4.37	5.93	46.50
	Overall	38.47	11.50	8.98	12.16	96.38
Total Travel Time	RV	15.07	5.29	5.02	8.37	49.72
	AV	15.00	5.47	5.28	8.00	49.95
	Overall	30.07	10.76	10.30	16.37	99.67
Total Travel Distance	RV	24.93	6.54	4.28	4.46	50.00
	AV	24.82	6.81	4.51	4.23	50.13
	Overall	49.75	13.35	8.79	8.68	100.13
90% Penetration Rate in AD Subnetwork (main variant)						
Total Travel Cost	RV	4.08	1.19	0.92	1.25	9.96
	AV	32.30	9.91	7.85	10.70	83.58
	Overall	36.38	11.10	8.78	11.95	93.54
Total Travel Time	RV	2.99	1.05	1.00	1.68	9.92
	AV	26.76	9.78	9.49	14.45	89.65
	Overall	29.75	10.83	10.49	16.13	99.55
Total Travel Distance	RV	4.99	1.31	0.86	0.89	10.00
	AV	44.67	12.26	8.11	7.61	90.24
	Overall	49.66	13.57	8.97	8.50	100.24

Note: Indexing is based on the values of 'all roads' column in the base case scenario and numbers for connectors are eliminated.

Table 2.5 Sensitivity analysis summary for 90% AV market penetration rate scenario in the main variant of AD subnetwork

Analyzed AV Parameter	Parameter Ratio (X_{AV}/X_{RV})	Vehicle Type	Total Travel Cost	Total Travel Time	Total Travel Distance	Other AV Parameters	
Passenger Car Unit (PCU)	**0.7**	RV	9.88	9.81	9.94	**VOT_{AV}/VOT_{RV} = 0.95**	**VOD_{AV}/VOT_{RV} = 0.85**
		AV	83.03	88.60	89.99		
		Overall	92.92	98.42	99.93		
	0.9	RV	9.93	9.94	9.94		
		AV	83.46	89.61	89.98		
		Overall	93.40	99.55	99.92		
	1.1	RV	10.02	10.12	9.94		
		AV	84.24	91.44	89.96		
		Overall	94.26	101.56	99.90		
Value of Time (VoT)	**0.7**	RV	9.93	9.94	9.94	**PCU_{AV}/PCU_{RV} = 0.90**	**VOD_{AV}/VOT_{RV} = 0.85**
		AV	78.34	89.65	90.14		
		Overall	88.28	99.57	100.08		
	0.9	RV	9.93	9.94	9.94		
		AV	82.44	89.63	90.01		
		Overall	92.38	99.55	99.95		
	1.1	RV	9.93	9.92	9.94		
		AV	86.52	89.59	89.89		
		Overall	96.45	99.53	99.82		
Fuel efficiency (VoD)	**0.7**	RV	9.93	9.94	9.94	**PCU_{AV}/PCU_{RV} = 0.90**	**VOT_{AV}/VOT_{RV} = 0.95**
		AV	78.21	89.63	90.13		
		Overall	88.15	99.57	100.06		
	0.9	RV	9.93	9.94	9.94		
		AV	85.21	89.61	89.93		
		Overall	95.14	99.53	99.87		
	1.1	RV	9.93	9.92	9.94		
		AV	92.17	89.57	89.75		
		Overall	102.10	99.51	99.69		

Note: Indexing is based on values of the base case

2.5 Discussion and Conclusions

In this chapter, different scenarios are used to gain insight into the impacts of a possible AD network configuration (AD subnetwork). A regular network with no AV market penetration is considered as the base case in order to provide a point of reference for the relative changes in each scenario. Also, a scenario where AD is allowed everywhere and all the vehicles in the network are AVs (i.e. 100% AV penetration) is simulated to illustrate the highest possible impacts.

Based on this study, the differences in impacts between AD everywhere and AD subnetwork with 100% penetration rate are not large. This means that AD subnetwork with high AV penetration rates can deliver a great share of benefits obtainable from AD everywhere (i.e., AD subnetwork with 100% penetration rate can deliver 64% of cost benefits of AD everywhere with 100% penetration rate). Given that AD everywhere is only possible for level-5 AVs and that AD subnetwork introduced here is suitable for level 3-4 AVs as well, it can be concluded that it is possible to realize most benefits of level-5 automation in urban regions with AD subnetwork only having level 3-4 AVs. It should be noted that only route choice is considered here and that changes in destination and mode choice due to usage of AVs can also affect network performance. As AVs become more appealing in time, they can attract users of other modes. Moreover, facilitating AD in certain parts of the network can affect location and destination choices in long term in favor of zones with better accessibility to AD-friendly parts of the network, especially given the relatively large impact on TTC (which is an input for choices of mode and destination).

The results support the expectation that AV market penetration rate is the dominating factor to affect the impacts. There is a sharp change in the impacts after 40% AV penetration rate (partially due to the changes in parameters) and the effects become more significant with higher AV penetration rates. Although, the overall impacts, even in the AD everywhere scenario, are rather low with maximum observed changes in TTC, TTT and TTD being 11.02%, 1.5% and 0.26%, respectively.

Sensitivity analysis shows that the parameters individually have limited impact at network level in urban regions, and their deviations, within a realistic range, do not affect the results significantly. It appears that only the combination of all three AD parameters (i.e. PCU, VoT, and VoD) along with the new considered route sets for AVs can lead to significant changes.

The observed patterns in the shift of traffic between different road types are expected to repeat themselves with AD subnetwork deployment in different network types since there is a clear rationale for this shift; AD subnetwork is more efficient and desirable for AVs and is expected to attract more traffic. However, main urban roads and regional roads within the AD subnetwork are used more than freeways due to their higher accessibility (i.e., for most origins and destinations, the closest access point to the AD subnetwork is via main urban roads and regional roads).

As for the AD subnetwork distance coverage, the variants including more road types (higher coverage) are shown to be more effective in terms of network performance, yet the increase in the performance is not necessarily linear. On the other hand, including more links in the AD subnetwork requires a higher adjustment cost. Therefore, finding the optimal trade-off between adjustment costs and benefits acquired from the AD subnetwork calls for optimization methods. Changes in general demand level (congestion level) can affect the optimal link choice as well.

Regarding the method proposed in this chapter, we believe the mechanisms are valid and generalizable for assessing the route choice impacts of AD at network level. Although, there are limitations to this approach and improvements to the model are possible through the following model components that constitute the research agenda for this topic:

- *Quantitative optimization methods:* the choice of links in this chapter is based on a qualitative analysis. Another alternative is to define feasible links with the same procedure and formulate a bi-level optimization problem to find the optimal link choice (i.e., upper level decisions) within feasible links in the AD subnetwork in equilibrium conditions (i.e., lower level optimization). In addition to travel cost, time, and distance, other criteria could be specified to analyze trade-offs between adjustment costs and benefits in the optimization problem.

- *Dynamic traffic assignment (DTA):* AVs are expected to cause changes in fundamental diagrams and flow-density relationships. These changes as well as queueing and spill back can accurately be captured in DTAs. Moreover, intersections are key elements in urban networks and can affect travel time; properly modeling the behavioral differences of RVs and AVs around intersections demands multi-class DTAs with elaborate node models. However, using these models for NDP is extremely challenging due to their computation time and data requirements.

- *Elastic demand:* AV demand and its adoption over time as a response to the quality of service in the network can be modeled using elastic demand as opposed to scenario based demand. This can include several travel choices such as AV ownership, location, destination, mode, and time of the day choice.

- *Time dimension considerations:* deployment of AD subnetwork (or any other network configuration) is a gradual and long-term process. It also depends on AD development path in the future which is uncertain. This development over time subject to different uncertainties needs to be taken into account for infrastructure investment decisions. An appropriate AD network design method should include the time dimension and proper stochastic models to deal with mentioned uncertainty.

2.6 Recommendations

Based on this study, we recommend municipalities and metropolitan regions to start considering the notion of facilitating AD and making urban regions AD-friendly to guarantee safety for all road users and to aid in having a smooth transition period to full automation. This requires answering the question of “on which roads do we facilitate AD?” In order to answer this question, trade-offs between adjustment costs and gains in network performance have to be taken into account. As mentioned in the research agenda, more comprehensive transport models are necessary to consider these trade-offs and to thoroughly investigate AD impacts on urban regions.

3 Automated-Vehicle-Ready Subnetwork Optimization

The main objective of this chapter is to develop a bi-level network design method for optimizing the tradeoff between deployment cost of AV-ready subnetworks and their societal benefits. This addresses the second research question of this dissertation. For the lower level, the traffic assignment model developed in chapter 2 is used. The formulation of the upper level problem, the formulation of the bi-level problem, and a solution method to solve the bi-level problem are presented in this chapter. A new solution method is essential since the new problem includes specific constraints that make existing optimization algorithms less suitable for solving it, particularly in case of large-scale networks. Furthermore, applicability of the model developed in this chapter is demonstrated using a case study.

The term “automated driving (AD) subnetwork” used in this chapter refers to the concept of automated-vehicle-ready (AV-ready) subnetwork.

This chapter is based on the following journal article:

Madadi, B., van Nes, R., Snelder, M., & van Arem, B. (2020). A bi-level model to optimize road networks for a mixture of manual and automated driving: An evolutionary local search algorithm. *Computer-Aided Civil and Infrastructure Engineering*, *35*(1), 80–96. https://doi.org/10.1111/mice.12498

3.1 Introduction

Automated driving (AD) is a trend that will evolve over time, both in the market penetration rate of automated vehicles (AVs) and the level of automation. According to (SAE International, 2018), level 5 is the ultimate automation level with an unlimited operating design domain (ODD), yet level 3-4 AVs have a limited ODD. This implies roads that can accommodate these vehicles must meet certain standards. On the other hand, level-5 AVs are not expected on the market in the near future (Shladover, 2016), and even after their market introduction, for at least several decades, a heterogeneous mix of traffic with vehicles of different automation levels and different ODDs will be inevitable.

By definition, having a limited ODD means there are certain types of situations based on the infrastructure that level 3-4 AVs cannot safely handle (SAE International, 2018). However, the conditions for these situations (exact differences in ODDs) have neither been clearly defined by automobile manufacturers nor by researchers. Some studies have advocated the need for infrastructure adjustments for AD (Courbon et al., 2016; Nitsche et al., 2014; Zhang, 2013). An overview of more studies that suggest specific infrastructure requirements for AD is provided in (Farah et al., 2018).

On the other hand, analyses of automated driving system (ADS) disengagement reports occurring during AV tests in the U.S. show that road surface conditions (including poor markings) is one of the main causes of disengagements (Lv et al., 2018). According to (Dixit et al., 2016), 9.98% of disengagements have happened due to road infrastructure (which is the second cause after system failure), 5% due to other road users, and 56.1% due to system failures which can also be associated with vehicle's disability to deal with the environment. Moreover, number of disengagements differ significantly based on road type (lower in motorways and higher in urban streets). According to (Favaro et al., 2017), 89% of the reported accidents involving an AV happened at an intersection, with 32% occurring in city roads. Finally, it is stated in (Favarò et al., 2018) that more than 10% of disengagements are attributed to external factors with infrastructure and other road users being the main categories. Less than 13% of all disengagements happened in motorways, freeways and arterial roads combined, and the rest of the disengagements (around 87%) happened in interstate roads and urban streets.

This signals the need for attention to infrastructure and its influence on safe operation of AVs. Guiding the AD flows towards specific parts of the network that are inherently safer, and upgrading the infrastructure in these parts to meet the needs of AVs can assist in ensuring safety for all road users and further promoting AD as well as realizing network performance benefits in the transition period. This requires developing a vision for urban road networks and a strategic plan for its implementation. So far, dedicated lanes and links for AVs as well as dedicated AV zones have been suggested and briefly studied in the literature (Chen et al., 2017, 2016; Ye and Wang, 2018). We suggest an alternative network configuration for urban regions. We envisage that AD for level 3-4 AVs will be possible only on certain designated roads which is consistent with SAE's definition of level 3-4 automation. These roads must be selected based on safety and quality considerations. This means either they meet certain standards or they have the potential to meet these standards after reasonable adjustments. For the remaining roads, manual driving will be mandatory (although supported by various assisting driving automation systems such as collision avoidance systems). On these designated roads, AVs will operate on the same lanes as regular vehicles (RVs). In order to facilitate safe and efficient AD in mixed traffic on these selected roads, investments are required to ensure that they meet the desirable design standards (e.g., machine-readable and uniform lane markings and road signs, high surface quality, digital maps and V2I communication infrastructure); hence there will be trade-offs between these investments and the benefits they provide. This necessitates a network design approach to decide which roads should be selected to facilitate AD in the transition period, and to assess the impacts of this selection.

In this chapter, we formulate this problem as a bi-level network design problem (NDP), define requirements for its solution methods, propose a new heuristic algorithm that meets those requirements, and compare its performance with two variations of a common algorithm in the literature for solving the NDP. Moreover, based on the findings of the model, we discuss practical considerations for implementation of this network configuration.

The rest of this chapter is organized as follows: section 3.2 provides a brief background for the problem; section 3.3 introduces the concept of subnetworks for AD and the problem formulation, section 3.4 describes solution algorithms; two case studies and numerical results are presented in section 3.5; and section 3.6 includes the discussion and concluding remarks.

3.2 Background: Network Design and Automated Driving

Strategic investment decisions regarding road networks are often considered within the concept of NDP (Farahani et al., 2013; Yang and Bell, 1998) where the objective is to optimize certain performance indicators for a network (e.g., total travel cost (TTC)) considering the travelers' reactions to the network performance (e.g., route choices). This can be modeled as a bi-level leader-follower Stackelberg game in which the leaders (i.e., transport authorities) supply the transport infrastructure aiming at optimizing their objective function (e.g., TTC) and the followers (i.e., travelers) react with their travel choices to optimize their own objective function (e.g., individual travel costs). However, classical NDP formulations do not consider AVs and their impacts on transport system and travelers' choices. AVs are expected to have far-reaching impacts on various aspects of mobility (Fagnant and Kockelman, 2015; Milakis et al., 2017b; Shladover et al., 2012). Some of these impacts, namely, changes in road capacity, fuel efficiency, and value of travel time (VoTT) need to be considered in the NDP in order to accurately capture the travel impacts of AVs. In the following subsections, we discuss these changes and different approaches for modeling and incorporating them into the NDP.

3.2.1 Capacity Changes due to AD

AVs can improve transportation systems, especially when combined with vehicle connectivity. Cooperative Adaptive Cruise Control (CACC) is one of the major potential benefits of AVs (Shladover, 2017; Shladover et al., 2015). According to (Nowakowski et al., 2010), CACC with vehicle to vehicle (V2V) communication can increase lane capacity by reducing the driving time headway from 1.4 seconds to approximately 0.6 seconds. Various microsimulation studies have similar (although less severe) conclusions about the increase in lane capacity after the introduction of AVs; see, for instance (Mahmassani, 2016; Shladover et al., 2012; van Arem et al., 2006). This has been utilized in some NDP studies that consider AVs via assumptions of increased capacity in presence of AVs (Chen et al., 2017, 2016; Ye and Wang, 2018) and some other studies of AV impacts using macroscopic traffic assignment models (Levin and Boyles, 2015; Madadi et al., 2019).

All above-mentioned macroscopic studies have used static traffic assignment models. Another approach for incorporating capacity changes as well as several other changes in traffic flow characteristics (e.g., traffic flow stability and throughput) in presence of AVs with more accuracy is via dynamic traffic assignment (DTA) models and adjusted fundamental diagrams. In order to model the differences in backwards wave speed and lane capacity in shared roads for AVs and RVs, Levin and Boyles (2016) proposed a multi-class cell transmission model (CTM) that was utilized later in (Patel et al., 2016) to evaluate the impacts of improved capacity due to AD efficiency and reservation controls on arterial and motorway networks. Other researchers have developed and used DTA models to study possible changes in traffic flow dynamics caused by AVs as well. However, using DTAs for NDP is computationally very challenging.

3.2.2 VoTT Changes due to AD

For trips in automated mode, changes in VoTT are likely due to the possibility of performing other activities while driving. Although this is not fully established and there is no consensus in the literature yet, there are studies that claim AD will lead to a lower VoTT (Correia et al., 2019; de Looff et al., 2018; Le Vine et al., 2015; Milakis et al., 2017a). This can be considered in the route choice component of macroscopic traffic assignment models via differences in generalized travel cost; see, for instance (Levin and Boyles, 2015; Madadi et al., 2019).

3.2.3 Fuel Efficiency Changes due to AD

Higher fuel efficiency is another expected advantage of CACC (Rios-Torres and Malikopoulos, 2017; Shida and Nemoto, 2009; Shladover et al., 2015). This can affect travelers' route choice behavior since generalized travel cost is the defining factor for route choice. This has been utilized in (Madadi et al., 2019; Puylaert et al., 2018) to explore the travel impacts of AVs.

3.2.4 Network Configurations for AD

Despite the rapidly-growing body of knowledge about various aspects of automated driving, optimal network configurations for AVs using NDP frameworks are still rare in the literature. Chen et al. (2016) modeled the optimal deployment of dedicated lanes for AVs over time as an NDP where the lower level included a multi-user class (MUC) deterministic user equilibrium (DUE) route choice with RVs and AVs as separate classes, and the upper level included optimal decisions regarding location, timing and number of exclusive lanes for AVs. Another optimal network configuration for AVs using NDP was introduced in (Chen et al., 2017) where the authors proposed exclusive AV zones for urban road networks; for the AV zones with no RV access, a system optimal route choice was suggested, and for the other parts of the network, an MUC equilibrium was applied. Lastly, Ye & Wang (2018) attempted to optimize urban road networks for AVs with a combination of exclusive AV links and congestion pricing for RVs using a bi-level NDP.

So far the only NDP studies that consider AD concepts, have suggested exclusive lanes, links and zones for AVs. A more realistic network configuration for the transition period with a heterogeneous mix of vehicles was introduced in chapter 2 of this thesis where we suggested promoting AD in mixed traffic only in specific parts of the network where the infrastructure is upgraded to enable safe and efficient AD. This leads to a subnetwork within the urban road network which is suitable for AVs, yet it is also accessible for other vehicles. The concept was referred to as AD subnetwork. An MUC stochastic user equilibrium (SUE) route choice based on the path-size logit (PSL) model was developed to assess the impacts of this subnetwork; however, the choice of links to include in the AD subnetwork was a priori. A scenario-based approach was used to compare the performance of several network configurations in absence of an optimization concept or a bi-level model. In this chapter, we model the problem as a bi-level NDP and develop a quantitative solution method for optimal construction of these subnetworks.

3.3 Optimal Subnetworks for Automated Driving

In this section, we outline AD subnetwork configuration as well as the problem formulation, and we suggest solution methods for the problem.

3.3.1 The Concept of AD Subnetworks

The concept of AD subnetworks entails the vision that AD will be facilitated on a subset of existing road networks, called AD subnetwork, within which AD will be allowed and utilized everywhere. In this chapter, we further develop the concept by formulating it as a bi-level NDP and developing a solution algorithm for its optimal deployment.

First, we specify a set of feasible links in the network that are assumed to be safe for AD after some adjustments based on sustainable safety principles (Wegman et al., 2008; Wegman and Aarts, 2006). These links include all roads with flow (mobility) function and some roads with distribution function. They correspond to motorways, expressways and main urban roads which have been shown to be the safest roads for AVs and include none or very few intersections, which are where most AV accidents and disengagements have happened (Favarò et al., 2018; Favaro et al., 2017).

Next, we assume certain adjustment costs based on road types to meet design requirements for AD. These requirements include having limited access, high quality (e.g. pavement, lane marking, traffic signs, and lights), segregated traffic (homogeneity of mass and speed for vehicles in each lane), and grade separated or clear at-grade intersections. Note that most roads considered feasible here, already meet many of the mentioned standards. For instance, motorways require minimum or no adjustments to meet the standards. Therefore, we have assumed different adjustment costs based on road types and their initial quality with roads that are higher in the road network hierarchy (i.e., motorways) requiring least amount of adjustments thereby having the least adjustment costs.

Finally, since investing on adjusting all the links in large networks can be infeasible and even unnecessary, the next step is to make an optimal selection (among feasible links) that guarantees the highest possible gain from the investment.

3.3.2 Operational Concepts and Assumptions

It is assumed in this model that all vehicles are allowed everywhere in the network but AD is only possible within the AD subnetwork in mixed traffic conditions for level-3 and level-4 ADS-equipped vehicles (which will be referred to as AVs for the remainder of this chapter). Vehicles with automation functions of levels 0-2 are referred to as RVs in this chapter. Outside the AD subnetwork, all vehicles drive manually.

3.3.3 Optimal AD Subnetwork Construction as a Bi-level NDP

We model the optimal AD subnetwork construction problem as a bi-level NDP where the upper level entails the choice of links to include in AD subnetwork in order to minimize total (generalized) travel cost plus total (discounted) adjustment cost (investment), and the lower level represents the user equilibrium route choice.

The following subsections introduce the problem formulation, but first, we provide some relevant definitions and explanations which are used throughout the chapter. All the following definitions are demonstrated in Figure 3.1.

Let $G(N, A)$ denote a directed graph representing the road network where N is the set of nodes (vertices) and A is the set of directed links (arcs) representing the road segments. $G_1(N_1, A_1)$ denotes the (directed) graph representing the AD subnetwork which is a subgraph of G and N_1 and A_1 represent nodes and links that are selected for the AD subnetwork. Conversely, $G_0(N_0, A_0)$ is the set of nodes and links in $G(N, A)$ that do not belong to $G_1(N_1, A_1)$. This represents the rest of the road network. $\hat{G}_1(N_1, \hat{A}_1)$ denotes an undirected graph that is obtained by replacing the directed links in G_1 with identical undirected links.

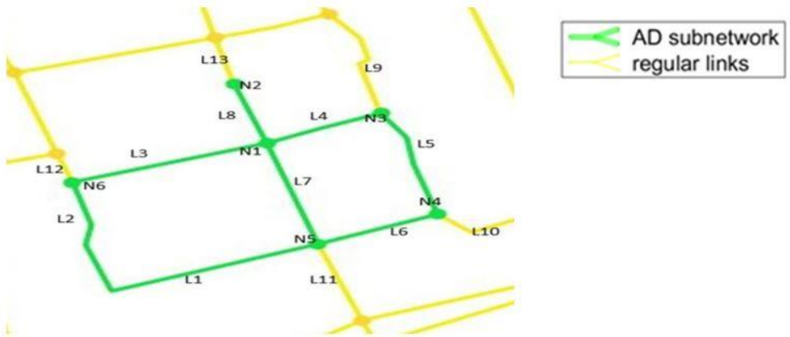


Figure 3.1 Definitions related to the AD subnetwork

Note: L1-8 and N1-6 are included in the AD subnetwork which is connected and the rest of links and nodes represent a part of the underlying road network. N2-6 are boundary nodes but not N1; degree of N1 is 4 which is equal to the degree of its underlying node while the degree of N5 is 3 which is less than the degree of its underlying node (i.e., 4). L9-13 are boundary links; they are incident to a boundary node but not included in the AD subnetwork.

Definition 3.1. Any node incident to at least one link included in an AD subnetwork, is included in that AD subnetwork.

Definition 3.2. The degree of a node is the number of links incident to that node within the graph in which that node is included.

Definition 3.3. A boundary node n_1^* for an AD subnetwork represented by a directed graph G_1 is a node included in G_1 with a degree less than the degree of the corresponding node n in the underlying graph G (i.e., the graph representing the original road network).

Definition 3.4. A boundary link a^* for an AD subnetwork is a link incident to a boundary node n_1^* of the AD subnetwork but not included in the graph G_1 representing the AD subnetwork. Note that boundary nodes of an AD subnetwork are included in the subnetwork while the corresponding boundary links are not included in that subnetwork (hence the difference in the subscripts).

Definition 3.5. An AD subnetwork represented by $G_1(N_1, A_1)$ is called connected if replacing all of its directed links A_1 with undirected links $\hat{A}_1$ produces a connected (undirected) graph $\hat{G}_1(N_1, \hat{A}_1)$. That is, for each two nodes in N_1, there is at least one path through links of $\hat{A}_1$ that connects these two nodes.

Lower level problem: multi-class equilibrium traffic assignment

For the lower level problem, we follow the formulation developed in chapter 2 where the lower level is modeled as an MUC SUE problem based on the PSL. Since during the transition period, there will be a mix of different vehicle types in the network, we believe system optimal routing is unlikely. Two user classes (AVs and RVs) with different generalized travel cost functions are distinguished. On links within the AD subnetwork, AVs are assumed to have a lower passenger car unit (PCU or equivalently PCE) to account for the capacity changes due to AD, a lower VoTT because travel time can be used for other activities, and a lower driving cost per kilometer (which we call value of travel distance (VoTD)) due to fuel efficiency of AD. The generalized travel cost function is the summation of travel time multiplied by VoTT, and travel distance multiplied by VoTD. Link travel time is based on a modified BPR function where the total flow is a weighted sum of class-specific flows to capture the correlation between link capacity and the proportion of AVs on the link. Weighting is based on (adjusted) PCU values. Values used along with other key parameters used in this chapter are reported in Table 3.3.

Note that since static traffic assignments do not model traffic flow dynamics with a high level of details, capacity changes as a result of AD have been incorporated in static traffic assignment models in two different (simplified) manners; the first one is via assigning higher capacity to links that accommodate AVs which is more appropriate when considering dedicated lanes and links for AVs (Chen et al., 2017, 2016; Ye and Wang, 2018). The alternative is to consider an AV-flow-specific link travel time that is dependent on both cumulative flow of RVs and AVs on the link, and the ratio of RVs to AVs on the link (Levin and Boyles, 2015; Madadi et al., 2019). This approach is more appropriate when considering a mix of traffic (RVs and AVs) on links. Therefore, we have opted for the second approach here. A mathematical formulation is presented below. The notation used throughout this chapter can be found in Table 3.1.

Table 3.1 Notation

Notation	Explanation
A	Set of all links a in the road network $A = A_0 \cup A_1$
A_0	Set of links a not in the AD subnetwork
A_1	Set of links a that belong to the AD subnetwork
W	Set of origin-destination (OD) pairs w
R^w	Set of routes between OD pair w
M	Set of user classes m (0: RVs, 1: AVs)
σ	Factor combining hourly to yearly travel cost conversion and discount rate over 10 years
l_a	Length of link a (kilometers)
ac_a	Adjustment cost of link a
η_m	VoTT for class m (0: RVs, 1: AVs)
γ_m	PCU value of class m
θ_m	VoTD for class m
$\delta_{m,a}^{w,r}$	Route-link incidence matrices (assignment maps)
D_m^w	Demand for OD pair w and class m
$PS_m^{w,r}$	Path-size penalty of route r between OD pair w and class m
β_m	Path-size correction parameter for class m
μ_m	Logit choice model parameter for class m
α_a, b_a	BPR function parameters for link a
Λ_a	Capacity of link a
$F_m^{w,r}$	Flow of route r between OD pair w for class m
Z_L	Objective function of the lower level mathematical problem (LLMP)
$\overline{f}_{m.a}$	Equilibrium flow of class m on link a (decision variable of LLMP)
$t_a(q_a)$	Travel time on link a based on total flow
q_a	Total flow of links a (PCU-equivalent)
Z_u	Objective function of the upper level mathematical problem (ULMP)
I_a	Binary decision variable (ULMP): 1 for AD subnetwork links (A_1), 0 otherwise (A_0)
$TTC(I_a)$	Total travel cost based on the value of I_a
$TAC(I_a)$	Total adjustment cost based on the value of I_a
$\lvert p^{s,t} \rvert$	Number of paths between source node s and sink node t traversing $\hat{G}_1$
n_κ	Number of connected components in graph
w_κ	Penalty weight for each disconnected part in a graph
n^p	Population size
n^g	Number of generations
m_i	Merging interval (ELS)
n_c^p	Number of candidates (ELS)
c_f	Crossover fraction (GA)
n_e^p	Elite size (GA)

LLMP:

Min $$Z_L = \sum_m \frac{1}{\mu_m} \sum_{w \in W} \sum_{r \in R^w} F_m^{w,r} \ln F_m^{w,r} - \sum_m \frac{1}{\beta_m} \sum_{w \in W} \sum_{r \in R^w} F_m^{w,r} \ln PS_m^{w,r}$$

$$+ \sum_{m \in M} \sum_{a \in A_0} \int_0^{q_a} \theta_0 l_a + \eta_0 t_a(x) dx$$

$$+ \sum_{a \in A_1} \int_0^{q_a} \theta_0 l_a + \eta_0 t_a(x) dx + \sum_{a \in A_1} \int_0^{q_a} \theta_1 l_a + \eta_1 t_a(x) dx, \qquad 3.1$$

s.t.

$$q_a = \gamma_0 (f_{0,a} + f_{1,a}), \qquad \forall a \in A_0, \qquad 3.2$$

$$q_a = \gamma_0 f_{0,a} + \gamma_1 f_{1,a}, \qquad \forall a \in A_1, \qquad 3.3$$

$$t_a(q_a) = t_a^0 [1 + \alpha_a (\frac{q_a}{\Lambda_a})^{b_a}], \qquad \forall a \in A, \qquad 3.4$$

$$\sum_{r \in R^w} F_m^{w,r} = D_m^w, \qquad \forall w \in W, \forall m \in M, \qquad 3.5$$

$$\sum_{w \in W} \sum_{r \in R^w} F_m^{w,r} \delta_{m,a}^{w,r} = f_{m,a}, \qquad \forall a \in A, \forall m \in M, \qquad 3.6$$

$$F_m^{w,r} \geq 0, \qquad \forall w \in W, \forall m \in M, \forall r \in R^w, \qquad 3.7$$

$$PS_m^{w,r} = \sum_{a \in r} (\frac{l_a}{l_r})(\frac{1}{\sum_{r \in R^w} \delta_{m,a}^{w,r}}), \qquad \forall w \in W, \forall m \in M, \forall r \in R^w. \qquad 3.8$$

Eq.3.1 presents the objective function of the lower level problem (LLMP). The first term is related to stochastic route flows, the second related to the PSL, and the next three terms present different generalized travel costs for each class on each subnetwork. *Eqs.3.2-4* show how total flows and travel times are calculated. *Eq.3.5* guarantees the demand is met for each class and each OD pair, *Eq.3.6* projects route flows on corresponding links, *Eq.3.7* prevents negative flows, and *Eq.3.8* shows how PS penalties are calculated.

Upper level problem: optimal design of AD subnetworks

The upper level objective function to be minimized includes TTC and discounted total adjustment cost (TAC) for upgrading links. The decision variables are binary integers representing links to be included in the AD subnetwork. The lower level equilibrium conditions are treated as constraints for the upper level optimization problem. A mathematical representation of the optimal AD subnetwork design is introduced below.

ULMP:

Min $$Z_u = TTC(I_a) + \frac{TAC(I_a)}{\sigma}, \quad 3.9$$

s.t.

$$TTC(I_a) = \sum_{a \in A} \{(1-I_a)[(\eta_0 \bar{t}_a + \theta_0 l_a)(\overline{f}_{0,a} + \overline{f}_{1,a})]$$

$$+ I_a[(\eta_0 \bar{t}_a + \theta_0 l_a)\overline{f}_{0,a} + (\eta_1 \bar{t}_a + \theta_1 l_a)\overline{f}_{1,a}]\}, \quad 3.10$$

$$TAC(I_a) = \sum_{a \in A} I_a ac_a, \quad 3.11$$

$$I_a(1-I_a) = 0, \qquad \forall a \in A, \quad 3.12$$

$$\left| p^{s,t}_{\hat{G}_1} \right| \geq 1, \qquad \forall s,t \in N_1, \quad 3.13$$

where $\bar{t}_a$, $\overline{f}_{0,a}$ and $\overline{f}_{1,a}$ are implicitly defined by solving the lower level mathematical problem (LLMP). *Eq.3.9* introduces the objective function of ULMP which is a summation of *TTC* and discounted *TAC*. Note that another representation of the objective function in *Eq.3.9* would be $\lambda TTC(I_a) + \frac{TAC(I_a)}{(1+\pi)^{\tau-1}}$, where π is the discount rate and λ converts the hourly travel cost to a yearly rate; however, for convenience of calculations, both terms were divided by λ, and $\lambda(1+\pi)^{\tau-1}$ was replaced by σ. *Eq.3.10* shows how *TTC* is calculated according to the link flows obtained in equilibrium, and *Eq.3.11* shows how *TAC* is calculated based on link adjustment costs and the value of decision variables. *Eq.3.12* represents the binary decision variables that take the value of one if the link is included in the AD subnetwork and zero otherwise. *Eq.3.13* ensures that the AD subnetwork represented by G_1 is connected (according to Definition 3.5). Solution algorithms are discussed in the following section.

3.4 Solution Algorithms

3.4.1 Requirements and Key Performance Indicators

To solve the lower level mathematical problem (LLMP), the MUC extension of the MSA algorithm introduced in (Wu et al., 2006) is used. Since the PCU-based weights used here for calculating total flow are similar to the approach used to account for trucks on links, LLMP has the same structure and mathematical properties as the multi-class mixed car and truck traffic network equilibrium problem. Therefore, it can be solved by MUC MSA algorithm mentioned earlier that is shown to be suitable for such problems. For more details, the reader is referred to (Madadi et al., 2019; Wu et al., 2006).

The upper level problem on the other hand, is an NP-hard problem that is non-convex and is known to be very challenging to solve, especially, when the lower level problem is an MUC SUE and the upper level decision variables are discrete. Several solution methods have been suggested in the literature. The reader is referred to (Magnanti and Wong, 1984; Yang and Bell,

1998) for reviews; however, within the last two decades, stochastic search and approximation heuristics have become the most preferred methods among transport researchers to solve NDPs (Migdalas, 1995; Poorzahedy and Rouhani, 2007), especially NDPs with discrete decision variables (DNDP). A more recent review study by (Farahani et al., 2013) shows that Genetic Algorithms (GA) introduced in (Holland, 1975) and elaborately reviewed in (Golberg, 1989) is one of the most commonly used methods for dealing with the complexity of DNDPs in recent years. Chakroborty (2003) concludes that GA-based procedures are highly effective for urban transit NDP which is a DNDP variation. This is perhaps due to their flexibility in modeling the underlying problem and their capability to produce nearly optimal solutions in reasonable times for large-scale problems. However, specific characteristics of AD subnetworks impose additional requirements on the solution algorithm that common GAs do not meet, such as graph connectivity requirements.

In this chapter, three major criteria for a suitable solution algorithm are considered, namely, effectiveness (degree of optimality), efficiency (computation time), and design quality (generating connected subnetworks). The first two criteria are relevant for the solutions of all DNDP variants. On the contrary, generating connected subnetworks has never been a requirement for DNDPs in the past. Therefore, the existing DNDP solution methods do not necessarily meet this criterion. However, in order to envisage an effective subnetwork within a road network that serves the mobility of AVs, it is crucial to guarantee at least a certain degree of connectivity within this subnetwork. Otherwise, switching frequently between automated and manual driving modes can compromise, most of all, safety, but also utility and efficiency gains of AD.

Since origins and destinations (centroids) are not located on feasible links for the AD subnetwork, guaranteeing a strongly connected subnetwork that provides accessibility from all origins to all destinations within the subnetwork is impossible, which is in fact due to the limited ODD of level 3-4 AVs. Yet, when a connected subnetwork for AD exists, AVs can start manually, reach the subnetwork and proceed in automated mode, and switch to the manual mode again when they approach their destination. In this manner, a safe and efficient navigation within the subnetwork in automated mode without disruptions is ensured for the middle part of the trip, which can be a large proportion of the trip. Then the design quality of an AD subnetwork refers to being connected, as defined in Definition 3.5, mathematically expressed in *Eq.3.13*, and illustrated in Figures 3.1 & 3.4. Moreover, we discuss the practical implications of this concept for optimal designs in section 3.5.3.

To introduce connectivity to the solution algorithm, different approaches can be considered. One approach is considering a penalty for switching from automated to manual driving mode in the utility function used in route choice; however, that can easily make the problem intractable since it requires a tremendous number of graph connectivity tests (i.e., one test per route for each fitness function evaluation). An alternative approach is to consider a penalty for disconnected designs. This is feasible since the number of connectivity checks are considerably less than the first option (one for each fitness function evaluation). Still, this may not be an efficient approach. Another alternative is to consider connectivity as a requirement and avoid disconnected designs all together. Yet there is no solution method in NDP literature that can do this. Therefore, in this chapter, we embarked on developing a new heuristic solution method (i.e., an evolutionary local search (ELS) algorithm) that produces connected designs efficiently. To demonstrate its performance, we also considered two alternative algorithms: a GA and a modified GA (MGA) that uses a penalty function to avoid disconnected designs based on the second approach mentioned earlier. In the following sections, we describe these three algorithms.

3.4.2 Genetic Algorithm (GA)

Genetic algorithms have become one of the favorite methods to solve DNDPs. They essentially make successive improvements over several generations of populations, each containing a set of strings (individual solutions) until finding a desirable solution. They use three general operations, namely, reproduction, crossover and mutation.

Reproduction is the process of selecting individuals that contribute to the next generation (parents). This process is random but biased based on individuals' fitness usually via a roulette wheel selection with slot sizes proportional to the individual's fitness value. Crossover operation includes constructing a mating pool based on roulette wheel, randomly matching individuals and swopping a portion of their characteristics. Mutation includes random perturbations on individuals to preserve diversity. Elitism is sometimes preferred in using GAs which includes selecting a certain number of individuals with best fitness (elites) and passing them to the next generation. This ensures that the good features in early generations survive to the later generations. A detailed explanation of GA operations is provided in (Golberg, 1989).

In this chapter, each individual design (chromosome) represented by I_a is coded as a binary string of bits, 0 and 1 (genes) with a length equal to the number of feasible links for the subnetwork where each link is represented by a 1 if it is included in the subnetwork (A_1) and a 0 otherwise (binary or bit string coding). Mutation is applied to 1% of genes in each chromosome and mutation genes are uniformly distributed over the range of genes. A multiple-point crossover function (sometimes referred to as uniform crossover) is used where each gene has an equal chance of coming from either parent. Fitness value in this chapter is the value of Z_u for each individual solution (*Eq.3.9*) where the equilibrium flows therein are obtained from solving LLMP.

The GA procedure used in this chapter is demonstrated in Figure 3.2. The number of offspring produced by each operation is determined by parameter values that define elite size (number of elites) and crossover fraction (reported in Table 3.3).

3.4.3 Modified Genetic Algorithm (MGA)

Since GAs generally are not tuned to produce connected designs, we use a modified GA (MGA) that penalizes disconnected designs in its fitness function evaluation in order to achieve connected optimal designs. The procedure of this MGA is similar to GA described in section 3.4.2 and depicted in Figure 3.2; the only difference here is that a penalty term $w_\kappa(n_\kappa-1)$ is added to the fitness function to aid the algorithm in finding connected designs. n_κ denotes the number of connected components and w_κ represents the penalty for each disconnected part. Note that in a connected graph the number of connected components is one. Therefore, the penalty term for connected graphs equals zero and with each disconnected part, the penalty increases. In this manner, the algorithm gets some feedback through improvements in fitness when moving from graphs with many separate connected components towards graphs with less connected components until (ideally) it reaches a graph with only one connected component. It is shown in section 3.5 that this approach is effective yet not necessarily efficient. Then instead of *Eq.3.9*, MGA uses *Eq.3.14* described below for its fitness function evaluations.

$$Z_u = TTC(I_a) + \frac{TAC(I_a)}{\sigma} + w_\kappa (n_\kappa - 1). \qquad 3.14$$

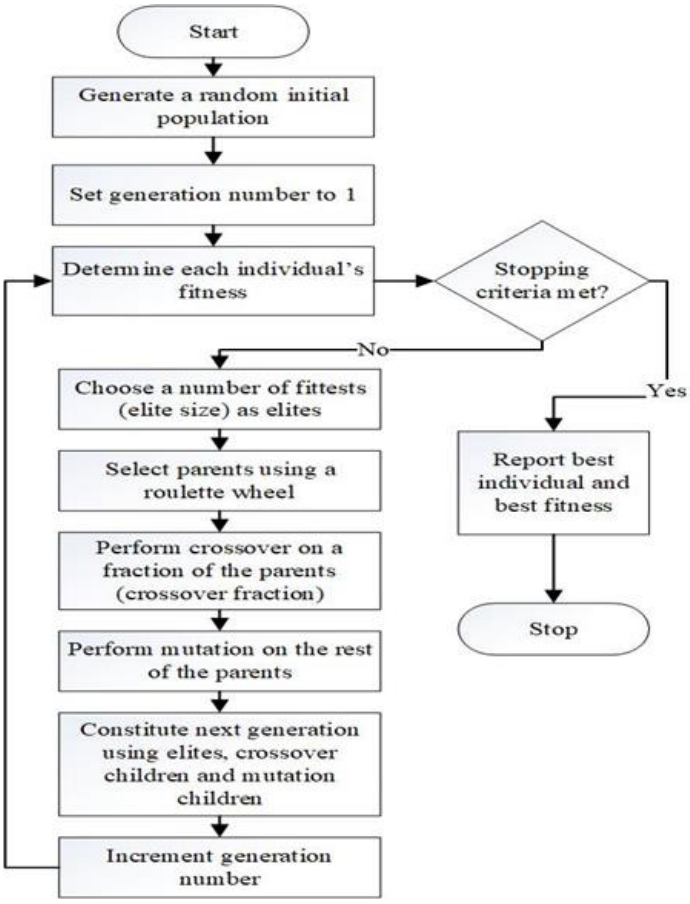


Figure 3.2 GA procedure for optimizing AD subnetwork

3.4.4 Evolutionary Local Search (ELS) Algorithm

It is acknowledged that testing connectivity of graphs is a computationally burdensome task (Even and Tarjan, 1975; Farahani and Miandoabchi, 2012). Besides, AD subnetworks are constructed from scratch. That gives us the opportunity to start with an initial (simple) connected graph and extend the graph only using operations that preserve connectivity; hence, a weakly connected graph is guaranteed without undertaking the burden of testing for connectivity at each iteration. On the other hand, hybrid metaheuristics can outperform single heuristics when tailored to a specific problem using the information available about the problem (Poorzahedy and Rouhani, 2007). Our ELS algorithm is inspired by local search algorithms such as hill climbing (step 2) that are generally efficient and evolutionary algorithms such as GAs (step 3) that have been proven effective for DNDP. It combines ideas from mentioned metaheuristics and takes advantage of the problem structure (building subnetworks from scratch) to generate connected AD subnetworks in an efficient manner.

ELS steps

ELS algorithm procedure consists of four major steps. Step 1 is initiation, step 2 is evolution to the next generation (i.e., adding links), step 3 is the merging process, and step 4 is the check for stopping criteria (i.e., termination). These steps are listed in detail below and depicted in Figure 3.3.

Step 1. Initiate n^p designs (subgraphs or individuals) for AD subnetwork in parallel, each one including one feasible link selected based on a roulette wheel where the odds are proportional to link capacities (this will favor links with higher capacities for initial designs).

Step 2. For each design:

- **Step 2.1.** find boundary nodes and boundary links
- **Step 2.2.** construct a set of n_c^p candidate new designs by adding a boundary link based on the roulette wheel selection (mentioned in step 1) to the existing design
- **Step 2.3.** for each candidate new design:
  - **Step 2.3.1.** measure fitness (evaluate the ULMP objective function (*Eq.3.9*) by solving LLMP)
- **Step 2.4.** replace the design with the relevant candidate new design that has the best fitness if it also exceeds the existing design in fitness.

Step 3. If the number of generations is a multiple of the merging interval m_i:

- **Step 3.1.** randomly match half the designs with the rest
- **Step 3.2.** eliminate the matches that do not have at least one node in common (to preserve connectivity)
- **Step 3.3.** construct new candidate designs by creating unions of matched designs (referred to as merged designs)
- **Step 3.4.** measure the fitness of each merged design
- **Step 3.5.** construct the new set of n^p designs by selecting the fittest n^p designs among the existing designs and the merged designs.

Step 4. Stop and return the fittest design if the stopping criterion is met (i.e., no improvement for 5 consecutive generations or no new candidate link), else go to step 2.

Note that while both ELS and GA are population-based algorithms, they have two fundamental differences.

GAs start with a random number of links selected for the subnetwork and iteratively add and remove (swap) links to reach better designs. While, ELS starts from designs with one link and grows the subnetwork only with the addition of links that improve the fitness till reaching the optimal design. GAs proceed with accepting disconnected designs as feasible solutions and reducing the probability of having disconnected designs via penalties (only MGA uses penalty). While, all ELS operations are designed to preserve connectivity. ELS starts with one link (which is a connected graph) and moves from one connected design to another via operating on boundary links and a conditional merging procedure that preserves connectivity.

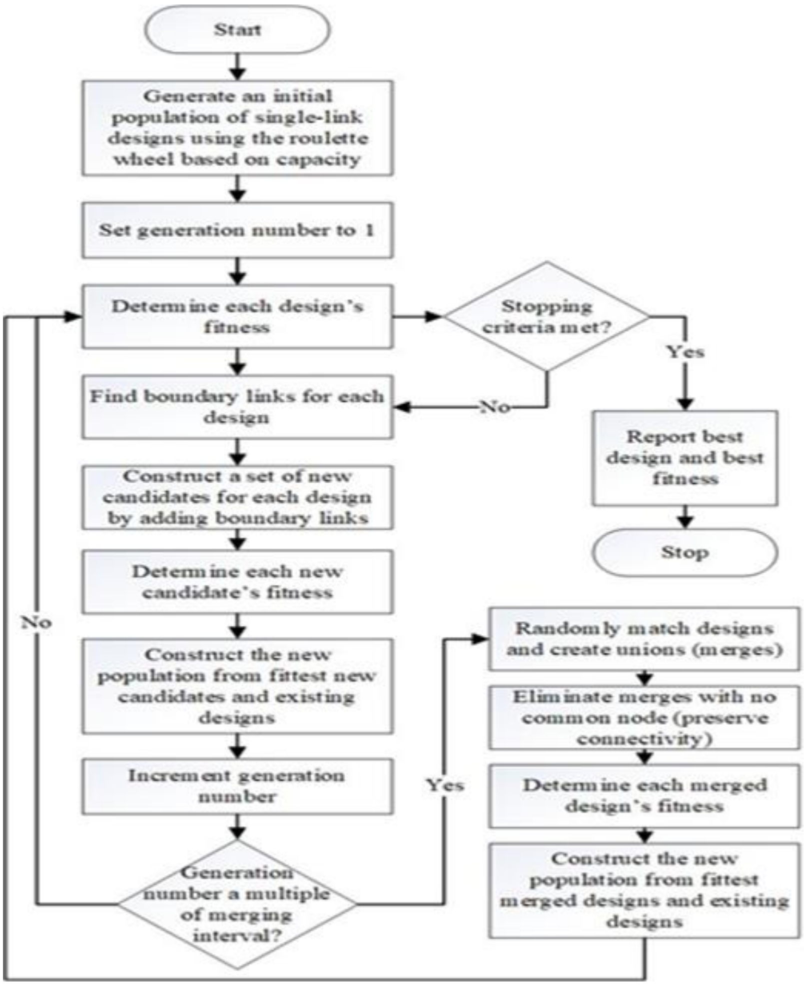


Figure 3.3 ELS procedure for optimizing AD subnetwork

3.5 Case Studies and Numerical Results

3.5.1 Case Study 1: Synthetic Network

In this section, a small synthetic network with 9 nodes and 24 links (12 pairs) is used as a proof of concept. We have considered local roads infeasible for AD subnetwork. Demand is assumed to be 280 v/h from each node to any other node. The adjustment cost is discounted over ten years ($\tau = 10$) (assumed effective life time for the infrastructure adjustments) with 4% discount rate ($\pi = 0.04$) which is the common value used for public investments in the Netherlands. Also, total travel cost for the peak hour is converted to a yearly basis. Peak hour travel cost is assumed to be 1/8 of 24 hour travel cost ($\lambda = 8 * 30 * 12$), therefore ($\sigma = \lambda(1+\pi)^{(\tau-1)} = 5945$). PS correction parameter value is set to one for this case. Logit route choice parameter values for RVs and AVs are respectively 1.25 and 2. The BPR function parameters (α & b) used are 0.15 and 4 respectively. A comprehensive description of the network, demand and optimal results obtained by each algorithm is provided in Table 3.2.

A full enumeration of all possible solutions is also performed to find the global optimum of this problem for the 50% penetration rate scenario. Note that this is only feasible for very small networks. Figure 3.4 exhibits the graph for the global optimum point of the problem and Table 3.2 reports network performance criteria related to this point.

All three algorithms easily find the global optimum point of the problem (Figure 3.4) on every run in less than a minute with an average common computer without serious parameter tuning. This confirms that these algorithms are capable of finding global optimum solutions of small size problems. Since this problem is trivial, parameter tuning and computational issues are not relevant here. These will be discussed in the next case study.

Table 3.2 Description of the demonstration case with a synthetic network

Network description					
Link pairs XY (node X to node Y)	**Capacity (v/h)**	**Length (km)**	**Speed (km/h)**	**Road type**	**Adjustment cost (€/km)**
12-21	500	3	40	Local road	infeasible
23-32	500	3	40	Local road	infeasible
36-63	4000	3	120	Motorway	300,000
69-96	4000	3	120	Motorway	300,000
98-89	500	3	40	Local road	infeasible
87-78	500	3	40	Local road	infeasible
74-47	4000	3	120	Motorway	300,000
41-14	4000	3	120	Motorway	300,000
45-54	2000	3	80	Expressway	1,200,000
56-65	2000	3	80	Expressway	1,200,000
25-52	2000	3	80	Expressway	1,200,000
58-85	2000	3	80	Expressway	1,200,000
Optimal results					
Objective function	**TTC**	**TTT**	**TTD**	**TAC**	**Connected**
48440	44807	2803	130559	21600000	Yes

Note: Demand is 280 v/h between each pair of nodes 1-9.

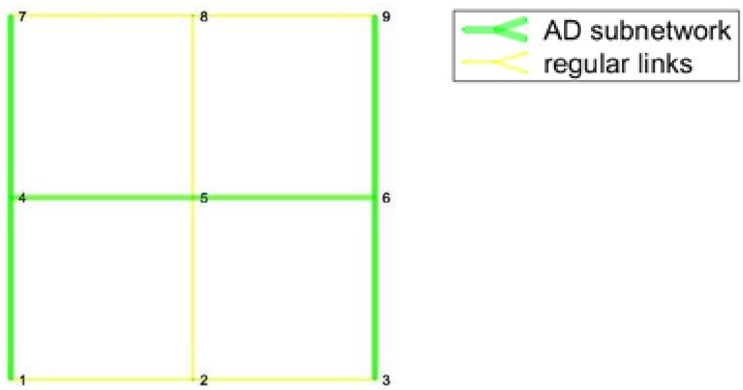
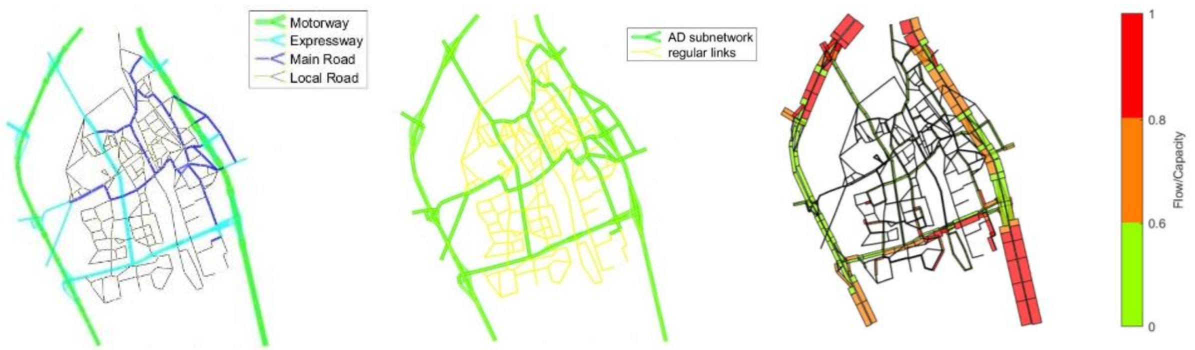


Figure 3.4 Optimal AD subnetwork for synthetic network and 50% penetration rate scenario

3.5.2 Case Study 2: Delft Network

In this section, we use another case study on a more complex network to compare the performance of different algorithms and to discuss relevant practical implications. A network similar to the road network of Delft, the Netherlands is used and the network data and the demand patterns are based on a peak hour in a tutorial project in OmniTRANS transport modeling software package. It includes 1,151 links, 494 nodes and 462 origin-destination pairs. The AV demand is determined assuming that the ratio of AVs to RVs available to travelers is the same as the AV market penetration rate and that the AVs available to travelers are homogeneously distributed in all zones. A subset of links corresponding to motorways, expressways and main roads are selected as feasible links for AD subnetwork. Analysis of the ADS disengagement reports mentioned in the introduction confirms that these are the most suitable roads for AVs. We consider different adjustment costs per road type which are proportional to the road hierarchy shown in Figure 3.5.a. The reason is that we want roads in AD subnetwork to meet certain design and quality standards. Motorways usually have highest standards and can be suitable for AD with minimal adjustments. Expressways and urban arterials can have high standards as well but they might include segments with unsegregated traffic and complex intersections that require adjustments given that 89% of the reported accidents involving an AV have happened at an intersection (Favaro et al., 2017). Since we are not aware of any study that provides accurate cost estimates for the adjustments mentioned in the introduction, we assumed the following costs and provide a sensitivity analysis to show how variations in the adjustment costs can affect the results: 100,000 €/per kilometer for local roads, 30,000 for expressways and 5,000 for motorways. The total number of feasible links is 421. We considered three demand scenarios with 10%, 50% and 90% penetration rate of AVs as well as a base case scenario with no AVs. Furthermore, a variation where all feasible links are included in the AD subnetwork is applied for each demand scenario to provide lower bounds for comparisons. The value of σ is the same as mentioned in the previous subsection.

Different road types, feasible roads for AD subnetwork and congestion patterns in the base case of Delft network are shown in Figure 3.5. Table 3.3 summarizes other important parameters used in this case study. ULMP parameters are determined after extensive trials to optimize each algorithm's performance. LLMP parameters are the same as those used in chapter 2 or within the sensitivity analysis bandwidth used therein.



(a) Road types (b) Feasible roads (c) Congestion pattern

Figure 3.5 Road types, feasible roads for AD subnetwork and congestion patterns in Delft case study (base case)

Table 3.3 Important input parameters used in Delft case study

Upper level problem (ULMP) parameters		
Algorithm	**Parameter**	**Value**
ELS	Population size (n^p)	10
	Merging interval (i_m)	20
	Number of candidates (n_c^p)	4
GA	Population size (n^p)	100
	Elite size (n_e^p)	20
	Maximum generations (n^g)	150
	Crossover fraction (c_f)	0.8
MGA	Population size (n^p)	300
	Elite size (n_e^p)	30
	Maximum generations (n^g)	200
	Crossover fraction (c_f)	0.8
	Penalty weight (w_κ)	2000

Lower level problem (LLMP) parameters			
Parameter	**Class**	**Penetration rate**	
		10%	**[50%-90%]**
PCU	RV	1	1
	AV	0.95	0.9
VoTT (€/h)	RV	9	9
	AV	7.2	7.2
VoTD (€/km)	RV	0.19	0.19
	AV	0.114	0.114

3.5.3 Numerical Results and Analysis

Since there is stochasticity in the algorithms, we used five independent runs for each scenario and reported the average results. For the objective function, a 95% confidence interval is reported to show the stability of the algorithms. The optimal graphs and convergence curves reported are from the runs with best results. In the following subsections, we compare the results based on each of three mentioned criteria.

Effectiveness

Since with heuristic algorithms, there is no guarantee for finding the global optimum, we consider the degree of optimality (i.e., how low the value of the objective function is) as the main indicator for effectiveness of an algorithm. As demonstrated in Table 3.4, in terms of degree of optimality, ELS narrowly outperforms GA and MGA in all scenarios in this problem. Although, both GA and MGA, especially MGA are competitive in terms of objective function value in all scenarios. Furthermore, the confidence intervals reported overlap in most cases but ELS shows more stability within different runs. An interesting observation regarding TTC is that MGA outperforms the other two in terms of TTC in all scenarios, yet this is achieved at price of significantly higher TAC that leads to less desirable objective function values. Regarding TTT, the differences in most cases are not significant. This can be explained partially by the fact that a portion of gains in TTC are due to the assumption of lower VoTD and VoTT for AVs. Moreover, congestion reduction gains of AD are, to some extent, counterbalanced by longer detours for routes using the AD subnetwork. Trends in TTD corroborate this explanation since less effective results in terms of TTC have lower TTD. That is, more effective designs cause more rerouting towards the AD subnetwork which lead to higher travel distances.

Another discernible pattern is related to the stretches of roads included in the best results. It can be seen in Figures 3.6-8 that the subnetworks obtained by all three algorithms, include almost identical main stretches of roads in all scenarios. This suggests plausibility for the designs acquired by these algorithms. Given that the TTT and TTC values are, in most cases, very similar to the ones in “all feasible links selected” variation that has been used as a lower bound for TTC and TTT.

Table 3.4 Comparison of algorithm performances in effectiveness (optimality)

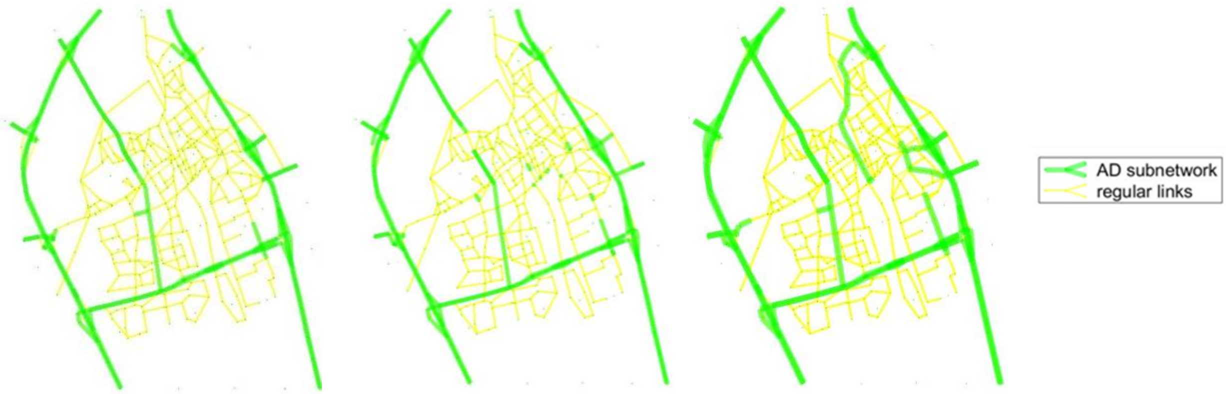
		10% penetration rate			
Algorithm	**Objective function**	**TTC (€)**	**TTT (h)**	**TTD (km)**	**TAC (€)**
ELS	100,982 ± 0.63	100,749	5,508	279,008	1,380,488
GA	100,998 ± 8.87	100,753	5,508	279,001	1,461,647
MGA	101,052 ± 6.16	100,708	5,508	279,006	2,047,384
All feasible links	101,449	100,561	5,508	279,006	5,281,545
"As is"	102,519	102,519	5,509	278,628	0
		50% penetration rate			
Algorithm	**Objective function**	**TTC (€)**	**TTT (h)**	**TTD (km)**	**TAC (€)**
ELS	93,371 ± 0.84	92,798	5,492	280,528	3,407,849
GA	93,437 ± 8.11	92,846	5,492	280,524	3,509,915
MGA	93,382 ± 5.01	92,790	5,492	280,527	3,519,558
All feasible links	93,546	92,658	5,492	280,525	5,281,545
"As is"	102,519	102,519	5,509	278,628	0
		90% penetration rate			
Algorithm	**Objective function**	**TTC (€)**	**TTT (h)**	**TTD (km)**	**TAC (€)**
ELS	85,673 ± 0.93	84,969	5,491	282,039	4,182,494
GA	85,783 ± 17.9	85,107	5,491	282,039	4,019,448
MGA	85,678 ± 4.11	84,962	5,491	282,039	4,257,011
All feasible links	85,789	84,901	5,491	282,038	5,281,545
"As is"	102,519	102,519	5,509	278,628	0

Note: Reported values are averages of 5 independent runs. ± signs denote standard errors for 95% confidence intervals.

Design quality

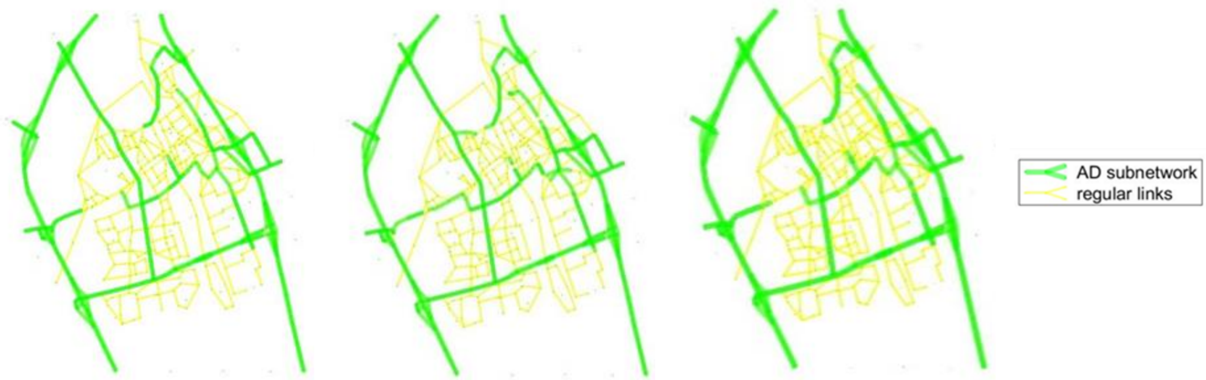
Since an AD subnetwork is primarily designed to serve the flow of AVs in automated mode with minimum possible disruptions, connectivity is a crucial factor for its quality and effectiveness. Yet a strong connectivity is unnecessary given that AD subnetworks are designed to facilitate mobility rather than accessibility. There is no need to reach every single node from every other node in both directions within the AD subnetwork. A weak connectivity (as defined in Definition 3.5) is sufficient to ensure that if a part of a trip is possible within the AD subnetwork, there is no interruption (i.e., switching between driving modes) in that part of the trip. Therefore, we considered connectivity as the criterion for comparing the design quality of the results.

It is evident from Figures 3.6-8 that all the optimal designs produced with ELS and MGA meet the connectivity criterion as defined in this chapter but most of the designs produced by GA do not meet this criterion. This was to be expected since GA does not include any mechanism to favor connected designs. Although, apart from a limited number of isolated links and stretches, GA has produced very similar designs compared to those produced by ELS. Also, all designs generated by MGA meet the connectivity criterion suggesting that the penalty function (*Eq.3.14*) has been effective. However, this has resulted in substantial additional computation times for MGA. Efficiency of the algorithms will be discussed in the next subsection.



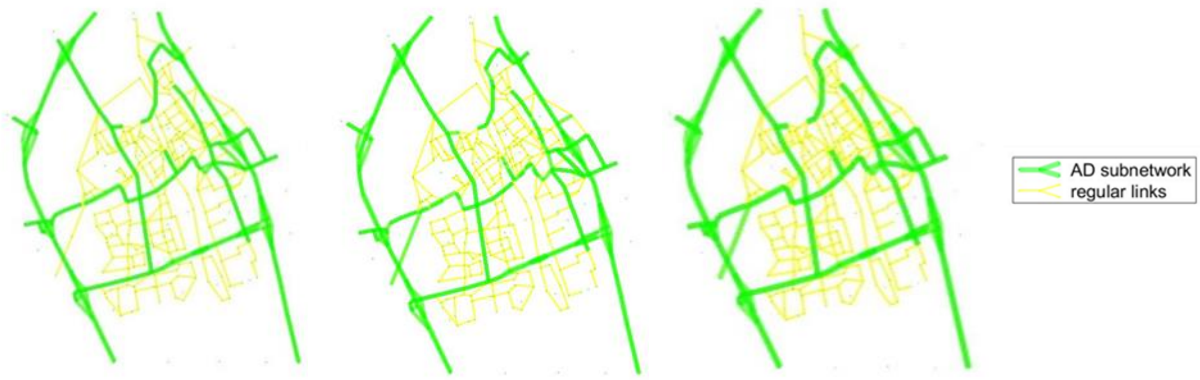
(a) ELS (b) GA (c) MGA

Figure 3.6 Optimal AD subnetworks (best results) for 10% penetration scenario



(a) ELS (b) GA (c) MGA

Figure 3.7 Optimal AD subnetworks (best results) for 50% penetration scenario



(a) ELS (b) GA (c) MGA

Figure 3.8 Optimal AD subnetworks(best results) for 90% penetration scenario

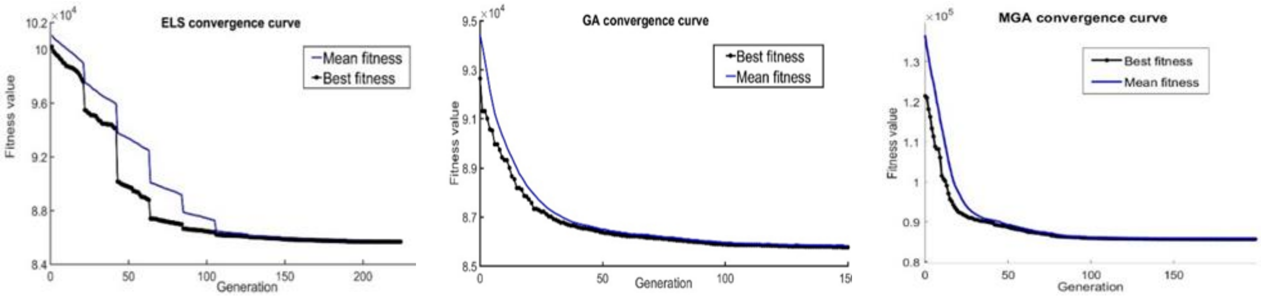
Efficiency and convergence

Table 3.5 summarizes the computation times for all algorithms in all scenarios. ELS has the lowest computation time in 10% penetration rate scenario but this value increases with the increase in penetration rate. The reason is that ELS extends one link at a time, so when the optimal AD subnetwork includes more links (with more AVs) computation times increase, yet this increase seems to be linear. GA shows stable computation times for different scenarios but they are higher compared to ELS in all scenarios. Finally, MGA shows stable computation times across scenarios as well. Although, its numbers are significantly higher compared to all others (more than three times the computation times of GA). This confirms that the penalty function, although effective in securing connected designs, is rather inefficient. This can become a major issue when applying MGA on large scale networks, especially since computation times tend to increase exponentially with the number of links.

Table 3.5 Average computation times for different market penetration rates in Delft case study

Algorithm	Average computation time (minutes)		
	10%	50%	90%
ELS	76	140	157
GA	271	274	270
MGA	964	968	966

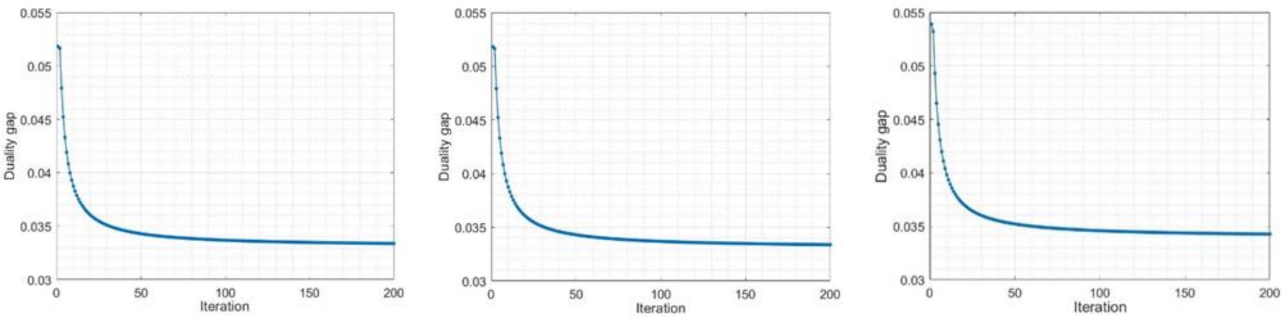
Regarding the ULMP convergence curves presented in Figure 3.9, GA and MGA display a common evolutionary behavior with the average population following the elite with a lag and algorithms settling down towards the end. The same pattern is noticeable for ELS with the addition of sudden significant improvements in fitness values that correspond to merging intervals (refer to Table 3.3). This implies that the merging operation (step 3 of ELS steps) substantially accelerates the convergence of ELS. Regarding MGA, two stages of improvements are noticeable from the graph. The first stage which include fast and substantial improvements in the beginning, shows how MGA is attempting to find designs with less connected components to reduce the penalty. The second stage denotes the search for more effective designs.



(a) ELS (b) GA (c) MGA

Figure 3.9 Convergence curves (ULMP) in 90% penetration scenario

Since ULMP is evaluated when LLMP is at its equilibrium, the convergence curves (duality gaps) of LLMP for optimal results of the 50% penetration rate scenario are shown in Figure 3.10 as an example. A reasonable duality gap is reached (even within the first few iterations) considering that LLMP is a SUE and that the duality gap for a SUE should not approach zero. For the intermediate iterations of ULMP, the solution of LLMP terminates when a threshold of 4% duality gap is reached which is very close to the lowest duality gap achieved after 200 iterations (3.3%). Also, given that all algorithms in all their fitness function evaluations use the same threshold, the comparison is fair.



(a) ELS (b) GA (c) MGA

Figure 3.10 Convergence curves of LLMP (duality gaps) for optimal results in 50% penetration scenario

Sensitivity analysis

An average peak hour traffic demand is used for the case studies here (as is the case in most NDPs). However, day to day variations in demand are common in transportation networks. Therefore a sensitivity analysis on demand is reported in Table 3.6 for 50% penetration rate scenario to show how variations in demand can cause variations in objectives. As it can be seen in Table 3.6, increases in demand and congestion levels have more impact on TTC, TTT and objective function compared to decreases. TTDs change proportional to demand and TACs show more sensitivity to reductions in demand after a certain point. One explanation is that with less demand, less investment is justified, yet significant increases in demand cause too much congestion that cannot be mitigated with extra investments. Therefore, proper demand estimation is crucial for practical applications.

Moreover, sensitivity analysis for variations in adjustment costs reveals that these changes have minor impacts on network performance indicators. Although, these variations affect the optimal configuration of the AD subnetwork, as evidenced by the changes in TACs whose sensitivities change with different proportions with respect to increases and decreases in the adjustment costs.

Table 3.6 Sensitivity analysis (variations in results) with respect to demand and adjustment cost variations in 50% penetration rate scenario (obtained by ELS)

Demand variation	Objective function	TTC	TTT	TTD	TAC
+20%	+31%	+31%	+41%	+20%	+7%
+10%	+14%	+14%	+18%	+10%	+5%
-10%	-13%	-13%	-15%	-10%	-3%
-20%	-24%	-24%	-28%	-20%	-20%
Adjustment cost variation					
+300%	+1.21%	+0.68%	+0.02%	0.00%	+87%
+100%	+0.55%	+0.33%	+0.02%	0.00%	+36%
-25%	-0.16%	-0.07%	0.00%	0.00%	-16%
-50%	-0.35%	-0.12%	0.00%	0.00%	-37%

3.6 Discussion and Conclusions

In this chapter, we considered the problem of optimizing urban road networks for a mixture of AVs and RVs. We modeled this problem as a bi-level DNDP and attempted to optimize the trade-offs between the costs and the benefits of a certain network configuration, namely the AD subnetwork. Moreover, we suggested a solution algorithm for this problem and benchmarked its performance against two solution algorithms for DNDP using a case study and considering three different performance criteria. The results reveal that the ELS algorithm presented in this chapter has a satisfactory performance in all three criteria considered, namely, effectiveness, efficiency and design quality.

Design quality criterion is specifically relevant to the formulation of NDP introduced in this chapter which enforces the connectivity constraint on the subnetworks. It is especially this constraint that makes the commonly used NDP solutions less suitable, while the MGA (i.e. the GA including a penalty for disconnected subnetworks) proves to be inefficient. The ELS algorithm developed in this chapter is better tailored to this new formulation of the NDP and proves to be efficient in dealing with connectivity constraints.

Moreover, our findings indicate that the optimal layout of AD subnetworks depends on demand and adjustment cost; lower penetration rates of AVs call for less dense subnetworks and less investment. Nonetheless, a large proportion of all possible network performance benefits are achievable with low-density AD subnetworks. This is, to some extent, due to the fact that more effective designs usually include roads with highest capacity first. Therefore, it is logical to redirect AD to designated parts of the network that are selected based on their inherent safety and optimized for network performance. On the other hand, higher penetration rates of AVs demand denser AD subnetworks and can deliver more benefits. A key observation was that an effective subnetwork can deliver a large proportion of the benefits obtained by upgrading all feasible links (which includes substantial costs), especially with higher penetration rates of

AVs. Constructing AD subnetworks has apparent safety and performance benefits since it includes redirecting AV traffic to specific parts of the network that are either naturally safer or have been adjusted to be safer, and are optimized for network performance. Therefore, it is recommendable for urban planners and transport authorities to consider such configurations for the transition period.

Demand for AVs will evolve gradually over time. So AD subnetworks should be extended gradually as well. It is practical to start investing on high-impact places first and gradually extending the AV-friendly part of the network with the increase in the market penetration rate of AVs to maintain the optimal trade-off between the investments and the benefits over time.

Therefore, a possible future research direction is considering the time dimension and modeling the problem as a tri-level NDP with timing as the third level and including AV demand elasticity. This can provide an optimal plan for gradually upgrading the AD subnetwork over a time horizon in reaction to AV demand.

Another possible extension of this problem is considering DTA models for the lower level equilibrium problem to model the behavioral differences of RVs and AVs more accurately. This can also allow for elaborate modeling of intersections, which are important elements in urban road networks, in presence of RVs and AVs. However, this can be extremely challenging due to the high computation requirements of DTA models and the fact that the literature on DTA models for AVs is still scarce.

4 Multi-Stage Automated-Vehicle-Ready Subnetwork Evolution

The aim of this chapter is to incorporate the time dimension into the bi-level model developed in the previous chapter in order to develop a multi-stage bi-level network design model to determine the optimal timing for the evolution of AV-ready subnetworks over time. This answers the third research question raised in chapter 1. The traffic assignment model developed in chapter 2 is extended to a multi-mode multi-class traffic assignment model in order to consider the elasticity of demand due to changes in the level of service in the network in time. A diffusion model is used to estimate the adoption rate of AVs over time as a result of changes in the level of service in transport networks. Three heuristic solution methods are proposed in this chapter to solve the problem. The applicability of the concepts developed in this chapter is demonstrated via a case study of a realistic road network of the Amsterdam metropolitan region.

This chapter is currently under review for journal publication.

4.1 Introduction

Automated vehicles (AVs) are on the horizon; however, it might take a long time before a large market share for highly automated vehicles can be observed. In the meantime, a heterogeneous mix of traffic with AVs and regular vehicles (RVs) on the roads is inevitable. According to SAE International (2018), there are five levels of vehicle automation where level 5 denotes fully automated vehicles, levels 3-4 denote conditionally and highly automated vehicles, and levels 1-2 denote partially automated vehicles, which are already available on the market. Although development of flawless level-5 AVs can take a long time, levels 3-4 might become a reality within the coming decade (Shladover, 2016). Based on SAE International (2018), the operating design domain (ODD) of levels 3-4 is limited. However, there is currently a lack of data to specify the exact ODD limitations of AVs. On the other hand, while ODDs are the accepted language of the automotive industry to define functional requirements for vehicle automation, there is no universally accepted standard for road operators describing the readiness of road network infrastructure to support automation functions. Furthermore, the interactions between AVs and infrastructure become crucial during the long transition period to full automation, since a mix of RVs and AVs is then expected on the roads. A proper infrastructure can support AVs' functionality, extend their ODD and improve safety for all road users, while lack of proper infrastructure can have negative impacts on these factors.

Some studies have specified infrastructure requirements for safe operation of AVs (see (Farah et al., 2018) for a review). However, establishing a correspondence between the requirements and different parts of road networks can be challenging. Moreover, these requirements can be idealistic, expensive and difficult to meet, especially in some road types, such as local distributors. Other researchers have suggested optimal networks of dedicated lanes (Chen et al., 2016), dedicated links (Ye and Wang, 2018), and dedicated zones for AVs (Chen et al., 2017) as well as AV-ready subnetworks for mixed traffic (chapter 2) to address this issue and promote the adoption of AVs via network design concepts. These are promising approaches; nonetheless, models representing these designs require more details and extensions to become operational. Furthermore, so far they have only been tested on small, often theoretical networks. Application of these concepts on large-scale real networks is a crucial but missing next step, especially since there are many practical issues and considerations involved with real networks that are not observed with theoretical networks, including network hierarchy and road type. Besides, dedicating parts of a network to one class of vehicles, can compromise accessibility of other classes and modes.

Moreover, designing optimal road networks for AVs is a gradual process that depends on the level of demand for AVs. Since the demand for AVs will increase over time, optimal networks for AVs should evolve over time as well. An efficient design for a network with a low level of AV demand is not necessarily an efficient design for the same network with a very high level of AV demand. In addition, demand for cars and the performance of road networks cannot be assessed without considering alternative modes.

On the other hand, analyses of automated driving system (ADS) disengagement and accident reports in the USA provide some insight into the suitability of various parts of road networks for AVs. Approximately, 10% of the ADS disengagements can directly be associated with road infrastructure (Dixit et al., 2016), and around 56% have been due to system failures, which can also be attributed to vehicle's inability to cope with its surrounding. Furthermore, around 87% of all disengagements have occurred at interstate roads and urban streets, while motorways, freeways and arterial roads combined are associated with less than 13% of all disengagements

(Favarò et al., 2018). According to Xu et al., (2019), 71.2% of the accidents involving AVs in California so far have occurred at intersections, 20.5% in urban streets and only 1.4% in highways. Moreover, based on the evidence of existing level-2 AVs, drivers are very likely to use the AD in freeways and unlikely to use it on rural and urban roads (Hardman et al., 2019). This shows that road type is an essential factor for safety of AVs.

In light of the discussion above, in this chapter, we select a set of roads based on their characteristics to define a (potentially) safe feasible region for the operation of AVs in AD in mixed traffic (on the same lanes as RVs). This region will be upgraded with physical and digital infrastructure investments to guarantee safety for all road users and to improve the efficiency of AD. However, investing in all feasible roads can be costly and unnecessary. Therefore, a next step is required to make a selection among the feasible roads that optimizes the trade-off between the investments and the societal benefits they provide in order to construct an efficient AV-ready subnetwork for mixed traffic. Therefore, we propose a multi-stage mathematical model to optimize the evolution of AV-ready subnetworks over time in response to the gradual development of AV demand considering the competing transport modes. Moreover, we use a case study of a large-scale real road network to demonstrate our proposed design concept and to discuss practical considerations related to its deployment. This chapter contributes to the existing literature and extends the authors' previous works on AV-ready subnetworks by the following.

1. Proposing the multi-mode multi-class formulation of the network equilibrium problem with a mix of AVs and RVs including asymmetric link costs (lower level problem)
2. Proposing a multi-stage (time-dependent) model for optimizing AV-ready subnetworks for mixed traffic over time with endogenous and time-varying demand for AVs
3. Proposing two heuristic algorithms to solve the problem, comparing their performance to a Genetic-algorithm-based solution procedure, and providing a rigorous analysis of their performance using extensive computational experiments
4. Demonstrating the applicability of the proposed methodology on a large-scale case study of the real network of the Amsterdam metropolitan region

The rest of this chapter is organized as follows: section 4.2 includes a brief problem background, section 4.3 presents the problem formulation and the solution methods, section 4.4 demonstrates the case study and numerical results, and section 4.5 entails the discussion and concluding remarks.

4.2 Problem Background

The problem under consideration is designing optimal AV-ready subnetworks in road networks and their evolution over a planning horizon considering a time-varying demand from different modes and vehicle types. In previous studies, these subnetworks have been referred to as automated driving (AD) subnetworks as well. Since models used to address such problems include many components, we provide an overview of the relevant studies in this section to place the problem within the literature, and discuss additional studies pertinent to each model component in the corresponding section.

In transport literature, strategic decisions regarding transport networks are considered within the framework of the well-known network design problem (NDP), which involves a large body of literature. For review studies, the reader is referred to (Farahani et al., 2013; Magnanti and Wong, 1984; Yang and Bell, 1998). NDPs are generally modeled as Stackelberg leader-follower games where the leader tries to make optimal decisions for transport infrastructure considering the followers' (travelers) response to the changes in the network by their travel choices. Mathematically, this results in a bi-level non-convex NP-Hard problem, which is very challenging to solve (Yang and Bell, 1998). The upper level includes optimal decisions for transport networks (e.g., building new streets and adding new lanes) considering their cost (investment) as well as their benefits (e.g., total user travel time saving), and the lower level includes predicting the travelers' response to these decisions via their travel choices, which is commonly perceived to follow Wardrop's equilibrium principles (Wardrop, 1952).

Accordingly, NDPs can be classified into several variants. Based on the upper level decision variables, discrete NDP (DNDP) (Chen and Alfa, 1991; Leblanc, 1975), continuous NDP (CNDP) (Davis, 1994) and mixed NDP (MNDP) (Cantarella et al., 2006) are recognized. Based on the upper level objective function, single objective (Cantarella et al., 2006) and multi-objective (Miandoabchi et al., 2012; Wang and Szeto, 2017) variants have been studied. Based on the lower lever equilibrium type, deterministic user equilibrium (DUE) (Leblanc, 1975), stochastic user equilibrium (SUE) (Davis, 1994), system optimal (Dantzig et al., 1979) and mixed (Chen et al., 2017) versions have been proposed. Regarding the lower level demand, fixed demand (Tobin and Friesz, 1988) and elastic demand (Yang, 1997), single class (Chen and Alfa, 1991) and multi-class (Chen et al., 2017), as well as unimodal (Chen et al., 2016) and multi-modal (Miandoabchi et al., 2012) versions have been considered. Another distinction is based on the number of decision stages (time periods) which leads to single stage (most NDPs) and multi-stage or time-dependent NDP (NDP-T) (Lo and Szeto, 2009; Ukkusuri and Patil, 2009). Finally, in recent years, a number of NDPs have considered special network configurations such as dedicated lanes (Chen et al., 2016), dedicated links (Ye and Wang, 2018), and dedicated zones for AVs (Chen et al., 2017) as well as AV-ready subnetworks for mixed traffic (chapter 3). This last category of NDPs is of special interest in this chapter, since they are closely related to this study due to proposing design concepts for AD. We refer to them as AD-NDP and briefly describe them in the following paragraph.

Chen et al. (2016) studied the problem of optimal deployment of dedicated AV lanes over time. The upper level included deciding where and when to deploy dedicated AV lanes to minimize the social cost and the lower level involved a multi-class DUE traffic assignment with fixed demand and a single mode (i.e., car). The model was tested on the (simplified version of) South Florida network with 232 links. To the best of our knowledge, this is the only time-dependent AD-NDP (AD-NDP-T) study so far. Chen et al. (2017) proposed optimizing dedicated AV zones in road networks with the upper level objective of minimizing social cost and a mixed route choice model for the lower level including system optimal routing for AVs and a deterministic routing for RVs. The model was demonstrated on a number of small synthetic networks (maximum 288 links). Ye and Wang (2018) studied the problem of optimizing dedicated AV links combined with congestion pricing for RV links where the upper level entailed minimizing total travel time cost as well as the link-based toll, and the lower level was a DUE including RVs and AVs. This model was tested on a small synthetic network with 18 links. It is worth noting that all three aforementioned configurations restrict RVs' access to some parts of the network since they dedicate some lanes, links and zones only to AVs. Finally, in chapter 3, we proposed optimal AV-ready subnetworks referred to as "AD subnetworks", which were accessible for all vehicles, yet adjusted for optimal ADS performance in mixed traffic (i.e., RVs and AVs on the same lanes). The upper level objective was to minimize total travel cost along with infrastructure adjustment cost of creating the subnetwork and the decision variables represented the links to be selected for the subnetwork. The lower level was a multi-class SUE with the path-size logit. This model was demonstrated on a semi-real network of Delft with 1,151 links.

In this chapter, we model the problem as a multi-stage (time-dependent) discrete NDP (AD-DNDP-T) with multi-mode and multi-class demand involving AVs. To the best of our knowledge, this is the first NDP-T with multi-mode multi-class demand with asymmetric link costs (for RVs and AVs) involving special network configurations for AVs with demonstrations on a large-scale real road network.

4.3 Multistage Design of Subnetworks for Automated Driving

The concept of AV-ready subnetworks or AD subnetworks for mixed traffic entails selecting a subset of (potentially safe) roads within road networks to form a subnetwork, and upgrading this subnetwork with necessary (physical and digital) infrastructure adjustments to meet higher quality standards to ensure uninterrupted, safe and efficient AD using ADS in mixed traffic conditions. The extent of these adjustments depends on available funds and authorities' commitment to facilitating safe AD. Several lists of possible adjustments under various scenarios based on expert interviews are provided by Lu et al. (2019). In this chapter, the selection of roads for adjustment occurs in two steps. The first step includes a preselection of feasible road segments (links) based on road type to guarantee that all selected links for the subnetwork have the potential to meet the desired standards after reasonable adjustments. It entails excluding roads with complex interactions between AVs and other road users. A more elaborate discussion on road selection for AV-ready subnetworks can be found in chapter 2 and a practical example is shown here in section 4.4 for a case study of Amsterdam. Since infrastructure adjustment projects tend to be costly and time consuming, a second step is required to select the best combination and timing of the adjustments that maximize their total societal benefits. The following subsections include the mathematical formulation of this problem. Table 4.1 provides the full notation used in this chapter.

4.3.1 General Assumptions

Let $G(N, A)$ denote a strongly connected graph representing the transportation network where N is the set of nodes and A is the set of directed arcs (links) representing the road and transit line segments. The planning horizon is divided into T equal decision stages each having a length of l years. For each stage τ, a decision is made to select a subset of roads denoted by A_1^τ for the AV-ready subnetwork represented by the graph $G_1^\tau(N_1^\tau, A_1^\tau)$ and the rest of the links are denoted by A_0^τ $(A = \{A_0^\tau \cup A_1^\tau\}, A_0^\tau \cap A_1^\tau = \varnothing)$. An example of AV-ready subnetwork graph is demonstrated in Figure 4.1. The effects of the construction period, which can be short given that the adjustments are not major, are ignored. All vehicles are allowed on all links in the network but level-3 and level-4 ADS-equipped vehicles (which will be referred to as AVs for the remainder of this chapter) can use their ADS only on the AV-ready subnetwork links, which are adjusted for safe and efficient use of ADS in mixed traffic. Vehicles with automation functions of levels 0-2 are referred to as RVs in this chapter. The travelers' response to these network design decisions is considered via a combined (simultaneous) mode and route choice user equilibrium, which is assumed to follow a hierarchical logit model. After each time stage, an AV diffusion model is used to estimate the market penetration rate of AVs in the next stage endogenously based on the level of service (expected satisfaction) of all available choices in the previous stage and the price difference between RVs and AVs.

Table 4.1 Notation

Notation	Definition
Sets	
W	Set of origin-destination (OD) pairs w
$R_m^{\tau,w}$	Set of routes r between OD pair w for mode m in year τ
M	Set of modes $m \in \{0,1,2\}$ (0 = PT, 1 = MD, 2 = AD)
K	Set of user classes $k \in \{0,1,2\}$ (0 = no access to vehicle, 1 = access to RV, 2 = access to AV)
A_0^τ	Set of links a not belonging to the AV-ready subnetwork in year τ
A_1^τ	Set of links a belonging to the AV-ready subnetwork in year τ
A	Set of all links a in the network; $A = \{A_0^\tau \cup A_1^\tau\}, \forall \tau \in \mathrm{T}$
Parameters	
μ_k	Logit route choice parameter for class k
θ_k	Logit mode choice parameter for class k
h_m	A constant representing attractiveness of mode m in the logit choice model
γ_m	PCU (PCE) of mode m
η_m	Value of travel time (VoTT) of mode m
$\bar{c}_{m,a}^\tau$	Fixed cost of mode m on link a in year τ (i.e., driving cost for RVs and AVs, and fare for PT)
t_a^0	Free flow travel time on link a
α_a, b_a	BPR function parameters of link a
Λ_a	Capacity of link a
$\delta_{m,a}^{w,r,\tau}$	Assignment map: 1 if route r between OD pair w for mode m includes link a, 0 otherwise
e_m^τ	Yearly expenses of using vehicles of mode $m \in \{1,2\}$
σ	Parameter converting the hourly travel cost to a yearly value
ρ_n	Diffusion function scale parameters
κ_a^τ	Adjustment cost of link a in year τ
ψ	Discount rate
$V^{\tau,w}$	Total number of vehicles (potential demand) between OD pair w in year τ
$\tilde{p}$	Market saturation rate of AVs
Variables	
$F_{m,r}^{\tau,w,k}$	(Route-based) flow of route r between OD pair w for class k and mode m in year τ
$F_m^{\tau,w,k}$	Flow of mode m between OD pair w for class k in year τ
$f_{m,a}^\tau$	(Link-based) flow of mode m on link a in year τ
q_a^τ	(Link-based) total flow (PCE-equivalent) on link a in year τ
$D^{\tau,w,k}$	Demand for (i.e., number of vehicles of) class k between OD pair w in year τ
t_a^τ	(Link-based) Travel time on link a in year τ
$c_{m,a}^\tau$	(Link-based) Travel cost of mode m on link a in year τ
$C_{m,r}^{\tau,w,k}$	(Route-based) travel cost of route r between OD pair w for class k and mode m in year τ
$\omega_m^{\tau,w,k}$	Expected satisfaction (logsum) of all routes for OD pair w for class k and mode m in year τ
TTC^τ	Total system travel cost in year τ
TTT^τ	Total system travel time in year τ
TTD^τ	Total system travel distance in year τ
TAC^τ	Total adjustment cost in year τ
x_a^τ	Binary variable taking the value 1 if link a is upgraded in year τ and 0 otherwise
X_a^τ	Binary variable with value of 1 if link a has been upgraded in year τ or before and 0 otherwise

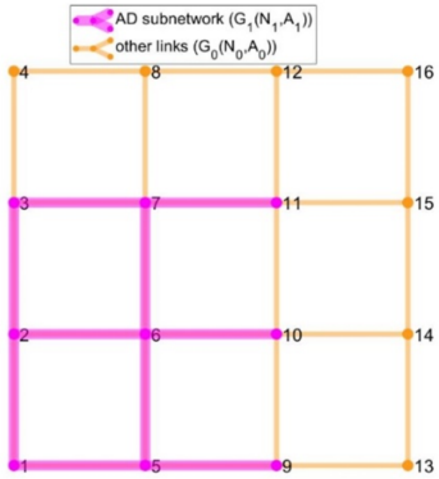


Figure 4.1 An example AV-ready subnetwork

Note: nodes 3-7-9-10-11 are boundary nodes, links (5,9)-(6,10)-(7,11) are inner boundary links, and links (9,13)-(10-14)-(11,15)-(11,12)-(7,8)-(3-4)-(9,10)-(10,11) are outer boundary links.

4.3.2 Lower Level Problem

The macroscopic static traffic assignment problem for RVs and AVs has been modeled using three general approaches so far in the literature. The first one is via the assumption of increased capacity for AVs (Chen et al., 2016; Ye and Wang, 2018) which is suitable when considering dedicated infrastructure for AVs. The next approach is considering different routing principles for RVs and AVs (Bagloee et al., 2017; Wang et al., 2019). This is suitable when those principles (e.g., system optimal routing) are applicable for travel behavior of AVs. A combination of system optimal routing and increased capacity for AVs is used in (Chen et al., 2017). Finally, the last approach is to consider AV-flow-specific (PCE-based) link travel times based on the ratio of AVs to RVs on each link along with assuming a different value of travel time (VoTT) for AD mode (Levin and Boyles, 2015; Liu and Song, 2019; Madadi et al., 2019). This approach is more common when considering a mixed traffic (RVs and AVs) on links. Note that system optimal routing is not realistic for mixed traffic conditions, particularly before reaching high market penetration rates of AVs. In this chapter, we follow the PCE-based approach and extend it for the multi-mode traffic assignment problem under consideration.

Three travel modes, namely automated driving (AD) mode, manual driving (MD) mode and public transport (PT) mode are considered here. Travelers are categorized within three separate classes based on their access to vehicles. Accordingly, each class has certain travel modes and their relevant routes available. Figure 4.2 depicts the travelers' choice tree for each class. The summation of the demand of all modes for each class for each origin-destination (OD) pair is fixed for each stage, but it can change over time and the proportion of demand assigned to each mode can change based on the travelers' mode choice. The following set of equations represents the behavioral rules considered in the network equilibrium model (NEM) for each stage τ.

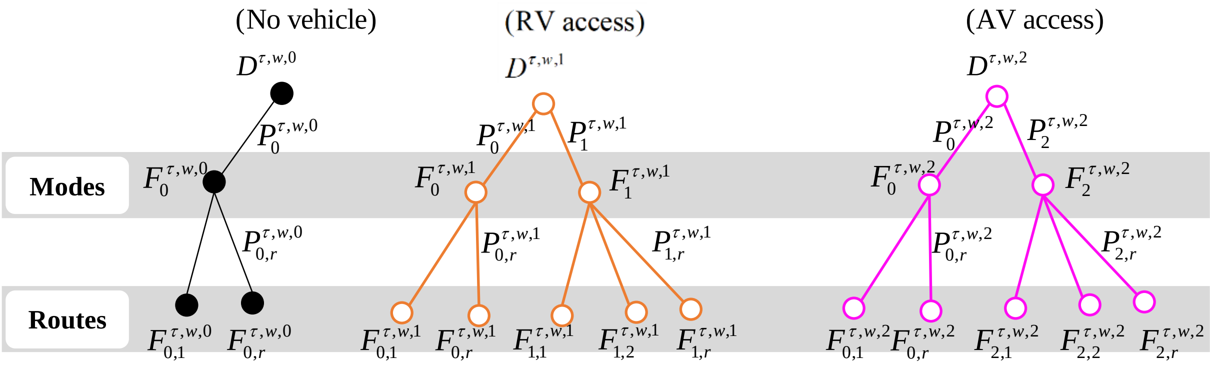


Figure 4.2 Travel decision choice tree

Note: Three classes of travelers represented in the figure are, from left, respectively, PT users (i.e., travelers without access to a vehicle (k = 0)), RV users (k = 1), AV users (k = 2).

NEM:

$$\sum_{w\in W}\sum_{k\in K}\sum_{r\in R_m^{w,\tau}} F_{m,r}^{\tau,w,k}\delta_{m,a}^{\tau,w,k} = f_{m,a}^{\tau}, \qquad \forall a\in A, \forall \tau\in \mathrm{T}, \forall m\in M, \qquad 4.1$$

$$\sum_{w\in W}\sum_{k\in K}\sum_{r\in R_m^{w,\tau}} C_{m,r}^{\tau,w,k}\delta_{m,r,a}^{\tau,w,k} = c_{m,a}^{\tau}, \qquad \forall a\in A, \forall \tau\in \mathrm{T}, \forall m\in M, \qquad 4.2$$

$$\sum_{r\in R_m^{w,\tau}} F_{m,r}^{\tau,w,k} = F_m^{\tau,w,k}, \qquad \forall \tau\in \mathrm{T}, \forall w\in W, \forall k\in K, \forall m\in M, \qquad 4.3$$

$$\sum_{m\in M} F_m^{\tau,w,k} = D^{\tau,w,k}, \qquad \forall \tau\in \mathrm{T}, \forall w\in W, \forall k\in K, \qquad 4.4$$

$$c_{m,a}^{\tau} = \bar{c}_{m,a}^{\tau} + \eta_m t_a^{\tau}, \qquad \forall a\in A, \forall \tau\in \mathrm{T}, \forall m\in M, \qquad 4.5$$

$$t_a^{\tau} = t_a^0[1+\alpha_a(\frac{q_a^{\tau}}{\Lambda_a^{\tau}})^{b_a}], \qquad \forall a\in A, \forall \tau\in \mathrm{T}, \qquad 4.6$$

$$q_a^{\tau} = \sum_{m\in M} f_{m,a}^{\tau}, \qquad \forall a\in A_0^{\tau}, \forall \tau\in \mathrm{T}, \qquad 4.7$$

$$q_a^{\tau} = \sum_{m\in M} \gamma_m f_{m,a}^{\tau}, \qquad \forall a\in A_1^{\tau}, \forall \tau\in \mathrm{T}, \qquad 4.8$$

$$F_{m,r}^{\tau,w,k}, F_m^{\tau,w,k} \ge 0, \qquad \forall \tau\in \mathrm{T}, \forall w\in W, \forall k\in K, \forall m\in M, \forall r\in R_m^{w,\tau}, \qquad 4.9$$

$$F_{1,r}^{\tau,w,0}, F_{2,r}^{\tau,w,0}, F_{2,r}^{\tau,w,1}, F_{1,r}^{\tau,w,2} = 0, \qquad \forall \tau\in \mathrm{T}, \forall w\in W, \forall r\in R_m^{w,\tau}, \qquad 4.10$$

where *Eq.4.1* stablishes the correspondence between route and links flows, *Eq.4.2* carries out the same for route costs and link costs, and *Eqs.4.3-4* guarantee flow conservation for routes and modes, respectively. *Eq.4.5* represents link travel costs, which include link travel time multiplied by VoTT for each class, plus a fixed cost (e.g., per kilometer driving cost for cars and fare for PT). *Eq.4.6* shows how link travel time is related to a BPR-type function based on the equivalent PCE-based flows of all classes shown in *Eqs.4.7-8* . *Eq.4.7* shows that the equivalent flow on regular links is the summation of mode flows, and *Eq.4.8* shows that on AV-ready subnetwork links, where AD is possible for the AV class, a lower PCE value is used for the AD mode which represents the shorter gaps between AVs. *Eq.4.9* guarantees feasible flows, and *Eq.4.10* restricts flows of classes in modes unavailable to them (i.e., owners of RVs do not have access to AD mode, owners of AVs drive in AD mode whenever available, and the class with no vehicle available does not have access to any car mode).

In the following parts, the equilibrium conditions of the NEM will be expressed in variational inequality (VI) and fixed-point (FP) formulations, and their equivalence will be shown. Once this is established, all solutions available for both formulations can be used to solve the NEM, while providing solution existence and uniqueness properties of one is sufficient for both.

Variational Inequality problem

Let the set Π defined by *Eqs.4.1-10* denote the admissible set of flows for the NEM under consideration (i.e., Π defines the feasible region of the NEM). It should be noted that Π is a non-empty, convex and compact set since the demand is non-zero and finite. Then the equilibrium condition for the VI problem is to find $\pi^* = [F_{m,r}^{*\tau,w,k}, F_m^{*\tau,w,k}] \in \Pi,$ such that $H(\pi^*)^T(\pi - \pi^*) \geq 0, \forall \pi \in \Pi,$ where $H(\pi) = [C_{m,r}^{\tau,w,k} + \frac{1}{\mu_m^k} \ln F_{m,r}^{\tau,w,k}, \frac{1}{\theta_m^k} \ln F_m^{\tau,w,k} - h_m^{\tau,w,k}]$, i.e., the VI problem is to solve the following:

$$\sum_{w \in W} \sum_{k \in K} \sum_{m \in M} \sum_{r \in R_m^{w,\tau}} [C_{m,r}^{\tau,w,k}(F_{m,r}^{*\tau,w,k}) - \frac{1}{\mu_m^k} \ln F_{m,r}^{*\tau,w,k}](F_{m,r}^{\tau,w,k} - F_{m,r}^{*\tau,w,k})$$

$$+ \sum_{w \in W} \sum_{k \in K} \sum_{m \in M} [\frac{1}{\theta_m^k} \ln F_m^{\tau,w,k} - h_m^{\tau,w,k}](F_m^{\tau,w,k} - F_m^{*\tau,w,k}) \geq 0 \,, \qquad \text{4.11.VI}$$

subject to *Eqs.4.1-10*.

Fixed-Point problem

The equilibrium condition for the FP problem is given below.

$$\text{Find } \pi^* \in \Pi, \text{satisfying } \pi^* = Proj_{\Pi}(\pi^* - \gamma H(\pi^*)), \forall \gamma > 0, \qquad \text{4.11.FP}$$

where $Proj_{\Pi}(x)$ denotes the orthogonal projection of point x on set Π with respect to the Euclidian norm. Then the FP problem can be expressed as:

solve *Eq.4.11.FP*, subject to *Eqs.4.1-10*.

Proposition 4.1. The VI problem defined by *Eq.4.11.VI* and the FP problem defined by *Eq.4.11.FP* are equivalent.

The proof is due to theorem 1.3 in (Nagurney, 1993).

4.3.3 Lower Level Solution

Daganzo (1983) has shown the existence and uniqueness of the solution (with respect to PCE-equivalent of link flows) for the FP formulation of the multi-mode multi-class SUE problem with asymmetric link costs and fixed demand. Cantarella (1997) has extended this to the elastic demand case. Zhou et al. (2009) have established existence and uniqueness of the solution for the VI formulation of the combined SUE problem (origin, destination, mode and route choice) with asymmetric link costs, which is a more general form of the NEM considered here. This is sufficient for confirming that the VI and FP formulations introduced above have a unique solution in terms of PCE-equivalent of link flows since the equivalence is already established. A comprehensive discussion on the equivalence of the mentioned formulations of the NEM is provided in (Florian and Hearn, 1995). Therefore, common solution methods for any of these problems can be used to solve the lower level problem (NEM) of the AD-NDP-T introduced here.

Regarding algorithms to solve VIs and FPs, the literature is vast. For review studies, see (Florian and Hearn, 1995; Harker and Pang, 1990). In this chapter, we use a linear approximation type algorithm introduced in (Wu et al., 2006) with step sizes according to the method of successive averages (MSA) and a parallel approach for updating class flows in each iteration. The algorithm was used by Wu et al. (2006) to solve a multi-class VI problem with asymmetric link costs where the efficiency and convergence of the algorithm were shown on a large-scale network with 99,867 links.

4.3.4 AV Diffusion Model

Diffusion models have been widely utilized in various fields as models to predict market penetration rates of new products and technologies. Recently, they have been used to predict the adoption rate of AVs (Chen et al., 2016; Lavasani et al., 2016; Nieuwenhuijsen et al., 2018). In this chapter, we use an adaptation of the model proposed by Yang & Meng (2001) and used by Chen et al. (2016). This model is suitable for our study since it relates the adoption rate of AVs in each stage to its adoption rate in the previous stage, net benefits provided by them during the previous stage, and their cost. This allows time-varying estimation of the number of AVs in each stage endogenously using the utility values obtained by the discrete choice model used for the lower level problem. Therefore, the number of AVs adopted by each OD pair in each stage is calculated as follows.

$$D^{\tau+1,w,2} = D^{\tau,w,2} + g(\phi^{w,\tau}) D^{\tau,w,2} \left(1 - \frac{D^{\tau,w,2}}{\tilde{p}\, V^{\tau+1,w}}\right), \qquad \forall \tau \in \mathrm{T}, \forall w \in W, \qquad 4.12$$

$$g(\phi^{w,\tau}) = \rho_1 \exp(\rho_2 [\phi^{w,\tau} - \bar{\phi}^{w,\tau}]), \qquad \forall \tau \in \mathrm{T}, \forall w \in W, \qquad 4.13$$

$$\phi^{w,\tau} = \sigma(\omega_1^{\tau,w,1} - \omega_2^{\tau,w,2}) - (e_1^{\tau} - e_2^{\tau}), \qquad \forall \tau \in \mathrm{T}, \forall w \in W, \qquad 4.14$$

$$\omega_m^{\tau,w,k} = \frac{1}{\mu_m^k} \ln \sum_{r \in R_m^{\tau,w}} \exp(\mu_m^k C_{m,r}^{\tau,w,k}), \qquad \forall \tau \in \mathrm{T}, \forall w \in W, \forall k \in K, \forall m \in M, \qquad 4.15$$

$$D^{\tau+1,w,1} + D^{\tau+1,w,2} = V^{\tau+1,w}, \qquad \forall \tau \in \mathrm{T}, \forall w \in W, \qquad 4.16$$

where *Eq.4.12* shows how AV demand in each stage is dependent on the AV demand in the previous stage, total number of vehicles, market saturation rate of AVs, and the function $g(\phi^{w,\tau})$. This function represents the intrinsic growth coefficient for the OD pair *w*. *Eqs.4.13-14* show how this coefficient is calculated based on the price difference between RVs and AVs and the difference in benefits (i.e., expected satisfactions) provided by each vehicle type based on the utilities of all routes available for that vehicle type during the previous stage. *Eq.4.15* indicates how the expected satisfactions are calculated according to the utility values obtained from the route choice model. *Eq.4.16* guarantees conservation of the total number of vehicles in each stage. This number is a model input. Note that the total number of travelers with and without access to cars for each OD pair throughout the planning horizon ($V^{\tau,w}$ and $D^{\tau,w,0}$) are assumed to be fixed here. Nevertheless, a general or an OD-based growth rate (in case that information becomes available) can easily be applied using a multiplier in *Eq.4.16*.

4.3.5 Upper Level Problem

The upper level problem involves deciding which links to upgrade and include in the AV-ready subnetwork in each decision stage. The objective is to minimize the sum of total discounted adjustment cost and total discounted travel cost over the whole planning horizon given the travelers' response to each network configuration captured by the NEM (i.e., the flow patterns used in the upper level problem to calculate the objective function value are obtained from solving the NEM). The following mathematical program represents the upper level problem.

Min

$$Z_U = \sum_{\tau \in [0,\mathrm{T}]} \left\{ \frac{\sigma TTC^{\tau} + TAC^{\tau}}{(1+\psi)^{l\tau}} + \sum_{j \in [1,l-1]} \left\{ \frac{\sigma TTC^{\tau}}{(1+\psi)^{l\tau+j}} \right\} \right\}, \quad l \geq 2, \qquad 4.17$$

s.t.

$$TTC^{\tau} = \sum_{a \in A} (1 - X_a^{\tau})[(\bar{c}_a^{\tau,1} + \eta^1 t_a^{\tau})(f_a^{*\tau,1} + f_a^{*\tau,2})]$$

$$+ \sum_{a \in A} X_a^{\tau}[(\bar{c}_a^{2,k} + \eta^2 t_a^{\tau}) f_a^{*\tau,2}], \qquad \forall \tau \in \mathrm{T}, \qquad 4.18$$

$$TAC^{\tau} = \sum_{a \in A} x_a^{\tau} \kappa_a^{\tau}, \qquad \forall \tau \in \mathrm{T}, \qquad 4.19$$

$$X_a^{\tau} = \sum_{\tau \in \mathrm{T}} x_a^{\tau}, \qquad \forall \tau \in \mathrm{T}, \forall a \in A, \qquad 4.20$$

$$X_a^{\tau} \leq 1, \qquad \forall \tau \in \mathrm{T}, \forall a \in A, \qquad 4.21$$

$$\left| p_{\hat{G}_1^{\tau}}^{s,s'} \right| \geq 1, \qquad \forall s, s' \in G_1^{\tau}, \forall \tau \in \mathrm{T}, \qquad 4.22$$

$$x_a^{\tau} \in \{0,1\}, \qquad \forall \tau \in \mathrm{T}, \forall a \in A, \qquad 4.23$$

where the objective function includes total discounted travel cost and total discounted adjustment cost. Since the length of each stage can be more than one year, a second term is added to the objective function to represent the net present value of the travel costs for years without any investment. Flows used in *Eq.4.18* are obtained by solving NEM. *Eq.4.19* represents total adjustment cost in each stage, *Eqs.4.20-21* ensure that each link can be upgraded only once and once it is upgraded, it stays part of the subnetwork, and *Eq.4.23* shows the decision variables of the upper level problem are binary. They assume the value of 1 for links that are chosen for the subnetwork and 0 otherwise (i.e., $X_a^\tau = 1, \forall a \in A_1^\tau$). *Eq.4.22* denotes the connectivity requirement for the AV-ready subnetwork. It entails that for any two nodes within the subnetwork, there should be at least one path within the (undirected) graph representing the subnetwork which connects those two nodes. The reason for inclusion of this constraint is to avoid subnetworks with separated parts. Since driving on AD mode is assumed to be allowed only on AV-ready subnetwork links, having subnetworks with separate components (i.e., disconnected subnetworks) leads to switching frequently between manual and AD mode, which is best to be avoided. This imposes extra requirements on solution methods, which are mentioned in the following section. The following definitions are necessary to describe how solution methods cope with the connectivity constraint. These definitions are demonstrated in Figure 4.1.

Definition 4.1. Any node n_1 incident to at least one link included in an AV-ready subnetwork (i.e., $G_1(N_1, A_1)$), is included in that AV-ready subnetwork.

Definition 4.2. The degree of a node is the number of links incident to that node within the graph in which that node is included.

Definition 4.3. A boundary node n_1^* for an AV-ready subnetwork represented by a directed graph G_1 is a node included in G_1 with a degree less than the degree of the corresponding node n in the underlying graph G (i.e., the graph representing the original road network).

Definition 4.4. An outer boundary link a_0^* for an AV-ready subnetwork is a link incident to a boundary node n_1^* of the AV-ready subnetwork but not included in the graph G_1 representing the AV-ready subnetwork.

Definition 4.5. An inner boundary link a_1^* for an AV-ready subnetwork is a link included in the graph G_1 representing the AV-ready subnetwork and incident only to boundary nodes with the degree of one.

Note that boundary nodes of an AV-ready subnetwork and its inner boundary links are included in the subnetwork while the corresponding outer boundary links are not included in that subnetwork (hence the difference in the subscripts).

4.3.6 Upper Level Solution

Any bi-level mathematical program is NP-Hard (Migdalas, 1995). In general, there is no exact solution procedure for bi-level NP-Hard problems. For single-stage DNDPs, genetic algorithm (GA) and simulated annealing have been the most common heuristic solutions so far (Farahani et al., 2013). DNDP-Ts on the other hand, have rarely been studied in the literature and are among the most difficult NDPs to solve. Therefore, in DNDP-T studies so far, either small case studies have been used that can be solved manually or approximate solutions have been used to solve the problem. Szeto et al., (2010) have used PREMIUM SOLVER PLATFORM (algorithm not reported) to solve a DNDP-T for a small synthetic network with four links. Miandoabchi et al., (2015) have used two metaheuristics, namely, non-dominated sorting GA and a B-cell algorithm on seven test networks with maximum 66 links to solve a multi-objective DNDP-T. O'Brien & Szeto (2007) have used a combination of branch and bound, and generalized reduced gradient for solving a DNDP-T. Finally, Chen et al. (2016) have solved a DNDP-T for the network of South Florida with 232 links using the active-set algorithm. It is crucial to notice that for combinatorial optimization problems such as the one considered here, computation times increase exponentially with respect to the problem size (i.e., number of integer decision variables). Therefore, for solving case studies with large-scale real networks, efficiency of the solution method is of paramount importance. On the other hand, the connectivity requirement of AV-ready subnetworks (*Eq.4.22*) makes most existing solutions ineffective since they are not designed to generate connected graphs. A generic remedy is using penalty functions but this can affect the efficiency of the solution method.

In this chapter, we solve the upper level problem using three evolutionary algorithms that are designed to deal with the complexity of the problem and the requirements mentioned earlier. We present two new algorithms, which were specially developed for this problem. They deal with the connectivity constraint via tailored operations that are inspired by evolutionary processes, yet adjusted to preserve connectivity of subnetworks. The third one is a modified GA that copes with the connectivity constraint via a penalty function. These three algorithms are described in the following subsections. Advantages and disadvantages of each algorithm are demonstrated using a case study with different scenarios.

Genetic algorithm (GA)

As mentioned earlier, GA introduced in (Holland, 1975) and elaborately discussed in (Golberg, 1989), is one of the most successful heuristic algorithms used for solving DNDPs. Therefore, we have used it here as a benchmark for the performance of existing solution methods. However, some modifications were necessary to guarantee the connectivity requirement of AV-ready subnetworks (*Eq.4.22*) and to accommodate the extra dimension (time) in decision variables. Instead of T sets of binary decision variables (one set for each time period), one set of integer decision variables in range of $[1, T+1]$ is used where the value specifies in which time period the link will be upgraded ($T+1$ represents never). The connectivity constraint is dealt with via a penalty function in fitness evaluation. GA operations are explained below and GA procedure is shown in Table 4.2.

GA initialization. GA here is initialized with a random configuration (i.e., a random selection of values within the feasible range for each decision variable).

GA mutation operation. A uniform mutation function is applied where 1% of genes (i.e., decision variables) in each chromosome (i.e., solution vector) selected based on a uniform probability distribution are perturbed to generate mutated offspring.

GA crossover operation. Crossover operation here is based on a multiple-point crossover function where multiple (uniformly selected) swapping points for chromosomes are used to exchange genes between crossover parents. This means each gene in each crossover offspring has an equal chance of being inherited from either crossover parent.

GA fitness evaluation. Fitness evaluation is according to the value of Z_U in *Eq.4.17* where the equilibrium flows used in there are obtained from solving NEM (lower level problem). In order to guide the algorithm to find connected designs (i.e., satisfy *Eq.4.22*), a penalty term is added to (17) for each subnetwork in each period with more than one connected component (i.e., designs that violate *Eq.4.17* are penalized in their fitness value). The exact value used for the penalty is 10^6 for each extra connected component. Note that there is exactly one connected component in any connected graph. Each fitness evaluation includes solving the NEM and checking for connectivity T times (once per each period) as well as running the diffusion model once after each period to specify demand for the next period.

Table 4.2 GA procedure

GA steps	
1	Initialize population (GA initialization) and measure fitness
2	**For** each generation **j**
3	**For** each individual **i**
4	Perform GA mutation operation
5	Measure fitness
6	**End**
7	Select parents using a binary tournament selection based on fitness
8	Perform GA crossover operation
9	Measure fitness
10	Select next generation from existing generation and produced offspring based on fitness
11	**If** stopping criteria met: **Terminate**
12	**End**

Evolutionary greedy search (EGS)

The EGS algorithm introduced here operates only on boundary links to preserve connectivity of produced subnetworks. It starts simple (one gene only) and the evolution process gradually adds complexity to the designs until there is no more gain from increasing complexity. To follow the general dynamic programming terminology, here we refer to an action as a decision for a single time stage and a policy as a series of decisions for the complete planning horizon. EGS operates on action space, starting from the first period it optimizes the design for that stage, and moves to the next stage once no further improvement is found for that stage. EGS operations are explained below and EGS procedure is shown in Table 4.3.

EGS initialization. EGS starts with a population of single links each one selected based on a roulette wheel prioritizing link capacity. Since there is no link elimination operation in EGS (i.e., once a link is added to the design, it stays in the design), starting with a larger number of links was found to be ineffective. Single links are sampled from the full set of feasible links, thus this process provides a sufficiently diverse (and connected) population pool to start the algorithm.

EGS extension operation. EGS operates on action space (i.e., optimizes one time stage at a time). This means EGS extension operation is performed only for one period at a time. Therefore, for each extension operation, first, the outer boundary links of the existing design for the active period are found. Next, a sample of outer boundary links are selected based on a roulette wheel with odds proportional to link capacity, and a number of them are added to the existing design. Adding only outer boundary links guarantees that the resulting designs are connected (i.e., they meet *Eq.4.22*). The number of candidate designs to evaluate and the number of links to add to each candidate are algorithm parameters.

EGS merging operation. To preserve diversity among the population of designs, and to expedite the search process, a merging operation is applied in EGS. It requires two parents who are first checked for compatibility. That is, parents are checked to determine whether they have a common node in their designs. Since both parents are connected designs, if they have a node in common, the union of their links will produce a connected graph. Note that checking for common nodes is computationally trivial compared to checking for connectivity after each merging. Then, if they pass the check, a merged design is generated from their union. Number of merged designs to produce in each generation is an EGS parameter.

EGS fitness evaluation. EGS fitness evaluation is based on the Z_U value in *Eq.4.17*. However, to reduce computation times and avoid unnecessary fitness evaluations, this operation is sliced into several pieces. In each time period, EGS starts with the optimal design obtained at the end of the previous time period and makes adjustments only on the active time period's designs. Therefore, EGS fitness evaluation only includes adding the fitness value of each design for the active period to the value of the optimal fitness at the end of previous period. In this manner, at the end of the planning horizon, the fitness values correspond to the objective function value in *Eq.4.17* while avoiding a large number of unnecessary fitness evaluations for inactive time stages. It will be shown in the next section that this approach is computationally very efficient.

Table 4.3 EGS procedure

EGS steps	
1	Initialize population (EGS initialization) and measure fitness
2	**For** each time stage **τ**
3	**For** each generation **j**
4	**For** each individual **i**
5	Perform EGS extension operation (on outer boundary links)
6	Measure fitness
7	**End**
8	Select parents using a binary tournament selection based on fitness
9	Perform EGS merging operation (on parents that can produce connected offspring)
10	Measure fitness
11	Select next generation from existing generation and offspring based on fitness
12	**If** stopping criteria met: **Go To** the next time stage
13	**End**
14	**End**

Evolutionary policy search (EPS)

EPS algorithm operates on policy space and uses operations inspired by evolutionary processes yet tailored to the problem to evolve to fitter designs while preserving connectivity. These operations include a context-aware merging and two types of mutations on boundary links. EPS operations are explained below and EPS procedure is shown in Table 4.4.

EPS initialization. EPS population is initialized with connected networks generated by the extension (similar to EGS extension) process without fitness evaluation. The number of starting links is a parameter. This means for each individual design, the algorithm starts with a random link, and adds links from the outer boundary links based on the roulette wheel explained earlier until a predefined number of links is added. The process provides the algorithm with a diverse set of connected designs to start the evolution process with minimal computational effort.

EPS extension operation. Since EPS operates on policy space, each time the extension operation is performed, first a time stage $\tau*$ is randomly selected as the active period for the operation. Then, an extension operation similar to EGS extension is performed on the active period's design. To guarantee *Eq.4.20*, once a set of links is selected to be added to a design on time stage $\tau*$, the same set is added to all following time stages (i.e., $[\tau*+1, T]$) of that design.

EPS reduction operation. This is the process of eliminating (ideally unwanted) links from designs to obtain better designs. However, to preserve connectivity of designs, candidate links for elimination are selected only from among inner boundary links. The number of links to be eliminated from each design and the number of candidate designs to consider for elimination are algorithm parameters. As in the extension operation, here the active time stage is randomly selected. However, the difference is that for the reduction operation, after the selection of candidates, all preceding time stages of candidate designs are modified accordingly. This guarantees satisfaction of *Eq.4.20*.

EPS merging operation. EPS merging is also similar to the EGS merging process including the check for nodes in common to ensure connected offspring. The difference is that first, an active period is randomly selected for EPS merging operation. If a common node is found for candidate parents in their active period's designs, then a merged design is generated from their union in the active period. The same design is used for the following periods, and the fitter parent contributes to designs of previous time stages for the merged offspring.

EPS fitness evaluation. EPS fitness evaluation is based on the Z_U value in *Eq.4.17*. Like GA, EPS measures the fitness once per period per design, with a diffusion function run after each period to define demand for the next period. However, no connectivity check or penalty is used for EPS since its operations guarantee connectivity. The optimal designs generated by EPS are checked for connectivity once after the optimization and as it will be shown later, they always meet the connectivity requirements of this problem.

Table 4.4 EPS procedure

EPS steps	
1	Initialize population (EPS initialization) and measure fitness
2	**For** each generation **j**
3	**For** each individual **i**
4	Randomly select the active time period τ^*
5	Perform EPS extension operation (on outer boundary
6	Measure fitness
7	Randomly select the active time period τ^*
8	Perform EPS reduction operation (on inner boundary
9	Measure fitness
10	**End**
11	Select parents using a binary tournament selection based on fitness
12	Perform EPS merging operation (on parents that can produce connected offspring)
13	Measure fitness
14	Select next generation from existing generation and produced offspring based on fitness
15	**If** stopping criteria met: **Terminate**
16	**End**

4.4 Case Study and Numerical Results

4.4.1 Description: Amsterdam Metropolitan Region

We demonstrate the multi-stage AV-ready subnetwork optimization concept on a case study of the Amsterdam metropolitan region. The network and demand data are obtained from the VENOM model (Kieft, 2013), which is based on the real network and demand patterns of Amsterdam, and is commissioned by the Amsterdam metropolitan region (Metropoolregio Amsterdam). It includes 52,812 links, 19,734 nodes, and 10,124 OD pairs (aggregated from 3722 original transportation zones). The study area (shown in Figure 4.3) includes 24,250 links and 6,642 OD pairs.

Demand data and availability of cars (thereby total number of users in each class) are extracted from the calibrated demand matrices per mode in VENOM model for the year 2004. The AV demand is determined assuming that the ratio of AVs to RVs available to travelers is the same as the AV market penetration rate and that the AVs available to travelers are homogeneously distributed in all zones. All transportation demand from, to, within and through the study area is included in the OD matrix and accounted for in the assignment, but network performance indicators are reported for the OD pairs within the study area. One morning peak hour on an average workday is modeled and a conversion rate of $\sigma = 10*12*30$ (30 days a month, 12 month a year, daily to peak hour travel time ratio of 10) is used to obtain yearly values for optimization. In the numerical results reported, all terms in *Eq.4.17* are divided by σ to avoid working with very large numbers. The traffic pattern of the base case (as is) is depicted in Figure 4.4.

Since the focus of this chapter is automated vehicles, in order to reduce computation times, a simplified assignment for public transport is used where for each OD pair, one (artificial) link represents expected satisfaction of all available routes with public transport. The data for calculating mentioned values is derived from (Brands, 2015) which has used the same network and demand data (VENOM model) for an NDP study with a focus on public transport.

Regarding the diffusion model, the starting AV penetration rate used is 5%, the potential market size for AVs (saturation rate) is 90%, and the annual cost difference between RVs and AVs is 2000 €.

Motorways, regional roads and main urban roads are considered as feasible links for the subnetwork, and (per kilometer) link adjustment costs used in the case study are 50,000 €/km for motorways, 75,000 €/km for regional roads and 100,000 €/km for main urban roads. The maintenance cost per period is 5% of the adjustment cost, which is added to the link adjustment costs. Net present values of economic benefits (i.e., total travel cost savings) and costs (i.e., total adjustment cost) up to one period after the planning horizon are calculated and added to the reference point. The effect of variations of adjustment costs from mentioned values is considered via sensitivity analysis reported in the next section. Total number of links selected as feasible links for the AV-ready subnetwork is 5,804 out of the 52,812 links (shown in Figure 4.3). This leads to a lower level problem (i.e., NEM) with 52,812 continuous decision variables and an upper level problem with 5,804 binary decision variables ($2^{5,804}$ possible solutions) for each period. Three different scenarios, described below, are considered to account for the effects of planning horizon and decision period length.

Scenario 1: six periods with preselection. A 30-year planning horizon is considered for all scenarios. In scenario 1, the planning horizon is divided into 6 periods (i.e., 6 viewing or decision points) with the length of 5 years for each period. It is assumed in this scenario that for period 0 (corresponding to the beginning of the planning horizon), all motorways are (pre)selected as part of the AV-ready subnetwork (i.e., the value of decision variables for period 0 are preselected and fixed). The main reason for this choice is that many experts believe motorways are the first places to facilitate AD; therefore, they should be included in any network configuration for AVs. Moreover, reducing the number of decision variables (of the upper level problem) can decrease computation times as well.

Scenario 2: two periods with preselection. In this scenario, a 30-year planning horizon with 2 time periods (each 15 years long) and the same link preselection for period 0 as scenario 1 is considered. Since in this scenario the length of the planning horizon is the same as scenario 1 yet investment decisions are made less frequently, the comparison aids in demonstrating the effects of investment frequencies.

Scenario 3: two periods without preselection. A 30-year planning horizon with 2 time periods (each 15 years long) and no link preselection for period 0 is considered in this scenario. Since the value of the objective function for period 0 is also included in all calculations for all scenarios, comparing scenario 3 with scenario 2 provides some insight into the impacts of preselecting motorways.

Hardware and software

The mathematical model and the solution algorithms were coded in MATLAB and ran on a Windows PC with a Core i5-8600 CPU @ 3.10GHz and 32 GB RAM. MATLAB parallel computation toolbox was utilized with 6 parallel computing units for efficient computations and dealing with the computational complexity arising from the problem size. It should be noted that population-based algorithms (such as evolutionary algorithms used in this chapter) can fully utilize the potential of parallel computation for efficiency. Moreover, sparse matrices in MATLAB were used for all algebraic operations on assignment maps to minimize the computation times of the MSA-based algorithm used to solve the lower level problem (NEM). Overall, 3 scenarios, 3 algorithms and 5 runs for each algorithm in each scenario led to a total of 45 optimization runs for the Amsterdam case study (excluding parameter tuning and sensitivity analysis), which culminated in approximately 72 days of computations.

Figure 4.3 Amsterdam case study: study area and feasible links for AV-ready subnetwork

Figure 4.4 Traffic patterns in Amsterdam case study: base case (as is)

Note: Bandwidth represents relative flow.

4.4.2 Numerical Results and Analysis

In this section, we assess the effects of deploying AV-ready subnetworks on network performance using three main network performance criteria, namely, total travel cost (TTC), total travel time (TTT) and total travel distance (TTD). Reported values for objective function (OF) are averages of 5 independent runs (replications) as well as bandwidths for 95% confidence intervals. Computation times (CTs) are reported in the same manner. TTC, TTT and TTD values are averages per stage per run. This allows comparisons with the base case. Note that due to the scaling of *Eq.17* explained in case study description, the values of TTC, TTT and TTD reported in Table 4.5 are hourly values rather than yearly values. Total adjustment cost (TAC) values are summed over all stages and averaged over all runs to show total investment values per scenario.

Since the global optimum point of the problem is unknown, and with heuristics, there is no guarantee to find this point, we have provided two additional variations as reference points for comparisons. First, the network performance criteria for a variation where all feasible links are included in the AV-ready subnetwork in all stages is reported (this is referred to as “all links” variation). Second, the values of the base case (as is) are reported to show network performance as it is before the changes. Comparing TTC values obtained by each algorithm with the base case provides a measure of how much improvement has been achieved using that algorithm, and the “all links” variation provides an optimistic lower bound for TTC values. Moreover, we discuss the performance of the (upper level) solution algorithms and the impacts of scenario settings with respect to several performance criteria. The solution algorithms are compared in terms of effectiveness (i.e., average OF value obtained), efficiency (i.e., average computation times), stability (i.e., within run variations), scalability (i.e., computation time to problem size ratio), and constraint satisfaction (i.e., producing connected graphs).

Network performance

As evidenced by Table 4.5, there is a notable network-wide decrease in (per stage) TTC values in all scenarios with all algorithms compared to the base case, which indicates that the AV-ready subnetwork concept has an overall positive impact on network performance in terms of TTC. In addition, TTC values (especially for EGS) are close to the values of the variation where all feasible links are included. Since the only two criteria considered in the upper level objective function are TTC and TAC, this suggests that the optimization results are near optimal. Note that the TTC values of “all links” variations are not necessarily reachable in reality, yet they provide optimistic lower bounds.

Regarding TTT, the trend is very similar to TTC with the exception of EGS in scenario 3 where it outperforms the “all links” scenario. This is logical, since the algorithm optimizes for TTC and TAC but not for TTT. Moreover, the tendency to choose longer routes via the AV-ready subnetwork in AD mode can cause more travel time with denser subnetworks while leading to lower TTC due to lower VoTT of AD mode. This also explains considerably higher TTDs in all optimal cases and “all links” scenarios compared to the base case. The traffic patterns of optimal cases are very similar to the base case patterns (Figure 4.4) with a slight increase in the volumes in the latter case. The explanation is that when AVs become more attractive in time with the existence of the AV-ready subnetworks, more travelers opt for cars. This leads to an increase in volumes on the network; nonetheless, the performance of the network is still favorable to the base case in terms of TTC and TTT due to the efficiency of AD mode.

Table 4.5 Summary of case study results

Algorithm	OF (€) (average)	TTC (€) (average)	TTT (h) (average)	TTD (km) (average)	TAC (€) (sum)	Connected	CT (h) (average)
Scenario 1: six periods with preselection							
GA	22,597,906 ± 3,635	1,150,222	97,140	1,373,965	245,869,501	No	154.92 ± 43
EGS	22,534,825 ± 33,797	1,146,307	97,051	1,378,132	250,380,768	Yes	5.53 ± 0.14
EPS	22,685,967 ± 1,626	1,156,838	97,289	1,366,779	240,203,049	Yes	144.06 ± 0.97
All links		1,145,451	97,038	1,379,113			
As is		1,185,655	97,208	1,314,145			
Scenario 2: two periods with preselection							
GA	25,143,063 ± 321	1,150,777	97,181	1,373,674	212,675,527	Yes	7.95 ± 0.08
EGS	25,023,282 ± 3	1,144,292	97,026	1,380,753	216,024,677	Yes	4.03 ± 0.03
EPS	25,139,950 ± 333	1,150,612	97,175	1,373,890	212,680,670	Yes	9.15 ± 0.09
All links		1,143,857	97,017	1,381,291			
As is		1,185,655	97,208	1,314,145			
Scenario 3: two periods without preselection							
GA	25,282,900 ± 243	1,159,160	97,106	1,357,159	198,926,645	Yes	8.78 ± 0.15
EGS	25,002,051 ± 1,138	1,144,221	96,920	1,380,716	131,099,802	Yes	6.72 ± 0.18
EPS	25,029,462 ± 1,400	1,145,238	97,026	1,379,269	209,199,259	Yes	3.78 ± 0.17
All links		1,143,907	97,014	1,381,163			
As is		1,185,655	97,208	1,314,145			

Note: ± signs denote standard errors for 95% confidence intervals.

Algorithm performance

In terms of effectiveness (i.e., average OF value obtained) EGS outperforms competing algorithms in all scenarios. Especially in scenario 1 when the number of possible solutions is considerably higher, the performance of the other two algorithms is not competitive to EGS at all. It appears they become ineffective when the problem becomes too large, even though their computation times are much higher than EGS. In this scenario, the greediness of EGS seems to serve well in terms of both efficiency and effectiveness. GA on the other hand, is outperformed in effectiveness by all other solutions in all scenarios, except for scenario 1 where, on average, it outperforms EPS.

When it comes to efficiency (i.e., average computation times), EGS is the best performer again with a major margin in most cases with the exception of scenario 3 where EPS has a lower computation time. EGS starts with a single link, and adds only a small number of links at each generation. Therefore, in scenario 3 where there is no link preselection, which leads to a higher number of decision variables, it takes longer than other scenarios, even longer than scenario 1.

As for stability, we reflect on within run variations captured by standard errors of both OF values and computation times (denoted by ± signs in Table 4.5). GA and EPS demonstrate notable stability in both OF values and computation times in scenario 2 and scenario 3. However, in scenario 1, EPS is somewhat stable around an undesirable OF value, and GA is highly unstable in terms of computation times and rather stable around a non-competitive OF value. Conversely, EGS shows persistent stability in computation times. Regarding OF values, EGS has the highest stability in scenario 2, followed by scenario 3 and scenario 1. It is worth noticing that although EGS shows a rather large within run variability for OF in scenario 1, its worse run is still superior to the other algorithms' best run (i.e., the upper bound of EGS confidence interval for OF is significantly lower that the lower bounds of GA and EPS confidence intervals for their OF value).

Regarding scalability (i.e., computation time to problem size ratio), computation times of EPS and GA grow exponentially and their performances deteriorate with an increase in the number of time periods, while EGS computation time grows rather linear and remains effective with more time periods. With the increase in the number of decision variables, EGS computation times grow rather linear while maintaining its level of effectiveness, whereas EPS and GA show less sensitivity to (a limited) increase in the number of decision variables. This relates to their structure and spaces on which they operate; EPS and GA operate on policy space, whereas EGS operates on action space. In policy space, the number of possible solutions grows exponentially with the number of time periods. While in action space, when the value of decision variables of previous stages are fixed in each stage, the number of possible solutions increase linearly with the increase in the number of time periods. On the other hand, operating in such a manner on action space in a multi-stage setting makes EGS a greedy algorithm, since the decisions on each stage are taken without considerations for the later stages. Yet that does not seem to deter its performance. This could relate to the nature of the problem, since another greedy algorithm has been shown to perform well for the AV-ready subnetwork optimization problem without the time dimension (chapter 3).

Constraint satisfaction here is related to satisfying *Eq.4.22*, i.e., producing connected AV-ready subnetworks as solutions. For a comprehensive discussion on the necessity of this constraint, the reader is referred to the chapter 3 of this thesis. As is shown in Table 4.5 and depicted in Figures 4.5-7, EGS and EPS generate connected subnetworks in all periods and all scenarios.

This was to be expected due to their operations, which are tailored to this purpose. On the contrary, GA does not always meet this criterion, especially in scenario 1. Even though a penalty is applied for each disconnected component in each period, and various values for this penalty were explored. It is shown in chapter 3 that GA with a similar penalty successfully finds connected designs for the single stage, unimodal problem. Likewise, GA with penalty satisfies the connectivity constraint in scenario 2 and scenario 3 of this chapter. However, with the increase in the number of stages, the penalty function proves to be ineffective for satisfying the connectivity constraint. This goes to show that the problem size is a defining factor for performance of the algorithms and it should be taken into consideration while selecting an appropriate solution method for such problems.

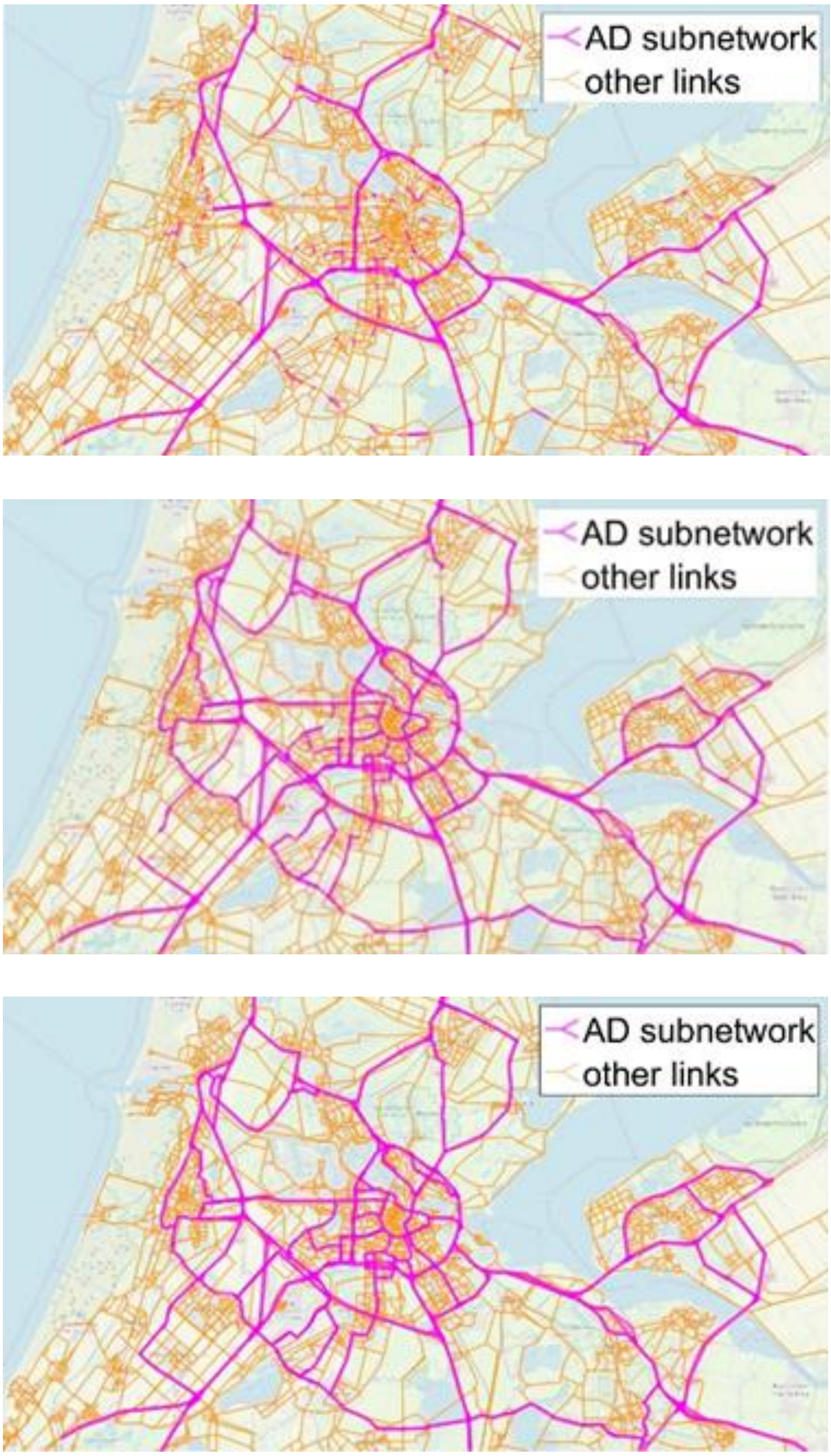
Sensitivity analysis

In this section, we discuss possible variations in model input parameters and their impacts on the output. The values reported in Table 4.6 are averages of 3 independent EGS runs for scenario 3. We analyzed 3 general categories of parameters, namely, diffusion model parameters, adjustment cost parameters and mode choice parameters.

Based on Table 4.6, the most sensitive model parameter is ρ_1, which represents the sensitivity of AV demand to its overall cost (i.e., generalized travel cost and ownership cost) in the diffusion model. This is also evident from Figure 4.8 where it has been demonstrated that the evolution of market penetration rate of AVs can take considerably different paths with different values of this parameter. This signifies the importance of accurate AV demand prediction for infrastructure planning decisions.

On the contrary, adjustment cost parameters are the least sensitive parameters. Although severe and rather proportional changes in TAC values are observed with variations in adjustment costs, which is natural, the changes in objective function value and network performance indicators are trivial in all cases. This indicates that the significance of proper adjustment cost estimation is for project budget but not for network performance.

Regarding mode choice parameters, an increase in the (absolute) value of the sensitivity of AVs to the level of service (expected satisfaction) has the least significant impact on the results. On the other hand, a decrease in the sensitivity of AVs to the level of service demands greater investment with less positive impact on the network performance. With RVs, the direction of changes is the opposite; less sensitivity to the level of service leads to better network performance and more sensitivity diminishes network performance.

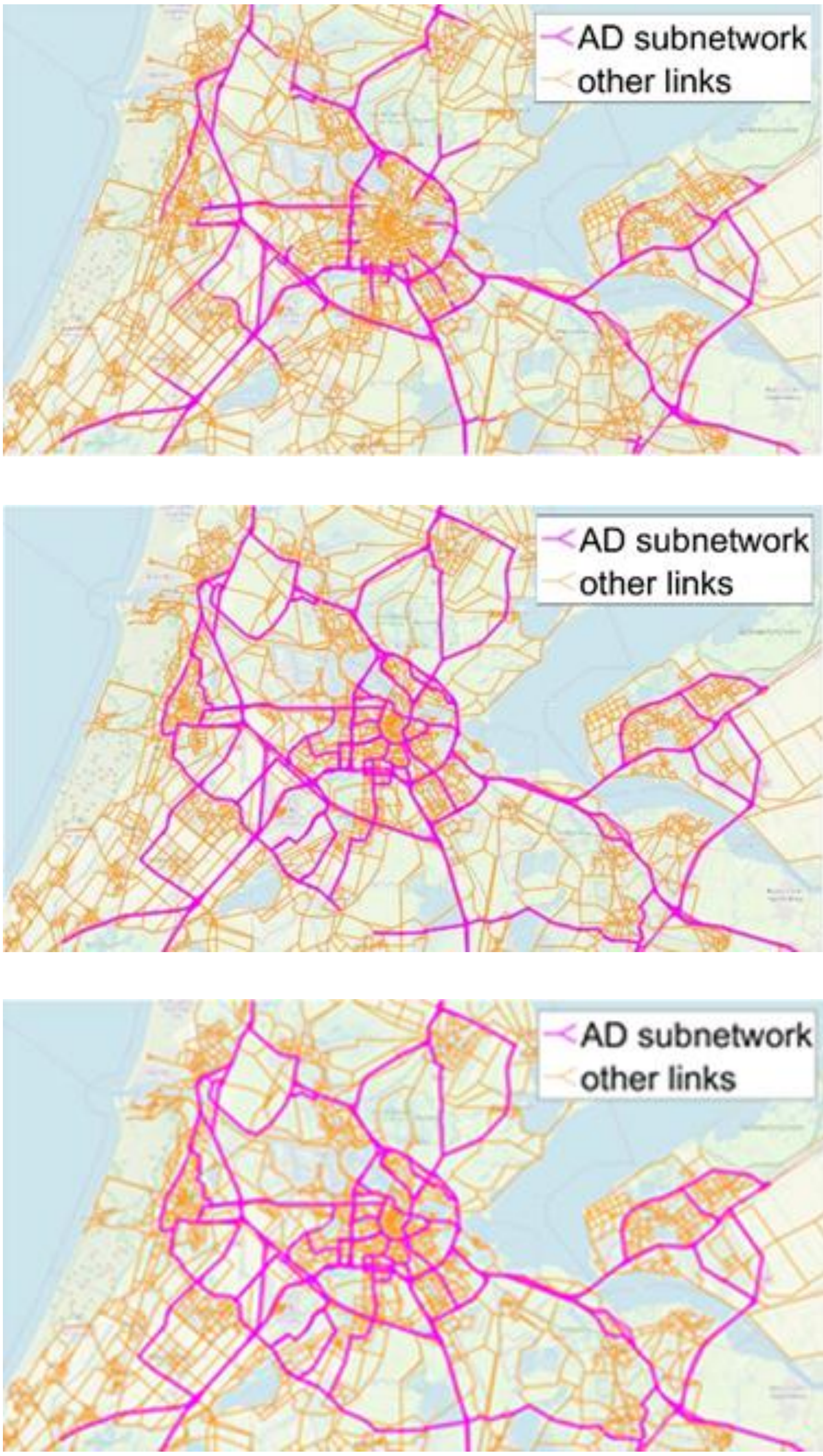


(a) Period 2

(b) Period 4

(c) Period 6

Figure 4.5 Optimal evolution of AV-ready subnetworks obtained by GA for Scenario 1

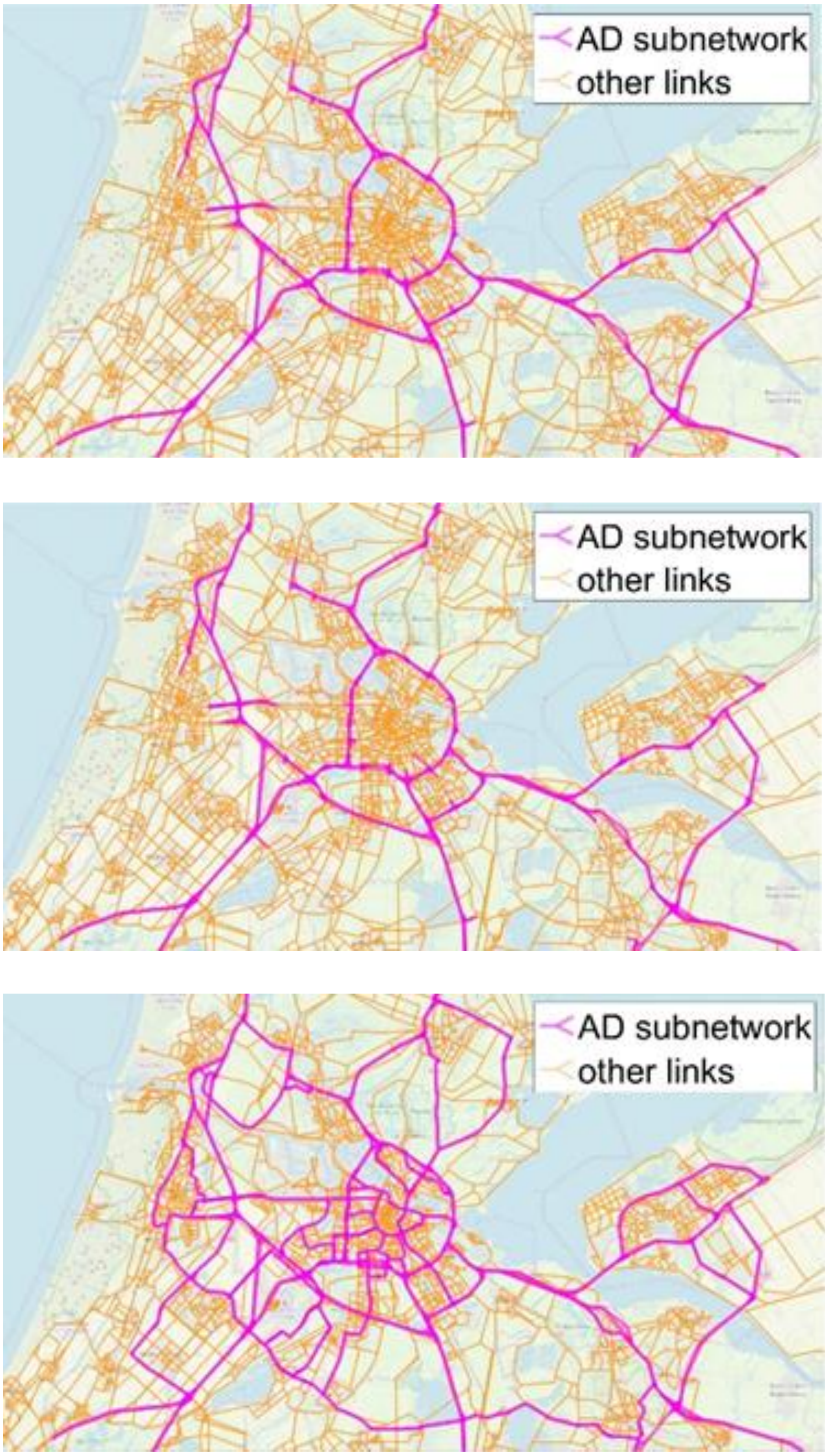


(a) Period 2

(b) Period 4

(c) Period 6

Figure 4.6 Optimal evolution of AV-ready subnetworks obtained by EGS for Scenario 1



(a) Period 2

(b) Period 4

(c) Period 6

Figure 4.7 Optimal evolution of AV-ready subnetworks obtained by EPS for Scenario 1

Table 4.6 Sensitivity analysis (% changes in scenario 3)

Category	Parameter	Variation	OF	TTC	TTT	TTD	TAC
Diffusion model	ρ_1	50%	-0.69%	-0.83%	-0.03%	1.15%	4.59%
	ρ_1	-50%	1.36%	2.02%	0.14%	-2.68%	-11.64%
	ρ_2	50%	0.00%	0.00%	0.00%	0.00%	-2.36%
	ρ_2	-50%	0.01%	0.00%	0.01%	-0.01%	8.02%
Adjustment cost	Cost ratios: (Motorway,Regional,Urban)	(-50%, -50%,-50%)	-0.04%	0.00%	0.01%	-0.01%	-45.58%
		(+50%,+50%,+50%)	0.04%	0.01%	0.00%	-0.01%	43.20%
		(+50%, 0% , -50%)	0.03%	0.02%	0.00%	-0.01%	17.51%
		(+100%,0%,-100%)	0.02%	-0.01%	-0.01%	0.01%	28.75%
Mode choice	θ_2	50%	0.00%	0.01%	-0.01%	0.00%	-2.58%
	θ_2	-50%	0.02%	0.02%	0.04%	-0.03%	8.86%
	θ_1	50%	0.01%	0.01%	0.02%	-0.02%	4.44%
	θ_1	-50%	-0.02%	-0.02%	-0.06%	0.03%	-10.99%

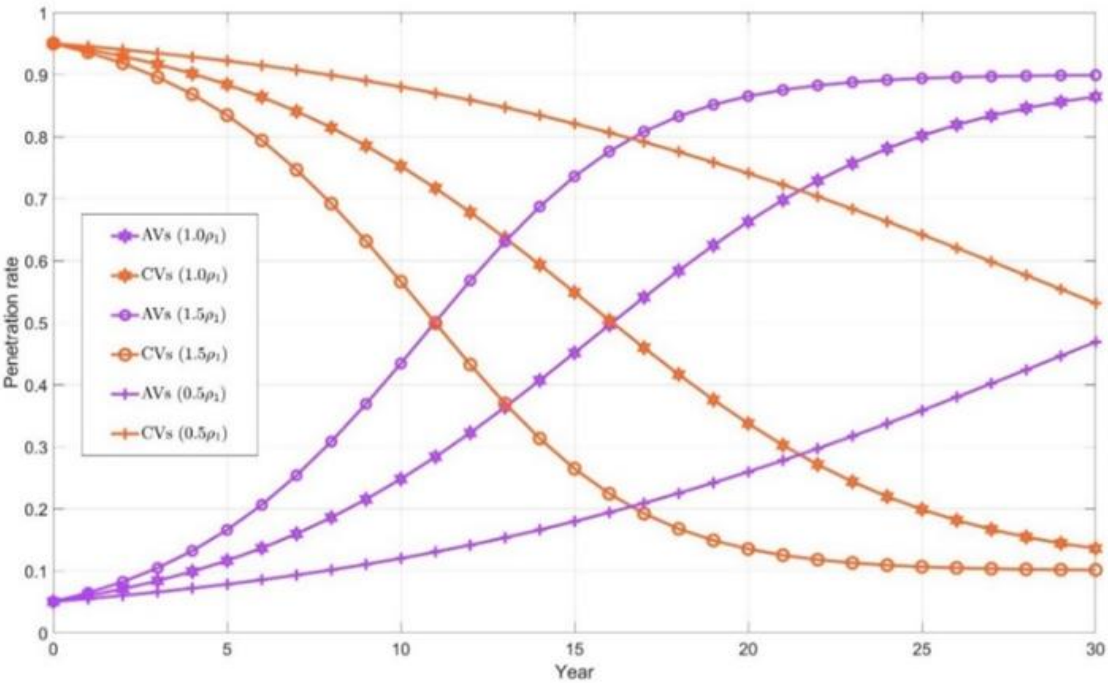


Figure 4.8 Sensitivity analysis for diffusion model parameters in scenario 3

4.5 Summary and Conclusions

In this chapter, we considered the problem of multi-stage optimization of AV-ready subnetworks within road networks with time-varying demand. We modeled the problem as a time-dependent bi-level NDP where the upper level denoted infrastructure decisions made by authorities in several stages over a planning horizon, and the lower level represented mode and route choices of different classes of travelers (with access to AVs, RVs or no vehicle) in each stage in response to the infrastructure supply. We presented VI and FP formulations of the lower level as a multi-class simultaneous mode and route choice UE using a hierarchical logit model, and solved it using a linear approximation type algorithm with step sizes based on MSA. The upper level problem was modeled as a mathematical program with binary decision variables and solved for near-optimal solutions using three evolutionary heuristics. It is crucial to notice that network configurations for AVs are sensitive to the level of AV demand, which is expected to evolve over time. Therefore, these configurations should evolve over time as well. This makes the multi-stage planning approach necessary, which adds tremendous complexity to the problem and calls for efficient solution methods. Furthermore, we used the real road network of the Amsterdam metropolitan region to demonstrate the concept and compare the performance of the solutions.

Two tailored evolutionary algorithms, namely EGS and EPS, were developed in this chapter and their performance was compared to a GA with a penalty function (to satisfy constraints) using the case study. Both EGS and EPS successfully satisfied the constraints in all scenarios while GA failed to meet this requirement with larger number of Stages. Regarding effectiveness and efficiency, EGS outperformed both competing algorithms overall. The advantage of GA over the proposed algorithms is that it is available in most optimization packages and can easily be applied to the problem with a rather standard penalty function. However, in this study, with the growth in problem size, performance of the GA became unsatisfactory.

It was shown that AV-ready subnetworks deliver significant benefits in terms of TTC and TTT. The extent of these benefits increased with higher AV penetration rates and more AV-ready roads. However, this was accompanied by slightly higher travel distances for AVs, which can cause higher emissions; though, AVs can be more energy efficient and cause less congestion.

4.5.1 Future Research Directions

Several shortcomings of this study and avenues for future research are discussed in the following paragraphs.

Macroscopic static traffic assignment models have been commonly used for the lower level of bi-level NDPs, even though dynamic traffic assignment models can capture the behavioral differences of RVs and AVs more accurately and allow for more elaborate intersection modeling. This is due to the general complexity of NDPs and high computation times of dynamic traffic assignment models. With advances in technology, computers with higher computation power are becoming available. However, application of dynamic traffic assignments for AD-NDP-T studies, which are computationally much more demanding compared to standard NDP studies, remains a challenge, especially for case studies of large-scale networks.

An interesting extension of this study is combining other network configurations, such as dedicated lanes and roads for AVs with AV-ready subnetworks for mixed traffic and developing a unified modeling framework for combinations of network design concepts for AVs. So far, these concepts have been modeled separately with incompatible frameworks. This makes it difficult to study them simultaneously using one model. Nevertheless, their combination can be relevant, particularly for large regions with various road types and jurisdictions.

The sensitivity analysis performed in this study indicated that the AV diffusion model parameters are the most sensitive parameters of the study. Therefore, fine-tuning these parameters can aid in accurate estimation of AV market penetration rate over time. However, since highly automated vehicles are not available on the market yet, fine-tuning the diffusion model parameters or validating its output is not possible at the moment. Moreover, models that can estimate market penetration rate of AVs and are compatible with discrete choice models are still rare in the academic literature. Nonetheless, an ideal AV demand estimation model for NDPs with AVs should be less reliant on scale parameters.

When a long planning horizon is considered, origin and destination choices might become relevant as well. This can add yet another layer of complexity to an already complex model. Nonetheless, assuming fixed OD pairs is a non-trivial simplification present in all NDP studies with a few exceptions.

Finally, a more detailed multimodal traffic assignment model including active modes, a more detailed representation of public transport, and combined modes such as park and ride can capture all available choices to the travelers and improve the accuracy of the model results.

5 A Unified Framework for Optimizing Road Networks for Automated Vehicles

The main objective of this chapter is to generalize the bi-level model of chapter 3 in order to develop a unified network design framework for AVs that combines viable network configurations for the transition period to full automation, namely AV-ready subnetworks, dedicated AV links and dedicated AV lanes, in one comprehensive model. This addresses the last research question of this thesis. The existing models for different network configurations to accommodate AVs are incompatible, making it difficult to compare and combine them. The model proposed in this chapter facilitates exploring and comparing different network configurations and their combinations for road networks before committing to major infrastructure investment projects using only one model for all AV network facilities. This chapter includes the formulation of the problem, a new solution method to solve the problem, a case study of a realistic road network to showcase the applicability of the model proposed, and a discussion on practical considerations related to the deployment of this network design method.

This chapter is currently under review for journal publication.

5.1 Introduction

Automated vehicles (AVs) are expected to deliver various benefits to transport systems, including traffic efficiency (Mahmassani, 2016; Shladover et al., 2012; van Arem et al., 2006). However, reaching a high market penetration rate (MPR) of fully automated vehicles is a gradual process that can take several decades. Thus for a long time, a heterogeneous mix of traffic with AVs and regular vehicles (RVs) on the roads will be inevitable. The mixed traffic in this transition period can reduce the traffic efficiency benefits expected to be provided by AVs and even negatively affect traffic flows (Calvert et al., 2017; Ghiasi et al., 2017). This period is referred to in (Luo et al., 2019) as the “dark age” of AVs.

On the other hand, according to (SAE International, 2018), the operating design domain (ODD) of highly automated vehicles (i.e., level-4 AVs) is limited. Therefore, these vehicles’ automated driving system (ADS) cannot be used in all roads and all driving conditions. In order for an ADS to operate without failures, it should only be activated within environments in which it can safely function. However, the exact extent of level-4 ODDs are not clear at the moment. Analyses of AV tests in the U.S. indicate that existing AVs are not capable of using their ADS (i.e., operating in AD mode) without failure everywhere on regular infrastructure yet (Boggs et al., 2020; Favarò et al., 2018).

In order to extend the ODD of AVs and to amplify their traffic efficiency benefits, various infrastructure-based solutions have been suggested. Carreras et al. (2018) have classified road infrastructure based on the level of support it can provide for AVs, and advocated cooperative automated driving enabled by infrastructure support elements such as static and dynamic digital maps, real-time microscopic traffic state data and vehicle movement information to improve traffic flow performance. Soteropoulos et al. (2020) have proposed an automated drivability index for different roads in the network of Vienna to determine which roads are suitable for level-4 AVs from a technological standpoint. They conclude that deployment of level-4 AVs in streets with low drivability index would only be possible with major infrastructure adjustments. Lu and Blokpoel (2017) have advocated intelligent infrastructure to support AVs. Rondinone et al. (2018) have discussed vehicle to infrastructure (V2I) communications to support cooperative automated driving. Lu et al. (2019) have provided lists of infrastructure requirements for AVs under four future scenarios based on expert opinion. Farah et al. (2018) have provided a review of studies indicating physical and digital infrastructure requirements to support AVs.

Another potential solution for the transition period is dedicated AV lanes. However, it has been shown that with a low percentage of AVs on a link, dedicating one or more lanes to AVs can lead to underutilization of dedicated lanes and a reduction of the overall link throughput (Mahmassani, 2016; van Arem et al., 2006). Recognizing this, Chen et al. (2016) considered the problem from a network perspective and proposed a model for deployment of a network of dedicated AV lanes over time. Ye and Wang (2018) suggested dedicated AV links (i.e., stretches of roads dedicated to AVs) and congestion pricing for RVs. Then again, this is only suitable when the MPR of AVs is very high. Chen et al. (2017) suggested dedicated AV zones; however, this is also relevant for networks with very high MPR of AVs. Moreover, dedicating parts of a network to one class of vehicles can compromise the accessibility of other vehicles and modes.

Li et al. (2020) proposed utilizing roadside units for V2I communication to overcome the connectivity gap in mixed traffic. This allows AVs to drive closer to their leading vehicles, which can increase link capacity in mixed traffic. In chapter 2, we suggested an AV-ready subnetwork where the infrastructure is enhanced to allow efficient use of ADS in mixed traffic. Furthermore, we proposed a model for optimal deployment of AV-ready subnetworks in chapter 3, and studied the evolution of these subnetworks over time with a multi-stage model in chapter 4. However, the potential link capacity increase with AVs in mixed traffic can be lower compared to configurations with segregated AV traffic.

So far, there seems to be an array of options available for transport planners and road authorities to accommodate AVs on road networks during the transition period to full automation. In order to make informed decisions, they need to test these solutions on their networks using transport models. However, the existing models for mentioned design concepts are incompatible. Therefore, exploring them requires several models, comparing them is difficult, and combining them is not possible with existing models. On the other hand, different AV MPRs and different environments (urban, rural, etc.) call for different network configurations. Moreover, the transition from low to high MPR of AVs will require evolution of road network designs to serve the AV demand appropriately. This means the best solution for the transition period might be a combination of different network configurations, which might change over time as well.

Therefore, this chapter proposes a unified formulation for combining AV-ready subnetworks for mixed traffic, dedicated AV links and dedicated AV lanes. The objective of the formulation is optimizing the costs and benefits of deploying the aforementioned network design concepts using a bi-level modeling framework. The upper level includes network-wide decisions with respect to the links to be selected as part of the AV-ready subnetwork, dedicated AV links and dedicated AV lanes. The lower level represents a macroscopic network equilibrium model developed to capture the travelers' route choice and the propagation of traffic throughout the network given a certain network topology for accommodating AVs. In order to solve the problem and find coherent network configurations, a heuristic solution procedure based on an evolutionary algorithm is developed in this study as well. Applicability of the proposed model is demonstrated on a realistic case study of the road network of the Amsterdam metropolitan region. Using extensive numerical experiments on this case study, we discuss important practical issues and considerations related to the deployment of the introduced design concept in real road networks, which are not observed in theoretical networks commonly used in academic studies. The model presented here allows planners to compare and combine different network configurations in order to find the most suitable design for their network based on the level of demand in different periods using only one model. To the best of our knowledge, this is the first time such a combination is studied in the literature.

The rest of this chapter is organized as follows. Section 5.2 presents the methodology, which includes the problem definition, a multi-class network equilibrium model, a deployment model of AV-ready subnetworks, dedicated AV links and dedicated AV lanes, and the solution procedure. Section 5.3 contains a description of a case study of the Amsterdam metropolitan region as well as numerical results and analysis. Finally, section 5.4 offers the concluding remarks.

5.2 Model Description

5.2.1 Problem Definition and Assumptions

For convenience, the notation used throughout this chapter is presented in Table 5.1.

Table 5.1 Notation

Notation	Explanation
Sets	
W	Set of origin-destination (OD) pairs *w*
R^w	Set of routes *r* between origin-destination (OD) pair *w*
M	Set of driving modes $m \in \{0,1\}$ (0: manual driving (MD), 1: automated driving (AD))
K	Set of user classes $k \in \{0,1\}$ (0: RV, 1:AV)
S	Set of subnetworks $s \in \{0,1,2,3\}$ (1: AV-ready link, 2: dedicated link, 3: dedicated lane)
A^s	Set of links *a* that belongs to the subnetwork *s*
A	Set of all original links *a* in the network; $A = \{A^0 \cup A^1 \cup A^2 \cup A^3\}$
Parameters	
μ^k	Multinomial logit route choice parameter for class *k*
γ_m	PCE value of driving in mode *m*
η_m	Value of travel time (VoTT) in mode *m*
$\bar{c}_{m,a}$	Fixed driving cost of mode *m* on link *a*
t_a^0	Free flow travel time on link *a* (identical for both modes)
Λ_a	Original (fixed) capacity of link *a*
Δ_a^s	Extra capacity proportion (capacity gain) of subnetwork *s*
l_a	Original number of lanes of link *a*
$\delta_{r,a}^{w,k}$	Assignment map: 1 if route *r* between OD pair *w* for class *k* includes link *a*, 0 otherwise
$D^{w,k}$	Demand of class *k* and OD pair *w*
σ	Parameter converting peak hour travel cost to a yearly basis
κ_a^s	Adjustment cost of link *a* for subnetwork *s*
π	Discount rate
t	Length of the planning horizon
Variables	
$C_r^{w,k}$	(Route-based) travel cost of route *r* between OD pair *w* for class *k*
$F_r^{w,k}$	(Route-based) flow of route *r* between OD pair *w* for class *k*
$c_{m,a}$	(Link-based) travel cost of mode *m* on link *a*
$f_{m,a}^k$	(Link-based) flow of class *k* in mode *m* on link *a*
$q_{m,a}$	Total flow (PCE-equivalent) in mode *m* on link *a*
$t_{m,a}$	Link travel time in mode *m* on link *a*
$\Lambda_{m,a}$	Capacity of link *a* for mode *m*
TTC	Total system travel cost
TTT	Total system travel time
TTD	Total system travel distance
TAC	Total adjustment cost
X_a	Binary variable: 1 if link *a* is upgraded for the AV-ready subnetwork ($s=1$), 0 otherwise
Y_a	Binary variable: 1 if link *a* is turned to a dedicated AV link ($s=2$), 0 otherwise
Z_a	Binary variable: 1 if a dedicated AV lane is included in link *a* ($s=3$), 0 otherwise
I_a	Integer variable: number of dedicated AV lanes on link *a*; $I_a \le l_a - 1$

Consider a road network represented by a directed graph $G(N, A)$ where N is the set of nodes and A is the set of directed arcs (links). There is at least one path (route) within the graph G that connects each origin-destination (OD) pair w. The set of links A^1 represents the AV-ready subnetwork, the set of links A^2 represents the subnetwork comprised of dedicated AV links, and the set of links A^3 represents the subnetwork with dedicated AV lanes. Binary variables $X_a, Y_a, Z_a \in \{0,1\}$ denote these decisions (i.e., $X_a = 1; \forall a \in A^1$, $Y_a = 1; \forall a \in A^2$, $Z_a = 1; \forall a \in A^3$). The number of dedicated AV lanes on link a is denoted by I_a. Each link can only be selected for one subnetwork (i.e., $X_a + Y_a + Z_a \leq 1; \forall a \in A$). The remaining links in the network are represented by A^0 (i.e., $X_a, Y_a, Z_a, I_a = 0; \forall a \in A^0$). The set $A_1 = \{A^1 \cup A^2 \cup A^3\}$ represents the set of links designated to ADSs to form a graph $G_1(N_1, A_1)$, which will be referred to as automated driving (AD) subnetwork throughout this chapter. Within G_1, level-3 and level-4 ADS-equipped vehicles (which will be referred to as AVs for the remainder of this chapter) activate their ADS, and outside of it, they drive manually. Vehicles with automation functions of levels 0-2 (which are referred to as RVs in this chapter) are not allowed on dedicated AV links and lanes, but they are allowed on AV-ready links since these links are designated to mixed traffic. It is essential to notice this distinction between AV-ready links designated to mixed traffic and links as well as lanes dedicated to AVs. On links where a dedicated lane is available, the use of that lane is assumed mandatory for AVs. The set of AD links represented by A_1 will be upgraded with roadside units and digital maps as well as clear and harmonized road signs and markings. This ensures safe operation of ADS within an extended ODD and allows AVs to keep shorter driving gaps while using their ADS, which improves traffic efficiency.

5.2.2 Multi-Class Network Equilibrium Model

Two classes of vehicles, namely RVs (class 0) and AVs (class 1) are considered here as well as two driving modes, namely manual driving (MD) denoted as mode 0 and AD denoted as mode 1. Using modes and classes is necessary since the class AV has both MD mode and AD mode available. RVs always drive in MD mode. AVs use the mode AD within G_1 and the mode MD within G_0. The travelers' route choice is captured via a multi-user class (MUC) stochastic user equilibrium (SUE), which is assumed to follow a multinomial logit model. Since one of the main impacts of the network configurations studied in this chapter is changes in link travel time, this is explicitly discussed in the next subsection. Then, the mathematical formulation and the equilibrium conditions are presented in the following parts.

Link travel time

Deployment of AV-ready subnetworks, dedicated AV links and dedicated AV lanes affects link travel times. The most common approach to derive the travel time function of links with AVs in mixed traffic for macroscopic static traffic assignment models is considering shorter driving gaps for AVs (Levin and Boyles, 2015; Liu and Song, 2019; Madadi et al., 2020; Noruzoliaee et al., 2018). In this approach, the Bureau of Public Roads (BPR) travel time function is used along with a total link flow equivalent based on passenger car equivalent (PCE) values, which is the sum of class flows (RVs and AVs) multiplied by a scaling parameter (i.e., the PCE value) representing their driving time headways.

Regarding dedicated links and lanes, the assumption of increased capacity has been the prevailing method. Ye and Wang (2018) assumed the capacity of a dedicated AV link to be triple the capacity of the regular link before conversion. Chen et al. (2016) assumed the capacity of a dedicated AV lane to be 2.5 times the per-lane capacity of a regular link. Both aforementioned approaches are utilized to derive the network equilibrium model presented below.

Mathematical formulation of the network equilibrium model

In order to represent new network topologies including $G_1(N_1, A_1)$ in mathematical terms, each link is virtually divided into two links, one per each mode of driving (mode 0 represents MD and mode 1 represents AD). Note that this does not change the actual performance of the network. On a link where a dedicated AV lane is present, $\Lambda_{0,a}$ represents the capacity of the regular part of the link with mode 0, and $\Lambda_{1,a}$ represents the capacity of the dedicated part of the link for mode 1. $f^0_{0,a}$ represents the flow of RVs (class 0) in MD mode (mode 0) on the regular part of the link, and $f^1_{1,a}$ represents the flow of AVs (class 1) in mode AD (mode 1) on the dedicated part of the link. The same holds for dedicated AV links but since RVs are not allowed in dedicated AV links, the capacity of these links for mode MD is set to 0, thereby excluding this part of the link from the network. For AV-ready links, the capacity for both parts of the link is the same since in this case, the systematic split between the two driving modes is artificial, and both driving modes should have the same travel time. Therefore, the flows on both parts of these links are aggregated in order to compute the travel time on these links. On regular links that do not belong to the AD subnetwork, the capacity for AD mode is set to 0 and both classes will have their flows only in MD mode since it is the only mode allowed on these links. This is demonstrated in Figure 5.1 for a simple 4-link network including a regular link, an AV-ready link, a dedicated AV link and a link with a dedicated AV lane.

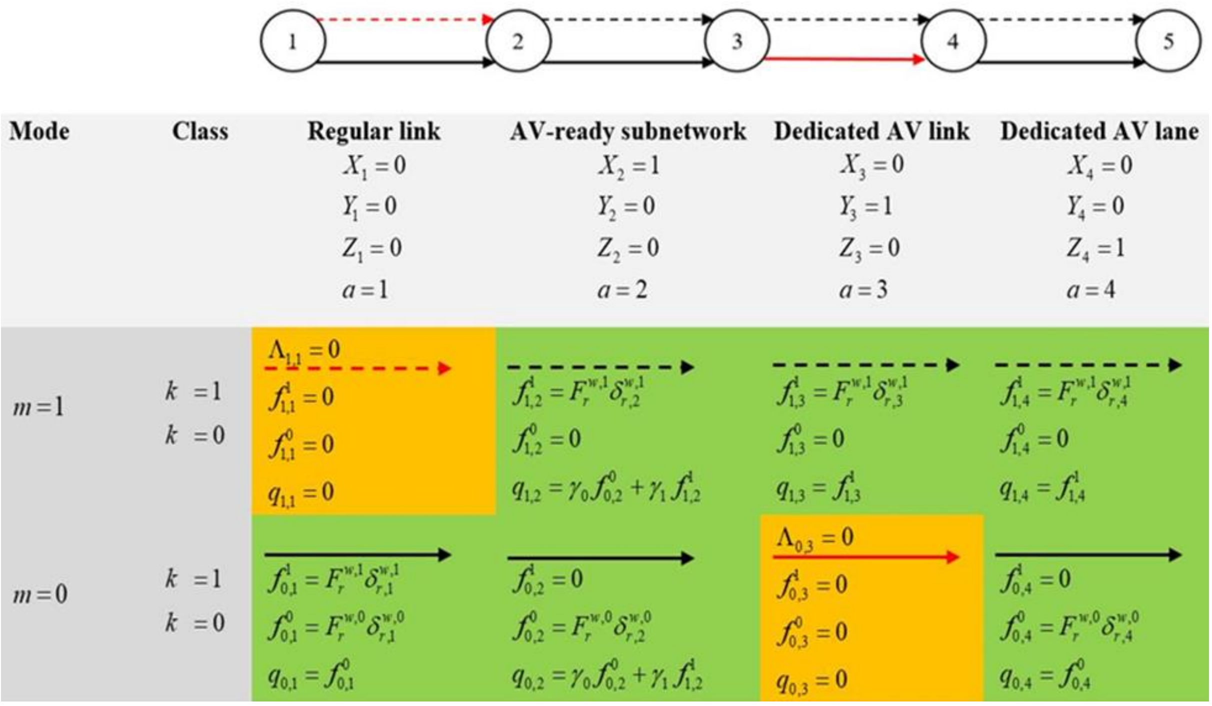


Figure 5.1 Summary of capacities and flows applied for each mode and class on each link type

The following set of equations operationalizes this concept in mathematical notations and represents the behavioral rules considered in our network equilibrium model (NEM) for any realization of $X_a, Y_a, Z_a \,\&\, I_a$ given $X_a + Y_a + Z_a \le 1, \forall a \in A$.

NEM:

$$\Lambda_{0,a} = \Lambda_a - Y_a \Lambda_a - Z_a (\frac{\Lambda_a}{l_a} I_a) + \varepsilon, \qquad \forall a \in A, \qquad 5.1$$

$$\Lambda_{1,a} = X_a \Delta^1 \Lambda_a + Y_a \Delta^2 \Lambda_a + Z_a \Delta^3 (\frac{\Lambda_a}{l_a} I_a) + \varepsilon, \qquad \forall a \in A, \qquad 5.2$$

$$q_{m,a} = (1 - X_a) \sum_{k \in K} f^k_{m,a} + (X_a) \sum_{k \in K} \sum_{j \in M} \gamma_j f^k_{j,a}, \qquad \forall a \in A, \forall m \in M, \qquad 5.3$$

$$t_{m,a} = t^0_a [1 + \alpha_a (\frac{q_{m,a}}{\Lambda_{m,a}})^{b_a}], \qquad \forall a \in A, \forall m \in M, \qquad 5.4$$

$$c_{m,a} = \bar{c}_{m,a} + \eta_m t_{m,a}, \qquad \forall a \in A, \forall m \in M, \qquad 5.5$$

$$\sum_{w \in W} \sum_{r \in R^w} F^{w,0}_r \delta^{w,0}_{r,a} = f^0_{0,a}, \qquad \forall a \in A, \qquad 5.6$$

$$\sum_{w \in W} \sum_{r \in R^w} F^{w,1}_r \delta^{w,1}_{r,a} = f^1_{1,a} (X_a + Y_a + Z_a) + f^1_{0,a} (1 - X_a + Y_a + Z_a), \quad \forall a \in A, \qquad 5.7$$

$$\sum_{w \in W} \sum_{r \in R^w} C^{w,0}_r \delta^{w,0}_{r,a} = c_{0,a}, \qquad \forall a \in A, \qquad 5.8$$

$$\sum_{w \in W} \sum_{r \in R^w} C^{w,1}_r \delta^{w,1}_{r,a} = c_{1,a} (X_a + Y_a + Z_a) + c_{0,a} (1 - X_a + Y_a + Z_a), \quad \forall a \in A, \qquad 5.9$$

$$\sum_{r \in R^w} F^{w,k}_r = D^{w,k}, \qquad \forall w \in W, \forall k \in K, \qquad 5.10$$

$$F^{w,k}_r \ge 0, \qquad \forall w \in W, \forall k \in K, \forall r \in R^w. \qquad 5.11$$

In this set of equations, *Eq.5.1* defines link capacities for mode 0 based on whether the link is assigned to G_1 or not. When $Y_a = 1$ (i.e., the link is selected as a dedicated AV link), the capacity for mode MD becomes 0. When $Z_a = 1$ (i.e., one or more lanes of the link are dedicated to AVs), based on the number of selected lanes a proportion of the link capacity will be subtracted from the link capacity for MD mode. In order to avoid undefined numbers in calculations when a capacity is set to 0, a very small value (ε) is added to all capacities. Note that this will force inadmissible flows be practically 0. *Eq.5.2* defines capacities for mode AD based on the type of subnetwork to which links belong. This is similar to *Eq.5.1* with the addition of multiplying capacities by Δ^s to increase them based on the subnetwork type. The value of Δ^1 is set to 1 to neutralize it but this parameter is included in the model for consistency. *Eq.5.3* shows how the PCE-equivalent of link flows is calculated to be used in the link travel time function for each mode. When $X_a = 1$ (i.e., the link becomes an AV-ready link), the total flow is a summation of AV flows in AD mode multiplied by PCE value of the AD mode (which is lower than 1 to represent shorter driving gaps in AD mode) and RV flows in MD mode multiplied by the PCE value of MD mode (which is 1). In this case, both parts of the link will have the same total flow, thereby the same travel time. When $X_a = 0$ (i.e., the link is not an AV-ready link), the total flow is a simple summation of flows of both classes; however, this leads to different total flows for each mode on each link type. On regular links, both classes only have non-zero flows in MD mode, so the total flow for MD mode will be the summation of class flows and the total flow in AD mode will be 0. On dedicated AV links, the only non-zero flow is the flow of AVs in AD mode. On links with dedicated AV lanes, RVs will have a non-zero flow in MD mode and AVs will have a non-zero flow in AD mode. This leads to the intended total flow on each link type for each mode. *Eq.5.4* shows how link travel time for each mode is calculated using the BPR function. Note that different combinations of capacities and total flows based on the subnetwork type and driving mode will be applied in *Eq.5.4*, which results in different travel times for MD and AD modes on dedicated AV lanes and links, but identical travel times for both modes in regular links and AV-ready links. Figure 5.1 summarizes capacities and flows applied for each mode and class on each link type. For computational reasons, infeasible modes (e.g., flow of RVs in MD mode on dedicated links to AVs) are modeled with very large travel times via very small capacity values. *Eq.5.5* is related to link travel cost for each mode, which is based on the link travel time for each mode multiplied by the value of travel time (VoTT) of that mode along with the fixed driving cost of the link in that mode. VoTT and fixed driving costs are assumed to be lower for AD mode. *Eq.5.6* and *Eq.5.7* denote the correspondence between link flows and route flows using a link-route incidence matrix (assignment map) for class RVs and class AVs separately. They ensure that the route flow of each class is assigned to the appropriate mode on each link based on the subnetwork type. On regular links (MD network), AV route flow is assigned to MD mode, and on the rest of the links (AD subnetwork), it is assigned to AD mode. *Eq.5.8* and *Eq.5.9* carry out the same task for the correspondence between link and route costs. *Eq.5.10* guarantees the travel demand in each class for each OD pair is satisfied by equating the sum of all route flows for each OD pair and each class to their relevant demand. *Eq.5.11* prohibits negative flows.

Equilibrium conditions

The equilibrium condition of the NEM based upon Brouwer's fixed-point theorem (Cantarella, 1997; Daganzo, 1983; Florian and Hearn, 1995) is to find an $\mathbf{F}^*$ that solves:

$$\mathbf{F} - \mathbf{P}(\mathbf{F})\mathbf{D} = 0, \qquad \forall \mathbf{F} = [F_r^{w,k}] \in \mathbf{\Pi}, \tag{5.12}$$

where $\mathbf{F}$ is the vector of route flow $F_r^{w,k}$. The set $\mathbf{\Pi}$ represents the set of flows $F_r^{w,k}$ that satisfy *Eqs.5.1-5.11*, i.e., $\mathbf{\Pi} = \{F_r^{w,k} | 5.1-5.11\}$ defines the feasible region of the NEM. Vector $\mathbf{D}$ is the vector of travel demand $D^{w,k}$, and $\mathbf{P}$ is the vector of route choice probability $P_{r|w,k}$ that represents the proportion of the travelers of class *k* between OD-pair *w* who take the route *r*. Solving *Eq.5.12* yields $P^*_{r|w,k} = \frac{F_r^{*w,k}}{D^{w,k}}$, which for the case of the multi-class multinomial logit SUE leads to the following well-known route choice probabilities (Oppenheim, 1995):

$$\frac{F_r^{*w,k}}{D^{w,k}} = P^*_{r|w,k} = \frac{\exp(\mu^k C_r^{*w,k})}{\sum_{r \in R^w} \exp(\mu^k C_r^{*w,k})}, \qquad \forall w \in W, \forall k \in K. \tag{5.13}$$

5.2.3 Deployment of Subnetworks for AVs

In transport literature, strategic decisions regarding road networks are commonly modeled as bi-level network design problems (NDPs) (Farahani et al., 2013; Yang and Bell, 1998). Studies that propose optimal network design concepts for AVs have followed this framework as well (Chen et al., 2016; Madadi et al., 2020; Ye and Wang, 2018). Therefore, we model the optimal deployment of AV-ready subnetwork, dedicated AV links and dedicated AV lanes as a bi-level NDP. The upper level problem investigates where AV-ready links, dedicated AV links and dedicated AV lanes should be deployed to maximize their societal benefits, while the lower level problem is the multi-class network equilibrium model based on the fixed-point formulation (NEM-FP) defined by *Eqs.5.1-5.12*, which captures the travelers' response to network topologies.

As explained earlier, the upper level decision variables are as follows. X_a is a binary decision variable taking the value 1 when link a is selected as an AV-ready link. Y_a is a binary decision variable taking the value of 1 when link a is selected as a dedicated AV link. I_a is an integer decision variable denoting the number of dedicated AV lanes on link a. Z_a is an auxiliary binary variable taking the value of 1 when link a contains one or more dedicated AV lanes. Mathematical formulation of the upper level problem is given below.

$$\textbf{Min} \quad Z_U = \sum_t \frac{\sigma TTC}{(1+\pi)^t} + TAC, \tag{5.14}$$

s.t.

$$(5.1)-(5.12),$$

$$TTC = \sum_{k\in K}\sum_{w\in W}\sum_{r\in R^w} F_r^{w,k} C_r^{w,k}, \tag{5.15}$$

$$TAC = \sum_{a\in A} X_a\kappa_a^1 + Y_a\kappa_a^2 + Z_a I_a \kappa_a^3, \tag{5.16}$$

$$I_a \le l_a - 1, \qquad \forall a\in A, \tag{5.17}$$

$$X_a + Y_a + Z_a \le 1, \qquad \forall a \in A, \tag{5.18}$$

$$(1-Z_a)I_a = 0, \qquad \forall a\in A, \tag{5.19}$$

$$Z_a \le I_a, \qquad \forall a \in A, \tag{5.20}$$

$$\left| p^{n,n'}_{G_1(N_1,A_1)} \right| \ge 1, \qquad \forall n,n' \in N_1, \tag{5.21}$$

$$X_a, Y_a, Z_a \in \{0,1\}, I_a \in \mathbb{Z}^{\ge 0}, \qquad \forall a\in A. \tag{5.22}$$

In this formulation, *Eq.5.14* represents the objective function (OF), which minimizes the sum of total adjustment cost (TAC) and the net present value of the total travel cost (TTC), given the entire planning horizon. One morning peak hour is considered in this study and TTC of this period is converted to a yearly basis with a parameter σ. This yearly value is discounted with a discount rate of π for t years, which is the useful lifetime of the infrastructure adjustments. The effects of the construction period itself are not considered here. The first set of constraints for the upper level problem is the lower level problem defined by *Eqs.5.1-12*, which means the upper level should be evaluated when the lower level is at equilibrium. *Eq.5.15* shows how TTC is calculated based on equilibrium flows and costs obtained from the lower level problem. TAC in *Eq.5.16* is the summation of link adjustment costs for all subnetworks. *Eq.5.17* ensures the number of dedicated AV lanes are less than the number of lanes on each link. *Eq.5.18* guarantees each link can be selected only for one subnetwork. *Eq.5.18*, *Eq.5.19* and *Eq.5.20* force Z_a to take the value of 1 when I_a is greater than 0, and 0 otherwise. *Eq.5.22* denotes admissible values for the upper level decision variables. Finally, *Eq.5.21* expresses a connectivity constraint, imposed in order to find coherent networks. It signifies that for each two nodes n and n' within the graph $G_1(N_1, A_1)$ where the AD mode is allowed, there must exist at least one undirected path connecting the two nodes. The purpose of this constraint is preventing the solution method from finding networks with disconnected components. Since the AD mode is only allowed within G_1, having networks with separate components (i.e., disconnected networks) leads to switching frequently between MD and AD mode, which is best to be avoided. This imposes extra requirements on the solution method, which are discussed below.

5.2.4 Solution Methods

Regarding the lower level problem, Daganzo (1983) has shown the existence and uniqueness of a solution for the fixed-point formulation of the multi-mode multi-class SUE problem with asymmetric link costs in terms of the PCE-equivalent of flows. This holds when the feasible region of a NEM is compact and convex, and all the functions are continuous, which is the case for the NEM-FP defined by *Eqs.5.1-12*. There are numerous solution methods available for solving FP problems. For a review, the reader is referred to (Florian and Hearn, 1995; Harker and Pang, 1990). In this study, we use an efficient sequential linear approximation algorithm with step sizes according to the method of successive averages (MSA) introduced in (Wu et al., 2006) to solve the lower level equilibrium problem.

The upper level problem on the other hand, is more challenging to solve. Bi-level NDPs with discrete upper level decision variables and multi-class SUE lower level problems are among the most challenging problems in transport literature and are often solved using heuristic procedures (Farahani et al., 2013; Yang and Bell, 1998). Solving such problems, particularly for large-scale networks, calls for efficient solution procedures. Moreover, the connectivity requirement denoted by *Eq.5.21* makes the problem even more difficult to solve, since existing methods for solving NDPs do not produce connected graphs. Therefore, we have developed a heuristic solution procedure based on an evolutionary algorithm in this study to solve the upper level problem. However, before introducing the solution procedure, we provide five definitions that are necessary in order to comprehend how the solution method operates and deals with the connectivity constraint. These definitions are illustrated in Figure 5.2.

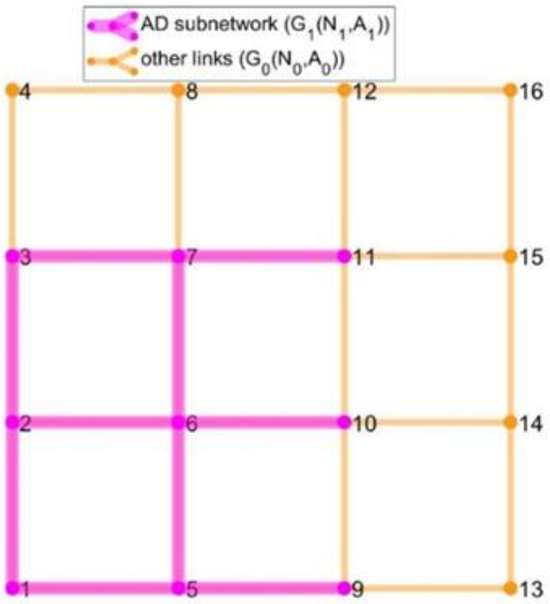


Figure 5.2 An example AD subnetwork

Note: Nodes 3-7-9-10-11 are boundary nodes, links (5,9)-(6,10)-(7,11) are inner boundary links, and links (9,13)-(10-14)-(11,15)-(11,12)-(7,8)-(3-4)-(9,10)-(10,11) are outer boundary links.

Definition 5.1. Any node n incident to at least one link included in an AD subnetwork represented by $G_1(N_1, A_1)$ is included in that AD subnetwork (i.e., $n \in N_1$).

Definition 5.2. The degree of a node belonging to each graph (e.g., G, G_0, G_1) is the number of links incident to that node within the graph in which that node is included.

Definition 5.3. A boundary node $n^* \in N_1$ for an AD subnetwork represented by a directed graph $G_1(N_1, A_1)$ is a node included in G_1 with a degree less than the degree of the corresponding node n in the underlying graph $G(N, A)$ (i.e., the graph representing the original road network).

Definition 5.4. An outer boundary link $a^* \in A_0$ for an AD subnetwork is a link incident to a boundary node $n^* \in N_1$ of the AD subnetwork but not included in the graph G_1 representing the AD subnetwork.

Definition 5.5. An inner boundary link $a^* \in A_1$ for an AD subnetwork is a link included in the graph G_1 representing the AD subnetwork and incident only to boundary nodes with the degree of 1.

Note that boundary nodes of an AD subnetwork and the inner boundary links are included in the subnetwork while the corresponding outer boundary links are not included in that subnetwork.

Evolutionary heuristic solution procedure

The evolutionary heuristic solution introduced here is inspired by evolutionary metaheuristics and takes advantage of this specific problem's structure to efficiently generate connected subnetworks and solve the problem. Using extensive computational experiments, it is shown in chapter 3 that an evolutionary local search algorithm tailored to the problem can efficiently find coherent and connected AV-ready subnetworks. Therefore, in this study, we develop another evolutionary algorithm that modifies and extends the procedure of mentioned algorithm to find coherent subnetworks including AV-ready links, dedicated AV links and dedicated AV lanes in an efficient manner.

The purpose of the solution procedure is to determine the values of $X_a, Y_a, Z_a \,\&\, I_a$ (which uniquely define a graph $G_1(N_1, A_1)$) in order to minimize *Eq.5.14* and satisfy *Eqs.5.15-22*. Since every AD subnetwork represented by $G_1(N_1, A_1)$ is built from scratch, we can start with a simple connected graph as an initial feasible solution, and iteratively modify this graph to improve the value of the OF in *Eq.5.14* using operations that preserve the connectivity requirement of *Eq.5.21* until no further improvement can be obtained by modifying the graph. To avoid termination of the procedure in local minima, a population of solutions is evolved through several generations to acquire a good solution at the end. The steps of the solution procedure are summarized in Table 5.2. This is followed by the description of the operations used in the procedure.

Table 5.2 Solution procedure

Solution steps	
1	Initialize population
2	Measure fitness of all individuals
3	**For** each generation **j**
4	Assign an operation to each individual **i** based on parameters (possible operations are extension, reduction & merging)
5	**For** each individual **i**
6	Perform the operation assigned to the individual **i**
7	Measure fitness of the individual **i**
8	**End**
9	Perform regeneration
10	Increment generation number
11	**If** stopping criteria met: **Terminate**
12	**End**

Evolutionary heuristic operations

Initialization

During the initialization process, a population of individual solutions is generated. Each individual solution includes one link selected via a roulette wheel prioritizing links with higher capacity. A subnetwork type *s* is randomly assigned to each link. With the addition of each link, the nodes incident to that link are added to the graph G_1 as well. Note that any graph including one link and its incident nodes satisfies *Eq.5.21*. This process will provide a pool of diverse and connected solutions to start the algorithm. The size of the population is an algorithm parameter.

Fitness evaluation

Fitness evaluation is based on the value of the upper level OF (Z_U) in *Eq.5.14*. This value is obtained by assigning a value to each (upper level) decision variable and solving the lower level problem (NEM-FP).

Extension operation

In this operation, first a subnetwork *s* is randomly selected for extension (e.g., dedicated AV link). Next, the outer boundary links of the graph (i.e., the set of links $a^* \in A_0$) are specified (see Definition 5.4). Then a sample of them is selected based on a roulette wheel prioritizing higher capacities and added to the graph. When the subnetwork type is dedicated AV lanes (i.e., $s = 3$), the number of lanes I_a is assigned at random. All nodes incident to any link added to G_1 are added to it as well. The sample size is an algorithm parameter. When the input of this operation meets *Eq.5.21*, the output is guaranteed to meet this constraint as well.

Reduction operation

In this operation, first the inner boundary links of the graph (i.e., the set of links $a^* \in A_1$) are determined (see Definition 5.5). Then a sample of them is selected based on a roulette wheel prioritizing links with lower capacity and removed from the graph. The corresponding nodes that are not incident to any other link in G_1 are eliminated as well. This sample size is also an algorithm parameter. This operation preserves the connectivity constraint of *Eq.5.21* as well, since it only removes links from the boundaries of a graph.

Merging operation

Unlike extension and reduction operations, which require an individual solution, merging operation first selects a merging partner from among the population based on a roulette wheel prioritizing fitness value. Then the merged graph is the union of the links and nodes in both graphs (merging parents) with their subnetwork type *s* unchanged. When there is a conflict (e.g., a link is included in both graphs and is a dedicated AV link in one graph but an AV-ready link in the other), it is resolved by inheriting from the fitter parent. That is, the values of $X_a, Y_a, Z_a \,\&\, I_a$ for a conflicting link *a* in a merging offspring is equal to those values of the merging parent with better fitness value.

Regeneration

At the end of each generation, the fittest individuals among the parents (from the previous generation) and the offspring (resulting from the operations in current generation) survive to re-populate the next generation.

5.3 Case Study: Amsterdam Metropolitan Region

5.3.1 Description

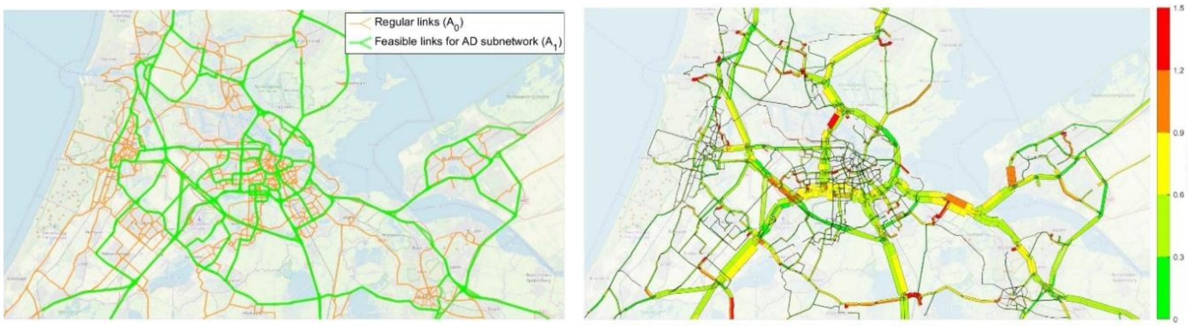
The concept of optimal AD subnetworks with AV-ready subnetwork, dedicated AV links and dedicated AV lanes is demonstrated in this section using a case study of the road network of the Amsterdam metropolitan region. The network and demand data of the VENOM model (Kieft, 2013) are used here. They are based on the real network and demand patterns of the Amsterdam metropolitan region. The original network includes 52,812 directed links, 19,734 nodes and 3,722 transportation zones. In this case study, the transportation zones are aggregated to 102 zones (10,124 OD pairs) and all the links that are not used in the traffic assignment of the base case scenario (in the morning peak) are eliminated. Therefore, the considered study area contains 12,781 links and 6,642 OD pairs. The network and the equilibrium flows in the base case are shown in Figure 5.3.a. The calibrated demand matrices in VENOM model for the year 2004 are used to extract the demand for cars. Demand for AVs is considered via four scenarios with 10%, 30%, 50%, and 90% MPR of AVs. It is assumed that the ratio of AVs to RVs available to travelers is the same as the AV market penetration rate and that the AVs available to travelers are homogeneously distributed in all zones. All demand from, to and through the study area is considered in the traffic assignment to calculate the equilibrium flows (NEM-FP). Network performance indicators are reported for the OD pairs inside the study area. One average morning peak hour is modeled in this study, which is assumed to account for 10% of the daily traffic and the value of TTC obtained is converted to a yearly basis using $\sigma = 10 \times 30 \times 12$ for calculation of the OF value in (14). The value of discount rate π is 4%, which is the common value used for public investments in the Netherlands, and the length of the planning horizon t is assumed to be 10 years. All OF values reported in this chapter are divided by 10^6 for convenience. The base case traffic pattern is shown in Figure 5.3.b.

The capacity gain from converting a regular link to a dedicated AV link is assumed to be 100% $(\Delta_a^2 = 2)$, and the capacity gain of converting a lane to a dedicated AV lane is assumed to be 50% $(\Delta_a^3 = 1.5)$. Studies that proposed deployment of networks of dedicated AV links (Ye and Wang, 2018) and lanes (Chen et al., 2016) have used the values 3 and 2.5 for Δ_a^2 and Δ_a^3, respectively. However, microsimulation studies of dedicated AV lanes suggest that these values are overly optimistic, see for instance (Mahmassani, 2016; van Arem et al., 2006). Therefore, we have selected the input values for capacity gains according to the results of mentioned microsimulation studies and performed a sensitivity analysis to investigate the effect of deviations from these values on model results. Converting a link to an AV-ready link for mixed traffic does not lead to any explicit capacity increase $(\Delta_a^1 = 1)$. Instead, the PCE value of driving in AD mode γ_1 on AV-ready links is assumed to be 0.9 for AV MPRs below 40% and 0.8 for above 40% AV MPR. The PCE value of driving in MD mode γ_0 is set to 1. VoTT for MD mode η_0 is 9 €/h and VoTT for AD mode η_1 is 80% of this value (Correia et al., 2019). Fixed driving cost per kilometer for MD mode $\overline{c}_{0,a}$ is 0.19 €/km and for the AD mode $\overline{c}_{1,a}$ is 80% of this value for below 40% AV MPR and 60% of this value for above 40% AV MPR (Shida et al., 2010; Shladover et al., 2015). These values are similar to those used in chapter 3 where a sensitivity analysis is also provided, which shows that minor deviations from these values will not significantly impact the results. The logit model route choice parameter μ^k value used for both classes is -0.5 (Brands, 2015).

Regarding the parameters of the upper level solution algorithm, the population size is 600, maximum number of generations is 500, the sample size for extension and reduction operations is 0.1% of the number of decision variables, and the population fractions assigned to operations extension, reduction and merging are 0.2, 0.2 and 0.6, respectively. The number of iterations used for the convergence of the lower level solution algorithm is 50.

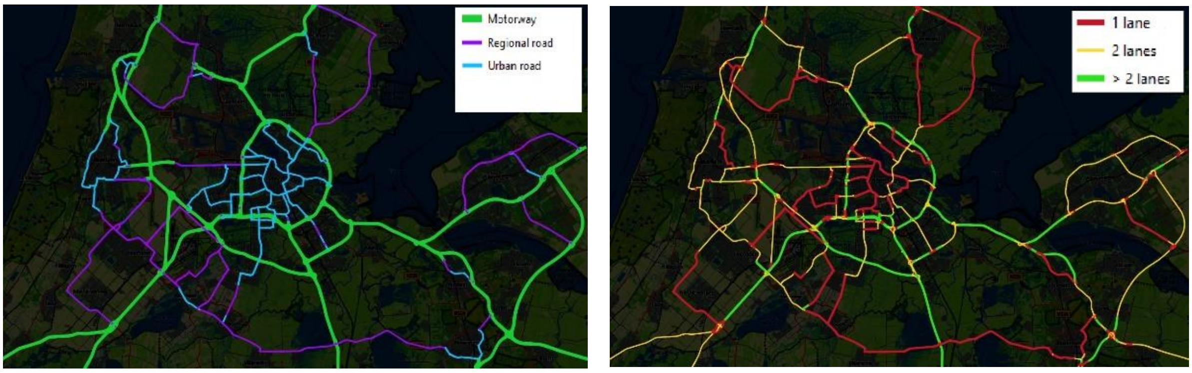
Motorways, major regional roads and main urban roads are considered feasible links for AD subnetwork in this case study. The purpose of this selection is to avoid complex interactions including AVs and vulnerable road users that might compromise safe operation of AVs in AD mode. This is consistent with the findings of AV accident reports, which indicate that around 90% of the accidents involving AVs have occurred at urban intersections (Favaro et al., 2017). Moreover, about half of the AV disengagements in California have happened in small urban streets (Favarò et al., 2018), which are not considered as feasible links in this study. A total of 3,402 links is considered feasible in this case study. They are also depicted in Figure 5.3.a.

As for the adjustment costs κ_a^s, different values per kilometer are used for each road type and each subnetwork type. The adjustment cost value for AV-ready links κ_a^1 is assumed to be 50,000 €/km for motorways, 75,000 for regional roads and 100,000 for urban roads. The reason for the increase in adjustment costs from motorways to regional roads and urban roads is that motorways generally have higher quality standards and require minimal interaction with other road users due to segregated traffic and off-grade intersections. Regional roads can have lower standards and require more interaction with other road users, thereby requiring higher adjustment costs to make them suitable for AVs. Finally, urban roads are the most challenging ones for AVs due to complex interactions between AVs and other road users. Therefore, they have the highest adjustment costs. Figure 5.4 depicts road types in Amsterdam case study. For dedicated AV links, κ_a^2 is twice the value of κ_a^1 for each road type. The adjustment cost values for dedicated AV lanes κ_a^3 are 50% higher than the values of κ_a^1 for each road type. In real applications, these values can be estimated in more detail by experts.



(a) Feasible links for AD Subnetwork (b) Base case traffic patterns

Figure 5.3 Feasible links for AD subnetwork and base case traffic patterns in Amsterdam case study



(a) Road types (b) Number of lanes

Figure 5.4 Road types and number of lanes for feasible AD subnetwork links in Amsterdam case study

5.3.2 Results and Analysis

In this section, we first discuss the consistency of AD subnetworks generated by the evolutionary heuristic solution method proposed in this study for real road networks. Then, we discuss the impacts of simultaneous deployment of AV-ready subnetworks, dedicated AV links and dedicated AV lanes on road network performance, the usage of different road types in the network and the usage of AD mode on different road types and subnetworks. In addition, we evaluate the distribution of TTC, TTT and TTD on each subnetwork. Computation times of the experiments are discussed at the end of this section.

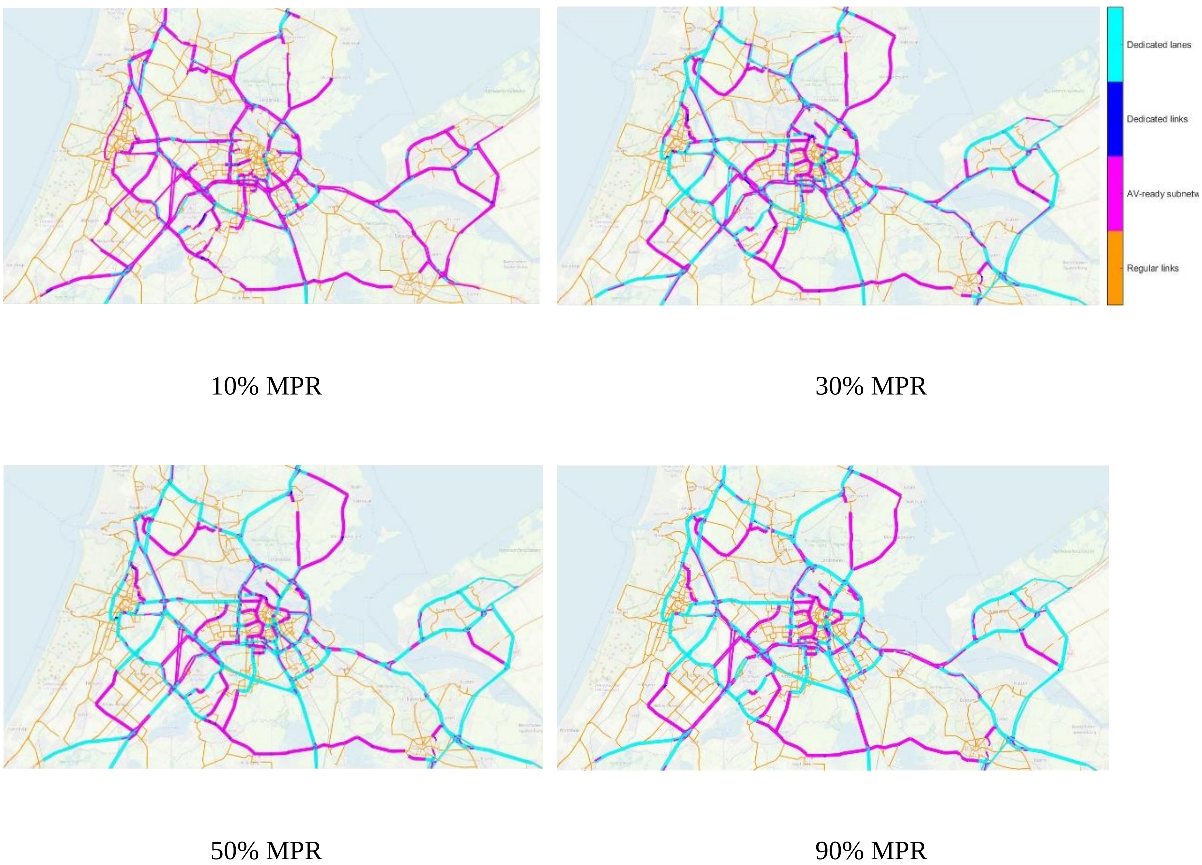
Consistency of AD subnetworks in real road networks

Figure 5.5 depicts the graphs obtained by the heuristic solution method introduced earlier as (near) optimal AD subnetworks in all four scenarios. In this section, we discuss three characteristics related to the logical consistency of AD subnetworks in real road networks, which are not observed in theoretical networks that are commonly used as case studies in scientific literature. These characteristics are connectivity, sub-optimality, and continuity in network hierarchy. We briefly discuss them in the following paragraphs.

Connectivity of a graph representing an AD subnetwork (*Eq.5.21*) is an undeniable requirement for efficiency of AD mode. However, common solution methods for NDP are not equipped to cope with this constraint effectively. It is shown in chapter 3 that without this constraint, incoherent subnetworks with numerous disconnected islands of links can be obtained as (near) optimal solutions. Tailored algorithms such as the evolutionary algorithm developed in this study can satisfy this constraint in an efficient manner, yet they do not necessarily find the global optimum of the problem.

Sub-optimality of the results is another characteristic that needs to be considered. The solutions provided by the heuristic algorithm introduced in this study are not necessarily optimal. The algorithm effectively copes with the connectivity constraint and generates plausible solutions with reasonable computation times, yet there is no guarantee for optimality of the solutions. This is caused by the well-known complexity of bi-level NDPs with discrete decision variables along with the large problem size due to the use of a realistic road network that makes the problem very challenging to solve for the global optimum. Therefore, this might introduce some degree of inconsistency in the resulting graphs as well.

On the other hand, when considering three types of subnetworks, it seems plausible to expect long stretches of roads selected for each single subnetwork, which are connected to each other at junctions. However, as demonstrated in Figure 5.4, road types (and consequently the network hierarchy) are not fully continuous in the network of Amsterdam; there are long stretches of roads that are classified as urban roads in some parts and as regional roads in some other parts. In addition, there are many stretches of roads in this network with fluctuating number of lanes. Perhaps these phenomena can be observed in all real road networks. Since each subnetwork is more suitable for certain types of roads with certain number of lanes, the discontinuity in road types and lane numbers causes discontinuity in subnetwork types. For instance, single-lane roads, which constitute a large proportion of roads in this network, as well as motorway on-ramps, off-ramps and main regional roads with occasional single-lane stretches in between are not feasible for dedicated AV lanes. These types of links are suitable for mixed traffic at any AV MPR during the transition period. This also explains the gaps with AV-ready links in between stretches of dedicated AV lanes. Comparing the (near) optimal graphs obtained by the solution method (Figure 5.5) with the map of the Amsterdam network including the number of lanes and road types (Figure 5.4) corroborates this notion.



10% MPR 30% MPR

50% MPR 90% MPR

Figure 5.5 AD subnetworks obtained by the evolutionary heuristic solution for the Amsterdam case study

Network performance

Table 5.3 summarizes the performance of the road network in presence of AV-ready subnetworks, dedicated AV links and dedicated AV lanes using three key performance indicators (KPIs), namely, TTC, TTT and TTD. Furthermore, three criteria for assessing the trade-off between costs and benefits of deploying AD subnetworks are reported. These criteria are OF, TAC and total discounted travel cost saving (TDTCS). TDTCS represents the net present value of TTC savings (compared to the base case) in each scenario over the entire planning horizon.

The values of TDTCS in all scenarios are substantially higher than TAC values (e.g., for 50% MPR, the TDTCS to TAC ratio is approximately 20). This means the benefits significantly outweigh the infrastructure adjustment costs. In general, the results show a steady decrease in TTC with the increase in MPR of AVs. The decrease in TTC is rather linear until 30% AV MPR but there is a sharp acceleration in the decrease starting 50% AV MPR. Moreover, the AV class observes a larger proportion of TTC benefits compared to the RV class in all scenarios; nonetheless, RVs are better off in terms of TTC compared to the base case in all scenarios. This indicates a major improvement in overall network performance with combined deployment of AV-ready subnetworks, dedicated AV links and dedicated AV lanes.

TTT values are also less than the base case values for all classes in all scenarios with the exception of 30% MPR where RVs observe a minor increase in TTT (in proportion to their market size). However, the general magnitude of the decrease in TTT is smaller compared to TTC. One interpretation for this phenomenon is that in addition to congestion relief and more traffic efficiency due to rerouting, a proportion of the gains in TTC is due to lower VoTT values for AD mode.

TTD for all vehicles is slightly decreasing until 30% MPR, and then it starts to increase. Nevertheless, RVs and AVs exhibit different trajectories in this regard. For RVs, in all scenarios, TTD values (proportional to their market size) are lower compared to the base case, while the values of TTD for AVs are slightly higher in all scenarios compared to the base case. This is mostly due to rerouting of AVs towards parts of the network with dedicated infrastructure. It is noteworthy that both classes have lower TTT in most scenarios despite the overall higher TTD.

The OF values steadily decrease with higher MPR of AVs. This is because larger reductions in TTC values are possible with higher AV MPR, while total adjustment cost (TAC) values stay rather steady after 30% MPR scenario. A remarkable observation is that the TAC value of the 90% MPR scenario is marginally lower compared to 50% MPR, even though OF and TTC values are lower. This means higher traffic efficiency benefits can become available with high MPR of AVs even with less investment on infrastructure.

Distribution of traffic among road types

The next factor we analyze in this chapter is the impact of AD subnetwork deployment on the usage of different road types. We focus on 50% MPR scenario to show the usage of different road types and compare statistics of RVs and AVs. Table 5.4 summarizes the distribution of network performance indicators on different road types indexed according to the base case values for the 50% MPR scenario. Therefore, comparing the values of RV and AV rows with

50% (and similarly the values of all cars rows with 100%) shows the percentage difference in each performance indicator on each road type compared to the base case scenario.

Table 5.4 shows that the TTD on motorways for the 50% MPR scenario is higher compared to the base case while regional roads and urban roads are used less intensively in this scenario compared to the base case. It is also evident that RVs have lower usages of all road types (i.e., they opt for shorter routes) compared to the base case scenario whereas AVs travel more using motorways and regional roads and less using urban roads (i.e., they opt for longer routes). This can logically be explained by the fact that most motorways and regional roads are part of the AD subnetwork. TTC and TTT values are lower in all road types for both classes in spite of higher TTD, with motorways and regional roads showing the highest impacts for AVs.

Since motorways generally are the safest roads, the shift of traffic towards motorways can provide traffic safety benefits. However, AD subnetworks attract AVs towards roads that are higher in network hierarchy and RVs towards lower ones, which leads to uneven distribution of traffic safety benefits between AV and RV classes.

Table 5.3 Key performance indicators for network of Amsterdam

0% MPR (base case)						
Class			**TTC (€)**	**TTT (h)**	**TTD (km)**	
All cars			611,704	35,656	1,530,514	
10% MPR						
Class	**OF**	**TDTCS (€)**	**TTC (€)**	**TTT (h)**	**TTD (km)**	**TAC (€)**
RV			89.52%	89.22%	89.85%	
AV			8.87%	9.91%	10.03%	
All cars	17,647	323,021,014	98.39%	99.13%	99.88%	73,918,289
30% MPR						
Class	**OF**	**TDTCS (€)**	**TTC (€)**	**TTT (h)**	**TTD (km)**	**TAC (€)**
RV			69.93%	70.12%	69.72%	
AV			26.25%	29.30%	30.12%	
All cars	17,288	766,422,530	96.18%	99.42%	99.84%	108,941,637
50% MPR						
Class	**OF**	**TDTCS (€)**	**TTC (€)**	**TTT (h)**	**TTD (km)**	**TAC (€)**
RV			49.14%	48.54%	49.79%	
AV			40.02%	48.23%	50.30%	
All cars	16,037	2,174,874,404	89.16%	96.77%	100.09%	111,967,171
90% MPR						
Class	**OF**	**TDTCS (€)**	**TTC (€)**	**TTT (h)**	**TTD (km)**	**TAC (€)**
RV			9.70%	9.42%	10.00%	
AV			73.92%	91.21%	90.45%	
All cars	15,045	3,288,394,048	83.61%	100.63%	100.45%	111,280,505

Note: Indexing is based on the values of the base case scenario

Table 5.4 Distribution of network performance indicators on different road types for 50% MPR

	Class	Motorway	Regional road	Urban road
TTC (€)	RV	48.98%	48.52%	49.31%
	AV	33.62%	37.58%	45.01%
	All cars	82.60%	86.10%	94.32%
	Base case	227,514	116,651	186,053
TTT (h)	RV	47.60%	47.12%	48.98%
	AV	48.16%	45.59%	48.51%
	All cars	95.76%	92.70%	97.48%
	Base case	8,821	6,078	13,314
TTD (km)	RV	49.71%	49.76%	49.92%
	AV	50.65%	50.20%	49.68%
	All cars	100.36%	99.96%	99.60%
	Base case	779,589	326,060	348,571

Note: Indexing is based on the values of the base case scenario

Distribution of ADS usage among different subnetworks

Table 5.5 summarizes the statistics related to performance indicators for the AV class and AD mode, as well as the usage of each subnetwork and its impacts on network performance indicators. The AD column is the summation of the values of all three subnetworks. In all scenarios, around 70% of TTD for AVs is in AD mode, which is very high given that the length of all feasible roads for AD subnetworks (A_1) is 26% of the length of all roads in the network. Yet this approximately 70% of TTD for the AV class in AD mode only accounts for approximately 43% of TTT and 50% of TTC of this class on average. This is partly due to the efficiency of AD mode and partly due to the compartmentalization of traffic among different road types as a result of the combined AD subnetwork deployment that draws AVs towards main roads having higher speeds and partially repels RVs from them.

Comparing the ratios of TTC and TTT to TTD for different subnetworks reveals that dedicated AV lanes provide more traffic efficiency benefits in comparison to the AV-ready subnetwork. The reason is the higher capacity gains of dedicated AV lanes. On the other hand, dedicated AV links are rarely selected in (near) optimal plans since they can compromise accessibility of RVs when there is no alternative route for them. Even when there are alternative routes available, rerouting all RV traffic can direct too much flow to certain links and lead to inefficient distribution of traffic throughout the network.

Regarding the distance traveled in AV-ready subnetwork and dedicated AV lanes, in 10% MPR scenario, the AV-ready subnetwork is chosen more frequently with 63% of the TTD of AV class versus 7% for dedicated AV lanes. However, as MPR of AVs increases, this balance shifts towards dedicated AV lanes until the 50% MPR scenario when 46% of the TTD of AVs is on these lanes while the share of AV-ready subnetwork drops to 28%. With the increase of AV MPR to 90%, the ratio changes to 39% versus 34% in favor of dedicated AV lanes. The transition from 50% to 90% can be explained by the fact that there are many two-lane roads in the network, thus assigning 50% of the capacity to 90% of traffic is not efficient considering that dedicated AV lanes are more expensive than AV-ready links. In this case, the extra capacity gain of dedicated lanes does not compensate for the loss of cost and efficiency.

In order for dedicated AV lanes to be efficient in terms of traffic distribution, the number of these lanes should be proportional to the share of traffic to which the lanes are dedicated. However, in real networks, situations such as the one mentioned earlier with 90% AV volume and only 50% of capacity (i.e., one out of two lanes) available for them are inevitable.

Table 5.5 Network performance indicators for the AV class in all subnetworks

	AV-ready subnetwork	Dedicated AV links	Dedicated AV lanes	AD mode	AV class
10% MPR					
TTC (€)	44.83%	0.06%	4.59%	49.47%	100.00%
TTT (h)	38.21%	0.06%	3.31%	41.55%	100.00%
TTD (km)	62.57%	0.08%	7.07%	69.72%	100.00%
30% MPR					
TTC (€)	28.82%	0.12%	23.43%	52.37%	100.00%
TTT (h)	26.58%	0.11%	16.94%	43.63%	100.00%
TTD (km)	37.52%	0.15%	35.52%	73.19%	100.00%
50% MPR					
TTC (€)	19.49%	0.11%	28.39%	47.99%	100.00%
TTT (h)	18.89%	0.11%	23.84%	42.84%	100.00%
TTD (km)	27.72%	0.15%	45.57%	73.44%	100.00%
90% MPR					
TTC (€)	22.36%	0.05%	26.57%	48.97%	100.00%
TTT (h)	20.20%	0.04%	24.92%	45.16%	100.00%
TTD (km)	34.09%	0.07%	39.15%	73.31%	100.00%

Note: Indexing is based on the values of the AV class column

Distance coverage of subnetworks

As it is shown in Table 5.6, considering all four scenarios, 51-59% of all roads selected for all subnetworks are motorways, 27-30% regional roads and 13-19% urban roads. Motorways constitute around two-third of dedicated AV lanes in all scenarios while regional roads have a slightly higher share of the other one-third compared to urban roads. For the AV-ready subnetwork, the ratios change drastically in different scenarios. At 10% MPR, more than half of AV-ready links are motorways, and regional roads cover around two-third of the rest. Starting from 30% MPR, regional roads have the highest share of AV-ready links with percentages slightly less than 50% of AV-ready links in most scenarios. Urban roads contribute to around 20% of AV-ready links in all scenarios except for 10% MPR when their share is slightly more than 10%.

The total length of the AV-ready subnetwork and dedicated AV lanes follows a similar trend to the distance traveled in each subnetwork, which was discussed in the previous subsection, with AV-ready links having a larger share at first (10% MPR), dedicated AV lanes reaching them at 30% MPR, and the ratio slightly balancing in 90% MPR again. Share of dedicated AV links is insignificant in all scenarios.

Therefore, regarding suitability of each subnetwork for different AV MPRs, it can be concluded that AV-ready subnetworks are more suitable for low MPRs of AVs. However, with the increase in AV MPR, dedicated AV lanes become more effective until a point where the share of AVs gets so high that their traffic efficiency benefits can be fully realized, even in mixed traffic. It is noteworthy that certain parts of road networks such as on-ramps, off-ramps and single-lane roads are exclusively suitable for mixed traffic as long as RVs are on the roads.

Table 5.6 Distance coverage of each subnetwork (km) in each road type for different MPRs

		10% MPR		
Road type	**AV-ready subnetwork**	**Dedicated AV links**	**Dedicated AV lanes**	**All subnetworks**
Motorways	46.43%	0.09%	12.03%	58.55%
Regional roads	25.09%	0.28%	2.35%	27.82%
Urban roads	10.90%	0.19%	2.63%	13.63%
All roads	82.42%	0.66%	17.01%	1,064 (100%)
		30% MPR		
Road type	**AV-ready subnetwork**	**Dedicated AV links**	**Dedicated AV lanes**	**All subnetworks**
Motorways	19.33%	0.15%	31.70%	51.18%
Regional roads	19.94%	0.38%	9.55%	29.87%
Urban roads	10.54%	0.46%	7.87%	18.95%
All roads	49.89%	0.99%	49.20%	1,309 (100%)
		50% MPR		
Road type	**AV-ready subnetwork**	**Dedicated AV links**	**Dedicated AV lanes**	**All subnetworks**
Motorways	14.18%	0.08%	36.81%	51.07%
Regional roads	18.52%	0.30%	10.98%	29.80%
Urban roads	9.91%	0.38%	8.77%	19.13%
All roads	42.61%	0.76%	56.55%	1,312 (100%)
		90% MPR		
Road type	**AV-ready subnetwork**	**Dedicated AV links**	**Dedicated AV lanes**	**All subnetworks**
Motorways	16.32%	0.08%	34.71%	51.11%
Regional roads	18.23%	0.23%	11.21%	29.67%
Urban roads	9.92%	0.15%	9.08%	19.22%
All roads	44.47%	0.46%	55.07%	1,311 (100%)

Sensitivity analysis

In this study, we used the knowledge available so far in academic literature to estimate the values of extra capacity gain and infrastructure adjustment costs for AV-ready subnetwork, dedicated AV lanes and dedicated AV links. However, these values are estimates with high uncertainties, and the real values might deviate from these estimates. Therefore, we perform a sensitivity analysis for these parameters to investigate the impacts of deviations from these values on results. We provide two sensitivity analysis variations for each subnetwork, namely, the optimistic case and the pessimistic case. In the optimistic case, the analyzed subnetwork is assumed to be more efficient with regard to the trade-off between capacity gain and adjustment cost compared to the original scenario in order to estimate the upper bound of the subnetwork's impacts. This is achieved via decreasing the link adjustment cost by 50% and increasing the link capacity by 50% (for AV-ready subnetwork the increase in capacity is realized via decreasing the PCE value for AVs within the subnetwork). For the pessimistic case, the analyzed subnetwork is assumed to be less efficient by means of 50% higher adjustment cost and 50% lower capacity gain compared to the original scenario in order to estimate the lower bound of the subnetwork's impacts. Table 5.7 summarizes the results of the sensitivity analysis for the 50% MPR scenario.

Notable changes in KPIs (i.e., TTC, TT and TTD) due to deviations in capacity and adjustment cost parameters are TTT and TTC increases (maximum 5% compared to the original 50% MPR scenario) in pessimistic variations of AV-ready subnetwork and dedicated AV lanes. This was to be expected, particularly since both capacity and link adjustment cost were modified simultaneously. The rest of the variations indicate no major deviation in KPIs. Conversely, for TAC, significant changes are observed in all variations, with AV-ready subnetwork showing the most sensitivity (37-49% change), dedicated AV lanes showing relatively lower sensitivity (around 25% change), and dedicated AV links displaying the least sensitivity (less than 5%) to deviations in adjustment cost and capacity changes. This means with reasonable deviations in values of input parameters, comparable gains in KPIs can be achieved, only with more or less budget depending on the direction of the deviations. Regarding the trade-off between costs and benefits, among all sensitivity analysis scenarios considered here, the lowest value of the TDTCS to TAC ratio, which is approximately 20 in the original 50% scenario, is around 18, which belongs to the pessimistic scenario for the AV-ready subnetwork. These are useful insights for planners since they show that when deploying AD subnetworks, the main uncertain factor is the amount of budget required for the project, yet the improvements in the performance of the network is rather stable with moderate deviations in input.

Table 5.7 Sensitivity analysis for 50% MPR scenario

Analyzed subnetwork	Parameter deviation			Performance deviation				
	Capacity (Δ_a^s)	Cost (κ_a^s)	PCE (γ_1)	OF	TTC	TTT	TTD	TAC
AV-ready subnetwork	-	-50%	-50%	-01.24%	-00.89%	-01.85%	+00.06%	-50.66%
	-	+50%	+50%	+03.66%	+03.41%	+04.83%	-00.06%	+37.08%
Dedicated AV links	+50%	-50%	-	+00.51%	+00.54%	+00.98%	00.00%	-03.61%
	-50%	+50%	-	+00.40%	+00.41%	+00.65%	+00.05%	-01.21%
Dedicated AV lanes	+50%	-50%	-	+00.03%	+00.20%	+00.11%	+00.07%	-23.70%
	-50%	+50%	-	+01.68%	+01.51%	+03.14%	-00.03%	+25.43%

A note on computations

The mathematical model and the solution algorithm introduced in this chapter were coded in MATLAB and ran on a Windows PC with a Core i5-8600 CPU @ 3.10GHz and 32 GB RAM. Since population-based algorithms such as the one developed in this study can be implemented more efficiently using parallel computation architectures, the Parallel Computation Toolbox in MATLAB with 6 computation units was used to utilize the efficiency of parallel computing in order to perform fitness evaluations for multiple solutions simultaneously. Moreover, sparse matrices in MATLAB were used for all matrix operations on assignment maps to minimize the computation times of the MSA-based algorithm used to solve the lower level problem. It is noteworthy that since assignment maps for large-scale networks include very few non-zero elements (i.e., they are very sparse), matrix operations using sparse matrices in MATLAB are significantly more efficient compared to regular matrix and loop operations. The lower level problem contained 12,781 continuous decision variables (12,781 links in the network) and the upper level problem included 3,402 integer decision variables (3,402 feasible links for the AD subnetwork). The average computation time for an optimization run in this chapter was approximately 28 hours.

5.4 Conclusions

In this chapter, we proposed a unified formulation for simultaneous deployment of AV-ready subnetworks, dedicated AV links and dedicated AV lanes, and modeled the problem as a bi-level network design problem. The upper level represents the decisions regarding links to be selected as part of mentioned subnetworks to optimize the trade-off between infrastructure adjustment cost and the total system travel cost, while the lower level includes a network equilibrium model that captures the travelers' responses to new network topologies with their route choices. We proposed a heuristic solution method to solve the problem and demonstrated the applicability of the model as well as the solution method on the large-scale road network of the Amsterdam Metropolitan Region.

The results indicated that simultaneous deployment of AV-ready subnetworks, dedicated AV links and dedicated AV lanes is effective and can deliver meaningful network performance improvements in terms of TTC and TTT while causing a slight increase in TTD, particularly for AVs. We also showed that TDTCS values significantly outweigh TAC values in all scenarios, especially after reaching 50% AV MPR where TDTCS can be as much as 20 times higher than TAC. Even in the pessimistic scenarios considered in the sensitivity analysis with higher link adjustment costs and lower capacity gains this ratio is never less than 18 for the 50% MPR scenario.

One essential finding of this chapter is that different subnetworks are relevant for different AV MPRs. For lower MPRs, AV-ready subnetworks, which accommodate AVs in mixed traffic, appear to be the most efficient configuration, but after 30% AV MPR, dedicated AV lanes prove to be more beneficial, particularly for motorways. This can be used as a guideline for planners to develop their strategies regarding road network infrastructure during the transition period to full automation. In addition, the results suggest that road types play a crucial role in the choice of network configuration as well. Motorway on-ramps and off-ramps, single-lane roads and major regional roads that include sections with a single lane are almost exclusively selected for mixed traffic in (near) optimal configurations, while the majority of dedicated AV lanes appear on motorways.

Dedicated AV links were rarely selected by our model since compromising the accessibility of RVs was penalized in this study. Furthermore, the radial structure of the Amsterdam network does not offer many alternative routes, which makes it more difficult to satisfy the RV demand when some links are dedicated to AVs and there is no alternative route for RVs. Even when there are alternative routes available for RVs, rerouting all RV traffic can direct too much flow to certain links and lead to inefficient distribution of traffic throughout the network. Nonetheless, dedicated AV links might be a viable option for urban areas with grid networks that provide more possibilities for compartmentalization of traffic. Moreover, these links might be useful as dedicated links for automated shuttles to be used as public transport in order to serve the demand of the class of travelers who do not have access to vehicles.

Based on the sensitivity analysis, it can be concluded that despite the uncertain development path of AVs, network performance benefits delivered by simultaneous deployment of AV-ready subnetworks, dedicated AV links and dedicated AV lanes are rather stable, and the most uncertain factor of the deployment project is the required budget for deployment. Furthermore, it was shown that the results depend more on the design concept (mixed traffic, dedicated links and lanes) rather than input parameters and the deployment cost of each concept, at least in 50% AV MPR scenario. This means that the inherent characteristics of the design concepts themselves have a stronger influence on the results compared to their deployment cost and capacity gains.

It was shown in this study that optimal deployment plan of AD subnetworks is dependent on the level of AV demand. Our results indicate that AV-ready subnetworks for mixed traffic are the most suitable configuration for low MPRs of AVs, and after 30% AV MPR, dedicated AV lanes become more relevant. Yet the results also indicate that for 90% AV MPR, AV-ready subnetworks are more suitable than dedicated AV lanes for about half of the roads in the network. This suggests that developing a multi-stage model that starts each stage with the network topology obtained during the previous stage can be beneficial as well. However, according to our model, there is no noticeable difference between the 50% MPR scenario and the 90% MPR scenario with respect to the location of links to include in the AD subnetwork. This implies that if the cost of changing dedicated AV lanes to AV-ready links for mixed traffic were to be relatively low, a scenario-based approach would be sufficient for real applications.

Possible improvements of the model and future research directions are listed below.

- Incorporation of a time dimension into the problem by considering a multi-stage bi-level network design model

- Considering other objectives in addition to TTC and TAC, such as traffic safety, emissions and accessibility

- Using dynamic traffic assignment models including detailed representation of intersections

- Modeling other travel modes including public transport, active modes and combined modes.

6 Conclusion and Recommendations

The overarching objective of this thesis was to develop a methodology for determining where and when to invest on road network infrastructure for AVs in order to optimize the network-level traffic performance during the transition period to full automation.

In order to reach this objective, a modular approach was used to develop and apply a comprehensive mathematical model through 4 steps. We started with an essential model to understand the fundamental properties of the problem and gradually added more complexity to the model by considering new dimensions of the problem. As a result, three traffic assignment models and three network design models were developed to realistically capture the main aspects of the problem at a desirable level.

The focus of this thesis was the transition period to full automation with a heterogeneous mix of vehicles of different automation levels. AVs of levels 3-4 have limited ODD. Therefore, their ODD should be clearly defined to specify a feasible realm of operation for their ADS, which can operate in mixed traffic or on dedicated infrastructure. We proposed upgrading this feasible region for the operation of ADSs with infrastructure adjustments to ensure safe usage of ADSs. We argued that not upgrading the road network infrastructure during the transition period to full automation can impose traffic safety risks, reduce traffic efficiency, and limit the operation of level 3-4 AVs to specific parts of motorways. On the other hand, upgrading all links in road networks with infrastructure adjustments to support AVs can impose substantial costs without offering sufficient benefits to justify such investments. Then the question is: where and when to invest on road network infrastructure during the transition period? Answering this question requires a vision for road networks and a strategic plan for its deployment.

We first introduced the concept of AV-ready subnetworks as a vision for urban road networks during the transition period. Then we presented a network design method to optimize the trade-off between its deployment cost and the network performance benefits provided by its deployment. Next, we proposed a multi-stage modeling framework to suggest optimal evolution paths for its deployment. Finally, we considered simultaneous deployment of AV-ready subnetworks, dedicated lanes and dedicated links for AVs and proposed a network design model for optimal deployment of this concept in road networks.

Main findings and scientific contributions of this thesis are summarized in 6.1. This is followed by overall conclusions and practical implications in 6.2, and future research directions in 6.3.

6.1 Main Findings and Scientific Contributions

Main findings of this thesis are summarized in this section and answers to all research questions raised in chapter 1 are provided here.

1. What is a suitable road network configuration for accommodating AVs and what are the impacts of this configuration on network performance? (**Chapter 2**)

We explored the concept of AV-ready subnetworks for vehicle automation levels 3-4 (according to the SAE classification) in an urban road network having mixed traffic and demonstrated its potential impacts. We assumed automated driving will be allowed on a selection of roads. For the remaining roads, manual driving (although supported by assisting driving automation systems) will be compulsory. Accordingly, we introduced an approach for road selection for AV-ready subnetworks and presented relevant operational concepts.

To evaluate the impacts of this configuration and model different vehicles' route choice behavior in mixed traffic, a static multi-class stochastic user equilibrium traffic assignment with a path-size logit route choice model and a Monte Carlo labeling route-set generation was adapted. Two user-classes were distinguished: vehicles with automation levels 0-2 and vehicles with automation levels 3-4 having a different passenger car unit to account for lower driving headways, a lower value of time, and a higher fuel efficiency.

The results showed a decrease in total travel cost with the increase in market penetration rate of higher automation levels, a decrease in total travel time, and a minor increase in total travel distance. Although in most cases vehicles with higher automation levels benefited more from the improvements, no deterioration in travel conditions was observed for the rest of the vehicles in any scenario. Furthermore, a noticeable shift of traffic from roads with access function towards roads with flow function and distributors was observed. Sensitivity analysis showed that the extent of changes in the impacts is not strongly dependent on the input parameters.

Our findings indicated that deploying AV-ready subnetworks has apparent safety and network performance benefits since it includes redirecting the AV traffic to specific parts of the network that are either inherently safer or have been adjusted with infrastructure enhancements to become safer. This selection was shown to improve the network performance as well.

2. How can network configurations for AVs be optimized and what are the impacts of the optimal configuration on network performance? (**Chapter 3**)

We formulated the problem as a network design problem and presented a bi-level model to optimize the trade-off between infrastructure adjustment costs and network performance benefits of AV-ready subnetwork deployment in road networks. The upper level included the choice of links to be upgraded as part of the AV-ready subnetwork and the lower level entailed the travelers' response to these decisions, which was captured via the multi-class network equilibrium model developed in chapter 2.

We defined solution requirements for the problem, suggested a solution algorithm that meets those requirements, and benchmarked its performance against two solution algorithms for network design problems considering three different performance criteria. Numerical examples for the network of Delft were presented to demonstrate the concept and solution algorithm performances.

The results revealed that the algorithm presented in this study has a satisfactory performance and outperforms competing algorithms in all three criteria considered, namely, effectiveness, efficiency and design quality. The design quality criteria is specifically relevant to the formulation of the network design problem introduced in this study, which enforces a connectivity constraint on AV-ready subnetworks. It is specifically this constraint that makes the commonly used solution methods for network design problems less suitable for this problem. The algorithm developed in this study is better tailored to this new formulation of the problem and proves to be efficient in dealing with connectivity constraints.

Furthermore, our findings indicated that the optimal layout of AV-ready subnetworks depends on the level of AV demand and infrastructure adjustment costs; lower penetration rates of AVs call for less dense subnetworks and less investment. Nonetheless, a high level of network performance benefit is achievable with low-density AV-ready subnetworks. This is, to some extent, due to the fact that more effective designs usually include roads with highest capacity first (i.e., motorways and regional roads that can handle large traffic volumes at high speeds). Therefore, it is logical to redirect automated driving to designated parts of the network that are selected based on their inherent safety and optimal traffic performance.

On the other hand, higher penetration rates of AVs demand denser AV-ready subnetworks and can deliver more benefits. A crucial observation was that an effective subnetwork can deliver a large proportion of the benefits obtained by upgrading all feasible links, which includes substantial costs, with significantly lower investment cost, especially with higher penetration rates of AVs. These findings are in accordance with findings of chapter 2.

3. How should the optimal road network configurations evolve over time and what are the impacts of network and AV demand evolution on network performance? (**Chapter 4**)

We proposed multi-stage deployment of AV-ready subnetworks in road networks and formulated the problem as a time-dependent network design problem, which is a bi-level mix-integer programming problem.

The upper level denoted infrastructure decisions made by authorities in several stages over a finite planning horizon to deploy and update the AV-ready subnetwork, that is, making a selection of links to upgrade as part of the AV-ready subnetwork in each decision stage. The objective of the formulation was to optimize total societal benefits, which was represented by the summation of total discounted adjustment cost and total discounted travel cost. The lower level involved a travel mode and route choice equilibrium model, which represented travel choices of different user classes in each stage in response to new road network topologies.

We presented variational inequality and fixed-point formulations of the lower level problem as a multi-class simultaneous mode and route choice user equilibrium by means of a hierarchical logit model, and solved it using a sequential linear approximation type algorithm. The upper level problem was modeled as a mathematical program with equilibrium constraints and solved for near-optimal solutions. The model was demonstrated on a case study using a realistic representation of the road network of the Amsterdam metropolitan region.

As it was shown in chapter 3, optimal network configurations for AVs are sensitive to the level of AV demand, which is expected to evolve over time. Therefore, these configurations should evolve over time as well. This makes the multi-stage planning approach necessary, which adds tremendous complexity to the problem, particularly for large-scale road networks, and calls for efficient solution methods. Therefore, we developed two efficient evolutionary heuristic algorithms that are tailored to the problem structure to solve the problem, and compared their performance to a genetic-algorithm-based heuristic solution method.

The results indicated that both proposed algorithms can efficiently solve a large-scale instance of the problem while satisfying constraints in all scenarios, whereas the genetic-algorithm-based solution procedure failed to meet some requirements in scenarios with larger number of time periods. The advantage of the genetic algorithm over proposed algorithms is that it is available in most optimization packages and can easily be applied to the problem with a rather standard penalty function. However, with the growth in problem size, its performance becomes unsatisfactory.

Regarding network performance, in all scenarios, multi-stage deployment of AV-ready subnetworks led to improvements in overall network performance in terms of total travel time and cost. Yet the improvements were always accompanied by increased total travel distance. Main factors affecting the extent of the changes in network performance were AV market penetration rate, AV-ready subnetwork density and timing of AV-ready subnetwork densification. This leaves planners with two choices in their approach regarding their AV infrastructure decisions. The first choice is to invest early and stimulate the adoption of AVs. The second choice is to react to the natural trend in the adoption of AVs. A comprehensive discussion on both approaches is provided in the next subsection.

4. How can different network configurations for accommodating AVs on road networks be combined and what are the impacts of combining different network configurations on network performance? (**Chapter 5**)

The literature suggests that dedicated infrastructure for AVs and enhanced infrastructure for mixed traffic (i.e., AVs on the same lanes with regular vehicles) are the main alternatives for accommodating AVs on road networks during the transition period to full automation. We utilized both alternatives and proposed a unified mathematical framework for optimizing road networks for AVs by simultaneous deployment of AV-ready subnetworks for mixed traffic, dedicated AV links and dedicated AV lanes.

We modeled the problem as a bi-level network design problem where the upper level represents the decisions regarding links to be selected as parts of mentioned subnetworks in order to optimize the trade-off between infrastructure adjustment cost and total system travel cost, and the lower level contains a network equilibrium model that captures the travelers’ responses to new network topologies with their route choices. An efficient heuristic solution method was introduced to solve the problem and find coherent network topologies in realistic networks. Applicability of the model on real road networks was demonstrated using a large-scale case study of the Amsterdam metropolitan region.

The results indicated that simultaneous deployment of AV-ready subnetworks, dedicated AV links and dedicated AV lanes is effective and can deliver meaningful network performance improvements while causing a slight increase in total travel distance, particularly for AVs.

According to the results, different subnetworks are relevant for different market penetration rates of AVs. For low market penetration rates, AV-ready subnetworks, which accommodate AVs in mixed traffic, were the most efficient configuration. After 30% market penetration rate, dedicated AV lanes appeared more often in optimal network topologies, but for very high market penetration rates, AV-ready subnetworks were selected more frequently again.

The results also suggested that road types play a crucial role in the choice of network configuration. Motorway on-ramps and off-ramps, single-lane roads and major regional roads including sections with a single lane were almost exclusively selected for mixed traffic in optimal configurations, while the majority of dedicated AV lanes appear on motorways. Dedicated AV links were rarely selected by the model since compromising the accessibility of regular vehicles was penalized.

6.2 Overall Conclusions and Implications for Practice

In this section, we present six overall conclusions of the thesis and discuss the practical implications pertinent to the deployment of the developed concepts. The methodological developments introduced in this dissertation along with the insights offered by the discussions provide transport planners with ample means to explore, evaluate and deploy the network design concepts presented here on their road networks.

6.2.1 Overall Conclusions

First overall conclusion of this thesis is that the AV market penetration rate is the dominating factor to affect the road network performance. This phenomenon has been observed in all chapters of this thesis. Although the mentioned network performance improvement is, to some extent, due to model parameters that depend on AV market penetration rate, the rate of changes in network performance benefits were sharper than a linear increase proportional to the changes in parameters with increase in market penetration rate of AVs. Furthermore, sensitivity analyses performed in different chapters corroborate that realistic changes in input parameters related to traffic efficiency of AVs do not have a major impact on network performance.

Second overall conclusion is that the bi-level modeling framework used in this thesis is essential for evaluating the impacts of infrastructure plans on network performance. This is due to the fact that the performance of infrastructure adjustment decisions depends on the travelers' response to these decisions via their travel choices. Both the lower level and the upper level can be modeled using different approaches (e.g., a simulation model instead of a mathematical model for the lower level), yet the interactions between the infrastructure decisions and the travel choices are crucial.

Third overall conclusion is that the connectivity of subnetworks representing designated infrastructure for AVs is an indispensable requirement. We established that this connectivity is necessary for finding coherent AD subnetworks, and we introduced four evolutionary solution methods that are tailored to the problem structure and can efficiently find connected AD subnetworks in case studies of large-scale road networks. We showed that general metaheuristics with penalty functions are not efficient for dealing with the connectivity constraint and that tailored algorithms have a superior performance. The numerical experiments and the computational analyses performed on the algorithms used in this project suggest that evolutionary algorithms tailored to the problem are the most suitable ones for solving the problems studied here and finding coherent AD subnetworks.

Fourth overall conclusion is that an effective AV-ready subnetwork including an appropriate selection of links to be upgraded with infrastructure adjustments to accommodate AVs can deliver a large proportion of the benefits obtainable from upgrading infrastructure on all links with significantly lower investment cost. Furthermore, deploying AV-ready subnetworks has apparent safety and network performance benefits during the transition period. It includes redirecting the AV traffic to specific parts of the network that are either inherently safe for automated driving or have been adjusted with infrastructure enhancements to become sufficiently safe. When the selection is made considering network performance as the main objective, it can ensure improved network performance as well. Therefore, it is recommendable for transport authorities and planners to consider this configuration for the transition period.

Fifth overall conclusion is that different network layouts for accommodating AVs in road networks are relevant for different market penetration rates of AVs, which was to be expected. For lower market penetration rates, AV-ready subnetworks, which accommodate AVs in mixed traffic, are found to be the most efficient configuration. However, starting from around 30% market penetration rate, dedicated AV lanes become relevant, and can efficiently host the AV traffic provided the capacity improvements due to their deployment can indeed be realized. This is, to some extent, due to the assumption that dedicated AV lanes will have a higher capacity compared to the AV-ready links. However, the natural characteristics of dedicated AV lanes such as separating the flows of different vehicle types play a part in the results as well. This is evident in very high market penetration rates of AVs (e.g., 90%) when these lanes get overcrowded and are not the most efficient alternative anymore, at least in some parts of road networks. For instance, on two-lane roads, the only option including dedicated AV lanes is dedicating one lane to AVs; when 90% of the flow belongs to AVs, assigning half of the lanes to them leads to unnecessary over congestion, and the capacity gain of dedicating a lane to AVs does not compensate for this over congestion. Additionally, at high AV market penetration rates, a great proportion of safety and traffic efficiency benefits of AVs can be realized in mixed traffic, thereby eliminating the need for separation of flows.

Sixth overall conclusion is that road types play a crucial role in the choice of network configuration as well. Motorway on-ramps and off-ramps, single-lane roads and major regional roads that include sections with a single lane have been shown to be appropriate for mixed traffic, while dedicated AV lanes are most suited for motorways. Therefore, an effective strategy for the transition period can be starting with AV-ready subnetworks for mixed traffic, gradually adding dedicated AV lanes in suitable places after reaching 30% AV market penetration rate, and rolling back to mixed traffic everywhere when reaching high AV market penetration rates. Alternatively, the addition of dedicated AV lanes can be selective and can be stopped before reaching high market penetration rates of AVs in order to avoid rolling back to mixed traffic later, particularly in case this transition turns out to be costly.

6.2.2 Implications for Practice

Applicability of the models developed in this thesis

The multi-class traffic assignment model presented in chapter 2 can be applied in exploratory studies of AV-ready subnetworks and their impacts on network performance. The bi-level network design model developed in chapter 3 can be used to determine the optimal location of AV-ready links in road networks given different demand scenarios for AVs. This model also serves as an intermediate step in developing models presented in chapter 4 and chapter 5. The model described in chapter 5 is useful when considering a combination of different network configurations including dedicated infrastructure for AVs and AV-ready subnetworks designated to mixed traffic, yet this model can be used for optimizing each one of the three subnetworks included individually as well. The multi-stage network design model developed in chapter 4 is useful for determining optimal AV-ready subnetwork evolution over long planning horizons. It explicitly considers the interactions between network performance and the adoption rate of AVs.

However, it is crucial to note that the uncertainties with respect to the AV adoption rate are large, and the estimates of AV market penetration rates using the existing models should be treated with caution. Therefore, the multi-stage approach might not perform well for long term decisions when market penetration rate of AVs is not estimated accurately. When this is the case, the scenario-based approaches introduced in chapter 3 and chapter 5 are more suitable for infrastructure planning.

When accurate estimates of the costs associated with changing one network layout to another (e.g., a dedicated AV lane to an AV-ready link) are available, a multi-stage network design model similar to the one developed in chapter 4 might be useful for identifying the optimal evolution of AD subnetworks (including different network layouts for AVs) over time as well. However, the model should include dedicated AV lanes in addition to AV-ready links, the possibility of changes from one concept to the other in time, and considerations for the cost associated with the change.

Systematic changes in traffic patterns in the future

The observed patterns in the shift of traffic among different road types after deployment of new network configurations for AVs repeated themselves in different settings and case studies throughout this thesis, indicating that this is a phenomenon to be expected in real applications. It includes AVs using the most efficient part of the network designated to them for automated driving, i.e., high-capacity roads with flow function such as motorways, freeways and arterials. This prompts AVs to opt for longer routes including these roads, thereby causing higher travel distances. The shift of AV traffic towards routes with flow function lifts the traffic burden from the rest of the roads and leaves them somewhat less congested for regular vehicles.

The main advantage of this traffic compartmentalization is that high-capacity roads are used more intensively yet mostly in automated mode, which is more efficient for traffic. Moreover, lower capacity roads are used less often but mostly by regular vehicles, which offers shorter routes to regular vehicles in most cases. One of the disadvantages of this traffic compartmentalization is slightly higher travel distances for AVs that may increase emissions. Another disadvantage of this compartmentalization is the more frequent use of roads with distribution and access function by regular vehicles; these road types are typically less safe compared to roads with flow function, so regular vehicles can become more exposed to traffic hazards compared to AVs. Nevertheless, human drivers are better equipped for driving on such roads. Also, when the flow on these roads is decreased compared to the reference case, they become relatively safer.

Accessibility implications of dedicated AV links

We showed that dedicated AV links can compromise accessibility of regular vehicles, particularly when there are not sufficient alternative routes without dedicated AV links available for regular vehicles. The somewhat radial structure of the Amsterdam network studied in this thesis characteristically offers very few (competitive) alternative routes. In addition, even in cases where one or two alternative routes were available, the extra traffic imposed on the alternative routes caused over congestion in many cases. Besides, in this thesis we penalized compromising accessibility of regular vehicles. Due to all these factors, dedicated AV links were not found to be beneficial with respect to network performance during the transition period, at least for the network of the Amsterdam metropolitan region.

Nonetheless, dedicated AV links might be a viable option for urban areas with grid-shaped networks, which offer more alternative routes and possibilities for compartmentalization of traffic, particularly if the authorities value promotion of AVs rather than accessibility of regular vehicles. Alternatively, these links can be useful as dedicated links for level-4 automated shuttles to be used as public transport in order to serve the demand of the class of travelers who do not have access to vehicles. These insights can be used as guidelines for planners to develop their strategies regarding road network infrastructure during the transition period to full automation.

Transport network design and urban livability

The last remark in this section is regarding the compatibility of network design concepts developed in this thesis with urban livability concepts, such as car-free zones. In the future, access to city centers and dense urban areas can be limited or banned for privately owned vehicles, and last mile trips can be undertaken using public transport, shared mobility, walking, biking and new micro mobility concepts. In that case, subnetworks for AVs can be deployed around the edges of the car-free zones with hubs as access points to these zones to allow smooth transfers to other modes. Underground parking areas for automated parking around access and egress points of these zones can be integrated into this scheme as well.

For instance, a home-work trip can start in a suburb with an AV on an AV-ready subnetwork until an access point to a car-free city center where the AV can be left to park itself in an underground automated parking area, and the trip to work can be completed by walking or using a bike that is available at the access point. Alternatively, shared (electric) AVs can be used on subnetworks designated to them in urban areas for the last mile of long-distance trips starting with public transport. Such concepts are feasible with level-4 AVs, which can have the functionality of level 5 within subnetworks dedicated to them. We showed that using a limited number of roads designated to AVs, large proportions of the lengths of the trips made by these vehicles can be completed in automated driving mode. Testing such concepts requires new modeling developments, yet they are compatible with the ones introduced in this thesis.

6.3 Future Research Directions

Several shortcomings of the research presented in this thesis as well as opportunities for future research in order to improve the methodology and to evaluate the applicability of the methodology are presented in this section.

6.3.1 Methodology

In all problems formulated in this thesis to find optimal designs, the objective of the optimization problem has been minimizing the sum of discounted infrastructure adjustment cost and total travel cost. While safety was implicitly considered by restricting the flow of AVs to feasible parts of networks for them, and we discussed several other criteria such as total travel distance, total travel time, accessibility and distribution of impacts among different classes of travelers in ex-post analyses of results, these criteria were not explicitly used for optimization. Future work can include multi-objective optimization methods that consider different objectives and suggest non-dominated solutions with respect to several objectives, e.g., investment cost, total travel cost, safety, environmental impacts, accessibility and social equity. However, quantifying all these criteria and measuring them accurately requires more detailed models, which can pose computational and methodological challenges for large-scale problems. Nonetheless, with the ever-advancing technology, more computation power might become available in the near future.

A multi-stage design method was used in chapter 4 for optimal evolution of AV-ready subnetworks over time, and (single-stage) simultaneous deployment of AV-ready subnetworks, dedicated AV lanes and dedicated AV links was optimized for different AV demand levels in chapter 5. However it was shown that the optimal network configurations depend on the level of AV demand; low levels of AV demand can be efficiently accommodated in mixed traffic, dedicated AV lanes become more effective after reaching 30% AV market penetration rate, and with high levels of AV demand, mixed traffic becomes more efficient again. Therefore, a multi-stage approach would be useful for modeling simultaneous deployment of the subnetworks mentioned above as well. However, the sensitivity analysis performed in chapter 4 indicated that the diffusion model parameters are the most sensitive parameters of the study. Therefore, AV diffusion models require more developments in order to provide a more robust estimate of AV market penetration rates. This is particularly important for multi-stage models with long planning horizons where inaccurate demand estimation can be consequential. Because highly automated vehicles are not available on the market yet, fine-tuning diffusion model parameters or validating their output is not possible at the moment. However, in general, AV demand estimation models that are less reliant on scale parameters and are compatible with discrete choice models can be more reliable for predicting the AV demand in multi-stage network design problems.

The behavioral differences of regular vehicles and AVs and the propagation of traffic throughout road networks were captured using macroscopic static traffic assignment models in this thesis. These models are the most suitable ones for network design problems in terms of computational requirements. They can accurately locate the bottlenecks in road networks and provide realistic estimates for travel times with reasonable computation times. Therefore, these models provide valuable insights into the problems considered in this thesis. Nonetheless, these models lack the resolution to represent all the changes caused by AVs on traffic phenomena. These changes as well as queueing and spill back can be modeled with a higher resolution using dynamic traffic assignment models. Furthermore, dynamic traffic assignment models allow for more elaborate modeling of intersections, which are one of the key elements of urban road networks. However, the concept of AV-ready subnetworks entails eliminating complex intersections and roads with access function from the subnetwork. Therefore, in an appropriate application of this concept, the number of at-grade intersections should be minimal or zero.

When a long planning horizon is considered, in addition to travel route and mode choices, the travelers' origin and destination choices might become relevant as well. This can add yet another layer of complexity to an already complex model. Nonetheless, assuming fixed origin-destination pairs is a non-trivial simplification present in all network design problem studies with a few exceptions.

A more detailed multimodal traffic assignment model including active modes, a more detailed representation of public transport, shared mobility, and combined modes such as park and ride can capture more available choices to the travelers and improve the accuracy of the model results. However this might add a substantial level of detail with limited consequences for the questions addressed in this dissertation. Unless a fundamental change in the use of private vehicles occurs in the future (e.g., ride-sharing becomes the norm), in which case non-trivial modeling developments are required to account for such phenomena.

6.3.2 Applicability

It was shown in chapters 3-5 that the optimal network designs obtained by solution algorithms might require minor adjustments to represent an infrastructure adjustment plan realistically. Perhaps this is always the case when realistic, complex and large-scale design problems are casted in a 'closed' mathematical framework. Yet this can be resolved by a post-processing procedure after running the model to guarantee applicable infrastructure projects. The procedure can be performed by experts who are familiar with the road network under consideration.

Methods and models developed in this thesis were tested on networks of the city of Delft and the Amsterdam metropolitan region, which are located in the Netherlands. Future work could validate and analyze robustness of our findings on case studies of road networks in different countries with different characteristics such as different network topologies and demand patterns.

Models developed in this thesis can be adjusted and extended to study other design concepts that might become relevant in the future and have similar characteristics to AV-ready subnetworks such as dedicated AV zones, optimal subnetworks for electric vehicle charging lanes and AV-ready subnetworks as cordons around car-free zones.

Finally, the bi-level modeling framework considered in this dissertation might be applied with a simulation model (microscopic or macroscopic) for the lower level, instead of the mathematical equilibrium models developed here, to model the behavior of AVs more accurately and to study the impacts of various infrastructure adjustment plans including adjustments that were not studied in this thesis.

Bibliography

Bagloee, S.A., Sarvi, M., Patriksson, M., Rajabifard, A., 2017. A Mixed User-Equilibrium and System-Optimal Traffic Flow for Connected Vehicles Stated as a Complementarity Problem. Computer-Aided Civil and Infrastructure Engineering 32, 562–580. https://doi.org/10.1111/mice.12261

Bel Geddes, N., 1940. Magic Motorways. New York.

Ben-Akiva, M., Bierlaire, M., 1999. Discrete Choice Methods and Their Applications to Short-Term Travel Decisions, in: Handbook of Transportation Science. Kluwer, Dordrecht, The Netherlands, pp. 5–33. https://doi.org/10.1007/0-306-48058-1_2

Ben-Akiva, M., Ramming, M.S., 1998. Lecture Notes: Discrete Choice Models of Traveler Behavior in Networks. Prepared for Advanced Methods for Planning and Management of Transportation Networks, Capri, Italy.

Boggs, A.M., Arvin, R., Khattak, A.J., 2020. Exploring the who, what, when, where, and why of automated vehicle disengagements. Accident Analysis and Prevention 136, 105406. https://doi.org/10.1016/j.aap.2019.105406

Brands, T., 2015. Multi-objective Optimisation of Multimodal Passenger Transportation Networks. University of Twente.

Buchanan, C., 1963. Traffic in Towns; A Study of the Long Term Problems of Traffic in Urban Areas. London.

Calvert, S.C., Schakel, W.J., van Lint, J.W.C., 2017. Will automated vehicles negatively impact traffic flow? Journal of Advanced Transportation 2017. https://doi.org/10.1155/2017/3082781

Cantarella, G.E., 1997. A General Fixed-Point Approach to Multimode Multi-User Equilibrium Assignment with Elastic Demand. Transportation Science 31, 107–128. https://doi.org/10.1287/trsc.31.2.107

Cantarella, G.E., Pavone, G., Vitetta, A., 2006. Heuristics for urban road network design: Lane layout and signal settings. European Journal of Operational Research 175, 1682–1695. https://doi.org/10.1016/j.ejor.2005.02.034

Carreras, A., Daura, X., Erhart, J., Ruehrup, S., 2018. Road infrastructure support levels for automated driving, in: 25th ITS World Congress. Copenhagen, Denmark.

Catalano, S.F., van der Zijpp, N., 2001. A Forecasting Model For Inland Navigation Based On Route Enumeration, in: The AET European Transport Conference. Cambridge, UK, pp. 1–11.

Chakroborty, P., 2003. Genetic algorithms for optimal urban transit network design. Computer-Aided Civil and Infrastructure Engineering 18, 184–200. https://doi.org/10.1111/1467-8667.00309

Chen, A., Pravinvongvuth, S., Xu, X., Ryu, S., Chootinan, P., 2012. Examining the scaling effect and overlapping problem in logit-based stochastic user equilibrium models. Transportation Research Part A: Policy and Practice 46, 1343–1358. https://doi.org/10.1016/j.tra.2012.04.003

Chen, M., Alfa, A.S., 1991. A Network Design Algorithm Using a Stochastic Incremental Traffic Assignment Approach. Transportation Science 25, 215–224. https://doi.org/10.2307/25768493

Chen, Z., He, F., Yin, Y., Du, Y., 2017. Optimal design of autonomous vehicle zones in transportation networks. Transportation Research Part B: Methodological 99, 44–61. https://doi.org/10.1016/j.trb.2016.12.021

Chen, Z., He, F., Zhang, L., Yin, Y., 2016. Optimal deployment of autonomous vehicle lanes with endogenous market penetration. Transportation Research Part C: Emerging Technologies 72, 143–156. https://doi.org/10.1016/j.trc.2016.09.013

Conceição, L., Correia, G., Tavares, J.P., 2017. The deployment of automated vehicles in urban transport systems: A methodology to design dedicated zones. Transportation Research Procedia 27, 230–237. https://doi.org/10.1016/j.trpro.2017.12.025

Conceição, L., de Almeida Correia, G.H., Tavares, J.P., 2020. The reversible lane network design problem (RL-NDP) for smart cities with automated traffic. Sustainability 12. https://doi.org/10.3390/su12031226

Correia, G., de Looff, E., van Cranenburgh, S., Snelder, M., van Arem, B., 2019. The impact of vehicle automation on the Value of Travel Time while performing work and leisure activities in a car: theoretical insights and results from a stated preference survey. Transportation research Part A 119, 359–382. https://doi.org/10.1016/j.tra.2018.11.016

Courbon, T., Scharningg, K., Techner, T., Innamaa, S., Kulmala, R., Alkim, T., Flament, M., Pilli-Sihvola, E., Bergareche, A., Rausch, J., 2016. Facilitating automated driving. The Hague, Netherlands.

Daganzo, C.F., 1983. Stochastic Network Equilibrium with Multiple Vehicle Types and Asymmetric , Indefinite Link Cost Jacobians. Transportation Science 17, 282–300. https://doi.org/10.1287/trsc.17.3.282

Dantzig, G.B., Harvey, R.P., Lansdowne, Z.F., Robinson, D.W., Maier, S.F., 1979. Formulating and solving the network design problem by decomposition. Transportation Research Part B 13, 5–17. https://doi.org/10.1016/0191-2615(79)90003-1

Davis, G.A., 1994. Exact Local Solution Of The Continuous Network Design Problem Via Stochastic User Equilibrium Assignment. Transportation Research Part B 28, 61–75.

de Looff, E., Correia, G.H. de A., van Cranenburgh, S., Snelder, M., van Arem, B., 2018. Potential changes in value of travel time as a result of vehicle automation : a case-study in the Netherlands, in: Transportation Research Board, 97th Annual Meeting.

Dixit, V. V., Chand, S., Nair, D.J., 2016. Autonomous vehicles: Disengagements, accidents and reaction times. PLoS ONE 11, 1–14. https://doi.org/10.1371/journal.pone.0168054

EU-EIP, 2016. Facilitating automated driving. The Hague, the Netherlands.

Even, S., Tarjan, R., 1975. Network flow and testing graph connectivity. SIAM journal on computing 4, 507–518.

Fagnant, D.J., Kockelman, K., 2015. Preparing a nation for autonomous vehicles: Opportunities, barriers and policy recommendations. Transportation Research Part A: Policy and Practice 77, 167–181. https://doi.org/10.1016/j.tra.2015.04.003

Farah, H., Bhusari, S., van Gent, P., Babu, F.A.M., Morsink, P., Happee, R., van Arem, B., 2020. An Empirical Analysis to Assess the Operational Design Domain of Lane Keeping System Equipped Vehicles Combining Objective and Subjective Risk Measures. IEEE Transactions on Intelligent Transportation Systems 1–10. https://doi.org/10.1109/TITS.2020.2969928

Farah, H., Erkens, S.M.J.G., Alkim, T., van Arem, B., 2018. Infrastructure for Automated and Connected Driving : State of the Art and Future Research Directions, in: Meyer, G., Beiker, S. (Eds.), Road Vehicle Automation 4. Springer International Publishing, pp. 187–197. https://doi.org/10.1007/978-3-319-60934-8

Farahani, R.Z., Miandoabchi, E., 2012. Graph Theory for Operations Research and Management: Applications in Industrial Engineering.

Farahani, R.Z., Miandoabchi, E., Szeto, W.Y., Rashidi, H., 2013. A review of urban transportation network design problems. European Journal of Operational Research 229, 281–302. https://doi.org/10.1016/j.ejor.2013.01.001

Favarò, F., Eurich, S., Nader, N., 2018. Autonomous vehicles' disengagements: Trends, triggers, and regulatory limitations. Accident Analysis and Prevention 110, 136–148. https://doi.org/10.1016/j.aap.2017.11.001

Favaro, F.M., Nader, N., Eurich, S.O., Tripp, M., Varadaraju, N., 2017. Examining accident reports involving autonomous vehicles in California. PLOS ONE 12, 1–20. https://doi.org/10.1371/ journal.pone.0184952

Fisk, C., 1980. Some developments in equilibrium assignment. Transportation Research Part B 14B, 243–255.

Florian, M., Hearn, D., 1995. Network Equilibrium Models and Algorithms. Handbooks in Operations Research and Management Science 8, 485–550. https://doi.org/10.1016/S0927-0507(05)80110-0

Ghiasi, A., Hussain, O., Qian, Z. (Sean), Li, X., 2017. A mixed traffic capacity analysis and lane management model for connected automated vehicles: A Markov chain method. Transportation Research Part B: Methodological 106, 266–292. https://doi.org/10.1016/j.trb.2017.09.022

Golberg, D.E., 1989. Genetic Algorithms in Search Optimization & Machine Learning. Addison-Wesley.

Gopalswamy, S., Rathinam, S., 2018. Infrastructure Enabled Autonomy: A Distributed Intelligence Architecture for Autonomous Vehicles, in: IEEE Intelligent Vehicles Symposium (IV). https://doi.org/10.1109/IVS.2018.8500436

Gyergyay, B., Gomari, S., Olstam, J., Johansson, F., Friedrich, M., Sonnleitner, J., Rupprecht, S., Backhaus, W., 2018. Automation-ready framework for urban transport planning, in:

7th Transport Research Arena (TRA). https://doi.org/10.5281/zenodo.1485761

Hardman, S., Lee, J.H., Tal, G., 2019. How do drivers use automation? Insights from a survey of partially automated vehicle owners in the United States. Transportation Research Part A: Policy and Practice 129, 246–256. https://doi.org/10.1016/j.tra.2019.08.008

Harker, P.T., Pang, J.-S., 1990. Finite-dimensional variational inequality and nonlinear complementarity problems: A survey of theory, algorithms and applications. Mathematical Programming 48, 161–220. https://doi.org/10.1007/BF01582255

Holland, J.H., 1975. Adaptation in Natural and Artificial Systems. University of Michigan Press, Ann Arbor, MI.

International Transport Forum, 2015. Automated and Autonomous Driving: Regulation under uncertainty.

Johnson, C., 2017. Readiness of the road network for connected and autonomous vehicles, RAC Foundation.

Keller, D.H., 1935. The Living Machine. Wonder Stories 1465–1511.

Kieft, S., 2013. Handboek Venom (in Dutch). The Netherlands, Stadsregio Amsterdam.

Kröger, F., 2016. Automated Driving in Its Social, Historical and Cultural Contexts, in: Maurer, M., Gerdes, J.C., Lenz, B., Winner, H. (Eds.), Autonomous Driving: Technical, Legal and Social Aspects. pp. 41–68. https://doi.org/10.1007/978-3-662-48847-8

Lavasani, M., Jin, X., Du, Y., 2016. Market Penetration Model for Autonomous Vehicles on the Basis of Earlier Technology Adoption Experience. Transportation Research Record: Journal of the Transportation Research Board 2597, 67–74. https://doi.org/10.3141/2597-09

Le Vine, S., Zolfaghari, A., Polak, J., 2015. Autonomous cars: The tension between occupant experience and intersection capacity. Transportation Research Part C: Emerging Technologies 52, 1–14. https://doi.org/10.1016/j.trc.2015.01.002

Leblanc, L., 1975. An algorithm for the discrete network design problem. Transportation Science 9, 183–199. https://doi.org/https://doi.org/10.1287/trsc.9.3.183

Levin, M.W., Boyles, S.D., 2016. A multiclass cell transmission model for shared human and autonomous vehicle roads. Transportation Research Part C: Emerging Technologies 62, 103–116. https://doi.org/10.1016/j.trc.2015.10.005

Levin, M.W., Boyles, S.D., 2015. Effects of autonomous vehicle ownership on trip, mode and route choice. Transportation Research Record: Journal of the Transportation Research Board 2493, 29–38. https://doi.org/10.3141/2493-04

Li, Y., Chen, Z., Yin, Y., Peeta, S., 2020. Deployment of roadside units to overcome connectivity gap in transportation networks with mixed traffic. Transportation Research Part C: Emerging Technologies 111, 496–512. https://doi.org/10.1016/j.trc.2020.01.001

Liu, Z., Song, Z., 2019. Strategic planning of dedicated autonomous vehicle lanes and autonomous vehicle/toll lanes in transportation networks. Transportation Research, Part C: Emerging Technologies 106, 381–403. https://doi.org/10.1016/j.trc.2019.07.022

Lo, H.K., Szeto, W.Y., 2009. Time-dependent transport network design under cost-recovery. Transportation Research Part B: Methodological 43, 142–158.

https://doi.org/10.1016/j.trb.2008.06.005

Lu, M., Blokpoel, R., 2017. A sophisticated intelligent urban road-transport network and cooperative systems infrastructure for highly automated vehicles, in: World Conference on ITS. pp. 1–11.

Lu, X., Madadi, B., Farah, H., Snelder, M., Annema, J.A., van Arem, B., 2019. Scenario-Based Infrastructure Requirements for Automated Driving, in: 19th COTA International Conference of Transportation Professionals. American Society of Civil Engineers (ASCE), pp. 5684–5695. https://doi.org/10.1061/9780784482292.489

Luo, Q., Saigal, R., Chen, Z., Yin, Y., 2019. Accelerating the adoption of automated vehicles by subsidies : A dynamic games approach 129, 226–243. https://doi.org/10.1016/j.trb.2019.09.011

Lv, C., Cao, D., Zhao, Y., Auger, D.J., Sullman, M., Wang, H., Dutka, L.M., Skrypchuk, L., Mouzakitis, A., 2018. Analysis of AD disengagements occurring during autonomous vehicle testing. IEEE/CAA Journal of Automatica Sinica 5, 58–68. https://doi.org/10.1109/JAS.2017.7510745

Madadi, B., van Nes, R., Snelder, M., van Arem, B., 2020. A bi-level model to optimize road networks for a mixture of manual and automated driving: An evolutionary local search algorithm. Computer-Aided Civil and Infrastructure Engineering 35, 80–96. https://doi.org/10.1111/mice.12498

Madadi, B., van Nes, R., Snelder, M., van Arem, B., 2019. Assessing the travel impacts of subnetworks for automated driving : An exploratory study. Case Studies on Transport Policy 7, 48–56. https://doi.org/10.1016/j.cstp.2018.11.006

Madadi, B., van Nes, R., Snelder, M., van Arem, B., 2018. Image-based assessment of road network readiness for automated driving: A Judgement Game. Delft, The Netherlands.

Magnanti, T.L., Wong, R.T., 1984. Network Design and Transportation Planning: Models and Algorithms. Transportation Science 18, 1:55.

Mahmassani, H.S., 2016. 50th Anniversary Invited Article — Autonomous Vehicles and Connected Vehicle Systems : Flow and Operations Considerations. Transportation Science 50, 1140–1162.

Miandoabchi, E., Daneshzand, F., Farahani, R.Z., Szeto, W.Y., 2015. Time-dependent discrete road network design with both tactical and strategic decisions. Journal of the Operational Research Society 66, 894–913. https://doi.org/10.1057/jors.2014.55

Miandoabchi, E., Farahani, R.Z., Dullaert, W., Szeto, W.Y., 2012. Hybrid Evolutionary Metaheuristics for Concurrent Multi-Objective Design of Urban Road and Public Transit Networks. Networks and Spatial Economics 12, 441–480. https://doi.org/10.1007/s11067-011-9163-x

Migdalas, A., 1995. Bilevel programming in traffic planning: Models, methods and challenge. Journal of Global Optimization 7, 381–405. https://doi.org/10.1007/BF01099649

Milakis, D., Snelder, M., van Arem, B., van Wee, B., Correia, G.H. de A., 2017a. Development and transport implications of automated vehicles in the Netherlands: scenarios for 2030 and 2050. EJTIR 17, 63–85.

Milakis, D., van Arem, B., van Wee, B., 2017b. Policy and society related implications of automated driving : a review of literature and directions for future research. Journal of Intelligent Transportation Systems 2450. https://doi.org/10.1080/15472450.2017.1291351

Nagurney, A., 1993. Network Economics: A Variational Inequality Approach. Kluwer Academic Publishers, Dordrecht/Boston/London.

Nieuwenhuijsen, J., Correia, G.H. de A., Milakis, D., van Arem, B., van Daalen, E., 2018. Towards a quantitative method to analyze the long-term innovation diffusion of automated vehicles technology using system dynamics. Transportation Research Part C: Emerging Technologies 86, 300–327. https://doi.org/10.1016/j.trc.2017.11.016

Nitsche, P., Mocanu, I., Reinthaler, M., 2014. Requirements on Tomorrow's Road Infrastructure for Highly Automated Driving, in: International Conference on Connected Vehicles and Expo. Vienna, Austria, p. 3. https://doi.org/10.1109/ICCVE.2014.7297694

Noriega, Y., Florian, M., 2007. Algorithmic Approaches for Asymmetric Multi-Class Network Equilibrium Problems with Different Class Delay Relationships. Quebec, Canada.

Noruzoliaee, M., Zou, B., Liu, Y., 2018. Roads in transition: Integrated modeling of a manufacturer-traveler-infrastructure system in a mixed autonomous/human driving environment. Transportation Research Part C: Emerging Technologies 90, 307–333. https://doi.org/10.1016/j.trc.2018.03.014

Nowakowski, C., O'Connell, J., Shladover, S.E., Cody, D., 2010. Cooperative Adaptive Cruise Control: Driver Acceptance of Following Gap Settings Less than One Second, in: Proceedings of the Human Factors and Ergonomics Society 54th Annual Meeting. pp. 2033–2037. https://doi.org/10.1177/154193121005402403

O'Brien, L., Szeto, W.Y., 2007. The Discrete Network Design Problem Over Time. HKIE Transactions 14.

Oppenheim, N., 1995. Urban Travel Demand Modeling: From Individual Choices to General Equilibrium. John Wiley & Sons, Inc., New York.

Patel, R., Levin, M.W., Boyles, S.D., 2016. Effects of Autonomous Vehicle Behavior on Arterial and Freeway Networks. Transportation Research Record: Journal of the Transportation Research Board 2561, 9–17. https://doi.org/10.3141/2561-02

Poorzahedy, H., Rouhani, O.M., 2007. Hybrid meta-heuristic algorithms for solving network design problem. European Journal of Operational Research 182, 578–596. https://doi.org/10.1016/j.ejor.2006.07.038

Puylaert, S., Snelder, M., van Nes, R., van Arem, B., 2018. Mobility impacts of early forms of automated driving - A system dynamic approach. Transport Policy 72, 171–179. https://doi.org/https://doi.org/10.1016/j.tranpol.2018.02.013

Rios-Torres, J., Malikopoulos, A.A., 2017. Automated and Cooperative Vehicle Merging at Highway On-Ramps. IEEE Transactions on Intelligent Transportation Systems 18, 780–789. https://doi.org/10.1109/TITS.2016.2587582

Rondinone, M., Walter, T., Blokpoel, R., Schindler, J., 2018. V2X Communications for Infrastructure-Assisted Automated Driving, in: 19th IEEE International Symposium on "a World of Wireless, Mobile and Multimedia Networks", WoWMoM 2018. IEEE, pp. 14–19. https://doi.org/10.1109/WoWMoM.2018.8449775

SAE International, 2018. Taxonomy and Definitions for Terms Related to Driving Automation

Systems for On-Road Motor Vehicles.

SAE International, 2016. Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles.

Shida, M., Doi, T., Nemoto, Y., Tadakuma, K., 2010. A Short-distance Vehicle Platooning System: 2nd Report, Evaluation of Fuel Savings by the Developed Cooperative Control, in: 10th International Symposium on Advanced Vehicle Control (AVEC). Loughborough, UK, pp. 719–723.

Shida, M., Nemoto, Y., 2009. Development of a small-distance vehicle platooning system, in: 16th ITS World Congress.

Shladover, S., Su, D., Lu, X.-Y., 2012. Impacts of Cooperative Adaptive Cruise Control on Freeway Traffic Flow. Transportation Research Record: Journal of the Transportation Research Board 2324, 63–70. https://doi.org/10.3141/2324-08

Shladover, S.E., 2017. Connected and automated vehicle systems: Introduction and overview. Journal of Intelligent Transportation Systems 0, 1–11. https://doi.org/10.1080/15472450.2017.1336053

Shladover, S.E., 2016. The Truth about "Self-Driving" Cars. Scientific American 314, 52–57. https://doi.org/10.1038/scientificamerican0616-52

Shladover, S.E., Nowakowski, C., Lu, X., Ferlis, R., 2015. Cooperative Adaptive Cruise Control (Cacc) Definitions and Operating Concepts, in: TRB 2015 Annual Meeting. https://doi.org/10.3141/2489-17

Soteropoulos, A., Mitteregger, M., Berger, M., Zwirchmayr, J., 2020. Automated drivability : Toward an assessment of the spatial deployment of level 4 automated vehicles. Transportation Research Part A 136, 64–84. https://doi.org/10.1016/j.tra.2020.03.024

Szeto, W.Y., Jaber, X., O'Mahony, M., 2010. Time-dependent discrete network design frameworks considering land use. Computer-Aided Civil and Infrastructure Engineering 25, 411–426. https://doi.org/10.1111/j.1467-8667.2010.00654.x

Tobin, R., Friesz, T., 1988. Sensitivity analysis for equilibrium network flow. Transportation Science 22, 242–250.

Ukkusuri, S. V., Patil, G., 2009. Multi-period transportation network design under demand uncertainty. Transportation Research Part B: Methodological 43, 625–642. https://doi.org/10.1016/j.trb.2009.01.004

Ulrich, S., Kulmala, R., Appel, K., Aigner, W., Penttinen, M., Laitinen, J., 2020. Consequences of automation functions to infrastructure.

van Arem, B., van Driel, C.J.G., Visser, R., 2006. The impact of cooperative adaptive cruise control on traffic-flow characteristics. IEEE Transactions on Intelligent Transportation Systems 7, 429–436. https://doi.org/10.1109/TITS.2006.884615

Wang, J., Peeta, S., He, X., 2019. Multiclass traffic assignment model for mixed traffic flow of human-driven vehicles and connected and autonomous vehicles. Transportation Research Part B: Methodological 126, 139–168. https://doi.org/10.1016/j.trb.2019.05.022

Wang, Y., Szeto, W.Y., 2017. Multiobjective Environmentally Sustainable Road Network Design Using Pareto Optimization. Computer-Aided Civil and Infrastructure Engineering

32, 964–987. https://doi.org/10.1111/mice.12305

Wardrop, J.., 1952. Some Theoretical Aspects of Road Traffic Research, in: Proceedings of the Institution of Civil Engineers.

Wegman, F., Aarts, L., 2006. Advancing sustainable safety, SWOV Institute for Road Safety Research. Leidschendam, The Netherlands. https://doi.org/10.1016/j.ssci.2007.06.013

Wegman, F., Aarts, L., Bax, C., 2008. Advancing sustainable safety National road safety outlook for The Netherlands for 2005-2020. Safety Science 46, 323–343. https://doi.org/10.1016/j.ssci.2007.06.013

Wu, J.H., Florian, M., He, S., 2006. An algorithm for multi-class network equilibrium problem in pce of trucks: Application to the scag travel demand model. Transportmetrica 2, 1–9. https://doi.org/10.1080/18128600608685656

Xu, C., Ding, Z., Wang, C., Li, Z., 2019. Statistical analysis of the patterns and characteristics of connected and autonomous vehicle involved crashes. Journal of Safety Research 71, 41–47. https://doi.org/10.1016/j.jsr.2019.09.001

Yang, H., 1997. Sensitivity analysis for the elastic-demand network equilibrium problem with applications. Transportation Research Part B: Methodological 31, 55–70. https://doi.org/http://dx.doi.org/10.1016/S0191-2615(96)00015-X

Yang, H., Bell, M., 1998. Models and algorithms for road network design: a review and some new developments. Transport Reviews 18, 257–278. https://doi.org/10.1080/01441649808717016

Yang, H., Meng, Q., 2001. Modeling user adoption of advanced traveler information systems : dynamic evolution and stationary equilibrium. Transportation Research Part A 35, 895–912.

Ye, Y., Wang, H., 2018. Optimal Design of Transportation Networks with Automated Vehicle Links and Congestion Pricing. Journal of Advanced Transportation. https://doi.org/https://doi.org/10.1155/2018/3435720

Zhang, Y., 2013. Adapting Infrastructure for Automated Driving.

Zhou, Z., Chen, A., Wong, S.C., 2009. Alternative formulations of a combined trip generation, trip distribution, modal split, and trip assignment model. European Journal of Operational Research 198, 129–138. https://doi.org/10.1016/j.ejor.2008.07.041

About the Author

Bahman Madadi was born in Tehran, Iran, in 1986.

He obtained his BSc. degree in Industrial Engineering (with distinction) from Istanbul Sehir University in Turkey. In 2016, he obtained his MSc. degree in Industrial and Systems Engineering. During his master studies, he held a teaching assistantship position and assisted in teaching activities of multiple courses. Later, he started a research assistantship position, supported by a grant from Scientific and Technological Research Council of Turkey (TUBITAK), and conducted research on vessel traffic in Istanbul Straits anchorage areas.

In August 2016, he joined the Transport & Planning Department of Delft University of Technology, as a PhD candidate. He was working on the research project "Network design and impacts of automated drving", as part of the project Spatial and Transport impacts of Automated Driving (STAD), supported by the Netherlands Organization for Applied Scientific Research (NWO).

At the Transport & Planning Department, he assisted in teaching activities of the master course "Transportation and Spatial Modeling", and supervised a master student. During his PhD project, Bahman presented his research at several national and international scientific and professional events, and organized several workshops with STAD project partners.

Bahman is an avid learner and has a wide range of academic as well as non-academic interests. His academic interests are focused on operations research, network science, transportation, and data analytics, while he has a keen interest in ethical philosophy, politics, economics, and journalism. In his free time, he enjoys rock climbing, mountaineering, traveling, and playing (European) football.

List of Publications

Contributions included in this PhD thesis

Chapter 2

- **Madadi, B.**, van Nes, R., Snelder, M., & van Arem, B. (2019). Assessing the travel impacts of subnetworks for automated driving : An exploratory study. *Case Studies on Transport Policy*, *7*(1), 48–56. https://doi.org/10.1016/j.cstp.2018.11.006

- **Madadi, B.**, van Nes, R., Snelder, M., & van Arem, B. (2018). Network Design and Impacts of Automated Driving: An Explorative Study. *In proceedings of the 97th Annual Meeting of the Transportation Research Board*.

Chapter 3

- **Madadi, B.**, van Nes, R., Snelder, M., & van Arem, B. (2020). A bi-level model to optimize road networks for a mixture of manual and automated driving: An evolutionary local search algorithm. *Computer-Aided Civil and Infrastructure Engineering*, *35*(1), 80–96. https://doi.org/10.1111/mice.12498

- **Madadi, B.**, van Nes, R., Snelder, M., & van Arem, B. (2018). Optimizing Urban Road Networks for Automated Driving. *In proceedings of the 7th Symposium of the European Association for Research in Transportation*.

Chapter 4

- **Madadi, B.**, van Nes, R., Snelder, M., & van Arem, B. (*under journal review*). Multi-Stage Optimal Design of Road Networks for Automated Vehicles with Elastic Multi-Class Demand.

- **Madadi, B.**, van Nes, R., Snelder, M., & van Arem, B. (2021). Multistage Optimal Design of Road Networks for Automated Vehicles with Multimode Multiclass Demand. *Accepted for the 22nd Conference of International Federation of Operations Research Societies*.

Chapter 5

- **Madadi, B.**, van Nes, R., Snelder, M., & van Arem, B. (*under journal review*). Optimizing road networks for automated vehicles with dedicated links, dedicated lanes and mixed-traffic subnetworks.

Other contributions prior and during the PhD project

- **Madadi, B.,** & Aksakalli, V. (2020). A Stochastic Approximation Approach to Spatio-Temporal Anchorage Planning with Multiple Objectives. *Expert Systems With Applications*, 146:1–14. https://doi.org/10.1016/j.eswa.2019.113170

- **Madadi, B.**, van Nes, R., Snelder, M., Farah, H., & van Arem, B. (2021). Image-based assessment of road network readiness for automated driving. *Accepted for the 6th International Symposium on Highway Geometric Design*.

- **Madadi, B.**, van Nes, R., Snelder, M., & van Arem, B. (2018). Image-based assessment of road network readiness for automated driving: A judgement game. *Delft University of Technology*. 1-59.

- Lu, X., **Madadi, B.**, Farah, H., Snelder, M., Annema, J.A., & van Arem, B. (2019). Scenario-Based Infrastructure Requirements for Automated Driving. *In proceedings of the 19th COTA International Conference of Transportation Professionals*. https://doi.org/10.1061/9780784482292.489

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.

Madadi, B., *Design and Optimization of Road Networks for Automated Vehicles*, T2021/3, January 2021, TRAIL Thesis Series, the Netherlands

Krabbenborg, L.D.M., *Tradable Credits for Congestion Management: support/reject?*, T2021/2, January 2021, TRAIL Thesis Series, the Netherlands

Castelein, B., *Accommodating Cold Logistics Chains in Seaport Clusters: The development of the reefer container market and its implications for logistics and policy*, T2021/1, January 2021, TRAIL Thesis Series, the Netherlands

Polinder, G.J., *New Models and Applications for Railway Timetabling*, T2020/18, December 2020, TRAIL Thesis Series, the Netherlands

Scharpff, J.C.D., *Collective Decision Making trough Self-regulation*, T2020/17, November 2020, TRAIL Thesis Series, the Netherlands

Guo, W. *Optimization of Synchromodal Matching Platforms under Uncertainties*, T2020/16, November 2020, TRAIL Thesis Series, the Netherlands

Narayan, J., *Design and Analysis of On-Demand Mobility Systems*, T2020/15, October 2020, TRAIL Thesis Series, the Netherlands

Gong, X., *Using Social Media to Characterise Crowds in City Events for Crowd Management*, T2020/14, September 2020, TRAIL Thesis Series, the Netherlands

Rijal, A., *Managing External Temporal Constraints in Manual Warehouses*, T2020/13, September 2020, TRAIL Thesis Series, the Netherlands

Alonso González, M.J., *Demand for Urban Pooled On-Demand Services: Attitudes, preferences and usage*, T2020/12, July 2020, TRAIL Thesis Series, the Netherlands

Alwosheel, A.S.A., *Trustworthy and Explainable Artificial Neural Networks for choice Behaviour Analysis*, T2020/11, July 2020, TRAIL Thesis Series, the Netherlands

Zeng, Q., *A New Composite Indicator of Company Performance Measurement from Economic and Environmental Perspectives for Motor Vehicle Manufacturers*, T2020/10, May 2020, TRAIL Thesis Series, the Netherlands

Mirzaei, M., *Advanced Storage and Retrieval Policies in Automated Warehouses*, T2020/9, April 2020, TRAIL Thesis Series, the Netherlands

Nordhoff, S., *User Acceptance of Automated Vehicles in Public Transport*, T2020/8, April 2020, TRAIL Thesis Series, the Netherlands

Winter, M.K.E., *Providing Public Transport by Self-Driving Vehicles: User preferences, fleet operation, and parking management*, T2020/7, April 2020, TRAIL Thesis Series, the Netherlands

Mullakkal-Babu, F.A., *Modelling Safety Impacts of Automated Driving Systems in Multi-Lane Traffic*, T2020/6, March 2020, TRAIL Thesis Series, the Netherlands

Krishnakumari, P.K., *Multiscale Pattern Recognition of Transport Network Dynamics and its Applications: A bird's eye view on transport*, T2020/5, February 2020, TRAIL Thesis Series, the Netherlands

Wolbertus, *Evaluating Electric Vehicle Charging Infrastructure Policies*, T2020/4, February 2020, TRAIL Thesis Series, the Netherlands

Yap, M.D., *Measuring, Predicting and Controlling Disruption Impacts for Urban Public Transport*, T2020/3, February 2020, TRAIL Thesis Series, the Netherlands

Luo, D., *Data-driven Analysis and Modeling of Passenger Flows and Service Networks for Public Transport Systems*, T2020/2, February 2020, TRAIL Thesis Series, the Netherlands

Erp, P.B.C. van, *Relative Flow Data: New opportunities for traffic state estimation*, T2020/1, February 2020, TRAIL Thesis Series, the Netherlands

Zhu, Y., *Passenger-Oriented Timetable Rescheduling in Railway Disruption Management*, T2019/16, December 2019, TRAIL Thesis Series, the Netherlands

Chen, L., *Cooperative Multi-Vessel Systems for Waterborne Transport*, T2019/15, November 2019, TRAIL Thesis Series, the Netherlands

Kerkman, K.E., *Spatial Dependence in Travel Demand Models: Causes, implications, and solutions*, T2019/14, October 2019, TRAIL Thesis Series, the Netherlands

Liang, X., *Planning and Operation of Automated Taxi Systems*, T2019/13, September 2019, TRAIL Thesis Series, the Netherlands

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

TRAIL

Summary

Relying on driving automation technology alone without infrastructure support might compromise the potential safety and efficiency of automated vehicles. Therefore, this thesis proposes a road network design method that can cope with the uncertain development path of automated vehicles in the future. Using a modular approach, a comprehensive model is developed to assess the impacts of network infrastructure decisions in presence of automated vehicles on travel behavior and network performance.

About the Author

Bahman Madadi conducted his PhD research at the Transport & Planning Department of Delft University of Technology. He holds a Master's degree in Industrial and Systems Engineering and a Bachelor's degree in Industrial Engineering.

TRAIL Research School ISBN 978-90-5584-272-8

Radboud University

UNIVERSITY OF TWENTE.