JOHANN HARTLEB

Public Transport and Passengers

Optimization Models that Consider Travel Demand

Public Transport and Passengers

Optimization Models that Consider Travel Demand

Public Transport and Passengers – Optimization Models that Consider Travel Demand

Openbaar vervoer en passagiers –
optimalisatiemodellen die reizigersvraag meenemen

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

Prof.dr. F.A. van der Duijn Schouten

and in accordance with the decision of the Doctorate Board.

The public defence shall be held on

Friday 17 September 2021 at 10:30 hours

by

Johann Hartleb
born in Coburg, Germany.

Erasmus University Rotterdam

Doctoral committee

Promotors:	Prof. dr. D. Huisman
	Prof. dr. M. Friedrich
Other members:	Prof. dr. R.A. Zuidwijk
	Prof. dr. A. Schöbel
	Dr. O. Cats
Co-promotor:	Dr. M.E. Schmidt

Erasmus Research Institute of Management - ERIM
The joint research institute of the Rotterdam School of Management (RSM)
and the Erasmus School of Economics (ESE) at the Erasmus University Rotterdam
Internet: www.erim.eur.nl

ERIM Electronic Series Portal: repub.eur.nl

ERIM PhD Series in Research in Management, 535
ERIM reference number: EPS-2021-535-LIS
ISBN 978-90-5892-609-8

Cover image: Marijke van Zuilen, marijkevanzuilen@gmail.com
Design: PanArt, www.panart.nl
Print: OBT bv, www.obt.eu

This publication (cover and interior) is printed on FSC® paper Magno Satin MC.

MIX
Papier van verantwoorde herkomst
FSC® C022586

Acknowledgements

'Johann, what are you going to do next?' I have been asked this question over and over again since I have reached adulthood: before my high school graduation, before obtaining my bachelor's and my master's degrees, and now, in the final phase of my doctoral studies, as I write these words. In most cases where I have been asked this, I have not yet had an answer. My next steps were often determined last minute and by coincidences. To the displeasure of those who always ask me, friends, family and colleagues that care about me, I will continue with this approach, mainly for a simple reason: my extremely good experiences with it. Doing a PhD and jointly working at RSM, VuV and π was a pleasure. I had the chance to work on really interesting projects with a great supervisory team and wonderful colleagues, in all of the departments. Also next to the research, I could enjoy the past five years to the fullest and will look back to this period with joy. Therefore, I would like to express my gratitude to everyone who contributed to it, directly or indirectly. Without all of your support I wouldn't have enjoyed this journey as much and couldn't have written this thesis.

First of all, I want to state I am very grateful for the guidance and support of my supervisors. I couldn't have imagined a better supervisory team and I hope that we will not part ways here yet but still work together on one or the other project.

Marie, thank you for always having an open ear for any kind of problem I encountered, no matter whether they are related to research or not. You have an impressive ability to ask the right questions which brought me back on track in difficult situations and gave me the good ideas to improve the projects. I appreciate your supervising style, your time, effort and enthusiasm you put into the projects with me, and the freedom you gave me to develop in this journey. Last but not least I'd also like to mention how happy I am about the good and relaxed interaction outside of the university. I am already looking forward to the next beer and climbing session!

Dennis, thank you very much for taking over the role as promotor after Leo's unexpected and unfortunate passing. For a moment I was uncertain about the future and direction of the project but all of my concerns were absolutely unnecessary. You always supported me in continuing with and diving deeper into my research and gave me good directions to define my next research projects. I am very grateful that you included me in the π department at NS. Thank you also for your support and patience, especially when teaching me to estimate time for my research projects - I am not there yet, but I have become better by leaps and bounds.

Markus, all my thanks for your support and your enthusiasm for this project, although I chose to spend more time in Rotterdam than in Stuttgart and brought a clear supply bias into this project. I still remember a kind of job interview in the train station of Göttingen, where you stunned me with the level of detail and sophistication of the travel demand models you build at the VuV. Despite its complexity, you managed to teach me the core mechanism behind demand estimation within short time, making a reasonable integration into the optimization models possible after all. I really enjoyed my time at the VuV, being the 'house mathematician' who got to solve the (at least to me) interesting problems, and joining you for the winter seminars.

I would also like to thank the committee members Oded Cats, Anita Schöbel and Rob Zuidwijk for assessing my thesis, giving valuable feedback and asking helpful questions. Special thanks go to Anita Schöbel, who I owe much of my knowledge about Operations Research and who - despite the fact that I'm a car enthusiast - managed to spark my interest in problems in public transport planning.

Thanks to the great colleagues, my time at RSM was really a pleasure. The TOM department has a very welcoming, inclusive and supportive atmosphere and a lot of effort was put to make the work for everybody fun and enjoyable. In particular, thanks to Lianne and Cheryl for not only organizing fun events greatly but also supporting us with a smile in the face and a joke on the lips. Also thanks to Carmen for her support with any finance issue and making sure our offices are always equipped with a variety of flowers and plants, and to Kim and Miho for the support from ERIM who made it happen that this thesis was ready in time.

I am happy to have made that many friendships during my time there. The regular bouldering and squash sessions, and especially the beers thereafter, with Anirudh, Sebastian and Micheal, clearly helped me to keep a healthy balance to all the brain work. I also really enjoyed going to to the office, which is no wonder when having such amazing office mates as Alberto, Liana and Masoud. Our mutual support and the jokes could also not be broken by a pandemic, since T09-48 is even on Teams! I am deeply grateful for the close friendship with my paranymphs Anirudh and Alberto. Your help and support, the endless small discussion about work, life, future and the meaningless, and the time I could spend with you not thinking about anything truly helped me to be where I am. Great thanks also to Arpan, Joydeep, Alp and Kaveh for the many helpful discussions on programming, coding styles and structures of our MIPs. Kaveh, thank you also for all your help, especially with the Bachelor theses supervision and in the battle against the unforeseen depths of hora finita. Thanks also to Gert-Jaap for the numerous train and car chats, Frankie for the fun discussions of the indiscussable, Joshua for helping me to convince others to tidy up my cupboard, everyday a little bit, Davide for pushing me further up the climbing wall (without actually pushing), Katha for astonishing me again and again with a very un-German dry humor, Sabine for making the organizational part of courses not only an easy job but even fun meetings, Ainara for cheering us up in our tax and insurance fighting, and Hobbs for introducing me to the beauty of languages. Hartelijk dank aan Fanny voor de geweldige en gezellige Nederlands lessen en het worden van een goede vriend in zo'n korte tijd. Many thanks go also to Alexandros, Amin, Anna, Atabak, Camill, Erik, Jun, Luna, Mo, Negin, Nienke, Timo, and many more great people at RSM. I would also like to thank the colleagues from ESE, Thomas, Rowan, Rolf, Twan, Paul, Shadi, Naut, Bart and many more for the regular exchange and especially the helpful feedback sessions we had in the Railway group improving each other's papers. Special thanks go to Robin for his help with the experiments. With all of you, I've had many great moments that I'll never forget, a lot of boozy nights that made us grow together, exciting trips, for example to an unforgettable wedding of Arpan and Jenny, visits in your (new) home, and the nice times on conferences.

I was also really lucky to regularly spend some time at the VuV in Stuttgart. Thank you so much for always giving me the feeling to be a full member of the team, for calling me a VuVler, although I spent the bigger part of the time 'in Rotterdam at the beach or elsewhere on holiday'. I really enjoyed my regular Stuttgart visits, having had almost each of you as an office mate for a while. Thanks to Max for the very smooth start, Jörg the tech nörd for keeping the tradition of low puns, Emely for a really smooth and enjoyable teamwork, as well as giving some algorithm relevance in the first place, Matthias for the really nice discussions on the ideal, Alex for endless car talks, and to all the other VuVlers Charlotte, Christoph, Eric, Jürgen, Magdalena, Steffi and Yannik for making working in Stuttgart an enjoyable and fun part of my PhD. I will miss the Monday morning hugs, the fun lunch discussions at the kitchen bar, the Feierabendbier in the sun and the boulder sessions. Also many thanks to Manfred for again and again making the impossible official, Alice for the great support and Micha for the really interesting insights into network architecture.

Even though I was not a very regular visitor in my 'third office', I have very good memories of my time at the π department at NS. I'd like to thank Joël and Niek for their time, the really helpful feedback in the kwartaalafspraken and providing me with data from NS. Your questions and suggestions helped shape my focus in the project and increase the practical relevance of my studies. Thanks to Gábor for the fruitful and funny discussions and for accepting me as a PIer although I never achieved anything close to a respectable nerd score. I am also grateful for the input from Ramon and Pieter-Jan, and the help of Anneke. The time at NS was really fun and in a nice environment, thanks to all the colleagues that contributed to it.

Doing a PhD is an intense time, and I only could make this journey since I am also blessed with so many friends outside of my academic world who are happy to see me, make me laugh and smile, laugh at my stupid jokes, who invite and visit me, keep texting me despite my absolutely inappropriate 1+ month response time, and are there for me when I need them. These acknowledgements already have become too long and there are too many names to be listed and too many stories to be told. Here is a humble attempt to name some of the groups I'd like to thank for the memories and fun time we spent together. Thank you for being there and making home feel like home since kindergarden time, Alice and my dear Stammtisch. It is also great to know that I have friends basically in any part of Germany. After our study time in Göttingen, we all went in different direction, making visits maybe more difficult but even more interesting. I am already looking forward to seeing you next time, either on our yearly hike or somewhere else. Special thanks also to the iaeste crew for the fun visits and trips we made, and to Alex, Daniel and Yulia for the great time in Stuttgart. I am incredibly happy that we still manage to see and feel close to each other as our Bandung family, after that many years living spread in different countries. Thank you all, I hope to see as many of you as possible soon and to spend some time together.

Big thanks also go to ZoefZoef, Alex, Jeanim, Kim, Marijke, Nicole, Ramona, and the previous members. I loved my time here, you made me feel home and I truly enjoyed being a ZoefZoefer and living with you. It breaks my heart to move out soon but I know this will not be the last time I see you. Special thanks goes to Marijke who made the amazing illustration on the thesis cover. I love it! During the corona restrictions, I wrote on my thesis in different home offices, among others in Bremen. Many thanks to Fritjof and Volker who always put up with me, it was a very fun time to stay with you.

Mama and Papa, I have always had difficulty finding the right words and expressing myself in written text, and also here I struggle to find words that could capture the appreciation for everything you gave me to take along. I cherish your attention, input, and unconditional love more than you know. Thanks for raising me as a positive, friendly and modest person, teaching me the important non-academic skills, and giving me the freedom to develop independently. None of my achievements would have been possible without your selfless and endless support. Especially, thank you very much for taking care of Molim and giving him a new appropriate home. Laura, I feel truly fortunate to have such a wonderful sister and I hope that we will always get along as well as we do now. Thanks also to rest of my large family for the support and good times.

Finally, I am incredibly grateful and happy to have you in my life, Malin. You made this journey a colorful, enjoyable and inspiring experience, you cheered me up in difficult times, you knew what I needed before I even realized that I needed something, and you made the sunny days even brighter. Although I am very sad to close this chapter, I am very much looking forward to the next one closer to you.

Johann Hartleb
Rotterdam, 2021

Table of contents

Chapter 1

Introduction

1.1 Background and motivation

It is impossible to imagine a world without public transport, and its contribution to our society is undisputed. First, public transport is a service of general interest. It increases mobility for the public and enables to also travel when no private means of transport are available. As such, public transport contributes to a just society with equal chances for all. Second, public transport is indispensable for a working passenger transport system in densely populated areas. In cities and other crowded places, the consequences of passenger traffic are often severe. Much space is dedicated to streets and parking areas and vehicles block each other, making traveling time-consuming and troublesome. Unlike private transport, public transport can carry a large number of passengers efficiently. By pooling passenger trips and utilizing vehicle fleets to their full capacity, urban travel can be accomplished in a reasonable amount of time and space. Third, public transport is a key element for our efforts to mitigate climate change. The *Paris Agreement* (2015), signed by more than 190 countries, aims to keep the increase in average global temperature below 2°C compared to pre-industrial levels. This can only be achieved by reducing the emissions in all sectors, in particular in the transport sector, which is responsible for approximately one-sixth of the emissions (Ritchie and Roser, 2020). Most of the emissions in this sector are produced by road traffic, while public transport has the least emissions per passenger-kilometer. Hence, public transport can help achieve the goals of the Paris agreement.

To reap its benefits, effective public transport services must be designed attracting large numbers of passengers. This requires public subsidies and constant improvements of the services. Research is necessary to improve the services by giving insights into operations and providing decision support for planning. Even though it is generally accepted that public transport must be of high quality, the goals and requirements concerning public transport are versatile and often even contradictory. We look at public transport from the perspectives of three different stakeholders, namely those of passengers, public transport authorities, and operators.

Passengers' perspective. People want or need to travel, for which they often have several modes of transport available to reach their destination. Depending on the availability, price, and convenience of the modes, they choose which one to use. To be an attractive alternative to travelers, public transport must provide a frequent service with reasonable travel times and, if possible, direct connections. In addition, fares must be low and the services should be punctual, reliable and close-by.

Public transport authorities' perspective. Public transport authorities have the task to efficiently use taxpayers' money for the common good. Hence, their goals concern mainly the consequences of transport, such as carbon footprint, accessibility and fairness for inhabitants, as well as the attractiveness of regions or cities. To achieve the desired conditions, they introduce constraints and incentives for operators and users of transport services. This often takes the form of subsidies for public transport fares, taxes on private transport, bans on certain technologies in selected areas, or the requirement of minimum service levels.

Public transport operators' perspective. In many parts of the world, public transport operators are privatized and the companies work profit-oriented. They aim for low operating costs and high revenues. The transport services they provide must unite the interests of all stakeholders. The service level should be high to attract passengers, while infrastructural and operational constraints, operational costs, as well as incentives and restraints from public transport authorities have to be considered.

1.2 Transport planning

In most cases, operators plan public transport services. Nevertheless, all stakeholders' constraints and objectives have to be considered, making the design of public transport services a complex task. The literature provides a plethora of models and methods for public transport design. A rough distinction is made between *demand-oriented* and *supply-oriented* approaches for transport planning (Cascetta, 2013).

1.2.1 Demand-oriented approaches

Demand-oriented approaches model the travel demand for a certain transport situation and evaluate the performance of the transport situation from several aspects. This evaluation is mostly done with travel demand models, following a step-wise approach (de Dios Ortúzar and Willumsen, 2011; McNally, 2010).

Figure 1.1 shows the main steps of a travel demand model. First, *activity choice* models are used to estimate the number of trips people make and *destination choice* models predict the places they want to travel to. This information is usually stored in an *origin-destination matrix* (OD matrix), where each entry represents the number of people who want to travel from an origin to a destination. The OD matrix may be time-dependent. Next, the mode of transport travelers use to reach their destination is estimated with *mode choice* models. Usually, the transport modes considered are

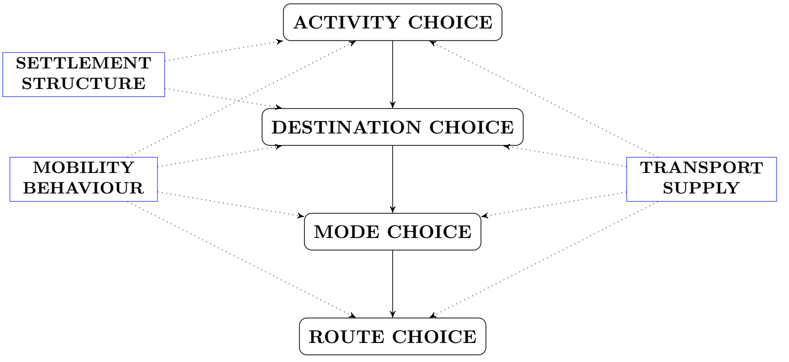


Figure 1.1: Travel demand model consisting of activity, destination, mode and route choice. Settlement structure, mobility behavior and a transport supply are assumed to be given.

walking, bike, public transport, and car, if available. This step results in several OD matrices, one for each mode of transport. Last, the individual routes for the people on their respective modes are predicted with *route choice* models. All available routes are evaluated according to different criteria such as travel time, cost, and convenience factors, among them the number of transfers or vehicle congestion. These models are influenced by the *settlement structure* and *mobility behavior* of people, and a *transport supply*, which includes available public transport services. This input is assumed to be given and fixed in most demand-oriented approaches.

Travel demand models are mainly applied to estimate the travel demand for a certain transport situation. This allows a thorough and detailed evaluation of the transport situation from various perspectives. For the design of a public transport service, a finite set of services are evaluated and compared in an experimental setup. Therefore, travel demand models do not provide potential transport designs, but rather provide decision support by allowing a high level of detail in the assessment of designs.

1.2.2 Supply-oriented approaches

Supply-oriented approaches are used for the design of public transport services. The services are designed to meet a certain passenger demand, which in most cases is assumed to be known before planning. The passenger demand is mostly specified as an OD matrix and in some cases includes passenger route decisions. Since certain design decisions are implemented for longer-term than others, public transport services are usually designed in a step-wise approach as depicted in Figure 1.2.

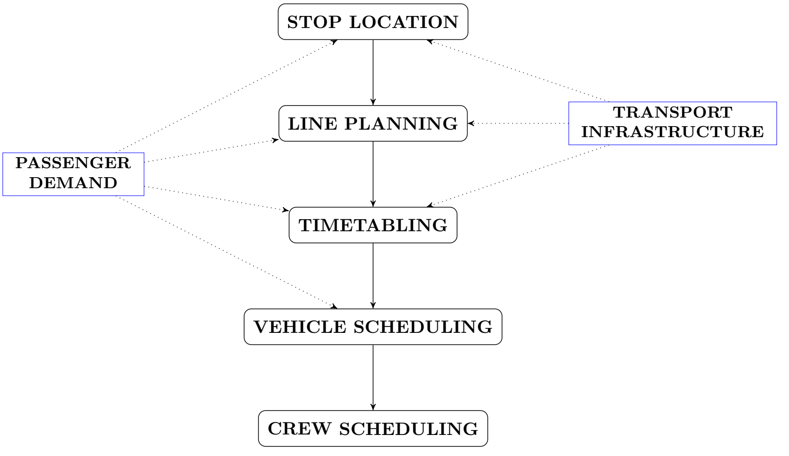


Figure 1.2: Step-wise approach for public transport design, consisting of stop location, line planning, timetabling, vehicle scheduling and crew scheduling. A transport infrastructure and passenger demand are assumed to be given.

To begin with, the optimal location of stops in the public transport infrastructure is determined using *stop location* models. Depending of the field of application, the infrastructure can be seen as a street or track network. The stops should be close to the passengers' destinations, but also low in number to keep installing and maintaining costs for the operator as well as waiting time for passengers at intermediate stops low. If a distinction is made between different service categories, such as regional or intercity transport, it is often also decided at which stop which service will stop. Afterward, in *line planning*, the aim is to find the number, routes, and frequencies of lines in the public transport infrastructure to serve the demand. The line operation should be cost-efficient for operators but provide reasonable travel times and direct connections for passengers. These first two steps concern long-term planning which is referred to as *strategical planning.*

Given the lines and their respective frequencies, optimal arrival and departure times for each vehicle on each line are determined in *timetabling.* Often, the focus is on spreading vehicle trips over time for a regular service, and on realizing short transfer times for passengers. Timetabling is classified as *tactical planning* as it considers medium-term decision making.

Having found solutions to these steps yields sufficient information for passengers about a public transport service. Afterward, the vehicles and drivers get assigned to previously determined trips in *vehicle* and *crew scheduling.* The aim is mostly to minimize operational costs, considering maintenance rhythms of vehicles and work regulations of staff. These steps of scheduling are denoted as *operational planning.*

There are many models and methods developed in the area of Operations Research for the steps of stop location, line planning, timetabling, and vehicle scheduling. An overview of Operations Research models in public transport planning is given in Huisman et al. (2005) or in Borndörfer et al. (2018).

1.2.3 Integration of demand-oriented into supply-oriented approaches

From Figures 1.1 and 1.2, it can be seen that the approaches require each other's results as input. For demand-focused models, a transport supply including public transport services is assumed to be given, whereas for supply-focused models, knowledge about the demand is assumed to be known. Since passenger demand and public transport supply are interdependent, they should be treated simultaneously. That means the passenger demand should be estimated during the design of the public transport services. Although supply and demand are known to influence each other, only few and basic combinations of these research fields are developed, probably due to the complexity of a simultaneous treatment.

In this thesis, we aim at investigating the potential of simultaneous passenger demand estimation and public transport design. We develop integrated models to optimize public transport services while estimating the corresponding passenger choices. The resulting public transport services are designed for the passenger demand they generate.

We focus on the interaction of *mode* and *route choice* from the demand-oriented models and *line planning* and *timetabling* from the supply-oriented models. Mode and route choice can be significantly impacted by changes in the public transport services. The activity and destination choice steps also depend on the availability of public transport, but are less reactive to (moderate) changes in the quality of public transport. The quality of public transport for passengers is primarily determined by the line plan and timetable. Therefore, the passengers' mode and route choices should be taken into account during line planning and timetabling. Stop location planning also significantly determines the people's choice to travel at all, and impacts their destinations. However, since the location of stops is seldom updated, it is not considered in the scope of this thesis. The remaining planning steps of vehicle and crew scheduling only have a minor impact on the passenger choices in public transport.

The challenge of estimating passenger choices during optimization is that demand models are very detailed and complex, necessitating simplifications. In Chapter 2, we investigate how simplifications in passenger modeling can impact the evaluation of timetables. Furthermore, mode and route choice models are usually non-linear and non-convex in the utility of the alternatives. Hence, the integration in an optimization framework quickly yields computationally intractable models. In Chapter 3, we discuss two linear route choice models within timetabling, and in Chapter 4, we present a mixed-integer linear program for line planning with integrated mode and route choice.

1.3 New forms of public transport

The previously discussed supply-oriented models mainly address *traditional* public transport, that is, a regular, scheduled service, operating on fixed lines. The recent development of technology enables new forms of public transport: large-scale and affordable *mobility on demand*. Instant information sharing, for example of vehicle occupancy rates or vehicle and passenger locations, allows a flexible approach without fixed stop locations. Online computing power and new algorithms enable efficient live planning of operations that do not rely on specified lines and schedules. With the usage of autonomously driving vehicles, also smaller-scale vehicles can be operated economically.

For passengers, on-demand services promise a fast and direct service at the time of their preference. The service adjusts to their needs and wishes, instead of requiring them to adjust to a rigid schedule. Due to the pooling of several passengers in a vehicle, the service quality might be inferior to a taxi service, however, this is compensated by considerably lower fares. In general, the launch of on-demand services is expected to improve the travel quality for passengers and they can easily test the services without any obligation.

For operators, the consequences of introducing on-demand services are less clear. Passenger acceptance and associated costs are difficult to estimate, but offering on-demand services requires large investments. Already now, many operators offer non-profitable on-demand services with a driver to learn about the operations and to collect data.

This possibility of data collection is not feasible for public transport authorities. Large-scale experiments to measure the impact of on-demand services on cities, regions and the environment are out of reach. Nevertheless, they need to estimate consequences for the managed region to react with regulatory measures in a timely manner. Hence, many transport scenarios included on-demand services need to be investigated and evaluated with the help of travel demand models.

A travel demand model requires a transport situation as input, including the available public transport service. However, the service level of on-demand services cannot be predetermined as it depends on the demand, which is to be estimated by the travel demand model. Hence, the service level of on-demand services has to be estimated within the travel demand model. In Chapters 5 and 6, we discuss how a travel demand model can be extended correspondingly. We present a heuristic and an exact solution approach to estimate the minimum vehicle fleet size and total distance traveled of on-demand services within a macroscopic travel demand model.

1.4 Thesis outline and contributions

This thesis is structured in two parts and seven chapters.

Part I of this thesis deals with the planning of public transport services as described in Section 1.2. Demand-oriented approaches are integrated into supply-oriented approaches with the goal to estimate passenger demand during public transport optimization. Chapter 2 compares different evaluation functions for consistency and gives further motivation for the integration of passenger choice models into optimization models. Chapters 3 and 4 present novel optimization models with integrated demand estimation for the steps of timetabling and line planning, respectively.

In Part II, we consider the determination of on-demand services within travel demand modeling, as motivated in Section 1.3. Both Chapters 5 and 6 present solution algorithms for a vehicle scheduling problem in the context of traffic estimation.

Figure 1.3 outlines the thesis structure and highlights which steps of demand-oriented and supply-oriented approaches are covered in the respective chapters. It is possible to read the chapters of this thesis independently, however, we recommend reading the chapters in each part in the given order. Chapter 7 concludes the main findings and implications of this thesis, and points out directions for future research. In the following, we give a brief summary of Chapters 2 to 6 and highlight the contributions.

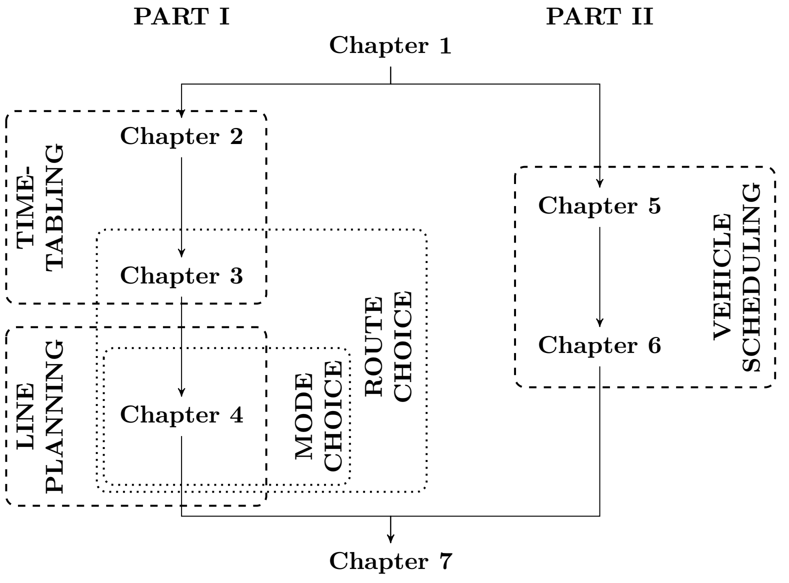


Figure 1.3: Structure of this thesis in two parts and 7 chapters. The dotted and dashed boxes indicate which steps of demand-oriented and supply-oriented approaches are covered in the chapters, respectively.

Chapter 2: Hartleb, Schmidt, Friedrich, and Huisman: "A good or a bad timetable: Do different evaluation functions agree?" In preparation for journal submission.

Estimating passenger demand instead of assuming a fixed demand level when designing public transport services has an impact on how the quality of solutions is assessed. In terms of optimization models, this means that the objective function is adjusted. To assess the extent to which solutions found under different objective functions can differ, we first examine the consistency of evaluation functions using public transport timetables as an example. The literature has established various ways to evaluate public transport timetables from the passengers' viewpoint. In Chapter 2, we investigate to what extent these evaluation functions agree on the quality of a timetable. First, we structure common timetabling evaluation functions and identify three components in which the functions differ from each other. Then, we use a novel method to empirically test the extent to which the evaluation functions are consistent. Our results show that the design of an evaluation function can have a significant impact on which timetable is considered optimal. Due to the structure of our experiments, we are further able to identify which components of evaluation functions influence the result of the evaluation most. This can help to design simple yet purposeful objective functions for Operations Research models.

Chapter 3: Hartleb and Schmidt: "Railway timetabling with integrated passenger distribution". Accepted for publication at the European Journal of Operational Research.

Timetabling for railway services often aims at optimizing travel times for passengers. At the same time, restricting assumptions on passenger behavior and passenger modeling are made. While research has shown that discrete choice models are suitable to estimate the distribution of passengers on routes, this has not been considered in timetabling yet. In Chapter 3, we investigate how multi-route passenger route choice can be integrated into a timetabling optimization framework and present two mixed-integer linear programs for this problem. Both approaches design timetables and simultaneously find a corresponding passenger distribution on available routes. One model uses a linear distribution model to estimate passenger route choices. The other model uses an integrated simulation framework to approximate a passenger distribution according to the logit model, a model commonly used in route choice. We compare the two new approaches with three timetabling approaches without multi-route search and a heuristic approach on a set of artificial instances and a partial network of Netherlands Railways (NS). Our experiments provide insights into the impact of considering multiple routes instead of a single route, and of integrated route choice instead of predetermined route assignment on the solution quality.

Chapter 4: Hartleb, Schmidt, Huisman, and Friedrich: "Modeling and solving line planning with integrated mode choice". Currently under review at a scientific journal

In Chapter 4, we present a mixed-integer linear program (MILP) for line planning with integrated mode and route choice. The model aims at finding line plans that maximize the profit for the public transport operator while estimating the corresponding passenger demand with choice models. Both components of profit, revenue and cost, are influenced by the line plan. More lines result in higher costs but also increase the level of service to passengers, which leads to higher passenger numbers and more revenue. Hence, the resulting line plans are not only profitable for operators but also attractive to passengers. The passengers' mode and route choices depend on the utility of the service, which includes travel time, number of transfers, and frequency of service. By suitable preprocessing of the utilities, we are able to apply any choice model for mode choices in a MILP. In contrast to existing approaches, the mode and route decisions are modeled according to the passengers' preferences while commercial solvers can be applied to solve the corresponding MILP. We provide and test means to improve the computational performance. In experiments on the Intercity network of the Randstad, a metropolitan area in the Netherlands, we show the benefits of our model compared to a standard line planning model with fixed passenger demand. Furthermore, we demonstrate with the help of our model the possibilities and limitations for operators when reacting to changes in demand. The results suggest that operators should regularly update their line plan in response to changes in travel demand and estimate their passenger demand during optimization.

Chapter 5: Hartleb, Friedrich, and Richter: "Vehicle Scheduling for On-demand Vehicle Fleets in Macroscopic Travel Demand Models". Accepted for publication at Transportation. An early version of this paper is published as Hartleb et al. (2021a).

The planning of on-demand services requires the formation of vehicle schedules consisting of service trips and empty trips. Chapter 5 presents a heuristic algorithm for building vehicle schedules that uses time-dependent demand as input and determines vehicle routes and the number of required vehicles as a result. The presented approach is intended for long-term, strategic transport planning. For this purpose, it provides planners with an estimate of vehicle fleet size and distance traveled by on-demand services. The algorithm can be applied to integer and non-integer demand matrices and is therefore particularly suitable for macroscopic travel demand models. An implementation of the algorithm is available online (Hartleb et al., 2020). We illustrate in two case studies potential applications of the algorithm and feature that on-demand services can be considered in macroscopic travel demand models.

Chapter 6: Hartleb and Schmidt: "A Rolling Horizon Heuristic with Optimality Guarantee for an On-Demand Vehicle Scheduling Problem". Published as Hartleb and Schmidt (2020).

In Chapter 6, we consider the same vehicle scheduling problem as in Chapter 5, which arises in the context of travel demand models. Given demanded vehicle trips, what is the minimum number of vehicles needed to fulfill the demand? In this chapter, we model the vehicle scheduling problem as a network flow problem. Since instances arising in the context of travel demand models are often so big that the network flow model becomes intractable, we propose using a rolling horizon heuristic to split huge problem instances into smaller subproblems and solve them independently to optimality. By letting the horizons of the subproblems overlap, it is possible to look ahead to the demand of the next subproblem. We prove that composing the solutions of the subproblems yields an optimal solution to the whole problem if the overlap of the horizons is sufficiently large. Our experiments show that this approach is not only suitable for solving extremely large instances that are intractable as a whole, but it is also possible to decrease the solution time for large instances compared to solving them as a whole.

Contributions

The main contribution of Chapters 2 to 6 is fivefold. First, we show in an empirical comparison in Chapter 2 that simplifications of passenger modeling can lead to different results. Furthermore, we identify which components of evaluation functions are crucial for the result for the example of timetable evaluation.

Second, we present novel optimization approaches for the design of public transport services with integrated passenger choice models. In Chapter 3 we estimate the passengers' route choice during timetabling in two different ways. First, we develop a linear distribution model resembling the characteristics of the targeted choice model and, second, we use a simulation framework to approximate it. With these two representations of the choice model, we develop two mixed-integer linear programs for timetabling. In Chapter 4, we estimate the passengers' route and mode choice during line planning. By preprocessing the utilities of routes and modes, we design a mixed-integer linear program for this problem.

Third, we develop solution algorithms to solve extremely large vehicle scheduling problems as they arise during demand estimation with a travel demand model. In Chapter 5, we present an efficient heuristic approach to estimate the fleet size of an on-demand service. For the same underlying vehicle scheduling problem, we develop another solution algorithm approach in Chapter 5. This algorithm is based on a rolling horizon framework with overlapping horizons and we provide an optimality guarantee for the solutions if the horizons overlap sufficiently.

Fourth, we test our approaches in experiments on artificial and real-world data. In Chapter 3, the developed timetabling models are compared to four methods motivated by the literature on a grid network and a partial network operated by Netherlands Railways. The line planning approach from Chapter 4 with integrated mode and route choice is tested on the Intercity network of the Randstad, a metropolitan area in the Netherlands. We test the heuristic for vehicle scheduling in Chapter 5 on a travel demand model of the Stuttgart region in Germany, and a model of the campus of the University of Stuttgart to estimate the impact of an electric scooter sharing system. The advantages of the rolling horizon heuristic are demonstrated on a set of randomly generated instances in Chapter 6.

Fifth, we provide valuable insights for public transport operators and public authorities generated with the developed models and methods. In Chapter 2, we show which components of timetabling evaluation functions are determining for differences in evaluation results. The other way around, this indicates how evaluation functions can be simplified without distorting the evaluation results. This is especially relevant for the design of timetables with Operations Research models where evaluation functions often have to be simplified to serve as an objective function in a tractable model. The line planning model with integrated mode choice in Chapter 4 stresses the importance of considering passenger behavior during public transport design. Our results show that modeling the choices of passengers during optimization yield line plans that are more profitable for operators and that have a higher level of service for passengers. Furthermore, our experiments show that the operators' profit is sensitive towards changes in total travel demand. This suggests that they should adapt their services regularly to maximize their profit. In Chapter 5, our experiments give insights into how the use of autonomous fleets affects the required fleet size and the vehicle distance traveled. Both are relevant figures for both operators as well as public transport authorities in estimating the impact of on-demand services.

1.5 Research statement

Chapter 5 is the result of joint work with Emely Richter from the University of Stuttgart. The author is responsible for algorithm development and implementation, as well as manuscript preparation. Emely Richter's contributions to this chapter are the integration of the developed algorithm into a travel demand model, analysis of results, and manuscript preparation. The research in all chapters except Chapter 5 was primarily conducted by the author of this thesis. There, the author is responsible for research design, modeling, and analysis of the results. The research questions in all chapters were developed and defined in fruitful discussions with the respective co-authors. Frequent discussions with and critical feedback from my doctoral advisors Marie Schmidt, Dennis Huisman, and Markus Friedrich greatly helped to improve the quality of the research in all chapters.

Part I

Public transport design considering travel demand

Chapter 2

A good or a bad timetable: Do different evaluation functions agree?

2.1 Introduction

When providing public transport, operators should aim for the highest possible quality from the passengers' viewpoint, respecting physical and monetary constraints. However, there are many different definitions for 'quality from the passengers' viewpoint'. The literature on public transport planning, both from Transport Engineering and Operations Research perspectives, as well as practitioners in railway companies, have come up with very different measures to evaluate quality. These range from very basic measures designed to be used in linear programming frameworks to sophisticated multi-variable models optimized to fit observed passenger behavior as well as possible.

In this chapter, we investigate the following question: Considering a situation characterized by demand for public transport, and different public transport services provided to satisfy this demand, to what extent do different evaluation functions agree on the quality of the provided transport services? That is, will the evaluation functions considered - all designed to measure 'quality from the passengers' perspective' - lead to the same evaluation of what is a good or a bad timetable?

We give an overview of different evaluation functions for timetables proposed in the literature and identify three components in which the functions differ from each other. Based on this, we classify the considered evaluation functions and design a set of representative evaluation functions that are different in the three components. These functions represent a wide range of the most commonly used evaluation functions in mathematical models, evaluation applications, or choice models. Moreover, their modular structure as a combination of the three components allows a purposeful analysis of their similarity.

To empirically compare these representative functions with each other and analyze how similar they are, we conduct three case studies. In each case study, we evaluate a set of timetables for a given demand situation with each of the representative evaluation functions. Two sets of timetables are defined for an artificial grid network and one set is defined for the real-world network of Netherlands Railways (NS). Since the sets of timetables are designed by different parties with varying methods and various objectives, the comparison of the functions should not be biased by the way the timetables were created.

Based on the resulting evaluation values of all timetables with respect to each evaluation function, we develop a method to quantify the degree to which the different evaluation functions coincide. The result of the method allows a pairwise comparison of the evaluation functions and can be interpreted as a measure of inconsistency, which we investigate in two ways. First, the pairwise inconsistency is interpreted directly, visualized with the help of heat maps and multidimensional scaling. This gives an overview of the extent of inconsistency between the evaluation functions and allows an immediate recognition of patterns of which evaluation functions are more or less consistent with each other. Second, we use cluster analysis to determine the strongest inconsistencies between the functions. The cluster analysis identifies groups of evaluation functions that are consistent while the evaluation functions in different groups are less consistent.

With this setting, we aim at empirically testing whether timetable evaluation functions agree on what is a good or bad timetable and to what extent they are consistent. In particular, it is not the purpose of the analysis to identify a 'good' or 'best' evaluation function. Instead, with the modular design of the timetable evaluation functions, we intend to identify which components of the functions have the most influence on differences in the evaluation results. Our intrinsic motivation is to show that the formulation of evaluation functions is crucial for the result of the evaluation. Furthermore, by identifying key components of the evaluation functions we want to provide information about which part to focus on when designing simple yet purposeful objective functions for Operations Research models.

The contribution of this chapter is twofold. First, we use a novel and structured method to compare multiple evaluation functions. In contrast to existing comparisons in literature, this is an empirical method that quantifies the difference between evaluation functions. Since the method is independent of the structure of the evaluation functions, it can be applied to empirically compare evaluation functions in other applications as well. Second, we provide a thorough comparison of timetable evaluation functions for passengers. Our analysis shows to what extent evaluation functions agree on what is a good or a bad timetable. Furthermore, we are able to identify whether and under which circumstances a component of a sophisticated evaluation function is crucial for the result of an evaluation. This can be used to either justify the simplifications made in current state-of-the-art optimization approaches to public transport planning or to point out which aspect is still lacking and needs to be incorporated to obtain objective functions providing a valid evaluation.

The remainder of this chapter is organized as follows. Section 2.2 gives an overview of evaluation functions that are commonly used to measure the quality of public transport from the passengers' viewpoint. Afterward, in Section 2.3, we structure the evaluation functions used in the literature and define a set of representative evaluation functions which we use for the analysis in this research. In Section 2.4, we describe the data which we use in the case studies. Section 2.5 introduces a novel measure of the inconsistency of evaluation functions and gives insight on the used method for comparison. We report on the main findings of our experiments in the same section. In Section 2.6, we demonstrate how the results can be used for the design of an evaluation function and conclude in Section 2.7.

2.2 Literature on evaluation functions

Naturally, research concerned with the design of public transport also deals with the corresponding evaluation. There are various evaluation functions proposed in different research areas. Since we focus on the evaluation of public transport from the passengers' point of view, we restrict ourselves to these evaluation functions. An overview of the most important factors of influence for timetable evaluation for passengers is given by Parbo et al. (2016). We consider only the planned case and neither disruptions nor robustness measures are considered, following the motto that "time savings are the single most important benefit of transport improvement projects all over the world" (de Dios Ortúzar and Willumsen, 2011). In this section, we give an overview of different evaluation functions for timetables structured by the different components of timetable evaluation.

2.2.1 Types of evaluation functions

First, there exist many different ways to evaluate public transport. These differ from each other in the incorporated characteristics and the structure of the functions. We distinguish between two principally different types of evaluation functions, where each of them can appear in different variations.

On the one hand, most commonly used are travel time-based evaluation functions. This is the default way of evaluation in both the research areas of Operations Research and Traffic Engineering. The key idea is to quantify the quality of public transport for passengers by a travel time equivalent. Travel time-based evaluation functions are typically linear functions of the passengers' travel time, but they vastly differ in the number and kind of incorporated characteristics (Hensher and Button, 2007). In Operations Research, timetabling models are mostly based on the periodic event scheduling problem introduced by Serafini and Ukovich (1989) and often use the absolute time passengers spend in public transport for evaluation, see for example Corman et al. (2017). In advanced evaluation functions, the travel time is usually "subdivided into walking time, waiting time, time on vehicle, transfer time, and concealed waiting time" (Flyvbjerg et al., 1986). Furthermore, travel time-based evaluation functions often take more influential factors into account, among them fare, frequency, or temporal spread of the connections offered to passengers. In this case, they are mostly referred to as perceived travel time, generalized cost, or disutility. Sometimes, also preferred departure or arrival times of passengers are modeled by penalizing early or late departures or arrivals. Kanai et al. (2011) considered late departures to be equivalent to waiting times for transfers and Robenek et al. (2016) introduced additional variables and penalty terms for the modeling of departure time preferences.

A comprehensive overview of generalized cost as evaluation functions can be found in de Dios Ortúzar and Willumsen (2011). Both in research and practice, the generalized costs are commonly used for evaluation purposes, although for a long time there have been many publications recommending to stop using them to evaluate the quality of timetables from the passengers' point of view. For instance, Grey (1978) discussed five aspects of why the generalized cost is unsuitable for evaluation, all following the same argument that depicting peoples' variety of perceptions in a single variable leads to an inaccurate representation.

On the other hand, we consider utility-based evaluation models that are mainly known from research in choice modeling. The difference to travel time-based evaluation models is that the evaluation value is not a travel time equivalent but follows the concept of passenger supplement. That means, each reasonably good available connection for passengers adds to their utility and thus improves the quality of the service. A comprehensive overview of utilities of alternatives is given in Ben-Akiva and Lerman (1985). Utility-based evaluation functions are still almost exclusively found in choice modeling, although several publications proposed to employ them for evaluation purposes as well. For example, de Jong et al. (2007) concluded that the 'logsum', a utility-based evaluation function, is well suited for evaluation and a probable reason for their little success is the seemingly complex theory behind it - in contrast to travel time-based evaluation functions.

2.2.2 Passenger distribution

Second, the assumed passenger distribution model is crucial for the evaluation. To evaluate the quality of a public transport service in a meaningful way, it is important to estimate how passengers will use it, that means, it is important to estimate how passengers distribute over available connections. The applied passenger distribution models in timetable-related research range from very simple assumptions to highly developed choice models. In Operations Research, it is often assumed that passenger routes are known before the timetable is fixed and most publications use a priori fixed passenger loads on the connections (Liebchen, 2018; Nachtigall, 1998). Recently, there is a change in the timetabling literature visible with more publications focusing on an integrated timetable-dependent passenger distribution. Since the connections passengers choose are not always reliably determinable beforehand, Sels et al. (2011) and Parbo et al. (2014) described an iterative approach for passenger assignment on shortest routes and timetable optimization. In further publications, the shortest path search was included in timetabling models. Schmidt and Schöbel (2015b) did that for the aperiodic case, Borndörfer et al. (2017) for the periodic case and Gattermann et al. (2016) also for the periodic case using a satisfiability formulation instead of a periodic event scheduling formulation. In these cases, the total travel time of all passengers on their shortest connections is evaluated, instead of the travel time on a previously defined connection. While it is often assumed that passengers only use a single route for each origin-destination pair, some timetabling papers specifically focus on a passenger distribution on multiple routes. For instance, Sels et al. (2015) described a passenger assignment to multiple available routes that are of reasonable quality for passengers. We are not aware of an integrated search for multiple routes, most probably due to the high complexity of such a model.

In contrast to that, research in Traffic Engineering primarily applies passenger distribution models including multiple routes for passengers. Depending on the preferred departure or arrival times, the rooftop model assigns passengers to a connection with shortest travel time (Guis and Nijënstein, 2015). The preferred departure or arrival times of passengers are assumed to be known. Indeed, van der Hurk et al. (2014) show in an analysis of smart card data from the Dutch railway network that many frequently found assumptions on route choice behavior do not hold in general. To also include unobserved preferences of passengers, choice models like the probit (Yang and W. Lam, 2006) or logit model are commonly applied for passenger route choice. The theory of choice models is explained in Ben-Akiva and Lerman (1985) and an overview of choice models suited for passenger route choice in transit networks can be found in de Dios Ortúzar and Willumsen (2011). It seems that the logit model is capable of depicting the passenger behavior best and it is therefore found most regularly. Friedrich et al. (2001) designed an efficient algorithm based on the logit model to compute passenger distributions in public transport networks. Recently, Espinosa-Aranda et al. (2018) proposed a new formulation with an estimation of a constrained nested logit model for connection choice in public transport. The successful application of the logit model is not limited to connection choice, C.-H. Wen et al. (2012) show that it is, for example, well suited to capture passenger behavior in mode choice as well.

2.2.3 Passenger preferences

Third, the evaluation of public transport services should be suited to the target group, that is the passengers. Therefore, the passengers' preferences must be reflected in the evaluation functions. This is commonly achieved by the use of parameters to tune the evaluation functions. Wardman and Toner (2018) showed in their analysis for the case of the generalized cost that choosing the correct parameters is essential for a correct evaluation. While research in Operations Research focuses on developing algorithmic methods to compute timetables and mainly uses given or estimated parameters, there is much research in Traffic Engineering and choice modeling on parameter identification.

Usually, the parameters are found by either stated preference or revealed preference approaches. In the first case, people are asked to make decisions in a survey and their theoretical choice is used to derive rules for passenger behavior. For example, Bradley and Gunn (1990) determined the value of travel time of the Dutch population by a stated preference survey. In the case of revealed preference studies, the actual decisions of passengers are generalized. Recently, with more data being available, more publications analyze passenger behavior with revealed preferences approaches. For example, Kusakabe et al. (2010) estimated passenger usage patterns from smart card data.

The most important parameters for public transport evaluation are of two different kinds, modeling passenger preferences and passenger behavior. The preference parameters specify how the different components of the passenger's journey are weighted. Different components include, but are not limited to waiting time, in-vehicle time, or transfer time. Dell'Olio et al. (2010) provides passenger preference parameters measured from a bus transport service and Schittenhelm (2013) lists preferences of passengers of the Copenhagen S-train. A collection of multiple parameter settings found in various publications is published in Wardman (2004). Part of the passenger preferences but usually researched individually is the value of time. Many publications determine values under certain conditions, see for example Wardman et al. (2012), and Mackie et al. (2001) study the circumstances and ways travel time values should be used within an evaluation. Parameters for passenger behavior refer to the parameters used in the passenger distribution model, for example, the logit parameter. The importance of correct parameter modeling for logit models is stressed in Swait and Louviere (1993).

2.2.4 Comparison of evaluation functions

Although there are various approaches to evaluate public transport from the passengers' point of view, there is only limited research comparing different evaluation functions. Most publications undertaking a comparison of evaluation functions compare only two evaluation functions, a newly introduced function and the state of the art. Usually, the purpose is either to illustrate the merits of the newly introduced evaluation function, as it was done in the previously discussed integrated shortest path search (Borndörfer et al., 2017; Gattermann et al., 2016; Schmidt and Schöbel, 2015b), or to better fit the evaluation to reality. As an example for the latter, de Jong et al. (2007) showed that in their case study a logsum based evaluation should be preferred to the currently applied evaluation since it is more precise in computing passenger surplus when changing the public transport service. Some publications undertake a comparison of multiple evaluation functions, however, these are limited to a theoretical comparison. For example, Parbo et al. (2016) provides a literature review on public transport evaluation and focuses on the conflict of passenger's versus operator's focus. We are not aware of an empirical comparison of public transport evaluation functions or of an investigation of their inconsistency, which are the topics of this chapter.

2.3 Timetable evaluation

From the literature review, it is apparent that researchers and practitioners have come up with many different evaluation functions to measure the quality of timetables from the passengers' viewpoint. We classify this multitude of different functions to define a set of representative evaluation functions. Each evaluation function is treated as a composition of a quality measurement, a passenger distribution model, and assumptions on passenger preferences and behavior. In this section, we explain how these three components are modeled and design a set of 16 evaluation functions to represent the evaluation functions in use.

For this purpose, we define terms that are important for the design of evaluation functions. All variables introduced are summarized in Appendix 2.A. Passenger demand is specified by a set of *origin-destination (OD) pairs* OD, where each of them is a directed pair of stations in the public transport network with time-dependent demand. We consider disjoint *time slices* $t \in T$ of one hour and define the hourly demand of passengers that want to depart in time slice $t \in T$ for each OD pair to be o_{od}^t. The sum of all hourly demand equals the daily demand o_{od} of each OD pair, i.e., $\sum_{t \in T} o_{od}^t = o_{od}$. To meet the demand of passengers, each timetable offers connections to the passenger. We use the term *connection* to denote a time-bound route for passengers using public transport services and denote a set of reasonable connections for each OD pair od with preferred departure time slice t by C_{od}^t. To evaluate timetables, we follow the usual approach to measure and aggregate the quality of available connections for passengers.

2.3.1 Quality measurements

We quantify five characteristics of a connection c as listed in Table 2.1. These characteristics are important factors of influence for a passenger's decision whether to travel on a connection c or not. Note, that we do not take the fare of connections into account. We assume a fare system where the fares depend on origin and destination only, as used, e.g., at Netherlands Railways (NS), the largest Dutch railway

IVT(c)	In-vehicle time	The time spent in public transport vehicles
WKT(c)	Walk time	The time spent walking between platforms for a transfer
TWT(c)	Transfer wait time	The time spent at a station waiting for the next connecting public transport vehicle
NTR(c)	Number of transfers	The number of transfers in the connection
DEP(c)	Departure time	The departure time at the origin

Table 2.1: Characteristics of connections

operator. Consequently, in such a system the ticket price for each OD pair *od* is constant and does not affect the attractiveness of connections.

Based on these five measured characteristics we define four quality measurements to represent the evaluation functions used in the literature. First, the *absolute travel time* ATT is defined as the total time spent in public transport from embarking the first vehicle until alighting the last vehicle during a passenger's connection,

$$\mathrm{ATT}(c) \coloneqq \mathrm{IVT}(c) + \mathrm{WKT}(c) + \mathrm{TWT}(c).$$

Second, the *perceived journey time* PJT applies a weighting of travel times of the different trip segments and includes a penalty for transfers,

$$\mathrm{PJT}(c) \coloneqq \mathrm{IVT}(c) + \alpha_{\mathrm{WKT}} \cdot \mathrm{WKT}(c) + \alpha_{\mathrm{TWT}} \cdot \mathrm{TWT}(c) + \alpha_{\mathrm{NTR}} \cdot \mathrm{NTR}(c). \tag{2.1}$$

As a weighted sum of travel times, the perceived journey time can be interpreted as a time equivalent expressing how long the public transport journey feels to passengers. With the coefficients $\alpha_{\mathrm{WKT}}, \alpha_{\mathrm{TWT}}$ and α_{NTR} it is possible to model passenger preferences.

Third, the *adapted journey time* AJT generalizes the perceived journey time by additionally considering departure time preferences of passengers. To model these preferences, we introduce the *adaption time* $\mathrm{ADT}^t(c)$ as the time a passenger has to deviate from their preferred departure time slice t to take connection c departing at $\mathrm{DEP}(c)$. The adaption time is further explained in Appendix 2.B. Including the adaption time, it is possible to model the impact of access time and the spread of available connections in the evaluation. We define the adapted journey time of a connection c for all passengers with preferred departure time slice t by

$$\begin{aligned} \mathrm{AJT}^t(c) \coloneqq \mathrm{IVT}(c) + \alpha_{\mathrm{WKT}} \cdot \mathrm{WKT}(c) + \alpha_{\mathrm{TWT}} \cdot \mathrm{TWT}(c) \\ + \alpha_{\mathrm{NTR}} \cdot \mathrm{NTR}(c) + \alpha_{\mathrm{ADT}} \cdot \mathrm{ADT}^t(c). \end{aligned} \tag{2.2}$$

This number quantifies how unattractive a certain connection is perceived by a passenger who wants to start traveling in time slice t.

We denote the passenger preferences by $\alpha \coloneqq (\alpha_{\mathrm{WKT}}, \alpha_{\mathrm{TWT}}, \alpha_{\mathrm{NTR}}, \alpha_{\mathrm{ADT}})$. Note, that for $\alpha = (1, 1, 0, 0)$ the adapted journey time equals the absolute travel time ATT, and for $\alpha = (\alpha_{\mathrm{WKT}}, \alpha_{\mathrm{TWT}}, \alpha_{\mathrm{NTR}}, 0)$ the adapted journey time equals the perceived journey time PJT.

	IVT	WKT	TWT	NTR	ADT
absolute travel time ATT	1	1	1	0	0
perceived journey time PJT	1	α_{WKT}	α_{TWT}	α_{NTR}	0
adapted journey time AJT	1	α_{WKT}	α_{TWT}	α_{NTR}	α_{ADT}
evaluated total utility ETU	*	*	*	*	*

Table 2.2: Each entry indicates which of the five characteristics (in-vehicle time IVT, walk time WKT, transfer wait time TWT, number of transfers NTR and adaption time ADT) are taken into account in the four quality measurements ATT, PJT, AJT and ETU. Linear dependencies are indicated by coefficients, non-linear by asterisks.

Fourth, there also exist utility-based evaluation functions in the literature that are derived from choice models. To represent these functions, we consider the *evaluated total utility* (ETU) of a connection as a number expressing how useful a connection is to a passenger with preferred departure time slice t. We define the evaluated total utility of a connection c to be

$$\text{ETU}^t(c) = e^{-\beta \cdot \text{AJT}^t(c)}, \tag{2.3}$$

based on the definition of the logit model as a passenger distribution model. The logit model and its associated parameter β are explained in detail in Section 2.3.3.

We refer to the four characteristics ATT, PJT, AJT and ETU as *quality measurements*. While the first three quality measurements are travel time equivalents, the evaluated total utility is a utility-based evaluation function, where each reasonably good connection for passengers adds to the utility and thus improves the quality of the service. Hence, we call ATT, PJT and AJT *travel time-based*, and ETU *utility-based* quality measurement.

Table 2.2 gives a summary and theoretical comparison of the four quality measurements. If a characteristic is included linearly in a quality measurement, the table shows the coefficient, if the dependency is non-linear, it is indicated by an asterisk whether the characteristic is taken into account.

The goal is to design evaluation functions for a timetable but, so far, just quality measurements for connections were defined. For evaluating a timetable, we aggregate the quality measurements of connections to derive a quality measurement for the whole timetable. To achieve this, we follow the approach commonly used in literature, divided into two steps. First, we aggregate the values of the quality measurement over all connections in C^t_{od} and over all time slices in T. The values are weighted by the demand o^t_{od} per time slice and the probability $p^t(c)$ that a connection is chosen which yields a quality measurement for each OD pair. Second, the values of quality measurements for OD pairs are averaged, weighted by their volume o_{od}, to obtain a quality measure for the whole timetable. The exact procedure used in our experiments with examples is described in Appendix 2.C.

2.3.2 Passenger distribution

The decision which connections passengers choose is dependent on the characteristics of the connections. There are two fundamentally different approaches for passenger distribution used in the literature. While research in Operations Research often assumes that all passengers travel on the shortest connection available, most publications from other research areas apply more realistic passenger distribution models when evaluating timetables. To investigate this difference, we consider two passenger distribution models.

On the one hand, we rely on the logit model to obtain a realistic distribution of the passengers on multiple connections (mc). We assume a set C_{od} of reasonably good connections for each OD pair od to be given. Then, the logit model can be interpreted as a function assigning a probability $p^t(c)$ to each connection $c \in C_{od}$ that is used by passengers with preferred departure time slice t. The logit model is defined by

$$p^t(c) = \frac{e^{-\beta \cdot \mathrm{AJT}^t(c)}}{\sum_{c' \in C_{od}} e^{-\beta \cdot \mathrm{AJT}^t(c')}} \qquad \forall c \in C_{od}, \tag{2.4}$$

where the parameter $\beta \in \mathbb{R}_{\geq 0}$ is used to adjust the model to a specific case study (Ben-Akiva and Lerman, 1985). Note, that the choice set of connections C_{od} is independent of the passengers' preferred departure time slice t. Since the logit model is based on the adapted journey time of all considered alternative connections, only connections departing in or close to the time slice t will be assigned a probability that is significantly larger than 0.

On the other hand, we consider a shortest connection (sc) strategy for the passengers. That means, passengers only take connections with lowest journey time departing within or close to their preferred departure time slice. Let C^t_{od} be the set of all connections with lowest adapted journey time for passengers of an OD pair od that want to depart in time slice t. Then, the share of passengers using connection $c \in C^t_{od}$

is

$$p^t(c) = \frac{1}{|C^t_{od}|} \qquad \forall c \in C^t_{od}.$$

That means, in case there are multiple shortest connections available, we assume that the passenger distribution on them is uniform. We want to point out that in this modeling both the probability p and the set of available connections C depend on the applied choice model.

2.3.3 Passenger preferences and behavior

We take different assumptions on passenger preferences and passenger behavior into account. To begin with, the definitions of perceived and adapted journey time as well as evaluated total utility in Equations (2.1), (2.2) and (2.3) depend on passenger preferences. The values of the coefficients $\alpha \in \mathbb{R}^4_{\geq 0}$ indicate how important in-vehicle time, walk time, transfer wait time, number of transfers, and adaption time are relative to each other to the passenger.

In addition, it is possible to adjust the logit model with the coefficient $\beta \in \mathbb{R}_{\geq 0}$ in Equation (2.4) to fit passenger behavior. This value indicates how sensitive passengers are to absolute differences in the adapted journey time of connections. For example, for $\beta = 0$ all connections in the choice set will be used by passengers equivalently and the logit model reduces to a uniform distribution. The higher the coefficient β, the more passengers will use the connections with lowest adapted journey time. This coefficient also influences the evaluated total utility of a public transport service, as defined in Equation (2.3).

Furthermore, the passengers' tolerance to deviations from their preferred departure times can be adjusted with a scaling parameter $\gamma \in \mathbb{N}$. A given value γ models that passengers prefer to depart in a $\frac{60}{\gamma}$ min time window within their departure time slice t. Hence, high values of γ indicate a low tolerance and vice versa.

To analyze the impact of modeling passenger preferences on the evaluation, we consider two user groups. These are represented by the two parameter settings

$$ps_1 = (\alpha, \beta, \gamma) \text{ with } \alpha = (1, 1, 5, 1),\ \beta = 0.13,\ \gamma = 1 \tag{2.5}$$

and

$$ps_2 = (\alpha, \beta, \gamma) \text{ with } \alpha = (2, 2, 10, 2),\ \beta = 0.22,\ \gamma = 60. \tag{2.6}$$

The first parameter setting models passengers that are mainly focused on journey time ($\alpha_{\mathrm{WKT}} = 1, \alpha_{\mathrm{TWT}} = 1$) and are relatively undeterred by transferring ($\alpha_{\mathrm{NTR}} = 5$). They would also make use of connections with higher adapted journey time ($\beta = 0.13$) and are rather flexible regarding departure time ($\alpha_{\mathrm{ADT}} = 1,\ \gamma = 1$), as long as connections are fast.

The second parameter setting models passengers that are more convenience-oriented. They prefer a public transport service that is suited to their needs with less and short transfers ($\alpha_{\mathrm{WKT}} = 2, \alpha_{\mathrm{TWT}} = 2, \alpha_{\mathrm{NTR}} = 10$), preferably use connections with low adapted journey time ($\beta = 0.22$) and are inflexible regarding their desired departure time ($\alpha_{\mathrm{ADT}} = 2,\ \gamma = 60$).

The parameters are chosen following recommendations from research and practice. For example, as of 2012, NS used a penalty of 10 min for each transfer (De Keizer et al., 2012). Wardman (2004) provides a thorough study of values of time, among them several values for the wait and walk time compared to in-vehicle time are listed. Usually, the coefficients for wait and walk time are around 2. The logit parameter β should be adjusted for each case study, but experience has shown that values of $\beta \in [0.13, 0.22]$ are a reasonable choice if minutes are used as time units. Values for the adaption time are chosen to fit the characteristics of the user groups modeled by the two parameter sets in Equations (2.5) and (2.6).

2.3.4 Evaluation functions

We define an *evaluation function* as a combination of a quality measurement, a passenger distribution model and an assumption on passenger preferences. That means, applying an evaluation function consists of two steps: Given a timetable with a connection choice set for passengers, the passengers are first distributed on the connections according to the distribution model and their preferences. Second, the quality of the timetable is evaluated with respect to the quality measurement, again using the passenger preferences. Many publications focus only on the second step when describing their evaluations. However, we believe that the distribution is an integral component of the evaluation that influences the evaluation results. Hence, we also investigate the extent of this influence.

When combining the four quality measurements defined in Section 2.3.1 with the two distribution models described in Section 2.3.2 and the two different assumptions on passenger preferences fixed in Section 2.3.3, we obtain 16 evaluation functions in total.

	sc	*mc*
ATT	Borndörfer et al. (2017)	Parbo et al. (2014)†
PJT	Wardman and Toner (2018)	Parbo et al. (2014)†
AJT	Kanai et al. (2011)	Robenek et al. (2016)
ETU	‡	de Jong et al. (2007)§

Table 2.3: Examples for the use of different evaluation functions in recent literature. We provide one publication for each cell, exemplifying the use of a quality measurement (absolute travel time ATT, perceived journey time PJT, adapted journey time AJT and evaluated total utility ETU) in combination with a shortest connection (*sc*) or multiple connection (*mc*) passenger distribution model

† Used ATT in evaluation and PJT in distribution

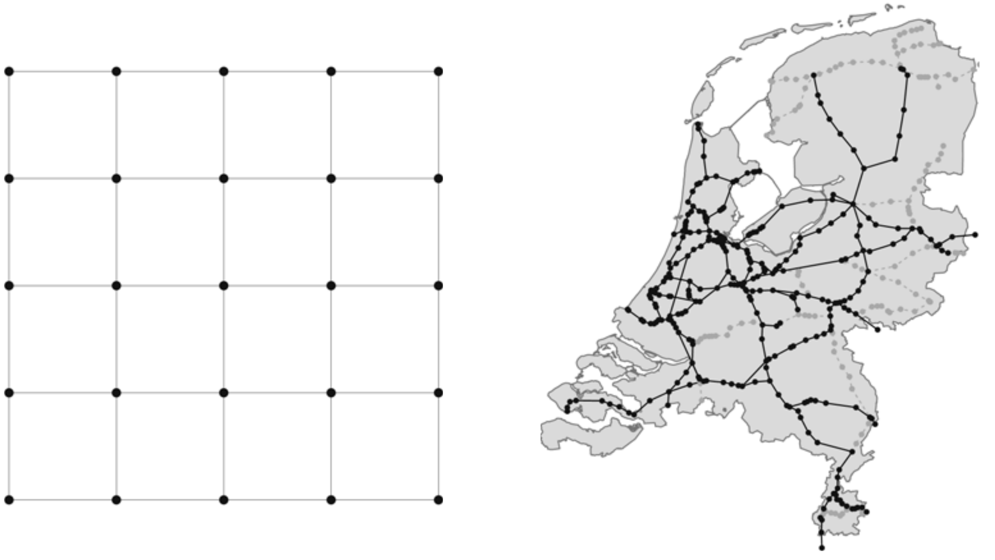
‡ ETU in combination with *sc* is not used since ETU does not require a passenger distribution

§ Used a slightly different utility-based evaluation function

This design of evaluation functions entails two advantages. First, these functions cover a wide range of commonly used evaluation functions in mathematical models, evaluation applications, and choice models as is indicated in Table 2.3. Second, their modular structure as a combination of quality measurement, distribution model, and assumptions on passenger preferences allows a purposeful analysis. Differences or similarities of evaluation functions can easily be traced down to components of the functions. We denote the set of the 16 evaluation functions by F.

2.4 Case studies

Our goal is to analyze how inconsistent the 16 different evaluation functions are by comparing their evaluation behavior on multiple public transport services. In this section, we describe three case studies in which we perform these evaluations. Each *case study* is characterized by a fixed public transport infrastructure, a demand situation on that infrastructure, and a set of services supplying the demand. A *public transport infrastructure* consists of stations and direct links between them and a *demand situation* is specified by a set of origin-destination (OD) pairs OD, where each of them is a directed pair of stations with time-dependent demand. For this demand situation, we consider several public transport services supplying this demand, for comparison. Each *public transport service* is formalized by a line plan and a timetable which together determine the potential connections and their quality. The procedure of how we derive connection choice sets is described in Appendix 2.D.



(a) Grid infrastructure

(b) Dutch railway infrastructure. Tracks and stations in black are operated by NS

Figure 2.1: The evaluation functions are compared on these two infrastructures

2.4.1 Case studies on a grid infrastructure

As a first infrastructure, we use an artificial 5×5 grid network[1] introduced by the research group FOR2083. The infrastructure consists of 25 stations and 40 direct links as depicted in Figure 2.1a. On this infrastructure, we consider two demand situations with multiple corresponding benchmark services available, each of them consisting of a line plan and a timetable. Both demand situations have an almost complete demand matrix with nearly 600 non-zero entries. Although they share the same infrastructure, we treat them as two different case studies due to the different data structures of demand and supply. The first demand situation has a typical daily demand pattern and 27 suitable services that are operated throughout the whole day. All of these services were designed by traffic engineers with established methods used in transport planning. We refer to the case study as GL. The second demand situation depicts a morning peak and 28 services operating only in the morning hours are available. These services were found with different optimization models by Operations Researchers and we denote the corresponding case study by GS.

[1] `https://github.com/FOR2083/PublicTransportNetworks/tree/master/Grid_5x5`, visited on November 12, 2018.

2.4.2 Case study on the Dutch railway infrastructure

The second infrastructure is the Dutch railway network with roughly 270 stations as it is operated by Netherlands Railways (NS). In Figure 2.1b a route map of the Dutch railway network is shown. The demand is given by a scientific demand set of more than 62000 non-zero OD pairs defined between the stations reflecting a realistic demand situation. For evaluation, we consider the yearly transport services that were operated by NS in the years 2012 till 2018. Note, that due to changes in the infrastructure in the Dutch railway network, not all public transport services are defined on the same network. That means, over the years some stations and tracks might have been introduced or abolished. However, we evaluate all different services with the same demand set between the same stations, therefore the evaluation is not directly affected by the slight changes of the infrastructure. We refer to this case study by NS.

2.5 Comparison of evaluation functions

We defined 16 evaluation functions for public transport services in Section 2.3 and introduced the infrastructures with corresponding demand situation and multiple services for the three case studies in Section 2.4. In this section, we describe a method to compare different evaluation functions and to set them into relation. Using this method, the 16 evaluation functions are investigated for their inconsistency in the three case studies.

The key idea is to compare the evaluation functions when applied to a number of services. We evaluate all public transport services $s \in S$ with each of the evaluation functions $f \in F$ and use the resulting evaluation values v_s^f to compare the functions in F. We evaluate the services with PTV Visum (PTV Group, 2018), a software package for macroscopic traffic analysis and forecasting. The complete results for all three case studies are provided in Appendix 2.E. To explain and demonstrate the used method, we discuss the results of the NS case study.

For the NS case study, Table 2.4 shows the evaluation values of the services operated between 2012 and 2018 for all 16 evaluation functions. At a first glance, all public transport services in Table 2.4 have very similar evaluation values, suggesting that the quality of the services is effectively the same. For example, the absolute travel time on the shortest connection evaluated with the first parameter setting (evaluation function 1) ranges for all seven services between 35.94 and 36.78 minutes, implying

	ATT				PJT				AJT				$100 \cdot$ ETU			
	ps_1		ps_2		ps_1		ps_2		ps_1		ps_2		ps_1		ps_2	
	sc	mc	sc	mc	sc	mc	sc	mc	sc	mc	sc	mc	sc	mc	sc	mc
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
NS12	36.78	37.43	36.79	37.60	37.99	38.90	41.04	42.22	38.64	40.49	58.37	53.30	38.31	55.98	1.29	1.77
NS13	36.30	36.96	36.32	37.13	37.44	38.38	40.27	41.58	38.07	39.97	57.60	52.63	34.19	53.60	1.11	1.70
NS14	36.30	36.94	36.31	37.09	37.44	38.37	40.31	41.60	38.03	39.96	56.88	52.22	36.83	56.50	1.21	1.80
NS15	36.22	36.90	36.24	37.07	37.36	38.33	40.22	41.51	37.98	39.93	56.88	52.01	36.68	56.25	1.20	1.77
NS16	36.23	36.91	36.26	37.06	37.38	38.32	40.24	41.52	37.99	39.92	56.90	51.97	38.28	55.93	1.28	1.75
NS17	36.03	36.77	36.04	36.96	37.25	38.29	40.28	41.67	37.87	39.90	56.85	52.03	40.44	57.67	1.33	1.80
NS18	35.94	36.71	35.95	36.89	37.14	38.22	40.14	41.56	37.78	39.83	56.96	51.75	39.72	59.16	1.31	1.85

Table 2.4: Evaluation values v_s^f in NS case study. Each column corresponds to one evaluation function $f \in F$ and each row to one public transport service s. The name of the services indicate the year in which this service was operated. The four topmost rows show the quality measurement, the used parameter setting and distribution model as introduced in Section 2.3 and lastly an index to identify the evaluation functions. The values for the travel time-based evaluation functions (ATT, PJT, AJT) show average travel time in minutes, the values of the utility-based evaluation functions (ETU) is dimensionless. For ease of exposition, all evaluation values of utility-based evaluation functions are multiplied with 100.

a difference of only 0.84 minutes. While this difference sounds negligible, it actually comprises considerable differences for individual OD pairs. A total gain of 0.84 minutes in absolute travel time corresponds to an improvement of 2.3% and could for example be achieved by decreasing the travel time on all connections of the 20 biggest OD pairs by 10 minutes. This improvement would affect more than 90,000 travelers every day.

Furthermore, Table 2.4 also shows that the best service regarding one evaluation function is not necessarily the best service regarding another evaluation function. For example, the best services regarding evaluation functions 7 and 8 do not coincide. While NS18 provides on average the shortest perceived journey time weighted with the second parameter set on the shortest connection, NS15 yields the shortest perceived journey time on multiple connections, indicating that the passenger distribution model has an influence on the evaluation in this case. Table 2.5 summarizes differences in ranking of all public transport services and all evaluation functions in a 'medal count', indicating how often the respective service is classified on a certain rank.

The highest numbers in Table 2.5 appear on, or close to the antidiagonal. This shows that the evaluation functions essentially agree that the services improved from NS12 to NS18, or equivalently, improved over the years. Taking an average over all evaluations, it seems to be conclusive which service is best. However, not all of the services could be unambiguously classified. Most of the services are ranked over

	1^{st}	2^{nd}	3^{rd}	4^{th}	5^{th}	6^{th}	7^{th}
NS12	0	0	2	0	2	0	12
NS13	0	0	0	2	1	9	4
NS14	0	1	2	0	11	2	0
NS15	1	1	7	5	0	2	0
NS16	0	2	4	8	0	2	0
NS17	3	10	0	1	1	1	0
NS18	12	2	1	0	1	0	0

Table 2.5: 'Medal count' from NS case study showing the number of times a public transport service is ranked on the n^{th} rank. Both row and column sum add up to 16, the number of considered evaluation functions.

a range of five, some even over six ranks. Using just one evaluation function, as it is often done in research, might yield a very different ranking than the average suggests. To draw inferences from this about the inconsistency of the evaluation functions, it is interesting to see whether the deviations in the ranking are due to some random dispersion or whether there is a structural connection between the rankings of evaluation functions.

2.5.1 Inconsistency of two evaluation functions

Even when the differences in the ranking are large, actual evaluation values may be very close to each other. To avoid fallacy when comparing the evaluation functions by rank, we focus on the relative differences in objective values. Since the evaluation values v_s^f depend on the evaluation function and, thus, are not directly comparable, we normalize the evaluation values. These normalized values are in the same number range and can be compared easily.

We define

$$V(f) \coloneqq \max_{s \in S} v_s^f - \min_{s \in S} v_s^f$$

to be the range of objective values of all public transport services with respect to evaluation function $f \in F$. For evaluation functions, for which smaller values are better, we define the normalized value of service $s \in S$ with respect to evaluation function $f \in F$ to be

$$\varphi_s^f \coloneqq \frac{v_s^f - \min_{s' \in S} v_{s'}^f}{V(f)}. \tag{2.7}$$

Equivalently, the normalized value for evaluation functions, for which larger values are better, is defined as

$$\varphi_s^f := \frac{\max_{s' \in S} v_{s'}^f - v_s^f}{V(f)}. \tag{2.8}$$

It is evident that these two definitions are equivalent since the right-hand sides of Equations 2.7 and 2.7 add to 1. The normalized values lie in the unit interval and indicate to what extent service s performs worse than the best service with respect to the same evaluation function considering the range of all other values. Therefore, the normalized values φ depend on the set of all considered services S of a case study.

The normalized values allow a comparison of the quality of public transport services regarding different evaluation functions. To compare the evaluation functions pairwise with each other, we define the *inconsistency* of two evaluation functions f_1 and f_2 as the mean difference in the normalized value, i.e.,

$$i_\varphi(f_1, f_2) := \frac{1}{|S|} \sum_{s \in S} |\varphi_s^{f_1} - \varphi_s^{f_2}|.$$

As the normalized values φ_s^f depend on the set of all considered services S of a case study, also the inconsistency i depends on the set S.

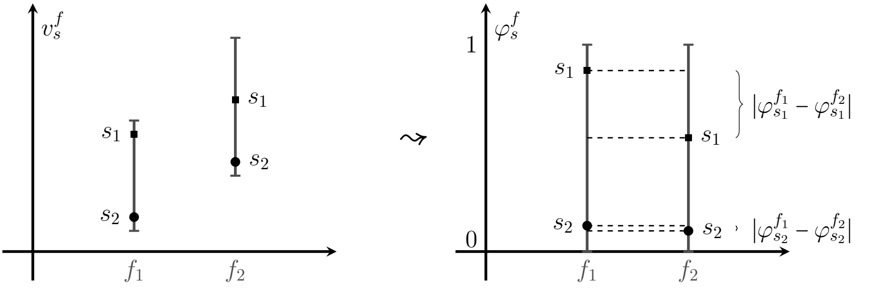


Figure 2.2: Normalization of evaluation values v_s^f for two evaluation functions and indication of computation of inconsistency $i_\varphi(f_1, f_2)$ for two public transport services

The normalization of evaluation values and the definition of the inconsistency as the mean difference in normalized values is depicted in Figure 2.2. The graph on the left shows the ranges of the evaluation values v_s^f of two evaluation functions f_1 and f_2 as vertical lines. On the lines, the evaluation values of two services s_1 and s_2 are marked. As it can be seen in this graph, the two evaluation functions yield different ranges of evaluation values, and therefore it is difficult to compare them. This is dealt with by the normalization of the evaluation values, which is depicted in the graph on the right. Both ranges of the two evaluation functions f_1 and f_2 cover exactly the unit interval and it is possible to compare the normalized evaluation values φ_s^f. This is shown with the same two services s_1 and s_2 from the left graph. It reveals that service s_1 is rated differently by f_1 and f_2 while the two evaluation functions nearly agree on the quality of service s_2. The vertical distance of the normalized evaluation values, averaged over all services, is defined to be the inconsistency of two evaluation functions in a certain case study.

One shortcoming of this approach is that the normalized evaluation values depend on the set of all considered services of a case study. As defined in Equations (2.7) and (2.8), all deviations in objective values between two services are compared relative to the largest differences between any services of the respective case study. That means, in case all services are almost identical in quality, different evaluation functions might be indicated as being inconsistent although they hardly show considerable differences in the evaluation. However, when considering services that do show differences in quality, such an incorrect indication of inconsistency cannot occur.

In the three case studies NS, GS, and GL we derive the pairwise inconsistencies between all 16 evaluation functions. Altogether, we find qualitatively similar results, which means, the inconsistencies of the studied evaluation functions are qualitatively alike in the different case studies. Only for very few pairs of evaluation functions, we observe a qualitative difference in the pattern of inconsistencies between the case studies. This indicates that the results are not dependent on the structure of the case study but indeed on the structure of the evaluation functions. Therefore, we discuss the findings independently of the case studies where this is applicable and just highlight differences in the results.

For a collective discussion we compute the weighted average of the inconsistencies between the evaluation functions over all case studies by

$$i(f_1, f_2) = \frac{\sum\limits_{I \in \{GS, GL, NS\}} |S_I| i_\varphi^I(f_1, f_2)}{\sum\limits_{I \in \{GS, GL, NS\}} |S_I|} \quad \forall f_1, f_2 \in F,$$

where $|S_I|$ is the number of services considered in case study I and $i_\varphi^I(f_1, f_2)$ is the inconsistency of evaluation functions f_1 and f_2 derived in case study I.

For better comprehensibility, the inconsistencies are presented in a heat map, a quadratic 16×16 matrix where each entry displays the inconsistency of two evalua-

	ATT				PJT				AJT				ETU			
	ps_1		ps_2		ps_1		ps_2		ps_1		ps_2		ps_1		ps_2	
	sc	*mc*	*sc*	*mc*	*sc*	*mc*	*sc*	*mc*	*sc*	*mc*	*sc*	*mc*	*sc*	*mc*	*sc*	*mc*
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0,00	3,63	1,37	4,73	7,70	7,52	12,84	12,80	9,72	8,42	17,92	18,58	20,51	19,99	20,12	19,36
2	3,63	0,00	3,49	2,03	8,08	7,33	12,97	12,59	9,59	8,18	17,24	17,83	20,17	20,03	19,55	18,84
3	1,37	3,49	0,00	4,17	7,24	7,14	12,39	12,34	8,97	7,76	17,22	17,86	20,87	20,49	20,55	19,73
4	4,73	2,03	4,17	0,00	7,45	6,81	12,26	12,03	8,57	7,48	16,55	16,72	20,92	21,02	20,38	19,78
5	7,70	8,08	7,24	7,45	0,00	3,32	5,32	5,47	4,01	3,21	13,26	15,99	23,42	23,26	23,47	22,70
6	7,52	7,33	7,14	6,81	3,32	0,00	6,44	5,56	4,48	1,77	12,89	15,94	23,46	23,40	23,13	22,56
7	12,84	12,97	12,39	12,26	5,32	6,44	0,00	3,38	5,99	5,54	11,98	16,89	26,01	25,73	26,02	25,31
8	12,80	12,59	12,34	12,03	5,47	5,56	3,38	0,00	5,78	5,02	11,15	16,66	25,77	25,49	25,69	24,93
9	9,72	9,59	8,97	8,57	4,01	4,48	5,99	5,78	0,00	3,45	10,46	14,48	23,61	23,67	23,57	22,92
10	8,42	8,18	7,76	7,48	3,21	1,77	5,54	5,02	3,45	0,00	12,01	15,21	24,07	23,99	23,79	23,19
11	17,92	17,24	17,22	16,55	13,26	12,89	11,98	11,15	10,46	12,01	0,00	13,16	27,64	28,41	27,23	27,28
12	18,58	17,83	17,86	16,72	15,99	15,94	16,89	16,66	14,48	15,21	13,16	0,00	31,70	31,75	30,73	29,97
13	20,51	20,17	20,87	20,92	23,42	23,46	26,01	25,77	23,61	24,07	27,64	31,70	0,00	3,16	2,37	4,16
14	19,99	20,03	20,49	21,02	23,26	23,40	25,73	25,49	23,67	23,99	28,41	31,75	3,16	0,00	4,35	2,97
15	20,12	19,55	20,55	20,38	23,47	23,13	26,02	25,69	23,57	23,79	27,23	30,73	2,37	4,35	0,00	2,93
16	19,36	18,84	19,73	19,78	22,70	22,56	25,31	24,93	22,92	23,19	27,28	29,97	4,16	2,97	2,93	0,00

Figure 2.3: Heat map showing the weighted average inconsistencies from all three case studies. For better depiction, all values are multiplied with 100.

tion functions. To make differences in inconsistencies easily identifiable, high values are indicated by a dark shading and low values have a light shading. Naturally, all diagonal values of the matrix are zero as each evaluation function is fully consistent with itself and the matrices are symmetric since $i_\varphi(f_1, f_2) = i_\varphi(f_2, f_1)$ holds. The weighted average inconsistencies of the three case studies are tabulated in the heat map in Figure 2.3.

The absolute values of the inconsistencies i allow an interpretation of the extent to which the evaluation functions agree in their assessment of the services. For example, an inconsistency of 19.36% between evaluation functions 1 and 16 can be found in the top right corner of Figure 2.3. This inconsistency implies that the normalized values of all services regarding these two evaluation functions deviate by 19.36% on average. Visualized in Figure 2.2, this would mean that the differences $|\varphi_s^1 - \varphi_s^{16}|$ are on average over all services s approximately one-fifth of the total range of normalized evaluation values.

The heat map in Figure 2.3 shows obvious patterns with dark and bright areas, indicating large and small differences in the inconsistencies between the evaluation functions. To provide a better intuition, we use multidimensional scaling to visualize the inconsistencies in Figure 2.4 as distances between the evaluation functions. That

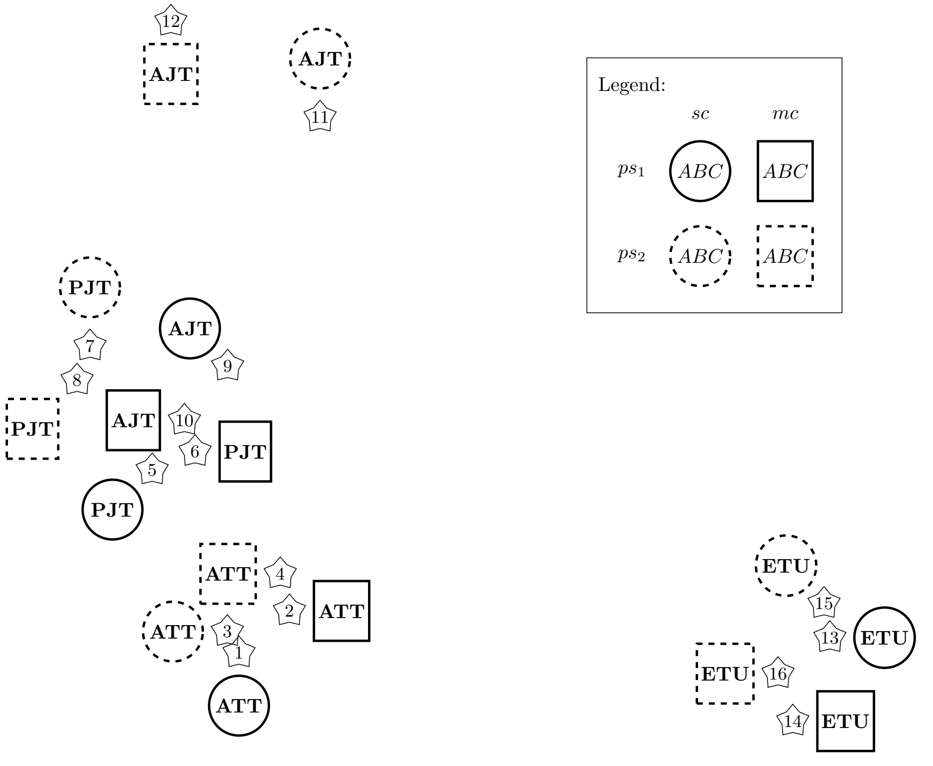


Figure 2.4: The inconsistencies of pairs of evaluation functions visualized as distances on the plane. Each star corresponds to one evaluation function displaying its id. The labels next to the stars explain how the evaluation function is constructed. The quality measurement ATT, PJT, AJT, or ETU is written in the labels. A round label shape indicates that passengers are distributed on the shortest connections (sc), while squared labels indicate the use of a passenger distribution model on multiple connections (mc). The used parameter setting is distinguishable by solid (ps_1) or dashed label edging (ps_2).

means, we depict each evaluation function f as a point $x_f \in \mathbb{R}^2$ on the plane such that the Euclidean distance $d(x_{f_1}, x_{f_2})$ between each two points is representative for the inconsistency $i(f_1, f_2)$ of the corresponding evaluation functions. This is ensured by minimizing the relative deviation of Euclidean distance from the inconsistency, i.e., we solve

$$\min_{x \in \mathbb{R}^{2|F|}} \frac{\sum_{f_1, f_2 \in F} \left(d(x_{f_1}, x_{f_2}) - i(f_1, f_2)\right)^2}{\sum_{f_1, f_2 \in F} i(f_1, f_2)^2}.$$

More on multidimensional scaling can be found in Borg and Groenen (2005). In general, the representation of inconsistencies as distances in Figure 2.4 allows a faster and easier interpretation but all observations can be confirmed with the derived inconsistencies in Figure 2.3.

Observations

It is obvious from Figure 2.4 that the four utility-based evaluation functions are separated from the travel time-based evaluation functions. This is independent of the chosen parameter setting or passenger distribution model. Also, the heat map indicates by a dark shading in the upper right (or equivalently lower left) part that the evaluation functions based on travel time are generally inconsistent with those based on utility. Furthermore, both figures suggest that the utility-based evaluation functions are rather consistent with each other, visible from low distances between pairs of utility-based evaluation functions in Figure 2.4 and also from light shading in the lower right corner of Figure 2.3. The utility-based evaluation functions are especially far from the functions of adapted journey time although ETU and AJT are the only two quality measurements that consider the adaption time besides other characteristics, see Table 2.2. This shows that the shape of an evaluation function is in this case more relevant for the inconsistency than the characteristics it takes into account in the evaluation.

A second group of evaluation functions that are consistent with each other but a bit separate from other groups is formed by the evaluation functions of absolute travel time. By design, this group of evaluation functions is least affected by different parameter settings and therefore it was expected that evaluation functions from this group are relatively consistent with each other. In line with this, a close inspection also shows that in our case studies the passenger distribution model has a higher impact on the inconsistency of evaluation functions of absolute travel time than the parameter setting. The group of evaluation functions of absolute travel time is far from the utility-based evaluation functions and closer to other travel time-based evaluation functions.

The closest group to the evaluation functions of absolute travel time are the four evaluation functions of perceived journey time and the two evaluation functions of adapted journey time with the first parameter setting. Especially with the first parameter setting this closeness is plausible since the first parameter setting is very similar to the fixed parameters of absolute travel time, see Equation (2.5). That the two evaluation functions of perceived journey time with the second parameter setting are a little further away indicates that the penalties for transfers and the weighting of transfer wait time have a measurable effect on the inconsistency of the evaluation functions.

In the top left corner of Figure 2.4 we find the two evaluation functions based on adapted journey time with the second parameter setting, separate from the other evaluation functions and also relatively far from each other. This is also reflected in the inconsistencies in the heat map in Figure 2.3 where both evaluation functions 11 and 12 show fairly high inconsistencies with all other evaluation functions. A plausible explanation for this is the adaption time. The adapted journey time is the only travel time-based quality measurement comprising the adaption time, and with the second parameter setting the adaption is penalized much higher than when using the first parameter setting.

A possible reason for the high inconsistency between the two evaluation functions of adapted journey time with the second parameter setting might be found in the set of services in our case studies; One kind of service provides no reasonably good alternative to the best connection(s) whereas the second kind of service additionally offers such alternatives. The evaluation of these two kinds of services is similar when considering the shortest connection since both offer comparable shortest connections. However, the adaption time in the second kind of service, which provides many comparably good connections for each OD pair, is drastically lower when considering multiple connections which leads to a different rating of the two kinds of services. The presence of both kinds of services in the case studies might account for the visible inconsistency between the two outliers for different passenger distribution models.

To summarize, Figure 2.4 suggests that there are three groups of evaluation functions that are close to each other, but far from functions of other groups. One group is formed by the four utility-based evaluation functions, one by the four evaluation functions of absolute travel time, and one by the evaluation functions of perceived journey time and adapted journey time with the first parameter setting. Additionally, the remaining two evaluation functions of adapted journey time with the second parameter setting seem to be two outliers apart from the three groups.

2.5.2 Cluster analysis

In addition to an investigation of the inconsistencies, we perform cluster analyses of the evaluation functions in each of the three case studies. These help to determine which of the evaluation functions are similar to each other and which are fundamentally different. With the cluster analyses we can, on the one hand, verify the group formation that is apparent in Figure 2.4 and, on the other hand, identify individual variations of the inconsistencies in the different case studies.

The evaluation functions $f \in F$ are clustered based on the normalized evaluation values φ_s^f of all considered services $s \in S$. For a given $k \in \mathbb{N}$, each evaluation function is assigned to exactly one of k clusters such that the sum of all distances between the evaluation functions and their cluster center is minimal. As distance measure between an evaluation function f and a cluster center m we use the rectilinear distance of the normalized evaluation values φ to the cluster center,

$$d(m, f) = \frac{1}{|S|} \sum_{s \in S} |\varphi_s^f - m_s|. \tag{2.9}$$

Note, that this distance $d(m, f)$ is consistent with the definition of the inconsistency $i_\varphi(f, m)$, in the sense that

$$d(f_1, f_2) = i_\varphi(f_1, f_2).$$

The complete mixed-integer program we use to solve the clustering problem is specified in Appendix 2.F. In each case study we cluster the set of 16 evaluation functions F into k clusters, for $k \in \{2, \ldots, 5\}$. Varying the number of clusters k helps to get a better understanding of the inconsistency of evaluation functions.

These 12 clusterings are summarized in Figure 2.5, each clustering represented by lines grouping several points. As before, each point corresponds to one evaluation function and for each cluster of evaluation functions, there is a line surrounding the corresponding points. The thickness of a line depends on the cumulative frequency of appearance of the cluster. Hence, the number and thickness of the lines separating two evaluation functions visualize how often these two functions were separated into different clusters. Note, that in Figure 2.5 the distances between evaluation functions are not representative of the inconsistencies.

Observations

In general, the cluster analysis confirms the observations made from a direct interpretation of the inconsistencies in Figure 2.4. Additionally, it contributes some kind of ranking of which inconsistencies are more substantial.

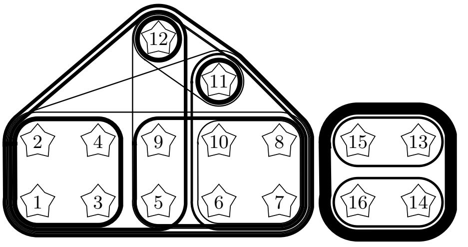


Figure 2.5: The accumulated clusterings of all evaluation functions

It can be seen that the strongest separation is between the utility-based evaluation functions and the travel time-based evaluation functions. In no case study two evaluation functions from the two different bases were found in the same cluster. This gives evidence that the decision of whether to use a travel time-based or a utility-based evaluation is most crucial in this setting. Also within the group of travel time-based evaluation functions, we observe that the visible inconsistencies in Figure 2.4 get confirmed by the cluster analysis. For the grouping of evaluation functions, it seems to be important whether the absolute travel time or a weighted travel time equivalent is used. In combination with the different passenger distribution models and assumptions on the passenger preferences, this can significantly influence how the evaluation functions are separated into different clusters. This is especially visible when comparing evaluation functions of the adapted journey time in combination with the second parameter setting to other travel time-based evaluation functions.

In addition to that, the cluster analysis adds a refinement of the previous observations and reveals coherences that are not or less visible in Figure 2.4. For example, the cluster analysis shows that there is a difference between utility-based evaluation functions for the different passenger distribution models. Functions of evaluated total utility are always clustered together when they use the same distribution model but are occasionally separated from each other when using different distribution models. This effect is mainly found in the NS case study and only visible in the cluster analysis since the three case studies are examined individually in contrast to an investigation of averaged values as in Figure 2.3. A probable explanation is that the services in this case study offer good alternative connections to the shortest connection for the main demand pairs. This affects the evaluation when considering all reasonable connections or the shortest connection only.

Figure 2.5 also shows that neither the parameter setting nor the choice of the distribution model is solely decisive for a clustering of the evaluation functions across the case studies. For some combinations of parameter settings and distribution models, evaluation functions of the different quality measurements are clustered together.

2.6 Implications

It is interesting to see that there are structural differences in the consistency of timetable evaluation functions. In addition to a mere statement that different evaluation functions might not agree on what is a good or a bad timetable, the structure of this study can identify and explain reasons for these differences. The analysis in Section 2.5 helps to determine which components of the functions have the most influence on the found inconsistencies. In this section, we give a brief indication of how this can be used for further research dealing with the evaluation of timetables.

Often, the design of evaluation functions is restricted for different reasons, such as unavailable data, imperfect knowledge about passenger behavior, or computational complexity. The observations from the inconsistencies and the cluster analysis allow implications on how to deal with these restrictions and which design element to focus on during the design or choice of an evaluation function.

On the one hand, the analysis can help to identify which simplifications of an evaluation function are justifiable. That means, it is possible to determine which simplifications have only a minor effect on the result of the evaluation. A simplification is justified if the desired evaluation function and its simplified version are rather consistent with each other, visible by not being separated into different clusters or by low values of inconsistency. For example, when designing an evaluation function based on absolute travel time without being aware of the precise parameters of the passenger preferences, approximate parameters will not drastically change the evaluation according to our case studies. This holds for both distribution models we tested, obvious from the low inconsistencies between evaluation functions 1 and 3, as well as between evaluation functions 2 and 4. Since, in the case of absolute travel time, the parameter settings for the passenger preferences affect only the connection choice, the validity of this simplification is expected and the analysis confirms that. This implies for the case of absolute travel time as the quality measurement that the negative impact of non-reflected modeling of passenger preferences can be disregarded as the resulting error is rather negligible.

On the other hand, this research helps to identify possibilities for improving a currently used evaluation function most effectively. Knowing that the evaluation function in use does not fully depict reality, it can be improved in various ways. The main categories of improvement are the quality measurement including which characteristics are considered, the modeling of passenger preferences and behavior, as well as the connection choice model. Since modifying an evaluation function often involves elaborate data acquisition or expensive remodeling, it is desirable to estimate the effects of possible modifications beforehand. For example, assume that a public transport operator applies the adapted journey time on a logit distribution for the evaluation of their services. To model passenger preferences of user groups, they use estimated parameters. In this case, it is highly recommended to identify the correct parameters for modeling the preferences and behavior of their customers properly. Using wrong parameters can lead to very different evaluation results as this research identified a high inconsistency between evaluation functions 10 and 12.

As mentioned, simplifying evaluation functions can be useful or necessary for several reasons. However, it is only reasonable if the evaluation results are consistent. It is therefore of utter importance to estimate the impact of a simplification on the evaluation. While this is important for any evaluation application, it is especially relevant when designing timetables. Using a wrong evaluation function as an objective in an optimization approach might not only give a wrong indication of what is a good or a bad timetable but can even misdirect the search for good solutions.

2.7 Conclusion

In this chapter, we structured evaluation functions for public transport timetables that are commonly used in the literature and identified three components in which the functions differ from each other. Based on this, we designed a set of evaluation functions representing a wide range of commonly used evaluation functions used in mathematical models, evaluation applications, and choice models.

Furthermore, we introduced and applied a novel method to quantify the inconsistency between evaluation functions. This is, unlike existing comparisons, an empirical approach based on the evaluation values of multiple timetables. Therefore, this definition is generally applicable for comparing evaluation functions and is not limited to the set of evaluation functions presented in this chapter.

With this method, we provided an analysis of the inconsistency of the designed evaluation functions. This analysis was conducted on three sets of timetables for an artificial grid network and the real-world network of Netherlands Railways. The findings are qualitatively similar for both infrastructures even though the networks and the timetables considered are structurally different. This suggests that a generalization of the results is possible.

In our experiments, we found that there are high inconsistencies between different evaluation functions although they are all designed to measure the same - the quality of timetables from the passengers' perspective. In all case studies, it appeared most crucial whether a travel time-based or a utility-based evaluation is used, which raises the question of why utility-based evaluation functions are commonly accepted for choice models but hardly used for evaluation. Furthermore, we observed that also within the group of travel time-based evaluation functions high inconsistencies can appear. It seemed most important which quality measurement is used but also different parameter settings and passenger distributions can significantly impact the inconsistency between evaluation functions. These inconsistencies can be used to validate simplifications of evaluation functions or to identify aspects of an evaluation function that need to be incorporated for a valid evaluation.

This research supports the impression that even within a set of evaluation functions which are all meant to evaluate the quality of timetables for passengers, the choice of the evaluation function can have a significant impact on the assessed quality of timetables, and thus also on which timetable is considered optimal. This observation is particularly crucial for Operations Research models in public transport as optimizing on the wrong objective function could make the world worse rather than better.

Appendix

2.A Notation

Greek letters

α	Scaling parameter for passenger preferences
β	Scaling parameter for logit model
γ	Scaling parameter for departure time tolerance
δ	Filter coefficient for ATT and PJT
ε	Filter parameter for ATT, PJT and NTR
φ	Normalized value of a service w.r.t. an evaluation function

Latin capitals

ADT	Adaption time
AJT	Adapted journey time
ATT	Absolute travel time
C	Set of connections
DEP	Departure time
ETU	Evaluated total utility
F	Set of evaluation functions
GL	Case study on grid infrastructure
GS	Case study on grid infrastructure
I	Index for case studies
IVT	In-vehicle time
NS	Case study on infrastructure of Netherlands Railways
NTR	Number of transfers
OD	Set of OD pairs
PJT	Perceived journey time
S	Set of public transport services
T	Analysis period
TWT	Transfer wait time
WKT	Walk time

Latin lower case letters

c	Index for connection
f	Index for evaluation function
i	Inconsistency
k	Number of clusters
mc	Distribution model on multiple connections
o	Passenger load
od	Index for OD pair
p	Probability for connection choice
ps	Parameter setting
s	Index for public transport service
sc	Distribution model on shortest connection
t	Index for time slice
v	Value of a service w.r.t. an evaluation function

2.B Definition of adaption time

The adaption time (ADT) is the time a passenger has to deviate from their preferred departure time slice t to take connection c. We use the adaption time to model the departure time preferences of passengers. Each time slice t corresponds to a one hour interval $[\underline{t}, \bar{t})$ of preferred departure time. Let $\hat{t} \in t$ be a time point in the time slice $t = [\underline{t}, \bar{t})$ and $\mathrm{DEP}(c)$ the departure time of connection c. Then, the adaption time is defined as

$$\mathrm{ADT}^t(c) \coloneqq \mathrm{ADT}^{[\underline{t}, \bar{t})}(c) \coloneqq \min_{\hat{t} \in [\underline{t}, \bar{t})} |\hat{t} - \mathrm{DEP}(c)|.$$

The adaption time could similarly be defined for arrival times, however, for the sake of simplicity we restrict ourselves to an adaption time at departures only. To model stronger departure time preferences we split each time slice t in $\gamma \in \mathbb{N}$ time windows t_j of equal length, with

$$t = \dot{\bigcup_{j=1}^{\gamma}} t_j.$$

Then, we assume that o^t_{od}/γ passengers want to depart in each of the γ time windows and the adaption time generalizes to the average adaption time to the γ time windows, i.e.,

$$\mathrm{ADT}^t(c) = \frac{1}{\gamma} \sum_{j=1}^{\gamma} \mathrm{ADT}^{t_j}(c).$$

2.C Aggregation of characteristics of connections

For the evaluation of timetables, we aggregate the characteristics of connections. As a first step, we aggregate the characteristics over all time slices $t \in T$ and connections $c \in C_{od}^t$ to obtain characteristic values for each OD pair. Let $p^t(c)$ be the probability that connection c is chosen by passengers with preferred departure time slice $t \in T$, i.e.,

$$\sum_{c \in C_{od}^t} p^t(c) = 1 \quad \forall t \in T$$

and

$$p^t(c) \geq 0 \quad \forall t \in T, c \in C_{od}^t.$$

How we derive meaningful values for this probability is outlined in Section 2.3.2. Let $X^t(c) \in \{\mathrm{ATT}(c), \mathrm{PJT}(c), \mathrm{AJT}^t(c)\}$ be a travel time-based characteristic of connection $c \in C_{od}^t$ with a value that possibly depends on the preferred departure time slice t. Then the average value of that characteristic over all time slices $t \in T$ and connections $c \in C_{od}^t$ for the OD pair od is derived by

$$X_{od} := \frac{\sum_{t \in T} \left(o_{od}^t \sum_{c \in C_{od}^t} p^t(c) \cdot X^t(c) \right)}{\sum_{t \in T} o_{od}^t}. \tag{2.10}$$

To compute the characteristic value for OD pairs, this value is weighted with the probability $p^t(c)$ that a connection c is chosen, given the preferred departure time slice t.

Furthermore, we define the evaluated total utility for passengers as

$$\mathrm{ETU}_{od} := \frac{\sum_{t \in T} \left(o_{od}^t \sum_{c \in C_{od}^t} \mathrm{ETU}_{od}^t(c) \right)}{\sum_{t \in T} o_{od}^t}. \tag{2.11}$$

This characteristic is not weighted with the passenger distribution $p^t(c)$ since the evaluated total utility of each connection $\mathrm{ETU}_{od}^t(c)$ is derived from the logit model which we use as the passenger connection choice model. However, note that the assumed passenger distribution model determines the set C_{od}^t of reasonably good alternatives. How the distribution model influences the set of alternatives is addressed in Section 2.3.2.

We define the characteristics of the public transport service X to be the weighted average of the characteristics for OD pairs, computed by

$$X \coloneqq \frac{\sum_{od \in OD} o_{od} \cdot X_{od}}{\sum_{od \in OD} o_{od}},$$

for $X_{od} \in \{\mathrm{ATT}_{od}, \mathrm{PJT}_{od}, \mathrm{AJT}_{od}, \mathrm{ETU}_{od}\}$. These aggregated quality measurements are used for evaluation of the public transport services.

2.D Derivation of a connection choice set

In all case studies, multiple services are considered, each of them consisting of a line plan and a timetable. The evaluation functions assume a set C^t_{od} of reasonable connections for each OD pair od with preferred departure time slice t to be given. In this section, we describe how we derive such sets from a given public transport service. To ensure better comparability of the evaluation, we derive the same choice sets for all evaluation functions within each case study.

In Section 2.3.2 we remark that two different connection choice sets are assumed, depending on the applied passenger distribution model. In the case of a distribution on multiple connections with the logit model, we assume that a set C_{od} of reasonably good connections for OD pair od is given. When all passengers are assigned to the shortest connections, we assume that the set C^t_{od} of all connections with lowest adapted journey time for passengers of OD pair od that want to depart in time slice t is given.

2.D.1 Choice set for logit model

To obtain a set with all reasonably good connections for an OD pair, we consider all connections with low absolute travel time, low perceived journey time, and a low number of transfers. The perceived journey time of the connections is compared using the fixed parameters

$$(\alpha_{\mathrm{WKT}}, \alpha_{\mathrm{TWT}}, \alpha_{\mathrm{NTR}}) = (1.5,\ 1.5,\ 7.5).$$

These values are the arithmetic mean of the values used for α in the two parameter settings ps_1 and ps_2. In addition, we use parameters δ_{PJT}, δ_{ATT}, $\varepsilon_{\mathrm{PJT}}$, $\varepsilon_{\mathrm{ATT}}$ and $\varepsilon_{\mathrm{NTR}}$ to decide whether a connection is good enough to be considered. Then, the choice set C_{od} contains

- all connections c which have at most an absolute travel time $\mathrm{ATT}(c)$ with

$$\mathrm{ATT}(c) < \delta_{\mathrm{ATT}} \cdot \mathrm{ATT}(c') + \varepsilon_{\mathrm{ATT}}$$

where c' is the connection with the lowest possible absolute travel time for OD pair *od*,

- all connections c which have at most a perceived journey time $\mathrm{PJT}(c)$ with

$$\mathrm{PJT}(c) < \delta_{\mathrm{PJT}} \cdot \mathrm{PJT}(c') + \varepsilon_{\mathrm{PJT}}$$

where c' is the connection with the lowest possible perceived journey time for OD pair *od* and

- all connections c which have at most $\mathrm{NTR}(c)$ transfers with

$$\mathrm{NTR}(c) < \mathrm{NTR}(c') + \varepsilon_{\mathrm{NTR}}$$

where c' is the connection with the lowest possible number of transfers for OD pair *od*.

For the derivation of choice sets for the analysis we use the values

$$\delta_{\mathrm{PJT}} \coloneqq 1.5,\ \delta_{\mathrm{ATT}} \coloneqq 1.5,\ \varepsilon_{\mathrm{PJT}} \coloneqq 10,\ \varepsilon_{\mathrm{ATT}} \coloneqq 10 \text{ and } \varepsilon_{\mathrm{NTR}} \coloneqq 1.$$

All dominated connections are removed from the choice sets. A connection $c \in C_{od}$ is dominated by another connection $c' \in C_{od}$ if

- connection c' starts simultaneously or later and arrives simultaneously or earlier than connection c, and
- connection c' has at most as many transfers as c, and
- the perceived journey time of connection c' is at most as high as the perceived journey time of connection c and
- at least one of the three conditions is a strict inequality

Since the search is independent of the time slice t, the choice set C_{od} contains all reasonably good connections for the OD pair during the whole analysis period T. As mentioned before, the logit model assigns a share of passengers significantly different from 0 only to those connections with low adaption time.

2.D.2 Choice set for shortest connections

For the assumption that all passengers use the shortest connections only, one choice set for each departure time slice t is required. We define these to be the subset of the choice set C_{od} with reasonably good connections, containing only the connections with minimal adapted journey time, i.e.,

$$C_{od}^t := \{c \in C_{od} : \ \mathrm{AJT}^t(c) \le \mathrm{AJT}^t(c') \quad \forall c' \in C_{od}\}.$$

2.E Results of case studies

We provide the normalized evaluation values, the medal counts, heat maps, and clusterings of all three case studies NS, GS, and GL in this section.

Explanation for the clusterings depicted in Tables 2.7, 2.10 and 2.13: The clusterings were found with the mixed-integer program described in Appendix 2.F. In the first column of each table is stated how many clusters are used. An asterisk indicates that the clustering is not proven to be optimal. The remaining columns contain the clusterings. The clusterings are separated by horizontal lines and in each row, one cluster is represented by the ids of the evaluation functions contained in the cluster.

2.E.1 NS case study

	ATT				PJT				AJT				ETU			
	ps_1		ps_2		ps_1		ps_2		ps_1		ps_2		ps_1		ps_2	
	sc	mc	sc	mc	sc	mc	sc	mc	sc	mc	sc	mc	sc	mc	sc	mc
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
NS12	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,34	0,57	0,20	0,51
NS13	0,43	0,35	0,44	0,35	0,35	0,23	0,14	0,10	0,34	0,22	0,49	0,57	1,00	1,00	1,00	1,00
NS14	0,43	0,31	0,43	0,29	0,35	0,22	0,19	0,13	0,29	0,20	0,02	0,31	0,58	0,48	0,56	0,34
NS15	0,34	0,27	0,35	0,26	0,26	0,17	0,09	0,00	0,23	0,15	0,02	0,17	0,60	0,52	0,60	0,49
NS16	0,35	0,28	0,38	0,24	0,28	0,16	0,10	0,01	0,25	0,14	0,03	0,15	0,35	0,58	0,24	0,66
NS17	0,12	0,08	0,11	0,11	0,13	0,11	0,15	0,23	0,11	0,11	0,00	0,18	0,00	0,27	0,00	0,29
NS18	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,07	0,00	0,00	0,07	0,00	0,12	0,00	0,10	0,00

Table 2.6: Normalized evaluation values φ in NS case study

	ATT				PJT				AJT				ETU			
	ps_1		ps_2		ps_1		ps_2		ps_1		ps_2		ps_1		ps_2	
	sc	*mc*	*sc*	*mc*	*sc*	*mc*	*sc*	*mc*	*sc*	*mc*	*sc*	*mc*	*sc*	*mc*	*sc*	*mc*
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0,00	5,52	0,65	6,08	4,76	11,24	15,33	21,43	6,58	12,12	18,70	10,22	26,68	22,90	29,89	25,66
2	5,52	0,00	5,94	1,26	1,28	6,35	10,73	16,83	1,75	7,49	15,51	7,95	31,11	28,42	32,37	28,44
3	0,65	5,94	0,00	6,50	5,30	11,67	15,87	21,97	7,01	12,62	18,90	10,53	26,80	22,48	30,01	25,13
4	6,08	1,26	6,50	0,00	1,77	5,16	9,47	15,57	0,75	6,22	15,02	7,14	32,37	28,98	32,60	29,00
5	4,76	1,28	5,30	1,77	0,00	6,93	10,57	16,67	2,27	7,81	16,68	7,75	31,27	27,21	32,57	27,54
6	11,24	6,35	11,67	5,16	6,93	0,00	4,38	10,48	4,66	1,13	13,05	7,22	37,46	34,14	37,51	34,17
7	15,33	10,73	15,87	9,47	10,57	4,38	0,00	6,10	8,98	3,25	12,71	10,06	41,84	37,31	41,89	37,33
8	21,43	16,83	21,97	15,57	16,67	10,48	6,10	0,00	15,08	9,35	10,94	15,14	45,92	41,04	45,97	41,06
9	6,58	1,75	7,01	0,75	2,27	4,66	8,98	15,08	0,00	5,74	14,84	7,04	32,86	29,49	33,33	29,51
10	12,12	7,49	12,62	6,22	7,81	1,13	3,25	9,35	5,74	0,00	12,72	7,77	38,59	35,02	38,65	35,05
11	18,70	15,51	18,90	15,02	16,68	13,05	12,71	10,94	14,84	12,72	0,00	12,60	38,10	39,85	38,16	39,87
12	10,22	7,95	10,53	7,14	7,75	7,22	10,06	15,14	7,04	7,77	12,60	0,00	32,79	27,25	32,85	27,27
13	26,68	31,11	26,80	32,37	31,27	37,46	41,84	45,92	32,86	38,59	38,10	32,79	0,00	14,64	4,23	17,80
14	22,90	28,42	22,48	28,98	27,21	34,14	37,31	41,04	29,49	35,02	39,85	27,25	14,64	0,00	17,85	4,87
15	29,89	32,37	30,01	32,60	32,57	37,51	41,89	45,97	33,33	38,65	38,16	32,85	4,23	17,85	0,00	21,01
16	25,66	28,44	25,13	29,00	27,54	34,17	37,33	41,06	29,51	35,05	39,87	27,27	17,80	4,87	21,01	0,00

Figure 2.6: Heat map showing inconsistencies in the normalized value $i_{\varphi}(f_1, f_2)$ in NS case study

	ATT				PJT				AJT				ETU			
	ps_1		ps_2		ps_1		ps_2		ps_1		ps_2		ps_1		ps_2	
	sc	mc	sc	mc	sc	mc	sc	mc	sc	mc	sc	mc	sc	mc	sc	mc
k	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
2	1	2	3	4	5	6	7	8	9	10	11	12				
													13	14	15	16
3	1	2	3	4	5				9			12				
						6	7	8		10	11					
													13	14	15	16
4	1	2	3	4	5				9			12				
						6	7	8		10	11					
													13		15	
														14		16
5	1	2	3	4	5				9			12				
						6	7	8		10						
											11					
													13		15	
														14		16

Table 2.7: Optimal clustering of the set of evaluation functions F into k clusters in the NS case study

2.E.2 GS case study

	ATT				PJT				AJT				ETU			
	ps_1		ps_2		ps_1		ps_2		ps_1		ps_2		ps_1		ps_2	
	sc	mc	sc	mc	sc	mc	sc	mc	sc	mc	sc	mc	sc	mc	sc	mc
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
GS1	0,67	0,64	0,64	0,58	0,53	0,52	0,50	0,48	0,45	0,48	0,24	0,29	0,70	0,73	0,68	0,68
GS2	0,61	0,64	0,58	0,60	0,50	0,52	0,46	0,49	0,43	0,49	0,29	0,33	0,84	0,84	0,83	0,80
GS3	0,62	0,64	0,59	0,61	0,51	0,52	0,47	0,50	0,45	0,49	0,36	0,36	0,88	0,87	0,87	0,83
GS4	0,93	0,95	0,89	0,88	0,96	1,00	0,95	0,98	0,81	0,91	0,44	0,61	0,79	0,82	0,78	0,79
GS5	0,93	0,88	0,89	0,81	0,96	0,95	0,97	0,96	0,79	0,87	0,36	0,57	0,61	0,65	0,60	0,61
GS6	0,71	0,84	0,69	0,88	0,62	0,72	0,54	0,61	0,57	0,66	0,43	0,45	1,00	0,96	0,97	0,91
GS7	0,66	0,73	0,63	0,76	0,55	0,62	0,48	0,54	0,50	0,58	0,42	0,43	0,88	0,91	0,86	0,87
GS8	0,48	0,52	0,46	0,50	0,38	0,43	0,33	0,37	0,35	0,40	0,38	0,38	0,92	0,95	0,89	0,89
GS9	0,50	0,52	0,48	0,51	0,39	0,41	0,32	0,35	0,36	0,39	0,30	0,44	0,88	0,92	0,85	0,86
GS10	0,91	0,87	0,87	0,80	0,68	0,67	0,56	0,55	0,57	0,62	0,31	0,34	0,84	0,87	0,86	0,86
GS11	0,91	0,87	0,87	0,80	0,68	0,67	0,56	0,55	0,57	0,62	0,31	0,34	0,84	0,87	0,86	0,86
GS12	0,32	0,53	0,30	0,47	0,48	0,62	0,52	0,66	0,81	0,57	0,71	0,34	0,83	0,79	0,77	0,78
GS13	0,80	0,85	0,81	0,87	0,64	0,70	0,51	0,57	0,57	0,66	0,43	0,54	0,98	0,98	0,98	0,97
GS14	0,65	0,71	0,63	0,71	0,49	0,54	0,39	0,45	0,44	0,50	0,39	0,42	1,00	1,00	0,99	0,96
GS15	0,00	0,02	0,00	0,02	0,00	0,03	0,00	0,04	0,00	0,04	0,10	0,29	0,54	0,55	0,53	0,54
GS16	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,34	0,00	0,00	0,00	0,00
GS17	0,48	0,47	0,46	0,45	0,39	0,39	0,39	0,40	0,36	0,37	0,42	0,67	0,93	0,98	0,92	0,93
GS18	0,70	0,67	0,67	0,62	0,55	0,55	0,51	0,49	0,48	0,51	0,45	0,40	0,79	0,79	0,80	0,77
GS19	1,00	0,96	0,95	0,89	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,92	0,97	0,94	0,96
GS20	0,66	0,63	0,63	0,59	0,53	0,52	0,50	0,48	0,56	0,54	0,65	0,56	0,82	0,83	0,83	0,81
GS21	0,97	0,94	0,93	0,88	0,71	0,72	0,58	0,58	0,77	0,74	0,69	0,64	0,97	1,00	1,00	1,00
GS22	0,38	0,51	0,36	0,47	0,51	0,61	0,55	0,65	0,61	0,59	0,71	0,51	0,82	0,81	0,81	0,82
GS23	0,98	1,00	1,00	1,00	0,77	0,82	0,64	0,67	0,82	0,83	0,78	0,76	0,94	0,97	0,97	0,98
GS24	0,94	0,91	0,90	0,84	0,70	0,69	0,58	0,57	0,67	0,67	0,56	0,52	0,91	0,95	0,93	0,94
GS25	0,01	0,06	0,01	0,03	0,02	0,09	0,02	0,05	0,04	0,10	0,10	0,00	0,19	0,13	0,13	0,13
GS26	0,48	0,48	0,46	0,45	0,43	0,45	0,41	0,43	0,43	0,43	0,50	0,52	0,87	0,90	0,87	0,88
GS27	0,59	0,56	0,57	0,51	0,48	0,47	0,45	0,44	0,46	0,46	0,44	0,38	0,78	0,82	0,78	0,79
GS28	0,56	0,52	0,53	0,48	0,47	0,45	0,44	0,42	0,45	0,45	0,47	0,28	0,15	0,17	0,12	0,12

Table 2.8: Normalized evaluation values φ in GS case study

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
GS16	10	5				1																						
GS15	5	5	2	4																								
GS25	1	2	13																									
GS28		4	1				2	4	1	2		1								1								
GS9				2	4	1			3	1				1	1	1			2									
GS8				2	2	2	2	1	1		2									3	1							
GS1				2		4				1	3	1	1		1	1	2											
GS27							1	1	5	2	4	1	1				1											
GS26					2	2	4	2								1	1		3		1							
GS2					2			1			3	4	2	3	1													
GS17				4		3	1	2						1							1	1	1		1	1		
GS12				2			2	2	1	2		1				1		2						1	1	1		
GS3										3		2	4	1	3	1	1	1										
GS22					2	2			1	1	1		1	1		1		2		1	1		1			1		
GS18							2		1	1			2	2	1	3		3	1									
GS20										1	1	5	1	1	2	3						1	1					
GS14						1			1	1	1	1	1	4			2								1		2	1
GS7													2		4	1	4	5										
GS10							2						1	1	1	1	1	1	2	1	5							
GS11							2						1	1	1	1	1	1	2	1	5							
GS5					4				1												1		2	3		3	2	
GS6																1	2		5		1	3	1	2				1
GS4								1	1	1	1							1					2	1	1	4	3	
GS13															1		1			6	2	1	1			4		
GS24																			1	2	2	4	4	1	2			
GS21																					1		1	5	4	2	1	2
GS23																								2	6		5	3
GS19																						1	2	1			3	9

Table 2.9: 'Medal count' from GS case study showing the number of times a public transport service is ranked on the n^{th} rank. Zeros are omitted for better visibility.

	ATT				PJT				AJT				ETU			
	ps_1		ps_2		ps_1		ps_2		ps_1		ps_2		ps_1		ps_2	
	sc	*mc*	*sc*	*mc*	*sc*	*mc*	*sc*	*mc*	*sc*	*mc*	*sc*	*mc*	*sc*	*mc*	*sc*	*mc*
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0,00	4,61	2,65	6,80	11,66	11,64	16,78	16,92	16,69	13,05	25,42	24,15	23,08	22,65	21,84	21,22
2	4,61	0,00	4,30	3,94	11,66	10,80	16,52	16,13	15,93	12,03	24,53	24,93	20,01	19,95	18,95	18,34
3	2,65	4,30	0,00	5,54	10,41	10,59	15,53	15,66	14,81	11,34	23,81	22,56	23,70	23,71	22,64	22,05
4	6,80	3,94	5,54	0,00	10,15	9,95	15,26	15,23	13,93	10,81	23,20	22,77	21,33	21,98	20,70	20,27
5	11,66	11,66	10,41	10,15	0,00	3,31	5,21	5,35	6,44	3,44	15,67	15,57	30,47	30,99	29,85	29,42
6	11,64	10,80	10,59	9,95	3,31	0,00	7,41	5,69	7,08	2,98	15,95	17,83	27,82	28,34	27,21	26,77
7	16,78	16,52	15,53	15,26	5,21	7,41	0,00	3,30	5,80	6,03	13,12	13,85	34,79	35,30	34,17	33,74
8	16,92	16,13	15,66	15,23	5,35	5,69	3,30	0,00	5,45	5,09	12,73	14,62	32,58	33,10	31,97	31,53
9	16,69	15,93	14,81	13,93	6,44	7,08	5,80	5,45	0,00	4,73	11,19	13,80	30,23	31,01	29,92	29,44
10	13,05	12,03	11,34	10,81	3,44	2,98	6,03	5,09	4,73	0,00	14,04	15,53	29,14	29,65	28,52	28,09
11	25,42	24,53	23,81	23,20	15,67	15,95	13,12	12,73	11,19	14,04	0,00	10,02	36,38	37,40	35,61	35,30
12	24,15	24,93	22,56	22,77	15,57	17,83	13,85	14,62	13,80	15,53	10,02	0,00	35,66	36,68	34,88	34,57
13	23,08	20,01	23,70	21,33	30,47	27,82	34,79	32,58	30,23	29,14	36,38	35,66	0,00	2,65	1,87	2,45
14	22,65	19,95	23,71	21,98	30,99	28,34	35,30	33,10	31,01	29,65	37,40	36,68	2,65	0,00	2,46	2,66
15	21,84	18,95	22,64	20,70	29,85	27,21	34,17	31,97	29,92	28,52	35,61	34,88	1,87	2,46	0,00	1,20
16	21,22	18,34	22,05	20,27	29,42	26,77	33,74	31,53	29,44	28,09	35,30	34,57	2,45	2,66	1,20	0,00

Figure 2.7: Heat map showing inconsistencies in the normalized value $i_\varphi(f_1, f_2)$ GS case study

	ATT				PJT				AJT				ETU			
	ps_1		ps_2		ps_1		ps_2		ps_1		ps_2		ps_1		ps_2	
	sc	mc	sc	mc	sc	mc	sc	mc	sc	mc	sc	mc	sc	mc	sc	mc
k	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
2	1	2	3	4	5	6	7	8	9	10	11	12				
													13	14	15	16
3	1	2	3	4												
					5	6	7	8	9	10	11	12				
													13	14	15	16
4	1	2	3	4												
					5	6	7	8	9	10						
											11	12				
													13	14	15	16
5*	1	2	3	4												
					5	6	7	8	9	10						
											11					
												12				
													13	14	15	16

Table 2.10: Optimal clustering of the set of evaluation functions F into k clusters in the GS case study. The asterisk indicates that the clustering is not proven to be optimal.

2.E.3 GL case study

	ATT				PJT				AJT				ETU			
	ps_1		ps_2		ps_1		ps_2		ps_1		ps_2		ps_1		ps_2	
	sc	mc	sc	mc	sc	mc	sc	mc	sc	mc	sc	mc	sc	mc	sc	mc
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
GL1	0,38	0,35	0,38	0,34	0,44	0,40	0,49	0,45	0,41	0,41	0,34	0,21	0,44	0,46	0,39	0,39
GL2	0,40	0,37	0,40	0,37	0,46	0,42	0,50	0,46	0,43	0,43	0,34	0,22	0,46	0,47	0,40	0,40
GL3	0,63	0,61	0,63	0,61	0,67	0,64	0,63	0,60	0,64	0,65	0,50	0,29	0,69	0,69	0,72	0,72
GL4	0,39	0,37	0,39	0,37	0,43	0,41	0,45	0,42	0,44	0,42	0,62	0,21	0,52	0,53	0,51	0,51
GL5	0,34	0,33	0,34	0,33	0,39	0,37	0,40	0,38	0,41	0,38	0,68	0,18	0,50	0,51	0,51	0,51
GL6	0,51	0,47	0,51	0,47	0,54	0,50	0,63	0,60	0,54	0,51	0,73	0,34	0,64	0,65	0,60	0,60
GL7	0,38	0,35	0,37	0,34	0,44	0,40	0,50	0,47	0,42	0,41	0,45	0,23	0,55	0,56	0,50	0,50
GL8	0,64	0,63	0,65	0,62	0,69	0,66	0,69	0,66	0,65	0,67	0,49	0,40	0,84	0,85	0,88	0,88
GL9	0,50	0,47	0,50	0,47	0,61	0,58	0,71	0,68	0,58	0,58	0,46	0,36	0,48	0,49	0,42	0,42
GL10	0,47	0,44	0,47	0,44	0,53	0,49	0,61	0,57	0,51	0,50	0,45	0,30	0,45	0,46	0,39	0,39
GL11	0,41	0,38	0,41	0,38	0,47	0,44	0,55	0,51	0,46	0,45	0,49	0,26	0,64	0,65	0,60	0,61
GL12	0,41	0,38	0,40	0,37	0,46	0,43	0,52	0,49	0,45	0,44	0,51	0,25	0,62	0,63	0,60	0,60
GL13	0,40	0,37	0,40	0,37	0,45	0,42	0,53	0,50	0,45	0,43	0,54	0,26	0,64	0,65	0,60	0,60
GL14	0,44	0,41	0,44	0,41	0,49	0,45	0,55	0,52	0,47	0,46	0,55	0,27	0,60	0,61	0,59	0,59
GL15	0,45	0,44	0,46	0,45	0,53	0,50	0,54	0,51	0,53	0,51	0,63	0,19	0,37	0,38	0,35	0,35
GL16	0,54	0,52	0,54	0,52	0,58	0,55	0,63	0,60	0,56	0,54	0,56	0,32	0,79	0,79	0,79	0,80
GL17	0,50	0,48	0,50	0,48	0,54	0,51	0,59	0,55	0,52	0,51	0,51	0,29	0,77	0,78	0,79	0,79
GL18	0,82	0,80	0,81	0,80	0,86	0,84	0,88	0,86	0,83	0,85	0,78	0,67	0,93	0,93	0,94	0,94
GL19	0,81	0,80	0,81	0,80	0,87	0,86	0,94	0,92	0,83	0,86	0,71	0,71	0,83	0,83	0,82	0,82
GL20	0,98	0,98	0,98	0,97	0,98	0,98	0,98	0,96	0,94	0,97	0,79	0,78	0,94	0,94	0,95	0,95
GL21	0,26	0,24	0,26	0,24	0,29	0,27	0,32	0,29	0,28	0,28	0,29	0,30	0,76	0,75	0,76	0,76
GL22	0,22	0,21	0,22	0,21	0,24	0,23	0,26	0,25	0,23	0,24	0,21	0,21	0,66	0,65	0,67	0,67
GL23	0,11	0,10	0,11	0,10	0,14	0,12	0,16	0,14	0,12	0,13	0,08	0,09	0,37	0,37	0,36	0,36
GL24	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
GL25	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
GL26	0,74	0,77	0,74	0,77	0,78	0,80	0,81	0,85	0,76	0,78	0,80	0,84	0,98	0,98	0,94	0,94
GL27	0,66	0,68	0,66	0,69	0,70	0,71	0,72	0,75	0,74	0,70	0,90	0,87	0,97	0,97	0,97	0,97

Table 2.11: Normalized evaluation values φ in GL case study

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
GL24	16																										
GL23		14	2																								
GL1				4	1	4	6	1																			
GL22			11				1									4											
GL5			1		1			2		2										1							
GL7						4	4	4	2	2																	
GL21				11												1		4									
GL2						5		3	3	4	1																
GL4					1	3		4	5	2								1									
GL13								2	5	1	3			4	1												
GL15		2	2	1								2		4	1	3			1								
GL12										4	7	4		1													
GL10					4		1							4	4	2	1										
GL11											1	9	4		2												
GL14											4		9	2		1											
GL9							4		1							4			4	1	2						
GL6													2		3	1	6	2	1			1					
GL17													1	1	4		4	2	4								
GL3												1			1		4		2	8							
GL16																	1	7	4	4							
GL8										1										2	9	4					
GL19																					5		1	4	6		
GL27																						1			2	4	
GL26																							1	2	2	2	
GL18																						1	5	6	4		
GL20																								4	2	1	
GL25																											16

Table 2.12: 'Medal count' from GL case study showing the number of times a public transport service is ranked on the n^{th} rank. Zeros are omitted for better visibility.

	ATT				PJT				AJT				ETU			
	ps_1		ps_2		ps_1		ps_2		ps_1		ps_2		ps_1		ps_2	
	sc	*mc*	*sc*	*mc*	*sc*	*mc*	*sc*	*mc*	*sc*	*mc*	*sc*	*mc*	*sc*	*mc*	*sc*	*mc*
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0,00	2,12	0,23	2,23	4,35	2,27	8,11	6,28	3,30	2,67	9,94	14,98	16,24	16,47	15,79	15,79
2	2,12	0,00	2,02	0,25	6,12	3,97	9,86	7,81	5,04	4,38	10,12	13,03	17,50	17,93	16,85	16,86
3	0,23	2,02	0,00	2,14	4,46	2,39	8,23	6,39	3,42	2,78	9,95	14,89	16,40	16,63	15,93	15,93
4	2,23	0,25	2,14	0,00	6,12	3,97	9,86	7,80	5,05	4,36	10,04	12,92	17,53	17,95	16,88	16,89
5	4,35	6,12	4,46	6,12	0,00	2,40	4,08	2,70	1,93	1,78	9,88	18,56	14,08	14,23	14,49	14,49
6	2,27	3,97	2,39	3,97	2,40	0,00	5,96	4,16	1,75	0,69	9,67	16,24	15,31	15,50	15,18	15,18
7	8,11	9,86	8,23	9,86	4,08	5,96	0,00	2,76	5,40	5,62	10,61	21,82	12,80	12,80	13,45	13,44
8	6,28	7,81	6,39	7,80	2,70	4,16	2,76	0,00	3,71	3,84	9,55	19,18	13,48	13,57	13,91	13,91
9	3,30	5,04	3,42	5,05	1,93	1,75	5,40	3,71	0,00	1,53	8,56	17,11	14,35	14,54	14,46	14,45
10	2,67	4,38	2,78	4,36	1,78	0,69	5,62	3,84	1,53	0,00	9,73	16,80	15,06	15,25	15,04	15,03
11	9,94	10,12	9,95	10,04	9,88	9,67	10,61	9,55	8,56	9,73	0,00	16,57	15,86	16,11	15,71	15,71
12	14,98	13,03	14,89	12,92	18,56	16,24	21,82	19,18	17,11	16,80	16,57	0,00	27,32	27,82	25,88	25,90
13	16,24	17,50	16,40	17,53	14,08	15,31	12,80	13,48	14,35	15,06	15,86	27,32	0,00	0,72	2,41	2,39
14	16,47	17,93	16,63	17,95	14,23	15,50	12,80	13,57	14,54	15,25	16,11	27,82	0,72	0,00	2,82	2,80
15	15,79	16,85	15,93	16,88	14,49	15,18	13,45	13,91	14,46	15,04	15,71	25,88	2,41	2,82	0,00	0,03
16	15,79	16,86	15,93	16,89	14,49	15,18	13,44	13,91	14,45	15,03	15,71	25,90	2,39	2,80	0,03	0,00

Figure 2.8: Heat map showing inconsistencies in the normalized value $i_\varphi(f_1, f_2)$ in GL case study

	ATT				PJT				AJT				ETU			
	ps_1		ps_2		ps_1		ps_2		ps_1		ps_2		ps_1		ps_2	
	sc	mc	sc	mc	sc	mc	sc	mc	sc	mc	sc	mc	sc	mc	sc	mc
k	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
2	1	2	3	4	5	6	7	8	9	10	11	12				
													13	14	15	16
3	1	2	3	4	5	6	7	8	9	10	11					
												12				
													13	14	15	16
4	1	2	3	4	5	6	7	8	9	10						
											11					
												12				
													13	14	15	16
5	1	2	3	4												
					5	6	7	8	9	10						
											11					
												12				
													13	14	15	16

Table 2.13: Optimal clustering of the set of evaluation functions F into k clusters in the GL case study

2.F Mixed-integer program for clustering problem

Let φ_s^f be the normalized evaluation values of a public transport service $s \in S$ with respect to evaluation function $f \in F$. Then, an optimal clustering of the set of evaluation functions F into k clusters can be found by solving the program

$$
\begin{array}{llcll}
\min & & & \sum_{f \in F} d(f) & \\
\text{s.t.} & \sum_{j=1}^{k} b_{j,f} & = & 1 & \forall f \in F \\
& d(f) & = & \sum_{j=1}^{k} d(m_j, f) \cdot b_{j,f} & \forall f \in F \\
& d(m_j, f) & = & \frac{1}{|S|} \sum_{s \in S} |\varphi_s^f - m_{j,s}| & \forall f \in F,\ \forall j = 1, \ldots, k \\
& m_{j,s} & = & \frac{1}{\sum_{f \in F} b_{j,f}} \sum_{f \in F} \varphi_s^f \cdot b_{j,f} & \forall s \in S,\ \forall j = 1, \ldots, k \\
& m_{j,s} & \in & \mathbb{R} & \forall s \in S,\ \forall j = 1, \ldots, k \\
& b_{j,f} & \in & \{0, 1\} & \forall f \in F,\ \forall j = 1, \ldots, k \\
& d(m_j, f) & \in & \mathbb{R} & \forall f \in F,\ \forall j = 1, \ldots, k \\
& d(f) & \in & \mathbb{R} & \forall f \in F
\end{array}
$$

The binary variable $b_{j,f}$ links the evaluation functions f to the clusters j and the first constraint ensures that each function is assigned to exactly one cluster. The second constraint assigns the distance of each evaluation function f to its cluster center m_j to the variable $d(f)$. The distance between the functions f and the cluster centers m_j are computed in the third constraint using the distance function $d(m, f)$ as defined in Equation (2.9). With the fourth constraint, the cluster centers are computed as the arithmetic mean of all evaluation functions that are assigned to the cluster. The objective is to minimize the total distance of all evaluation functions to their respective cluster center. We solve a linearized version of this clustering problem.

Chapter 3

Railway timetabling with integrated passenger distribution

3.1 Introduction

Public transport is important to our society for various reasons, such as increased mobility for the general public or lower air pollution compared to individual transport. Especially the potential of public transport to reduce emissions is recently much discussed in the context of climate change. To be considered an alternative to individual transport, public transport has to be as attractive as possible to passengers. For decades both researchers and practitioners have been working on the improvement of public transport from different perspectives using various approaches. Most of them follow the same pattern and design public transport sequentially. First, long-term planning decisions are taken, such as stop location planning and, in the case of railways, infrastructure design. Afterward, the line routes are designed and the corresponding frequencies of lines are fixed. On the tactical level, a timetable is determined, based on the results of the previous steps. Finally, vehicles and crews are scheduled.

Finding a good timetable is an integral step for providing high-quality public transport services to passengers. Next to the driving times of vehicles, the timetable determines the transfer times and thereby the travel times of passengers. Since transfer and travel times have a significant effect on the chosen routes of passengers and also their satisfaction with public transport, timetabling is a relevant problem with high practical impact. Moreover, from an algorithmic perspective timetabling is an interesting task since finding a feasible periodic timetable is NP-complete. For this reason, research often focuses on efficient solution strategies. In recent years, many publications deal with the question of how passenger travel time can be used as an objective to guide the search for timetables of high quality.

When designing public transport services, a good compromise must be made between service quality and the costs of operating a public transport service. Since costs are mainly determined by the line plan as well as the vehicle and crew schedule, many optimization approaches for timetabling only aim at providing the best quality to passengers. Even though the focus is on the quality for passengers, strong assumptions on passenger demand are made. Among them, two assumptions are commonly found: First, all passengers travel on their shortest available route. Second, a predetermined passenger assignment to routes is sufficient to estimate passenger loads in the public transport network. In this context, a passenger route defines when and on which lines passengers travel. As summarized in Table 3.1, the impact of each of these two assumptions has been studied individually. Improvements could be

	Predetermined route choice	Integrated route choice
Single route	Nachtigall and Opitz (2008) Pätzold and Schöbel (2016)[1] Liebchen (2018)	Gattermann et al. (2016) Borndörfer et al. (2017) Löbel et al. (2019)
Distribution	Parbo et al. (2014) Sels et al. (2015) Robenek et al. (2016)	this chapter

Table 3.1: Selection of timetabling publications, categorized by (1) whether a predetermined route choice is assumed or a route choice model is integrated and (2) whether it is assumed that passengers use a single route only or distribute on multiple routes. The mentioned publications are discussed together with other related literature in Section 3.2.

achieved by considering a passenger distribution on multiple routes and by integrating a shortest-route search into optimization, respectively.

Motivated by these improvements, we relax both assumptions at the same time. We study the problem of finding a travel-time minimal timetable under the assumption that passengers' route choice can be modeled using a discrete choice model. To our best knowledge, this is the first time that a choice model is used to derive a passenger distribution within a timetable optimization model.

Depending on the quality of all available routes, discrete choice models estimate the probability that a route is chosen by passengers. This route choice corresponds to a passenger distribution in the network. We use the logit model, a commonly used passenger route choice model in transport applications, to estimate passenger distributions on available routes and incorporate it in an optimization framework for timetabling. Due to the non-linear structure of the logit model, the mathematical program for this problem requires reformulation to be solved exactly. We present two ways to integrate a passenger distribution on multiple routes into a timetabling model as a linear formulation. Our first model uses a novel linear distribution model. This distribution model is designed to have the same characteristics as the logit

[1]Publications with predetermined route choice mostly assume a passenger weight to be given without explicitly mentioning which distribution was used to obtain the weights. The authors know from conference presentations and personal conversations that almost always a shortest path routing is used. If a passenger distribution is derived from a choice model or historic data, this is usually reported. Therefore, we assume that publications with a predetermined route choice applied an assignment to a single route unless explicitly stated otherwise. This matches with reports of other authors, see for example Siebert and Goerigk (2013), Schmidt and Schöbel (2015b) or P. Schiewe and Schöbel (2020).

model. Due to its linear formulation, it can easily be incorporated into an optimization model. The second model relies on a simulation of the logit distribution of passengers. By considering multiple scenarios, the distribution of passengers according to a logit model can be approximated within an optimization model that is linear in all variables.

We aim at maximizing the quality of timetables for passengers. Researchers and practitioners developed a variety of ways to evaluate timetables from the passengers' perspective. Due to their design, not all of these evaluations are suitable as an objective function in an optimization program. Chapter 2 shows in an empirical comparison that different evaluation methods do not necessarily yield a consistent evaluation of timetables. To best reflect the quality of the found solutions, we evaluate all timetables in our experiments with multiple evaluation functions. As an objective function, the first model uses the absolute travel time to minimize the time spent in the public transport system, which matches common practice in timetabling literature. In the second model, simulated travel times are minimized to incorporate passengers' preferences that are not captured by absolute travel times only. These preferences can include any kind of non-modeled factors of influence, from differently perceived transfer times through to a popular ice cream shop at a certain transfer station. We discuss the theoretical properties of the chosen objective functions of the two models and analyze their influence on the resulting timetable in the experiments. This discussion suggests that the absolute travel time, although commonly used in literature, might not be suitable for evaluating timetables when considering multiple alternative routes for passengers.

We compare our models for timetabling with integrated passenger distribution with four timetabling approaches motivated from the literature. Two of these approaches assume that a passenger assignment to routes is fixed before optimizing the timetable, using either a single route for all passengers traveling between the same stations or a distribution on multiple routes. Another approach finds optimal timetables based on the assumption that passengers use the shortest available routes. A fourth approach solves the problem of timetabling with integrated passenger distribution heuristically by iterating between assigning passengers to routes according to the logit model and finding optimal timetables. The experiments show that the two proposed models are capable of finding better solutions than the benchmark approaches. The found timetables performed better concerning some evaluation functions while being of comparable quality concerning other evaluation functions when compared to the timetables found by existing approaches. These improvements come at the expense of increased complexity of the models. From this, we conclude that the integration of a passenger distribution model has the potential to find better timetables for passengers, but more efficient solution strategies have to be developed.

We want to highlight two contributions of this chapter: First, we present a novel timetabling model with an integrated choice model to derive a passenger distribution on multiple routes. We provide and discuss alternative representations of the passenger distribution and develop two mixed-integer linear timetabling programs. Second, we show on multiple artificial instances and a partial real-world network the advantages and disadvantages of the novel approaches when compared to state-of-the-art methods. In particular, our experiments provide insight into (1) how considering multiple routes for passengers instead of a single route, and (2) how integrating route choice instead of a predetermined route assignment affects solution quality.

The remainder of this chapter is structured as follows. We summarize the literature on passenger distribution models, on optimization approaches for timetabling, and on the evaluation of timetables in Section 3.2. In Section 3.3, the basic models relevant for this chapter are introduced and the problem is defined. In Section 3.4, we develop and discuss two mixed-integer linear timetabling programs with an integrated passenger distribution model. Section 3.5 describes the experimental setup, such as considered instances, benchmark methods, and used evaluation functions. We report and discuss our results of the experiments in Section 3.6 and conclude in Section 3.7.

3.2 Related literature

3.2.1 Passenger Route choice

State-of-the-art discrete choice models provide appropriate solutions for describing passengers' behavior concerning mode and route choices (de Dios Ortúzar and Willumsen, 2011). A choice model estimates which alternative is chosen by an individual given the utilities of all alternatives. Ben-Akiva and Lerman (1985) give in their book a comprehensive overview of the theory of choice models. In aggregate form, the chosen routes of individual passengers correspond to a distribution of all passengers in the public transport network. For estimating passenger distributions in public transport applications, the logit model is most commonly applied. To adjust to specific requirements, the logit model is continually developed further. For example, Espinosa-Aranda et al. (2018) propose a constrained nested logit model to model passenger distributions on routes in public transport. Since recently, choice models in general and the logit model in specific are applied in optimization approaches for public transport applications. Canca et al. (2019) use it to estimate a passenger distribution and mode choice in the context of transit network planning. They solve the resulting non-linear program with a neighborhood search-based matheuristic. Due to the non-linear structure, exact solution approaches rely mostly on a linearization of the logit model. De-Los-Santos et al. (2017) developed a linear approximation by using that one alternative with fixed utility is available. An overview of common linearizations of the logit model is given by Haase and Müller (2014).

One interpretation of choice theory is that each alternative is perceived differently by people. This is usually modeled by adding an error term to the deterministic utility of alternatives. The error terms are used as an unknown part of the utility in many choice models. They model different sources of uncertainty and imperfect knowledge of analysts, such as unobserved route attributes, unobserved passenger preferences, or measurement errors (Ben-Akiva and Lerman, 1985). The distribution of the error terms determines the choice model. For example, independent and identical Gumbel distributed error terms yield a logit model. By drawing random terms from a specific distribution, the corresponding choice model can be simulated (Train, 2009). Pacheco et al. (2016) described such a simulation framework to compute optimal pricing strategies for different parking options while considering passenger behavior.

3.2.2 Timetabling

Timetabling approaches for public transport applications are usually classified into periodic and aperiodic cases. As we aim at finding a periodic timetable, we focus on the periodic timetabling literature. Most formulations are based on the periodic event scheduling problem (PESP) as introduced by Serafini and Ukovich (1989) or the cyclic periodicity formulation (CPF), which is a further development of the PESP model by Nachtigall (1994). While the PESP has one variable for each event modeling points in time, the CPF uses one variable for each activity expressing a time duration.

Serafini and Ukovich (1989) showed that the problem of finding a periodic timetable is NP-complete. Many publications focus on finding efficient ways to solve periodic timetabling. Schrijver and Steenbeek (1994) developed a constraint propagation algorithm which later on served as a basis for one of the first successful implementations of a timetable found with methods of Operations Research (Kroon et al., 2009). A powerful heuristic to solve the PESP model is the modulo network simplex algorithm developed by Nachtigall and Opitz (2008). The algorithm is inspired by the simplex algorithm for solving linear programs where a feasible solution is improved in each iteration by exchanging a basis and a non-basis variable. Pätzold and Schöbel (2016) proposed a promising matching-based heuristic that could find timetables in short computation times. Their algorithm was designed for a reduced PESP model with fixed drive and dwell times for vehicles. Liebchen (2018) described how to exploit the specific structure of a PESP instance to derive effective preprocessing techniques that reduce the complexity of the timetabling problem. An overview of models and solution methods for railway timetabling is given in Borndörfer et al. (2018).

Initially introduced as a feasibility program, the PESP model was quickly extended by objective functions to guide the optimization. Recent publications often aim at designing timetables with minimal passenger travel time or with the lowest energy consumption during operation. We refer to Scheepmaker et al. (2017) for a summary of energy-efficient timetabling approaches and focus on passenger travel time. However, to model the objective of passenger travel time, two restrictive assumptions on passenger behavior are usually made. These assumptions have been shown to distort the search for an optimal solution.

First, passengers are usually assigned to routes in the transport network before the timetable optimization. With this passenger assignment to routes, the arcs in the network are assigned weights to consider passenger routes during optimization in a heuristic way. Many publications have challenged this assumption and shown that the routes passengers use depend on the timetable (Schmidt, 2014) and, therefore, cannot be reliably determined beforehand. To take passengers' reactions on the designed public transport into consideration, Nachtigall (1998) and Siebert and Goerigk (2013) experimented with iterative approaches. They alternately assigned passengers to shortest routes and optimized the timetable given the updated passenger routes. Schmidt and Schöbel (2015b) integrated a shortest-route search for passengers into the timetabling optimization model and further improved the quality of timetables found. They used that the exact route of passengers does not need to be known in the aperiodic case since start and end events contain sufficient information for travel time computation. With this trick, the resulting timetabling model with integrated passenger assignment to shortest routes could be solved efficiently. Borndörfer et al. (2017) developed a general timetable optimization model that allows the implementation of different passenger routing models. They discussed theoretical bounds for four passenger routing models: a lower-bound routing model where passenger routes are found before knowing the timetable; a shortest path routing model where passengers use the shortest path depending on the timetable; a capacitated multi-path routing model where passengers distribute on several paths to avoid violation of capacity constraints; and a capacitated unsplittable path routing model, where all passengers between the same origin and destination travel on the same path while respecting capacity constraints. Their results include the finding that, for different objectives, the travel time on a timetable optimized with predetermined passenger routes can be arbitrarily higher than the travel time on a timetable optimized with integrated passenger routing. Next to theoretical gaps, Borndörfer et al. (2017) also compared the lower-bound routing model with the shortest path routing model in experiments and found significantly improved transfer waiting times for passengers by integration of the passenger routing model. A different solution approach to periodic timetabling with integrated shortest-route search was described in Gattermann et al. (2016). They used time slices to model departure time preferences and defined a translation of the integrated model to a satisfiability problem. P. Schiewe and Schöbel (2020) provide a heuristic approach for the timetabling problem with an integrated shortest-route search that considers only a small share of the OD pairs for timetable-dependent routes. Depending on whether the remaining OD pairs are assigned to fixed routes or not, upper or lower bounds for the exact solution can be found. Together with a preprocessing procedure that reduces the problem size by eliminating unnecessary routing variables, they are able to find improved solutions for close-to-real-world instances. Recently, Löbel et al. (2019) proposed an adjustment of the modulo simplex algorithm to incorporate a shortest-route search during optimization. Assuming that passengers always take the next available train in a high-frequency network, Polinder et al. (2019) and Polinder et al. (2020) integrated a route selection of passengers in a PESP model.

Second, for the design of a majority of timetable objective functions, it is assumed that passengers only travel on the shortest route. Van der Hurk et al. (2014) concluded from their study based on smart card travel data that this is one of the common misassumptions on passenger behavior. Many publications challenged this assumption and proposed enhanced models to develop better timetables for passengers. As input to their timetabling model, Sels et al. (2015) described a passenger assignment to routes that are at most 20% longer than the potentially shortest route. Robenek et al. (2016) used estimates for utilities of available connections as defined for choice models together with time-dependent demand structures to estimate the distribution of passengers. A similar approach was used by Parbo et al. (2014) for deriving passenger distributions, who updated the passenger distribution after each timetable computation. As mentioned in the literature review on passenger choice models in Section 3.2.1, first choice models were integrated into optimization approaches of other public transport applications. To the best of our knowledge, other choice models for passenger route choice than a shortest-route search were not integrated into an optimization framework for timetabling, which we do in this chapter.

3.2.3 Timetable evaluation

As discussed in Section 3.2.2, the majority of publications in Operations Research use the absolute travel time of passengers on predetermined routes as objective. This evaluation function is suitable for optimization because of its simple structure. In other research areas, timetables are usually evaluated differently. For evaluation purposes in Transport Engineering, the perceived travel time is often used. That is a weighted travel time equivalent that incorporates more factors of influence, such as penalties for transfers, fares, or adaption time (de Dios Ortúzar and Willumsen, 2011). In contrast to that, commonly applied choice models use an evaluated utility to measure the quality of a timetable for passengers. An evaluated utility is usually a non-linear function of a weighted travel time equivalent, such as the perceived travel time. Recently, evaluated utilities are often proposed as a replacement for established evaluation functions. For example, de Jong et al. (2007) summarized the literature on 'logsums', an evaluated utility, and showcased the advantages of this evaluation in a case study on high-speed trains in the Netherlands. Indeed, Chapter 2 showed that timetable evaluation functions do not yield consistent evaluation results, although they are all designed to evaluate the quality of timetables for passengers. This suggests that timetables should be evaluated from different perspectives.

3.3 Problem definition

In this section, we define the problem of timetabling with an integrated passenger distribution on multiple routes. To this end, we give a basic formulation for both problems: timetabling assuming that a passenger assignment to routes is given, and route choice modeling assuming that a timetable is known.

All formulations are based on an event activity network $\mathcal{N} = (\mathcal{E}, \mathcal{A})$ with a set of events $\mathcal{E}$ and a set of activities $\mathcal{A}$. In this context, an event $i \in \mathcal{E}$ denotes an arrival or a departure of a vehicle at a station, and an activity $ij \in \mathcal{A}$ represents a drive or a wait activity of a vehicle between two events $i \in \mathcal{E}$ and $j \in \mathcal{E}$. Activities can be used to model more than vehicle actions, for example, transfer activities of passengers and headway or synchronization constraints between vehicles (Liebchen and Möhring, 2007).

3.3.1 Passenger distribution

Discrete choice models can be used to describe passengers' behavior concerning route choice when a timetable is known. We use the logit model to estimate the distribution of passengers on their routes. The passenger routes in the public transport network are represented by paths in the event activity network. A path $p = (i_1, \ldots, i_{m_p})$ is a sequence of events i in the event activity network such that two consecutive events are connected by a drive, wait, or transfer activity. We denote the perceived travel time of path p by t_p, which is a weighted linear combination of the influencing factors such as travel time and the number of transfers. The perceived travel time is often interpreted as a negative utility of the path p and assumed to be given in the context of choice modeling.

Let a set P of alternative paths with perceived travel times t_p for all paths $p \in P$ be given. Then, the logit model can be interpreted as a probability function w_p^{lm} that assigns a probability to alternative p, based on the utility of all considered alternatives;

$$w_p^{lm}((t_q)_{q \in P}) = \frac{e^{\beta t_p}}{\sum_{q \in P} e^{\beta t_q}}, \tag{3.1}$$

where $(t_q)_{q \in P}$ is a vector containing the utilities of all paths in the set P. With the scalar $\beta \in \mathbb{R}$, the logit model can be adjusted to suit the specific instance.

3.3.2 Timetabling

In the literature, an instance $\mathcal{I} = (\mathcal{N}, l, u, OD)$ for a timetabling problem usually consists of an event activity network $\mathcal{N}$ with lower and upper bounds l and u on the activity durations and a demand matrix OD that indicates how many passengers wish to travel from each origin to destination. It remains to find arrival and departure times for each line at each station. We focus on the cyclic periodicity formulation for periodic timetabling problems, as described in Nachtigall (1994). This integer linear formulation is based on an event activity network with constraints ensuring that the duration $\delta_{ij} \in \mathbb{Z}_+$ of each activity $ij \in \mathcal{A}$ is between a given lower $l_{ij} \in \mathbb{Z}_+$ and upper bound $u_{ij} \in \mathbb{Z}_+$, i.e.,

$$l_{ij} \leq \delta_{ij} \leq u_{ij} \quad \forall ij \in \mathcal{A}. \tag{3.2}$$

We assume that the timetable has an accuracy of one time unit and the duration between events is integer-valued. To ensure that the durations can be transformed into a feasible timetable that assigns a point in time to each event, cycle constraints need to be added to the model (Nachtigall, 1994). It is sufficient to include cycle constraints for each cycle c in an integral cycle basis $\mathcal{C}$ of the event activity network (Liebchen and Peeters, 2009). We add

$$\Gamma_c \delta = T \cdot \mu_c \quad \forall c \in \mathcal{C} \tag{3.3}$$

to the constraints, using an integer cycle variable $\mu_c \in \mathbb{Z}$. The vector Γ_c indicates all forward or backward edges in cycle c, and T denotes the length of the period. The objective of most timetabling formulations is to minimize the total travel time of passengers. Mostly, this is achieved with the help of passenger weights x_{ij} on each activity ij and by minimizing

$$\sum_{ij \in \mathcal{A}} x_{ij} \cdot \delta_{ij}. \tag{3.4}$$

Note that the passenger weights x_{ij} are predetermined by assigning passengers to routes before optimization. The cyclic periodicity formulation for timetabling with predetermined passenger routes uses Constraints (3.2) and (3.3) and is given by

$$\begin{array}{lrcll}
\min & \multicolumn{4}{l}{\sum_{ij \in \mathcal{A}} x_{ij} \cdot \delta_{ij}} \\
\text{s.t.} & \delta_{ij} & \geq & l_{ij} & \forall ij \in \mathcal{A} \\
 & \delta_{ij} & \leq & u_{ij} & \forall ij \in \mathcal{A} \\
 & \Gamma_c \delta & = & T \cdot \mu_c & \forall c \in \mathcal{C} \\
 & \delta_{ij} & \in & \mathbb{Z}_+ & \forall i \in \mathcal{E} \\
 & \mu_c & \in & \mathbb{Z}_+ & \forall c \in \mathcal{C}
\end{array}$$

3.3.3 Integration of passenger distribution and timetabling

Section 3.3.1 defines the logit model to estimate passengers' route choice for a given timetable, and Section 3.3.2 provides a standard model to optimize a timetable for a predetermined passenger route choice. Since the result of one model is the input for the other and vice versa, we aim at developing a model integrating both aspects.

We assume that for each OD pair k a finite choice set P_k of n_k possible paths is given. Each path $p \in P_k$ is a sequence of events in the event activity network that could be taken by the passengers of OD pair k. Since these paths are defined in the event activity network, two paths for one OD pair can be different although they might use the same stations and tracks. In fact, such two paths do not need to have a single event in common. The passenger weight x_{ij} on each activity ij is not assumed to be predetermined as in the timetabling program introduced in Section 3.3.2, but we derive it from the distribution on the paths. To this end, we compute the respective length

$$t_p = \sum_{ij \in p} \delta_{ij} \quad \forall p \in P_k, \ \forall k \in OD \tag{3.5}$$

of each path for all OD pairs as the sum of durations of the activities. Note that the definition of t_p can easily be extended by additional external influencing factors such as a fare for taking path p or a penalty for each transfer included in path p. Since the path choice sets for OD pairs are assumed to be given, fares or transfer penalties can be determined in a preprocessing step for each path and are constant in the model formulation. As these constants added to t_p do not affect the structure of the model, they are omitted in the problem formulation for ease of exposition. Given the path lengths t_p, we can use the logit distribution w_p^{lm} to compute a share of each OD pair using the path p. Multiplied by the number of passengers o_k of OD pair k, this yields the number of passengers on each activity ij using path p, which we denote by

$$x_{ij}^p = w_p^{lm}((t_q)_{q \in P_k}) \cdot o_k \quad \forall ij \in p, \ \forall p \in P_k, \ \forall k \in OD. \tag{3.6}$$

This is an expected value and not necessarily integral. By aggregating these numbers over all paths p for each OD pair, we obtain the number of passengers on each activity

$$x_{ij} = \sum_{k \in OD} \sum_{p \in P_k} x_{ij}^p \quad \forall ij \in \mathcal{A}. \tag{3.7}$$

As in the timetabling formulation from Section 3.3.2, this number is used in the objective function to find a travel time minimal timetable. We formulate a general optimization problem for timetabling assuming that a passenger distribution can be modeled with a logit model:

$$
\begin{array}{llcll}
\min & & & \sum_{ij \in \mathcal{A}} x_{ij} \cdot \delta_{ij} & \\
\text{s.t.} & \delta_{ij} & \geq & l_{ij} & \forall ij \in \mathcal{A} \\
& \delta_{ij} & \leq & u_{ij} & \forall ij \in \mathcal{A} \\
& \Gamma_c \delta & = & T \cdot \mu_c & \forall c \in \mathcal{C} \\
& t_p & = & \sum_{ij \in p} \delta_{ij} & \forall p \in P_k,\ \forall k \in OD \\
& x_{ij}^p & = & w_p^{lm}((t_q)_{q \in P_k}) \cdot o_k & \forall ij \in p,\ \forall p \in P_k,\ \forall k \in OD \\
& x_{ij} & = & \sum_{k \in OD} \sum_{p \in P_k} x_{ij}^p & \forall ij \in \mathcal{A} \\
& \delta_{ij} & \in & \mathbb{Z}_+ & \forall i \in \mathcal{E} \\
& \mu_c & \in & \mathbb{Z}_+ & \forall c \in \mathcal{C} \\
& x_{ij} & \in & \mathbb{R}_+ & \forall ij \in \mathcal{A} \\
& x_{ij}^p & \in & [0, o_k] & \forall ij \in \mathcal{A},\ \forall p \in P_k,\ \forall k \in OD \\
& t_p & \in & \mathbb{Z}_+ & \forall p \in P_k,\ \forall k \in OD
\end{array}
$$

Note that the variables δ and t can be relaxed to be continuous since the lower l and upper bounds u are integer and $\mathcal{C}$ is an integral cycle basis. No matter which domain is chosen, this formulation cannot be solved efficiently due to the passenger distribution function w_p^{lm}. Furthermore, the objective is non-linear in the variables since the passenger loads x are modeled to be dependent on the durations δ.

3.4 Models

Already Parbo et al. (2014) argued that the problem from Section 3.3.3 is "extremely difficult to solve mathematically, since the timetable optimization is a non-linear non-convex mixed-integer problem, with passenger flows defined by the route choice model, where the route choice model is a non-linear non-continuous mapping of the timetable." In this section, we describe two different representations of the route choice model. Using these, we introduce two linear formulations for the problem of finding travel-time minimal routes under the assumption that passengers' routes choice can be modeled using a logit model.

3.4.1 Model 1 - Timetabling with linear distribution model

The model from Section 3.3.3 is not tractable because of the integration of the non-linear formulation of the logit model to derive a passenger distribution. In a first model, we use a novel linear passenger distribution model that is developed inspired by characteristics of the logit model. Furthermore, the quadratic objective is linearized. We address these two details in the following and provide a mixed-integer linear program for timetabling with an integrated passenger distribution model.

Linear distribution model

Using the non-linear analytic expression of the logit distribution from Equation (3.1) as a distribution model in the program of Section 3.3.3 yields an intractable optimization program. The literature provides multiple linearizations of the logit model for applications in Operation Research. To our best knowledge, these linearizations can be classified into two cases. Either, just the utility of a single alternative is variable while the utilities of all remaining alternatives are fixed. Or, the utilities of all alternatives for customers are fixed and the decision is whether to offer alternatives or not. Since in our case all alternative paths are always available and their utility depends on the timetable, these linearizations are not appropriate.

Therefore, we develop a linear distribution model to approximate the logit model by requiring appropriate characteristics for the linear functions. Our model allows all utilities to be flexible in their domain, i.e., $t_p \in [\underline{m}_k, \overline{m}_k]\ \forall p \in P_k$, and satisfies the probability characteristics. For each OD pair $k \in OD$, we require the following five characteristics.

Distribution characteristics

$$w_p((t_q)_{q\in P_k}) \in [0,1] \text{ and } \sum_{p\in P_k} w_p((t_q)_{q\in P_k}) = 1 \tag{3.8}$$

Monotonicity Let $|P_k| > 1$, let $\varepsilon > 0$ and let e_p be the unit vector with a 1 at the position of path p. Then

$$w_p((t_q)_{q\in P_k} + \varepsilon \cdot e_p) < w_p((t_q)_{q\in P_k}) \tag{3.9}$$

Uniform distribution on equivalent alternatives

$$w_p((t_q)_{q\in P_k}) = \frac{1}{|P_k|}, \tag{3.10}$$

if all paths have the same length, that is, $t_p = t_q\ \forall q \in P_k$.

Independence of order Let $\pi_p: P_k \to P_k$ be any permutation on a set of paths P_k that keeps path p constant, i.e., $\pi_p(p) = p$. Then

$$w_p((t_q)_{q\in P_k}) = w_p((t_{\pi_p(q)})_{q\in P_k}) \tag{3.11}$$

Logit characteristic: absolute utility differences determine probability

$$w_p((t_q + \hat{t})_{q\in P_k}) = w_p((t_q)_{q\in P_k}), \tag{3.12}$$

where $\hat{t} \in \mathbb{R}_+$ is a constant.

This yields a family of linear distribution functions.

Lemma 3.1. *Let $n_k = |P_k|$ be the number of alternative paths and let $\underline{m}_k$ and $\overline{m}_k$ be the minimal and maximal possible length of any considered path in the event activity network for OD pair k, respectively. Then all linear distribution functions fulfilling the five characteristics* (3.8) *to* (3.12) *can be characterized according to the three following cases:*

I $n_k = 1$:
If there is just one path p for OD pair k given, then $P_k = \{p\}$ and

$$w_p((t_p)) = 1. \tag{3.13}$$

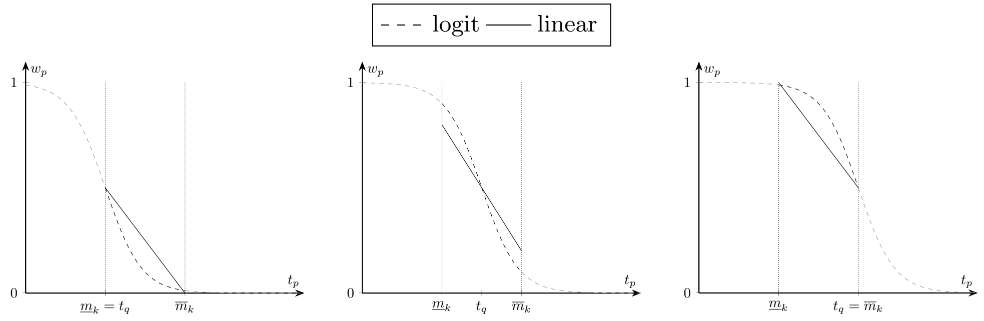
II $n_k \neq 1$ and $\underline{m}_k = \overline{m}_k$:
If $\underline{m}_k = \overline{m}_k$, all paths have the same fixed length, i.e., $t_p = t_q \;\; \forall p, q \in P_k$. Then,

$$w_p((t_q)_{q\in P_k}) = w_p((t_p, \dots, t_p)) = \frac{1}{n_k}. \tag{3.14}$$

III $n_k \neq 1$ and $\underline{m}_k \neq \overline{m}_k$:
In the general case all linear functions with the required characteristics have the form

$$w_p((t_q)_{q\in P_k}) = \frac{\alpha}{n_k(\underline{m}_k - \overline{m}_k)} \left(t_p - \frac{1}{n_k - 1} \sum_{\substack{q\in P_k \\ q\neq p}} t_q \right) + \frac{1}{n_k} \tag{3.15}$$

with $\alpha \in (0, 1]$.



(a) Probability that path p is chosen, given that length t_q of path q is as short as possible

(b) Probability that path p is chosen, given that length t_q of path q is between bounds $\underline{m}_k$ and $\overline{m}_k$

(c) Probability that path p is chosen, given that length t_q of path q is as long as possible

Figure 3.1: Probabilities of a logit and a linear distribution model that path p is chosen, given an alternative path q with fixed length t_q

A constructive proof for Lemma 3.1 is given in Appendix 3.B. We replace the logit model by the linear distribution functions (3.13), (3.14), and (3.15) in their respective cases in the model from Section 3.3.3. This yields a linearly constrained feasible region of the optimization problem and further ensures that the five characteristics (3.8) to (3.12) hold.

The linear distribution function is defined in the range $[\underline{m}_k, \overline{m}_k]$ for the length t_q of each path q. The slope of the function in that domain can be adjusted with the parameter $\alpha \in (0, 1]$. For example, for $\alpha \to 0$, we approximate the uniform distribution, independent of the path lengths. The higher α, the more do passengers react to differences in path lengths. In experiments, we learned that the linear distribution function from Lemma 3.1 tends to distribute passengers more evenly on paths than a logit distribution. Therefore, we use a value of $\alpha = 1$ to scale the linear distribution function in all experiments.

Figure 3.1 visualizes the probabilities that path p is chosen according to a logit and a linear distribution model, given a second path q with fixed length t_q. To better demonstrate the linear distribution model, three cases for the fixed path length t_q are considered. For example, in Figure 3.1a it is assumed that the length of the alternative path q is as short as possible, i.e., $t_q = \underline{m}_k$. Then, the probability that path p is chosen is at most 0.5 since it cannot be shorter than path q. The higher the length of path q, the higher the probability that path p is chosen, see Figures 3.1b and 3.1c.

This figure also illustrates how the probability of the logit distribution can be overestimated or underestimated by the linear distribution model. Knowing the length t_q of the alternative path q, a better linear approximation of the logit model is possible. However, since the utilities of all alternatives depend on the timetable, a linear distribution model can only rely on the bounds $\underline{m}_k$ and $\overline{m}_k$.

Reformulation of the model

Using the linear distribution functions from Lemma 3.1 instead of the logit distribution allows us to express the number of passengers on each activity x_{ij} as a linear function of the durations δ_{ij}. We obtain the quadratic integer program for timetabling with **I**ntegrated passenger **D**istribution according to a **LIN**ear distribution model (ID-LIN):

$$
\begin{array}{llcll}
\min & \delta^\dagger A\delta + b^\dagger\delta & & & \\
\text{s.t.} & \delta_{ij} & \geq & l_{ij} & \forall ij \in \mathcal{A} \\
 & \delta_{ij} & \leq & u_{ij} & \forall ij \in \mathcal{A} \\
 & \Gamma_c\delta & = & T\cdot\mu_c & \forall c \in \mathcal{C} \\
 & \delta_{ij} & \in & \mathbb{Z}_+ & \forall i \in \mathcal{A} \\
 & \mu_c & \in & \mathbb{Z}_+ & \forall c \in \mathcal{C}
\end{array}
$$

where $\delta^\dagger$ and $b^\dagger$ denote the transpose of the column vectors δ and b, respectively. The coefficients in the objective function are defined as

$$
A_{ij,i'j'} \coloneqq \sum_{k\in OD^*} \frac{\alpha\cdot o_k}{n_k(\underline{m}_k - \overline{m}_k)} \left(\sum_{\substack{p\in P_k:\\ ij,i'j'\in p}} 1 - \sum_{\substack{p\in P_k:\\ ij\in p}} \sum_{\substack{q\neq p\in P_k:\\ i'j'\in q}} \frac{1}{n_k - 1} \right), \quad \forall ij, i'j' \in \mathcal{A} \tag{3.16}
$$

and

$$
b_{ij} \coloneqq \sum_{k\in OD} \sum_{p\in P_k: ij\in p} \frac{o_k}{n_k}, \quad \forall ij \in \mathcal{A}.
$$

In equation (3.16), OD^* denotes the set of OD pairs k with $n_k > 1$ and $\overline{m}_k \neq \underline{m}_k$. This means, that only OD pairs with multiple paths contribute to the matrix A, and thus add to the quadratic part of the objective function. OD pairs with only one path, or with multiple paths of fixed length, only add to the linear part of the objective. The derivation of the coefficient matrix A and vector b can be found in Appendix 3.C.

(ID-LIN) is a minimization program and the coefficient matrix A can be proven to be negative semi-definite, see Appendix 3.D. That means the objective function is concave and standard methods for quadratic programs are not expected to be efficient. We therefore apply a linearization to the objective function. To this end, we express the integer variables δ_{ij} as a sum of binary variables,

$$\delta_{ij} = l_{ij} + \sum_{m=0}^{\lfloor \log(u_{ij}-l_{ij}) \rfloor} 2^m \sigma_{ij}^m$$

and linearize the products of binaries $\sigma_{ij}^m \cdot \sigma_{i'j'}^{m'}$. The corresponding linearization of the optimization program (ID-LIN) as used for the experiments can be found in Appendix 3.E.

3.4.2 Model 2 - Simulation of logit model

In a second model, we integrate a simulated passenger distribution into the timetabling framework. The simulation is based on an alternative way to compute the logit probabilities. According to Train (2009), it holds that

$$w_p((t_q)_{q \in P_k}) = \frac{e^{\beta t_p}}{\sum_{q \in P_k}^{n} e^{\beta t_q}} = Prob\left(t_p + \varepsilon_p \leq \min_{q \in P_k}(t_q + \varepsilon_q)\right), \tag{3.17}$$

where the ε_p are independent and identically Gumbel distributed. That means the logit probability that alternative p is chosen equals the probability that the length of path p, deferred by some random value ε_p, is shorter than the length of any alternative path q, deferred by some random value ε_q. Following similar steps as Pacheco et al. (2016), we use the representation in Equation (3.17) to simulate the logit model by drawing random values for ε. That means we consider several scenarios $r \in R$, draw a random value ε_{pr} for each path p in each scenario r, and add these to the path lengths. This yields a different, randomized path length in each scenario, which we denote by

$$t_{pr} = \sum_{ij \in p} \delta_{ij} + \varepsilon_{pr} \quad \forall k \in OD,\ \forall p \in P_k,\ \forall r \in R.$$

Note that similar to the path length computation in Equation (3.5), this modeling can easily be extended by additional factors like fares or a penalty for each transfer as well. Then, we choose the shortest path in each scenario for each OD pair and denote the travel time for OD pair k in scenario r by

$$t_{kr} = \min_{p \in P_k} t_{pr} \quad \forall k \in OD,\ \forall r \in R. \tag{3.18}$$

This discrete choice of the shortest path in each scenario r yields a distribution of the passengers of OD pair k over the available paths in the path choice set P_k. Since we choose the random terms ε_{pr} to be independent and identically Gumbel distributed, this distribution converges towards a logit distribution for an increasing number of scenarios, see Equation (3.17).

Using a binary choice variable z_{pr} that is set to one if and only if path p is the shortest in scenario r, constraint (3.18) can be linearized to

$$
\begin{aligned}
t_{kr} &\le t_{pr} && \forall k \in OD,\ \forall p \in P_k,\ \forall r \in R \\
t_{kr} &\ge t_{pr} - (1 - z_{pr}) M_{kr} && \forall k \in OD,\ \forall p \in P_k,\ \forall r \in R \\
\sum_{p \in P_k} z_{pr} &= 1 && \forall k \in OD,\ \forall r \in R
\end{aligned}
$$

where

$$M_{kr} = \max_{p \in P_k} \left(\sum_{ij \in p} u_{ij} + \varepsilon_{pr} \right) - \min_{p \in P_k} \left(\sum_{ij \in p} l_{ij} + \varepsilon_{pr} \right)$$

is sufficiently large.

Note that if in a scenario two paths are the shortest, this modeling will do a random assignment of the passenger choice. We obtain the model for timetabling with an **I**ntegrated passenger **D**istribution by **SIM**ulation of the logit model (ID-SIM):

$$
\begin{array}{llcll}
\min & \displaystyle\sum_{k \in OD} o_k \frac{1}{|R|} \sum_{r \in R} t_{kr} & & & \\
\text{s.t.} & \delta_{ij} & \ge & l_{ij} & \forall ij \in \mathcal{A} \\
 & \delta_{ij} & \le & u_{ij} & \forall ij \in \mathcal{A} \\
 & \Gamma_c \delta & = & T \cdot \mu_c & \forall c \in \mathcal{C} \\
 & t_{pr} & = & \sum_{ij \in p} \delta_{ij} + \varepsilon_{pr} & \forall k \in OD,\ \forall p \in P_k,\ \forall r \in R \\
 & \sum_{p \in P_k} z_{pr} & = & 1 & \forall k \in OD,\ \forall r \in R \\
 & t_{kr} & \le & t_{pr} & \forall k \in OD,\ \forall p \in P_k,\ \forall r \in R \\
 & t_{kr} & \ge & t_{pr} - (1 - z_{pr}) M_{kr} & \forall k \in OD,\ \forall p \in P_k,\ \forall r \in R \\
 & \delta_{ij} & \in & \mathbb{Z}_+ & \forall ij \in \mathcal{A} \\
 & \mu_c & \in & \mathbb{Z} & \forall c \in \mathcal{C} \\
 & t_{pr} & \in & \mathbb{R}_+ & \forall k \in OD,\ \forall p \in P_k,\ \forall r \in R \\
 & t_{kr} & \in & \mathbb{R}_+ & \forall k \in OD,\ \forall r \in R \\
 & z_{pr} & \in & \{0, 1\} & \forall k \in OD,\ \forall p \in P_k,\ \forall r \in R
\end{array}
$$

The constraints and the objective function of this formulation are linear in the variables.

There is a trade-off between the solvability of the MILP model (ID-SIM) and the accuracy of the simulation. Considering only a few scenarios results in a small model which, however, yields a random solution because a path could be privileged or disadvantaged by chance. With an increasing number of scenarios, we expect the passenger distribution on paths to converge and the solution to stabilize, but the model size and hence solution time to increase. To choose a setting that balances solvability and accuracy, we ran preliminary experiments with varying numbers of scenarios. Based on this, we choose to use a low number of $|R| = 10$ scenarios and pick the best solution of 10 repetitions instead of using a large number of scenarios. In our experiments, this has been shown to yield a good trade-off between computation time and a high probability to find a solution of high quality. Another advantage of solving each instance multiple times with a small number of scenarios over considering large scenario sets is the independence of repetitions that can easily be parallelized.

3.4.3 Illustration of model differences

In this section, we compare the two models (ID-LIN) and (ID-SIM) concerning their objective functions. The objective function of (ID-LIN) is the sum of the absolute travel times of all passengers on their respective paths, which are chosen based on the linear distribution function introduced in Lemma 3.1. This distribution assumption implies that not everyone travels on a shortest path, but passengers make use of paths with slightly longer travel times than the shortest as well. Combining the distribution of passengers on multiple paths with the objective to minimize absolute travel time can have undesirable consequences, as can be seen in the following example.

Example 1. *Consider a network consisting of two stations A and B and one OD pair k that wants to travel from A to B. Assume, there are two available paths, p and q, with respective bounds* $[10, 22]$ *and* $[11, 21]$. *The example network is illustrated in Figure 3.2.*

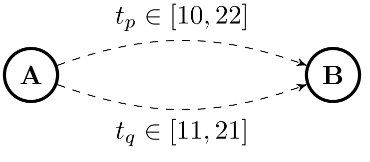


Figure 3.2: Example network

	(t_p, t_q)	Shortest path	Logit distribution	Linear distribution
t^1	$(11, 11)$	11	11	11
t^2	$(10, 13)$	10	11.02	11.13
t^3	$(10, 21)$	10	10.90	10.46

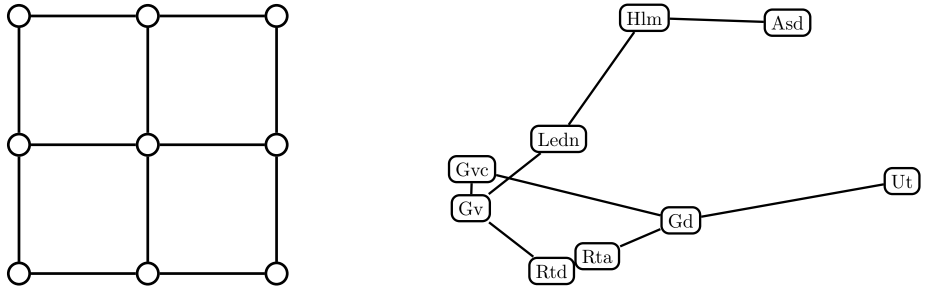
Table 3.2: Comparison of travel times of three timetables w.r.t. different distributions

We compare three different timetables t^1, t^2 and t^3. The first timetable offers two equally good paths, these are $t^1_p = 11 = t^1_q$. The second and third timetable has one short path and one longer alternative. These are $t^2_p = 10$, $t^2_q = 13$ and $t^3_p = 10$, $t^3_q = 21$, respectively.

We evaluate the three timetables with the travel time on the passengers' respective shortest path, the travel time when assuming that passengers distribute according to a logit distribution, and the travel time when assuming that passengers distribute according to the linear distribution model from Lemma 3.1. For the logit and linear distribution models, we use the parameters $\beta = -0.22$ and $\alpha = 1.0$, respectively. The objective values of one passenger of OD pair k can be found in Table 3.2.

We find, as expected, that the travel time on the shortest path is best in timetables t^2 or t^3, regardless of the length of alternative q. Regarding travel time according to a linear or logit distribution, timetable t^2 is worse than timetable t^1. This result is open for discussion as none of the two timetables is obviously better than the other. However, it is striking that timetable t^3 is better than timetable t^2 according to a linear or logit distribution. This result is undesired for evaluation purposes and might be unexpected at first glance, but it has a simple explanation: The worse the travel time t_q of alternative q, the more probable it is that passengers choose to travel via path p, which yields a lower total travel time.

The objective of the second model (ID-SIM) is to minimize the weighted sum of randomized shortest path lengths t_{kr} instead of the absolute travel time as used in the first model (ID-LIN). In this formulation, a path only enters the objective function if it is perceived better than any alternative in at least one scenario. Hence, no considered path in the path choice set can deteriorate, but only improve the objective value. That means the objective function of the program (ID-SIM) does not have a bias towards the undesirable effects demonstrated in Example 1.



(a) 3x3 grid infrastructure (b) Partial network of Netherlands Railways

Figure 3.3: The methods are compared on instances defined on these two infrastructures

3.5 Experimental setup

3.5.1 Instances

To test and compare our approaches, we run experiments on a number of instances. Each instance $\mathcal{I}$ consists of an event activity network $\mathcal{N}$ with lower and upper bounds l and u and a demand situation. The event activity network is derived from information about the public transport network, i.e., stations and tracks, as well as a line plan. Both models (ID-LIN) and (ID-SIM) assume a choice set of paths P_k for each OD pair k to be given. The paths are defined in the event activity network and can be interpreted as a sequence of line trips. Depending on the line plan, there can be multiple paths on the same geographical route or just a single path although origin and destination are connected by several geographical routes. Hence, a track network can indicate but does not determine the number of passenger paths for an OD pair. How we preprocess the instances and derive a path choice set is described in Appendix 3.F.

Instances on grid network

We consider 32 instances defined on a 3×3 grid network, which is depicted in Figure 3.3a. On this network, we consider four different demand situations, and for each of them, several line plans with corresponding event activity networks. The number of events and activities in the corresponding timetabling instances range from 120 to 208 and 326 to 760, respectively. The instances are partial instances of a bigger grid network introduced by Friedrich et al. (2017a) and made available in an online repository[2]. The grid infrastructure has several geographically different routes of comparable length for passengers. Depending on the line plan, this provides good conditions to find multiple passenger paths in the event activity network. On the 32 instances, there are on average 1.7 paths for each OD pair with a maximum of 8 paths for one OD pair across all instances. On average, 29.9 OD pairs have more than one path.

Instance on Dutch railway network

To test our approaches on a real-world instance, we consider a part of the Dutch railway network operated by Netherlands Railways (NS). The partial network includes the stations Amsterdam Centraal (Asd), Den Haag Centraal (Gvc), Den Haag HS (Gv), Gouda (Gd), Haarlem (Hlm), Leiden Centraal (Ledn), Rotterdam Alexander (Rta), Rotterdam Centraal (Rtd), and Utrecht Centraal (Ut) in the Randstad, a metropolitan region in the Netherlands. The track network is depicted in Figure 3.3b. We consider eight Intercity lines operating between the stations, yielding 128 events and 357 activities for the timetabling model. Based on this, 1 to 7 paths are available per OD pair, with an average of 2.4 paths. On the Dutch railway network, 40 OD pairs have multiple available paths. Both the number of OD pairs with multiple paths and the average number of paths per OD pair are higher than in the grid network although this network contains fewer cycles. This indicates that the optimization problem for the Dutch instance is larger and thus potentially harder to solve.

3.5.2 Timetabling approaches

We compare the timetabling models with integrated passenger distribution (ID-LIN) and (ID-SIM) with three state-of-the-art methods for timetabling: two methods (PS) and (PD) assume a predetermined passenger assignment to routes, and one method (IS) has an integrated passenger routing on the shortest paths. Besides the timetabling models (ID-SIM) and (ID-LIN) that integrate the passenger distribution, we also test and compare a heuristic solution approach (ID-ITR) for timetabling with passenger distribution. These approaches are described in more detail below.

(PS) First, a timetabling model with **P**redetermined passenger assignment on a **S**ingle path is considered. In this model, the passengers' routes are fixed before the optimization step. We assign passengers to the shortest

[2] https://github.com/FOR2083/PublicTransportNetworks/tree/master/Grid_5x5

route using the average bounds $\frac{1}{2}(l_{ij} + u_{ij})$ on edges in the event activity network. This basic version of the timetabling model is the subject of many publications since the development of the PESP model, see for example Nachtigall and Opitz (2008) or Liebchen (2018). An integer programming formulation is given by Equations (3.2) to (3.4), as described in Section 3.3.2.

(PD) Second, we consider another model with **P**redetermined passenger routes. In contrast to the model (PS), passengers are **D**istributed on multiple paths according to a logit model with the parameter $\beta = -0.22$ and using average bounds on edges. In consultation with traffic engineers, the value of β is chosen similar to values that are typically found when fitting the logit model on instances with similar travel distances. We are not aware of a published timetabling approach that explicitly states a predetermined passenger distribution according to a logit model. Still, this strategy can be compared to those made in Parbo et al. (2014) or Robenek et al. (2016), where passenger distributions were derived from utilities of alternative routes. The underlying integer programming model is all the same as the one in (PS), only the passenger weights are predetermined differently.

(IS) Third, we consider a timetabling model with **I**ntegrated **S**hortest path search. The timetable is optimized with the objective of minimizing passenger travel times for passengers that choose the shortest path based on the timetable. This approach resembles the idea of the integrated shortest path models described in Siebert and Goerigk (2013), Gattermann et al. (2016) and Borndörfer et al. (2017), for example. An integer programming formulation of this model is attached in Appendix 3.G.1.

(ID-ITR) Fourth, we consider a heuristic approach for timetabling with **I**ntegrated passenger **D**istribution that **IT**e**R**ates between timetable design and passenger distribution. To compute the passenger distribution based on a fixed timetable, we use the logit model with the parameter $\beta = -0.22$. The initial passenger loads are determined by using the average bounds as edge lengths. In all following iterations, the realized edge lengths of the timetable are used. This yields fixed passenger loads on each edge in the event activity network in each iteration and a standard timetabling model assuming a predetermined passenger distribution can be solved with the given loads. We iterate until the solution value does not change significantly between two iterations or a maximum number of iterations is reached. Similar iterative approaches for timetabling and passenger route choice are described in Sels et al. (2011) or Parbo et al. (2014), for example. The pseudocode for this method can be found in Appendix 3.G.2.

We refer to these benchmark models by (PS), (PD), (IS), and (ID-ITR), respectively. Table 3.3 indicates whether the route choice is integrated into the methods as well as which kind of route choice model is assumed.

	Predetermined route choice	**I**ntegrated route choice
Single route	(PS)	(IS)
Distribution	(PD)	(ID-ITR), (ID-LIN), (ID-SIM)

Table 3.3: Summary indicating which solution approach (1) assumes a predetermined route choice or has an integrated route choice and (2) assumes that passengers use a single route only or distribute on multiple routes

With a comparison of the models (ID-LIN) and (ID-SIM) with the heuristic approach (ID-ITR) and the three benchmark models (PS), (PD), and (IS), we can identify the benefits of integrating (1) passenger route search and (2) simultaneous modeling of a passenger distribution.

3.5.3 Implementation

To reduce the size of the search space, the domain of the variables μ_c is constrained in all models with the following inequalities.

$$\left\lceil \frac{1}{T} \sum_{ij \in c^+} l_{ij} - \sum_{ij \in c^-} u_{ij} \right\rceil \leq \mu_c \leq \left\lfloor \frac{1}{T} \sum_{ij \in c^+} u_{ij} - \sum_{ij \in c^-} l_{ij} \right\rfloor \quad \forall c \in \mathcal{C}.$$

Here, c^+ and c^- denote the set of edges in cycle c in forward and backward direction, respectively, and l_{ij} and u_{ij} are the lower and upper bounds of activity ij. These well-established inequalities were first described in Odijk (1996).

All mixed-integer linear programs are solved with the general-purpose solver Fico Xpress 8.5 on a laptop with 32 GB RAM and an Intel® Core™ i7-6700HQ. A time limit of one hour is used for the grid instances and no time limit for the Dutch railway instance. For all experiments, we use a start solution to warm start the optimization. This start solution consists of an initial timetable for the instance and, if applicable, a corresponding passenger routing according to the passenger distribution model of the solution approach used.

3.5.4 Evaluation of timetables

Different research areas apply different measures to evaluate timetables from the passengers' perspective. We could see in Example 1 that different evaluation functions can yield different results on small networks. This small example suggests two features: First, although travel time is commonly used to evaluate timetables, it might not be suitable when considering a passenger distribution on multiple routes. Second, different evaluation measures may consider different timetables to be better, although the functions are commonly accepted to serve for the evaluation of timetables. Chapter 2 compared multiple timetable evaluation functions for passengers on different instances and indeed found that these functions are often not consistent in their evaluation. We learn that there is no default objective function to be used when optimizing timetables with an integrated passenger distribution. To avoid misinterpretation of the results due to a simplistic or biased evaluation, we evaluate all resulting timetables with four different evaluation functions. As before, we denote the total passenger load of OD pair k with o_k and the length of path p with t_p, as defined in Equation (3.5). Let P_k be a set of available paths for OD pair k. The used evaluation functions are

tt_{sp} The total **t**ravel **t**ime of all passengers on their **s**hortest **p**ath:

$$tt_{sp} = \sum_{k \in OD} o_k \sum_{p \in P_k} w_p^{sp} \cdot t_p,$$

where w_p^{sp} is the probability that passengers choose path p assuming that all passengers use their shortest paths only.

tt_{mp} The total **t**ravel **t**ime of all passengers when distributed on **m**ultiple **p**aths according to the logit model:

$$tt_{mp} = \sum_{k \in OD} o_k \sum_{p \in P_k} w_p^{lm} \cdot t_p,$$

where w_p^{lm} is the probability that passengers choose path p assuming that all passengers distribute on their paths according to a logit distribution.

ut_{sum} The evaluated total **ut**ility for all passengers, defined as the weighted **sum** of all logit denominators:

$$ut_{sum} = \sum_{k \in OD} o_k \sum_{p \in P_k} e^{\beta t_p},$$

with $\beta = -0.22$. Derived from the logit model, this measure indicates utility of a public transport service for passengers. The utility of a path p is weighted with the passenger load o_k.

ut_{log} The **log**sums, a **ut**ility-based evaluation function, defined as the weighted sum of the logarithm of all logit denominators:

$$ut_{log} = \sum_{k \in OD} o_k \cdot \ln\left(\sum_{p \in P_k} e^{\beta t_p}\right),$$

with $\beta = -0.22$. Similar to the evaluated total utility, the logsums are a measure of utility for passengers. Due to the logarithm in this evaluation function, the effect of changing travel time t_p on the evaluation ut_{log} depends not only on the passenger load but also on the number of paths of that OD pair. For example, changing the travel time on one path of an OD pair with many good paths might affect the value of ut_{log} less than changing the travel time of one path of an OD pair with just a few and bad paths, even if the passenger load in the first case is higher than in the second case. That means OD pairs have different weights relative to each other, as contrasted with the evaluated total utility.

All four functions evaluate the quality of timetables from the passengers' perspective. Note that these functions are commonly used for evaluation, but due to their structure, not all are suitable as objective functions in an optimization program. The first two evaluation functions are *travel time-based* and thus to be minimized while the latter two evaluation functions are *utility-based* and hence to be maximized. Considering all four evaluation functions allows a thorough investigation and comparison of the timetables and, in this way, of the proposed timetabling methods.

For better comparability, we present the relative solution values when compared to an *ideal solution*. In an ideal solution, it is assumed that the travel time on each path for each OD pair is equal to the length of the path using the lower bounds on all edges. This is also called *lower-bound routing* of passengers, see Borndörfer et al. (2017). For most instances, such an ideal solution does not exist, but it is a common measure to see how close solutions are to perfect conditions. More details about ideal solutions and about how they are used in practice can be found in Caimi et al. (2017).

3.6 Results

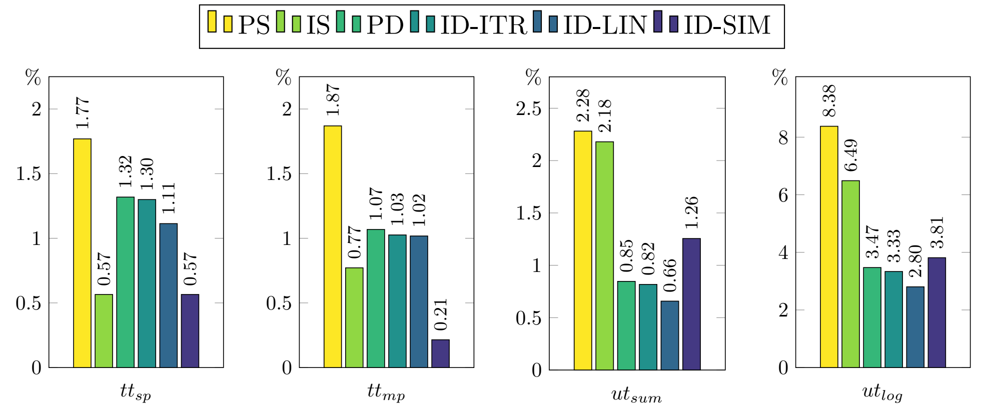
In the experiments, we showcase the benefits and drawbacks of the timetabling models with integrated passenger distribution (ID-LIN) and (ID-SIM) when compared to existing timetabling approaches.

3.6.1 Experiments on 32 instances on the grid network

We conduct experiments on 32 instances on the grid network as depicted in Figure 3.3a. On seven instances, all six methods find an ideal solution, and on another four instances, the model (ID-LIN) could not find an optimal solution or could not prove optimality in tests with a time limit of ten hours. Therefore, we exclude these 11 instances from the discussion. In Figure 3.4, we present the evaluation values of the solutions found by the different approaches averaged over the remaining 21 instances on the grid network. This figure shows the average performance of the six methods on the four considered evaluation functions introduced in Section 3.5.4. All values are given in percent, relative to the evaluation value of an ideal solution.

The relative evaluation values can be read as follows. For example, a relative value of 1.77 for tt_{sp} in Figure 3.4a of the model (PS) means that the travel time on the shortest connection in the solution of (PS) is, on average, 1.77 percent longer than the travel time on the shortest connection in an ideal solution. Comparing this to the relative travel time on a shortest connection of the model (IS), 0.57, shows that (IS) performs, on average, better than (PS) regarding the travel time on the shortest path. In general, the relative evaluation values show to what extent a solution is worse than an ideal solution, according to the used evaluation function. We discuss the results per evaluation function.

Figure 3.4a When evaluating timetables with travel time on the shortest path tt_{sp}, on average, the methods (IS) and (ID-SIM) provide the best solutions. This is expected for the method (IS) since its objective is to minimize the total travel time of passengers on their shortest paths. To simulate a logit distribution in the model (ID-SIM), in each scenario the shortest path is chosen, as modeled in Equation (3.18). It seems that in many scenarios the same path is chosen, which in turn gets assigned high weights in the objective function. The model



(a) Relative evaluation values with respect to travel time of all passengers on shortest paths

(b) Relative evaluation values with respect to travel time of all passengers on multiple paths

(c) Relative evaluation values with respect to evaluated total utility for all passengers

(d) Relative evaluation values with respect to total logsums for all passengers

Figure 3.4: The bars show the evaluation values of the six different methods relative to those of an ideal solution, averaged over 21 instances on the grid network.

(ID-LIN) finds solutions with travel times on the shortest route that are, on average, higher than those of methods (IS) and (ID-SIM) and only slightly lower than those of methods (PD) and (ID-ITR). As discussed in Section 3.4.1, the linear distribution model in (ID-LIN) tends to distribute passengers more evenly on paths than the logit model. Thus, the weights assigned to the shortest paths are lower compared to those in the models (IS) and (ID-SIM). This could explain the worse performance of (ID-LIN) regarding travel time on the shortest path. The remaining three methods, (PS), (PD), and (ID-ITR) perform worse according to travel time on the shortest path. Compared to the best found solutions, their respective travel times are up to three times as far away from an ideal solution.

Figure 3.4b In the case of evaluating travel time using a logit distribution tt_{mp}, the method (ID-SIM) performs best, which is presumably due to the simulated logit distribution of passengers. The model (ID-LIN) performs, on average, worse than (ID-SIM) and finds solutions that are only as good as those found by (PD) and (ID-ITR). This indicates that the passenger distribution of the linear distribution model used in (ID-LIN) is different from the distribution according to a logit model, which is used for evaluation. Furthermore, we can observe that the method (IS) finds better solutions than (PD) and (ID-ITR) averaged over all 21 instances. This is surprising since the methods (PD) and (ID-ITR) consider a passenger distribution according to a logit model, whereas (IS) does not consider any alternatives to the shortest route.

We identify the combination of tt_{mp} as an evaluation function and a passenger distribution on multiple routes as the reason for this observation. In the model (IS), alternative routes might get assigned high travel times, which implies a low utilization of these routes in a subsequent distribution of passengers according to the logit model. As shown in Example 1 with the comparison of timetables t^2 and t^3, this can result in lower total travel times for passengers than providing low travel times on all alternative routes. Indeed, with all six methods, we find solutions on certain instances with negative relative evaluation values for tt_{mp}, implying that the found solutions are 'better' than an ideal solution. As in Example 1, this finding appears unexpected at first glance and is undesired for evaluation. This questions whether the total (or average) travel time of passengers, while assuming that passengers distribute over multiple routes in the network, is a valid evaluation function for public transport timetables.

Figure 3.4c The evaluation with the evaluated total utility ut_{sum} shows a different pattern. The methods (PD), (ID-ITR), (ID-LIN), and (ID-SIM) outperform the methods (PS) and (IS). The gap to the evaluation value of an ideal solution is more than halved. On average, the method (ID-LIN) finds the best solutions, almost halving the gap to the ideal solution once more compared to the model (ID-SIM). This is contrary to the observations made with the travel time-based evaluation functions tt_{sp} and tt_{mp} where (ID-SIM) performs better than (ID-LIN), see Figures 3.4a and 3.4b. A similar observation can be made for the model (IS). While it performs very well on the travel time-based evaluation functions, (IS) yields solutions that are among the worst according to the evaluated total utility.

Figure 3.4d We make similar observations with the total logsums ut_{log} as the evaluation function. Also here, the methods (PD), (ID-ITR), (ID-LIN), and (ID-SIM) find better solutions than the methods (PS) and (IS). However, when evaluating the found timetables with the total logsums, the gaps to an ideal solution are by far larger. Furthermore, the solutions of (IS) are, on average, rated better than those of (PS), which is not visible with the other utility-based evaluation function ut_{sum} in Figure 3.4c.

Cross-figure discussion As indicated in Table 3.3, we consider four different categories of modeling passengers in optimization approaches for timetabling. They result from a combination of (1) whether a predetermined route choice is assumed or a route choice model is integrated into optimization and (2) whether passengers are assumed to use a single route only or to distribute on multiple routes.

With the utility-based evaluation functions, ut_{sum} and ut_{log}, our experiments show that the quality of timetables can be considerably improved by considering multiple routes instead of a single route for passengers. All four methods that consider a passenger distribution on multiple routes find solutions with a significantly lower gap to an ideal solution than the two models that assume passengers to use a single route only. In comparison, the integration of a passenger route choice model, as opposed to a predetermined route assignment, did not help to improve the quality of the found timetables according to the utility-based evaluation functions. Only the solutions of (IS) are, on average, slightly better than those of (PS), but the others were not in comparison to (PD).

Regarding the travel time-based evaluation functions, tt_{sp} and tt_{mp}, the methods with an integrated route choice model find better timetables than the corresponding single or multiple route methods that assume a predetermined route choice. Especially the models (IS) and (ID-SIM) could find timetables with significantly better travel time on the shortest path and the latter also on a logit distribution. Considering multiple routes for passengers during optimization instead of only one route yields better solutions for tt_{mp}, but not necessarily for tt_{sp} since there just the shortest path is considered for evaluation. Moreover, although the method (PD) finds, on average, better solutions than (PS), (PD) is outperformed by all other methods regarding travel time-based evaluation functions. In our experiments, considering multiple routes for passengers is not sufficient to find timetables with best travel times.

We find that considering a passenger distribution on multiple routes mainly improves the utilities, and integrating a passenger route choice model mainly improves the travel times of the found timetables. Furthermore, by integrating a passenger distribution model, it is possible to find solutions with multiple good routes that yield both

Method	CPU time	no. instances
(PS)	0.4s	21/21
(IS)	5.6s	21/21
(PD)	0.8s	21/21
(ID-ITR)	0.8s	21/21
(ID-LIN)	1952.9s	17/21 (5.68)
(ID-SIM)	1184.0s	19/21 (1.17)

Table 3.4: Average CPU times and the number of instances that were solved within one hour. The remaining gap to the best bound after one hour is given in parentheses.

good travel times and high utilities for passengers on the considered instances. The model (ID-SIM) provided the best solutions regarding the travel time-based evaluation functions and comparable solutions with respect to one utility-based evaluation function. The model (ID-LIN) could not perform as well as one state-of-the-art approach according to the travel time-based evaluation functions but provided the solutions with the best utilities. Thus, by integrating a passenger distribution model, it is possible to find better timetables than the benchmark methods regarding some evaluation functions while maintaining the quality regarding some other evaluation functions.

These improvements by the integration of a passenger distribution model come at the expense of significantly larger models. Table 3.4 shows the average solution times of the six different methods on the discussed 21 instances on the grid network. From the computation times, it is apparent that the two proposed models (ID-LIN) and (ID-SIM) need by far the most time for solving the instances. It took almost 20 minutes to solve the model (ID-SIM) and more than 30 minutes to solve the model (ID-LIN), on average. The other methods were solved within a few seconds.

The second column displays the number of instances that were solved within one hour. The model (ID-LIN) could only find optimal solutions for 17 of the 21 instances and (ID-SIM) provided optimal solutions for 19 instances. After one hour, the model (ID-LIN) had, on average, a gap of more than 5% to the best bound, whereas the simulation-based model was close to an optimal solution with a remaining gap of a little more than 1%. The other four methods were always able to terminate within one hour.

Method	CPU time
(PS)	1.4s
(IS)	9.8s
(PD)	30.3s
(ID-ITR)	35.7s
(ID-LIN)	14007.6s
(ID-SIM)	9150.2s

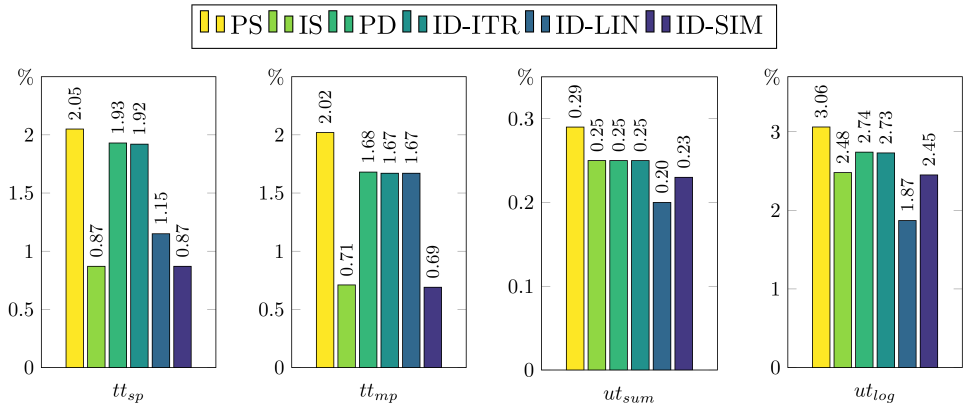
Table 3.5: CPU times for solving the Dutch railway instance with the six different methods.

3.6.2 Experiments on Dutch railway network

We also compare the six different methods on a part of the network of Netherlands Railways as depicted in Figure 3.3b. Table 3.5 shows that the solution times for the Dutch railway instance are generally higher compared to the solution times of the instances on the grid network. Model (ID-LIN) required almost four hours to be solved to optimality, and model (ID-SIM) took on average two and a half hours for solving, where three random scenarios could be solved in less than one hour.

In Figure 3.5, the evaluation values of all methods are given relative to those of an ideal solution. We observe in Figures 3.5a and 3.5b that two models with an integrated passenger route choice model, (IS) and (ID-SIM), perform best. The gap to an ideal solution is significantly lower compared to the other methods. This is in line with the observation made in the evaluation by the travel time-based evaluation functions on the grid instances and demonstrates once more the benefits of integrating a passenger route choice model into timetabling optimization. The model (ID-LIN) provides a solution with higher travel times, but it has notably shorter travel times than the remaining methods on the shortest path and comparable travel times assuming a passenger distribution.

The relative evaluation values of the utility-based evaluation functions in Figures 3.5c and 3.5d suggest that the method (ID-LIN) performs best, as it was observed on the grid instances. In contrast to the instances on the grid network, there seems to be no visible advantage of the methods that consider a passenger distribution on multiple routes over the methods that assume that passengers use only a single route. Instead, the method (IS) performs better than the two methods (PD) and (ID-ITR) with respect to the logsums.



(a) Relative evaluation values with respect to travel time of all passengers on shortest paths

(b) Relative evaluation values with respect to travel time of all passengers on multiple paths

(c) Relative evaluation values with respect to evaluated total utility for all passengers

(d) Relative evaluation values with respect to total logsums for all passengers

Figure 3.5: The bars show the evaluation values of the six different methods relative to those of an ideal solution on a partial network of Netherlands Railways

We find that the solutions found by (IS) and (ID-SIM) dominate the solutions found by all other methods regarding the travel time-based evaluation functions, while the consideration of multiple routes brings only a slight advantage to the model (ID-SIM). According to the utility-based evaluation functions, the solution found by (ID-LIN) dominates all other solutions. Moreover, the results in Figure 3.5 demonstrate the importance of a thorough evaluation with multiple evaluation functions. Together with the results on the grid network, these experiments illustrate that an evaluation with a single evaluation function is likely to falsify the interpretation.

3.7 Conclusion

In this chapter, we study the problem of finding a travel time minimal timetable under the assumption that the distribution of passengers on available routes can be modeled using a discrete choice model. We use the logit model to estimate a passenger distribution and formulate this problem as a mixed-integer program. Based on this, we develop two mixed-integer linear programs proposing different ways to model the interaction of passenger route choice and timetable design. In the first model, we incorporate a novel multidimensional linear passenger distribution model that resembles the logit model. Our second model approximates a logit distribution of the passengers from an integrated simulation framework.

We compare the two timetabling models with integrated passenger distribution with three state-of-the-art methods and a heuristic approach that iterates between timetabling and passenger routing to find travel time-optimal timetables for passengers. The experiments are conducted on a set of artificial instances and a part of the network of Netherlands Railways. We provide a thorough comparison of all solutions with four structurally different evaluation functions.

With the integration of a passenger distribution model into a timetabling framework, we were able to find better timetables for passengers than the considered state-of-the-art methods. The gap to an ideal solution for passengers could be significantly reduced for some evaluation functions while performing similarly according to other evaluation functions. In general, the experiments give insight into how two model decisions for passenger distribution on routes affect the solution quality. The first decision examined is whether to consider multiple routes or a single route for passengers, and the second is whether route choice is integrated or the assignment of passengers to routes is predetermined.

It is interesting to observe that the different evaluation functions yield different results for the considered methods. This supports the impression that a comprehensive evaluation with multiple functions is useful and necessary to make clear statements about the quality of methods. In particular, we address observations that a commonly used evaluation function for timetables, the total travel time of passengers, in combination with a passenger distribution model might yield an undesired assessment of the timetable. Our results and a simple example raise the question of whether this function is suitable for evaluation or as an objective function when considering a distribution of passengers on multiple paths.

The integration of a passenger distribution model in both timetabling models comes at the expense of significantly higher solution times. Future research could deal with the development of solution approaches to be able to solve large instances.

Appendix

3.A Notation

Greek letters

α	Scaling parameter for linear distribution function
β	Scaling parameter for logit model
δ_{ij}	Duration of activity $ij \in \mathcal{A}$
$\varepsilon_{p,r}$	Random term to vary the path length of path p in scenario r
Γ	Indicator matrix for forward and backward edges in cycles
μ_c	Auxiliary variable for cycle constraints
σ_{ij}^m	Variable for binary representation of δ_{ij}
$\gamma_{ij,i'j'}^{m,m'}$	Variable for linearization of product $\sigma_{ij}^m \cdot \sigma_{i'j'}^{m'}$

Latin upper case letters

A	Coefficient matrix of program (ID-LIN)
$\mathcal{A}$	Set of activities in the event activity network
$\mathcal{C}$	Integral cycle basis in the event activity network
$\mathcal{E}$	Set of events in the event activity network
$\mathcal{I}$	Timetabling instance
(ID-ITR)	Iterative heuristic for timetabling problem with passenger distribution
(ID-LIN)	Timetabling model with integrated linear passenger distribution model
(ID-SIM)	Timetabling model with simulated passenger distribution model
(IS)	Timetabling model with integrated shortest path search
M_{ij}	Set of indices for binary representation of δ_{ij}
M_{kr}	Auxiliary number for linearization of choice constraints
$\mathcal{N}$	Event activity network
OD	Set of OD pairs
P_k	Set of alternative paths for OD pair k
(PD)	Timetabling method with predetermined passenger distribution
(PS)	Timetabling method with predetermined passenger assignment to one path
R	Set of scenarios
T	Length of period

Latin lower case letters

b	Coefficient vector of program (ID-LIN)
c	Index for cycles in the event activity network
i, j	Indices for events in the event activity network
ij	Index for activity from events i to event j in the event activity network
k	Index for OD pairs
l_{ij}	Lower bound on activity $ij \in \mathcal{A}$
m	Index for binary representation of δ_{ij}
$\underline{m}_k$	Length of shortest path for OD pair k w.r.t lower bounds
$\overline{m}_k$	Length of longest path for OD pair k w.r.t upper bounds
n_k	Number of alternative paths for OD pair k
o_k	Number of passengers of OD pair k
p, q	Indices for paths in the event activity network
r	Index for scenarios
t_p	Length of path p
t_{pr}	Length of path p in scenario r
t_{kr}	Length of shortest path for OD pair k in scenario r
tt_{mp}	Evaluation function: travel time assuming a logit distribution of passengers
tt_{sp}	Evaluation function: travel time assuming shortest paths for passengers
u_{ij}	Upper bound on activity $ij \in \mathcal{A}$
ut_{log}	Evaluation function: total logsums
ut_{sum}	Evaluation function: evaluated total utility
w_p	Probability that path p is chosen
x_{ij}	Passenger load on activity $ij \in \mathcal{A}$
x_{ij}^p	Passenger load of path p on activity $ij \in \mathcal{A}$
z_{pr}	Binary variable indicating whether path p is the shortest in scenario r

3.B Linear distribution function with characteristics of logit model

To prove Lemma 3.1, we consider the three cases

I $n_k = 1$:

It obviously follows by the property 'certain event' in Equation (3.8) that

$$w_p((t_p)) = \sum_{p \in P_k} w_p((t_p)) = 1.$$

This does not conflict with any other required characteristic.

II $n_k \neq 1$ and $\underline{m}_k = \overline{m}_k$:

If $\overline{m}_k = \underline{m}_k$, all paths have to have the same fixed length, i.e., $t_p = t_q \ \forall p, q \in P_k$. Then, it follows by the property 'uniform distribution on equivalent alternatives' in Equation (3.10) that

$$w_p((t_q)_{q \in P_k}) = w_p((t_p, \ldots, t_p)) = \frac{1}{n_k}.$$

This also does not conflict with any other required characteristic.

III $n_k \neq 1$ and $\underline{m}_k \neq \overline{m}_k$:

To show that all linear functions with the five desired characteristics are of the stated shape, we take a linear function

$$w_p((t_q)_{q \in P_k}) = \alpha_0^p + \sum_{q \in P_k} \alpha_q^p t_q \quad \forall p \in P_k$$

in its general form and restrict it by adding the desired characteristics to it.

Logit characteristic: absolute utility differences determine probability, Equation (3.12)

To obtain a linear distribution function with the characteristics of a logit distribution, we require that the probabilities do not depend on the values of the utilities but on their absolute differences only. Thus, we get for each path p

$$\begin{aligned} & w_p((t_q + \hat{t})_{q \in P_k}) = w_p((t_q)_{q \in P_k}) && \forall \hat{t} \in \mathbb{R} \\ \Leftrightarrow \quad & \alpha_0^p + \sum_{q \in P_k} \alpha_q^p (t_q + \hat{t}) = \alpha_0^p + \sum_{q \in P_k} \alpha_q^p t_q && \forall \hat{t} \in \mathbb{R} \\ \Leftrightarrow \quad & \sum_{q \in P_k} \alpha_q^p \hat{t} = 0 && \forall \hat{t} \in \mathbb{R} \\ \Leftrightarrow \quad & \sum_{q \in P_k} \alpha_q^p = 0 \end{aligned}$$

That means, to obtain a linear distribution function with the logit characteristic, all coefficients of the utilities t_p have to sum up to zero.

Uniform distribution on equivalent alternatives, Equation (3.10)

We add the requirement that the distribution function should be a uniform distribution in case all alternatives have the same utility. We therefore require that

$$
\begin{aligned}
& w_p(t,\ldots,t) = \frac{1}{n_k} \quad \forall p \in P_k \ \forall t \in \mathbb{R} \\
\Leftrightarrow \quad & \alpha_0^p + \sum_{q \in P_k} \alpha_q^p t = \frac{1}{n_k} \quad \forall p \in P_k \ \forall t \in \mathbb{R}
\end{aligned}
$$

Plugging in the result that the coefficients sum up to zero, this leaves the second condition

$$\alpha_0^p = \frac{1}{n_k} \quad \forall p \in P_k.$$

Since this equation has to hold for all paths $p \in P_k$, it follows that

$$\alpha_0^p = \alpha_0^q \quad \forall p, q \in P_k$$

and we define $\alpha_0 := \alpha_0^p$ for any path $p \in P_k$. This yields that all linear functions with the characteristics in Equations (3.10) and (3.12) are of the form

$$w_p((t_q)_{q \in P_k}) = \sum_{q \in P_k} \alpha_q^p t_q + \frac{1}{n_k}$$

with

$$\sum_{q \in P_k} \alpha_q^p = 0.$$

Independence of order, Equation (3.11)

Next, we consider the condition 'independence of order' of alternatives. Assume the probability of path $p \in P_k$ is to be determined. Then, its probability depends on the quality of all alternatives, but it should be independent of which alternative takes which of these values. We consider any permutation π_p that permutes two paths $q_1, q_2 \neq p \in P_k$ and keeps all other paths constant. Then,

$$
\begin{aligned}
& w_p((t_q)_{q \in P_k}) = w_p((t_{\pi_p(q)})_{q \in P_k}) \\
\Leftrightarrow \quad & \sum_{q \in P_k} \alpha_q^p t_q + \frac{1}{n_k} = \sum_{q \in P_k} \alpha_q^p t_{\pi_p(q)} + \frac{1}{n_k} \\
\Leftrightarrow \quad & \alpha_{q_1}^p t_{q_1} + \alpha_{q_2}^p t_{q_2} = \alpha_{q_1}^p t_{q_2} + \alpha_{q_2}^p t_{q_1} \\
\Leftrightarrow \quad & \alpha_{q_1}^p (t_{q_1} - t_{q_2}) = \alpha_{q_2}^p (t_{q_1} - t_{q_2}) \\
\Leftrightarrow \quad & \alpha_{q_1}^p = \alpha_{q_2}^p
\end{aligned}
$$

The last equivalence holds since we require this for all values of the perceived travel times t_{q_1} and t_{q_2} of paths q_1 and q_2. Since we require this to hold for all permutations π_p, in particular for those that permute two arbitrary paths, the coefficients α_q^p have to be equal for all $q \in P_k$, i.e.,

$$\alpha_{q_1}^p = \alpha_{q_2}^p \quad \forall q_1, q_2 \neq p \in P_k.$$

This condition is sufficient to obtain the independence of order with any permutation π_p. Together with the characteristic

$$\sum_{q \in P_k} \alpha_q^p = 0$$

from the property 'uniform distribution on equivalent alternatives' we get

$$\alpha_p^p = -\sum_{\substack{q \in P_k \\ q \neq p}} \alpha_q^p = -(n_k - 1)\alpha_q^p$$

$$\Leftrightarrow \alpha_q^p = \frac{-\alpha_p^p}{(n_k - 1)}.$$

Note that this is well defined since we discuss here only the cases where $n_k \neq 1$. This means, we can express all α_q^p by the single parameter α_p^p. Defining $\alpha_p \coloneqq \alpha_p^p$ we can plug this into the linear probability function and get

$$\begin{aligned} w_p((t_q)_{q \in P_k}) &= \alpha_p t_p - \sum_{\substack{q \in P_k \\ q \neq p}} \frac{\alpha_p t_q}{(n_k - 1)} + \frac{1}{n_k} \quad \forall p \in P_k \\ &= \alpha_p \left(t_p - \frac{1}{(n_k - 1)} \sum_{\substack{q \in P_k \\ q \neq p}} t_q \right) + \frac{1}{n_k} \quad \forall p \in P_k \end{aligned}$$

Monotonicity, Equation (3.9)
With the monotonicity of the distribution function follows

$$\begin{aligned} & w_p((t_q)_{q \in P_k} + \varepsilon \cdot e_p) < w_p((t_q)_{q \in P_k}) \\ \Leftrightarrow \quad & \alpha_p \left(t_p + \varepsilon - \frac{1}{(n_k - 1)} \sum_{\substack{q \in P_k \\ q \neq p}} t_q \right) + \frac{1}{n_k} < \alpha_p \left(t_p - \frac{1}{(n_k - 1)} \sum_{\substack{q \in P_k \\ q \neq p}} t_q \right) + \frac{1}{n_k} \\ \Leftrightarrow \quad & \alpha_p \varepsilon < 0 \end{aligned}$$

$$\Leftrightarrow \quad \alpha_p < 0$$

Note that for $\alpha_p = 0$ we obtain the uniform distribution

$$w_p((t_q)_{q \in P_k}) = \frac{1}{n_k} \quad \forall p \in P_k$$

which would not fulfill the required monotonicity.

Note that this only implies that $\alpha_p < 0$ has to be chosen in the general case III. In particular, case II, where all paths have the same fixed length, does not contradict the required monotonicity although a uniform distribution is applied. This is apparent when comparing two inputs $((t_q)_{q \in P_k}$ and $(t_q)_{q \in P_k} + \varepsilon \cdot e_p))$. For at least one of the inputs case III applies and monotonicity holds if $\alpha_p < 0$.

Distribution characteristics, Equation (3.8)
Finally, we add the distribution characteristics

$$w_p((t_q)_{q \in P_k}) \in [0, 1] \text{ and } \sum_{p \in P_k} w_p((t_q)_{q \in P_k}) = 1.$$

To fulfill the first characteristic, we consider the cases with the highest and lowest possible probability. The lowest probability for path p is achieved, if path p is as long as the upper bound $\overline{m}_k$ and all other paths $q \neq p \in P_k$ are as short as as the lower bound $\underline{m}_k$(due to monotonicity and the probability of the certain event). Then, we require

$$\begin{aligned}
& w_p((t_q)_{q \in P_k}) \geq 0 && \forall t_q \in [\underline{m}_k, \overline{m}_k],\ \forall q \in P_k \\
\Leftrightarrow \quad & \alpha_p \left(t_p - \frac{1}{(n_k - 1)} \sum_{\substack{q \in P_k \\ q \neq p}} t_q \right) + \frac{1}{n_k} \geq 0 && \forall t_q \in [\underline{m}_k, \overline{m}_k],\ \forall q \in P_k \\
\Leftrightarrow \quad & \alpha_p \left(\overline{m}_k - \frac{1}{(n_k - 1)} \sum_{\substack{q \in P_k \\ q \neq p}} \underline{m}_k \right) + \frac{1}{n_k} \geq 0 \\
\Leftrightarrow \quad & \alpha_p \left(\overline{m}_k - \underline{m}_k \right) + \frac{1}{n_k} \geq 0 \\
\Leftrightarrow \quad & \alpha_p \geq \frac{-1}{n_k \left(\overline{m}_k - \underline{m}_k \right)}
\end{aligned} \tag{3.19}$$

Similarly, the highest probability for path p is achieved, if path p is as short as possible and all other paths $q \neq p \in P_k$ are as long as possible, i.e., $t_p = \underline{m}_k$ and $t_q = \overline{m}_k \ \forall q \neq p \in P_k$. Then, we require

$$\begin{aligned} & w_p((t_q)_{q\in P_k}) \leq 1 && \forall t_q \in [\underline{m}_k, \overline{m}_k],\ \forall q \in P_k \\ \Leftrightarrow \quad & \alpha_p \left(\underline{m}_k - \frac{1}{(n_k-1)} \sum_{\substack{q\in P_k \\ q\neq p}} \overline{m}_k \right) + \frac{1}{n_k} \leq 1 \\ \Leftrightarrow \quad & \alpha_p \left(\underline{m}_k - \overline{m}_k \right) + \frac{1}{n_k} \leq 1 \\ \Leftrightarrow \quad & \alpha_p \geq \frac{-(n_k-1)}{n_k\left(\overline{m}_k - \underline{m}_k\right)} && (3.20) \end{aligned}$$

Note that both lower bounds (3.19) and (3.20) on α_p are negative and well defined as we consider the case where $\overline{m}_k \neq \underline{m}_k$. The second lower bound (3.20) is less strict, equality can only be achieved for $n_k = 2$. This also implies that we always have $w_p < 1$ by construction of the probability function, unless $n_k = 2$.

In total we obtain

$$\alpha_p \in \left[\frac{-1}{n_k(\overline{m}_k - \underline{m}_k)}, 0 \right)$$

In this range for α_p the function w_p can be tuned.

The second distribution characteristic requires that the probability of the certain event equals one. This is

$$\begin{aligned} & \sum_{p\in P_k} w_p((t_q)_{q\in P_k}) = 1 \\ \Leftrightarrow \quad & \sum_{p\in P_k} \left(\alpha_p \left(t_p - \frac{1}{(n_k-1)} \sum_{\substack{q\in P_k \\ q\neq p}} t_q \right) + \frac{1}{n_k} \right) = 1 \\ \Leftrightarrow \quad & \sum_{p\in P_k} \alpha_p t_p - \sum_{p\in P_k} \frac{\alpha_p}{(n_k-1)} \sum_{\substack{q\in P_k \\ q\neq p}} t_q + 1 = 1 \\ \Leftrightarrow \quad & \sum_{p\in P_k} t_p \left(\alpha_p - \frac{1}{(n_k-1)} \sum_{\substack{q\in P_k \\ q\neq p}} \alpha_q \right) = 0 \end{aligned}$$

Here, we factorized the path length t_p. Since the sum has to vanish for all $t_p \in [\underline{m}_k, \overline{m}_k]$, each of the addends has to vanish, which yields

$$\Leftrightarrow \quad \alpha_p - \frac{1}{(n_k - 1)} \sum_{\substack{q \in P_k \\ q \neq p}} \alpha_q = 0 \qquad \forall p \in P_k$$

$$\Leftrightarrow \quad \alpha_p = \alpha_q \quad \forall p, q \in P_k$$

That means, for each OD pair k, the parameter α_p are equal for all probability functions w_p with $p \in P_k$. We therefore define $\alpha_k := \alpha_p$ for any $p \in P_k$. This yields the linear distribution functions

$$w_p((t_q)_{q \in P_k}) = \alpha_k \left(t_p - \frac{1}{(n_k - 1)} \sum_{\substack{q \in P_k \\ q \neq p}} t_q \right) + \frac{1}{n_k} \quad \forall p \in P_k$$

As from now, we use a scaling factor $\alpha \in (0, 1]$ and write

$$\alpha_k = \frac{\alpha}{n_k(\underline{m}_k - \overline{m}_k)} \quad \forall k \in OD.$$

For $n_k > 1$ and $\overline{m}_k \neq \underline{m}_k$, all linear probability functions fulfilling the given criteria can be written as

$$w_p((t_q)_{q \in P_k}) = \frac{\alpha}{n_k(\underline{m}_k - \overline{m}_k)} \left(t_p - \frac{1}{n_k - 1} \sum_{\substack{q \in P_k: \\ q \neq p}} t_q \right) + \frac{1}{n_k}$$

with $\alpha \in (0, 1]$.

3.C Derivation of A and b

For the derivation of the coefficient matrix A and the coefficient vector b of the program (ID-LIN), we split the set of OD pairs into three disjoint sets:

$$OD = OD^* \mathbin{\dot\cup} OD_1 \mathbin{\dot\cup} OD_=, \tag{3.21}$$

where OD_1 denotes the set of OD pairs k with $n_k = 1$, $OD_=$ denotes the set of OD pairs k with $n_k > 1$ and $\overline{m}_k = \underline{m}_k$, and OD^* denotes the set of remaining OD pairs k with $n_k > 1$ and $\overline{m}_k \neq \underline{m}_k$. Then, we plug the Constraints (3.5), (3.6) and (3.7) into the objective function (3.4). In Constraint (3.6) we use the respective linear distribution for the OD sets OD_1, $OD_=$ and OD^* as given in Lemma 3.1.

$$
\begin{aligned}
&\sum_{(i,j)\in\mathcal{A}} x_{ij}\delta_{ij} \\
\overset{(3.7)}{=}\; &\sum_{(i,j)\in\mathcal{A}} \sum_{k\in OD} \sum_{\substack{p\in P_k:\\ ij\in p}} x_{ij}^p \delta_{ij} \\
\overset{(3.21)}{=}\; &\sum_{(i,j)\in\mathcal{A}} \sum_{k\in OD^*} \sum_{\substack{p\in P_k:\\ ij\in p}} x_{ij}^p \delta_{ij} + \sum_{(i,j)\in\mathcal{A}} \sum_{k\in OD_1} \sum_{\substack{p\in P_k:\\ ij\in p}} x_{ij}^p \delta_{ij} + \sum_{(i,j)\in\mathcal{A}} \sum_{k\in OD_=} \sum_{\substack{p\in P_k:\\ ij\in p}} x_{ij}^p \delta_{ij} \\
\overset{(3.6)}{=}\; &\sum_{(i,j)\in\mathcal{A}} \sum_{k\in OD^*} \sum_{\substack{p\in P_k:\\ ij\in p}} \left[\frac{\alpha}{n_k(\underline{m}_k - \overline{m}_k)} \left(t_p - \frac{1}{n_k - 1} \sum_{\substack{q\in P_k:\\ q\neq p}} t_q \right) + \frac{1}{n_k} \right] o_k \delta_{ij} \\
&\quad + \sum_{(i,j)\in\mathcal{A}} \sum_{k\in OD_1} \sum_{\substack{p\in P_k:\\ ij\in p}} o_k \delta_{ij} \\
&\quad + \sum_{(i,j)\in\mathcal{A}} \sum_{k\in OD_=} \sum_{\substack{p\in P_k:\\ ij\in p}} \frac{o_k}{n_k} \delta_{ij} \\
=\; &\sum_{(i,j)\in\mathcal{A}} \sum_{k\in OD^*} \sum_{\substack{p\in P_k:\\ ij\in p}} \frac{\alpha}{n_k(\underline{m}_k - \overline{m}_k)} \left(t_p - \frac{1}{n_k - 1} \sum_{\substack{q\in P_k:\\ q\neq p}} t_q \right) o_k \delta_{ij} \\
&\quad + \sum_{(i,j)\in\mathcal{A}} \sum_{k\in OD^*} \sum_{\substack{p\in P_k:\\ ij\in p}} \frac{1}{n_k} o_k \delta_{ij} \\
&\quad + \sum_{(i,j)\in\mathcal{A}} \sum_{k\in OD_1} \sum_{\substack{p\in P_k:\\ ij\in p}} \frac{o_k}{n_k} \delta_{ij} \\
&\quad + \sum_{(i,j)\in\mathcal{A}} \sum_{k\in OD_=} \sum_{\substack{p\in P_k:\\ ij\in p}} \frac{o_k}{n_k} \delta_{ij} \\
=\; &\underbrace{\sum_{(i,j)\in\mathcal{A}} \sum_{k\in OD^*} \sum_{\substack{p\in P_k:\\ ij\in p}} \frac{\alpha}{n_k(\underline{m}_k - \overline{m}_k)} \left(t_p - \frac{1}{n_k - 1} \sum_{\substack{q\in P_k:\\ q\neq p}} t_q \right) o_k \delta_{ij}}_{=:quadr(\delta_{ij})} \\
&\quad + \underbrace{\sum_{(i,j)\in\mathcal{A}} \sum_{k\in OD} \sum_{\substack{p\in P_k:\\ ij\in p}} \frac{1}{n_k} o_k \delta_{ij}}_{=:lin(\delta_{ij})}
\end{aligned}
$$

Here, we can see that the quadratic part $quadr(\delta_{ij})$ of the objective function only depends on the OD pairs $k \in OD^*$ that have multiple paths with possibly different lengths at choice. All OD pairs $k \in OD_1$ with just one optional path per period and all OD pairs $k \in OD_=$ with multiple paths of the same fixed lengths are uniformly distributed on their respective path(s) and their total travel time is added to the objective function. These addends

$$\sum_{(i,j)\in\mathcal{A}} \sum_{k\in OD_1} \sum_{\substack{p\in P_k:\\ ij\in p}} \frac{o_k}{n_k}\delta_{ij}$$

and

$$\sum_{(i,j)\in\mathcal{A}} \sum_{k\in OD_=} \sum_{\substack{p\in P_k:\\ ij\in p}} \frac{o_k}{n_k}\delta_{ij}$$

are hidden in the linear part $lin(\delta_{ij})$ of the objective function. To derive a closed form for the coefficients, we will consider the quadratic and the linear part of the objective function separately.

$$
\begin{aligned}
&quadr(\delta_{ij})\\
&= \sum_{ij\in\mathcal{A}} \sum_{k\in OD^*} \sum_{\substack{p\in P_k:\\ ij\in p}} \frac{\alpha}{n_k(\underline{m}_k - \overline{m}_k)} \left(t_p - \frac{1}{n_k - 1} \sum_{\substack{q\in P_k:\\ q\neq p}} t_q \right) o_k \delta_{ij}\\
&\overset{(3.5)}{=} \sum_{ij\in\mathcal{A}} \sum_{k\in OD^*} \frac{\alpha o_k}{n_k(\underline{m}_k - \overline{m}_k)} \sum_{\substack{p\in P_k:\\ ij\in p}} \left(\sum_{i'j'\in p} \delta_{i'j'} - \frac{1}{n_k - 1} \sum_{\substack{q\in P_k:\\ q\neq p}} \sum_{i'j'\in q} \delta_{i'j'} \right) \delta_{ij}\\
&= \sum_{ij\in\mathcal{A}} \sum_{k\in OD^*} \frac{\alpha o_k}{n_k(\underline{m}_k - \overline{m}_k)} \sum_{\substack{p\in P_k:\\ ij\in p}} \sum_{i'j'\in p} \delta_{i'j'}\delta_{ij}\\
&\quad + \sum_{ij\in\mathcal{A}} \sum_{k\in OD^*} \frac{\alpha o_k}{n_k(\underline{m}_k - \overline{m}_k)} \sum_{\substack{p\in P_k:\\ ij\in p}} \left(-\frac{1}{n_k - 1} \sum_{\substack{q\in P_k:\\ q\neq p}} \sum_{i'j'\in q} \delta_{i'j'}\delta_{ij} \right)\\
&= \sum_{ij\in\mathcal{A}} \sum_{k\in OD^*} \frac{\alpha o_k}{n_k(\underline{m}_k - \overline{m}_k)} \sum_{i'j'\in\mathcal{A}} \sum_{\substack{p\in P_k:\\ ij\in p,\\ i'j'\in p}} \delta_{i'j'}\delta_{ij}
\end{aligned}
$$

$$-\sum_{ij\in\mathcal{A}}\sum_{k\in OD^*}\frac{\alpha o_k}{n_k(\underline{m}_k-\overline{m}_k)}\frac{1}{n_k-1}\sum_{\substack{p\in P_k:\\ ij\in p}}\sum_{i'j'\in\mathcal{A}}\sum_{\substack{q\in P_k:\\ q\neq p,\\ i'j'\in q}}\delta_{i'j'}\delta_{ij}$$

$$=\sum_{ij\in\mathcal{A}}\sum_{i'j'\in\mathcal{A}}\sum_{k\in OD^*}\frac{\alpha o_k}{n_k(\underline{m}_k-\overline{m}_k)}\left(\sum_{\substack{p\in P_k:\\ ij,i'j'\in p}}1-\frac{1}{n_k-1}\sum_{\substack{p\in P_k:\\ ij\in p}}\sum_{\substack{q\neq p\in P_k:\\ i'j'\in q}}1\right)\delta_{i'j'}\delta_{ij}$$

This yields the proposed matrix A as given in Equation (3.16).

$$A_{ij,i'j'}=\sum_{k\in OD^*}\frac{\alpha o_k}{n_k(\underline{m}_k-\overline{m}_k)}\left(\sum_{\substack{p\in P_k:\\ ij,i'j'\in p}}1-\sum_{\substack{p\in P_k:\\ ij\in p}}\sum_{\substack{q\neq p\in P_k:\\ i'j'\in q}}\frac{1}{n_k-1}\right)$$

Next, the linear term will be computed.

$$lin(\delta_{ij})=\sum_{ij\in\mathcal{A}}\sum_{k\in OD}\sum_{\substack{p\in P_k:\\ ij\in p}}\frac{1}{n_k}o_k\delta_{ij}$$

This yields directly the proposed vector b.

$$b_{ij}=\sum_{k\in OD}\sum_{p\in P_k:ij\in p}\frac{o_k}{n_k}$$

3.D Proof of negative semi-definiteness of coefficient matrix A

To prove that the coefficient matrix A is negative definite, it is sufficient to prove that $\delta^\dagger A\delta\le 0$ holds for all $\delta\in\mathbb{R}^{|\mathcal{A}|}$. Here, $\delta^\dagger$ denotes the transpose of δ. For this purpose, we rewrite the term $\delta^\dagger A\delta$ using the definition of the path length t_p.

$$\delta^\dagger A\delta$$

$$=\sum_{ij\in\mathcal{A}}\sum_{i'j'\in\mathcal{A}}\sum_{k\in OD}\frac{\alpha o_k}{n_k(\underline{m}_k-\overline{m}_k)}\left(\sum_{\substack{p\in P_k:\\ ij,i'j'\in p}}1-\frac{1}{n_k-1}\sum_{\substack{p\in P_k:\\ ij\in p}}\sum_{\substack{q\in P_k:\\ q\neq p,\\ i'j'\in q}}1\right)\delta_{i'j'}\delta_{ij}$$

$$= \sum_{k\in OD} \frac{\alpha o_k}{n_k(\underline{m}_k - \overline{m}_k)} \left(\sum_{p\in P_k} \sum_{\substack{ij\in\mathcal{A}\\ ij\in p}} \sum_{\substack{i'j'\in\mathcal{A}\\ i'j'\in p}} \delta_{i'j'}\delta_{ij} - \frac{1}{n_k - 1} \sum_{p\in P_k} \sum_{\substack{q\in P_k:\\ q\neq p}} \sum_{\substack{ij\in\mathcal{A}:\\ ij\in p}} \sum_{\substack{i'j'\in\mathcal{A}\\ i'j'\in q}} \delta_{i'j'}\delta_{ij} \right)$$

$$= \sum_{k\in OD} \frac{\alpha o_k}{n_k(\underline{m}_k - \overline{m}_k)} \left(\sum_{p\in P_k} \underbrace{\sum_{ij\in p} \delta_{ij}}_{t_p} \underbrace{\sum_{i'j'\in p} \delta_{i'j'}}_{t_p} - \frac{1}{n_k - 1} \sum_{p\in P_k} \sum_{\substack{q\in P_k:\\ q\neq p}} \underbrace{\sum_{ij\in p} \delta_{ij}}_{t_p} \underbrace{\sum_{i'j'\in q} \delta_{i'j'}}_{t_q} \right)$$

$$= \sum_{k\in OD} \frac{\alpha o_k}{n_k(\underline{m}_k - \overline{m}_k)} \left(\sum_{p\in P_k} t_p^2 - \frac{1}{n_k - 1} \sum_{p\in P_k} \sum_{\substack{q\in P_k:\\ q\neq p}} t_p t_q \right)$$

$$= \sum_{k\in OD} \frac{\alpha o_k}{2n_k(n_k - 1)(\underline{m}_k - \overline{m}_k)} \left(2 \sum_{p\in P_k} (n_k - 1)t_p^2 - 2 \sum_{p\in P_k} \sum_{\substack{q\in P_k:\\ q\neq p}} t_p t_q \right)$$

$$= \sum_{k\in OD} \frac{\alpha o_k}{2n_k(n_k - 1)(\underline{m}_k - \overline{m}_k)} \left(\sum_{p\in P_k} (n_k - 1)t_p^2 - \sum_{p\in P_k} \sum_{\substack{q\in P_k:\\ q\neq p}} 2t_p t_q + \sum_{p\in P_k} (n_k - 1)t_p^2 \right)$$

$$\stackrel{(*)}{=} \sum_{k\in OD} \frac{\alpha o_k}{2n_k(n_k - 1)(\underline{m}_k - \overline{m}_k)} \left(\sum_{p\in P_k} (n_k - 1)t_p^2 - \sum_{p\in P_k} \sum_{\substack{q\in P_k:\\ q\neq p}} 2t_p t_q + \sum_{p\in P_k} \sum_{\substack{q\in P_k:\\ q\neq p}} t_q^2 \right)$$

$$= \sum_{k\in OD} \frac{\alpha o_k}{2n_k(n_k - 1)(\underline{m}_k - \overline{m}_k)} \sum_{p\in P_k} \left((n_k - 1)t_p^2 - \sum_{\substack{q\in P_k:\\ q\neq p}} 2t_p t_q + \sum_{\substack{q\in P_k:\\ q\neq p}} t_q^2 \right)$$

$$= \sum_{k\in OD} \frac{\alpha o_k}{2n_k(n_k - 1)(\underline{m}_k - \overline{m}_k)} \sum_{p\in P_k} \sum_{\substack{q\in P_k:\\ q\neq p}} \left(t_p^2 - 2t_p t_q + t_q^2 \right)$$

$$= \sum_{k\in OD} \frac{\alpha o_k}{2n_k(n_k - 1)(\underline{m}_k - \overline{m}_k)} \sum_{p\in P_k} \sum_{\substack{q\in P_k:\\ q\neq p}} (t_p - t_q)^2,$$

where we used in the starred equation that

$$\begin{aligned} &\sum_{p\in P_k} (n_k - 1)t_p^2 \\ =&n_k \sum_{p\in P_k} t_p^2 - \sum_{p\in P_k} t_p^2 \\ =& \sum_{p\in P_k} \sum_{q\in P_k} t_q^2 - \sum_{p\in P_k} t_p^2 \end{aligned}$$

$$= \sum_{p \in P_k} \sum_{\substack{q \in P_k: \\ q \neq p}} t_q^2.$$

Since $(t_p - t_q)^2$, o_k and α are positive, n_k is at least 2, and $(\underline{m}_k - \overline{m}_k)$ is negative, it follows that

$$\delta^\dagger A \delta \leq 0.$$

In our application, δ_{ij} represents the duration of each activity $ij \in \mathcal{A}$ in the event activity network, so we require them to be positive. However, in general no such restriction is required and the implication $\delta^\dagger A \delta \leq 0$ holds for all $\delta \in \mathbb{R}^{\mathcal{A}}$. This implies that the matrix A is negative semi-definite for all practical relevant instances and the objective function is concave.

3.E Linearized formulation of Model (ID-LIN)

Using the binary representation

$$\delta_{ij} = l_{ij} + \sum_{m \in M_{ij}} 2^m \sigma_{ij}^m$$

with $M_{ij} := \{0, \ldots, \lfloor \log_2(u_{ij} - l_{ij}) \rfloor\}$ and replacing the product of the binary variables $\sigma_{ij}^m \cdot \sigma_{i'j'}^{m'}$ by $\gamma_{ij,i'j'}^{m,m'}$, we obtain a linearized version of the model (ID-LIN)

$$\begin{aligned}
\min \quad & \sum_{ij \in \mathcal{A}} \sum_{i'j' \in \mathcal{A}} A_{ij,i'j'} \Bigg(\sum_{m' \in M_{i'j'}} 2^{m'} l_{ij} \sigma_{i'j'}^{m'} + \sum_{m \in M_{ij}} 2^m l_{i'j'} \sigma_{ij}^m \\
& \qquad + \sum_{m \in M_{ij}} \sum_{m' \in M_{i'j'}} 2^{m+m'} \gamma_{ij,i'j'}^{m,m'} \Bigg) + \sum_{ij \in \mathcal{A}} b_{ij} \delta_{ij}
\end{aligned}$$

$$\begin{array}{llcll}
\text{s.t.} & \delta_{ij} & \geq & l_{ij} & \forall ij \in \mathcal{A} \\
& \delta_{ij} & \leq & u_{ij} & \forall ij \in \mathcal{A} \\
& \Gamma_c \delta & = & \mu_c \cdot T & \forall c \in \mathcal{C} \\
& \gamma_{ij,i'j'}^{m,m'} & \leq & \sigma_{ij}^m & \forall ij, i'j' \in \mathcal{A}: A_{ij,i'j'} < 0,\ \forall m \in M_{ij},\ \forall m' \in M_{i'j'} \\
& \gamma_{ij,i'j'}^{m,m'} & \leq & \sigma_{i'j'}^{m'} & \forall ij, i'j' \in \mathcal{A}: A_{ij,i'j'} < 0,\ \forall m \in M_{ij},\ \forall m' \in M_{i'j'} \\
& \gamma_{ij,i'j'}^{m,m'} & \geq & \sigma_{ij}^m + \sigma_{i'j'}^{m'} - 1 & \forall ij, i'j' \in \mathcal{A}: A_{ij,i'j'} > 0,\ \forall m \in M_{ij},\ \forall m' \in M_{i'j'} \\
& \sigma_{ij}^m & \in & \{0,1\} & \forall ij \in \mathcal{A},\ \forall m \in M_{ij} \\
& \gamma_{ij,i'j'}^{m,m'} & \in & \{0,1\} & \forall ij, i'j' \in \mathcal{A},\ \forall m \in M_{ij},\ \forall m' \in M_{i'j'} \\
& \mu_c & \in & \mathbb{Z}_+ & \forall c \in \mathcal{C}
\end{array}$$

For the sake of simplicity we omitted the constant term $\sum_{ij\in\mathcal{A}}\sum_{i'j'\in\mathcal{A}} A_{ij,i'j'} l_{ij} l_{i'j'}$ in the objective function.

To prove that $\gamma_{ij,i'j'}^{m,m'} = \sigma_{ij}^{m} \cdot \sigma_{i'j'}^{m'}$ in an optimal solution to the ILP, two cases are considered

- If $A_{ij,i'j'} < 0$, then $\gamma_{ij,i'j'}^{m,m'} \leq \min\{\sigma_{ij}^{m}, \sigma_{i'j'}^{m'}\}$. That means, as soon as $\sigma_{ij}^{m} = 0$ or $\sigma_{i'j'}^{m'} = 0$ it follows that $\gamma_{ij,i'j'}^{m,m'} = 0$. Otherwise $\gamma_{ij,i'j'}^{m,m'} = 1$ can be chosen, which is preferable since $A_{ij,i'j'} < 0$.
- If $A_{ij,i'j'} > 0$, then $\gamma_{ij,i'j'}^{m,m'} \geq \max\{0, \sigma_{ij}^{m} + \sigma_{i'j'}^{m'} - 1\}$. That means, as soon as $\sigma_{ij}^{m} = 1$ and $\sigma_{i'j'}^{m'} = 1$ it follows that $\gamma_{ij,i'j'}^{m,m'} = 1$. Otherwise $\gamma_{ij,i'j'}^{m,m'} = 0$ can be chosen, which is preferable since $A_{ij,i'j'} > 0$.

It follows that $\gamma_{ij,i'j'}^{m,m'} = \sigma_{ij}^{m} \cdot \sigma_{i'j'}^{m'}$ for $A_{ij,i'j'} \neq 0$. Note that for this linearization only constraints for non-zero coefficients $A_{ij,i'j'}$ have to be installed. By definition of $A_{ij,i'j'}$, this can only happen if two activities ij and $i'j'$ are contained in paths for the same OD-pair.

To decrease the search space in the computations we additionally add the constraints

$$
\begin{array}{rcll}
\gamma_{ij,i'j'}^{m,m'} & = & \gamma_{i'j',ij}^{m',m} & \forall ij < i'j' \in \mathcal{A}\colon A_{ij,i'j'} \neq 0,\ \forall m \in M_{ij},\ \forall m' \in M_{i'j'} \\
\gamma_{ij,ij}^{m,m'} & = & \gamma_{ij,ij}^{m',m} & \forall ij \in \mathcal{A}\colon A_{ij,ij} \neq 0,\ \forall m < m' \in M_{ij} \\
\gamma_{ij,ij}^{m,m} & = & \sigma_{ij}^{m} & \forall ij \in \mathcal{A}\colon A_{ij,ij} \neq 0,\ \forall m \in M_{ij}
\end{array}
$$

3.F Path choice sets

In this section, we describe how we preprocess the instances to derive a path choice set based on the line plan and the line frequency.

The quality of the path choice sets is of major importance for the quality of the results. There are two possibilities when designing a choice set. One should either take all alternatives into account and let 'the choice model decide that the choice probabilities of unrealistic options are low or zero' or take 'into account only subsets of the options which are effectively chosen in the sample' (or a heuristic approximation of that) (de Dios Ortúzar and Willumsen, 2011). Since a large path choice set implies a large number of variables and constraints in both models, the first option of including all alternatives is impractical in this setting. However, if the choice sets are too small, they might not contain all routes that are important for passengers, and the timetable will be constructed neglecting some relevant connections.

The literature provides multiple path generation algorithms to derive realistic path choice sets. Although most of them focus on road networks, the concept is often applicable to public transport networks as well. An overview of different algorithms is given in Bekhor et al. (2006). The number of alternative paths in such choice sets turned out to be very large, complicating the solution process of the models. Since these choice sets often contain many routes that are identical to some extent, we use a simple heuristic to find small choice sets of independent paths that are of comparable quality, as described in Sels et al. (2015). For each OD pair, we iteratively add the shortest path in the event activity network to the path choice set and delete all visited vertices. This procedure is repeated until origin and destination are not connected anymore, or the found path is too long or too inconvenient to be a possible alternative. Furthermore, we set the lengths of all transfer activities to the upper bounds and apply the same procedure to the event activity network again to find a choice set of paths with a minimum number of transfers. We take the union of these two sets and remove dominated paths.

This heuristic has shown to provide a small choice set for each OD pair with a representative selection of paths. The size of the choice sets is small enough for usage in an optimization framework. At the same time, all paths are good alternatives for the passengers and they are independent of each other. Since the paths are generated in the event activity network, there can be several independent paths for an OD pair if multiple lines are available. These paths might follow the same geographical route, which means, they use the same stations and tracks, but they are independent of each other in the sense that they have no event nor activity in common.

3.G Reference methods

3.G.1 Integer programming formulation for model (IS)

We implement the following mixed-integer linear program for timetabling with integrated shortest-route search for passengers.

$$
\begin{array}{llcll}
\min & \multicolumn{3}{l}{\sum_{k \in OD} o_k t_k} & \\
\text{s.t.} & \delta_{ij} & \geq & l_{ij} & \forall ij \in \mathcal{A} \\
& \delta_{ij} & \leq & u_{ij} & \forall ij \in \mathcal{A} \\
& \Gamma_c \delta & = & T \cdot \mu_c & \forall c \in \mathcal{C} \\
& t_p & = & \sum_{ij \in p} \delta_{ij} & \forall k \in OD,\ \forall p \in P_k \\
& t_k & = & \min_{p \in P_k} t_p & \forall k \in OD,\ \forall p \in P_k, \\
& \delta_{ij} & \in & \mathbb{Z}_+ & \forall ij \in \mathcal{A} \\
& \mu_c & \in & \mathbb{Z} & \forall c \in \mathcal{C} \\
& t_p & \in & \mathbb{R}_+ & \forall k \in OD,\ \forall p \in P_k \\
& t_k & \in & \mathbb{R}_+ & \forall k \in OD
\end{array}
$$

The minimum is linearized with the following set of constraints

$$
\begin{array}{rcll}
t_{kr} & \leq & t_p & \forall k \in OD,\ \forall p \in P_k \\
t_{kr} & \geq & t_p - (1 - z_p) M_k & \forall k \in OD,\ \forall p \in P_k \\
\sum_{p \in P_k} z_p & = & 1 & \forall k \in OD \\
z_p & \in & \{0, 1\} & \forall k \in OD,\ \forall p \in P_k
\end{array}
$$

where

$$M_k = \max_{p \in P_k} \sum_{ij \in p} u_{ij} - \min_{p \in P_k} \sum_{ij \in p} l_{ij}.$$

3.G.2 Pseudocode for method (ID-ITR)

We give the pseudocode of the algorithm used for the iterative method (ID-ITR).

Algorithm 3.1: Pseudocode for method (ID-ITR)

Input: Instance $\mathcal{I} = (\mathcal{N}, l, u, OD)$, period T

Output: Durations δ, cycle numbers z and passenger distribution x

Initialize: Set edge lengths to average bounds $\delta_{ij} \coloneqq \frac{1}{2}(l_{ij} + u_{ij})$;

while *Termination criterion not reached* **do**

  Compute passenger weight on edges with Equations (3.5), (3.6) and (3.7) and the logit model in Equation (3.1) as distribution function;

  Find timetable with given weights with model from Equations (3.2), (3.3) and (3.4);

Return (δ, z, x);

The termination criterion is reached if one of the following two conditions is met:

1. The improvement in objective value between two iterations is marginal, i.e., if

$$\frac{v_i - v_{i-1}}{v_i} < 0.01,$$

   where v_i denotes the objective value of the i-th iteration.

2. The number of iterations exceeds 4.

Chapter 4

Modeling and solving line planning with integrated mode choice

4.1 Introduction

In previous years, public transport operators around the globe recorded a continuous increase in passenger numbers. There is a variety of reasons that might explain changes in passenger numbers, such as a shift in passenger interests and behavior, the development of regions, or policy measures. Since the onset of the Covid-19 pandemic in early 2020, however, the number of passengers in many countries has dropped dramatically to a fraction of its original size. The pandemic is likely to have long-term effects on travel behavior, as, for example, working from home is more accepted by many companies. In the case of the Netherlands Railways (NS), the expectation is that passenger numbers will not reach their pre-pandemic level before 2024[1], but increase slowly over time.

Before and after the pandemic, railway operators like NS were and will be constantly faced with changing travel demand. To adapt their service to small fluctuations, operators can make adjustments on the level of tactical planning. For example, small adaptions in the timetables and rolling stock schedules can be implemented relatively spontaneously and with comparatively little effort. However, such adjustments are not suitable to cope with greater and longer-term changes in demand. Instead, this issue needs to be approached from a strategic planning perspective and the line plan needs to be adjusted from time to time.

When designing a public transport line plan, it is important to distinguish between passenger and traveler demand. With *traveler demand* we refer to the total number of people who want to travel. Travelers may choose to use any available mode of transport, such as train or car. *Passenger demand* includes only those travelers who choose to use the public transport service. The decision of travelers for their mode of transport, and thus the number of passengers, depends to a certain extent on the quality of the service offered. This poses an interesting yet complex situation for operators: public transport services have to be designed to provide sufficient capacity for passenger demand, which in turn depends on the service.

In this chapter, we consider the problem of finding a line plan and simultaneously estimating the corresponding passenger demand based on a prognosis for traveler demand. The aim is to estimate both the share of travelers deciding to use public transport (mode choice) and the passenger distribution in the network (route choice). The travelers' choices depend on the quality of the service offered: they value a ser-

[1] `https://nos.nl/artikel/2344006-ns-lijdt-185-miljoen-verlies-door-corona.html`, visited on February 19, 2021.

vice with fast, direct, and frequent connections. A line plan offering such connections between two stations will attract more passengers between those stations, while passengers between stations with slow and infrequent connections will be inclined to turn to other modes of transport. Considering travelers' decisions allows an accurate estimation of passenger demand during optimization, and the resulting line plans are aligned with the demand they generate.

Although many approaches state that passenger demand and line plans are interdependent, we identify two reasons why demand estimation is mostly not modeled accurately. First, travelers' mode choice is in most cases neglected. Second, if a passenger distribution on routes is considered, usually one of the following two simplifications is applied: either all passengers traveling between two stations are required to use the same route, or the model can assign passengers to routes in favor of a system optimum, rather than considering passenger preferences. An imprecise demand estimate is obstructive to the search for efficient line plans and carries the risk of insufficient seating capacity. Only a few publications deal with the integration of passenger choice models in line planning. However, these approaches are usually not computationally tractable and the quality of solutions found with heuristic approaches is hard to assess. In Section 4.2 we discuss the related line planning literature in detail.

In this chapter, we present a mixed-integer linear programming (MILP) formulation for finding a line plan and corresponding passenger loads from given traveler demand. The passenger loads depend on the line plan and are estimated with passenger choice models. By suitable preprocessing of the utilities for the passengers' mode and route decisions, the choice models can be linearized and commercial solvers can be used to find solutions. Approaching the problem from the operator's perspective, our objective is profit maximization, where profit is equal to the revenue from serving passenger demand minus the operating costs. This yields line plans that attract many passengers, while being efficient with respect to operational costs.

We test and analyze the model in experiments on the Intercity network of the Randstad region that is operated by NS, the largest Dutch railway operator. Additional constraints and branching strategies are tested to improve the computational performance. The model is compared to a basic line planning model with predetermined passenger loads which highlights the advantages of demand estimation during optimization. Furthermore, the integration of passenger decision models enables to conduct a sensitivity analysis of the service level and operator profit on fluctuations in traveler demand. This gives valuable insights for operators into their business models and the optimal solutions provide concepts for how the line plans should be adjusted in response to demand changes.

Summarizing, our main contributions are fourfold. First, we present a novel line planning model that considers both route and mode choice from a customer's perspective. In contrast to existing optimization approaches, passengers are not assigned to routes according to a system optimum but distribute on the best routes in our model. Second, we develop a linear formulation for this model which allows the usage of commercial solvers and we provide means to improve the computational performance. Third, we show in experiments that operators should include an estimation of mode and route choice *during* optimization to achieve the best possible profit and passenger shares. Fourth, we show the impact of drastic changes in travel demand on the modal split and the financial performance of the public transport operator.

The remainder of this chapter is structured as follows. In Section 4.2, related line planning approaches are summarized. The modeling of line planning and passenger demand estimation is described in Section 4.3. This section discusses the used choice models in detail and the assumptions made in order to linearize them. Section 4.4 gives information about the experimental setup, used data sets, and parameters. The experiments are described and discussed in Section 4.5, followed by practical insights for operators in Section 4.6. The chapter concludes in Section 4.7 with a summary of the findings.

4.2 Related literature

The goal of the line planning problem is to find a set of lines with corresponding frequencies such that conditions on operating costs and passenger service level are satisfied. In this context, lines are defined as a sequence of stations that are served by a vehicle. Schöbel (2012) summarizes different modeling approaches and solution methods for the line planning problem in public transport and identifies several variations: In some formulations, the task is to select lines from a given pool of lines (Gattermann et al., 2017), while in others the line routes are constructed during optimization (Borndörfer et al., 2007). There also exist different objectives for line planning. On the one hand, cost-oriented objectives aim at minimizing operational costs while ensuring a certain passenger service level (Claessens et al., 1998; Friedrich et al., 2017b; Goossens et al., 2006). On the other hand, passenger-oriented objectives mostly consider a budget constraint on operational costs to maximize passenger service level, represented by the share of passengers with direct connections (Bussieck, 1998; Bussieck et al., 1997; Schöbel and Scholl, 2006) or by passenger journey times (Goerigk and Schmidt, 2017; Schöbel and Scholl, 2006).

The impact of the service on passengers is often neglected. Most of the existing approaches have the assumption in common that (an estimate of) the passenger demand is known before the line plan is found. This means the number of passengers between each station pair is assumed to be fixed. In addition, passengers are in many cases assigned a priori to paths in the network to estimate the required capacity between stations. However, both the number of passengers and the passenger paths depend on the line plan and the corresponding passenger service level (de Dios Ortúzar and Willumsen, 2011).

Most existing approaches that have included passenger route choice either applied a single (shortest) route search (Guan et al., 2006; D. Liu et al., 2019; Nachtigall and Jerosch, 2008) or a distribution according to a system optimum (Borndörfer et al., 2007; Borndörfer and Karbstein, 2012). Both strategies are unlikely to accurately estimate a passenger distribution, bearing the risk for operators of crowded or underutilized vehicles. In a cross-entropy heuristic for integrated line planning and timetabling presented by Kaspi and Raviv (2013), passengers are distributed on shortest paths for evaluation, which serves as the basis to refine the search for an updated solution in the next iteration. Schmidt and Schöbel (2015a) and Friedrich et al. (2017b) present generic line planning models with integrated passenger route choice and discuss complexity and bounds. A passenger-optimal route search was introduced in Schmidt (2014) and Goerigk and Schmidt (2017) where sufficient seating capacity is ensured assuming that passengers distribute over the available shortest routes. This approach overcomes the problem that passengers are assigned to suboptimal routes by the model and prevents capacity conflicts during operation. A. Schiewe et al. (2019) propose a game-theoretical approach where passengers are individual players choosing their routes with the highest travel quality.

All of the approaches discussed above consider a flexible passenger to route assignment or search but assume the total passenger demand to be fixed. Only a few publications consider the mode choice of travelers during line planning to estimate the number of passengers attracted by the solution. The integrated stop location and line planning approaches discussed in Laporte et al. (2005) and Laporte et al. (2007) aim at a maximum trip coverage. Similarly, Klier and Haase (2015) maximize the number of expected passengers and estimate the mode choice with the logit model as a traveler's decision between the best available route and an alternative mode. Bertsimas et al. (2021) have the same objective to maximize ridership of the public transport service. In their model, each additional line increases the modal split by a predefined percentage, independent of which other lines are selected and their frequencies. De-Los-Santos et al. (2017) use the logit model for the mode choice as well and approximate it with a piecewise linear function. For the specific case of Intercity buses, Steiner and Irnich (2018) consider different passenger demand levels depending on departure and travel time in a combined optimization approach for stop selection on a line and timetabling. A comprehensive model integrating network design, line planning, traveler mode choice, passenger route choice, and fleet investment is discussed in Canca et al. (2016). Later, the authors provide in Canca et al. (2017) an adaptive large neighborhood search metaheuristic for this model, limiting the passenger route search to a shortest route. For a revised model including an integrated passenger distribution on routes, Canca et al. (2019) present a two-level local search matheuristic which was successfully used to find solutions for real-world sized instances.

These models include a modeling of travelers' mode choice during line planning, but they contain at least one of the two limitations. Either, the passenger distribution on routes, although relevant for seat capacity estimation, does not reflect passenger preferences. Or, the quality of the solutions found by the applied metaheuristics is in many cases hard to assess.

4.3 Modeling

In this chapter, we develop a mixed-integer linear programming model for finding a line plan that is tailored to the corresponding passenger demand. Lines are selected from a pool of potential lines. The passenger demand is estimated with a discrete choice model and the distribution on routes is modeled according to passengers' choices. This modeling of passengers provides a basis for an accurate estimation of the required seat capacity and the expected revenues from ticket sales. The objective is to find a profit-optimal line plan, which means, the difference between revenue and cost is maximized. In the following sections, we discuss all components of the line planning problem, the underlying assumptions made, and how we model it. Throughout this chapter, we use the terminology of public rail transport, however, the proposed model applies to public transport services of any kind.

4.3.1 Network structure and line selection

We consider a public transport network with a set of *stations* $\mathcal{S}$ as nodes, a set of *tracks* $\mathcal{T}$ as direct connections between the stations, and symmetric traveler demand between the stations. Let k denote an *origin-destination (OD) pair*, that is, an unordered pair of stations s_1 and $s_2 \in \mathcal{S}$, and let $\mathcal{OD}$ be the *set of OD pairs.* The *traveler demand*, that is, the number of persons wanting to travel between OD pair $k \in \mathcal{OD}$, is denoted by δ_k. We assume a *line pool* $\mathcal{P}$ to be given, where each *line* $l \in \mathcal{P}$ is an undirected sequence of stations $(s_1, s_2, \ldots, s_{n_l})$.

The aim is to find a *line plan* $\mathcal{L} \subseteq \mathcal{P}$, a subset of lines from the line pool, for a regular, symmetric service. When selecting a line from the pool $\mathcal{P}$, it is assumed to operate in both directions.

The selection of lines is modeled using binary decision variables

$$z_l \in \{0, 1\} \text{ with } z_l = \begin{cases} 0, & \text{Line } l \text{ is not selected,} \\ 1, & \text{Line } l \text{ is selected.} \end{cases}$$

This constitutes the basis of the line planning problem: Lines have to be selected from a line pool to meet passenger demand. Higher line frequencies can be achieved by selecting multiple lines with the same itinerary. This indirect modeling of frequencies is accepted to unambiguously link lines and passenger routes in Section 4.3.3. In the following, the objective of the line planning model is defined, the procedure to estimate the passenger demand is explained, and, based on that, the required capacity is calculated.

4.3.2 Profit maximization

Our objective is to find a line plan that is optimal with respect to profit, that means, the difference of revenue generated by passenger fares and costs for installing and operating lines should be as high as possible. The *costs* c_l for installation and operation of line l comprise acquisition and maintenance of rolling stock as well as personnel and energy expenses. We assume an OD pair-dependent passenger pricing as it is applied in the Netherlands. There, passengers are charged based on the locations of their origin and destination only, and ticket prices are independent of the chosen route. Let p_k denote the *ticket price* for a passenger of OD pair k and let w_k denote the *share of travelers* of OD pair k that choose to use public transport. Then, the following function describes the profit made by a line plan $\mathcal{L}$.

$$\sum_{k \in \mathcal{OD}} p_k \cdot \delta_k \cdot w_k - \sum_{l \in \mathcal{P}} c_l \cdot z_l \tag{4.1}$$

The total cost calculation is defined as the sum of costs for all selected lines. The product of the total number of travelers δ_k, the share of travelers using public transport w_k and the ticket price p_k gives the total revenue from passengers of OD pair k. Note that next to the cost, also the revenue is affected by the selected lines. With a better service, more passengers decide to use public transport and generate more revenue. Therefore, the share of travelers deciding to use public transport w_k is a variable in this context. How the value of $w_k \in [0, 1]$ is estimated based on the selected lines is described in the next section.

4.3.3 Demand estimation

Only travelers that decide to use public transport because of the quality of the line plan generate revenue for the operator. Travelers that choose to travel with an alternative mode, such as private car, do not contribute to the objective function (4.1). Mode choices of travelers can be estimated with discrete choice models (de Dios Ortúzar and Willumsen, 2011). We assume a traveler demand instead of a passenger demand to be given and include a mode choice based on the quality of the line plan.

Furthermore, passengers that choose to travel with public transport distribute over available routes. A *route* is a sequence of consecutive line segments, where each line segment corresponds to a track in the network operated by a certain line. This concept of a route corresponds to a path in the *change&go network* introduced by Schöbel and Scholl (2006). Estimating which routes passengers use is important to determine the required seating capacity correctly.

Route choice

Passengers distribute over multiple available routes if they are of reasonably good quality. A route is *available* if all lines that are used on the route are selected. Since the availability of routes is only known after the line plan is known, we consider a *choice set* $\mathcal{C}_k$ of routes for each OD pair k to be given. The availability of a route $r \in \mathcal{C}_k$

is modeled with the binary auxiliary decision variable

$$y_r \in \{0,1\} \text{ with } y_r = \begin{cases} 0, & \text{Route } r \text{ is not available,} \\ 1, & \text{Route } r \text{ is available.} \end{cases}$$

The selection of lines and availability of routes can be linked with the following set of constraints

$$y_r \le z_l \qquad \forall k \in \mathcal{OD},\ r \in \mathcal{C}_k,\ l \in \mathcal{P}_r \tag{4.2a}$$

$$y_r \ge \sum_{l \in \mathcal{P}_r} z_l - |\mathcal{P}_r| + 1 \qquad \forall k \in \mathcal{OD},\ r \in \mathcal{C}_k \tag{4.2b}$$

where $\mathcal{P}_r$ is the set of lines in the pool used on route r.

In practice, we observe that passengers predominantly distribute over the best available routes. Hence, we restrict the choice sets $\mathcal{C}_k$ to contain only the best routes. The quality of a route is determined by the driving time as well as the number of transfers. Let j_r be the *journey time* of route r, that is, the approximate driving time including dwell times plus a transfer penalty for each transfer. Then, we consider only those passenger routes for OD pair k that are among the journey-time shortest routes. By restricting the choice sets to the shortest routes, all alternative routes for one OD pair are of very similar quality to the passengers. As a consequence, we assume that passengers of one OD pair distribute uniformly on available routes.

Note that this assumption of a uniform distribution on routes of similar journey time does not perfectly represent passenger behavior. The distribution of passengers on routes also depends on the temporal spread of departure and arrival times of the available routes. Since the timetable is not known yet at the stage of line planning, it is only possible to estimate the passenger distribution based on the quality of the available routes. Based on the information available, the uniform distribution on the journey-time shortest routes is a good approximation. In any case, we believe that this modeling comes closer to an actual distribution than the common assumptions that all passengers of an OD pair use a single (shortest) route or can be assigned to routes according to a system optimum. Hence, this distribution is expected to give a better estimate of required seating capacity than existing approaches.

To assume that passengers distribute uniformly on routes is an essential component to achieve a tractable model. It is therefore important to consider comparable routes for each OD pair. In this chapter, we use the journey time to compare routes, but in principle, any definition of quality that is independent of the timetable can be used.

Let the variable w_r denote the *share of travelers using route* r. Then,

$$w_r \in [0,1] \text{ with } w_r = \begin{cases} 0, & y_r = 0, \\ \frac{w_k}{\sum_{r' \in \mathcal{C}_k} y_{r'}}, & y_r = 1 \end{cases} \qquad \forall k \in \mathcal{OD},\ r \in \mathcal{C}_k \tag{4.3}$$

The values of w_r are the same for all available routes for OD pair k and they sum up to the share of travelers using public transport w_k. To linearize the computation of w_r in Equation (4.3), the binary auxiliary variable

$$b_k^i \in \{0,1\} \text{ with } b_k^i = \begin{cases} 0, & \text{OD pair } k \text{ has less than } i \text{ available routes,} \\ 1, & \text{OD pair } k \text{ has at least } i \text{ available routes} \end{cases}$$

to count the number of available routes of OD pairs is introduced. Using variable b_k^i, the route choice constraints (4.3) can be linearized with the following set of constraints

$$w_r \le y_r \qquad \forall k \in \mathcal{OD},\ r \in \mathcal{C}_k \tag{4.4a}$$

$$w_r \le \frac{w_k}{i} - b_k^i + 1 \qquad \forall k \in \mathcal{OD},\ r \in \mathcal{C}_k,\ i \in \{1,\dots,|\mathcal{C}_k|\} \tag{4.4b}$$

$$\sum_{r \in \mathcal{C}_k} w_r = w_k \qquad \forall k \in \mathcal{OD} \tag{4.4c}$$

Constraints (4.4a) ensure that w_r is positive only if route r is available, that is, if $y_r = 1$. Constraints (4.4b) impose an upper bound of $\frac{w_k}{i}$ to w_r if at least i routes are available for OD pair k, that is, if $b_k^i = 1$. For increasing i, $\frac{w_k}{i}$ decreases and these constraints get tighter. If less than i routes are available and $b_k^i = 0$, the right-hand side is greater than 1 and the constraints are redundant. Constraints (4.4c) ensure that the shares w_r of passengers of OD pair k using route r sum up to the share of travelers using public transport w_k. Together, constraints (4.4) model that the passengers of OD pair k distribute uniformly on all available routes from the set $\mathcal{C}_k$. Note that both the number of passengers and the number of available routes depend on the line plan.

The number of available routes can be counted within the model with the following constraints

$$\sum_{r \in \mathcal{C}_k} y_r \geq i \cdot b_k^i \qquad \forall k \in \mathcal{OD},\ i \in \{1, \ldots, |\mathcal{C}_k|\} \tag{4.5a}$$

$$b_k^i \geq \frac{1}{|\mathcal{C}_k|} \left(\sum_{r \in \mathcal{C}_k} y_r - i + 1 \right) \qquad \forall k \in \mathcal{OD},\ i \in \{1, \ldots, |\mathcal{C}_k|\} \tag{4.5b}$$

Constraints (4.5a) force b_k^i to 0 if less than i routes are available for OD pair k, and constraints (4.5b) ensure that b_k^i equals to 1 if at least i routes are available.

Mode choice

It remains to estimate the modal share w_k, that means, the share of travelers deciding to use public transport. The travelers' mode choice depends on both the utility of public transport and the utility of alternatives. We consider one alternative mode representing individual road transport, such as driving by private car. In general, changes in the public transport service affect the modal split, with this the congestion on roads, and eventually the utility of the alternative mode. However, as long as the utility of the public transport service is not changed substantially, the impact on the utility of the alternative mode is negligible. Hence, we assume that the utility of the alternative mode is independent of the utility of public transport in this research. As a consequence, the utility of the alternative mode is constant and the travelers' decision whether to take public transport or not solely depends on the utility of public transport.

Since a line plan is designed, not all information about the public transport service determining its utility is available. For example, departure and arrival times, as well as transfer durations are only known after a timetable is found. Nevertheless, the line plan determines the most important factors of influence for the travelers' decision about their mode of transport: the number of available routes, the approximate driving time, and the number of transfers.

While the number of available routes depends on the line plan, the approximate driving time and the number of transfers can be predetermined for each route $r \in \mathcal{C}_k$. Both are combined in the route journey time j_r that is known for each route. Since all routes in the choice set $\mathcal{C}_k$ have a very similar journey time, the journey time j_k for OD pair k is very close to the route journey time j_r for any route $r \in \mathcal{C}_k$ and does not depend on which routes are available. Hence, it can be predetermined as well and we define

$$j_k = \frac{1}{|\mathcal{C}_k|} \sum_{r \in \mathcal{C}_k} j_r \qquad \forall k \in \mathcal{OD}.$$

The journey time j_k captures the duration needed for OD pair k when traveling with public transport. It is a travel time equivalent comprising drive and dwell times, and possibly transfer penalties. The number of available routes indicates the frequency of the service for the passengers. The more routes are available within a time period, the fewer passengers have to adapt to the schedule of the public transport. This is quantified in the *adaption time* a_k, the time passengers need to deviate from their preferred departure or arrival times. With a more frequent service, the expected adaption time decreases inversely proportional.

We define the utility of the public transport for OD pair k as the sum of the journey time and the adaption time.

$$u_k = j_k + a_k \qquad \forall k \in \mathcal{OD} \tag{4.6}$$

The utility u_k does not depend on *which* routes are available, but only on the *number* of available routes. Thus, the number of available routes is the main factor of influence on travelers' mode choice that cannot be predetermined. However, it is possible to express the utility of public transport, and consequently the mode choice of travelers for each OD pair as a function of the number of available routes.

Figure 4.1 shows the expected share of travelers using public transport for different numbers of routes, exemplified for one OD pair. In this chapter, we use a logit model to estimate the travelers' mode choice and to derive the modal split. The logit model is a discrete choice model that is commonly used to estimate travelers' choices. The probability that an alternative is chosen depends on the utilities of all available alternatives. For the mode choice of travelers, just two alternatives are considered: traveling by public transport and traveling by individual transport such as by private car. Let u_k be the utility of public transport for OD pair k from Equation (4.6) and $\hat{u}_k$ be the utility of the alternative mode. Then, the logit model estimates the modal share for public transport as

$$w_k = \frac{e^{\beta u_k}}{e^{\beta u_k} + e^{\beta \hat{u}_k}},$$

where β is the logit coefficient to tune the model.

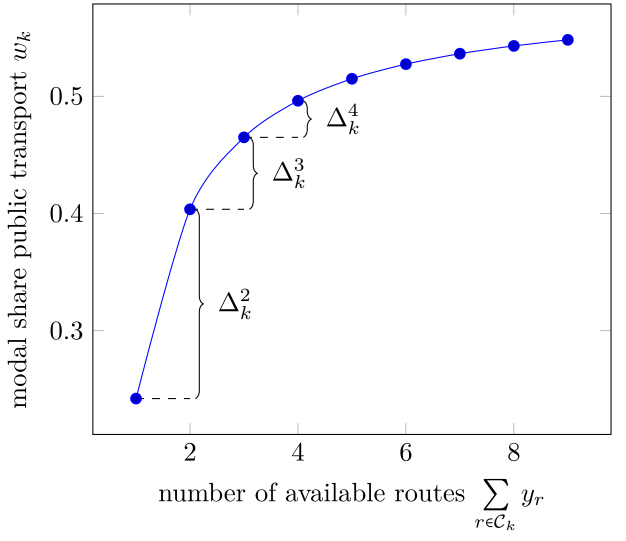


Figure 4.1: Estimated modal split for one OD pair k as a function of the number of available routes

The *increments in modal share* Δ_k^i indicated in Figure 4.1 express the additional share of travelers of OD pair k deciding to use public transport if i instead of $(i-1)$ routes are available. Using these values, the mode choice of travelers can be modeled linearly with the constraints

$$w_k = \sum_{i=1}^{|\mathcal{C}_k|} \Delta_k^i \cdot b_k^i \qquad \forall k \in \mathcal{OD} \tag{4.7}$$

As discussed for the passenger route choice in Section 4.3.3, the binary variables b_k^i equal 1 if OD pair k has at least i routes. Hence, constraints (4.7) set the modal share of public transport w_k for OD pair k dependent on the number of routes.

Note that the coefficients Δ_k^i can be predetermined for each OD pair k and each possible number of available routes $1 \le i \le |\mathcal{C}_k|$. Hence, the Δ_k^i are constant parameters in the model formulation. This framework allows the integration of travelers' mode choice according to any choice model using the linear constraints (4.7). Our model is thus not limited to the logit model which is used in this chapter.

4.3.4 Operational requirements

Capacity constraints

The strength of the demand modeling in Section 4.3.3 is that the number of passengers on each route can be estimated accurately. It should be ensured by the operator that there is sufficient seating capacity on each available route r for the number of passengers that is expected to choose route r. We model this as one capacity constraint per *line segment*, the part of a line l traversing a track t. Let κ_l be the *seating capacity* of line l, $\mathcal{P}(t)$ be the set of lines in the line pool that operate on track t and $\mathcal{C}_k(l,t)$ be the choice set of routes for OD pair k using line l on track t. The following constraints ensure that on each line segment sufficient capacity is provided by the line plan for all passengers on their chosen routes.

$$\sum_{k\in\mathcal{OD}} \sum_{r\in\mathcal{C}_k(l,t)} \delta_k \cdot w_r \leq \kappa_l \cdot z_l \qquad \forall t \in \mathcal{T},\ l \in \mathcal{P}(t) \tag{4.8}$$

Note that the presented model uses individual capacity constraints for each line segment and considers a distribution on routes according to passenger preferences. The combination of individual capacity constraints and passenger-optimal routes is important for accurate capacity estimation. It achieves that passengers use the same routes in the model as they would choose in real life, thus avoiding potential conflicts with capacity constraints. Existing line planning models often use only one capacity constraint per track t, aggregated over all lines operating on that track, or they assign passengers on a single route or according to a system optimum. Both can cause capacity conflicts unless passengers accept additional transfers to make space for other passengers, which is unrealistic.

Minimum service requirement

In addition to their aim to meet the passenger demand, most operators are required to offer a minimum level of service in certain parts of the networks. This ensures that all passengers have access to public transport, also in sparsely-populated areas. We model this as an additional set of constraints ensuring that at least f_t vehicles service track t.

$$\sum_{l\in\mathcal{P}(t)} z_l \geq f_t \quad \forall t \in \mathcal{T} \tag{4.9}$$

(LPwMC)

$$
\begin{array}{llcll}
\text{max profit = revenue - cost} & & & & \\
\max & \displaystyle\sum_{k\in\mathcal{OD}} p_k \cdot \delta_k \cdot w_k - \sum_{l\in\mathcal{P}} c_l \cdot z_l & & & \\
\text{link between line and route variables} & & & & \\
\text{s.t.} & y_r & \le & z_l & \forall k \in \mathcal{OD},\ r \in \mathcal{C}_k,\ l \in \mathcal{P}_r \\
 & y_r & \ge & \displaystyle\sum_{l\in\mathcal{P}_r} z_l - |\mathcal{P}_r| + 1 & \forall k \in \mathcal{OD},\ r \in \mathcal{C}_k \\
\text{passenger route choice} & & & & \\
 & w_r & \le & y_r & \forall k \in \mathcal{OD},\ r \in \mathcal{C}_k \\
 & w_r & \le & \dfrac{w_k}{i} - b_k^i + 1 & \forall k \in \mathcal{OD},\ r \in \mathcal{C}_k, \\
 & & & & i \in \{1,\dots,|\mathcal{C}_k|\} \\
 & \displaystyle\sum_{r\in\mathcal{C}_k} w_r & = & w_k & \forall k \in \mathcal{OD} \\
\text{counting of available routes} & & & & \\
 & \displaystyle\sum_{r\in\mathcal{C}_k} y_r & \ge & i \cdot b_k^i & \forall k \in \mathcal{OD},\ i \in \{1,\dots,|\mathcal{C}_k|\} \\
 & b_k^i & \ge & \dfrac{1}{|\mathcal{C}_k|}\left(\displaystyle\sum_{r\in\mathcal{C}_k} y_r - i + 1\right) & \forall k \in \mathcal{OD},\ i \in \{1,\dots,|\mathcal{C}_k|\} \\
\text{traveler mode choice} & & & & \\
 & w_k & = & \displaystyle\sum_{i=1}^{|\mathcal{C}_k|} \Delta_k^i \cdot b_k^i & \forall k \in \mathcal{OD} \\
\text{capacity constraints per line segment} & & & & \\
 & \displaystyle\sum_{k\in\mathcal{OD}} \sum_{r\in\mathcal{C}_k(l,t)} \delta_k \cdot w_r & \le & z_l \cdot \kappa_l & \forall t \in \mathcal{T},\ l \in \mathcal{P}(t) \\
\text{minimum service requirement} & & & & \\
 & \displaystyle\sum_{l\in\mathcal{P}(t)} z_l & \ge & f_t & \forall t \in \mathcal{T} \\
\text{domains of variables} & & & & \\
 & z_l & \in & \{0,1\} & \forall l \in \mathcal{P} \\
 & y_r & \in & \{0,1\} & \forall k \in \mathcal{OD},\ r \in \mathcal{C}_k \\
 & b_k^i & \in & \{0,1\} & \forall k \in \mathcal{OD},\ r \in \mathcal{C}_k, \\
 & & & & i \in \{1,\dots,|\mathcal{C}_k|\} \\
 & w_r & \in & [0,1] & \forall k \in \mathcal{OD},\ r \in \mathcal{C}_k \\
 & w_k & \in & [0,1] & \forall k \in \mathcal{OD}
\end{array}
$$

4.3.5 Line planning model with integrated mode choice

In this section, we give the mixed-integer linear program for line planning with integrated mode choice (LPwMC). As described in Section 4.3.2, the objective is to maximize profit, defined as revenue minus cost in Equation (4.1). The first part deals with the demand estimation explained in Section 4.3.3 including auxiliary modeling constraints. The linking constraints between route and line variables are discussed in Equations (4.2) and the uniform passenger distribution on available routes is modeled with Equations (4.4). The number of available routes is counted within the model with Equations (4.5) and the traveler mode choice according to any choice model is estimated with Equations (4.7). The second part deals with the operational requirements from Section 4.3.4. It is ensured by Equations (4.8) that sufficient seating capacity is available for the expected number of passengers on each line segment. This links the line selection and estimation of passenger demand in the model. Furthermore, the minimum service requirement in Equations (4.9) ensures a minimum frequency on each part of the network. The last part defines the domains of the variables.

4.4 Experimental setup

We solve the model (LPwMC) with the branch and bound method implemented in the Fico® Xpress Optimizer version 35.01. All experiments are conducted on the Lisa cluster[2] operated by SURFsara with a time limit of one hour per model run. In the following sections, the instances are introduced, the derivation of passenger routes is explained and the choices for parameters described.

4.4.1 Instances

We consider the Intercity network of the Randstad, a metropolitan area in the Netherlands. This is a partial network of the network operated by the largest Dutch railway operator, Netherlands Railways (NS). The network contains 21 stations connected by 31 direct tracks between them. The network is depicted in Figure 4.2 and denoted by $IC21$, indicating the number of stations in the network. The line pool $\mathcal{P}$ contains 107 lines, 43 of which are duplicates in order to be able to model higher line frequencies. The pool contains all relevant lines that currently operate in the considered area and is given in Appendix 4.B. A reference line plan is available that is used as a feasible start solution. The reference line plan is a solution of a line planning model with fixed passenger assignment. Based on this reference line plan, a competing mode such as driving by car, and passenger count data from NS, traveler demand was estimated with the logit model. This resulted in 174 undirected OD pairs with positive traveler demand.

[2] https://userinfo.surfsara.nl/systems/lisa

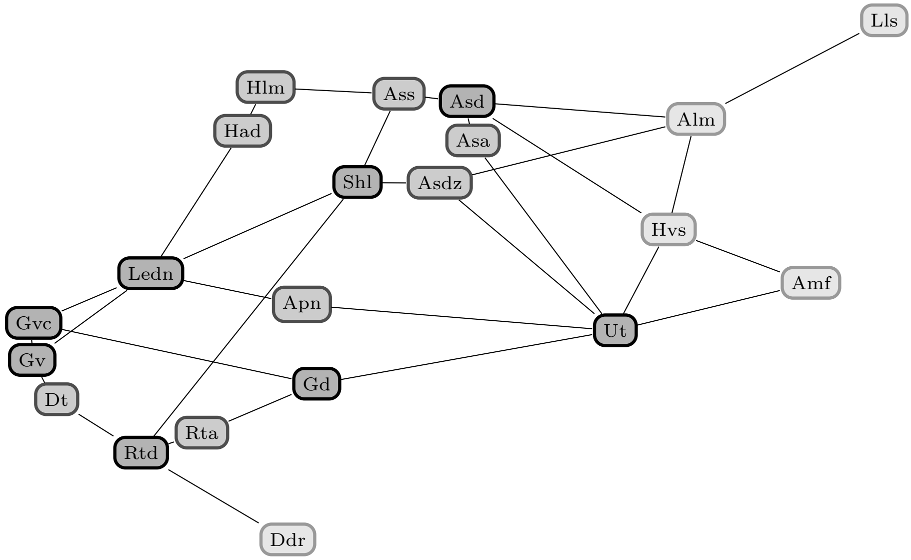


Figure 4.2: IC Network of the Randstad region in the Netherlands

To obtain a variety of instances in some experiments, we consider two additional instances $IC08$ and $IC16$ and randomized demand situations for all networks. These instances have 8 and 16 stations, respectively, and are sub-networks of network $IC21$. In Figure 4.2 the stations in network $IC08$ are marked by a dark node color scheme. The stations that are additionally in network $IC16$ have a lighter grey color scheme. The remaining stations with the lightest color scheme are only contained in network $IC21$. The *randomized demand situations* are generated by multiplying the number of travelers δ_k with a random number between 0.5 and 1.5 for each OD pair k.

4.4.2 Passenger routes

In a preprocessing step, a choice set of passenger routes $\mathcal{C}_k$ for each OD pair k is determined and used as input to the model (LPwMC). As described in Section 4.3.3, we consider the journey-time shortest passenger routes for the route choice sets. In our experiments, we model this with a tolerance coefficient α and a tolerance addend ε to limit the maximum acceptable journey time. First, we derive the shortest possible journey time $\hat{j}_k$ for each OD pair based on an extensive line pool for each network. Then, a route r is in the set $\mathcal{C}_k$, if and only if its journey time j_r is at most $\alpha\hat{j}_k + \varepsilon$. Only routes that are at most 5% ($\alpha = 1.05$) and $\varepsilon = 10\,\text{min}$ longer than the shortest possible journey time were accepted. The journey time of a route comprises drive and dwell times and a transfer penalty, if applicable. Average driving times per track are used and at all stations a dwell time of 4 min is assumed. A penalty of 20 min is added to the journey time of routes including a transfer. In the experiments, we restrict the route choice sets to routes with at most one transfer.

This yields a set $\mathcal{C}_k$ of passenger routes of comparable journey time for each OD pair k. For the Intercity network of the Randstad $IC21$ from Figure 4.2, a total of 6391 routes are considered, which is on average 36.7 routes per OD pair.

4.4.3 Parameters

The monetary values for ticket price p_k for a passenger of OD pair k, and cost c_l related to line operation were chosen to represent a simplified situation for the Dutch railway operator NS. All lines are operated by trains with a capacity of $\kappa_l = 1000$ seats.

For estimating the mode choice, the logit model is used with a logit parameter of $\beta = -0.1$. We estimate the adaption time based on the assumption that demand is uniformly distributed over the period, and that route departures are spread evenly over the hour. We therefore arrive at an adaption time of half of the considered period of 60 min, divided by the number of available routes. The alternative mode resembling individual transport does not have an adaption time. Hence, the utility $\hat{u}_k$ of the alternative mode is quantified by the journey time only. We use the SAQ method (FGSV, 2008) to estimate the journey time based on the Euclidean distance between stations.

For the minimum service requirement (4.9), a minimum track frequency of $f_t = 2$ for all tracks $t \in \mathcal{T}$ is used. This is in line with the requirements for the Dutch railway operator NS.

4.5 Comparison and analysis

To test the line planning model with integrated mode choice, we conduct experiments on the Randstad network. We test means to improve the computational performance and compare our model (LPwMC) with a standard line planning model with fixed passenger demand to investigate differences in solution quality.

4.5.1 Improvement of computational performance

We test the impact of adding order constraints (*order*), adding symmetry breaking constraints (*sym*), and setting priorities for branching (*prio*) on the solution-finding process. We observe the CPU time until an optimal solution is found, and the gap to the best bound found in case the time limit of 1 hour is exceeded. The settings *order*, *sym* and *prio* are explained in the following.

order The auxiliary variables b_k^i are used for counting the number of available routes for each OD pair. By definition, b_k^i equals 1 if for OD pair k at least i routes are available, and 0 otherwise. This implies that b_k^i can only be 1 if b_k^{i-1} equals 1. The other way around, b_k^i can only be 0 if b_k^{i+1} equals 0. This relation can be modeled with the order constraints

$$b_k^i \geq b_k^{i+1} \quad \forall k \in \mathcal{OD},\ \forall i \in \{1, \dots, |\mathcal{C}_k| - 1\}$$

sym The binary decision variables z_l model whether line l is selected in the line plan or not. To model higher frequencies of a line, duplicates are considered in the line pool that can be selected independently. To break the symmetry implicated by this setting, we consider constraints enforcing an order of selection of a line and its duplicates.

prio The solution of the line planning problem is uniquely determined by the line plan $\mathcal{L}$, that means, the solution values for the line selection variables z_l. The corresponding solution values of all other variables can be reconstructed from the solution values of z_l. However, model (LPwMC) uses three different sets of binary variables, z_l, y_r, and b_k^i, where y_r and b_k^i are auxiliary variables to model the availability and the number of routes. By default, any of these variables can be used for branching. We test whether branching first on the variables z_l is preferable to the standard branching strategy of Fico Xpress. The Xpress Optimizer offers the option to set the branching priority of a variable between 0 and 1000, where always a variable with a lower priority number will be selected for branching. We set a high branching priority (1) for variables z_l, medium branching priority (500) for variables y_r, and low branching priority (999) for variables b_k^i.

The tests are conducted on the three networks $IC21$, $IC16$, and $IC08$ with 10 randomized demand situations per network. Table 4.1 gives the CPU times in seconds and the gaps to the best bound for solving the model (LPwMC). The CPU times

	average CPU time[s]			average gap[%]		
	$IC08$	$IC16$	$IC21$	$IC08$	$IC16$	$IC21$
none	166	3600	3600	-	30.5(10)	8.9(10)
order	144	3600	3600	-	11.4(10)	17.2(10)
sym	31	303	3194	-	-	1.6(5)
prio	1034	3600	3600	-	2.5(10)	9.3(10)
all	32	248	1305	-	-	-

Table 4.1: CPU times in seconds and gaps in percent to the best bound for model (LPwMC) in different settings. The number in brackets gives the number of cases where the time limit of 1 hour is exceeded

in Table 4.1 are averaged over ten randomized demand situations on each network. Only those runs that exceed the time limit are included for computing the average gap. The number in brackets behind the average gap shows how often the time limit of 1 hour is exceeded. The first and last row show results for the settings where *none* (reference case) or *all* three options are used, respectively.

Adding order constraints (setting *order*) reduces the average CPU time for the smallest network and the average gap for the midsize network. However, it increases the average gap for the largest network compared to the reference case (setting *none*). The reason might be that the additional constraints increase the problem size. This could initially make the search for a feasible solution more difficult, but accelerate the solution process at a later point in time. The symmetry-breaking constraints (setting *sym*) significantly reduce the CPU times and gaps for all instance sizes. For the largest instances with 21 stations, the time limit is exceeded in only five out of ten cases and the resulting average gap is with 1.6% very small. Setting branching priorities (setting *prio*) drastically increases the CPU time for $IC08$, showing that the solver was able to find better branching strategies for the small network. In contrast to that, the average gap for the medium network size is significantly reduced by setting the branching priorities, and the found solutions were close to the optimum. For the large instances, no improvement is found with the setting *prio*. The gaps are slightly higher than in the reference case without branching priorities. Tests with all strategies combined (setting *all*) yield by far the best CPU times and all instance sizes can be solved to optimality within the time limit of one hour. The largest instance considered with 21 stations is solved within an average CPU time of less than 22 minutes. Therefore, we keep this setting with all options (*order*, *sym* and *prio*) for further experiments.

4.5.2 Comparison of (LPwMC) with line planning without mode choice

To investigate the added value of estimating passenger loads during optimization, we compare model (LPwMC) with a line planning model (LP) without integrated mode or route choice. Such a model requires a passenger assignment to tracks as input and assumes that passenger demand and distribution are independent of the solution found. The objective is to find a line plan meeting all demand with minimal cost. Since the passenger demand is assumed to be fixed, the revenue is constant and this objective corresponds with finding a profit-optimal line plan in model (LPwMC). Similar to model (LPwMC), model (LP) considers a minimum service requirement and seating capacity constraints. In contrast to model (LPwMC), the capacity constraints are aggregated per track. Individual constraints per line segment are not feasible since the passengers are assigned to the network before the line plan is found. The MIP formulation for the line planning model (LP) used in the experiments can be found in Appendix 4.C.

To obtain a passenger assignment for model (LP), we assume a fixed percentage of travelers to use public transport and distribute them uniformly on the routes in the choice set $\mathcal{C}_k$. This yields a passenger load on each track in the network. We test model (LP) with an assigned modal share for public transport ranging from 40% to 90%.

Table 4.2 shows the modal share MS for public transport in decimals, and the revenue R, cost C, and profit P of the found line plans. The values are given for model (LP) with different assigned passenger shares and for model (LPwMC). The column $P^{(\mathrm{LP})}$ gives the objective value of model (LP), that is, the profit assuming the assigned passenger demand used as input. This value is only available for model (LP) with assigned passenger shares. All monetary values are given in relation to the profit of the line plan found by model (LPwMC), which is normalized to the value 100. That means, a value of 110 implies that the corresponding monetary value is 10% higher than the profit of the solution of model (LPwMC).

The numbers in brackets in the first column give the share of travelers in decimals that are *assigned to use* public transport in the input for model (LP). Based on the found solution, we estimate the expected share of travelers that *decide to use* public transport in a subsequent distribution of passengers with the logit model. This average modal share MS coincides with the integrated modal share estimate

	MS	R	C	P	$P^{(\mathrm{LP})}$
(LP) (0.4)	0.36	174.1	115.3	59.2	98.4
(LP) (0.5)	0.36	176.0	115.3	60.7	151.7
(LP) (0.6)	0.42	205.0	128.7	76.3	191.6
(LP) (0.7)	0.42	203.4	129.3	74.1	244.5
(LP) (0.8)	0.44	211.2	133.6	77.6	293.8
(LP) (0.9)	0.48	229.9	155.5	74.5	325.2
(LPwMC)	0.49	240.5	140.5	100.0	-

Table 4.2: Results of models (LP) with different passenger shares and (LPwMC). The modal share is given as decimal and the monetary values are normalized such that the profit of the solution of (LPwMC) equals 100

in model (LPwMC) and is given in the second column of Table 4.2. For all tests with model (LP), the *assigned* share of travelers is higher than the *estimated* modal share MS. While the modal share is comparable for a low number of assigned travelers, the assigned and estimated modal share significantly differ for high numbers. This shows that in our experiments, the solutions of model (LP) do not attract as many passengers as they were planned for.

In particular, line planning models without an integrated estimation of mode choice are not suitable for operators that strive for increasing their modal share. It is striking that the highest modal share is achieved by the solution of model (LPwMC).

The monetary observation variables revenue R, cost C, and profit P are based on the modal share MS as estimated with the logit model. Both revenue and cost increase with the given passenger share for model (LP) but the profit stagnates at around 75% of the profit generated by the solution of model (LPwMC). The last column gives the anticipated profit $P^{(\mathrm{LP})}$ based on the *assigned* share of travelers, i.e., the objective value of model (LP). Especially for high values of the assigned passenger share, the estimated profit $P^{(\mathrm{LP})}$ is significantly higher than the corresponding profit based on estimated passenger numbers. By fixing the modal split before making the line plan, line planning models such as (LP) are prone to drastically overestimating the number of passengers and, with this, their revenue and profit. This causes these models to choose for solutions in which many lines are established, without attracting sufficient passengers to be profitable in the end.

The experiments clearly show the advantages of our model (LPwMC) over line planning models without an integrated mode choice. Due to an accurate estimation of passenger demand, profitable line plans of high quality can be designed that attract a high number of travelers.

4.6 Practical insights

Over years, travel demand is constantly changing and public transport operators need to react with an adjustment of their service. For operators, it is important to understand the impact of changes in travel demand on the quality of their service and the generated profit. Due to the integration of route and mode choices, model (LPwMC) is capable of computing profit-optimal line plans for different levels of travel demand. In this section, we outline which insights this can provide for operators.

The integrated traveler mode choice allows conducting a sensitivity analysis of the travel demand on passenger service level and operator performance. The different levels of traveler numbers are obtained by multiplying the original traveler demand with a factor ranging from 0.5 to 1.5. Figure 4.3 shows the results per traveler factor. The corresponding data is given in Table 4.4 in Appendix 4.D. We analyze the number of lines $|\mathcal{L}|$ in the line plan, the number of available routes $|\mathcal{R}|$, the modal share MS, and the profit P of the line plans found by model (LPwMC) for different levels of traveler demand. The profit P is normalized such that the profit of the solution for the traveler factor 1.0 equals 100. The number of available routes $|\mathcal{R}|^{OD}$ averaged per OD pair and the number of available routes $|\mathcal{R}|^{pax}$ averaged per passenger are examined separately.

With higher traveler numbers, more lines $|\mathcal{L}|$ can be installed and the service for passengers improves. Accordingly, the number of available routes for OD pairs $|\mathcal{R}|^{OD}$ and passengers $|\mathcal{R}|^{pax}$ increase. Large OD pairs mostly have direct routes in the route choice sets, while for small OD pairs we often observe transfer routes. Since the number of lines increases, many more line combinations form feasible transfer routes for passengers. Hence, the number of available routes increases proportionally more for small OD pairs than for large OD pairs, which explains the steeper increase of $|\mathcal{R}|^{OD}$ than of $|\mathcal{R}|^{pax}$.

The higher number of available routes implies an improved service level for passengers and with this a higher modal share for public transport. The increase in modal share from 44% to 52% is rather moderate, considering that the overall number of travelers as specified in the input data triples, and the average number of routes per OD pair almost doubles. The low effect on the modal split can be explained with the structure

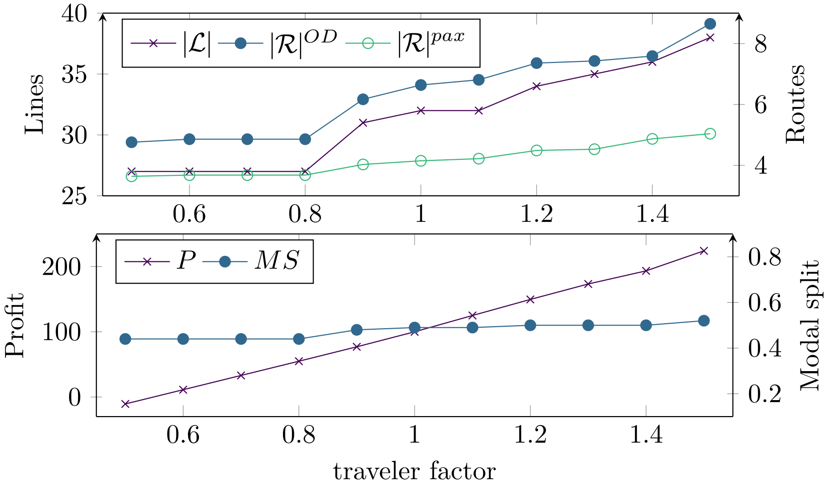


Figure 4.3: Results of model (LPwMC) for varying passenger load

of the choice model that we use to estimate the travelers' mode choice. As indicated in Figure 4.1, the increase in modal share flattens for a higher number of available routes. This is in line with observations in the real world where the modal split is hardly affected by improvements in service for passengers, once a certain service level is reached.

Nevertheless, the profit increases approximately linearly from approximately negative 10% to 225% of the reference profit for an increasing number of travelers. The negative profit values for low passenger numbers are a result of the minimum level of service required by governmental regulations, which enforces to operate two lines on each track, even if this cannot be done profitably. This increase in profit can be explained by the high costs for operating a basic line plan on the whole network. For up to 80% of the travelers, the capacity of such a basic line plan is sufficient on the considered instance. Until then, the number of lines, the number of routes, and the modal split stay almost constant. Only for more travelers, and thus for more passengers, it pays off to install more lines. It is interesting to see that the slightly higher modal share and thus the higher revenues offset the costs for additional lines and, overall, lead to an approximately linear increase in profit.

The evaluation shows that the operator's profit is very sensitive to changes in the traveler demand. In case of declining traveler demand, operators cannot prevent losses even if they react to demand changes in an optimal way. This is in line with recent observations where operators incur tangible losses due to considerably lower traveler demand caused by the Covid-19 pandemic despite efforts to reduce the service level. Conversely, operators can profit greatly from growing traveler demand when exploiting the full potential of passenger demand estimation during line planning.

4.7 Conclusion and outlook

4.7.1 Conclusion

The line plan significantly determines the service level of public transport for passengers. It has an impact on how many travelers decide to use public transport, and which routes passengers use. In this chapter, we present a line planning model with integrated mode and route choice models. In contrast to most existing approaches, both choices are modeled from a passenger's perspective and are not driven by a system optimum. This allows an accurate estimation of passenger demand during optimization, resulting in line plans that are tailored for the demand they generate.

Considering passenger choice models during line plan optimization is a complex and hard-to-solve problem. In order to obtain a tractable model, we assume (1) that passengers distribute uniformly on the best available routes and (2) that the utilities of alternative modes are independent of the designed line plan. By considering only shortest routes for passengers and making these two assumptions, the presented model can be linearized and solved with existing branch and bound methods.

The mixed-integer linear program presented in this chapter may be combined with any choice model to estimate the mode choice of travelers. In particular, the choice model does not need to be linear. Due to the two assumptions made, the traveler mode choice can be preprocessed using the preferred choice model without affecting the solving of the line planning model. In the chapter, a logit model is used to estimate the mode choice.

We discuss additional constraints and branching priorities for improving the computational performance and show their effectiveness in experiments. The advantages of integrated passenger choice models are outlined in a comparison with a standard line planning model that relies on predetermined passenger loads. Based on the results of this comparison, considering demand estimation during line planning is strongly recommended. Integrated demand estimation yields line plans that are well-suited for the demand they generate. They are more profitable for operators and feature a higher level of service for passengers compared to line plans found based on fixed passenger demand. Furthermore, we analyze the sensitivity of the public transport service level and operator profit on fluctuations in travel demand. This gives valuable insights into the business models of operators and suggests that operators should react to changes in travel demand regularly.

4.7.2 Outlook

In additional experiments, we have noticed that the profit of two line plans can be very similar while their respective costs and revenues are different from each other. Indeed, two line plans with the same profit do not need to have any line in common. This is an interesting observation from both an algorithmic and an application point of view.

On the one hand, this implies that the concept of 'neighborhood' of solutions, although a key element in many (heuristic) approaches to solve line planning models, is less useful than in other contexts. On the other hand, for operators, it can be very valuable to see different solutions with similar profit. It would allow them to choose between solutions with different modal shares. This particularly motivates the search for multiple good solutions instead of one optimal solution only. One approach to obtain multiple solutions could be to modify the ticket price p_k for passengers or the operational costs c_l of lines to shift the weights between revenue and cost in the objective function.

Future research should address customized solution approaches for solving model (LPwMC) to find line plans for larger instances, and possibly with the feature to compute multiple good solutions.

Appendix

4.A Notation

Sets

$\mathcal{C}$	Route choice set
$\mathcal{L}$	Line plan
$\mathcal{OD}$	OD pairs
$\mathcal{P}$	Line pool
$\mathcal{R}$	Available routes
$\mathcal{S}$	Stations
$\mathcal{T}$	Tracks between stations

Indices

i	Iterator used to count available routes
l	Line
k	OD pair
r	Route
s	Station
t	Track between two stations

Parameters

a_k	Adaption time for OD pair k
β	Logit parameter
c_l	Cost for operator for installing line l
δ_k	Traveler demand of OD pair k
Δ_k^i	Increment in modal share if i instead of $i-1$ routes are available
f_t	Minimum frequency on track t
j_k	Journey time for OD pair k
κ_l	Capacity of line l
p_k	Ticket price for OD pair k for using public transport
u_k	Utility of public transport for OD pair k
$\hat{u}_k$	Utility of alternative mode for OD pair k

Variables

b_k^i	$\in \{0,1\}$: OD pair k has at least i available routes or not
w_k	$\in [0,1]$: Share of travelers of OD pair k using public transport
w_r	$\in [0,1]$: Share of passengers using route r
y_r	$\in \{0,1\}$: Route r is available or not
z_l	$\in \{0,1\}$: Line l is selected or not

Observation variables

C	Cost for operator
MS	Modal share according to logit model
P	Profit (= revenue - cost)
R	Revenue generated by transporting passengers

4.B Line pool

count	sequence of stations									
2	Alm	Asd								
1	Alm	Asd	Ass	Hlm						
1	Alm	Asdz	Shl	Ledn						
1	Alm	Asdz	Shl	Ledn	Gv					
1	Alm	Asdz	Shl	Ledn	Gvc					
1	Alm	Asdz	Shl	Rtd						
1	Alm	Asdz	Shl	Rtd	Ddr					
1	Alm	Hvs	Amf							
2	Alm	Hvs	Ut							
1	Alm	Lls								
2	Amf	Hvs	Asd							
1	Amf	Hvs	Alm	Lls						
1	Amf	Hvs	Asd	Ass						
2	Amf	Hvs	Asd	Ass	Hlm					
1	Amf	Ut								
1	Amf	Ut	Apn	Ledn						
2	Amf	Ut	Gd	Gvc						
4	Amf	Ut	Gd	Rta	Rtd					
1	Amf	Ut	Gd	Rta	Rtd	Ddr				
1	Asd	Alm	Lls							
1	Asd	Ass	Hlm	Had	Ledn					
3	Asd	Ass	Hlm	Had	Ledn	Gvc				
2	Asd	Ass	Shl	Ledn	Gv					
1	Asd	Ass	Shl	Ledn	Gvc					
1	Asd	Ass	Shl	Ledn						
1	Asd	Ass	Shl	Rtd						
1	Asd	Ass	Shl	Rtd	Ddr					
1	Asd	Shl	Gv	Rtd						
3	Asd	Shl	Rtd							
1	Asd	Shl	Rtd	Ddr						
2	Asdz	Shl	Ledn	Gvc						
1	Ass	Asd	Alm	Lls						
4	Ass	Asd	Asa	Ut						
1	Ass	Shl	Ledn	Gv						
2	Ass	Shl	Ledn	Gvc						
1	Ass	Shl	Rtd	Ddr						
1	Ddr	Rtd								
1	Ddr	Rtd	Dt	Gv						
1	Ddr	Rtd	Dt	Gv	Gvc					
1	Ddr	Rtd	Dt	Gv	Ledn					
1	Ddr	Rtd	Dt	Gv	Ledn	Had	Hlm			
2	Ddr	Rtd	Dt	Gv	Ledn	Shl	Ass	Asd	Alm	Lls
2	Ddr	Rtd	Rta	Gd	Ut					
2	Ddr	Rtd	Shl	Asdz	Alm	Lls				
1	Gv	Ledn								
1	Gv	Ledn	Had	Hlm						
1	Gv	Ledn	Shl	Asdz	Alm	Lls				
2	Gv	Ledn	Shl	Ass						
5	Gvc	Gd	Ut							
3	Gvc	Gv	Dt	Rtd						
1	Gvc	Ledn								
1	Gvc	Ledn	Had	Hlm						
2	Gvc	Ledn	Shl	Asdz	Alm	Lls				
2	Gvc	Ledn	Shl	Ass	Asd	Alm	Lls			
2	Hlm	Ass	Asd	Alm	Lls					
1	Hlm	Had	Ledn							
2	Hlm	Had	Ledn	Gv	Dt	Rtd				
3	Ledn	Apn	Ut							
1	Ledn	Gv	Dt	Rtd						
3	Ledn	Shl	Asdz	Alm	Lls					
1	Lls	Alm	Asdz	Shl	Rtd					
2	Lls	Alm	Hvs	Ut						
4	Rtd	Rta	Gd	Ut						
4	Shl	Asdz	Ut							

Table 4.3: Line pool for Intercity network of the Randstad. The first column gives the number of occurrences of the line in the pool and the remaining columns give the sequence of stations on the line

4.C MIP formulation for a standard line planning model with fixed passenger demand

max profit = revenue - cost

$$\max \quad \sum_{k \in \mathcal{OD}} p_k \cdot \delta_k \cdot w_k - \sum_{l \in \mathcal{P}} c_l \cdot z_l$$

capacity constraints per track

$$\sum_{k \in \mathcal{OD}} \sum_{r \in \mathcal{C}_k(t)} \delta_k \cdot w_r \leq \sum_{l \in \mathcal{P}(t)} z_l \cdot \kappa_l \quad \forall t \in \mathcal{T}$$

minimum service requirement

$$\sum_{l \in \mathcal{P}(t)} z_l \geq f_t \quad \forall t \in \mathcal{T}$$

domains of variables

$$z_l \in \{0,1\} \quad \forall l \in \mathcal{P}$$

The objective is to maximize profit, defined as revenue minus cost. Note that the mode choice of travelers w_k is assumed to be known and fixed in this model. Hence, the revenue is constant and the objective is equivalent to minimizing cost. Similarly, the passenger assignment w_r to routes is predetermined and constant in the model. Consequently, the constraints for determining passenger loads are omitted and only the constraints ensuring sufficient capacity and a minimum level of service on each track remain. Passengers are assigned to routes in the route choice set to obtain the approximate passenger load on each track between to stations. Since it is not known yet which routes will be available in the solution, only aggregated capacity constraints per track can be applied. The set $\mathcal{C}_k(t)$ denotes the set of all routes in the choice set for OD pair k via track t.

4.D Results sensitivity analysis on travel demand

	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5
$\|\mathcal{L}\|$	27	27	27	27	31	32	32	34	35	36	38
$\|\mathcal{R}\|^{OD}$	4.76	4.86	4.86	4.86	6.17	6.64	6.81	7.36	7.43	7.59	8.65
$\|\mathcal{R}\|^{pax}$	3.64	3.68	3.68	3.68	4.03	4.15	4.22	4.49	4.53	4.87	5.04
MS	0.44	0.44	0.44	0.44	0.48	0.49	0.49	0.5	0.5	0.5	0.52
P	-10.6	11.2	33.1	54.9	77.0	100.0	125.0	149.6	173.3	193.4	224.1

Table 4.4: Results of model (LPwMC) for varying passenger load. The number of lines and routes are counts. The modal share is given as decimal and the monetary values are normalized such that the profit of the solution with traveler factor 1.0 equals 100

Part II

Supply estimation in travel demand modeling

Chapter 5

Vehicle Scheduling for On-demand Vehicle Fleets in Macroscopic Travel Demand Models

5.1 Introduction

In densely populated areas public transport is more efficient than private means of transport for two reasons. First, it pools the trips of several people. This leads to a high occupancy rate and thus reduces the total vehicle distance traveled. Second, service trips are concatenated to vehicle schedules. This reduces the number of required vehicles.

In contrast to traditional public transport, on-demand services in the form of car-sharing or ridesplitting (definitions from Feigon and Murphy (2016)) have neither a fixed route nor a predetermined timetable. The actual vehicle schedules are only known at the end of an operating day. For the planning of on-demand services, travel demand models are used to either determine the number of served passengers for a given fleet size or to determine the required fleet size for a given demand situation.

So far, primarily microscopic travel demand models are used for modeling on-demand services (see literature review in Section 5.2). Microscopic models simulate the demand of individuals and the operational processes at the level of individual vehicles using agent-based approaches. Typical input values are trip requests coming from a demand model, fleet size, vehicle capacity, and service parameters, e.g. maximum detour factor for passengers. Using some type of vehicle scheduling process, the models deliver as results indicators describing the service quality from the perspective of passengers (e.g. waiting time, in-vehicle-time, detour factor, number of passengers not served) and operators (e.g. empty and loaded vehicle kilometers, occupancy rates, revenues).

In this chapter, we present an algorithm for the vehicle scheduling process of on-demand services, which can be embedded in macroscopic travel demand models. The presented approach is intended for long-term, strategic transport planning. For this purpose, it provides planners with an estimate of vehicle fleet size and distance traveled by on-demand services. The model aims to extend the four-step model such that public transport authorities can use it for estimating vehicle fleet size and infrastructure utilization by on-demand services. Using a macroscopic travel demand model has advantages and disadvantages compared to a microscopic approach. Important advantages include the following:

- Many cities, regions, and states use macroscopic travel demand models to quantify the impacts of supply changes on travel demand. Moeckel et al. (2019) report that most states in the U.S. operate macroscopic travel demand models.

A survey by Rieser et al. (2018) finds that most travel demand models operated on regional and state level in German speaking countries are macroscopic models.

- A macroscopic travel demand model replicates the demand of an average day in one model run. It works with probabilities so that every model run produces the same solution. A microscopic model simulates a certain day and requires multiple simulations to obtain results for an average day.
- Macroscopic model implementations are usually faster.

The main disadvantage of macroscopic models is probably that they can reproduce the traffic-related decision processes of activity choice, destination choice, mode choice, departure time choice, and route choice only in a simplified way. Microscopic models can capture a more complex decision process considering household constraints, vehicle ownership, and temporal constraints coming from an activity schedule. In case average results are obtained with multiple simulations, microscopic models also give information about the variability of the results. In case average results are obtained with multiple simulations, microscopic models also give information about the variability of the results.

Looking at the pooling and vehicle scheduling processes in connection with on-demand services, macroscopic travel demand models bring a further challenge: Demand is not represented as discrete trips of individuals but as a probability. This leads to a non-integer demand that is stored in demand matrices. For modeling fluctuations in travel demand over the course of a day, demand is divided into time intervals, e.g. 96 intervals of 15 minutes. This rather abstract representation of travel demand requires specific methods for integrating on-demand services into a macroscopic travel demand model. Friedrich et al. (2018) describe an algorithm to pool macroscopic travel demand to demanded vehicle trips (=service trips).

In this chapter, we present an algorithm for the vehicle scheduling problem that uses these time-dependent vehicle trips as input. As a result, the algorithm determines the number of required vehicles and empty trips for vehicle relocation per time interval. The efficient algorithm design makes it suitable for solving large instances in short computation time, which is crucial for the use in travel demand models. Furthermore, it can be applied for both integer and non-integer demand matrices, which also allows an application to microscopic models with integer demand. A python implementation can be found in Hartleb et al. (2020).

The contribution of this chapter is twofold: First, we develop a vehicle scheduling algorithm to estimate the vehicle fleet size of on-demand services efficiently. Second, we show in two case studies that the algorithm is suitable for the vehicle scheduling problem as it arises in macroscopic travel demand models.

The remainder of this chapter is structured in the following way: In Section 5.2, the presented approach is compared to existing research on on-demand services in travel demand models and on vehicle scheduling approaches. Section 5.3 defines the vehicle scheduling problem in a formalized way, followed by a description of the basic algorithm in Section 5.4. In Section 5.5, extensions of the basic algorithm are discussed. Section 5.6 illustrates the applicability of the algorithm in two case studies. A conclusion and outlook complete this chapter in Section 5.7.

5.2 State of the art and related work

In section 2.1, we relate our work to existing research on on-demand services in travel demand models. The differences of macroscopic and microscopic travel demand models, i.e. agent-based models, are highlighted. In section 2.2, we report on related vehicle scheduling approaches and their solution techniques.

5.2.1 On-demand services in travel demand models

Travel demand models replicate the decision-making process of travelers, which is triggered by the need of people to participate in activities. According to Friedrich et al. (2016), “these decisions range from long-term to short-term decisions. Long-term decisions cover decisions concerning the place of residence and the workplace. These decisions influence subsequent medium-term decisions regarding the purchase of a car or a season ticket for public transport, which then affect later decisions on the activity locations and the transport modes. Short-term decisions on departure time, a certain route or a certain lane are taken within a short time horizon.” Most transport models cover only some of these decisions or replicate some decisions in a simplified way. Many macroscopic travel demand models capture the decisions associated with the pursuit of activities within the framework of the four-step algorithm. This framework distinguishes the steps trip generation, destination choice, mode choice and route choice (de Dios Ortúzar and Willumsen, 2011; McNally, 2010). To consider temporal travel patterns, macroscopic models are supplemented by a step for departure time choice. This step requires a model implementation, which distinguishes trip matrices not only by trip purpose but rather by activity pairs (e.g. Home-Work, Home-Education, Home-Shopping, Work-Shopping). For each activity pair observed temporal distributions are used to compute time-dependent trip tables. Figure 5.1a provides a schematic flow chart of a standard travel demand model.

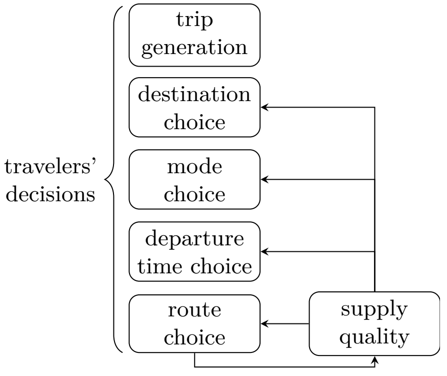
While there are many approaches to model pooling and scheduling of on-demand services in general, only a limited number of modeling approaches replicate the impacts of on-demand services on travel demand, especially on destination and mode choice. Almost all of these are microscopic approaches. In their overview of demand models including one-way carsharing services, Vosooghi et al. (2017) come to the same conclusion.

Examples of microscopic approaches including on-demand services in existing travel demand models are presented by Azevedo et al. (2016) for SimMobility, Maciejewski (2016) and Hörl (2017) for MATSim, Heilig et al. (2018) and Wilkes et al. (2019) for mobiTopp and Martínez et al. (2017) for an agent-based model for Lisbon. A macroscopic approach is described by Richter et al. (2019).

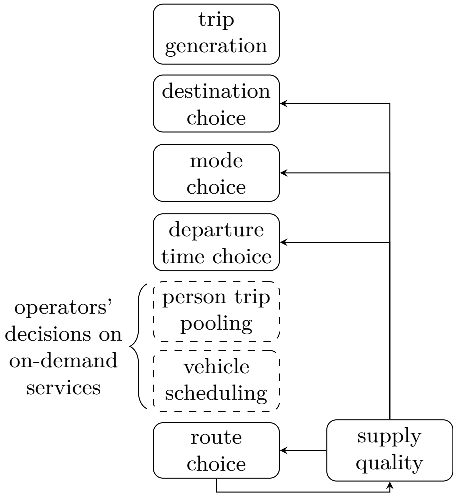
All approaches need to deal with the challenge that the transport supply provided by on-demand services depends on the demand and is not given as model input. Although traditional models capture the impact of demand on the supply in form of volume-delay functions, the spatial and temporal structure of the supply remains fixed. For on-demand services, however, the spatial and temporal supply structure must be adapted to the demand.

Therefore, to include on-demand services in the four-step algorithm, the travel demand model needs to be extended by an additional set of steps determining the on-demand supply. These steps replicate short-term decisions of operators which schedule the on-demand supply. Hence, the structure and availability of the on-demand supply are established in response to the trip requests of travelers. Figure 5.1b extends the algorithm of Figure 5.1a to include the additional steps. The short-term decisions of operators can be categorized into two parts: First, pooling of passenger trip requests into vehicle trips and, second, scheduling of vehicles to serve the vehicle trips.

The pooling step converts person trip requests into vehicle trip requests. This step is only required for on-demand services which aim at pooling several independent travelers into one vehicle, i.e. for ridepooling services. On-demand carsharing does not require a pooling step as users directly request a vehicle trip. Microscopic ap-



(a) Four-step algorithm supplemented by departure time choice



(b) Extended four-step algorithm to model the impact of on-demand services

Figure 5.1: Standard and extended travel demand model

proaches to pooling algorithms can be found in Zhang et al. (2015), Bischoff et al. (2017) or Engelhardt et al. (2020), for example. A macroscopic approach is proposed by Friedrich et al. (2018).

The focus of this chapter is on the vehicle scheduling step, where vehicle trip requests are assigned to specific vehicles. This step either determines the number of vehicles needed for serving a given demand or defines the demand which can be served by a given vehicle fleet. The scheduling step also identifies empty vehicle trips which are required for vehicle relocation. Approaches to replicate this step differ for microscopic and macroscopic travel demand models as discussed in the following.

Microscopic or agent-based travel demand models simulate discrete choices of persons using probability distributions. Each model run replicates the demand situation of a specific day. This modelling approach is described for example by de Dios Ortúzar and Willumsen (2011) or Horni et al. (2016). Commonly, persons are assigned daily plans which are processed chronologically. In this chronological processing, the problem of vehicle scheduling can be defined as a dynamic vehicle routing problem (Maciejewski et al., 2017). At the start of the analyzed time period, not all trip requests are known. Instead, requests come in over time.

Macroscopic models, in contrast, aim to replicate an average day. This is achieved by using average trip rates for trip production and by assigning probabilities to each choice of a choice set. The results are non-integer values that represent the demand situation of a recurrent average day. As on-demand services are designed to adapt to a specific demand situation varying from day to day, it is not helpful for planning purposes to replicate the trip requests of one specific day. Instead, an average demand situation should be used for the service design. Furthermore, it seems reasonable to assume that information on all trip requests is available at the beginning of the vehicle scheduling step. This makes the problem more similar to traditional vehicle scheduling in timetable-based public transport.

Due to the respective ways they calculate travel demand, microscopic and macroscopic approaches tend to answer different research questions: Microscopic approaches rather answer the question of how well a certain vehicle fleet can satisfy a given demand (Maciejewski et al., 2017; Marczuk et al., 2015; PTV Group, 2020). Macroscopic approaches rather take the reverse approach and determine the required fleet size to fully satisfy a given demand. Nevertheless, it is important to note that although the model types are prone to the mentioned uses, it is also possible to use them in the opposite way. Boesch et al. (2016), Wang et al. (2018) and Fagnant and Kockelman (2018), for example, confirm this by using microscopic approaches to calculate the number of vehicles needed.

5.2.2 Vehicle scheduling

The literature on vehicle scheduling provides many approaches to find schedules with a minimum number of vehicles. Standard models and solution approaches for vehicle scheduling in public transport are summarized in Bunte and Kliewer (2009). Recent vehicle scheduling approaches usually incorporate problem-specific aspects such as variable timetables (Desfontaines and Desaulniers, 2018; Lan et al., 2019) or a limited range of electric vehicles and recharging strategies (Rogge et al., 2018; M. Wen et al., 2016). To be able to find good schedules for realistic instances, elaborate solution methods are proposed. For example, Desfontaines and Desaulniers (2018) rely on column generation, and Lan et al. (2019) combine Benders decomposition with a branch-and-price approach. With these methods, instances with up to 2100 vehicle trips could be solved within less than one hour to optimality or close to optimality. The considered instances in M. Wen et al. (2016) contain up to 500 vehicle trips and are solved with an adaptive large neighborhood search within 20 minutes. Rogge et al. (2018) develop a genetic algorithm and provide results for instances with up to 200 vehicle trips.

A. Lam et al. (2016) formulate a vehicle scheduling problem specifically for a ridesharing setting with automated vehicles. Their model includes admission control and pooling of passengers, as well as a maximum route length due to a restrictive battery capacity. They use different instances constructed from taxi data from Boston with 100 trips and propose a genetic algorithm to find vehicle schedules. Lin et al. (2012) use a simulated annealing approach to find vehicle movements that are both cost-efficient and convenient for passengers in a ridesharing setting for taxis. They find that both the mileage as well as the number of vehicles can be reduced significantly by ridesharing, however, only results of a single and relatively small instance with 29 trips were discussed.

Most vehicle scheduling contributions consider an operational setting and aim at providing an optimal solution for a certain demand situation. In contrast to these approaches, our algorithm is designed for usage in an extended four-step algorithm as depicted in Figure 5.1b. We intend to provide good estimates for the required fleet size and the impact on the traffic volume within very short computation times. This is suitable for long-term strategic transport planning. Furthermore, most solution methods exploit that each planned trip has to be covered exactly once. Since this does not necessarily hold for macroscopic demand models, a generalized approach is required. Similar to early approaches as presented in Bodin (1983), we model the vehicle scheduling problem as a flow problem. This design choice is motivated by the huge demand data of realistic instances considered in this chapter that include up to 100 million vehicle trips.

We formulate the problem as minimum-cost circulation with lower bounds. A polynomial solution algorithm based on a gradual convergence by iteratively finding better routes for vehicles is described in Schrijver (2003): First, an initial circulation is found that is not necessarily optimal. Then, the solution is iteratively improved by identifying a directed circuit with negative cost in the residual graph. The circulation is adjusted correspondingly along this circuit. To identify a directed circuit, a flow problem has to be solved. This yields a time complexity of $\mathcal{O}(|Z|^8|T|^5 \log(|Z||T|))$ for this approach (Schrijver, 2003), where Z and T correspond to a discretization of space and time, respectively.

In a project, many vehicle schedules have to be computed since often many scenarios are considered and a feedback loop between demand estimation and supply design is common. Hence, we propose a simple heuristic approach for macroscopic on-demand problems to realize short solution times for huge instances. Our approach presented in Section 5.4 has a time complexity of $\mathcal{O}(|Z|^2|T|^2)$ and meets the requirements of an application in travel demand models.

5.3 Problem definition

In this section, the problem of finding a vehicle scheduling with the minimum fleet size is formalized. To this end, the required input is described and the underlying network for the presented solution algorithm is introduced.

5.3.1 Input and notation

Passenger demand is given as demanded vehicle trips, aggregated in time and space. For ridesharing applications, passenger trips are pooled to vehicle trips in a preceding step.

The analysis period is split into time intervals of equal length. Time intervals are indexed with t and the set of time intervals is denoted by $T = \{1, 2, \ldots, |T|\}$. The examination area is divided into traffic zones, the set of traffic zones is Z. A traffic zone is denoted by z, or, when considering a traffic zone as origin or destination zone, by z_o and z_d, respectively.

The number of demanded vehicle trips from an origin zone z_o to a destination zone z_d starting in time interval t is denoted by $d_{z_o z_d t}$. In this setting, these requested vehicle trips are composed of pooled passenger trips and called *service trips*. It is assumed that the pooling of passenger trips is done in a previous step, which is not discussed here. Further, distances $\delta_{z_o z_d}$ between traffic zones are given as multiples of time intervals. They result from the travel time j between traffic zones and the duration of a time interval l, $\delta_{z_o z_d} = \left\lceil \frac{j_{z_o z_d}}{l} \right\rceil$. For the presentation of the algorithm in this chapter, two assumptions are made. First, the travel time between zones is independent of the time of day. To consider the asymmetric nature of congestion, the distance matrix can be extended by a third dimension representing the departure time interval. Second, all trips within one zone require a travel time of at most one time interval, that is $\delta_{zz} = 1\ \forall z \in Z$. If this assumption does not apply, the model can be extended to distinguish between waiting and traveling within a zone.

The input of an instance $\mathcal{I} = (\delta, d)$ consists of a distance matrix δ and a demand situation d. By concatenating the demanded vehicle trips to vehicle schedules, the presented algorithm determines the number of required vehicles as well as empty trips for vehicle relocation per time interval as a result. The aim is to serve the entire demand with as few vehicles as possible.

5.3.2 Underlying network

For a simpler representation of the algorithm, a time-space network $G = (V, E)$ is introduced. Nodes can be interpreted as traffic zones at the beginning of time intervals and arcs as potential time-bound vehicle trips between zones. In the network, we depict traffic zones on the vertical and time intervals on the horizontal. The presented vehicle scheduling algorithm is designed to find a feasible vehicle flow in this network with as few vehicles as possible so that demand is met on all edges. An example network with three traffic zones and four time intervals is shown in Figure 5.2a.

Formally, we introduce the set of nodes $V = V_Z \cup V_{Z,\overline{T}}$ with

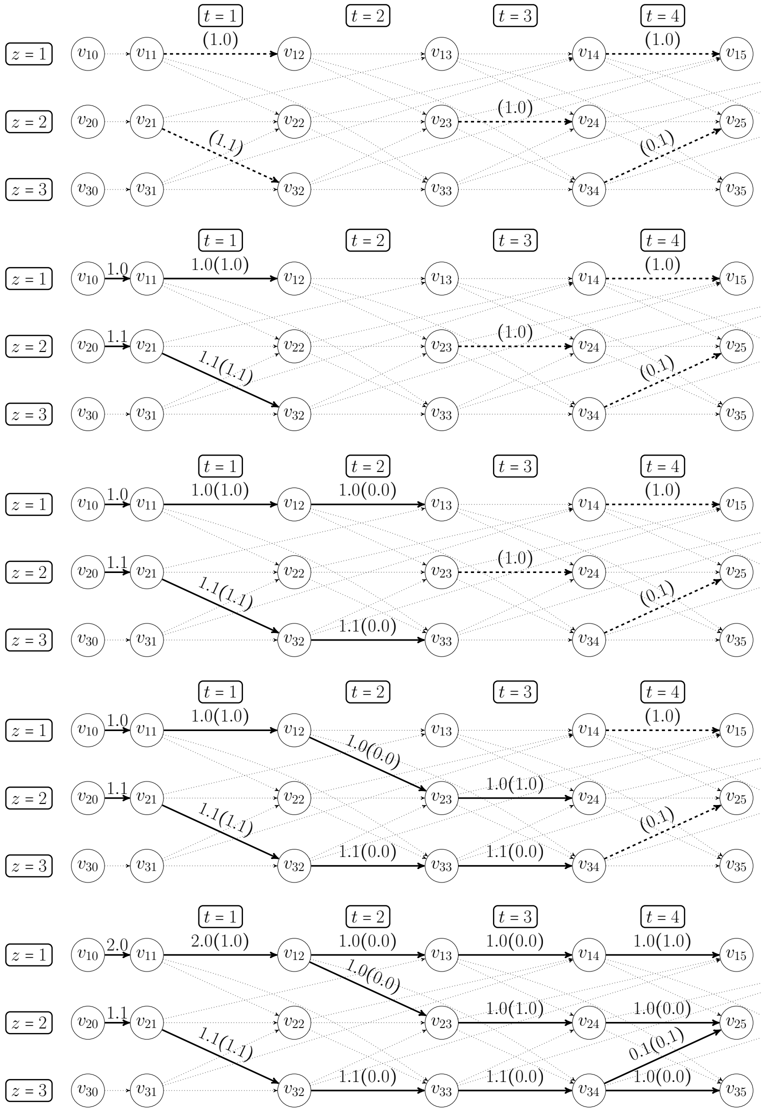
$$V_Z = \{v_{z0} \colon\ z \in Z\} \text{ and } V_{Z,\overline{T}} = \{v_{zt} \colon\ z \in Z,\ t \in \overline{T}\},$$

where $\overline{T} = T \cup \{|T| + 1, \ldots, |T| + \max_{z_o, z_d \in Z} \delta_{z_o z_d}\}$ is an extended set of time intervals. For each traffic zone $z \in Z$, there is a node v_{z0} in the network G at the beginning of the analysis period. Moreover, there is a node v_{zt} representing each traffic zone $z \in Z$ at the beginning of each time interval $t \in \overline{T}$. The nodes in V are connected by directed edges in $E = E_Z \cup E_{Z,T}$, where

$$E_Z = \{(v_{z0}, v_{z1}) \colon\ z \in Z\} \text{ and } E_{Z,T} = \{(v_{z_o t}, v_{z_d (t + \delta_{z_o z_d})}) \colon\ z_o, z_d \in Z,\ t \in T\}.$$

From each node v_{z0} there is a directed edge to the node v_{z1}, which represents the traffic zone z at the beginning of the first time interval. There are also $|T|$ edges that connect each pair of origin zone z_o and destination zone z_d. These edges start in the time intervals $t \in T$ and end in $t + \delta_{z_o z_d} \in \overline{T}$, where $\delta_{z_o z_d}$ corresponds to the distance between the two traffic zones.

The demand $d_{z_o z_d t}$ can be interpreted as a lower bound on the edges $e \in E_{Z,T}$, defining a minimum flow on these edges. The distances $\delta_{z_o z_d}$ between the traffic zones z_o and z_d are modeled by the horizontal length of the edges. A trip from the first to the second traffic zone in the example of Figure 5.2a can be covered in one



(a) Input to algorithm: Example network with indicated demand.

(b) First time interval processed: 2.1 vehicles are inserted to meet all demand starting from first time interval.

(c) Second time interval processed: No demand, hence vehicles are idle in their traffic zone.

(d) Third time interval processed: 1.0 vehicles are relocated to meet demand in different zone.

(e) Last time interval processed: 1.0 additional vehicles are inserted to meet the demand.

Figure 5.2: Step-wise construction of a vehicle schedule on a network with 3 traffic zones and 4 time intervals. Figure 5.2a shows the input situation without vehicle flow. The schedule construction in Figures 5.2b to 5.2e is described in detail in Section 5.4. In each figure, the rectangular labels display the traffic zones on the vertical and the time intervals on the horizontal. The nodes in V are represented with round node shapes, reading the node label v_{zt}. For better presentation, nodes v_{zt} are omitted for $t > 5$. The directed edges in E are represented with arcs in the network, distinguished in three cases. Grey dotted lines show potential vehicle trips without demand or vehicle flow, dashed lines indicate demand on edges, and solid lines depict edges with vehicle flow. The numbers written at the arcs read the flow values f, and, in brackets, the demand values d.

time interval, the return trip needs two time intervals. Asymmetries can be caused by one-way streets or differing traffic volumes in the network.

5.3.3 Vehicle scheduling

The flow variables $f_{z_o z_d t} \in \mathbb{R}_+$ are introduced to represent a vehicle flow on the edges $e \in E_{Z,T}$. The value $f_{z_o z_d t}$ can be interpreted as the amount of vehicles driving from traffic zone z_o to z_d, starting in time interval t. To ensure that the total demand is served, the flow on each edge must be at least as large as the demand,

$$f_{z_o z_d t} \geq d_{z_o z_d t} \quad \forall z_o, z_d \in Z, \ \forall t \in T. \tag{5.1}$$

For the flow to be feasible, it must also be ensured that the total number of arriving and departing vehicles in each node v_{zt} is equal,

$$\sum_{\substack{z_o \in Z: \\ t-\delta_{z_o z} \geq 1}} f_{z_o z (t-\delta_{z_o z})} = \sum_{z_d \in Z} f_{z z_d t} \quad \forall z \in Z, \ \forall t \in T \setminus \{1\}. \tag{5.2}$$

This ensures that the flow is preserved in every node v_{zt}, that means that no vehicles "appear" or "disappear" in traffic zone z at time t. A feasible flow f in the network G is called a vehicle schedule. Next, the variables $x_z \in \mathbb{R}_+$ are introduced to model the vehicle flow on the edges E_Z. These correspond to the total number of vehicles leaving the traffic zone $z \in Z$ in the first time interval, defined as

$$x_z = \sum_{z_d \in Z} f_{z z_d 1} \quad \forall z \in Z.$$

x_z can be interpreted as the number of vehicles that must be available in the traffic zone z at the beginning of the analysis period. The aim is to serve the demand with as few vehicles as possible, which corresponds to minimizing the sum of vehicles leaving traffic zones in the first time interval $\sum_{z \in Z} x_z$. Equations (5.1) and (5.2) ensure that any demand is met and the vehicle schedule is feasible.

5.4 Algorithm

5.4.1 Description

The basic structure of the algorithm is simple: The nodes v_{zt} in the network are processed chronologically and the vehicle flow is expanded step by step on the outgoing edges. Figures 5.2b to 5.2e illustrate the construction of a vehicle schedule in the example network in Figure 5.2a. In each step, it is ensured that the demand is met and that the vehicle flow is feasible at all processed nodes. Thus, the design of the algorithm ensures that equations (5.1) and (5.2) are fulfilled step by step. While the algorithm constructs the vehicle flow chronologically, that is, from left to right in the network in Figure 5.2, the flow in the previous time intervals can be amended. To perform this amendment efficiently, we maintain node labels a storing the current number of vehicles at each node during flow construction.

For a simpler representation of the vehicle scheduling algorithm, we split it into three nested parts. The basic structure is given in Algorithm 5.1. This part specifies that the nodes in the network are considered in chronological order and that at each considered node all demand on outgoing edges is met and the flow conservation holds. The flow conservation, which ensures that there is the same number of incoming and outgoing vehicles at each node, is specified in Algorithm 5.2. There, three cases are considered. First, the number of incoming vehicles is sufficient for the number of demanded vehicles on outgoing edges. Second, vehicles in other zones are available and can be relocated to the current traffic zone. Third, additional vehicles need to be added to the vehicle flow under construction. While the first and third cases are easy to handle, the relocation of vehicles in the second case requires an amendment of the flow in previous time intervals. This amendment is described in Algorithm 5.3.

The nested structure means that Algorithm 5.1 calls Algorithm 5.2 to ensure the flow conservation, which in turn calls Algorithm 5.3 for vehicle relocation, if necessary. In the following, the pseudocode of the three algorithms is described and exemplified with the flow construction in Figure 5.2.

Basic structure

In Algorithm 5.1 the basic structure of the vehicle scheduling algorithm is given as pseudocode. The loops in lines 4 and 5 scroll through the nodes v_{zt} in chronological order. Starting from the considered node, the demand is served on each outgoing edge, see line 6. This step ensures that there is sufficient vehicle flow on the demanded

Algorithm 5.1: *VehScheduling($\mathcal{I}$)*

1 **Input:** Instance $\mathcal{I} = (\delta, d)$ with distance matrix and demand
2 **Output:** Number of required vehicles x and vehicle flow f
3 **Initialisation:** Required vehicles per station $x_z \leftarrow 0 \quad \forall z \in Z$, available vehicles per traffic zone and time interval $a_{zt} \leftarrow 0 \quad \forall z \in Z, t \in \overline{T}$;
`# Process all nodes` v_{zt} `in the network chronologically;`
4 **for** $t \in T$ **do**
5   **for** $z \in Z$ **do**
        `# Fix minimal flow on all outgoing edges, satisfies equation` (5.1);
6         $f_{zz_dt} \leftarrow d_{zz_dt} \quad \forall z_d \in Z$;
        `# Update label: Mark vehicles as available in destination zone;`
7         $a_{z_d(t+\delta_{zz_d})} \leftarrow a_{z_d(t+\delta_{zz_d})} + d_{zz_dt} \quad \forall z_d \in Z$;
        `# ensure feasible flow, satisfies equation` (5.2);
8   *FlowConservation*(z, t, a, x, f, $\mathcal{I}$);
9 **return** (x, f);

edges in the network, see for example Figure 5.2b where a flow of 1.0 and 1.1 vehicles is set between nodes v_{11} and v_{12}, and between nodes v_{21} and v_{32}, respectively, to meet the demand. Then, in line 7 labels are updated at the nodes indicating how many vehicles are available in the traffic zones at the beginning of the time intervals. After the first time interval is processed in Figure 5.2b, there are 1.0 and 1.1 vehicles available at nodes v_{12} and v_{32}, respectively. Finally, calling the function *FlowConservation()* in line 8 ensures that the number of arriving and departing vehicles at the considered node v_{zt} are equal and, thus, that the vehicle flow is feasible.

Flow conservation

Algorithm 5.2 is called at every node v_{zt} to ensure flow conservation. This is necessary since in Algorithm 5.1 only the vehicle flow on outgoing edges was set in order to meet demand. In Algorithm 5.2, sufficient incoming flow is ensured to match the outgoing flow, or the outgoing flow is increased if the incoming flow is predominant. To match the number of incoming and outgoing vehicles in that node, vehicles from three different sources are considered in the following priority.

1. The first step is to try to satisfy as much demand as possible with vehicles available at the current node v_{zt}. Vehicles are considered available at a node v_{zt} if they are idle in the traffic zone z at the beginning of the time interval t. In the algorithm, the number of available vehicles at each node is stored in the

Algorithm 5.2: *FlowConservation(z, t, a, x, f, $\mathcal{I}$)*

1 **Input:** Traffic zone z, time interval t, available vehicles a, number of required vehicles x, vehicle flow f and instance $\mathcal{I} = (\delta, d)$

`# 1. If sufficient vehicles are available, these are used;`

2 **if** $a_{zt} \geq \sum_{z_d \in Z} d_{zz_dt}$ **then**

3 $\quad a_{zt} \leftarrow a_{zt} - \sum_{z_d \in Z} d_{zz_dt}$;

$\quad$`# Other available vehicles are waiting in the zone;`

4 $\quad$**if** $a_{zt} > 0$ **then**

5 $\quad\quad f_{zzt} \leftarrow f_{zzt} + a_{zt}$;

$\quad\quad$`# Update label: Mark vehicles in destination zone as available;`

6 $\quad\quad a_{z(t+1)} \leftarrow a_{z(t+1)} + a_{zt}$;

`# Otherwise additional vehicles are needed;`

7 **else**

$\quad$`# Define` n `as number of additional vehicles required;`

8 $\quad n \leftarrow \sum_{z_d \in Z} d_{zz_dt} - a_{zt}$;

$\quad$`# Update Label: All available vehicles are used;`

9 $\quad a_{zt} \leftarrow 0$;

$\quad$`# 2. Relocate as many available vehicles as possible from other zones;`

10 $\quad n \leftarrow$ *VehRelocation(z, t, a, n, f, $\mathcal{I}$)*;

$\quad$`# 3. Insert vehicles if still needed after relocation;`

11 $\quad$**if** $n > 0$ **then**

$\quad\quad$`# Increase number of required vehicles per zone accordingly;`

12 $\quad\quad x_z \leftarrow x_z + n$;

$\quad\quad$`# Increase vehicle flow, vehicles wait in the zone until demanded;`

13 $\quad\quad f_{zzt'} \leftarrow f_{zzt'} + n \quad \forall t' < t$;

14 **return**;

label a_{zt}. If more vehicles are available than needed, they wait in the traffic zone and the labels at the nodes are adjusted, see lines 3, 5, and 6 in Algorithm 5.2. Both usage of available vehicles and waiting in the traffic zone can be observed at the node v_{34} in Figure 5.2e, for example. There, 0.1 vehicles are sent to node v_{25} to meet demand, and the remaining 1.0 available vehicles wait in the third traffic zone.

2. If there are not enough vehicles available, the algorithm tries to relocate vehicles from other traffic zones z_o to traffic zone z. For a permissible relocation, the vehicles must be available already δ_{z_oz} time intervals before the considered time interval t. Only in that case, they can be relocated in time to meet demand at the beginning of time interval t in traffic zone z. The relocation is designed in

such a way that demand will continue to be met on all previously considered edges and that flow will continue to be preserved in all previously considered nodes.

By relocation, it is possible to find good vehicle schedules requiring few vehicles only at the expense of empty vehicle kilometers. In Section 5.6, the number of required vehicles and the length of empty trips are compared in scenarios with and without vehicle relocation. The exact procedure of vehicle relocation is described in Algorithm 5.3, which is called in line 10 of Algorithm 5.2 if there are not enough vehicles available.

3. If after the relocation of vehicles from other traffic zones the total demand on outgoing edges of the considered node v_{zt} is not met, further vehicles are necessary for a feasible vehicle flow. These vehicles are inserted in the traffic zone z by increasing the variable x_z and are idle until time interval t, see lines 12 and 13 in Algorithm 5.2. In the example network, this happens at the beginning of the analysis period, see Figure 5.2b, and when processing the last time interval, see Figure 5.2e. In the former, 1.0 and 1.1 vehicles are inserted in the first and the second traffic zone, respectively. In the latter, another 1.0 vehicles are inserted in the first traffic zone. There, it is possible to see how all flow variables within this zone are increased, indicating that the vehicles are idle until demanded in the fourth time interval.

Vehicle relocation

Algorithm 5.3 describes how the relocation of vehicles is performed and the flow in previous time intervals is amended. First, it is calculated how many vehicles can be relocated, see lines 5 to 7. Then, the previously set vehicle flow is undone and the corresponding labels are updated, see lines 8 to 12. Finally, the empty vehicle trip for relocation is added to the vehicle flow, see line 13. Figures 5.2c and 5.2d show the relocation of vehicles from the first to the second traffic zone. Initially, 1.0 vehicles wait in the first traffic zone during the second time interval. When processing the third time interval, this flow is undone and the vehicles are relocated from the first to the second traffic zone during the second time interval to meet demand. While the basic structure in Algorithm 5.1 works chronologically, the relocation of vehicles in Algorithm 5.3 can be seen as a backward correction.

Algorithm 5.3: *VehRelocation*(z, t, a, n, f, $\mathcal{I}$)

```
Input: Traffic zone z, time interval t, available vehicles a, needed vehicles n,
       vehicle flow f and instance I = (δ, d)
Output: Number of vehicles that are still needed n
# Check all other zones for available vehicles;
for z_o ∈ Z do
    # Relocation can only start within the analysis period;
    if t − δ_{z_o z} ≥ 1 then
        # Define a_{z_o} as maximum number of relocatable vehicles from zone z_o;
        a_{z_o} ← min_{t': t − δ_{z_o z} ≤ t' < t} a_{z_o t'};
        # Relocate at most as many vehicles as needed;
        a_{z_o} ← min{n, a_{z_o}};
        # Update number of needed vehicles;
        n ← n − a_{z_o};
        # Reset previously set vehicle flow;
        f_{z_o z_o t'} ← f_{z_o z_o t'} − a_{z_o}   ∀(t − δ_{z_o z}) ≤ t' < t;
        # Update label:  Reset number of available vehicles;
        a_{z_o t'} ← a_{z_o t'} − a_{z_o}   ∀(t − δ_{z_o z}) ≤ t' ≤ t;
        # If node v_{z_o t} has been edited in Algorithm 5.1, reset flow and label;
        if z_o < z then
            f_{z_o z_o t} ← f_{z_o z_o t} − a_{z_o};
            a_{z_o (t+1)} ← a_{z_o (t+1)} − a_{z_o};
        # Relocate vehicles;
        f_{z_o z (t − δ_{z_o z})} ← f_{z_o z (t − δ_{z_o z})} + a_{z_o};
return n;
```

Summary

The presented vehicle scheduling algorithm is designed such that the vehicle flow is feasible at each node and the demand is served on each edge. The relocation of vehicles preserves these properties at nodes and edges that have already been processed. Therefore, the solution of the algorithm is a feasible vehicle flow f, which requires as few vehicles x as possible. An implementation of the presented algorithm is available in Hartleb et al. (2020).

5.4.2 Solution quality

The algorithm is deterministic and provides the same solution in every call. However, it is a heuristic procedure that does not necessarily find an optimal solution. This can be seen in the example in Figure 5.2. At node v_{23} not enough vehicles are available to serve the outgoing demand. Therefore, attempts are made to relocate vehicles from other traffic zones, see Figure 5.2d. In this case, there are enough vehicles available in the first traffic zone, that are relocated within the second time interval. As a result of this relocation, no vehicles are available at node v_{14} in the fourth time interval. Additional vehicles must be inserted, increasing the total number of required vehicles, see Figure 5.2e.

In the solution of the algorithm, a total of 3.1 vehicles is required to meet the entire demand. In an optimal solution, however, only 2.1 vehicles are needed, for example, by relocating vehicles from the third instead of the first traffic zone to node v_{23}. This shows that the solution quality depends, among other things, on the order of traffic zones from which vehicles are relocated. In the example described, the algorithm finds a solution that is almost 50 percent worse than an optimal one. However, preliminary tests have shown that the solution quality on both randomly generated and real networks is significantly higher than in this contrived example. In most practical applications the deviation from the optimal number of required vehicles was smaller than the deviation due to other modeling errors.

5.4.3 Complexity

In a case study, many vehicle schedules need to be computed because usually several transport scenarios are examined and a feedback loop between demand estimation and supply design is applied within each scenario. Therefore, a heuristic approach with short computation times is most practical. The vehicle scheduling algorithm presented in this chapter performs $|Z||T|(|Z| + |Z| + |Z| + |Z| + |Z| + |Z|(\overline{\delta} + \overline{\delta} + \overline{\delta}) + |T|)$ operations, where $\overline{\delta} \coloneqq \max_{z_o,z_d \in Z} \delta_{z_o z_d}$ is the maximum distance between two time intervals. Hence, the time complexity is in $\mathcal{O}(|Z|^2|T|\overline{\delta} + |Z||T|^2)$. Since $\overline{\delta}$ is bounded by the number of time intervals $|T|$, the presented algorithm is strongly polynomial with complexity $\mathcal{O}(|Z|^2|T|^2)$. This low complexity is achieved by locally improving the solution during its construction. The network has to be traversed only once.

Existing exact approaches are based on an iterative improvement of an initial circulation by identifying a directed circuit with negative cost in the residual graph. The circulation is adjusted correspondingly along this circuit. In these approaches, the graph has to be traversed multiple times, which leads to a time complexity of $\mathcal{O}(|Z|^8|T|^5 \log(|Z||T|))$ (Schrijver, 2003).

5.5 Extensions

In this section, multiple extensions are discussed that enhance the basic algorithm. They aim at improving the solution quality or the computation time of the algorithm. All extensions are implemented and in the following paragraphs is stated how they are used in the experiments.

Consideration of the neighborhood of traffic zones: The relocation of vehicles from other traffic zones results in empty trips, which should be kept as short as possible for cost reasons. This can be taken into account by adjusting the order of the neighboring traffic zones in row 3 in Algorithm 5.3. For all experiments, the set of traffic zones Z is replaced by a sorted neighborhood $N(z)$ of the considered traffic zone z. This means that vehicles are first requested from the closest traffic zones. This favors short empty vehicle trips and implicitly takes operating costs into account. The total number of vehicles required can be influenced positively or negatively.

Limitation of the relocation distance: Further, it is possible to not only sort the set of all neighboring traffic zones but also to limit it. This can, for example, prevent particularly long empty vehicle trips. This restriction can result in more vehicles being needed to meet the overall demand. In return, the length of empty trips will decrease. The trade-off between the number of vehicles and empty trips is discussed in Section 5.6.

Scanning of future time intervals: Vehicles can be relocated if they have been available in another traffic zone for a sufficient number of time intervals. Still, it may be better to not relocate the vehicles, for example, if they are needed shortly thereafter in their current traffic zone. With the scanning of future time intervals, it is possible to prevent such unnecessary vehicle relocation. However, both future incoming and outgoing edges at the nodes should be taken into account. Since this entails a high calculation effort for each additional time interval, in the current implementation only one future time interval is scanned. The total number of required vehicles can either increase or decrease, but operating costs are reduced.

Termination criterion: In the current implementation, the vehicle relocation in Algorithm 5.3 is terminated as soon as enough vehicles have been found. This significantly reduces the computation time of the algorithm.

5.6 Applications

The applicability of the algorithm is illustrated in two case studies. Both case studies show that it is possible to consider on-demand services in macroscopic demand models by the use of the vehicle scheduling algorithm. In the first case study, the number of vehicles required for a regional carsharing system in a region in southern Germany is determined. It is assumed that carsharing is used for all private car journeys. This is not a realistic assumption, but it demonstrates the capability of the algorithm in large networks with a very large number of demanded vehicle trips. In the second case study, the required fleet size of an electric scooter rental system for operation on a university campus is computed. This case study emphasizes the importance of appropriate time interval durations for models with small spatial dimensions as well as the influence of demand symmetry on the number of vehicles needed.

5.6.1 Regional carsharing

The Stuttgart Region covers an area of 80×80 kilometers with 2.7 million inhabitants living in urban and rural surroundings. The regional travel demand model is used to determine the fleet size of a regional carsharing system. The model is a macroscopic travel demand model covering the four model steps of trip generation, destination choice, mode choice, and route choice in person transport. It calculates the demand on workdays for the modes car driver, car passenger, public transport, bicycle and walking with a tour-based model. The model includes $|Z| = 1013$ traffic zones in the examination area of the Stuttgart region. The baseline scenario assumes a situation without carsharing, which describes more or less the current state, where sharing is a rare event

From this baseline scenario, three scenarios are derived for comparison, all assume that private car journeys will be carried out with carsharing vehicles. The scenarios *S02* and *S03* require automated vehicles allowing driverless relocation of the vehicles. The following scenarios are distinguished:

S00 Baseline scenario without carsharing, private cars only.

S01 Carsharing rides replace car rides,
Carsharing without relocation.

S02 Carsharing rides replace car rides,
Carsharing with relocation aiming at a minimum number of vehicles,
Duration of empty trips is not limited.

S03 Carsharing rides replace car rides,
Carsharing with relocation aiming at a minimum number of vehicles,
Duration of empty trips must not exceed 15 min.

The same demand situation is assumed in all scenarios. It considers 3.6 million private car trips that have their origin and destination in the region. Two input variables are passed to the scheduling algorithm:

1. Day-time dependent demand d: By using trip-purpose specific temporal distributions, the demand for car trips is divided into 96 time intervals of 15 min. This demand defines the service trips in the network.

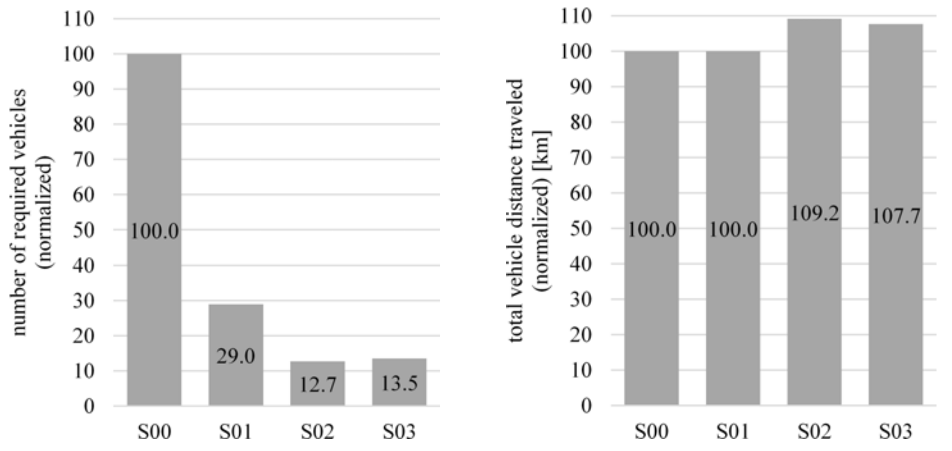
2. Distance matrix δ: This matrix describes the travel time between the traffic zones as multiples of time intervals. The values of the matrix are based on the car travel times in the congested road network. For service trips and empty trips the same travel times are assumed.

The algorithm calculates a vehicle schedule and outputs the number of vehicles required. The vehicle schedule contains all necessary information about empty trips which are needed for relocating vehicles. An assignment of the service trips and empty trips to the road network provides the vehicle distance traveled.

From the results of the German national travel survey 2017 (infas et al., 2017) it can be deduced that only about two-thirds of all private cars in Germany are moved on an average working day. On average, these moving vehicles perform 3.2 trips per working day. This leads to about 1.1 million vehicles (without not moving vehicles) in the baseline scenario *S00*.

This number of vehicles is normalized in Figure 5.3a to the value 100 and serves as a reference for the calculated fleet sizes of scenarios *S01* to *S03*. While demand remains constant, the number of vehicles in scenario *S01* can be reduced to less than one-third of the private cars required in the baseline scenario although no vehicle relocation is allowed. When including vehicle relocation in scenarios *S02* and *S03*, the number of vehicles drops to about one-eighth of the vehicles required in the baseline scenario. The limitation of the empty trip duration to 15 min in *S03* implies a comparatively small increase in the number of required vehicles.

In the scenarios *S00* and *S01* the vehicle kilometers traveled are identical since there are no empty trips. In the scenarios *S02* and *S03*, however, the vehicle kilometers traveled increase due to empty vehicle trips by 9.2 and 7.7 percent, respectively (see



(a) Change in number of required vehicles (b) Change in vehicle distance traveled

Figure 5.3: Number of required vehicles and total vehicle distance traveled per scenario. Normalization: *S00* = 100.

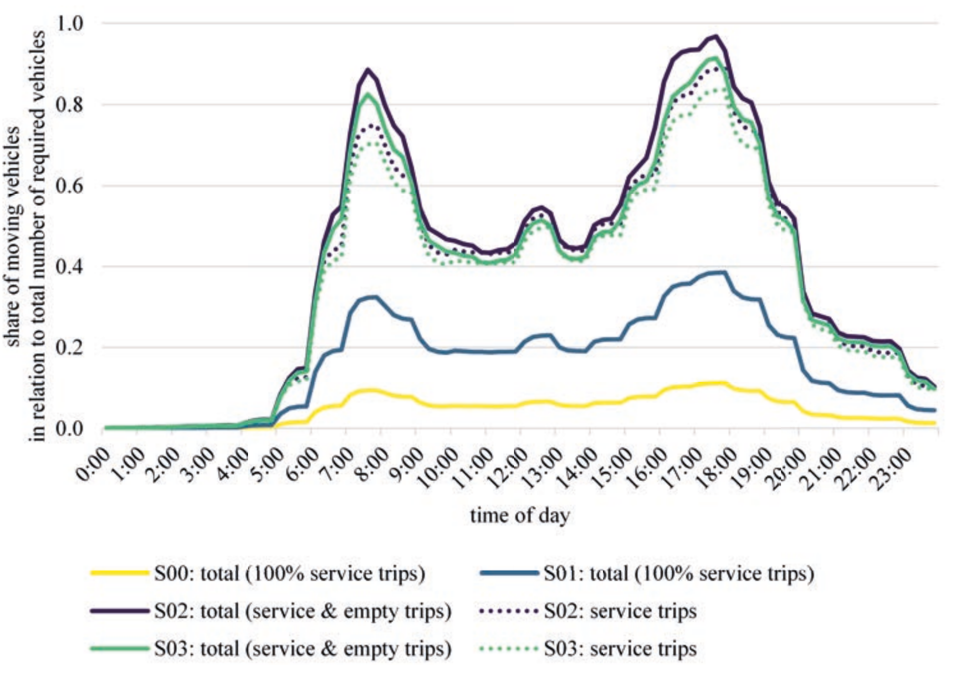


Figure 5.4: Share of moving vehicles per time interval in relation to the total number of required vehicles for each scenario.

Figure 5.3a). In scenario *S02* an average carsharing vehicle travels about 230 km per day. By limiting the empty trip duration to 15 min in *S03*, this daily distance goes down to 215 km. This corresponds to an average reduction of the total vehicle kilometers traveled by about 55 km per additional vehicle.

Figure 5.4 shows the time series of the moving vehicles in relation to the total number of required vehicles per scenario. Carsharing increases the occupancy rate of the vehicle fleet considerably, especially during peak hours. In both scenarios *S02* and *S03*, the maximum share of empty vehicle trips per time interval is 20 percent. Similar to the service trips, the empty trips take place mainly during peak hours.

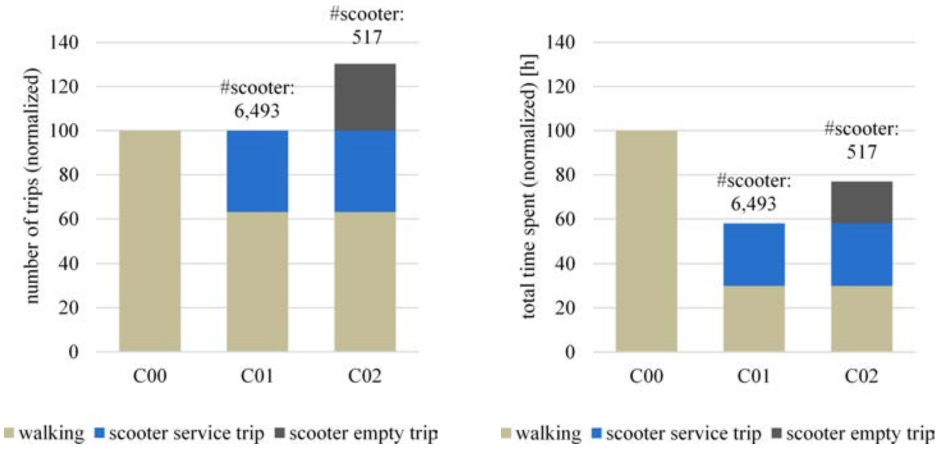
5.6.2 Sharing of electric scooters on a university campus

The University of Stuttgart plans to introduce a campus-wide shared scooter service with autonomous electric scooters. Autonomous electric scooters still require a human to drive, but they are able to carry out driverless empty trips to relocate or to drive to a charging station (Wenzelburger and Allgöwer, 2020). To estimate the required size of such an electric scooter fleet, the demand for pedestrian traffic between bus stops, parking lots, and buildings is determined for the campus of the University of Stuttgart. The basis for this estimation is a travel survey recording the choices of students and employees regarding their mode of transport (car, public transport), the exit stop or the destination car park, and the time of day for their trips to and from the campus. In a baseline scenario *C00*, all movements between car parks or stops and university buildings are walking trips. In scenarios *C01* and *C02* it is assumed that walking trips longer than 400 m are no longer covered by foot, but with electric scooters. Since the demand at a university shows considerable peaks at the beginning and end of lectures, many scooters are required in the respective load direction. An automated relocation of autonomous scooters could reduce the number of scooters. This results in the following three scenarios:

C00 Baseline scenario without scooter, only walking.

C01 Scooter rides replace walking for trip lengths from 400 m,
Scooters are not relocated.

C02 Scooter rides replace walking for trip lengths from 400 m,
Scooters are relocated aiming at a minimum number of vehicles,
Duration of empty trips is not limited.



(a) Change in number of trips

(b) Change in total time spent

Figure 5.5: Number of trips and total time spent for all scenarios. Normalization: *C00* = 100.

For walking trips, a speed of 4 km/h, and for scooter rides, a speed of 10 km/h is assumed. With this speed, a distance matrix is created for the 150 locations on campus. Since the average time of one scooter trip on campus is only 4 min, trip times would be greatly overestimated when using time intervals of 15 min. Therefore, the demand is divided into 288 time intervals of 5 min each. Using these input variables, the algorithm can be applied as in the carsharing scenarios to find scooter schedules for the three campus scenarios.

Figure 5.5 shows the number of person and vehicle trips made as well as the total time spent in each of the three scenarios. On an average workday, there are almost 40,000 walking trips to and from the buildings. In scenario *C00*, the trips are completely covered by foot. In the scenarios *C01* and *C02* about a third of the trips are performed with scooters. This reduces the total time spent traveling by approximately 40 percent. However, in scenario *C01* without relocation nearly 6500 scooters are required. In scenario *C02* the number of scooters can be reduced to about 500 scooters by relocation. The vehicle numbers correspond to roughly 2.2 vehicle trips per scooter and day in *C01*, whereas in scenario *C02* a scooter is used for about 51.0 vehicle trips, of which 23.0 are empty trips.

A test shows the importance of the selected time interval length: If the demand and the distance matrix are divided into 15 min time intervals instead of 5 min time

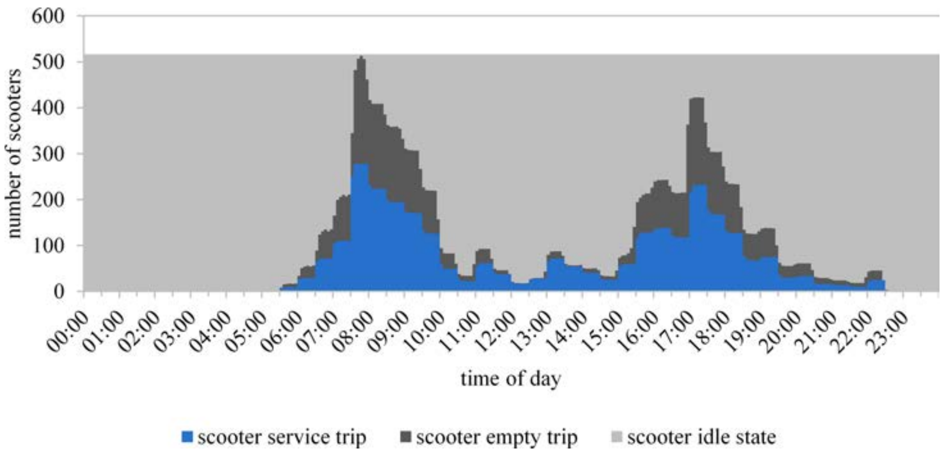


Figure 5.6: Number of scooters in scenario *C02* per time interval, subdivided into service trips, empty trips and idle.

intervals, the vehicle scheduling algorithm only finds a solution with 1400 scooters for scenario *C02* instead of 500 scooters due to the overestimated travel times.

Figure 5.5 and Figure 5.6, which distinguishes the number of scooters in scenario *C02* by activity (service trip, empty trip and idle) per time interval, show that the share of empty scooter trips on campus is considerably higher compared to the regional carsharing scenarios discussed in Section 5.6.1. This can be explained by the demand structure on a university campus with strongly pronounced load directions. The more asymmetrical the demand, the more vehicles or empty trips are required to serve the same number of trip requests.

5.7 Conclusion and Outlook

In this chapter, we presented an efficient heuristic for the vehicle scheduling problem (available at Hartleb et al. (2020)). The aim of the heuristic is to find a vehicle schedule serving a given demand with as few vehicles as possible.

In contrast to most existing vehicle scheduling approaches, the presented algorithm is suitable for integration into existing macroscopic travel demand models to estimate the required vehicle fleet size and the corresponding traffic volumes of on-demand services. The presented algorithm provides two advantages compared to standard applications of vehicle scheduling.

The first advantage is the problem size that it can handle. Due to the simple procedure of the presented algorithm, vehicle schedules for large instance sizes can be found in short computation times. This allows the analysis of large-scale on-demand services with a high number of trip requests. The second advantage is the usability for integer as well as non-integer demand values. Macroscopic models deal with non-integer demand structures, which can be handled by the presented algorithm. With a high temporal resolution, it can also be used in microscopic travel demand models.

In two case studies, we illustrated the applicability of the algorithm in a macroscopic travel demand model with on-demand services. In the first case study, the algorithm was used to determine the number of required vehicles and the vehicle distance traveled including empty trips for a regional carsharing system. The results show that the number of required vehicles can be reduced drastically by using carsharing as a substitute for private cars. The second case study quantified the impacts of autonomous scooters on the number of required scooters necessary for a shared scooter service. The results indicated considerable potential for reducing the required fleet size by relocating the scooters.

One limitation of the algorithm is that travel costs can be considered only to a limited extent. We discussed extensions to the algorithm to implicitly account for these costs and illustrate the trade-off between the number of vehicles and empty trips. In further extensions of the algorithm, time-of-day dependent travel times and different speeds for service trips and empty trips could be considered. Both extensions allow a more detailed modeling but potentially have a negative effect on the computation time.

Furthermore, the algorithm does not provide a lower bound on the number of required vehicles and thus on the solution quality. Although the solution quality was good in preliminary tests, the gap to the optimal value can be quite large as shown in the example of Figure 5.2. Further research should address the development of solution strategies to solve the extremely large instances to optimality, which is done in Chapter 6.

Appendix

5.A Notation

Greek letters

$\delta_{z_o z_d}$	Distance between two traffic zones z_o and z_d, in time intervals
$\overline{\delta}$	Maximum distance between any two traffic zones in an instance

Latin capitals

E	Set of edges in the network
$\mathcal{I}$	Instance
$N(z)$	Set of neighboring traffic zones of zone z
T	Set of time intervals
V	Set of nodes in the network
Z	Set of traffic zones

Latin lower case letters

a_{zt}	Number of available vehicles at a node v_{zt}
$d_{z_o z_d t}$	Demand from traffic zone z_o to z_d starting at time interval$t$
e	Index for edge in the network
$f_{z_o z_d t}$	Flow from traffic zone z_o to z_d starting at time interval$t$
$j_{z_o z_d}$	Travel time between traffic zones z_o and z_d, in minutes
l	Duration of a time interval
n	Number of additionally needed vehicles for relocation
v_{zt}	Node representing traffic zone z at beginning of time interval t
t	Index for time interval
x_z	Number of vehicles starting in traffic zone z
z	Index for traffic zone
z_o	Index for origin traffic zone
z_d	Index for destination traffic zone

Chapter 6

A rolling horizon heuristic with optimality guarantee for an on-demand vehicle scheduling problem

6.1 Introduction

On-demand transport services are becoming more and more popular among travelers and they have the potential to replace a significant part of the traditional public transport services in the near future. To be able to react to and regulate such services in a meaningful way, it is important for infrastructure managers and public transport authorities to model and estimate the impact of on-demand services on the utilization of streets. Recently, many microscopic approaches (Bischoff et al., 2017; Fagnant and Kockelman, 2018; Heilig et al., 2017) have been proposed to model on-demand services. These rely on the simulation of individual agents to obtain a virtual traffic volume and estimate the impact of on-demand vehicles on the infrastructure. In contrast to that, macroscopic approaches such as Richter et al. (2019) model vehicle and traveler movements as flows to estimate the utilization of streets.

In this chapter, we focus on macroscopic approaches and discuss a simple vehicle scheduling model for on-demand vehicles: Given demanded vehicle trips, what is the minimum number of vehicles needed to fulfill the demand? The resulting vehicle schedule describes vehicle itineraries and yields both the required size of the vehicle fleet and the positions of the vehicles over time. With this information, the utilization of streets can be estimated.

Most existing vehicle scheduling approaches are developed for operational purposes to find an assignment of vehicles to planned trips (El-Azm, 1985; Bunte and Kliewer, 2009; Foster and Ryan, 1976). Recent vehicle scheduling approaches focus on the integration of further operational aspects, for example, the integration of related planning steps (Carosi et al., 2019; Schöbel, 2017) or the integration of recharging strategies of battery electric vehicles (T. Liu and Ceder, 2020; M. Wen et al., 2016). Compared to these approaches, the application of vehicle scheduling to estimate the impact of on-demand services in macroscopic models brings two differences: (1) Demanded vehicle trips are not planned trips but correspond to expected passenger demand in a macroscopic model. Since these are expected values, both passenger demand and resulting vehicle fleet size might be fractional. (2) Compared to scheduled public transport, the amount of on-demand vehicle trips can be extremely large. Especially the second difference makes many existing optimization approaches unsuitable as corresponding problems easily exceed the size of tractable instances. In Chapter 5, the vehicle scheduling problem is modeled as a network flow problem that computes the size of the necessary vehicle fleet and their itineraries. Due to the large instances in realistic applications, in Chapter 5 the vehicle scheduling model is solved with a simple heuristic that constructs the vehicle schedule chronologically. While this heuristic scales well and is thus also able to solve very large instances, no guarantee on the solution quality is given. Therefore, it is unknown whether a vehicle schedule computed with the method from Chapter 5 is optimal or how far it is from an optimal solution.

The contribution of this chapter is a rolling horizon approach to solve the vehicle scheduling model introduced in Chapter 5 to optimality. The rolling horizon approach is a heuristic that splits instances into tractable subproblems and solves them independently. By enforcing overlap of the horizons of these subproblems, it is possible to look ahead to the demand of the next horizon and include that information while solving the current subproblem. As a consequence, the decisions taken in the current subproblem are well suited for the next subproblem and the overall solution quality can be improved. For a sufficient length of the overlap, we prove that a globally optimal solution can be found by composing the locally optimal solutions for the horizons. In numerical experiments, we could solve instances from Chapter 5 to optimality that were too large to be solved as a whole with a commercial solver. Furthermore, we could show that using the rolling horizon approach can bring a speed-up in solution time for large instances with millions of trips, compared to solving them as a whole to optimality.

The remainder of this chapter is structured as follows. In Section 6.2, the vehicle scheduling problem is introduced and in Section 6.3, we describe the rolling horizon heuristic in detail as our proposed solution approach. For a sufficiently long overlap of the horizons, we provide an optimality guarantee for the rolling horizon heuristic in Section 6.4. By modifying the formulation, we can strengthen the conditions for optimality. In Section 6.5, we show in numerical experiments that using the rolling horizon heuristic can help to speed up the solution process for large instances. Section 6.6 concludes the chapter.

6.2 Problem setting

In this section, we provide a detailed problem description of the vehicle scheduling problem following from the application in Chapter 5. We want to find a optimal vehicle schedule, that is, a feasible routing of a minimum number of vehicles meeting all given demand. A macroscopic model is considered in which neither the demand nor the resulting size of the vehicle fleet need to be integer.

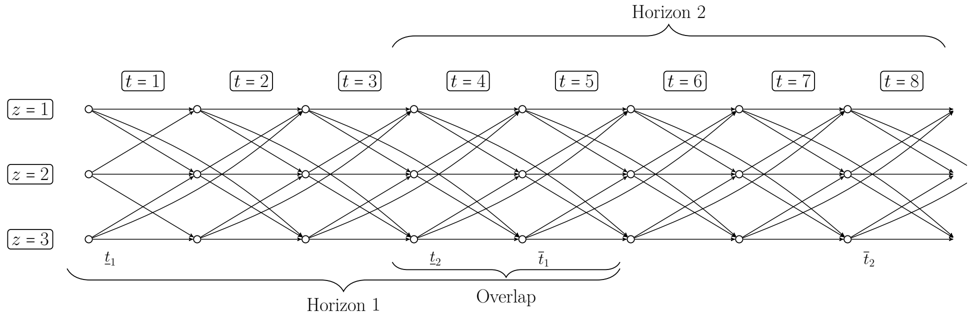


Figure 6.1: Instance with 3 zones and 8 time intervals. A possible assignment of time intervals to two overlapping horizons $\{\underline{t}_1, \ldots, \overline{t}_1\} = \{1, \ldots, 5\}$ and $\{\underline{t}_2, \ldots, \overline{t}_2\} = \{4, \ldots, 8\}$ is indicated.

In this setting, the considered time frame and observation area are discretized into a set of *time intervals* T and a set of *traffic zones* Z. The *distance* $\delta_{z_o z_d}$ between two zones is given as multiples of time intervals, i.e., $\delta_{z_o z_d} = n$ means that driving from zone z_o to z_d can be done in n time intervals. It is assumed that vehicle trips within a zone can be performed in one time interval, i.e., $\delta_{zz} = 1 \ \forall z \in Z$. The passenger demand is given aggregated to *demanded vehicle trips* $d_{z_o z_d t}$ between origin z_o and destination zone z_d, starting at the beginning of time interval t. The trips end at the beginning of time interval $t + \delta_{z_o z_d}$, determined by the distance between origin and destination zone and the start time. Vehicle trips either correspond to demanded person trips or, in applications with trip pooling, comprise multiple person trips. For carsharing or ridesharing applications, we assume that the pooling of person trips was done in a preceding step, for example, by the approach of Friedrich et al. (2018).

We denote an instance consisting of a set of zones Z, a set of time intervals T, a distance function $\delta_{z_o z_d}$ and aggregated demand $d_{z_o z_d t}$ by $\mathcal{I} = (Z, T, \delta, d)$. For an instance $\mathcal{I}$, the aim is to compute a vehicle schedule with a minimum number of vehicles that meet the demand. The fleet size and the vehicle routes in the schedule can be used to estimate the infrastructure utilization. Deadheading trips are allowed to relocate vehicles, and vehicles waiting in a traffic zone can be modeled as empty trips within a zone.

6.2.1 Network representation

As described in Chapter 5, this problem can be visualized in a time-space network. Figure 6.1 shows an instance with 3 zones and 8 time intervals. For each combination of z and t in the planning horizon, a node v_{zt} is introduced. These nodes are displayed in a grid structure with time intervals t on the horizontal and traffic zones z on the vertical axis. Each node v_{zt} represents a traffic zone z at the *beginning of a time interval* t. The distance $\delta_{z_o z_d}$ between two zones z_o and z_d is represented by the horizontal length of the arcs between the corresponding nodes. An arc $e_{z_o z_d t}$ between two nodes represents a possible vehicle trip between two zones z_o and z_d, starting at time interval t. The arrival time $t + \delta_{z_o, z_d}$ results from the start time t and the distance δ_{z_o, z_d}. For readability, the arrival time in the notation of an arc is omitted. The demanded vehicle trips $d_{z_o z_d t}$ are modeled as lower bounds on the arc $e_{z_o z_d t}$. A minimum flow in this network corresponds to a vehicle schedule with a minimum number of vehicles.

6.2.2 Model

In Chapter 5, it is proposed to model the vehicle scheduling problem as a network flow problem (Ahuja et al., 1988; Bunte and Kliewer, 2009; Schrijver, 2003) on the time-space network as described in Section 6.2.1. The following linear program (VS) for finding vehicle tours with a minimum number of vehicles was formulated in Chapter 5.

$$
\begin{aligned}
\min \quad & \sum_{z_o \in Z} \sum_{z_d \in Z} f_{z_o z_d 1} && && \text{(6.1a)} \\
\text{s.t.} \quad & f_{z_o z_d t} \geq d_{z_o z_d t} && \forall z_o, z_d \in Z,\ \forall t \in T && \text{(6.1b)} \\
& \sum_{\substack{z_o \in Z: \\ t - \delta_{z_o z} \geq 1}} f_{z_o z (t - \delta_{z_o z})} = \sum_{z_d \in Z} f_{z z_d t} && \forall z \in Z,\ \forall t \in T \setminus \{1\} && \text{(6.1c)} \\
& f_{z_o z_d t} \in \mathbb{R}_+ && \forall z_o, z_d \in Z,\ \forall t \in T && \text{(6.1d)}
\end{aligned}
$$

The flow variables $f_{z_o z_d t} \in \mathbb{R}_+$ denote the number of vehicle trips from zone z_o to zone z_d, starting at the beginning of time interval t. The objective (6.1a) is to minimize the total number of vehicles, expressed by the number of vehicle trips starting in the first time interval. The resulting number of vehicles for a flow f is also referred to as flow value $|f|$. The first set of constraints (6.1b) ensures that the demand is satisfied. If $f_{z_o z_d t} > d_{z_o z_d t}$, that is, if there are more vehicle trips than demanded, this can be interpreted as empty trips for vehicle relocation or waiting in a traffic zone if $z_o = z_d$. To obtain a feasible vehicle flow, the second set of constraints (6.1c) requires that the number of vehicles is preserved in each zone z at the beginning of each time interval t. The domains of the flow variables f in (6.1d) show that the vehicle trips do not need to be integer-valued.

6.2.3 Difficulty of the problem

The problem (VS) is a continuous linear program and can therefore be solved efficiently with available solvers for moderately sized instances. The coefficient matrix is totally unimodular, hence, the problem of finding integer flows is polynomially solvable for integer demand $d_{z_o z_d t} \in \mathbb{Z}$ (see Garey and Johnson (1979), problem [ND37], second comment). However, to determine the impact of on-demand vehicles on traffic and infrastructure in realistic cases of application, the number of time intervals and traffic zones may be enormous and yield intractable instances.

In Chapter 5, an application instance for the city area of Stuttgart is discussed. In this instance, the observation area is separated into $|Z| = 1175$ traffic zones and the time frame of one full day is segmented in $|T| = 96$ time intervals of 15 min. The numbers of variables and constraints of this instance are in the order of 10^8 and the corresponding optimization model could not be built with the general-purpose solver Fico Xpress 8.8 on a laptop with 32GB RAM[1].

To handle extremely large instance sizes, in Chapter 5 a simple heuristic is proposed that chronologically processes the nodes in the network and gradually constructs a vehicle schedule. By backtracking and repairing the vehicle schedule during construction, good solutions for huge instances can be achieved. However, the algorithm does not provide an approximation guarantee for the constructed vehicle schedules (see Williamson and Shmoys (2011) for more information on approximation algorithms). That means, it cannot be guaranteed how close the solution is to an optimal one. Furthermore, no optimality gap is provided by the design of the algorithm.

6.3 Rolling horizon heuristic

In this chapter, we propose using a rolling horizon heuristic to solve the model (VS). The idea is to divide the considered time frame into shorter time horizons and solve one subproblem for each time horizon. The solutions to the subproblems can be composed to a solution for the entire time frame.

[1]Hardware: Intel® Core™ i7-6700HQ CPU with 32GB of RAM; OS: Windows 10 Enterprise 2015 64-bit

6.3.1 Generalization of the linear program

To be able to compose the solutions of the subproblems to a feasible global solution, the optimization model (VS) from Chapter 5 is generalized. We consider an additional input of vehicles a_{zt} that become available in zone z at the beginning of time interval t. Available vehicles to be considered in one subproblem are a result of the flow fixed in previous subproblems. We denote the generalized input by $\overline{\mathcal{I}} = (Z, T, \delta, d, a)$ and generalize the optimization program to $(\overline{\text{VS}})$:

$$
\begin{aligned}
\min \quad & \sum_{z_o \in Z} \sum_{z_d \in Z} f_{z_o z_d 1} - a_{z_o 1} && && \text{(6.2a)}\\
\text{s.t.} \quad & f_{z_o z_d t} \geq d_{z_o z_d t} && \forall z_o, z_d \in Z,\ \forall t \in T && \text{(6.2b)}\\
& \sum_{\substack{z_o \in Z:\\ t-\delta_{z_o z} \geq 1}} f_{z_o z (t-\delta_{z_o z})} + a_{zt} = \sum_{z_d \in Z} f_{z z_d t} && \forall z \in Z,\ \forall t \in T \setminus \{1\} && \text{(6.2c)}\\
& \sum_{z_d \in Z} f_{z z_d 1} \geq a_{z1} && \forall z \in Z && \text{(6.2d)}\\
& f_{z_o z_d t} \in \mathbb{R}_+ && \forall z_o, z_d \in Z,\ \forall t \in T && \text{(6.2e)}
\end{aligned}
$$

As in the program (VS), the objective (6.2a) is to minimize the total number of vehicles needed to serve the demand. Since the flow variables f comprise all moving vehicles (including those that are given as available vehicles), available vehicles a in the first time interval are subtracted in the objective. This corresponds to minimizing the number of *additional vehicles* needed for serving the demand. The constraint (6.2b) ensuring that all demand is satisfied remains unchanged and is the same as constraint (6.1b). It is necessary to generalize the flow conservation constraints (6.2c) by treating available vehicles a_{zt} as incoming flow in nodes v_{zt}. Furthermore, an additional set of constraints (6.2d) ensures that the outgoing flow in the first interval considers all available vehicles a since this time interval is not covered in the flow conservation constraints (6.2c). Note, that for $a = 0$ the program $(\overline{\text{VS}})$ coincides with the program (VS).

6.3.2 Overlapping horizons

The idea of the rolling horizon heuristic is to divide the time frame into horizons and solve one subproblem for each horizon. By letting the horizons overlap it is further possible to look ahead to the demand of the next horizon. That means the demand in the overlap with the next horizon is considered in the subproblem of the current horizon. However, the vehicle trips to satisfy this demand are not yet fixed in the solution to the current subproblem, but in the solution to the next subproblem.

The longer the overlap, the more demand can be considered, which allows moving vehicles to positions that are well-suited for meeting demand in the next horizon. These positions of vehicles are considered as available vehicles a_{zt} in the subproblem for the next horizon.

A possible division of a time frame into two overlapping horizons is indicated in the example network in Figure 6.1. In this example, the first horizon $\{1, \ldots, 5\}$ spans over the first five time intervals. In the first subproblem, all demand starting in these five intervals is considered. However, only the flow starting before the overlap, that is, starting in the first three time intervals is fixed in the solution of the first subproblem. The flow in the overlap is fixed by solving the subproblem of the second horizon $\{4, \ldots, 8\}$.

6.3.3 Algorithm

With the generalized optimization model ($\overline{\text{VS}}$) we can define the rolling horizon heuristic. Its general idea is to solve the problem for smaller horizons that may be overlapping and compose the partial solutions to a solution for the full problem. Let h denote the number of time intervals in each horizon and let o denote the number of overlapping time intervals in the rolling horizon heuristic. Naturally, we require $0 \leq o < h$.

The pseudocode for the heuristic is presented in Algorithm 6.1. This algorithm processes one horizon $\{\underline{t}_i, \ldots, \overline{t}_i\}$ after another (Line 4), with the first horizon starting in the first time interval (Line 3). The horizons span h time intervals (Line 8) and each two consecutive horizons have an overlap of o time intervals (Line 7). For each iteration, the subproblem corresponding to the current horizon is called (Line 5 & 11), which is explained in detail in Algorithm 6.2. Afterward, the available vehicles are updated to communicate information from the solution of one subproblem to the next (Line 6). Available vehicles are a mean to model fixed vehicle trips that started before the overlap. By definition, these trips end $\max \delta_{z_o z_d} - 1$ time intervals after the beginning of the overlap at the latest.

In Algorithm 6.2, the optimization model ($\overline{\text{VS}}$) is called (Line 5) to find an optimal vehicle schedule for the subinstance $\overline{\mathcal{I}}$ that is constrained to the current horizon

Algorithm 6.1: *Rolling horizon heuristic*

Input: Instance $\mathcal{I} = (Z, T, \delta, d)$, length of horizon h, horizon overlap o

Output: Vehicle flow f

`# Initialize first horizon from time interval 1 to` h`, initialize variables for available vehicles` a `and flow` f `with 0;`

Initialize: $i \leftarrow 1$, $\underline{t}_i \leftarrow 1$, $\overline{t}_i \leftarrow h$, $a_{zt} \leftarrow 0 \quad \forall z \in Z, t \in T$, $f_{z_o z_d t} \leftarrow 0 \quad \forall z_o, z_d \in Z, t \in T$;

`# Iterate through horizons until end of time frame is reached;`

while $\overline{t}_i < |T|$ **do**

- `# Solve the subproblem corresponding to the current horizon` i `and get flow` f`;`
- $f \leftarrow$ *Solve subproblem*$(\mathcal{I}, \{\underline{t}_i, \dots, \overline{t}_i\}, a, f)$;
- `# Update number of available vehicles from fixed flow for next horizons;`
- $a_{zt} \leftarrow \sum_{\substack{z_o \in Z: \\ 1 \le t - \delta_{z_o z} < \overline{t}_i - o}} f_{z_o z (t - \delta_{z_o z})} \quad \forall z \in Z,\ t \in \{\overline{t}_i - o, \dots, \overline{t}_i - o + \max \delta_{z_o z_d} - 1\}$;
- `# Update bounds of next horizon and goto next horizon by increasing counter` i`;`
- $\underline{t}_{i+1} \leftarrow \underline{t}_i + h - o$;
- $\overline{t}_{i+1} \leftarrow \underline{t}_{i+1} + h$;
- $i \leftarrow i + 1$;

`# When reached end of time frame, truncate last horizon at` $|T|$`;`

$\overline{t} \leftarrow |T|$;

`# Solve the subproblem corresponding to the last horizon` i `and get flow` f`;`

$f \leftarrow$ *Solve subproblem*$(\mathcal{I}, \{\underline{t}_i, \dots, \overline{t}_i\}, a, f)$;

Return f;

Algorithm 6.2: *Solve subproblem*

Input: Instance $\mathcal{I}$, horizon $\{\underline{t}, \dots, \overline{t}\}$, available vehicles a, (partial) vehicle flow f

Output: Updated vehicle flow f

`# Initialize sub-instance` $\overline{\mathcal{I}}$ `constrained to the horizon;`

Initialize: $T' \leftarrow \{\underline{t}, \dots, \overline{t}\}$, $\overline{\mathcal{I}} \leftarrow (Z, T', \delta|_{T'}, d|_{T'}, a|_{T'})$;

do

`# Solve generalized optimization problem and get optimal flow` f' `for horizon` T'`;`

$f' \leftarrow (\overline{\mathrm{VS}})(\overline{\mathcal{I}})$;

`# Update total vehicle flow with solution from subproblem;`

$f_{z_o z_d t} \leftarrow f'_{z_o z_d t} \quad \forall z_o, z_d \in Z,\ t \in T'$;

`# Add additional vehicles to time intervals before the horizon to conserve flow;`

$f_{zzt} \leftarrow f_{zzt} + \min\{\sum_{z_d \in Z} f'_{z z_d \underline{t}} - a_{z\underline{t}}, 0\} \quad \forall z \in Z,\ t < \underline{t}$;

Return f;

(Line 3). After each optimization call, the total flow f is updated with the partial solution found (Line 6). Note, that this overwrites the vehicle flow in the overlap with the previous horizon. This procedure is equivalent to considering all demand in the current horizon, but just fixing the flow before the overlap, as described in Section 6.3.2. The flow in the overlap is discarded and then fixed with the solution of the next subproblem. If more vehicles were necessary to serve the demand in the current horizon than in the previous horizons, additional vehicles are added to the flow f (Line 7). This can be interpreted as introducing waiting vehicles in a traffic zone during all previous horizons.

Algorithm 6.1 keeps the structure of the rolling horizon heuristic, and for each horizon, the total flow is extended by the vehicle flow found in Algorithm 6.2. In the end, Algorithm 6.1 returns a vehicle flow for the whole time frame that is composed by the optimal partial flows for the horizons.

6.4 Quality of the solution

In this section, we prove that the vehicle schedules found by the rolling horizon heuristic in Algorithm 6.1 are optimal for certain choices of the overlap o. We start with the argument that the rolling horizon heuristic finds a feasible solution.

Definition 6.1. *A flow f is called* feasible *for an instance $\mathcal{I}$ if it satisfies all demand and fulfills the flow conservation in each vertex, i.e., if constraints* (6.1b) *and* (6.1c) *hold.*

Since the partial solutions are optimal and hence feasible solutions to the flow problems per horizon, the demand is satisfied by the composed vehicle flow. By carrying over vehicles to the next horizon with the help of available vehicles a, and by introducing additional waiting vehicles in previous horizons, the flow conservation holds.

Observation 6.2. *Hence, the composed vehicle schedule found by the rolling horizon algorithm is a feasible vehicle schedule for the whole time frame of an instance $\mathcal{I}$.*

Theorem 6.3. *The rolling horizon heuristic finds an optimal solution for an instance $\mathcal{I}$ if*

$$o \geq 2 \cdot \max_{z_o, z_d \in Z} \delta_{z_o z_d} - 1.$$

Sketch of the proof. The main idea of the proof for Theorem 6.3 is simple: Since the overlap is long enough, any vehicle trip that was fixed in the solution of a previous horizon can be corrected by another vehicle trip in any desired direction, if necessary.

We prove Theorem 6.3 by induction and start with the observation that the flow f^1 for the first horizon is optimal.

Next, we consider a flow for the first i horizons, denoted by f^i, and assume that it is optimal for the first i horizons. That means, it is optimal for the instance $\mathcal{I} = (Z, T^i, \delta, d)$ with restricted time frame $T^i := \{1, \ldots, \bar{t}_i\}$. Hence, the vehicle schedule f^i is a feasible flow that meets all demand starting at the latest at time interval $\bar{t}_i$, the end of the i^{th} horizon. Remember that the flow in the overlap $\{\bar{t}_i - o, \ldots, \bar{t}_i\}$ is not fixed yet but will be overwritten in the next iteration of the rolling horizon heuristic. The key is to show that fixing vehicle trips that start up to $t < \bar{t}_i - o$, the beginning of the overlap, does not prevent the rolling horizon heuristic to find an optimal solution f^{i+1} for the first $i+1$ horizons if $o \geq 2 \cdot \max \delta_{z_o z_d} - 1$.

To do the induction step, we first consider the demand in the overlap, and, afterward, vehicles that are not necessary to meet demand in the overlap. Since f^i is a feasible flow for the first i horizons, all demand in the overlap is satisfied. Of course, this demand is also met in any feasible solution for $i+1$ horizons. With some adaptions on the flow in the overlap, it can be shown that any optimal solution f^i for i horizons can be extended by any optimal solution after the overlap. Since these adaptions require extensive notation, we give a detailed proof for this in Appendix 6.B. The underlying idea of the proof is, that it is not important *which* vehicles meet the demand, but it is ensured that *sufficient* vehicles are available to meet the demand in the overlap.

For the remaining vehicles, we focus on vehicle trips in f^i that start before, and end at or after the beginning of the overlap $\bar{t}_i - o$. Vehicle trips that end earlier do not interfere with the next horizon, and vehicle trips that start later are overwritten by the solution for the next horizon. Hence, these trips in f^i do not restrict the solution for the $(i+1)^{\text{st}}$ horizon.

The trips under consideration start before the overlap, hence at $t \leq \bar{t}_i - o - 1$ and end at latest at $t \leq \bar{t}_i - o - 1 + \max \delta_{z_o z_d}$. Relocating the vehicles that have executed these trips to an arbitrary traffic zone from their current location takes at most another $\max \delta_{z_o z_d}$ time intervals. Hence, these vehicles are able to meet demand starting from any zone at time interval $t \leq \bar{t}_i - o - 1 + 2 \cdot \max \delta_{z_o z_d}$. For $o \geq 2 \cdot \max \delta_{z_o z_d} - 1$, the vehicles are able to meet demand just after the overlap at $t \leq \bar{t}_i$.

Together, this shows that fixing vehicle trips starting before the overlap in the solution of one subproblem and carrying this decision over to the next subproblem by the means of available vehicles, does not prevent finding a globally optimal solution.

There are sufficient vehicles to meet demand starting in the overlap and the remaining vehicles can be relocated to any zone to meet demand after the overlap. □

Theorem 6.3 gives a lower bound on the overlap to ensure finding optimal solutions. It is further possible to show that this lower bound is minimal.

Lemma 6.4. *For* $o < 2 \cdot \max \delta_{z_o z_d} - 1$*, optimality of the rolling horizon heuristic cannot be guaranteed.*

Proof. We consider an example instance containing two zones and five intervals. The maximum distance between two zones is $\delta_{12} = \delta_{21} = 2$. There is only demand of 1.0 vehicle trips within the first zone starting in time intervals 1 and 5, i.e., $d_{111} = d_{115} = 1.0$ and $d_{z_o z_d t} = 0.0$ otherwise. An outline of the underlying network can be found in Figure 6.2a.

One optimal solution is to use one vehicle that stays within the first zone all the time and satisfies the demand during the first and the fifth time interval. This solution is

$$f_{11t} = 1.0 \ \forall t \in \{1, \ldots, 5\} \text{ and } f_{z_o z_d t} = 0.0 \text{ otherwise},$$

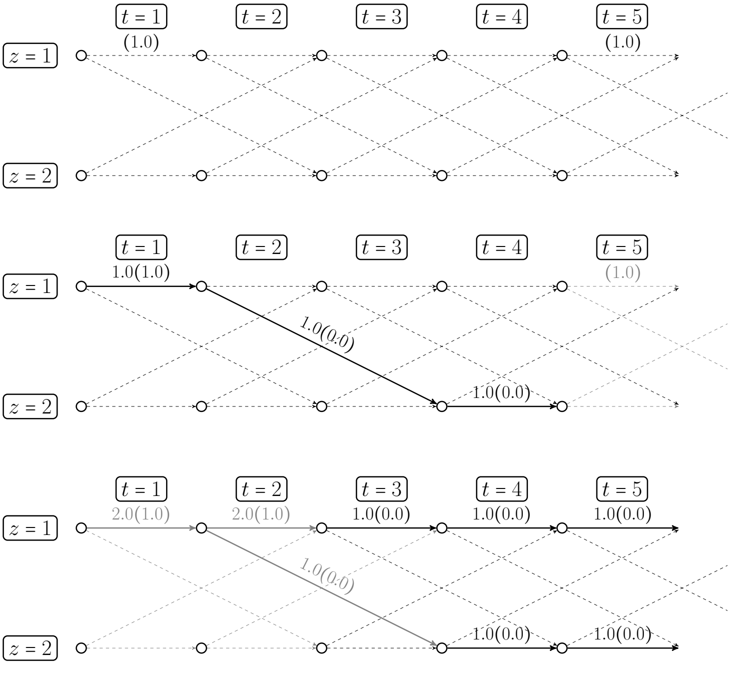
with an objective value of $|f| = 1.0$. Applying the rolling horizon heuristic to this instance with a horizon length of $h = 4$ and an overlap of $o = 2 = 2 \max \delta_{z_o z_d} - 2$, optimality cannot be guaranteed.

When considering the first horizon $\{1, \ldots, 4\}$, demand $d_{115} = 1.0$ lies outside the horizon and is not considered yet. Therefore, routing 1.0 vehicles to the second zone after meeting demand d_{111} is an optimal solution to the first subproblem. This partial solution

$$f_{111} = f_{122} = f_{224} = 1.0 \text{ and } f_{z_o z_d t} = 0.0 \text{ otherwise}$$

with an objective value of $|f| = 1.0$ is depicted in Figure 6.2b. Given this partial solution, it is impossible to satisfy the demand in the fifth time interval with the same vehicles. When considering the second horizon $\{3, \ldots, 5\}$, the vehicle trip f_{122} to the second zone cannot be reverted since it started before the overlap, and it is impossible to send the 1.0 vehicles back to the first zone in time. In this case, another 1.0 vehicles have to be added to satisfy demand d_{115} in the fifth interval, yielding a suboptimal global solution. This solution

$$f_{111} = 2.0, \ f_{11t} = 1.0 \ \forall 2 \le t \le 5, \ f_{122} = f_{224} = f_{225} = 1.0 \text{ and } f_{z_o z_d t} = 0.0 \text{ otherwise}$$



(a) Instance: Network with 2 zones and 5 time intervals, indicated demand and visualized distances.

(b) An optimal solution to first horizon $\{1, 2, 3, 4\}$. This prevents finding a globally optimal solution.

(c) An optimal solution to second horizon $\{3, 4, 5\}$. Previously fixed flow enforces use of additional vehicles.

Figure 6.2: Example network with 2 zones and 5 time intervals and suboptimal solution found by the rolling horizon heuristic for an overlap of $o = 2$ time intervals. Dashed lines indicate potential vehicle trips, thick edges positive flow, and grey lines and vertices are outside of the considered horizon. Numbers in brackets on edges show demand, numbers without brackets show vehicle flow.

has an objective value of $|f| = 2.0$ and can be seen in Figure 6.2c.

Note, that this example can be generalized to provide a counterexample for any maximum distance between two zones. Consider a network with the same pattern: two zones and a distance of $\delta_{12} = \delta_{21}$ between these two zones. Define the demand by $d_{111} = d_{11(2\max\delta_{z_oz_d}+1)} = 1.0$ and $d_{z_oz_dt} = 0.0$ otherwise. Then, the rolling horizon with a setting of $h = 2\max\delta_{z_oz_d}$ and $o = 2\max\delta_{z_oz_d} - 2$ might fail to find the optimal solution with the same argumentation. This generalization shows that a choice of $o = 2\max\delta_{z_oz_d} - 1$ is indeed the smallest value for the horizon overlap that ensures an optimal solution for Algorithm 6.1. □

The counterexample in the proof of Lemma 6.4 abuses the fact that an unreasonable decision to route an empty vehicle to another zone can appear in an optimal solution.

By preventing this kind of unreasonable vehicle trip, the condition for the optimality guarantee of the rolling horizon heuristic can be strengthened. In this context, a vehicle trip to a different zone is considered unreasonable if it does not satisfy any demand $d_{z_o z_d t}$ or if it is not performed to satisfy any demand in the zone of its destination. Staying within zones is never considered to be unreasonable as it is also used to model waiting vehicles during a time interval. This is formalized in the following definition.

Definition 6.5. *A flow f is called **unreasonable** if for an arc $e_{z_o z_d t}$ with $z_o \neq z_d$ none of the two conditions holds*

1. *Flow $f_{z_o z_d t}$ satisfies demand $d_{z_o z_d t}$, i.e., $f_{z_o z_d t} = d_{z_o z_d t}$.*

2. *Flow $f_{z_o z_d t}$ is performed to have enough vehicles available to meet demand $d_{z_d z t'}$ starting in zone z_d at $t' \coloneqq t + \delta_{z_o z_d}$, and otherwise there were too few vehicles, i.e.,*

$$f_{z_o z_d t} + \sum_{\substack{z_o \neq z \in Z: \\ t' - \delta_{z z_d} \geq 1}} f_{z z_d (t' - \delta_{z z_d})} = \sum_{z \in Z} d_{z_d z t'}.$$

As a consequence, if it is ensured that no flow is unreasonable, all vehicles stay in the destination zone z_d of their last satisfied demand $d_{z_o z_d t}$ unless they are needed to satisfy demand. In particular, sending 1.0 vehicles from the first to the second zone in the counterexample is unreasonable. Preventing unreasonable flow helps to improve the condition for an optimality guarantee.

Theorem 6.6. *If it is ensured in each iteration that no vehicle flow is unreasonable, the rolling horizon heuristic finds an optimal solution for the whole time frame if*

$$o \geq \max_{z_o, z_d \in Z} \delta_{z_o z_d}.$$

Sketch of the proof. The idea for the proof of Theorem 6.6 follows the structure of the one for Theorem 6.3. In this case, the vehicle trips that start before the overlap end at the beginning of the overlap $t = \bar{t}_i - o$, unless they meet demand. Then, after relocating them, the vehicles are able to meet demand at $t \leq \bar{t}_i - o + \max \delta_{z_o z_d}$, i.e., at $t \leq \bar{t}_i$ for $o \geq \max \delta_{z_o z_d}$.

Since the flow is not unreasonable, the fixed vehicle trips that end after $\bar{t}_i - o$ either meet demand or relocate vehicles to meet demand in the zone of destination. Hence, they are not a restriction to finding a globally optimal solution as these trips have to be performed in any feasible solution. With the same argumentation as in the proof of Theorem 6.6, it follows that the rolling horizon heuristic finds an optimal solution under the given conditions. □

Again, it is possible to show that this value is minimal.

Lemma 6.7. *For $o < \max \delta_{z_o z_d}$ optimality of the rolling horizon heuristic cannot be guaranteed, even if it is ensured in each iteration that no vehicle flow is unreasonable.*

Proof. We again provide a counterexample to show that the rolling horizon heuristic might not be optimal if $o = \max \delta_{z_o z_d} - 1$. This instance has two zones and $\max \delta_{z_o z_d} + 2$ intervals, where the case of $\max \delta_{z_o z_d} = \delta_{12} = 2$ is depicted in Figure 6.3a. The only demand in this instance is $d_{111} = d_{22(\max \delta_{z_o z_d}+2)} = 1.0$. An optimal solution is

$$f_{111} = f_{122} = f_{22(\max \delta_{z_o z_d}+2)} = 1.0 \text{ and } f_{z_o z_d t} = 0.0 \text{ otherwise}$$

with an optimal objective value of 1.0.

With the assumption of no unreasonable flow, the rolling horizon heuristic with parameter setting $h = \max \delta_{z_o z_d} + 1 = 3$ and $o = \max \delta_{z_o z_d} - 1 = 1$ cannot find an optimal solution for this instance. In the first horizon $\{1, \dots, \max \delta_{z_o z_d} + 1\}$, demand $d_{22(\max \delta_{z_o z_d}+2)}$ is not considered and flow $f_{111} = f_{112} = \dots = f_{11(\max \delta_{z_o z_d}+1)} = 1.0$ is fixed, see Figure 6.3b for the case $\max \delta_{z_o z_d} = \delta_{12} = 2$. In the second horizon $\{3, \dots, \max \delta_{z_o z_d} + 2\}$ it is not possible any more to route the available vehicle from the first zone to the second zone to satisfy demand $d_{22(\max \delta_{z_o z_d}+2)} = 1.0$. This situation can be seen in Figure 6.3c. It is necessary to introduce 1.0 additional vehicles in the second zone, which yields the suboptimal solution

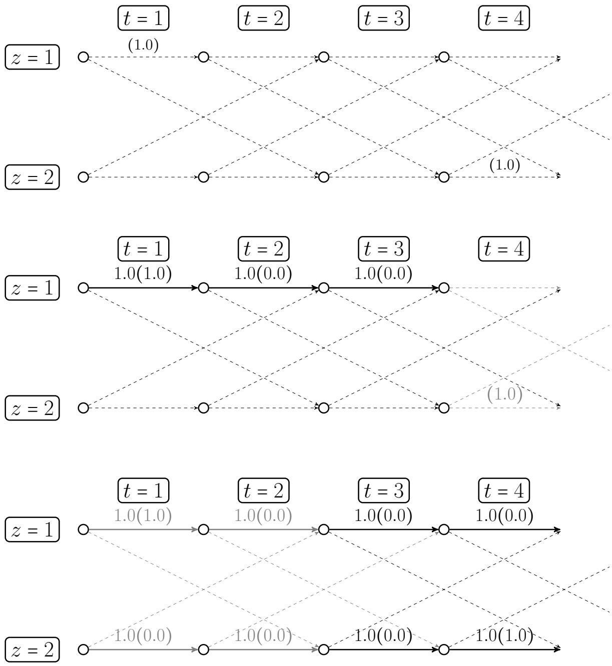
$$f_{11t} = f_{22t} = 1.0 \ \forall t \in T \text{ and } f_{z_o z_d t} = 0.0 \text{ otherwise}$$

with the objective value 2.0. This shows that in case of no unreasonable flow the rolling horizon heuristic only is guaranteed to find an optimal solution for $o \geq \max \delta_{z_o z_d}$. □

Unreasonable flow can be avoided, for example, by using the objective function (6.3):

$$\sum_{z_o \in Z} \sum_{z_d \in Z} f_{z_o z_d 1} - a_{z_o 1} + \sum_{z_o \in Z} \sum_{z_d \in Z} c_{z_o z_d} \sum_{t \in T} f_{z_o z_d t}. \tag{6.3}$$

with artificial routing costs c. By setting $c_{zz} = 0$ and $0 < c_{z_o z_d} < \frac{1}{|T|} \ \forall z_o \neq z_d \in Z$, waiting in a zone is always preferred to an unreasonable flow. The upper bound



(a) Instance: Network with 2 zones and 4 time intervals, indicated demand and visualized distances.

(b) An optimal solution to first horizon $\{1, 2, 3\}$. This prevents finding a globally optimal solution.

(c) An optimal solution to second horizon $\{3, 4\}$. Previously fixed flow enforces use of additional vehicles.

Figure 6.3: Example network with 2 zones and 4 time intervals and suboptimal solution found by the rolling horizon heuristic for an overlap of $o = 1$ time interval. Dashed lines indicate potential vehicle trips, thick edges positive flow, and grey lines and vertices are outside of the considered horizon. Numbers in brackets on edges show demand, numbers without brackets show vehicle flow.

on $c_{z_o z_d}$ ensures that never additional vehicles are acquired to save artificial routing costs. This means, using objective (6.3) with that cost setting minimizes the number of vehicles and at the same time prevents unreasonable flow.

6.5 Numerical experiments

Chapter 5 discusses instances with up to 10^8 vehicle trips and a similar amount of variables. While it was not possible to build an optimization model (VS) for such huge instances with the solver FICO Xpress on a laptop with 32GB RAM, optimal solutions for these instances could be found with the rolling horizon heuristic on the same machine.

p		1	2	3	4	5	6	7	8	9	10	11	
h		96	53	39	32	28	25	23	21	20	19	18	
$\|Z\|$	trips						rel CPU						abs CPU [s]
20	$3.8\,10^4$	**1.0**	1.17	1.27	1.43	1.40	1.53	1.69	1.84	1.90	2.05	2.15	1.0
40	$1.5\,10^5$	1.02	1.01	**1.0**	1.10	1.16	1.23	1.25	1.36	1.37	1.42	1.65	4.3
60	$3.5\,10^5$	1.31	1.06	**1.0**	1.06	1.09	1.15	1.15	1.22	1.26	1.44	1.41	10.7
80	$6.1\,10^5$	1.42	1.11	**1.0**	1.03	1.04	1.14	1.12	1.19	1.20	1.21	1.27	22.1
100	$9.6\,10^5$	1.49	1.08	1.01	**1.0**	1.02	1.03	1.06	1.09	1.16	1.19	1.27	38.4
120	$1.4\,10^6$	1.78	1.18	**1.0**	1.00	1.10	1.04	1.12	1.16	1.15	1.21	1.20	62.2
140	$1.9\,10^6$	1.72	1.17	1.04	**1.0**	1.05	1.06	1.07	1.08	1.12	1.13	1.15	95.8
160	$2.5\,10^6$	1.76	1.18	1.03	1.03	**1.0**	1.05	1.04	1.06	1.10	1.14	1.12	142.2
180	$3.1\,10^6$	1.82	1.18	1.04	**1.0**	1.03	1.02	1.04	1.02	1.04	1.05	1.10	194.0
200	$3.8\,10^6$	1.95	1.42	1.17	1.12	1.07	**1.0**	1.04	1.09	1.05	1.07	1.11	241.7
220	$4.6\,10^6$	1.93	1.25	1.08	1.10	1.07	1.03	**1.0**	1.05	1.08	1.10	1.08	316.1
240	$5.5\,10^6$	1.81	1.19	1.02	**1.0**	1.03	1.04	1.07	1.06	1.09	1.05	1.07	437.2

Table 6.1: Solution times for varying instance sizes (number of zones $|Z|$) and length of horizon h. The top two rows indicate the length per horizon h and the corresponding number of subproblems p. The first two columns state the number of zones $|Z|$ and the resulting number of vehicle trips. The last column gives the best absolute solution time in seconds per instance size. The remaining columns show the solution times relative to the best solution time per instance size.

Besides the fact that huge instances become tractable, splitting the problem into subproblems can speed up the solution process for tractable instance sizes. We conduct experiments on randomly generated instances with $|T| = 96$ time intervals, a maximum distance of $\max \delta_{z_o z_d} = 10$ time intervals between zones, and a varying number of zones $|Z|$. The rolling horizon heuristic is used with the adjusted objective function (6.3) and a minimum overlap of $o = 10$ that ensures finding an optimal solution. For each instance size, that means, for each number of zones $|Z|$, five randomly generated instances are solved with various settings for the horizon length h. The horizon length h and the overlap o determine the number of subproblems p that need to be solved during the rolling horizon heuristic. Applying the rolling horizon heuristic with a horizon length of 96 time intervals means solving the whole problem at once and is considered as the base case.

Table 6.1 shows relative and best absolute solution times for finding a globally optimal solution, averaged over five random instances for each instance size. Both the number of trips and the average absolute solving time increase exponentially with the number of zones, indicating that large instances are hard to solve.

A value of 1.0 in the top left corner indicates that it is fastest to solve the instances with 20 zones at once, i.e., with a horizon length of $h = 96$. With decreasing length of the horizon, and thus increasing number of subproblems, the solution times increase. For example, solving the same instances by splitting them up into 11 horizons spanning 18 time intervals each, thus solving 11 (smaller) subproblems, takes more than twice as long as the fastest option.

The larger the instances, the more it pays off to solve a larger number of small subproblems instead of only a few but large subproblems. With the tendency to increase further, solving instances with a number of trips in the order of magnitude of 10^6 at once took almost twice as long as solving them with the rolling horizon approach in the best setting. Comparably low computation times could be achieved with various settings for the horizon length.

6.6 Conclusion and Outlook

6.6.1 Conclusion

This chapter presented an alternative way to solve a simple vehicle scheduling problem as it occurs, for example, in the context of traffic estimation. The aim of the presented solution approach is to meet the given demand with the least amount of vehicles possible. For certain applications such as on-demand services, the number of demanded trips can be extremely large, making real-world instances intractable.

We proposed a rolling horizon heuristic to solve large instances of this problem. The principle is to split the considered time frame into small horizons and solve a vehicle scheduling problem for each horizon. For a sufficient overlap of the horizons, we proved that a solution composed of the partial solutions of the horizons is globally optimal. By introducing artificial routing costs, we could further relax the condition on the optimality criterion which makes finding optimal solutions less expensive.

In experiments, we found that the rolling horizon approach has a computation time advantage over solving a full model already for moderately sized instances, which illustrates the benefit of our approach also for instances of medium size.

6.6.2 Outlook

The presented rolling horizon approach was motivated with and developed for the application of vehicle scheduling in macroscopic demand models. However, the underlying theory of the solution approach is more general. The vehicle scheduling problem (VS) was modeled as a general network flow problem on a directed cycle-free graph. Hence, the presented rolling horizon approach is also applicable to a wider set of applications that can be modeled similarly.

Furthermore, it would be interesting to investigate whether the basic idea of the proof can be adjusted to be used in an even broader range of applications. The key ingredient of the proof is that decisions do not influence the remote future, which is the case in many applications with time-space networks, for example. Therefore, it might be possible to prove that a rolling horizon solution approach is capable of finding optimal solutions in other applications as well. This could be especially interesting for applications of online optimization where information is revealed successively.

Appendix

6.A Notation

Greek letters

γ	Vehicle duty
$\delta_{z_o z_d}$	Distance between two traffic zones z_o and z_d, in time intervals
φ	Value of flow f
τ	Vehicle tour

Latin capitals

$\mathcal{I}$	Instance
T	Set of time intervals
$\mathcal{T}$	Set of vehicle tours
W	Set of nodes v with waiting vehicles
Z	Set of traffic zones

Latin lower case letters

a_{zt}	Number of available vehicles at node v_{zt}
$c_{z_o z_d}$	Travel cost per vehicle from traffic zone z_o to z_d
$d_{z_o z_d t}$	Demand from traffic zone z_o to z_d starting at time interval$t$
$e_{z_o z_d t}$	Edge from traffic zone z_o to z_d starting at time interval t
$f_{z_o z_d t}$	Flow from traffic zone z_o to z_d starting at time interval t
h	Number of time intervals in a horizon
i	Iterator used to index horizons
k	Iterator used to index vehicle tours
o	Number of overlapping time intervals in the rolling horizon heuristic
p	Number of subproblems
t	Index for time interval
v_{zt}	Node representing traffic zone z at beginning of time interval t
w_{zt}	Waiting vehicles at node v_{zt}
x	Number of vehicles
z	Index for traffic zone
z_o	Index for origin traffic zone
z_d	Index for destination traffic zone

6.B Proof of optimality

To proof Theorem 6.3, some additional definitions and observations are helpful. From Observation 6.2 we know that a composed solution for an instance $\mathcal{I}$ found by the rolling horizon heuristic is feasible. It remains to prove that the solution does not require more vehicles than an optimal solution to the program (VS). To this end, we introduce necessary notation to examine the vehicle flow per vehicle tour.

Definition 6.8. *A* φ-vehicle tour *is a sequence*

$$\tau = ((z_{o1} z_{d1} t_1), \dots, (z_{on} z_{dn} t_n))$$

of n *consecutive vehicle trips with a positive flow of value* $\varphi \in \mathbb{R}_+$*. Consecutive trips are characterized by*

$$z_{di} = z_{oi+1} \text{ and } t_i + \delta_{z_{oi} z_{di}} = t_{i+1} \qquad \forall 1 \leq i < n.$$

A φ-vehicle tour can be imagined as a tour that is driven by exactly φ vehicles. Obviously, a vehicle schedule consists of many vehicle tours:

Observation 6.9 (Schrijver, 2003)**.** *A feasible flow* f *can be decomposed into a finite set of vehicle tours* $\{\tau_k\}_k$ *such that the sum of all vehicle tour values* $\sum_k \varphi_k$ *equals the total flow* $|f|$*. Each of the vehicle tours spans the whole time frame, i.e.,*

$$t_1 = 1 \text{ and } t_n + \delta_{z_{on} z_{dn}} > |T|.$$

Such a decomposition is not unique.

Next, we introduce vehicle duties to keep track of which vehicle tour serves which demand.

Definition 6.10. *Let* $\mathcal{I} = (Z, T, \delta, d)$ *be an instance and let* f *be a feasible vehicle schedule, decomposed to a set of vehicle tours. A mapping* γ *from a vehicle tour* τ *and a vehicle trip* (z_o, z_d, t) *to a positive value,*

$$\gamma : (\tau, (z_o, z_d, t)) \mapsto \mathbb{R}_+$$

is called vehicle duty *if the following three conditions hold:*

1. *The value is only positive if the vehicle trip is in the tour,*

$$\gamma(\tau, (z_o, z_d, t)) > 0 \Rightarrow (z_o, z_d, t) \in \tau.$$

2. *The value is at most the value φ of the vehicle tour,*

$$\gamma(\tau, (z_o, z_d, t)) \le \varphi.$$

3. *The sum of values for one vehicle trip (z_o, z_d, t) sum up to the demand $d_{z_o z_d t}$ on that trip,*

$$\sum_{\tau:(z_o,z_d,t)\in\tau} \gamma(\tau, (z_o, z_d, t)) = d_{z_o z_d t} \quad \forall (z_o, z_d, t).$$

A vehicle duty can be interpreted as assigning all demand to vehicle trips that meet the demand.

Observation 6.11. *For each feasible flow f decomposed to a set of vehicle tours, there exists a vehicle duty such that the tour value φ of each vehicle tour equals the value of the last positive demand assigned to the tour. Demand $d_{z_o z_d t}$ is called the last demand assigned to the tour if there is no other demand $d_{z_o' z_d' t'}$ assigned to that tour with $t' > t$.*

For any vehicle duty, this can easily be constructed by iteratively splitting each φ-vehicle tour not fulfilling this criterion into two φ_1 and φ_2 vehicle tours with the same sequence of vehicle trips where at least one tour fulfills the criterion.

Definition 6.12. *Such a vehicle duty is called* maximal vehicle duty*.*

Proof of Theorem 6.3. We show that a composed vehicle schedule of the rolling horizon heuristic is optimal for the whole time frame by induction over the number of horizons.

Induction basis

It is easy to see that the optimization program ($\overline{\text{VS}}$) finds an optimal solution f^1 for the first horizon $\{\underline{t}_1, \ldots, \overline{t}_1\} = \{1, \ldots, h\}$.

Induction hypothesis

We consider a solution f^i of the rolling horizon algorithm for the first i horizons and as an induction hypothesis, we assume that the solution is optimal. That means, it is not possible to satisfy all demand $d_{z_o z_d t}$ for $t \le \overline{t}_i$ with less than $x^i := |f^i|$ vehicles. This flow f^i is fixed up to the beginning of the overlap and may not be changed by the solution of a future horizon. The flow in the overlap, $f^i_{z_o z_d t}$ for $t \ge \overline{t}_i - o$ is overwritten by the solution of the next horizon and may change.

Induction step

Let f^* be a optimal solution for $i+1$ horizons, for example found by solving the optimization model (VS). Our aim is to show that a solution from the next iteration of the rolling horizon heuristic with an overlap of $o \geq 2 \cdot \max \delta_{z_o z_d} - 1$ is optimal for $i+1$ horizons. This is done by constructing a feasible flow f^{i+1} for the first $i+1$ horizons that is identical to the flow f^i before the overlap and uses $x^* := |f^*|$ vehicles. Since we can construct such a solution, Algorithm 6.2 in the rolling horizon heuristic will find a solution that is at least as good.

First, we consider a decomposition of the flow f^i into finitely many vehicle tours τ^i, and a maximal vehicle duty γ^i assigning all demand to the vehicle tours. We 'cut off' each vehicle tour τ^i

I after meeting the last demand that starts in the overlap and that is assigned to that tour, or else,

II if no demand starting in the overlap is assigned to that tour in the vehicle duty, after the first vehicle trip that ends *in* the overlap.

To formalize, let

$$\tau^i = ((z_{o1}, z_{d1}, t_1), \ldots, (z_{ok}, z_{dk}, t_k), (z_{ok+1}, z_{dk+1}, t_{k+1}), \ldots, (z_{on}, z_{dn}, t_n))$$

be a vehicle tour in flow f^i. Let (z_{ok}, z_{dk}, t_k) be the last vehicle trip starting in the overlap with demand assigned to the tour τ^i, or else, be the first vehicle trip that ends in the overlap. Then, the rear part after this vehicle trip, starting with $(z_{ok+1}, z_{dk+1}, t_{k+1})$, is cut off, which yields the incomplete tour

$$\tau' = ((z_{o1}, z_{d1}, t_1), \ldots, (z_{ok}, z_{dk}, t_k)).$$

This can be interpreted as letting all vehicles from vehicle tour τ^i *wait* in zone z_{dk} at time interval $t_k + \delta_{z_{ok} z_{dk}}$.

I We denote the number of all vehicles waiting in node v_{zt} after meeting demand that starts in the overlap by w^{I}_{zt} and initialize the set of vertices where these vehicles are waiting as

$$W^{\mathrm{I}} = \{v_{zt} : \; w^{\mathrm{I}}_{zt} > 0, \; z \in Z, \; \bar{t}_i - o < t \leq \bar{t}_i + \max \delta_{z_o z_d}\}.$$

II Equivalently, we denote the number of all vehicles waiting in node v_{zt} after the first vehicle trip ending in the overlap by w^{II}_{zt}.

This leaves us with incomplete vehicle tours that start in the first time interval and end sometime after the beginning of the overlap with waiting vehicles.

Next, we use these incomplete vehicle tours as a basis for the flow f^{i+1} that we want to construct as a solution for the first $i+1$ horizons. We set

$$f^{i+1}_{z_o z_d t} \coloneqq \sum_{\tau' : (z_o, z_d, t) \in \tau'} \varphi(\tau') \quad \forall z_o, z_d \in Z,\ t \in \{1, \dots, \bar{t}_i\}$$

where $\varphi(\tau')$ is the flow value of vehicle tour τ'. We want to highlight three characteristics of f^{i+1}:

1. Since the tours τ' are not cut off before the start of the overlap, f^{i+1} is identical to flow f^i up to the beginning of the overlap. This is required for the construction of f^{i+1} since all vehicle trips before the overlap are fixed by the design of the rolling horizon heuristic.

2. Since it is defined by incomplete tours, f^{i+1} is *not* a feasible flow (yet). The flow conservation does not hold at some nodes. In this proof, we show that it is possible to extend it to a feasible flow at these nodes.

3. Since the tours τ' are cut off after meeting the last demand, f^{i+1} does meet all demand starting up to the end of the overlap.

Our goal is to show that we can complete f^{i+1} to a feasible flow for $i+1$ horizons while using x^* vehicles. The sum of all flow values of the incomplete tours is x^i, equal to the flow value of f^i. It holds that $x^i \leq x^*$, otherwise f^i is not optimal for the first i horizons as $f^*|_i$ would be a better solution, which contradicts the induction hypothesis. In case that $x^i < x^*$ we add $(x^* - x^i)$ more vehicles to f^{i+1} at an arbitrary zone, for example by letting them stay in the first zone until the beginning of the overlap:

$$f^{i+1}_{11t} \coloneqq f^i_{11t} + (x^* - x^i) \quad \forall t < \bar{t}_i$$

This increases the number of waiting vehicles $w^{\mathrm{II}}_{1\bar{t}_i}$ in node $(1, \bar{t}_i)$ by $(x^* - x^i)$. Then, the sum of all flow values in f^{i+1} is x^*, as in any optimal flow f^*.

Next, we consider an arbitrary but fixed optimal solution f^* for $i+1$ horizons, decomposed into finitely many vehicle tours τ^*, and a vehicle duty γ^* assigning all demand to the vehicle tours τ^*. Let $\mathcal{T}$ be the set of all tours τ^* in f^*. We aim to extend the incomplete tours in f^{i+1} with the rear parts of the tours in f^*.

I First, we consider waiting vehicles w^{I}_{zt} that met demand starting in the overlap. We extend f^{i+1} according to the following procedure:

While W^{I} is not empty, we choose an arbitrary node $v_{zt} \in W^{\mathrm{I}}$. By definition of W^{I}, at least w^{I}_{zt} demanded vehicle trips end in node v_{zt}. Hence, there exist vehicle tours $\tau^* \in \mathcal{T}$ that this demand was assigned to, otherwise f^* was infeasible. We extend f^{i+1} at node v_{zt} with these tours τ^* from f^* until there are no more waiting vehicles w^{I}_{zt} in node v_{zt}:

While $w^{\mathrm{I}}_{zt} > 0$, we take such a tour τ^* with tour value $\varphi(\tau^*)$ and remove it from the set $\mathcal{T}$. If $\varphi(\tau^*) > w^{\mathrm{I}}_{zt}$, we split the tour τ^* into two tours with the same sequence of vehicle trips as τ^*, one tour τ^*_w with flow value w^{I}_{zt}, and one tour $\tau^*_{\varphi-w}$ with flow value $\varphi(\tau^*) - w^{\mathrm{I}}_{zt}$. Else, for $\varphi(\tau^*) \leq w^{\mathrm{I}}_{zt}$, we take the tour with the full value $\varphi(\tau^*)$ and define $\tau^*_w \coloneqq \tau^*$.

We extend f^{i+1} at node v_{zt} with tour τ^*_w and put $\tau^*_{\varphi-w}$ back into the set $\mathcal{T}$. Extending f^{i+1} with tour τ^*_w means, we increase the flow value $f^{i+1}_{{z_o}'{z_d}'t'}$ for each vehicle trip $({z_o}', {z_d}', t')$ in tour τ^*_w after node v_{zt} by the value $\varphi(\tau^*_w)$,

$$f^{i+1}_{{z_o}'{z_d}'t'} = f^{i+1}_{{z_o}'{z_d}'t'} + \varphi(\tau^*_w) \quad \forall ({z_o}', {z_d}', t') \in \tau^*_w\colon\ t' \geq t.$$

Based on this extension of f^{i+1}, we update the number of waiting vehicles:

At the current node v_{zt}, there are $\varphi(\tau^*_w)$ waiting vehicles less, hence, we set $w^{\mathrm{I}}_{zt} \coloneqq w^{\mathrm{I}}_{zt} - \varphi(\tau^*_w)$. Moreover, it might be that some further demand $d_{{z_o}'{z_d}'t'}$ starting in the overlap after node v_{zt} was assigned to vehicle tour τ^*_w in the optimal flow f^*, that means, $\gamma^*(\tau^*_w, ({z_o}', {z_d}', t')) > 0$ for $t < t' \leq \bar{t}_i$. Then, we assign this demand to the newly extended tour in f^{i+1} as well. In particular, we undo the assignment of this demand to another tour τ^{i+1} in f^{i+1}.

If demand $d_{{z_o}'{z_d}'t'}$ was the last demand assigned to tour τ^{i+1}, this has two consequences: First, it caused $w^{\mathrm{I}}_{{z_d}'(t'+\delta_{{z_o}'{z_d}'})}$ waiting vehicles after the demanded vehicle trip $({z_o}', {z_d}', t')$. We remove these waiting vehicles since the demand is met by the newly extended vehicle tour in f^{i+1} as well:

$$w^{\mathrm{I}}_{z_d'(t'+\delta_{z_o'z_d'})} \coloneqq \max\{w^{\mathrm{I}}_{z_d'(t'+\delta_{z_o'z_d'})} - \gamma^*(\tau^*_w, (z_o', z_d', t')), 0\}.$$

Second, since the assignment of the last demand to tour τ^{i+1} is undone, either another demand $d_{z_o''z_d''t''}$ with $\bar{t}_i - o \leq t'' < t$ is the last demand, or no other demand that starts in the overlap is assigned to tour τ^{i+1}. In the first case, we increase the number of waiting vehicles $w^{\mathrm{I}}_{z_d''(t''+\delta_{z_o''z_d''})}$ after that demand by $\gamma^*(\tau^*_w, (z_o', z_d', t'))$ since it is now the last demand. In the second case, we increase the number of waiting vehicles $w^{\mathrm{II}}_{\hat{z}\hat{t}}$ by $\gamma^*(\tau^*_w, (z_o', z_d', t'))$, where node $(\hat{z}, \hat{t})$ is the first node of tour τ^{i+1} in the overlap.

If $w^{\mathrm{I}}_{zt} = 0$ for any node v_{zt} after updating of the number of waiting vehicles, we remove it from W^{I} and continue with the next node in W^{I}.

This procedure extends f^{i+1} with vehicle trips from f^* until there are no more waiting vehicles w^{I}_{zt} at node v_{zt}. During this construction, also the waiting vehicles at other nodes might be changed. We want to emphasize that this procedure is well-defined and finite. There exist sufficient vehicle tours in the set $\mathcal{T}$ to be chosen from in the procedure. For each node v_{zt}, at most vehicle tours with a total flow value of incoming demand at v_{zt} are requested from set $\mathcal{T}$. Since all demand is met by the solution f^*, these tours exist. Furthermore, we reduce the waiting vehicles in all future nodes by the flow value of a tour, if a tour with assigned demand is removed from $\mathcal{T}$. Hence, taking a tour τ^* with assigned demand ending in v_{zt} from set $\mathcal{T}$ is well-defined. In each update of the number of waiting vehicles, the total number of waiting vehicles never increases. Furthermore, it is impossible to process waiting vehicles caused by the same demand twice, which makes the procedure finite.

This procedure is applied to all nodes with waiting vehicles w^{I} until the set of waiting vehicles W^{I} is empty. After this procedure, we obtain an incomplete flow f^{i+1} with some complete tours that start in the first time interval and reach the end of the horizon, and some waiting vehicles w^{II} after the beginning of the overlap that were not treated yet.

II Second, we consider these waiting vehicles w^{II} after the beginning of the overlap. We start with determining the amount of waiting vehicles w^{II}: Let x denote the sum of the flow values of the complete tours constructed in case I. Since these tours are based on flow f^i and extended with tours from the optimal solution f^*, the sum of the flow values of the vehicle tours left in the set of tours $\mathcal{T}$ is equal to the total number of waiting vehicles w^{II}, namely $(x^* - x)$. The waiting vehicles w^{II}_{zt} are present at nodes v_{zt} after the first vehicle trip that starts before and ends in the overlap. Hence, the vehicles are waiting in zone z at the beginning of time interval t with

$$t \leq \bar{t}_i - o - 1 + \max \delta_{z_o z_d}.$$

It is important to note that all demand in the overlap is met by the complete tours constructed in case I and we do not need to take care of this.

We disconnect the remaining vehicle tours in the set $\mathcal{T}$ at the first node $v_{z't'}$ after the overlap into two incomplete tours. Then, it is possible relocate the waiting vehicles to zone z' within at most $\max \delta_{z_o z_d}$ time intervals, that means the vehicles can be available in zone z' at latest at

$$t \leq \bar{t}_i - o - 1 + 2 \cdot \max \delta_{z_o z_d} \leq \bar{t}_i \leq t'.$$

That means, it is possible to relocate the waiting vehicles w^{II} within the overlap and extend f^{i+1} with the rear parts of the disconnected vehicle tours from f^*.

As a result, we obtain a feasible flow f^{i+1} for the first $i+1$ horizons that uses x^* vehicles. This flow is identical to flow f^i before the overlap, and identical to flow f^* after the overlap. For the time intervals in the overlap, we constructed f^{i+1} in such a way that it connects f^i and f^*. By construction, it is ensured that all demand is satisfied and with the help of waiting vehicles we could connect the flows ensuring flow conservation at each node.

Since it is possible to construct a flow f^{i+1} with these characteristics, the rolling horizon algorithm will find a vehicle schedule for $i+1$ horizons that is at least as good. The theorem follows by induction. $\square$

Chapter 7

Conclusion

In this thesis, we investigated the interaction of the estimation of passenger demand and the planning of public transport services. Existing research on public transport planning can in most cases be attributed to either the demand-oriented or the supply-oriented approaches. In demand-oriented approaches, a transport supply is assumed to be fixed and the corresponding passenger demand is evaluated. In this setting, a new or an adapted version of a public transport service is compared to the status quo and then adjusted. In supply-oriented approaches, a public transport service is designed for a given demand. Both approaches neglect the interaction of demand and supply. In particular, if a public transport service changes, the corresponding demand is likely to change as well. The other way around, when the demand structure changes, the service should be adjusted to the new demand. We developed novel optimization models to estimate the passenger demand during public transport service optimization. The focus is on the design of the timetable and line plan on the supply side, and the estimation of the passengers' route and mode choices on the demand side.

Furthermore, we considered a vehicle scheduling problem as it occurs in the context of on-demand services in travel demand modeling. Travel demand models require a transport supply to be given. However, in contrast to scheduled public transport, on-demand services cannot be planned beforehand and be used as input. Instead, the service depends on and changes with the demand. Hence, to estimate the demand, also the on-demand service has to be estimated simultaneously. We used a vehicle scheduling model to estimate the fleet size and vehicle distance traveled of the on-demand service and develop two solution approaches to solve the problem for really large instances.

In this chapter, we summarize the main findings of this thesis, highlight implications from our research, reflect on the limitations of this thesis, and discuss directions for future research.

7.1 Main findings and implications

Enhancing supply-oriented models by integrating passenger demand estimation implies that the considered objective functions have to be modified. Before modifying objective functions, we first used the example of timetables in Chapter 2 to examine the extent to which evaluation functions agree on the quality of solutions. To this end, we classified established evaluation functions from the literature and identified three components in which the functions differ from each other. Based on the components, we defined a set of representative timetable evaluation functions for the commonly used functions in the literature. All of the considered functions are designed to assess the quality from the passengers' perspective.

We investigated to what extent the different timetable evaluation functions agree on the quality of timetables. For this purpose, we described a novel method to quantify the inconsistency between evaluation functions. This method allows an analysis of the inconsistency of the defined evaluation functions. The findings of this analysis are qualitatively similar for sets of timetables on an artificial grid network and the railway network of Netherlands Railways. Although the evaluation functions are defined to assess the quality of timetables from the passengers' perspective, we found high inconsistencies between them. Due to the structure of the evaluation functions, it is possible to identify which components are responsible for differences in evaluation results. Vice versa, this structure also gives insights into how evaluation functions can be simplified without distorting the evaluation results.

Most notably was the found difference between travel time-based and utility-based evaluation functions. While travel time-based functions are mostly used for evaluation, utility-based functions are shown to be suitable for choice modeling of passengers. This raises the question of why utility-based functions are not used more for evaluation. Furthermore, also within travel time-based evaluation functions high inconsistencies were found, caused by different quality measures, parameter settings, and assumptions on the passenger distribution. Our findings support the supposition that timetable evaluation functions can be inconsistent, even if they are all designed for evaluation from the passenger perspective. The inconsistencies implicate that, depending on which evaluation function for assessing the quality for passengers is used, different timetables might be considered optimal. This finding is particularly crucial for Operations Research models where the evaluation functions are used as objectives to guide the search for solutions.

In Chapter 3 we studied how a passenger distribution can be estimated during timetabling. The public transport timetable determines travel and transfer times for passengers on each possible route and hence has an impact on which routes passengers choose. Given the quality of the routes, the passenger distribution can be estimated with the logit model. Since the logit model is non-linear and non-convex in the quality of routes, we investigated two linear representations of a passenger distribution within timetabling. For the first representation, we developed a novel multidimensional linear passenger distribution model which resembles the logit model. The second representation is based on a simulation framework to approximate a logit distribution within the optimization model.

In experiments on a set of artificial instances and a part of the Dutch Railway network, we compared the two novel formulations with four state-of-the-art approaches. All solutions were compared using four different evaluation functions, including both travel time-based and utility-based functions. In general, we were able to find better timetables for passengers using the novel models with an integrated passenger route choice model than with the considered state-of-the-art approaches. Compared to the state-of-the-art approaches, the new methods significantly reduced the gap to an ideal solution for passengers according to some evaluation functions, while achieving similarly good results according to other evaluation functions. In addition, our experiments provide insight into (1) how considering multiple routes for passengers instead of a single route, and (2) how integrating route choice instead of a predetermined route assignment affects solution quality.

In Chapter 4, we investigated the interaction between line planning and mode and route choices of travelers. Travelers only choose to use a public transport service, if it offers a good service quality. In particular, the service should be frequent, fast, and with as few transfers as possible. These factors of influence for both the mode and route choice of travelers are determined by the line plan. Hence, we considered mode and route choice during line planning to get a good estimate of the number of passengers and design the service for the corresponding passenger demand.

We developed a problem formulation for line planning with integrated choice models for mode and route choice. By suitable assumptions and preprocessing of the utilities of routes for passengers, we are able to provide a mixed-integer linear program for this problem. The mixed-integer linear program can be solved with available general-purpose solvers and we provide and test means to improve the computational performance. The model can be combined with any choice model to estimate the mode choice that does not need to be linear.

In experiments on the Intercity network of the Randstad, a metropolitan area in the Netherlands, we used the logit model to estimate the mode choice. We compared the developed model with a standard line planning model that assumes a fixed passenger demand. We found that integrated demand estimation yields line plans that are well-suited for the demand they generate. The line plans found with the developed model generated higher profits for the operator and provided a higher level of service to the passengers. These experiments suggest that the passenger demand should be estimated during line planning, otherwise the solution quality can be limited by the assumed passenger demand. Furthermore, a sensitivity analysis showed that the operators' profit reacts sensitively to the total travel demand. Consequently, operators should respond regularly to changes in travel demand and adjust their services. This helps them to obtain the highest possible profit and to offer a high-quality service to their passengers.

In Chapter 5 we considered a vehicle scheduling problem for on-demand services within a macroscopic travel demand model. The travel demand model estimates the travel demand for a given transport supply. However, on-demand services depend on the demand and the service level cannot be determined before the demand is known. Hence, the service level and its impact on the transport network have to be estimated during demand estimation. Important factors are the required fleet size and the vehicle distance traveled. These values can be found with a vehicle schedule for the on-demand service. Since the required vehicle trips are not necessarily integer in a macroscopic model, and the number of required trips is considerably larger than in scheduled public transport applications, a new approach has to be developed.

We modeled this problem in Chapter 5 as a network flow problem and provided a heuristic solution algorithm to construct vehicle schedules. The algorithm can be applied to both integer and non-integer demand values and is therefore especially suited for an application in macroscopic models. Our heuristic scales well and we were able to find solutions to huge instances in short computation times. We demonstrated in two case studies how the algorithm can be applied to estimate the vehicle fleet size and the vehicle distance traveled of on-demand services within a macroscopic travel demand model. In both case studies, our results show that the number of required vehicles can be drastically reduced by relocating vehicles. Furthermore, we provide extensions to the algorithm to implicitly consider the travel costs of the vehicles. The trade-off between the number of required vehicles and vehicle distance traveled is illustrated in experiments.

In Chapter 6, we considered the same vehicle scheduling problem as in Chapter 5. The solution algorithm developed in Chapter 5 is able to find good vehicle schedules in short computation times, but it does not provide a guarantee on the solution quality. It is not known whether the found solution is optimal or how far it is from an optimal solution.

Therefore, we developed an alternative solution algorithm for the vehicle scheduling problem of on-demand services within a macroscopic travel demand model. The algorithm is based on a rolling horizon framework where the considered time frame is split into smaller horizons. For each horizon, a sub-problem can be solved to optimality and the partial solutions can be composed to a solution for the whole problem. We generalized the network flow problem such that the horizons can overlap and provided an optimality guarantee for the composed solution if the horizons overlap sufficiently. With this framework, we were able to find optimal solutions for instances that were too large to be solved at once. In this way, Chapter 6 helped to understand that the heuristic approach in Chapter 5 provides good solutions also for very large instances. We also showed in experiments that the rolling horizon framework brings a speed-up for solving large instances with millions of trips, compared to solving them as a whole.

Moreover, the presented solution strategy in Chapter 6 is more general. We developed it for a vehicle scheduling problem modeled as a network flow problem on a directed cycle-free graph. Hence, it is possible to be applied to a wider set of applications that are based on a similar network flow problem than only to the presented vehicle scheduling problem. This means the rolling horizon algorithm with overlapping horizons and the optimality guarantee can be used to find optimal solutions for extremely large instances for other applications as well.

7.2 Limitations and future research

Solution approaches. The research in this thesis was motivated by the need for better public transport services that are tailored to the corresponding demand. We developed novel public transport optimization models with integrated choice models for passenger demand estimation. In experiments, we showed the benefits of these approaches and gained valuable insights for operators and public transport authorities. However, for the experiments in Chapters 2 and 3 we relied on the solver Fico® Xpress for finding solutions. As Xpress is a general-purpose solver, it does not utilize the special structure of the adjusted timetabling and line planning problems. As a result, only instances of moderate size could be solved, and solution times were high. Therefore, further research should deal with the development of suitable solution algorithms or the adjustment of special purpose solvers to solve larger instances and reduce computation times.

For the timetabling problem with integrated route choice, it may be possible to extend existing methods, such as the modulo network simplex (Nachtigall and Opitz, 2008). An example of an extension to the modulo network simplex was recently proposed in Löbel et al. (2019) to consider passenger routes on a shortest path. The structure of their approach should allow a generalization to also consider a passenger distribution on multiple routes during optimization.

Similarly, for the line planning model with integrated mode and route choice, more advanced solution techniques should be investigated. The objective of the model is to find profit-maximizing line plans, which facilitates that completely different solutions are rated very similar or even the same. Hence, specialized solution methods should, unlike many existing ones, not rely on the similarity of solutions or use the concept of neighborhoods. Furthermore, this motivates that a solution method should ideally provide multiple solutions of high quality. This would allow operators to choose from similarly profitable solutions with potentially different modal splits.

Integrating further steps. In this research, we focus on *horizontal* integration, that is, the integration of models from both demand-oriented and supply-oriented approaches. The literature provides approaches for *vertical* integration, that means, integration of several supply-oriented models (Schöbel, 2017), or of several demand-oriented models (Dugge, 2006). Considering multiple steps can be beneficial for the quality of the solution. In most cases, multiple steps of the demand-oriented and supply-oriented approaches are dependent on each other. Hence, it seems reasonable to develop models integrating several steps of both demand-oriented and supply-oriented approaches. For example, a next step could be to develop integrated line planning and timetabling approaches with simultaneous mode and route choice estimation.

Robustness. Public transport operations are subject to disruptions that propagate through the network with potentially significant delays for a large group of passengers. One limitation of this thesis is that the presented models only consider the nominal planning case and disregard the performance of the found services under realistic circumstances with unforeseen events. To minimize the propagation of disruptions and delays, a stream of research focuses on finding robust public transport services (Ahuja et al., 2009). When considering a passenger distribution on multiple routes during line planning or timetabling, the corresponding solutions at least provide multiple passenger routes which might be beneficial in case of disruptions. However, the operations are not designed to be robust, with possibly severe consequences for operators. Hence, the aspect of robustness should be considered when designing public transport services, also in models that focus on estimating passenger choices.

Practical applicability. The research in this thesis aims to provide insights for practitioners on the extent to which integrating passenger demand estimation into public transport optimization models improves the quality of designed services. While we think that the models adequately consider the relevant requirements for strategical and tactical planning, they are not capable of providing solutions that can be implemented for operation. The presented models are limited by the level of detail considered. For example, the found timetables and line plans might conflict with the available platforms in stations or tracks between intersections. This should be investigated in coordination with practitioners.

Generalization of rolling horizon algorithm. The rolling horizon algorithm was developed to solve a vehicle scheduling problem for on-demand vehicles as it occurs in the context of travel demand modeling. We provided an optimality guarantee for the found solution if the horizons overlap sufficiently. The proof for this finding mainly uses the structure of the cycle free directed graph and that the decisions taken in one horizon do not influence the remote future. Many applications can be modeled similarly, for example, many problems relying on time-space networks. Hence, it would be interesting to investigate whether this solution approach including optimality guarantee is also applicable to a wider set of problems with a similar structure. Similar to the presented case, this solution procedure could help to solve extremely large instances or decrease solution times for large instances. It also might be very valuable in the case of online optimization, where information is usually revealed successively. There, the optimality guarantee would ensure the quality of solutions of a rolling horizon approach, even if not yet all information is available.

References

Ahuja, R., T. Magnanti, and J. Orlin (1988). “Network flows”.

Ahuja, R., R. Möhring, and C. Zaroliagis (2009). *Robust and online large-scale optimization: models and techniques for transportation systems*. Vol. 5868. Springer.

Azevedo, C., K. Marczuk, S. Raveau, H. Soh, M. Adnan, K. Basak, H. Loganathan, N. Deshmunkh, D.-H. Lee, E. Frazzoli, and M. Ben-Akiva (2016). “Microsimulation of Demand and Supply of Autonomous Mobility On-Demand”. *Transportation Research Record: Journal of the Transportation Research Board* 2564.1, pp. 21–30.

El-Azm, A. (1985). “The minimum fleet size problem and its applications to bus scheduling”. *Computer scheduling of public transport 2*, pp. 493–512.

Bekhor, S., M. Ben-Akiva, and M. Ramming (2006). “Evaluation of choice set generation algorithms for route choice models”. *Annals of Operations Research* 144.1, pp. 235–247.

Ben-Akiva, M. and S. Lerman (1985). *Discrete choice analysis: theory and application to travel demand*. Vol. 9. MIT press.

Bertsimas, D., Y. Ng, and J. Yan (2021). “Data-Driven Transit Network Design at Scale”. *Operations Research*.

Bischoff, J., M. Maciejewski, and K. Nagel (2017). “City-wide shared taxis: A simulation study in Berlin”. *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, pp. 275–280.

Bodin, L. (1983). “Routing and scheduling of vehicles and crews, the state of the art”. *Computers & Operations Research* 10.2, pp. 63–211.

Boesch, P., F. Ciari, and K. Axhausen (2016). “Autonomous vehicle fleet sizes required to serve different levels of demand”. *Transportation Research Record* 2542.1, pp. 111–119.

Borg, I. and P. Groenen (2005). *Modern multidimensional scaling: Theory and applications*. Springer Science & Business Media.

Borndörfer, R., M. Grötschel, and M. Pfetsch (2007). “A column-generation approach to line planning in public transport”. *Transportation Science* 41.1, pp. 123–132.

Borndörfer, R., H. Hoppmann, and M. Karbstein (2017). “Passenger routing for periodic timetable optimization”. *Public Transport* 9.1-2, pp. 115–135.

Borndörfer, R. and M. Karbstein (2012). “A direct connection approach to integrated line planning and passenger routing”. *OASIcs-OpenAccess Series in Informatics*. Vol. 25. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.

Borndörfer, R., T. Klug, L. Lamorgese, C. Mannino, M. Reuther, and T. Schlechte (2018). *Handbook of Optimization in the Railway Industry*. Vol. 268. Springer.

Bradley, M. and H. Gunn (1990). “Stated preference analysis of values of travel time in the Netherlands”. *Transportation Research Record* 1285.

Bunte, S. and N. Kliewer (2009). “An overview on vehicle scheduling models”. *Public Transport* 1.4, pp. 299–317.

Bussieck, M. (1998). “Optimal lines in public rail transport”. PhD thesis. Citeseer.

Bussieck, M., P. Kreuzer, and U. Zimmermann (1997). “Optimal lines for railway systems”. *European Journal of Operational Research* 96.1, pp. 54–63.

Caimi, G., L. Kroon, and C. Liebchen (2017). “Models for railway timetable optimization: Applicability and applications in practice”. *Journal of Rail Transport Planning & Management* 6.4, pp. 285–312.

Canca, D., A. De-Los-Santos, G. Laporte, and J. Mesa (2016). “A general rapid network design, line planning and fleet investment integrated model”. *Annals of Operations Research* 246.1, pp. 127–144.

Canca, D., A. De-Los-Santos, G. Laporte, and J. Mesa (2017). “An adaptive neighborhood search metaheuristic for the integrated railway rapid transit network design and line planning problem”. *Computers & Operations Research* 78, pp. 1–14.

Canca, D., A. De-Los-Santos, G. Laporte, and J. Mesa (2019). “Integrated Railway Rapid Transit Network Design and Line Planning problem with maximum profit”. *Transportation Research Part E: Logistics and Transportation Review* 127, pp. 1–30.

Carosi, S., A. Frangioni, L. Galli, L. Girardi, and G. Vallese (2019). “A matheuristic for integrated timetabling and vehicle scheduling”. *Transportation Research Part B: Methodological* 127, pp. 99–124.

Cascetta, E. (2013). *Transportation systems engineering: theory and methods*. Vol. 49. Springer Science & Business Media.

Claessens, M., N. van Dijk, and P. Zwaneveld (1998). “Cost optimal allocation of rail passenger lines”. *European Journal of Operational Research* 110.3, pp. 474–489.

Corman, F., A. D'Ariano, A. Marra, D. Pacciarelli, and M. Samà (2017). "Integrating train scheduling and delay management in real-time railway traffic control". *Transportation Research Part E: Logistics and Transportation Review* 105, pp. 213–239.

De Keizer, B., K. Geurs, and G. Haarsman (2012). "Interchanges in timetable design of railways: A closer look at customer resistance to interchange between trains". *Proceedings of the European Transport Conference, Glasgow*, pp. 8–10.

De Dios Ortúzar, J. and L. Willumsen (2011). *Modelling transport.* John Wiley & Sons.

De Jong, G., A. Daly, M. Pieters, and T. van der Hoorn (2007). "The logsum as an evaluation measure: review of the literature and new results". *Transportation Research Part A: Policy and Practice* 41.9, pp. 874–889.

Dell'Olio, L., A. Ibeas, and P. Cecín (2010). "Modelling user perception of bus transit quality". *Transport Policy* 17.6, pp. 388–397.

Desfontaines, L. and G. Desaulniers (2018). "Multiple depot vehicle scheduling with controlled trip shifting". *Transportation Research Part B: Methodological* 113, pp. 34–53.

Dugge, B. (2006). "Ein simultanes Erzeugungs-, Verteilungs-, Aufteilungs- und Routenwahlmodell". PhD thesis. Technische Universität Dresden.

Engelhardt, R., F. Dandl, and K. Bogenberger (2020). *Speed-up Heuristic for an On-Demand Ride-Pooling Algorithm: Presented at Informs Annual Meeting 2019.*

Espinosa-Aranda, J.L., R. García-Ródenas, M.L. López-García, and E. Angulo (2018). "Constrained nested logit model: formulation and estimation". *Transportation* 45.5, pp. 1523–1557.

Fagnant, D. and K. Kockelman (2018). "Dynamic ride-sharing and fleet sizing for a system of shared autonomous vehicles in Austin, Texas". *Transportation* 45.1, pp. 143–158.

Feigon, S. and C. Murphy (2016). *Shared mobility and the transformation of public transit.* TCRP research report 188. Washington, DC: The National Academies Press.

FGSV (2008). *Richtlinien für integrierte Netzgestaltung (RIN).* (in German). Forschungsgesellschaft für Straßen-und Verkehrswesen, Köln.

Flyvbjerg, B., K. Kahr, P. Petersen, and J. Vibe-Petersen (1986). "Evaluation of public transport: method for application in open planning". *Transportation* 13.1, pp. 23–52.

Foster, B. and D. Ryan (1976). “An integer programming approach to the vehicle scheduling problem”. *Journal of the Operational Research Society* 27.2, pp. 367–384.

Friedrich, M., M. Hartl, and C. Magg (2018). “A modeling approach for matching ridesharing trips within macroscopic travel demand models”. *Transportation* 45.6, pp. 1639–1653.

Friedrich, M., M. Hartl, A. Schiewe, and A. Schöbel (2017a). “Angebotsplanung im öffentlichen Verkehr–planerische und algorithmische Lösungen”. *Heureka’17* 9.

Friedrich, M., M. Hartl, A. Schiewe, and A. Schöbel (2017b). “Integrating Passengers’ Assignment in Cost-Optimal Line Planning”. *17th workshop on algorithmic approaches for transportation modelling, optimization, and systems (atmos 2017)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.

Friedrich, M., I. Hofsaess, and S. Wekeck (2001). “Timetable-based transit assignment using branch and bound techniques”. *Transportation Research Record: Journal of the Transportation Research Board* 1752, pp. 100–107.

Friedrich, M., F. Leurent, I. Jackiva, V. Fini, and S. Raveau (2016). “From Transit Systems to Models: Purpose of Modelling”. *Modelling Public Transport Passenger Flows in the Era of Intelligent Transport Systems: COST Action TU1004 (TransITS)*. Ed. by G. Gentile and K. Noekel. Cham: Springer International Publishing, pp. 131–234.

Garey, M. and D. Johnson (1979). *Computers and intractability: a guide to NP-completeness*. Vol. 174. WH Freeman and Company, San Francisco.

Gattermann, P., P. Großmann, K. Nachtigall, and A. Schöbel (2016). “Integrating Passengers’ Routes in Periodic Timetabling: A SAT approach”. *OASIcs-OpenAccess Series in Informatics*. Vol. 54. Schloss Dagstuhl-Leibniz-Zentrum für Informatik.

Gattermann, P., J. Harbering, and A. Schöbel (2017). “Line pool generation”. *Public Transport* 9.1, pp. 7–32.

Goerigk, M. and M. Schmidt (2017). “Line planning with user-optimal route choice”. *European Journal of Operational Research* 259.2, pp. 424–436.

Goossens, J.-W., S. van Hoesel, and L. Kroon (2006). “On solving multi-type railway line planning problems”. *European Journal of Operational Research* 168.2, pp. 403–424.

Grey, A. (1978). “The generalised cost dilemma”. *Transportation* 7.3, pp. 261–280.

Guan, J., H. Yang, and S. Wirasinghe (2006). “Simultaneous optimization of transit line configuration and passenger line assignment”. *Transportation Research Part B: Methodological* 40.10, pp. 885–902.

Guis, N. and S. Nijënstein (2015). “Modelleren van klantvoorkeuren in dienstregelingsstudies (in Dutch)”. *Colloquium Vervoersplanologisch Speurwerk. Antwerpen: NS.*

Haase, K. and S. Müller (2014). “A comparison of linear reformulations for multinomial logit choice probabilities in facility location models”. *European Journal of Operational Research* 232.3, pp. 689–691.

Hartleb, J., M. Friedrich, and E. Richter (2021a). “Umlaufbildung für On Demand-Fahrzeugflotten in makroskopischen Nachfragemodellen”. *HEUREKA’21: Optimierung in Verkehr und Transport.*

Hartleb, J., M. Friedrich, E. Richter, and C. Canales Martinez (2020). *Vehicle Scheduling for On-demand Vehicle Fleets.* Version v1.0.0. URL: `https://github.com/jhartleb/VehicleScheduling/tree/v1.0.0`.

Hartleb, J. and M. Schmidt (2019). “Railway timetabling with integrated passenger distribution”. *Available at SSRN 3505167.*

Hartleb, J. and M. Schmidt (2020). “A Rolling Horizon Heuristic with Optimality Guarantee for an On-Demand Vehicle Scheduling Problem”. *20th Symposium on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS 2020).* Schloss Dagstuhl-Leibniz-Zentrum für Informatik.

Hartleb, J., M. Schmidt, M. Friedrich, and D. Huisman (2019). “A good or a bad timetable: Do different evaluation functions agree?” *ERIM Report Series.*

Hartleb, J., M. Schmidt, D. Huisman, and M. Friedrich (2021b). “Modeling and solving line planning with integrated mode choice”. *Available at SSRN 3849985.*

Heilig, M., T. Hilgert, N. Mallig, M. Kagerbauer, and P. Vortisch (2017). “Potentials of autonomous vehicles in a changing private transportation system - a case study in the Stuttgart region”. *Transportation research procedia* 26, pp. 13–21.

Heilig, M., N. Mallig, O. Schröder, M. Kagerbauer, and P. Vortisch (2018). “Implementation of free-floating and station-based carsharing in an agent-based travel demand model”. *Travel Behaviour and Society* 12, pp. 151–158.

Hensher, D. and K. Button (2007). *Handbook of transport modelling.* Emerald Group Publishing Limited.

Hörl, S. (2017). “Agent-based simulation of autonomous taxi services with dynamic demand responses”. *Procedia Computer Science* 109, pp. 899–904.

Horni, A., K. Nagel, and K.W. Axhausen, eds. (2016). *The Multi-Agent Transport Simulation MATSim.* London: Ubiquity Press.

Huisman, D., L. Kroon, R. Lentink, and M. Vromans (2005). “Operations research in passenger railway transportation”. *Statistica Neerlandica* 59.4, pp. 467–497.

infas, DLR, IVT, and infas 360 (2017). *Mobilität in Deutschland - MiD 2017 (im Auftrag des BMVI)*. URL: `http://www.mobilitaet-in-deutschland.de/MiT2017.html` (visited on July 5, 2019).

Kanai, S., K. Shiina, S. Harada, and N. Tomii (2011). "An optimal delay management algorithm from passengers' viewpoints considering the whole railway network". *Journal of Rail Transport Planning & Management* 1.1, pp. 25–37.

Kaspi, M. and T. Raviv (2013). "Service-oriented line planning and timetabling for passenger trains". *Transportation Science* 47.3, pp. 295–311.

Klier, M. and K. Haase (2015). "Urban public transit network optimization with flexible demand". *OR Spectrum* 37.1, pp. 195–215.

Kroon, L., D. Huisman, E. Abbink, P.-J. Fioole, M. Fischetti, G. Maróti, A. Schrijver, A. Steenbeek, and R. Ybema (2009). "The new Dutch timetable: The OR revolution". *Interfaces* 39.1, pp. 6–17.

Kusakabe, T., T. Iryo, and Y. Asakura (2010). "Estimation method for railway passengers' train choice behavior with smart card transaction data". *Transportation* 37.5, pp. 731–749.

Lam, A., Y.-W. Leung, and X. Chu (2016). "Autonomous-vehicle public transportation system: scheduling and admission control". *IEEE Transactions on Intelligent Transportation Systems* 17.5, pp. 1210–1226.

Lan, Z., S. He, R. Song, and S. Hao (2019). "Optimizing vehicle scheduling based on variable timetable by benders-and-price approach". *Journal of Advanced Transportation* 2019.

Laporte, G., Á. Marín, J. Mesa, and F. Ortega (2007). "An integrated methodology for the rapid transit network design problem". *Algorithmic methods for railway optimization*. Springer, pp. 187–199.

Laporte, G., J. Mesa, F. Ortega, and I. Sevillano (2005). "Maximizing trip coverage in the location of a single rapid transit alignment". *Annals of Operations Research* 136.1, pp. 49–63.

Liebchen, C. (2018). "On the Benefit of Preprocessing and Heuristics for Periodic Timetabling". *Operations Research Proceedings 2017*. Springer, pp. 709–714.

Liebchen, C. and R. Möhring (2007). "The modeling power of the periodic event scheduling problem: railway timetables - and beyond". *Algorithmic methods for railway optimization*. Springer, pp. 3–40.

Liebchen, C. and L. Peeters (2009). "Integral cycle bases for cyclic timetabling". *Discrete Optimization* 6.1, pp. 98–109.

Lin, Y., W. Li, F. Qiu, and H. Xu (2012). "Research on optimization of vehicle routing problem for ride-sharing taxi". *Procedia-Social and Behavioral Sciences* 43, pp. 494–502.

Liu, D., P. Vansteenwegen, G. Lu, and Q. Peng (2019). "An iterative approach for profit-oriented railway line planning". *Linköping University Electronic Press*, pp. 806–825.

Liu, T. and A. Ceder (2020). "Battery-electric transit vehicle scheduling with optimal number of stationary chargers". *Transportation Research Part C: Emerging Technologies* 114, pp. 118–139.

Löbel, F., N. Lindner, and R. Borndörfer (2019). *The Restricted Modulo Network Simplex Method for Integrated Periodic Timetabling and Passenger Routing.* eng. Tech. rep. 19-36. Takustr. 7, 14195 Berlin: ZIB.

De-Los-Santos, A., G. Laporte, J. Mesa, and F. Perea (2017). "The railway line frequency and size setting problem". *Public Transport* 9.1-2, pp. 33–53.

Maciejewski, M. (2016). "Dynamic Transport Services". *The Multi-Agent Transport Simulation MATSim.* London: Ubiquity Press, pp. 145–152.

Maciejewski, M., J. Bischoff, S. Hörl, and K. Nagel (2017). "Towards a Testbed for Dynamic Vehicle Routing Algorithms". *Highlights of Practical Applications of Cyber-Physical Multi-Agent Systems: International Workshops of PAAMS 2017, Proceedings.* Vol. 722. Cham and s.l.: Springer International Publishing, pp. 69–79.

Mackie, P., S. Jara-Dıaz, and A. Fowkes (2001). "The value of travel time savings in evaluation". *Transportation Research Part E: Logistics and Transportation Review* 37.2-3, pp. 91–106.

Marczuk, K., H. Hong, C. Azevedo, M. Adnan, S. Pendleton, E. Frazzoli, and D.-H. Lee (2015). "Autonomous mobility on demand in SimMobility: Case study of the central business district in Singapore". *2015 IEEE 7th International Conference on Cybernetics and Intelligent Systems (CIS 2015) and IEEE Conference on Robotics, Automation and Mechatronics (RAM 2015).* IEEE, pp. 167–172.

Martínez, L., G. Correia, F. Moura, and M. Mendes Lopes (2017). "Insights into carsharing demand dynamics: Outputs of an agent-based model application to Lisbon, Portugal". *International Journal of Sustainable Transportation* 11.2, pp. 148–159.

McNally, M. (2010). "The Four-Step Model". *Handbook of transport modelling.* Ed. by D.A. Hensher. Vol. 1. Handbooks in transport. Amsterdam: Elsevier, pp. 35–53.

Moeckel, R., R. Donnelly, and J. Ji (2019). "Statewide Transportation Models in the U.S.: A Review of the State of Practice". *Transportation Research Board 98th Annual Meeting.*

Nachtigall, K. (1994). *A branch and cut approach for periodic network programming.* Tech. rep. 29. Hildesheimer Informatik-Berichte.

Nachtigall, K. (1998). "Periodic network optimization and fixed interval timetables". *Habilitationsschrift, Deutsches Zentrum für Luft- und Raumfahrt, Institut für Flugführung, Braunschweig, Germany.*

Nachtigall, K. and K. Jerosch (2008). "Simultaneous network line planning and traffic assignment". *8th Workshop on Algorithmic Approaches for Transportation Modeling, Optimization, and Systems (ATMOS'08).* Schloss Dagstuhl-Leibniz-Zentrum für Informatik.

Nachtigall, K. and J. Opitz (2008). "Solving periodic timetable optimisation problems by modulo simplex calculations". *8th Workshop on Algorithmic Approaches for Transportation Modeling, Optimization, and Systems (ATMOS'08).* Schloss Dagstuhl-Leibniz-Zentrum für Informatik.

Odijk, M. (1996). "A constraint generation algorithm for the construction of periodic railway timetables". *Transportation Research Part B: Methodological* 30.6, pp. 455–464.

Pacheco, M., S. Sharif Azadeh, and M. Bierlaire (2016). "A new mathematical representation of demand using choice-based optimization method". *16th Swiss Transport Research Conference*. EPFL-CONF-218673.

Parbo, J., O. Nielsen, and C. Prato (2014). "User perspectives in public transport timetable optimisation". *Transportation Research Part C: Emerging Technologies* 48, pp. 269–284.

Parbo, J., O. Nielsen, and C. Prato (2016). "Passenger perspectives in railway timetabling: a literature review". *Transport Reviews* 36.4, pp. 500–526.

Paris Agreement (Dec. 2015). UN Treaty [Treaty No. XXVII-7-d]. United Nations. URL: `https://treaties.un.org/pages/ViewDetails.aspx?src=TREATY&mtdsg_no=XXVII-7-d&chapter=27&clang=_en` (visited on Apr. 5, 2021).

Pätzold, J. and A. Schöbel (2016). "A matching approach for periodic timetabling". *16th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS 2016).* Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.

Polinder, G.-J., M. Schmidt, and D. Huisman (2019). "A new approach to periodic railway timetabling". *Rail Norrköping 2019. 8th International Conference on Railway Operations Modelling and Analysis (ICROMA), Norrköping, Sweden, June 17th–20th, 2019.* 069. Linköping University Electronic Press, pp. 998–1014.

Polinder, G.-J., M. Schmidt, and D. Huisman (2020). *Timetabling for strategic passenger railway planning*. Tech. rep.

PTV Group (2018). *PTV Visum 18 Online Help*. English. URL: https://cgi.ptvgroup.com/vision-help/VISUM_18_ENG/ (visited on Nov. 12, 2018).

PTV Group (2020). *PTV Visum 2020 Online Help*. English. see demand responsive transport. URL: https://cgi.ptvgroup.com/vision-help/VISUM_2020_ENG/ (visited on Nov. 24, 2020).

Richter, E., M. Friedrich, A. Migl, and J. Hartleb (2019). "Integrating ridesharing services with automated vehicles into macroscopic travel demand models". *MT-ITS 2019*. IEEE.

Rieser, N., B. Tasnády, M. Friedrich, E. Pestel, N. Vries, M. Rothenfluh, and R. Fischer (2018). *Quality assurance for transport models and their applications*. Accessed April 9, 2021. URL: https://www.mobilityplatform.ch/fileadmin/mobilityplatform/normenpool/21728_1645_Inhalt.pdf.

Ritchie, H. and M. Roser (2020). "CO_2 and greenhouse gas emissions". *Our world in data*. URL: https://ourworldindata.org/co2-and-other-greenhouse-gas-emissions (visited on Apr. 5, 2021).

Robenek, T., Y. Maknoon, S. Sharif Azadeh, J. Chen, and M. Bierlaire (2016). "Passenger centric train timetabling problem". *Transportation Research Part B: Methodological* 89, pp. 107–126.

Rogge, M., E. van der Hurk, A. Larsen, and D. Sauer (2018). "Electric bus fleet size and mix problem with optimization of charging infrastructure". *Applied Energy* 211, pp. 282–295.

Scheepmaker, G., R. Goverde, and L. Kroon (2017). "Review of energy-efficient train control and timetabling". *European Journal of Operational Research* 257.2, pp. 355–376.

Schiewe, A., P. Schiewe, and M. Schmidt (2019). "The line planning routing game". *European Journal of Operational Research* 274.2, pp. 560–573.

Schiewe, P. and A. Schöbel (2020). "Periodic Timetabling with Integrated Routing: Toward Applicable Approaches". *Transportation Science* 54.6, pp. 1714–1731.

Schittenhelm, B. (2013). "Quantitative Methods for Assessment of Railway Timetables". PhD thesis. Technical University of Denmark, Transport.

Schmidt, M. (2014). *Integrating routing decisions in public transportation problems*. Springer.

Schmidt, M. and A. Schöbel (2015a). "The complexity of integrating passenger routing decisions in public transportation models". *Networks* 65.3, pp. 228–243.

Schmidt, M. and A. Schöbel (2015b). “Timetabling with passenger routing”. *OR Spectrum* 37.1, pp. 75–97.

Schöbel, A. (2012). “Line planning in public transportation: models and methods”. *OR Spectrum* 34.3, pp. 491–510.

Schöbel, A. (2017). “An eigenmodel for iterative line planning, timetabling and vehicle scheduling in public transportation”. *Transportation Research Part C: Emerging Technologies* 74, pp. 348–365.

Schöbel, A. and S. Scholl (2006). “Line planning with minimal traveling time”. *5th Workshop on Algorithmic Methods and Models for Optimization of Railways (ATMOS’05).* Schloss Dagstuhl-Leibniz-Zentrum für Informatik.

Schrijver, A. (2003). *Combinatorial optimization: polyhedra and efficiency.* Vol. 24. Springer Science & Business Media.

Schrijver, A. and A. Steenbeek (1994). “Dienstregelingontwikkeling voor Railned (timetable construction for Railned), Rapport CADANS 1.0”. *CWI Center for Mathematics and Computer Science, Amsterdam. In Dutch.*

Sels, P., T. Dewilde, D. Cattrysse, and P. Vansteenwegen (2011). “Deriving all passenger flows in a railway network from ticket sales data”. *4th International Seminar on Railway Operations Modelling and Analysis (IAROR) : RailRome 2011, Proceedings.* Rome, Italy: Ghent University, Department of Industrial Management.

Sels, P., K. Meisch, T. Moller, J. Parbo, T. Dewilde, D. Cattrysse, and P. Vansteenwegen (2015). “Towards a Better Train Timetable for Denmark Reducing Total Expected Passenger Time”. *Proceedings of the 13th Conference on Advanced Systems in Public Transport (CASPT).*

Serafini, P. and W. Ukovich (1989). “A mathematical model for periodic scheduling problems”. *SIAM Journal on Discrete Mathematics* 2.4, pp. 550–581.

Siebert, M. and M. Goerigk (2013). “An experimental comparison of periodic timetabling models”. *Computers & Operations Research* 40.10, pp. 2251–2259.

Steiner, K. and S. Irnich (2018). “Schedule-based integrated intercity bus line planning via branch-and-cut”. *Transportation Science* 52.4, pp. 882–897.

Swait, J. and J. Louviere (1993). “The role of the scale parameter in the estimation and comparison of multinomial logit models”. *Journal of marketing research*, pp. 305–314.

Train, K.E. (2009). *Discrete choice methods with simulation.* Cambridge university press.

Van der Hurk, E., L. Kroon, G. Maróti, and P. Vervest (2014). "Deduction of passengers' route choices from smart card data". *IEEE Transactions on Intelligent Transportation Systems* 16.1, pp. 430–440.

Vosooghi, R., J. Puchinger, M. Jankovic, and G. Sirin (2017). "A critical analysis of travel demand estimation for new one-way carsharing systems". *IEEE ITSC 2017*. Piscataway, NJ: IEEE, pp. 199–205.

Wang, B., S. Ordonez Medina, and P. Fourie (2018). "Simulation of Autonomous Transit On Demand for Fleet Size and Deployment Strategy Optimization". *Procedia Computer Science* 130, pp. 797–802.

Wardman, M. (2004). "Public transport values of time". *Transport policy* 11.4, pp. 363–377.

Wardman, M., P. Chintakayala, G. de Jong, and D. Ferrer (2012). "European wide meta-analysis of values of travel time". *ITS, University of Leeds, Paper prepared for EIB.*

Wardman, M. and J. Toner (2018). "Is generalised cost justified in travel demand analysis?" *Transportation*, pp. 1–34.

Wen, C.-H., W.-C. Wang, and C. Fu (2012). "Latent class nested logit model for analyzing high-speed rail access mode choice". *Transportation Research Part E: Logistics and Transportation Review* 48.2, pp. 545–554.

Wen, M., E. Linde, S. Ropke, P. Mirchandani, and A. Larsen (2016). "An adaptive large neighborhood search heuristic for the electric vehicle scheduling problem". *Computers & Operations Research* 76, pp. 73–83.

Wenzelburger, P. and F. Allgöwer (2020). *A first step towards an autonomously driving E-Scooter: Presented at 21. IFAC World Congress.* URL: `https://www.ist.uni-stuttgart.de/institute/team/pdf/PW/IFAC20_E-Scooter.pdf` (visited on Nov. 24, 2020).

Wilkes, G., L. Briem, M. Heilig, T. Hilgert, M. Kagerbauer, and P. Vortisch (2019). "Identifying service provider and transport system related effects of different ridesourcing service schemes through simulation within the travel demand model mobiTopp". *International Conference on Mobility as a Service (iCoMaaS).*

Williamson, D. and D. Shmoys (2011). *The design of approximation algorithms.* Cambridge university press.

Yang, L. and W. Lam (2006). "Probit-type reliability-based transit network assignment". *Transportation Research Record: Journal of the Transportation Research Board* 1977, pp. 154–163.

Zhang, W., S. Guhathakurta, J. Fang, and G. Zhang (2015). “The performance and benefits of a shared autonomous vehicles based dynamic ridesharing system: An agent-based simulation approach”. *Transportation Research Board 94th Annual Meeting.*

Summary

Public transport services are important to our society for several reasons. They provide mobility for the public, constitute an efficient transport system in crowded places, and help reduce traffic emissions compared to other modes of transport. To make the most of the benefits of public transport, the services must be an attractive alternative for passengers. This requires the design of efficient and high-quality services and their continuous improvement. In particular, public transport services should provide accessible service with reasonably short travel times and as few and short transfers as possible.

For the design of public transport, the literature provides numerous demand-oriented and supply-oriented approaches. Demand-oriented approaches estimate the travel demand for a certain transport scenarios and allow a thorough evaluation of the supply. This is particularly useful when comparing a number of transport services and assessing their strength and weaknesses. Supply-oriented approaches design a public transport service for a given demand situation. The input of demand-oriented approaches is the output of supply-oriented approaches and vice versa. Although the interdependence of supply and demand is known, only a few and basic combinations of these approaches have been developed. This thesis examines approaches that integrate public transport supply and demand.

In Part I of this thesis, we investigate the potential of integrated demand-oriented and supply-oriented models by incorporating passenger demand estimation into public transport service design. In Chapter 2, we compare timetabling evaluation functions designed to measure the quality of timetables from the perspective of passengers and examine how consistent their evaluation results are. In this comparison, we find that the design of an evaluation function can have a significant impact on which timetable is considered optimal. Furthermore, we identify which components impact the

results of the evaluation most. In Chapter 3, we develop two timetabling models for finding travel time-optimal timetables while estimating the passenger distribution on different routes. The passenger distribution is approximated with the logit model, a discrete choice model commonly applied to estimate passengers' route choices. We provide two linear representations for the logit model to integrate it into an optimization framework. Our comparison with four state-of-the-art timetabling methods shows that integrating a passenger distribution model has the potential to find better timetables for passengers, but more efficient solution strategies have to be developed. In Chapter 4, we provide a model to find profit-optimal line plans while estimating passengers' mode and route choices. By suitable preprocessing of the utilities for the passengers' mode and route decisions, the choice models can be linearized and commercial solvers can be used to find solutions. In experiments, we show that estimating the mode choice of travelers during optimization yields line plans with higher profits for operators and higher service levels for passengers. Furthermore, our experiments suggest that operators should regularly react to changes in travel demand and update their line plans.

In contrast to scheduled public transport, recent technological developments enable the implementation of more flexible public transport services: large-scale and affordable mobility on demand. The operations of these services are planned online and do not follow fixed lines or timetables. As on-demand services are not yet operating on a large scale, the impact of these services on cities, traffic, and the environment is difficult to assess. To be able to estimate the consequences of the implementation of on-demand services, public transport authorities investigate and evaluate several potential transport scenarios using travel demand models. This leads to the following problem: On the one hand, travel demand models are demand-oriented approaches that estimate passenger demand and require service levels of the supply as input. On the other hand, the service level of on-demand services depends on passenger demand and cannot be determined until the demand is known. Hence, to be able to consider on-demand services in travel demand models, the models have to be extended.

In Part II of this thesis, we investigate how to estimate the service level of on-demand services within a travel demand model. We use vehicle scheduling to determine the required vehicle fleet size of on-demand services. Our models also provide insight into the vehicle distance traveled and the vehicle location over time. This allows for a thorough examination of the impact of on-demand services on traffic and the environment. In Chapter 5, we provide a simple heuristic to find solutions to the vehicle scheduling problem of on-demand vehicles in short computation times. It is capable of handling non-integer demand values and extremely large instance sizes, both of which are characteristics of the application. We illustrate in two case studies how the presented algorithm can be applied to estimate the service level of on-demand services within macroscopic travel demand models. The algorithm presented in Chapter 5 is able to find good solutions in short computation times but it does not provide a guarantee on the solution quality. It is not known whether the found solutions are optimal or how far they are from an optimal solution. In Chapter 6, we present a solution algorithm for the vehicle scheduling problem of on-demand services that is capable of finding optimal solutions. This solution algorithm is based on a rolling horizon framework and we construct a solution by composing the solutions for the individual horizons. By overlapping the horizons, it is possible to look ahead to the demand of the next horizon and improve the solution quality. For a sufficiently large overlap of the horizons, we show that the composed solution is optimal for the whole problem. In experiments, we show that this approach is suitable to solve very large instances to optimality and brings a speed-up for large instances compared to a comprehensive approach.

In summary, we develop and investigate approaches that integrate demand-oriented into supply-oriented models. We show that the integration is possible and that the found solutions provide a higher service level to the passengers. However, the integrated models come at the cost of higher complexity and further research should address specialized solution approaches.

Samenvatting

Er zijn meerdere redenen waarom openbaar vervoer (ov) belangrijk is voor onze maatschappij. Ov zorgt ervoor dat mensen mobiel zijn en helpt vervoersemissies te verminderen in vergelijking met andere vervoersmiddelen. Daarnaast draagt het ov bij een efficiënt vervoerssysteem in drukke gebieden. Het ov moet een aantrekkelijk alternatief zijn voor reizigers. Dit vereist de inrichting van efficiënte en hoogwaardige voorzieningen en een continue verbetering van deze voorzieningen. Ov-systemen hebben als belangrijkste taak het bieden van toegankelijke vervoer met redelijk korte reistijden, zo min mogelijk overstappen en zo kort mogelijke overstaptijden.

Voor het inrichten van het openbaar vervoer biedt de literatuur talrijke vraaggerichte en aanbodgerichte benaderingen. Vraaggerichte benaderingen schatten de reisvraag voor een bepaald vervoersscenario in en maken een grondige evaluatie van het aanbod mogelijk. Dit is vooral handig bij het vergelijken van een aantal vervoerssystemen en de beoordeling van hun sterke en zwakke punten. Aanbodgerichte benaderingen ontwerpen een ov-systeem voor een gegeven vraagsituatie. De input van vraaggerichte benaderingen is de output van aanbodgerichte benaderingen en vice versa. Ook al is het bekend dat vraag en aanbod van elkaar afhankelijk zijn, er zijn slechts enkele basale combinaties van deze benaderingen ontwikkeld. In deze thesis worden benaderingen onderzocht die vraag en aanbod binnen het openbaar vervoer integreren.

In deel I van deze thesis onderzoeken we het potentieel van geïntegreerde vraaggerichte en aanbodgerichte modellen door een inschatting van de reizigersvraag in het ontwerp van het ov-aanbod op te nemen. In hoofdstuk 2 vergelijken we evaluatiefuncties voor dienstregelingen die ontworpen zijn om de kwaliteit van dienstregelingen te meten vanuit het perspectief van reizigers, en onderzoeken we in hoeverre de evaluatieresultaten hiervan overeenkomen. Door deze vergelijking ontdekken we dat het ontwerp van een evaluatiefunctie een significant effect kan hebben op welke

dienstregeling als beste wordt gezien. Vervolgens stellen we vast welke componenten de evaluatieresultaten het meest beïnvloeden. In hoofdstuk 3 ontwikkelen we twee dienstregelingmodellen om met het oog op reistijd optimale dienstregelingen te vinden en tegelijkertijd in te schatten wat de reizigersdistributie over de verschillende routes is. Er wordt een benadering van de reizigersdistributie gemaakt op basis van het logitmodel, een discreet keuzemodel dat vaak wordt toegepast om routekeuzes van reizigers in te schatten. We geven twee lineaire representaties voor het logitmodel om het in een optimalisatiekader te integreren. Onze vergelijking met vier moderne dienstregelingmethoden laat zien dat het meenemen van een reizigersdistributiemodel het potentieel heeft om betere dienstregelingen voor reizigers te vinden, maar ook dat er efficiëntere oplossingsstrategieën moeten worden ontwikkeld. In hoofdstuk 4 presenteren we een model voor het vinden van een optimale lijnvoering met als doel winstmaximalisatie en waarmee tegelijkertijd kan worden ingeschat welke vervoersmiddel- en routekeuze reizigers zullen maken. Door voorafgaand de voorzieningen voor de vervoersmiddel- en routekeuze van reizigers op een passende manier te verwerken, kunnen de keuzemodellen worden gelineariseerd en kunnen er met commerciële solvers oplossingen worden gevonden. Met behulp van experimenten tonen we aan dat het inschatten van de vervoersmiddelkeuze van reizigers tijdens de optimalisatie linvoeringen oplevert met grotere winsten voor vervoerders en hogere serviceniveaus voor reizigers. Ook komt uit onze experimenten naar voren dat vervoerders regelmatig op veranderingen in de reisvraag moeten inspelen en hun linvoering overeenkomstig moeten aanpassen.

Naast het openbaar vervoer met dienstregeling maken technologische ontwikkelingen de uitvoering van flexibelere ov-voorzieningen mogelijk: grootschalige en betaalbare mobiliteit on demand. De uitvoer van deze voorzieningen wordt online gepland en maakt geen gebruik van vaste lijnen of dienstregelingen. Omdat on-demand voorzieningen nog niet op grote schaal worden toegepast, is het moeilijk te beoordelen wat het effect is van deze voorzieningen op steden, het verkeer en het milieu. Om de gevolgen van het invoeren van on-demand voorzieningen te kunnen inschatten, onderzoeken en evalueren ov-autoriteiten meerdere potentiële vervoersscenario's met gebruik van reizigersvraagmodellen. Dit leidt tot het volgende probleem: aan de ene kant zijn reisvraagmodellen vraaggerichte benaderingen die de reizigersvraag inschatten en de voorzieningenniveaus van het aanbod als input vereisen. Aan de andere kant is het voorzieningenniveau van on-demand voorzieningen afhankelijk van de reizigersvraag en kan dit niet worden vastgesteld voordat de vraag bekend is.

Als we dus on-demand voorzieningen in reisvraagmodellen willen kunnen meenemen, moeten de modellen worden uitgebreid.

In deel II van deze thesis onderzoeken we hoe het voorzieningenniveau van on-demand voorzieningen kan worden ingeschat binnen een reisvraagmodel. We gebruiken voertuigplanning om de vereiste wagenparkgrootte voor on-demand voorzieningen te bepalen. Onze modellen geven ook inzicht in de afgelegde afstand en de locatie van een voertuig over een bepaalde periode. Dit maakt een grondige bestudering van het effect van on-demand voorzieningen op het verkeer en het milieu mogelijk. In hoofdstuk 5 presenteren we een simpele heuristiek om met korte rekentijden oplossingen te vinden voor het voertuigplanningsprobleem voor on-demand voertuigen. Hiermee is het mogelijk te rekenen met vraagwaarden die niet uit gehele getallen bestaan en met gevallen van extreem grote omvang; dit zijn beide kenmerken van de toepassing. We illustreren in twee casestudy's hoe het gepresenteerde algoritme kan worden toegepast om het voorzieningenniveau van on-demand voorzieningen in te schatten binnen macroscopische reisvraagmodellen. Het in hoofdstuk 5 gepresenteerde algoritme kan met korte rekentijden goede oplossingen vinden, maar het biedt geen garantie op de oplossingskwaliteit. Het is niet bekend of de gevonden oplossingen ook optimaal zijn of hoever ze zijn verwijderd van een optimale oplossing. In hoofdstuk 6 presenteren we een oplossingsalgoritme voor het voertuigplanningsprobleem van on-demand voorzieningen waarmee optimale oplossingen kunnen worden gevonden. Dit oplossingsalgoritme is gebaseerd op een rollende horizon en we construeren een oplossing door de oplossingen voor de individuele horizons samen te brengen. Door overlap van de horizons is het mogelijk om vooruit te kijken naar de vraag van de volgende horizon en de oplossingskwaliteit te verbeteren. Bij een overlap van de horizons die groot genoeg is, tonen we aan dat de samengestelde oplossing optimaal is voor het gehele probleem. Met behulp van experimenten tonen we aan dat deze benadering geschikt is om ook bij zeer grote instanties optimale oplossingen te vinden.

Samengevat ontwikkelen en onderzoeken we methoden die vraag en aanbodgerichte modellen integreren. We tonen aan dat de integratie mogelijk is en dat de gevonden oplossingen een hoger kwaliteit voor reizigers opleveren. De geïntegreerde modellen brengen echter ook een grotere complexiteit met zich mee en vervolgonderzoek zou zich moeten richten op gespecialiseerde oplossingsmethoden.

Zusammenfassung

Der öffentliche Verkehr ist aus mehreren Gründen wichtig für unsere Gesellschaft. Er sichert die Mobilität der Menschen und hilft, die Verkehrsemissionen im Vergleich zu anderen Verkehrsmitteln zu reduzieren. Der öffentliche Verkehr trägt auch in Ballungsräumen zu einem effizienten Verkehrssystem bei. Um die Vorteile des öffentlichen Verkehrs bestmöglich zu nutzen, müssen die öffentlichen Verkehrsangebote für die Fahrgäste eine attraktive Alternative zu anderen Verkehrsmitteln darstellen. Dies erfordert die Gestaltung von effizienten und qualitativ hochwertigen Angeboten und deren kontinuierliche Verbesserung. Insbesondere sollten öffentliche Verkehrssysteme ein zugängliches Angebot mit angemessen kurzen Fahrzeiten und möglichst wenigen und kurzen Umstiegen bieten.

Für die Gestaltung von öffentlichen Verkehrsangeboten bietet die Literatur zahlreiche nachfrageorientierte und angebotsorientierte Ansätze. Nachfrageorientierte Ansätze schätzen die Reisenachfrage für ein bestimmtes Verkehrsszenario und ermöglichen eine gründliche Bewertung des Angebots. Dies ist besonders nützlich, wenn mehrere Verkehrssysteme miteinander verglichen und ihre Stärken und Schwächen beurteilt werden sollen. Angebotsorientierte Ansätze entwerfen ein öffentliches Verkehrsangebot für eine gegebene Nachfragesituation. Eingangsgrößen für nachfrageorientierte Ansätze sind die Ausgabegrößen von angebotsorientierten Ansätzen und andersherum. Obwohl bekannt ist, dass Angebot und Nachfrage voneinander abhängen, wurden bisher nur einige grundlegende Kombinationen dieser Ansätze entwickelt. In dieser Arbeit werden Ansätze untersucht, die Angebot und Nachfrage im öffentlichen Verkehr tiefergehend integrieren.

In Abschnitt I dieser Arbeit untersuchen wir das Potenzial von integrierten nachfrage- und angebotsorientierten Modellen, indem wir die Passagiernachfrage in Entwurfsmodellen für öffentliche Verkehrsangebote schätzen. In Kapitel 2 vergleichen wir

Bewertungsfunktionen für Fahrpläne, die die Qualität von Fahrplänen aus der Sicht der Reisenden messen sollen, und untersuchen, inwieweit ihre Bewertungsergebnisse übereinstimmen. Bei diesem Vergleich stellen wir fest, dass das Design einer Bewertungsfunktion einen erheblichen Einfluss darauf haben kann, welcher Fahrplan als der beste angesehen wird. Außerdem ermitteln wir, welche Komponenten der Bewertungsfunktionen die Bewertungsergebnisse am meisten beeinflussen. In Kapitel 3 entwickeln wir zwei Fahrplanmodelle, um reisezeitoptimale Fahrpläne zu finden und gleichzeitig die Fahrgastverteilung auf verschiedenen Strecken zu schätzen. Die Fahrgastverteilung wird mit dem Logit-Modell geschätzt, einem diskreten Wahlmodell, das häufig zur Schätzung der Routenwahl von Fahrgästen verwendet wird. Wir stellen zwei lineare Repräsentationen für das Logit-Modell zur Verfügung, um es in ein Optimierungsmodell zu integrieren. Unser Vergleich mit vier modernen Fahrplangestaltungsmethoden zeigt, dass das Einbeziehen eines Fahrgastverteilungsmodells das Potenzial hat, bessere Fahrpläne für Passagiere zu finden. Jedoch ist dafür noch die Entwicklung effizienterer Lösungsstrategien erforderlich. In Kapitel 4 stellen wir ein Modell vor, um gewinnoptimale Linienpläne zu finden, während die Verkehrsmittel- und Routenwahl der Fahrgäste geschätzt wird. Durch eine vorherige Bestimmung des Nutzen von Routen, der für die Entscheidungen der Reisenden wichtig ist, können die Wahlmodelle linearisiert und kommerzielle Solver zur Lösungsfindung eingesetzt werden. In Experimenten zeigen wir, dass die Schätzung der Verkehrsmittelwahl der Reisenden während der Optimierung zu Linienplänen mit höherem Gewinn für die Betreiber und höherem Servicelevel für die Fahrgäste führt. Darüber hinaus legen unsere Experimente nahe, dass Betreiber regelmäßig auf Änderungen der Reisenachfrage reagieren sollten und ihre Linienpläne entsprechend aktualisieren sollten.

Im Gegensatz zu fahrplangebundenem öffentlichen Verkehr ermöglichen die jüngsten technologischen Entwicklungen die Umsetzung von flexibleren öffentlichen Verkehrsangeboten: erschwingliche und flächendeckende On-Demand-Dienste. Der Betrieb dieser Dienste wird live geplant und folgt keinen festen Linien- oder Fahrplänen. Da On-Demand-Dienste noch nicht in großem Umfang in Betrieb sind, sind ihre Auswirkungen auf Städte, Verkehr und Umwelt nur schwer abzuschätzen. Um die Folgen der Einführung von On-Demand-Diensten abschätzen zu können, untersuchen und bewerten Verkehrsbehörden mehrere mögliche Verkehrsszenarien mit Hilfe von Verkehrsnachfragemodellen. Dies führt zu folgendem Problem: Einerseits sind Verkehrsnachfragemodelle nachfrageorientierte Ansätze, die die Fahrgastnachfrage abschätzen und den Servicelevel des Angebots als Input benötigen. Andererseits hängt der Servicelevel von On-Demand-Diensten von der Fahrgastnachfrage ab und kann erst bestimmt werden, wenn die Nachfrage bekannt ist. Um On-Demand-Dienste in Reisenachfragemodellen berücksichtigen zu können, müssen die Modelle daher erweitert werden.

In Abschnitt II dieser Arbeit untersuchen wir, wie man den Servicelevel von On-Demand-Diensten innerhalb eines Reisebedarfsmodells abschätzen kann. Wir verwenden ein Fahrzeugumlegungsmodell, um die erforderliche Flottengröße der On-Demand-Dienste zu bestimmen. Unsere Modelle geben auch Aufschluss über die zurückgelegten Fahrzeugstrecken und die Fahrzeugstandorte im zeitlichen Verlauf. Dies ermöglicht eine gründliche Untersuchung der Auswirkungen von On-Demand-Diensten auf Verkehr und Umwelt. In Kapitel 5 stellen wir eine einfache Heuristik vor, um Lösungen für das Fahrzeugumlegungsproblem für On-Demand-Fahrzeuge in kurzen Berechnungszeiten zu finden. Die Heuristik ist in der Lage, mit nicht-ganzzahligen Nachfragewerten und mit sehr großen Instanzgrößen umzugehen, was beides übliche Merkmale der Anwendung sind. Wir illustrieren in zwei Fallstudien, wie die vorgestellte Heuristik zur Abschätzung des Servicelevel von On-Demand-Diensten innerhalb makroskopischer Verkehrsnachfragemodelle angewendet werden kann. Die in Kapitel 5 vorgestellte Heuristik kann gute Lösungen in kurzen Berechnungszeiten finden, gibt aber keine Optimalitätsgarantie. Es ist also nicht bekannt, ob die gefundenen Lösungen optimal sind oder wie weit sie von einer optimalen Lösung entfernt sind. In Kapitel 6 stellen wir einen Lösungsalgorithmus für das Fahrzeugumlegungsproblem von On-Demand-Diensten vor, mit dem sich optimale Lösungen finden lassen. Dieser Lösungsalgorithmus basiert auf dem Rolling Horizon-Prinzip und wir konstruieren eine Lösung, indem wir die Lösungen für die einzelnen kürzeren Horizonte zusammensetzen. Durch Überlappen der Horizonte ist es möglich, die Nachfrage aus dem nächsten Horizont zu berücksichtigen und die Lösungsqualität dadurch zu verbessern. Wenn die Horizonte weit genug überlappen, zeigen wir, dass die zusammengesetzte Lösung optimal für das gesamte Problem ist. In Experimenten zeigen wir, dass der Ansatz geeignet ist, optimale Lösungen für sehr große Instanzen zu finden. Außerdem konnten wir einen Laufzeitvorteil für große Instanzen feststellen, verglichen mit dem Ansatz, das gesamte Problem auf einmal zu lösen.

Zusammenfassend entwickeln und erforschen wir Methoden, die Nachfrage- und Angebotsmodelle integrieren. Wir zeigen, dass die Integration möglich ist und dass die gefundenen Lösungen Fahrgästen eine höhere Angebotsqualität bieten. Allerdings gehen die integrierten Modelle mit einer höheren Komplexität einher und künftige Forschung sollte sich mit spezialisierten Lösungsansätzen befassen.

About the author

Johann was born in Coburg, Germany in 1990. He studied Mathematics at the University of Göttingen and obtained his M.Sc. degree from there in 2016. During his studies, Johann visited the Bandung Institute of Technology in Bandung, Indonesia as a freemover in 2013-2014, and worked as a research intern at Technion - Israel Institute of Technology in Haifa, Israel in 2015. In 2016, Johann started as a PhD candidate at the Erasmus Research Institute of Management under the supervision of Prof. Leo Kroon, Prof. Dennis Huisman, Prof. Markus Friedrich, and Dr. Marie Schmidt. During his doctoral studies, Johann worked and conducted research at both the department of Technology and Operations Management, Rotterdam School of Management and the department for Transportation Planning and Traffic Engineering, University of Stuttgart, Germany. The work on his interdisciplinary and international project was also carried out in close collaboration with the department of Process quality and Innovation (π) of Netherlands Railways.

Johann's research interests include the development of optimization methods to advance public transport. His project dealt with the integration of passenger demand models and public transport optimization models. His work has been presented at several international conferences, including IFORS, EURO, ATMOS, and HEUREKA. In 2021, Johann was awarded the HEUREKA recognition award for his work on estimating the service level of on-demand vehicles within travel demand models. Part of his work is published in various conference proceedings and accepted for publication in scientific journals such as EJOR or Transportation.

Portfolio

Publications in Conference Proceedings
E. Richter et al. (2019). "Integrating ridesharing services with automated vehicles into macroscopic travel demand models". *MT-ITS 2019*. IEEE
J. Hartleb and M. Schmidt (2020). "A Rolling Horizon Heuristic with Optimality Guarantee for an On-Demand Vehicle Scheduling Problem". *20th Symposium on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS 2020)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik
J. Hartleb et al. (2021a). "Umlaufbildung für On Demand-Fahrzeugflotten in makroskopischen Nachfragemodellen". *HEUREKA'21: Optimierung in Verkehr und Transport*

Technical reports
J. Hartleb et al. (2019). "A good or a bad timetable: Do different evaluation functions agree?" *ERIM Report Series*
J. Hartleb, M. Friedrich, E. Richter (2020). "Vehicle Scheduling for On-demand Vehicle Fleets in Macroscopic Travel Demand Models". (accepted for publication at Transportation)
J. Hartleb and M. Schmidt (2019). "Railway timetabling with integrated passenger distribution". *Available at SSRN 3505167*. (accepted for publication at the European Journal of Operational Research)
J. Hartleb et al. (2021b). "Modeling and solving line planning with integrated mode choice". *Available at SSRN 3849985*. (submitted)

Prizes and awards

HEUREKA Anerkennungspreis 2021, Award of the HEUREKA foundation for the vehicle scheduling algorithm presented at the HEUREKA'21 conference, available at Hartleb et al. (2021a)

Software

J. Hartleb et al. (2020). *Vehicle Scheduling for On-demand Vehicle Fleets.* Version v1.0.0. URL: `https://github.com/jhartleb/VehicleScheduling/tree/v1.0.0`

Conferences Attended

Timetabling Workshop 2017, Göttingen, Germany
IFORS 2017, Quebec, Canada
DFG Annual Meeting 2017, Dortmund, Germany
EURO 2018, Valencia, Spain
RDTM 2018, Darmstadt, Germany
EURO 2019, Dublin, Ireland
ATMOS 2020, Pisa, Italy (online)
HEUREKA'21, Stuttgart, Germany (online)
EURO 2021, Athens, Greece (online, upcoming)
INFORMS TSL Workshop 2021, Rotterdam, The Netherlands (online, upcoming)
IFORS 2021, Seoul, South Korea (online, upcoming)

PhD Courses

Algorithms and Complexity	Choice-Based Optimization
Column Generation School 2018	Constraint Programming
Convex Analysis for Optimization	Cooperative Games
Deep Learning Demystified	English (CPE certificate)
Interior Point Methods	Networks and Polyhedra
OR-Games	Philosophy of Science
Publishing Strategy	Transport Logistics Modeling
Scientific Integrity	

Teaching

Advances in Supply Chain Management, Rotterdam School of Management. Guest Lecturer (2017)

Master Seminar on Transport Engineering, Institute for Road and Transport Science, University of Stuttgart, Germany. Guest Lecturer (2017-18, 2020)

Research Methods and Skills, Rotterdam School of Management. Master Thesis Coach & Co-reader (2018, 2020-21)

Quantitative Methods and Techniques: Mathematics, Rotterdam School of Management. Coordinator (2019) & Lecturer (2020)

Supply Chain Simulation, Rotterdam School of Management. Workshop and Lab Instructor & Lecturer (2019-21)

Research Training & Bachelor Thesis, Rotterdam School of Management. Bachelor Thesis Instructor (2020-21)

Masterclass: Discrete-Event Simulation, Rotterdam School of Management. Guest Lecturer (2021)

The ERIM PhD Series

The ERIM PhD Series contains PhD dissertations in the field of Research in Management defended at Erasmus University Rotterdam and supervised by senior researchers affiliated to the Erasmus Research Institute of Management (ERIM). All dissertations in the ERIM PhD Series are available in full text through the ERIM Electronic Series Portal: `https://repub.eur.nl/pub`. ERIM is the joint research institute of the Rotterdam School of Management (RSM) and the Erasmus School of Economics (ESE) at the Erasmus University Rotterdam (EUR).

Dissertations in the last four years

Ahmadi, S., *A motivational perspective to decision-making and behavior in organizations*, Promotors: Prof. J.J.P. Jansen & Dr. T.J.M. Mom, EPS-2019-477-S&E, `https://repub.eur.nl/pub/116727`

Akemu, O., *Corporate Responses to Social Issues: Essays in Social Entrepreneurship and Corporate Social Responsibility*, Promotors: Prof. G.M. Whiteman & Dr. S.P. Kennedy, EPS-2017-392-ORG, `https://repub.eur.nl/pub/95768`

Albuquerque de Sousa, J.A., *International stock markets: Essays on the determinants and consequences of financial market development*, Promotors: Prof. M.A. van Dijk & Prof. P.A.G. van Bergeijk, EPS-2019-465-F&A, `https://repub.eur.nl/pub/115988`

Alserda, G.A.G., *Choices in Pension Management*, Promotors: Prof. S.G. van der Lecq & Dr. O.W. Steenbeek, EPS-2017-432-F&A, `https://repub.eur.nl/pub/103496`

Anantavrasilp, S., *Essays on Ownership Structures, Corporate Finance Policies and Financial Reporting Decisions*, Promotors: Prof. A. de Jong & Prof. P.G.J. Roosenboom, EPS-2021-516-F&E, `https://repub.eur.nl/pub/134947`

Arampatzi, E., *Subjective Well-Being in Times of Crises: Evidence on the Wider Impact of Economic Crises and Turmoil on Subjective Well-Being*, Promotors: Prof. H.R. Commandeur, Prof. F. van Oort & Dr. M.J. Burger, EPS-2018-459-S&E, `https://repub.eur.nl/pub/111830`

Arslan, A.M., *Operational Strategies for On-demand Delivery Services*, Promotors: Prof. R.A. Zuidwijk & Dr. N.A.H. Agatz, EPS-2019-481-LIS, `https://repub.eur.nl/pub/126463`

Aydin, Z., *Mobile Consumers and Applications: Essays on Mobile Marketing*, Promotors: Prof. G.H. van Bruggen & Dr. B. Ataman, EPS-2021-519-MKT, `https://repub.eur.nl/pub/135352`

Azadeh, K., *Robotized Warehouses: Design and Performance Analysis*, Promotors: Prof. M.B.M. de Koster & Prof. D. Roy, EPS-2021-515-LIS, `https://repub.eur.nl/pub/135208`

Avci, E., *Surveillance of Complex Auction Markets: a Market Policy Analytics Approach*, Promotors: Prof. W. Ketter, Prof. H.W.G.M. van Heck & Prof. D.W. Bunn, EPS-2018-426-LIS, `https://repub.eur.nl/pub/106286`

Balen, T.H. van, *Challenges of Early Stage Entrepreneurs: the Roles of Vision Communication and Team Membership Change*, Promotors: Prof. J.C.M. van den Ende & Dr M. Tarakci, EPS-2019-468-LIS, `https://repub.eur.nl/pub/115654`

Bansraj, S.C., *The Principles of Private Equity: Ownership and Acquisitions*, Promotors: Prof. J.T.J. Smit & Dr. V. Volosovych, EPS-2020-507-F&A, `https://repub.eur.nl/pub/132329`

Bavato, D., *With New Eyes: The recognition of novelty and novel ideas*, Promotors: Prof. D.A. Stam & Dr. S. Tasselli, EPS-2020-500-LIS, `https://repub.eur.nl/pub/134264`

Bernoster, I., *Essays at the Intersection of Psychology, Biology, and Entrepreneurship*, Promotors: Prof. A.R. Thurik, Prof. I.H.A. Franken & Prof. P.J.F. Groenen, EPS-2018-463-S&E, `https://repub.eur.nl/pub/113907`

Blagoeva, R.R., *The Hard Power Of Soft Power: A behavioral strategy perspective on how power, reputation, and status affect firms*, Promotors: Prof. J.J.P. Jansen & Prof. T.J.M. Mom, EPS-2020-495-S&E, https://repub.eur.nl/pub/127681

Bouman, P., *Passengers, Crowding and Complexity: Models for Passenger Oriented Public Transport*, Promotors: Prof. L.G. Kroon, Prof. A. Schöbel & Prof. P.H.M. Vervest, EPS-2017-420-LIS, https://repub.eur.nl/pub/100767

Breugem, T., *Crew Planning at Netherlands Railways: Improving Fairness, Attractiveness, and Efficiency*, Promotors: Prof. D. Huisman & Dr. T.A.B. Dollevoet, EPS-2020-494-LIS, https://repub.eur.nl/pub/124016

Bunderen, L. van, *Tug-of-War: Why and when teams get embroiled in power struggles*, Promotors: Prof. D.L. van Knippenberg & Dr. L. Greer, EPS-2018-446-ORG, https://repub.eur.nl/pub/105346

Burg, G.J.J. van den, *Algorithms for Multiclass Classification and Regularized Regression*, Promotors: Prof. P.J.F. Groenen & Dr. A. Alfons, EPS-2018-442-MKT, https://repub.eur.nl/pub/103929

Chammas, G., *Portfolio concentration*, Promotor: Prof. J. Spronk, EPS-2017-410-F&E, https://repub.eur.nl/pub/94975

Chan, H.Y., *Decoding the consumer's brain: Neural representations of consumer experience*, Promotors: Prof. A. Smidts & Dr. M.A.S. Boksem, EPS-2019-493-MKT, https://repub.eur.nl/pub/124931

Couwenberg, L., *Context dependent valuation: A neuroscientific perspective on consumer decision-making*, Promotors: Prof. A. Smit, Prof. A.G. Sanfrey & Dr. M.A.S. Boksem, EPS-2020-505-MKT, https://repub.eur.nl/pub/129601

Dalmeijer, K., *Time Window Assignment in Distribution Networks*, Promotors: Prof A.P.M. Wagelmans & Dr. R. Spliet, EPS-2019-486-LIS, https://repub.eur.nl/pub/120773

Dennerlein, T., *Empowering Leadership and Employees' Achievement Motivations: the Role of Self-Efficacy and Goal Orientations in the Empowering Leadership Process*, Promotors: Prof. D.L. van Knippenberg & Dr. J. Dietz, EPS-2017-414-ORG, https://repub.eur.nl/pub/98438

Dolgova, E., *On Getting Along and Getting Ahead: How Personality Affects Social Network Dynamics*, Promotors: Prof. P.P.M.A.R. Heugens & Prof. M.C. Schippers, EPS-2019-455-S&E, https://repub.eur.nl/pub/119150

Duijzer, L.E., *Mathematical Optimization in Vaccine Allocation*, Promotors: Prof. R. Dekker & Dr. W.L. van Jaarsveld, EPS-2017-430-LIS, https://repub.eur.nl/pub/101487

Eijlers, E., *Emotional Experience and Advertising Effectiveness: on the use of EEG in marketing*, Promotors: Prof. A. Smidts & Prof. M.A.S. Boksem, EPS-2019-487-MKT, https://repub.eur.nl/pub/124053

El Nayal, O.S.A.N., *Firms and the State: An Examination of Corporate Political Activity and the Business-Government Interface*, Promotors: Prof. J. van Oosterhout & Dr. M. van Essen, EPS-2018-469-S&E, https://repub.eur.nl/pub/114683

Fasaei, H., *Changing the Narrative: The Behavioral Effects of Social Evaluations on the Decision Making of Organizations*, Promotors: Prof. J.J.P. Jansen, Prof. T.J.M. Mom & Dr. M.P. Tempelaar, EPS-2020-492-S&E, https://repub.eur.nl/pub/129598

Feng, Y., *The Effectiveness of Corporate Governance Mechanisms and Leadership Structure: Impacts on strategic change and firm performance*, Promotors: Prof. F.A.J. van den Bosch, Prof. H.W. Volberda & Dr. J.S. Sidhu, EPS-2017-389-S&E, https://repub.eur.nl/pub/98470

Frick, T.W., *The Implications of Advertising Personalization for Firms, Consumer, and Ad Platfroms*, Promotors: Prof. T. Li & Prof. H.W.G.M. van Heck, EPS-2018-452-LIS, https://repub.eur.nl/pub/110314

Fytraki, A.T., *Behavioral Effects in Consumer Evaluations of Recommendation Systems*, Promotors: Prof. B.G.C. Dellaert & Prof. T. Li, EPS-2018-427-MKT, https://repub.eur.nl/pub/110457

Gai, J., *Contextualized Consumers: Theories and Evidence on Consumer Ethics, Product Recommendations, and Self-Control*, Promotors: Prof. S. Puntoni & Prof. S.T.L. Sweldens, EPS-2020-498-MKT, `https://repub.eur.nl/pub/127680`

Ghazizadeh, P., *Empirical Studies on the Role of Financial Information in Asset and Capital Markets*, Promotors: Prof. A. de Jong & Prof. E. Peek, EPS-2019-470-F&A, `https://repub.eur.nl/pub/114023`

Giurge, L., *A Test of Time; A temporal and dynamic approach to power and ethics*, Promotors: Prof. M.H. van Dijke & Prof. D. de Cremer, EPS-2017-412-ORG, `https://repub.eur.nl/pub/98451`

Gobena, L., *Towards Integrating Antecedents of Voluntary Tax Compliance*, Promotors: Prof. M.H. van Dijke & Dr. P. Verboon, EPS-2017-436-ORG, `https://repub.eur.nl/pub/103276`

Groot, W.A., *Assessing Asset Pricing Anomalies*, Promotors: Prof. M.J.C.M. Verbeek & Prof. J.H. van Binsbergen, EPS-2017-437-F&A, `https://repub.eur.nl/pub/103490`

Hanselaar, R.M., *Raising Capital: On pricing, liquidity and incentives*, Promotors: Prof. M.A. van Dijk & Prof. P.G.J. Roosenboom, EPS-2018-429-F&A, `https://repub.eur.nl/pub/113274`

Harms, J.A., *Essays on the Behavioral Economics of Social Preferences and Bounded Rationality*, Promotors: Prof. H.R. Commandeur & Dr. K.E.H. Maas, EPS-2018-457-S&E, `https://repub.eur.nl/pub/108831`

Hendriks, G., *Multinational Enterprises and Limits to International Growth: Links between Domestic and Foreign Activities in a Firm's Portfolio*, Promotors: Prof. P.P.M.A.R. Heugens & Dr. A.H.L. Slangen, EPS-2019-464-S&E, `https://repub.eur.nl/pub/114981`

Hengelaar, G.A., *The Proactive Incumbent: Holy grail or hidden gem? Investigating whether the Dutch electricity sector can overcome the incumbent's curse and lead the sustainability transition*, Promotors: Prof. R.J.M. van Tulder & Dr. K. Dittrich, EPS-2018-438-ORG, `https://repub.eur.nl/pub/102953`

Jacobs, B.J.D., *Marketing Analytics for High-Dimensional Assortments*, Promotors: Prof. A.C.D. Donkers & Prof. D. Fok, EPS-2017-445-MKT, `https://repub.eur.nl/pub/103497`

Jia, F., *The Value of Happiness in Entrepreneurship*, Promotors: Prof. D.L. van Knippenberg & Dr. Y. Zhang, EPS-2019-479-ORG, `https://repub.eur.nl/pub/115990`

Kahlen, M.T., *Virtual Power Plants of Electric Vehicles in Sustainable Smart Electricity Markets*, Promotors: Prof. W. Ketter & Prof. A. Gupta, EPS-2017-431-LIS, `https://repub.eur.nl/pub/100844`

Kampen, S. van, *The Cross-sectional and Time-series Dynamics of Corporate Finance: Empirical evidence from financially constrained firms*, Promotors: Prof. L. Norden & Prof. P.G.J. Roosenboom, EPS-2018-440-F&A, `https://repub.eur.nl/pub/105245`

Karali, E., *Investigating Routines and Dynamic Capabilities for Change and Innovation*, Promotors: Prof. H.W. Volberda, Prof. H.R. Commandeur and Dr. J.S. Sidhu, EPS-2018-454-S&E, `https://repub.eur.nl/pub/106274`

Keko. E., *Essays on Innovation Generation in Incumbent Firms*, Promotors: Prof. S. Stremersch & Dr. N.M.A. Camacho, EPS-2017-419-MKT, `https://repub.eur.nl/pub/100841`

Kerkkamp, R.B.O., *Optimisation Models for Supply Chain Coordination under Information Asymmetry*, Promotors: Prof. A.P.M. Wagelmans & Dr. W. van den Heuvel, EPS-2018-462-LIS, `https://repub.eur.nl/pub/109770`

Khattab, J., *Make Minorities Great Again: a contribution to workplace equity by identifying and addressing constraints and privileges*, Promotors: Prof. D.L. van Knippenberg & Dr. A. Nederveen Pieterse, EPS-2017-421-ORG, `https://repub.eur.nl/pub/99311`

Kim, T.Y., *Data-driven Warehouse Management in Global Supply Chains*, Promotors: Prof. R. Dekker & Dr. C. Heij, EPS-2018-449-LIS, `https://repub.eur.nl/pub/109103`

Klitsie, E.J., *Strategic Renewal in Institutional Contexts: The paradox of embedded agency*, Promotors: Prof. H.W. Volberda & Dr. S. Ansari, EPS-2018-444-S&E, `https://repub.eur.nl/pub/106275`

Koolen, D., *Market Risks and Strategies in Power Systems Integrating Renewable Energy*, Promotors: Prof. W. Ketter & Prof. R. Huisman, EPS-2019-467-LIS, `https://repub.eur.nl/pub/115655`

Kong, L., *Essays on Financial Coordination*, Promotors: Prof. M.J.C.M. Verbeek, Dr. D.G.J. Bongaerts & Dr. M.A. van Achter, EPS-2019-433-F&A, https://repub.eur.nl/pub/114516

Korman, B., *Leader-Subordinate Relations: The Good, the Bad and the Paradoxical*, Promotors: S.R. Giessner & Prof. C. Tröster, EPS-2021-511-ORG, https://repub.eur.nl/pub/135365

Kyosev, G.S., *Essays on Factor Investing*, Promotors: Prof. M.J.C.M. Verbeek & Dr. J.J. Huij, EPS-2019-474-F&A, https://repub.eur.nl/pub/116463

Lamballais Tessensohn, T., *Optimizing the Performance of Robotic Mobile Fulfillment Systems*, Promotors: Prof. M.B.M. de Koster, Prof. R. Dekker & Dr. D. Roy, EPS-2019-411-LIS, https://repub.eur.nl/pub/116477

Leung, W.L., *How Technology Shapes Consumption: Implications for Identity and Judgement*, Promotors: Prof. S. Puntoni & Dr. G. Paolacci, EPS-2019-485-MKT, https://repub.eur.nl/pub/117432

Li, W., *Competition in the Retail Market of Consumer Packaged Goods*, Promotors: Prof. D. Fok & Prof. Ph.H.B.F. Franses, EPS-2021-503-MKT, https://repub.eur.nl/pub/134873

Li, X., *Dynamic Decision Making under Supply Chain Competition*, Promotors: Prof. M.B.M. de Koster, Prof. R. Dekker & Prof. R. Zuidwijk, EPS-2018-466-LIS, https://repub.eur.nl/pub/114028

Liu, N., *Behavioral Biases in Interpersonal Contexts*, Promotors: Prof. A. Baillon & Prof. H. Bleichrodt, EPS-2017-408-MKT, https://repub.eur.nl/pub/95487

Maas, A.J.J., *Organizations and their external context: Impressions across time and space*, Promotors: Prof. P.P.M.A.R. Heugens & Prof. T.H. Reus, EPS-2019-478-S&E, https://repub.eur.nl/pub/116480

Maira, E., *Consumers and Producers*, Promotors: Prof. S. Puntoni & Prof. C. Fuchs, EPS-2018-439-MKT, https://repub.eur.nl/pub/104387

Manouchehrabadi, B., *Information, Communication and Organizational Behavior*, Promotors: Prof. G.W.J. Hendrikse & Dr. O.H. Swank, EPS-2020-502-ORG, https://repub.eur.nl/pub/132185

Matawlie N., *Through Mind and Behaviour to Financial Decisions*, Promotors: Prof. J.T.J. Smit & Prof. P. Verwijmeren, EPS-2020-501-F&A, https://repub.eur.nl/pub/134265

Mirzaei, M., *Advanced Storage and Retrieval Policies in Automated Warehouses*, Promotors: Prof. M.B.M. de Koster & Dr. N. Zaerpour, EPS-2020-490-LIS, https://repub.eur.nl/pub/125975

Nair, K.P., *Strengthening Corporate Leadership Research: The relevance of biological explanations*, Promotors: Prof. J. van Oosterhout & Prof. P.P.M.A.R. Heugens, EPS-2019-480-S&E, https://repub.eur.nl/pub/120023

Nullmeier, F.M.E., *Effective contracting of uncertain performance outcomes: Allocating responsibility for performance outcomes to align goals across supply chain actors*, Promotors: Prof. J.Y.F. Wynstra & Prof. E.M. van Raaij, EPS-2019-484-LIS, https://repub.eur.nl/pub/118723

Okbay, A., *Essays on Genetics and the Social Sciences*, Promotors: Prof. A.R. Thurik, Prof. Ph.D. Koellinger & Prof. P.J.F. Groenen, EPS-2017-413-S&E, https://repub.eur.nl/pub/95489

Peng, X., *Innovation, Member Sorting, and Evaluation of Agricultural Cooperatives*, Promotor: Prof. G.W.J. Hendriks, EPS-2017-409-ORG, https://repub.eur.nl/pub/94976

Petruchenya, A., *Essays on Cooperatives: Emergence, Retained Earnings, and Market Shares*, Promotors: Prof. G.W.J. Hendriks & Dr. Y. Zhang, EPS-2018-447-ORG, https://repub.eur.nl/pub/105243

Plessis, C. du, *Influencers: The Role of Social Influence in Marketing*, Promotors: Prof. S. Puntoni & Prof. S.T.L.R. Sweldens, EPS-2017-425-MKT, https://repub.eur.nl/pub/103265

Pocock, M., *Status Inequalities in Business Exchange Relations in Luxury Markets*, Promotors: Prof. C.B.M. van Riel & Dr. G.A.J.M. Berens, EPS-2017-346-ORG, https://repub.eur.nl/pub/98647

Polinder, G.J., *New Models and Applications for Railway Timetabling*, Promotors: Prof. D. Huisman & Dr. M.E. Schmidt, EPS-2020-514-LIS, https://repub.eur.nl/pub/134600

Pozharliev, R., *Social Neuromarketing: The role of social context in measuring advertising effectiveness*, Promotors: Prof. W.J.M.I. Verbeke & Prof. J.W. van Strien, EPS-2017-402-MKT, https://repub.eur.nl/pub/95528

Qian, Z., *Time-Varying Integration and Portfolio Choices in the European Capital Markets*, Promotors: Prof. W.F.C. Verschoor, Prof. R.C.J. Zwinkels & Prof. M.A. Pieterse-Bloem, EPS-2020-488-F&A, https://repub.eur.nl/pub/124984

Reh, S.G., *A Temporal Perspective on Social Comparisons in Organizations*, Promotors: Prof. S.R. Giessner, Prof. N. van Quaquebeke & Dr. C. Troster, EPS-2018-471-ORG, https://repub.eur.nl/pub/114522

Riessen, B. van, *Optimal Transportation Plans and Portfolios for Synchromodal Container Networks*, Promotors: Prof. R. Dekker & Prof. R.R. Negenborn, EPS-2018-448-LIS, https://repub.eur.nl/pub/105248

Romochkina, I.V., *When Interests Collide: Understanding and modeling interests alignment using fair pricing in the context of interorganizational information systems*, Promotors: Prof. R.A. Zuidwijk & Prof. P.J. van Baalen, EPS-2020-451-LIS, https://repub.eur.nl/pub/127244

Schie, R.J.G. van, *Planning for Retirement: Save More or Retire Later?*, Promotors: Prof. B.G.C. Dellaert & Prof. A.C.D. Donkers, EOS-2017-415-MKT, https://repub.eur.nl/pub/100846

Schneidmüller, T., *Engaging with Emerging Technologies: Socio-cognitive foundations of incumbent response*, Promotors: Prof. H. Volberda & Dr. S.M. Ansari, EPS-2020-509-S&E, https://repub.eur.nl/pub/131124

Schouten, K.I.M., *Semantics-driven Aspect-based Sentiment Analysis*, Promotors: Prof. F.M.G. de Jong, Prof. R. Dekker & Dr. F. Frasincar, EPS-2018-453-LIS, https://repub.eur.nl/pub/112161

Sihag, V., *The Effectiveness of Organizational Controls: A meta-analytic review and an investigation in NPD outsourcing*, Promotors: Prof. J.C.M. van den Ende & Dr. S.A. Rijsdijk, EPS-2019-476-LIS, https://repub.eur.nl/pub/115931

Slob, E., *Integrating Genetics into Economics*, Promotors: Prof. A.R. Thurik, Prof. P.J.F. Groenen & Dr. C.A. Rietveld, EPS-2021-517-S&E, https://repub.eur.nl/pub/135159

Smolka, K.M., *Essays on Entrepreneurial Cognition, Institution Building and Industry Emergence*, Promotors: P.P.M.A.R. Heugens, & Prof. J.P. Cornelissen, EPS-2019-483-S&E, https://repub.eur.nl/pub/118760

Straeter, L.M., *Interpersonal Consumer Decision Making*, Promotors: Prof. S.M.J. van Osselaer & Dr. I.E. de Hooge, EPS-2017-423-MKT, https://repub.eur.nl/pub/100819

Stuppy, A., *Essays on Product Quality*, Promotors: Prof. S.M.J. van Osselaer & Dr. N.L. Mead, EPS-2018-461-MKT, https://repub.eur.nl/pub/111375

Subaşi, B., *Demographic Dissimilarity, Information Access and Individual Performance*, Promotors: Prof. D.L. van Knippenberg & Dr. W.P. van Ginkel, EPS-2017-422-ORG, https://repub.eur.nl/pub/103495

Suurmond, R., *In Pursuit of Supplier Knowledge: Leveraging capabilities and dividing responsibilities in product and service contexts*, Promotors: Prof. J.Y.F. Wynstra & Prof. J. Dul, EPS-2018-475-LIS, https://repub.eur.nl/pub/115138

Toxopeus, H.S., *Financing sustainable innovation: From a principal-agent to a collective action perspective*, Promotors: Prof. H.R. Commandeur & Dr. K.E.H. Maas, EPS-2019-458-S&E, https://repub.eur.nl/pub/114018

Turturea, R., *Overcoming Resource Constraints: The Role of Creative Resourcing and Equity Crowdfunding in Financing Entrepreneurial Ventures*, Promotors: Prof. P.P.M.A.R. Heugens, Prof. J.J.P. Jansen & Dr. I. Verheuil, EPS-2019-472-S&E, https://repub.eur.nl/pub/112859

Valboni, R., *Building Organizational (Dis-)Abilities: The impact of learning on the performance of mergers and acquisitions*, Promotors: Prof. T.H. Reus & Dr. A.H.L. Slangen, EPS-2020-407-S&E, https://repub.eur.nl/pub/125226

Vandic, D., *Intelligent Information Systems for Web Product Search*, Promotors: Prof. U. Kaymak & Dr. F. Frasincar, EPS-2017-405-LIS, https://repub.eur.nl/pub/95490

Verbeek, R.W.M., *Essays on Empirical Asset Pricing*, Promotors: Prof. M.A. van Dijk & Dr. M. Szymanowska, EPS-2017-441-F&A, https://repub.eur.nl/pub/102977

Visser, T.R., *Vehicle Routing and Time Slot Management in Online Retailing*, Promotors: Prof. A.P.M. Wagelmans & Dr. R. Spliet, EPS-2019-482-LIS, https://repub.eur.nl/pub/120772

Vlaming, R. de, *Linear Mixed Models in Statistical Genetics*, Promotors: Prof. A.R. Thurik, Prof. P.J.F. Groenen & Prof. Ph.D. Koellinger, EPS-2017-416-S&E, https://repub.eur.nl/pub/100428

Vongswasdi, P., *Accelerating Leadership Development: An evidence-based perspective*, Promotors: Prof. D. van Dierendonck & Dr. H.L. Leroy, EPS-2020-512-ORG, https://repub.eur.nl/pub/134079

Vries, H. de, *Evidence-Based Optimization in Humanitarian Logistics*, Promotors: Prof. A.P.M. Wagelmans & Prof. J.J. van de Klundert, EPS-2017-435-LIS, https://repub.eur.nl/pub/102771

Wang, R., *Corporate Environmentalism in China*, Promotors: Prof. P.P.M.A.R. Heugens & Dr. F. Wijen, EPS-2017-417-S&E, https://repub.eur.nl/pub/99987

Wang, R., *Those Who Move Stock Prices*, Promotors: Prof. P. Verwijmeren & Prof. S. van Bekkum, EPS-2019-491-F&A, https://repub.eur.nl/pub/129057

Wasesa, M., *Agent-based inter-organizational systems in advanced logistics operations*, Promotors: Prof. H.W.G.M. van Heck, Prof. R.A. Zuidwijk & Dr. A.W. Stam, EPS-2017-LIS-424, https://repub.eur.nl/pub/100527

Wessels, C., *Flexible Working Practices: How Employees Can Reap the Benefits for Engagement and Performance*, Promotors: Prof. H.W.G.M. van Heck, Prof. P.J. van Baalen & Prof. M.C. Schippers, EPS-2017-418-LIS, https://repub.eur.nl/pub/99312

Wiegmann, P.M., *Setting the Stage for Innovation: Balancing Diverse Interests through Standardisation*, Promotors: Prof. H.J. de Vries & Prof. K. Blind, EPS-2019-473-LIS, https://repub.eur.nl/pub/114519

Wijaya, H.R., *Praise the Lord!: Infusing Values and Emotions into Neo-Institutional Theory*, Promotors: Prof. P.P.M.A.R. Heugens & Prof. J.P. Cornelissen, EPS-2019-450-S&E, https://repub.eur.nl/pub/115973

Williams, A.N., *Make Our Planet Great Again: A Systems Perspective of Corporate Sustainability*, Promotors: Prof. G.M. Whiteman & Dr. S. Kennedy, EPS-2018-456-ORG, https://repub.eur.nl/pub/111032

Witte, C.T., *Bloody Business: Multinational investment in an increasingly conflict-afflicted world*, Promotors: Prof. H.P.G. Pennings, Prof. H.R. Commandeur & Dr. M.J. Burger, EPS-2018-443-S&E, https://repub.eur.nl/pub/104027

Wu, J., *A Configural Approach to Understanding Voice Behavior in Teams*, Promotors: Prof. D.L. van Knippenberg & Prof. S.R. Giessner, EPS-2020-510-ORG, https://repub.eur.nl/pub/132184

Ye, Q.C., *Multi-objective Optimization Methods for Allocation and Prediction*, Promotors: Prof. R. Dekker & Dr. Y. Zhang, EPS-2019-460-LIS, https://repub.eur.nl/pub/116462

Yuan, Y., *The Emergence of Team Creativity: a social network perspective*, Promotors: Prof. D.L. van Knippenberg & Dr. D.A. Stam, EPS-2017-434-ORG, https://repub.eur.nl/pub/100847

Zhang, Q., *Financing and Regulatory Frictions in Mergers and Acquisitions*, Promotors: Prof. P.G.J. Roosenboom & Prof. A. de Jong, EPS-2018-428-F&A, https://repub.eur.nl/pub/103871

Public transport is an indispensable part of our society. It increases mobility for all, enables efficient transportation in densely populated areas, and protects the environment with lowest emissions per passenger kilometer. To reap its benefits, an effective public transport system must be created attracting large numbers of passengers.

In the first part of this thesis, we present integrated models to optimize public transport services while estimating the corresponding passenger choices. The first study compares different timetable evaluation functions for consistency and gives further motivation for the integration of passenger choice models into optimization models. In the next two studies, we present novel optimization models with integrated demand estimation for the steps of timetabling and line planning, respectively. The resulting public transport services are designed for the passenger demand they generate.

The second part of this thesis deals with new and more flexible forms of public transport: mobility on demand. In order to assess the consequences of large-scale on-demand services on cities and regions, travel demand models need to be extended to determine the service level of on-demand services. Both studies in this part present solution algorithms for a vehicle scheduling problem of on-demand services to estimate the required vehicle fleet size and distance traveled.

ERIM

The Erasmus Research Institute of Management (ERIM) is the Research School (Onderzoekschool) in the field of management of the Erasmus University Rotterdam. The founding participants of ERIM are the Rotterdam School of Management (RSM), and the Erasmus School of Economics (ESE). ERIM was founded in 1999 and is officially accredited by the Royal Netherlands Academy of Arts and Sciences (KNAW). The research undertaken by ERIM is focused on the management of the firm in its environment, its intra- and interfirm relations, and its business processes in their interdependent connections.

The objective of ERIM is to carry out first rate research in management, and to offer an advanced doctoral programme in Research in Management. Within ERIM, over three hundred senior researchers and PhD candidates are active in the different research programmes. From a variety of academic backgrounds and expertises, the ERIM community is united in striving for excellence and working at the forefront of creating new business knowledge.

ERIM

ERIM PhD Series
Research in Management

Erasmus University Rotterdam (EUR)
Erasmus Research Institute of Management
Mandeville (T) Building
Burgemeester Oudlaan 50
3062 PA Rotterdam, The Netherlands

P.O. Box 1738
3000 DR Rotterdam, The Netherlands
T +31 10 408 1182
E info@erim.eur.nl
W www.erim.eur.nl