Aalto University
School of Science

Master's programme in Mathematics and Operations Research

Optimization Approaches for Line Planning in Linear Railway Systems

Viljami Uusihärkälä

Master's Thesis
2024

Aalto University
School of Science

Author Viljami Uusihärkälä

Title Optimization Approaches for Line Planning in Linear Railway Systems

Degree programme Mathematics and Operations Research

Major Systems and Operations Research

Supervisor Asst. Prof. Philine Schiewe

Advisor Asst. Prof. Philine Schiewe

Date 12.3.2024 **Number of pages** 57 **Language** English

Abstract

The relevance of public transportation has only increased in the past decades given its counteracting effect on climate change. The positive effects of public transit are increased when more passengers perceive it as an attractive option compared to alternative means of travel, which gives motivation to design the service such that the passengers' convenience is maximized. The problem becomes more complex when the operator's interest in making a profit is taken into account. This thesis concentrates on the line planning stage of public transportation planning, in which the objective is to determine which lines are operated and how regularly. Line planning is studied with an emphasis on examining various ways of modeling passengers' convenience, specifically on linear public transportation networks.

Three passenger-oriented line planning models are developed, where the following passenger convenience metrics are considered: congestion, number of direct passengers, and initial waiting time. The models are formulated as mixed-integer linear programs, where the operator's interest in exploiting a minimal fleet is regarded as a constraint. Analogously, models where the operator's interest is maximized and the passengers' convenience is constrained are also studied. The models are specifically designed for linear railway networks, which excludes the need to account for various passenger routes. Moreover, the models are augmented with the possibility to limit the number of terminal stations, which allow for lines to begin and end. These models are experimentally evaluated in numerous experiments in which data generated for both the network and passenger demand is employed.

The experimental evaluation indicates that the models are successful in improving the passengers' convenience. For instance, almost all passengers can travel directly when the appropriate model is solved and if the solution is feasible. The results depend both on the fleet size and terminal stations, where some terminals are identified as more influential for improving passengers' convenience. The line concepts obtained with the different models can be rather conflicting in the sense that they provide poor performance when evaluated on the other metrics. This result suggests further development for regarding the different objectives simultaneously. In addition, the models require further evaluation with real-world networks and passenger data since the results obtained in this thesis are based on only generated data.

Keywords line planning, public transportation, mathematical programming

Aalto-yliopisto
Perustieteiden
korkeakoulu

Tekijä Viljami Uusihärkälä

Työn nimi Optimointimenetelmiä lineaarisen rautatieverkon linjasuunnitteluun

Koulutusohjelma Mathematics and Operations Research

Pääaine Systems and Operations Research

Työn valvoja Asst. Prof. Philine Schiewe

Työn ohjaaja Asst. Prof. Philine Schiewe

Päivämäärä 12.3.2024 **Sivumäärä** 57 **Kieli** englanti

Tiivistelmä

Julkisen liikenteen merkitys on kasvanut entisestään viimeisimpinä vuosikymmeninä, mikä johtuu osaltaan sen mahdollisuuksista torjua ilmastonmuutosta. Julkisen liikenteen hyödyt korostuvat erityisesti silloin, kun useammat matkustajat pitävät sitä houkuttelevana vaihtoehtona muihin matkustustapoihin verrattuna. Näin ollen joukkoliikenne kannattaa suunnitella siten, että matkustajien mukavuus maksimoituu. Ongelmasta tulee vielä monimutkaisempi, kun otetaan lisäksi huomioon palveluntarjoajan tavoite tehdä voittoa. Tässä diplomityössä tarkastellaan yhtä julkisen liikenteen suunnitteluvaihetta linjaoptimointia, jossa päätökset koskevat linjojen valintaa sekä vuorojen tiheyttä. Työn tavoitteena on tutkia erilaisia tapoja mallintaa matkustajien mukavuutta keskittyen erityisesti lineaarisiin rautatieverkkoihin.

Tässä työssä kehitetään kolme matkustajalähtöistä linjasuunnittelumallia, joiden tavoitefunktiot ottavat huomioon ruuhkautumisen, suorien matkojen lukumäärän sekä odotusajan lähtöasemilla. Mallit muotoillaan sekalukutehtäviksi, joissa palveluntarjoajan tavoite käyttää mahdollisimman pientä määrää junia otetaan huomioon rajoitteena. Työssä tutkitaan vastaavasti myös malleja, joissa palveluntarjoajan tavoitetta maksimoidaan ja matkustajien tavoitetta rajoitetaan. Lisäksi mallit räätälöidään lineaarisille rataverkoille, mikä poistaa tarpeen huomioida matkustajien useampia mahdollisia reittejä. Malleihin lisätään myös mahdollisuus rajoittaa pääteasemien määrää, jotka voivat toimia linjojen päätepysäkkeinä. Näitä malleja arvioidaan kokeellisesti hyödyntäen generoitua dataa sekä rautatieverkon että matkustajamäärien osalta.

Tulokset osoittavat, että mallit pystyvät lisäämään matkustajien mukavuutta. Melkein kaikki matkustajat voivat matkustaa ilman vaihtoja, mikäli käytetään kyseistä mallia ja kapasiteettirajoitteet täyttyvät. Tulokset ovat myös melko riippuvaisia käytettävissä olevien junien lukumäärästä sekä pääteasemista, joista osalla on suuri merkitys matkustajien mukavuuden lisäämisessä. Eri malleilla optimoidut linjaratkaisut voivat olla melko ristiriitaisia keskenään. Tulokset voivat olla heikkoja, mikäli niitä arvioidaan muiden kuin optimoinnissa käytettyjen metriikkojen perusteella. Nämä tulokset viittaavat siihen, että jatkotutkimuksessa pitäisi keskittyä eri tavoitefunktioiden samanaikaiseen optimointiin. Lisäksi malleja pitäisi käyttää linjasuunnitelmien tekemiseen oikealla rataverkko- ja matkustajamäärädatalla, koska tässä työssä esitellyt tulokset nojaavat pelkästään generoituun dataan.

Avainsanat linjasuunnittelu, julkinen liikenne, matemaattinen optimointi

Contents

Abstract 3

Abstract (in Finnish) 4

Contents 5

Symbols and abbreviations 6

1 Introduction 7

2 Background 9
- 2.1 Public transportation planning . . . 9
- 2.2 The global process . . . 9
  - 2.2.1 Network design . . . 11
  - 2.2.2 Timetabling and scheduling . . . 11
  - 2.2.3 Integrated planning . . . 12
- 2.3 Line planning . . . 13
  - 2.3.1 Line planning models . . . 14
  - 2.3.2 Ongoing research . . . 15

3 Modeling 17
- 3.1 Concepts and notation . . . 17
  - 3.1.1 The ε-constraint method . . . 17
- 3.2 Base model . . . 19
  - 3.2.1 Terminal station constraints . . . 21
- 3.3 Passenger convenience models . . . 22
  - 3.3.1 Congestion model . . . 22
  - 3.3.2 Direct passengers model . . . 23
  - 3.3.3 Initial waiting time model . . . 26
- 3.4 Model size . . . 29

4 Experimental evaluation 31
- 4.1 Data generation . . . 31
- 4.2 Pareto solutions . . . 33
- 4.3 Solving times . . . 35
- 4.4 Terminal importance . . . 37
- 4.5 Fleet size results . . . 44
- 4.6 Intercorrelations . . . 48

5 Conclusions 53

References 55

Symbols and abbreviations

Symbols

C	Capacity of a train
d_{uvl}	Number of direct passengers from u to v using line l
E	Set of edges (between consecutive stations)
e_{uv}	First edge in the path from u to v
F_{ek}	Binary variable indicating if sum of line frequencies on edge e equals k
$F^{e}_{\max}$	Upper bound for k in F_{ek}
$\mathcal{F}$	Fleet size
f_l	Frequency of line l
$f_{\max}$	Upper bound for frequency of any line
h_l	Headway of consecutive trains on line l
L	Line pool
L^{uv}	Set of lines that contain stations u and v
$\mathcal{L}$	Set of lines in the line concept
$(\mathcal{L}, f)$	Line concept
l^{uv}	Subpath of line l from u to v
λ	Minimum edge load factor
λ_e	Edge load factor of edge e
$\lambda_{\min}$	Lower bound for minimum edge load factor λ
OD	Origin-destination-matrix
p	Percentage requirement for direct passengers
T_l	Round trip time of line l
τ	Planning period
V	Set of nodes (stations)
V^T	Set of terminal stations
V^T_l	Both end stations of line l
V^T_{PTN}	Both end stations of the PTN
$W_{\max}$	Upper bound for initial waiting time
w_e	Edge-wise passenger loads
x_l	Number of trains allocated to line l
z_t	Binary variable for including terminal t
$z_{\max}$	Upper bound for number of terminals

Abbreviations

PTN	Public transportation network

1 Introduction

Operations research has long been interested in the optimization of public transportation, due to its complex and dynamic nature, as well as its broad relevance globally. Historically, the primary goals of public transit have involved mitigating the growing population and urbanization as well as providing everyone with affordable options for traveling. Simultaneously, public transportation decreases the negative effects related to individual transportation, including traffic congestion and environmental pollution in urban areas. More recently, the counteracting impact on climate change has further emphasized the importance of public transportation and, consequently, its optimization. To reduce individual traffic, public transportation needs to offer a sufficiently attractive alternative for travel by improving the passengers' convenience while keeping the expenses low. Meanwhile, it must remain profitable for the operators to provide these services. The differing preferences of both passengers and operators hence present a natural opportunity for optimization of public transportation, out of which line planning is the component of primary interest in this thesis.

Public transportation planning is conventionally divided into subproblems based on the time span and hierarchical order of the decisions [Desaulniers and Hickman, 2007]. In the first stage, the public transportation network (PTN) is designed, whereas in the later subproblems, the decisions are related to timetables and schedules. In this thesis, the focus is on the second phase, line planning, where the network obtained in the first phase is considered as input. Line planning is responsible for making the long period decisions about which lines are operated and how regularly [Schöbel, 2012]. The literature on line planning is very extensive, which is due to its relatively long history and vast number of objectives. The focus in line planning research has been mainly oriented towards passengers' perspective objectives, while operator's preferences are mostly regarded in the operation costs. From the passengers' perspective, the objectives include metrics that measure aspects like travel time, waiting time, and transfers between vehicles. This thesis continues this research by comparing elements of passenger convenience in terms of waiting and transfers but also includes less studied features such as congestion in the vehicles. Also, instead of costs, the operator's side is studied through the minimization of the fleet size. Finally, this thesis adds to the existing research on line planning by focusing specifically on linear public transportation networks, which are often found in railway and metro transportation.

The major goal of this thesis is to develop three passenger-oriented line planning models aiming to improve the convenience of the line structure for the passengers. The objectives for these models are: minimizing congestion, minimizing initial waiting time, and maximizing the number of direct passengers. These models are specifically tailored for linear public transportation networks, and therefore, studying the advantages associated with modeling these networks is also recognized as a significant objective. The models are formulated as mixed-integer linear programs where the passengers' perspective objective is optimized and the operator's objective is regarded as a constraint, by means of the ε-constraint method [Ehrgott, 1999]. However, the perspectives can be swapped, leading to the operator's perspective models, which are also studied. To analyze the models, data is generated for both the input PTN and the passenger demand, reflecting the scale of real-world applications. These datasets, and their variations, are employed to compare the passenger convenience metrics and to investigate the models' properties. Particularly, it is examined how sensitive the models are to changes in the line pool via the change of terminals as well as how crucial the fleet size is for passengers' convenience. In addition, the passenger metrics are compared to one another to determine the degree of consistency between the models' outputs.

This thesis is structured as follows. Section 2 provides context for line planning by presenting relevant literature on the field and explaining its key concepts. Section 3 introduces the three line planning models developed in this thesis and discusses the aspects of modeling linear PTNs. The results of the evaluations are presented in Section 4, which is divided into subsections based on the themes of the experiments. Finally, conclusions are drawn in Section 5.

2 Background

2.1 Public transportation planning

Public transportation planning, also known as transit planning, is a wide research area aiming to efficiently plan transportation systems, from designing the infrastructure to scheduling crew shifts. One aspect contributing to the vast literature on transportation planning arises from the long history of public transportation. More importantly, transit planning is relevant to operations research due to the complex nature of the problems it involves. The complexity of the field originates from the large problem instances as well as the numerous variables and constraints related to dealing with moving vehicles, passengers with personal preferences, and drivers, all interacting with each other [Desaulniers and Hickman, 2007]. Moreover, considering that every major city worldwide has some form of public transportation infrastructure, the relevance of the field extends globally.

Public transportation planning is driven by the preferences of the two major sides, passengers and operators. In concise terms, passengers would like to travel as conveniently and quickly as possible, while operators are interested in minimizing costs and maximizing profits. Various factors affect the passengers' perceived convenience. Firstly, passengers value a short travel time. In the public transportation context, in addition to the actual time spent in the vehicle, the total travel time includes waiting and transfer times. Both the waiting and transfer time for a given journey are highly dependent on the configuration of the public transportation system. Secondly, passengers would like to travel directly from their origin to their destination, thus multiple transfers for a given journey will likely cause the passengers to choose alternative means of travel. In addition, congestion in vehicles and stations is an aspect responsible for reducing the attractiveness of public transportation among the users.

While passengers have diverse requirements for convenient traveling via public transportation, the operators are responsible for providing these services using limited resources. The operator's interest is to maximize profits by appealing to more customers while simultaneously minimizing costs emerging from the investments to the infrastructure and running the operations. Thus the operators seek to optimize for a reasonable price-quality ratio to prevent losing customers to alternative transport options [Blanco et al., 2020]. Alternatively, public transportation operators are often funded by public capital and hence the operators need to provide the services utilizing a limited budget. In conclusion, the problems within public transportation planning deals with contradicting objectives characterized by the passengers' and operator's preferences. The following section introduces the various phases of public transportation planning.

2.2 The global process

As mentioned in the previous section, public transportation planning is responsible for solving problems in various parts and stages within the transportation system. The most fundamental problem involves decisions about where to build the rails

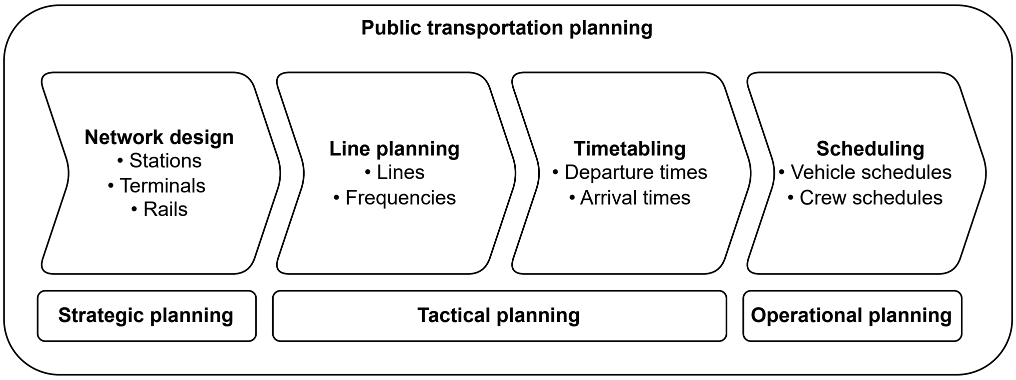


Figure 1: Stages in public transportation planning. Figure adapted from [Goossens et al., 2006].

and stations while scheduling as well as allocating drivers and vehicles occur at the opposite end of the spectrum. The ensemble of these steps is called the global process in public transportation planning [Guihaire and Hao, 2008]. Given the diverse problem description of the global process, the input of the problem ranges from the topological features of the area to the passenger demand, and finally to the available fleet, personnel, and budget. Albeit solving the global problem at once would guarantee an optimal outcome, the global problem is excessively complex to be solved in practice.

To solve the global process of public transportation planning, the problem is divided into subproblems, which are solved in a sequential manner. The subproblems naturally emerge from the hierarchy of the steps in the planning process. That is, constructing a timetable requires having the lines for which to set the arrival times. Furthermore, the lines are derived based on the rail infrastructure, which is modeled in the first phase. The literature identifies four main subproblems in transit planning: network design, line planning, timetabling, and scheduling, with some room in the definitions and further subdivisions [Desaulniers and Hickman, 2007], [Schöbel, 2017b], [Bussieck et al., 1997b]. In addition, post-planning-phase optimization steps, such as rescheduling and delay management are also subject to study [Schöbel, 2017b].

Figure 1 depicts the four main subproblems along with their conventional decision outcomes in the order in which the subproblems are typically solved. The figure also presents the division of these subproblems into strategic, tactical, and operational planning problems, aiming to emphasize the planning horizon of these steps. Strategic planning includes the network design phase, where the decisions are effective for decades since rails and roads have long-lasting lifetimes and the investments to those are substantial. Line planning and timetabling stages are included in the tactical planning category since they comprise decisions that might be subject to change on a seasonal basis. That is, there might be occasional updates to lines or frequencies, which requires solving the problems more frequently. Finally, operational planning contains vehicle and driver scheduling, which are solved often, given the changing passenger demand, and personnel availability. The division into strategic, tactical, and operational planning also varies slightly in the literature, the partition in Figure 1 aligns with [Desaulniers and Hickman, 2007].

2.2.1 Network design

The first step in public transportation planning is network design, where the aim is to design a railway or road network in such a way that it would allow the passengers to travel conveniently through the network in a cost-efficient manner for the operator. The output of the network design stage is a public transportation network, PTN, which serves as a foundation for the consecutive phases in the transit planning framework [Ceder and Wilson, 1986]. In essence, a PTN contains all relevant information on the infrastructure of the public transportation system including the stations and terminals, and the roads or rails connecting them. In practice, PTNs are often, quite intuitively, modeled as graphs, which also demonstrates the network design problems' close relation to general network flow problems [Schöbel, 2017b].

Optimizing for the passengers' perspective objectives, such as travel time minimization, requires information about the passenger demand within the public transportation network. The central input data concept is the origin-destination-matrix (OD-matrix), which contains estimated or gathered data on the number of passengers willing to travel between each station within the PTN or between zones in the area, during a specific time period.

The scope of questions in network design extends beyond the mere creation of a network system from scratch, such that often existing infrastructure serves as a starting point for the planning. It is a common task to improve an existing PTN by expanding the network to cover broader areas as demonstrated in [Blanco et al., 2011]. Additionally, there are instances where a predefined set of possible rails are available, and hence it is the model's responsibility to select the best possible subset of lines to be built, as detailed by [Drezner and Wesolowsky, 2003]. In general, the tasks in network design include a variety of activities, all with the common goal of establishing an optimal base for the implementation of the line planning stage. Line planning is covered in more detail in Section 2.3.

2.2.2 Timetabling and scheduling

Following the determination of lines and frequencies in the line planning phase, a timetable is generated during the subsequent planning stage, timetabling. The resulting timetable defines the departure and arrival times for each line. A line is a path in the PTN, on which vehicles travel from one end to another. As a result, a timetable also gives approximate stopping times for each station. As in public transportation planning in general, timetabling also aims to minimize operator side costs and improve passenger comfort by minimizing waiting times. In addition, one objective in timetabling is to establish a timetable that is robust enough to withstand unexpected delays in the schedule [Kroon et al., 2009].

A timetable can be either periodic or non-periodic, wherein the former option the departure times occur in regular time intervals. Periodic timetables excel in situations where the service frequency is high and the demand is sufficiently stable [Kroon et al., 2009]. In addition, a periodic timetable is more user-friendly due to the easily memorable departure times. However, the regular departure times are not necessarily optimal in the case of dynamic demand, which usually arises from commuter traffic. A non-periodic timetable accounts for the fluctuating demand, which results in a more efficient utilization of resources and reduced waiting times during peak hours. [Barrena et al., 2014] introduce an exact algorithm for the timetabling problem in the dynamic demand environment, where no specific shape is assumed for the demand function.

Once the timetable is set, specific vehicles need to be assigned for specific lines and departures. Conducting this procedure efficiently yields the vehicle scheduling stage in the public transportation planning framework. In vehicle scheduling, the objective is generally cost minimization, considering both fixed and operational costs related to fulfilling the timetable [Bunte and Kliewer, 2009]. In addition to assigning the vehicles, personnel shifts are also to be planned. The crew scheduling phase includes assigning the drivers and conductors to specific departures. The additional rules and restrictions regarding staff utilization make the crew scheduling problem challenging. That is, the problem must involve constraints related to breaks and driving limits of drives as well as considerations for holidays and weekends [Schöbel, 2017b].

2.2.3 Integrated planning

As discussed in previous sections, the global transit planning problem is divided into sequentially solved subproblems to reduce the vast complexity of the whole procedure. The sequential solving approach implies that the solution of the previous phase is used as an input for the following phase. This process can be interpreted as a greedy method for solving the global problem, where each solution represents a local optimum for the given stage [Schöbel, 2017a]. However, the solutions obtained from the subproblems may not collectively yield an optimal solution for the global problem. For instance, in the long run, the primary expenses in the public transit system are generated on the operational side, which are decided in the final stages of the planning scheme. Consequently, the operational costs depend heavily on the earlier solutions, even though they were not considered at that time. This issue motivates the research for developing models to solve multiple stages simultaneously, which is referred to as integrated planning.

Currently, there exist models that can solve two or three stages of the global problem in an integrated manner. [Blanco et al., 2020] propose a model that integrates line planning and timetabling phases in the general metro transportation setting. In the approach by [Michaelis and Schöbel, 2009], line planning, timetabling, and vehicle scheduling are integrated by rearranging the conventional order of the steps to incorporate the operational costs into line planning considerations. Moreover, [Schöbel, 2017a] introduces a single model for solving the same steps in one go, while also providing an iterative heuristic to reduce the complexity of the integrated model. Finally, in the model by [Liebchen and Möhring, 2007], the timetabling problem is enriched with decisions involving network design, line planning, and vehicle scheduling.

2.3 Line planning

The decisions made in the line planning stage are often effective for rather extended periods and thus line planning is part of tactical planning, as illustrated in Figure 1. The goal in line planning is to define the paths on which the vehicles travel as well as how often they operate. In transit planning terms, this means constructing the lines and setting their frequencies. Sometimes the lines are created from scratch, but often the lines are selected from a predefined pool of possible lines. The resulting set of selected lines with their operation frequencies is called a line concept.

Line planning builds upon the public transportation network that is acquired in the earlier stage of the planning process, i.e., network design. Conventionally, PTNs are modeled as graphs (V,E), where the set of nodes V represents the stations or stops, and the set of edges E represents the links (rails or roads) between them. In addition to the underlying PTN, the line planning step employs information regarding passenger demand within the PTN as its input. Like in network design, in line planning, the passenger demand is also expressed as an origin-destination matrix. However, in some line planning models, it is more useful to have the demand aggregated on the edge level. This poses a challenge since some information on the passengers' route choices within PTN must be estimated, as highlighted in [Schöbel, 2012].

Among the diverse set of objectives in line planning models, one goal is formulating a line concept that is feasible in the sense that the operation is possible as well as that all passenger demand can be transported. Usually, the passenger demand is estimated for a certain planning period for which the frequencies are also determined. The frequencies of the selected lines should be high enough to allow transporting the passengers to their destinations considering the capacities of the available vehicles. Moreover, in numerous line planning models, the frequencies are explicitly constrained per edge [Goerigk and Schmidt, 2017]. Lower bounds for edge frequencies are simply included to ensure minimum service for certain parts of the PTN. Upper bounds, however, could be used to restrict traffic around busy areas to guarantee the safe operation of the system. That is, keeping sufficiently large headways (time intervals) between consecutive vehicles, which is especially important in rail transport.

Like with public transportation planning in general, line planning is also subject to dealing with conflicting objectives from the passengers' and operator's sides. In the more conventional approach, a line planning model handles one of the perspectives in the objective function, for instance, by minimizing some passenger convenience metric. Meanwhile, it regards the other perspective on the constraint side. In the literature, the operator's perspective is regularly considered through the operational costs that a given line concept would generate [Schöbel, 2012]. Because the operator's benefit is usually measured solely through costs, these models are often referred to as cost-oriented models. In the passengers' perspective, however, the alternatives for measuring convenience are considerably broader. Passenger convenience metrics in line planning usually relate to how long it takes to travel as well as how directly the passengers can travel to their destinations, that is, how often the passengers need to transfer vehicles during their journey. In addition to the time spent in the vehicles, waiting times and transfer times are also acknowledged, when measuring the passenger convenience. The objective functions for both perspectives are studied in further detail in Section 2.3.1. There also exist models where both perspectives are simultaneously regarded in the objective functions. For instance, [Blanco et al., 2020] employ scalarizations in the objective function to find a balance between the two perspectives.

2.3.1 Line planning models

Cost-oriented line planning models aim for designing a line concept that can be operated with minimal expenses to the operator, while still satisfying requirements for service quality. The model proposed by [Claessens et al., 1998] serves as a foundation for models focused on minimizing costs. Since the model allows for multiple different types of railway transportation, the costs are separately estimated for different vehicle types. In addition, the model distinguishes between fixed costs for adding more cars as well as variable costs related to the distance traveled by each vehicle type and car. The nonlinear objective function is then linearized with the introduction of binary variables indicating the inclusion of a certain vehicle and car for each line and frequency separately. Feasibility and service quality are considered through frequency and capacity constraints. In the work by [Bussieck et al., 2004], the model by [Claessens et al., 1998] is further developed to allow for faster solving times. Particularly, the objective function is reformulated to obtain a linear expression. In addition, the solving procedure is enhanced via a heuristic that reduces the solution space by fixing variables related to sub-optimal routes to zero.

Shifting attention to passenger-oriented models, the direct passengers approach has been of great interest in the earlier literature. In the direct passengers (or travelers) model by [Bussieck et al., 1997a], the objective is to maximize the number of passengers that can travel directly from their origin to destination without the need for intermediate transfer between lines. The model assumes that all passengers select the shortest path available for their journey, which is a only realistic assumption in a long-distance public transportation setting. Furthermore, the model is constrained to require that all direct passengers can be transported with the resulting line concept. However, this condition is not required for non-direct passengers. A revised version of the direct passengers model is introduced in [Bussieck, 1998], where the frequency constraints are relaxed to allow for more flexibility in the solutions. Despite the intuitive and quite straightforward nature of the objective function, the maximization of the number of direct passengers brings some issues to the model. That is, when increasing the number of passengers that can travel directly, the lengths of the lines tend to increase. As denoted by [Goossens et al., 2006], this results in unused capacity in the less busy parts of the lines, since more and more passengers travel using fewer lines, which increases the edge-wise capacity requirements. Also, optimizing merely for direct passengers omits the fact that often transferring lines once during a trip might reduce the traveling time significantly. This conclusion gives motivation to passenger-oriented models that consider traveling time in the objective function.

In addition to the directness of the journey, the time spent traveling from origin to destination is also a significant element affecting passengers' convenience in public transportation. The total journey time consists of multiple parts, including the waiting time in the origin station, possible waiting for connecting lines as well as all the time spent traveling in the vehicles. All combinations of these components have been studied in the literature, however, it is noteworthy to discuss the work of [Schöbel and Scholl, 2006] for their novel approach to include the transferring aspect into the traveling time. To model transfers, the passengers' routes through the PTN need to be determined. This task is included in the line planning model with the introduction of a change-and-go network, which models the transfers via links between lines and associates time penalties to them. That is, minimizing the travel time, the model then allows passengers to prefer transfers if it reduces the overall time spent traveling. Nevertheless, the transfer times are only approximations, since the exact values cannot be determined without the timetable, which is planned after the line planning stage. To overcome this limitation, [Kaspi and Raviv, 2013] integrate timetabling into line planning, which yields more precise comparisons between different route options. Moreover, the integrated model also covers the waiting that occurs before boarding the first vehicle at the origin station to further improve the total journey time examination. In the model proposed by [Cancela et al., 2015], waiting time for the next line, at different stations, is regarded as a decision variable. This way, the nonlinear relationship between frequencies and waiting times can be controlled, and the waiting time can be minimized directly in the linear objective function.

In the recent model by [Wu et al., 2023], capacity utilization is regarded as both the operator's and the passengers' side constraint. Particularly, the model has lower and upper bounds for load factor, which measures the ratio between the number of passengers using a line and the capacity provided by the line. In the passenger perspective, the upper bound ensures that the service quality remains reasonable as high congestion is restricted. On the other hand, the lower bound for the load factor assures that the utilization of resources is effective, which is of interest to the operator. Also, the work of [Wu et al., 2023] is interesting for their daily approach to line planning. While acknowledging the tactical and long-term nature of the line planning stage, the authors propose a line planning model that adjusts a reference line plan for daily demand fluctuations. That is, the daily-run model suggests small changes in the frequencies of the lines by adding more vehicles when the demand is higher than usual, and vice versa. This way, the operator can minimize costs, but also provide better service when demand is increased momentarily. The potential changes proposed by the model are constrained with a reference line plan, which is calculated, conventionally, based on long-term demand. This restriction to the daily maneuver ensures that the adapted line plans remain robust and convenient for the passengers.

2.3.2 Ongoing research

One of the more recent topics of research in line planning is the line pool generation phase. The models discussed thus far have been based on a predefined line pool, out of which the line concept is constructed. Therefore, the outcomes of these models are heavily influenced by the line pool. Since considering all possible lines is impossible, due to the exponential amount of possibilities, some procedure is needed to regard only potential lines. [Gattermann et al., 2017] developed an algorithm for generating feasible line pools to be used as inputs in the line planning phase. Moreover, the line pools provided by the algorithm are reported to allow fast running times for the subsequent line planning models. The generation of a line pool can also be integrated into the line planning algorithm itself, as proposed by [Borndörfer et al., 2007]. The authors developed a line planning model that, instead of taking a line pool as input, generates the lines concurrently with their frequencies. This approach allows routing the passengers independently of the travel mode, such that transfers between different transportation types can be regarded in the line planning phase.

In addition to line pool generation and integrated planning (discussed in Section 2.2.3), the notion of stopping patterns is a subject that is currently being researched in the context of line planning. Stop planning allows for trains to bypass some stations, and hence the decision of at which stations certain line halts is included in line planning. The resulting plan for halting stations is then referred to as a stopping pattern. The possibility of avoiding stopping at some stations increases the flexibility of the line concept, in the sense that express lines can only stop at busy stations for reducing travel time, whereas slower lines can stop at all stations thus ensuring service at all stations. [Goossens et al., 2006] first introduced a cost-minimizing line planning model that incorporated different stopping patterns. The model combines different types of trains in the same line planning problem, such that the faster trains utilize a smaller, predetermined subset of stops that are considered when optimizing the line plan. In the integrated model by [Burggraeve et al., 2017], line planning is combined with timetabling, with robustness as one of the main targets. The model also allows for different stopping patterns for lines, which is regarded as one of the standard decisions of the line planning phase, along with the line frequencies. Finally, [Li et al., 2019] also integrate decisions on stopping locations and timetables in the model that incorporates both express and local railway transportation. It is noteworthy, that the express trains are allowed to overtake the local trains, which stop at every station. The overtaking stations are known in advance because they need to contain multiple rails for overtaking to be physically possible. The timetabling aspect must be included in the line planning phase since exact overtaking times have to be coordinated between the express and local trains as the locations for overtaking are limited.

3 Modeling

In this section, the main contributions of this thesis are presented. Particularly, three passenger-oriented line planning models, tailored for linear PTNs, are detailed. Prior to focusing on the models, the basic concepts in line planning are formally introduced in Section 3.1, along with the notation used to describe them. The section also addresses the ε-constraint method, which is utilized in this thesis to handle the conflicting objectives between passengers and the operator. Thereafter, in Section 3.2, a base line planning model is introduced to serve as a foundation for the passenger-oriented line planning models. These three models are derived and discussed in Section 3.3. Finally, Section 3.4 delves into an exploration of the models' size.

3.1 Concepts and notation

The notation used to formulate the concepts and models in this thesis follows mostly the notation used in [Schöbel, 2012]. The underlying public transportation network (PTN) is given as an undirected graph PTN = (V, E), where V is the set of nodes used to model the metro stations and E is the set of edges representing the rails between the stations. Since the focus of the models developed in this thesis is on linear metro networks, it follows that the PTNs are inherently linear graphs. In addition to the metro network, a line pool, denoted with L, is given as an input to the models. The line pool consists of potential lines l, which are predetermined paths in the PTN. The task is to associate the lines with the operating frequencies f_l, which indicate how often service is offered along line l within the planning period τ. Ultimately, the set of selected lines $\mathcal{L}$ and their frequencies $f = (f_l)_{l \in \mathcal{L}}$ together form a line concept $(\mathcal{L}, f)$, which is the primary output of a line planning model.

The passenger demand data is given in the format of an origin-destination-matrix, OD. The OD-matrix specifies the number of passengers willing to travel from origin station u to destination station v, for all $u, v \in V$. The estimate of the demand must be given for the same planning period τ for which the line concept $(\mathcal{L}, f)$ is planned. In this thesis, an additional demand data representation is also utilized. Namely, the edge-wise traffic loads w_e are needed to ensure that the train capacity limit is not exceeded in any segment (edge) in the metro network during the planning period. In general, the edge loads cannot be derived from the OD-matrix alone since passengers can select diverse routes through the PTN and hence the edge loads can vary based on passengers' behaviour. However, in the case of a linear PTN, edge loads w_e can be explicitly determined from the OD-matrix because there exists only one simple path between any pair of stations in V. For this specific case, the load on edge e is calculated by considering the cumulative passenger demands in both directions separately and then selecting the higher value as the resulting edge load w_e.

3.1.1 The ε-constraint method

Preceding the formulation of the base model in the following section, some aspects regarding multi-criteria optimization are discussed next. As addressed in Section 2.3,

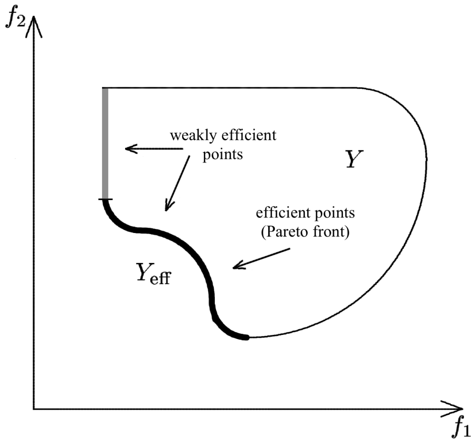


Figure 2: Illustration of the Pareto front and the weakly efficient points in a bi-objective setting. Figure adapted from [Ehrgott, 1999].

line planning involves optimizing for the conflicting objectives of passengers and the operator and hence an approach for dealing with multiple objectives is needed.

Optimality in multi-criteria optimization is characterized by the notion of *Pareto optimality*. The subsequent definitions are based on the lecture notes by [Ehrgott, 1999], in which the following general multi-objective problem is considered

$$\min \quad \{f_1(x), ..., f_k(x)\} \tag{1a}$$

$$\text{s.t.} \quad x \in X, \tag{1b}$$

where $k \geq 2$ objective functions $f_i : \mathbb{R}^n \to \mathbb{R}$, $i \in \{1, ..., k\}$, are minimized in the feasible region X. Next, two definitions are presented to formulate Pareto optimality and efficiency, along with their less strict counterparts.

Definition 1 A point $x^* \in X$ is called *Pareto optimal*, if there is no $x \in X$ such that $f_i(x) \leq f_i(x^*)$ for all $i = 1, ..., k$ and $f_j(x) < f_j(x^*)$ for some $j \in \{1, ..., k\}$. If x^* is Pareto optimal, then $f(x^*)$ is called *efficient*, where $f(x)$ denotes the vector of objective functions $f_1(x), ..., f_k(x)$. The set of all Pareto optimal $x^* \in X$ is X_{Par}. Let $Y = f(X)$. The set of all efficient points $y = f(x^*) \in Y$ is Y_{eff}.

Definition 2 A point $x^* \in X$ is called *weakly Pareto optimal* if there is no $x \in X$ such that $f_i(x) < f_i(x^*)$ for all $i = 1, ..., k$. If x^* is weakly Pareto optimal, then $f(x^*)$ is called *weakly efficient*.

The definitions suggest that Pareto optimal points are solutions such that they cannot be further improved in any criterion without sacrificing any of the remaining criteria. The objective function values for such solutions form the set of efficient solutions. These efficient solutions are often referred to as the Pareto front since they form a frontier in the feasible objective region Y, as depicted in Figure 2. The figure also illustrates the set of weakly efficient points, which share similar properties with the efficient points that make up the Pareto front. The weakly efficient points differ from the efficient points such that they allow the existence of solutions where some of the objective functions could be improved without worsening the others, as long as there exists at least one objective for which such condition does not hold. The Pareto front representing the efficient solutions suggests that there remains a set of solutions that are not superior to one another. That is, moving along the Pareto front results in trade-offs between the objectives.

In this thesis, the line planning problems are solved by employing the ε-constraint method, described by [Ehrgott, 1999]. In this method, only one objective function is minimized, while the other objectives are formulated as constraints. The ε-method is obtained by expressing the Problem 1 as

$$\min \quad f_i(x) \tag{2a}$$

$$\text{s.t.} \quad f_j(x) \leq \varepsilon_j \forall j = 1, ..., k, \; j \neq i \tag{2b}$$

$$x \in X. \tag{2c}$$

Using the ε-constraint method thus implies that the trade-offs are given explicitly as the constants ε_j for all objective functions but one. It is shown in [Ehrgott, 1999] that the solutions acquired with the ε-constraint method are weakly Pareto optimal. Furthermore, if the solution is unique, a Pareto optimal solution is obtained. The problems in this thesis have two objective functions, one describing the passengers' convenience and the other related to the operator's interest. Therefore, utilizing the ε-constraint method leads to optimizing the objective of one perspective while imposing a constraint on the other via a constant ε.

3.2 Base model

This section introduces a base model for the line planning problem that is further augmented later to incorporate different aspects of passenger convenience. The characteristics of the base model are carried along to these convenience-specific models. Next, the assumptions needed to formulate the models are discussed briefly.

As discussed earlier, the models are developed to solve line planning problems specific to linear metro networks and thus some modeling decisions do not generalize to settings with more complex PTNs. Moreover, only a single type of train is considered such that there are no speed or capacity differences between the trains. This also implies that stopping patterns are not regarded in these models and hence all lines stop at every station they include. The lines in this work are modeled as round trips starting from the origin station and then driving to the end station and back. A constant T_l is used to denote the round trip time (in minutes) of line $l \in L$. It is also assumed that the track segments are equipped with rails in both directions, eliminating the need to address conflicts with oncoming traffic. The planning period τ is set to one hour. Thus the demand data, given in OD-matrix (OD) and edge load vector $w = (w_e)_{e \in E}$, is also given for an hour period. As is common in line planning models, the assumption here is that passengers arrive at stations uniformly over the planning period. Finally, the models are simplified such that all trains are set to have a fixed capacity of $C = 600$ people, throughout this work.

In the base model, the objective function is denoted with an abstract passenger inconvenience function $P(y)$, where y is a vector of decision variables related to the passenger inconvenience. The passenger inconvenience function is minimized in the passenger perspective base model, whereas the operator's interest is acknowledged in the constraints side, as in the ε-constraint approach. Commonly in line planning models, the operator's preferences are modeled with costs related to operating the line plan, as discussed in Section 2.3.1. Specifically, in this thesis, the operator's interest is in minimizing the number of trains required to operate the line concept. The number of trains is referred to as fleet size and denoted with $\mathcal{F}$. Hence, the costs are considered indirectly as the fleet size is a large contributor to the overall costs. Also, instead of regarding the line frequencies f_l as decision variables, the number of trains allocated to each line x_l is decided. When $x_l = 0$, it indicates that line l is not included in the line concept. The line frequencies can be derived from the decision variables x_l using the following formula

$$f_l = \frac{\tau}{T_l} x_l, \tag{3}$$

which suggests that one train offers service τ / T_l times in the planning period τ. When multiplied by the number of trains, the frequencies are obtained.

Finally, the base model has a capacity constraint that is shared with the passenger perspective models. Particularly, the models require that all passenger demand, given in OD-matrix, or alternatively in edge load vector, can be transported within the planning period. With these constraints and the objective function, the base model can be formulated as a mixed-integer linear program. The base model (in the passengers' perspective) reads

$$\begin{aligned}
\min_{x,\, y} \quad & P(y) && \text{(4a)}\\
\text{s.t.} \quad & \sum_{l \in L} x_l \leq \mathcal{F} && \text{(4b)}\\
& \sum_{l \in L : e \in l} C \frac{\tau}{T_l} x_l \geq w_e \text{ for all } e \in E && \text{(4c)}\\
& x_l \in \mathbb{N}_0 \text{ for all } l \in L. && \text{(4d)}
\end{aligned}$$

In Problem 4, the abstract passenger inconvenience function $P(y)$ is minimized subject to the fleet size constraint (4b), which adds the allocated trains over all lines $l \in L$ to ensure that the fleet size limit $\mathcal{F}$ is not exceeded. In addition, the model is constrained by the capacity constraints (4c) for every edge $e \in E$. The sum over $l \in L : e \in l$ adds the capacity over all lines operating along edge e to satisfy the passenger demand w_e on that edge.

The Problem 4 is referred to as the base model in the passengers' perspective since the objective function concerns the passengers' inconvenience. However, based on the

ε-constraint method, the base model can be expressed in the operator's perspective as well. In this case, the passengers' inconvenience function is constrained with a constant $P_{\max}$, and the objective function is replaced by the minimization of the fleet size. Hence, the operator's perspective base model can be formulated as

$$\min_{x,\, y} \quad \sum_{l \in L} x_l \tag{5a}$$

$$\text{s.t.} \qquad P(y) \le P_{\max} \tag{5b}$$

$$\sum_{l \in L : e \in l} C \frac{\tau}{T_l} x_l \ge w_e \text{ for all } e \in E \tag{5c}$$

$$x_l \in \mathbb{N}_0 \text{ for all } l \in L. \tag{5d}$$

In this thesis, models of both perspectives will be studied in detail. Prior to introducing the different passenger convenience models, the base model is enriched with a possibility to consider only a subset of terminal stations, where the lines can end.

3.2.1 Terminal station constraints

It is assumed in this work, that the lines cannot terminate at every station. Instead, there is a predefined set of stations, where the trains have the ability to turn around. This assumption is reasonable since turning around requires additional rails and thus it would not be feasible to build such infrastructure at every station. The set of these terminal stations is denoted with V^T. Nevertheless, the formulations of the base models don't explicitly include this set since information on the terminals is indirectly contained in the line pool L. That is, the line pool contains only predetermined lines such that the terminals are regarded already in the line pool generation phase.

Next, a set of constraints is added to the base models to enable consideration of only a subset of terminals and lines associated with them. For this purpose, two more sets are needed to model the terminals. The set V_l^T includes both end stations (terminals) of line l, whereas the set V_{PTN}^T includes the two ends of the linear PTN. Moreover, the decision variables $z_t \in \{0, 1\}$ are added to signify whether or not a terminal $t \in V^T$ is included in the line concept. Excluding a terminal t means that none of the lines starting or ending at terminal t can be established. The following set of constraints is introduced to model the limitation of terminals.

$$\sum_{t \in V^T} z_t \le z_{\max} \tag{6}$$

$$x_l \le M z_t \quad \text{for all } l \in L,\ t \in V_l^T \tag{7}$$

$$z_t \ge 1 \quad \text{for all } t \in V_{\text{PTN}}^T \tag{8}$$

$$z_t \in \{0, 1\} \quad \text{for all } t \in V^T. \tag{9}$$

Above, Constraint 6 restricts the number of terminals that can be established with the upper bound $z_{\max} \in \{2, ..., |V^T|\}$. Constraint 7 is responsible for preventing the allocation of trains to lines that do not have designated terminals. The Big M method is utilized to enforce the condition that x_l must be set to zero when a terminal t, serving as either end of line l, is not selected. In this context, setting $M = \sum_{u,v \in V} OD_{uv}$ ensures that the constraint is effective in all scenarios since the number of trains required to establish the line concept can never exceed the total number of passengers. Finally, Constraint 8 requires that both ends of the PTN remain as terminal stations to maintain operation in the outermost edges. However, this constraint becomes redundant, if a nonzero passenger load is assumed between every edge because the capacity constraints (4c and 5c) guarantee service on all edges with some demand. It is worthwhile to note that the terminal constraints (6 - 8) are the same for both perspectives of the base model and for all passenger convenience models discussed in the next section.

3.3 Passenger convenience models

In this section, the base model is expanded by replacing the abstract passenger inconvenience function $P(y)$ with varied metrics for the passengers' satisfaction and convenience, which will be maximized or minimized, depending on the formulation of the metric. In addition to specifying the objective function, additional constraints specific to the passenger convenience metrics are introduced. Three different approaches for minimizing passengers' inconvenience are considered in the subsequent sections.

3.3.1 Congestion model

The first approach in improving the passengers' convenience in line planning is related to the less researched aspect of congestion. The congestion model is based on the idea that crowding in train cars should be maintained at acceptable levels or minimized. Particularly, the capacity of trains should be allocated such that there would be enough available seats on every edge of the PTN. An intuitive way to minimize congestion would be to minimize the ratio between passenger load and available seats on every edge to obtain a value between 0 and 1. However, since the number of available seats is directly proportional to the number of trains on each edge (the decision variable), the objective function would become nonlinear. Therefore, the inverse of that ratio is maximized. To formulate this ratio, a measure called edge availability factor, denoted by $\lambda_e \in [1, \infty)$, is introduced. The edge availability factor is defined as

$$\lambda_e = \frac{\text{available seats on edge } e}{\text{passenger load on edge } e} \quad \text{(edge availability factor of edge } e\text{).} \tag{10}$$

The definition of λ_e can be expressed using the Constraint 4c from the base model, which states that the available capacity (number of seats) should be greater or equal to the load, on every edge $e \in E$. Hence, the edge availability factor for edge e is given by

$$\lambda_e = \frac{1}{w_e} \sum_{l \in L: e \in l} C \frac{\tau}{T_l} x_l. \tag{11}$$

To maximize passenger convenience, the edge availability factors should be maximized since they are defined as ratios between seat availability and demand. Instead of maximizing the edge-wise metrics separately in the objective function, the minimum edge availability factor, λ, is maximized. That is, maximize $\lambda = \min_{e \in E} \lambda_e$. In the case of maximizing or bounding λ, it is enough to require that $\lambda \leq \lambda_e$ for all $e \in E$. These constraints can be expressed using the definition in Equation 11, as follows

$$\sum_{l \in L: e \in l} C \frac{\tau}{T_l} x_l \geq \lambda w_e \qquad \text{for all } e \in E, \tag{12}$$

where setting $\lambda = 1$ returns the capacity constraints (4c).

Ultimately, the congestion model is derived from the base model (4 and 5) by replacing the inconvenience function and the capacity constraint accordingly. From the passengers' perspective, the congestion model takes the form

$$\max_{x, \lambda} \quad \lambda \tag{13a}$$

$$\text{s.t.} \quad \sum_{l \in L} x_l \leq \mathcal{F} \tag{13b}$$

$$\sum_{l \in L: e \in l} C \frac{\tau}{T_l} x_l \geq \lambda w_e \qquad \text{for all } e \in E \tag{13c}$$

$$x_l \in \mathbb{N}_0 \qquad \text{for all } l \in L \tag{13d}$$

$$\lambda \in [1, \infty). \tag{13e}$$

The operator's perspective congestion model is obtained by minimizing the fleet size $\sum_{l \in L} x_l$ and constraining the minimum edge availability factor as $\lambda \geq \lambda_{\min}$. Later, in Section 4, both the objective function metric λ and its more intuitive counterpart $1/\lambda$ are employed when analyzing the models.

3.3.2 Direct passengers model

In this section, a second line planning model is introduced, with an alternative way of increasing passenger convenience. The goal of this model is to allow as many passengers as possible to have a direct connection from their origin to their destination. Hence, the model is titled the direct passengers model. As discussed in Section 2.3.1, the direct passenger's approach is explored quite extensively in the literature, and the models introduced in them are far developed. However, by concentrating on models specifically designed for linear PTNs, certain challenges associated with the direct passengers approach can be resolved.

In the general case, where there can exist multiple paths between two locations within the PTN, defining a direct passenger is not straightforward. Namely, even if a direct line existed for some origin-destination pair (u, v), a passenger might still want to choose a route with one transfer, if it reduces the travel time significantly. Thus it would not make sense to maximize direct travelers since it would actually decrease the convenience for some passengers. For this reason, the definition of a direct passenger should require that the direct route is also the preferred one. Deciding which direct routes are preferred and which are not, poses a challenge in the general case. In a linear PTN, however, there exists only one simple path between any two stations $u, v \in V$. Therefore, the definition of a direct passenger is not ambiguous, and hence, a direct passenger can be regarded as someone who can travel from their origin to destination using only a single line.

A naive solution to the direct passengers problem would be to establish a line that would cover the whole PTN from one end to another. With that solution, all passengers could travel directly if enough capacity was provided for that line. Nevertheless, transporting all passengers on one line would require an extensive amount of capacity since the edge loads would accumulate, leading to inefficient utilization of resources. Also, this amount of capacity is typically either unavailable or too expensive, highlighting the relevance of line planning. Due to the limited capacity, not all passengers can be transported directly even if the traveling distance is shorter and a direct line exists. Particularly, if demand exceeds the direct line capacity for a route from u to v, some passengers need to use alternative lines. This aspect must be taken into account in the decision variables.

To count the direct passengers in the objective function, the direct passengers are regarded as decision variables along with x_l. Moreover, direct passengers traveling from u to v need to be assigned to specific lines. This allocation is necessary to follow the usage of various lines and, ultimately, to constrain it. Therefore, the decision variables related to direct passengers are denoted by d_{uvl}, which indicates the amount of direct passengers traveling from origin u to destination v using line l. The objective function is then to maximize the number of direct passengers over all origin-destination pairs ($u, v \in V$) and lines $l \in L^{uv}$ as

$$\sum_{u,v \in V} \sum_{l \in L^{uv}} d_{uvl}, \tag{14}$$

where $L^{uv} \subseteq L$ denotes the set of all lines that enable to travel directly between stations u and v. Analogously, the notation l^{uv} is used later to denote the subpath of line $l \in L^{uv}$ from u to v.

The decision variables d_{uvl} are subject to a few additional constraints. Firstly, it is enforced that the quantity of direct passengers does not surpass the demand for any origin-destination pair, as quantified by the OD-matrix. This constraint is formulated as

$$\sum_{l \in L^{uv}} d_{uvl} \leq OD_{uv} \qquad \text{for all } u, v \in V. \tag{15}$$

The second constraint ensures that the number of direct passengers using line l on edge e does not exceed the capacity of that line. This constraint is considered separately for each line $l \in L$ and for each $e \in E$ such that e is contained in l. Furthermore, both directions are also regarded independently since passengers traveling in opposite directions use separate trains on the same edge. Note that the directions are not considered separately for the capacity constraints (4c) because the edge loads w_e

provide the maximum load between the two directions. The direct passenger capacity constraints specific to edges and lines, in one direction, are expressed as follows

$$\sum_{\substack{u,v\in V: l\in L^{uv},\\ e\in l^{uv}, u>v}} d_{uvl} \leq C\frac{\tau}{T_l}x_l \qquad \text{for all } e\in E, l\in L \text{ with } e\in l, \tag{16}$$

where the directions are designated with $u > v$ and $u < v$. This formulation assumes that the stations are indexed sequentially. Including the capacity and fleet size constraints from the base model, yields the direct passengers model. From the passengers' perspective, the model reads

$$\max_{x,d} \quad \sum_{u,v\in V}\sum_{l\in L^{uv}} d_{uvl} \tag{17a}$$

$$\text{s.t.} \quad \sum_{l} x_l \leq \mathcal{F} \tag{17b}$$

$$\sum_{l\in L^{uv}} d_{uvl} \leq OD_{uv} \quad \text{for all } u,v\in V \tag{17c}$$

$$\sum_{l\in L: e\in l} C\frac{\tau}{T_l}x_l \geq w_e \qquad \text{for all } e\in E \tag{17d}$$

$$\sum_{\substack{u,v\in V: l\in L^{uv},\\ e\in l^{uv}, u>v}} d_{uvl} \leq C\frac{\tau}{T_l}x_l \quad \text{for all } e\in E, l\in L \text{ with } e\in l \tag{17e}$$

$$x_l \in \mathbb{N}_0 \qquad \text{for all } l\in L \tag{17f}$$

$$d_{uvl} \in \mathbb{N}_0 \qquad \text{for all } u,v\in V, l\in L, \tag{17g}$$

where the direction $u < v$ is handled analogously to Constraint 17e. The operator's side direct model would be straightforwardly obtained by exchanging the objective function and the fleet size requirement, similar to what is done with the congestion model. Particularly, fleet size would be minimized and some minimum requirement for the number of direct passengers would be imposed. However, it might not be simple to come up with a specific number of direct passengers that should be able to travel directly. Therefore, the operator's perspective ε-constraint is reformulated such that a percentage p of passengers is required to travel directly. This constraint is formulated as

$$\sum_{u,v\in V}\sum_{l\in L^{uv}} d_{uvl} \geq p \sum_{u,v\in V} OD_{uv}, \tag{18}$$

where the sum over the origin-destination-matrix gives the total number of passengers. Thus, replacing the Constraint 17b with 18 and minimizing the fleet size returns the operator's perspective direct passengers model, where the direct passengers requirement can be controlled with percentage p.

3.3.3 Initial waiting time model

In the third, and final passenger convenience-focused line planning model, the objective is to minimize the time that passengers have to wait for the first vehicle on their journey. Different aspects of waiting time have been considered in the literature, as outlined in Section 2.3.1, but little research has been done with the initial waiting time specifically in focus. Like with the direct passengers model in the previous section, the linear transportation network setting provides some simplification to the modeling. Because the metro network is linear, passengers' paths are again known in advance, and hence the subset of lines to be considered in calculating the waiting time is predetermined for each origin-destination pair. In this model, the objective function counts the expected waiting times at each station and factors in the corresponding passenger demand originating from each station.

To formulate the objective function, several assumptions are taken into account. First, it is assumed that a passenger boards the first train that arrives on the platform, provided it is heading in the same direction as the passenger. With linear PTN, this assumption is plausible since all trains use the same path. To model this behavior, all lines operating toward the passenger's direction need to be known. This is done by considering the first edge in the passenger's journey from u to v and collecting the lines operating on that edge. The first edge in the passenger's direction is denoted with e_{uv}. Then, the second assumption is related to the arrival of trains. That is, all trains operating on edge e_{uv} are assumed to arrive in even intervals at the station u, regardless of their frequency. With this assumption, it is possible to calculate the combined headway between consecutive trains (over all lines) on edge e_{uv} to estimate the waiting time. However, the combined headway is an approximation, as the exact value would require knowledge of the timetables. Finally, the passengers are assumed to arrive at stations evenly distributed. Specifically, the number of passengers specified in OD is assumed to arrive at even time intervals, similar to the trains. If the combined headway of potential trains is known, the assumption of evenly distributed passengers implies that the expected waiting time is half of the headway. Exploiting all the assumptions mentioned above, the objective function can be outlined as

$$\sum_{u \in V} \left(\sum_{v \in V} OD_{uv} \frac{1}{2} (\text{combined headway on edge } e_{uv}) \right), \tag{19}$$

which minimizes the expected initial waiting time of all passengers, weighted by the demand.

Next, the combined headway in (19) is formulated. The headway, h_l, for a single line l is defined as the round trip time divided by the number of vehicles allocated, $h_l = T_l / x_l$. Substituting the definition of line frequency f_l in (3), the headway can be expressed as $h_l = \tau / f_l$. To get the combined headway, the reciprocal of the sum of the reciprocals of the headways is considered. Hence, the combined headway on edge e_{uv} can be expressed as

$$\left(\sum_{l\in L:e_{uv}\in l}\frac{1}{h_l}\right)^{-1}=\left(\sum_{l\in L:e_{uv}\in l}\frac{f_l}{\tau}\right)^{-1}=\tau\left(\sum_{l\in L:e_{uv}\in l}f_l\right)^{-1}. \tag{20}$$

Using the formulation of combined headway in (20), the objective function becomes

$$\sum_{u\in V}\left(\sum_{v\in V}OD_{uv}\frac{1}{2}\tau\left(\sum_{l\in L:e_{uv}\in l}f_l\right)^{-1}\right). \tag{21}$$

The objective function in (21) suggests that in addition to x_l, also the frequencies f_l are regarded as decision variables. However, the objective function is nonlinear in f_l.

In order to obtain a linear objective function, a third set of decision variables is introduced. Let the binary variable F_{ek} indicate if the sum of frequencies of all lines operating on edge e equals k. The domain of the index k, representing the sum of frequencies can, in theory, become arbitrarily large since the number of lines and their frequencies are not limited. To reduce the number of variables F_{ek}, some limits are set to the decision variables. Firstly, the domain of the frequencies are limited to $f_l \in \{0, ..., f_{\max}\}$, where $f_{\max}$ is the upper limit of frequency of any line $l \in L$. Later in this thesis, a value of $f_{\max} = 60$ is used. Together with $\tau = 60$ (minutes), meaning that a single line can depart from a station every minute, the limitation of frequencies is reasonable. The maximum frequency of any line $f_{\max}$ can be used to derive an upper bound for k in F_{ek} as well. Since F_{ek} denotes the sum frequencies of lines operating on edge e, the upper bound $F^e_{\max}$ for k in F_{ek} is given by

$$F^e_{\max}=\sum_{l\in L:e\in l}f_{\max}. \tag{22}$$

With the binary variables F_{ek}, the objective function in (21) can be replaced by a linear alternative, as

$$\sum_{u\in V}\left(\sum_{v\in V}OD_{uv}\frac{1}{2}\tau\left(\sum_{k}^{F^{e_{uv}}_{\max}}\frac{1}{k}F_{e_{uv}k}\right)\right). \tag{23}$$

Apart from the constraints inherited from the base model (4), additional restrictions are necessary. This is because the decision variables f_l and F_{ek} need to be constrained and connected to the already existing decision variables x_l. Namely, the connection between F_{ek} and f_l is expressed as an equality constraint

$$\sum_{k}^{F^e_{\max}}kF_{ek}=\sum_{l\in L:e\in l}f_l \qquad \text{for all } e\in E. \tag{24}$$

In addition, it must be explicitly constrained, that only one k in F_{ek} is active at once. This is managed with the constraint

$$\sum_{k}^{F^e_{\max}}F_{ek}=1 \qquad \text{for all } e\in E. \tag{25}$$

Finally, the frequencies f_l and the number of trains x_l are linked. The connection between these variables is exact, as given in Equation 3. Nonetheless, in the waiting time model, the frequencies f_l are decision variables, and given the dependency with F_{ek} in Constraint 24, the frequencies are defined as integer variables $f_l \in \{0, \ldots, f_{\max}\}$. To maintain the dependency of Equation 3, it is formulated as an inequality constraint

$$f_l \leq \frac{\tau}{T_l} x_l \qquad \text{for all } l \in L, \tag{26}$$

where f_l is allowed to be smaller than otherwise determined by x_l. This inequality has one impact on the model such that it slightly overestimates the waiting time in the objective function. However, the inequality constraint doesn't affect the capacity constraint since it is expressed in terms of x_l. Particularly, it is guaranteed that all passengers can be transported with a feasible solution of the model. With these additional constraints, the passengers' perspective initial waiting time model is given by

$$\begin{aligned}
\min_{x,F,f} \quad & \sum_{u \in V} \left(\sum_{v \in V} OD_{uv} \frac{1}{2} \tau \left(\sum_{k}^{F_{\max}^{e_{uv}}} \frac{1}{k} F_{e_{uv}k} \right) \right) && && \text{(27a)} \\
\text{s.t.} \quad & \sum_{l} x_l \leq \mathcal{F} && && \text{(27b)} \\
& \sum_{l \in L: e \in l} C \frac{\tau}{T_l} x_l \geq w_e && \text{for all } e \in E && \text{(27c)} \\
& \sum_{k}^{F_{\max}^{e}} k F_{ek} = \sum_{l \in L: e \in l} f_l && \text{for all } e \in E && \text{(27d)} \\
& \sum_{k}^{F_{\max}^{e}} F_{ek} = 1 && \text{for all } e \in E && \text{(27e)} \\
& f_l \leq \frac{\tau}{T_l} x_l && \text{for all } l \in L && \text{(27f)} \\
& x_l \in \mathbb{N}_0 && \text{for all } l \in L && \text{(27g)} \\
& f_l \in \{0, \ldots, f_{\max}\} && \text{for all } l \in L && \text{(27h)} \\
& F_{ek} \in \{0, 1\} && \text{for all } e \in E, k \in \{1, \ldots, F_{\max}^{e}\}. && \text{(27i)}
\end{aligned}$$

Consistent with the previous models, the operator's perspective of the initial waiting time model is obtained by minimizing the fleet size and constraining the expected initial waiting time with a bound $W_{\max}$ (in minutes). Note that the expected waiting time in the objective function gives the total waiting time for all passengers. Thus, the value should be divided by the number of all passengers, given by the sum of elements in OD, to get the average expected initial waiting time for a single passenger.

Table 1: Number of variables and constraints with a linear PTN comprising 20 stations and 8 terminals. The line pool consists of all possible lines between 8 terminals ($|L| = 28$). In the initial waiting time model $f_{\max} = 60$.

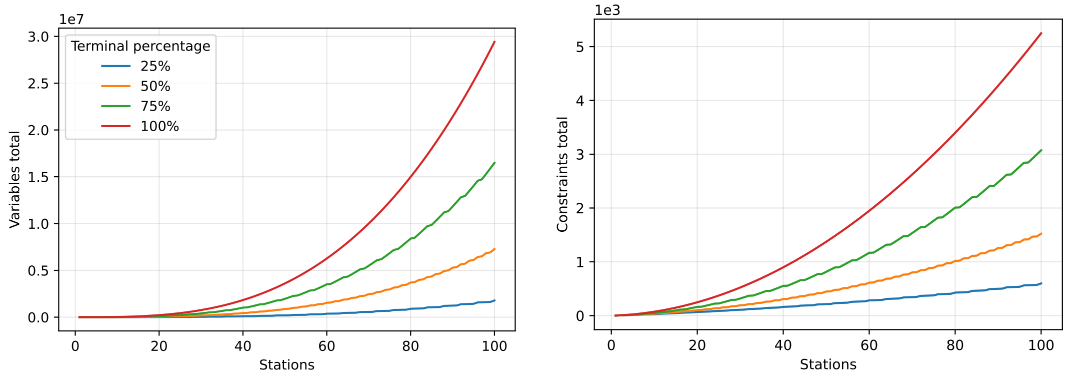
Model	Terminals	Variables	Constraints
Congestion	All 8	29	20
Congestion	Constrained to 4	37	79
Direct passengers	All 8	10668	838
Direct passengers	Constrained to 4	10676	897
Initial waiting time	All 8	13196	86
Initial waiting time	Constrained to 4	13204	145

3.4 Model size

The formulation of the decision variables F_{ek} in the previous section suggests that the initial waiting time model is prone to having a large number of variables. Hence, the models' size, in terms of variables and constraints, is shortly discussed next. These numbers are collected in Table 1, where the sizes of all three models are compared. The number of variables and constraints are calculated for a situation, where line planning is conducted for a linear PTN, which consists of 20 stations, out of which 8 are terminal stations. The line pool consists of all possible lines that can be established between the 8 terminals. Particularly, there exists $\frac{1}{2}|V^T|(|V^T| - 1) = 28$ lines to choose from. In addition to evaluating the situation with all terminals available, the terminal constraints, introduced in Section 3.2.1, are exploited to constrain the number of terminals to 4. Moreover, in the initial waiting time model, the parameter $f_{\max}$ controls the dimensionality of F_{ek} and it thus affects the model's size. Therefore, the value for the maximum line frequency $f_{\max}$ is set to 60, and it remains fixed throughout the remainder of this thesis.

From Table 1 it is apparent that the congestion model is the smallest model out of the three, by a few orders of magnitude. The direct passengers model has the most constraints, mainly due to the Constraints 17e, which cover all edges and the lines operating on them separately. On the other hand, the initial waiting time model is the largest in terms of variables. The large number of variables is a result of the auxiliary variables F_{vk}, whose size increases as a function of the number of edges and lines. Furthermore, the addition of the terminal constraints increases the number of variables and constraints by a constant amount, independent of the model. That is, the terminal constraints yield 8 new decision variables, one for each terminal $t \in V^T$. Also, 59 new constraints are added in this case.

Next, the size of the initial waiting time model is studied slightly more in detail. In particular, the model's response to increasing input PTN size is considered. The upper bounds for the number of variables and constraints are calculated such that it is again assumed that all possible lines between terminals are included in the line pool. In addition, it is assumed that all lines operate on all edges to obtain robust upper bounds for the numbers. Using these assumptions, the number of lines is given



(a) Number of variables

(b) Number of constraints

Figure 3: Number of variables and constraints in the initial waiting time model with a variable size linear PTN. The line pool consists of all possible lines. The terminal percentage indicates how many stations act as terminals. Terminal constraints are not applied. $f_{\max} = 60$.

by $|L| = \frac{1}{2}|V^T|(|V^T| - 1)$, the number of variables x and f is $|L|$. With the robust approach, the upper bound for F variables is given by $|E||L|f_{\max}$. In the linear PTN, the number of edges is given by $|E| = |V| - 1$. On the constraints side, the number of constraints for (27c - 27e) is equal to $|E|$. The number of constraints for (27f) is equal to $|L|$. Figure 3 presents these results as a function of stations, up to 100 stations. Also, the estimates are given for four different terminal percentages, indicating which proportion of all stations are also terminals. From the figure, it is obvious that the size of the initial waiting time model increases quadratically as a function of stations, both in terms of variables and constraints. However, the scale for the number of variables is vastly larger, compared to that of the constraints. The quadratic growth on the variables side is caused by the F variables since their number increases in both $|E|$ and $|L|$. On the constraints side, on the other hand, the growth is mostly due to the Constraints 27f, as the set of lines grows quadratically as a function of stations. The curves, where the subset of stations are regarded as terminals, exhibit fluctuations. These small inconsistencies arise because the number of terminals is derived from the number of stations and converted into integers using the floor function, allowing multiple consecutive estimates to yield the same numerical value. Despite the quadratic growth of the model size, as depicted in Figure 3, it can be also observed that with a reasonable input size of 20-40 stations and terminal percentages less than 75%, the number of variables remains below million. Furthermore, the size of the initial waiting time model can be further reduced by controlling the $f_{\max}$ variable, if necessary. The models' size assessment is continued later in Section 4.3, where the solving times are evaluated as a function of the PTN size.

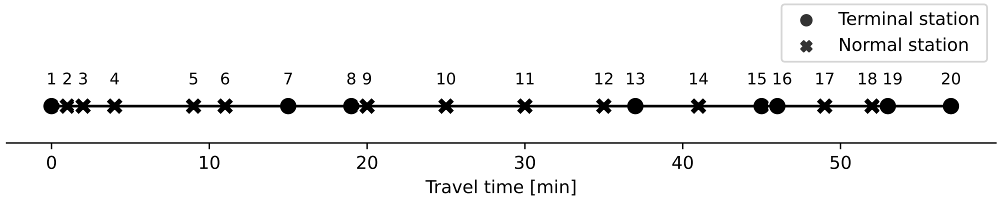


Figure 4: 20-station PTN with default terminals. The indices of the stations are labeled above the markers.

4 Experimental evaluation

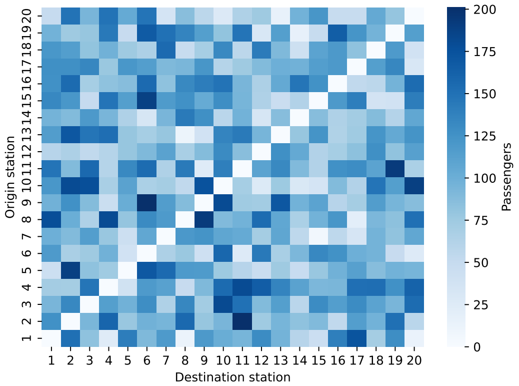
In this section, the line planning models developed in this thesis are examined in various aspects using generated data. This section begins with a description of the process for generating demand and infrastructure data in Section 4.1, which is followed by a discussion of obtaining Pareto solutions in Section 4.2. Thereafter, the computational requirements of the models are assessed through the exploration of solving times in Section 4.3. Finally, the three models are evaluated with the generated datasets. The main focus areas covered in the analysis include aspects such as the significance of terminals for passengers' convenience (Section 4.4), how the models respond to changes in the fleet size (Section 4.5), and the alignment of the different objectives between the models (Section 4.6).

To conduct the analyses in the forthcoming sections, the LinTim framework is utilized [Schiewe et al., 2023]. LinTim is an open-source library for algorithms and datasets tailored to the needs of public transportation optimization. The models, along with the generated datasets, are implemented within the LinTim framework and solved using the Gurobi Optimizer [Gurobi Optimization, LLC, 2023]. In the next section, the datasets and their generation are described in detail.

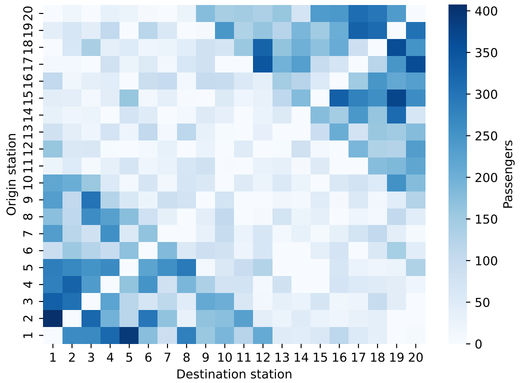
4.1 Data generation

In this thesis, both the PTN and the passenger demand data OD are generated. Algorithms are developed such that it is possible to generate arbitrarily sized linear PTNs and demand data to conduct experiments with variable-sized settings. Generating a linear PTN is straightforwardly obtained by creating a linear graph with a desired amount of stations. The algorithm also allows for specifying the number of terminals. Given this number, the algorithm randomly decides which stations act as terminals using a uniform distribution. Both ends of the PTN are fixed as terminals and are included in the total number of terminals, given in the input. The line pool is generated based on the terminals in the sense that all possible lines between them are included. Finally, each edge is given a duration, uniformly from $\{1, \ldots, 5\}$ (minutes), indicating the time it takes for a train to traverse that particular edge. For each line, the durations of the edges are added to obtain the round trip times T_l.

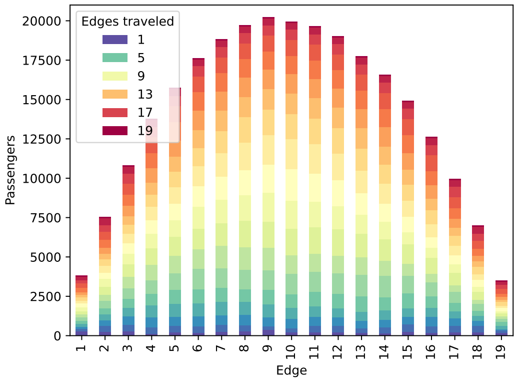
While the algorithm is employed for generating PTNs ad hoc, the experiments also involve studying a fixed PTN. This PTN, also generated with the algorithm,



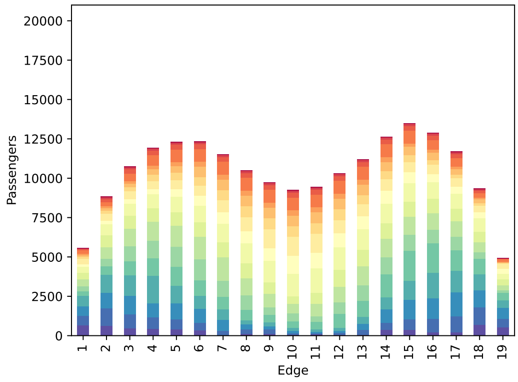
(a) OD-matrix of the unicentric demand profile



(b) OD-matrix of the bicentric demand profile



(c) Edge loads of the unicentric demand profile



(d) Edge loads of the bicentric demand profile

Figure 5: Visualization of the fixed unicentric and bicentric demand profiles for a PTN with 20 stations. Both demand profiles comprise 37833 passengers.

contains 20 stations and 8 terminals, indexed $\{1, 7, 8, 13, 15, 16, 19, 20\}$ concerning the ascending indices of the 20 stations. Moving forward, the fixed PTN is referred to as the 20-station PTN. The 20-station PTN is illustrated in Figure 4, where the lengths of the edges represent their travel times in minutes. In addition, a modification of the 20-station PTN is considered, where the terminals are spaced evenly along the linear network. That is, the set of terminals are indexed $\{1, 4, 7, 10, 13, 16, 19, 20\}$ to allow for comparisons with the unevenly spaced counterpart. However, unless otherwise specified, the 20-station PTN with unevenly spaced terminals is utilized. Note also that both variants of the 20-station PTN share the same passenger demand profile (OD), but the line pools are distinct, given their direct dependency on the terminals.

Similar to PTNs, an algorithm is developed for generating passenger demand, namely, origin-destination-matrices. Two variations are considered: unicentric and bicentric demand profiles. The names of these demand profiles describe the shape of the cumulative edge loads when visualized in the corresponding order of the PTN. Examples of both profiles are given in Figures 5c and 5d. The unicentric demand profile represents a situation where the PTN is centered around a single point of interest such as a city, whereas the bicentric demand profile has two demand peaks towards the opposite ends of the PTN. An OD-matrix producing a unicentric demand profile can be generated by drawing a uniformly distributed random variable to each origin-destination pair in the OD-matrix. For this purpose, distribution $\mathcal{N}(100, 40)$ is utilized, where negative values are replaced by zero. To acquire an OD-matrix with bicentric demand, the demand needs to be greater on both ends of the PTN and less in the middle. The following function is used to produce demand values that are larger within the extremities of the PTN:

$$OD_{uv} \sim \begin{Bmatrix} \mathcal{N}(300, 50), & u + v > \frac{7}{4}|V| \ \vee \ u + v < \frac{1}{4}|V| \\ \mathcal{N}(170, 70), & \frac{7}{4}|V| > u + v > \frac{7}{10}|V| \ \vee \ \frac{1}{4}|V| < u + v < \frac{3}{10}|V| \\ \mathcal{N}(40, 40), & \text{else} \end{Bmatrix}, \tag{28}$$

where the highest passenger demand values are obtained when the sum of the indices of the origin and destination stations u and v is either close to $2|V|$ or close to 0. Once the OD-matrices are generated, the cumulative edge loads w_e are determined. As discussed in Section 3.1, the edge loads are explicitly derived from the OD-matrix in the linear PTN setting.

The demand-generating algorithm is also utilized to produce two fixed OD-matrices for the 20-station PTN, one with a unicentric demand profile and another with a bicentric. First, the unicentric OD-matrix is generated, yielding 37833 passengers in total. The OD-matrix is depicted in Figure 5a and the corresponding demand profile in Figure 5c, where the colors denote the length of the passengers' journeys in number of edges. Then, the bicentric OD-matrix is generated until the total number of passengers is close to that of the unicentric OD-matrix. Finally, some of the OD-values in the bicentric OD-matrix are slightly modified to obtain exactly 37833 passengers. The resulting bicentric demand data is presented in Figures 5b and 5d. Comparing the two demand profiles in Figure 5, it can be observed that, with the bicentric demand, passengers travel significantly shorter distances within the PTN. Moreover, the cumulative demand directed to the whole PTN is significantly less in the bicentric case, despite the same total number of passengers, since passengers travel more locally in the PTN and hence the demand is not accumulated in the middle. In conclusion, the two fixed demand profiles, along with the 20-station PTN, are both employed in this thesis to conduct the experiments with variable scenarios.

4.2 Pareto solutions

In the pursuit of achieving a balance between the passengers' and operator's interests, this section delves into the visualization and application of the ε-constraint method. As discussed in Section 3.1.1, the ε-constraint method is employed to handle the two conflicting objectives. Therefore, the Pareto frontier obtained with the initial waiting time model is discussed next.

To outline the Pareto frontier depicting trade-offs between the initial waiting time and the fleet size, the initial waiting time model is solved with varying ε-constraints

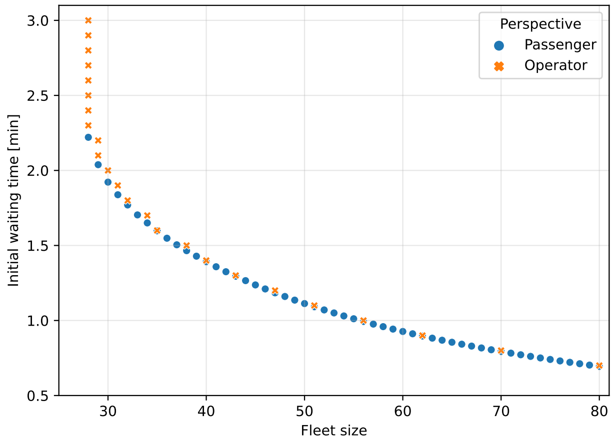
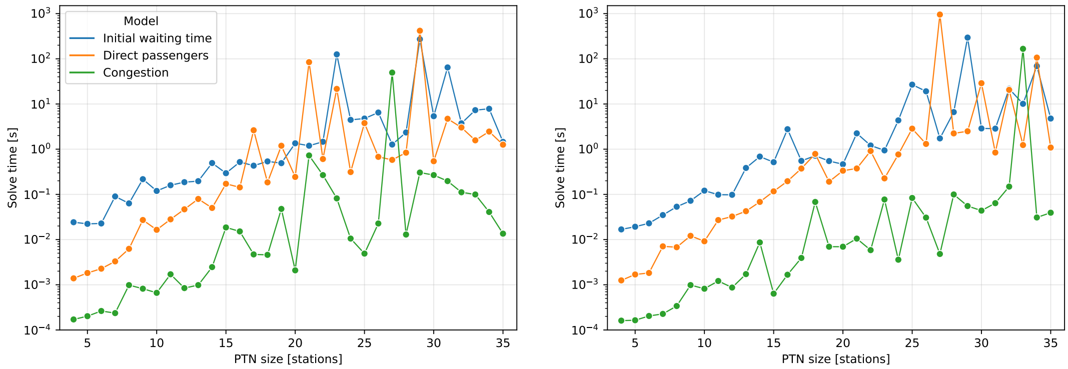


Figure 6: Outline of the Pareto frontier given by the solutions of initial waiting time model in both perspectives, using the ε-constraint method with varying bounds. The unicentric demand profile on the 20-station PTN is used.

in both perspectives. In particular, the passengers' perspective model is optimized by alternately constraining the fleet size in the range $[1, 80]$, while the operator's perspective model is optimized such that the constraint for initial waiting time (for an individual) is varied in the range $[0.1, 3.0]$ with a 0.1 step size. This process is carried out on the 20-station PTN with the unicentric demand profile, and no terminal constraints are considered. The optimal solutions acquired with this method are collected in Figure 6, where (some of) the points form the Pareto frontier. The figure clearly illustrates the trade-offs between the objectives, where increasing the fleet size results in reduced waiting time and vice versa. Only feasible solutions are plotted in the figure, and thus it can be observed that 28 trains are required for the solution to be feasible. That is, with less than 28 trains, the capacity constraint (27c) is not satisfied and the solution becomes infeasible. Regarding Pareto optimality, the solutions obtained by the ε-method are guaranteed to be weakly efficient, as shown in [Ehrgott, 1999]. Moreover, since the passengers' perspective solutions are strictly decreasing as a function of fleet size, as evident in Figure 6, and all values in the range [1,80] are covered as the fleet size is constrained in integer values, the passengers' perspective solutions form the exact Pareto frontier in that range. Conversely, the figure suggests that the solutions in the operator's perspective are only weakly efficient since some points are clearly above the points given by the passengers' perspective. Hence, the objective can be further improved in the passengers' perspective, which contradicts Pareto optimality. This condition is the most apparent with fleet size 28, where the solutions from the operator's perspective form a vertical line above the efficient point given by the passengers' perspective model. In this experiment, the operator's perspective model gives only weakly efficient points since the initial waiting time is a continuous variable and the step size of 0.1 is rather large. Specifically, with the step size of 0.1, it is quite improbable to consider an ε-constraint for initial waiting time that would be the optimal solution to the passengers' perspective model with some integer fleet size constraint. To conclude, the passengers' perspective



(a) Unicentric demand

(b) Bicentric demand

Figure 7: Solving times of the models versus the PTN size. Operator's perspective model solved with ε-constraints $W_{\max} = 2.0$, $p = 0.8$ and $\lambda_{\min} = 1.2$, respectively.

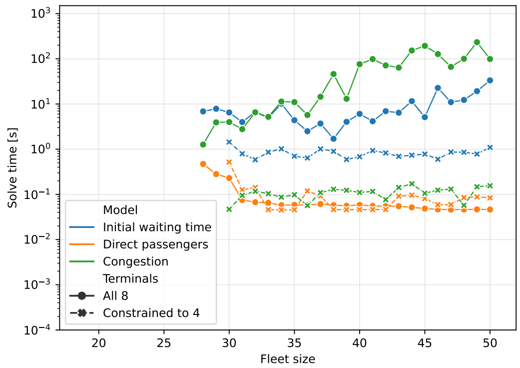
model is guaranteed to yield Pareto optimal solutions, but the solutions acquired from the operator's perspective might not be Pareto optimal. However, if the passengers' convenience ε-constraint is strict enough, the solutions between both perspectives are very close, as visible in Figure 6.

4.3 Solving times

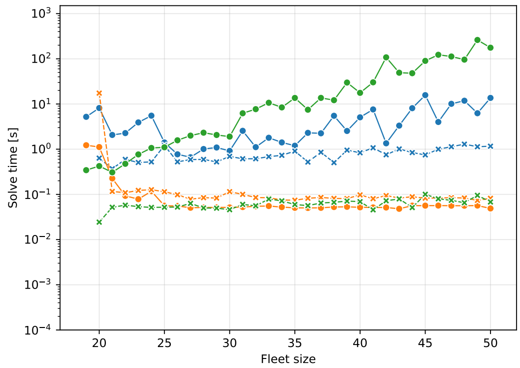
In what follows, the models' computational requirements are assessed through the analysis of solving times. The effects of various features on the models' running times are studied. These aspects include the two perspectives of the models, the demand profile, the PTN size, the fleet size, and the incorporation of the terminal constraints.

First, it is investigated how the size of the linear PTN affects the solving times. To conduct this experimentation, PTNs ranging from 4 to 35 stations in size are generated along with unicentric and bicentric demand data. The PTNs are set to contain a number of terminals that is obtained as a function of the PTN size given by $|V^T| = \lfloor 2\sqrt{|V|} \rfloor$. With this function, for instance, a PTN with 20 stations yields 8 terminals. The three line planning models are then solved in the operator's perspective with the following ε-constraints. The minimum requirement for the initial waiting time is set to $W_{\max} = 2.0$ minutes, the ratio requirement for direct passengers is set to $p = 0.8$, and finally, the lower bound for the minimum edge availability factor is set to $\lambda_{\min} = 1.2$, which means that at most 83% of the capacity is allowed to be utilized on any edge.

The solving times measured in the first investigation are presented in Figure 7, where the results of both demand profiles are separated into individual figures. The vertical axis in the figures is set to a logarithm with base 10 since the running times of the 4-station and the 35-station PTNs differ by several orders of magnitude. The figure suggests that independent of the demand profile, the initial waiting time model is computationally the most demanding problem to solve. However, the direct passengers model shares similar running times for larger PTNs. The congestion model, on the



(a) Unicentric demand



(b) Bicentric demand

Figure 8: Solving times of the models versus the fleet size constraint $\mathcal{F}$. Passengers' perspective model solved with and without the terminal constraints on the 20-station PTN.

other hand, is significantly less demanding to solve with approximately 10 to 100 times less time required, compared to the other models. Moreover, the figure indicates that the solving times increase exponentially when the PTNs are expanded. The growth is similar between the two demand profiles but the unicentric profile seems to require slightly more time to solve the problems, on average. Finally, the running times appear to be very sensitive to changes in the demand data, as especially apparent in Figure 7a. Examining the figure reveals that with PTNs within the range of 25 to 35 stations, the time required to solve the congestion model can vary by as much as four orders of magnitude. Also, the direct passengers and the waiting time models exhibit this behavior. In addition to the changes in the number stations and terminals, also the demand data changes between the PTNs. Therefore, the problems are very different from each other, which can explain much of the fluctuations in the solving times. Nevertheless, the overall increasing trend in still apparent in the figures.

In addition to the PTN size, the fleet size is varied to study the effects on the solving times. This time, the passengers' perspective models are solved by gradually altering the fleet size constraint $\mathcal{F}$ from 1 to 50 trains and optimizing for the corresponding passenger convenience function. This process is conducted on the 20-station PTN, where both demand profiles are regarded separately. Additionally, the results are further convoluted by factoring in the terminal constraints such that the models are solved first with all 8 terminals, and then the number of terminals is constrained to 4. The feasible solutions obtained using this procedure are displayed in Figure 8, where an identical vertical axis is utilized as previously. The figure immediately reveals that compared to the 28 trains required to operate a line concept for the unicentric demand, the problem with the bicentric demand profile can be feasibly solved by employing just 19 trains. This is because, with the bicentric demand, passengers travel shorter distances within the PTN and thus fewer edges are traversed in total. Hence, the accumulation of the demand in the middle of the PTN is reduced, leading to a decreased need for trains to meet the transportation demand. Similar to the

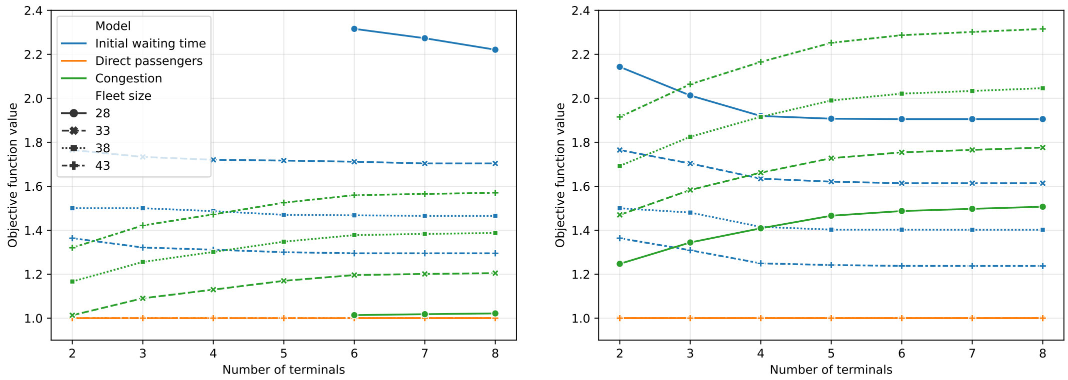
PTN size, the relationship between the fleet size and solving times is exponential for the initial waiting time and the congestion models. However, Figure 8 displays that computational requirements for the direct passengers model are independent of the fleet size. Moreover, this behavior is also present in the other two models when the terminal constraints are included. The terminal constraints seem not to have any effect on the direct passengers model, but for the initial waiting time and congestion models, the constraints reduce the computational effort significantly. The reason for the reduced running times might arise from the fact that the terminal constraints reduce the feasible region of the problem, and consequently, the search space is diminished causing reduced solving times. Furthermore, the figures reveal that solving for very low fleet sizes is also quite challenging. With the unicentric demand profile, the solving time curves have peaks around 28-30 trains, whereas with the bicentric demand, the peaks are around 19-21 trains. This finding suggests that when the number of available trains is barely enough to obtain a feasible line concept, the problem is more demanding compared to when a few more trains are added, after which the computational demand momentarily drops.

Comparing Figures 7 and 8, it becomes obvious that the perspective of the model has substantial importance on the solving times of the models. For instance, changing the perspective from the operator to the passengers leads to the congestion model becoming the computationally most demanding model, while in the operator's perspective, the congestion model is the least demanding one. The congestion model also has the highest variation in the running times between the constrained terminals and the utilization of all lines. One possible reason for the congestion model's sensitivity to changes in the perspective and the terminals might be due to the decision variable λ in the passengers' perspective model. Namely, the minimum edge availability factor λ is a variable that is directly connected to all constraints in the model, except for one. For this reason, even slight changes in the input data could cause significant changes in the computational demand. To conclude, the analyses conducted in this section indicate that the solving times are significantly influenced by the constraints associated with the inclusion of the terminals. Motivated by these findings, the terminals' effects on the convenience of the passengers, as well as the mutual importance of the terminals are investigated in the following section.

4.4 Terminal importance

With the congestion and direct passengers models, it was observed that constraining the total number of terminals reduces the computational demand substantially. This result is explained by the fact that removing terminals from the PTN leads to reducing the size of the line pool. This section delves into the investigation of how the selection of terminals affects the passengers' convenience and whether some terminal stations are more important for constructing an optimal line concept.

To study the effects on the passengers' convenience, the passengers' perspective models are solved for varying terminal constraints on the 20-station PTN with both demand profiles. The 8 terminals are constrained such that all number of terminals between 2 and 8 are considered. In addition, the results are calculated for fleet sizes



(a) Unicentric demand

(b) Bicentric demand

Figure 9: Objective function values of the three models versus the constraint for the number of terminals. Passengers' perspective models solved on the 20-station PTN with both demand profiles and four fleet size constraints. The objective function values of the direct passengers and the initial waiting time models are divided by the total number of passengers, resulting in the ratio of direct passengers and the average initial waiting time for a single passenger, respectively.

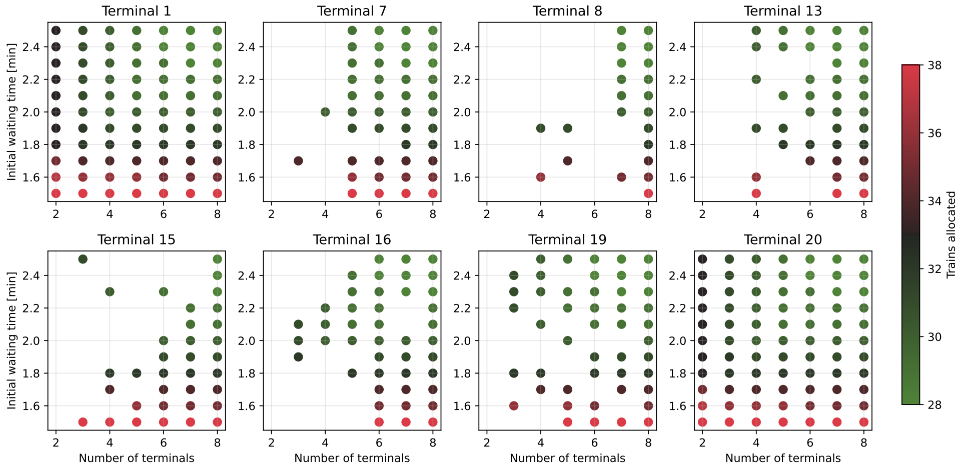
28, 33, 38, and 43. The results obtained in this experimentation are presented in Figure 9, where only the feasible solutions are plotted. Starting with feasibility, it can be observed that all solutions obtained in the process are feasible, with the exception of those generated with 2 to 5 terminals, unicentric demand, and a fleet size of 28. For the unicentric demand profile, the fleet size of 28 was recognized as the threshold for feasibility also in the earlier results, in Figures 6 and 8a. Overall, the effects of restricting the number of terminals are quite similar between the two demand profiles. That is, with the initial waiting time and direct passengers models, the objective worsens as the number of terminals is reduced. This deterioration is due to the reduced size of the line pool, as the lines ending at the excluded terminals cannot be considered and hence the flexibility of the line concept is decreased. With the bicentric demand, the objective values are better since the accumulated demand is distributed more evenly along the PTN, which results in a better utilization of the fixed fleet. Furthermore, comparing the shapes of the curves between different fleet sizes, it can be concluded that the shape of the dependency between the objective value and the number of terminals is unchanged.

It is especially evident in Figure 9b, that the significant decrease in the initial waiting time stops after allowing 4 or more terminal stations in the line concept. The phenomenon is similar across all fleet sizes, while with the unicentric demand profile, the effect is less obvious. The abrupt change in the rate of decrease of the initial waiting time, especially with the bicentric demand, could be explained by the change of the restricting constraint. That is, with just 2 or 3 terminals, the line pool is very limited in size and hence there are not many lines to choose from. Therefore, the trains cannot be allocated evenly along all edges and thus waiting times might increase on some stations. When more than three terminals can be established, the flexibility of the line pool is no longer the constraining factor, but the waiting time is more constrained by the fleet size, and hence no significant improvements can be made between 4 and 8 terminals. With the congestion model, a similar phenomenon is present, but the change is not as abrupt. Also, the minimum edge availability factor can be still improved when 6, 7, or 8 terminals are considered. This observation could be explained by the way the minimum edge availability factor λ is defined. While the initial waiting time model minimizes the initial waiting time, which is a weighted average of all passengers, the congestion model maximizes the minimal edge availability factor such that the change in just one edge can lead to big differences in the whole objective function. Thus, the congestion model benefits from having more lines to choose from, as the capacity can be more evenly distributed, leading to increased minimum edge availability.

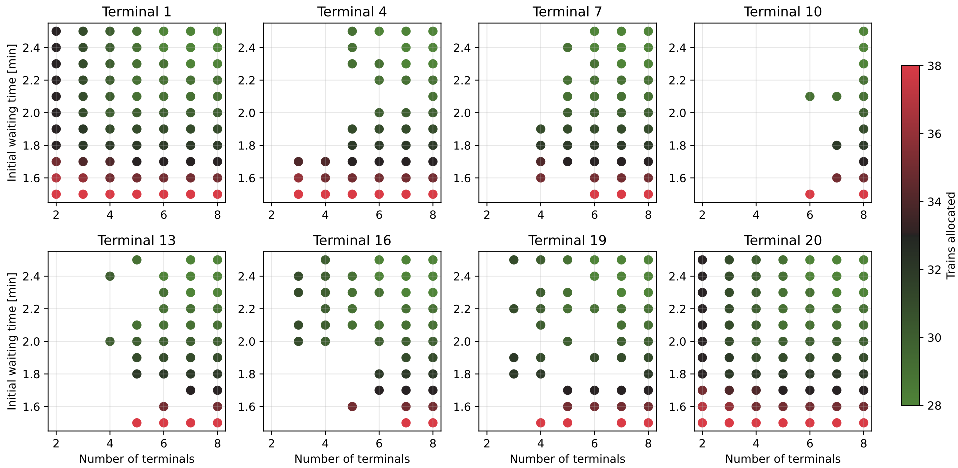
Finally, Figure 9 also reveals interesting properties of the direct passengers model. Particularly, the figure suggests that regardless of the number of terminals or the fleet size, it is always possible to transport all passengers directly. It is not visible in the figure, but for the fleet size of 28, the solutions are very slightly less than one such that in all three solutions, three passengers out of the 37833 need to transfer during their journey. However, this exception is negligible. Within this experimentation, it can be concluded that regardless of the constraints, all passengers can travel directly in the direct passengers model. This observation is plausible since the PTN is linear and relatively short such that lines operating through the whole PTN can be included in the line concept. The capacity constraints are then responsible for allocating enough capacity for all passengers, which is guaranteed to provide a direct connection to (almost) all passengers if instructed by the objective function.

Next, the focus is shifted from the objective function values to the terminal stations themselves. To study the mutual importance of the terminals, the operator's perspective initial waiting time model is solved with varying number of terminals as well as varying initial waiting time requirements. Namely, the number of terminals is varied in the range [2,8], and the initial waiting time in the range [1.5,2.5] minutes, with a step size of 0.1 minutes. In this experiment, the 20-station PTN is considered with three different variations. In addition to the unicentric and bicentric demand profiles, the unicentric demand is also solved with the 20-station PTN with evenly spaced terminals, which is described in Section 4.1. With this setting, the terminals' mutual importance can be analyzed by observing how often a given terminal is included in the optimal solution.

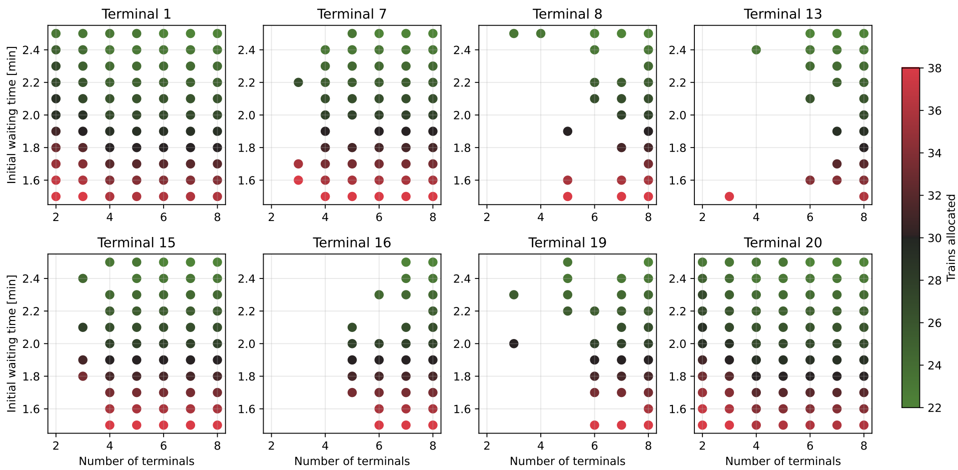
This process yields 77 solutions for each of the three input variations. These 77 runs are visualized on a terminal level in Figure 10, where a dot indicates that the terminal is included in the solution, characterized by the requirements in the horizontal and vertical axes. Furthermore, the colors indicate the objective function values, namely, the minimized fleet sizes. The figure indicates that both ends of the PTN, terminals 1 and 20, are included in all solutions, as required by the explicit terminal constraints (8). In general, the results acquired with the three settings conclude that the subsets of the terminals are subject to quite significant variation. Moreover, even slight changes in initial waiting time requirements can result in completely different selection terminals. The variability of the inclusion between the terminals is dependent on the terminal, however. For instance, Terminal 7 in Figure 10a and Terminal 15 in



(a) Unicentric demand

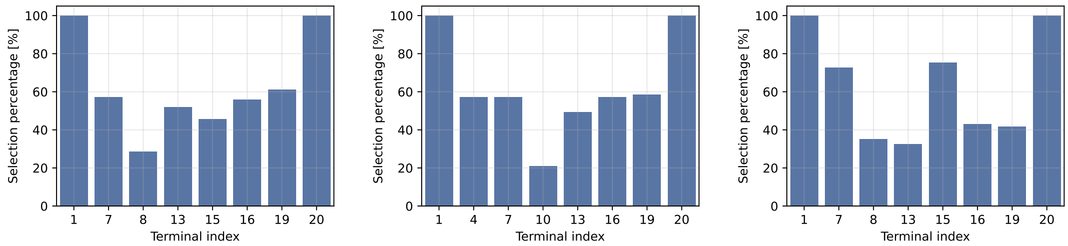


(b) Unicentric demand, evenly spaced terminals



(c) Bicentric demand

Figure 10: The inclusion of terminals in the optimal solutions. The initial waiting time model solved in the operator's perspective on the 20-station PTN with both demand profiles.



(a) Unicentric demand

(b) Unicentric demand, evenly spaced terminals

(c) Bicentric demand

Figure 11: Percentage of runs for which the terminals are included in the line concept. Results aggregated from Figure 10. The initial waiting time model solved in the operator's perspective on the 20-station PTN with both demand profiles.

Figure 10a are quite consistent in the inclusion, whereas Terminal 19 in Figure 10a and Terminal 8 in Figure 10a seem to have no pattern for when they are included in the solution. Also, the figure reveals, that no significant patterns related to the initial waiting time requirement can be found. That is, there are no terminals that are more likely to be included in the solution if the initial waiting time requirement is higher, or vice versa. However, it is obvious that some terminals are more important than others in the sense that they are included more frequently. This aspect is discussed in more detail later in this section.

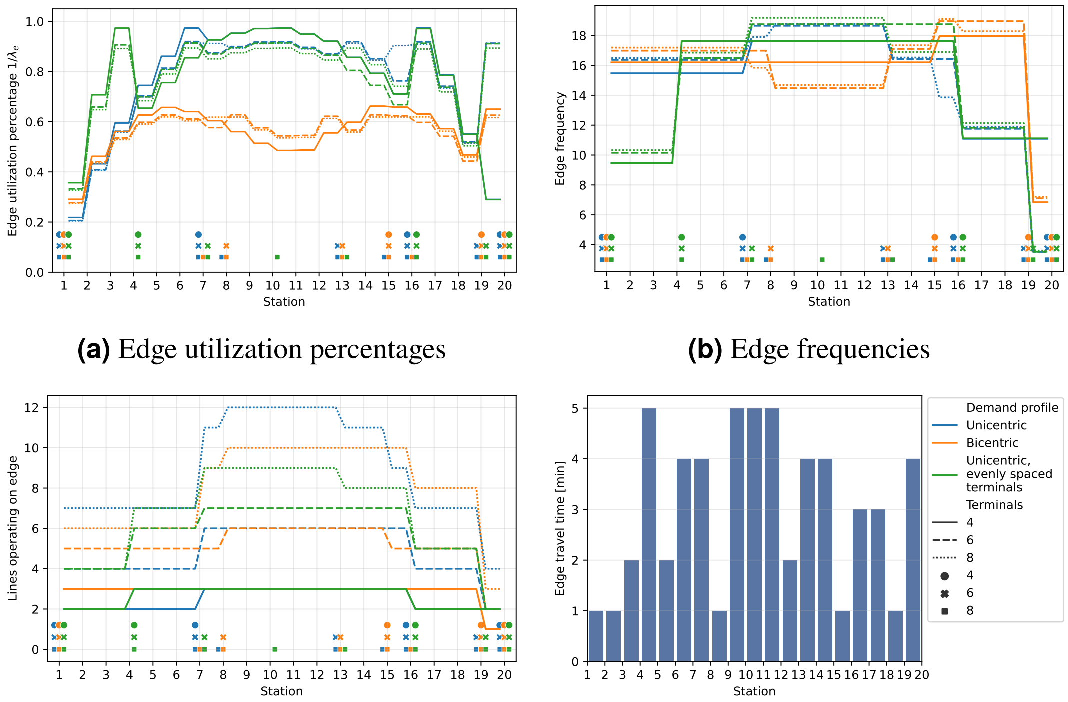
Figure 10 also implies that 38 trains are required to satisfy the lowest initial waiting time requirement, regardless of the demand profile or the terminal spacing. However, with the highest initial waiting time requirement, the bicentric demand setting can be operated with just 22 trains while the settings with unicentric demand require 28 trains. This finding is again explained by the capacity constraints. With the relatively loose initial waiting time requirement of 2.5 minutes, fewer trains are required to fulfill this requirement, and the optimal solution reflects the capacity requirement for the trains to operate the given line concept. As discussed earlier, the requirement for the fleet size is 28 for the unicentric demand and 19 for the bicentric demand (see Figure 8). Thus, the bicentric demand can fulfill the requirement with less than 28 trains. In the more demanding initial waiting time requirement of 1.5 minutes, the capacity constraint is no longer the active constraint, but all settings require 38 trains to bring the initial waiting time this low, which results in the initial waiting time requirement becoming the active constraint.

To visualize the terminal importances from another perspective, the results presented in Figure 10 are aggregated in Figure 11, which depicts the percentage of runs for which the terminals are included in the line concept. The distribution of terminal inclusion for the unicentric demand in Figure 11a resembles a U-shape such that the terminals closer to the ends of the PTN are considered more important as they are included in more solutions. Terminal 8 is located the closest to the center of the PTN and it has the lowest selection percentage compared to the other terminals. This result is expected, as there are approximately equal number of passengers departing at each station, and hence the accumulated edge loads become the defining factor for selecting lines, and ultimately the terminals. The lines halting at the center of the PTN are then not desirable since they cannot account for a significant proportion of the demand willing to travel past the PTN center. With the bicentric demand in Figure 11c, after the PTN ends, terminals 7 and 15 are the most important. These two terminals are roughly aligned with the demand peaks illustrated in Figure 5d and hence they are important as they allow selecting lines that operate locally around the two demand peaks. Finally, the distribution of the terminal selection for the evenly spaced terminals, presented in Figure 11b, portrays an almost symmetrical shape, which is anticipated given the symmetrical demand and terminal spacing. Slight variation is due to Terminal 19 which doesn't have a counterpart on the opposite side of the PTN. The evenly spaced terminal setting also further highlights the conclusions made earlier about Terminal 8. Particularly, in this setting, Terminal 10 is at the center of the PTN and it has an even lower inclusion percentage than Terminal 8, which was considered rather unimportant earlier. Observing Figure 10b concludes that Terminal 10 is selected almost only in the case when all 8 terminals can be included in the line concept. The minor significance of Terminal 10 is explained by its central location as with Terminal 8 earlier.

The investigation of the terminals' importance is concluded with a third experimentation, where the congestion model is analyzed in greater detail. From Figure 9 it was observed that the congestion model benefits from having a higher number of terminals, as compared to the initial waiting time model. Therefore, the solutions from the congestion model, acquired with various terminal configurations, are studied on the edge and station level. To conduct this analysis, the passengers' perspective congestion model is solved for the 20-station PTN using again both demand profiles, and in addition, the evenly spaced terminals PTN with the unicentric demand profile. These configurations are solved by constraining the number of terminals to 4, 6, and 8. A fleet size of 30 trains is utilized throughout this process.

Figure 12 summarizes the results obtained in the congestion model investigation. All four subfigures display the results on the edge level, while the stations are marked on the figures. Consequently, the edges are inherently represented by the spaces between the stations. With this visualization, the edges will have the same indices as the stations to the left of them. Together, the four subfigures entail a vast amount of information on the different solutions. In Figure 12a, instead of the edge availability factors λ_e, their inverses are plotted. That is, the so-called edge utilization percentages $1/\lambda_e$ are presented for their more intuitive interpretation. The values of $1/\lambda_e$ are defined in [0,1] and higher values imply higher congestion which leads to easier readability of the figure. In the subsequent figures, also the sum of frequencies and number of unique lines are plotted on each edge, respectively. Finally, in Figure 12d, the travel times of the edges are provided to give context. Note that the travel times are identical among all configurations since neither the demand profile nor the terminal spacing affects the travel times as the PTN is fixed in this experiment.

Concentrating first on Figure 12a, it reveals that the shapes of the congestion, or the utilization percentage, over all edges, coincide quite precisely with the corresponding shapes of the demand profiles, as depicted in Figures 5c and 5d. The exceptions from the shapes of the demand profiles, especially with the evenly spaced terminals, are discussed later in this section. Otherwise, the correspondence of the shapes between



(a) Edge utilization percentages

(b) Edge frequencies

(c) Number of lines on each edge

(d) Edge travel times

Figure 12: Terminals' effects on edge utilization, edge frequencies, and number of lines on edges. The congestion model solved in the passengers' perspective on the 20-station PTN with both demand profiles. Fleet size is constrained to 30 trains. The included terminals are visualized as markers on the bottom of the figures, on the corresponding stations.

the edge utilization and the demand emerges from the definition of the edge availability factor, given in Equation 10. Namely, because the inverses of the edge availability factors are considered, the edge utilization percentage is directly proportional to the accumulated demand. To compensate for the increased demand in certain parts of the PTN, more trains are allocated there, resulting in increased edge frequency (Figure 12b), and ultimately, in more unique lines being employed (Figure 12c).

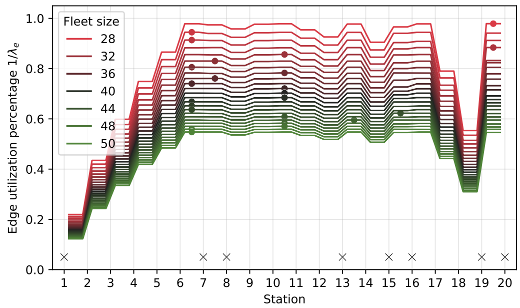
When the number of terminals is reduced, the variation in the shapes of the edge utilization percentages is increased. Contrarily, with a larger set of terminals, the profiles of the edge utilization percentages are more planar, indicating better distribution of capacity utilization across the PTN. Because more terminals are available, also more diverse set of lines can be established, and hence the line and frequency composition better corresponds to the passenger demand. Ultimately, the better distribution of resources leads to lower maximum edge utilization, which is being minimized. These findings are supported in Figures 12b and 12c in the sense that more precise changes in the shapes are introduced when more terminals are available, compared to the rather flat and elementary shapes associated with 4 terminals.

Between stations 16 and 17, the edge utilization percentage momentarily increases to the same level as in the middle of the PTN for both variations of the unicentric demand profile. Moreover, on the opposite end of the PTN, the solutions of the evenly spaced terminals variation exhibit the same behavior also between stations 3 and 4. For the evenly spaced terminals, the peaks occur on both sides of the PTN, on the edges that are adjacent to the first terminals after those terminals that serve as the ending stations of the PTN. The next edges towards the center of the PTN have again substantially lower congestion rates. This finding is explained by terminals 4 and 16, which allow increasing the frequency for the innermost edges within the PTN. From Figure 12b it can be observed that the edge frequencies are doubled between terminals 4 and 16, which leads to the abrupt decrease in the utilization for edges that follow these terminals, moving towards the PTN center. The inclusion of terminals 4 and 16 is important for the evenly spaced terminals setting since it allows for better distribution of the trains. In fact, terminals 4 and 16 are always included in the solution, even if only four terminals are available. The unicentric demand with the original terminals spacing shares the phenomenon on the right side of the PTN, on Edge 16 since Terminal 16 is also included in it. However, due to the lack of Terminal 4, the congestion peak is not present on the left side of the PTN. The absence of Terminal 4 results in relatively high edge frequencies on edges 1 to 6 and low congestion as service cannot be distributed according to the demand. Terminal 7 is the first terminal that can stabilize the situation, and it is therefore included in all solutions of the unicentric demand with the original terminals. Finally, similar conclusions to those made about the peaks on edges 3 and 16, also apply to Edge 19, which exhibits high variation in the utilization levels depending on the terminal configuration. This variation can be analogously explained by the inclusion of the adjacent terminal, with index 19 in this case.

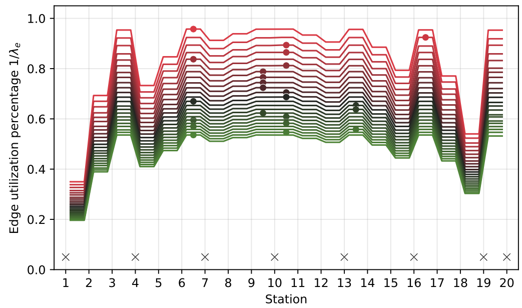
Comparing the edge travel times, presented in Figure 12d, to the results acquired in this experimentation, it can be observed that the edge-wise measures are independent of the travel times. This is expected since the travel times are only regarded in the round trip times of the lines T_l as aggregate values. Increased edge travel times are then reflected in decreased frequencies per train, which leads to more required capacity overall. However, these requirements apply to the line level and hence the edge-wise metrics are not affected, as evident in Figure 12. The investigation of the congestion model concludes this section, where the terminals have been the main area of interest. Moving forward, the focus is shifted towards the fleet size, which is already discussed shortly together with the solving times and the terminal stations, but in the next section, the fleet size is the primary topic under examination.

4.5 Fleet size results

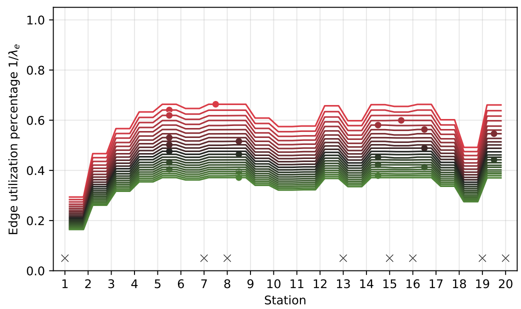
In Section 4.2 it was observed, that the trade-off between the operator's and passengers' interests changes rapidly between fleet sizes from the feasible 28 trains to around 40 trains. With the initial waiting time model, this finding implies that increasing the fleet size just slightly leads to substantially reduced waiting times. In this section, the trade-offs between the perspectives are studied in the context of the congestion model. Moreover, the previous section revealed that the shape of the edge utilization along the PTN is heavily influenced by the selection of terminals. Therefore, the



(a) Unicentric demand



(b) Unicentric demand, evenly spaced terminals



(c) Bicentric demand

Figure 13: Edge utilization percentages with various fleet size constraints. The congestion model solved in the passengers' perspective on the 20-station PTN with both demand profiles. The round markers denote the maximum edge utilization percentage for each solution. Terminal stations are denoted with crosses.

edge-wise utilization percentages are studied also in this section. Finally, the solutions are compared by examining the results on the line level.

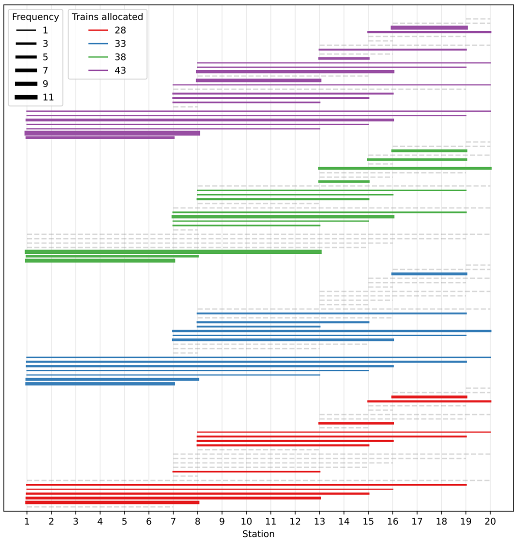
In this section, the following experimental setup is considered. The congestion model is solved in the passengers' perspective such that the fleet size is constrained in the range [28,50], consecutively. The process is repeated for the three input configurations familiar from the previous section. In this investigation, the terminal constraints are not included and thus all eight terminals, along with the lines associated with them, are available. Firstly, the results are visualized in Figure 13 as the edge-wise utilization percentages $1/\lambda_e$, similar to the previous section.

Figure 13 immediately exposes an interesting property of the relationship between the edge-wise congestion profiles and the fleet sizes. Particularly, regardless of the fleet size, the shape of the utilization percentage is unchanged. The shape is only scaled according to the fleet size in the sense that larger fleets reduce the congestion proportionally along the whole PTN. This phenomenon is shared by all three configurations. With the unicentric and bicentric demand profiles in Figures 13a and 13c, however, very minor deflections from the general shape are visible around edges 15 and 19, which are both edges located between two terminals. Nevertheless, the figure suggests that there exists an optimal shape for the edge utilization that provides the minimized maximum edge utilization percentage, regardless of the fleet size. To understand this phenomenon better, the solutions are visualized on the line level later in this section. Furthermore, the persistence of the utilization shape implies that the congestion peaks, discussed in the previous section, remain present despite the variation in the fleet size. This observation is most evident in Figure 13b, on edges 3 and 16.

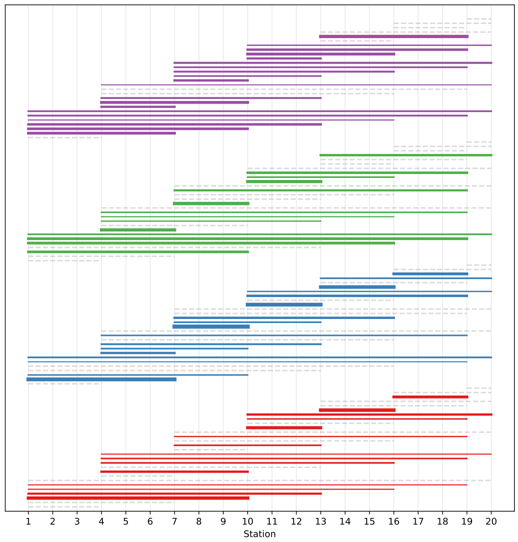
Figure 13 also illustrates the vast differences between the demand profiles. As discussed earlier, the bicentric demand profile can be operated with a significantly smaller fleet compared to the 28 trains required by the unicentric demand. With 28 trains, the unicentric demand yields utilization percentages very close to 100% on many edges, whereas the corresponding percentages are less than 70% for the bicentric demand profile, which is due to the more evenly distributed passengers. However, starting with 28 trains, the congestion rates for the unicentric demand begin to decrease at a quicker rate when the fleet size is increased, compared to the bicentric demand. Increasing the fleet size further slows down this pace, nevertheless. The decreasing rate of improvement in congestion is due to a decreased proportion of the added capacity. When the number of trains is increased, the individual contribution of each successive train to the overall capacity decreases. Thus each added train brings a reduced amount of improvement to the solution. This trend is also apparent in Figure 6, where the rate of improvement in the initial waiting time is reduced as the fleet size is increased.

For every solution acquired in this experiment, Figure 13 denotes the respective maximum edge utilization percentages with round markers. These markers then indicate which edges are limiting the improvement in the objective function. Interestingly, for each setting, there exist multiple edges that alternately exhibit the highest, limiting edge utilization. Furthermore, these edges are not uniformly scattered along the PTN, but there are three to five edges per setting that can serve as the binding edge. In general, these edges with maximum congestion seem to be either those that have the highest accumulated demand, as given in Figures 5c and 5d, or those that are adjacent to important terminals, as discussed in the previous section. For instance, with the unicentric demand in Figure 13a, Edge 6 is located just before the first important terminal, Terminal 7, which is critical for stabilizing the congestion in the middle of the PTN. Before this terminal, however, the congestion grows locally so high, that Edge 6 becomes the limiting edge for the solution. Also, Edge 10 is limiting the maximum congestion since it is located in the middle of the PTN, where the accumulated edge demand is the highest for the unicentric demand profile. Similar conclusions can be made about Edge 5 in Figure 13c and Edge 6 in Figure 13b.

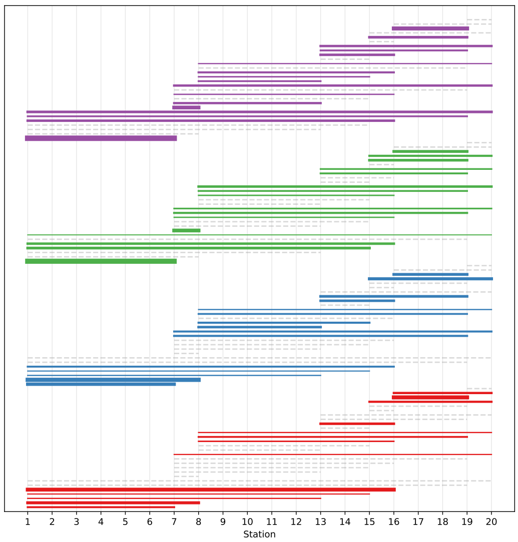
To obtain more information on the underlying line compositions of the solutions discussed in this section, the line concepts for fleet sizes 28, 33, 38, and 43 are visualized in Figure 14. Observing the line concepts, the figure reveals that the line compositions for different fleet sizes are quite diverse even though the edge utilization profiles are almost fixed. Particularly, there exist lines that are present with the smallest fleet but are not established for fleet sizes of the 30s and are then again introduced with the 43-train fleet. For instance, with the bicentric demand, the line from Station 13 to Station 16 is utilized with all other fleet sizes but 38 trains. Moreover, the lines that remain in the line concept regardless of the fleet size might still have high fluctuations



(a) Unicentric demand



(b) Unicentric demand, evenly spaced terminals



(c) Bicentric demand

Figure 14: Line concepts $(\mathcal{L}, f)$ with various fleet sizes. The congestion model solved in the passengers' perspective on the 20-station PTN with both demand profiles.

in their frequencies with different fleets. An example of such a line is given by the line from Station 1 to Station 8 in Figure 14a. In general, all line concepts utilize a rather large amount of unique lines, which is a fact that becomes even more pronounced as fleet size increases. As more trains are available, also shorter lines are introduced. With the bicentric demand, even a line containing just one edge is included in the solution. This line, covering Edge 7, is located close to the first demand peak of the bicentric demand profile such that it allows for reducing the accumulated demand impact on the overlapping lines.

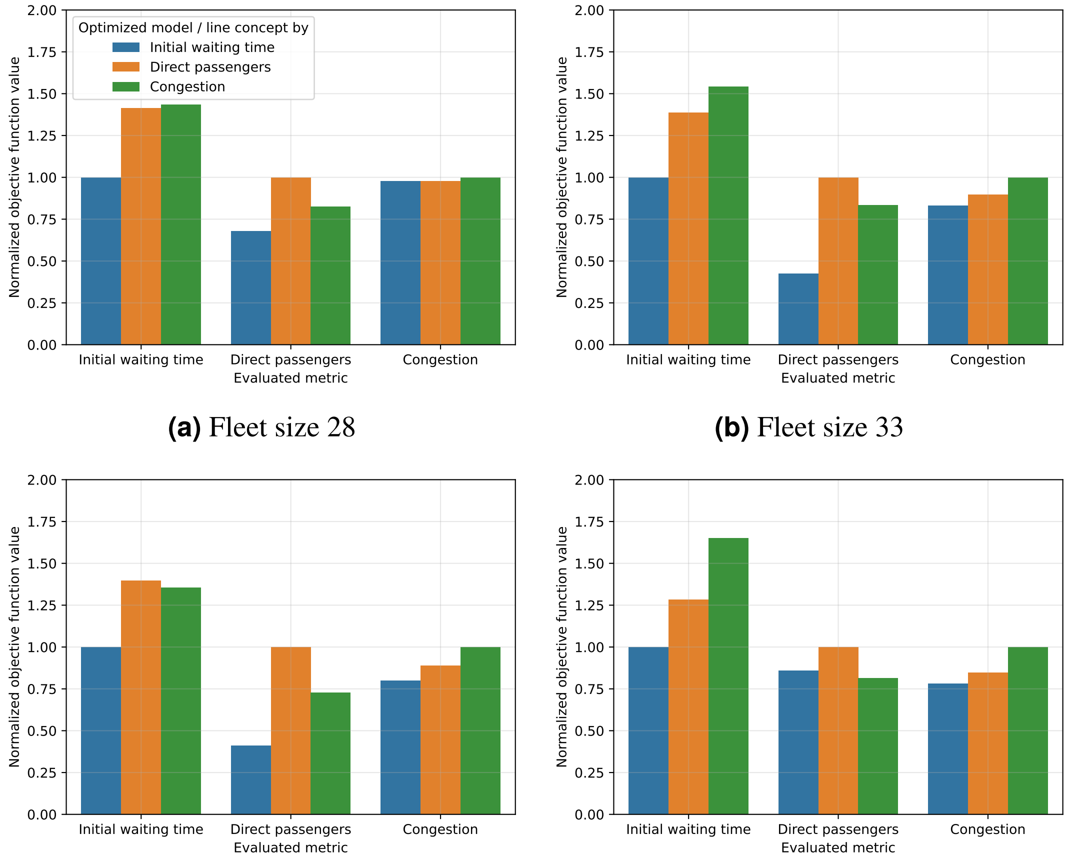
One explanation for the great variety in the line concepts is given by the high number of lines to choose from in the line pool L. In this thesis, the line pool consists of all possible lines between the terminal stations. Because the line pool is so diverse, the line planning models can select a complex subset of lines to improve passenger convenience, dictated by the objective function. This diversity leads to a line concept with a generous amount of unique lines. However, too big a variety in the line concept is not desired by the passengers as the line concept becomes too complex to comprehend and memorize. Line concepts of more than 10 unique lines, as evident in Figure 14, are certainly too complex in this context. Nevertheless, using a sophisticated combination of lines, the congestion model can maintain the optimal shape of the edge-wise utilization, as in Figure 13, to bring the congestion down steadily with increasing fleet size.

Finally, it is discussed, how the line concepts depicted in Figure 14 uncover the primary differences between the settings, namely, the demand profiles and the terminal layouts. First, in the settings where the terminals are not evenly distributed, the lines operating from Station 1 to stations 7 and 8 are allocated a high number of trains leading to frequencies more than 10 on those lines. As discussed earlier, these lines are crucial for the operation due to the lack of terminals between stations 1 and 7. Second, the bicentric demand profile employs the shorter lines on the right side of the PTN the most, compared to other configurations, to account for the latter demand peak of the bicentric demand profile. Similarly, with the unicentric demand, the lines operating merely in the middle of PTN are given great importance. Lastly, given the symmetrical demand and evenly distributed terminals, it can be observed that the line concepts illustrated in Figure 14b represent the most regular structure. This analysis concludes this section, and the final part of Section 4 follows. In the upcoming section, all three models are solved and the solutions' intercorrelations are studied.

4.6 Intercorrelations

In the concluding experimentation, the line planning models developed in this thesis are evaluated against each other. Special emphasis is placed on the examination of how correlating or conflicting the different objectives are. This investigation is conducted by first solving the line planning problem with one of the models and then employing the obtained line concept to evaluate the remaining two objective functions. Repeating this process for all models yields results that are here referred to as intercorrelations.

In this section, the intercorrelations are calculated for all three models in the passengers' perspective. The 20-station PTN is again considered with the unicentric demand profile and the original terminals, indexed $\{1, 7, 8, 13, 15, 16, 19, 20\}$. Furthermore, the results are calculated for fleet sizes 28, 33, 38, and 43. The intercorrelations are collected in Figure 15, where the colors indicate the model that is optimized and whose line concept is used to evaluate the different objective functions. The objective function values presented in the figure are normalized such that the optimized model gets the value of 1 and the other two objectives, evaluated with the same line concept, get scaled values. With the initial waiting time model, the scaled values are above 1 since the model minimizes initial waiting time. Analogously, with the other two models,



(a) Fleet size 28

(b) Fleet size 33

(c) Fleet size 38

(d) Fleet size 43

Figure 15: Intercorrelations of the models with various fleet sizes. All three models solved in the passenger's perspective on the 20-station PTN with unicentric demand profile and the original terminals $\{1, 7, 8, 13, 15, 16, 19, 20\}$.

the evaluated metrics are less than 1 given the maximization of direct passengers and minimum edge availability factor λ.

In general, Figure 15 suggests that there is some consistency in the dependencies of the objective functions. For instance, when congestion is evaluated the line concept acquired with the direct passengers model provides better values than the line concept given by the initial waiting time model, regardless of the fleet size. This finding suggests that congestion and direct passengers models are more intercorrelated than the congestion and initial waiting time models. Likewise, the other metrics share similar characteristics, but there are exceptions to the dependencies. Particularly, for fleet size 38, the congestion line concept provides shorter waiting times compared to that of the direct passengers model, while in general, it is the other way around. Similarly, the trend for direct passengers metric is disturbed with fleet size 43, in Figure 15d, where initial waiting time bypasses congestion by becoming the better line concept in the direct passengers' perspective. In conclusion, the initial waiting time line concept seems to be conflicting more with the other two metrics compared to how the direct passengers and congestion models intercorrelate with each other,

Table 2: Maximum edge utilization ratio $1/\lambda$ evaluated on solutions of all three models. The models solved in the passenger's perspective on the 20-station PTN with unicentric demand profile and the original terminals $\{1, 7, 8, 13, 15, 16, 19, 20\}$.

Fleet size **Optimized model**	28	33	38	43
Congestion	0.979	0.830	0.721	0.637
Direct passengers	1.000	0.925	0.810	0.753
Initial waiting time	1.000	1.000	0.902	0.816

despite the few exceptions.

Figure 15 also reveals other noteworthy properties of the models. In Section 4.4 it was found that it is always possible to transport almost all passengers directly to their destinations, regardless of the terminals and the fleet size as long as the number of direct passengers is maximized. However, the results acquired in this section suggest that if some other passenger convenience metric is considered in the objective function, the resulting line concept is far from being capable of transporting all passengers directly. In worst cases, only approximately 40% of passengers can travel directly, as denoted by the line concept of the initial waiting time model in Figure 15c. This amount of direct passengers is considerably small, given the linear PTN. Observing the direct passengers metric still, Figure 15 suggests that optimizing for one metric might yield arbitrary values for the other metrics. Particularly, comparing the number of direct passengers as given by the line concept of the initial waiting time model, the values are relatively high for fleet sizes 28 and 43 but quite low for 33 and 38. This finding further emphasizes the conflicting nature of objectives between the initial waiting time and direct passengers models.

As mentioned already, the congestion metric provides consistent results in the sense that, after the congestion model itself, the line concept by the direct passengers model yields consistently better minimum edge availability factor values compared to the initial waiting time model. Also, the congestion metric has the least variation between the models and the fleet sizes. To study the congestion values of the solutions more, the maximum edge utilization percentages $1/\lambda$, evaluated on all three models, are collected in Table 2. The table reveals that with a fleet size of 28, the maximum edge utilization percentages are either equal or close to 1. Because 28 is the smallest feasible fleet size, as constrained by the capacity constraints, the maximum edge utilization percentage is just barely below 1 when optimizing for congestion. Due to the capacity constraints, all line concepts must utilize almost all capacity on the busiest edges leading to similar congestion metrics regardless of the line concept. With larger fleets, there is more room for maneuver and thus the congestion model can bring the maximum edge utilization percentage down. The other two models can also achieve lower congestion values when more trains are available, but the effect is more subtle compared to the congestion model. For instance, with a fleet size of 33, the initial waiting time model still utilizes all capacity at least on one edge, as depicted in Table 2.

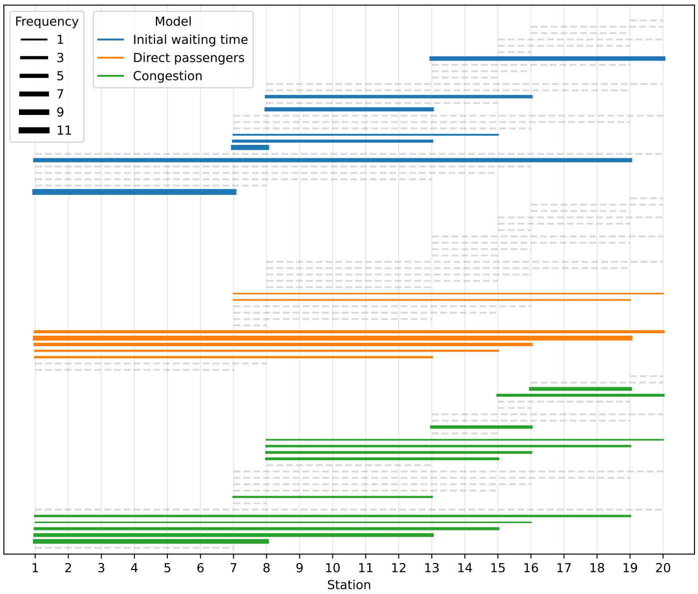


Figure 16: Line concepts $(\mathcal{L}, f)$ obtained with the three models. Models solved in the passengers' perspective on the 20-station PTN with unicentric demand profile and the original terminals $\{1, 7, 8, 13, 15, 16, 19, 20\}$. Fleet size is constrained to 28 trains.

To gain a deeper comprehension of the findings related to the intercorrelations, the solutions of the models are visualized on the line level, similarly as in the previous section. Figure 16 presents the line concepts obtained with the three models using a fleet of 28 trains on the 20-station PTN with the unicentric demand profile. The figure uncovers the quite significant differences in the line compositions between the three models. Starting with the initial waiting time model, it can be observed that the model yields a relatively small number of unique lines but with higher frequencies. Given the objective function, the model tries to allocate more trains and thus higher frequencies to edges adjacent to stations with more departing passengers. With the unicentric demand profile, this objective function results in higher effective frequencies towards the center of the PTN, as evident in the figure. The line concept also suggests that the mere minimization of initial waiting time can lead to suboptimal results concerning the other passenger convenience metrics. For instance, the line concept includes the edge from Station 7 to Station 8 and does so by allocating a rather high frequency to it. This strategy yields an optimal solution to the initial waiting time model but results in many passengers needing to transfer during their journey, as the frequency of the line operating just Edge 7 makes up approximately half of the service on Edge 7. Furthermore, many lines are halting at Station 7 and many lines starting at Station 8, which further reduces the number of direct passengers. Figure 15a confirms that only some 70% of the passengers can travel directly in the line concept given by the initial waiting time model.

Also, the line concept given by the direct passengers model is quite representative of the model's objective function, as illustrated in Figure 16. Rather intuitively, the line concept consists of only the longer lines, where the longest lines are given the highest frequencies. The solution is also the only line concept among the three models, which includes the line operating through the whole PTN to allow direct service for passengers between stations 1 and 20. Given the longer lines of the direct passengers model, the solution also has the smallest number of unique lines. Conversely, the line concept acquired by the congestion model has the biggest number of unique lines. As discussed in Section 4.4, a larger set of unique lines is needed to allocate the capacity along the PTN and to obtain the optimal edge utilization shape, as present in Figure 13. In conclusion, Figure 16 shines light on the reasons behind the differences and intercorrelations of the passenger convenience metrics, discussed earlier in this section, by revealing the vast differences in the characteristics of line concepts.

5 Conclusions

In conclusion, this thesis developed three optimization models for solving the line planning problem, where each model optimizes for one of the following passenger convenience metrics: congestion in the trains, number of direct passengers, and initial waiting time on platforms. While the models contribute to the vast literature of line planning research by considering the less studied objectives of congestion and initial waiting time, the models are also developed specifically for linear PTNs, which is also a rather novel approach in the line planning literature. Customizing the models for linear PTNs simplifies their complexity by removing the need to consider multiple routing options. This makes it easier to model the aforementioned passenger convenience metrics compared to the general case. The multi-objective nature of the line planning problem, arising from the conflicting preferences of the passengers and operator, was managed by exploiting the ε-constraint method, where one perspective is considered in the constraints side. The models developed in this thesis were evaluated on generated datasets by studying how the passenger convenience metrics respond to changes in the following aspects: passenger demand profile, size of the train fleet, and the set of terminal stations. Finally, it was studied how solutions obtained with the different models differ from each other.

In general, the models developed in this thesis were successful in improving the passengers' convenience with various experimental settings. However, it was discovered that different models yield quite diverse line concepts such that an optimal solution concerning one convenience metric might be far from a decent solution for the other models. Particularly, the congestion model thrived to distribute the available capacity evenly according to an optimal edge utilization profile specific to the corresponding demand profile. The utilization profile was observed to be independent of the fleet size, but dependent on the line pool, dictated by the set of terminal stations. For the direct passengers model, the results revealed that in all scenarios considered in the experiments, almost all passengers could be transported directly. Finally, the experiments showed that the initial waiting time model is prone to selecting lines and frequencies that are often unsuitable for the other convenience metrics since the model only regards the passengers' origin station and not the destination. In addition, the solving time evaluations highlighted that some of the models require exponential running time as a function of both the PTN and the fleet size. Nevertheless, all models could be solved in a short time with datasets of realistic size.

The results acquired in this thesis suggest that the models are functional for solving line planning problems in the given perspectives. However, the results are based on generated datasets with only a few variations. Therefore, this thesis provides an opportunity for further evaluation of the models with real passenger data and rail networks. Given that the models are tailored for linear PTNs, the models could be evaluated locally for linear subgraphs of a bigger public transportation network. Moreover, the models developed in this thesis are rather elementary in the sense that they lack crucial constraints and variables needed for practical applications. For instance, the line concepts proposed by the models often contained 10 to 20 unique lines, which is unreasonable for realistic use cases. Adding limitations to the models, such as constraining the number of lines, would lead to more useful results. The augmentation of the models to better correspond to real-world demands is identified as one aspect of future research. Finally, the evaluations identified a major shortcoming, specifically the substantial differences in line concepts between the models. For instance, the initial waiting time model provided solutions where the majority of the passengers would need to transfer, despite the setting of a linear PTN. On the other hand, the direct passengers model easily found solutions that allowed for direct travel such that, without a secondary objective, the line concept could become quite arbitrary. Combining the passenger convenience metrics into a single multi-objective model would thus provide an interesting topic for additional research. One option would be to further employ the ε-constraint method by considering one of the metrics in the objective function and constraining the other two passenger convenience metrics, along with the fleet size. The ε-constraints could be fine-tuned by iteratively solving the models with various values and by possibly consulting the operator on the acceptable levels of different passenger convenience metrics. Nevertheless, simultaneous consideration of the metrics would provide a line concept that takes multiple aspects of passenger convenience into account, also eliminating the possible ambiguity in the case of metrics that are easily optimized, such as the direct passengers percentage.

References

Barrena, E., Canca, D., Coelho, L. C., and Laporte, G. (2014). Exact formulations and algorithm for the train timetabling problem with dynamic demand. *Computers & Operations Research*, 44:66–74.

Blanco, V., Conde, E., Hinojosa, Y., and Puerto, J. (2020). An optimization model for line planning and timetabling in automated urban metro subway networks. a case study. *Omega*, 92:102165.

Blanco, V., Puerto, J., and Ramos, A. B. (2011). Expanding the spanish high-speed railway network. *Omega*, 39(2):138–150.

Borndörfer, R., Grötschel, M., and Pfetsch, M. E. (2007). A column-generation approach to line planning in public transport. *Transportation Science*, 41(1):123–132.

Bunte, S. and Kliewer, N. (2009). An overview on vehicle scheduling models. *Public Transport*, 1(4):299–317.

Burggraeve, S., Bull, S. H., Vansteenwegen, P., and Lusby, R. M. (2017). Integrating robust timetabling in line plan optimization for railway systems. *Transportation Research Part C: Emerging Technologies*, 77:134–160.

Bussieck, M. (1998). *Optimal lines in public rail transport*. PhD thesis, Univ.-Bibl.

Bussieck, M. R., Kreuzer, P., and Zimmermann, U. T. (1997a). Optimal lines for railway systems. *European Journal of Operational Research*, 96(1):54–63.

Bussieck, M. R., Lindner, T., and Lübbecke, M. E. (2004). A fast algorithm for near cost optimal line plans. *Mathematical Methods of Operations Research*, 59:205–220.

Bussieck, M. R., Winter, T., and Zimmermann, U. T. (1997b). Discrete optimization in public rail transport. *Mathematical programming*, 79:415–444.

Cancela, H., Mauttone, A., and Urquhart, M. E. (2015). Mathematical programming formulations for transit network design. *Transportation Research Part B: Methodological*, 77:17–37.

Ceder, A. and Wilson, N. H. (1986). Bus network design. *Transportation Research Part B: Methodological*, 20(4):331–344.

Claessens, M., van Dijk, N. M., and Zwaneveld, P. J. (1998). Cost optimal allocation of rail passenger lines. *European Journal of Operational Research*, 110(3):474–489.

Desaulniers, G. and Hickman, M. D. (2007). Public transit. *Handbooks in operations research and management science*, 14:69–127.

Drezner, Z. and Wesolowsky, G. O. (2003). Network design: selection and design of links and facility location. *Transportation Research Part A: Policy and Practice*, 37(3):241–256.

Ehrgott, M. (1999). Multicriteria optimization. *Lecture Notes*.

Gattermann, P., Harbering, J., and Schöbel, A. (2017). Line pool generation. *Public Transport*, 9:7–32.

Goerigk, M. and Schmidt, M. (2017). Line planning with user-optimal route choice. *European Journal of Operational Research*, 259(2):424–436.

Goossens, J.-W., van Hoesel, S., and Kroon, L. (2006). On solving multi-type railway line planning problems. *European Journal of Operational Research*, 168(2):403–424.

Guihaire, V. and Hao, J.-K. (2008). Transit network design and scheduling: A global review. *Transportation Research Part A: Policy and Practice*, 42(10):1251–1273.

Gurobi Optimization, LLC (2023). Gurobi Optimizer Reference Manual.

Kaspi, M. and Raviv, T. (2013). Service-oriented line planning and timetabling for passenger trains. *Transportation Science*, 47(3):295–311.

Kroon, L., Huisman, D., Abbink, E., Fioole, P.-J., Fischetti, M., Maróti, G., Schrijver, A., Steenbeek, A., and Ybema, R. (2009). The new dutch timetable: The or revolution. *Interfaces*, 39(1):6–17.

Li, Z., Mao, B., Bai, Y., and Chen, Y. (2019). Integrated optimization of train stop planning and scheduling on metro lines with express/local mode. *IEEE Access*, 7:88534–88546.

Liebchen, C. and Möhring, R. H. (2007). The modeling power of the periodic event scheduling problem: railway timetables—and beyond. In *Algorithmic Methods for Railway Optimization: International Dagstuhl Workshop, Dagstuhl Castle, Germany, June 20-25, 2004, 4th International Workshop, ATMOS 2004, Bergen, Norway, September 16-17, 2004, Revised Selected Papers*, pages 3–40. Springer.

Michaelis, M. and Schöbel, A. (2009). Integrating line planning, timetabling, and vehicle scheduling: a customer-oriented heuristic. *Public Transport*, 1:211–232.

Schiewe, P., Schöbel, A., Jäger, S., Albert, S., Baumgart, U., Biedinger, C., Grafe, V., Roth, S., Schiewe, A., Spühler, F., Stinzendörfer, M., and Urban, R. (2023). *LinTim: An integrated environment for mathematical public transport optimization. Documentation for version 2023.12*.

Schöbel, A. (2012). Line planning in public transportation: models and methods. *OR spectrum*, 34(3):491–510.

Schöbel, A. (2017a). An eigenmodel for iterative line planning, timetabling and vehicle scheduling in public transportation. *Transportation Research Part C: Emerging Technologies*, 74:348–365.

Schöbel, A. (2017b). Optimization models in public transportation. *Lecture Notes*.

Schöbel, A. and Scholl, S. (2006). Line planning with minimal traveling time. In *5th Workshop on Algorithmic Methods and Models for Optimization of Railways (ATMOS'05)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik.

Wu, J., Shan, X., Sun, J., Weng, S., and Zhao, S. (2023). Daily line planning optimization for high-speed railway lines. *Sustainability*, 15(4):3263.